Methods for simulation, planning, and operation of Aquifer Thermal Energy Storage under deep uncertainty

Jaxa-Rozen, Marc

DOI
10.4233/uuid:2b9d9501-3c1a-44e2-a14e-86578f62c5b4

Publication date
2019

Document Version
Final published version

Citation (APA)
Jaxa-Rozen, M. (2019). *Methods for simulation, planning, and operation of Aquifer Thermal Energy Storage under deep uncertainty*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:2b9d9501-3c1a-44e2-a14e-86578f62c5b4

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Methods for simulation, planning, and operation of Aquifer Thermal Energy Storage under deep uncertainty

Methods for simulation, planning, and operation of Aquifer Thermal Energy Storage under deep uncertainty

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Tuesday 15 January 2019 at 15:00 o'clock

by

Marc JAXA-ROZEN

Master of Science in Engineering and Policy Analysis,
Delft University of Technology, The Netherlands
born in La Pocatière, Canada

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr.ir. P.M. Herder	Delft University of Technology, promotor
Dr.ir. J.H. Kwakkel	Delft University of Technology, copromotor

Independent members:

Prof. dr. P. Blum	Karlsruhe Institute of Technology, Germany
Prof. dr. M. Gibescu	Utrecht University
Prof. dr. ir. T.N. Olsthoorn	Delft University of Technology
Prof. dr. P.M. Reed	Cornell University, United States
Dr. F. Pianosi	University of Bristol, United Kingdom
Prof. dr. K. Blok	Delft University of Technology, reserve member

The research presented in this dissertation was supported by the Netherlands Organisation for Scientific Research (NWO) under the project Aquifer Thermal Energy Storage Smart Grids (grant number 408-13-030), as part of the program Uncertainty Reduction in Smart Energy Systems (URSES). Additional support was provided by EIT Climate-KIC.

Cover art: Amovitania | Dreamstime
Printed by Gildeprint.

ISBN 978-94-6366-124-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Acknowledgements

This thesis concludes my work as part of the Aquifer Thermal Energy Storage Smart Grids (ATES-SG) project at TU Delft. Before presenting the results of this work, I would first like to thank the people who have turned these four years of PhD research – and my six years in the Netherlands – into one of the most rewarding and stimulating periods of my life.

I could not have wished for a better supervision team than my copromotor Jan Kwakkel and promotor Paulien Herder, who were inspiring by their academic creativity and rigor, but also in their commitment to balancing life and work. Jan left me the freedom to pursue random detours and avenues in my research, steered me back in the right direction when I needed some prodding to turn my work into actual papers, and always encouraged me with clear feedback. Paulien gave me invaluable comments for the direction and structure of my thesis, tolerated my highly adaptive writing schedule, and left me reassured after every single one of our meetings thanks to her cheerful pragmatism.

I will very much miss the stimulating, diverse, and supportive environment of the Policy Analysis section, and all my colleagues there who inspired me and made PhD work more enjoyable. I would especially like to thank my officemates Abby and Sharlene for having been so great to work with, the aquarium crew for fueling my writing through the last year of this project with humor and late-night snacks, and Monique and Marlies (as well as Laura at ESS) for helping me navigate cross-departmental red tape. I am also especially grateful to my MSc supervisor Erik Pruyt, who encouraged me to continue towards a PhD – directly by telling me about this opportunity, and indirectly by his enthusiastic teaching and supervision during my time in EPA. Still, my MSc thesis made me realize how much I missed the teamwork I often took for granted in earlier work, and one of my reasons for applying for the ATES Smart Grids project was the way in which it brought together multiple researchers across faculties. This gave me the opportunity to work alongside my fellow PhD candidates Martin Bloemendal (CiTG) and Vahab Rostampour (DCSC), and as we learned to bridge our mutual cultures both professional and personal, I was glad to see our work turn into a fruitful collaboration and friendship. Martin's insights and feedback (and generosity in letting me crash his conference hotel rooms), combined with Vahab's creativity in the work we present

in Chapter 7, were invaluable towards successfully finishing my part of the project. I would like to thank Tamas Keviczky for his management of our project over these four years, and the members of our users' committee for their constructive comments over the course of our progress meetings. The members of BodemenergieNL also helped me ground this research through their feedback at workshops, and by participating in our scoping survey at the onset of this project. And looking back even earlier, I would like to acknowledge the encouragement I received from Mathias Glaus and Robert Hausler during my time at École de technologie supérieure, as their courses and research sparked my interest in sustainability and eventually led me to pursue graduate studies.

I am grateful to the external members of the doctoral committee for taking the time to evaluate this thesis, and for their clear and constructive feedback. I would especially like to thank Prof. Patrick Reed for hosting me for a summer visit at his group at Cornell, during which I was able to refine the work presented in Chapter 4, as well as all the members of his research team for welcoming me in Ithaca and sharing their knowledge. Similarly, I am grateful to Dr. Robert Lempert for letting me visit Pardee RAND Graduate School earlier in my project and for sharing his insights, as well as to Ms. Laurie Rennie for facilitating my stay.

Finally, I would like to thank my friends for helping me feel at home in the Netherlands over these last six years, taking my mind off my work when I needed it, and for all the adventures at (Delfts)'bleau and elsewhere; and my family for their support and understanding, even when no one (beginning with myself) quite knew where I was going with this. I would like to dedicate this thesis to my parents for their unconditional support – I'm glad that my European detour eventually brought us closer together, *i mam nadzieję że tu znajdziecie trochę wyjaśnienia.*

Contents

1

Introduction

The built environment is one of the most important components in a transition towards sustainable modes of energy consumption. At the global scale, buildings currently account for approximately one-third of final energy use, and one-fifth of energy-related greenhouse gas (GHG) emissions (Lucon et al., 2014). This contribution may be compounded by economic and demographic trends – such as urbanization and shifts away from informal housing in developing countries – which point towards a major increase in construction over the upcoming decades. Under current practices, this may double or triple global energy use from buildings by the middle of the century.

However, these trends also open a window of opportunity to apply best practices for energy-efficient building design and operation, which could significantly contribute towards objectives for GHG reductions. A wide range of cost-effective technologies are currently available, but their adoption remains hindered by market barriers – such as a lack of appropriate legislation or financing instruments – which could lead to suboptimal practices becoming "locked in" over the lifetime of new buildings. Taking advantage of this window of opportunity will require a suitable combination of building technologies and policy interventions.

As the largest single end use of energy in the built environment, space heating is a crucial component of building energy efficiency. Geothermal energy technologies are increasingly popular for this application. The subsurface can for instance be used to store energy for the space heating and cooling of buildings, using "shallow" systems for seasonal Underground Thermal Energy Storage (UTES; reviewed in Lanahan and Tabares-Velasco, 2017). While cooling is currently less important on a global scale than heating, its use of energy may increase significantly due to growing demand in developing countries as well as climate change (Lucon et al., 2014). UTES technologies which can address both heating and cooling demand are therefore particularly promising.

Aquifer Thermal Energy Storage (ATES) is a rapidly developing form of UTES, in which thermal energy is seasonally stored in natural aquifer formations. In combination with a heat pump, ATES can reduce energy demand for heating and cooling by more than half in larger buildings (Tomasetta et al., 2015), and consequently reduce operating costs for building owners. Furthermore, when combined with supply-side trends towards renewable power generation, ATES can yield additional GHG benefits by supporting the electrification of building energy systems. This technology is thus increasingly popular in Northern Europe. Furthermore, the climate and subsurface conditions required for ATES can be found across Europe, Asia and North America (Bloemendal et al., 2015).

With approximately 2500 active ATES systems in 2017, the Netherlands are currently a world leader for ATES technology, due to a combination of easily accessible aquifer resources, high urbanization, and increasing demand for energy-efficient technologies. As such, the use of energy from ATES in the Netherlands is expected to further grow threefold by 2023 (Agterberg, 2016). However, this development has already highlighted a number of issues with the operation and management of urban ATES systems. Due to their relative complexity, ATES systems are challenging to operate effectively; for instance, uncertain aquifer characteristics and changes in building energy demand contribute to unpredictable operational performance. Building owners thus often save less energy than expected using ATES.

In response to these uncertainties, and considering the sensitivity of aquifer resources, the management of ATES technology has typically been relatively conservative. In the Netherlands, a revised policy framework (*Wijzigingsbesluit bodemenergiesystemen*, or WBBE) was implemented in 2013, reflecting new research on the environmental risks of ATES technology and practical experience with the development of ATES systems. Early assessments of this framework have been mixed in regards to its impact on ATES development (de Graaf et al., 2016; Agterberg, 2016), and point to several points of concern for researchers and practitioners.

In this context, the following section will present current challenges for the development of ATES technology, from the perspective of policymakers and system operators. This will lead towards the research problem treated in this thesis, which is described in section 3. Section 4 summarizes the approach followed to address this problem, while section 5 outlines the structure of the thesis.

1.1. Background

Shallow geothermal systems are currently the fastest-growing application of geothermal energy (Bayer et al., 2012). These systems include technologies for Underground Thermal Energy Storage (UTES), which rely on the subsurface to seasonally extract or store thermal energy at depths of less than 500m. Aquifer Thermal Energy Storage (ATES) is an increasingly common variant of this approach; ATES systems can be

used to directly pump groundwater for seasonal energy storage. As such, these systems typically involve at least one pair of coupled wells, which simultaneously infiltrate and extract groundwater from different locations to avoid net changes in the groundwater stored in the aquifer.

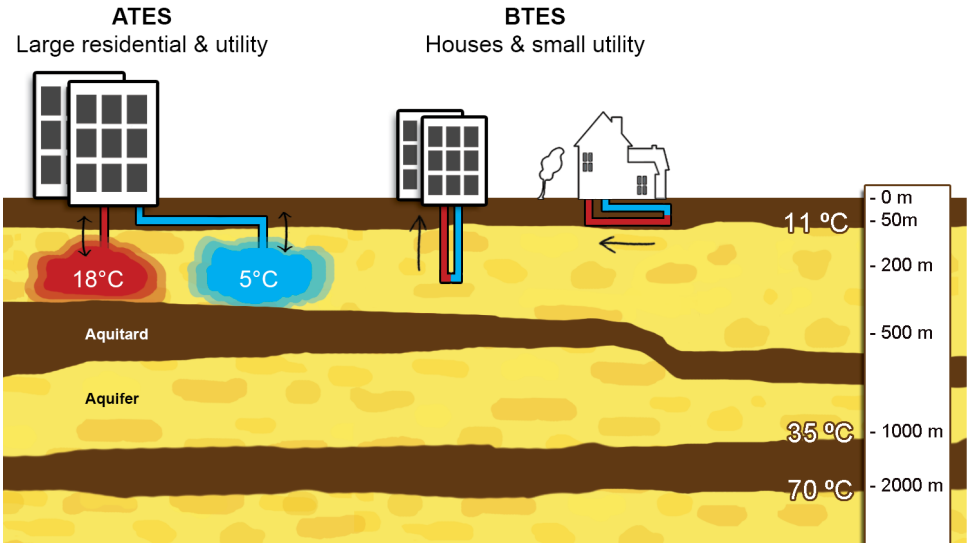


Figure 1.1: Working principle of Aquifer Thermal Energy Storage (ATES) and Borehole Energy Storage (BTES) (Bloemendal, 2018)

In winter conditions, groundwater is extracted from a "warm" well, then circulated through a heat exchanger to provide heating in combination with a heat pump. This process reduces the temperature of the extracted water, which is then re-injected into the opposite "cold" well at a temperature of 5-10°C. Under summer conditions, this process is reversed: the cooler water which was injected during the winter is extracted, used for cooling, and re-injected at a temperature of 15-25°C. These injection temperatures represent typical practices for "low temperature" storage systems; higher storage temperatures may cause adverse environmental impacts on the subsurface, so that maximum injection temperatures are usually set by local legislation. This topic is an active area of research, as a sustainable application of higher storage temperatures could significantly improve energy performance. Over time, these pumping patterns lead to the formation of warm and cold zones in the groundwater around each well, which should ideally represent equivalent amounts of thermal energy to maintain the thermal balance of the subsurface. This implies that ATES is largely limited to temperate areas in which buildings require both heating and cooling over a typical year.

Although ATES is increasingly popular in Northern Europe, the most common UTES technology is currently Borehole Thermal Energy Storage (BTES). This tech-

nology typically relies on a series of U-shaped, vertically oriented pipes, which carry a thermal working fluid and transfer heat to the surrounding soil medium. As with ATES, a heat exchanger transfers heat between the building and the storage system. Compared to ATES, BTES can be scaled down more economically for smaller residential or utility buildings, but drilling costs can make it less attractive for larger buildings. The thermal performance of BTES systems is also typically lower than for ATES. However, BTES does not require an aquifer, which makes it more geographically versatile. Its thermal interactions with the subsurface are also more localized, so that thermal balance between heating and cooling is less critical.

1.1.1. Issues for the planning and management of urban ATES systems

The development of thermal storage zones in the subsurface is a crucial factor for the performance and management of ATES systems: thermal interferences caused by an insufficient distance between warm and cold wells will lead to energy losses, while neighbouring wells of the same type may have a positive mutual thermal influence (Bakr et al., 2013). However, the monitoring of these thermal zones is technically challenging, and their evolution is tightly linked to local geohydrological conditions – which are themselves difficult to assess. These characteristics yield significant uncertainties in regards to thermal subsurface dynamics, and therefore to the resulting performance of ATES systems.

These uncertainties are compounded by variable weather conditions and building occupancy patterns. Under existing practices for the control and operation of ATES systems, these factors make it difficult to predict the pumping rates required to maintain both building comfort and thermal balance in the subsurface. Actual pumping rates are therefore likely to differ from the expected values which are used to design and plan the systems. However, excessive imbalances between heating and cooling may compromise the long-term viability of storage systems through unexpected thermal interactions, or by changing ambient groundwater temperatures; the low levels of thermal dissipation in the subsurface – which make aquifers attractive for thermal storage – also mean that imbalances may linger for decades. Calje (2010) for instance found that the current performance of ATES systems near Amsterdam's Dam square is likely to be reduced by a warm "bubble" in the subsurface, which was created in the 1990s by a building using groundwater for cooling. In dense urban environments, ambient aquifer temperatures may also be increased by anthropogenic (e.g. "heat island") surface effects (Zhu et al., 2010), which could for instance in turn affect the performance of systems used for cooling.

Thermal balance and interferences are therefore key elements to consider for the sustainable development of shallow geothermal energy (Haehnlein et al., 2010). The risk of negative interferences particularly introduces additional constraints on the spatial planning of ATES systems in urban areas, which also needs to account for existing

subsurface functions such as groundwater extraction, or other infrastructures such as utility piping or wiring. These competing functions may therefore yield a scarcity of available space for ATES systems in urban areas.

Other environmental issues to consider include chemical, hydrological, and microbiological impacts, which were reviewed by Bonte et al. (2011) and Hähnlein et al. (2013). Chemical impacts may be driven by the mixing of different types of groundwater and by local temperature changes; for instance, vertical groundwater mixing can have adverse impacts on water quality, while several processes (such as changes in mineral solubility or reaction kinetics) may affect the long-term performance of ATES systems by promoting corrosion or clogging of the well screens. Chemical impacts may also relate to contaminants present in the aquifer, e.g. by affecting their biodegradation. In parallel, hydrological impacts related to the local extraction and infiltration of water may lead to soil subsidence due to a change in groundwater levels, which for example affected early groundwater cooling systems used in China in the 1960s (Morofsky, 2007); ATES flow patterns can also change the capture zones of nearby groundwater supply wells. Hydrological risks can also be created by design or construction flaws which lead to undesirable flow paths. For instance, in several cases reported in Germany (Fleuchaus and Blum, 2017), poorly backfilled boreholes used for heating eventually caused significant swelling and surface uplifts due to unexpected contact between groundwater and anhydrite layers, damaging nearby buildings. Finally, temperature changes may have a significant impact on the composition and diversity of the bacterial ecosystems present in groundwater (Brielmann et al., 2009); while these impacts appear to be limited at infiltration temperatures which are typical of low-temperature storage, further research on microbiological impacts would be needed in the context of high-temperature systems.

In response to these issues, design and management methods for ATES technology have typically followed the precautionary principle (Haehnlein et al., 2010). Morofsky (2007) provided a set of guidelines for the design and operation of UTES systems, in light of common environmental concerns; Hähnlein et al. (2013) proposed a broader framework for the management of shallow geothermal energy, which would acknowledge the technical, environmental, and social aspects of sustainability. This framework would include a technical assessment of geological conditions and interference effects, followed by a case-by-case review of environmental impacts – taking into account mineral and microbiological impacts on groundwater, as well as potential interactions with anthropogenic aquifer contamination. Similarly, Vienken et al. (2014) proposed an adaptive development workflow for shallow geothermal energy, which would emphasize improved modelling and monitoring methods to increase system performance and mitigate environmental impacts.

1.1.2. The current situation for ATES in the Netherlands

In the Netherlands, revised policies for shallow geothermal energy were implemented in July 2013 under the WBBE framework, with the dual objectives of promoting the development of the technology while minimizing its impacts on the environment. These new policies notably require permits from provincial authorities for the construction of new ATES systems, which are granted for a given pumping volume on a "first come, first served" basis. Furthermore, design guidelines specify minimal relative distances between neighbouring wells to avoid thermal interactions between systems. These policies also include a geographic distinction between different classes of ATES development areas, giving provinces and municipalities the authority to designate specific development regions in dense areas. The planning of new systems in these regions can instead be addressed through collective master plans, which take into account the layout of neighbouring systems to manage thermal interactions. In addition, ATES development is largely excluded in areas with existing subsurface functions – such as the extraction of drinking water – while limiting injection temperatures to a range of 5-25°C to minimize chemical and microbiological impacts.

In this context, recent studies have taken different perspectives to assess the changes introduced by the WBBE framework. de Graaf et al. (2016) thus interviewed 46 stakeholders from government and industry about their perception of the WBBE. The revised policies were perceived to provide an adequate level of protection for the subsurface, with a fairly comprehensive management of environmental risks from ATES and BTES. In parallel, there were no clear indications that the WBBE had significantly influenced the adoption rate of thermal storage systems, with other factors likely being more influential – such as legal requirements for energy performance in buildings, and increasing competition from air-source heat pump systems. However, several stakeholders perceived new regulations (notably the permitting process) to be overly restrictive for the development of smaller ATES systems, due to technical requirements as well as administrative overhead. Notably, the objective of avoiding any thermal interferences between ATES systems was perceived as being at odds with an optimal use of the subsurface for thermal storage, particularly in high-potential areas. In the short term, high upfront costs and complex operation were perceived as key barriers for the further development of ATES.

These results were consistent with a survey of 53 current users of ATES systems (Bloemendal and Jaxa-Rozen, 2016), which was performed in early 2016 with the support of the BodemenergieNL platform for geothermal energy in the Netherlands. This survey focused on evaluating perceived barriers for ATES use in the Netherlands, as well as the perceived performance of ATES compared to conventional building systems across several dimensions. The results are summarized in Figure 1.2.

Over half of respondents thus perceived regulations to be a "high" or "very high" barrier to adoption. While significant, this fraction was overshadowed by the nearly

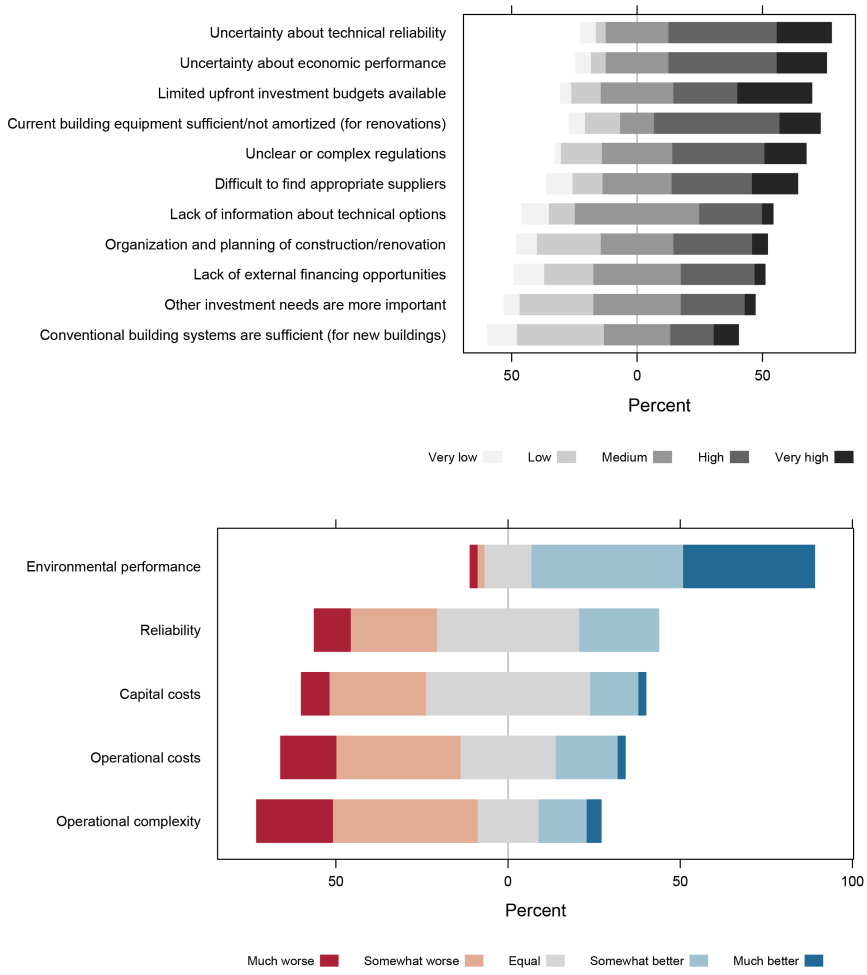


Figure 1.2: Top panel: Perceived barriers for the adoption of ATES in the Netherlands. Bottom panel: Perception of ATES relative to conventional building energy systems.

two-thirds of respondents who perceived technical reliability to be a "high" or "very high" barrier to ATES adoption, with about the same fraction describing ATES as "worse" or "much worse" than conventional building systems in terms of operational complexity. Agterberg (2016) similarly cited the difficult management of ATES as a major issue for further development in the short term; this has in fact led to new business opportunities for engineering contractors who offer specialized diagnostic services, or even purchase and redevelop ATES systems on behalf of building owners.

These factors are likely reflected in Willemsen (2016)'s survey of operational data

for 125 ATES systems. As such, these systems were found to only have used 41% of their average permitted pumping capacity over the 2010-2014 period, with changes in weather and building operation leading to annual thermal imbalances of around ±25%. With delivered energy being largely proportional to the pumped volume, the energy provided by ATES would therefore be significantly less than expected by operators and policymakers. This under-utilization can be linked to operational issues, but also to the current approach to permits: without a feedback mechanism between ATES operation and licensing, ATES operators have an incentive to overestimate the expected pumping volume in permit applications, so that they have more flexibility to manage variations in energy demand.

In certain areas – such as the city center of Utrecht – these technical and planning issues already combine to create a scarcity of space, with a lack of subsurface volume available to match the demand for new systems. Given current forecasts for the development of ATES, under which energy use from ATES may increase threefold by 2023 (Agterberg, 2016), the scarcity of space available for new systems is thus likely to emerge as a critical barrier for the further development of ATES in the Netherlands. However, under current practices for planning and operation, much of this allocated volume is likely to remain unused – artificially limiting the contribution of ATES to energy savings.

In sum, the current situation yields suboptimal outcomes for system operators as well as policymakers: for the former, the WBBE framework applies restrictions on spatial planning – as well as administrative overhead – which may not always be warranted; for the latter, ATES contributes less than expected towards targets for energy efficiency. Further revisions to current planning methods will likely be needed to better align these private and public interests, and fulfil the technical potential of ATES in the longer term. This problem is particularly relevant given the increasing demand for ATES technology, as well as the expected development of improved control systems which could mitigate operational challenges for ATES (e.g. Rostampour and Keviczky, 2017).

The design of improved control systems and planning policies will also require refinements in the geohydrological simulation models which are used to support ATES development. Although numerical simulation models are commonly used both in research and in practice to evaluate interferences between systems (Bakr et al., 2013; Li, 2014; Sommer et al., 2015), these engineering models focus on subsurface conditions. As such, they typically take a simplified view of "above-ground" dynamics for the adoption of ATES systems, or short-term operational uncertainties. Given the importance of operational factors and thermal interactions for ATES planning and performance, modelling and simulation methods should be able to represent feedbacks between buildings and the subsurface to properly acknowledge the multiple dimensions of ATES development.

It can be noted that these issues are currently specific to the Netherlands, due to relatively higher ATES adoption rates and dense urbanization. Although the markets for ATES technology in most European countries remain immature, a combination of appropriate legislation and technical development should lead to increased demand over the upcoming decade (Climate-KIC - E-USE, 2014). Early experience with these issues in the Netherlands can therefore help develop best practices for the successful large-scale management of ATES technology.

1.2. Research objective and questions

This discussion leads to the main research objective to be addressed in this thesis:

To assess the role of improved methods for the planning and operation of urban ATES systems towards a better alignment of private and public interests.

This objective will be addressed by answering five research questions, detailed below:

1. *How can ATES operation and spatial planning options be represented more realistically within ATES simulation models?* Current methods for the modelling and simulation of ATES systems are unable to represent the full complexity of the problem. To support the evaluation of different ATES control designs and planning policies, answering this question will require the design of a coupled simulation architecture which integrates geohydrological, socio-technical, and control model components.

2. *How can this simulation approach be used to efficiently explore different options for ATES operation and planning under uncertainty?* Using the coupled simulation architecture to assess ATES control schemes and planning policies will require an understanding of model sensitivities. Due to technical issues such as model runtime, this question will in turn involve a refinement of existing tools for the sensitivity analysis of complex environmental models.

3. *How do the operational uncertainties of ATES systems affect the design of suitable spatial planning arrangements?* The coupled simulation architecture will be used to simulate plausible scenarios for the future development of urban ATES systems, accounting for operational uncertainties under current planning arrangements, and under revised guidelines. To compare different levels of scale and complexity, this question will be answered by considering an idealized case study, as well as a case study for the Utrecht city center.

4. *How can information exchange between ATES systems lead to a more efficient trade-off between individual system performance and collective interests?* Improved control systems

can potentially help reconcile private and public interests for the future development of ATES systems, by mitigating the impact of operational uncertainties through the cooperative operation of neighbouring systems. This question will aim to better understand the conditions under which cooperative operation would be beneficial, and the implications of cooperative operation in regards to planning options.

5. *To what extent can improved methods for ATES planning and operation contribute to the large-scale potential of ATES for energy savings in the built environment?* The issues with ATES planning and operation discussed in this thesis are largely limited to the Netherlands, as the country is currently the primary market for the technology. To better understand how the results of the thesis can be translated to a larger scale, the last question will be answered by taking a global view of ATES development to assess the long-term potential of ATES for urban energy efficiency, as well as the conditions under which the improved planning and operating methods discussed in this thesis would be relevant.

1.3. Scientific and societal relevance

This project is expected to deliver several contributions to the state of current research in relation to ATES systems, and to the coupled modelling of social-ecological systems. As described in the literature review, current planning policies for ATES systems are a barrier for the future development of the technology in urban areas, and will likely need to be revised to accompany the design of more effective control systems. From a methodological standpoint, existing simulation tools for the model-supported planning of ATES systems are unable to represent time-dependent, endogenous adoption patterns and operational control. The thesis will contribute towards both of these issues, by using a coupled simulation architecture to address the full complexity of the problem and generate practical policy insights, and by making this architecture easily accessible to other researchers and practitioners.

From a broader perspective, the results of this research will also be transferrable to other cases involving social-ecological systems or large-scale distributed control problems — such as groundwater management or geothermal heat recovery projects. Although agent-based modelling is increasingly used for the study of social-ecological systems, existing research typically considers environmental dynamics in a simplified manner, which compromises the extent to which the models can be used for policy analysis. Furthermore, coupled models of socio-technical or social-ecological systems typically do not systematically consider model sensitivities and uncertainties. This thesis introduces methods which can help modellers and analysts address these issues.

1.4. Research approach

The general approach of this research will draw upon three related perspectives: complex adaptive systems, socio-technical systems, and social-ecological systems. As defined by Holland (1992), the behaviour of complex adaptive systems emerges from interactions between the agents which compose the system, and which are themselves able to learn or adapt through these interactions. This emergent behaviour may be unpredictable due to non-linear causal relationships or feedbacks operating at multiple time scales. Complex adaptive systems may also present coherent behaviours as a result of cooperation or competition between agents. For instance, a system may self-organize towards a stable structure through the repeated interactions of its components, without imposed external constraints (Nikolic and Kasmire, 2013).

Socio-technical systems (STSs) are a class of complex adaptive systems, which combine physical networks of technical elements interwoven with networks of social entities (de Bruijn and Herder, 2009). These technical components interact through physical relationships, while their operation and management is driven by multi-actor relationships within the social layer. These systems may be found at different scales of complexity and geographic scope, such as production plants, regional power grids, or international supply chains. Similarly, social-ecological systems (SESs) are complex adaptive systems composed of interdependent but relatively separable social and biophysical subsystems (Ostrom, 2009), such as a natural resource system, a social system of resource users, and a governance system of institutional arrangements.

The modelling and simulation of STSs and SESs can help understand how internal feedbacks – along with the structural characteristics of each system – drive the overall behaviour of the coupled system (Schlüter et al., 2012). However, SES and STS models face a range of conceptual and technical challenges due to the complexity of the underlying systems, and their varied disciplinary backgrounds. SES models thus need to synthesize methods from social and natural sciences to properly represent the coupled dynamics of social and ecological subsystems, while STS models may need to integrate an engineering model of technical subsystems. As emphasized by Tavoni and Levin (2014), although this multidisciplinary perspective is challenging, it may ultimately be required to portray the complexity of environmental and economic systems, and increase the policy relevance of academic models.

In the context of this research, the development of urban ATES systems will be perceived as a complex adaptive system, which combines socio-technical and social-ecological elements: the performance of ATES systems is a function of environmental conditions in the subsurface, which are themselves affected by the operation of the systems. Furthermore, ATES operators may interact directly through governance mechanisms within the social subsystem, or indirectly through thermal interactions within the environmental subsystem. Understanding these dynamics under different policy options will therefore require a modelling and simulation approach which can account

for the socio-technical and social-ecological dimensions of the problem. This approach is described below in a stepwise fashion, to match each of the research questions: first by developing techniques for the coupled simulation and analysis of appropriate models, then by implementing and testing these models at increasing levels of scale and complexity.

1.4.1. Coupled agent-based/geohydrological modelling

Agent-based modelling will be the basic paradigm used to model and simulate the planning and operation of ATES systems. Agent-based models represent complex systems by decomposing them into individual, heterogeneous agents; the decision-making of these agents and their interactions then lead to the emergence of collective outcomes over time (Epstein, 2006). This approach is increasingly popular for the "bottom-up" simulation of systems such as decentralized energy technologies, in which individual heterogeneity and spatial properties play a crucial role.

This bottom-up perspective is an intuitive choice for modelling ATES adoption and operation at the level of individual buildings: an agent-based approach can for instance make it easier to represent realistic operation dynamics by simulating a dedicated building control model for each agent, and by simulating the exchange of information aross buildings in the context of a cooperative operating scheme. However, ATES performance depends on thermal subsurface dynamics – which then affect the use of ATES by building owners. The agent-based component will thus need to be coupled with a numerical model of the subsurface to include these feedbacks. While analytical methods can provide an efficient alternative to numerical modelling for simpler cases (Banks, 2011; Pophillat et al., 2018), in particular for "scoping" studies which aim to evaluate the presence of thermal interactions in dense environments, they may not adequately quantify the effect of these interactions on system performance, so that they would be less relevant for the purposes of this research.

The first step of the research will therefore entail the design of a suitable simulation architecture which can exchange information between the model components. This simulation architecture will be based on the Python language, using the pyNetLogo connector developed as part of this research as a key component to link Python with the NetLogo agent-based modelling software; the following chapter of this thesis will describe pyNetLogo in more detail. This choice of software is motivated by the increasing popularity of Python as an environment for the modelling and analysis of groundwater problems, for instance using the FloPy pre/post-processing library for the MODFLOW/SEAWAT geohydrological modelling software (Bakker et al., 2016). The Python architecture will therefore be interfaced with a MODFLOW/SEAWAT model of the subsurface using FloPy, as described in Chapter 3.

1.4.2. Sensitivity analysis for complex simulation models

Sensitivity analysis aims to identify the parameters or combinations of parameters which contribute significantly towards uncertainties in the output of a model (Saltelli and Annoni, 2010). This technique will be needed to quantify the sensitivities of the different model components, and better understand their role for overall system outcomes. Methods for global sensitivity analysis (GSA) are increasingly accepted as a standard for the analysis of complex environmental models, in order to reliably evaluate the impact of uncertain parameters; these methods can be used to cover the full domain of uncertain inputs and capture the impact of interactions between parameters. However, most existing GSA techniques are highly computationally expensive or otherwise restrictive, which makes them difficult to apply in the case of complex simulation models – such as the coupled models developed in this research. To mitigate these restrictions and facilitate sensitivity analysis for the simulation case studies presented later in the thesis, this step will refine existing statistical learning techniques, and demonstrate their use for the sensitivity analysis of complex simulation models in Chapter 4.

1.4.3. Exploratory modelling and analysis

As described by Bankes (1993), exploratory modelling and analysis (EMA) can help understand complex systems by using computational experiments to clarify the implications of multiple uncertainties. By representing a broad set of scenarios and hypotheses, EMA can identify counterintuitive outcomes or leverage points; in particular, techniques for scenario discovery (Bryant and Lempert, 2010) can be used from an EMA perspective by exploring the specific situations in which a policy performs inadequately, or the circumstances under which a system may present a specific behaviour. This approach can contribute to the design of policies which are more robust, by performing well over a broad range of uncertain futures (Lempert et al., 2006). This is especially relevant in the presence of deep uncertainties, i.e. uncertainties for which probability distributions or structural relationships are unknown (Lempert et al., 2003).

Given the various uncertainties which influence the different model components – such as energy prices or aquifer properties – the case studies will be developed and interpreted from the overarching perspective of EMA. The first case study will depict an idealized "proof-of-concept" case, in Chapter 5 of the thesis; this will be followed in Chapter 6 by a second case representing scenarios for ATES development in the city center of Utrecht, using data on existing ATES systems and subsurface conditions.

1.4.4. Mechanisms for coordinated ATES operation

The representation of planning options in the agent-based model will first draw upon current practices for permits and master plans. Alternative spatial planning policies will be then be simulated in the agent-based model, in parallel with more advanced ATES control schemes. These control schemes follow a model predictive control (MPC)

approach (Rostampour and Keviczky, 2016, 2017), in order to depict a more realistic integration of ATES within building energy systems in Chapter 7. For each case study, these controllers will be simulated across different scenarios for building energy demand and spatial planning, starting from individual building operation, then introducing coupling constraints which aim to dynamically manage thermal interactions by exchanging information across neighbouring ATES systems.

1.4.5. Integrated assessment of ATES energy potential

To broaden the scope of this research beyond the Netherlands, the last step – presented in Chapter 8 – will use a spatially-explicit analysis to evaluate the technology's worldwide potential towards urban energy savings. Given that the feasibility and performance of ATES depends on subsurface and climate conditions, this analysis will integrate existing data sources for aquifer, climate, and building properties, to assess plausible energy savings from ATES in 556 major urban areas. These savings will be estimated across a set of scenarios for climate conditions at the 2050 horizon, along with multiple scenarios for building energy performance and ATES adoption. The analysis will include BTES as a reference technology, to better understand the relative performance of ATES across subsurface and climate conditions.

1.5. Outline

The following three chapters present the main methodological contribution of this thesis. Chapter 2 describes the pyNetLogo interface, which links the NetLogo agent-based modelling software with the Python programming language. This interface is used in Chapter 3 to support the design of a coupled simulation architecture for agent-based/geohydrological modelling. Chapter 4 then describes an alternative approach for the sensitivity analysis of complex simulation models, which applies statistical learning techniques to assess the importance of uncertain model parameters, and facilitates the analysis of the coupled models used in this work.

The next chapters then apply these modelling and analysis techniques to address the main research problem of this thesis. Chapters 5 and 6 introduce model-based ATES case studies which explore the interplay between operational uncertainties and spatial planning, respectively for an idealized case and for the city center of Utrecht. These cases are revisited in Chapter 7 using a model predictive control approach for the coordinated operation of ATES systems (Rostampour and Keviczky, 2016, 2017). Chapter 8 takes a broader view of ATES development, evaluating the long-term worldwide potential of ATES for energy savings in urban areas.

The methological and applied findings are synthesized in Chapter 9, which presents policy recommendations for the future development of urban ATES systems, interprets the model-based case studies in the context of the worldwide assessment of ATES potential, and outlines directions for future research.

2

pyNetLogo: Linking NetLogo with Python

This chapter is based on Jaxa-Rozen and Kwakkel (2018a), with minor revisions to fit the structure of this thesis.

2.1. Introduction

Agent-based models (ABMs) are a well-established method for the study of complex adaptive systems, in which the interactions of heterogeneous entities yield emergent large-scale behaviors. As such, this approach has been applied across a wide variety of fields such as economics, ecology, or socio-technical systems (e.g. Judd, 2006; Grimm and Railsback, 2012; Nikolic and Kasmire, 2013).

However, the computational nature of ABMs can make them more difficult to understand and communicate than analytical models (Grimm et al., 2006). Without the use of standard frameworks to structure their analysis and documentation, ABMs may yield *ad hoc*, poorly reproducible results (Thiele, 2015). Different initiatives are attempting to address this gap, such as the ODD and TRACE protocols for documentation (Grimm et al., 2006; Schmolke et al., 2010).

In practice, these documentation protocols are easier to apply when supported by suitable computational tools – for instance to generate experimental designs for uncertain inputs, visualize output data, or apply standard statistical methods. While many agent-based modelling platforms include basic analysis tools, these are typically not sufficient to meet the requirements of a comprehensive analysis and documentation process. Conversely, using standalone analysis software to process input and output data files can quickly become unwieldy for complex models – making the analysis workflow more difficult to reproduce.

The literature therefore presents different connectors to directly interface agent-based modelling software with analysis environments. In particular, the popular open-source NetLogo modelling software can be linked at runtime with Mathematica (Bakshy and Wilensky, 2007) and R (Thiele et al., 2012b), which allows modellers to use the comprehensive analysis and visualization functionalities available in these programming languages.

As a complement to these connectors, this chapter introduces the pyNetLogo library, which can be used to control NetLogo through the Python programming language. Python is a general-purpose language which is consistently ranked as one of the five most popular languages on the TIOBE Programming Community index (TIOBE, 2017); it offers a variety of libraries which can support ABM development and testing, and is increasingly used for scientific computing. In the following chapters of this thesis, Python will thus be used to link NetLogo with a groundwater model, relying on pyNetLogo as a key building block to design a coupled simulation architecture. It should be emphasized that pyNetLogo is not intended as a replacement for the existing R and Mathematica connectors, or as a comment on the suitability of these various environments for ABM analysis. However, given the popularity of the Python language, pyNetLogo extends the benefits of a specialized analysis environment to a broader audience.

The following section of this chapter describes the different software platforms used in this work. A software implementation section then introduces pyNetLogo and its key features, and illustrates these mechanisms for a simple predator-prey model. As an example of the analysis workflow which is enabled by pyNetLogo, this model is controlled interactively from a Python environment, then tested using a global sensitivity analysis.

2.2. Software description

2.2.1. NetLogo

NetLogo (Wilensky, 1999) is an open-source environment for the design and testing of agent-based models. While NetLogo was initially intended as an educational tool, its ease of use, robust performance and active user community have made it a pragmatic choice for a wide range of research applications (Kravari and Bassiliades, 2015; Railsback et al., 2006). It has therefore established itself as a leading platform for agent-based modelling (Thiele, 2015).

NetLogo is primarily implemented in Java and Scala, and includes a range of functions and methods to support the rapid development of spatially-explicit agent-based models. Railsback et al. (2017) further discuss strategies and techniques to improve the performance of more complex NetLogo models. In addition to connectors for Mathematica and R, different extension modules are available, for instance to inter-

face NetLogo models with GIS datasets. In particular, an extension for Python (Head, 2017) offers a converse functionality to the pyNetLogo connector, by allowing Python code to be executed from a NetLogo model.

2.2.2. Python

Python is a widely used high-level, general-purpose open source programming language that supports various programming paradigms. Python places a strong emphasis on code readability and code expressiveness. A large collection of libraries for many typical programming tasks is readily available. Python is increasingly popular for scientific computing purposes due to the rapidly expanding scientific computing ecosystem available for Python.

This ecosystem includes NumPy (Walt et al., 2011) and pandas (McKinney, 2010) for data manipulation, SciPy (Jones et al., 2001) for general numerical tasks, Matplotlib (Hunter, 2007) for plotting and visualization, as well as Jupyter and IPython (Pérez and Granger, 2007) for interactive analysis. These libraries are pre-packaged in several scientific distributions for Python, such as Continuum Anaconda. Additional libraries can be installed through standard package managers such as pip and conda.

Python is often used as a "glue" language, meaning that it connects pieces of software written in different languages together into a bigger application. For instance, the JPype library (Menard and Nell, 2014) can be used to access Java class libraries through interfacing the Python interpreter and the Java Virtual Machine. PyNetLogo therefore relies on JPype for interacting with NetLogo.

2.3. Software implementation

This section first describes basic interactions between the Python environment and a NetLogo model, using the pyNetLogo connector. These interactions are demonstrated using the simple wolf-sheep predation example which is available in NetLogo's model library. This functionality is then extended to illustrate a typical model analysis workflow, using the SALib Python library (Herman and Usher, 2017) to perform a global sensitivity analysis.

The model files used for these examples are available from the pyNetLogo repository at `https://github.com/quaquel/pyNetLogo`, along with interactive Jupyter notebooks which replicate the analysis and visualizations presented in this chapter. Detailed documentation and installation notes for pyNetLogo are provided at `http://pynetlogo.readthedocs.io`.

The pyNetLogo connector can be installed using the pip package manager, using the following command from a terminal or command prompt: `pip install pynetlogo` The pyNetLogo connector has been tested with NetLogo 5.2, 5.3 and 6.0 using the 64-bit Continuum Anaconda 2.7 and 3.6 Python distributions. Using these distributions, pyNetLogo requires the additional installation of JPype (available through the conda

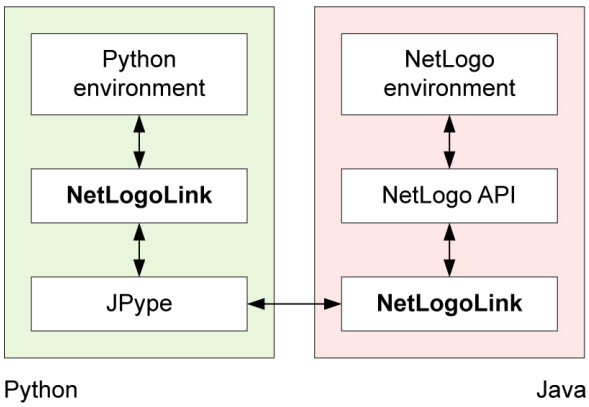


Figure 2.1: Interactions between Python and NetLogo

package manager). The pyNetLogo connector is currently also included in the Exploratory Modeling Workbench Python package (Kwakkel, 2017), which offers support for experiment design and exploratory modeling and analysis.

2.3.1. Controlling NetLogo through Python with pyNetLogo

The pyNetLogo package is composed of a Python module (core.py) and a Java JAR file (netlogolink.jar). The Python module defines a `NetLogoLink` class; an instance of this class is used to handle interactions on the Python side. The Python and Java environments are linked with the JPype package through the Java Native Interface (JNI). On the Java side, the JAR file provides a corresponding `NetLogoLink` Java class in two versions, for NetLogo 5.x and 6.0. An instance of the appropriate Java class in turn communicates with the NetLogo API. This allows for bidirectional data exchanges between a Python environment (which can for instance be an interactive Jupyter notebook) and a NetLogo model at runtime, with appropriate data type conversions between the two environments.

Table 2.1 summarizes the basic methods available through the `NetLogoLink` Python class. These are intended to provide "building blocks" for the interactive analysis of NetLogo models with Python, and largely replicate the basic capabilities of the RNetLogo connector for the R environment (Thiele et al., 2012a).

To illustrate the functionality of pyNetLogo, a simple example follows below, using the wolf-sheep predation model which is included in the NetLogo 6.0 example library. Table 2.2 summarizes the model based on its original documentation. The Jupyter notebook available from the pyNetLogo repository replicates this example and demonstrates the key methods of the pyNetLogo connector in more detail, using a slightly modified version of the model to test a broader range of data types.

First, a link to NetLogo is instantiated. This involves starting a Java VM, followed

Table 2.1: pyNetLogo methods

Name	Description	Arguments	Returns
`load_model()`	Load a NetLogo model file	Model path (string)	-
`kill_workspace()`	Close the NetLogo instance	-	-
`report()`	Return the value of a NetLogo reporter	Valid NetLogo reporter (string)	-
`patch_report()`	Return values for an attribute of the NetLogo patches	Valid NetLogo patch attribute (string)	pandas DataFrame of patch values (column labels and row indices following NetLogo patch coordinates)
`patch_set()`	Set NetLogo patch attributes from a pandas DataFrame	• Valid NetLogo patch attribute (string) • pandas DataFrame with same dimensions as NetLogo world, containing attribute values to be set	-
`repeat_command()`	Execute a given command a number of times in the NetLogo environment	• Valid NetLogo command (string) • Number of repetitions (integer)	-
`repeat_report()`	Return the values of one or multiple NetLogo reporters over a given number of ticks	• Valid NetLogo reporter (string or list of strings) • Number of ticks (integer) • NetLogo command used to execute the model (string, 'go' by default)	pandas DataFrame of reported values with columns for each reporter, indexed by NetLogo ticks
`write_netlogo_attriblist()`	Update a set of NetLogo agents of the same type with multiple attributes	• pandas DataFrame containing attribute values to be set, indexed by agent • Valid NetLogo agent breed (string)	-

Purpose	This model explores the stability of predator-prey ecosystems. Such a system is called unstable if it tends to result in extinction for one or more species involved. In contrast, a system is stable if it tends to maintain itself over time, despite fluctuations in population sizes.
Design concepts	In this variant of the model, wolves and sheep wander randomly around the landscape, while the wolves look for sheep to prey on. Each step costs the wolves energy, and they must eat sheep in order to replenish their energy - when they run out of energy they die. To allow the population to continue, each wolf or sheep has a fixed probability of reproducing at each time step. In this variant, sheep do not either gain or lose energy by eating or moving. This variation produces interesting population dynamics, but is ultimately unstable. This variation of the model is particularly well-suited to interacting species in a rich nutrient environment, such as two strains of bacteria in a petri dish.
Parameters	*initial-number-sheep*: The initial size of sheep population *initial-number-wolves*: The initial size of wolf population *wolf-gain-from-food*: The amount of energy wolves get for every sheep eaten *sheep-reproduce*: The probability of a sheep reproducing at each time step *wolf-reproduce*: The probability of a wolf reproducing at each time step

Table 2.2: Summary of the wolf-sheep predation model (Wilensky, 1999).

by starting NetLogo. All interactions with NetLogo are handled by an instance of the `NetLogoLink` class. Note that when using Linux, the NetLogoLink class requires the `netlogo_home` and `netlogo_version` parameters to be set manually. If these parameters are not set on Mac or Windows, the class will attempt to identify and use the most recent NetLogo version found in the default program directory.

Next, we can load a model using the `load_model` method, followed by basic commands to set up the model and run it for 100 ticks. The report method is then used to return NumPy arrays to Python, containing the NetLogo coordinates of the *sheep* agents, and the energy attribute of the *sheep* and *wolf* agents. These arrays can then for instance be used with conventional Python functions to plot the coordinates of the agents, or the distribution of energy across agents (Figure 2.2).

```
import pyNetLogo
netlogo = pyNetLogo.NetLogoLink(gui=True) #Show NetLogo GUI
netlogo.load_model(r'WolfSheepPredation.nlogo')
netlogo.command('setup')

netlogo.repeat_command('go', 100)

x = netlogo.report('map [s -> [xcor] of s] sort sheep')
y = netlogo.report('map [s -> [ycor] of s] sort sheep')
energy_sheep = netlogo.report('map [s -> [energy] of s] sort sheep')
energy_wolves = netlogo.report('map [w -> [energy] of w] sort wolves')
```

Building on this functionality, the `repeat_report` method returns a pandas DataFrame containing reported values over a given number of ticks, for one or multiple NetLogo reporters. The DataFrame is structured using columns for each reporter, and indexed

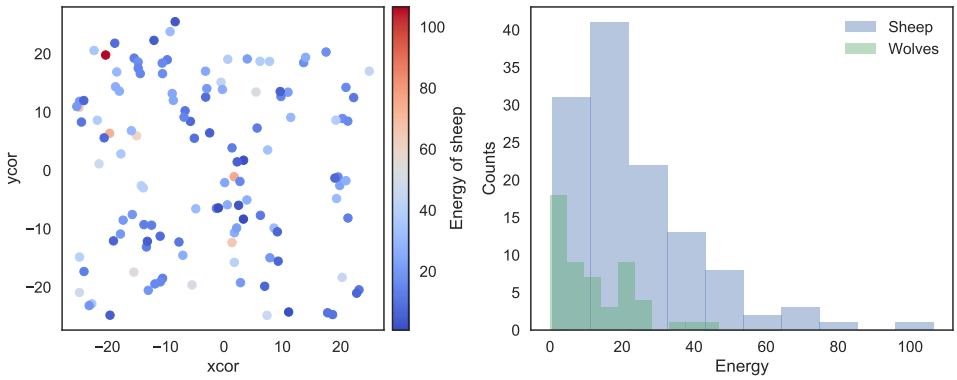


Figure 2.2: Basic plots generated in Python: agent coordinates (left); distribution of energy attribute across agents (right)

by NetLogo ticks (i.e. time steps). By default, this assumes the model is executed with the NetLogo `go` command; this command can be changed by specifying an optional `go` argument when calling the method.

In this case, we can first track the count of both agent types over 200 ticks. The outcomes are first plotted as a function of time on the left panel of Figure 2.3. On the right panel, the number of sheep agents is then plotted as a function of the number of wolf agents, to approximate a phase-space plot.

The `repeat_report` method can also be used with reporters that return a NetLogo list. In this case, the list will be converted into a NumPy array, which is formatted according to the data type returned by the reporter (i.e. numerical or string data):

```
counts = netlogo.repeat_report(['count wolves','count sheep'], 200, go='go')
```

In addition to these reporting methods, the `patch_report` method can be used to return a DataFrame which contains a given patch attribute (in this case, the countdown attribute):

```
patch_df = netlogo.patch_report('countdown')
```

This DataFrame (visualized in Figure 2.4) essentially replicates the NetLogo environment, with column labels corresponding to the pxcor patch coordinates, row indices following the pycor coordinates, and values from the specified patch attribute. The DataFrames can be manipulated with any of the existing pandas functions, for instance by exporting to an Excel file. The `patch_set` method provides the inverse functionality to patch_report, and updates the NetLogo environment from a DataFrame.

2.3.2. Using Python for global sensitivity analysis on a NetLogo model

The Python environment enables access to a wide variety of packages to support the development and analysis of NetLogo models. As an example, this subsection uses the

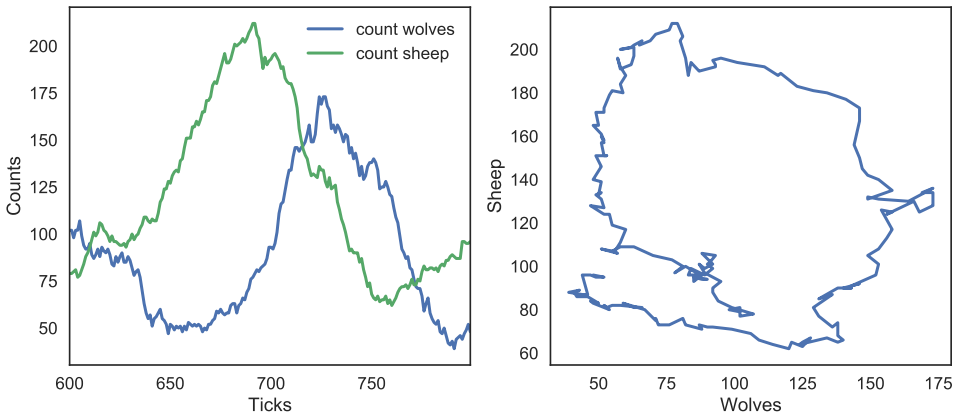


Figure 2.3: Python plots using repeat_report method: number of agents as a function of time (left); number of sheep agents as a function of wolf agents (right)

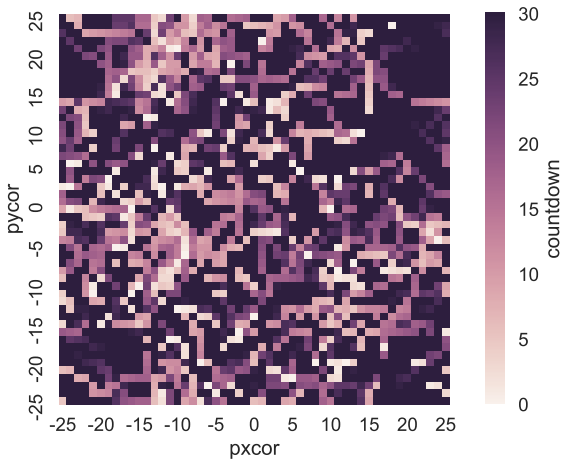


Figure 2.4: Python plot using patch_report method: distribution of the countdown patch attribute across the NetLogo environment

SALib Python library for a global sensitivity analysis (GSA) on the wolf-sheep predation model presented earlier. The full code used for the analysis and visualizations can be found in the Jupyter notebook available from the pyNetLogo repository.

GSA aims to capture the behavior of the model across the full domain of uncertain inputs (see e.g. Saltelli et al. (2008) for a comprehensive overview, and Saltelli and Annoni (2010) for a comparison of GSA with typical "one-at-a-time" approaches). This is especially useful for models in which interactions between parameters can be expected to be significant. A simple example of GSA would be generating a Monte Carlo sample of all uncertain inputs, then applying a multiple linear regression to the model output.

For more complex, non-linear models, variance-based approaches such as Sobol indices Sobol (2001) can accurately capture each parameter's contribution to the variance of model output. Sobol indices are defined using variance decomposition; first-order and total indices respectively estimate the fractional contribution of each input to output variance on its own, and inclusive of interactions with other inputs. Second-order indices can also be computed to estimate the contribution of pairwise variable interactions towards output variance. However, this type of variance-based analysis requires specific techniques for input sampling and output analysis.

In this context, the SALib library provides sampling and analysis modules for methods including Sobol indices, Morris elementary effects (Campolongo et al., 2007; Morris, 1991), and derivative-based global sensitivity measures (Sobol' and Kucherenko, 2009). Integrating these methods within a NetLogo workflow significantly extends the functionality of NetLogo's BehaviorSpace tool, which has limited sampling options. This example will use SALib to estimate Sobol indices; although these indices accurately represent input importances, their calculation may require a large input sample size to yield stable results. For complex models which may be too time-consuming to simulate over such an ensemble of experiments, the Morris elementary effects technique can instead be used from SALib to "screen" non-influential variables at a smaller sample size, while still accounting for parameter interactions and non-linearities by systematically sampling the input space. Ayllón et al. (2016) describe an application of this method for a complex NetLogo model.

SALib relies on a problem definition dictionary (i.e., a key-value map), which contains the number of input parameters to sample, their names (which should here correspond to a NetLogo global variable), and the sampling bounds:

```
problem = {
  'num_vars': 6,
  'names': ['random-seed','grass-regrowth-time','sheep-gain-from-food',
         'wolf-gain-from-food','sheep-reproduce','wolf-reproduce'],
  'bounds': [[1, 100000], [20., 40.], [2., 8.],
  [16., 32.], [2., 8.], [2., 8.]]
  }
```

The SALib sampler will then generate an appropriate experimental design based on the analysis technique to be used. To calculate first-order, second-order and total Sobol sensitivity indices, this gives a sample size of $n(2p+2)$, where p is the number of input parameters, and n is a baseline sample size which should be large enough to stabilize the estimation of the indices.

For this example, we use $n = 1000$, for a total of 14000 experiments. The next subsection will demonstrate the use of ipyparallel to parallelize the simulations and reduce runtime.

```
from SALib.sample import saltelli
from SALib.analyze import sobol

n = 1000
# Generates an input array of shape (n*(2p+2), p) with rows for each
# experiment and columns for each input
param_values = saltelli.sample(problem, n, calc_second_order=True)
```

Assuming we are interested in the mean number of sheep and wolf agents over a timeframe of 100 ticks, we first create an empty DataFrame to store the results. We then simulate the model over the 14000 experiments, by reading input parameters from the *param_values* array generated by SALib and using the `repeat_report` method to track the outcomes of interest over time.

```
results = pd.DataFrame(columns=['Avg. sheep', 'Avg. wolves'])

for run in range(param_values.shape[0]):
    # Set the input parameters
    for i, name in enumerate(problem['names']):
        if name == 'random-seed':
            # The NetLogo random seed requires a different syntax
            netlogo.command('random-seed {}'.format(param_values[run,i]))
        else:
            # Otherwise, assume the input parameters are global variables
            netlogo.command('set {0} {1}'.format(name, param_values[run,i]))

    netlogo.command('setup')
    # Run for 100 ticks and return the number of sheep and wolf agents at
    # each time step
    counts = netlogo.repeat_report(['count sheep','count wolves'], 100)

    # For each run, save the mean value of the agent counts over time
    results.loc[run, 'Avg. sheep'] = counts['count sheep'].values.mean()
    results.loc[run, 'Avg. wolves'] = counts['count wolves'].values.mean()
```

We can then proceed with the analysis, first using a histogram to visualize output distributions for each outcome as shown in Figure 2.5:

Bivariate scatter plots can be useful to visualize relationships between each input parameter and the outputs. Taking the outcome for the average sheep count as an example, we obtain Figure 2.6, using SciPy to calculate the Pearson correlation coefficient (r) for each parameter. This indicates a positive correlation between the *sheep-*

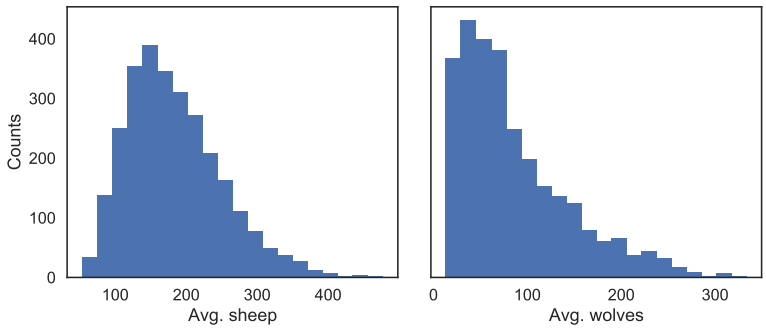


Figure 2.5: Output distributions for the average number of sheep agents (left) and wolf agents (right) over 100 ticks

gain-from-food parameter and the mean sheep count, and negative correlations for the *wolf-gain-from-food* and *wolf-reproduce* parameters.

We can use SALib to calculate first-order (S1), second-order (S2) and total (ST) Sobol indices, to estimate each input's contribution to the variance of the average sheep count. By default, 95% confidence intervals are also estimated for each index. The analysis function returns a Python dictionary.

```
Si = sobol.analyze(problem, results['Avg. sheep'].values)
```

As a simple example, Figure 2.7 visualizes the first-order and total indices and their confidence bounds (shown as error bars) using the default pandas plotting functions, after converting the dictionary returned by SALib to a DataFrame:

The *sheep-gain-from-food* parameter has the highest S1 and ST indices, indicating that it contributes roughly 40% of output variance on its own, and over 50% when accounting for interactions with other parameters. However, the first-order confidence bounds are overly broad due to the relatively small n value used for sampling (i.e. 1000), so that a larger sample would be required for reliable results.

We can use a more sophisticated visualization to include the second-order pairwise interactions between inputs, shown in Figure 2.8. The size of the ST and S1 circles correspond to the normalized total and first-order indices, and the width of connecting lines between variables indicates the relative importance of their pairwise interactions on output variance.

In this case, the *sheep-gain-from-food* variable has strong interactions with the *wolf-gain-from-food* and *wolf-reproduce* inputs in particular, as indicated by their thicker connecting lines.

2.3.3. Using ipyparallel for parallel simulation

ipyparallel is a standalone package (available through the conda package manager) which can be used to interactively run parallel tasks from IPython on a single PC, but also on

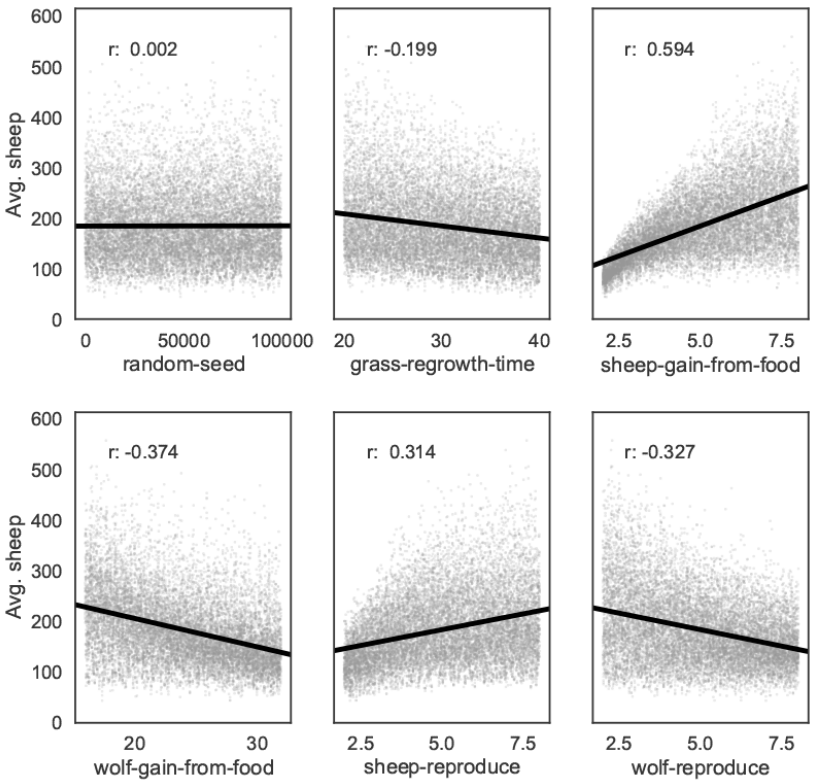


Figure 2.6: Scatter plots with linear trendlines for the average number of sheep agents as a function of each input parameter

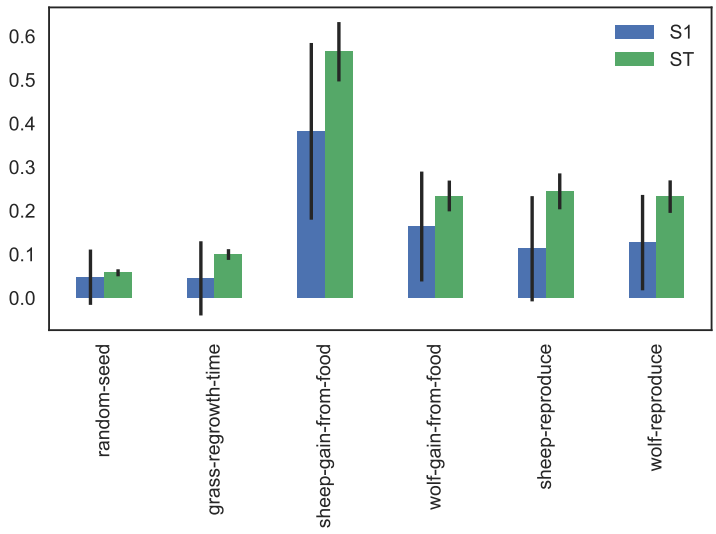


Figure 2.7: First-order and total Sobol indices with confidence bounds, for the average number of sheep agents

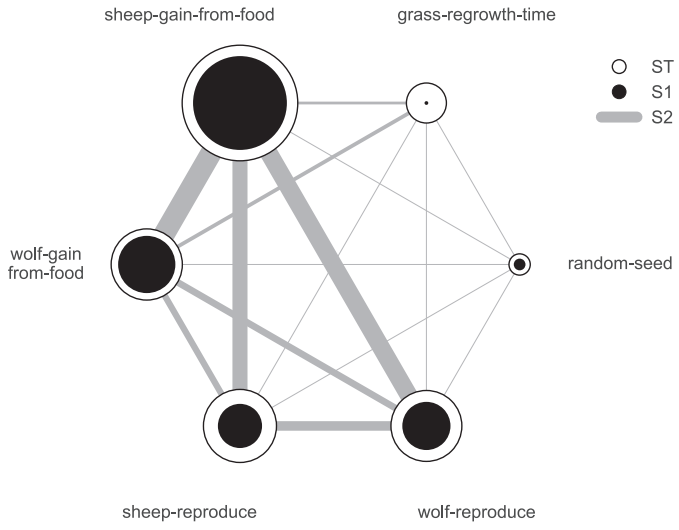


Figure 2.8: First-order, second-order and total Sobol indices for the average number of sheep agents

multiple computers. On machines with multiple cores, this can significantly improve performance: for instance, the multiple simulations required for a sensitivity analysis are easy to run in parallel. This subsection will repeat the global sensitivity analysis presented in the previous subsection, this time using ipyparallel to distribute the simulations across multiple cores on a single computer. The code fragments assume the analysis is executed from a Jupyter notebook; as with the previous examples, the full notebook is available from the pyNetLogo repository.

ipyparallel first requires starting a controller and multiple engines, which can be done from a terminal or command prompt with the following:

```
ipcluster start -n 4
```

The optional -n argument specifies the number of processes to start (4 in this case). By default, the number of logical processor cores will be used.

Next, we can connect the interactive notebook to the cluster by instantiating a client (within a notebook), and checking that client.ids returns a list of 4 available engines.

```
client = ipyparallel.Client()
print(client.ids)
```

After defining the SALib problem dictionary and input sample as in the previous subsection, we can then set up the engines so that they can run the simulations, using a "direct view" that accesses all engines. We first need to ensure the engines can access the current working directory in order to find the NetLogo model. We can then also pass the SALib problem definition dictionary to the engines.

The `%%px` command can be added to a notebook cell to run it in parallel on each of the engines. Here the code first involves some imports and a change of the working directory. We then start a link to NetLogo, and load the example model (assumed to be in the working directory) on each of the engines.

```
import os
os.chdir(cwd)

import pyNetLogo
import pandas as pd

netlogo = pyNetLogo.NetLogoLink(gui=False)
netlogo.load_model(r'WolfSheepPredation_v6.nlogo')
```

We can then use ipyparallel's map functionality to run the sampled experiments, now using a "load balanced" view to automatically handle the scheduling and distribution of the simulations across the engines. This is useful when simulations may take different amounts of time.

We first slightly modify the simulation code used previously, setting up a simulation function that takes a single experiment (i.e. a vector of input parameters) as an argument, and returns the outcomes of interest in a pandas Series.

```
def run_simulation(experiment):
    # Set the input parameters
    for i, name in enumerate(problem['names']):
        if name == 'random-seed':
            # The NetLogo random seed requires a different syntax
            netlogo.command('random-seed {}'.format(experiment[i]))
        else:
            # Otherwise, assume the input parameters are global variables
            netlogo.command('set {0} {1}'.format(name, experiment[i]))

    netlogo.command('setup')
    counts = netlogo.repeat_report(['count sheep','count wolves'], 100)

    results = pd.Series([counts['count sheep'].values.mean(),
                         counts['count wolves'].values.mean()],
                        index=['Avg. sheep', 'Avg. wolves'])
    return results
```

We then create a load balanced view, and run the simulation with the view's `map_sync()` method. This method takes a function and a Python sequence as arguments, applies the function to each element of the sequence, and returns results once all computations are finished. In this case, we pass the simulation function and the array of experiments (*param_values*), so that the function will be executed for each row of the array.

The DataFrame constructor is used to immediately build a DataFrame from the results (which are returned as a list of Series).

```
lview = client.load_balanced_view()
results = pd.DataFrame(lview.map_sync(simulation, param_values))
```

We can then proceed with the analysis as in the previous subsection. Figure 2.9 compares the runtimes obtained with ipyparallel and a sequential simulation (using an Intel i7-6700HQ CPU) for 14000 experiments. The elapsed parallel runtime is approximately one-third of the sequential runtime; given that we were using 4 engines, this is slightly more than could be expected from a perfectly parallel computation, due to the overhead involved in data exchanges, etc.

2.4. Conclusions

The analysis and communication of agent-based models can benefit from the comprehensive analysis features which are available in specialized software environments. To this end, this chapter first introduced the pyNetLogo connector, which interfaces the NetLogo agent-based modelling software with a Python environment. This connector provides basic command and reporting functionalities similar to the RNetLogo package in R. These features were illustrated using one of NetLogo's sample models. As an example of the more complex analyses which are enabled by a Python interface, the SALib Python library was then used for a Sobol variance-based global sensitivity analysis of the model. This analysis was performed using sequential simulations, then parallelized for improved performance using the ipyparallel library.

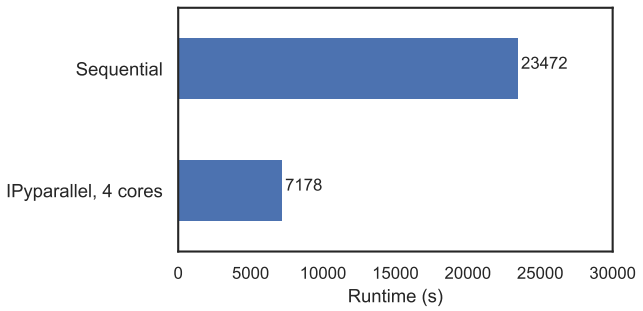


Figure 2.9: Comparison of runtimes for sensitivity analysis (14000 total experiments), using sequential and parallel simulations

The current implementation of pyNetLogo relies on a Java Native Interface (JNI) through the JPype library, which allows Java classes (and thus NetLogo) to be called from Python. However, this does not support a bidirectional linkage through which a NetLogo model could also directly execute Python code. For applications in which this functionality would be helpful (for instance by using more advanced statistical or geospatial functions in NetLogo models), the Python extension for NetLogo can instead be used to execute Python code from NetLogo through a JSON interface. As a complement to existing interfaces which link NetLogo with R or Mathematica, the combination of these tools thus allows modellers to extend NetLogo's capabilities with Python's extensive ecosystem for scientific computing. This approach will be illustrated in the following chapter, by using pyNetLogo as a building block for the design of a Python-based coupled simulation architecture.

3

A coupled simulation architecture for agent-based/geohydrological modelling with NetLogo and MODFLOW

This chapter is based on Jaxa-Rozen, Kwakkel, and Bloemendal (2017a).

3.1. Introduction

Agent-based models (ABMs) are an increasingly popular complement to conventional analytical approaches for the study of environmental problems, by allowing for the simulation of systemic outcomes which emerge from the behavior of individual entities. The bottom-up perspective offered by ABMs is especially relevant in the context of social-ecological systems (SESs), which are complex adaptive systems driven by interacting but relatively distinct social and biophysical subsystems (Ostrom, 2009). These interactions typically include time-dependent feedbacks between environmental and social variables; these feedbacks may lead to a regime shift, i.e. a persistent, significant change in the state of the coupled system (Scheffer et al., 2001), which may otherwise not arise from changes in a single subsystem. However, these interactions and their contribution to transient system behaviour may be overly simplified or left out of scope by conventional modelling approaches, such as computable equilibrium models (Filatova et al., 2016).

ABMs provide an intuitive framework for the study of SESs (Hare and Deadman, 2004). By accounting for interactions across heterogeneous decision-makers as well

as interactions across system levels, they can yield insights regarding the impact of cross-system feedbacks on coupled system behaviors (Schlüter et al., 2012). ABMs can similarly be combined with other paradigms: Vincenot et al. (2011) discuss the complementarities of agent-based models and System Dynamics models, in the case of systems which combine "divisible" and "whole" components – which is typical of coupled socio-environmental systems. However, regardless of the modelling paradigm used, quantitative SES models generally face a range of conceptual and technical challenges due to the complexity of the underlying systems, and the different disciplinary perspectives involved (e.g. Filatova et al., 2013; Voinov and Shugart, 2013). SES models may thus need to combine methods from social and natural sciences to properly represent the coupled dynamics of social and ecological subsystems. As described by Tavoni and Levin (2014), this interdisciplinary approach may be required to fully acknowledge the complexity of environmental and socio-economic systems, and increase the policy relevance of academic models.

To this end, different studies have focused on coupling physical models of environmental systems together with agent-based simulations of socio-economic processes (e.g. Bithell and Brasington, 2009; Kelly et al., 2013; Reeves and Zellner, 2010). Groundwater resources are a particularly relevant case for coupled agent-based/physical simulation: these resources are widely exploited and often scarcely available, and their management can involve complex interactions between heterogeneous users, making ABMs an appropriate modelling option. Furthermore, groundwater resources are often difficult to track and monitor in the subsurface. Numerical modelling can therefore provide useful insights for their management.

However, coupled agent-based/groundwater models may present particular challenges due to the long time scales and model runtimes which are typically involved. Castilla-Rho et al. (2015) describe four issues with conventional approaches for the coupled modelling of such groundwater systems: the limitations of simplified lumped models, the technical complexity introduced by the use of linked software packages and data-exchange libraries, a lack of flexibility in developing scenarios, and the impracticality of performing sensitivity analysis on separate models. To address these limitations, they introduce an interactive environment which directly implements groundwater flow equations in the popular NetLogo agent-based platform. Given an ongoing trend towards increasingly complicated agent-based models of human-environmental systems (Sun et al., 2016), which correspondingly become more difficult to design, this has the advantage of providing a user-friendly environment for the design and use of agent-based groundwater management models.

Nonetheless, this NetLogo-based approach still has drawbacks for those groundwater management problems where geohydrological models are already available and could be directly reused, or where detailed geohydrological modelling is required. This includes problems related to aquifer pollution, in which the transport of contami-

nants plays a role, or studies in which groundwater conditions are subject to significant changes in salinity or temperature – for example due to coastal saltwater intrusion or energy storage, which are complex coupled processes of flow in porous media, chemical reactions, transport and/or heat transfer.

To address this gap, this chapter introduces a simple coupled simulation architecture which can be used to connect the NetLogo platform with the MODFLOW/SEAWAT geohydrological simulation packages, using the Python object-oriented language and the pyNetLogo connector described in Chapter 2 of the thesis. This approach retains NetLogo's simplicity while allowing users to account for complex hydrological processes, or to directly interface the agent-based model with existing MODFLOW/SEAWAT models. Furthermore, the issues raised by Castilla-Rho et al. in relation to scenario design and sensitivity analysis are addressed by relying on existing Python libraries, to establish an easily repeatable workflow for the analysis of coupled agent-based/geohydrological models.

Section 3.2 follows this introduction by describing Aquifer Thermal Energy Storage (ATES) as an example of a social-ecological system which requires detailed geohydrological modelling, motivating the development of an appropriate simulation architecture. This is then placed in the context of recent work related to the coupled agent-based simulation of social-ecological systems. Section 3.3 introduces the different software platforms used in this work; section 3.4 then describes the object-oriented architecture which is used to couple NetLogo and MODFLOW/SEAWAT. This architecture is applied in section 3.5 for a simplified case study of Aquifer Thermal Energy Storage. Section 3.6 summarizes the chapter, along with recommendations for further work.

3.2. Background

3.2.1. Aquifer Thermal Energy Storage as a social-ecological system

ATES systems are used to seasonally store thermal energy in aquifers, which – in combination with a heat pump – can significantly reduce the energy demand of buildings for heating and cooling in temperate climates. These systems involve at least one pair of coupled wells, which inject and extract groundwater at different locations or depths of the aquifer; in winter conditions, relatively warmer water is thus extracted from one well and passed through a heat exchanger for heating, then re-injected into a "cold" well at a lower temperature (typically 5-10°C). Conversely, in summer conditions, the flow across the wells is reversed – so that the cooler water injected in winter is used for cooling, then reinjected into the "warm" well at a temperature of 15-25°C. This process is illustrated on the left of Figure 3.1. This eventually creates thermal zones around each well, which can have a radius of a few dozen meters (shown in plan view for a typical urban layout on the right of Figure 3.1).

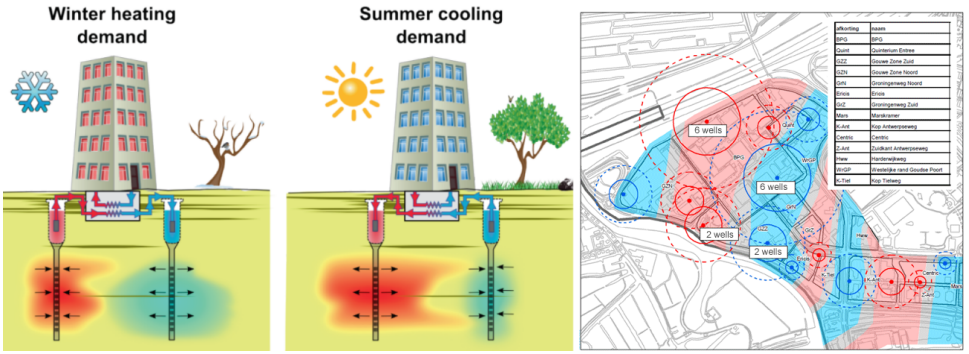


Figure 3.1: Basic ATES operation (left); ATES thermal zones (right) (Bonte, 2013)

The properties of these thermal zones are crucial for the performance and management of ATES systems. They are affected by local geohydrological conditions, such as the porosity of the aquifer or the presence of a regional groundwater flow; thermal interferences between neighbouring systems can also reduce thermal recovery if cold and warm wells are located too closely, while wells of similar temperatures can have beneficial interactions by reducing dissipation from the stored thermal volumes to the ambient medium. These geohydrological and operational factors cause significant uncertainties regarding subsurface conditions and the resulting performance of ATES systems. Furthermore, at the building level, the demand for heating or cooling is difficult to forecast due to variations in building occupancy, weather conditions, or long-term changes in climate. The actual operation of ATES systems (and their associated use of subsurface space) can therefore differ significantly from the expected conditions which are used for permitting and design. For example, although the cold and warm wells would ideally be used symmetrically over the seasons for cooling and heating, ATES systems often have a significant level of thermal imbalance in practice.

Given that local temperature disturbances can persist in the subsurface over a period of decades, these geohydrological and operational variations can cause unforeseen long-term changes in aquifer temperature distributions – which, in turn, can affect the performance of ATES systems, and eventually their continued adoption by building owners. The use of the subsurface for thermal storage can essentially be perceived as a common-pool resource (CPR) problem, due to the subtractable yield of subsurface storage and to the relatively difficult exclusion of potential users (in the absence of appropriate institutional arrangements). As such, some of the problems facing ATES development and planning – e.g. the layout of systems in dense urban areas, or thermal imbalances and interferences – can be related to generic CPR issues such as crowding effects and resource overuse (Kunneke and Finger, 2009). This makes ATES particularly relevant as a case for coupled agent-based/geohydrological simulation: technol-

ogy adoption dynamics and CPR management are classic applications for agent-based modelling, while the complexity of the underlying subsurface processes requires a full-featured geohydrological simulation approach.

3.2.2. Coupled agent-based models for social-ecological simulation

As a complement to analytical methods and empirical studies, the ability of agent-based models to link individual decision processes with aggregate system outcomes has made them increasingly relevant for the study of SESs (Janssen, 2006). These systems are complex and uncertain, and involve extensive feedbacks between social and environmental changes – factors which may not be fully recognized by traditional methods used in policy analysis (Schlüter et al., 2012). By representing different hypotheses related to social and economic decision processes within environmental models, agent-based simulation can be used to explore SES dynamics and contribute to the design of appropriate policies. This may for instance foster a more participative approach to policymaking by providing clear assumptions about user behavior (Matthews et al., 2007). An (2012) extensively reviews decision models for agent-based models of social-ecological systems, covering microeconomic, psychosocial, institutional, participatory and heuristic approaches. Existing frameworks for the study of SESs, such as the IAD and SES frameworks, can also be applied to the conceptualization of agent-based models (Ghorbani, 2013).

CPR problems have been a core application of agent-based models of SESs. The institutional arrangements which are used for the management or self-governance of CPRs involve relationships between multiple system levels, at different temporal and spatial scales – which makes agent-based models a useful tool for their study (Janssen and Ostrom, 2006). As such, Deadman et al. (2000) and Jager et al. (2000) considered the influence of individual decision-making heuristics on collective outcomes in CPR experiments. Other authors have focused on specific case studies, notably in the field of agricultural water management (Becu et al., 2003; Berger, 2001; Schlüter and Pahl-Wostl, 2007).

The agent-based modelling of social-ecological systems can benefit from an accurate representation of environmental dynamics using specialized physical models, but this integration entails additional challenges for modellers. Matthews et al. (2005) review different approaches and challenges for the development of such coupled agent-based models; from a technical perspective, a potential drawback is the complexity of the resulting architecture, making the models more difficult to test and interpret. Similarly, the design of appropriate interfaces and data exchange processes can lead to an overly complex and impractical software architecture.

More fundamentally, addressing different temporal and spatial scales in the social and environmental components is likely to be a key challenge for coupled models. The choice of spatial scale can on its own significantly affect the behavior of a SES model:

larger scales may for instance reduce the relative importance of local interactions and heterogeneity across agents (Gotts and Polhill, 2010). Furthermore, while reconciling spatial and temporal scales across models may be a relatively simple technical issue, it may be more difficult to meaningfully exchange information across models designed for different purposes and scales (Voinov and Shugart, 2013). This can for instance involve aggregating lower-level results (in space or time), at the risk of ignoring important feedbacks. A starting point towards addressing this challenge may be to design models with enough flexibility to test the implications of different choices of scale on the coupled system's behavior.

From this perspective, Bithell and Brasington (2009) recommend a stepwise approach for the development of coupled SES models, with additional detail being added as necessary to describe critical processes. Examples of this approach include Bithell and Brasington (2009)'s coupling of an agent-based decision model, an individual-based forestry model, and a spatially explicit hydrological model, to study spatial dynamics in subsistence farming. Similarly, Reeves and Zellner (2010) coupled a groundwater model with an agent-based layer for the study of land-use changes in Michigan, although this approach only included unidirectional communication between the model components.

While these coupled modelling methods can help capture the complex behaviors of social-ecological systems, the use and interpretation of the models should acknowledge the uncertainties which are present at different levels of the system. On a technical level, model runtimes, large parameter spaces and interactions between components can make it difficult to perform sensitivity analysis on coupled models. This may affect the practical usefulness of the models for decision support or policy analysis, and ultimately their credibility (Saltelli and Annoni, 2010).

Modellers should also acknowledge the fundamentally unpredictable nature of complex adaptive systems. For instance, a typical groundwater management problem will include conventional probabilistic uncertainties such as aquifer heterogeneity, which can be modelled using geostatistical methods, then assessed with a sensitivity analysis. However, the behavior of the coupled system will also be driven by "deep" uncertainties (i.e. uncertainties for which probability distributions or structural relationships are unknown Lempert et al., 2003), which are not amenable to a probabilistic treatment. These include exogenous drivers such as long-term climate conditions, or structural assumptions about decision-making in the social subsystem – which can ultimately produce significantly different emergent outcomes from equally justifiable assumptions. These uncertainties imply that the models would be invalid for predictive purposes, so that their use for decision-making requires a different approach.

Under such conditions, exploratory modelling (e.g. Bankes et al., 2013) can help understand the behavior of the coupled system by using the models for computational experiments, for instance by generating a wide ensemble of plausible models to as-

sess the effect of different uncertainties and modelling assumptions. By representing a broad set of hypotheses about parameters or relationships, exploratory modelling can help identify counterintuitive outcomes, as well as key sensitivities which may usefully guide the collection of empirical data. Furthermore, techniques for scenario discovery can be used as a complement to sensitivity analysis to explore the conditions under which a system may present a specific behavior – for instance, to identify assumptions which would lead to the failure of a simulated policy (Bryant and Lempert, 2010). This approach can contribute to the design of policies which are more robust, i.e. which perform acceptably over a broad range of uncertain futures (Lempert et al., 2006; Rosenhead et al., 1972), rather than attempting to maximize performance under an a priori best-estimate set of uncertain conditions.

3.3. Software description

3.3.1. NetLogo

NetLogo (Wilensky, 1999) is an open-source environment for the design, implementation and analysis of agent-based models, which has become a leading platform for this purpose due to its user-friendliness and active user community. This tool is primarily implemented in Java and Scala, and includes a range of functions and methods to support the rapid development of spatially-explicit agent-based models. Different extension modules are also available, for instance to allow an interface with GIS datasets, or to link NetLogo with the R package (Thiele et al., 2012b); for this work, the pyNetLogo connector presented in Chapter 2 will be used as a link to the Python programming language. As such, NetLogo offers a suitable starting point for the purposes of this work, by facilitating the design of agent-based models and enabling users to focus on the properties of the system under study, rather than on the technical details of software implementation.

3.3.2. MODFLOW/SEAWAT

MODFLOW (Harbaugh, 2005) is a standard code for the simulation of steady and transient groundwater flow in the subsurface, using a finite-difference approach to solve the three-dimensional flow equations for a rectangular grid. It allows for the simulation of representative subsurface conditions (e.g. heterogeneous hydraulic conductivities and transmissivities), as well as external stresses such as precipitation and flows through wells and drains. Additionally, the SEAWAT version (Langevin et al., 2008) couples MODFLOW with the MT3DMS code (Zheng and Wang, 1999); the latter provides a multi-species transport model for the simulation of advection, dispersion, and sorption. This coupling enables the simulation of groundwater flow with variable density and viscosity, and can be applied to study the transport of solutes and heat. This makes the SEAWAT version especially relevant for problems related to aquifer contamination

(Zhang et al., 2013a), or for the study of open or closed geothermal systems (e.g. Bakr et al., 2013; Hecht-Méndez et al., 2010). The simulation architecture described in this chapter is currently compatible with MODFLOW-2005 and SEAWAT 4.

3.3.3. Python

Python is a general-purpose, object-oriented programming language which is increasingly popular for scientific and engineering applications. An extensive set of libraries is available for general data manipulation and analysis, such as Numpy (Walt et al., 2011) and pandas (McKinney, 2010), as well as interfaces with specific software packages and other environments. As such, the pyNetLogo connector is used to interactively communicate with the NetLogo API from Python. In addition, the FloPy library (Bakker et al., 2016) is used for pre/post-processing MODFLOW/SEAWAT input and output files, by interfacing these files with standard Python data structures. The coupled simulation architecture is executed using the EMA Workbench Python package (Kwakkel, 2017). This package can be used to design experiments (e.g. for sensitivity analysis) and provides different features for exploratory modeling and analysis, such as parallel simulation of multiple experiments and built-in visualization.

3.4. An object-oriented architecture for coupling NetLogo and MODFLOW

The Python modules used in this work will be made available under the following repository: `https://github.com/quaquel/pyNetLogo`. These modules have been tested with a standard distribution for scientific Python (Continuum Anaconda 3.6); using this distribution, the modules require the additional installation of the pyNetLogo and Flopy Python packages, which are available with the standard `pip` package manager. More information on pyNetLogo is provided in Chapter 2 of the thesis.

The basic functionality of the pyNetLogo connector can be combined with Python's object-oriented environment to create a link with MODFLOW/SEAWAT models, with Python objects being used as a common interface between the two model components. In the context of groundwater management, interactions are likely to involve stresses such as well flows; using the `agent_functions` module described in this section, these interactions can be mediated through Python objects representing wells, which are "mapped" to corresponding NetLogo agents using the pyNetLogo functions. Parameters such as well flows or injection temperatures can thus be determined in the agent-based model, then passed to the geohydrological model. Figure 3.2 below presents an overview of the overall coupled architecture.

As such, after each step of the NetLogo model, the Python well objects are updated based on the actions taken by NetLogo agents, and generate input files for the geohydrological model using the FloPy library (Bakker et al., 2016). The geohydrolog-

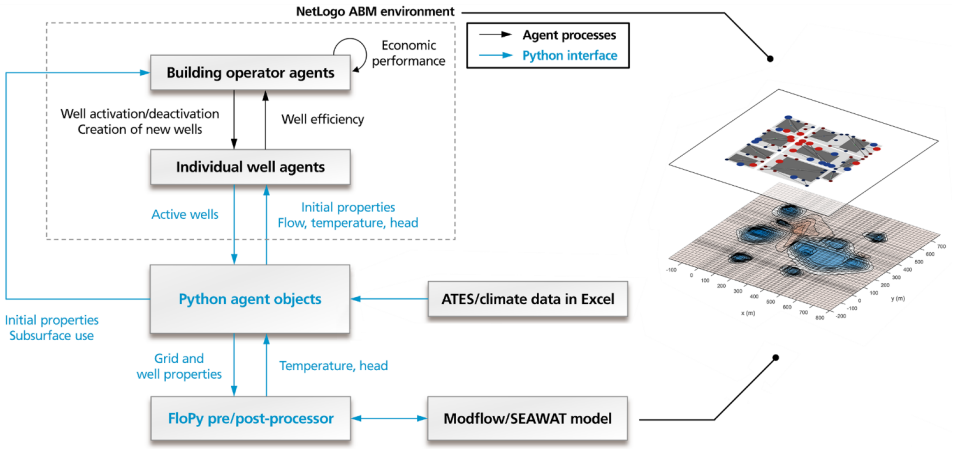


Figure 3.2: General overview of the coupled simulation architecture.

ical model is in turn executed for one period, after which the Python objects process the resulting binary output files using FloPy's utility functions to obtain arrays for hydraulic head and temperature, or any other simulated concentration (e.g. salinity). By default, the execution periods for each model respectively correspond to one NetLogo "tick" and one MODFLOW stress period. A specified subset of the results (such as the effective head and temperature at the grid location of each well) is then passed to NetLogo, so that the geohydrological output can be used as an input for the decision-making routines of agents. Using Antle et al. (2014)'s terminology, the Python objects and NetLogo agents are thus "closely" coupled at runtime, while the Python objects are "loosely" coupled through data exchanges with the MODFLOW/SEAWAT geohydrological model.

The core classes and methods of the `agent_functions` module are described in Figure 3.3 below. In addition to the PyAgent generic agent class, the `PyGrid` class is used to track the properties of the MODFLOW/SEAWAT simulation grid, including spatial and temporal discretization parameters, and output arrays for cell conditions (e.g. hydraulic head, salinity or temperature).

As discussed by Voinov and Shugart (2013), the choice of temporal and spatial resolution is a core question for the development of coupled models of social-ecological systems. By default, the architecture assumes that the models share the same resolution. However, to allow processes to be represented at different resolutions in each model, the time resolution can be modified by setting the `tmult` attribute of the PyGrid object to a desired multiplier for the NetLogo temporal resolution, relative to MODFLOW/SEAWAT. For instance, a value of 2 implies that NetLogo will be run for two time steps in each MODFLOW/SEAWAT time step; conversely, a value of 0.5 will execute MODFLOW/SEAWAT twice for each NetLogo step. Similarly, the

Class/method name	Description	Arguments	Returns
PyAgent	Generic agent class		
create_NetLogo_agent()	Create a NetLogo agent corresponding to the Python object	Object attributes to be passed to the NetLogo agent (list of strings)	-
update_NetLogo_agent()	Update an existing NetLogo agent	Object attributes to be passed to the NetLogo agent (list of strings)	-
update_Python_object()	Update the Python object with properties from the corresponding NetLogo agent	NetLogo attributes to be passed to the Python object (list of strings)	-
PyWell	ATES well class		
calc_LRC()	Locate the well in a PyGrid object to set layer (L) / row (R) / column (C) grid coordinates for MODFLOW / SEAWAT	PyGrid instance	-
PyGrid	Grid class		
make_grid()	Generate grid properties for MODFLOW / SEAWAT – can be used with existing model input files, or with a list of well objects to dynamically refine the grid	-List of well objects (for dynamic grid refinement) -MODFLOW .dis filename (optional)	-
clean_grid()	Dynamically refine the MODFLOW / SEAWAT simulation grid using a list of well objects	-List of well objects (for dynamic grid refinement) -MODFLOW .dis filename (optional)	-
boundaries()	Generate grid boundary properties for MODFLOW / SEAWAT	-MODFLOW .dis filename (optional)	-
update_runtime_objectlist()	Compare active NetLogo agents with the existing Python objects, and create / remove Python objects as needed	- List of objects to compare with the NetLogo agents - NetLogo attributes to be passed to newly created Python objects (list of strings)	Updated list of Python objects
create_obj_from_NetLogo()	Create a list of Python objects corresponding to NetLogo agents	- Python object class to be instantiated - NetLogo attributes to be passed to Python objects (list of strings)	List of Python well objects
write_NetLogo_attriblist()	Update a set of NetLogo agents with a list of attributes from a list of Python objects	List of Python objects for which to update corresponding NetLogo agents	-

Figure 3.3: Basic `agent_functions` classes and methods

`smult` attribute sets a conversion factor for spatial resolution between MODFLOW grid cells and NetLogo environment patches, with values larger than 1 implying a coarser NetLogo resolution.

The UML diagrams in Figure 3.4 summarizes the class structure and action sequence for this example.

3.5. Case study

3.5.1. Case description: a simplified study of ATES

This section uses a simple model which depicts an urban application of Aquifer Thermal Energy Storage (ATES) systems. A more detailed ATES case study is described in Jaxa-Rozen et al. (2015a); for the purposes of this chapter, the model was simplified to minimize runtimes while illustrating typical interactions between the model components, and the plausible impacts on coupled system behavior which these feedbacks

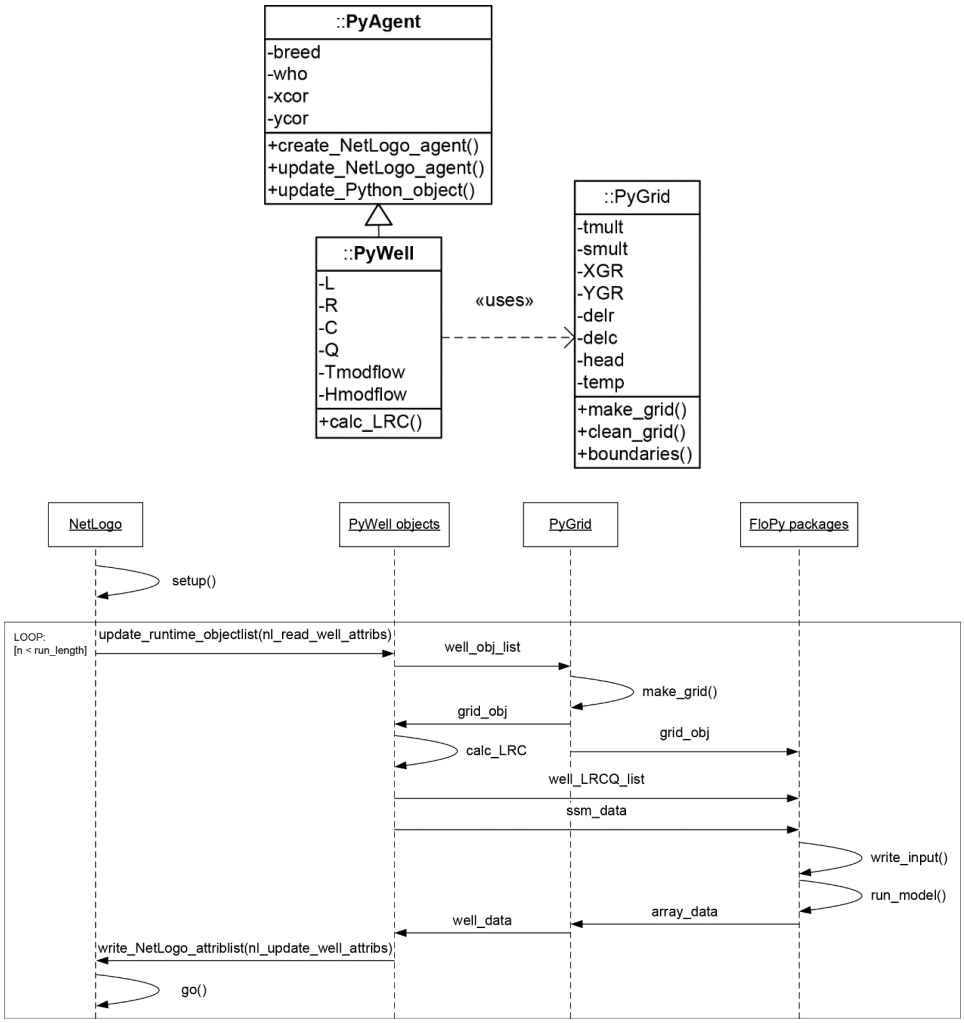


Figure 3.4: Simplified class and sequence diagrams

may cause.

For the present case, a set of 10 NetLogo agents represents simulated building owners which are able to create warm and cold ATES storage wells (also defined in NetLogo as agents of a different type) over a 120-month period. These building and well agents are randomly located in a 1000m x 1000m environment, with a 20m nominal NetLogo patch resolution. The wells follow a predefined pumping pattern over time which corresponds to typical seasonal storage cycles. These well flows are computed in NetLogo at a monthly time resolution, then passed to a single-layer confined aquifer model in MODFLOW / SEAWAT. As discussed by Lo Russo et al. (2014), a

monthly discretization should offer reasonable accuracy when simulating typical ATES flow patterns. For this idealized case, this resolution is therefore chosen as a practical compromise between the seasonal discretization used in previous work on ATES (e.g. Bakr et al., 2013), which may lead to inaccurate temperature estimations, and the daily or weekly resolution needed to represent more complex operating patterns at the building level (e.g. Bloemendal and Hartog, 2018; Rostampour et al., 2016), which would increase model runtimes.

The geohydrological model extents are set to 1400m x 1400m to provide sufficient clearance for temperature distributions to stabilize around the borders of the NetLogo environment, with a 20m thickness and 4m nominal spatial resolution. This resolution ensures that the expected radius of thermal influence around each well covers at least five grid cells, following recommendations by Sommer et al. (2015) for the numerical study of ATES systems in MODFLOW / SEAWAT; in typical conditions, this resolution allows the estimated thermal efficiency to converge to <1%. After executing the MODFLOW / SEAWAT model over one monthly simulation period, the NetLogo well agents are then updated with the effective hydraulic head and temperature at their location.

Figure 3.5 summarizes the main parameters used for the NetLogo and MODFLOW / SEAWAT components; the ranges indicated for certain parameters will be used in the next subsections for runtime evaluation and sensitivity analysis. Parameters for the geohydrological model are derived from typical operating conditions for ATES systems in the Netherlands (Calje, 2010).

The ATES wells are operated in coupled pairs (or doublets) of cold and warm wells, which inject water at each monthly simulation period t with the Q_t^c and Q_t^w rates given below (with negative values corresponding to an extraction of water from the aquifer). As shown in Figure 3.6, the period of the flows is chosen to approximate seasonal storage patterns, with one injection and extraction cycle per 12 months. The total annual pumped volume is representative of a typical commercial building using ATES in the Netherlands. Due to the coupling between wells, these flow patterns do not cause any net extraction from the aquifer over a full annual storage cycle.

These flows can be converted to an equivalent amount of injected or retrieved thermal energy per doublet of wells:

$$\begin{aligned} E_t^{out} &= \begin{cases} Q_t^c C_w (T_{out,t}^w - T_{in}^c)\Delta t, & Q_t^c > 0 \\ Q_t^c C_w (T_{out,t}^c - T_{in}^w)\Delta t, & Q_t^c < 0 \end{cases} \\ E_t^{in} &= Q_t^w C_w (T_{in}^w - T_{in}^c)\Delta t \end{aligned} \tag{3.5.1}$$

The injection temperatures T_{in}^c, T_{in}^w for cold and warm wells are constant and provided in Figure 3.5. The extraction temperatures $T_{out,t}^c$, $T_{out,t}^w$ are updated at each simulation period from the MODFLOW / SEAWAT simulation grid, and are assumed to correspond directly to the temperature $T_{k,t}$ of the grid cell k in which

Parameter	Component	Value or range	Unit	Symbol
Adoption rate	NetLogo	0.02 – 0.1	period^{-1}	α
Distance policy -- min. distance between wells	NetLogo	50 – 200	m	d
Efficiency threshold for adoption	NetLogo	0.8	-	η_a
Efficiency threshold for deactivation	NetLogo	0.7	-	η_d
Grid cell (patch) resolution	NetLogo	4 – 20	m	Δ_p
Length of each simulation period	NetLogo/SEAWAT	30	day	Δt
Maximum ATES well flow	NetLogo	1000	m^3/day	Q_{max}
Maximum number of ATES wells	NetLogo	20 – 300	agents	n
Random seed	NetLogo	-	-	θ
Ambient aquifer temperature	SEAWAT	10	°C	T_{amb}
Ambient groundwater flow	SEAWAT	0 – 20	m/year	v
Aquifer depth	SEAWAT	20	m	Δ_z
Aquifer porosity	SEAWAT	0.1 – 0.5	-	p
ATES injection temperature (cold wells)	SEAWAT	5	°C	T^c_{in}
ATES injection temperature (warm wells)	SEAWAT	15	°C	T^w_{in}
Bulk density of aquifer medium	SEAWAT	1760	kg/m^3	ρ_b
Grid cell resolution	SEAWAT	4 – 20	m	Δ_g
Hydraulic conductivity	SEAWAT	10 – 60	m/day	K
Volumetric heat capacity of water	SEAWAT	4.18*10^6	J/m^3/K	C_w

Figure 3.5: Model parameters for simplified ATES case study

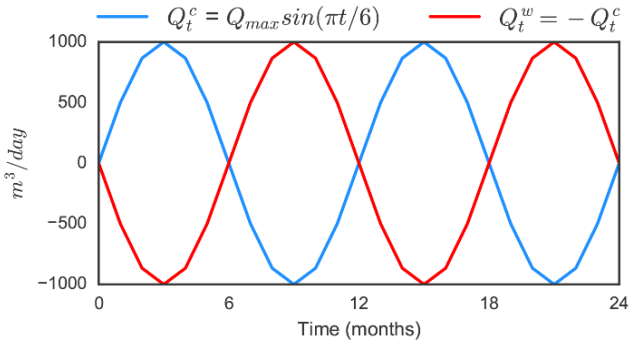


Figure 3.6: Simplified seasonal ATES flows

each well is located. The NetLogo agents then compute the thermal efficiency of each doublet η_t they own, using the ratio of the cumulative retrieved and injected thermal energy:

$$\eta_t = \frac{\sum_{1}^{t} E_t^{out}}{\sum_{1}^{t} E_t^{in}} \tag{3.5.2}$$

The temperatures of the grid cells k can also be used to calculate an indicator ϵ_t, for the fraction of the simulated subsurface volume which presents a significant temperature change relative to the average aquifer temperature T_{amb}. Given that the MODFLOW / SEAWAT cell volumes are uniform, and taking a threshold of 0.1K:

$$\varepsilon_t = \frac{|\{k \mid |T_{k,t} - T_{amb}| > 0.1\}|}{|\{k\}|} \tag{3.5.3}$$

This indicator thus tracks the intensity at which the aquifer is being used for thermal storage over time.

The building agents use simple decision heuristics to build additional wells or deactivate existing wells, based on the thermal efficiency computed from the SEAWAT results using eq. 3.5.2 (after an initialization period of 36 months to let temperature distributions stabilize). If the average thermal efficiency of the wells owned by a building is above a certain threshold η_a, the building agent has a given adoption probability α of adding a pair of coupled warm and cold wells at each time step, until a given maximum number n of ATES wells is reached in the simulation. These new wells are located randomly within the simulated area, at a given minimal distance d from existing wells. If the average thermal efficiency is under another threshold η_d (for instance due to excessive thermal interferences), the agents are assumed to deactivate the wells.

3.5.2. Computational runtime evaluation

Although NetLogo's accessibility has made it popular as a "prototyping" environment for the development of simple agent-based models, it has also successfully been used for more complex models; these can offer comparable performance to base programming languages such as Java when efficiently implemented (Railsback et al., 2017). To support these capabilities, the coupled simulation architecture should therefore be usable with more sophisticated agent-based models without overly increasing runtimes in relation to the individual models. This subsection therefore tests a variant of the ATES case study under different parameterizations for the spatial resolution of the agent-based and geohydrological models, as well as different fixed numbers of ATES well agents (i.e. ignoring the decision heuristics otherwise used to create or deactivate well agents).

The NetLogo model is tested with three different environment resolutions from 4m – 20m (corresponding to 2500, 22,500 and 62,500 patches), while the MODFLOW

/ SEAWAT aquifer grid is tested on five different resolutions in the same resolution interval (ranging from 4900 to 122,500 cells). In addition, the NetLogo model is tested with five numbers of ATES well agents, from 20 – 300. To account for cases in which the MODFLOW / SEAWAT grid would be recomputed over time due to changes in agent locations, the aquifer model grid is recomputed at each time step. At finer resolutions, the MODFLOW grid is representative of a typical complex groundwater case (such as the large-scale ATES case study presented by Bakr et al. (2013), which used 650,000 grid cells).

Figure 3.7 presents the total runtime (in seconds) which is attributed to each model component under these parameterizations, using 30 monthly time periods in each case. For clarity, the figure presents each resolution level as the corresponding total number of NetLogo patches and MODFLOW / SEAWAT grid cells, given that runtimes are more likely to be proportional to these values rather than resolution. Figure 3.8 presents these results expressed as a fraction of total runtime across the three simulation components.

It can be observed from Figure 3.7 that the runtime attributed to the Python architecture scales proportionally to the number of agents, and is mostly independent of the resolutions used in NetLogo and MODFLOW / SEAWAT. In parallel, NetLogo runtime scales roughly proportionally to the number of environment patches, while MODFLOW runtime increases more than proportionally to the number of grid cells. As such, although the fraction of total runtime attributed to the simulation architecture is significant in Figure 3.8 when combined with a large number of agents and coarse model resolutions, it becomes largely negligible at finer resolutions (e.g. 1-3% of total runtime using a 4m resolution in NetLogo and MODFLOW / SEAWAT).

Figure 3.9 further breaks down the runtime performance of the Python architecture in a given parameterization, using 200 agents with 2500 NetLogo patches and 122,500 MODFLOW grid cells. The Python runtime is mostly attributed to interactions between NetLogo and Python through the `update_runtime_objectlist()` and `write_NetLogo_attriblist()` functions, with the grid and agent processes (i.e. methods of the PyGrid and PyAgent classes) being relatively negligible.

Based on these results, the coupled simulation architecture is therefore unlikely to significantly increase computational costs, compared to the runtimes which would be associated with each individual model component – in particular given that the cost of data exchanges scales proportionally to the number of agents, whereas the runtime of more complex agent-based models may scale much more quickly due to interactions or links between agents (Railsback et al., 2017).

3.5.3. Analysis under uncertainty

The use of models for decision support under uncertainty can be facilitated by an integrated environment for experimental design and analysis (e.g. Hadka et al., 2015). To

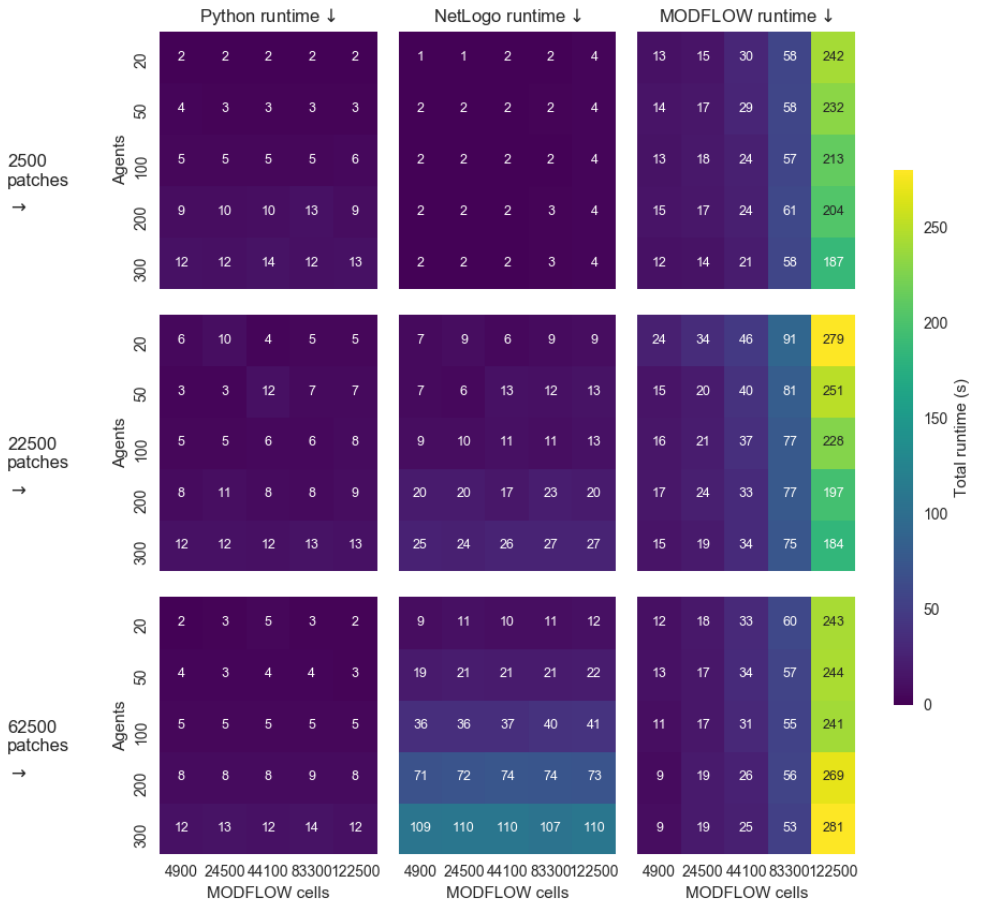


Figure 3.7: Runtime for each model component (left column: Python architecture; middle column: NetLogo model; right column: MODFLOW / SEAWAT model). Each row of subplots corresponds to a different resolution of the NetLogo environment. The rows and columns of each subplots correspond to different NetLogo agent counts and MODFLOW / SEAWAT grid resolutions.

highlight the relevance of such an approach for the coupled agent-based/geohydrological models, this subsection uses the EMA Workbench Python package (Kwakkel, 2017) to test the ATES case under parametric uncertainty. The EMA Workbench provides features for experimental design (for instance using Monte Carlo or Latin Hypercube

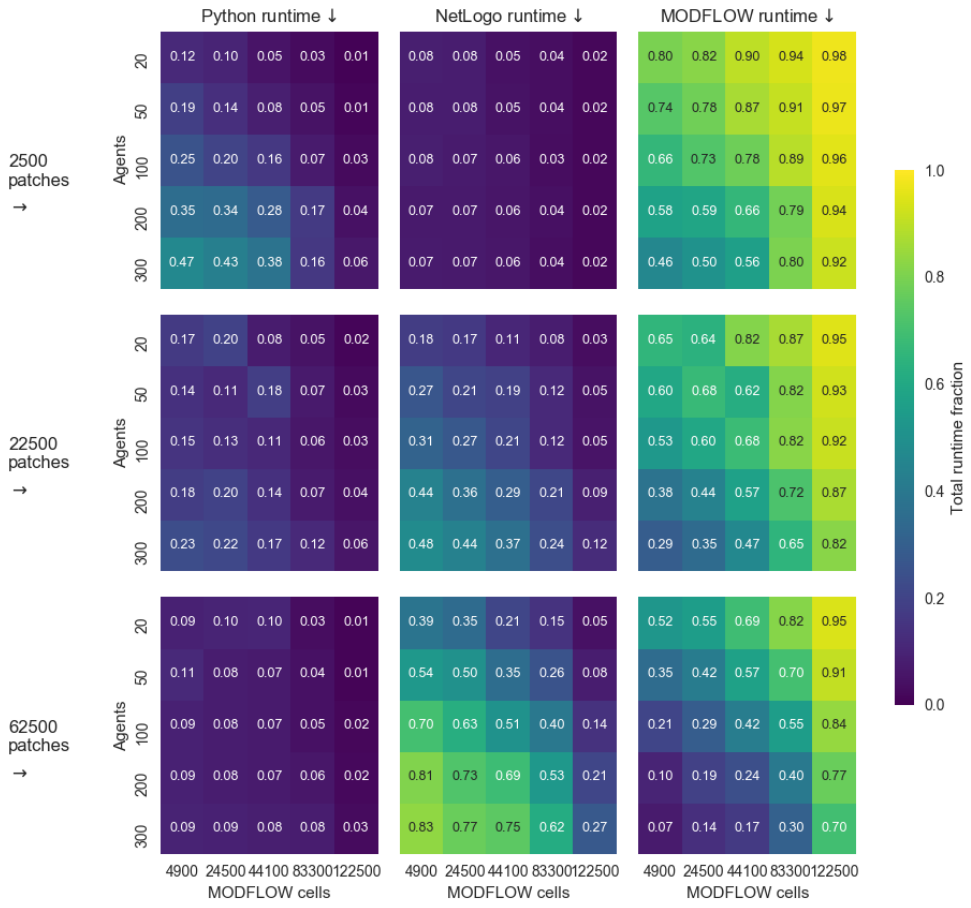


Figure 3.8: Fraction of total runtime for each model component (left column: Python architecture; middle column: NetLogo model; right column: MODFLOW / SEAWAT model). Each row of subplots corresponds to a different resolution of the NetLogo environment. The rows and columns of each subplot correspond to different NetLogo agent counts and MODFLOW / SEAWAT grid resolutions.

sampling), the parallel execution of simulation runs, and exploratory analysis with techniques for sensitivity analysis or scenario discovery. In the context of this work, it enables a consolidated approach for the analysis of the coupled models, for instance by assigning parameters to both model components through a common model interface,

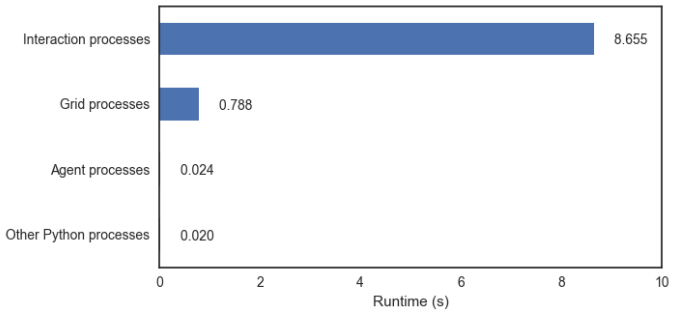


Figure 3.9: Distribution of runtime across Python processes (for 200 agents with 2500 NetLogo patches and 122,500 MODFLOW grid cells)

and by collecting results from both components within a single data structure.

The IPython Notebook provided with this chapter documents this analysis, including the integration of the coupled models within the EMA Workbench, the sampling of uncertain parameters in the NetLogo and geohydrological components, and the post-processing of the results.

Basic exploration

Figure 3.10shows an example of output from the coupled models, showing randomly located ATES well agents in the NetLogo environment, and the corresponding temperature distribution in the PyGrid object.

Using a Latin Hypercube sample of 512 experiments to adequately sample the uncertainty ranges listed in Figure 3.5, the coupled simulation yields a broad range of behaviors, illustrated in Figure 3.11. The figure presents three key outcomes: the number of active ATES wells over time (left graph), the mean thermal efficiency of the ATES systems (middle graph), and the fraction of the simulated subsurface volume which shows a significant temperature change (right graph). The number of active ATES wells and the subsurface usage respectively summarize the states of the agent-based and aquifer model, with thermal efficiency as a key link between the models due to its feedback effect on agent decision-making. The graphs show time series for a random subset of 16 experiments (colored in shades of blue) after an initialization period of 36 months, chosen to stabilize thermal distributions after three annual storage cycles. The shaded envelopes show the minimum and maximum values for each outcome over time, and a kernel density estimator (in the panels to the right of the graphs) illustrates the distribution of experiments within this envelope at the end of the simulation.

The graphs point to different modes of behavior for the coupled models, which are affected by interactions across the NetLogo and MODFLOW / SEAWAT com-

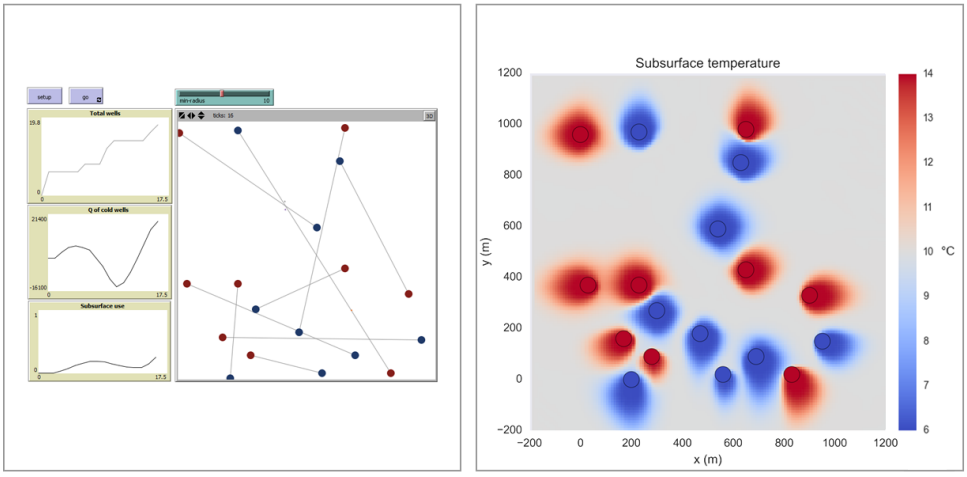


Figure 3.10: Example of coupled model output: NetLogo model (left); SEAWAT temperature distribution (right)

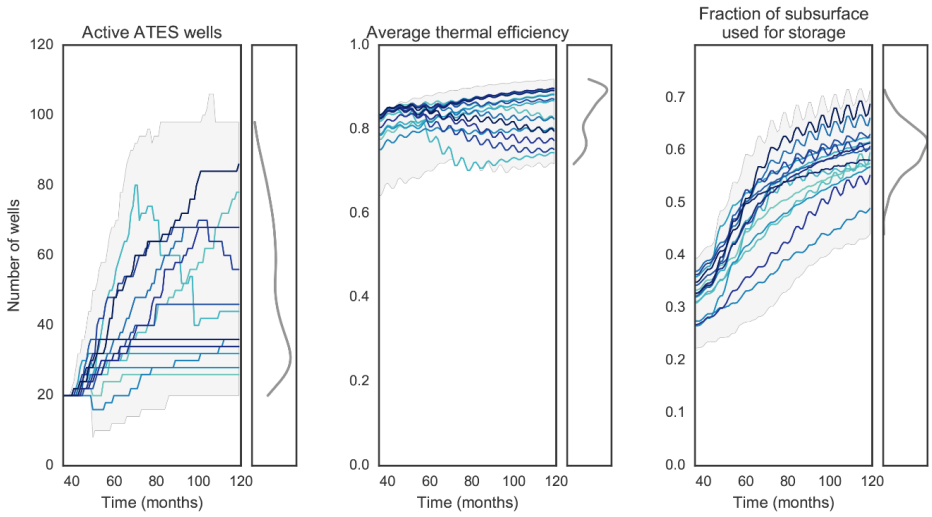


Figure 3.11: Coupled model outcomes over time, for a random subset of 16 out of 512 experiments, after an initialization period of 36 months.

ponents: for instance, under some conditions, a rapid increase in the number of active wells can reduce their average thermal efficiency, due to interactions across neighboring wells – which then leads building agents to deactivate some of the systems. The scatter

plots in Figure 3.12 indicate basic relationships between the three indicators for all 512 experiments, with the number of active wells in NetLogo being associated positively with the used MODFLOW / SEAWAT subsurface volume and negatively with mean thermal efficiency, while the latter has a weaker relationship with the used subsurface volume.

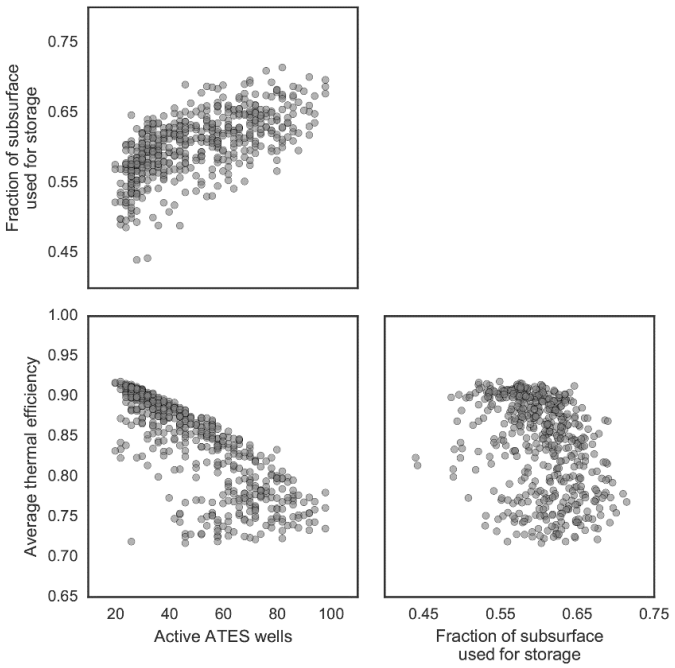


Figure 3.12: Scatter plots for outcomes at the end of the simulation.

Global sensitivity analysis

To better understand the impact of uncertain parameters on the model outcomes and illustrate a typical analysis workflow, this subsection applies a simple global sensitivity analysis to the coupled models using the SALib library (Herman and Usher, 2017). This library can be directly called from the EMA Workbench to generate appropriate sampling designs for common sensitivity analysis techniques (such as Morris elementary effects, Fourier amplitude sensitivity testing, or Sobol indices) and to analyze the model outcomes.

For this example, SALib was used to sample a set of experiments with the Morris technique, using five uncertain parameters across both models with an additional value for the random seed of the NetLogo model (which corresponds to stochastic uncertainty, and here drives the choice of a random location for newly created wells).

Figure 3.13 summarizes the sensitivity analysis results using $u*$ indices (Campolongo et al., 2007) obtained from 160 replications, for a total of 160*(N+1) = 1120 samples. The magnitude of the $u*$ values estimates the relative influence of each parameter on model outcomes; the sum of the $u*$ values can thus be normalized to 1 at each time step for each indicator, so that the area graphs below show the relative importance of each variable over time.

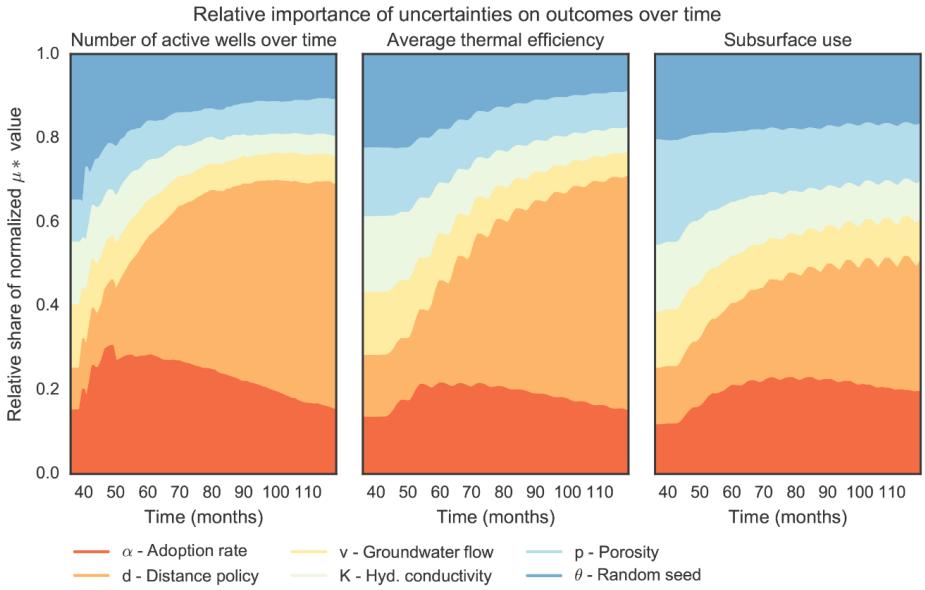


Figure 3.13: Morris sensitivity results over time.

This approach enables the comparison of sensitivities over time and across model components, on a common basis. For instance, the number of active wells over time is mostly driven by the NetLogo parameters for adoption rate and distance policy, with the former being predominant earlier in the simulation, while the latter is most influential on values at the end of the simulation (given that it largely determines how many wells can be built in the simulated area). Similarly, the MODFLOW / SEAWAT parameter for aquifer porosity is most influential on the fraction of subsurface volume used for storage early in the simulation (when the number of active wells is still similar across the experiments), but is then overtaken by the parameters which cause different pathways for ATES adoption in the NetLogo model. Although the geohydrological uncertainties remain significant, this implies that the assumptions made in the NetLogo model are more influential for the overall behavior of the coupled models. It should be noted that this is at least partly driven by the choice of uncertainty ranges: for instance, while the simulated groundwater flow values were typical of average conditions

in the Netherlands, approximately 20% of systems in the country are subject to a higher ambient groundwater flow relative to their storage capacity than was simulated in this idealized case (Bloemendal and Hartog, 2018). These systems may encounter significant losses in thermal efficiency, which could lead to different conclusions regarding the relative importance of agent-related or geohydrological uncertainties.

Nonetheless, this finding is broadly consistent with observations made in relation to more realistic models of social-ecological systems (e.g. Schlüter et al., 2014): although environmental processes may themselves be significantly affected by deep uncertainties, the design choices made in the conceptualization and formalization of agent decision-making almost inevitably have a substantial impact on the outcomes of SES models. Without resolving these uncertainties, an exploratory modelling approach can at least help clarify the implications of these design choices for the behavior of the system.

Scenario discovery

Scenario discovery (Bryant and Lempert, 2010; Kwakkel and Jaxa-Rozen, 2016) aims to identify the combinations of input uncertainties which tend to be associated with given regions of the model output space, using statistical techniques such as the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) or classification and regression trees (CART). This approach is similar to the "factor mapping" setting discussed in the sensitivity analysis literature (Saltelli et al., 2008), and complements the study of variable importances by focusing on specific outcomes or behaviors of interest.

To complete this example, this subsection therefore applies PRIM on the Latin Hypercube sample of 512 experiments previously used for basic exploration, with a given scenario of interest. This scenario is assumed to correspond to a high usage of the aquifer for thermal storage, defined as cases in which the fraction of used subsurface volume is in the top quintile at the end of the simulation; these cases of interest are highlighted in the full ensemble of results in Figure 3.14 below.

Figure 3.15 shows an example of a "box" – or combination of uncertainty ranges – identified by the PRIM algorithm for this scenario. The values next to each variable name indicate the estimated p-value of each parameter, following a binomial test for its significance in this combination. The mass, coverage and density values respectively give the fraction of the total experiments which are within this box, the fraction of all cases of interest for the scenario which are described by the box, and the fraction of experiments within the box which are of interest.

The scenario for a high usage of subsurface volume thus tends to be associated with a low range of the d distance parameter (i.e. between 7.5 and 11 NetLogo patches), a high range for the adoption rate α, and a fairly low range of aquifer porosity p in the SEAWAT model. Based on the density metric, 93.9% of the experiments which are within the given box would be in the top quintile of subsurface use. By identifying influential combinations of uncertainties across the coupled models, scenario discovery

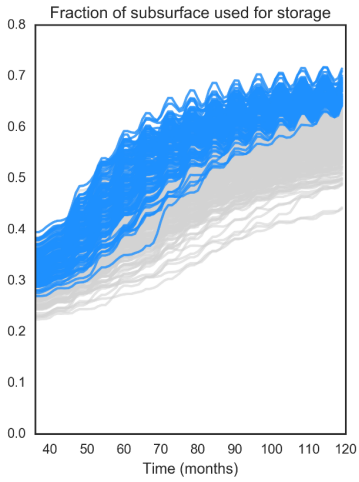


Figure 3.14: Cases of interest for scenario discovery.

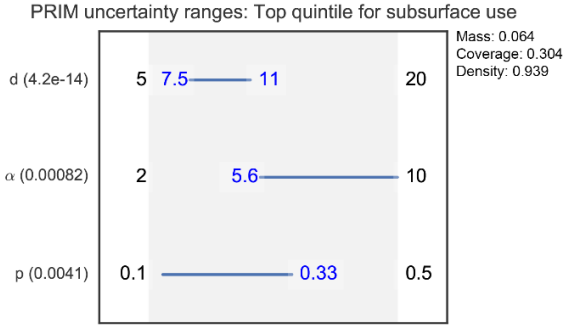


Figure 3.15: PRIM uncertainty ranges identified for cases of interest.

thus provides useful additional information to complement conventional sensitivity analysis techniques. This is particularly relevant for models which present broad ranges of plausible outcomes or different modes of behavior, and for the study of systems which may need to meet explicit performance thresholds.

3.6. Conclusions

This chapter introduced a coupled simulation architecture which interfaces NetLogo and the MODFLOW/SEAWAT geohydrological simulation codes, using Python's object-oriented features. This architecture was applied for a simplified case study of Aquifer Thermal Energy Storage; the operation and deployment of this technology relies on dynamic agent behavior as well as relatively complex subsurface processes (such as

heat transport and sorption), which justified the development of a full-featured simulation architecture which can account for feedbacks between human and groundwater components. As a simple example of the possibilities of this approach, the case study included basic decision heuristics under which simulated ATES owners endogenously adjust the use of the systems based on realized performance. These feedbacks can for instance support more realistic case studies for the long-term adoption of ATES in urban areas, in which adoption patterns are affected by expected technical and economic performance. In parallel, NetLogo's spatial modelling features allow for an intuitive representation of ATES spatial planning, which was here depicted by a minimal distance policy for neighboring storage wells. The following chapters on the thesis will present more comprehensive case studies which build on this approach, for instance by adding a GIS layer to the NetLogo model to account for surface-level location constraints in Chapter 6, and interfacing the NetLogo agents with realistic building operation models in Chapter 7.

As described by Castilla-Rho et al. (2015), the typical drawbacks of coupled agent-based/groundwater modelling include technical complexity, a lack of flexibility in scenario design, and the difficulty of performing coupled sensitivity analysis. The first two of these challenges were here addressed by relying on a simple object-oriented design, which extends the NetLogo model component in an intuitive fashion. Although the coupling requires the external pyNetLogo and FloPy libraries, these libraries are actively supported and at a relatively stable stage of development, which should reduce future compatibility issues. Furthermore, we expect that the possibility of reusing existing MODFLOW/SEAWAT models, while benefiting from the user-friendliness of the NetLogo platform, helps manage the complexity of the overall model development process. An evaluation of the computational costs associated with the different simulation components, under different parameterizations which should be representative of the typical scope of a groundwater management study, also showed that the coupled simulation architecture is unlikely to significantly increase total runtimes relative to the individual models.

The use of the Python language also addresses issues with the coupled analysis of the models, by enabling the straightforward integration of the simulation architecture with different open-source packages available in Python. This approach was demonstrated by using the coupled models with the EMA Workbench; this package can be used to design experiments and analyze the behavior of the coupled models through a common model interface. A typical simulation workflow was illustrated with a global sensitivity analysis of the coupled models, along with a scenario discovery analysis.

As such, while we acknowledge the benefits of a fully integrated platform such as the FlowLogo environment (Castilla-Rho et al., 2015) in terms of technical complexity, we believe that the capabilities of a comprehensive geohydrological modelling environment justify a coupled simulation approach in the case of more complex groundwater

management problems. This approach would indeed be required for studies of aquifer pollution or contamination, or the management of coastal aquifers in which saltwater intrusion is significant (which represents an increasingly pressing issue in the context of climate change adaptation, e.g. Rawlani and Sovacool, 2011). To facilitate future work using these features, we have therefore ensured that the architecture modules are available through an online repository, along with interactive notebooks which replicate the analyses presented in the chapter.

From a broader view, the results of these analyses highlighted the importance of an integrated view for the treatment of model uncertainties, which can be supported by an approach like exploratory modelling: while this case only involved simple behavioral assumptions in the agent-based component, the behavior of the coupled models was sensitive to different parametric values and combinations across the agent-based and geohydrological components. A separate treatment of these uncertainties would have made it more difficult to identify important relationships between the models. Such a consolidated process for the analysis of coupled models can ultimately help analysts better understand the interactions and feedbacks between socio-technical and environmental variables, and contribute to the design of more robust policies for the management of social-ecological systems.

4

Tree-based ensemble methods for sensitivity analysis of environmental models

This chapter is based on Jaxa-Rozen and Kwakkel (2018b).

4.1. Introduction

Sensitivity analysis (SA) is recognized as a key step for analyses which involve the assessment and propagation of uncertainty in mathematical models (Frey and Patil, 2002; Helton and Oberkampf, 2004). In particular, techniques for global sensitivity analysis (GSA) have become an accepted standard for the evaluation of the impact and interactions of uncertain inputs in complex environmental models (as described by e.g. Saltelli and Annoni 2010; Nossent et al. 2011; Pianosi and Wagener 2015). These techniques consider the output behaviour of the model over the full domain of uncertain inputs; specifically, this implies that the full distribution of each input parameter should be evaluated, and that the importance of each input should be evaluated across the domain of all other parameters (Liu and Homma, 2009). This is in contrast to typical applications of "one-at-a-time" (OAT) sensitivity analysis which may only focus on response to changes in individual inputs, and may for instance inadequately capture non-additive responses caused by interactions between input parameters. These properties make GSA particularly relevant for applications such as integrated assessment models, which frequently combine a large number of highly uncertain inputs with a non-linear, non-additive structure. In these conditions, a OAT analysis can lead to an incomplete or misleading interpretation of model uncertainty. As such, GSA can help analysts and decision-makers better understand and communicate the results of complex models,

and ultimately make these models more credible in a decision support context. However, the computational cost of existing GSA methods can quickly become prohibitive with complex simulation models.

This chapter therefore draws on the statistical learning literature to evaluate the performance of decision tree-based ensemble methods, when applied to typical sensitivity analysis problems. These methods rely on ensembles of decision trees which match partitions of the input space with a predicted output, and are commonly implemented using the random forests and Extra-Trees algorithms (Breiman, 2001; Geurts et al., 2006). These techniques perform well at relatively small sample sizes for non-linear regression or classification problems in which input interactions are significant; they are also able to handle both numerical and categorical inputs (Louppe, 2014). Building on previous investigations of decision tree methods for sensitivity analysis (e.g. Harper et al., 2011), this chapter will show that these methods can replicate some of the key insights of GSA by estimating relative variable importances and interactions, at a much smaller computational cost.

In the context of GSA, Saltelli et al. (2008) summarize four analysis objectives, or "settings": *i)* factor prioritization, which identifies inputs (or groups of inputs) which contribute the most towards output uncertainty; *ii)* factor fixing, which conversely identifies inputs which have a negligible contribution to output uncertainty and may thus be fixed at a given value; *iii)* variance cutting, which investigates the assumptions on input values under which output uncertainty can be reduced below a given threshold; and *iv)* factor mapping, which identifies regions of the input space associated with a given output space. Factor prioritization is especially valuable for identifying uncertain inputs on which additional data collection and modelling efforts should be focused, while factor fixing can make models easier to test and interpret by discarding non-influential inputs. These two settings are arguably the most common for sensitivity analysis in environmental modelling. Variance cutting can be applied in risk and reliability analysis, in which analysts may need to meet a certain tolerance (e.g. Plischke et al., 2013; Saltelli and Tarantola, 2002), while factor mapping can be related to techniques for scenario discovery (e.g. Bryant and Lempert, 2010; Kwakkel and Jaxa-Rozen, 2016; Guivarch et al., 2016).

GSA results are typically interpreted through quantitative importance indices, which can be used to compare the uncertain inputs in the context of the desired setting (e.g. factor prioritization or factor fixing). Liu and Homma (2009) and Saltelli (2002b) describe several features of an ideal uncertainty importance index. Notably, the measure should be *i)* unconditional, in the sense of the index being independent of assumptions about the input value (so that the sensitivity metric of an input is not conditional on a given baseline value); *ii)* easy to interpret, for instance by representing an input's proportional contribution to output uncertainty; *iii)* easy to compute numerically; *iv)* stable across different samples (e.g. robust to bootstrapped resamples); and *v)* model-

free, so that the indices are independent from structural properties of the model such as linearity and additivity. Borgonovo (2007) and Pianosi and Wagener (2015) further propose *vi)* moment independence as a criteria, so that the influence of the entire input distribution can be assessed on the output distribution independently of the shape of the latter, without being conditional on a specific moment of the output distribution.

In practice, the estimation of these indices often presents analysts with a trade-off between computational cost, and the information gained from the sensitivity analysis. Variance-based GSA (Sobol, 2001; Saltelli, 2002b) is arguably the most prominent approach in the literature. This technique can be used under factor prioritization or factor fixing settings to directly assess the contribution of uncertain inputs to unconditional output variance. A typical application of the Sobol technique provides first-order and total indices, which respectively describe the fraction of output variance contributed by each factor on its own, and by the sum of first-order and all higher-order interactions for each factor. Additional terms which decompose these higher-order interactions, such as pairwise second-order interactions between variables, can be computed at an additional computational cost. These indices satisfy the above requirements except for moment independence (by relying on variance as a proxy for output uncertainty – which may cause issues with multimodal or skewed distributions, e.g. Pianosi and Wagener, 2015). Given their clear mathematical interpretability and straightforward computation, Sobol indices have for example been increasingly applied for hydrological and integrated assessment models (Tang et al., 2006; Pappenberger et al., 2008; Nossent et al., 2011; Herman et al., 2013; Butler et al., 2014). The indices can also be extended to cover non-scalar inputs – e.g. "switches" for structural model uncertainties – in addition to scalar input ranges (Baroni and Tarantola, 2014). However, the use of variance-based GSA can be difficult for models with a large number of input parameters. In principle, the model evaluations N required to calculate Sobol indices grow linearly with the number of input parameters p, so that $N = n(p + 2)$ for the calculation of first-order and total indices (where n is a baseline sample size). In practice, this baseline sample size also tends to increase significantly for complex models with multiple parameters, and may vary from 100 to 10,000 or more (e.g. Butler et al., 2014, in which $n > 130,000$ was needed for a simulation model with 30 inputs). The computational cost of variance-based GSA may therefore prevent its use for models with a significant runtime.

The literature presents a variety of alternative methods which can be used under such circumstances to reproduce some of the insights of variance-based GSA, at a smaller number of model evaluations. These are often used in a factor fixing setting to screen non-influential variables (see e.g. Kleijnen, 2009 for a review of screening techniques). The elementary effects method (Morris, 1991; Campolongo et al., 2007) is commonly applied to estimate sensitivity measures, using an efficient sampling design to cover the domain of uncertain inputs with a set of sampling trajectories. However,

while elementary effects indices can be used to rank inputs based on their influence on model output, the interpretation of the indices is essentially qualitative rather than quantitative, as the values do not have a clear quantitative meaning (such as the contribution to output variance estimated by Sobol indices). In addition, the sampling trajectories assume continuous inputs, so that these indices are unsuitable for models with categorical or non-scalar inputs; they also do not provide information about specific interactions between variables. The specific input sampling required for elementary effects is also a drawback: for instance, this prevents the use of model datasets which may have been generated from a typical uncertainty analysis, and which could be reused for SA under a "given data" approach (Borgonovo et al., 2017; Plischke et al., 2013). A generic input sampling can otherwise support a multi-method framework which covers complementary aspects of model sensitivity at the same computational cost (such as Pianosi et al. (2017)'s framework for the estimation of first-order indices, density-based indices, and interactions using a Latin Hypercube sample). Under Liu and Homma (2009)'s criteria, the elementary effects indices would therefore be suboptimal in terms of interpretability and ease of computation.

These sensitivity analysis methods have largely been developed and applied in the context of model-based risk analysis and environmental science. However, a parallel domain of research has also focused on the problem of feature selection in statistical learning, which offers some useful analogies to the factor fixing setting in sensitivity analysis. As described by Guyon and Elisseeff (2003), feature selection aims to reduce the dimensionality of the input data used in a learning problem by selecting a subset of the original variables, and eliminating variables which are not relevant. This process offers several advantages, such as making output data easier to analyze, making the prediction model more understandable, or improving the accuracy of the prediction model by avoiding overfitting. Several definitions of variable relevance (described more extensively in e.g. Blum and Langley, 1997; Kohavi and John, 1997) can be followed, leading to different paths for feature selection. For instance, the feature selection literature describes *wrapper* methods, in which variables are assessed based on their relevance for a given predictor (Kohavi and John, 1997). In this application, feature selection aims to select a subset of variables which maximizes the accuracy of a predictor, which is considered as a "black box". When combined with a suitable predictor, this approach enables a more flexible analysis, for instance by relaxing assumptions on input types or distributions (Lazar et al., 2012). Decision trees are a popular example of such a predictor, which combine several desirable properties for statistical learning in general, and for feature selection in particular. As such, these predictors can represent arbitrary relations between inputs and outputs, without prior assumptions about inputs or structural relationships (Louppe, 2014). They can also be used for non-linear problems with heterogeneous input data (such as continuous or categorical parameters), and implicitly account for variable interactions. Decision trees are therefore a popular option for

feature selection (Guyon and Elisseeff, 2003); they are commonly used within ensemble methods which combine multiple decision trees to improve performance, such as random forests and Extra-Trees (see e.g. Hapfelmeier and Ulm, 2013 for a review of random forests in a feature selection context).

It is therefore interesting to assess whether insights from the literature on feature selection can be transferred to the sensitivity analysis of complex environmental models. In particular, decision tree-based predictors may mitigate some of the drawbacks of common screening techniques, as they can be applied with generic input sampling designs and categorical uncertainties, while supporting the study of variable interactions. This work builds on past applications of decision tree-based predictors in the environmental modelling literature, such as Harper et al. (2011); this study combined random forests and individual trees to evaluate variable importances and interactions in a model of cottonwood dynamics. Similarly, Almeida et al. (2017) and Singh et al. (2014) used individual classification trees to study critical thresholds in a factor mapping setting, for a hydro-climactic watershed modelling framework and for a model of slope stability, respectively. Given the demonstrated performance and widespread availability of the random forests and Extra-Trees ensemble predictors, this chapter will focus on comparing both of these methods with the reference Sobol and elementary effects techniques, using typical model cases.

Section 4.2 provides more background about the Sobol and elementary effects methods for global sensitivity analysis, and describes the selected decision tree-based methods. Section 4.3 then compares the performance of the tree-based ensemble methods against reference GSA results, for three cases: an Ishigami test function, a H1N1 flu pandemic model, and the CDICE integrated assessment model. Section 4.4 discusses the results and describes potential avenues for future work.

4.2. Methods

4.2.1. Reference methods for global sensitivity analysis

Variance-based Sobol indices

The Sobol technique for global sensitivity analysis uses variance decomposition to establish the contribution of each uncertain input to the unconditional output variance of a model, which can be non-linear and non-additive (e.g. Sobol, 2001; Homma and Saltelli, 1996). Given a model output Y and a set $X = (x_1, ..., x_p)$ of independent parameters, the corresponding function $f(X)$ can be decomposed into a finite number of terms of increasing order:

$$Y = f(X) = f(x_1, ..., x_p) \tag{4.2.1}$$

$$f(x_1, ..., x_p) = f_0 + \sum_{j=1}^{p} f_j(x_j) + \sum_{j=1}^{p} \sum_{k=j+1}^{p} f_{jk}(x_j, x_k) + ... + f_{1,...,p}(x_1, ..., x_p) \tag{4.2.2}$$

This decomposition (detailed in Sobol, 2001) assumes that $f(X)$ and the individual terms are square-integrable. The unconditional variance $V(Y)$ can correspondingly be decomposed into partial variances, where e.g. V_j and V_{jk} represent the variances of f_j and f_{jk}, respectively:

$$V(Y) = \int_{\Omega} f^2(X)\, dX - f_0^2 \tag{4.2.3}$$

$$V(Y) = \sum_{j=1}^{p} V_j + \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} V_{jk} + ... + V_{1,...,p} \tag{4.2.4}$$

Using these partial variances, the first-order, second-order and total Sobol sensitivity indices can then be defined in relation to the total variance:

$$S_j = \frac{V_j}{V(Y)} = \frac{V_{x_j}\left[E_{X\sim j}(Y|\, x_j)\right]}{V(Y)} \tag{4.2.5}$$

$$S_{jk} = \frac{V_{jk}}{V(Y)} = \frac{E_{X\sim x_j, x_k}\left[V_{x_j, x_k}(Y|\, X_{\sim x_j, x_k})\right]}{V(Y)} - S_j - S_k \tag{4.2.6}$$

$$ST_j = 1 - \frac{V_{\sim j}}{V(Y)} = \frac{E_{X\sim x_j}\left[V_{x_j}(Y|\, X_{\sim x_j})\right]}{V(Y)} \tag{4.2.7}$$

The first-order index S_j , or main effect, represents the fraction by which the output variance would be reduced on average by fixing x_j within its range. The second-order index S_{jk} then represents the fraction of output variance linked to inputs x_j and x_k which is not captured by the superposition of each input's first-order index, and thus corresponds to interaction effects in a non-additive model. Finally, the total effect ST_j includes the contribution of the first-order effect and the sum of all higher-order interaction effects. For a non-additive model, the difference $ST_j - S_j$ thus indicates the importance of interaction effects, which can be directly assessed for pairwise interactions using the second-order S_{jk} index. These indices can be used

for factor prioritization, in which the input parameters with the highest main effect S would be assessed as the most influential. Conversely, for factor fixing, input parameters with $ST \approx 0$ can be judged to be non-influential and discarded from the analysis, given that they do not contribute to output variance either through their main effect or through interactions (Saltelli et al., 2008). As shown by Baroni and Tarantola (2014), these indices can similarly be applied to assess the contribution of non-scalar inputs (such as structural model "switches") to output variance.

In practice, the unconditional variance $V(Y)$ typically needs to be estimated using Monte Carlo integrals rather than an analytical form. Saltelli (2002a) for instance presents an input sampling strategy which can be used to estimate the first-order, second-order and total indices at a cost of $N = n(2p + 2)$ evaluations. This sampling design has been implemented in a variety of software packages; for the purposes of this work, the Python SALib library (Herman and Usher, 2017) is used to generate input samples, and to calculate the resulting Sobol indices with bootstrapped confidence intervals.

Morris elementary effects

For models with a large number of uncertain inputs and/or a high computational cost, the elementary effects method is used as a standard screening technique for factor fixing (Morris, 1991; Campolongo et al., 2007). The method relies on a systematic sampling of the input space to generate a randomized ensemble of "one-at-a-time" experiments. Taking a set $X = (x_1, ..., x_p)$ of independent input parameters transformed so as to be uniformly distributed in the interval [0,1], a certain number r of sampling "trajectories" of $(p + 1)$ points are then constructed by varying one input at a time, across k levels of the [0,1] input domain. Starting from a given value of X and taking $\Delta \in \{1/(k - 1), ..., 1 - 1/(k - 1)\}$, the elementary effect of x_j is given by:

$$EE_j(X) = \left(\frac{f(x_1, ..., x_{j-1}, x_j + \Delta, x_{j+1}, ..., x_p) - f(X)}{\Delta} \right) \tag{4.2.8}$$

The distribution F_j of this elementary effect can then be obtained by sampling multiple initial values of X. Morris (1991) originally proposed using the mean μ and standard deviation σ of this distribution to respectively assess the overall influence of the variable on output, and the magnitude of higher-order effects due to non-linearities and interactions. However, the μ measure was shown to be vulnerable to type II error (i.e. potentially ignoring influential variables) in the case of non-monotonic models, as elementary effects may cancel each other out at different points of the input set X. Campolongo et al. (2007) thus introduced a measure $\mu*$, which takes the mean of the distribution of the absolute values of the elementary effects. This index was shown to acceptably estimate the ST indices obtained from a variance-based global sensitivity

analysis. The elementary effects technique can thus reliably be used to identify factors which have a negligible influence, and which may be discarded from the analysis. However, the index σ is more difficult to interpret; it combines the effect of interactions as well as non-linearities, so that specific interactions between pairs of variables cannot be evaluated. The assumption of scalar inputs X also makes the indices unsuitable for categorical inputs.

As with the Sobol technique, the SALib library will be used to sample input trajectories (with the efficient trajectories introduced by Campolongo et al., 2007) and to estimate the elementary effect indices.

4.2.2. Decision tree-based ensemble methods

Decision trees are a simple and well-established general approach for statistical learning; such trees aim to identify the splitting criteria which describe the relationship between a set of input combinations, and regions of the output space (graphically illustrated for an idealized case in Figure 4.1). Decision trees can be fitted through several specific algorithms, such as classification and regression trees (CART; Breiman et al., 1984). The right panel of Figure 4.1 presents a simple example of a regression tree for a test case $y = f(X)$. The tree is fitted to an output vector $y = (y_1, ..., y_i)^T$, with vectors of predictor values $x_j = (x_{1,j}, ..., x_{i,j})^T$ for $j \in \{1, 2\}$, forming the matrix $X = (x_1, x_2)$. The depth of the tree is here artificially constrained to create a small number of nodes t; each of the terminal nodes t_3, t_4, t_5, t_6 corresponds to a rectangular region of the input space shown in the left panel of the figure. The predicted value $\hat{y_t}$ at each node (i.e. for each corresponding combination of ranges for the predictor values) is then the mean of the output values in each node, $y_i \in t$.

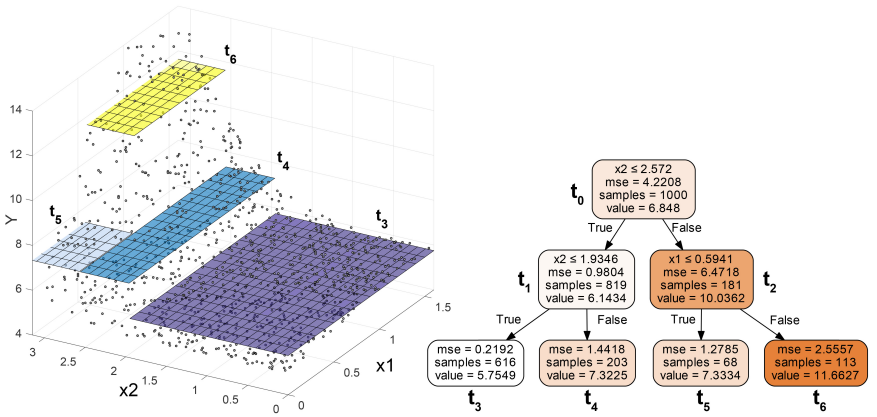


Figure 4.1: Graphical representation of a regression tree. Left panel: two-dimensional partition of a feature space; right panel: decision tree corresponding to the partition.

Starting from the root node t_0, the tree is "grown" using an optimization procedure to search over all possible binary splits $s = (x_j \leq c)$, and identify the splitting point c across the values of variable x_j which leads to the greatest reduction in the impurity of the resulting "child" nodes (typically using Gini impurity for classification, or mean square error for regression). We let t_L and t_R represent the left and right child nodes obtained when partitioning node t with a binary split. The reduction in impurity from split s at node t is then:

$$\Delta i(s,t) = i(t) - \frac{N_{tL}}{N_t} i(t_L) - \frac{N_{tR}}{N_t} i(t_R) \tag{4.2.9}$$

where N_t, N_{tL}, N_{tR} are the number of samples in the parent node and the left and right child nodes. For regression, we use the mean square error as a measure of impurity, considering the predicted value $\hat{y_t}$ at each node:

$$i_R(t) = \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - \hat{y_t})^2 \tag{4.2.10}$$

In the example shown in Figure 4.1, this leads to the selection of a splitting value $c_0 = 2.572$ on x_2 in the root node. This splitting procedure is repeated until a stopping criterion is reached, which can be e.g. the depth of the tree, or the maximum number of samples to be found in a terminal node.

Individual decision trees will typically display high levels of variance, so that small changes in the selected input data may cause significant changes in the structure of the fitted tree. As such, ensemble methods – in which multiple, randomly generated instances of an estimator are aggregated – can improve the performance of decision trees on classification and regression tasks. The most popular of these has been the random forests algorithm (RF; Breiman, 2001), in which multiple CARTs are fitted to bootstrap samples of the data and aggregated (or "bagged"). This bagging step will tend to reduce the variance of the resulting estimator, making it more robust than individual trees. The trees are randomized by selecting a subset of the input variables as candidates for splitting at each node. Their predictions are then simply averaged for a regression problem, or taken as a majority vote for classification. Geurts et al. (2006) add an additional randomization step for the construction of "extremely randomized trees" (or Extra-Trees), in which the random selection of variables for splitting is combined with randomized cutting points at each node (typically using the full input set, rather than bootstrap samples). This step can improve accuracy as well as computational performance. This work will thus focus on the RF and Extra-Trees (ET) algorithms, due to the demonstrated accuracy and versatility of these techniques for non-linear regression problems with heterogeneous inputs (Hastie et al., 2009; Louppe, 2014).

The performance of random forests and Extra-Trees can be tuned with parameters which control the construction of the ensemble. The most significant of these are *i)*

the number of trees T used in the ensemble, *ii)* the size of the candidate subset m of the input variables p which is assessed for each split of the individual trees, and *iii)* the depth to which the trees are grown (which can be controlled with the same criteria described above for individual trees, such as the minimum number of samples N_{leaf} to be left in the nodes created after a split).

Increasing the number of trees T used in the ensemble will in principle reduce prediction error, with the methods being robust to overfitting (Geurts et al., 2006). In practice, the size of the ensemble is likely to be driven by computational constraints, with a trade-off between accuracy and time. The size of the subset of variables m will affect correlation between the trees within the ensemble, with a smaller value increasing randomness; in the extreme case of $m = 1$, each split is determined by a single random input, and the trees are said to be totally randomized. The choice of this parameter depends on the problem, with $m = p/3$ as a starting point for regression (Hastie et al., 2009). Finally, the depth of the trees will affect generalization error: fully developed trees may overfit the data, while smaller trees will typically have larger bias. The empirical results presented by Geurts et al. (2006) suggest a value of $N_{min} = 5$ as a robust starting point for regression, for the minimum number of samples required to split a node.

Variable importance metrics

Different measures can be used to assess the importance (or predictive strength) of input variables in random forests and Extra-Trees. The most common metrics are *Mean Decrease Impurity* (MDI) and *Mean Decrease Accuracy* (MDA) (Breiman, 2001; detailed in Louppe, 2014). MDI relies on the criterion used to select an optimal split in CART (defined in eq. 4.2.11), extending it across the ensemble of trees. The MDI importance of a variable x_j can thus be computed from the total decrease in node impurity (across the trees in the ensemble) which is obtained when x_j is used for splitting. A variable associated with a large decrease in impurity is then influential. We use the definition given by Louppe (2014), with an ensemble of T trees:

$$MDI(x_j) = \frac{1}{T} \sum_{\tau=1}^{T} \sum_{t \in \varphi_\tau} 1(j_t = x_j) \left[\frac{N_t}{N} \Delta i(s_t, t) \right] \tag{4.2.11}$$

where the change in impurity $\Delta i(s_t, t)$ is summed in tree φ_τ over all nodes t in which x_j is used for splitting, weighted by the fraction of total samples present in the node (N_t/N); j_t is the variable used for splitting at node t. This value is averaged over all trees φ_τ in the ensemble.

An alternate measure is given by the MDA (or permutation) importance, in which the change in prediction accuracy of the ensemble is assessed after randomly permuting the input values for variable x_j. When using bootstrapping, MDA can be estimated on the out-of-bag (OOB) samples at each tree, i.e. the samples which were not part

of the bootstrapped training set for each tree. Following Strobl et al. (2008), we compare prediction accuracy on the OOB samples for the original vector of input values $x_i = (x_{i,1}, ..., x_{i,p})$, and for a vector ${x^{\pi j}}_i$ in which the values of x_j are permuted across the observations i: ${x^{\pi j}}_i = (x_{i,1}, ..., x_{\pi j(i),j}, x_{i,j+1}, ..., x_{i,p})$. An influential variable will cause a large decrease in prediction accuracy, while a non-influential variable would not significantly change the performance of the ensemble. The mean square error is typically used as a measure of prediction accuracy for regression. Taking $\hat{y}_{\tau,i} = \varphi_\tau(x_i)$ as the prediction given by tree φ_τ for observation i, averaging over each observation in the set of OOB samples B^τ , then averaging over the ensemble of trees, we obtain:

$$MDA(x_j) = \frac{1}{T}\sum_{\tau=1}^{T} \frac{\sum_{i \in B^\tau} (y_i - \varphi_\tau({x^{\pi j}}_i))^2 - (y_i - \varphi_\tau(x_i))^2}{|B^\tau|} \tag{4.2.12}$$

These variable importance measures have been extensively studied and refined in the context of RF and feature selection (e.g. Ishwaran, 2007; Strobl et al., 2007, 2008; Wright et al., 2016; Bureau et al., 2005). An advantage of the measures is their implicit consideration of interactions across variables, which follows from the tree induction process. This makes RF importance measures a potential candidate for approximating the total effect indices obtained through global sensitivity analysis. In feature selection, Qi et al. (2006) for instance found that RF outperformed five other classifier methods for the detection of interactions in large datasets. However, the MDI metric tends to be biased towards especially salient variables, due to the underlying bias of the splitting criterion. In the case of categorical variables, MDI also tends to be biased towards inputs with a larger number of categories (Strobl et al., 2007). The bias of the MDA measure was less obvious in the results discussed by Strobl et al. (2007) but can nonetheless affect the reliability of the measures, particularly in the case of correlated predictors. Strobl et al. (2009) and Altmann et al. (2010) thus introduced revised metrics to address these characteristics. For the purposes of this work, the relative values of the ST and $\mu*$ indices obtained from the Sobol and elementary effects techniques will be compared to the standard MDI importance index, which offers better computational performance than MDA on large datasets. The revised metrics of e.g. Strobl et al. (2009) are less relevant for this application, due to the typical assumptions on uncorrelated parameters which are used when sampling inputs for sensitivity analysis.

In addition to MDI, we use a variant of the MDA metric (Bureau et al., 2005) to directly estimate the effect of pairwise interactions between variables, by permuting both of the corresponding input samples across observations in a vector ${x^{\pi j,k}}_i$, and

subtracting individual MDA importances. For variables x_j, x_k, the pairwise MDA is then given by:

$$MDA(x_j, x_k) = \left[\frac{1}{T}\sum_{\tau=1}^{T} \frac{\sum_{i \in B^\tau} (y_i - \varphi_\tau(x^{\pi j,k}{}_i))^2 - (y_i - \varphi_\tau(x_i))^2}{|B^\tau|}\right] - MDA(x_j) - MDA(x_k) \tag{4.2.13}$$

To assess the stability of the MDI indices, we use a convergence criterion presented by Touzani and Busby (2014) (eq. 4.2.14), where $V_N = (v_1, \dots, v_p)$ is the vector of estimated variable importance indices at a sample size of N observations. The criterion considers the Euclidean norm $\|\cdot\|$ of the vector rather than individual indices, so that more influential indices have a greater effect on measured convergence. We compute the indices sequentially over an increasing sample size at intervals of ΔN total samples. The convergence criterion κ_N is then computed backwards from N over t intervals, with t and ΔN being specified for each case study. This criterion will also be used to ensure the stability of the reference vector of Sobol ST indices, ST.

$$\kappa_N = \frac{1/t \sum_{s=1}^{t} \|V_N - V_{N-s\Delta N}\|}{\|V_N\|} \tag{4.2.14}$$

Finally, the accuracy of the proportional estimated variable importances is assessed with the root mean square error and mean bias error of V_N, relative to ST. As the indices measure different quantities (e.g. the decrease in mean square error for MDI, and fraction of output variance for ST), the values are not directly comparable; however, by first rescaling each vector relative to its maximum value across all p variables, we can compare the proportional importances estimated by each method. We avoid normalizing the estimated importances over $[0, 1]$ to preserve negative values which may indicate numerical artifacts. RMSE is used as an overall indicator of accuracy, while MBE provides information about the average over-estimation or under-estimation of variable importances.

$$RMSE = \frac{\|ST/\max(ST) - V_N/\max(V_N)\|}{\sqrt{p}} \tag{4.2.15}$$

$$MBE = \frac{\sum (ST/\max(ST) - V_N/\max(V_N))}{p} \tag{4.2.16}$$

4.2.3. Software availability

The model cases are tested in the Python environment using the Exploratory Modeling Workbench (Kwakkel, 2017). This library provides an interface for sensitivity analysis using the *scikit-learn* implementation of the random forests and Extra-Trees algorithms (Pedregosa et al., 2011), as well as the Sobol and Morris techniques through the *SALib* library (Herman and Usher, 2017). These libraries are available through the *pip* package manager for Python. Alternative implementations of the tree-based methods can be found in the R environment, with the *party* and *extraTrees* packages (Hothorn et al., 2017; Simm and de Abril, 2015).

4.3. Model cases

This section will present model case studies in increasing order of complexity, using the benchmark Ishigami-Homma function (Ishigami and Homma, 1990), an exploratory model of the A(H1N1)v swine flu epidemic (Pruyt and Hamarat, 2010), and the CDICE simulation version of the DICE-2007 integrated assessment model (Butler et al., 2014; Nordhaus, 2007). Each case will first present reference sensitivity analysis results with the Sobol and Morris techniques. These results will then be compared with the MDI and pairwise MDA variable importances, as estimated from the random forests and Extra-Trees ensemble techniques.

4.3.1. Ishigami test function

The first test case is the Ishigami-Homma function (Ishigami and Homma, 1990), which is a common test case for sensitivity analysis due to its analytical tractability and non-additive properties:

$$Y = sin(x_1) + \alpha sin(x_2)^2 + \beta x_3^4 sin(x_1) \tag{4.3.1}$$

where x_1, x_2, x_3 are uniformly distributed in $[-\pi, \pi]$, with $\alpha = 7$ and $\beta = 0.1$. Using a Latin Hypercube sample with $N = 1500$ yields the output distribution shown in Figure 4.2.

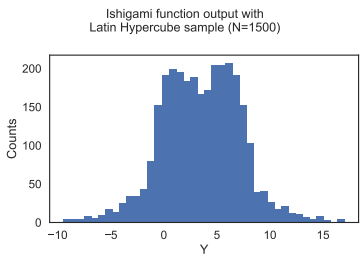


Figure 4.2: Output distribution for Ishigami function.

Figure 4.3 presents the convergence of the Sobol (top panel, left) and elementary effects (middle panel, left) indices as a function of the total number of input samples, and the relationships between the key indices provided by each technique (right panels). The Sobol sample size of $N = 15000$ is chosen to achieve a convergence criterion of $\kappa_N < 0.01$ (using intervals of $\Delta N = 400$ samples and $t = 4$ intervals); the shaded envelopes present 95% confidence bounds for the indices. The relationship between the first-order and total Sobol indices indicates higher-order interactions for x_1 and x_3, as expected from the structure of the function, while S and ST are identical for x_2.

Using $k = 8$ levels, with $\Delta = k/[2(k-1)]$ as recommended by Campolongo et al. (2007), the variable ranking obtained from the $\mu*$ elementary effects converges at a relatively small number of samples. However, the ranking does not match the order of the ST indices, underestimating the relative importance of x_1. This is illustrated in the bottom panel of the figure by plotting the proportional values of the $\mu*$ and ST indices against each other; the values of each group of indices are scaled relatively to the maximum value in each group, for $N = 5000$ and $N = 15000$ respectively. x_1 and x_3 show relatively higher values of the σ index, compared to their values for the $\mu*$ index. This could potentially be related to their interaction effects (which, in this case, can be inferred from the structure of the model), but the contribution of interactions towards the value of σ cannot be distinguished from the contribution of non-linearities (Saltelli and Annoni, 2010). This is indeed highlighted by x_2, which has approximately the same value on σ as x_3; although it has a non-linear impact, it does not interact with other variables in the model structure.

Figure 4.4 shows the convergence of the mean MDI importance indices for the random forests (top panel, left) and Extra-Trees (bottom panel, left) techniques over a Latin Hypercube sample, using 50 bootstrap resamples to estimate confidence bounds (shown by shaded envelopes which contain the full range of estimated values). Appendix A presents detailed convergence results, indicating that both algorithms stabilize below $\kappa_N < 0.02$ around $N = 3000$ samples, similarly to the Morris indices. Both algorithms are parameterized with $T = 100$ trees, $m \approx p/3 = 1$ (so that the trees are totally randomized), and a stopping criterion of $N_{leaf} = 2$. The right panels compare the mean estimated MDI importances (scaled relative to the highest MDI value at $N = 5000$), against the scaled reference ST indices.

For both techniques, Appendix A shows the root mean square error (RMSE) and mean bias error (MBE) estimated over all scaled MDI values, compared to scaled ST values (where positive bias is linked with an underestimation of relative variable importances compared to ST; eq. 4.2.16). Compared to the Morris $\mu*$ results, both ensemble techniques rank the input variables consistently with ST values; compared to random forests, Extra-Trees show quicker convergence, and a lower error compared to the relative ST values.

A potential drawback of the ensemble techniques is the requirement of choosing

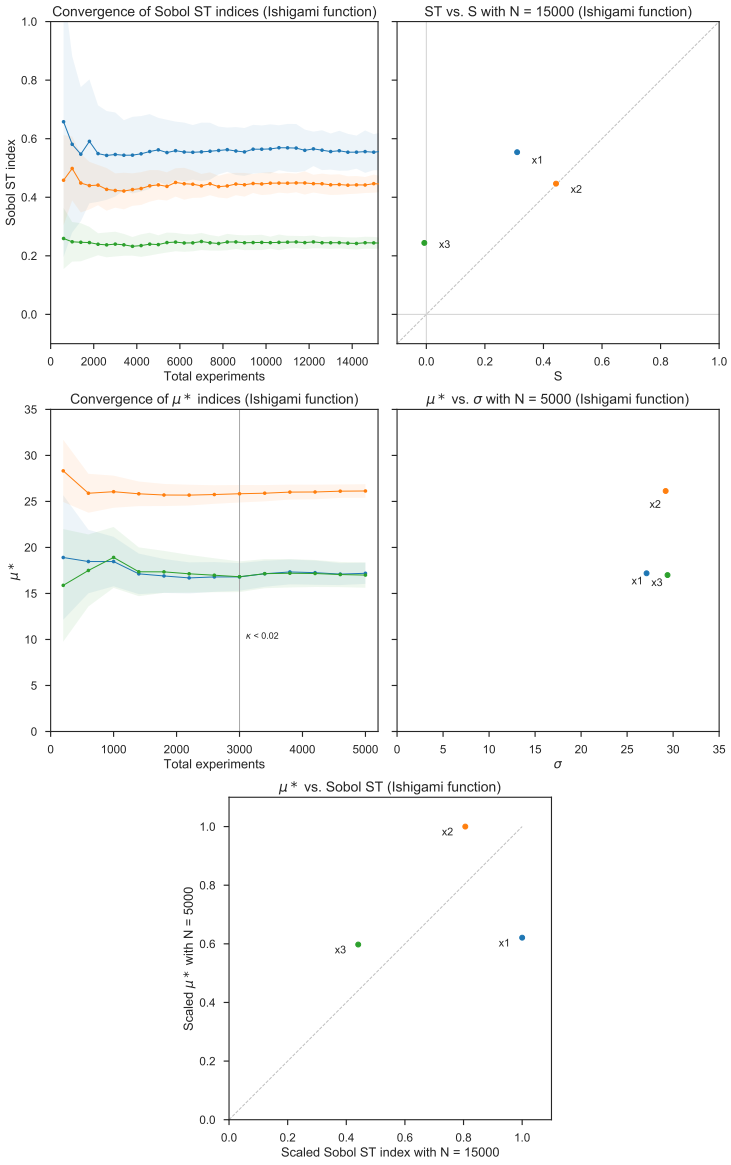


Figure 4.3: Results of Sobol (top panel) and elementary effects (middle panel) methods for the Ishigami test function. The vertical line indicates the $\kappa < 0.02$ convergence criterion for the $\mu*$ indices.

suitable tuning parameters. Focusing on the Extra-Trees technique due to its favorable performance, Figure 4.5 shows the RMSE (relative to scaled ST values) for scaled estimated importances, bootstrapped confidence interval on RMSE across 50 resam-

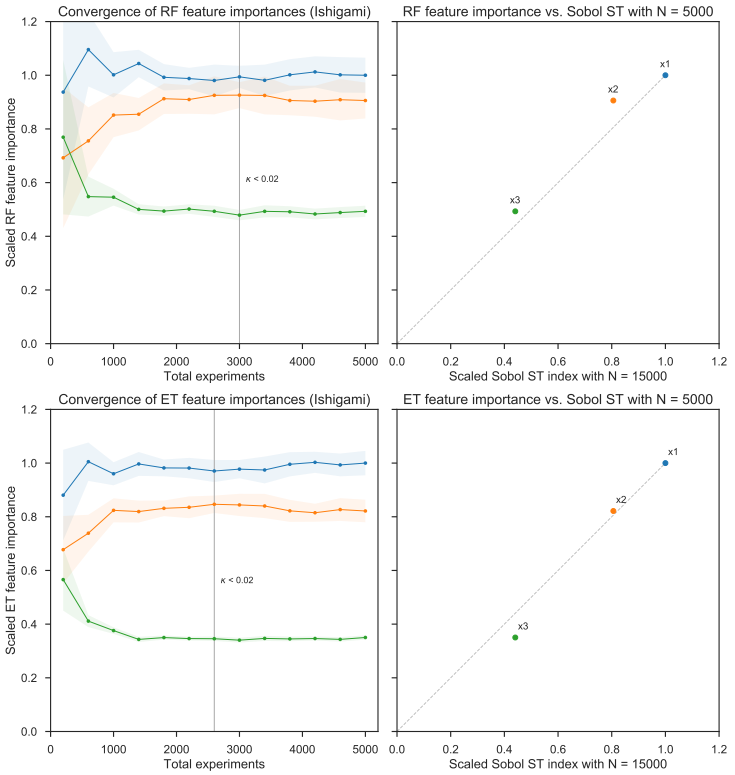


Figure 4.4: Estimation of MDI variable importances with the random forests (top panel) and Extra-Trees (bottom panel) techniques for the Ishigami test function. Vertical lines indicate the $\kappa < 0.02$ convergence criterion.

ples, and MBE. These metrics are presented across a range of values for the number of trees T, the number of splitting features m (subplot rows) and the minimum number of samples per node N_{leaf} (subplot columns).

RMSE appears robust to the number of trees T. The combination of m and N_{leaf} has a significant influence on RMSE; the starting point of $m \approx p/3$ suggested by Hastie et al. (2009) provides good results on RMSE, when combined with a small value for N_{leaf} (which controls the depth of the trees). N_{leaf} has a significant influence on MBE at a given value of m, which is particularly relevant for sensitivity analysis: a positive value indicates that relative variable importances are underestimated compared to ST, which could lead to a type II error in a screening setting (i.e. discarding potentially influential variables). Smaller trees appear more vulnerable to this error, which emphasizes higher-ranked variables (x_1 and x_2). This can be compensated by increasing m to decrease the randomness of the trees; however, at smaller values of N_{leaf} (e.g. $m = 3$ and $N_{leaf} = 1$), this increases RMSE.

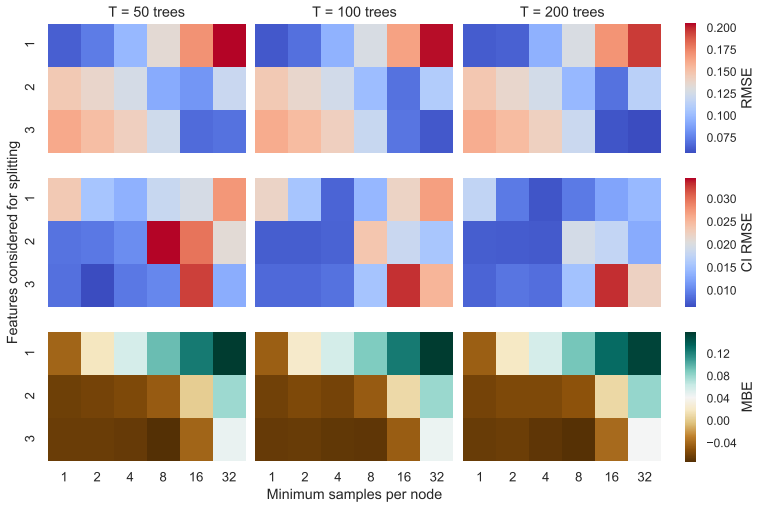


Figure 4.5: Extra-Trees performance relative to ST across key tuning parameters, for Ishigami function (N=3000). The figure shows RMSE (top three panels), bootstrapped confidence interval on RMSE (middle), and MBE (bottom panels) across a range of values for the number of trees T, the number of features considered for splitting m (subplot rows), and the minimum number of samples per node N_{leaf} (subplot columns).

As indicated by the relative values of ST and S for x_1 and x_3, the interaction between these variables contributes significantly to the output behavior. The left panel of Figure 4.6 shows the pairwise interaction importances estimated by the second-order Sobol S2 indices; the right panel presents MDA interaction importances estimated with Extra-Trees (averaged over 50 bootstrap resamples, on a 1500 sample set). The analytical relationship between x_1 and x_3 is therefore identified by both techniques, with other pairwise importances being negligible.

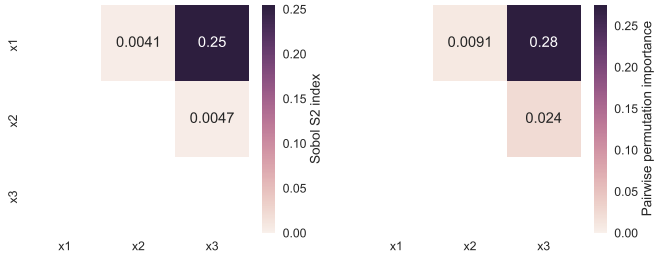


Figure 4.6: Comparison of pairwise variable interactions in the Ishigami function, using Sobol S2 indices (left) and Extra-Trees MDA pairwise permutation importances (right).

4.3.2. H1N1 swine flu epidemic model

Pruyt and Hamarat (2010) present a simple exploratory system dynamics model of the 2009 swine flu epidemic, based on a two-region susceptible-infected-recovered (SIR) population structure. This provides a more complex test case for sensitivity analysis due to the larger number of input variables (with 17 continuous inputs and two structural switches), and a broad output distribution. Table 4.1 shows the input variables and their bounds, assuming uniform distributions for all continuous variables. Figure 4.7 presents the resulting output distribution on the outcome of interest (defined as the number of fatalities in region 1 of the model) with a Latin Hypercube sample.

Table 4.1: Input variables for H1N1 flu model

Name	ID	Min.	Max.
Structural switch on immunity	immunity_switch	{0,1}	
Structural switch on contact rate lookup function	lookup_switch	{0,1,2,3}	
Additional seasonal immune population fraction - region 1	x11	0.1	0.5
Additional seasonal immune population fraction - region 2	x12	0.1	0.5
Fatality rate - region 1	x21	0.01	0.1
Fatality rate - region 2	x22	0.01	0.1
Initial immune fraction of the population - region 1	x31	0.1	0.5
Initial immune fraction of the population - region 2	x32	0.1	0.5
Normal interregional contact rate	x41	0.1	0.9
Permanent immune population fraction - region 1	x51	0.1	0.5
Permanent immune population fraction - region 2	x52	0.1	0.5
Recovery time - region 1	x61	0.1	0.8
Recovery time - region 2	x62	0.1	0.8
Root contact rate - region 1	x81	1	10
Root contact rate - region 2	x82	1	10
Infection rate - region 1	x91	0.01	0.1
Infection rate - region 2	x92	0.01	0.1
Normal contact rate - region 1	x101	10	100
Normal contact rate - region 2	x102	10	70

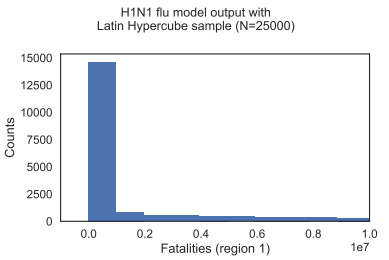


Figure 4.7: Output distribution for H1N1 model.

For this example, the Sobol technique requires $N > 150,000$ for a stable estimation of variable rankings, as shown in the top panel of Figure 4.8. A reference value of $N = 800,000$ was chosen by setting the convergence criterion to $\kappa_N < 0.01$ (using

intervals of $\Delta N = 40,000$ samples and $t = 4$ intervals). The relationship between ST and S indicates that higher-order interactions are present for most of the variables, with a group of 7 variables contributing significantly to output behavior.

The middle panel shows Morris results with $k = 8$ levels and $\Delta = k/[2(k-1)]$. While the same group of 7 variables is identified by the $\mu*$ indices, they require a relatively large sample size for a stable estimation. Appendix A presents a convergence analysis with $\Delta N = 10,000$ samples and $t = 4$ intervals, which requires approximately 190,000 samples for a stable convergence at $\kappa_N < 0.02$. Although both of the structural "switch" uncertainties are ranked consistently with ST values by the $\mu*$ indices, their relative estimated importance is less stable than the continuous uncertainties across sample sizes.

Figure 4.9 shows the convergence of the MDI variable importances for the random forests (top panel, left) and Extra-Trees (bottom panel, left) techniques over a Latin Hypercube sample, using 30 bootstrap resamples to estimate confidence bounds. Both algorithms are parameterized with $T = 100$ trees, $m \approx p/3 = 6$, and a stopping criterion of $N_{leaf} = 6$. The right panels compare the mean estimated importances (scaled relative to the highest MDI at $N = 150,000$) with the scaled reference ST indices.

As with the Ishigami function, Extra-Trees are more accurate than random forests and the Morris $\mu*$ indices for approximating ST; under the parameterization used, random forests present a higher error relative to ST than the Morris $\mu*$ indices. ET and random forests converge more quickly than the $\mu*$ indices (in particular for the "switch" uncertainties), with a largely stable variable ranking for $N > 10,000$, and a convergence criterion $\kappa_N < 0.02$ above 80,000 samples.

Figure 4.10 shows the influence of the tuning parameters on RMSE and MBE, compared to the reference scaled ST values. In this application, totally randomized ($m = 1$), fully grown ($N_{leaf} = 1$) trees perform significantly worse. The assumption of $m \approx p/3 = 6$ provides consistent performance, and the mean bias can be tuned by adjusting the value of N_{leaf} in a range of approximately 1 to 16 without introducing a larger error. N_{leaf} has a similar effect as in the Ishigami-Homma test case, with relatively smaller trees having a smaller negative bias.

As evidenced by the large difference between the ST and S indices, higher-order interactions are influential for output behavior. The left panel of Figure 4.11 shows second-order interaction importances as estimated by the Sobol S2 indices, for the same sample size of $N = 800,000$. The right panel presents the mean pairwise MDA interaction importances estimated with Extra-Trees (with 30 bootstrap resamples on a 50,000 sample set). These estimated importances for each interacting pair are plotted in the bottom panel, after scaling relatively to the highest value in each set (S2 and MDA).

The interpretation of these results should take into account the numerical sensitivity of the reference S2 results. As shown on the left panel, each of the second-order

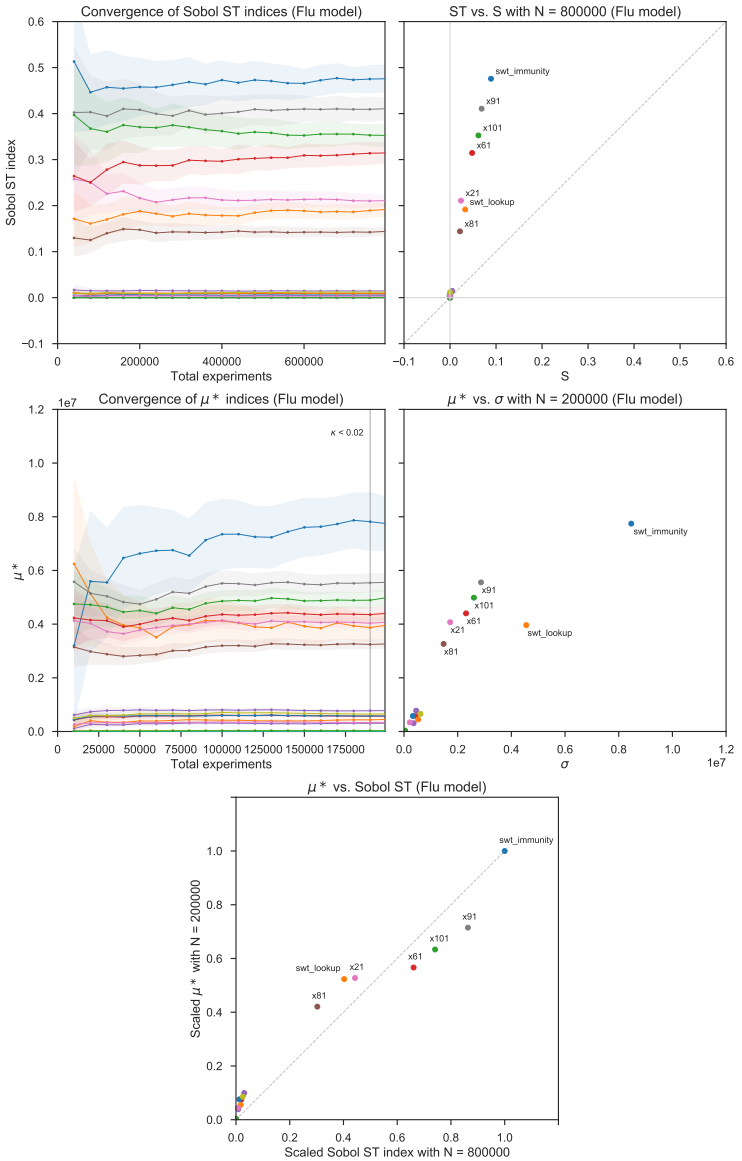


Figure 4.8: Results of Sobol (top panel) and elementary effects (middle panel) methods for H1N1 flu model. The vertical line indicates the $\kappa < 0.02$ convergence criterion for the $\mu*$ indices.

interaction terms only contributes a small portion of variance, which is typically smaller than the 95% confidence interval provided by SALib. This remains the case at significantly larger sample sizes ($N > 1e6$). The bottom panel illustrates this result with

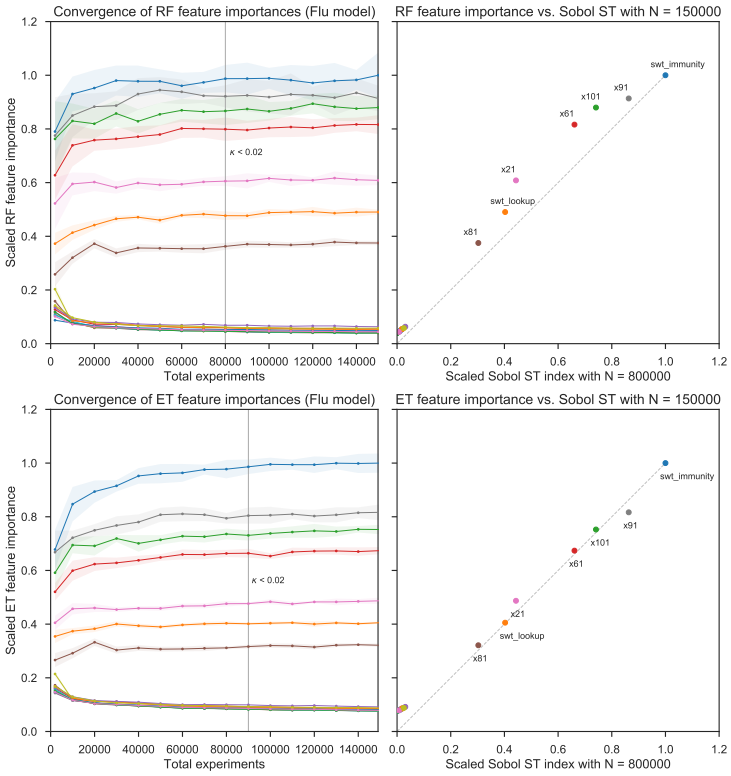


Figure 4.9: Estimation of MDI variable importances with the random forests (top panel) and Extra-Trees (bottom panel) techniques for H1N1 flu model. Vertical lines indicate the $\kappa < 0.02$ convergence criterion.

light gray markers for values of the S2 indices which are smaller than the estimated confidence interval, and are therefore likely to be unreliable. Nonetheless, the pairwise permutation generally performs well for identifying more significant interactions, for which the S2 index is outside the confidence interval (e.g. between the infection rate *x91* and other parameters to which it is structurally related in the model, such as the normal contact rate *x101* and the structural switch on immunity).

4.3.3. CDICE integrated assessment model

The last case study uses the CDICE model (Butler et al., 2014), which replicates the outcomes of the globally-aggregated DICE-2007 integrated assessment model (Nordhaus, 2007) under given policy scenarios. This model represents a simplified global economy, coupled with a 3-reservoir carbon cycle model and a 2-reservoir climate model; the feedbacks between these components lead to highly non-linear outputs. When used in an optimization setting, DICE yields an optimal policy for the time series of GHG

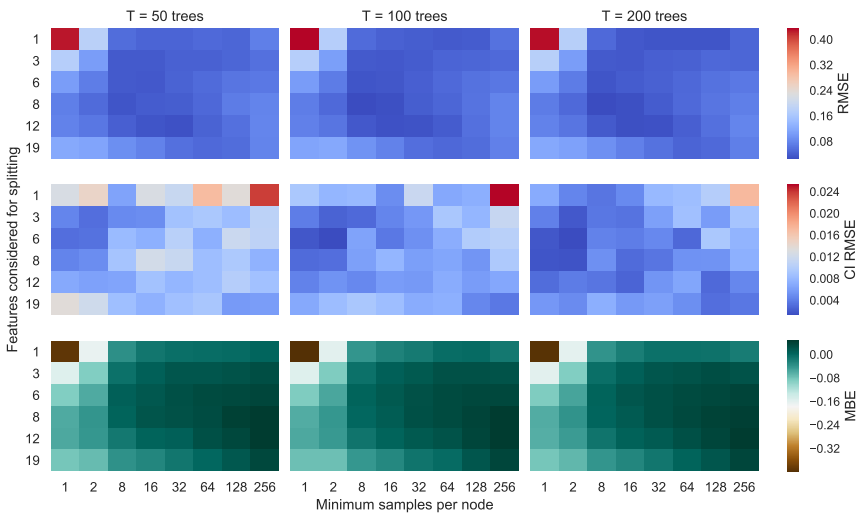


Figure 4.10: Extra-Trees performance relative to ST across key tuning parameters, for H1N1 flu model (N=50,000). The figure shows RMSE (top three panels), bootstrapped confidence interval on RMSE (middle), and MBE (bottom panels) across a range of values for the number of trees T, the number of features considered for splitting m (subplot rows), and the minimum number of samples per node N_{leaf} (subplot columns).

emission control rates and investments that maximize the discounted utility of consumption over the modelled time frame. Conversely, the CDICE simulation version introduced by Butler et al. (2014) can be used to evaluate the impact of exogenous uncertainties on the performance of policy scenarios.

The full version of this model uses 31 exogenous input variables, shown with their input ranges in Table 4.2. With uniform input distributions, these assumptions yield the output distribution shown in Figure 4.12 for the net present value (NPV) of abatement costs. This outcome will be used for the analysis due to its relatively quicker convergence with Sobol measures.

Figure 4.13 shows the convergence of ST (top panel) and Morris $\mu*$ (middle panel) indices, for the NPV of abatement costs. Due to the large number of parameters, the Sobol indices require $N > 9e6$ for a stable ranking. As shown by the low values of the first-order S indices relative to ST, higher-order interactions are significant for the behavior of this outcome.

The Morris results use a sampling of $k = 10$ levels and $\Delta = k/[2(k-1)]$, yielding a mostly stable estimation of variable rankings above $N > 150,000$. However, the bottom panel of Figure 4.13 shows several inconsistencies in the variable rankings given by $\mu*$ compared to scaled ST values.

Figure 4.14 shows the convergence of the MDI variable importances for the ran-

Table 4.2: Input variables for CDICE model (baseline scenario)

ID	Min.	Max.
popasym	5000	13000
gpop0	0.2	0.35
ga0	0.092	0.2
dela	0.001	0.016
sig0	0.13364	0.15273
gsigma	-0.16	-0.07
dsig	0.001	0.003
dsig2	0	0.0002
eland0	9	15
dtree	0.05	0.2
b12	0.155288	0.223288
b23	0.025	0.1
fex0	-0.3	0
fex1	-0.2	0.5
t2xco2	1	8
fco22x	3.6	3.9
c1	0.2	0.24
c3	0.27	0.33
c4	0.045	0.055
a1	0	0.001
a2	0.002255	0.003123
a3	1.5	3
pback0	0.6	3
theta2	2.6	3
backrat	1.5	2.5
gback	0.045	0.055
partfrac1	0.1	1
partfrac2	0.25372	1
partfracn	0.5	1
dpartfrac	0	0.25
saverate0	0.2	0.24

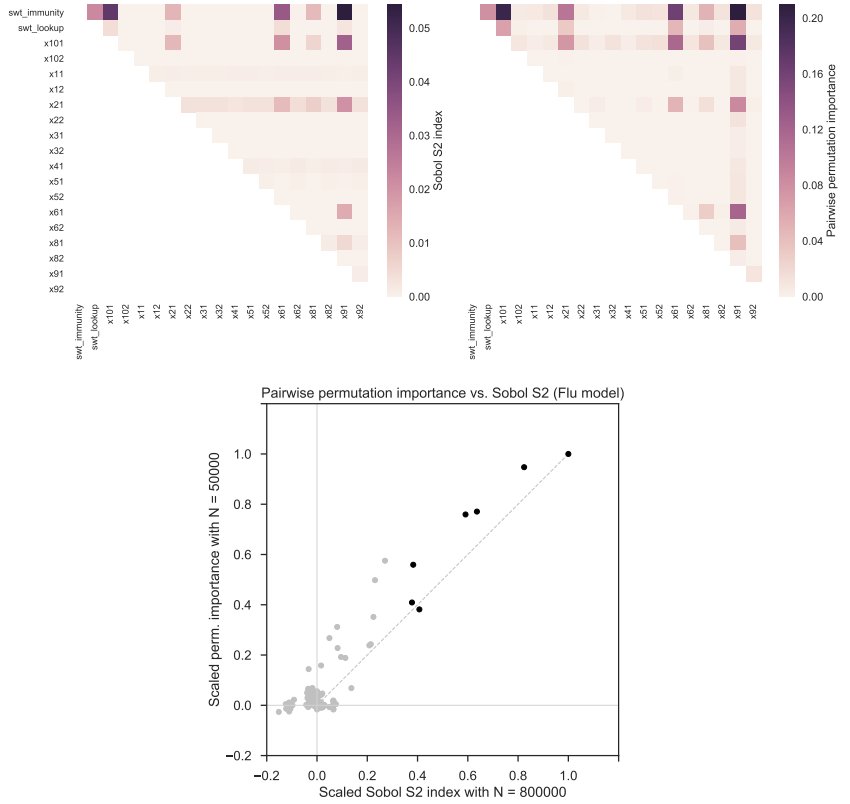


Figure 4.11: Comparison of pairwise variable interactions in the H1N1 flu model, using Sobol S2 indices (left) and Extra-Trees MDA pairwise permutation importances (right). The bottom panel plots scaled Sobol S2 and Extra-Trees interaction importances against each other, with light gray markers corresponding to S2 values which are within the confidence bounds estimated by SALib.

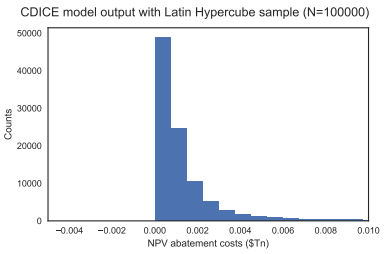


Figure 4.12: Output distribution for CDICE model (NPV of abatement costs).

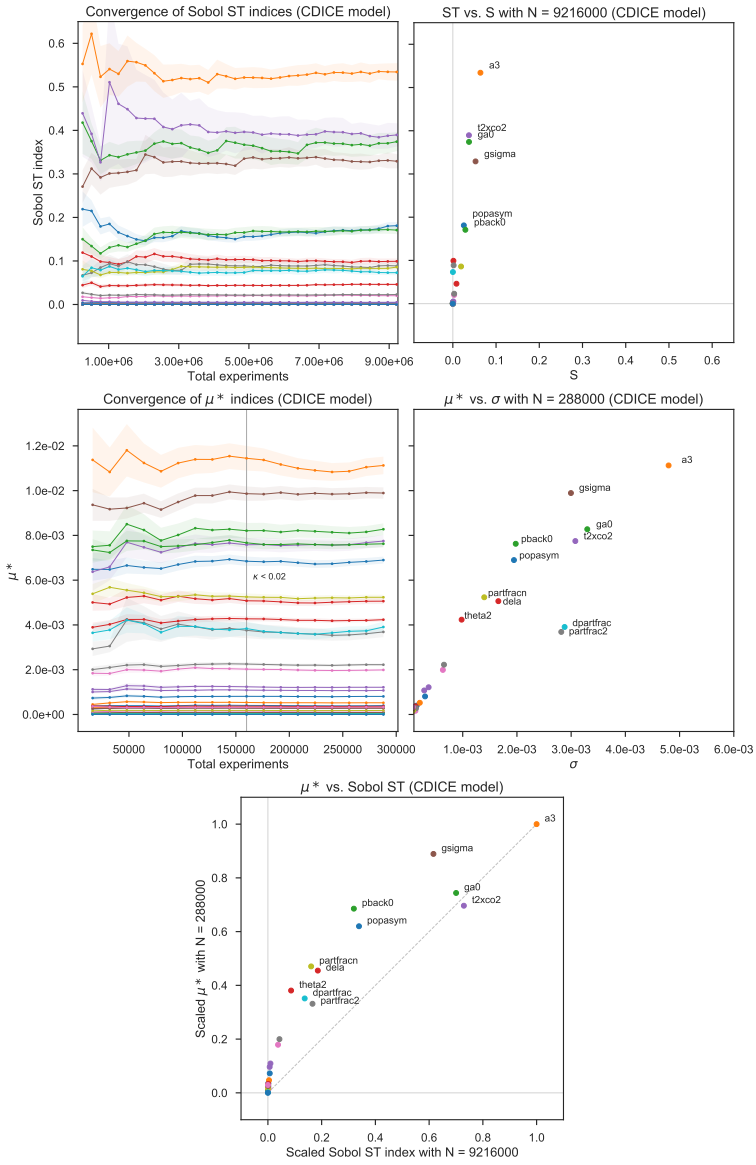


Figure 4.13: Results of Sobol (top panel) and elementary effects (middle panel) methods for the CDICE model (NPV of abatement costs). The vertical line indicates the $\kappa < 0.02$ convergence criterion for the $\mu*$ indices.

dom forests (top panel, left) and Extra-Trees (bottom panel, left) techniques over 30 resamples on a Latin Hypercube sample, for the same outcome. The algorithms are parameterized with $T = 100$ trees, $m \approx p/3 = 10$, and $N_{leaf} = 8$. The right panels

compare the scaled mean estimated importances with the scaled reference ST indices.

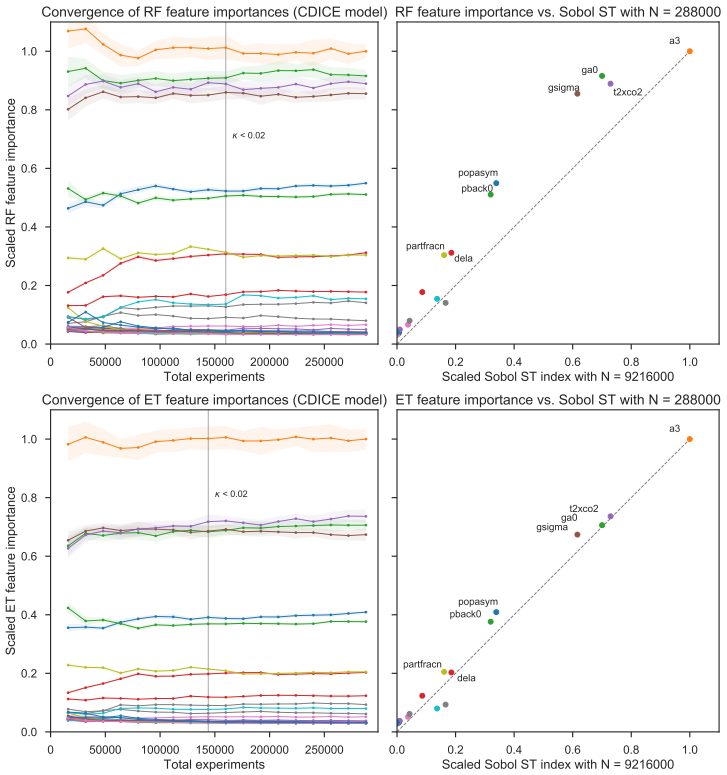


Figure 4.14: Estimation of MDI variable importances with random forests (top panel) and Extra-Trees (bottom panel) for the CDICE model (NPV of abatement costs). Vertical lines indicate the $\kappa < 0.02$ convergence criterion.

The Extra-Trees variable rankings mostly stabilize for $N > 100,000$, with a better approximation of relative ST values than the Morris $\mu*$ indices. For random forests, however, the variable ranking shows some discrepancies with the ST results. Appendix A presents a convergence analysis with $\Delta N = 16,000$ samples and $t = 4$ intervals; both ensemble methods, as well as Morris indices, reach a criterion of $\kappa_N < 0.02$ with approximately 150,000 samples.

Figure 4.15 shows the performance of the Extra-Trees estimation across the tuning parameters, compared to the scaled relative ST values. As with the H1N1 flu model, highly randomized and fully developed trees do not perform adequately, but the estimated importances are robust in a range of m/p of 0.3 to 0.6 ($m = 9$ to $m = 18$). The MBE metric also presents a comparable pattern to the H1N1 model results, with larger values of m/p leading to a negative bias unless compensated by a larger stopping criterion N_{leaf}.

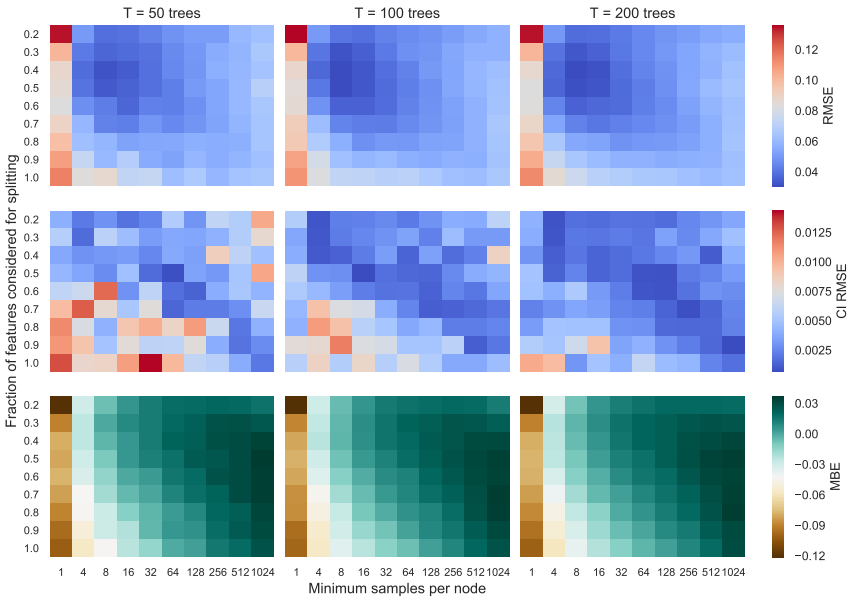


Figure 4.15: Extra-Trees performance relative to ST across key tuning parameters, for CDICE model (NPV of abatement costs, N=100,000). The figure shows RMSE (top three panels), bootstrapped confidence interval on RMSE (middle), and MBE (bottom panels) across a range of values for the number of trees T, fraction of features considered for splitting m/p (subplot rows), and minimum number of samples per node N_{leaf} (subplot columns).

The left panel of Figure 4.16 shows second-order interaction importances estimated by the Sobol S2 indices, with the same sample of $N = 9.22e6$. The right panel presents the mean pairwise MDA interaction importances estimated with Extra-Trees (with 30 bootstrap resamples, on a 100,000 sample set).

The scaled estimated importances for each interacting pair are plotted against each other in the bottom panel. The most significant pairwise interactions appear to be identified by the permutation measure, such as the interactions involving the *a3* exponent of the model's climate damage function. As with the H1N1 flu model, however, the S2 indices may be numerically unreliable due to relatively large confidence bounds. It can be noted that some of the S2 indices present negative values, which is clearly a numerical artifact. The analysis was in this case limited by the computational costs of the larger input samples which would be required for a stable estimation of S2 indices.

4.4. Discussion and conclusions

This chapter assessed the performance of decision tree-based ensemble methods for the estimation of global sensitivity analysis measures, focusing on the random forests

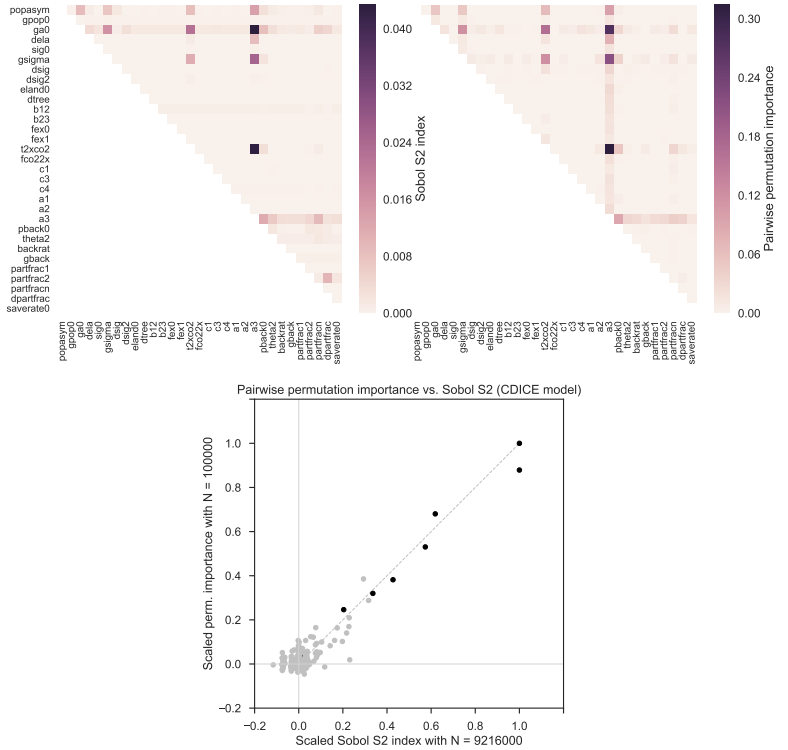


Figure 4.16: Comparison of pairwise variable interactions in the CDICE model (NPV of abatement costs), using Sobol S2 indices (left) and Extra-Trees pairwise permutation importances (right). The bottom panel plots scaled Sobol S2 and Extra-Trees interaction importances against each other, with light gray markers corresponding to S2 values which are within the confidence bounds estimated by SALib.

and Extra-Trees algorithms. Compared to the Morris elementary effects method which is commonly used for screening non-influential variables, the Extra-Trees technique in particular performed well to estimate relative Sobol ST total effect indices, using the Mean Decrease Impurity (MDI) metric for variable importance. Across the three case studies presented in the work, Extra-Trees therefore outperformed the Morris $\mu*$ indices on measures of RMSE and variable ranking error relative to the proportional values of ST indices. For the more complex H1N1 and CDICE cases, a sample size of less than 10% of the Sobol sample size was sufficient for a stable estimation of variable rankings. Furthermore, a pairwise Mean Decrease Accuracy (MDA) permutation metric allowed for the study of variable interactions with Extra-Trees. While the more common random forests algorithm performed well on the benchmark Ishigami-Homma test function, it was less reliable in the more complex cases.

The Extra-Trees estimation of variable importances was systematically assessed across a range of tuning parameters for the algorithm. Based on the case studies and previous literature (Hastie et al., 2009), values of $T = 100$ trees and a number of splitting features $m \approx p/3$ appear to be suitable starting points. The choice of a stopping criterion significantly affects bias, which is especially relevant for a screening application. In order to avoid possible type II errors, a conservative guideline would be to use fully developed trees ($N_{leaf} = 1$) for $N \approx 1000$, then to introduce a stopping criterion $N_{leaf} \propto \sqrt{N}$ for larger samples. Values of 6 and 8 for N_{leaf} thus performed well for $N = 50,000$ and $N = 100,000$ with the H1N1 and CDICE models.

The variable importance metrics provided by the tree-based methods can be assessed in relation to the criteria summarized by Pianosi and Wagener (2015) for an "ideal" sensitivity metric. As such, the MDI and MDA metrics largely meet these criteria, by being suitable for global sampling designs, independent of model structure, relatively easy to implement numerically, and stable across sample sizes and bootstrap resamples. Compared to Sobol indices, a downside of these metrics is the lack of a straightforward mathematical interpretation, as they only provide information about the relative importance of inputs, rather than their direct effect on output variance. However, for practical purposes, the accurate estimation of relative total effects should be sufficient for a factor fixing application. Compared to the $\mu*$ indices for elementary effects (which share this limitation on mathematical interpretability), MDI more accurately estimates the relative values of ST indices, is suitable for non-scalar inputs, and appears more stable at smaller sample sizes. MDA additionally estimates relative pairwise interaction effects, which are not identified by the elementary effects σ indices. MDI and MDA can also be computed from generic Latin Hypercube or Monte Carlo sampling designs. This makes it easier to reuse existing datasets which may have been generated from an uncertainty analysis, or to combine the ensemble methods with other analysis techniques in a multi-method analysis framework.

In parallel, Appendix A compares the total runtime required to compute importance metrics (as well as the total model evaluation runtime), for the more complex H1N1 and CDICE cases. With the software libraries used in this work, the MDI and Morris indices have a similar computation runtime at a given sample size, with the computation runtime largely scaling in proportion to sample size N. The pairwise MDA metric is slightly costlier and scales with the square of the number of input variables p. In the presented cases, the analysis runtime for these metrics was typically small relative to the total evaluation runtime required by the simulation models; it was also significantly smaller than the computation time for Sobol indices. In an analysis setting focused on estimating the relative importance of variables and their interactions, the smaller sample size required by the MDI and MDA metrics can therefore significantly reduce the overall computational cost of the analysis.

In a broader perspective, however, it should be noted that the techniques evaluated in this chapter followed a variance-based approach to global sensitivity analysis, by directly calculating indices with the Sobol method or by approximating the proportional importance of the latter with tree-based methods. As described by Pianosi and Wagener (2015), variance may not be an appropriate measure of uncertainty for multi-modal or highly skewed output distributions; in these cases, an approach based on the probability density function of the output may be preferable. This property was demonstrated by the authors with a simple non-linear model, for which variance-based GSA did not properly distinguish variable importances. This has clear implications for the cases studied in this work, as the outputs of the H1N1 and CDICE models showed highly skewed distributions under the uncertainty ranges used to generate input samples. Given the possible limitations of variance-based methods under such conditions, it would be useful to compare the reference Sobol results with a density-based method, and to evaluate the performance of Extra-Trees across a wider range of output distribution shapes.

5

Trade-offs and endogenous dynamics for the planning of Aquifer Thermal Energy Storage systems

This chapter is based on Jaxa-Rozen, Kwakkel, and Bloemendal (2017b), with revisions to the problem definition to avoid overlap with previous chapters of the thesis.

5.1. Introduction

Improving the energy performance of the building sector is a key step towards national and international objectives for reductions in greenhouse gas (GHG) emissions. Geothermal energy has become an increasingly accessible option for displacing the use of fossil fuels in the built environment; in particular, "shallow" systems can reduce energy demand for space heating and cooling, by seasonally storing energy in the subsurface. Aquifer Thermal Energy Storage (ATES) has emerged as a leading form of shallow geothermal energy storage, which can reduce overall energy use by more than half in large commercial or institutional buildings (Vanhoudt et al., 2011). The Netherlands are currently the main market for ATES technology, due to a combination of suitable climactic and hydrological conditions as well as increasingly strict energy regulations for buildings; recent research suggests that approximately half of the world's urban population lives in areas which would be potentially appropriate for ATES (Bloemendal et al., 2015). Increasing demand for energy-efficient technologies is therefore likely to lead to a significant growth in the market for ATES, which will need to be matched by appropriate planning policies to safeguard the long-term sustainability of aquifers as a common-pool resource for thermal storage.

Due to the potential environmental impacts of subsurface energy storage, the precautionary principle appears to be a useful guide for the governance of ATES technology (Haehnlein et al., 2010). In the Netherlands, layout guidelines are for example intended to prevent thermal interactions between adjacent ATES systems, which could potentially lead to degraded performance. However, recent studies indicate that this approach may be overly conservative, particularly from the point of view of collective urban energy savings – which could be increased by relaxing design guidelines to allow for a greater number of ATES wells in a given area (Sommer et al., 2015). From this perspective, public authorities may need to adapt their approach to ATES management in order to fully benefit from the energy-saving potential of the technology, without compromising the long-term potential of the subsurface for thermal storage or other functions. The interplay between the adoption of ATES technology and the state of aquifer resources leads to complex, time-dependent dynamics: subsurface conditions affect the performance of ATES systems, which may lead to different pathways for the adoption and use of ATES – in turn influencing temperature distributions in the subsurface. This forms a classic example of a social-ecological system (SES), in which overall system outcomes emerge from the interactions between relatively distinct social and environmental subsystems (Ostrom, 2009). However, this complexity has typically not been fully acknowledged by current practices for ATES modelling and planning, which rely on geohydrological models and leave adoption and operation dynamics out of scope (Li, 2014). These elements are subject to both exogenous and endogenous uncertainties -– for instance, energy prices and the decision-making heuristics of ATES investors, respectively. Understanding the different plausible behaviors of the system therefore requires a modelling approach which considers the full complexity of the problem, as well as the uncertainties involved. This work thus applies the coupled simulation architecture introduced in Chapter 3 of this thesis, in order to study the plausible dynamics of an idealized ATES-aquifer system by linking an agent-based component with a finite-difference aquifer model. The agent-based model represents the investment behavior of ATES users and follows existing research on energy-efficient technology adoption. The geohydrological component provides a realistic model of the transient temperature distributions created by ATES well flows in the subsurface. This coupled architecture is applied across a range of parametric uncertainties and spatial planning policies to investigate the dynamics of ATES adoption and performance, with a focus on policy-relevant trade-offs between private and public outcomes. The decision tree-based sensitivity analysis method described in Chapter 4 of the thesis is used to better understand the key sensitivities driving the model outcomes.

Section 5.2 structures the problem based on the XLRM framework (Lempert et al., 2003), followed by results for different policies in section 5.3. Section 5.4 discusses the results in light of previous work, and section 5.5 concludes with policy implications.

5.2. Methods and problem definition

Table 5.1 synthesizes the problem definition following the XLRM framework (Lempert et al., 2003). The following subsections discuss the models and relationships (R) used to represent the problem, the performance measures (M) on which different ATES layout policies are assessed, the planning parameters which can be used as policy levers by authorities (L), and the external uncertainties (X) which affect the performance of the system.

Uncertainties	**Policy levers**
Socio-technical: • Electricity price • Gas price • Discount rate used by ATES investors • Investment cost of conventional and ATES systems • Energy demand Geohydrological: • Aquifer conditions (conductivities, porosity)	Distance between ATES wells, defined as a multiplier of average thermal radius
Models and relationships	**Measures**
Agent-based model of ATES adoption: • ATES adoption dynamics driven by economic performance within imposed spatial planning constraints, with 10 simulated ATES operator agents Geohydrological model: • 1000m x 1000m x 20m aquifer model; monthly time resolution Control model: • ATES well flows computed based on representative climate data and building energy demand	Main performance indicators: • Cumulative and annual cost savings realized by ATES systems, relative to equivalent conventional energy system • Cumulative and annual GHG savings relative to conventional energy • Equivalent GHG abatement cost Secondary indicators: • Average ATES thermal efficiency • Number of active ATES wells

Figure 5.1: XLRM summary

5.2.1. Models and relationships

The development of ATES systems is driven by interactions and feedbacks between the technical, environmental and economic performance of systems, their adoption by building owners, and the use of subsurface resources. These interactions are mediated by spatial planning policies and can plausibly lead to several archetypal dynamics for the adoption and performance of ATES systems over time – such as an S-shaped adoption curve in the number of active wells, as expected economic returns provide an incentive for building operators to use ATES wells; or a growth-and-collapse pattern in thermal efficiency, caused by interferences which may lead systems to become unprofitable. These basic relationships are summarized in the causal loop diagram in Figure 5.2:

These conceptual dynamics were translated into a simulation model using the coupled simulation architecture described in Chapter 3 of this thesis, which couples an agent-based model of ATES adoption and planning, a geohydrological aquifer model, and a control model for ATES operation. This simulation model is parameterized to

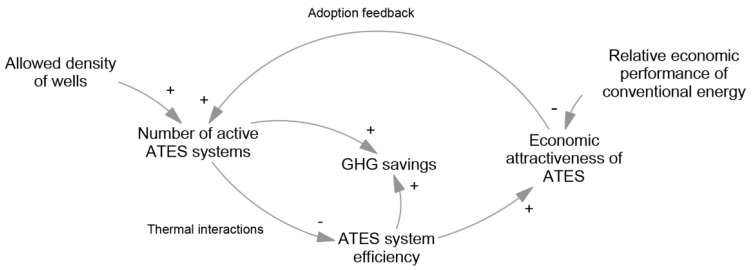


Figure 5.2: Idealized ATES adoption dynamics

represent a stylized case of ATES operation and adoption. Appendix B describes the model and software implementation in more detail, following the ODD+D protocol (Müller et al., 2013).

5

5.2.2. Measures

A particular challenge for ATES planning is the lack of a consistent framework which can be used to assess the performance and sustainable use of urban ATES systems, and therefore to define metrics and objectives for analysis. Several methods have been applied to assess the thermal and economic performance of single ATES systems (Rosen and Dincer, 2003; Sommer et al., 2015); however, an integrated approach to assess the long-term efficiency and sustainable use of an aquifer with ATES systems has not yet been implemented in practice. In the Netherlands, for instance, the assessment of the energy benefits of ATES development is often left out of scope of planning methods (Bloemendal et al., 2018). This is becoming an increasingly important issue for the public actors involved in ATES governance, such as municipalities and regional authorities. These actors have a twofold interest in maximizing the energy-saving potential of ATES technology, while preserving the long-term technical and economic viability of the technology. Based on the dynamics outlined previously, this analysis starts from two main aggregate performance objectives which are likely to present a policy-relevant trade-off:

- The average economic performance of ATES systems, which should maximize annualized energy cost savings (or minimize the payback period) relative to an equivalent conventional energy system;
- The contribution of ATES systems towards collective targets for GHG reductions, which should maximize the operational GHG emissions avoided in comparison to a conventional energy system.

This formulation assumes that GHG reductions are an adequate proxy for collective performance, and implies that chemical, hydrological or microbiological impacts on the

environment are left out of scope. In turn, this assumes an appropriate legal framework which e.g. protects drinking water extraction zones and regulates injection temperatures, in order to mitigate the environmental impacts reviewed by e.g. Bonte et al. (2011). The current WBBE framework used in the Netherlands appears to be suitable from this perspective (de Graaf et al., 2016); however, a broader problem formulation would for instance be needed to plan and assess higher-temperature ATES systems, or systems operating in contaminated aquifers.

These two performance indicators can be related by taking the ratio between annualized energy savings (or costs) and annual GHG reductions, and expressing this ratio as an equivalent GHG abatement revenue (or cost). Marginal abatement cost (MAC) curves are frequently used to compare the economic efficiency of different GHG mitigation options (e.g. McKinsey & Company, 2009); however, these indicators are in practice often hindered by the aggregations needed to represent technologies with a single average cost level, and by an incomplete treatment of uncertainty (Kesicki and Ekins, 2012). This is particularly relevant for decentralized energy technologies such as ATES, where costs and performance may be location-sensitive and time-dependent. Given the importance of individual heterogeneity and transient dynamics for the adoption and performance of ATES systems, the analysis will consider these outcomes for individual systems over time, in addition to the aggregate indicators. Other intervening outcomes are also recorded to track the number of active ATES wells over time as well as their thermal efficiency, which indicates the fraction of energy retrieved relative to the energy injected into the aquifer.

5.2.3. Policy levers

Under current approaches for the planning and governance of ATES systems, the primary policy levers available to public authorities relate to spatial planning parameters. These parameters are typically based on the minimal required distance between neighbouring wells, which is itself defined as a multiplier of the average thermal radius R_{th} (illustrated in Figure 5.3). This value corresponds to the expected radius of thermal influence, based on the analytical solution for heat transport in porous media. The prevailing Dutch guidelines for ATES system design require a distance of 3 R_{th} between neighbouring wells to prevent thermal interactions. As shown in Figure 5.3, this distance could theoretically be reduced to $\sqrt{2}$ R_{th} in idealized conditions.

As previously noted, recent research suggests that the current guidelines used in the Netherlands may be overly conservative. The analysis will thus explore a broader range of design parameters.

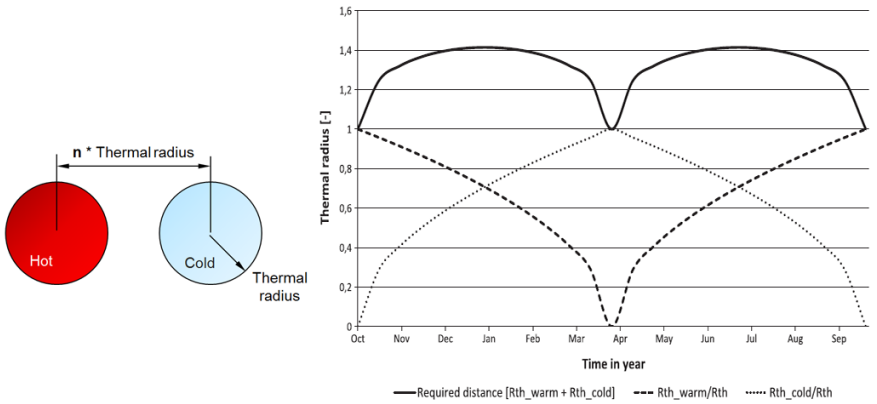


Figure 5.3: Spatial layout parameters

5.2.4. Uncertainties

Each of the model components presents a set of specific uncertainties, which are summarized in Table 5.4[1]. Uncertainty ranges for the socio-technical and geohydrological parameters were defined in consultation with research partners, and based on representative ATES/aquifer data for the Netherlands. Given the simple decision heuristics used in the agent-based model, this component is primarily driven by uncertain exogenous values for energy prices, which are assumed to remain constant over the time frame of the simulation.

	Min.	Max.	Unit
Socio-technical			
Electricity price	41	82	EUR/GJ
Gas price	11	21	EUR/GJ
Discount rate	0.04	0.15	-
Cost for conventional heating machine (Hmcost)	60	100	EUR/kW
Cost for conventional cooling machine (Cmcost)	130	220	EUR/kW
Cost factor for ATES wells	34000	50000	EUR/m^3/h
Critical payback period for adoption	Drawn from distribution		years
Geohydrological			
Aquifer porosity	0.2	0.4	-
Hydraulic conductivity (horiz.)	10	60	m/day

Figure 5.4: Uncertain parameters

The results shown in subsections 5.3.2, 5.3.3 and 5.3.4 are based on ensembles of 512 experiments which were sampled from these uncertainty ranges, using Latin Hypercube sampling with uniform distributions. Subsection 5.3.1 uses ensembles of 24

[1]Energy price ranges used in the original work are here presented in units of EUR/GJ for consistency with other chapters of the thesis.

replications with best-estimate values for the uncertain parameters, to illustrate the impact of the stochastic uncertainty caused by the distribution of critical payback periods across the simulated agents.

The agent-based model assumes that operators attempt to use all of their allowed storage capacity, so that the effective thermal radius approximates the theoretical expected value. Within each of the experiments sampled from Table 5.4 and used in subsections 5.3.2 and 5.3.3, each individual pair of simulated wells is in turn assigned a random annual storage capacity, sampled from a uniform distribution in a range of 40,000-100,000 m^3/yr. Based on this value, the calculation of ATES well flows uses a simple weather-dependent approach to compute storage flows over time, on the basis of the difference between outside temperature and representative setpoints for heating and cooling. In addition to the allowed storage volume, the computation assumes that warm and cold storage flows are required to be balanced over a moving 5-year period. More details are given in Bloemendal and Hartog (2018). To generate representative demand profiles, this study uses a realization of the KNMI W+ climate scenario over the 2010-2045 period, for the weather station of De Bilt in the Netherlands. Under this simplified approach, the weather profiles have a relatively limited impact on resulting well flows, due to the restrictions which are imposed on thermal balance and capacity usage; operational uncertainties will be explored further in Chapter 6 of the thesis, for instance by directly imposing thermal imbalances.

5.3. Results

Based on this setup, the model was simulated over a period of 15 years (or 180 monthly periods) using a 1000x1000x20m grid. As detailed in the ODD+D table, the model is parameterized to represent an idealized case which disregards features such as urban geographic constraints, or specific geohydrological features; the geohydrological model also does not incorporate ambient groundwater flow. The simulated timeframe was chosen to let adoption patterns and temperature distributions stabilize, while balancing runtime constraints. The following subsections present the simulation results with a focus on the impact of well distance parameters.

5.3.1. Visualization of model outcomes over time

In order to illustrate the basic dynamics observed, Figure 5.5 shows a selected subset of model outcomes over time, with the model being parameterized following best-estimate values for the uncertain parameters. The graphs present ensembles of 24 replications which only include the stochastic uncertainty related to the random distribution of critical payback periods. These ensembles were tested over four different well distance policies, from 2 R_{th} to 3.5 R_{th}. The shaded envelopes correspond to the minimum and maximum values observed at each point in time under each policy, with the panel on the right showing the Gaussian kernel density estimator for the final distribution of

outcomes within each envelope at the end of the simulation.

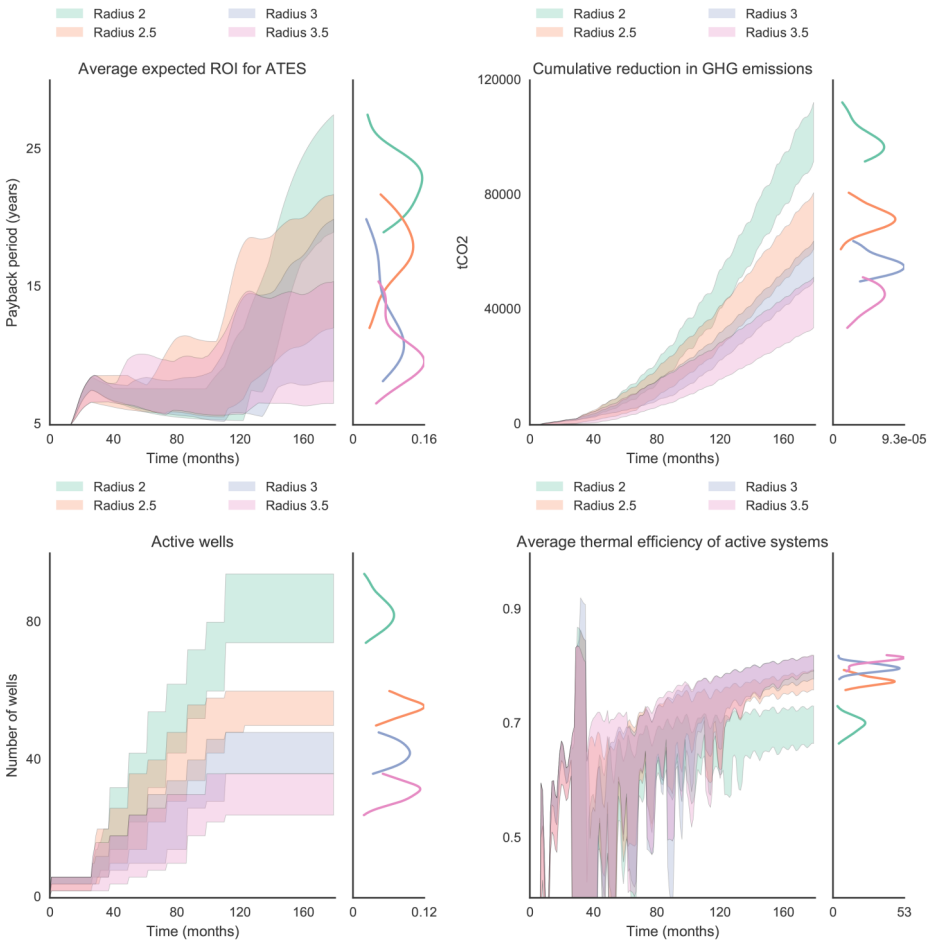


Figure 5.5: Basic model outcomes over time under stochastic uncertainty

The top two graphs of Figure 5.5 point towards a trade-off between public and private interests: although smaller distance policies increase the realized GHG savings by allowing for a greater number of wells to be built, they also penalize economic performance (here illustrated by the payback period), in comparison to policies with greater distances. This can be explained by considering the average thermal efficiency: due to negative thermal interactions between neighboring systems, which decrease the energy which can be effectively retrieved from wells, the average system efficiency stabilizes at a markedly lower value for the 2 R_{th} policy.

5.3.2. Impact of well distance policies under uncertainty

To extend the analysis under uncertainty, 512 experiments were sampled from the parametric ranges listed in Table 5.4, using a Latin Hypercube design. Each of these experiments was then tested across a set of 8 different policies for the minimal clearance between new ATES wells, covering the interval [1.75, 3.5] R_{th}. Figure 5.6 presents boxplots for these policies at the final time of the simulation. To represent the uncertainty introduced by the range of plausible discount rates, economic performance is given by the total annualized energy costs incurred by the ATES operator agents compared to a conventional energy system. Taking the ratio of this indicator with the corresponding annual GHG savings yields an equivalent abatement cost, which estimates the efficiency of each distance policy as a GHG mitigation option.

The basic trade-off between private and public interests which was suggested by Figure 5.5 also seems to appear under the full range of uncertainties, with smaller well distances increasing total GHG reductions at the price of greater (and more variable) costs to operators. A Kruskal-Wallis non-parametric test was applied to check for a difference in medians across policies (after rejecting the hypothesis that the outcomes are normally distributed by using a Shapiro-Wilk test), followed by Dunn's multiple comparisons test. Figure 5.7 shows the resulting p-values for pairwise comparisons between policies, for the outcomes shown in Figure 5.6 (where $p > 0.05$ indicates a non-significant difference).

In relation to total annual GHG savings and average thermal efficiency, well distance policies thus cause a significant difference in the median outcome values for all pairs. In the case of total annualized energy costs and average abatement costs, it is interesting to note that the pair comparisons for distances greater than 3 R_{th} are not significant at $\alpha = 0.05$, which implies that these distances would not be more economically efficient. The p-values for the pair comparison between 2.75 R_{th} and 3 R_{th} are also not significant for economic outcomes at this threshold, potentially indicating that well distances could be reduced from the current Dutch guideline of 3 R_{th} without significantly affecting the economic returns of ATES users – while benefiting from the GHG savings provided by a greater density of wells.

5.3.3. Visualization of model sensitivities

The socio-technical and geohydrological uncertainties lead to a broad ensemble of potential outcomes, particularly with tighter distance policies in which thermal interactions between wells are significant. To better understand the influence of these uncertainties, a Random Forest (RF) non-linear regressor (Breiman, 2001) was first used to assess the most influential parameters in relation to economic outcomes. As described in Chapter 4 of the thesis, RF generates a set of classification or regression trees from bootstrap samples of the data, and can be used to rank and quantify the importance of parameters towards a given output. In this application, this method is essentially intended

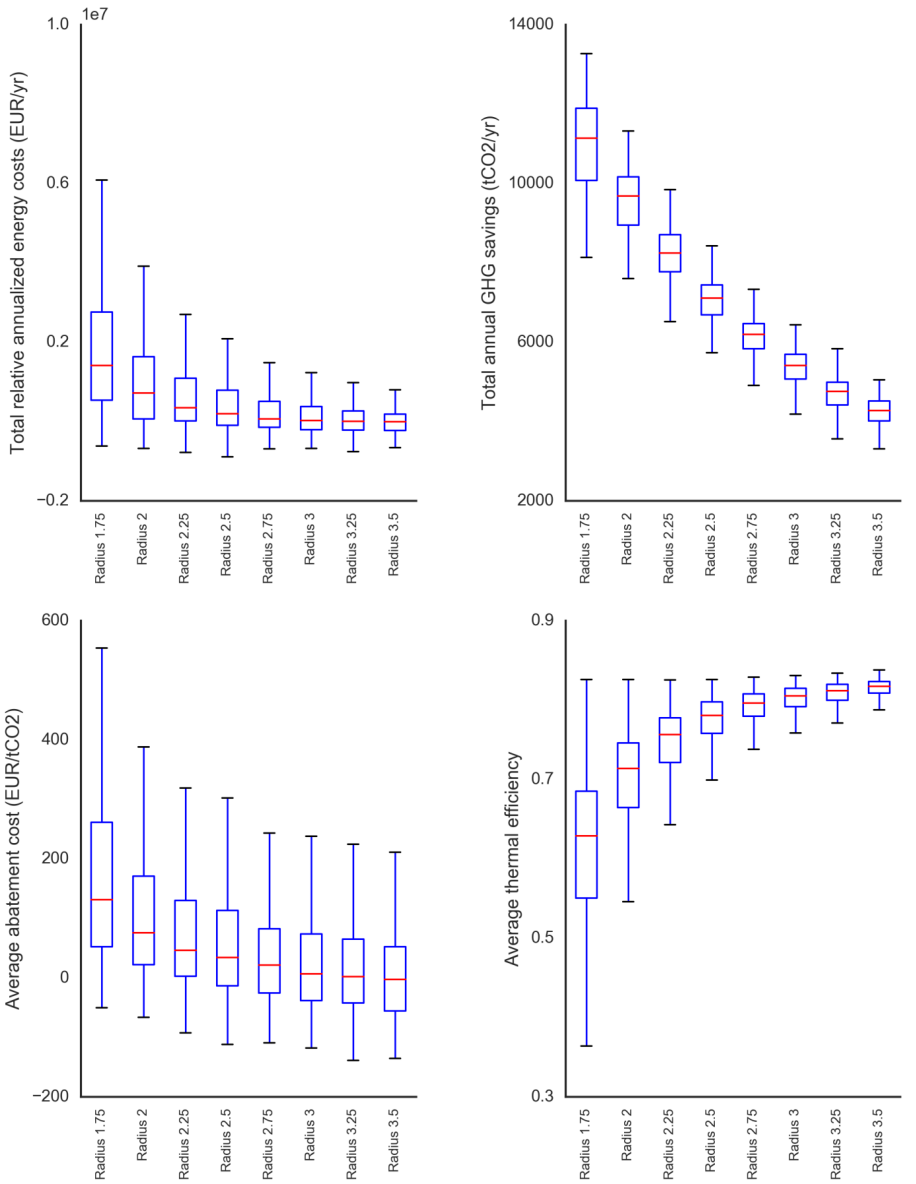


Figure 5.6: Impact of well distance policies on final model outcomes under uncertainty

to approximate the output of a traditional global sensitivity analysis at a smaller computational cost, while including categorical uncertainties for the well distance policies. Figure 5.8 shows the resulting ranking for influential variables, in relation to the relative annualized energy costs and the average GHG abatement cost at the end of the

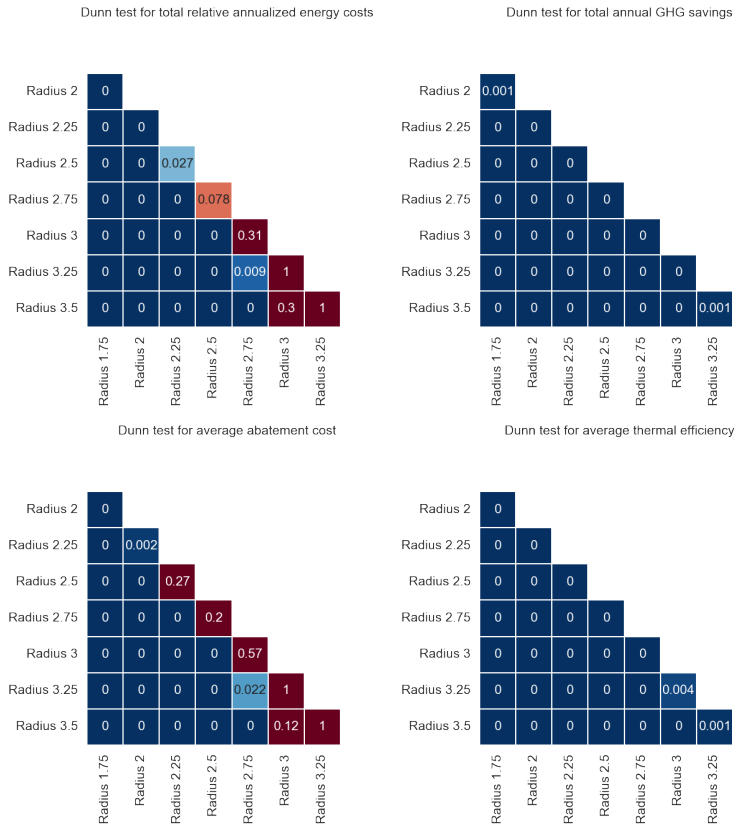


Figure 5.7: P-values for Dunn post-hoc tests on impact of well distance policies for model outcomes

simulation.

Gas price and the well distance policy are the most influential parameters for these two outcomes, with the other economic parameters for ATES capital cost, discount rate and electricity price also being relatively important. According to this classification, and under the assumptions followed when defining the uncertainty ranges, the geohydrological parameters for horizontal conductivity and aquifer porosity would only have a minor influence on annualized energy costs and effective GHG abatement cost.

5.3.4. Impact of time-dependent dynamics on system performance

The results presented in the previous subsections are based on aggregate outcomes across the simulated ATES operator agents. However, the performance of ATES systems is closely linked to local aquifer conditions which may be affected by the actions of neighbouring systems over time, and by the delayed feedbacks between expected

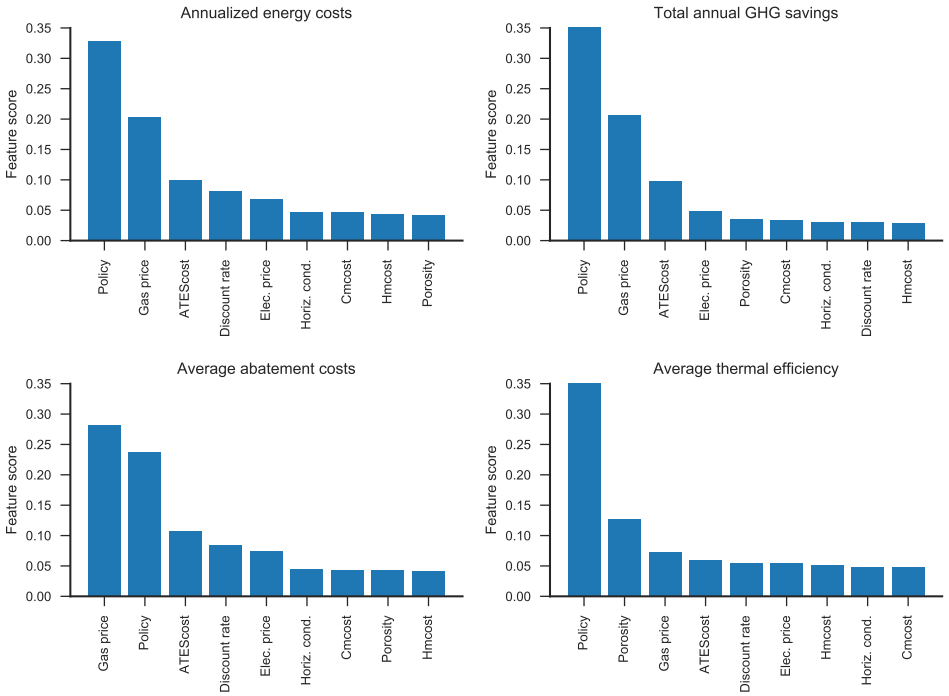


Figure 5.8: Ranking of influential variables for model outcomes

performance and adoption and operation. Further insights may therefore be gained by analysing the outcomes for individual agents over time. In particular, given that current ATES planning and permitting policies are typically based on a "first come, first served" principle, the links between the time of adoption and realized system performance are particularly policy-relevant. As such, Figure 5.9 and Figure 5.10 show the result of discretizing the simulation time frame in four periods; the boxplots correspond to the final outcomes for thermal efficiency and abatement cost, as obtained by the simulated ATES systems which first activated their wells within each of these periods. Given the variation of adoption rates across experiments, this provides a basis for comparing time-dependent dynamics across different experiments and policies. The plots for instance illustrate that "early adopter" agents – which activated their wells within the first half of the time interval – tend to obtain a higher final thermal efficiency, compared to agents which adopted ATES wells in the last half of the simulation. The inset tables provide the p-values for Dunn's post-hoc tests across the four periods, for each policy. This analysis indicates that adopters in the first quarter of the time interval benefit from a significantly higher thermal efficiency for the 1.75 and 2.5 R_{th} policies, but not in the 3 R_{th} case.

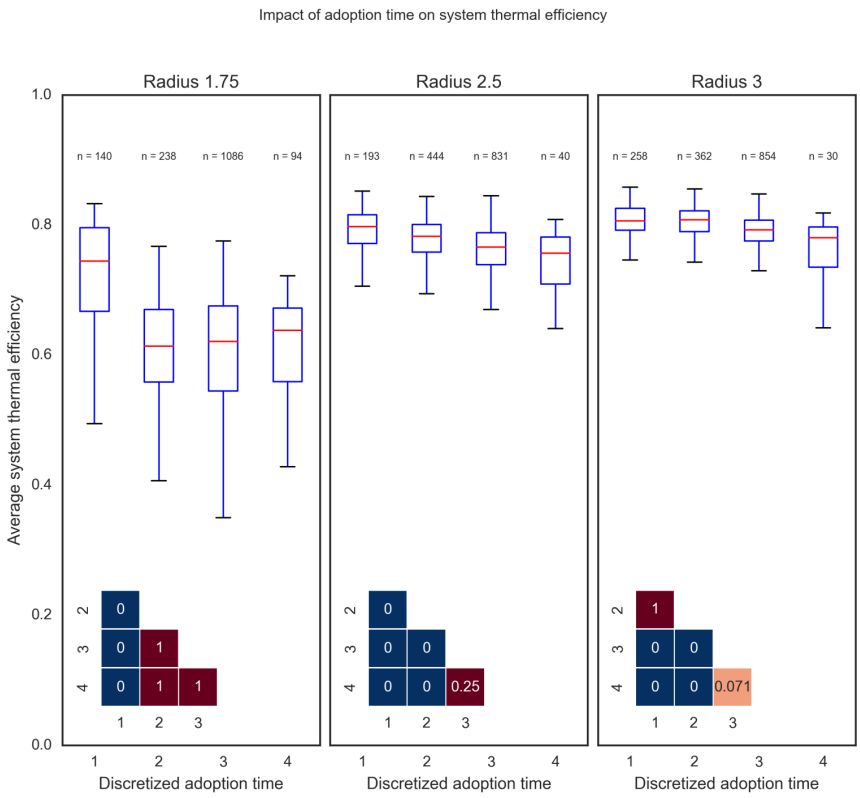


Figure 5.9: Adoption time effects and Dunn's post-hoc test for average thermal efficiency

The same conclusion can be obtained by considering the effective GHG abatement cost in Figure 5.10, for which "early adopter" systems in the first half of the simulation have significantly improved outcomes. In this case, outcomes for the final quarter of adopters are somewhat exaggerated by the high variability of the calculated abatement cost prior to the stabilization of temperature distributions in the subsurface. In both outcomes, the results indicate that larger distance policies – which delay and reduce thermal interactions – appear somewhat less sensitive to adoption time effects.

5.4. Discussion

Subsection 5.3.1 illustrates basic adoption patterns in which the simulated adoption of ATES technology follows a stylized S-shaped diffusion curve, which is typical of new technologies in general (Geroski, 2000); this adoption process underlies the economic and environmental outcomes which are the main focus of the analysis. Under the assumptions of the model, the adoption of ATES technology over time is directly

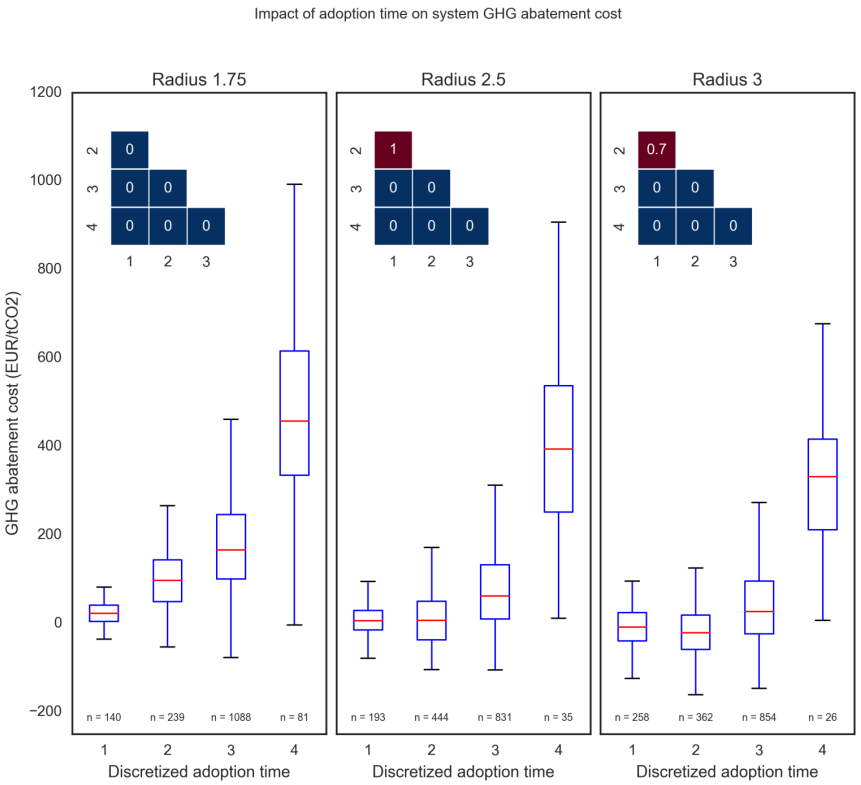


Figure 5.10: Adoption time effects and Dunn's post-hoc test for GHG abatement cost

driven by the distribution of economic adoption thresholds across agents. This functionally corresponds to analytical probit models in the economic literature (e.g. Davies, 1979). However, this approach does not explicitly consider factors which may be relevant for new energy technologies, such as technical uncertainty and risk aversion, or the diffusion of information across adopters. Given that ATES technology is relatively novel and relies on multidisciplinary expertise for the design and operation of systems, the empirical data which was used as a plausible assumption for economic adoption thresholds (Blok et al., 2004) may need to be refined to account for the specificities of ATES. With less conservative policies for ATES well distances, these adoption patterns lead to the development of thermal interferences over time, which eventually degrade the average efficiency and economic performance of systems. These results point towards the risk of a "tragedy of the commons" for urban ATES systems under improper planning practices, as the economic returns of operators are reduced by excessive interferences – which is exacerbated by delayed feedbacks between thermal interactions, economic performance, and ATES adoption. Furthermore, it is important

to note the presence of a general trade-off between economic returns and overall energy savings: overly conservative location policies will maximize individual returns but will artificially limit the space available for new wells, and therefore the energy-saving potential of the technology on an urban scale. On the other hand, since thermal interactions with smaller well distances tend to increase energy losses between neighbouring systems, individual ATES systems would be penalized by policies which perform best for collective GHG reductions or subsurface use. This conflict between individual interests and systemic outcomes will need to be acknowledged by planning authorities. Subsection 5.3.2 analyzes this trade-off under uncertainty, indicating that differences in energy savings and abatement cost may not be significant when reducing well distances to 2.75 R_{th} from the current Dutch guideline of 3 R_{th}. However, such a reduction could have a significant positive impact on collective GHG savings. This is generally consistent with results found by Li (2014), who suggested that system performance does not significantly increase at well distances above 2.5 R_{th}. Sommer et al. (2015) similarly found that relatively denser well layouts will increase the total energy retrieved from a given aquifer volume, despite higher losses. These relationships are in practice compounded by operational factors which yield a greater effective distance between wells, or a smaller effective thermal radius (such as other constraints on well locations in urban environments, and operators not using their full permitted capacity). This issue could be mitigated through feedback mechanisms which would track the effectively used capacity of wells, and periodically adjust the effective required clearances to avoid a waste of subsurface space. Subsection 5.3.4 focused on time-dependent dynamics, in regards to the relationship between adoption time and system performance. Generic order effects have been addressed by previous economic work on technology diffusion (e.g. Ireland and Stoneman, 1986; Karshenas and Stoneman, 1993); these may for instance allow early adopters of a technology to benefit from easier access to important production inputs or favourable geographic sites. In the context of ATES adoption, these temporal effects are explained by thermal processes in the subsurface: the development of thermal bubbles tends to increase the thermal efficiency of ATES systems over time, by reducing conduction losses to confining layers and increasing the fraction of energy which can be retrieved (Calje, 2010). Early adopters will therefore typically experience higher thermal performance within a given time frame. This has further implications for the transient development of positive or negative interactions between systems: for example, the optimal distance between neighbouring systems is likely to be affected by each system's operation time, and systems which have been operating long enough to develop stable thermal bubbles may be more robust to interactions with newer systems. These order effects will be an important subject for further study, as existing governance schemes lack the flexibility to manage these issues.

5.5. Conclusions

ATES technology can contribute significantly to reductions in energy consumption within the built environment. However, the sustainable management of this technology will require a good understanding of the interactions and trade-offs between the adoption of ATES, the economic performance of ATES systems, and the environmental dynamics which support the function of ATES systems. These interactions are often not taken into account by current practices for ATES modelling, which neglect endogenous dynamics for adoption and operation. To contribute to this goal, this work relied on a simulation architecture combining an agent-based ATES adoption and operation model, and a geohydrological aquifer model. This architecture was used to compare different spatial layout configurations for ATES systems under uncertainty. A comparison of different policies for the minimal distance between ATES wells evidenced a general trade-off between the economic performance of individual systems, and overall reductions in operational GHG emissions. However, this trade-off is not linear, and the results tend to support previous research in suggesting that existing planning practices in the Netherlands may artificially restrict the adoption of ATES systems. Under the assumptions of the model, current design guidelines used in the Netherlands could be revised to allow for a smaller distance of 2.75 R_{th} between wells –- and thus for a greater amount of wells to be built -– without significantly degrading economic performance. When further reducing well distances, however, the development of thermal interferences could plausibly lead to a "tragedy of the commons", which over time would compromise the average performance of ATES systems. An important point for the further improvement of planning policies would concern the inclusion of feedback mechanisms to adjust the required clearances between wells, according to their actual use: under current practices in the Netherlands, systems typically only use a fraction of their permitted pumping volume, which results in a smaller effective thermal radius -– and therefore in a waste of available space for ATES systems. An analysis of model sensitivities indicated that the natural gas price and well distance policy are most influential for the annualized energy costs incurred by ATES operators in comparison to conventional energy, and for the equivalent GHG abatement cost. Although conservative planning policies could help mitigate the effect of low gas prices on the relative economic performance of ATES, the latter remains primarily driven by exogenous factors.

It should otherwise be noted that the parameterization of the socio-technical and geohydrological models reflected idealized conditions for ATES in the Netherlands, but the simulation results may not necessarily be directly applicable in different geographic contexts – for example in the case of aquifers which present a high groundwater flow, high heterogeneity, or a significant salinity gradient. Spatial planning methods are simulated under a broader set of aquifer parameterizations in Bloemendal et al. (2018), while Bloemendal and Hartog (2018) further explore the impact of groundwater flow.

Similarly, the problem formulation used in this work focused on operational GHG emissions as an indicator of collective and environmental performance, but a broader definition of environmental impacts (e.g. Hähnlein et al., 2013) would for instance be needed to assess systems operating at higher storage temperatures, or in areas with significant aquifer contamination or drinking water extraction. The GHG emissions indicators could also be refined by accounting for lifecycle emissions, rather than only operational emissions (Tomasetta et al., 2015). The idealized model will be extended in the next chapter for a full case study of ATES development in the city centre of Utrecht, in the Netherlands.

6

Spatial planning for ATES in the city center of Utrecht

This chaper is based on Bloemendal, Jaxa-Rozen, and Rostampour (2017), with significant changes to the presentation, analysis and discussion of the simulation results; the simulation experiments and analysis in the original work had been carried out by the second author. The case study description was also revised for consistency with Chapter 7 of this thesis.

6.1. Introduction

There is a growing demand for energy-efficient technologies in buildings, which can meet heating and cooling needs under increasingly strict targets for energy use and greenhouse gas (GHG) emissions. Aquifer Thermal Energy Storage (ATES), which combines a heat pump with seasonal thermal storage in the subsurface, is an emerging technology which can markedly reduce the energy consumption of larger buildings. The potential for using ATES systems depends on climate and subsurface conditions; based of these conditions, the use of ATES is technically feasible in many temperate areas over the world (Bloemendal et al., 2015), and is therefore expected to rise in the future. Although this technical potential remains undeveloped in many parts of the world, practical experience with ATES systems has already been achieved in several European countries and elsewhere (Blum et al., 2010; Eugster and Sanner, 2007; Fry, 2009; Verbong et al., 2001; Fleuchaus et al., 2018). Particularly in the Netherlands, the number of ATES systems has grown rapidly in the past decade, often alongside the (re)development of urban areas.

ATES systems are typically clustered in urban areas where many large utility buildings – such as offices and commercial or institutional buildings – are concentrated on

top of a suitable aquifer. The spreading of warm and cold groundwater originating from the storage cycles depends on aquifer properties, ambient groundwater velocity, and the energy demand of the associated buildings. Under typical conditions in the Netherlands, this spreading may vary in a radius between 20-150 m around the storage wells and, therefore, often crosses beyond the plot of a building owner. As interaction between wells of "opposite" types (i.e. warm and cold) reduces the thermal efficiency of the systems, overlapping warm and cold zones are to be prevented.

However, the use of subsurface volume by ATES wells is variable and hard to predict. This is largely driven by daily and seasonal variations in the energy demand of the building, for instance due to weather, building occupancy, and changes in the operating conditions of building systems over their lifetime. The integration of the ATES system with other components of the building's heating, ventilation, and air conditioning (HVAC) systems can also be challenging, so that the ATES system may be difficult to operate consistently. At the same time, the spreading of warm and cold groundwater in the subsurface is not visible, and difficult and expensive to monitor. In these circumstances, to ensure that negative interactions are avoided between ATES systems, current practices in the Netherlands lead to wells being kept at a large mutual distance. This results in an underutilization of subsurface volume, and a loss of potential GHG savings from ATES systems. This underutilization is compounded by variations in the operation of the systems: Willemsen (2016) for instance found that ATES users in the Netherlands typically pump less than half of their permitted storage capacity on an annual basis.

To maximize the collective reduction in GHG emissions which could be achieved through the use of ATES, it is crucial to minimize the claim on the subsurface by individual systems, to allow accommodation of the largest possible number of ATES systems, while optimizing with respect to thermal recovery efficiency. As in many common-pool resource (CPR) problems (Ostrom, 1990), there is a trade-off between collective and individual performance; accommodation of more ATES systems in an aquifer reduces the total GHG emissions of all the buildings in that area, but this may reduce the efficiency of individual systems (Bloemendal et al., 2018). In current practice, concerns over the latter prevail, which results in conservative policies creating a scarcity in the subsurface.

In that respect, although the energy performance of ATES wells is important, the continued adoption of ATES will require resolving the scarcity of subsurface space which is emerging in areas such as the city of Utrecht – which is one of the densest areas for ATES development in the Netherlands, and where there is already a lack of space for additional systems in the city center. In parallel, as found by Willemsen (2016), the operation of systems typically differs significantly from the nominal design values which are used in their permitting and planning. As systems are used less than expected, this implies that the scarcity of subsurface space may be at least partly artifi-

cial. Subsurface conditions may also differ from expectations due to significant yearly changes in the thermal balance between warm and cold storage, as a result of variable weather; injection temperatures into the warm and cold storage wells can also vary widely, depending on the operation of the building's ATES and HVAC systems.

There is still a lack of understanding in regards to the interactions between these operational uncertainties, the spatial planning policies which are used for ATES in dense urban areas, and the resulting performance of ATES systems. To explore this issue, this chapter will introduce a simulation case study of ATES development in the Utrecht city center. This case study will be used to investigate the performance of ATES systems across different scenarios for spatial planning as well as technical and economic uncertainties, from the perspective of system operators and policymakers.

The following section describes the simulation methods, as well as the case study; this is followed by simulation experiments, then by a discussion of the results and directions for future work.

6.2. Methods

6.2.1. Simulation environment

The adoption and operation of ATES technology essentially represents a complex adaptive system, due to the feedbacks and interactions between aquifer conditions and ATES/building operation. To simulate the development of ATES and the resulting performance trade-offs, this work therefore relies on a coupled simulation architecture in which geohydrological dynamics are modelled using the MODFLOW / MT3DMS codes, while an agent-based model of ATES adoption and operation is implemented using NetLogo. These model components are linked through an object-oriented architecture using the Python language. The coupled models are then simulated using the EMA Workbench package for exploratory modelling (Kwakkel, 2017), which allows for the generation of ensembles of computational experiments and for their analysis. The simulation architecture and the NetLogo agent-based model are detailed further in Chapters 3 and 5 of this thesis, respectively.

6.2.2. Performance objectives and ATES parameters

To represent the potential trade-off between individual and collective interests for ATES operation, the main indicators used to assess ATES performance are the aggregate operational cost savings of the simulated ATES wells (Δ_C), as well as the GHG savings realized (Δ_{GHG}), both relative to a conventional building energy system which would provide equivalent heating and cooling. Additional measures track the specific cost savings per cubic meter of water pumped by ATES systems (ν_C), and the specific GHG savings realized per cubic meter of subsurface space allocated to ATES thermal zones (ν_{GHG}). These measures provide an indication of the individual and collective

efficiency of the systems. As an additional reference for individual system performance, the average thermal recovery efficiency (η_{tot}) of the systems is also tracked. Appendix C summarizes the computation of these indicators.

The variables used in these calculations are given in Table 6.1; the technical parameters for ATES operation, as well as energy prices, are sampled across uncertainty ranges which represent typical operating conditions in the Netherlands (Willemsen, 2016; Eurostat, 2018a,b). The grid emission factor for electricity similarly corresponds to current values in the Netherlands (Moro and Lonza, 2017). Cooling delivered from ATES is assumed to be direct (i.e. free) cooling, while a heat pump is used for heating. The energy demand profile of the buildings is driven by an exogenous temperature time series, based on historical temperature data for the De Bilt weather station for 1998-2014, then by synthetic data corresponding to the KNMI W+ scenario for 2014-2030 (KNMI, 2014a). Equivalent ATES well flows are then computed, assuming that warm and cold wells are required to be balanced over a moving 5-year period. The resulting nominal ATES well flows are here modified by a flow multiplier Q_{mult}, and by a thermal imbalance factor Q_{imb}. The former can be used to represent a partial usage of the nominal well capacity, while the latter shifts the nominal thermal balance between warm and cold wells; starting from nominal well flows, a value of 0.2 for Q_{imb} would for instance yield an average annual imbalance of 20% towards cooling.

Parameter	Value or range	Unit	Symbol
ATES nominal temperature difference	4–8	[K]	ΔT
ATES pump efficiency	0.25	[-]	η_p
Boiler efficiency	0.95	[-]	η_b
COP chiller	3–5	[-]	COP_c
COP heat pump	3–5	[-]	COP_{hp}
Effective flow multiplier	0.4–1.0	[-]	Q_{mult}
Annual thermal imbalance towards cooling	-0.2–0.2	[-]	Q_{imb}
Emission factor (electricity)	0.157	$[tCO_2/GJ]$	f_e
Emission factor (natural gas)	0.056	$[tCO_2/GJ]$	f_g
Price for electricity	15–60	$[EUR/GJ]$	C_e
Price for natural gas	5–25	$[EUR/GJ]$	C_g
Distance multiplier between wells	[2.5, 3]	[-]	d
Nominal capacity of new simulated wells	40,000-200,000	$[m^3/yr]$	Q

Table 6.1: Parameter values used in the simulation study.

6.2.3. Model setup

This analysis relies on simulation models for ATES adoption and operation which represent ATES systems in a 2500 m x 2500 m region of the city centre of Utrecht. Data

for a set of 89 existing and planned wells was obtained from the Utrecht provincial database, and combined with GIS data from the TOP10NL cadaster dataset; this spatial data was used to represent building plots and spatial constraints on ATES well location (e.g. roads or water features) within the NetLogo agent-based environment. The models are simulated over 300 monthly time steps (i.e. 25 years), starting in 1998 with 23 active wells; additional wells then become active over time based on their historical construction date, until 2017. After 2017 and until the end of the simulation, the agent-based model adds new ATES wells on available building plots, within the well layout policies which are described in the next subsection.

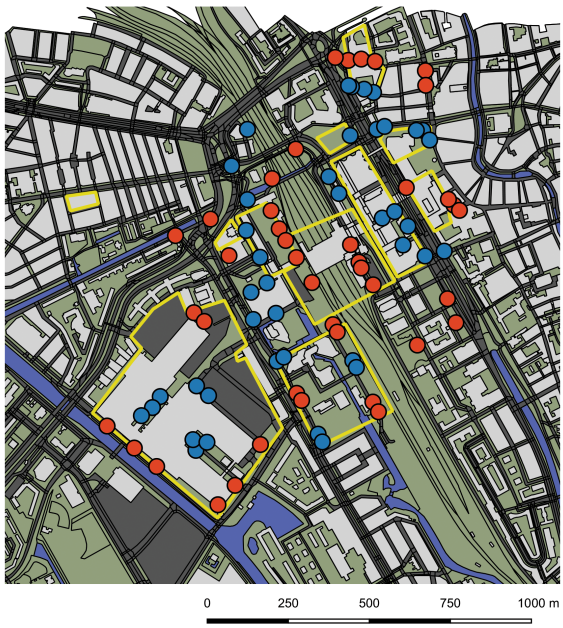


Figure 6.1: Area simulated in the agent-based model, showing building plots available for development (in yellow) and warm and cold ATES wells (red and blue).

The model used to represent the geohydrological setting is a cut-out of the regional MODFLOW (Langevin et al., 2003) groundwater model, developed under the authority of the water board *Hoogheemraadschap De Stichtse Rijnlanden* of the Utrecht district. More details on the model conceptualization and parameterization are provided in Borren (2009) and Gunnink (2004).

The subsurface of Utrecht is heterogeneous and consists of an alternation of permeable sandy aquifers and non-permeable clay aquitards. The upper layer is anthropogenically influenced by excavations and building construction, and is therefore strongly mixed. The first (upper) aquifer consists of several horizontal layers, varying from fine

sand to gravel, with a depth from 4 to 45 m. ATES systems are placed in this upper aquifer. A 25 m thick aquitard then separates the first aquifer from the deeper second aquifer. Faults are present in the original model, but not in the cropped part of the model which is used in this work. Constant head boundaries are set up with the corresponding heads from the original model, with initial heads set to represent the ambient groundwater flow. To minimize simulation runtimes, only ATES-relevant processes are simulated; the MODFLOW packages for recharge and evaporation (RCH) and surface runoff (SOF) are not used in this study. As this study focuses on confined deeper aquifers with a relatively short time horizon, these effects can be disregarded without loss of accuracy. The model consists of eight layers, up to about 200 m below surface level. To improve accuracy of the calculated temperature field around the well, the regular grid was refined around the ATES wells following the recommendations of Sommer (2015). Additional parameterization for the simulation of transport processes using MT3DMS can be found in Phernambucq (2015).

6.2.4. Scenarios for ATES layout

The spatial planning of ATES systems is typically based on guidelines for the minimal distance between neighbouring wells, which is itself defined as a given multiplier (d) of the average thermal radius R_{th} representing the footprint of the thermal cylinder created around the well in the subsurface (Figure 5.3). This value corresponds to the expected radius of thermal influence, based on the analytical solution for heat transport in porous media; the prevailing Dutch guidelines for ATES system design require a distance of 3 R_{th} to avoid thermal interactions. However, these guidelines may be overly conservative (Sommer et al., 2015), particularly when combined with other geographic restrictions on well layout. The simulations will therefore test a tighter distance policy of 2.5 R_{th} in addition to the current policy.

In addition, Willemsen (2016) showed that ATES systems on average only use less than half of their permitted capacity. To evaluate how this "over-claiming" affects total energy savings, individual performance, and effective spatial layout, a multiplier on the theoretical capacity is added to test the effect of smaller effective flows (Q_{mult} in Table 6.1). Furthermore, an adaptive policy measure is evaluated. For the city centre of Utrecht, the permitted volumes of yearly groundwater storage and recovery are obtained from provincial data; starting from this baseline, the simulated permit capacity is adapted every two years based on the actual use of existing ATES systems, and the equivalent well footprint is recomputed. New ATES systems may then be placed in the subsurface space which becomes available from unused claims.

6.3. Simulation results

The uncertainty ranges presented in Table 6.1 were sampled using uniform distributions and a Latin Hypercube design, for 512 experiments under each of the four policy

scenarios: static or adaptive simulated permits, for distances of 2.5 and 3 R_{th}, for a total of 2048 simulations. To illustrate the basic behaviour of the model, Figure 6.2 shows a Gaussian kernel density estimate for the distribution of key outcomes at the end of the simulated period, grouped by policy scenario. As an additional reference, the total number of wells under each policy is also presented. As can be expected, the highest number of wells typically occurs with the adaptive scenarios and the denser layout scenarios; the economic indicators present a broad distribution due to the wide ranges of plausible energy prices which were specified.

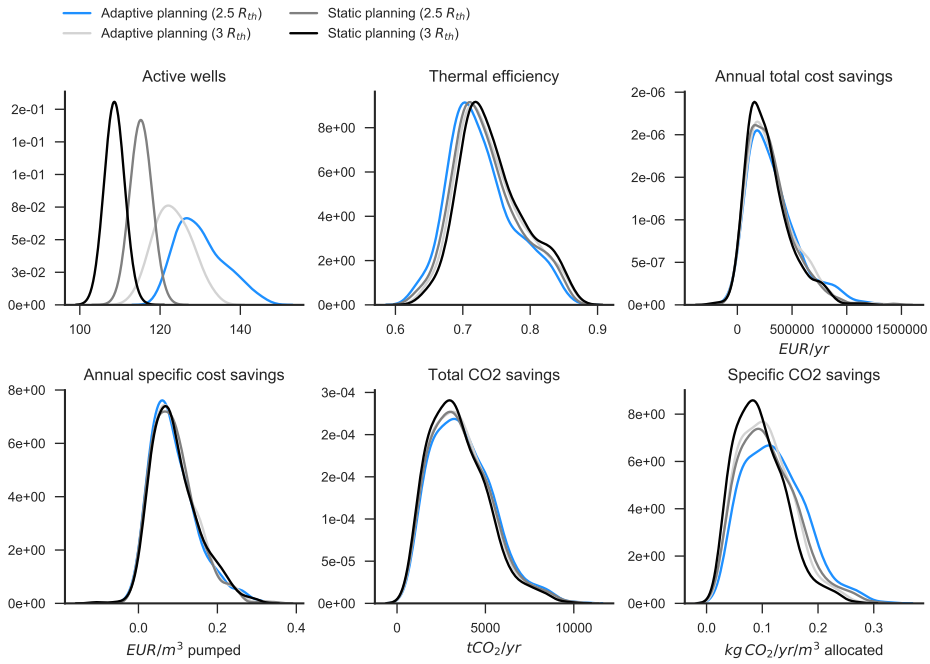


Figure 6.2: Distribution of the key performance indicators at the end of the simulation, grouped by policy scenario.

Table 6.2 summarizes sensitivity analysis results for the five main performance indicators, using the Random Forests technique (Breiman, 2001) for nonlinear regression to approximate the results of a global sensitivity analysis at a smaller computational cost, as described in Chapter 4 of the thesis. The table shows the five most influential parameters for each output, based on values at the end of the simulation. The results show that the gas price is a key uncertainty for the total and specific cost savings, while heat pump COP has a smaller impact; the Q_{mult} and ΔT operational uncertainties are predominant for total and specific GHG savings, with the latter (which represents the effective temperature difference between the wells) being significant across all in-

dicators. Figure 6.3 provides an additional visualization for the impact of these two operational uncertainties on the economic and GHG indicators, using a parallel coordinate plot which for instance clearly shows the positive relationship between Q_{mult} and GHG savings. While the policy scenarios are fairly influential on specific GHG savings, they are not ranked in the five most influential variables for the economic indicators, and are less influential than the technical/operational uncertainties in regards to the thermal efficiency of systems.

	Thermal efficiency		Total cost savings		Specific cost savings		Total CO_2 savings		Specific CO_2 savings	
Rank	**Variable**	**Estimated importance**	**Variable**	**Estimated importance**	**Variable**	**Estimated importance**	**Variable**	**Estimated importance**	**Variable**	**Estimated importance**
1	ΔT	0.37	C_g	0.49	C_g	0.57	Q_{mult}	0.40	ΔT	0.35
2	Q_{tmb}	0.26	Q_{mult}	0.13	ΔT	0.11	ΔT	0.33	Q_{mult}	0.29
3	Q_{mult}	0.07	ΔT	0.09	C_e	0.07	COP_{hp}	0.06	**Policy**	0.15
4	**Policy**	0.06	C_e	0.06	COP_{hp}	0.05	COP_C	0.06	COP_C	0.05
5	-	-	COP_{hp}	0.05	COP_C	0.04	**Policy**	0.04	COP_{hp}	0.05

Table 6.2: Sensitivity of varied parameters for performance indicators.

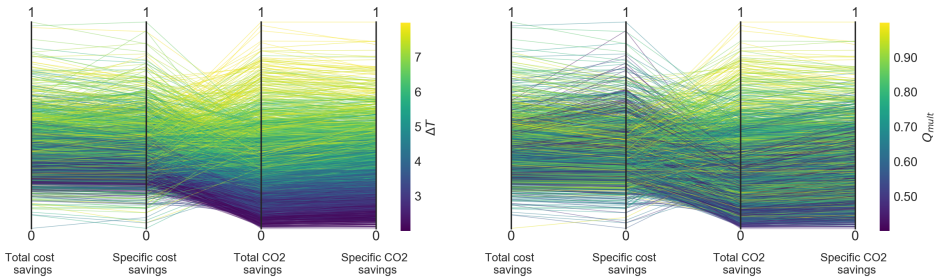


Figure 6.3: Normalized performance across values of ΔT (left) and Q_{mult} (right).

Table 6.3 summarizes the ensemble results for each policy; compared to the baseline policy, this indicates that the adaptive 2.5 R_{th} policy leads to significant average improvements on GHG indicators and total cost savings: due to the additional systems which can be built in this scenario, more groundwater is utilized, resulting in higher overall savings on GHG emissions and costs. The additional use of the subsurface may increase thermal interferences over time, so that thermal efficiency and specific cost savings per pumped unit of water are slightly reduced. However, the difference across policies on this indicator is relatively minor (<4%), compared to the difference on the GHG indicators (8% for total GHG savings and 27% on specific GHG savings), due to the greater efficiency in allocation of subsurface volume under the adaptive 2.5

R_{th} policy. The adaptive 3.0 R_{th} and static 2.5 R_{th} policies occupy a middle ground on these indicators, with slightly different trade-offs: the former offers relatively better economic performance, and the latter performs better on GHG indicators.

	Thermal efficiency		Total cost savings		Specific cost savings		Total CO_2 savings		Specific CO_2 savings	
Policy	**Mean**	**%**	**Mean (EUR/yr)**	**%**	**Mean (EUR/m³)**	**%**	**Mean (tCO_2/yr)**	**%**	**Mean (kg CO_2/yr/m³)**	**%**
Adaptive 2.5 Rth	0.730	96.5	325454	116.1	0.089	96.5	3713	108.0	0.123	127.0
Adaptive 3.0 Rth	0.746	98.6	309160	110.3	0.091	98.3	3578	103.6	0.106	110.4
Static 2.5 Rth	0.744	98.3	297068	106.0	0.089	96.8	3607	104.5	0.111	114.8
Static 3.0 Rth	0.756	100	280239	100	0.092	100	3452	100	0.097	100

Table 6.3: Performance indicators for each policy, averaged across the ensembles (% = relative to Baseline 3.0 R_{th}).

To better understand the implications of uncertain parameters and policies towards these trade-offs, a scenario discovery approach (e.g. Bryant and Lempert, 2010) was used to study the combinations of influential parameters which may cause outcomes of particular interest. A Gaussian mixture model (GMM; described in e.g. McLachlan and Peel, 2000) was first applied to cluster the simulation experiments across the economic and GHG indicators used in the analysis, using outcome values at the end of the simulation. Thermal efficiency was not retained for this step, as it largely correlates to specific cost savings and is not influenced by the energy price uncertainties.

The GMM technique essentially assumes that the outcome values are generated by a mixture of an arbitrary number of Gaussian distributions; expectation maximization can then be used to assign each sample to the distribution to which it most likely belongs, generating a finite number of clusters. Using the Bayesian information criterion to assess the quality of fit of the possible models, we select a model composed of four Gaussian components (or clusters), resulting in the groups shown in Figure 6.4. The graphs present the normalized axes used for clustering, so that values of 0 and 1 respectively correspond to the lowest and highest value across the 2048 experiments. For instance, Cluster 0 groups experiments which have average total cost savings, relatively high specific cost savings, and relatively low total and specific GHG savings.

These clusters can then individually be studied by using the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) to identify the combinations of uncertainties which lead to a given cluster classification. Expressing the model output as $y = f(x)$, with x corresponding to a vector of uncertain input parameters of length M, the sets of cases of interest corresponding to each cluster C can be defined as $I_j = \{x^I | f(x^I) \in C_j\}$ (Bryant and Lempert, 2010). Each of these sets of cases of interest can be described by one or more sets of limiting constraints $B_k = \{a_n \leq x_n \leq b_n, n \in L_k\}$, applied to the ranges of a subset of the input

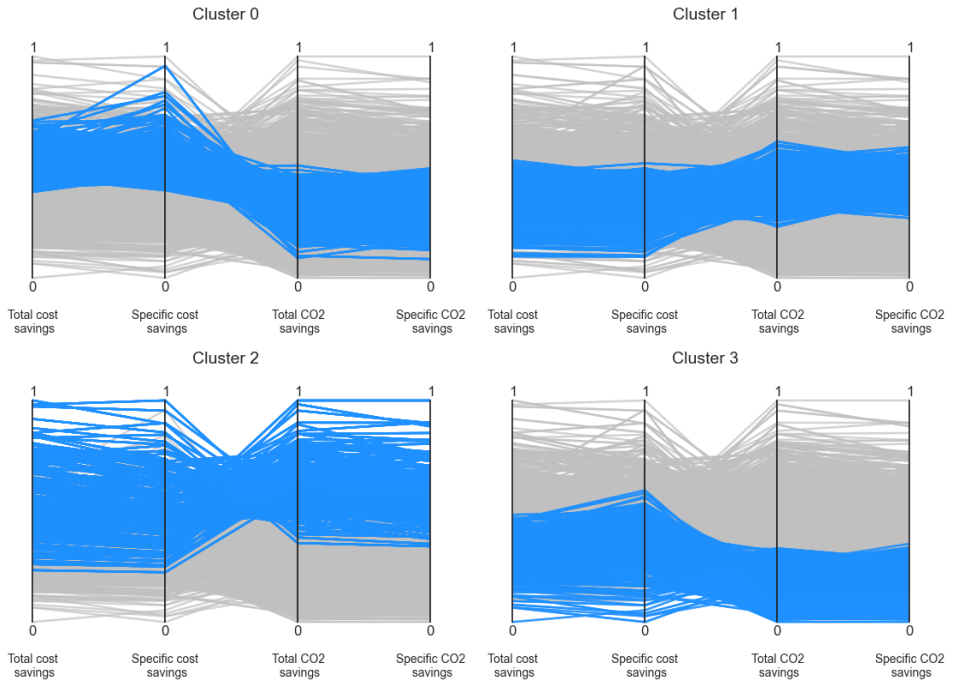


Figure 6.4: GMM clustering results.

parameters: $L_k \subseteq \{1, ..., M\}$. Each set B_k then corresponds to a "box" within the multidimensional input space which is associated with a given output set of cases of interest; the set of boxes B can in turn be interpreted as a scenario.

This definition assumes that the cases of interest are entirely described by a (hyper-) rectangular region of the uncertainty space, which is rarely the case in practice. To support the identification of interpretable scenarios, the PRIM technique therefore aims to maximize the coverage and density of a box set B. The coverage equals the ratio between the total number of cases of interest within a box set, and the total number of cases of interest; the density corresponds to the ratio between the total cases of interest in a box set, and the total number of cases in the box set. Following Bryant and Lempert (2010), these indicators are expressed as:

$$\text{Coverage} = \frac{\sum_{x_i \in B} y_i}{\sum_{x_i \in x^I} y_i} \quad ; \quad \text{Density} = \frac{\sum_{x_i \in B} y_i}{\sum_{x_i \in B} 1} \tag{6.3.1}$$

$$\text{with} \begin{cases} y_i = 1 & \text{if } x_i \in I \\ y_i = 0 & \text{otherwise} \end{cases}$$

Cluster 2 (which has relatively high economic performance and the highest GHG savings) and Cluster 3 (which has the lowest performance on all indicators) are of particular interest. Table 6.4 shows the parametric combinations which tend to be associated with these two clusters following the PRIM analysis. The Min./Max. values indicate the parametric ranges identified for each cluster; the p-value estimates the significance of each parameter within this combination using a one-sided binomial test, so that low values indicate a high likelihood that a parametric range is significant in the combination. For each cluster, the PRIM coverage metric gives the fraction of experiments within the cluster which is described by the specified combination of uncertainties; conversely, the PRIM density gives the fraction of experiments which share these combinations which are within the given cluster.

	Cluster 2 - Best performance (176 experiments in cluster)			**Cluster 3 – Worst performance (546 experiments in cluster)**		
Parameter	**Min.**	**Max.**	**p-value**	**Min.**	**Max.**	**p-value**
ΔT	6.6	8	~0	4	5.7	~0
Q_{mult}	0.79	1	0.002	0.4	0.8	~0
COP_{hp}	3.5	5	0.007	-	-	-
C_g	-	-	-	5	19	0.008
Policy	[Baseline 2.5, Adaptive 2.5, Adaptive 3.0]		0.047	[Baseline 2.5, Baseline 3.0, Adaptive 3.0]		0.063
PRIM coverage	0.57			0.61		
PRIM density	0.81			0.87		

Table 6.4: PRIM results for cluster 2 and 3.

For example, for Cluster 2, this means that 81% of the cases in which ΔT is between 6.6 and 8, Q_{mult} is between 0.79 and 1, the effective COP is between 3.5 and 5, and the planning policy is either adaptive or denser, are classified in the high-performing cluster (although it should be noted the adaptive policies would be less influential with these relatively high values for Q_{mult}). Conversely, the results for Cluster 3 show that a combination of low ΔT, low Q_{mult}, low gas price, and static or sparser planning policies, tends to cause poor performance. In both clusters, the policies are relatively less significant than the parametric uncertainties.

6.4. Discussion and conclusions

The results of the analysis show that the relative economic performance of ATES systems is dominated by energy prices – in particular for natural gas – as well as two operational factors which vary significantly in practice: the effective pumped volume, and the effective injection temperature difference between warm and cold wells (Q_{mult},

ΔT). Under typical practices, this implies that adaptive and/or tighter permit policies would be relatively less influential towards operational costs for ATES users. This was supported by the mean values for the specific cost savings per unit volume of pumped water, which showed a relatively minor reduction of 3.5% for the adaptive 2.5 R_{th} policy relative to the baseline case. Conversely, the policies have a more significant impact on total and specific GHG savings; due to the additional systems which can be allocated to subsurface volume which would otherwise be reserved but not used, more groundwater can be utilized under adaptive and denser policies – increasing overall economic and GHG savings. As such, the adaptive 2.5 R_{th} policy would lead to an average 8% increase in total GHG savings, and a 27% increase in specific GHG savings per unit of subsurface volume. This is despite the path-dependence of ATES development, which is already extensive in the Utrecht city center: such revised policies would likely be more beneficial in a "green-field" development context, where wells could be located more densely from the start.

The interaction of spatial planning and operational uncertainties is particularly relevant in the case of the Q_{mult} multiplier, which corresponds to the actual usage of permitted storage capacity. Figure 6.5 presents the effective spatial planning density across three values of Q_{mult}, using a kernel density estimate of the distribution of minimal relative distances to other wells of the opposite temperature, across the 89 existing and planned wells simulated in the analysis. For instance, with a full usage of the nominal storage capacity, this indicates that the median minimal distance is 3.37 R_{th} – or slightly above the current design guideline, which is reasonable considering geographic limitations on well layout. However, with a value of 0.4 for Q_{mult} (which was approximately the mean value found by Willemsen (2016)), this median is increased to 5.33 R_{th}. This indicates that the effective spatial layout of wells in urban areas is likely much sparser than would be expected under nominal design conditions. Furthermore, considering that the guideline of 3 R_{th} is in itself conservative, this adds an additional and unneeded safety factor in regards to thermal interferences – and ultimately confirms that the scarcity of space for further development of ATES may be largely artificial. In these circumstances, an adaptive permitting approach could allocate subsurface volume more efficiently.

However, in the context of current conditions for ATES in the Netherlands – where permitting requirements are already perceived by ATES users to be fairly restrictive for the development of new ATES systems (de Graaf et al., 2016; Bloemendal and Jaxa-Rozen, 2016) – an adaptive permitting method may be difficult to implement in practice; this approach would entail additional administrative overhead for system operators, as well as provincial or municipal authorities. Systematically revoking or adjusting the licenses of existing systems would be particularly difficult, given that system operators could reasonably expect to maintain their licensed storage volumes. On the other hand, "grandfathering" these existing systems would limit the benefits of adaptive permitting

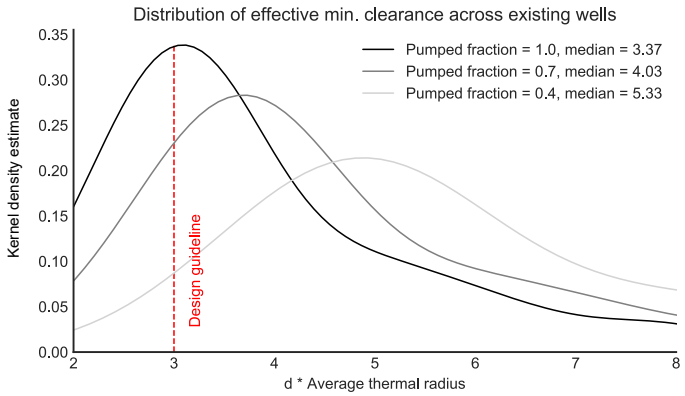


Figure 6.5: Impact of used storage fraction on effective well layout.

for areas where subsurface space is already scarce. As such, introducing adaptive permitting could potentially prove counterproductive should these requirements discourage prospective ATES adopters, by providing less flexibility for the usage of systems, and by introducing uncertainty in regards to the regulatory framework for ATES. A more efficient compromise may ultimately be found through the exchange of information across systems, to maximize their use of storage capacity while dynamically managing thermal interactions. This approach will be assessed in the next chapter of this thesis. We note that this analysis was limited by some simplifications in the operational behaviour of the ATES systems, particularly relating to the simple heat pump/COP assumptions. The calculation method for energy performance assessment was also relatively straightforward. The ATES/building system model should thus be improved by taking into account a cut-off temperature for free cooling, the use of "peak" conventional equipment for both heating and cooling, and the effect of partial load operation of the heat pump. While these aspects may have a relatively limited impact on ATES use, a more detailed building model would help endogenize some of the parameters which were here taken as exogenous, such as the ΔT temperature difference and the Q_{imb} storage imbalance. This is currently the topic of further research in the context of the URSES+ research program, which will include a more detailed case study for ATES operation in Amsterdam's Van Gogh museum.

7

Smart Grids for Aquifer Thermal Energy Storage

This chapter is based on Rostampour, Jaxa-Rozen[1], Bloemendal, Kwakkel, and Keviczky (2018b), with the addition of an idealized simulation case study.

7.1. Introduction

Aquifer Thermal Energy Storage (ATES) is an innovative building technology which can be used on a large scale to store thermal energy in natural subsurface formations. In combination with a heat pump, ATES can reduce energy use for heating and cooling by more than half in larger buildings (Tomasetta et al., 2015; Vanhoudt et al., 2011), while supporting the electrification of building energy systems. This has made the technology increasingly popular in Northern Europe. For instance, it is currently used in approximately 10% of new commercial and institutional buildings in The Netherlands, where ATES has been identified as a key technology towards long-term targets for greenhouse gas (GHG) emissions reductions in the built environment. Furthermore, the conditions required for ATES are relatively widespread across the globe; by the middle of the century, roughly half of the world's urban population is expected to live in areas with suitable subsurface and climate conditions for ATES (Bloemendal et al., 2015). These areas include large parts of China and North America, where the large-scale deployment of ATES could eventually reduce total energy demand for space heating and cooling by up to 10% in large residential, commercial and institutional buildings (as will be detailed in the next chapter of the thesis).

[1]The first two authors have contributed equally to this work.

However, the progressive adoption of this technology in Europe has already evidenced some issues of concern for policymakers. The spatial layout and planning of ATES systems is a key aspect for the management of the technology, as thermal interferences between neighboring systems which share an aquifer can affect their technical and economic performance. Current regimes for ATES management have therefore followed a highly conservative approach, to minimize the risk of a "tragedy of the commons" (Hardin, 1968) – in which the overly dense development of ATES systems could lead to thermal interferences, and eventually degrade the potential of aquifers for thermal storage. However, recent research suggests that current planning methods are essentially incompatible with long-term objectives for the development of ATES; for instance, several urban areas in the Netherlands already lack the space to accommodate further demand for ATES (Bloemendal et al., 2018), although the technology's total market share is still only one-fifth of national policy objectives (Agterberg, 2016). This situation presents a trade-off between public and private interests: while relaxed planning guidelines could contribute to GHG mitigation efforts by increasing the adoption of ATES, this could reduce the economic performance of the technology – and ultimately its long-term attractiveness as an energy-efficient option for building owners.

Improved methods for the management and operation of ATES systems will therefore be needed to better align the interests of policymakers and building owners, and fulfill the technical potential of ATES. To this end, this chapter presents distributed ATES control as a starting point towards an improved regime for the management of ATES in urban areas. This approach enables the dynamic management of thermal interactions in the subsurface – thus allowing a significantly denser spatial layout for ATES without affecting system performance, and increasing feasible energy savings from a given subsurface volume. To illustrate this approach, we first implement a generic dynamical energy management framework based on model predictive control (MPC) principles (Rostampour and Keviczky, 2016). This framework is tested using an idealized case study, which assumes a complete exchange of information to coordinate ATES operation across three neighboring buildings. This case study is simulated using a coupled agent-based/geohydrological environment. We then extend this concept to a more realistic configuration, which addresses computational issues arising at a larger scale through a distributed control approach. To this end, we implement a distributed probabilistic dynamical energy management framework (Rostampour and Keviczky, 2017). This framework is simulated under a range of scenarios for spatial planning and ATES operation in a case study for the city of Utrecht, in The Netherlands.

Section 7.2 will introduce the methods used to implement this control approach, by summarizing model predictive control principles, as well as the coupled agent-based/geohydrological simulation architecture which is used to test the control framework. Section 7.3 then describes the two case studies, followed by simulation results for each case in section 7.4. Section 7.5 discusses the results in the context of the ATES

simulation studies presented earlier in this thesis, and section 7.6 concludes with policy recommendations.

7.2. Methods

7.2.1. Model predictive control for ATES systems

Model predictive control (MPC) is a widely used modern optimal control strategy, which typically offers an attractive trade-off between optimality and computational cost. The concept of MPC is simple: predict the behaviour of a system given its model and measurements of the current state of the system, and given a hypothetical control input trajectory or feedback control policy. The control inputs are parametrized by a finite number of variables which denote a finite number of degrees of freedom. The predicted cost of the problem is optimized over these variables, using a given cost function. The control input is then applied to the system in a receding horizon fashion, wherein only the first element of the predicted control input sequence is applied to the system at the current time instant. The horizon is shifted at the next time instant, and the optimization problem is carried out again to obtain a new sequence of control inputs.

The receding horizon strategy is instrumental in reducing the gap between the predicted response and the actual response of the system; this strategy also provides a certain amount of robustness to uncertainty that can arise in the system. This uncertainty arises in the form of uncertain model parameters – which is known as multiplicative model uncertainty – and in the form of additive disturbances appearing from external sources, which is known as additive uncertainty; typical formulations for these uncertainties are for instance described in Farina et al. (2016). MPC has the ability to handle operations of processes within well-defined operating constraints, which is not always a given with other methods, but which allows e.g. equipment limits to be represented realistically. These constraints are handled systematically during the design and the implementation of the controller. MPC can respond to structural changes such as actuator and sensor failures, or changes in system parameters, by adapting the control strategy at every time step of execution of the algorithm. For these reasons, MPC has evolved from a basic multivariable process control technology, to a technology that has become widely accepted in industry – which includes the operation of building energy systems or smart thermal grids (Ma et al., 2012; Farahani et al., 2016; Patel et al., 2016; Mirakhorli and Dong, 2016). Compared to conventional methods such as PID controllers, the ability of MPC to handle large-scale dynamical systems under strict constraints offers several advantages for these applications.

Centralized control

Rostampour and Keviczky (2016) presented a MPC formulation for a building climate comfort (BCC) system combining ATES with conventional heating/cooling equip-

ment (i.e. a boiler and chiller). This formulation was first expressed as a finite-horizon, mixed-integer quadratic optimization problem for a single building; as such, this single-agent formulation aims to match the demand and production of thermal energy in the building, while minimizing operation costs and satisfying physical constraints for heating and cooling capacity. This problem was then extended to a centralized multi-agent formulation for multiple buildings. This maintains the individual optimization problems for each building and adds coupling constraints between neighbouring buildings, in order to avoid mutual interactions between ATES systems. These formulations build on a robust randomized approach to remain computationally tractable while meeting a desired level of reliability in regards to the constraints.

The coupling constraints rely on the single-agent state variables of the individual ATES systems, which are modelled using first-order difference equations to represent the water volume and thermal energy stored by each building. These equations assume that each ATES system is composed of one warm well and one cold well, which are physically linked; the control variable for the pump flow rate in heating and cooling modes (H and C) is given by $u^H_{a,k}$ and $u^C_{a,k}$ $[m^3h^{-1}]$ respectively, for each sampling time $k = 1, 2, ...$ Taking $\tau[h]$ as the sampling period, the usable volume of water stored in the warm and cold ATES wells, V^H_a and V^C_a, is then given by:

$$\begin{aligned} V^H_{a,k+1}[m^3] &= V^H_{a,k} - \tau(u^H_{a,k} - u^C_{a,k}) \\ V^C_{a,k+1}[m^3] &= V^C_{a,k} - \tau(u^H_{a,k} - u^C_{a,k}) \end{aligned} \tag{7.2.1}$$

Assuming the stored volumes can be approximated by a cylinder with a height equal to the well screen length $L[m]$, and taking a constant aquifer porosity n, this yields hydraulic radii r^H_h, r^C_h for the warm and cold wells:

$$\begin{aligned} r^H_{h,k}[m] &= \sqrt{\frac{V^H_{a,k}}{n\pi L}} \\ r^C_{h,k}[m] &= \sqrt{\frac{V^C_{a,k}}{n\pi L}} \end{aligned} \tag{7.2.2}$$

The equivalent thermal radii r^H_{th} and r^C_{th} of the warm and cold wells are then given by:

$$\begin{aligned} r^H_{th,k}[m] &= \sqrt{\frac{c_w V^H_{a,k}}{c_{aq}\pi L}} \\ r^C_{th,k}[m] &= \sqrt{\frac{c_w V^C_{a,k}}{c_{aq}\pi L}} \end{aligned} \tag{7.2.3}$$

where c_w and c_{aq} are the specific heat capacity for water and for the aquifer, respectively. For each building agent i, constraints can then be added to the optimization problem in order to avoid overlap between neighboring well radii, so that:

$$(r^H_{th,k})_i + (r^C_{th,k})_j \leq \theta d_{i,j}, \quad j \in \mathcal{N}_i \tag{7.2.4}$$

where $\mathcal{N}_i$ is the set of neighboring agents of agent i, $d_{i,j}[m]$ is a given distance between wells for agents i and j, and θ is a constant which can be used to adjust the influence of the constraint (so that a larger value will tend to relax the coupling constraint). The coupling constraint can equivalently be applied to hydraulic radii by choosing a different θ parameter.

These decoupled and centralized multi-agent formulations (i.e. without and with ATES coupling constraints; hereafter DS and CS) will be used in the idealized 3-building ATES simulation study, with an hourly sampling time and a day-ahead prediction horizon.

Distributed control

In practice, the centralized formulation can become computationally too costly to solve for a larger number of agents; in these conditions, distributed control offers improved performance. Distributed MPC (described in more detail in e.g. Camponogara et al., 2002) aims to replace large-scale centralized optimization problems with several smaller-scale problems, which can be solved in parallel. These small-scale problems make use of partial information from other subsystems to implement a distributed solution. In the presence of uncertainties, however, the main challenge in formulating a distributed MPC is the design of a suitable communication scheme to exchange this information between subsystems. Rostampour and Keviczky (2017) provide an appropriate technique to decompose a large-scale scenario-based MPC problem into distributed problems, which exchange a certain number of samples with each other to compute local decisions.

This approach implements the same ATES well coupling constraints as the centralized formulation, using a hierarchical scheme; Rostampour and Keviczky (2017) showed this formulation to be practically equivalent to the CS coupling formulation. An upper control layer thus applies the coupling constraints to coordinate the operation of neighboring ATES systems, with weekly time steps and a 3-month prediction horizon. A lower layer then implements the same individual control problem for each building as in Rostampour and Keviczky (2016). This method can be applied efficiently for larger sets of agents, with computational runtimes scaling $\mathcal{O}(n)$ in proportion to the number of agents. This formulation (hereafter DSMPC) will therefore be used for the large-scale ATES simulation study.

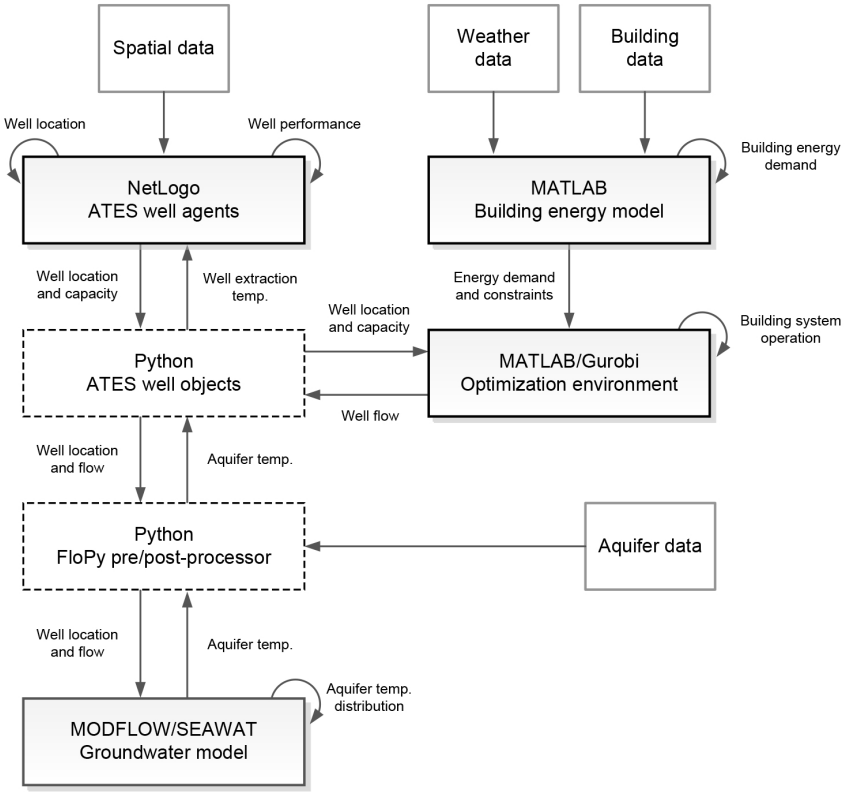


Figure 7.1: Simulation architecture

7.2.2. Coupled building/geohydrological simulation

This work relies on the coupled simulation architecture described in Chapter 3, which links an agent-based model of ATES planning in NetLogo (Wilensky, 1999), a geohydrological model of groundwater dynamics in MODFLOW/SEAWAT (Harbaugh, 2005; Langevin et al., 2008), and a building/control model in MATLAB. The three model components are linked through an object-oriented Python architecture, so that Python objects form the interface between the three models. Figure 7.1 illustrates the basic architecture and shows the data exchanges, which are facilitated by the FloPy pre/post-processor for MODFLOW/SEAWAT (Bakker et al., 2016).

The control formulations for the decoupled, centralized and distributed cases are implemented in MATLAB 2016a, using the YALMIP interface (Lofberg, 2004) with the Gurobi 8.0 solver (Gurobi Optimization, 2016). The controllers are simulated using given energy demand time series for the building agents, which are generated by a

stochastic version of the Low Energy Architecture (LEA) simulation model (detailed in van Vliet (2013); Ananduta (2016); Rostampour et al. (2017)). This energy balance model accounts for weather conditions, building characteristics, and occupancy patterns, and generates the heating and/or cooling demand profiles which are required to maintain a desired indoor temperature. In this application, time series for energy demand, and the corresponding control action for ATES pumping rates of each agent, are computed *ex ante* at an hourly resolution. The ATES pumping rates are then aggregated at a weekly scale and simulated as equivalent ATES well flows in the MODFLOW/SEAWAT groundwater model.

Climate properties

Changes in the energy demand of buildings over time are a key component of operational uncertainties for ATES. These changes may be caused by daily variations in weather conditions or building occupancy, but also by longer-term trends in climate which may eventually affect the balance between heating and cooling demand. To evaluate the performance of the CS formulation under variable building energy demand, a set of representative weather profiles was therefore derived from the four KNMI'14 climate scenarios (KNMI, 2014b).

These scenarios (G_L, G_H, W_L, W_H) are defined by moderate or high increases in global mean temperature (G and W respectively, with an increase of 1°C and 2°C by 2040), and low or high changes in air circulation patterns (L and H, respectively). The change in temperature under the G and W scenarios is roughly consistent with the standard RCP 4.5 and RCP 8.5 emissions scenarios (Van Vuuren et al., 2011).

In order to simulate building energy demand under these scenarios, a 3-year hourly time series was first obtained for observed air temperature (2011-01-01/2014-01-01) at the De Bilt weather station. In parallel, synthetic daily time series were generated using the KNMI time series transformation tool (KNMI, 2014a), to simulate air temperature at the same location under each of the climate scenarios. These series were generated for a 3-year period under reference conditions in 2010, and under scenarios for 2040. To simulate energy demand under these synthetic series at an hourly resolution, the observed hourly time series was then modified using a shifting/stretching method (Chan, 2011) to match the long-term trends described by the synthetic series. The method is described by Equation 7.2.1, where ΔTMAX_m, ΔTMIN_m and ΔT_m are the predicted changes in monthly maximum, minimum and mean temperature, using the synthetic series for 2040 relative to 2010. $(t_{0max})_m$ and $(t_{0min})_m$ are the observed monthly mean daily maximum and minimum temperatures, while t_0 is the observed hourly temperature time series. The scaled hourly temperature series t is then:

$$\alpha = \frac{\Delta\mathrm{TMAX}_m - \Delta\mathrm{TMIN}_m}{(t_{0max})_m - (t_{0min})_m}$$
$$t = t_0 + \Delta\mathrm{T}_m + \alpha \cdot (t_0 - (t_0)_m) \tag{7.2.5}$$

Following this method to generate hourly time series for each of the climate scenarios, Figure 7.2 presents the simulated 3-year temperature time series (left panel; smoothed over a 30-day moving average), and the corresponding average annual thermal imbalance towards heating for building 1 of the idealized case study, over the last two years of the simulated time series (right panel; defined as $(H - C)/(H + C)$, where H and C are the building energy demand for heating and cooling, respectively). This indicates that the climate scenarios shift building energy demand from primarily heating-driven (+8% imbalance under G_L), to cooling-driven (-5% under W_H).

These trends can be compared with typical year-to-year variations in ATES thermal balance which are caused by weather fluctuations. Willemsen (2016) presents results for a sample of 125 Dutch ATES systems over the 2010-2014 period, with the average ATES thermal balance showing a positive relationship with the deviation of heating degree-days from the average. As such, the average annual imbalance varied between -16% and +23%, following variations in degree-days from -29% to +20% of the average. While the magnitude of long-term climate trends may thus remain less significant than annual weather variations for ATES thermal balance by 2040, a consistent trend towards warming would make it more difficult to meet balance requirements. These trends should therefore be acknowledged in the planning and operation of ATES systems.

7.3. Case studies

7.3.1. ATES performance assessment

The simulation case studies use key performance indicators which build on an earlier assessment framework (Bloemendal et al., 2018), to evaluate the performance of the DS and DSMPC control methods from the perspectives of ATES system owners and policymakers. For the former perspective, three indicators will be used:

- The average thermal efficiency of ATES systems (η_{tot}), i.e. the fraction of injected thermal energy which is recovered from the subsurface over a given number of storage cycles;
- The average effective coefficient of performance (COP) of ATES systems, i.e. the ratio between the energy delivered from ATES systems to buildings over a given number of storage cycles, and the energy used to operate the ATES systems;

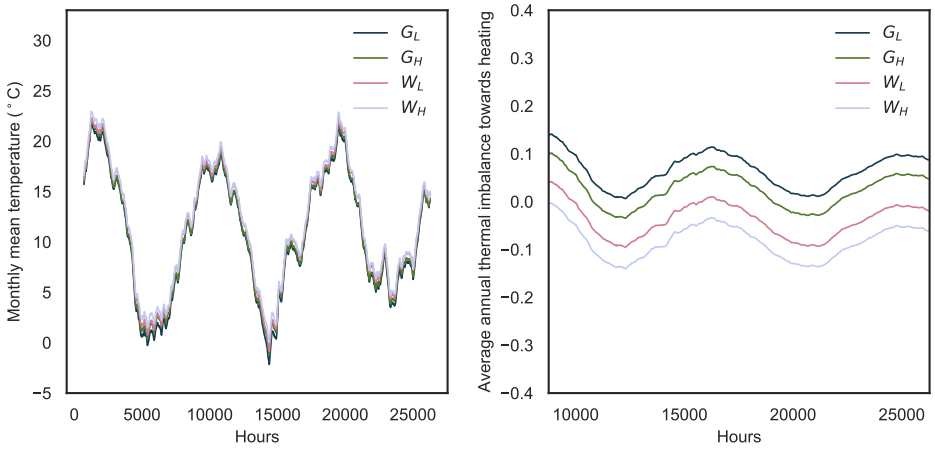


Figure 7.2: Left panel: Simulated time series for hourly average temperatures under each KNMI climate scenario at the De Bilt weather station, smoothed over a 30-day moving average. Right panel: Simulated annual thermal imbalance towards heating, under each KNMI climate scenario (Building 1 of the idealized case study)

- The average economic efficiency of ATES systems (νC), defined as specific energy cost savings per unit of water pumped by ATES, relative to a conventional building system delivering the same quantity of heating and cooling energy.

Three additional indicators will be used from the perspective of policymakers:

- The total GHG savings (Δ_{GHG}) obtained by ATES systems, relative to conventional building energy systems which would deliver equivalent heating and cooling energy;
- The average subsurface usage efficiency of ATES systems (ν_{GHG}), defined as the specific GHG savings obtained per unit of subsurface volume which is allocated for thermal storage;
- The average equivalent GHG abatement cost corresponding to ATES use (α_{GHG}), defined as the ratio between ATES operating costs compared to equivalent conventional building systems, and total GHG savings.

Details for the computation of each indicator are presented in Appendix C, using the parameters summarized in Table 7.1.

Parameter	Value or range	Unit	Symbol
ATES nominal temperature difference	6	[K]	ΔT
ATES pump efficiency	0.3	[-]	η_p
Boiler efficiency	0.9	[-]	η_b
Chiller coefficient of performance	3	[-]	COP_c
Heat pump coefficient of performance	4	[-]	COP_{hp}
Grid emission factor for electricity	0.157	$[tCO_2/GJ]$	f_e
Combustion emission factor for natural gas	0.056	$[tCO_2/GJ]$	f_g
Price for electricity	12–60	$[EUR/GJ]$	C_e
Price for natural gas	5–25	$[EUR/GJ]$	C_g

Table 7.1: Parameters used for the assessment of the case studies.

7.3.2. Idealized 3-building case

The first case study presents a simple test case for cooperative ATES control, which assumes a perfect exchange of information across three neighboring building agents to dynamically manage thermal interactions between ATES wells. To test the influence of the well coupling constraints, the CS MPC formulation is therefore compared to the decoupled DS formulation.

The building agents represent generic large office/commercial buildings which are typical of ATES usage in the city center of Utrecht. Their key parameters are summarized in Table 7.2, with further details in Ananduta (2016). Agents 1 and 3 are decomposed into sub-buildings in the building simulation model. These sub-buildings are assumed to share a single ATES system, so that each of the building agents operates one warm well and one cold well. The *ATES fraction* column indicates the approximate fraction of annual building energy demand which corresponds to the nominal ATES storage capacity, under the parameters used in the case study.

Building	Floor area $[m^2]$	Shell area $[m^2]$	Floors	ATES capacity $[m^3/yr]$	ATES fraction
Agent 1, building 1	4275	2664	4	107000	0.48
Agent 1, building 2	4275	5328	8		
Agent 1, building 3	2513	20991	29		
Agent 2	19125	14000	4	64000	0.4
Agent 3, building 1	2909	16612	22	129000	0.42
Agent 3, building 2	2909	16612	22		
Agent 3, building 3	2778	6641	9		
Agent 3, building 4	3889	7875	9		

Table 7.2: Building characteristics for idealized case study.

For this case, the control model is configured to prioritize ATES to fulfil heating and cooling demand, using the full nominal storage capacity of the ATES wells (within the coupling constraints used to manage thermal interactions in the CS formulation). Figure 7.3 shows a representative output of the building/control simulation for agent 3 over a two-year period, presenting the energy delivered by each supply type (smoothed

over a 24-hour window) and the corresponding volumes of water stored in the ATES wells. The controller thus uses the ATES wells until they are depleted, at which point the boiler and chiller provide heating and cooling.

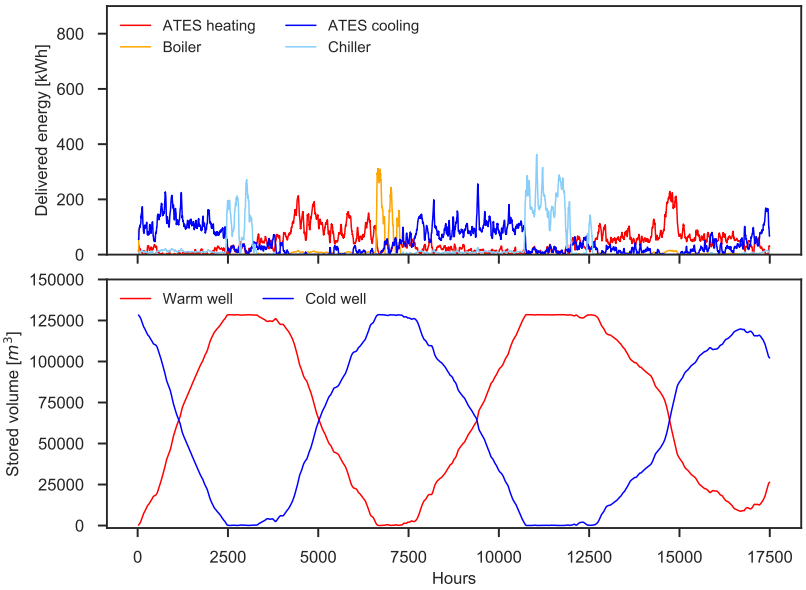


Figure 7.3: Representative output of the building/control model for agent 3.

The performance of ATES systems is then tested across four building energy demand profiles derived from the standard KNMI climate scenarios, and across six spatial planning policies representing the density at which neighboring ATES wells are laid out. The current layout guidelines as a reference (i.e. a distance of 3 R_{th}, where R_{th} is the average maximum expected thermal radius of neighboring wells based on permitted capacity). Denser policies are then tested at 2.75, 2.5, 2.38, 2.25 and 2.125 R_{th}. These different layout guidelines are illustrated in Figure 7.4, with the thermal diameters of the warm (*H*) and cold (*C*) wells of each agent. The spatial layout is represented in the NetLogo agent-based model on a 1000m x 1000m grid; this grid is used to compute spatial indicators such as the fraction of allocated subsurface volume.

For the CS control formulation, suitable coupling constraints are thus imposed for the well pairs *H1/C2*, *H2/C3*, and *H3/C1*.

Aquifer properties

The MODFLOW/SEAWAT model simulates three-dimensional aquifer flow and heat transport and heat recovery, using a finite-difference approach. This model is used to determine the thermal performance of the ATES wells by accounting for thermal

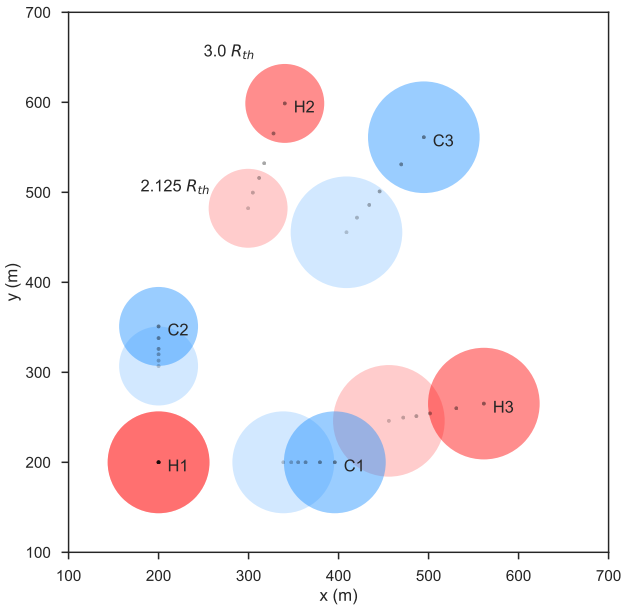


Figure 7.4: Spatial layout of ATES wells for the idealized case study.

and/or hydraulic interactions between neighbouring ATES systems, as well as heat losses between the storage aquifer and confining layers.

In this case study, the aquifer model is parameterized to represent typical conditions for ATES use in the Netherlands, based on (Bloemendal and Hartog, 2018). The model grid is discretized at a 5m resolution in the horizontal plane; this horizontal resolution is approximately 10% of the nominal thermal radii of the wells, within Sommer et al. (2015)'s recommendations for a reliable estimation of temperature fields. The ATES well screens are located in a single confined storage layer with a thickness of 15m. Input and output time series, such as ATES flows and extraction temperatures, are processed at a weekly time resolution following Bloemendal et al. (2018). It should however be noted that the transport processes are solved by SEAWAT at a smaller variable time step, based on a Courant number of 1.0.

To avoid boundary effects, the groundwater model extends 500 m beyond the boundaries of the NetLogo spatial grid, for total dimensions of 2000m x 2000m. The initial and boundary hydraulic heads were set to be uniform over the model grid, so that the model does not include ambient groundwater flow. Table 7.3 summarizes the other key model parameters.

Parameter	Value
Porosity	0.3
Horizontal conductivity	40 $[m/day]$
Vertical anisotropy	5
Longitudinal dispersion	1 $[m]$
Transversal dispersion	0.1 $[m]$
Bulk density	1890 $[kg/m^3]$
Bulk thermal diffusivity	0.16 $[m^2/day]$
Solid heat capacity	880 $[J/kg-K]$
Thermal conductivity of aquifer	2.55 $[W/m-K]$

Table 7.3: Aquifer characteristics for idealized case study.

7.3.3. Utrecht case

The large-scale case study represents a more realistic application for the dynamic management of thermal interactions between ATES systems, by *i)* testing a DSMPC control formulation which can be scaled more easily to multiple buildings, and *ii)* by simulating existing ATES systems and geohydrological conditions in the city center of Utrecht, in the Netherlands. For this case, the building/control models are configured to represent operational uncertainties through a variable demand for ATES storage. The effective maximum storage capacity is thus set by applying a multiplier Q to the nominal annual capacity, with $0.6 \leq Q \leq 1.1$. The value of this upper bound is set to maintain computational tractability: for larger values of Q, the pumping patterns of the ATES wells are primarily driven by the coupling constraints, significantly increasing runtimes under typical solution parameters for the Gurobi solver. Three spatial planning scenarios are used to compare different pathways for the future development of ATES in the area. Following an earlier case study presented in Bloemendal et al. (2017), these scenarios assume that ATES wells are built on 9 building plots on which ATES is currently used in the area, starting from a set of 82 currently active or planned wells. The NetLogo agent-based model is used to locate these wells in a 2000m x 2500m grid, on which GIS data is overlaid to generate exclusion areas corresponding to roads, buildings, and waterways. Scenario 1 represents future development under current layout guidelines (i.e. well distances of 3.0 R_{th}). Scenarios 2 and 3 then simulate revised, denser layout guidelines of 2.5 R_{th} and 2.25 R_{th}, respectively. In all scenarios, new wells are located in available development areas as long as sufficient space is available under the layout guidelines. These scenarios are illustrated in Figure 7.5. The dashed lines between wells illustrate the coupling constraints for the DSMPC formulation, which are in this case applied to wells of opposite types with a relative distance below 2.75 R_{th}. This distance approximately represents the maximum distance at which thermal interactions can be expected to be significant. Inset plots show a bivariate Gaussian kernel density estimate for the distribution of nominal well capacities in each scenario (distinguishing between existing or planned wells, and simulated new wells), and for the distribution of

the minimum relative distance to any neighboring well. For instance, this indicates that the simulated new wells in all scenarios have capacities below 100,000 m^3/year, as there is already a lack of space to accommodate larger wells on existing building plots. In parallel, the new wells built in Scenario 3 all have minimum distances of approximately 2.25 R_{th} with other wells, while the sparser guidelines in Scenario 1 lead to a broader distribution on this indicator; we can for instance expect that wells which are not within 3.0 R_{th} of other wells would not be significantly affected by the dynamic management of thermal interactions. Given that the simulated new wells in Scenarios 2 and 3 are located under different layout guidelines than the existing or planned wells, the management of thermal interactions is likely to impact these groups of wells in different ways – potentially leading to a different distribution of benefits across incumbent ATES users and new adopters. The case study will thus use a simple game-theoretical approach, to test the combinations of ATES usage and energy prices under which cooperation (i.e. participation in an information exchange scheme) would be a Pareto-optimal Nash equilibrium. This analysis assumes that the ATES wells are grouped into "incumbent" and "new" systems.

Aquifer properties

This case study relies on a 3000m x 3000m cutout of the Hydromedah regional groundwater model for the area of Utrecht (Borren, 2009; Gunnink, 2004); this model had previously been adapted to include ATES wells, as detailed in (Bloemendal et al., 2017). As such, the grid was rediscretized in the horizontal plane to refine cells around the ATES wells, with a grid size varying from 8m at the center of the wells, to 16m at the border of the model; the corresponding arrays for horizontal conductivity and groundwater head were rediscretized using bilinear interpolation. The ATES wells are located in a confined layer with an average thickness of 26m. In addition, standard MT3DMS packages were parameterized using the same assumptions as the idealized aquifer model, in order to include relevant transport processes. For the purposes of this work, the Utrecht case presents two key differences from the idealized aquifer model: the ambient groundwater flow (approximately 10 m/y) may influence recovery performance, while the greater thickness of the aquifer reduces the "footprint" of the wells in the horizontal plane and its variation over storage cycles, so that the effect of well couplings may be smaller.

7.4. Results

7.4.1. Idealized case with centralized control

System performance

To illustrate the impact of the well coupling constraints, the left graphs of Figure 7.6 first present the sum of the hydraulic radii for each coupling (*H1/C2*, *H2/C3*, and *H3/C1*), for three simulated annual storage cycles under the G_L climate scenario and the 2.125 R_{th} spatial layout policy. The dashed lines show the distance $\theta d_{i,j}$ which

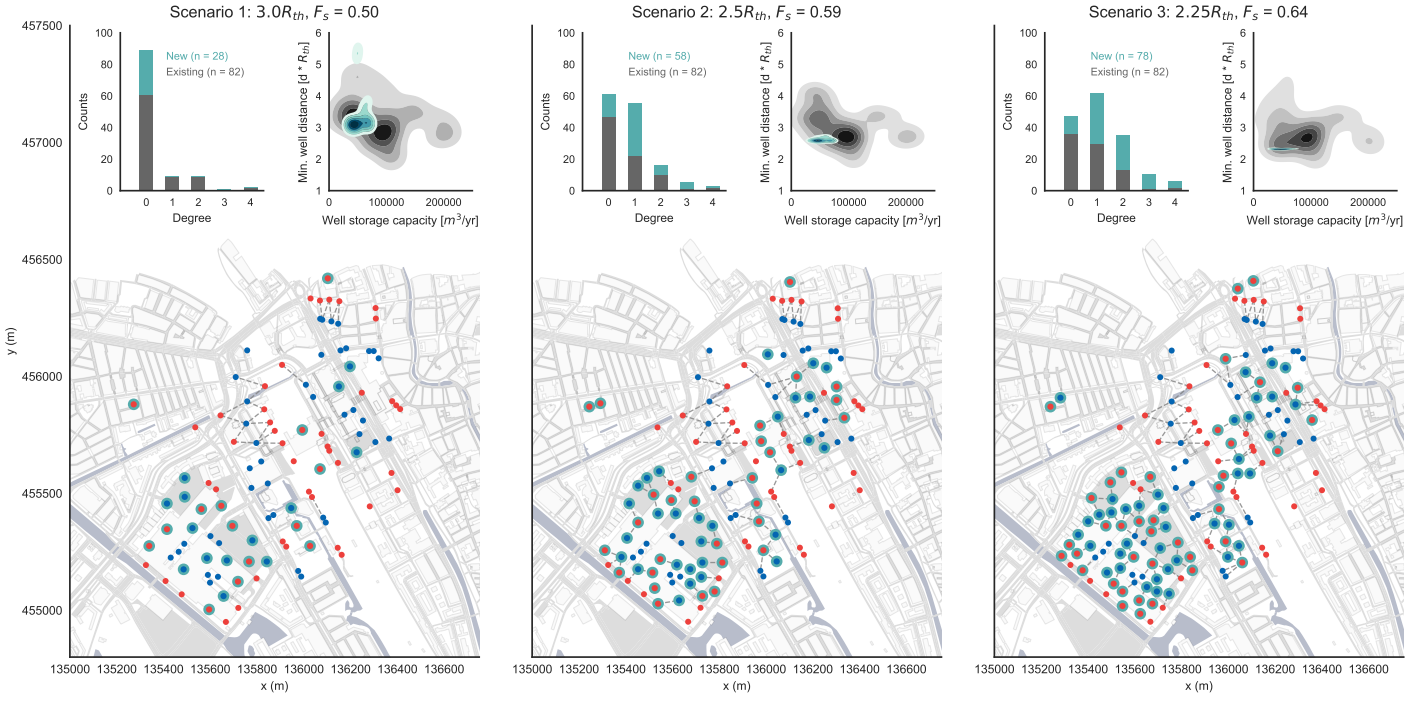


Figure 7.5: Spatial planning scenarios for Utrecht case study.

corresponds to each coupling constraint in the CS formulation, taking $\theta = 0.78$.

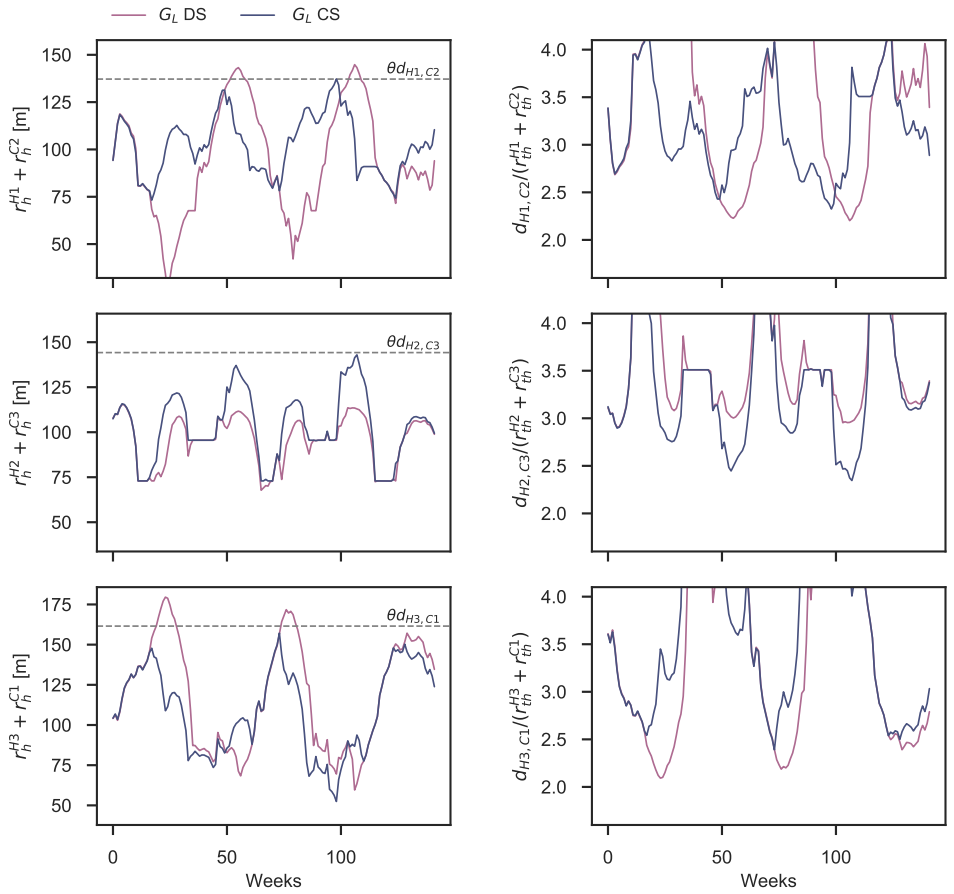


Figure 7.6: Left panels: sum of the hydraulic radii for each well coupling. Right panels: equivalent relative distance between the coupled wells, as a multiplier of their average thermal radius.

For the decoupled DS formulation, the sum of the coupled radii exceeds $\theta d_{i,j}$ for the first and third couplings (*H1/C2*, *H3/C1*), while the wells corresponding to the second coupling are used relatively less. The CS formulation meets the constrained distance as expected, and increases the volume stored in the *H2/C3* pair. The right graphs present the effect of the constraints on the equivalent relative distance between the wells. For the first and third couplings, the DS formulation yields minimal values which are consistent with the spatial layout policy, but the lower usage of the second coupling results in higher effective distances. The CS formulation again yields a more

equal distribution of minimal distances across the couplings; the choice of the θ value limits the effective minimal distance to approximately 2.5 R_{th}, which is an acceptable threshold to avoid thermal interferences (Bloemendal et al., 2018). Figure 7.7 focuses on the impact of the climate scenarios on ATES operation, only showing the CS formulation for clarity. The left graphs again present the sum of the coupled hydraulic radii for each coupling; for each climate scenario, the coupling constraints are met within a tolerance of 3%. This tolerance is related to the choice of a reliability level for the solution of the constraints, which entails a trade-off between computational cost and reliability. Given that the constraint is only exceeded over a maximum of two timesteps in any of the scenarios, this reliability level can be considered to be acceptable. In parallel, the right graphs show the volume of cold water stored by each agent. The climate scenarios particularly have an impact on the water stored by Agent 1: the cold well is depleted more quickly over the first storage cycle in scenarios W_L and W_H, which increase cooling demand; conversely, the cold well is replenished more slowly in winter conditions, due to lower heating demand. The climate scenarios have a smaller effect on the other agents, as the relatively smaller capacity of their ATES wells (relative to building energy demand) ensure that the wells are depleted more quickly.

Figure 7.8 summarizes the key indicators for average ATES performance as a function of spatial layout, across each climate scenario and control formulation. The choice of $\theta = 0.78$ means that the CS formulation only shows minimal differences in thermal efficiency with the decoupled DS case for $R_{th} \geq 2.75$; however, at smaller distances, the coupling constraints largely stabilize thermal efficiency until 2.25 R_{th}, while the DS formulation presents significantly reduced efficiency due to the development of thermal interactions. Due to computational limitations, the values are presented after two annual storage cycles only. However, as shown in D.1 for a subset of scenarios, this time horizon is sufficient to stabilize the relative performance trends for each control formulation. The average COP and cost savings both slightly increase until 2.25 R_{th} in the CS formulation, as the coupling constraints lead to shifts in ATES use over time and more efficient pumping schedules. In the DS formulation, both indicators present a downward trend, similarly to thermal efficiency. Appendix D presents these results as "regret" values, relative to the DS formulation with $R_{th} = 3.0$.

Collective performance

Figure 7.9 presents the key collective performance indicators, in the form of total and specific GHG savings. As with COP and cost savings, the more efficient pumping schedules under coupling constraints yield higher total GHG savings. With both control formulations, specific GHG savings improve significantly at smaller well distances, primarily due to the decrease in allocated subsurface volume; the CS formulation provides additional benefits by maintaining system performance at smaller well distances.

Based on these results, the CS formulation could thus support a significant im-

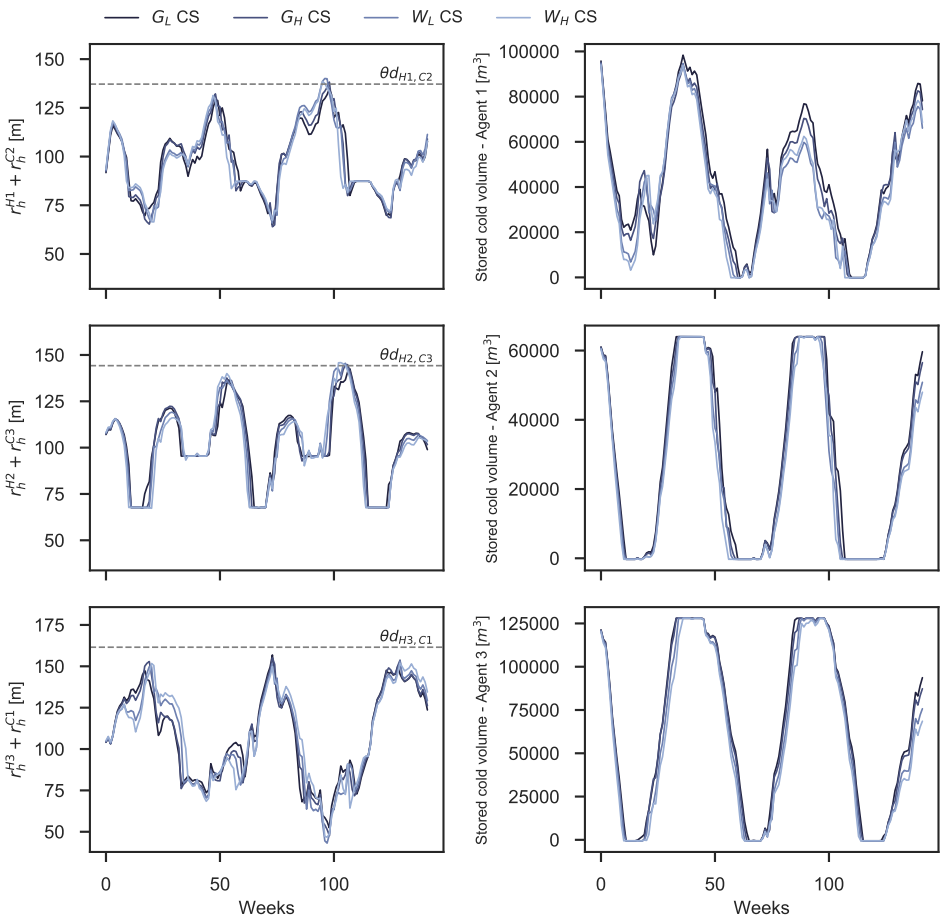


Figure 7.7: Left panels: sum of the hydraulic radii for each well coupling, across climate scenarios. Right panels: volume of water stored in each agent's cold well.

provement in the efficiency of ATES spatial planning: relative to current planning guidelines, specific GHG savings increase by over 75% at $R_{th} = 2.25$ by reducing allocated subsurface volume – without significantly affecting thermal efficiency, and potentially improving COP and cost savings.

7.4.2. Utrecht case with distributed control

System performance and distributional effects

Figure 7.10 presents the average system performance indicators for each spatial planning scenario, as a function of the total volume of water pumped by ATES systems. The DS formulation yields a consistent trend, in which performance tends to decrease

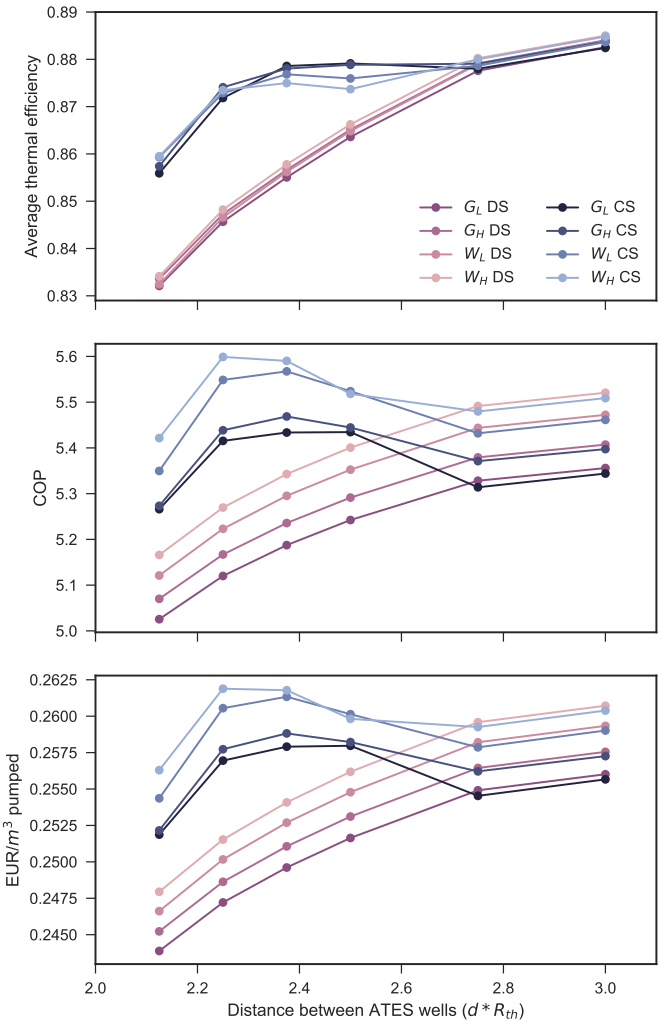


Figure 7.8: ATES performance indicators for idealized case, as a function of spatial layout policy: average thermal efficiency (top panel), coefficient of performance (middle panel), and cost savings relative to conventional energy per pumped unit of water (bottom panel).

within each planning scenario (i.e. with an increase in the allowed pumped fraction Q), and across the planning scenarios (i.e. with an increase in well density). In parallel, the total pumped volume of water increases in proportion to the allowed pumped fraction Q, as expected.

The DSMPC formulation presents significantly different behavior: the total pumped

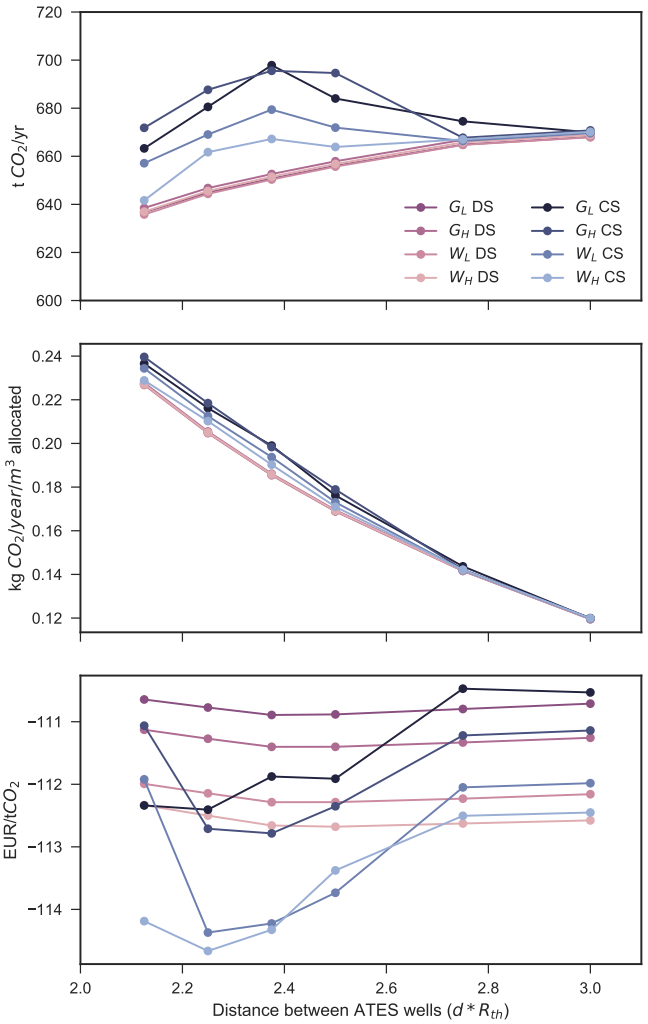


Figure 7.9: Collective performance indicators for idealized case, as a function of spatial layout policy: total annual GHG savings (top panel); specific annual GHG savings per allocated unit of subsurface volume (middle panel); equivalent GHG abatement cost (bottom panel).

volume of water tends to saturate at $Q = 1.0$, then drops for $Q = 1.1$. This is accompanied by an improvement in the system performance indicators. This behavior can be explained by the decrease in usage of wells which are subject to multiple coupling constraints with neighboring wells, and which typically have a lower thermal efficiency due to this density; with an increase in allowed pumped capacity, these "marginal"

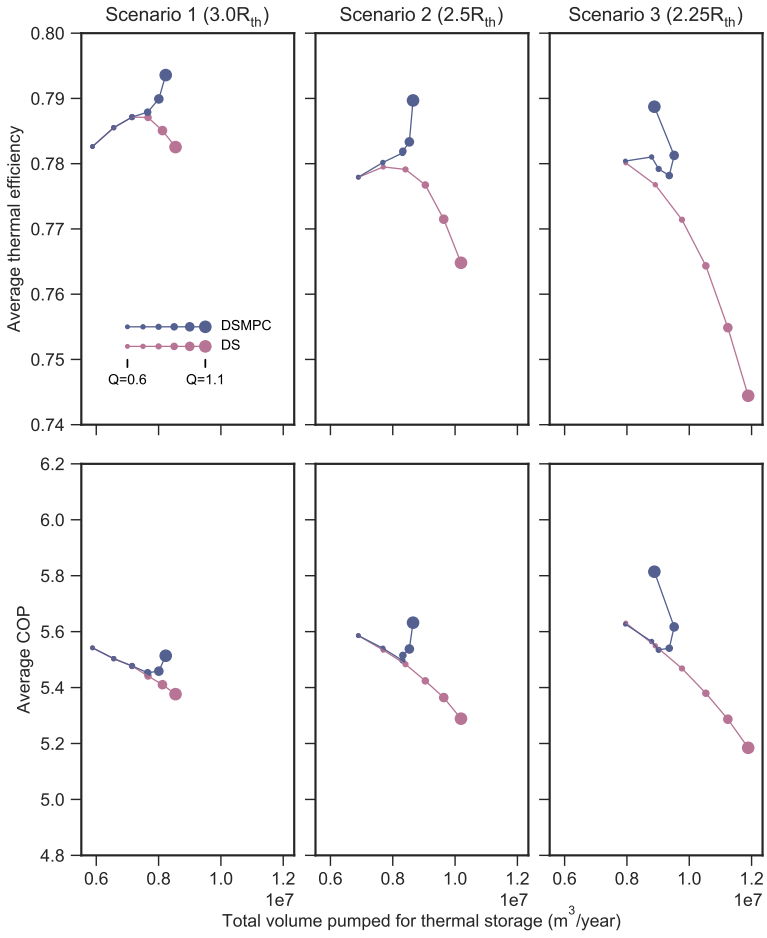


Figure 7.10: ATES performance indicators for Utrecht case, for each spatial planning scenario: average thermal efficiency (top panel) and coefficient of performance (bottom panel).

wells tend to be used less in order to meet the coupling constraints – thus increasing average performance. Figure D.2 in Appendix D details this effect, by plotting the relative usage of the well pairs simulated in Scenario 3, as a function of their decoupled thermal efficiency. Less efficient wells thus tend to be used less when applying the coupling constraints. For values of $Q = 1.0$, the performance of the systems typically remains at least equal to performance at $Q = 0.6$; Table D.2 summarizes these results, computed as "regret" values relative to the DS formulation in Scenario 1 (i.e. $R_{th} = 3.0$). However, assumptions on energy prices may further complicate the situation; economic performance does not always exactly correlate with thermal

efficiency, depending on the relative costs of energy for heating or cooling. As such, whereas high gas prices directly increase cost savings from ATES relative to a conventional boiler, electricity prices have a more complex effect, by affecting the operating cost of ATES well pumps and the building heat pump, as well as cost savings relative to a conventional chiller. Figure 7.11 presents the relative economic performance of different control/layout combinations, based on the specific energy cost savings per unit of water pumped by ATES in each case . For Scenario 3, this implies that the DSMPC formulation would outperform the DS formulation across all tested energy price combination, which are based on typical non-household energy prices for electricity and natural gas in the European Union (Eurostat, 2018a,b)). However, energy price combinations which tend to make ATES more relatively profitable overall (i.e. high gas prices and low electricity prices) would slightly penalize the relative performance of the DSMPC/Scenario 3 combination, compared to "business-as-usual" development (DS, Scenario 1). Notably, in conditions where ATES would be less profitable overall, the DSMPC formulation retains a relative advantage due to more efficient pumping schedules.

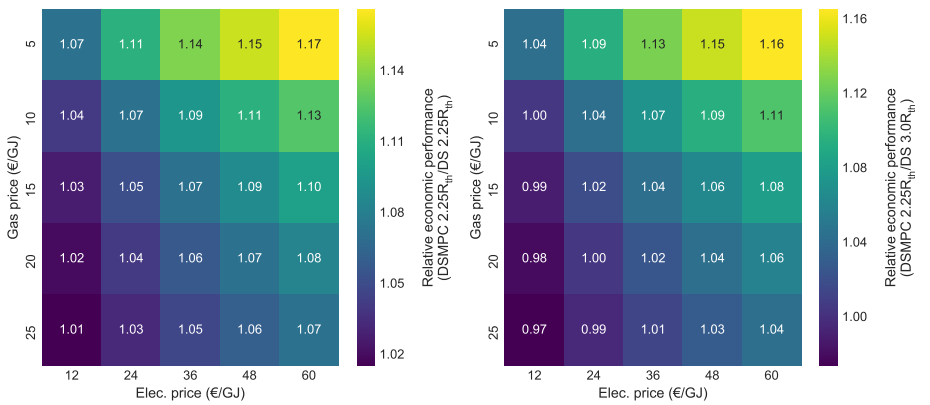


Figure 7.11: Relative economic performance across different control/layout combinations. Left panel: DSMPC compared to DS, for Scenario 3 ($2.25R_{th}$). Right panel: DSMPC for Scenario 3, compared to DS for Scenario 1.

The analysis has so far considered average results over the full set of active wells. However, Scenarios 2 and 3 in particular may lead to different benefits for existing (or "incumbent") wells, and simulated wells created under different layout guidelines. In this situation, new and incumbent wells may decide to partially cooperate, i.e. to exchange information across their subset of wells only. To evaluate the conditions under which these two groups of users could be assumed to have an incentive to cooperate in exchanging information, Figures 7.12 and 7.13 thus present the effect of different

plausible combinations of energy prices and Q values, on the average cost savings of each group of users for Scenario 3. The markers present four possible courses of action: *full cooperation* (i.e. the DSMPC formulation), *only incumbent wells coupled*, *only new wells coupled*, and *no cooperation* (i.e. the DS formulation). The energy cost savings for each action are then considered as payoffs in a 2-player game; shaded subplots indicate that full cooperation is not a Nash equilibrium in a given combination.

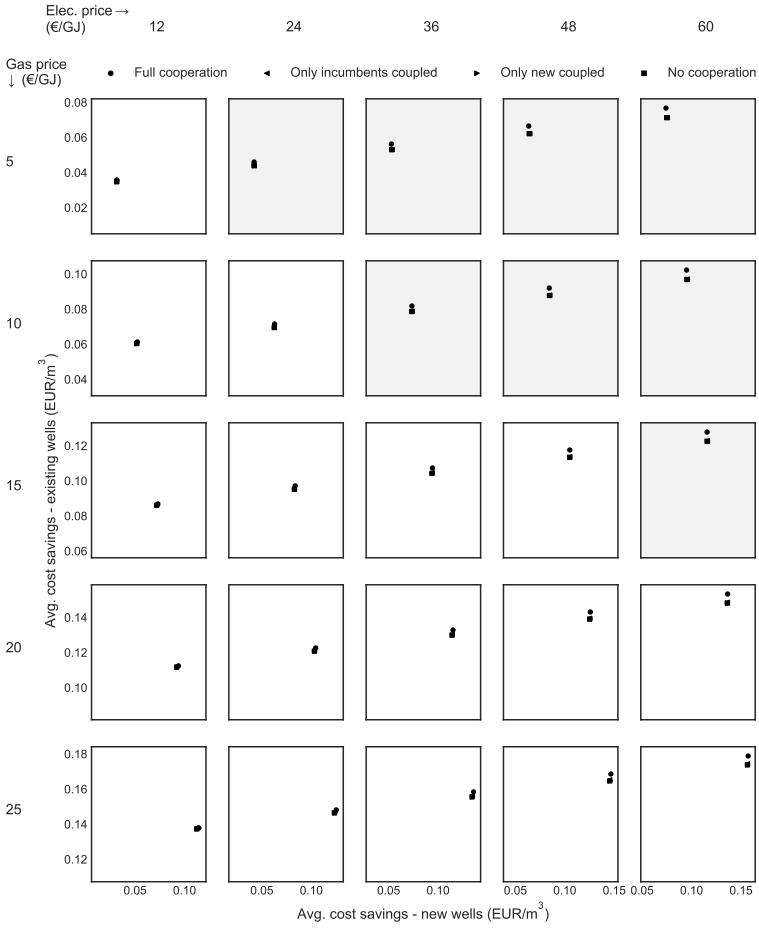


Figure 7.12: Game-theoretical analysis for cooperation between incumbent and new wells, as a function of gas and electricity prices, for $Q = 0.7$ and Scenario 3. Shaded subplots indicate that full cooperation is not a Nash equilibrium in a given combination of energy prices.

Figure 7.12 thus indicates that relatively high electricity prices may prevent cooperation from being a Nash equilibrium, when combined with a lower ATES usage ($Q = 0.7$) and low gas prices; this combination makes ATES relatively less economi-

cally attractive. In this situation, incumbent wells would benefit from full cooperation, while the Pareto-optimal decision for new wells would be to only partially cooperate within their own subset. However, for $Q \geq 1.0$, the greater overall cost savings from a larger usage of ATES then yield a cooperative Nash equilibrium in all of the energy price combinations, as illustrated in Figure 7.13.

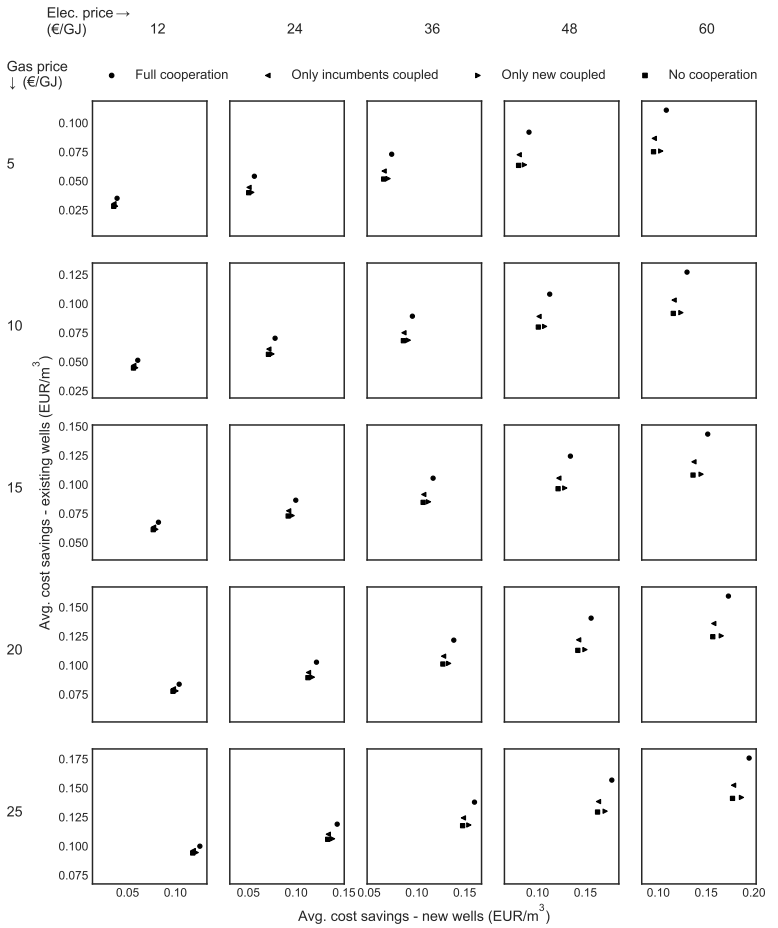


Figure 7.13: Game-theoretical analysis for cooperation between incumbent and new wells, as a function of gas and electricity prices, for $Q = 1.1$ and Scenario 3.

Collective performance

Given that GHG savings are largely driven by the total pumped volume, a similar pattern holds for the GHG indicators shown in Figure 7.14 as for system performance. As such, in the DS formulation, the value of both indicators increases monotonically with

the allowed pumped fraction Q, but reaches a maximum at $Q = 1.0$ for the DSMPC case.

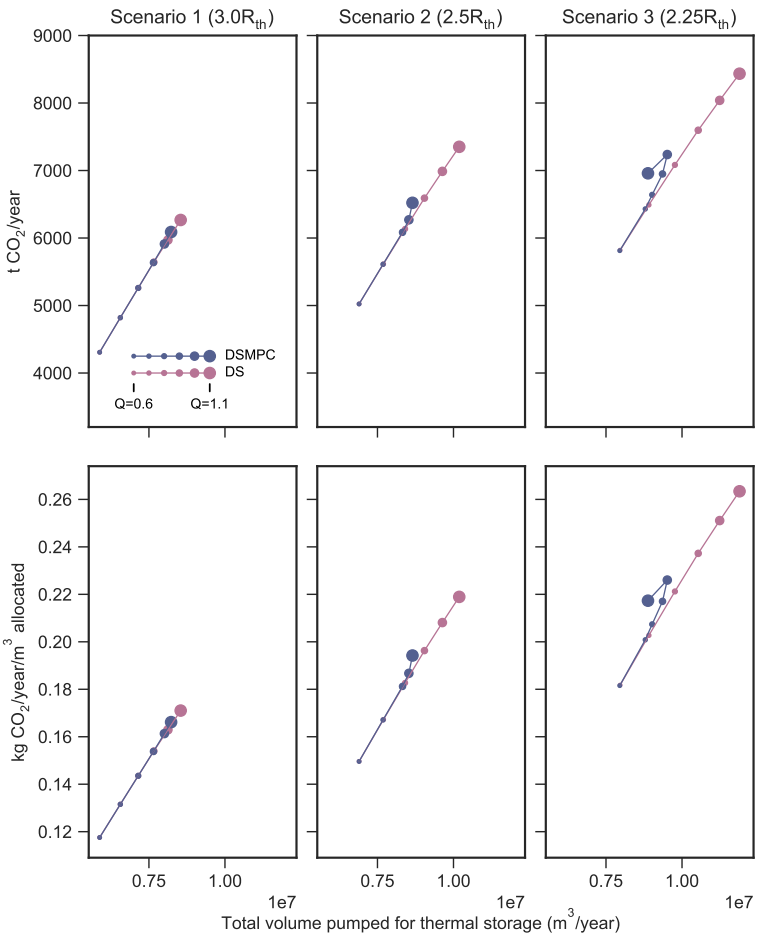


Figure 7.14: Collective performance indicators for Utrecht case, for each spatial planning scenario: total annual GHG savings (top panels); specific annual GHG savings per allocated unit of subsurface volume (bottom panels).

As indicated in the tabular results shown in Appendix D, the increase in specific GHG savings remains significant for denser layout guidelines (e.g. 37% with the DSMPC formulation in Scenario 3), but smaller than for the idealized case presented in supplementary material. The highest specific GHG savings are obtained with the DS formulation in Scenario 3, which maximizes the total pumped water volume; however, the associated decrease in system performance would likely be unacceptable for ATES

owners.

This trade-off can be expressed through the equivalent marginal GHG abatement cost, which relates energy cost savings and GHG savings. The left panel of Figure 7.15 presents the GHG abatement cost for the DSMPC formulation in Scenario 3; the negative values indicate that the additional development of ATES which would be allowed in this case would nonetheless yield additional cost savings. In parallel, the right panel compares the DSMPC formulation in Scenario 3, with the "business-as-usual " case (DS, Scenario 1). As previously shown in Figure 7.11, a higher gas price combined with a relatively lower electricity price would tend to make the DSMPC approach relatively less economically attractive. However, given the overall improvements in GHG savings which would be supported by this approach, the equivalent GHG abatement cost is relatively low; this cost is for instance below typical carbon prices for the European Union Emissions Trading System for the 2017-2018 period. Under the assumptions used to parameterize the models, this implies that denser ATES development with a DSMPC approach would remain an economically attractive GHG abatement option for policymakers, even under an unfavorable combination of energy prices. Furthermore, the opposite combination of energy prices (i.e. low gas price and high electricity price, which tends to make ATES more economically sensitive to pumping schedules) would favor the DSMPC approach, yielding negative marginal abatement costs.

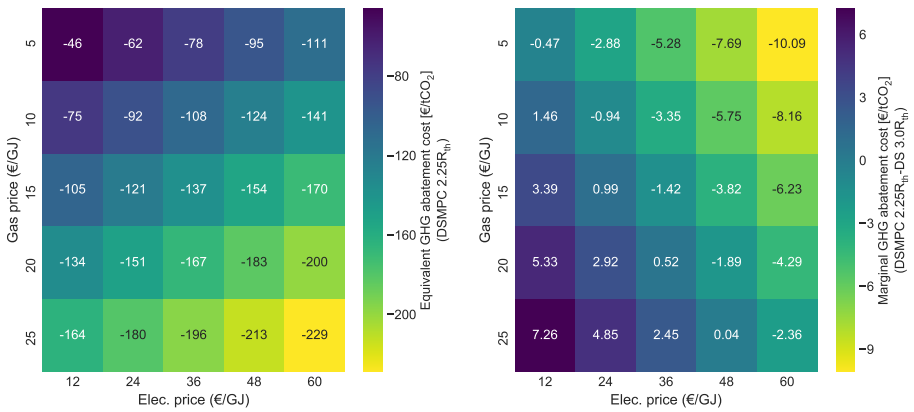


Figure 7.15: Equivalent marginal GHG abatement cost across different control/layout combinations. Left panel: DSMPC for Scenario 3 ($2.25R_{th}$). Right panel: relative GHG abatement cost for DSMPC in Scenario 3, compared to "business-as-usual" development (DS, Scenario 1).

7.5. Discussion

The DSMPC results for the Utrecht case study can first be compared with the results of Chapter 6, which simulated an "adaptive" permitting method in which the permitted storage capacity of wells (and, consequently, their allocated footprint) was periodically adjusted based on actual usage. This feedback mechanism could help counteract the over-allocation of subsurface space under conditions where ATES systems would only use a fraction of their permitted storage capacity, as unused subsurface volume would then become available for additional systems; the adaptive permits therefore allocated subsurface volume more efficiently than a baseline static planning approach. Figure 7.16 combines results from this earlier case study with results from this work, focusing on specific GHG savings only; the earlier case study used a different representation of building system operation, so that the ATES system performance indicators would be difficult to compare on an equal basis.

The two boxplots on the left illustrate the full range of model output for the previously simulated adaptive permit scenarios, which used a broader parameterization of 512 scenarios to include technical uncertainties such as ΔT and heat pump COP. Due to the computational cost of the detailed building models used in this chapter, simulating similar ensembles would not have been practical, so that these technical uncertainties were here left out of scope. The gray markers illustrate parameterizations within these ensembles which were consistent with the assumptions of this work, with $0.6 \leq Q \leq 1.0$. With the adaptive permits, specific GHG savings are relatively more stable across different values for Q, as the allocated subsurface volume is adjusted consequently to the used storage fraction. However, for $Q = 1.0$, the DSMPC formulation for Scenario 2 ($2.5R_{th}$) yields an additional improvement of 9% on this indicator, or 27% for Scenario 3 ($2.25R_{th}$), while better maintaining system performance.

It can be noted that the average system performance values for the Utrecht case study were significantly lower than for the idealized case. This can be attributed to a combination of ATES well design parameters as well as aquifer properties such as ambient groundwater flow, which reduces recovery efficiency due to advection losses; the formulation of the coupling constraints could be extended by adding a time-dependent term to represent this effect, for instance using the analytical expression provided by Bloemendal and Hartog (2018). In parallel, the dynamic management of thermal interactions is likely to offer greater benefits with relatively shorter well screen lengths (such as the model used in the idealized case study). In these conditions – and with the assumption of a cylindrical storage volume – seasonal pumping patterns will intuitively yield larger variations in the hydraulic radii of storage wells. The dynamic management of these radii through well coupling constraints would then in turn allow for more precise management of thermal interactions between neighboring thermal volumes, as the magnitude of the changes in stored radii through direct transport (which can be directly controlled) would be relatively greater in relation to the magnitude of conduction

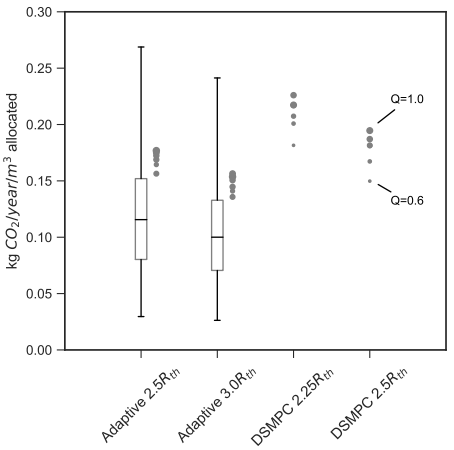


Figure 7.16: Specific GHG savings for Scenarios 2 and 3 of the Utrecht case study, compared with simulated adaptive permits in Bloemendal et al. (2017). The gray markers illustrate comparable parameterizations in each case study.

and dispersion processes between the stored thermal volume and the ambient aquifer medium (which are largely uncontrolled, and driven by aquifer properties). Due to the relatively larger resulting footprint of wells in the horizontal plane, shallower aquifers are also more likely to experience a scarcity of space for new ATES systems, making the management of interactions particularly relevant in these cases.

More broadly, although the case studies represented different plausible settings for the dynamic management of thermal interactions between ATES doublets, well design characteristics can vary significantly in practice; Figure 7.17 compares the characteristics used in the case studies with a dataset of 331 Dutch ATES wells studied by Bloemendal and Hartog (2018). As shown in the left panel, the simulated wells tended to be relatively small, in terms of annual storage capacity as well as screen length; the right panel presents the distribution of the ratio between the well screen length and thermal radius (L/R_{th}), which is a key design attribute. Doughty et al. (1982) presented this indicator as a proxy for thermal recovery efficiency, recommending a range of $1 \leq L/R_{th} \leq 4$. Bloemendal and Hartog (2018) similarly suggested a range of $0.5 \leq L/R_{th} \leq 3$ in the absence of ambient groundwater flow, or a narrower range of $0.6 \leq L/R_{th} \leq 1.2$ for an ambient flow of 10 m/year representative of the Utrecht case. These ranges aim to minimize losses from conduction, dispersion, and advection.

While the simulated wells in the Utrecht case are reasonably consistent with this recommendation, the L/R_{th} ratio for wells in the idealized case is below the recommended optimal range, and would be around the 25th percentile of the reference dataset. A smaller ratio will tend to increase thermal losses between the stored thermal

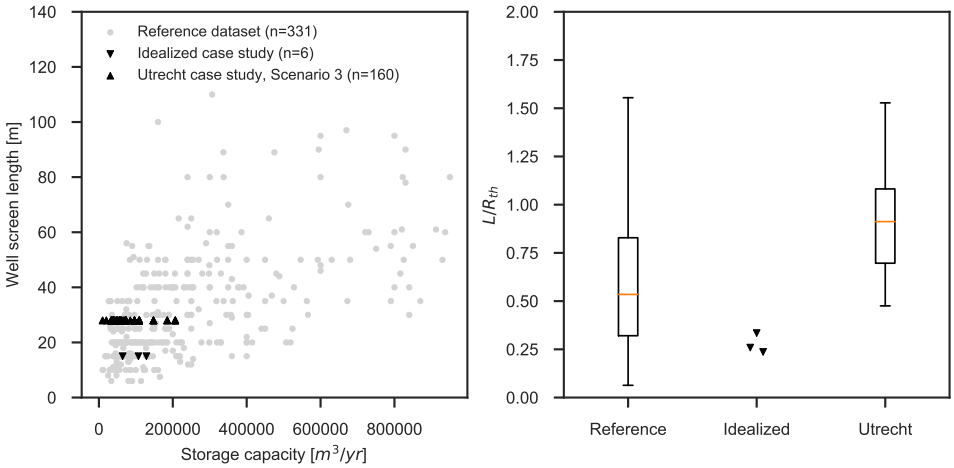


Figure 7.17: Well characteristics as simulated in the idealized case study and Utrecht case study, compared to the reference dataset used in Bloemendal and Hartog (2018). Left panel: well screen length vs. nominal storage capacity; right panel: distribution of the L/R_{th} ratio across the wells in each sample.

volumes and the ambient aquifer: these losses are typically dominated by conduction, which can be reduced by minimizing the surface area of the stored thermal cylinder relative to its volume (Doughty et al., 1982). An optimal value for this relation can be found at $L/R_{th} = 2$. Below this optimum, a smaller L/R_{th} ratio yields a larger ratio between the area and volume, increasing conduction losses. While the simulated thermal efficiency of the wells remained relatively high in the idealized case study, this has additional implications in the context of a DSMPC approach: the dynamic management of thermal interactions is likely to be more useful with smaller L/R_{th} values, which increase the horizontal footprint of the wells and yield relatively greater variations in thermal radii. However, this may require a compromise with conduction losses.

7.6. Conclusions

Current methods for the planning and operation of ATES systems lead to an inefficient trade-off between private and public interests, by limiting the deployment of the technology – and thus energy savings – in the dense urban areas which account for a growing portion of energy use in the built environment. This situation is motivated by the risk of a "tragedy of the commons" which could be caused by uncontrolled thermal interferences between ATES systems sharing an aquifer. As a starting point towards an improved management regime which could resolve this trade-off, this work assessed an approach based on the distributed control of ATES systems, in which information

exchange would support the dynamic management of thermal interactions between neighboring ATES systems. Compared to a reference scenario (i.e. decoupled operation under current planning guidelines of $3.0R_{th}$), information exchange combined with denser layout guidelines of $2.25R_{th}$ improved specific GHG savings by 75% for an idealized case study, and by 38% for a case study of the Utrecht city center. In both cases, system performance (as measured by thermal efficiency and effective COP) was comparable or greater than in the reference scenario.

A coordinated approach to ATES operation could therefore restructure the trade-off between private and public interests: under plausible operating conditions, the exchange of information across ATES systems could lead to a "win-win" situation for policymakers and operators, by increasing collective GHG savings without penalizing economic performance, and without introducing new administrative requirements (which would for instance be required for an adaptive permitting strategy). This approach could eventually support a self-organized approach to the management of ATES systems: by relying on coordination to avoid the adverse impacts of denser development and to adapt to changes in the operation of systems (such as thermal imbalances or a partial use of the planned capacity), provincial or municipal authorities could streamline the current permitting and reporting process, and delegate part of their management authority to local clusters of ATES users.

However, this approach would lead to a different compromise, under which ATES operators would trade off the implicit value of information about their use of ATES and other building energy systems. The privacy implications of smart energy systems have drawn increased scrutiny in the literature (Milchram et al., 2018; McKenna et al., 2012; Véliz and Grunewald, 2018; Kabalci, 2016); in the case of industrial energy users, thermal demand profiles could for instance be used to infer sensitive information about production processes Samad and Kiliccote (2012). Similarly, in the case of residential users, aggregation across multiple sources and levels of energy usage may make it impossible for participating individuals to offer genuinely informed consent towards the use of their data (Véliz and Grunewald, 2018). Although dedicated research is needed to assess these issues in the specific context of ATES, we note that the DSMPC approach is entirely compatible with differential privacy methods, under which the required information is pre-processed to maintain a level of privacy for the participating agents (Rostampour et al., 2018a). This would then yield a new trade-off between the bandwidth of exchanged information, and its reliability towards the management of thermal interactions.

The simulated case studies covered a range of well configurations which adequately represents current practice in the Netherlands, as defined by the L/R_{th} ratio, so that the DSMPC approach should be applicable in most geographic areas which are otherwise suitable for ATES. However, additional work would be useful to explore the potential trade-off between conduction losses and thermal radius management which

is driven by the L/R_{th} ratio, and to simulate DSMPC management with larger systems (in terms of storage capacity and well screen length) which would be more representative of ATES development in thicker aquifers. In such cases, the DSMPC approach – which implies the use of doublet wells – should also be compared to the use of monowells, which were not applicable in the simulated case studies due to the relative shallowness of the aquifers.

Finally, it should be emphasized that taking advantage of the dynamic management of thermal interactions will also require revised, denser spatial planning policies: the current guidelines used to plan ATES systems in the Netherlands are effective at avoiding thermal interactions between systems, which implies there would be little benefit in their management. This was supported by the results of both simulated case studies, in which a coordinated approach showed limited gains under current layout guidelines of $3.0R_{th}$; layout guidelines of $2.25R_{th}$ would instead enable denser development and would suitably complement coordinated operation. Crucially, the current rapid deployment of ATES provides a window of opportunity to apply such improved methods for the planning and operation of systems, as policymakers and ATES operators may otherwise become locked into suboptimal practices.

8

Assessing the worldwide potential of Underground Thermal Energy Storage (UTES) for energy savings

This chapter is based on Jaxa-Rozen, Bloemendal, and Kwakkel (2018), with changes to the introduction and problem background to avoid overlap with earlier chapters of the thesis.

8.1. Introduction

Underground Thermal Energy Storage (UTES) is a set of building technologies which can be used to seasonally store thermal energy in the subsurface. Variants of UTES include Aquifer Thermal Energy Storage (ATES), which relies on natural aquifer formations to store water at different temperatures, and Borehole Thermal Energy Storage (BTES), in which boreholes are drilled in soil or rock to store heat through a working fluid. In combination with a heat pump, UTES can reduce energy use for heating and cooling by more than half in larger buildings (Tomasetta et al., 2015), while supporting the electrification of building energy systems. These technologies are thus increasingly popular in Northern Europe. Furthermore, the climactic and subsurface conditions required for a technically viable use of UTES can be found across Europe, Asia and North America.

Buildings currently account for approximately one-third of total global final energy use (Lucon et al., 2014), so that technologies such as UTES can have a major impact towards reductions in worldwide energy use and greenhouse gas (GHG) emissions. This is compounded by economic and demographic trends, such as urbanization and shifts away from informal housing in developing countries; these trends have significantly

increased construction activity, and may eventually double or triple global energy use from buildings by the middle of the century – yet they also open a window of opportunity for energy-efficient technologies in new buildings.

Although energy use from ground-source heat pumps and BTES have been the focus of several regional reviews (e.g. Bayer et al., 2012; Blum et al., 2010; Ni et al., 2015), the long-term worldwide potential of ATES and BTES for urban energy savings has not yet been evaluated in the literature. Given that UTES performance is highly site-specific and sensitive to climate conditions, this requires a spatially-explicit analysis which accounts for local properties, as well as plausible future changes in operating conditions. This chapter therefore aims to synthesize existing data sources for climate, building, and subsurface properties, to estimate the long-term energy savings which could be achieved from ATES and BTES across 556 urban areas worldwide at the 2050 horizon. The analysis builds on an assessment of ATES suitability presented by Bloemendal et al. (2015); this earlier work constructed a subsurface suitability index for ATES, and used it to estimate the global applicability of ATES based on the fraction of urban population found in suitable areas. In this work, we combine subsurface suitability with a broader set of scenarios for climate and building energy use, in order to translate technical suitability into plausible urban energy savings for ATES as well as BTES. This approach aims to enable analysts and policymakers to assess the potential of UTES on an equal basis with other energy-efficient technologies.

The development and performance of UTES is subject to significant uncertainties, which are acknowledged in the analysis by considering sets of scenarios for climate change, building energy performance, and UTES performance and adoption. As such, local energy demand for heating and cooling was first estimated using a subset of the Climate Model Intercomparison Project phase 5 (CMIP5) model ensemble, under three Representative Concentration Pathways (RCPs) for GHG concentrations. These results were combined with existing forecasts for the energy intensity of building heating and cooling as a function of region and building type, under two scenarios for building performance. Finally, energy savings from UTES were estimated for two scenarios for thermal performance, across nine plausible scenarios for ATES and BTES adoption; thermal performance and adoption were assumed to be driven by climate and subsurface properties. These performance and adoption scenarios were primarily derived from current UTES data for the Netherlands.

Section 8.2 of this chapter presents the working principle of UTES variants. Section 8.3 then describes the data sources used in this analysis for climate conditions, building performance, and UTES properties. Section 8.4 synthesizes these data sources to present regionally-aggregated results for UTES suitability and potential energy savings. Sections 8.5 and 8.6 conclude by discussing these results in the context of current UTES usage, and by indicating directions for future work.

8.2. Background

Buildings can be efficiently heated and cooled using UTES technologies (reviewed by Lanahan and Tabares-Velasco, 2017) to seasonally store and retrieve thermal energy in the subsurface. ATES and BTES are particularly versatile forms of UTES, and can be used for both heating and cooling. This analysis will therefore focus on these two technologies. Figure 8.1 presents a schematic view of ATES and BTES.

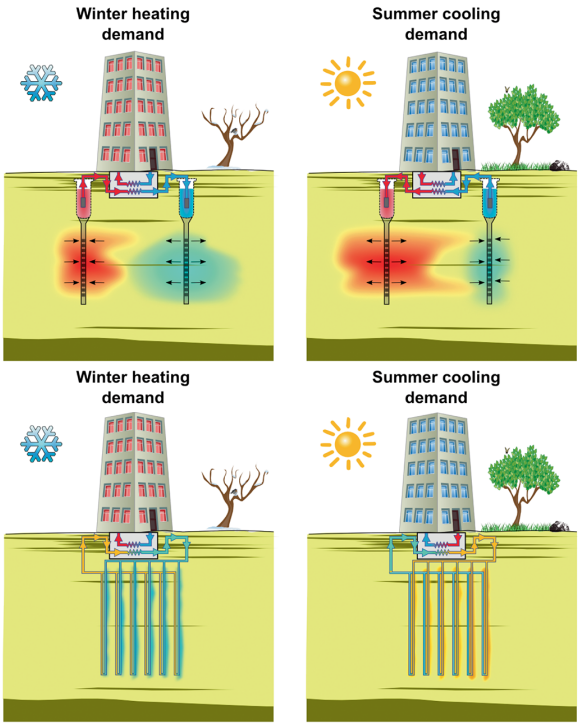


Figure 8.1: Schematic representation of ATES (top panels) and BTES (bottom panels), and their seasonal operating modes for heating and cooling (Bonte, 2013).

ATES is particularly effective for reducing the primary use of energy for heating and cooling in larger buildings (Tomasetta et al., 2015). It commonly relies on at least one pair of coupled wells, which store "warm" and "cold" water (relative to the ambient groundwater temperature) in an aquifer. During winter, groundwater is extracted from the warm well, and together with a heat pump and heat exchanger, provides heating to the associated building. The same groundwater is simultaneously reinjected at around 5-8°C in the cold well. Conversely, cooling is provided in summer by using groundwater from the cold well; cooling the building typically warms up the groundwater to about 15-20°C, which is then stored in the warm well. Next to the heat exchange with the groundwater, no other facilities are typically required during the cooling mode of

operation, i.e. "free cooling". In heating mode, the heat pump produces both heating and cooling capacity, with the latter being stored in the cold well. Depending on building properties and climate conditions, the annual heating and cooling demand may be somewhat mismatched (Bloemendal and Hartog, 2018).

However, balancing the long-term storage and extraction of thermal energy is essential to ensure the sustainable use of ATES. The changes in temperature distributions caused by excessive thermal imbalances can otherwise degrade the performance of other ATES systems sharing the aquifer (Calje, 2010). This implies that a certain level of both heating and cooling demand is needed for ATES. In addition, the large-scale applicability of the technology is limited by geohydrological conditions, such as the presence and accessibility of a suitable aquifer. The market for ATES is currently largely limited to Northern Europe, with the Netherlands being the most active market; the share of new utility buildings using ATES in the country was approximately 8% in 2015 (Bloemendal, 2018). A comprehensive review of the history and current international status of ATES is provided by Fleuchaus et al. (2018).

BTES is another common application of UTES, which tends to be more geographically flexible (Lanahan and Tabares-Velasco, 2017); this technology typically relies on a series of U-shaped, vertically oriented pipes which carry a thermal working fluid and transfer heat to the surrounding soil medium. A typical configuration of BTES is technically similar to ground-source heat pump systems (e.g. Sanner et al., 2003), but it is specifically operated for thermal storage (Hähnlein et al., 2013). As with ATES, a heat exchanger and heat pump are used to transfer heat between the building and storage. Relative to ATES, BTES can be scaled down more economically for smaller residential or utility buildings, but drilling costs can become less attractive for the multiple boreholes used in the case of larger buildings (Lanahan and Tabares-Velasco, 2017). The thermal recovery efficiency of BTES systems is also typically lower than for ATES. However, BTES does not require an aquifer, which makes it more geographically versatile. In addition, its thermal interactions with the subsurface are more localized, so that thermal balance between heating and cooling is less critical. For these reasons, while the current contribution of BTES to overall building energy use remains limited, the technology is increasingly deployed in European markets and has significant potential for further growth (Lund and Boyd, 2016; Bayer et al., 2012; Blum et al., 2010).

8.3. Data sources

The performance of UTES systems is driven by three key aspects which require a spatially-explicit analysis: local climate conditions, building properties, and geohydrological characteristics. Where possible, this work relies on existing data sources to assess the influence of these properties. Figure 8.2 summarizes the structure of the analysis and its main data sources. The technology-specific assessment of ATES and BTES energy potential will be described in subsection 8.4.1 of the analysis, leading to the

evaluation of energy savings in subsection 8.4.2. This assessment will assume the two technologies have similar plausible pathways for adoption, but differ in their thermal performance and their building application sector. The spatial focus of the analysis is on 556 urban areas worldwide with a population greater than 750,000 people, using projected demographic data until 2050 (Nordpil and UN Population Division, 2010). The local outputs illustrated in Figure 8.2 are computed at the level of these cities. These locations are grouped within the 11 standard GEA RC11 regions (International Institute for Applied Systems Analysis (2012); illustrated in supplementary material, Figure E.1), to match the regional building energy forecasts provided by Urge-Vorsatz et al. (2012) and McNeil et al. (2008). These forecasts yield intermediate regional outputs in Figure 8.2. The year 2050 will be used as a reference throughout the analysis. This time horizon matches the demographic forecasts and other key data sources, and covers a sufficiently long time frame to evaluate medium-term consequences of climate change for UTES systems.

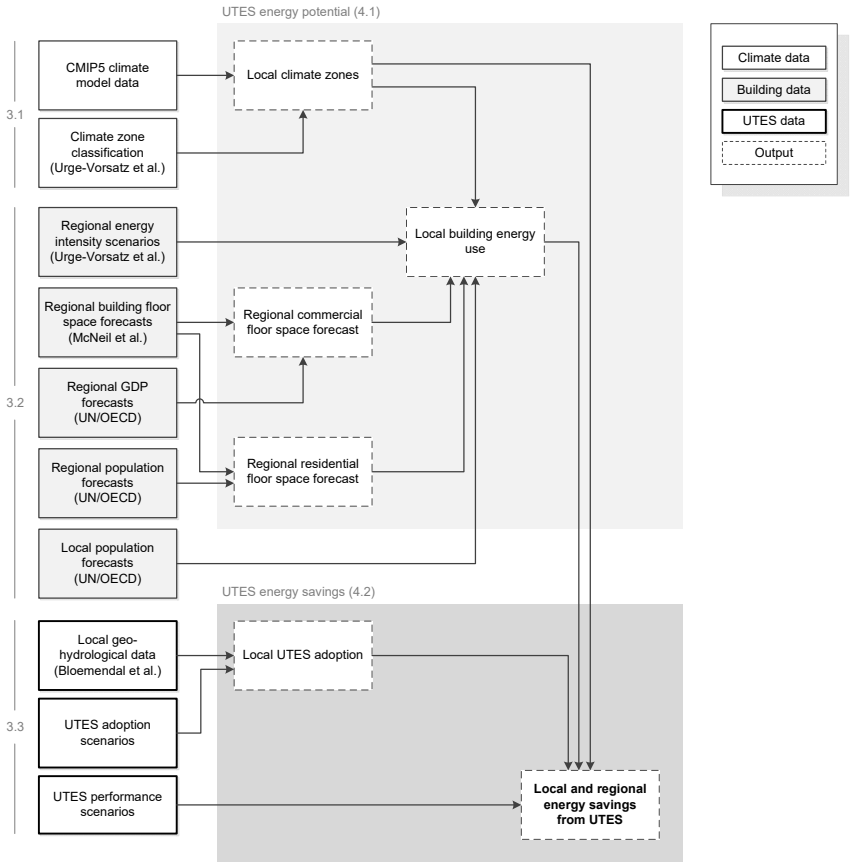


Figure 8.2: Overview of data sources used in the analysis.

8.3.1. Climate data

Local climate properties have a significant effect on the operation of ATES systems in particular, which require the average building heating and cooling demand to be roughly equal to avoid thermal imbalances in the subsurface. While BTES is less sensitive to thermal imbalances, it requires a certain minimal heating demand in order to be economically viable. In addition to regional variations which affect the suitability of ATES, this implies that climate change may impact the performance of UTES over the lifetime of existing buildings.

As a starting point for the analysis, a subset of the multi-model ensemble of CMIP5 climate simulation data is used to approximate heating and cooling demand for the 556 urban areas studied, over the 2040-2050 period. Future conditions are represented under three RCPs for GHG concentrations: RCP 2.6 (i.e. aggressive GHG mitigation measures), RCP 4.5 (under which anthropogenic GHG emissions peak around 2040),

and RCP 8.5 ("business-as-usual" GHG emissions). Table E.1 lists the eight models used in the analysis.

Based on the temperature indices, heating and cooling demand is represented with a degree-day approach (i.e. by integrating deviations over time from reference temperatures for heating and cooling). For each climate model, daily heating and cooling degree-days are computed at each grid point from the near-surface temperature time series in each RCP scenario, following the UK Meteorological Office approach (Day, 2006), then interpolated to the location of each urban area. Further details are provided in Appendix E. Following the estimation of mean heating and cooling demand, each urban area is then assigned to a given climate zone under each RCP, following the classification of Urge-Vorsatz et al. (2012) which is presented in Figure E.2. Scenarios for UTES performance (subsection 8.3.3) assume that ATES requires a minimal threshold of 1000 degree-days for both heating and cooling, to ensure thermal balance; climate zones below this threshold are therefore assumed to be unsuitable and not retained for the ATES analysis. The thermal balance requirement is less relevant for BTES, so that the technology is assumed to be applicable with a heating demand of at least 1000 degree-days, with any level of cooling demand.

8.3.2. Building data

This work primarily relies on data from the 3CSEP HEB model (Urge-Vorsatz et al., 2012) to estimate building energy use and feasible savings from UTES. This model takes a bottom-up perspective to assess global building energy use at the 2050 horizon, across several policy scenarios. For the purposes of this analysis, part of the model output is used to represent the combined energy intensity of space heating and cooling (e.g. as units of energy per unit of floor space). The model results are disaggregated according to geographic region, climate zone, and building type, so that variations in energy intensity across geographic regions, and across climate zones within a given region, can both be accounted for. In addition, this analysis will include energy intensity results from the 3CSEP HEB model under a reference building efficiency scenario, and under a moderate building efficiency scenario which represents accelerated renovations under stricter building codes. This scenario reflects comprehensive policy efforts to improve the energy efficiency of the building stock (through e.g. building envelope standards), but not a large-scale deployment of state-of-the-art building technologies. Under both of these scenarios, UTES would therefore have a structurally similar contribution to energy savings, by replacing conventional HVAC systems.

The data from the 3CSEP HEB model data is combined with regional projections for residential and commercial floor space, following Urge-Vorsatz et al. (2012) and using additional data from the BUENAS model (McNeil et al., 2008). These baseline datasets are available through the GBPN data API (Global Buildings Performance Network, 2018). This data is updated with recent demographic forecasts (United Nations,

Department of Economic and Social Affairs, Population Division, 2017). The regional floor space forecasts are scaled down to each urban area in proportion to its population, then combined with the energy intensity data to estimate building energy use in each urban location, according to its geographic region and climate zone.

The distinction between building types provided by the 3CSEP HEB data is used to assign each UTES technology to a given building application sector. ATES is assumed to be applied in two building types: large residential (multi-family) buildings, and utility buildings (i.e. commercial and public buildings). In parallel, BTES is applied for single-family buildings. In practice, there is significant overlap between the applicability of the technologies: BTES is commonly used for building sizes of 100-10,000 m^2, while ATES is technically applicable at sizes above 1000 m^2 (Agterberg, 2016). The assessment will therefore tend to underestimate the energy potential of BTES, and overestimate the contribution of ATES. However, distinguishing between ATES and BTES adoption within a building type (e.g. utility buildings) would require more detailed data than is currently available, so that the analysis will instead focus on highlighting the relative importance of each building sector for UTES.

Equations 8.3.1 and 8.3.2 illustrate the calculation of urban energy use; for a given building type (*build*, i.e. multi-family, single-family or utility), the estimated floor area [m^2] in each city is given by the product of the estimated city population p, and the estimated specific floor area f [$m^2/person$] for the corresponding building type in the city's geographic region:

$$Area_{build,city} = p_{city} \cdot f_{build,region} \tag{8.3.1}$$

For each RCP and building efficiency scenario (eff), the energy used by buildings of a given type in each city is then the product of floor area by the energy intensity EI [GJ/m^2], with the latter corresponding to the 3CSEP HEB estimate for each geographic region, climate zone (computed for each RCP), and efficiency scenario:

$$E_{used,build,city,RCP,eff} = Area_{build,city} \cdot EI_{build,region,climate,eff} \tag{8.3.2}$$

For each building type and combination of scenarios, E_{used} then corresponds to the total energy use towards which the UTES technologies can contribute in each urban area.

8.3.3. UTES properties

Geohydrological suitability

In parallel to climate conditions, the design and performance of UTES systems is strongly dependent on subsurface properties. Bloemendal et al. (2015) described four aspects which are specifically relevant for ATES:

1. Water quality: the injection and extraction of groundwater for thermal storage results in mixing across the depth of the aquifer, which can cause undesirable chemical reactions – for instance, the mixing of aerobic water with reduced iron-rich water creates precipitants, which may eventually affect the performance of storage wells by clogging the well screens.

2. Water salinity: freshwater aquifers are usually preferable for thermal storage, as saline aquifers require more costly maintenance and design to avoid corrosion of the equipment. In addition, salinity stratifications in thicker aquifers result in density differences, which can decrease thermal efficiency. However, saline or brackish coastal aquifers offer largely unused potential for thermal storage, as these aquifers are mostly unsuitable for domestic or agricultural use.

3. Groundwater flow: areas with a high ambient flow will cause higher losses through advection, i.e. the transport of stored thermal energy away from the capture zones of the wells. An ambient flow of 25 m/year, which is typical of several areas of the Netherlands, can be significant in relation to the thermal radius of an average thermal storage well (25-100 m) (Bloemendal and Hartog, 2018). Areas with a lower groundwater flow are therefore better suited for thermal storage.

4. Structural composition and thickness of the aquifer: complex geohydrological structures are typically less suitable for thermal storage. Faults and fractures may make it more challenging to control the development of stable thermal storage zones in the subsurface; similarly, high hydraulic conductivities may yield larger advection losses. The effect of thickness on suitability presents a trade-off: wells will be cheaper to drill and install in shallow aquifers, but they may be more vulnerable to contamination and stratification.

To evaluate the effect of these properties on thermal storage potential, Bloemendal et al. (2015) used the WHYMAP (Richts et al., 2011) and IGRAC (BGR/UNESCO, 2008) subsurface databases to construct a subsurface suitability index for thermal storage. The WHYMAP transboundary aquifer maps provide data on aquifer composition and annual recharge, while the IGRAC database provides country-by-country data on subsurface/aquifer composition, productivity, recharge, and groundwater abstraction. The two databases were used to assess aquifer and groundwater characteristics; these characteristics were then quantified based on their impact on thermal storage suitability, to yield a worldwide suitability index scaled from 0 to 10. The value of this index will be used in this analysis as a measure of the local geohydrological potential for thermal storage, at each city.

Bloemendal et al. primarily identified this measure for ATES, but the index can also be used to estimate suitability for BTES. BTES systems do not require suitable aquifers,

and can technically be applied in rock/sandstone formations, although unconsolidated depositions are likely to be preferable to reduce drilling costs (Lanahan and Tabares-Velasco, 2017). In general, the technical and economic factors driving the adoption of BTES can be assumed to be less sensitive to subsurface conditions than ATES. A study of 1100 ground-source heat pump systems in the German state of Baden-Württemberg for instance found that capital costs were largely driven by local market conditions, rather than subsurface properties (Blum et al., 2011). More research would be needed to assess these properties in different markets; similarly, the small sample size available for ATES makes it difficult to systematically evaluate the impact of subsurface conditions on investment costs, even within a single country such as the Netherlands (Agterberg, 2016). Considering that there is still a lack of large-scale data regarding the economic performance of ATES and BTES, this analysis will consider the implications of different scenarios for the impact of subsurface conditions on the adoption of each technology.

UTES adoption

With the exception of a small number of national markets such as the Netherlands, there is currently still a lack of knowledge about technology adoption dynamics for UTES. However, the validation process used by Bloemendal et al. (2015) indicated that current adoption patterns for ATES were generally consistent with the subsurface suitability index, with highly suitable areas (when combined with appropriate climate zones) typically having more active markets for thermal storage. This follows from the technical and economic assumptions made in the construction of the index, as thermal storage systems will be more technically difficult and costly to install in unsuitable areas.

Given that subsurface suitability can be assumed to be a reasonable proxy for the technical and economic drivers of thermal storage adoption, this analysis uses Bloemendal et al.'s suitability index as the key factor driving thermal storage use in each location. However, the maximum long-term market potential of thermal storage, as well as the relationship between subsurface suitability and adoption, both remain highly uncertain. The analysis therefore uses a set of nine scenarios to represent plausible adoption values in 2050, illustrated in Figure 8.3. These scenarios cover different maximum market shares for thermal storage at the 2050 horizon (L, M and H in the figure, for low, medium and high adoption), along with different relationships for the sensitivity of the market share A as a function of geohydrological suitability (1, 2 and 3, in decreasing order of sensitivity). Market share is here defined as the fraction of the applicable building stock (large urban residential and commercial buildings for ATES, single-family urban residential for BTES) which uses each UTES technology. The same scenarios are used for ATES and BTES, but will be applied separately to each technology's building sector.

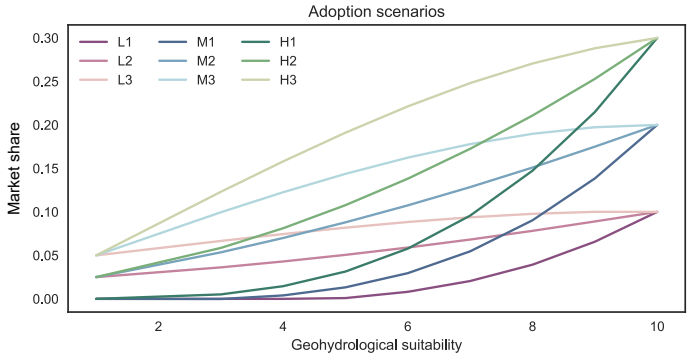


Figure 8.3: UTES adoption scenarios, defined as market share (A) as a function of geohydrological suitability.

Although the time dynamics of adoption are not explicitly represented in the scenarios, the maximum market shares (L, M and H) are based on a simplified building stock model which uses current data for the Netherlands to project plausible adoption values over a 35-year period. Additional details for the derivation of these scenarios are presented in Appendix E.

UTES energy performance

The energy savings obtained from UTES depend on multiple factors aside from climate conditions alone, such as effective integration with building control systems; energy savings may thus vary widely across buildings with similar energy demand profiles. Nonetheless, energy savings for ATES and BTES can be approximated from existing studies on the expected and realized performance of storage systems (de Graaf et al., 2016; Tomasetta et al., 2015; Willemsen, 2016). These studies indicate energy savings varying between 40 to 80%, which is consistent with studies of the effective coefficient of performance (COP) for heating and cooling with UTES. These values typically range from about 2-4 and 15-30 respectively, in each operational mode. Two performance scenarios will be used in the analysis to represent different levels of thermal performance for UTES. Table E.4 presents the COP values used in each scenario for each technology, and for a reference conventional building energy system (Kalaiselvam and Parameshwaran, 2014). Based on these values and estimated climate conditions in each urban area, average COP values are then computed for each technology, city, RCP, and thermal performance scenario, following equations E and E.

The effective energy savings realized from UTES should also acknowledge the additional lifecycle emissions associated with the installation of ATES and BTES systems; effective savings can thus be significantly lower than expected from primary energy use alone. Following the lifecycle assessment presented by Tomasetta et al. (2015), the aver-

age energy savings rate Δ is then given by the following equation, where LCA_{UTES} is the correction factor for lifecycle emissions of each UTES technology (conservatively set to 50% and 40% for ATES and BTES respectively, for a balanced heating/cooling demand). $\overline{COP}$ is the average COP value computed in equation E for each technology, city, RCP, and thermal performance scenario (*perf*).

$$\Delta_{UTES,city,RCP,perf} = 1 - \frac{\overline{COP}_{conv,city,RCP}}{\overline{COP}_{UTES,city,RCP,perf} \cdot LCA_{UTES}} \tag{8.3.3}$$

For each UTES technology, this savings rate is multiplied by the average market share (A) and by the estimated total annual energy use of applicable buildings in each urban area (E_{used}, eq. 8.3.2), to yield estimated energy savings. Eq. 8.3.4 presents this computation, which is applied for each city, RCP, building efficiency scenario (*eff*), and thermal performance scenario (*perf*):

$$\begin{aligned} E_{saved,build,city,RCP,eff,perf} = & E_{used,build,city,RCP,eff} \cdot A_{UTES,city} \\ & \cdot \Delta_{UTES,city,RCP,perf} \end{aligned} \tag{8.3.4}$$

The ratio of energy savings to total energy use will also be used to estimate fractional energy savings, and isolate the relative impact of UTES across scenarios in which energy use varies significantly (such as building efficiency scenarios).

8.4. Analysis

This section first evaluates the worldwide and regional energy potential for ATES and BTES, based on climate and geohydrological suitability as well as each technology's application sector. Subsection 8.4.1 presents suitability maps for each technology under different climate scenarios. The energy potential of each technology is then assessed in the context of its respective building sector in subsection 8.4.1. Subsection 8.4.2 extends this assessment using scenarios for adoption and thermal performance, to evaluate the feasible contribution of ATES and BTES towards urban energy savings.

8.4.1. Worldwide and regional energy potential for UTES

Suitable climate zones for UTES

The climate prerequisites for UTES are first combined with the CMIP5 climate model data to generate suitability maps across climate scenarios, based on the degree-day thresholds detailed in subsection 8.3.1 and Figure E.2. Figure 8.4 presents global maps for the RCP 2.6 and RCP 8.5 climate scenarios. The intermediate RCP 4.5 scenario is discarded for this analysis, as it does not yield significantly different results from RCP 2.6 at the 2050 horizon.

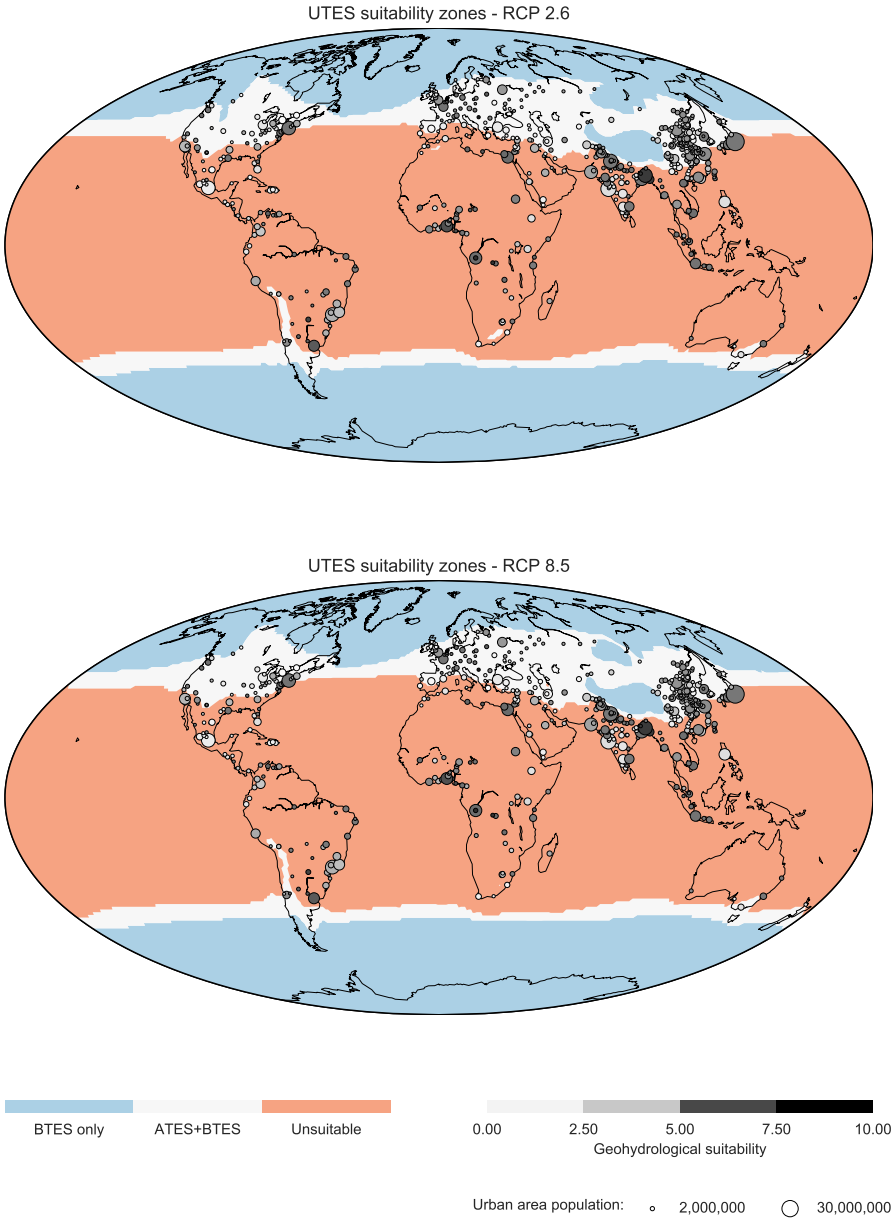


Figure 8.4: Worldwide suitability maps for UTES, for RCP 2.6 (top panel) and RCP 8.5 (bottom panel). Urban areas are shaded as a function of geohydrological suitability, and sized as a function of population.

These results are largely consistent with the assessment provided by Bloemendal et al. (2015), so that the use of ATES is mostly limited to temperate regions. Some discrepancies can be found for tropical and sub-tropical areas which were classified as suitable in Bloemendal et al. (2015), e.g. sub-Saharan Africa; while the earlier study used a simple climate suitability mapping based on average temperatures and seasonal variations, these areas would fall outside the degree-day criteria used in this work (which more explicitly consider thermal balance for ATES, and require a minimal heating demand in the case of BTES). The suitability regions shift slightly away from the equator in RCP 8.5, which e.g. increases suitable zones for ATES in the Northwestern U.S., Canada and Russia relative to RCP 2.6. Figure 8.5 displays the same suitability maps with a focus on Asia, which contains a large number of cities within the suitability bounds. These include major areas such as Beijing, Shanghai and Tokyo. While the relatively coarse resolution of the CMIP5 models and their inherent uncertainty make this approach unsuitable for local forecasts, these results indicate that enough heating and cooling demand should be present to allow ATES systems to maintain thermal balance.

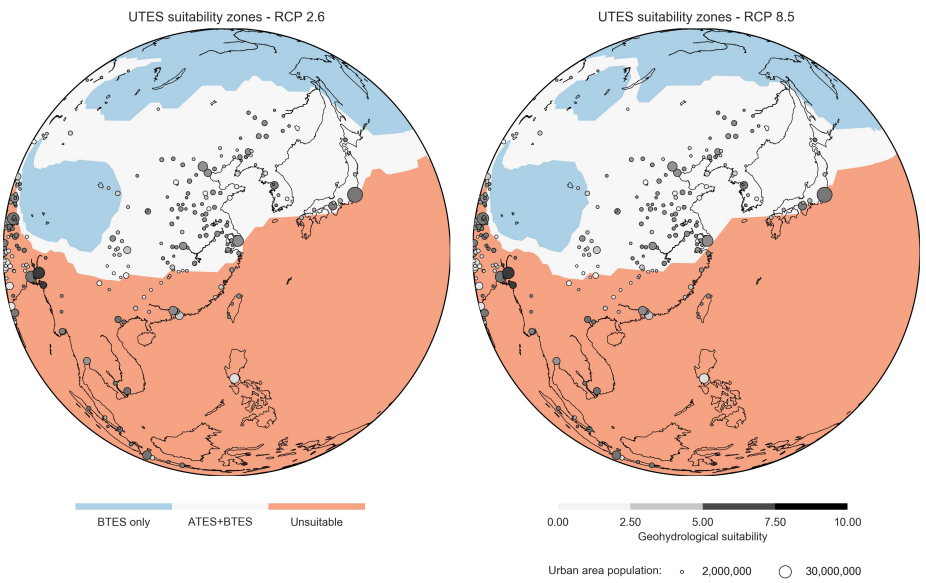


Figure 8.5: Suitability maps for UTES in Asia, for RCP 2.6 (left panel) and RCP 8.5 (right panel). Urban areas are shaded as function of geohydrological suitability, and sized as a function of population.

Building energy use in UTES climate zones

Figure 8.6 presents the worldwide distribution of total energy use in suitable climate zones (for the building sectors applicable for ATES and BTES) in each building efficiency scenario, as a function of subsurface suitability. This distribution corresponds to the energy usage towards which UTES could feasibly contribute, as computed from eq. 8.3.2 for each urban area. The top panels show that energy use under RCP 2.6 is largely clustered in areas with a medium subsurface score (4-7), with the building efficiency scenarios having a limited effect on the distribution of energy use. The bottom panels present the change in energy use under RCP 8.5, indicating that energy demand is slightly increased in areas with a very low subsurface score (1-2), and mostly decreased in medium to medium-high scoring areas – particularly for ATES.

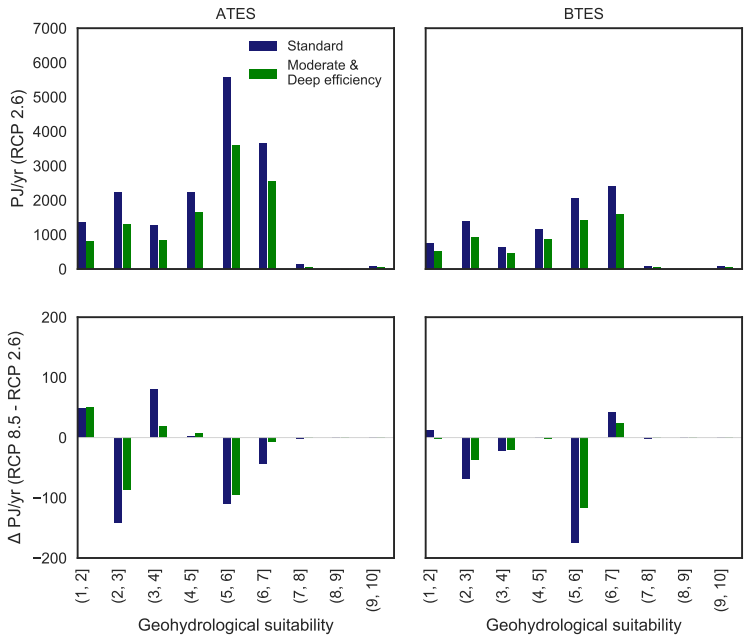
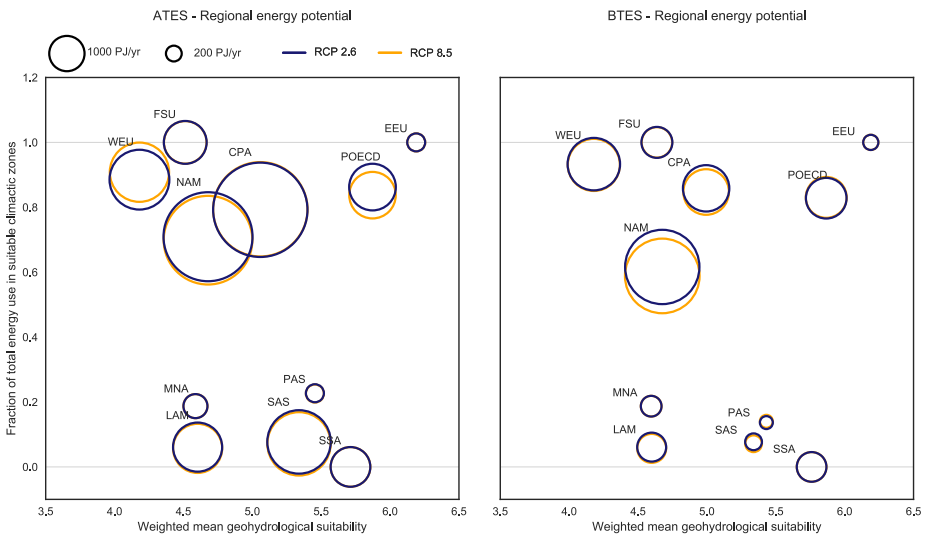


Figure 8.6: Worldwide distribution of estimated energy use in large urban areas in suitable climate zones and relevant building sectors, as a function of geohydrological suitability and building efficiency scenarios, for ATES (left panels) and BTES (right panels). Top panels: RCP 2.6; bottom panels: change in energy use under RCP 8.5 relative to RCP 2.6.

This change is driven by two effects: the change in building energy use linked to shifts in heating and cooling demand (which directly follows the input data on building energy intensity across climate zones), and the geographical shift in climate zones suitable for UTES. RCP 8.5 may therefore reduce overall feasible energy savings from UTES as certain areas fall outside of the suitability areas, although effective energy sav-

ings (subsection 8.4.2) are also mediated by the link between local climate and UTES performance. Figures E.5 and E.6 in Appendix E presents aggregate results for feasible energy savings (i.e. energy use found within suitable climate zones, in applicable building sectors), across climate and building energy scenarios. Feasible energy savings from ATES are significantly larger than for BTES, mainly due to the assumption on large residential and utility buildings being reserved for ATES; in practice, both technologies would be applicable for most buildings in these sectors. However, this result highlights that the large residential and utility market offers more energy potential for the future development of UTES. Figures E.5 and E.6 also break down these results regionally as a function of subsurface suitability, indicating that energy use in the CPA (Centrally-Planned Asia, i.e. largely China) region is clustered in a medium subsurface suitability range, while high suitability scores are only found in the European regions (WEU, EEU and FSU).



ID	CPA	EEU	FSU	LAM	MNA	NAM	PAS	POECD	SAS	SSA	WEU
Description	Centrally-Planned Asia and China	Central and Eastern Europe	Former Soviet Union	Latin America and Caribbean	Middle East and North Africa	North America	Other Pacific Asia	Pacific OECD countries	South Asia	Sub-Saharan Africa	Western Europe

Figure 8.7: Fraction of energy use in each geographic region which is in suitable climate zones (standard energy efficiency scenario), as a function of weighted mean subsurface suitability of each region. Markers sized by total energy use in relevant building sectors in each geographic region. Left panel: ATES; right panel: BTES.

To visualize the attractiveness of each geographic region for UTES development, Figure 8.7 (with ATES on the left, and BTES on the right) presents the fraction of urban energy use within each region which is found in suitable climate zones, as a

function of the mean regional subsurface score (weighted by the energy use of each urban area in the region). This indicates that Central and Eastern Europe (EEU) may offer particular opportunities for UTES: all cities simulated in the region have a suitable climate for UTES, combined with the highest mean subsurface score. The Pacific OECD (POECD) region also performs well on these indicators, although RCP 8.5 somewhat decreases ATES opportunities in the region by making some cities climactically unsuitable. Similarly, with BTES and RCP 8.5, some areas in North America (NAM) and Centrally-Planned Asia (CPA) fall below the threshold for heating demand which was assumed to be required for an economically viable application. Although South Asia (SAS) and sub-Saharan Africa (SSA) have relatively high energy usage and subsurface scores, the cities simulated in these regions are almost entirely outside of the suitable climate zones.

8.4.2. Energy savings from UTES

Building on the assessment of energy potential for UTES across climate and building efficiency scenarios, this subsection adds scenarios for the adoption and performance of ATES and BTES, to evaluate energy savings from each technology (using a full factorial experiment design, for a total of 108 scenarios). These savings are evaluated within specific building sectors for each technology: large urban residential and utility buildings for ATES, and urban single-family residential buildings for BTES. The energy savings are computed for each urban area using eq. 8.3.4 and aggregated. Figure 8.8 shows boxplots for worldwide urban total and fractional energy savings (i.e. as a fraction of total building energy use) for each technology, grouped by climate scenarios. The boxplot whiskers indicate the full range of model results. Compared to RCP 2.6, mean total and fractional energy savings from ATES both slightly increase in RCP 4.5, then decrease in RCP 8.5, as some urban areas move outside of the climate suitability zones. In the case of BTES, the estimated total and fractional energy savings improve monotonically with increased climate change, as geographic effects are compensated by relatively improved thermal performance under a cooling imbalance. However, for both technologies, the effect of climate scenarios on mean energy savings remains small in relation to the uncertainty created by other input scenarios. The top panels of Figure 8.9 displays the regional breakdown of total energy savings for both technologies; the highest savings for ATES are found in the CPA and NAM regions. This is consistent with the regional potential visualized in Figure 8.7, as these regions have reasonably high subsurface scores and the largest total energy use. For BTES, the relatively low floor space and energy intensity of single-family housing in the CPA region decrease overall potential in the area, compared to ATES. Maximum energy savings from ATES show a slight downward trend in the NAM region with increased climate change, although this may be caused by changes in energy intensity rather than by ATES use.

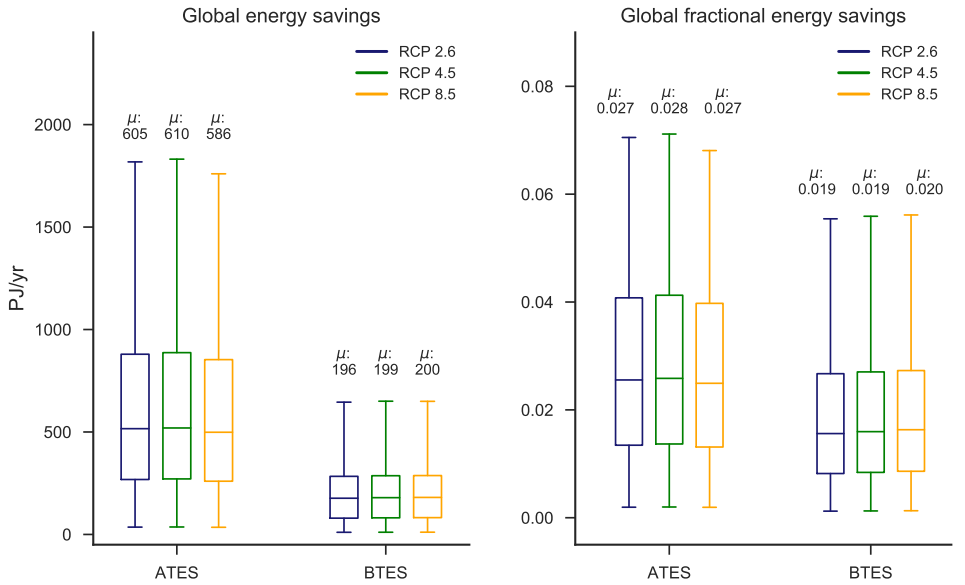


Figure 8.8: Worldwide total (top panels) and fractional (bottom panels) energy savings from ATES and BTES, for large urban areas.

8

To isolate the effect of climate scenarios on the relative performance of ATES and BTES, the bottom panels of Figure 8.9 present the corresponding fractional energy savings (defined as a fraction of total energy use in the building sectors applicable for each technology). We observe that RCP 4.5 and RCP 8.5 improve fractional savings in the European regions: EEU, former Soviet Union (FSU) and Western Europe (WEU). Greater cooling demand makes UTES more broadly applicable in these regions, and improves the thermal performance of both technologies. However, fractional savings for ATES in the CPA region decrease under RCP 8.5 compared to RCP 2.6, as certain urban areas move outside of suitable climate zones. Fractional ATES savings in the NAM region largely stay constant, so that the decrease in total energy savings can be attributed to lower total energy use (consistently with results presented by Petri and Caldeira (2015) for energy demand under climate change in the United States).

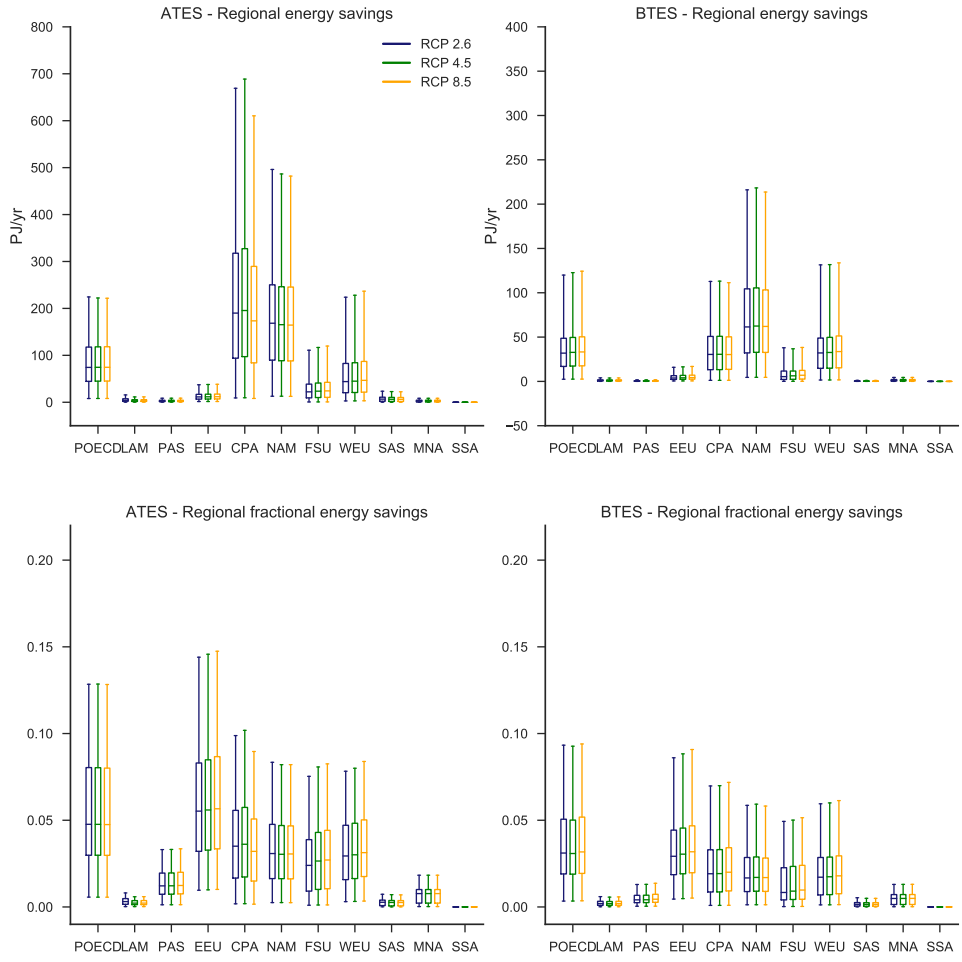


Figure 8.9: Top panels: total energy savings from ATES (left) and BTES (right) for large urban areas across geographic regions, grouped by climate scenario. Bottom panels: fractional energy savings from ATES (left) and BTES (right).

As evidenced by the dispersion of UTES performance within each climate scenario, a significant portion of the output uncertainty is driven by other input scenarios, including adoption and performance. Figure 8.10 focuses on the impact of adoption scenarios on global fractional energy savings (relative to each technology's building application sector). This visualization compares the effect of maximum market share (in the L, M and H scenario groups), relative to the sensitivity of adoption to geohydrological suitability (1, 2 and 3, within each scenario group). The latter is more influential, so that the L2 and L3 scenarios yield comparable or higher savings than the M1 and H1 scenarios, despite a lower maximum market share. This is consistent with the results

presented in Figure 8.6: given that most of the simulated urban energy use is in areas with medium subsurface suitability, stimulating UTES adoption in these areas would offer the largest potential benefits. This pattern is present for both ATES and BTES.

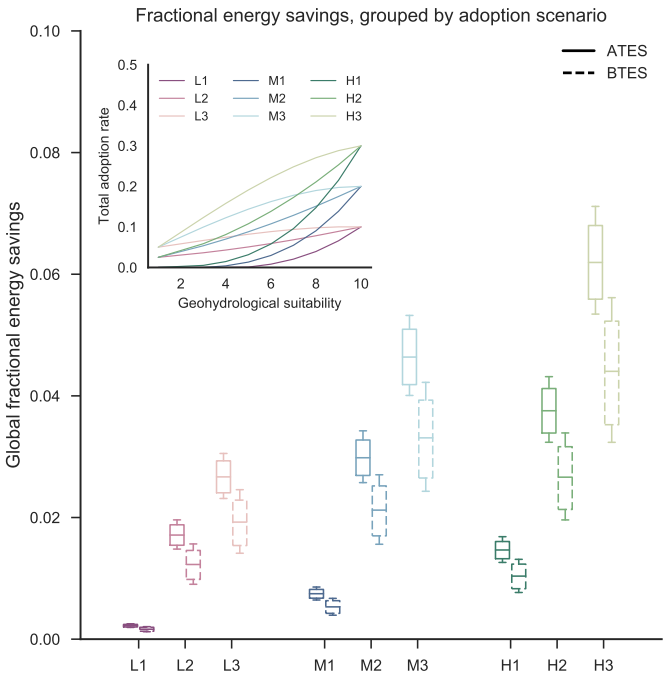


Figure 8.10: Worldwide fractional energy savings from ATES (solid boxplots) and BTES (dashed boxplots), grouped by adoption scenario. Inset: adoption scenarios.

A global sensitivity analysis is presented in Appendix E to support the interpretation of each scenario's relative importance; this decomposes the contribution of each input scenario (climate, COP performance, building efficiency, and adoption) towards the variance of estimated energy savings. Due to its relatively lower estimated thermal performance, BTES is somewhat more sensitive to the COP scenarios. However, overall energy savings with both technologies remain largely driven by adoption scenarios.

8.5. Discussion

The magnitude of the estimated energy savings can be compared to Lund and Boyd (2016)'s assessment of ground-source heat pump (GSHP) energy use. The latter indicates a worldwide use of 330 PJ/year from geothermal heat pumps in 2015, largely from BTES. It should be emphasized that the current analysis was limited to large urban areas, which underestimates the overall potential of UTES at the national or regional scale, and which would make direct comparisons difficult. In addition, the lifecycle

factor added to COP calculations significantly reduces energy savings for a given primary energy usage. However, the comparison of this analysis with current GSHP use does indicate regional patterns, presented in Figure 8.11; current use in Western Europe (approximately 110 PJ/year) is relatively close to the upper bound found in this analysis for ATES and BTES combined (380 PJ/year in urban areas), while China and North America should present much greater unexploited technical potential. Although a detailed comparison with other energy sources would be outside of the scope of this work, it can for example be noted that energy savings under a high-adoption scenario in China would be in the same order of magnitude as current energy production from solar photovoltaic and solar thermal technologies in the country, respectively 420 and 890 PJ in 2017 (China Energy Portal, 2018; International Energy Agency, 2017).

While this estimated technical potential depends on the assumptions used in the underlying input data for e.g. energy intensity, these should offer a reasonable starting point for the estimation of energy savings; for instance, using demographic data for 2015 instead of the projections used in the analysis, the estimated energy potential in Beijing and Shanghai would be respectively 355 PJ/yr and 280 PJ/yr, in the standard energy efficiency scenario. These estimations are consistent with the ranges estimated by Huo et al. (2018) for the energy consumption of urban residential and commercial buildings in the corresponding metropolitan regions, assuming a typical breakdown of end-use intensity for space heating and cooling of 40-70% in the residential and commercial sectors (Li, 2016; Zhou and Lin, 2008). It is also interesting to note that ATES was used at a relatively large scale in the Shanghai area in the 1980s, with an annual usage of over 1 PJ of cooling energy (approximately one-quarter of the current total usage in the Netherlands); due to issues with e.g. well clogging – which were common with early implementations of ATES – these systems were eventually abandoned (Fleuchaus et al., 2018). This early experience nonetheless suggests that suitable aquifer conditions are present in the area, and that improved technical practices could enable large-scale adoption. In parallel, current GSHP use in the Pacific OECD region (largely Japan) is almost nonexistent, although the region appears to have significant potential for UTES. This may be related to the increasing uptake of air source heat pumps in the region, which compete with UTES in the same building sectors.

Given the relatively established market for UTES in the Netherlands, the country provides a suitable reference to discuss the results of the urban analysis in the context of a national market. As such, the analysis included the cities of Amsterdam and Rotterdam, which together account for roughly 20% of national GDP and energy use. Using these indicators to extrapolate estimated UTES energy savings to a national scale, Figure 8.12 compares the current analysis with the national UTES policy target for 2023 (21 PJ/year), and with the historical use of UTES systems (4 PJ in 2016; Centraal Bureau voor de Statistiek, 2017). This extrapolation can be justified for the Netherlands, given the relatively uniform subsurface and climate conditions in the country.

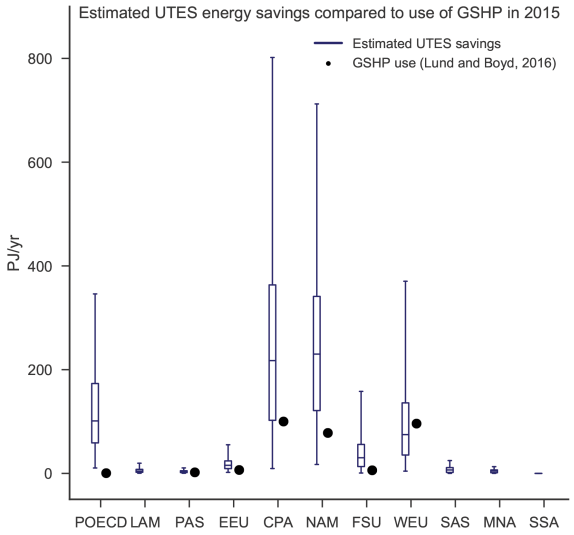


Figure 8.11: Regional comparison for estimated energy savings for UTES (boxplots), relative to ground source heat pump use in 2015.

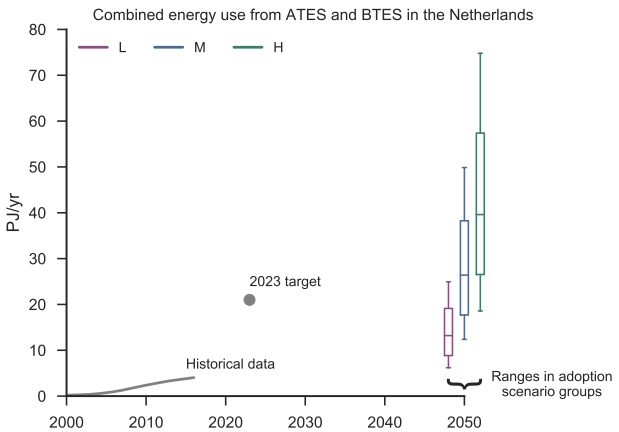


Figure 8.12: Historical data for ATES and BTES energy in the Netherlands (2000-2016), compared with policy target for UTES (2023), and estimated energy savings extrapolated from the current analysis (2050). Estimated future energy savings are grouped by adoption scenario.

The lower bound of the analysis output (i.e. low COP, low adoption, and high building energy efficiency scenarios) approximately reflects the current situation for UTES. This is consistent with assumptions underlying the low-adoption scenario, which projects current adoption rates over several decades of turnover in the building stock.

In parallel, the policy target for 2023 falls into the second quartile of energy savings in scenarios for medium adoption. For the Netherlands, this indicates that the magnitude of potential energy savings should be plausible at the national level – and, conversely, that the policy target is technically achievable under fairly conservative assumptions. However, given the inertia inherent to the building stock, reaching the policy target by 2023 would require much greater retrofit and/or adoption rates than was assumed in the adoption scenarios.

Finally, we note that the scarcity of available subsurface space is increasingly emerging as a barrier for the adoption of ATES in some areas of the Netherlands, such as the city of Utrecht (Bloemendal et al., 2017). This scarcity is largely a function of the spatial density of energy demand from buildings, the spatial layout guidelines imposed by local regulations for ATES, and physical factors such as the thickness of local aquifers; ATES wells are thus more challenging to accommodate in shallower aquifers due to the larger resulting footprint of the stored thermal volumes. To evaluate whether this scarcity may eventually become generalized across other markets for ATES, Figure 8.13 presents the cumulative distribution of simulated energy density across the cities included in this assessment which would be in suitable climate zones. For a given city, energy density is defined as the ratio between the estimated energy use for heating and cooling in buildings suitable for ATES (i.e. large residential and utility buildings), and the area of the city. This metric can therefore offer an indication of the spatial density at which ATES would be used. The vertical dashed line corresponds to the estimated energy density for Utrecht (assuming that local energy intensity would be the same as in Amsterdam), which indicates that the latter would be close to the median across all simulated cities.

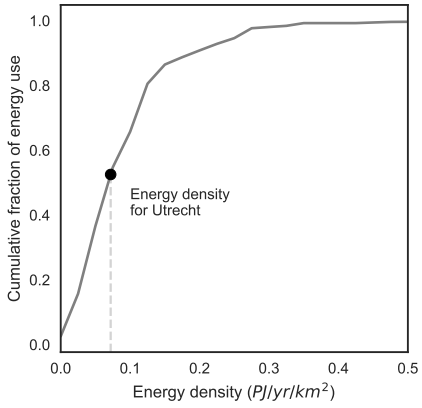


Figure 8.13: Cumulative distribution of the simulated energy density of cities in suitable ATES climate zones (RCP 2.6, with reference building energy efficiency scenario).

While this result follows from a high-level estimation of energy use in the built environment, it nonetheless indicates that – depending on future adoption pathways – a scarcity of subsurface space for ATES is unlikely to be unique to the Netherlands in the longer term: Dutch cities are typically geographically dense, but they also combine energy-efficient buildings with relatively low heating and cooling loads. Given that the fixed physical layout of ATES systems leads to a strongly path-dependent development, local authorities in new markets for ATES should therefore account as early as possible for the possibility of a scarcity of space when implementing guidelines for the spatial layout of systems.

8.6. Conclusions

This analysis combined existing data sources for climate conditions, building performance, and geohydrological properties, to evaluate the potential of ATES and BTES for urban energy savings at the 2050 horizon. Based on the climate requirements of each technology, suitability maps were first generated under the RCP 2.6 and RCP 8.5 climate scenarios. For ATES, these results are largely consistent with Bloemendal et al. (2015)'s assessment. While climate scenarios are unlikely to significantly affect UTES suitability by the middle of the century, increased climate change may eventually make some temperate areas unsuitable for ATES due to a mismatch between heating and cooling – particularly in East Asia. To a certain extent, this shift may be compensated by increased cooling demand in the Northwestern U.S., Canada and Russia. However, as these areas are more sparsely populated, the total energy use in suitable climate zones was found to decrease for both technologies under the RCP 8.5 scenario, through a combination of changes in climate zones and building energy intensity. BTES is more widely applicable and less sensitive to climate conditions than ATES, as it was only assumed to require a minimal level of heating demand.

Under the assumptions made for each technology's applicable building sector, the overall market for ATES (i.e. large urban residential and utility buildings) would be larger than the single-family urban residential market for BTES. In practice, the market for large residential and small utility buildings is likely to be split between ATES and BTES, depending on site-specific factors which would be outside of the scope of this study, and BTES would be technically applicable across all of the analyzed building sectors. However, in general, urban multi-family residential and utility buildings clearly offer greater energy potential for UTES than single-family housing.

For both technologies, Eastern Europe and the Pacific OECD regions should offer attractive prospects for future development, due to a combination of high geohydrological and climactic suitability. Although North America and China have relatively less suitable subsurface conditions for ATES, they are nonetheless key markets for UTES due to their high energy use. While BTES is increasingly common in these areas, the use of ATES remains marginal despite its potentially greater energy potential. Future

work could focus on detailing subsurface suitability in these regions, and evaluating the viability of ATES in large urban centers which appear suitable for the technology (e.g. Beijing, Tokyo or Shanghai). This assessment was combined with scenarios for thermal performance and adoption, to estimate global and regional energy savings from ATES and BTES in urban areas. The largest estimated total energy savings from ATES were found in China and North America, with mean values of 220 PJ/yr and 180 PJ/yr, respectively. Based on assumptions on thermal performance under demand imbalances, BTES consistently benefits from increased climate change, but has lower overall potential as a fraction of sectoral building energy use.

More broadly, although UTES has significant technical potential for energy efficiency, it remains an emergent technology, and its effective contribution towards urban energy savings will depend on future adoption pathways. This was highlighted in the analysis by the predominant role of adoption scenarios towards total energy savings, with scenarios for climate change or building efficiency being relatively less important towards this outcome. Long-term prospects for the technology are still uncertain; current experience in European markets suggests that ATES in particular is still held back by barriers such as inconsistent legislation and a lack of technical standardization, which make the technology more challenging for building owners (Agterberg, 2016). Future research could focus on better understanding how the characteristics of UTES affect its adoption and diffusion, compared to established energy-efficient building technologies. In the case of ATES, policymakers should also account for the possibility of a scarcity of space when planning the future development of systems in dense areas, as this scarcity may become problematic at relatively low total rates of adoption depending on local conditions.

Finally, although different scenarios were used to explore plausible values for the performance and adoption of the technologies, the analysis should be regarded as an exploratory assessment rather than a forecast, given the significant uncertainties which will affect future UTES development. The analysis also shares the limitations of the building model data which was used to represent building energy performance. For instance, energy intensity estimates for space heating and cooling were combined into a single value, as insufficient data was available to disaggregate these values across building sectors and regions. The analysis also takes a demand-side view of energy consumption in buildings, focusing on energy savings rather than GHG reductions (following Lucon et al., 2014). The latter will be determined by future emissions factors, themselves driven by supply-side changes in electricity generation. As such, the uncertainty created by future decarbonization pathways would make it difficult to draw meaningful conclusions in regards to GHG savings within this study. However, decarbonization is crucial to maximize GHG savings from UTES: as indicated by Bayer et al. (2012), ground-source heat pumps – including UTES systems – may ultimately only offer limited GHG benefits if they are powered by a carbon-intensive grid.

9

Conclusions and recommendations

These conclusions will first synthesize the findings of each individual chapter, in order to answer the research questions which guided the thesis. These findings will lead to a broader reflection on the methodological approach and its limitations, pointing towards avenues for future work. Finally, a set of policy recommendations concludes the thesis.

9.1. Conclusions

After reviewing key challenges for the future development of Aquifer Thermal Energy Storage, Chapter 1 formulated the main research objective for this thesis:

To assess the role of improved methods for the planning and operation of urban ATES systems towards a better alignment of private and public interests.

This objective was addressed through five research questions, which were first answered by developing an improved methodological toolkit for the modelling of ATES systems, then by applying this toolkit to increasingly realistic case studies. These questions are revisited in order below.

1. *How can ATES operation and spatial planning options be represented more realistically within ATES simulation models?*

   The adoption and operation of ATES technology can be understood as a complex adaptive system, in which collective outcomes – such as adoption pathways for the technology – emerge from an interplay between the use of the technology by building owners, external factors such as spatial planning policies or energy prices, and physical processes in the subsurface. These feedbacks are not addressed by the groundwater simulation models which are typically used to design and assess ATES systems (Bakr

et al., 2013; Li, 2014; Sommer et al., 2015), and which commonly leave "above-ground" dynamics out of scope – for instance by assuming constant or predefined pumping patterns for ATES systems.

As described in Chapter 1, agent-based modelling is a suitable alternative starting point towards representing the full complexity of ATES adoption and operation. However, this implied the design of a coupled simulation architecture which could bridge the gap between the modelling of ATES use at the level of buildings, and the modelling of physical processes in the subsurface.

As a key building block for this architecture, Chapter 2 described the pyNetLogo connector, which interfaces the popular NetLogo agent-based modelling software with a Python environment. This connector complements existing interfaces which link NetLogo with R or Mathematica, and extends NetLogo's capabilities through Python's extensive software ecosystem for scientific computing. The key features of pyNetLogo were illustrated by performing a global sensitivity analysis on a simple test model, and by parallelizing its execution for improved performance. Chapter 3 then introduced a Python-based object-oriented simulation architecture, which used pyNetLogo to link NetLogo with MODFLOW/SEAWAT. As a proof of concept, the coupled simulation architecture was used for a simplified case study of ATES, which included bidirectional feedbacks between subsurface conditions and the use of ATES by simulated building owners.

In the context of this research, this coupled simulation architecture therefore enables a more realistic simulation of ATES operation and spatial planning, by supporting separate modelling paradigms which are suited to the key processes involved – i.e. an agent-based model of ATES operation, and a finite-difference solution of the differential equations which govern transport processes in the subsurface. The co-simulation of these models then maintains a suitable level of detail for each individual model component, while including key feedbacks between ATES users and subsurface processes.

Compared to other methods for agent-based/groundwater modelling such as the integrated FlowLogo environment (Castilla-Rho et al., 2015), we note that this architecture supports the full capabilities of the MODFLOW/SEAWAT codes; beyond ATES applications, this enables modellers to address other complex groundwater management problems which include transport processes, such as aquifer contamination or saltwater intrusion. The architecture can also be used to add an agent-based component to existing MODFLOW/SEAWAT models. In parallel, this coupled approach preserves the user-friendliness of the NetLogo platform and its comprehensive features for spatial simulation. More complex agent processes can be handled through the object-oriented architecture by directly using the Python agent objects, or by introducing an additional software component through one of the various interfaces available through Python. This method was for instance illustrated in Chapter 7, by adding a MATLAB-based

model predictive control component to compute ATES flows. To facilitate future work using this coupled simulation approach, the architecture modules will be available through an online repository under `http://www.github.com/quaquel/pynetlogo`, along with interactive notebooks which replicate the analysis presented in Chapter 3.

2. *How can this simulation approach be used to efficiently assess different options for ATES operation and planning under uncertainty?*

   The "toy" case of ATES dynamics presented in Chapter 3 highlighted the importance of an integrated view for the treatment of uncertainties across the coupled models: while this case only involved simple behavioral assumptions in the agent-based component, the behavior of the coupled models was highly sensitive to different parametric values and interactions between the agent-based and geohydrological components – which included socio-technical uncertainties such as the response of ATES operators to system performance, and physical uncertainties such as aquifer properties. Using sensitivity analysis to better understand the implications of these uncertainties for ATES operation and planning was therefore a key element of the case studies presented in Chapters 5 and 6.

   However, this first required the choice of a suitable method for sensitivity analysis. Conventional methods for sensitivity analysis entail a trade-off between computational cost, and the information gained from the analysis: for instance, Sobol variance-based sensitivity indices (Sobol, 2001) accurately estimate the contribution of uncertain inputs to output variance as well as their interactions, but would have been computationally impractical for the simulation models used in this research. Conversely, screening methods such as Morris elementary effects (Morris, 1991; Campolongo et al., 2007) impose several restrictions on the analysis, as they require the use of continuous (rather than categorical) input parameters with a specific sampling design, and do not directly cover variable interactions.

   In the context of this trade-off, the problem of feature selection in the statistical learning literature offered several previously unexplored parallels to the sensitivity analysis of simulation models. Based on this literature, Chapter 4 identified the Random Forests (RF) and Extra-Trees (ET) decision tree-based algorithms (Breiman, 2001; Geurts et al., 2006) as alternative candidate methods for sensitivity analysis. These techniques were particularly promising for this application: they can be used with generic input sampling designs and categorical uncertainties, while supporting the study of variable importances for individual inputs as well as pairwise interactions.

   The decision tree-based algorithms were compared to Sobol indices and to the Morris elementary effects method, across three case studies of increasing complexity. The ET algorithm in particular accurately estimated the ranking and relative importance of Sobol total effect indices, outperforming the Morris method at a smaller or equiva-

lent computational cost. For the more complex case studies, which included a System Dynamics model of the H1N1 flu pandemic and the CDICE integrated assessment model, a sample size of less than 10% of the Sobol sample size was sufficient to reliably rank variable importances with ET. In parallel, a pairwise permutation metric allowed for the study of variable interactions. The decision tree-based algorithms can thus replicate some of the key insights of a global sensitivity analysis, for models in which conventional analysis techniques would be either computationally intractable, or insufficiently informative. More broadly, this can help modellers and analysts better understand the behaviour of complex environmental models, and contribute to a more transparent and credible modelling process.

For the purposes of this research, the decision tree-based method was thus applied under an exploratory modelling approach for the ATES case studies in Chapters 5 and 6, in order to assess the impact of uncertain parameters and policy options. Due the computational costs of the coupled agent-based/groundwater models, these exploratory studies were based on relatively small ensembles of 512 experiments; this sample size was nonetheless sufficient for a stable ranking of variable importances with the RF algorithm – which was for instance not the case with Morris elementary effects (Jaxa-Rozen et al., 2015b). By combining an efficient technique for sensitivity analysis with an integrated treatment of uncertainties across the agent-based and groundwater models, this exploratory modelling approach thus led to a better understanding of the impact of policy options and uncertainties.

3. *How do the operational uncertainties of ATES systems affect the design of suitable spatial planning arrangements?*

   The simulation and analysis methods presented in Chapters 2, 3 and 4 were used to simulate different scenarios for the future development of urban ATES systems. To compare different levels of scale and complexity, Chapter 5 thus explored an idealized case study – representing ATES adoption and operation in a representative aquifer without geographic constraints – while Chapter 6 depicted a more realistic case study for the Utrecht city center.

   In the idealized case, a comparison of different policies for the minimal distance between ATES wells evidenced a general trade-off between the economic performance of individual systems, and overall reductions in GHG emissions. However, this trade-off was not linear; under the assumptions of the model, the current design guidelines of 3.0 R_{th} could be revised to allow for a smaller distance of 2.75 R_{th} between wells –- and thus for a greater amount of wells to be built — without a statistically significant degradation of economic performance. When further reducing well distances, however, the development of thermal interferences could plausibly lead to a "tragedy of the commons", in which the economic performance of systems would be affected to a degree which could compromise the adoption of the technology by building owners.

In parallel, a sensitivity analysis indicated that the well distance guidelines and natural gas prices were the most influential factors for the annualized energy costs incurred by ATES operators relative to conventional energy; similarly, well distance was more influential for thermal efficiency and total GHG savings than geohydrological factors, or than energy prices (which would affect efficiency through their feedback effect on system adoption).

These results were obtained under idealized assumptions for the operation of ATES systems, in which the simulated systems would use their full nominal storage capacity, and meet thermal balance requirements following current guidelines in the Netherlands. However, technical and operational uncertainties led to significantly different conclusions for the more realistic case study presented in Chapter 6. This case simulated the future development of ATES in the Utrecht city center, starting from an initial set of 89 existing or planned wells. This development was simulated with static spatial planning, using well distances of 2.5 R_{th} or 3.0 R_{th} based on the nominal storage capacity of systems, and with a simple adaptive permitting approach which would include feedback from the actual usage of systems to optimize the allocation of aquifer volume. These planning policies were simulated across ensembles of 512 experiments to account for technical and operational uncertainties.

For this case study, the effective temperature difference (ΔT) between warm and cold storage wells, as well as the thermal balance between heating and cooling energy demand, were both found to be more influential for the thermal efficiency of systems than spatial planning policies. Similarly, ΔT and the fraction of used storage capacity had a greater impact on total and specific GHG savings than spatial planning. These results can be explained by several factors: spatial layout guidelines become relatively less influential in a urban environment with significant geographic constraints, which already limit the available space for new wells; due to its dense existing use of ATES and limited space for further development, the city center of Utrecht is already an example of partial "lock in". In parallel, the actual operation of systems can differ significantly from nominal design values, with systems typically being operated much less than expected.

As such, while the safety margins provided by current guidelines are reasonably conservative under idealized conditions, they are in practice further compounded by geographic constraints and by operational factors – so that the effective clearance between wells is likely to be significantly greater. At an average usage of 70% of the nominal storage capacity, which remains relatively high compared to typical practice (Willemsen, 2016), the median minimum clearance between simulated wells in Utrecht was thus approximately 4.0 R_{th}. The scarcity of space for new wells in the area may therefore largely be artificial.

This situation could be to some extent mitigated by relaxing the current layout guide-

lines to allow for well distances of 2.5 R_{th}, and by including feedback from operating conditions to adjust the allocation of subsurface space; compared to a "business as usual" case, this approach increased specific GHG savings by an average of 27% across the ensemble of simulation experiments, against a reduction of 3% in average thermal efficiency. However, maximizing the actual usage of the storage systems, as well as ΔT, would remain crucial towards increasing GHG savings from ATES.

4. *How can information exchange between ATES systems lead to a more efficient trade-off between individual system performance and collective interests?*

   The Utrecht case study highlighted the influence of operational uncertainties for the individual and collective performance of ATES systems. This implied that improved methods for the control of ATES systems could mitigate the impact of these uncertainties, and potentially better reconcile private and public interests for the future development of ATES systems. Chapter 7 therefore assessed an approach based on the distributed control of ATES systems – or ATES "smart grids" – in which information exchange would support the dynamic management of thermal interactions between neighboring ATES systems. This analysis revisited the case studies explored in Chapters 5 and 6, using a centralized model predictive control framework (Rostampour and Keviczky, 2016) for the idealized case, and a more computationally efficient distributed framework (Rostampour and Keviczky, 2017) for the Utrecht case. For both control frameworks, coupling constraints were applied to neighboring ATES wells to prevent overlap between the stored thermal volumes, while maintaining the thermal balance of individual systems. Due to computational constraints, the frameworks were simulated across a limited number of scenarios for energy use and spatial planning, with nominal values for technical and operational parameters. Notably, the simulation of building energy demand across standard KNMI climate scenarios indicated that climate change may already have an impact on the annual thermal balance of systems by 2040, shifting thermal balance towards cooling by approximately 10% for a representative office building between the G_L and W_H climate scenarios.

   Compared to a reference scenario (i.e. decoupled operation, under current planning guidelines of 3.0 R_{th}), information exchange combined with denser layout guidelines of 2.25 R_{th} improved specific GHG savings by 75% for the idealized case, and by 38% for the Utrecht case. In both cases, system performance (as measured by thermal efficiency and effective coefficient of performance) was comparable or greater than in the reference scenario, with the gains in specific GHG savings largely due to a smaller allocation of subsurface volume for the same energy use. Conversely, and consistently with the previous versions of the case studies, decoupled operation with these denser layout guidelines would significantly affect the efficiency of systems due to thermal interferences.

   These results indicate that the coordinated operation of ATES systems could lead to a

"win-win" for policymakers and operators, by enabling further adoption and increasing specific and total GHG savings without penalizing economic performance. However, this is conditional on the use of revised, denser layout guidelines for ATES systems – given that current guidelines are effective at avoiding the thermal interactions which could otherwise be managed through coupling constraints. Furthermore, the exchange of information introduces an alternative trade-off, under which participating ATES operators would trade off the implicit value of information about their use of thermal storage. This has particular implications for industrial users, and will require additional research to evaluate.

5. *To what extent can improved methods for ATES planning and operation contribute to the large-scale potential of ATES for energy savings in the built environment?*

   The case studies presented in Chapters 5, 6 and 7 depicted typical conditions for ATES use in the Netherlands, and covered a range of well configurations which adequately represents current practice in the country (as defined by e.g. the capacity and screen length of storage wells). In these conditions, the "smart grid" case studies point towards a promising opportunity to restructure the trade-off between private and public interests for the planning and operation of ATES, by combining information exchange with a denser spatial layout. However, this approach would be most relevant for cases where the demand for ATES is sufficient to otherwise lead to a scarcity of subsurface space. This scarcity can be expected to be a function of the spatial density of energy demand from buildings, as well as physical factors such as the thickness of local aquifers; ATES wells are more challenging to accommodate in shallower aquifers due to the larger resulting footprint of the stored thermal volumes, making the management of interactions particularly relevant in these cases.

   To better understand how the results of the case studies could be translated to a different geographical context, Chapter 8 assessed the long-term worldwide potential of ATES and BTES for urban energy efficiency. This analysis combined existing data sources for climate conditions, building performance, and geohydrological properties, to evaluate the potential of ATES towards energy savings in 556 major cities by 2050. Based on expected climate requirements for the minimum heating and cooling demand which would be viable for each technology, suitability maps were first generated under the RCP 2.6 and RCP 8.5 climate scenarios. This climate data was then combined with adoption scenarios generated by a simple building stock model.

   The analysis indicated that Eastern Europe and the Pacific OECD regions (e.g. Japan) should offer attractive prospects for future ATES development, due to a combination of high geohydrological and climactic suitability. In addition, while North America and China typically have relatively less suitable subsurface conditions for ATES, they are nonetheless key markets due to their high energy use. For instance, in a high-adoption scenario, which would apply technology adoption rates similar to the current situation

in the Netherlands, ATES could save up to 700 PJ/year in China – or approximately 10% of energy use for large residential and utility buildings, within large urban areas. Chinese cities suitable for ATES would include Beijing and Shanghai, where an early version of the technology was already used in the 1980s (Fleuchaus et al., 2018). For the Netherlands, the assessment indicated that the magnitude of energy savings which is targeted by current policy objectives for ATES and BTES (i.e. a total of 21 PJ/year by 2023) would technically be feasible under fairly conservative assumptions by 2050; however, achieving this level by 2023 would require a significant increase in retrofit and/or adoption rates.

In addition, the urban data was used to estimate the energy density of cities in suitable climate zones, as a potential indication of future demand for ATES, and therefore of demand for subsurface space. Under the assumptions for energy demand which were used for the analysis, the average energy density of major cities in the Netherlands would thus be lower than cities such as New York, Shanghai, or Tokyo – which have suitable climate and subsurface conditions for ATES, and significantly greater seasonal cooling and/or heating loads. In particular, Utrecht would be close to the median of the 556 simulated cities; as discussed in the context of the simulation case studies, the scarcity of space for ATES is already emerging as a barrier for adoption in this city and in certain other areas of the Netherlands – despite an average adoption level which currently corresponds to less than one-fifth of short-term policy targets. It is therefore plausible that the high-potential areas highlighted in Chapter 8 could eventually experience similar pressures, depending on local aquifer conditions which were beyond the scope of this research; this emphasizes the importance of improved methods for planning and operation in order to increase the density of ATES development. The methods discussed in this thesis could thus find a much broader application beyond the Netherlands.

9.2. Reflection

9.2.1. Methodological limitations and future outlook

This thesis followed a primarily quantitative approach to assess short-term challenges for the planning and operation of ATES, and outline a longer-term perspective in which ATES "smart grids" could better align public and private interests. The design of the research project unavoidably left some relevant aspects out of scope. As such, one of the original goals of this project was to evaluate the extent to which improved control systems (such as the centralized and distributed control frameworks tested in Chapter 7) could stimulate the adoption of ATES by building owners, by improving the efficiency of the systems and therefore their economic performance. However, as shown in Chapter 6 and in parallel research (Bloemendal et al., 2018), the recovery efficiency of ATES under current practices is affected more significantly by choices made in the design of

the wells, and by operational uncertainties (such as ΔT) which were out of the scope of the control frameworks; these uncertainties would have required a more detailed simulation of building energy systems to assess endogenously. From this perspective, as dynamically managing thermal interactions would not significantly improve efficiency, the gains from ATES "smart grids" were primarily associated with the greater density at which systems could be laid out without compromising their performance. The research then mostly focused on interpreting the repercussions of ATES "smart grids" for policymakers, rather than ATES operators.

This focus had several implications. For instance, Chapter 7 assumes that adoption decisions would not be affected by the introduction of ATES "smart grids"; in reality, considerations related to privacy and trust may play a significant role towards the perception and acceptance of smart energy systems (Milchram et al., 2018). This may especially be relevant for industrial users (Samad and Kiliccote, 2012; Ma et al., 2018). Given that ATES users frequently raised concerns about operational complexity in the initial scoping survey used for this thesis, the perceived complexity of a "smart grid" approach may also pose a barrier. Operational complexity was generally left out of scope of the control models, which assumed that the operation of building systems was purely an economic optimization problem. However, discussions with stakeholders made it clear that the daily operation of ATES is driven by comfort and reliability rather than economic considerations, and presented more complex dynamics than could be included in the building control model. This yields some of the operational outcomes which were treated as exogenous uncertainties in the case studies, such as thermal imbalances and a partial usage of the allowed storage capacity.

The implications of privacy considerations and operational complexity would have required dedicated research, taking a more stakeholder-focused perspective on the adoption and operation of systems. Considering the current challenges facing ATES in the Netherlands – such as a negative perception of the technology's reliability (Agterberg, 2016) – such research would be useful to understand ATES adoption dynamics in general. This future research could also improve the depiction of adoption dynamics in simulation models of ATES development. While discrete-choice analysis is increasingly used in combination with agent-based technology models, the highly site-specific nature of ATES would make it more difficult to apply in this context, compared to more easily standardized energy technologies. However, attitude-based models such as the theory of planned behavior (Ajzen, 1991) offer a promising avenue for the study of ATES adoption across different control paradigms, and have been applied to similar problems (such as Sopha et al. (2011)'s study of sustainable heating system adoption). In parallel, operational decisions could be depicted more realistically in more comprehensive building control models, which would represent the building's physical systems at a higher level of detail. Heat pump performance is for instance a key aspect of ATES operation, which was here left out of scope by assuming constant coefficients of perfor-

mance. The integration of realistic building equipment models is presently the focus of a follow-up research project as part of the URSES+ research program, which includes an ATES case study for Amsterdam's Van Gogh museum.

On a technical level, leaving adoption dynamics out of the scope of Chapter 7 also means that the research essentially side-stepped an issue which is frequently found in computationally expensive models for socio-technical and social-ecological systems – namely reconciling the time scales which are required to depict short-term operational decisions (such as the use of storage systems at an hourly or daily scale), and the long-term dynamics of investment or environmental processes (which, in the case of ATES, can imply an horizon of several decades). The latter were addressed in the longer-term simulation case studies of Chapters 5 and 6, but at a temporal resolution which would be too coarse to accurately simulate the dynamic management of thermal interactions.

The most practical resolution of this compromise would likely involve a meta-modelling approach, to enable a computationally tractable simulation of the control frameworks over multiple decades. Preliminary work with a single-building model pointed towards gradient-boosted regression trees (Friedman, 2001) as a promising option to approximate ATES well flows as well as extraction temperatures, based on a given building energy demand profile. Due to the regular rectilinear grid scheme imposed by MODFLOW, the discretization used to accurately model temperature profiles around ATES wells leads to computationally inefficient grids. A reliable meta-model could thus significantly reduce the runtimes associated with the MODFLOW/SEAWAT groundwater model, in addition to the control component. In turn, this meta-modelling approach could have supported a more comprehensive treatment of uncertainty in Chapter 7, by using exploratory modelling to simulate and interpret a broader range of scenarios for e.g. building energy demand or spatial planning, as in Chapters 5 and 6.

Another direction of research could explore alternatives to the co-simulation approach which was used in this thesis to couple models implemented in different formalisms – for instance by instead translating the models into a shared "common denominator" formalism. As noted by Rizzoli et al. (2008), the translation and reimplementation of the individual models under such a common formalism would entail significant development effort. Nonetheless, this approach has some precedent for groundwater modelling. For instance, Castilla-Rho et al. (2015) implemented the groundwater flow equations using NetLogo as a finite-difference solution scheme, while the iMOD (Vermeulen and Minnema, 2015) implementation of MODFLOW's finite-difference approach improves performance by avoiding the processing of input/output files. The discrete event system specification (DEVS) formalism (Zeigler, 1990) would provide a technically suitable "common denominator" for this purpose, with benefits for the scalability and interoperability of the resulting system model. Extensions of this formalism have been used to support the simulation of large-scale

multi-agent systems (Zhang et al., 2013b), and the solution of ordinary or partial differential equations under a quantization scheme (Vangheluwe et al., 2002; Bolduc and Vangheluwe, 2002); these could address the key model components which were used in this thesis. Relevant variants have also successfully been applied for the hybrid modelling of complex environmental systems (ElSawah et al., 2012; Filippi and Bisgambiglia, 2004).

We note that model runtimes also limited the use of more sophisticated methods for decision making under deep uncertainty, such as many-objective robust decision making or robust optimization (e.g. Kasprzyk et al., 2013; Beyer and Sendhoff, 2007). The latter method was tested for the idealized case study, in order to estimate a Pareto-optimal set of optimized well layout policies across a subset of uncertainties (Jaxa-Rozen et al., 2016). This work yielded promising results – for instance by pointing towards the role of positive mutual thermal interactions for making systems more robust to economic uncertainties – but the optimization was ultimately limited by computational resources. To address this issue, improved scenario sampling techniques could for instance facilitate the use of optimization to search for robust policy options in a multi-scenario, many-objective robust decision making framework (Watson and Kasprzyk, 2017; Eker and Kwakkel, 2018; Bartholomew et al., 2018).

9.2.2. ATES in the energy system

The research adopted a relatively narrow view of ATES in relation to the broader transition towards more sustainable modes of energy use in the built environment. Chapters 5 and 6 thus assumed a binary adoption decision between ATES or conventional energy systems (i.e. a natural gas boiler and compression chiller), and the assessment framework used in the case studies focused on relative economic and energy performance compared to conventional energy. In reality, the diffusion of ATES is subject to multiple pressures, such as competition from increasingly efficient air-source heat pumps, which are easier to integrate with conventional building systems. This is particularly relevant for renovated buildings, which are a key component towards long-term energy targets for the built environment due to the inertia of the building stock. Agterberg (2016) presented a simple comparative analysis of ATES/BTES and air-source heat pumps, noting that the superior environmental performance of ATES is currently not reflected in the business case for the choice of building energy systems – leading to a market failure which could potentially be addressed by higher carbon pricing.

This research also primarily interpreted challenges for ATES spatial planning by focusing on the management of thermal interactions between ATES systems, and – for the Utrecht case study – on the impact of surface-level restrictions on system layout. However, as described by Bonte et al. (2011), ATES spatial planning in dense urban environments also requires balancing competing uses of the subsurface, as well as potential opportunities for e.g. heat recovery from subsurface structures or the re-

mediation of groundwater contamination (e.g. Slenders et al., 2010). Further research could take a broader view of underground thermal storage in the context of urban planning, and investigate best practices towards identifying and including such cross-sectoral constraints and opportunities in the spatial planning of ATES. The expected future development of higher-temperature ATES systems will also require revisiting the criteria used to assess the collective or environmental performance of ATES, as higher storage temperatures can be expected to magnify the chemical or microbiological impacts on groundwater which were left out of the scope of this thesis.

More broadly, the development of ATES "smart grids" would occur in the context of a rapidly evolving energy system, in which the distributed generation and storage of thermal energy are becoming increasingly relevant (see e.g. Howell et al. (2017) for a comprehensive review of different paradigms for distributed energy systems). As discussed by Rostampour and Keviczky (2017), the distributed stochastic control framework used in Chapter 7 can easily be extended to cover different sources and sinks of thermal energy – for instance in the context of a smart thermal grid, in which building operators could directly exchange energy through piped connections in parallel to the coordinated operation of ATES wells. Improved methods for ATES operation can help provide more flexibility in this setting, by complementing short-term buffer storage with seasonal storage and by integrating additional sources of heat to be stored, such as solar thermal collectors. Further work in this direction could find a useful case study in "district" ATES systems (Velvis and Buunk, 2017), in which several networked buildings share common ATES wells, and which are increasingly used in the Netherlands.

As an alternative approach to distributed control, we note that the use of electronic equilibrium markets to match thermal supply and demand has drawn increasing attention in the literature. For instance, the HeatMatcher concept (van Pruissen et al., 2014; Booij et al., 2013), in which autonomous agents networked through a thermal grid can bid to supply or demand a given amount of energy, was successfully tested in a small-scale pilot project. The coupling constraints used to coordinate ATES operation in Chapter 7 would *a priori* be compatible with such market mechanisms; however, an interesting option to explore could be to replace the "hard" constraints on well overlap which were used in this work, with soft constraints which would instead represent thermal interferences through an increasing marginal cost of supply for ATES.

Finally, the greatest potential of coordinated ATES operation is perhaps as a key building block towards a self-organized regime for the management of ATES "smart grids". Self-organization could provide a more flexible approach for the operation and planning of ATES systems, under which provincial authorities could for instance delegate much of their current administrative responsibilities to groups of users, and rely on coordination to manage the impacts of ATES development. However, this research did not investigate the institutional factors which would be required to suc-

cessfully transition to such a regime. For instance, to accommodate changes in ATES use over time, self-organization would need to rely on structured feedback mechanisms and compensation arrangements between users (which could technically be supported by distributed control and/or market mechanisms). The simple game-theoretical analysis used in Chapter 7 sketched out some of these issues, but conceptualizing a self-organized approach will require dedicated research centered on institutional aspects.

In this regard, Ostrom (2009)'s framework for social-ecological systems (SESs) would be a useful guide for further research, given that urban ATES systems incorporate the key features of a SES (such as dependence on a natural resource). This framework was designed to identify the basic elements and relationships which drive SESs, and provides a common set of variables for their multidisciplinary analysis. As such, it could help synthesize the work presented in this thesis with further qualitative or quantitative research on ATES adoption and operation, and draw parallels with best practices in regards to the sustainable management of SESs (e.g. Bal, 2015).

9.3. Policy recommendations

As implemented under the WBBE policy framework, current planning methods for ATES in the Netherlands follow a conservative approach, with a strong emphasis on protecting existing functions of the subsurface and avoiding negative interactions between systems. This framework was a suitable starting point for large-scale ATES management in the Netherlands; it also generally provides an appropriate reference for other countries where the technology is still at an emerging stage (Fleuchaus et al., 2018). However, revisions to this framework will eventually be needed to match the transition of the technology towards a more mature stage of market development. As such, at the onset of this research, there was already some evidence that current policies may hinder the further development of ATES systems in urban areas – particularly at smaller storage capacities – due to a combination of restrictive spatial planning and administrative overhead. While the latter aspect was largely out of scope, the simulation case studies supported this assessment of the current spatial planning guidelines. Given that the current use of ATES in the Netherlands is less than one-fifth of the level which would be required to meet national targets for the technology by 2023, there is thus clearly a need to ensure that the policy framework for ATES will be consistent with longer-term objectives for the technology by promoting its further adoption.

In the short term, the scarcity of subsurface space for new systems is increasingly becoming a barrier to further adoption in areas which combine a relatively shallow aquifer with dense energy demand, such as the Utrecht city center. As shown in Chapters 5 and 6, and by parallel research on ATES well design parameters (Bloemendal et al., 2018), this scarcity is partly artificial and could be addressed in two ways: first, nominal well layout guidelines can be relaxed to a well distance of 2.5 R_{th}, with a minimal impact on thermal and economic performance under idealized operating conditions.

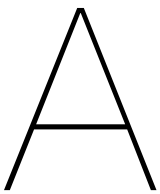
Secondly, as discussed in Chapter 6, well layouts in realistic conditions may effectively be much sparser than these nominal guidelines, due to geographic constraints and operational uncertainties – e.g. the actual use of storage capacity and/or imbalances between heating and cooling. This could be addressed through an adaptive permitting approach (which is currently used on an *ad hoc* basis in some areas) to periodically revise allocated storage capacities, following actual usage. This would ensure that building owners have less of an incentive to overestimate the required storage capacity in permit applications. An adaptive approach should nonetheless leave sufficient leeway for operators to meet annual variations in energy demand, as well as longer-term demand trends: as shown in Chapter 7, climate change may lead to a sustained shift towards cooling demand in the Netherlands by 2040, which is within the expected lifetime of systems currently being installed. These trends should be taken into account in the planning and design of systems, as they will make it more difficult to meet thermal balance requirements and may for instance require transferring excess heat to surface water.

In the longer term, coordinated operation in the context of ATES "smart grids" can lead to a more efficient resolution of the trade-off between private and public interests, by allowing a denser use of the subsurface and ultimately by supporting a more streamlined, self-organized management regime. This would yield a "win-win" for system owners, as well as policymakers: the former could preserve the technical and economic performance provided by current policies, while benefiting from greater operational flexibility; the latter could increase the contribution of ATES to objectives for energy efficiency and GHG reductions, while avoiding the administrative overhead of an adaptive permitting strategy.

However, a shift towards self-organization will imply a fundamental revision of the policy framework for ATES, and still requires significant research on technical and institutional aspects. As ATES increases its contribution towards targets for sustainable energy and moves to a more mature stage in the Netherlands, this revision would also yield an opportunity to revisit ATES-specific policies in relation to broader sectoral policies for heating, energy efficiency, and urban planning. In the interim, municipal, provincial, and national authorities can play an important role by stimulating the development of ATES "smart grid" demonstration projects to raise awareness, and explore the broader complementarities of coordinated ATES operation with smart thermal grids. In parallel, these demonstration projects could support applied research in order to explore the legal and institutional conditions under which ATES "smart grids" would be acceptable.

While the Netherlands are currently the most developed market for ATES, these recommendations are at least equally relevant for emergent markets: the adoption of ATES yields a limited window of opportunity to apply improved methods for planning and operation, as the development of the technology is strongly path-dependent due to the physical layout of systems – so that policymakers should strive to avoid "lock-in"

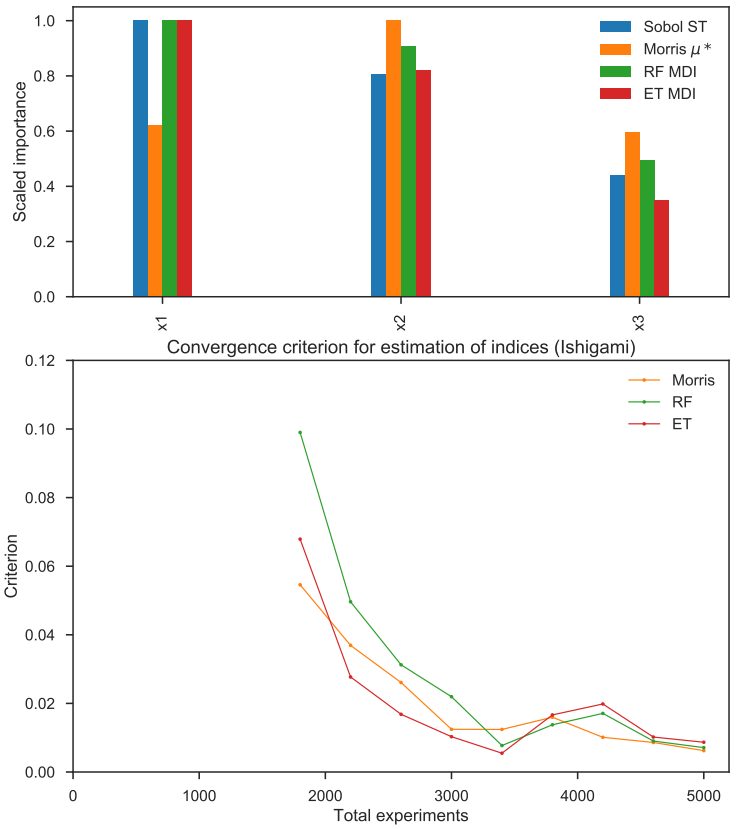
under suboptimal methods. In particular, as shown in Chapter 8, the spatial density of demand for ATES observed in the Netherlands – and the resulting pressure on available subsurface space – is eventually likely to be found across other areas which are suitable for ATES; policymakers should thus account for this possibility when designing a framework for the local management of the technology. This will enable new markets for ATES to benefit from improved practices for planning and operation from an early stage of development.



Appendix for Chapter 4: Additional sensitivity analysis results

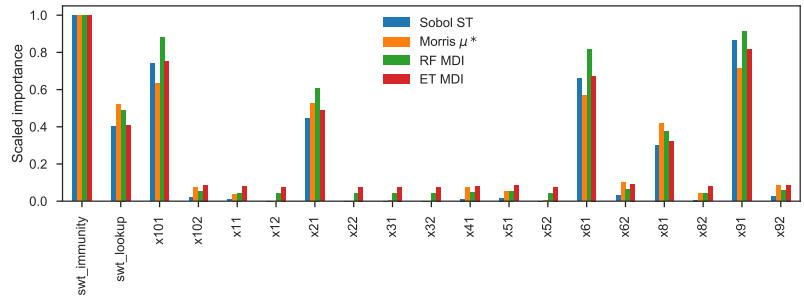
Variable importance metrics

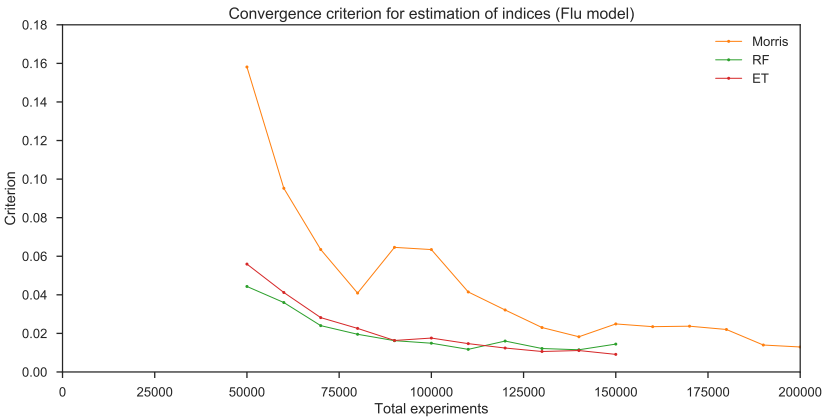
Figures A.1, A.2 and A.3 present detailed results for the estimation of scaled variable importances in each case study.



	Sobol ST (N=17000)	Morris $\mu*$ (N=5000)	RF MDI (N=5000)	ET MDI (N=5000)
RMSE	-	0.250	0.069	0.051
MBE	-	-0.016	-0.054	0.021

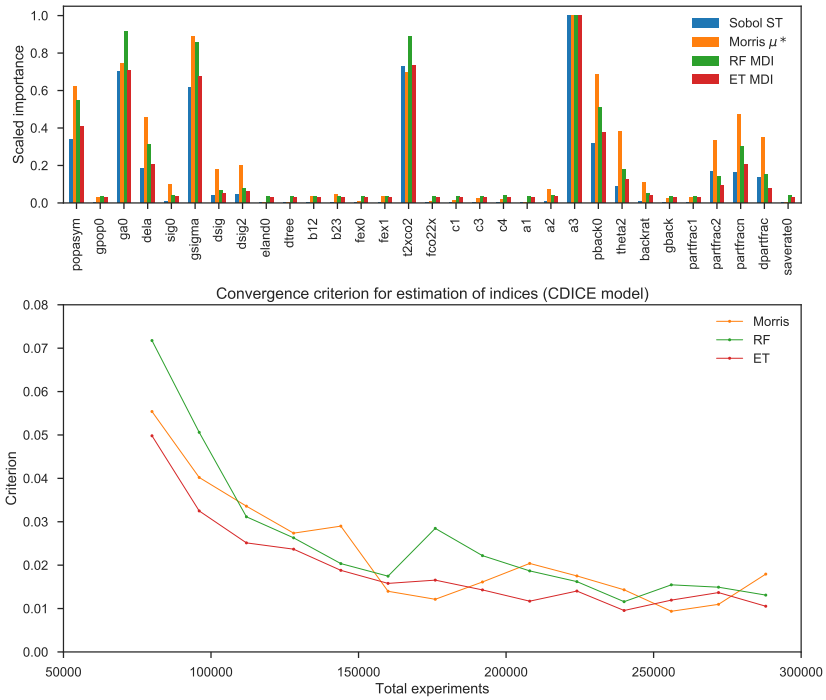
Figure A.1: Scaled variable importances and error measures for Ishigami function.





	Sobol ST (N=8e5)	Morris $\mu*$ (N=2e5)	RF MDI (N=1.5e5)	ET MDI (N=1.5e5)
RMSE	-	0.081	0.084	0.056
MBE	-	0.017	-0.056	-0.041

Figure A.2: Scaled variable importances and error measures for H1N1 flu model.



	Sobol ST (N=9.2e6)	Morris $\mu*$ (N=2.9e5)	RF MDI (N=2.9e5)	ET MDI (N=2.9e5)
RMSE	-	0.149	0.094	0.035
MBE	-	-0.099	-0.064	-0.023

Figure A.3: Scaled variable importances and error measures for CDICE model.

Comparison of analysis runtimes

Table A.1 shows typical runtimes for each of the key analyses, using the EMA Workbench 1.1 library to sample and simulate experiments, and SALib 1.1.3 (for Sobol and Morris) and scikit-learn 0.18.1 (for Extra-Trees) to compute sensitivity indices. The analyses were performed on an Intel Xeon E5-2620 CPU with the Anaconda 3.6 Python 64-bit distribution.

	H1N1 flu model		CDICE model	
	Model evaluation (s)	Analysis (s)	Model evaluation (s)	Analysis (s)
Sobol (S1, S2, ST)	8778 (N=8e5)	105	4661 (N=9.22e6)	735
Morris ($\mu*$, σ)	2131 (N=2e5)	5.8	126 (N=2.88e5)	7.8
ET (MDI importances)	1614 (N=1.5e5)	6.4	128 (N=2.88e5)	9.6
ET (pairwise MDA)	1059 (N=5e4)	16.4	42.4 (N=1e5)	51.8

Table A.1: Typical runtimes for the H1N1 and CDICE test cases, and for the computation of sensitivity indices. The sample size used in each case is indicated in parentheses.

B

Appendix for Chapter 5: ODD+D documentation for idealized model

ODD+D model documentation

Outline		Guiding questions	Description
I) Overview	I.i Purpose	I.i.a What is the purpose of the study?	Aquifer Thermal Energy Storage is an increasingly popular form of shallow geothermal energy, which relies on underground aquifers to store thermal energy for the heating and cooling of buildings. However, the rapid growth of this technology in urban areas has raised problems with interferences between neighboring systems as well as long-term thermal imbalances; the thermal storage potential of the subsurface can be perceived as a common-pool resource, the long-term sustainability of which may be threatened under current planning policies for ATES. The study therefore aims to explore plausible interactions between the adoption of ATES systems and subsurface conditions. The model is currently parameterized to represent an idealized case, and will later be extended into realistic case studies to support ongoing ATES development in the city centers of Amsterdam and Utrecht.
		I.ii.b For whom is the model designed?	Specialists of shallow geothermal energy, as well as researchers interested in common-pool resource governance and technology diffusion.
	I.ii Entities, state variables, and scales	I.ii.a What kinds of entities are in the model?	Agent-based model layer: ○ ATES system operators ○ ATES wells ○ Land parcels Geohydrological model layer: ○ Aquifer model grid ○ ATES wells
		I.ii.b By what attributes (i.e. state variables and parameters) are these entities characterized?	ATES system operators (ABM layer): ○ Geographic coordinates ○ Set of land parcels owned by the operator ○ Set of ATES wells owned by the operator ○ ATES adoption status: current adopter, potential adopter, previous adopter ○ Adoption threshold (ROI period) for construction of ATES wells ○ Current thermal performance of ATES wells ○ Memory of previous thermal performance ATES wells (ABM layer): ○ Physical properties: geographic coordinates, type (warm/cold), temperature, flow ○ Design temperatures, setpoints, and setpoint calculation period for heating / cooling ○ Thermal performance: energy injected/lost/recovered ○ Aquifer properties at own location: temperature, head Land parcels (ABM layer): ○ Geographic coordinates

			○ Status: occupied by building, available for any wells, available only for cold wells, or available only for warm wells Aquifer model grid (geohydrological layer): ○ Geohydrological properties: horizontal and vertical conductivities, porosity ○ Temperature and head distributions ATES wells (geohydrological layer): ○ Physical properties read from properties of ATES wells in the ABM layer (geographic coordinates, temperature, flow)
		I.ii.c What are the exogenous factors / drivers of the model?	○ Daily temperature (based on KNMI W+ 2010-2045 climate scenario) ○ Design temperatures for heating and cooling ○ Planning policies (minimal distances between ATES wells of the same type or opposite type) ○ Cost data for construction and operation of ATES or conventional heating/cooling system ○ Electricity and gas prices
		I.ii.d If applicable, how is space included in the model?	ATES system operators and ATES wells are spatially located within the agent-based model layer. The wells are mapped to corresponding locations within the aquifer model.
		I.ii.e What are the temporal and spatial resolutions and extents of the model?	The coupled simulation is executed for 180 periods of 30 days. Each spatial layer is modelled as follows: Agent-based model layer: • Rectangular grid area of 1000m x 1000m, discretized in land parcels of 10m x 10m Geohydrological model layer: • Rectangular grid with dynamic discretization around ATES wells, with cell sizes varying from 5m to 20m. Dynamic extents to provide a minimal allowance of 200m around ATES wells.
	I.iii Process overview and scheduling	I.iii.a What entity does what, and in what order?	For each monthly period: 1) Agent-based model layer (asynchronous updating with random execution order): ○ ATES wells calculate their flow based on climate data, design temperature, and heating/cooling setpoints ○ ATES wells calculate their efficiency and effective energy cost based on geohydrological model results from previous period ○ ATES system operators calculate the overall thermal efficiency of their ATES wells and the corresponding ROI period, taking into account the annualized cost of the ATES system relative to a conventional energy system ○ ATES system operators build new wells if their expected system performance exceeds an individual adoption threshold (i.e. is below a given ROI period). Existing wells

			are activated if they are more profitable than an equivalent conventional energy system ○ ATES system operators update their expected system performance based on current performance 2) Hydrologic layer: ○ If the set of active ATES wells has changed: re-discretize the simulation grid ○ Based on the new well properties, update the temperature and head distributions within the aquifer
II) Design concepts	II.i Theoretical and Empirical Background	II.i.a Which general concepts, theories or hypotheses are underlying the model's design at the system level or at the level(s) of the submodel(s) (apart from the decision model)? What is the link to complexity and the purpose of the model?	The heterogeneity of ATES system operators is based on diffusion theory (Rogers, 2010). In parallel, the dynamic hypothesis for the system's behavior under certain parameterizations (e.g. the possible emergence of a "tragedy of the commons") follows the literature on common-pool resource governance (e.g. Hardin, 1968; Ostrom, 2009).
		II.i.b On what assumptions is/are the agents' decision model(s) based?	The decision model assumes that ATES system operators are boundedly rational (Simon, 1955); the adoption of new ATES wells is based on "satisficing" thresholds and past performance without explicit foresight, and with limited knowledge of subsurface conditions.
		II.i.c Why is a/are certain decision model(s) chosen?	There is currently a lack of specific data on ATES adoption processes. However, as commonly described in the literature (DeCanio 1993, 1998; Wustenhagen and Menichetti, 2011), organizational energy investments typically follow an imperfect approximation of rational economic theory. Furthermore, system performance is affected by thermal interactions between wells, which are driven by adoption patterns (and which are themselves an emergent and unforeseeable property of the system). Bounded rationality thus provides a useful framework for the adoption processes.
		II.i.d If the model / a submodel (e.g. the decision model) is based on empirical data, where does the data come from?	The distribution of adoption thresholds across ATES operators approximates the empirical results of Blok et al. (2004) for the critical payback periods expected by firms investing in energy-efficient technologies.
		II.i.e At which level of aggregation were the data available?	Individual industrial firms
	II.ii Individual Decision Making	II.ii.a What are the subjects and objects of decision-making? On which level of aggregation is decision-making modeled? Are multiple levels of decision making included?	Decision-making is modelled at the level of ATES systems, who are assumed to correspond to individual building operators. When deciding to activate wells, the operators uniformly change the status of all wells under their control.

		II.ii.b What is the basic rationality behind agents' decision-making in the model? Do agents pursue an explicit objective or have other success criteria?	ATES system operators attempt to satisfy a given payback period for the construction of new wells, by comparing the expected annualized investment and operational costs of an ATES system with a conventional energy system.
		II.ii.c How do agents make their decisions?	ATES system operators compare their expected system performance (expressed as a ROI period) with their adoption thresholds. Depending on their adoption status, they may then decide to build additional wells if the expected ROI period is below the adoption threshold, or activate their existing wells based on expected profitability compared with a conventional energy system. For current adopters, the expected system performance is an exponential moving average of the realized ROI period, based on current operational costs for the ATES system and investment costs. For potential adopters, this value is assumed to be the average expected performance of all current adopters.
		II.ii.d Do the agents adapt their behavior to changing endogenous and exogenous state variables? And if yes, how?	ATES well agents update their flow at each period in order to maintain thermal balance over the setpoint calculation period, given the exogenous temperature input. ATES system operators decide whether to activate existing wells or build new wells depending on their expected system performance, which is driven by endogenous aquifer conditions and exogenous system costs and energy prices.
		II.ii.e Do social norms or cultural values play a role in the decision-making process?	N/A
		II.ii.f Do spatial aspects play a role in the decision process?	New ATES wells may only be built on available land parcels which are owned by the system operator. The subset of available parcels is further restricted by planning policies for the minimum clearances between wells, defined as a multiplier of the average radius around each well within which aquifer temperature is significantly affected.
		II.ii.g Do temporal aspects play a role in the decision process?	The expected system performance is calculated using a given exponential smoothing factor, which is assumed to be uniform across all ATES operators. Delayed feedbacks are present in the aquifer model (due to the evolution of temperature distributions over time), but are not considered in the decision process of agents.
		II.ii.h To which extent and how is uncertainty included in the agents' decision rules?	Uncertainty is not explicitly considered in the decision rules.
	II.iii Learning	II.iii.a Is individual learning included in the decision process? How do individuals change their decision rules over time as consequence of their experience?	N/A

		II.iii.b Is collective learning implemented in the model?	N/A
	II.iv Individual Sensing	II.iv.a What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions? Is the sensing process erroneous?	ATES operators perceive aquifer conditions (temperature and head) at the location of the wells under their control, as well as exogenous energy prices and cost data. Well agents perceive the exogenous time series for outside temperature. Sensing errors are not explicitly modelled; however, as operators only perceive aquifer conditions at each well, they only have limited information about subsurface conditions.
		II.iv.b What state variables of which other individuals can an individual perceive?	Potential ATES adopters can perceive the expected system performance of current adopters.
		II.iv.c What is the spatial scale of sensing?	Local for physical properties; global for expected performance
		II.iv.d Are the mechanisms by which agents obtain information modeled explicitly, or are individuals simply assumed to know these variables?	Information sharing mechanisms are not explicitly modelled for this case study, and the expected system performance is assumed to be shared without error.
		II.iv.e Are costs for cognition and costs for gathering information included in the model?	N/A
	II.v Individual Prediction	II.v.a Which data uses the agent to predict future conditions?	ATES operators use past system performance as an indicator for adoption. Foresight is not explicitly modelled.
		II.v.b What internal models are agents assumed to use to estimate future conditions or consequences of their decisions?	N/A
		II.v.c Might agents be erroneous in the prediction process, and how is it implemented?	The realized system performance will differ from the expected performance due to variable weather conditions (which lead to variable well flows), and due to thermal interaction effects between wells.
	II.vi Interaction	II.vi.a Are interactions among agents and entities assumed as direct or indirect?	ATES operators interact directly through the sharing of information about expected performance in the agent-based model layer. They also interact through the geohydrological model layer, due to positive or negative thermal interactions between wells.
		II.vi.b On what do the interactions depend?	Thermal interactions depend on the location and flow properties of ATES wells, and on the geohydrological properties of the aquifer.
		II.vi.c If the interactions involve communication, how are such communications represented?	N/A

		II.vi.d If a coordination network exists, how does it affect the agent behaviour? Is the structure of the network imposed or emergent?	N/A
	II.vii Collectives	II.vii.a Do the individuals form or belong to aggregations that affect, and are affected by, the individuals? Are these aggregations imposed by the modeller or do they emerge during the simulation?	ATES operators may transition between adoption status groups over the course of the simulation (i.e. from potential adopter to adopter to previous adopter). ATES wells are grouped into sets belonging to a given ATES operator.
		II.vii.b How are collectives represented?	N/A
	II.viii Heterogeneity	II.viii.a Are the agents heterogeneous? If yes, which state variables and/or processes differ between the agents?	All attributes listed under I.ii.b are heterogeneous across agents. However, the agents share common decision rules.
		II.viii.b Are the agents heterogeneous in their decision-making?	All ATES system operators use the same decision model, but with heterogeneous adoption thresholds.
	II.ix Stochasticity	II.ix.a What processes (including initialization) are modeled by assuming they are random or partly random?	Decision thresholds for system operators are initialized using a given random distribution to differentiate thresholds across agents. New well pairs are created with a random flow capacity, and at random locations (within the rules for minimal distances between wells).
	II.x Observation	II.x.a What data are collected from the ABM for testing, understanding, and analyzing it, and how and when are they collected?	ATES system operators ○ Number of active wells ○ Realized system performance / effective energy price ○ Reduction in CO_2 emissions ATES wells ○ Thermal performance: energy injected/lost/recovered Aquifer ○ Fraction of total aquifer volume used for thermal storage ○ Temperature and head distributions The data are collected at each time step in each model layer.
		II.x.b What key results, outputs or characteristics of the model are emerging from the individuals? (Emergence)	ATES adoption dynamics are affected by the minimal required distance between wells. Insufficient distances may result in rapid adoption followed by a collapse in performance, as interference between wells eventually reduces thermal efficiency. Larger distances will increase the individual efficiency of systems but may penalize collective performance in terms of total energy savings, as operators may be unable to find suitable locations for new wells.

III) Details	III.i Implementation Details	III.i.a How has the model been implemented?	Agent-based model layer: NetLogo 5.1.0 Geohydrological model layer: Modflow/MT3DMS/SEAWAT The two layers are linked through an object-oriented architecture developed using Python 2.7, with NetLogo agents mapped to corresponding Python objects. The FloPy library (Bakker et al., 2013) provides a pre-/post-processing interface between Python and the geohydrological model. The hybrid architecture is executed using the EMA Workbench suite for experiment design and exploratory modelling and analysis (Kwakkel, 2015).
		III.i.b Is the model accessible and if so where?	N/A
	III.ii Initialization	III.ii.a What is the initial state of the model world, i.e. at time t=0 of a simulation run?	The model is initialized with 10 randomly located ATES operator agents. Two randomly selected operators are assumed to already be active adopters at the start of the simulation, with an initial set of 6 wells.
		III.ii.b Is initialization always the same, or is it allowed to vary among simulations?	The operator agents are initialized with random adoption thresholds, drawn from a distribution of adoption thresholds which approximates the results of Blok et al. (2004); agent locations are random.
		III.ii.c Are the initial values chosen arbitrarily or based on data?	The initial data for wells and systems is intended to approximate a typical configuration for ATES systems in a dense urban environment (e.g. Bloemendal et al., 2014).
	III.iii Input Data	III.iii.a Does the model use input from external sources such as data files or other models to represent processes that change over time?	An external Excel file provides initial data for wells and systems, as well as a time series for daily temperature (based on the KNMI W+ climate scenario).
	III.iv Submodels	III.iv.a What, in detail, are the submodels that represent the processes listed in 'Process overview and scheduling'?	○ The setpoint module calculates heating and cooling setpoints for each well on an annual basis, in order to maintain the thermal balance of inflows and outflows over a given period. ○ The well flow module then calculates the average daily flow for each well at each simulation period, based on the setpoints and on the exogenous temperature input.
		III.iv.b What are the model parameters, their dimensions and reference values?	The model parameters are set in an external spreadsheet with representative parameters for ATES systems in urban areas.
		III.iv.c How were submodels designed or chosen?	The setpoint and well flow submodels are adapted from the mfLab suite (Olsthoorn, 2014). These will eventually be replaced by a distributed Model Predictive Control design implemented in Matlab.

C

Appendix for Chapter 6: Assessment framework for Utrecht case study

This framework describes key performance indicators for ATES systems, building on an earlier version presented in Bloemendal et al. (2018).

Energy use and emissions of ATES systems

The energy balance of the heat pump is used to trace back the heating and cooling demand (E_h, E_c) of the associated buildings and the energy consumption by the heat pump. The total heating capacity for the building provided by the heat pump is described by two basic relations:

$$P_h = P_{ATES} + P_e \quad \& \quad COP_{hp} = \frac{P_h}{P_e} \tag{C.0.1}$$

where P_h [W] is the heating capacity deliverable to the building; P_{ATES} [W] is the thermal heating power retrieved from the groundwater, P_e [W] the electrical power consumed by the heat pump, and COP_{hp} the coefficient of performance of the heat pump. Equation E.0.3 shows that all electric power fed to the heat pump contributes to the heat output. When it is assumed that 100% of the heating and cooling demand of the building is delivered by the ATES system, the heating capacity and total heat energy ($E_{h,ATES}$) from the groundwater between times t and t_0 equals

$$E_{h,ATES}(t_0 \rightarrow t) = c_w \int_{t_0}^{t} P_{ATES} dt = c_w \Delta \bar{T}_h \int_{t_0}^{t} Q dt = c_w \Delta \bar{T}_h V_h \tag{C.0.2}$$

with

$$P_{ATES} = c_w Q(T_w T_c) = c_w Q \Delta T_h \tag{C.0.3}$$

The integration is done for the whole heating season ($t_0 \rightarrow t$). V_h [m^3] is the given seasonal volume of groundwater required for heating. ΔT_h [K] is the instantaneous temperature difference between the warm (T_w) and cold (T_c) well, $\Delta \bar{T}_h$ is the average temperature difference during heating season, Q [m3/h] is the groundwater flow from the warm well to the cold well and c_w [J/m^3/K] is the volumetric heat capacity of the water. With V_h substituted in equations E.0.3 and C.0.3, equation C.0.4 yields the heat E_h [J] delivered to the building over the heating season:

$$E_h = c_w \Delta \bar{T}_h V_h \frac{COP_{hp}}{COP_{hp} - 1} \tag{C.0.4}$$

The cooling delivered to the building is calculated using the same equations, while distinguishing between free cooling and heat pump cooling. An absolute temperature threshold of 9°C was set for the cold well above which no free cooling is assumed possible. When the extraction temperature of the cold well surpasses this threshold, the heat pump is used to meet the cooling demand and resulting heat is transferred to the warm well via the condenser of the heat pump. The total cooling delivered to the building then follows from:

$$E_c = c_w \Delta \bar{T}_{c,fc} V_{c,fc} + c_w \Delta \bar{T}_{c,hp} V_{c,hp} \frac{COP_{hp} - 1}{COP_{hp} - 2} \tag{C.0.5}$$

in which $V_{c,fc}$ and $V_{c,hp}$ are the groundwater volumes required for free cooling and cooling by the heat pump and $\Delta T_{c,fc}$ and $\Delta T_{c,hp}$ are the average temperature differences between the warm and cold well for free cooling and cooling by the heat pump, respectively. Note that the heat pump COP is 1 lower during cooling. The total energy consumption of the ATES system (E_{ATES}) is completed by including the pump energy consumption. Substituting equations E.0.3 into C.0.4 and C.0.5 yields:

$$E_{ATES} = \frac{E_h}{COP_{hp} - 1} + \frac{E_{c,hp}}{COP_{hp} - 2} + \frac{(V_h + V_{c,fc} + V_{c,hp})\Delta p}{\eta_p} \tag{C.0.6}$$

where Δp is the lifting pressure generated by the groundwater pump and η_p its nominal efficiency. The effective coefficient of performance of the ATES systems corresponds to the ratio between the quantities of energy which are delivered and used:

$$COP = \frac{E_h + Ec}{E_{ATES}} \tag{C.0.7}$$

In parallel, the energy efficiency (η) of a well over the simulation period is calculated in weekly steps by dividing the extracted amount of thermal energy by the infiltrated amount of thermal energy. The thermal efficiency taken over all the wells in the model (η_{tot}) is the average of the individual efficiencies, weighted by the individual total storage volume of the wells ($V_i = V_{h,i} + V_{c,fc,i} + V_{c,hp,i}$).

$$\eta_{tot} = \frac{\sum_{i=1}^{n} \eta_i V_i}{\sum_{i=1}^{n} V_i} \tag{C.0.8}$$

The equivalent GHG emissions [tCO_2] are retrieved by calculating the CO_2 emissions of the considered ATES systems:

$$GHG_{ATES} = \sum_{i=1}^{n} E_{ATES,i} f_e \tag{C.0.9}$$

where f_e [tCO_2/GJ] is the grid emission factor for electricity, E_{ATES} [GJ] is the electricity consumption of the ATES system, and n the number of active ATES wells. The calculation assumes a representative emissions factor for delivered electricity in the Netherlands.

Energy use and emissions of reference boiler/chiller systems

As a reference for technical and economic performance of ATES systems, the calculation considers a conventional climate control installation which would deliver the same amount of heating E_h [GJ] and cooling energy E_c [GJ] to the building. It is assumed that natural gas is used for heating in a boiler with combustion efficiency η_b, and that electricity is used for a cooling machine operating at a constant coefficient of performance COP_c. The energy consumption and greenhouse gas emissions [tCO_2] for these buildings then equal:

$$E_{boiler} = \frac{E_h}{\eta_b} \quad \& \quad E_{chiller} = \frac{E_c}{COP_c} \tag{C.0.10}$$

$$GHG_{conv} = \sum_{j=1}^{n} (E_{boiler,j} f_g + E_{chiller,j} f_e) \tag{C.0.11}$$

in which f_e [tCO_2/GJ] is the emission factor for gas and m the number of active conventional systems (which we here consider to be equal to the number of ATES systems).

Economic parameters

Operational costs for ATES and conventional systems can be computed similarly to GHG emissions, using the electricity price C_e [EUR/GJ] and natural gas price C_g [EUR/GJ] instead of the emission factors:

$$C_{ATES} = \sum_{i=1}^{n} E_{ATES,i} C_e \tag{C.0.12}$$

$$C_{conv} = \sum_{j=1}^{n} (E_{boiler,j} C_g + E_{chiller,j} C_e) \tag{C.0.13}$$

The economic efficiency of ATES can then be expressed as cost savings per total volume of water used for storage [EUR/m^3]:

$$\nu C = \frac{C_{conv} - C_{ATES}}{\sum_{i=1}^{n} V_{h,i} + V_{c,fc,i} + V_{c,hp,i}} \tag{C.0.14}$$

We note that this analysis focuses on operational costs only rather than upfront investment costs, given the high variability of fixed costs for ATES across different sites and buildings Agterberg (2016).

Collective performance indicators

The simulated GHG savings Δ_{GHG} [tCO_2] correspond to the difference between the emissions of conventional energy systems and ATES systems, for a given amount of delivered energy:

$$\Delta_{GHG} = GHG_{conv} - GHG_{ATES} \tag{C.0.15}$$

As a measure of the efficiency with which subsurface volume is used for thermal storage, these greenhouse gas savings can be expressed in relation to the aquifer volume allocated to ATES wells, using the distance policy d, well screen length L_i [m] and the total nominal storage volume of the wells V_i [m^3/yr]:

$$\nu_{GHG} = \frac{\Delta_{GHG}}{\sum_{i=1}^{n} \pi d R_{th,i}^2} = \frac{c_{aq}}{c_w} \frac{\Delta_{GHG}}{\sum_{i=1}^{n} \frac{V_i}{L_i}} \tag{C.0.16}$$

Finally, cost savings and GHG savings can be related as an equivalent GHG abatement cost α_{GHG} [EUR/tCO_2], which will be negative if cost and GHG savings from ATES are both positive:

$$\alpha_{GHG} = (C_{ATES} - C_{conv})/\Delta_{GHG} \tag{C.0.17}$$

D

Appendix for Chapter 7: Additional simulation results

R_{th}	2.125		2.25		2.375		2.5		2.75		3.0	
	CS	DS	CS	DS	CS	DS	CS	DS	CS	DS	CS	DS
Thermal eff.	0.972	0.942	0.988	0.957	0.992	0.969	0.991	0.978	0.994	0.994	1.0	1.0
COP	0.978	0.936	1.014	0.954	1.017	0.967	1.009	0.978	0.992	0.994	0.998	1.0
Tot. GHG savings	0.983	0.952	1.002	0.965	1.017	0.974	1.006	0.982	0.998	0.995	1.003	1.0
Spec. GHG savings	1.961	1.898	1.778	1.713	1.621	1.552	1.448	1.413	1.188	1.185	1.002	1.0

Table D.1: Results for idealized case study, relative to DS 3.0 R_{th}.

	Scenario 1 (3.0 Rth)		Scenario 2 (2.5 Rth)		Scenario 3 (2.25 Rth)	
	DSMPC	DS	DSMPC	DS	DSMPC	DS
Thermal efficiency	1.007	1.0	1.001	0.971	0.998	0.941
COP	1.010	1.0	1.015	0.968	1.034	0.949
Total GHG savings	0.990	1.0	1.060	1.159	1.212	1.290
Specific GHG savings	0.990	1.0	1.157	1.265	1.388	1.518

Table D.2: Results for Utrecht case study with $Q = 1.0$, relative to DS Scenario 1.

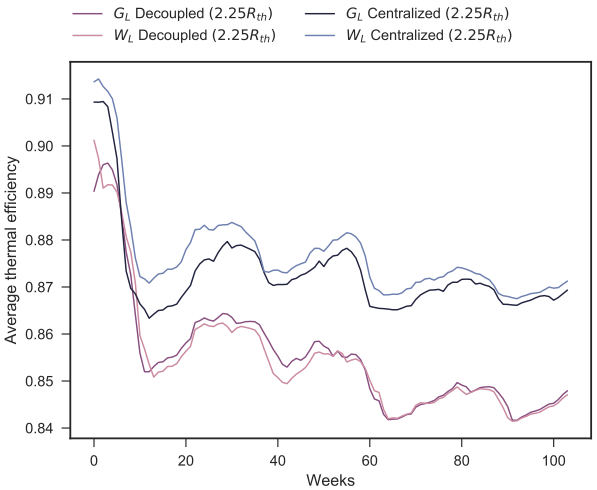


Figure D.1: Time series for average thermal efficiency over two annual storage cycles for idealized case.

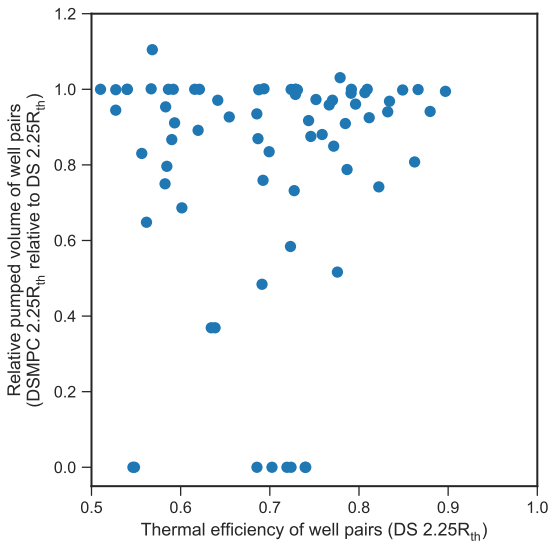


Figure D.2: Effect of well coupling constraints on pumped volume, as a function of decoupled thermal efficiency in Scenario 3 for Utrecht case. Each of the markers corresponds to a simulated well pair.

E

Appendix for Chapter 8: Additional methods and assessment results

Regional and climate data

Table E.1 lists the eight CMIP5 models used in the analysis, which were chosen based on the availability of daily minimum and maximum near-surface air temperature time series over the 2040-2050 period, for the three RCPs. Datasets for the first realization (r1) of each model were obtained through the Earth System Grid Federation-LLNL node (Williams et al., 2015).

Table E.1: CMIP5 models used to evaluate heating and cooling demand

Model	Institution	Resolution (lon. × lat., °)
CanESM2	Canadian Centre for Climate Modelling and Analysis	2.81×2.79
IPSL-CM5A-MR	Institut Pierre-Simon Laplace, France	2.5×1.25
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	1.41×1.39
MIROC-ESM		2.81×1.77
MIROC-ESM-CHEM		2.81×1.77
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	1.875×1.85
MPI-ESM-MR		1.875×1.85
NorESM1-M	Norwegian Climate Centre	2.5×1.875

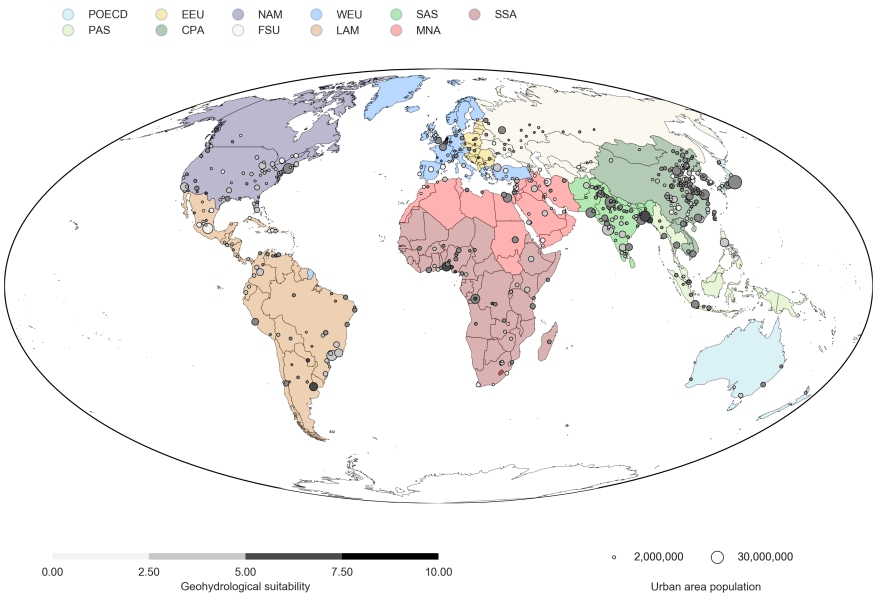
Based on the temperature indices, heating and cooling demand is represented with a degree-day approach (i.e. by integrating deviations over time from reference temperatures for heating and cooling). For each climate model, daily heating and cooling degree-days are computed at each grid point from the near-surface temperature time series in each RCP scenario, following the UK Meteorological Office approach (Day, 2006). To match data presented by Urge-Vorsatz et al. (2012) for building energy and climate zones, the reference temperatures Tb are set to 18°C and 10°C for heating and cooling degree-days (HDD18 and CDD10), respectively:

$$HDD = \begin{cases} T_b - \frac{1}{2}(T_{max} + T_{min}) & if\ T_{max} \leq T_b \\ \frac{1}{2}(T_{min} - T_b) - \frac{1}{4}(T_{max} - T_b) & if\ T_{min} < T_b\ \&\ (T_{max} - T_b) < (T_b - T_{min})\ \&\ T_{max} > T_b \\ \frac{1}{4}(T_b - T_{min}) & if\ T_{max} > T_b\ \&\ (T_{max} - T_b) > (T_b - T_{min})\ \&\ T_{min} < T_b \\ 0 & if\ T_{min} \geq T_b \end{cases} \tag{E.0.1}$$

$$CDD = \begin{cases} \frac{1}{2}(T_{max} + T_{min}) - T_b & if\ T_{min} \geq T_b \\ \frac{1}{2}(T_{max} - T_b) - \frac{1}{4}(T_b - T_{min}) & if\ T_{max} > T_b\ \&\ (T_{max} - T_b) > (T_b - T_{min})\ \&\ T_{min} < T_b \\ \frac{1}{4}(T_{max} - T_b) & if\ T_{min} < T_b\ \&\ (T_{max} - T_b) < (T_b - T_{min})\ \&\ T_{max} > T_b \\ 0 & if\ T_{max} \leq T_b \end{cases} \tag{E.0.2}$$

To estimate heating and cooling demand in each urban area, the HDD18 and CDD10 values are summed over the 2040-2050 period at each grid point and averaged on an annual basis, then interpolated to the coordinates of each urban area using

bilinear interpolation (following the approach of Petri and Caldeira (2015)). For each RCP scenario, the final heating and cooling demand estimates are then respectively given by the mean of the annual HDD18 and CDD10 values across all models in the ensemble. Figure E.3 presents the coefficient of variation of the total energy demand estimate (HDD18+CDD10) across models at each location. The standard deviation of the combined demand estimate at most locations is under 20% of the mean, indicating that the estimation is reasonably stable.



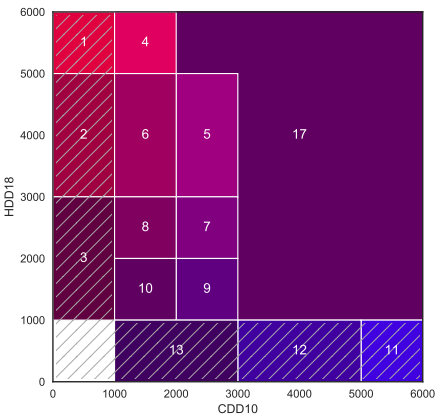
ID	CPA	EEU	FSU	LAM	MNA	NAM	PAS	POECD	SAS	SSA	WEU
Description	Centrally-Planned Asia and China	Central and Eastern Europe	Former Soviet Union	Latin America and Caribbean	Middle East and North Africa	North America	Other Pacific Asia	Pacific OECD countries	South Asia	Sub-Saharan Africa	Western Europe

Figure E.1: Standard GEA RC11 zones, and urban geohydrological suitability (IIASA, (2012); Bloemendal et al. (2015))

Following the estimation of mean heating and cooling demand, each urban area is then assigned to a given climate zone under each RCP scenario, following the classification of Urge-Vorsatz et al. (2012) which is presented in Figure E.2. Figure E.4 presents the robustness of the climate zone estimation across models, with most locations showing a consistent climate classification under at least half of the models.

To generate suitability maps for UTES under the different RCP scenarios, the HDD18 and CDD10 values over each model grid were regridded on a common 240×121

grid (1.5°×1.5° resolution) with a first-order conservative remapping (Jones, 1999), using the Climate Data Operators suite (Schulzweida et al., 2011). The demand values were then averaged across models at each grid point.



ID	Description	Suitability
1	*Only heating (very high heating demand)*	BTES only
2	*Only heating (high heating demand)*	
3	*Only heating (low and moderate heating demand)*	
4	Heating and cooling (very high heating, mostly low cooling)	ATES + BTES
5	Heating and cooling (high heating, mostly moderate cooling)	
6	Heating and cooling (high heating, low cooling)	
7	Heating and cooling (moderate heating, moderate cooling)	
8	Heating and cooling (moderate heating, low cooling)	
9	Heating and cooling (low heating, moderate cooling)	
10	Heating and cooling (low heating, low cooling)	
11	*Only cooling (very high cooling demand)*	Unsuitable
12	*Only cooling (high cooling demand)*	
13	*Only cooling (low and moderate cooling demand)*	
17	Heating and cooling and dehumidification	ATES + BTES

Figure E.2: Climate zone classification, based on Urge-Vorsatz et al. (2012)

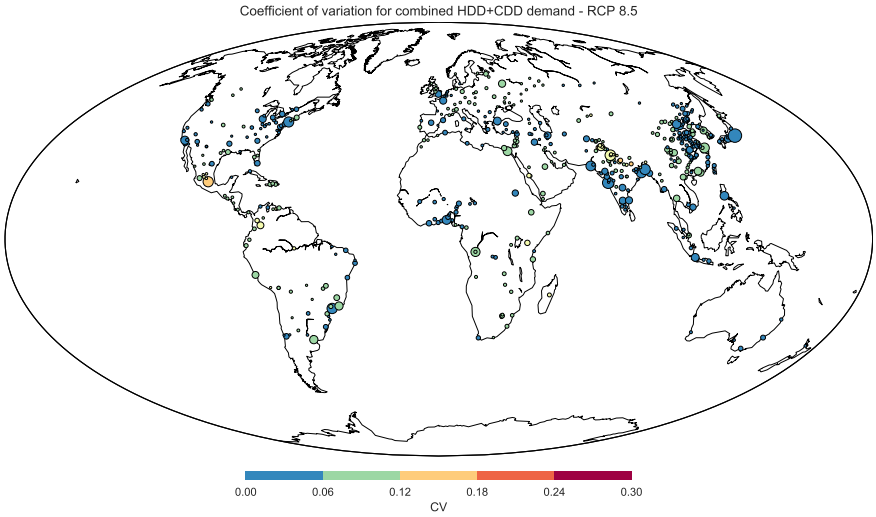


Figure E.3: Coefficient of variation for combined heating and cooling demand across climate models, at each urban area

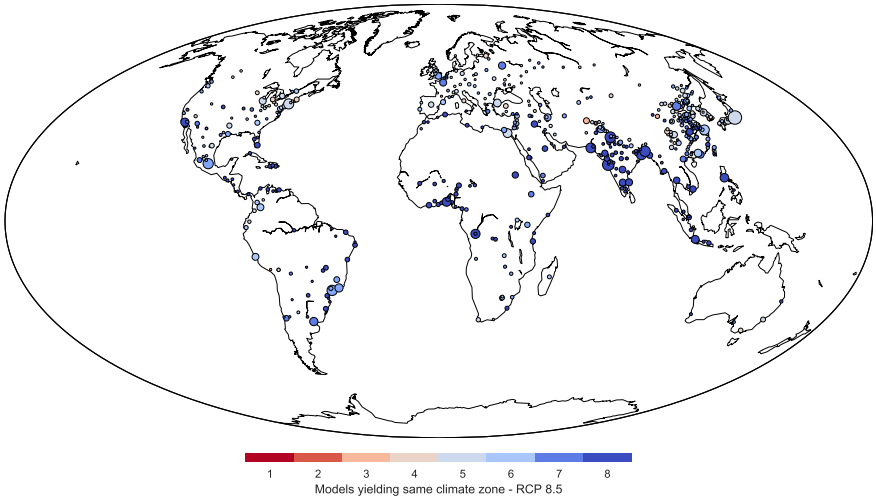


Figure E.4: Consistency of climate zone classification across climate models, at each urban area

Global and regional energy use patterns

		ATES - Feasible savings (PJ/yr)	BTES - Feasible savings (PJ/yr)
Standard eff.	RCP 2.6	16467	8470
	RCP 4.5	16410	8423
	RCP 8.5	16302	8258
Moderate & deep eff.	RCP 2.6	10787	5817
	RCP 4.5	10768	5798
	RCP 8.5	10675	5660

Table E.2: Impact of climate and building scenarios on feasible energy savings from ATES and BTES (defined as energy use within relevant building sectors in suitable climate zones)

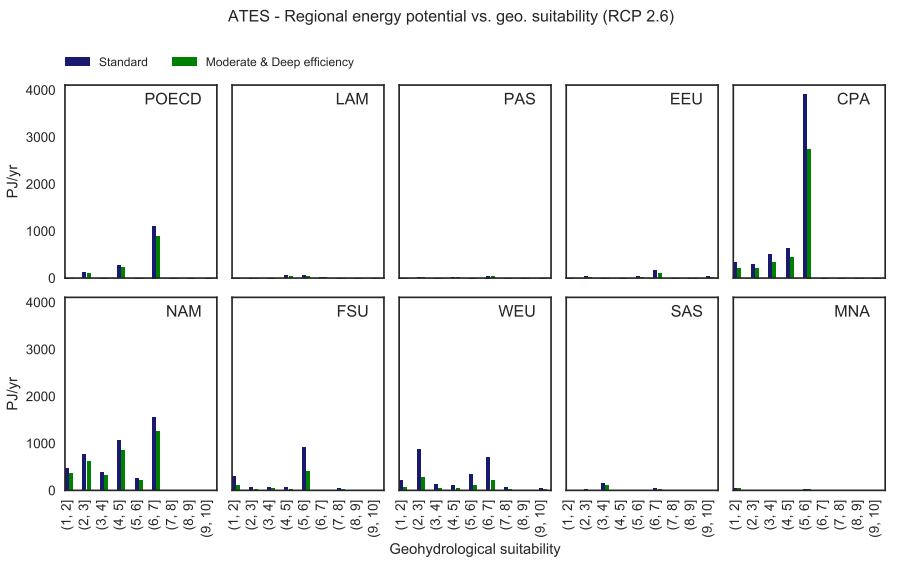


Figure E.5: Distribution of total energy use in suitable climate zones as a function of geohydrological suitability, for each geographic region (RCP 2.6 climate scenario)

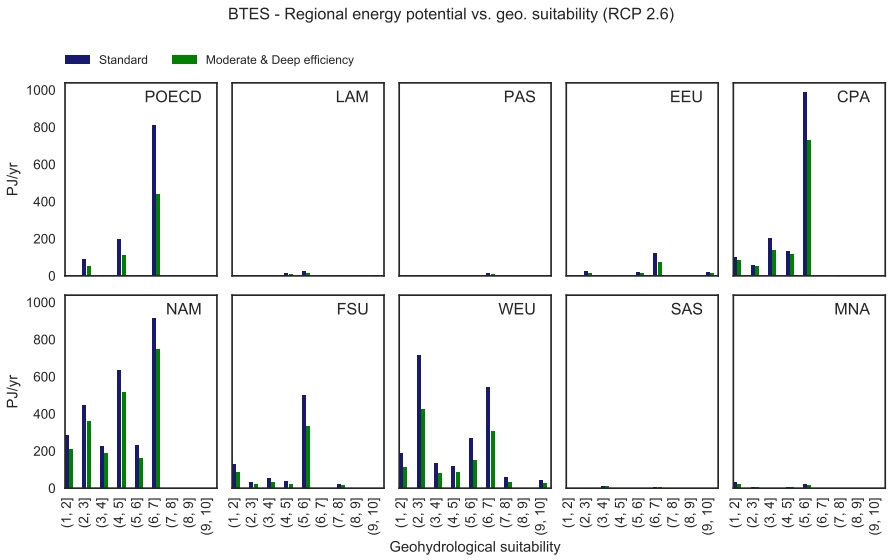


Figure E.6: Distribution of total energy use in suitable climate zones as a function of geohydrological suitability, for each geographic region (RCP 2.6 climate scenario)

Thermal storage adoption scenarios

Recent market data for thermal storage in the Netherlands was presented by Agterberg (2016) and Bloemendal and Hartog (2016). For ATES and BTES combined, this data indicates an adoption rate of 20% of newly constructed utility buildings, and a market share of 3% when considering the entire utility building stock. For ATES specifically, the adoption rate in new utility buildings is approximately 8% for the 2010-2015 period. The residential market shows lower adoption rates, with thermal storage being adopted by 5% of new dwellings in 2016, and a market share of approximately 1% over the entire residential building stock. In addition, Agterberg (2016) estimated that retrofit activities correspond to roughly 25% of the total market for thermal storage.

This data was combined with a highly simplified building stock model to generate future market penetration scenarios for the Netherlands at the 2050 horizon. Although a detailed forecast would be outside of the scope of this work, these results provide a plausible starting point for the adoption scenarios used in the analysis, taking the Netherlands as a "best-case" market for thermal storage due to the country's high geohydrological suitability score. These values are therefore used to determine maximum adoption rates in the *L*, *M* and *H* scenarios (low, medium and high adoption).

The building stock model is based on a simple "coflow" structure (e.g. Sterman (2000)), illustrated in a System Dynamics stock-flow diagram in Figure E.7. The corresponding equations for the building stock $B(t)$ and total thermal storage adoption

$S(t)$ follow below, with the average market share A across the building stock being given by the ratio $A = S(t)/B(t)$.

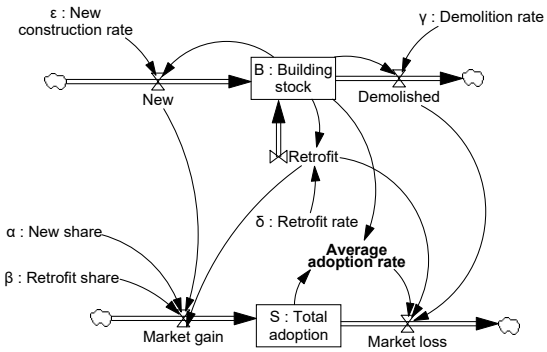


Figure E.7: Stock-flow diagram of idealized building stock model

$$\frac{dB(t)}{dt} = \epsilon B(t) - \gamma B(t) \tag{E.0.3}$$

$$\frac{dS(t)}{dt} = \alpha\epsilon B(t) + \beta\delta B(t) - \frac{S(t)(\gamma B(t) + \delta B(t)}{B(t)} = B(t)(\alpha\epsilon + \beta\delta) - S(t)(\gamma + \delta) \tag{E.0.4}$$

The model is parameterized using approximate values from Agterberg (2016) shown in Table E.3, for three different thermal storage adoption rates (corresponding to the current situation in the Netherlands in the L scenario, as well as higher assumptions for future adoption in the M and H scenarios). The building stock $B(t)$ is expressed as a fraction of the initial building stock; under this parameterization, its value remains constant over time (i.e. 1). These values yield the adoption curves shown in Figure E.8 as a function of time, which yield approximate final market shares A of 10%, 20% and 30% of the building stock in the L, M and H scenarios.

Table E.3: Building stock model parameters

Parameter	Value
α: Thermal storage adoption rate for new construction	{0.2, 0.35, 0.5}
β: Thermal storage adoption rate for retrofits	{0.1, 0.2, 0.3}
γ: Building demolition rate	1/50
δ: Building retrofit rate	1/30
ϵ: Building construction rate	1/50
$B(0)$: Initial building stock	1
$S(0)$: Initial thermal storage market share	0.03

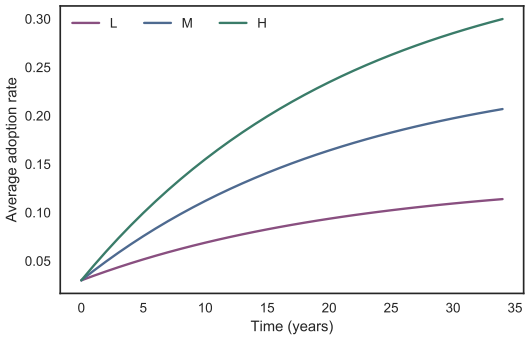


Figure E.8: Baseline adoption scenarios as a function of time

Thermal performance scenarios

The following table presents the COP values used in each scenario for each technology, distinguishing between COP values for cooling, heating, and under cooling or heating imbalance (COP_c, COP_h, COP_i), for ATES, BTES, and a reference conventional system:

Table E.4: Thermal performance scenarios

Scenario		COP_c	COP_h	COP_i with (CDD-HDD) > 0	COP_i with (CDD-HDD) < 0
High COP	ATES	20	3	15	2
	BTES	10	2.5	10	1.5
Low COP	ATES	15	2	10	1
	BTES	8	1.5	8	1
Reference	Conventional system	3	1	-	-

Based on these values for each thermal performance scenario (*perf*), and on the yearly degree-day values for heating and cooling demand in each RCP, the following equations then yield average weighted COP values for each UTES technology and for conventional systems across operational modes, in each urban area:

$$\overline{COP}_{UTES,city,RCP,perf} = \frac{CDD \cdot COP_c + HDD \cdot COP_h + |HDD - CDD| \cdot COP_i}{CDD + HDD + |HDD - CDD|} \quad \text{(E.0.5)}$$

$$\overline{COP}_{conv,city,RCP} = \frac{CDD \cdot COP_c + HDD \cdot COP_h}{CDD + HDD} \quad \text{(E.0.6)}$$

Sensitivity analysis

Figures E.9 (for ATES) and E.10 (for BTES) present the estimated relative importance of each type of scenario input towards the variance of energy savings, using variance decomposition (Sobol, 1993) to compute the relative total sensitivity index of each input. For instance, this indicates that approximately 65% of the variance of total ATES energy savings in the FSU region is explained by adoption scenarios, with approximately 25% being explained by building efficiency scenarios, and less than 5% being related to the climate scenarios. Adoption scenarios therefore have a predominant impact on output uncertainty.

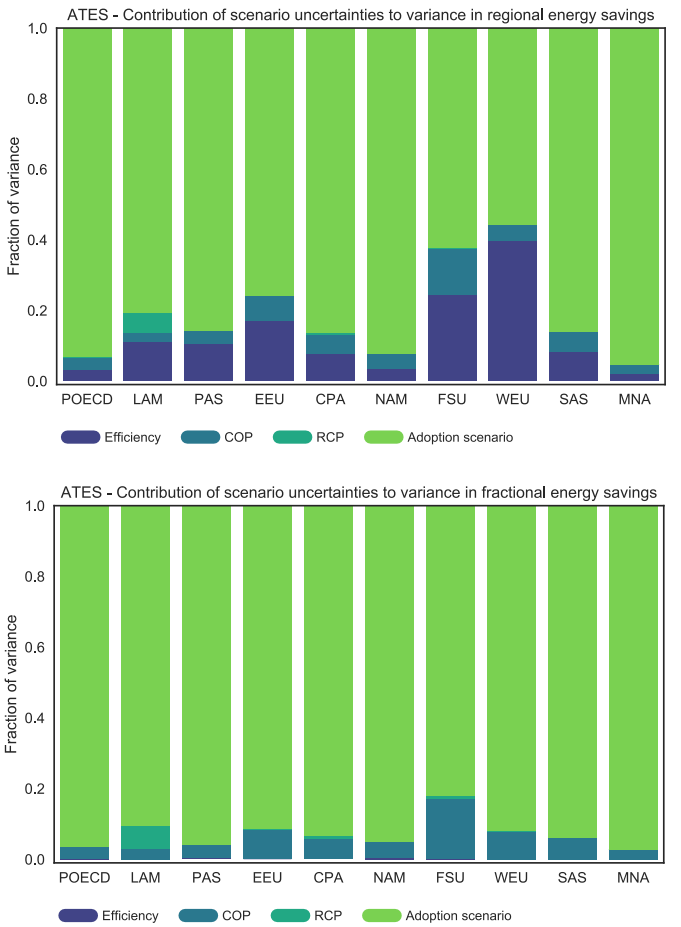


Figure E.9: Relative contribution of scenario uncertainties (building efficiency, ATES COP, RCP climate scenario, and adoption scenario) to the variance of total (top panel) and fractional (bottom panel) ATES energy savings, in each region

It can be noted that thermal performance (COP scenarios) is more significant for BTES than ATES, particularly in regions with a relatively higher heating demand, such as FSU; due to the relatively lower performance of BTES in these conditions, the COP scenarios have a greater impact on realized savings.

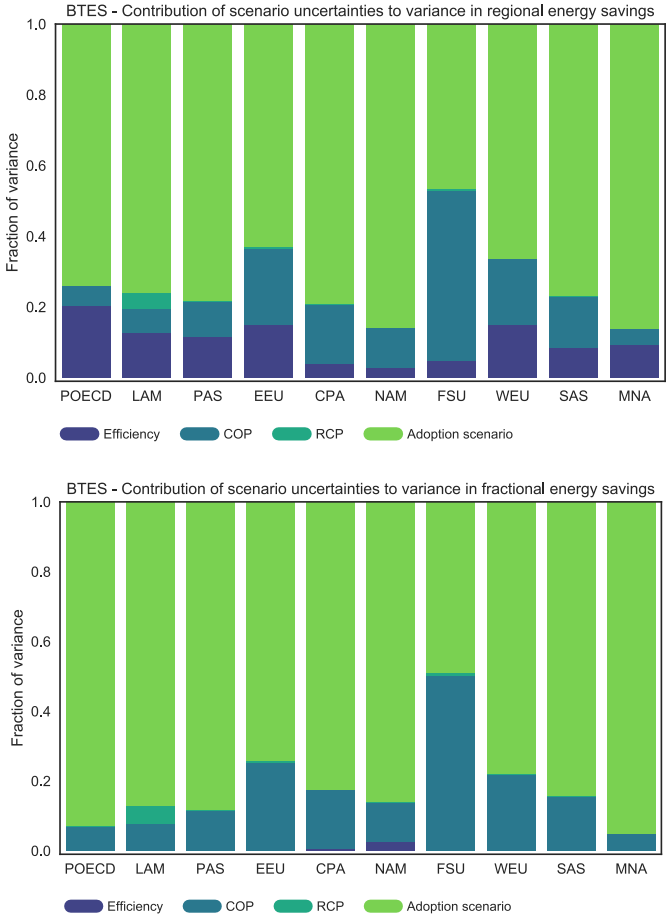


Figure E.10: Relative contribution of scenario uncertainties (building efficiency, BTES COP, RCP climate scenario, and adoption scenario) to the variance of total (top panel) and fractional (bottom panel) BTES energy savings, in each region

Bibliography

Agterberg, F. (2016). Developing strategic options for the Dutch subsoil energy sector. Master's thesis, TIAS School for Business and Society. Available (in Dutch) at https://www.rvo.nl/sites/default/files/2017/04/2016%20Startdocument%20RM%20BE_vertaald_def.pdf.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Almeida, S., Holcombe, E. A., Pianosi, F., and Wagener, T. (2017). Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change. *Natural Hazards and Earth System Sciences*, 17(2):225.

Altmann, A., Tolosi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347.

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229:25–36.

Ananduta, W. W. (2016). Distributed Energy Management in Smart Thermal Grids with Uncertain Demands. Master's thesis, Delft University of Technology.

Antle, J. M., Capalbo, S. M., Elliott, E. T., Hunt, H. W., Mooney, S., and Paustian, K. H. (2014). Research Needs for Understanding and Predicting the Behavior of Managed Ecosystems: Lessons from the Study of Agroecosystems. *Ecosystems*, 4(8):723–735.

Ayllón, D., Railsback, S. F., Vincenzi, S., Groeneveld, J., Almodóvar, A., and Grimm, V. (2016). InSTREAM-Gen: Modelling eco-evolutionary dynamics of trout populations under anthropogenic environmental change. *Ecological Modelling*, 326(Supplement C):36–53.

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., and Fienen, M. N. (2016). Scripting MODFLOW Model Development Using Python and FloPy. *Groundwater*, pages n/a–n/a.

Bakr, M., van Oostrom, N., and Sommer, W. (2013). Efficiency of and interference among multiple Aquifer Thermal Energy Storage systems; A Dutch case study. *Renewable Energy*, 60:53–62.

Bakshy, E. and Wilensky, U. (2007). NetLogo-Mathematica Link. http://ccl.northwestern.edu/netlogo/mathematica.html.

Bal, M. (2015). *Social-Ecological System Framework*. PhD thesis, Erasmus University Rotterdam.

Bankes, S. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, 41(3):435–449.

Bankes, S., Walker, W. E., and Kwakkel, J. H. (2013). Exploratory Modeling and Analysis. In Gass, S. I. and Fu, M. C., editors, *Encyclopedia of Operations Research and Management Science*, pages 532–537. Springer US.

Banks, D. (2011). The application of analytical solutions to the thermal plume from a well doublet ground source heating or cooling scheme. *Quarterly Journal of Engineering Geology and Hydrogeology*, 44(2):191–197.

Baroni, G. and Tarantola, S. (2014). A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling & Software*, 51:26–34.

Bartholomew, E., Kwakkel, J., and Jaxa-Rozen, M. (2018). On the role of scenarios in designing robust strategies: a comparison of MORDM, multi-scenario MORDM and Robust Optimization. In *9th International Congress on Environmental Modelling and Software*, Fort Collins, CO, United States.

Bayer, P., Saner, D., Bolay, S., Rybach, L., and Blum, P. (2012). Greenhouse gas emission savings of ground source heat pump systems in Europe: A review. *Renewable and Sustainable Energy Reviews*, 16(2):1256–1267.

Becu, N., Perez, P., Walker, A., Barreteau, O., and Page, C. L. (2003). Agent based simulation of a small catchment water management in northern Thailand: Description of the CATCHSCAPE model. *Ecological Modelling*, 170(2–3):319–331.

Berger, T. (2001). Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics*, 25(2–3):245–260.

Beyer, H.-G. and Sendhoff, B. (2007). Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218.

BGR/UNESCO (2008). Groundwater Resources of the World. *Hannover, Paris*.

Bithell, M. and Brasington, J. (2009). Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution. *Environmental Modelling & Software*, 24(2):173–190.

Bloemendal, J. and Hartog, N. (2016). After the boom: Evaluation of Dutch ates-systems for energy efficiency. In *European geothermal congress 2016*.

Bloemendal, M. (2018). *The hidden side of cities: Methods for governance, planning and design for optimal use of subsurface space with ATES*. PhD thesis, Delft University of Technology.

Bloemendal, M. and Hartog, N. (2018). Analysis of the impact of storage conditions on the thermal recovery efficiency of low-temperature ATES systems. *Geothermics*, 71:306–319.

Bloemendal, M. and Jaxa-Rozen, M. (2016). Results of ATES users survey on drivers and barriers for investing in ATES. Technical report, Delft University of Technology.

Bloemendal, M., Jaxa-Rozen, M., and Olsthoorn, T. (2018). Methods for planning of ATES systems. *Applied Energy*, 216:534–557.

Bloemendal, M., Jaxa-Rozen, M., and Rostampour, V. (2017). Improved performance of heat pumps helps to use full potential of subsurface space for Aquifer Thermal Energy Storage. In *12th IEA Heat Pump Conference*.

Bloemendal, M., Olsthoorn, T., and van de Ven, F. (2015). Combining climatic and geo-hydrological preconditions as a method to determine world potential for aquifer thermal energy storage. *Science of The Total Environment*, 538:621–633.

Blok, K., de Groot, H., Luiten, E., and Rietbergen, M. (2004). *The Effectiveness of Policy Instruments for Energy-Efficiency Improvement in Firms: The Dutch Experience*. Kluwer Academic Publishers.

Blum, A. L. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1):245–271.

Blum, P., Campillo, G., and Kölbel, T. (2011). Techno-economic and spatial analysis of vertical ground source heat pump systems in Germany. *Energy*, 36(5):3002–3011.

Blum, P., Campillo, G., Munch, W., and Kolbel, T. (2010). CO2 savings of ground source heat pump systems - A regional analysis. *Renewable Energy*, 35(1):122–127.

Bolduc, J.-S. and Vangheluwe, H. (2002). Expressing ODE models as DEVS: Quantization approaches. In *Proceedings of the AIS'2002 Conference (AI, Simulation and Planning in High Autonomy Systems), April 2002, Lisboa, Portugal/F. Barros and N. Giambiasi (eds.)*, pages 163–169.

Bonte, M. (2013). *Impacts of shallow geothermal energy on groundwater quality*. PhD thesis, Vrije Universiteit Amsterdam, Amsterdam.

Bonte, M., Hulsmann, A., Beelen, P. V., and Stuyfzand, P. J. (2011). Underground Thermal Energy Storage: Environmental Risks and Policy Developments in the Netherlands and European Union.

Booij, P., Kamphuis, V., van Pruissen, O., and Warmer, C. (2013). Multi-agent control for integrated heat and electricity management in residential districts. In *4th International Workshop on Agent Technologies for Energy Systems (ATES), a workshop of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Minnesota*.

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., and Hill, M. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black box. *Water Resources Research*, 53(9):7933–7950.

Borren, W. (2009). Ontwikkeling HDSR hydrologisch modelinstrumentarium – HYDROMEDAH. Deelrapport 1: Beschrijving MODFLOW model. Technical report, Deltares, Utrecht.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees.

Brielmann, H., Griebler, C., Schmidt, S. I., Michel, R., and Lueders, T. (2009). Effects of thermal energy discharge on shallow groundwater ecosystems. *FEMS Microbiology Ecology*, 68(3):273–286.

Bryant, B. P. and Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, 77(1):34–49.

Bureau, A., Dupuis, J., Falls, K., Lunetta, K. L., Hayward, B., Keith, T. P., and Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology*, 28(2):171–182.

Butler, M. P., Reed, P. M., Fisher-Vanden, K., Keller, K., and Wagener, T. (2014). Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. *Environmental Modelling & Software*, 59:10–29.

Calje, R. (2010). Future use of Aquifer Thermal Energy Storage below the historic centre of Amsterdam. Master's thesis, Delft University of Technology.

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518.

Camponogara, E., Jia, D., Krogh, B. H., and Talukdar, S. (2002). Distributed model predictive control. *IEEE Control Systems*, 22(1):44–52.

Castilla-Rho, J. C., Mariethoz, G., Rojas, R., Andersen, M. S., and Kelly, B. F. J. (2015). An agent-based platform for simulating complex human–aquifer interactions in managed groundwater systems. *Environmental Modelling & Software*, 73:305–323.

Centraal Bureau voor de Statistiek (2017). Aardwarmte en bodemenergie; onttrekking van warmte en koude. Technical Report (StatLine dataset), Den Haag/Heerlen.

Chan, A. L. S. (2011). Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong. *Energy and Buildings*, 43(10):2860–2868.

China Energy Portal (2018). 2017 electricity and other energy statistics. `https://chinaenergyportal.org/en/2017-electricity-energy-statistics/`.

Climate-KIC - E-USE (2014). Europe-wide Use of Sustainable Energy from Aquifers. Technical report, Climate-KIC.

Davies, S. (1979). *The Diffusion of Process Innovations*. Cambridge University Press.

Day, T. (2006). Degree-days: theory and application. *The Chartered Institution of Building Services Engineers, London*, 106.

de Bruijn, H. and Herder, P. (2009). System and Actor Perspectives on Sociotechnical Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39(5):981–992.

de Graaf, A., Heijer, R., and Postma, S. (2016). Evaluatie Wijzigingsbesluit bodemenergiesystemen. Technical report, Ministry of Infrastructure and Environment, Cothen, The Netherlands.

Deadman, P., Schlager, E., and Gimblett, R. (2000). Simulating Common Pool Resource Management Experiments with Adaptive Agents Employing Alternate Communication Routines. *Journal of Artificial Societies and Social Simulation*, 3(2).

Doughty, C., Hellström, G., Tsang, C. F., and Claesson, J. (1982). A dimensionless parameter approach to the thermal behavior of an aquifer thermal energy storage system. *Water Resources Research*, 18(3):571–587.

Eker, S. and Kwakkel, J. H. (2018). Including robustness considerations in the search phase of Many-Objective Robust Decision Making. *Environmental Modelling & Software*, 105:201–216.

ElSawah, S., Haase, D., Delden, H. V., Pierce, S., Elmahdi, A., Voinov, A. A., and Jakeman, A. J. (2012). Using system dynamics for environmental modelling: Lessons learnt from six case studies. *International Congress on Environmental Modelling and Software*.

Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press.

Eugster, W. and Sanner, B. (2007). Technological status of shallow geothermal energy in Europe. In *European Geothermal Congress 2007*, Unterhaching, Germany.

Eurostat (2018a). Electricity prices for non-household consumers - bi-annual data. Technical Report NRG_PC_205.

Eurostat (2018b). Gas prices for non-household consumers - bi-annual data. Technical Report NRG_PC_203.

Farahani, S. S., Lukszo, Z., Keviczky, T., De Schutter, B., and Murray, R. M. (2016). Robust model predictive control for an uncertain smart thermal grid. In *Control Conference (ECC), 2016 European*, pages 1195–1200.

Farina, M., Giulioni, L., and Scattolini, R. (2016). Stochastic linear Model Predictive Control with chance constraints – A review. *Journal of Process Control*, 44:53–67.

Filatova, T., Polhill, J. G., and van Ewijk, S. (2016). Regime shifts in coupled socio-environmental systems: Review of modelling challenges and approaches. *Environmental Modelling & Software*, 75:333–347.

Filatova, T., Verburg, P. H., Parker, D. C., and Stannard, C. A. (2013). Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental Modelling & Software*, 45:1–7.

Filippi, J.-B. and Bisgambiglia, P. (2004). JDEVS: an implementation of a DEVS based formal framework for environmental modelling. *Environmental Modelling & Software*, 19(3):261–274.

Fleuchaus, P. and Blum, P. (2017). Damage event analysis of vertical ground source heat pump systems in Germany. *Geothermal Energy*, 5(1):10.

Fleuchaus, P., Godschalk, B., Stober, I., and Blum, P. (2018). Worldwide application of aquifer thermal energy storage–A review. *Renewable and Sustainable Energy Reviews*, 94:861–876.

Frey, H. C. and Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, 22(3):553–578.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189–1232.

Friedman, J. H. and Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2):123–143.

Fry, V. A. (2009). Lessons from London: regulation of open-loop ground source heat pumps in central London. *Quarterly Journal of Engineering Geology and Hydrogeology*, 42(3):325–334.

Geroski, P. A. (2000). Models of technology diffusion. *Research Policy*, 29(4–5):603–625.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.

Ghorbani, A. (2013). *Structuring Socio-technical Complexity: Modelling Agent Systems Using Institutional Analysis*. PhD thesis, Delft University of Technology.

Global Buildings Performance Network (2018). Data API. `http://www.gbpn.org/databases-tools/mrv-tool/data-api`.

Gotts, N. M. and Polhill, J. G. (2010). Size matters: large-scale replications of experiments with fearlus. *Advances in Complex Systems*, 13(04):453–467.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., and DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1):115–126.

Grimm, V. and Railsback, S. F. (2012). Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1586):298–310.

Guivarch, C., Rozenberg, J., and Schweizer, V. (2016). The diversity of socio-economic pathways and CO2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, 80:336–353.

Gunnink, J. L. (2004). Deklaagmodel en geohydrologische parametrisatie voor het beheersgebied van het Hoogheemraadschap "De Stichtse Rijnlanden". Technical report, TNO, Utrecht.

Gurobi Optimization, I. (2016). Gurobi Optimizer Reference Manual.

Guyon, I. and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.*, 3:1157–1182.

Hadka, D., Herman, J., Reed, P., and Keller, K. (2015). An open source framework for many-objective robust decision making. *Environmental Modelling & Software*, 74:114–129.

Haehnlein, S., Bayer, P., and Blum, P. (2010). International legal status of the use of shallow geothermal energy. *Renewable and Sustainable Energy Reviews*, 14(9):2611–2625.

Hapfelmeier, A. and Ulm, K. (2013). A new variable selection approach using Random Forests. *Computational Statistics & Data Analysis*, 60:50–69.

Harbaugh, A. W. (2005). *MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process*. U.S. Geological Survey.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(3859):1243–1248.

Hare, M. and Deadman, P. (2004). Further towards a taxonomy of agent-based simulation models in environmental management. *Mathematics and Computers in Simulation*, 64(1):25–40.

Harper, E. B., Stella, J. C., and Fremier, A. K. (2011). Global sensitivity analysis for complex ecological models: a case study of riparian cottonwood population dynamics. *Ecological Applications: A Publication of the Ecological Society of America*, 21(4):1225–1240.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). 15 - Random Forests. In *The Elements of Statistical Learning*, Springer Series in Statistics, pages 587–604. Springer New York.

Head, B. (2017). NetLogo Python extension. `https://github.com/qiemem/PythonExtension`.

Hecht-Méndez, J., Molina-Giraldo, N., Blum, P., and Bayer, P. (2010). Evaluating MT3dms for Heat Transport Simulation of Closed Geothermal Systems. *Ground Water*, 48(5):741–756.

Helton, J. C. and Oberkampf, W. L. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety*, 85(1):1–10.

Herman, J. and Usher, W. (2017). SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software*, 2(9).

Herman, J. D., Reed, P. M., and Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, 49(3):1400–1414.

Hähnlein, S., Bayer, P., Ferguson, G., and Blum, P. (2013). Sustainability and policy for the thermal use of shallow geothermal energy. *Energy Policy*, 59:914–925.

Holland, J. H. (1992). Complex Adaptive Systems. *Daedalus*, 121(1):17–30.

Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17.

Hothorn, T., Hornik, K., Strobl, C., and Zeileis, A. (2017). party. `https://cran.r-project.org/web/packages/party/index.html`.

Howell, S., Rezgui, Y., Hippolyte, J.-L., Jayan, B., and Li, H. (2017). Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources. *Renewable and Sustainable Energy Reviews*, 77:193–214.

Hunter, J. D. (2007). Matplotlib: A 2d Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95.

Huo, T., Ren, H., Zhang, X., Cai, W., Feng, W., Zhou, N., and Wang, X. (2018). China's energy consumption in the building sector: A Statistical Yearbook-Energy Balance Sheet based splitting method. *Journal of Cleaner Production*, 185:665–679.

International Energy Agency (2017). Solar Heat Worldwide. Technical report, AEE INTEC/IEA Solar Heating & Cooling Programme.

International Institute for Applied Systems Analysis (2012). Global Energy Assessment - Toward a Sustainable Future. Technical report, Vienna, Austria.

Ireland, N. and Stoneman, P. (1986). Technological Diffusion, Expectations and Welfare. *Oxford Economic Papers*, 38(2):283–304.

Ishigami, T. and Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. In *First International Symposium on Uncertainty Modeling and Analysis, 1990. Proceedings*, pages 398–403.

Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics*, 1:519–537.

Jager, W., Janssen, M. A., De Vries, H. J. M., De Greef, J., and Vlek, C. A. J. (2000). Behaviour in commons dilemmas: Homo economicus and Homo psychologicus in an ecological-economic model. *Ecological Economics*, 35(3):357–379.

Janssen, M. (2006). Chapter 30 Governing Social-Ecological Systems. In Judd, L. T. a. K. L., editor, *Handbook of Computational Economics*, volume 2, pages 1511–1548. Elsevier.

Janssen, M. and Ostrom, E. (2006). Empirically based, agent-based models. *Ecology and Society*, 11(2):37.

Jaxa-Rozen, M., Bloemendal, M., and Kwakkel, J. (2018). Assessing the worldwide potential of Underground Thermal Energy Storage (UTES) for energy savings in the built environment. Working paper under review at Earth's Future.

Jaxa-Rozen, M., Bloemendal, M., Rostampour, V., and Kwakkel, J. (2016). Assessing the sustainable application of aquifer thermal energy storage. In *European Geothermal Congress*, Strasbourg, France.

Jaxa-Rozen, M. and Kwakkel, J. (2018a). PyNetLogo: Linking NetLogo with Python. *Journal of Artificial Societies and Social Simulation*, 21(2):4.

Jaxa-Rozen, M. and Kwakkel, J. (2018b). Tree-based ensemble methods for sensitivity analysis of environmental models: A performance comparison with Sobol and Morris techniques. *Environmental Modelling & Software*.

Jaxa-Rozen, M., Kwakkel, J., and Bloemendal, M. (2015a). The adoption and diffusion of common-pool resource-dependent technologies: The case of aquifer Thermal Energy Storage systems. In *2015 Portland International Conference on Management of Engineering and Technology (PICMET)*, pages 2390–2408.

Jaxa-Rozen, M., Kwakkel, J., and Bloemendal, M. (2015b). Planning ATES systems under uncertainty. In *6th European Geothermal PhD Day*, page 35.

Jaxa-Rozen, M., Kwakkel, J., and Bloemendal, M. (2017a). A coupled simulation architecture for agent-based/geohydrological modelling with NetLogo and MODFLOW. Working paper under review at Environmental Modelling & Software.

Jaxa-Rozen, M., Kwakkel, J., and Bloemendal, M. (2017b). Trade-offs and endogenous dynamics for the planning of Aquifer Thermal Energy Storage systems. Working paper under review at Journal of Urban Technology.

Jones, E., Oliphant, T., Peterson, P., and others (2001). *SciPy: Open source scientific tools for Python*.

Jones, P. W. (1999). First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. *Monthly Weather Review*, 127(9):2204–2210.

Judd, K. L. (2006). Chapter 17 Computationally Intensive Analyses in Economics. In Judd, L. T. a. K. L., editor, *Handbook of Computational Economics*, volume 2, pages 881–893. Elsevier.

Kabalci, Y. (2016). A survey on smart metering and smart grid communication. *Renewable and Sustainable Energy Reviews*, 57:302–318.

Kalaiselvam, S. and Parameshwaran, R. (2014). *Thermal energy storage technologies for sustainability*. Elsevier.

Karshenas, M. and Stoneman, P. L. (1993). Rank, Stock, Order, and Epidemic Effects in the Diffusion of New Process Technologies: An Empirical Model. *The RAND Journal of Economics*, 24(4):503–528.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., and Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, 42:55–71.

Kelly, R. A., Jakeman, A. J., Barreteau, O., Borsuk, M. E., ElSawah, S., Hamilton, S. H., Henriksen, H. J., Kuikka, S., Maier, H. R., Rizzoli, A. E., van Delden, H., and Voinov, A. A. (2013). Selecting among five common modelling approaches for integrated environmental assessment and management. *Environmental Modelling & Software*, 47:159–181.

Kesicki, F. and Ekins, P. (2012). Marginal abatement cost curves: a call for caution. *Climate Policy*, 12(2):219–236.

Kleijnen, J. P. (2009). Factor Screening in Simulation Experiments: Review of Sequential Bifurcation. In *Advancing the Frontiers of Simulation*, pages 153–167. Springer.

KNMI (2014a). KNMI Klimaatscenario's Transformatie tijdreeksen KNMI'14. `http://climexp.knmi.nl/scenarios_knmi14_form.cgi`.

KNMI (2014b). KNMI'14: Climate Change scenarios for the 21st Century – A Netherlands perspective. Technical Report KNMI WR 2014-01, KNMI, De Bilt, The Netherlands.

Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1):273–324.

Kravari, K. and Bassiliades, N. (2015). A Survey of Agent Platforms. *Journal of Artificial Societies and Social Simulation*, 18(1):11.

Kunneke, R. and Finger, M. (2009). The Governance of Infrastructures as Common Pool Resources. Fourth Workshop on the Workshop (WOW4), Bloomington, IN.

Kwakkel, J. H. (2017). The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, 96:239–250.

Kwakkel, J. H. and Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, 79:311–321.

Lanahan, M. and Tabares-Velasco, P. C. (2017). Seasonal Thermal-Energy Storage: A Critical Review on BTES Systems, Modeling, and System Design for Higher System Efficiency. *Energies*, 10(6):743.

Langevin, C. D., Shoemaker, W. B., and Guo, W. (2003). *MODFLOW-2000, the US Geological Survey Modular Ground-Water Model–Documentation of the SEAWAT-2000 Version with the Variable-Density Flow Process (VDF) and the Integrated MT3DMS Transport Process (IMT)*. US Department of the Interior, US Geological Survey.

Langevin, C. D., Thorne Jr, D. T., Dausman, A. M., Sukop, M. C., and Guo, W. (2008). SEAWAT Version 4: a computer program for simulation of multi-species solute and heat transport. Technical report, U.S. Geological Survey.

Lazar, C., Taminau, J., Meganck, S., Steenhoff, D., Coletta, A., Molter, C., Schaetzen, V. d., Duque, R., Bersini, H., and Nowe, A. (2012). A Survey on Filter Techniques for Feature Selection in Gene Expression Microarray Analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(4):1106–1119.

Lempert, R., Popper, S., and Bankes, S. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. Technical Report MR-1626, RAND Corporation, Santa Monica, CA.

Lempert, R. J., Groves, D. G., Popper, S. W., and Bankes, S. C. (2006). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science*, 52(4):514–528.

Li, J. (2016). Energy performance heterogeneity in China's buildings sector: A data-driven investigation. *Renewable and Sustainable Energy Reviews*, 58:1587–1600.

Li, Q. (2014). Optimal use of the subsurface for ATES systems in busy areas. Master's thesis, Delft University of Technology.

Liu, Q. and Homma, T. (2009). A new computational method of a moment-independent uncertainty importance measure. *Reliability Engineering & System Safety*, 94(7):1205–1211.

Lo Russo, S., Gnavi, L., Roccia, E., Taddia, G., and Verda, V. (2014). Groundwater Heat Pump (GWHP) system modeling and Thermal Affected Zone (TAZ) prediction reliability: Influence of temporal variations in flow discharge and injection temperature. *Geothermics*, 51:103–112.

Lofberg, J. (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289.

Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice*. PhD thesis, Universite de Liege.

Lucon, O., Urge-Vorsatz, D., Ahmed, A., Akbari, H., Bertoldi, P., Cabeza, L., Eyre, N., Gadgil, A., Harvey, L., Jiang, Y., and others (2014). *Chapter 9: Buildings. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.

Lund, J. W. and Boyd, T. L. (2016). Direct utilization of geothermal energy 2015 worldwide review. *Geothermics*, 60:66–93.

Ma, Y., Borrelli, F., Hencey, B., Coffey, B., Bengea, S., and Haves, P. (2012). Model predictive control for the operation of building cooling systems. *IEEE Transactions on control systems technology*, 20(3):796–803.

Ma, Z., Asmussen, A., Jørgensen, B., Ma, Z., Asmussen, A., and Jørgensen, B. N. (2018). Industrial Consumers' Smart Grid Adoption: Influential Factors and Participation Phases. *Energies*, 11(1):182.

Matthews, R., Gilbert, N., Roach, A., Polhill, J., and Gotts, N. (2007). Agent-based land-use models: a review of applications. *Landscape Ecology*, 22(10):1447–1459.

Matthews, R., Polhill, J., Amin, N., and Roach, A. (2005). Integrating agent-based social models and biophysical models. pages 12–15.

McKenna, E., Richardson, I., and Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41:807–814.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Walt, S. v. d. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56.

McKinsey & Company (2009). Pathways to a low-carbon economy: Version 2 of the global greenhouse gas abatement cost curve. Technical report.

McLachlan, G. and Peel, D. (2000). Mixtures of Factor Analyzers. In *Finite Mixture Models*, pages 238–256. John Wiley & Sons, Inc.

McNeil, M. A., Letschert, V., de la Rue du Can, S., and others (2008). Global Potential of Energy Efficiency Standards and Labeling Programs. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Menard, S. and Nell, L. (2014). Jpype. `https://pypi.python.org/pypi/JPype1`.

Milchram, C., van de Kaa, G., Doorn, N., Künneke, R., Milchram, C., van de Kaa, G., Doorn, N., and Künneke, R. (2018). Moral Values as Factors for Social Acceptance of Smart Grid Technologies. *Sustainability*, 10(8):2703.

Mirakhorli, A. and Dong, B. (2016). Occupancy behavior based model predictive control for building indoor climate—A critical review. *Energy and Buildings*, 129:499–513.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., and Schwarz, N. (2013). Describing human decisions in agent-based models – ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, 48:37–48.

Moro, A. and Lonza, L. (2017). Electricity carbon intensity in European Member States: Impacts on GHG emissions of electric vehicles. *Transportation Research Part D: Transport and Environment*.

Morofsky, E. (2007). HISTORY OF THERMAL ENERGY STORAGE. In Paksoy, H. ., editor, *Thermal Energy Storage for Sustainable Energy Consumption*, NATO Science Series, pages 3–22. Springer Netherlands.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174.

Ni, L., Dong, J., Yao, Y., Shen, C., Qv, D., and Zhang, X. (2015). A review of heat pump systems for heating and cooling of buildings in China in the last decade. *Renewable Energy*, 84:30–45.

Nikolic, I. and Kasmire, J. (2013). Theory. In Dam, K. H. v., Nikolic, I., and Lukszo, Z., editors, *Agent-Based Modelling of Socio-Technical Systems*, number 9 in Agent-Based Social Systems, pages 11–71. Springer Netherlands.

Nordhaus, W. (2007). Accompanying Notes and Documentation on Development of DICE-2007 Model: Notes on DICE-2007. delta. v8 as of September 21, 2007. *Yale University, New Haven, NE, USA*.

Nordpil and UN Population Division (2010). World database of large urban areas, 1950-2050. `https://nordpil.com/resources/world-database-of-large-cities/`.

Nossent, J., Elsen, P., and Bauwens, W. (2011). Sobol sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12):1515–1525.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, 325(5939):419–422.

Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, 31(1):1–14.

Patel, N. R., Risbeck, M. J., Rawlings, J. B., Wenzel, M. J., and Turney, R. D. (2016). Distributed economic model predictive control for large-scale building temperature regulation. In *American Control Conference (ACC), 2016*, pages 895–900.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Petri, Y. and Caldeira, K. (2015). Impacts of global warming on residential heating and cooling degree-days in the United States. *Scientific Reports*, 5.

Phernambucq, I. (2015). Contaminant spreading in high density SATES system areas. Master's thesis, Utrecht University.

Pianosi, F., Iwema, J., Rosolem, R., and Wagener, T. (2017). Chapter 7 - A Multimethod Global Sensitivity Analysis Approach to Support the Calibration and Evaluation of Land Surface Models. In Petropoulos, G. P. and Srivastava, P. K., editors, *Sensitivity Analysis in Earth Observation Modelling*, pages 125–144. Elsevier.

Pianosi, F. and Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67:1–11.

Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550.

Pophillat, W., Attard, G., Bayer, P., Hecht-Méndez, J., and Blum, P. (2018). Analytical solutions for predicting thermal plumes of groundwater heat pump systems. *Renewable Energy*.

Pérez, F. and Granger, B. E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, 9(3):21–29.

Pruyt, E. and Hamarat, C. (2010). The influenza H1n1v pandemic: an exploratory system dynamics approach. In *Proceedings of the 28th International Conference of the System Dynamics Society, Seoul, Korea, 25-29 July 2010*.

Qi, Y., Bar-Joseph, Z., and Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins*, 63(3):490–500.

Railsback, S., Ayllón, D., Berger, U., Grimm, V., Lytinen, S., Sheppard, C., and Thiele, J. (2017). Improving Execution Speed of Models Implemented in NetLogo. *Journal of Artificial Societies and Social Simulation*, 20(1):3.

Railsback, S., Lytinen, S., and Jackson, S. (2006). Agent-based Simulation Platforms: Review and Development Recommendations. *SIMULATION*, 82(9):609–623.

Rawlani, A. K. and Sovacool, B. K. (2011). Building responsiveness to climate change through community based adaptation in Bangladesh. *Mitigation and Adaptation Strategies for Global Change*, 16(8):845–863.

Reeves, H. W. and Zellner, M. L. (2010). Linking MODFLOW with an Agent-Based Land-Use Model to Support Decision Making. *Ground Water*, 48(5):649–660.

Richts, A., Struckmeier, W. F., and Zaepke, M. (2011). WHYMAP and the Groundwater Resources Map of the World 1:25,000,000. In Jones, J. A. A., editor, *Sustaining Groundwater Resources: A Critical Element in the Global Water Crisis*, pages 159–173. Springer Netherlands, Dordrecht.

Rizzoli, A. E., Leavesley, G., Ascough, J. C., Argent, R. M., Athanasiadis, I. N., Brilhante, V., Claeys, F. H. A., David, O., Donatelli, M., Gijsbers, P., Havlik, D., Kassahun, A., Krause, P., Quinn, N. W. T., Scholten, H., Sojda, R. S., and Villa, F. (2008). Integrated Modelling Frameworks for Environmental Assessment and Decision Support. In Jakeman, A. J., Voinov, A. A., Rizzoli, A. E., and Chen, S. H., editors, *Developments in Integrated Environmental Assessment*, volume 3 of *Environmental Modelling, Software and Decision Support*, pages 101–118. Elsevier.

Rosen, M. and Dincer, I. (2003). Exergy methods for assessing and comparing thermal storage systems. *International Journal of Energy Research*, 27(4):415–430.

Rosenhead, J., Elton, M., and Gupta, S. K. (1972). Robustness and Optimality as Criteria for Strategic Decisions. *Operational Research Quarterly (1970-1977)*, 23(4):413–431.

Rostampour, V., Bloemendal, J. M., and Keviczky, T. (2017). A model predictive framework of Ground Source Heat Pump coupled with Aquifer Thermal Energy Storage System in heating and cooling equipment of a building. *12th IEA Heat Pump Conference*.

Rostampour, V., Bloemendal, M., Jaxa-Rozen, M., and Keviczky, T. (2016). A Control-Oriented Model For Combined Building Climate Comfort and Aquifer Thermal Energy Storage System. In *European Geothermal Congress*, Strasbourg, France.

Rostampour, V., Ferrari, R., Teixeira, A., and Keviczky, T. (2018a). Privatized Distributed Anomaly Detection for Large-Scale Nonlinear Uncertain Systems. *Submitted to IEEE Transactions on Automatic Control*.

Rostampour, V., Jaxa-Rozen, M., Bloemendal, M., Kwakkel, J., and Keviczky, T. (2018b). Aquifer Thermal Energy Storage (ATES) Smart Grids: Large-Scale Seasonal Energy Storage as A Distributed Energy Management Solution. Working paper under review at Applied Energy.

Rostampour, V. and Keviczky, T. (2016). Probabilistic energy management for building climate comfort in smart thermal grids with seasonal storage systems. *arXiv preprint arXiv:1611.03206*.

Rostampour, V. and Keviczky, T. (2017). Distributed Stochastic Model Predictive Control Synthesis for Large-Scale Uncertain Linear System. *arXiv preprint arXiv:1703.06273*.

Saltelli, A. (2002a). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297.

Saltelli, A. (2002b). Sensitivity Analysis for Importance Assessment. *Risk Analysis*, 22(3):579–590.

Saltelli, A. and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Saltelli, A. and Tarantola, S. (2002). On the Relative Importance of Input Factors in Mathematical Models. *Journal of the American Statistical Association*, 97(459):702–709.

Samad, T. and Kiliccote, S. (2012). Smart grid technologies and applications for the industrial sector. *Computers & Chemical Engineering*, 47:76–84.

Sanner, B., Karytsas, C., Mendrinos, D., and Rybach, L. (2003). Current status of ground source heat pumps and underground thermal energy storage in Europe. *Geothermics*, 32(4):579–588.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413(6856):591.

Schlüter, M., Hinkel, J., Bots, P. W. G., and Arlinghaus, R. (2014). Application of the SES Framework for Model-based Analysis of the Dynamics of Social-Ecological Systems. *Ecology and Society*, 19.

Schlüter, M., Mcallister, R. R. J., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., Milner-Gulland, E., Müller, B., Nicholson, E., Quaas, M., and Stöven, M. (2012). New Horizons for Managing the Environment: A Review of Coupled Social-Ecological Systems Modeling. *Natural Resource Modeling*, 25(1):219–272.

Schlüter, M. and Pahl-Wostl, C. (2007). Mechanisms of Resilience in Common-pool Resource Management Systems: an Agent-based Model of Water Use in a River Basin. *Ecology and Society*, 12(2):4.

Schmolke, A., Thorbek, P., DeAngelis, D. L., and Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. *Trends in Ecology & Evolution*, 25(8):479–486.

Schulzweida, U., Kornblueh, L., and Quast, R. (2011). *Climate data operators (CDO), user guide, Version 1.5. 2.*

Simm, J. and de Abril, I. (2015). ExtraTrees. `http://github.com/jaak-s/extraTrees`.

Singh, R., Wagener, T., Crane, R., Mann, M. E., and Ning, L. (2014). A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA. *Water Resources Research*, 50(4):3409–3427.

Slenders, H. L. A., Dols, P., Verburg, R., and Vries, A. J. d. (2010). Sustainable remediation panel: Sustainable synergies for the subsurface: Combining groundwater energy with remediation. *Remediation Journal*, 20(2):143–153.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280.

Sobol', I. M. and Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79(10):3009–3017.

Sommer, W. (2015). *Modelling and monitoring of Aquifer Thermal Energy Storage : impacts of soil heterogeneity, thermal interference and bioremediation.* PhD thesis, Wageningen University, Wageningen.

Sommer, W., Valstar, J., Leusbrock, I., Grotenhuis, T., and Rijnaarts, H. (2015). Optimization and spatial pattern of large-scale aquifer thermal energy storage. *Applied Energy*, 137(2015):322–337.

Sopha, B., Klöckner, C. A., and Hertwich, E. G. (2011). Exploring policy options for a transition to sustainable heating system diffusion using an agent-based simulation. *Energy Policy*, 39(5):2722–2729.

Sterman, J. (2000). *Business Dynamics*. McGraw-Hill.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8:25.

Strobl, C., Hothorn, T., and Zeileis, A. (2009). Party on! A new, conditional variable importance measure available in the party package. *The R Journal*, (2):14–17.

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., and Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. *Environmental Modelling & Software*, 86:56–67.

Tang, T., Reed, P., Wagener, T., and Van Werkhoven, K. (2006). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences Discussions*, 3(6):3333–3395.

Tavoni, A. and Levin, S. (2014). Managing the climate commons at the nexus of ecology, behaviour and economics. *Nature Climate Change*, 4(12):1057–1063.

Thiele, J. C. (2015). *Towards Rigorous Agent-Based Modelling*. PhD thesis, Georg-August-Universität Göttingen, Göttingen.

Thiele, J. C., Kurth, W., and Grimm, V. (2012a). Agent-Based Modelling: Tools for Linking NetLogo and r. *Journal of Artificial Societies and Social Simulation*, 15(3):8.

Thiele, J. C., Kurth, W., and Grimm, V. (2012b). RNETLOGO: an R package for running and exploring individual-based models implemented in NETLOGO. *Methods in Ecology and Evolution*, 3(3):480–483.

TIOBE (2017). TIOBE Programming Community index. `https://www.tiobe.com/tiobe-index/`.

Tomasetta, C., Van Ree, C., and Griffioen, J. (2015). Life cycle analysis of underground thermal energy storage. In *Engineering Geology for Society and Territory-Volume 5*, pages 1213–1217. Springer.

Touzani, S. and Busby, D. (2014). Screening Method Using the Derivative-based Global Sensitivity Indices with Application to Reservoir Simulator. *Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles*, 69(4):619–632.

United Nations, Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision. Technical report.

Urge-Vorsatz, D., Petrichenko, K., Antal, M., Staniec, M., Labelle, M., Ozden, E., and Labzina, E. (2012). Best Practice Policies for Low Energy and Carbon Buildings. A Scenario Analysis. Technical report, Center for Climate Change and Sustainable Policy (3CSEP), Central European University.

van Pruissen, O., van der Togt, A., and Werkman, E. (2014). Energy Efficiency Comparison of a Centralized and a Multi-agent Market Based Heating System in a Field Test. *Energy Procedia*, 62:170–179.

van Vliet, E. (2013). Flexibility in heat demand at the TU Delft campus smart thermal grid with phase change materials. Master's thesis, Delft University of Technology.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., and others (2011). The representative concentration pathways: an overview. *Climatic change*, 109(1-2):5.

Vangheluwe, H., DE LARA, J., and MOSTERMAN, P. (2002). An introduction to multi-paradigm modelling and simulation. In *Proceedings of the AIS'2002 Conference (AI, Simulation and Planning in High Autonomy Systems), April 2002, Lisboa, Portugal/F. Barros and N. Giambiasi (eds.)*, pages 9–20.

Vanhoudt, D., Desmedt, J., Van Bael, J., Robeyn, N., and Hoes, H. (2011). An aquifer thermal storage system in a Belgian hospital: Long-term experimental evaluation of energy and cost savings. *Energy and Buildings*, 43(12):3657–3665.

Velvis, H. and Buunk, R. J. (2017). District Aquifer Thermal Energy Storage (DATES). In *12th IEA Heat Pump Conference*, Rotterdam.

Verbong, G., Van Selm, A., Knoppers, R., and Raven, R. (2001). *Een Kwestie van Lange Adem. De Geschiedenis van Duurzame Energie in Nederland 1970-2000*. Aeneas, Boxtel.

Vermeulen, P. and Minnema, B. (2015). iMOD user manual. version: 3.01.

Vienken, T., Schelenz, S., Rink, K., and Dietrich, P. (2014). Sustainable Intensive Thermal Use of the Shallow Subsurface—A Critical View on the Status Quo. *Groundwater*, pages n/a–n/a.

Vincenot, C. E., Giannino, F., Rietkerk, M., Moriya, K., and Mazzoleni, S. (2011). Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology. *Ecological Modelling*, 222(1):210–218.

Véliz, C. and Grunewald, P. (2018). Protecting data privacy is key to a smart energy future. *Nature Energy*, page 1.

Voinov, A. and Shugart, H. H. (2013). 'Integronsters', integral and integrated modeling. *Environmental Modelling & Software*, 39(Supplement C):149–158.

Walt, S. v. d., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2):22–30.

Watson, A. A. and Kasprzyk, J. R. (2017). Incorporating deeply uncertain factors into the many objective search process. *Environmental Modelling & Software*, 89:159–171.

Wilensky, U. (1999). NetLogo.

Willemsen, N. (2016). Rapportage bodemenergiesystemen in Nederland. Technical report, RVO / IF Technology, Arnhem, The Netherlands.

Williams, D. N., Balaji, V., Cinquini, L., Denvil, S., Duffy, D., Evans, B., Ferraro, R., Hansen, R., Lautenschlager, M., and Trenham, C. (2015). A Global Repository for Planet-Sized Experiments and Observations. *Bulletin of the American Meteorological Society*, 97(5):803–816.

Wright, M. N., Ziegler, A., and Konig, I. R. (2016). Do little interactions get lost in dark random forests? *BMC Bioinformatics*, 17.

Zeigler, B. (1990). *Object-oriented simulation with hierarchical, modular models: intelligent agents and endomorphic systems*. Academic Press.

Zhang, H., Xu, W. L., and Hiscock, K. M. (2013a). Application of MT3dms and Geographic Information System to Evaluation of Groundwater Contamination in the Sherwood Sandstone Aquifer, UK. *Water, Air, & Soil Pollution*, 224(2):1438.

Zhang, M., Seck, M., and Verbraeck, A. (2013b). A DEVS-based M&S Method for Large-scale Multi-agent Systems. In *Proceedings of the 2013 Summer Computer Simulation Conference*, SCSC '13, pages 3:1–3:8, Vista, CA. Society for Modeling & Simulation International.

Zheng, C. and Wang, P. P. (1999). MT3dms: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide.

Zhou, N. and Lin, J. (2008). The reality and future scenarios of commercial building energy consumption in China. *Energy and Buildings*, 40(12):2121–2127.

Zhu, K., Blum, P., Ferguson, G., Balke, K.-D., and Bayer, P. (2010). The geothermal potential of urban heat islands. *Environmental Research Letters*, 5(4):044002.

Summary

Research objective

The building sector currently accounts for approximately one-third of the global demand for energy, and one-fifth of all energy-related greenhouse gas emissions (GHG). The development and adoption of energy-efficient technologies in this sector is therefore a key element towards efforts for the mitigation of climate change. In particular, heating is the single largest end use of energy in buildings; basic trends towards urbanization, as well as climate change, are also expected to significantly increase the demand of energy for cooling by the middle of the century. Energy technologies which can address both of these aspects are thus particularly promising.

In this context, Aquifer Thermal Energy Storage (ATES) is an increasingly popular shallow geothermal energy technology. This method uses natural aquifer formations to seasonally store energy for heating and cooling, using "warm" and "cold" storage wells combined with a heat pump. This approach can reduce energy demand by more than half in larger buildings. ATES is used in nearly one-tenth of new commercial and utility buildings in the Netherlands, where suitable aquifers – combined with increasing demand for energy-efficient technologies – make the technology especially competitive.

However, this growth has already evidenced some issues with the policy framework for ATES in the Netherlands. Given the sensitivity of groundwater resources and the technical uncertainties which still affect ATES, this framework follows a conservative approach: for instance, spatial planning guidelines for ATES systems aim to avoid any thermal interferences which could develop between the stored thermal volumes of ATES systems which share an aquifer, and which could reduce their performance.

In practice, however, the operation of ATES systems can differ significantly from expectations, and systems are typically used less than planned – so that subsurface volume may be allocated, but not used. This pattern currently leads to a growing scarcity of subsurface space for new ATES systems in certain areas, such as the city center of Utrecht. These issues affect the performance of ATES from the perspective of building owners and operators. In turn, this limits the technology's contribution towards GHG targets. As such, the current policy framework may need to be reviewed in the context of longer-term objectives for the development of the technology; current policy targets for the Netherlands foresee a five-fold increase in adoption by the middle of the next decade, which would put significant pressure on subsurface space. This situation led to the main research objective for this thesis: *to assess the role of improved methods for the planning and operation of urban ATES systems towards a better alignment of private and public interests.*

The planning and operation of ATES systems involves complex, uncertain dynamics which link physical processes in the subsurface, and the use of the technology by building owners. Representing these dynamics first required a suitable methodological toolkit for the modelling and simulation of ATES systems. Agent-based modelling was a suitable starting point to represent ATES adoption and operation. However, this implied the design of a simulation ar-

chitecture which could link the dynamics of ATES use at level of buildings, and the physical processes of heat storage in the subsurface. To this end, Chapter 2 described the pyNetLogo connector, which interfaces the NetLogo agent-based modelling software with a Python environment. Chapter 3 then introduced a Python-based simulation architecture, using pyNetLogo to link NetLogo with the MODFLOW/SEAWAT codes for geohydrological modelling. As a proof of concept, this architecture was used for a simplified case study of ATES, including feedbacks between subsurface conditions and the use of ATES by simulated building owners. In parallel, Chapter 4 explored the use of decision tree-based algorithms for the sensitivity analysis of simulation models, showing that the random forests and Extra-Trees algorithms could replicate some of the key insights of a conventional global sensitivity analysis at a much smaller computational cost.

This toolkit was then applied to increasingly realistic case studies. Chapter 5 first used an idealized model of ATES development; Chapters 6 then simulated the development of ATES in the city center of Utrecht – which has been one of the most active areas of ATES use in the Netherlands, due to suitable aquifer conditions and a dense built environment. Chapter 7 revisited these case studies by exploring cooperative mechanisms for the operation of ATES, which could help dynamically manage thermal interferences between systems.

Results and policy implications

At the onset of this research, there was already some evidence that the current policy framework for ATES in the Netherlands may hinder the further development of systems in urban areas due to a combination of restrictive spatial planning and administrative overhead. While the latter aspect was largely out of scope of this research, the simulation case studies supported this assessment of the current spatial planning guidelines.

In the short term, the scarcity of subsurface space for new systems is becoming a barrier in areas which combine a relatively shallow aquifer with dense energy demand, such as the Utrecht city center. As shown in Chapters 5 and 6, this scarcity could be addressed in two ways: first, nominal well layout guidelines based on the thermal radius R_{th} of storage wells can be relaxed to a minimal distance of 2.5 R_{th}, with a minimal impact on thermal and economic performance under idealized operating conditions. Secondly, as discussed in Chapter 6, effective well layouts may be much sparser than these nominal guidelines, due to geographic constraints and operational uncertainties. This could be addressed through an adaptive permitting approach to periodically revise allocated storage capacities. This would ensure that building owners have less of an incentive to overestimate the required storage capacity in permit applications. An adaptive approach should however leave enough flexibility for operators to meet annual variations in energy demand, and longer-term demand trends due to climate change – which, in the Netherlands, may become significant for ATES energy balance by 2040.

In the longer term, Chapter 7 showed that coordinated operation in the context of ATES "smart grids" could lead to a more efficient trade-off between private and public interests, by supporting the exchange of information between ATES systems to avoid thermal interferences between systems. Combined with denser ATES spatial planning, this approach would allow a denser use of the subsurface, and could ultimately support a more streamlined, self-organized management regime. This would yield a "win-win" for system owners, as well as policymakers: the former could preserve the system performance provided by current policies, and benefit from greater flexibility; the latter could increase the contribution of ATES to objectives for

energy efficiency and GHG reductions, and avoid the administrative complexity of adaptive permits. However, a shift towards self-organization will imply a fundamental revision of ATES policy, and still requires significant research on technical and institutional aspects.

While the Netherlands are currently the most developed market for ATES, these recommendations are at least equally relevant for emergent markets: the adoption of ATES yields a limited window of opportunity to apply improved methods for planning and operation, as the technology is strongly path-dependent due to the physical layout of systems – so that policymakers should try to avoid "lock-in" under suboptimal methods. In particular, as shown in Chapter 8, the spatial density of demand for ATES observed in the Netherlands may eventually be found across other areas which are suitable for ATES. Policymakers should therefore consider this possibility when designing a framework for the local management of the technology. This will enable new markets for ATES to benefit from improved practices for planning and operation from an early stage of development.

Samenvatting

Onderzoeksdoel

De gebouwde omgeving is verantwoordelijk voor ongeveer een-derde van de wereldwijde energievraag, wat zich door vertaald naar een-vijfde van alle uitstoot van broeikasgassen in de wereld. Omdat gebouwen zo'n groot aandeel in de wereldwijde uitstoot hebben, is het ontwikkelen en toepassen van energiezuinige technieken in deze sector van cruciaal belang voor het voorkomen van klimaatverandering. Vooral het verwarmen gebouwen vraagt de meeste energie. Maar ook het koelen van gebouwen draagt voor een belangrijk deel bij aan het energiegebruik van gebouwen, de bijdrage van koeling zal naar verwachting ook toenemen door verdere urbanisatie en klimaatverandering. Kortom: technieken die gebouwen op een duurzame manier kunnen koelen en verwarmen zijn veelbelovend voor het beperken van de uitstoot van broeikasgassen.

In deze context is het seizoensmatig opslaan en terugwinnen van warmte in aquifers een steeds populairdere vorm van bodemenergie. Bij toepassing van deze techniek wordt met behulp van een warmtepomp en warme en koude grondwaterbronnen warmte opgeslagen en weer onttrokken in natuurlijke zandlagen (aquifers). Warmte- Koude opslag (WKO) is goed toepasbaar bij (grotere) gebouwen met een warmte en koelvraag en zorgt voor een reductie van het energiegebruik van meer dan de helft. Omdat de aquifers er erg geschikt zijn wordt WKO bij ongeveer 1/10de van alle nieuwe utiliteitsgebouwen in Nederland toegepast.

Ondanks de beperkte toepassing van WKO zorgt het beleidskader voor toepassing van deze systemen voor problemen. Zo zijn er veel onzekerheden over energievraag van gebouwen en daarmee het ruimtegebruik in de bodem, daarom is het beleid conservatief en erop gericht om onderlinge interactie tussen warme en koude bronnen te voorkomen.

In de praktijk blijkt het gebruik van de bodem echter sterk af te wijken van het beeld in de ontwerpfase, waarbij systemen structureel veel minder ruimte in de bodem gebruiken dan verwacht. Dat resulteert erin dat er ruimte in de bodem is gealloceerd in vergunningen, maar niet wordt gebruikt, terwijl er nog wel gebouwen staan die ook een bodemenergie systeem willen (individueel belang). Deze situatie leidt er dus toe dat er (onterecht) schaarste ontstaat in de ruimte om bodemenergie toe te passen. En daarmee beperkt het ook de bijdrage die het kan leveren aan het verminderen van de uitstoot van broeikasgassen (maatschappelijk belang). Door de energietransitie wordt een vervijfvoudiging van het aantal bodemenergiesystemen verwacht in de komende 10 jaar, wat dit probleem verder vergroot. Daarom is het nodig om het beleidskader rondom het plaatsen en vergunnen van bodemenergiesystemen te verbeteren, zodat de langer termijn doelen t.a.v. de toepassing van bodemenergie op een duurzame manier kan worden uitgevoerd. Deze situatie heeft geleid tot de volgende doelstelling voor dit onderzoek: *evalueer verbeterde methoden voor planning en beheer van bodemenergiesystemen zodat het maatschappelijke belang van bodemenergie en ook het belang voor de individuele gebruikers van bodemenergie beter wordt gediend.*

De planning en het beheer van bodemenergiesystemen wordt gekenmerkt door veel onzekerheden en is een complexe dynamiek die fysieke processen in de ondergrond verbindt met het

gebruik van het gebouw. Het representatief in beeld brengen van deze interacties en afhankelijkheden vereist een geschikt model instrumentarium voor het modelleren en simuleren van de bodemenergiesystemen. Zogenaamde agent-based modellen zijn daarvoor een geschikt middel. Hiervoor is een simulatiearchitectuur nodig die de dynamiek van het gebruik van het bodemenergiesysteem op het niveau van gebouwen en de fysieke processen van warmteopslag in de ondergrond kan verbinden. In hoofdstuk 2 is de pyNetLogo-connector beschreven, deze connector koppelt de op NetLogo agent-based modelleringssoftware aan de Python-omgeving. Vervolgens is in hoofdstuk 3 een op Python gebaseerde simulatiearchitectuur beschreven, waarbij pyNetLogo wordt ingezet om NetLogo te koppelen aan de MODFLOW / SEAWAT-codes voor de temperatuur modellering in de ondergrond. Als onderbouwing van de werking van dit modelinstrumentarium is deze in hoofdstuk 4 gebruikt voor een vereenvoudigde casestudy, waarbij de onderlinge interactie tussen de bronnen en het daaruit volgende rendement voor de gesimuleerde gebouweigenaren representatief in beeld is gebracht.

Deze ontwikkelde toolkit is vervolgens toegepast op stapsgewijs steeds realistischer wordende casestudies. In hoofdstuk 5 is als eerste de generieke ontwikkeling van bodemenergie geanalyseerd onder geïdealiseerde omstandigheden. In hoofdstukken 6 en 7 is vervolgens de ontwikkeling van bodemenergie in de binnenstad van Utrecht gesimuleerd. Er is voor Utrecht gekozen omdat het een van de meest drukke gebieden met het gebruik van de bodemenergie is in Nederland, vanwege een relatief dunne aquifer en de hoge bebouwingsdichtheid.

Resultaten en implicaties voor beleid

Bij de start van dit onderzoek was het al duidelijk dat het huidige beleidskader voor bodemenergie de verdere ontwikkeling van de techniek in stedelijke gebieden in Nederland zou beperken door een combinatie van te restrictieve ruimtelijke planning en grote administratieve lasten. Ondanks dit laatste aspect grotendeels buiten de scope van dit onderzoek valt ondersteunen de resultaten van de casestudies deze beoordeling van het huidige beleidskader.

Op korte termijn is schaarste aan ondergrondse ruimte een barrière voor toepassing nieuwe bodemenergiesystemen in gebieden met een relatief dunne aquifer en een hoge thermische energievraag, zoals in het stadscentrum van Utrecht. In de hoofdstukken 5 en 6 is onderbouwd dat op twee manieren met deze drukte in de ondergrond kan worden omgegaan: ten eerste kunnen de standaard afstanden tussen bronnen van het tegenovergestelde type worden versoepeld tot 2,5 keer de straal van het thermische beïnvloedingsgebied (R_{th}) zonder dat dat significante impact heeft op de thermische en economische prestaties. Deze verkleining van de afstand helpt ook bij het vinden van geschikte bronlocaties omdat de inpassing van deze locaties wordt beperkt door de aanwezigheid van gebouwen wegen en ondergrondse infrastructuur. Ten tweede is een adaptief vergunningsbeleid waarbij de vergunde hoeveelheid periodiek wordt herzien een effectief middel om ervoor te zorgen dat er geen ruimte in de ondergrond wordt geclaimd maar niet gebruikt. Een dergelijke adaptieve aanpak moet de bodemenergie gebruikers wel voldoende flexibiliteit bieden om om te kunnen gaan met a) de jaarlijkse variaties in de energievraag en b) energievraagontwikkelingen op langere termijn als gevolg van bijvoorbeeld verander gebruik van het gebouw of klimaatverandering.

In hoofdstuk 7 is aangetoond dat uitwisseling van informatie tussen de bodemenergiesystemen gecoördineerd beheer mogelijk maakt om zo negatieve onderlinge interactie te voorkomen. Op de lange termijn leidt een dergelijk "smart-grid" van bodemenergiesystemen tot een efficiëntere afweging tussen particuliere en publieke belangen. In combinatie met een dichtere

ruimtelijke planning van bodemenergiesystemen zou deze benadering een intensiever gebruik van de ondergrond mogelijk maken en een zelf organiserend beheersregime kunnen ondersteunen. Dit zou een " win-win " opleveren voor systeemeigenaren, evenals beleidsmakers. De eerstgenoemden kunnen de systeemprestaties behouden die in het huidige beleid worden beschermd en profiteren daarnaast van meer flexibiliteit. Voor de beleidsmakers is het belangrijkste voordeel dat de bijdrage van bodemenergie aan de doelstellingen voor de reductie van broeikasgasemissies worden vergroot en dat de administratieve complexiteit van adaptieve vergunningen worden vermeden. Een verschuiving naar zelforganisatie impliceert echter een fundamentele herziening van het bodemenergiebeleid en vereist voordat het kan worden geïmplementeerd nog vervolg onderzoek naar zowel technische als ook institutionele aspecten.

De grote vraag naar bodemenergie die in Nederland in veel steden wordt ervaren, zal uiteindelijk ook optreden in andere landen met gebieden die geschikt zijn voor bodemenergie. Beleidsmakers in deze landen kunnen voor het ontwerpen van een beleidskader voor bodemenergiesystemen gebruik maken van de resultaten uit dit proefschrift. Hoewel Nederland momenteel de meest ontwikkelde markt voor bodemenergie is, zijn de aanbevelingen uit dit proefschrift ook relevant voor opkomende markten: want de fysieke lay-out van de bronnen van bodemenergiesystemen die in een gebied worden aangelegd heeft grote invloed op de hoe systemen die later komen kunnen worden ingepast. Een "lock-in" van suboptimale toepassing kan worden vermeden als vanaf het eerste systeem rekening wordt gehouden met potentiele grote drukte. Hierdoor kunnen nieuwe markten voor bodemenergie vanaf een vroeg stadium van ontwikkeling profiteren van praktijkervaring en laatste inzichten uit Nederland voor hun planning en exploitatie van bodemenergiesystemen.

Publications

Peer-reviewed journal papers

- **Jaxa-Rozen, M.**, & Kwakkel, J. (2018). Tree-based ensemble methods for sensitivity analysis of environmental models: A performance comparison with Sobol and Morris techniques. Environmental Modelling & Software, 107, 245-266.
- **Jaxa-Rozen, M.**, & Kwakkel, J. (2018). pyNetLogo: Linking NetLogo with Python. Journal of Artificial Societies and Social Simulation, 21(2), 4.
- Bloemendal, M., **Jaxa-Rozen, M.**, Olsthoorn, T.N. (2018). Methods for planning of ATES systems. Applied Energy, 216, 534-557.
- Rostampour, V., **Jaxa-Rozen, M.**, Bloemendal, M., & Keviczky, T. (2016). Building Climate Energy Management in Smart Thermal Grids via Aquifer Thermal Energy Storage Systems. Energy Procedia, 97, 59-66.
- Kwakkel, J., & **Jaxa-Rozen, M.** (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. Environmental Modelling & Software, 79, 311-321.

Submitted papers

- **Jaxa-Rozen, M.**, Bloemendal, M., Kwakkel, J. (2018). Assessing the worldwide potential of Underground Thermal Energy Storage (UTES) for energy savings in the built environment. Working paper under revision at Earth's Future.
- Rostampour, V.*, **Jaxa-Rozen, M.***, Bloemendal, M., Kwakkel, J., Keviczky, T. (2018). Aquifer Thermal Energy Storage (ATES) Smart Grids: Large-Scale Seasonal Energy Storage as a Distributed Energy Management Solution. Working paper under review at Applied Energy. (*: Equal contribution)
- **Jaxa-Rozen, M.**, Kwakkel, J., Bloemendal, M. (2017). A coupled simulation architecture for agent-based / geohydrological modelling with NetLogo and MODFLOW. Working paper under review at Environmental Modelling and Software.
- **Jaxa-Rozen, M.**, Kwakkel, J., Bloemendal, M. (2017). Trade-offs and endogenous dynamics for the planning of Aquifer Thermal Energy Storage (ATES) systems. Working paper under review at Urban Technology.

Peer-reviewed conference papers

- **Jaxa-Rozen, M.**, Rostampour, V., Herrera, H., Bloemendal, M., Kwakkel, J. & Keviczky, T. (2017). Integrated Building Energy Management Using Aquifer Thermal Energy Storage (ATES) in Smart Thermal Grids. Proceedings of BuildSys'17. ACM, New York, NY, USA.
- Bloemendal, M., **Jaxa-Rozen, M.**, Rostampour, V. (2017) ATES Smart Grids research project overview and first results. Proceedings of the 12th IEA Heat Pump Conference, Rotterdam, The Netherlands.
- Rostampour, V., Bloemendal, M., **Jaxa-Rozen, M.**, Keviczky, T. (2017). Combined Building and Aquifer Thermal Energy Storage Model. Proceedings of the 12th IEA Heat Pump Conference, Rotterdam, The Netherlands.
- Bloemendal, M., & **Jaxa-Rozen, M.** (2017). Improved performance of Heat Pumps helps to use full potential of subsurface space for Aquifer Thermal Energy Storage. Proceedings of the 12th IEA Heat Pump Conference, Rotterdam, The Netherlands.
- **Jaxa-Rozen, M.**, Bloemendal, M., Rostampour, V., Kwakkel, J. (2016). Assessing the Sustainable Application of Aquifer Thermal Energy Storage. European Geothermal Congress, Strasbourg, France.
- Rostampour, V., Bloemendal, M., **Jaxa-Rozen, M.**, Keviczky, T. (2016). A Control-Oriented Model for Combined Building Climate Comfort and Aquifer Thermal Energy Storage System. European Geothermal Congress, Strasbourg, France.
- **Jaxa-Rozen, M.**, Kwakkel, J., Bloemendal, M. (2015). The adoption and diffusion of common-pool resource-dependent technologies: the case of Aquifer Thermal Energy Storage systems. Proceedings of the Portland International Conference on Management of Engineering and Technology, Portland, United States.

Curriculum vitae

Marc Jaxa-Rozen was born in La Pocatière, Canada, on Feb. 9th, 1985. Following undergraduate studies in mechanical engineering at École de technologie supérieure, in Montreal, he gained experience in the field of energy efficiency consulting. This led him to develop an interest in a broad range of issues related to sustainable development, which he then pursued through graduate studies in Engineering and Policy Analysis at TU Delft, graduating with a MSc thesis on the financing of solar portable lighting in Cameroon.

Under the supervision of prof.dr.ir. Paulien Herder and dr.ir. Jan Kwakkel, his PhD research focused on long-term uncertainties in smart energy systems, using case studies for the management of Aquifer Thermal Energy Storage (ATES) in The Netherlands. This project combined qualitative stakeholder research, agent-based modelling, and geohydrological simulation models, to test how different ATES planning and operation schemes may affect the adoption of the technology and the sustainability of aquifer resources. Over the course of his PhD research, he had the opportunity to conduct research visits at Pardee RAND Graduate School and Cornell University.