Model-Based Hydrodynamic Leveling

An Impact Study on the European Vertical Reference Frame

Afrasteh, Y.

DOI
10.4233/uuid:e39336b5-6943-47b4-909f-9fc83e215b7c

Publication date
2024

Document Version
Final published version

Citation (APA)
Afrasteh, Y. (2024). *Model-Based Hydrodynamic Leveling: An Impact Study on the European Vertical Reference Frame*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:e39336b5-6943-47b4-909f-9fc83e215b7c

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

MODEL-BASED HYDRODYNAMIC LEVELING

AN IMPACT STUDY ON THE EUROPEAN VERTICAL REFERENCE FRAME

MODEL-BASED HYDRODYNAMIC LEVELING

AN IMPACT STUDY ON THE EUROPEAN VERTICAL REFERENCE FRAME

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 Friday 19 January 2024 at 10:00 hour

by

Yosra AFRASTEH

Master of Science in Surveying Engineering – Geodesy, University of Tehran, Iran,
born in Tehran, Iran.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. Dr. Ing. habil. R. Klees,	Delft University of Technology, promotor
Prof. Dr. Ir. M. Verlaan,	Delft University of Technology, promotor
Dr. Ir. D.C. Slobbe,	Delft University of Technology, copromotor
Independent members:	
Prof. Dr. O. Andersen	Technical University of Denmark
Prof. Dr. G. Vergos	Aristotle University of Thessaloniki
Prof. Dr. Ir. R.F. Hanssen	Delft University of Technology
Dr. Ing. L. Sánchez-Drewes	Technical University of Munich
Prof. Dr. Ir. P. Teunissen	Delft University of Technology, reservelid

Keywords: Model-based hydrodynamic leveling, Hydrodynamic model, Height system realization, European Vertical Reference System, European Vertical Reference Frame, Tide gauge, Empirical noise model

Printed by: RidderPrint (https://www.ridderprint.nl/)

Front & Back: Water level variation

ISBN 978-94-6384-525-0

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

CONTENTS

SUMMARY

Establishing an accurate global unified vertical reference frame (VRF) is a long-standing objective of geodesy. However, that objective has still not been achieved. One particular application where the lack of such a VRF is evident, is the improvement of hydrodynamic models by assimilating total water levels acquired by tide gauges. Indeed, to facilitate a straightforward assimilation requires that both the observed and modeled water levels refer to the same vertical datum. The required accuracy is high; it is expected to be in the order of 1 centimeter. The best alternative VRF for the area of interest, the northwest European continental shelf, is the European Vertical Reference Frame 2019 (EVRF2019). The EVRF2019, however, still lacks complete coverage and the required accuracy. The key reason is that it is solely based on geopotential differences from spirit leveling/gravimetry, which are not available between benchmarks separated by large water bodies. This thesis exploits model-based hydrodynamic leveling to provide these differences. The specific objective is to assess the potential of including these data in realizing of European Vertical Reference System (EVRS).

Using geodetic network analyses, Chapter 2 investigates the potential impact of using model-based hydrodynamic leveling data on the quality (i.e., precision and reliability) of the EVRF2019. In doing so, we used variance information from the latest Unified European Leveling Network (UELN) adjustment. The model-based hydrodynamic leveling data were assumed to be obtained from not-yet existing hydrodynamic models covering either all European seas surrounding the European mainland or parts of it that provide the required mean water level with uniform precision. A heuristic search algorithm was implemented to identify the set of hydrodynamic leveling connections that provide the lowest median of the propagated height standard deviations. In the scenario which considered only connections between tide gauges located in the same sea basin, all having a precision of 3 cm, the median of the propagated height standard deviations improved by 38% compared to the spirit leveling-only solution. Except for the countries around the Black Sea, coastal countries benefit the most with a maximum improvement of 60% for Great Britain. We also found increased redundancy numbers for the observations in the coastal areas and over the entire Great Britain. Allowing for connections between tide gauges among all European seas increased the impact to 42%. Lowering the precision of the hydrodynamic leveling data lowers the impact. The results show, however, that even in case the assumed precision is 5 cm, the overall improvement is still 29%. We also identified which tide gauges have the largest impact. Our results show that these are the ones located in Sweden in which most height markers are located. The impact, however, hardly depends on the geographic location of the tide gauges within a country.

Chapter 3 develops and analyzes an empirical noise model for the model-derived coastal summer mean water levels (SMWLs) used to establish hydrodynamic leveling connec-

tions. The noise model was subsequently used to obtain a more realistic quality impact of combining hydrodynamic leveling and UELN data in realizing the EVRS. We considered three state-of-the-art hydrodynamic models for the northwest European continental shelf; AMM7, DCSMv6-ZUNOv4, and 3D DCSM-FM. Moreover, we assessed the spatio-temporal performance of these three models in representing coastal SMWLs. The empirical noise models were determined from the differences between observation- and model-derived SMWLs at coastal tide gauges. All three noise models show that the model noise is indeed correlated over sea distances as large as hundreds of kilometers. At the same time, they all show a relatively large discontinuity at the origin (i.e., nugget effect); between 12.1 cm^2 (3D DCSM-FM) and 16.3 cm^2 (DCSMv6-ZUNOv4). The variance (i.e., covariance at zero sea distance) for these two models is 15.3 cm^2 and 21.7 cm^2, respectively. Averaging the water levels over three summers, lowered the variance and nugget effect for 3D DCSM-FM to 12.7 cm^2 and 10.0 cm^2, respectively. Our analysis also showed that between 30% and 50% of the variance can be attributed to errors in the vertical referencing of the tide gauges. We lacked the information to assess what proportion of the observed noise covariances should be attributed to these errors. The performance assessments revealed significant variations over both space and time as well as among the three hydrodynamic models. The results suggest that there is still room for model improvement. In the final experiments, we used the noise model of the best overall performing model (i.e., 3D DCSM-FM) to reassess the quality impact of combining hydrodynamic leveling and UELN data in realizing the EVRS. The results suggest that ignoring noise covariances leads to an overestimation of the total quality impact by 7% and 8%, when we average the water levels over one and three summer periods, respectively.

Finally, Chapter 4 presents the first realization of the EVRS which combines geopotential differences from spirit leveling/gravimetry and model-based hydrodynamic leveling. The model-derived coastal SMWLs used in computing the hydrodynamic leveling connections were obtained from the Nemo-Nordic (Baltic Sea) and 3D DCSM-FM (northwest European continental shelf) hydrodynamic models. The impact of model-based hydrodynamic leveling on the European Vertical Reference Frame (EVRF) is significant, especially for France and Great Britain. Compared to a solution which only uses spirit leveling/gravimetry, the differences in these countries reach tens to hundreds of kgal·mm. We also observed an improved agreement with normal heights obtained by differencing GNSS and the European gravimetric quasi-geoid 2015 (EGG2015) heights. In Great Britain, the south-north slope of 48 mm deg^{-1} present in the solution which uses only spirit leveling/gravimetry data reduced to 2.2 mm deg^{-1}. In France, the improvement is confined to the southwest. The choice of the period over which water levels are averaged has an impact on the results as it determines, among others, the set of tide gauges available to establish the hydrodynamic leveling connections. To underscore this, an experiment was conducted in which an averaging period was used that can be considered as the least preferred choice based on three established criteria. Using this choice, the positive impact for France has gone. For Great Britain, the estimated south-north slope became 12.6 mm deg^{-1}. This is larger than the slope obtained using the most preferred averaging period but still substantially lower compared to the slope associated with a solution that uses only spirit leveling/gravimetry.

The work presented in this thesis is expected to contribute to the further exploitation of model-based hydrodynamic leveling in the realization of vertical reference systems (VRSs). Not only the EVRS or another regional VRS, the technique also has potential when it comes to the realization of the international height reference system (IHRS). Unlocking this potential requires a close collaboration between geodesy and oceanography.

SAMENVATTING

De realisatie van een nauwkeurig wereldwijd verticaal referentiestelsel (VRF) is een al lang bestaand doel van de geodesie. Dat doel is echter nog steeds niet bereikt. Een specifieke toepassing waarbij het ontbreken van een dergelijke VRF duidelijk is, is de verbetering van hydrodynamische modellen door het assimileren van totale waterstanden verkregen door getijmeters. Om een rechttoe-rechtaan assimilatie mogelijk te maken, moeten zowel de waargenomen als de gemodelleerde waterstanden verwijzen naar hetzelfde verticale datum. De vereiste nauwkeurigheid is hoog; het zal naar verwachting in de orde van grootte van 1 centimeter zijn. Het beste alternatieve VRF voor het gebied van interesse, het Noordwest-Europese continentale plat, is het European Vertical Reference Frame 2019 (EVRF2019). De EVRF2019 mist echter nog steeds volledige dekking en de vereiste nauwkeurigheid. De belangrijkste reden hiervoor is dat het uitsluitend gebaseerd is op geopotentiaalverschillen bepaald uit de combinatie van waterpassen en gravimetrie. Deze data zijn niet beschikbaar tussen hoogtemerken die gescheiden zijn door grote wateren. Dit proefschrift maakt gebruik van 'modelgebaseerde hydrodynamische waterpassing' om deze verschillen te bepalen. Het specifieke doel is om het potentieel te beoordelen van het opnemen van deze data bij het realiseren van het Europees verticaal referentiesysteem (EVRS).

Met behulp van geodetische netwerkanalyses onderzoekt Hoofdstuk 2 de potentiële impact van het gebruik van modelgebaseerde hydrodynamische waterpasgegevens op de kwaliteit (zowel precisie als betrouwbaarheid) van de EVRF2019. Daarbij hebben we variantie-informatie gebruikt uit de laatste vereffening van het Unified European Leveling Network (UELN). Aangenomen werd dat de modelgebaseerde hydrodynamische waterpasdata afkomstig waren van nog niet bestaande hydrodynamische modellen die ofwel alle Europese zeeën rond het Europese vasteland bestrijken, ofwel delen daarvan. Verder is aangenomen dat deze modellen de benodigde gemiddelde waterstanden leveren met uniforme precisie. Er is een heuristisch zoekalgoritme geïmplementeerd om de set hydrodynamische waterpasverbindingen te identificeren die de laagste mediaan van de voortgeplante standaarddeviaties van de hoogte oplevert. In het scenario waarin we alleen verbindingen toestaan tussen getijdenmeters die zich in hetzelfde zeegebied bevinden, die allemaal een precisie hebben van 3 cm, verbeterde de mediaan van de voortgeplante standaarddeviaties van de hoogtes met 38% ten opzichte van de oplossing verkregen op basis van uitsluitend UELN data. Behalve de landen rond de Zwarte Zee profiteren de kustlanden het meest met een maximale verbetering van 60% voor Groot-Brittannië. We vonden ook hogere redundantiegetallen voor de waarnemingen in de kustgebieden en over heel Groot-Brittannië. Door verbindingen tussen getijdenmeters tussen alle Europese zeeën mogelijk te maken, nam de impact toe tot 42%. Het verlagen van de precisie van de hydrodynamische waterpasgegevens verlaagt de impact. De resultaten laten echter zien dat zelfs als de veronderstelde precisie 5 cm is, de algehele verbetering nog steeds 29% is. We hebben ook vastgesteld welke getijdenmeters qua impact het meest voordeling zijn. Onze resultaten laten zien dat dit de getijdenmeters

in Zweden zijn, waar de de dichtheid van het UELN netwerk het hoogst is. De impact hangt echter nauwelijks af van de geografische locatie van de getijdenmeters binnen een land.

Hoofdstuk 3 ontwikkelt en analyseert een empirisch ruismodel voor de modelgebaseerde zomergemiddelde kustwaterstand die worden gebruikt om hydrodynamische waterpasverbindingen tot stand te brengen. Het ruismodel werd vervolgens gebruikt om een meer realistische kwaliteitsimpact te verkrijgen van het combineren van modelgebaseerde hydrodynamische waterpasdata en UELN data bij het realiseren van het EVRS. We hebben drie geavanceerde hydrodynamische modellen voor het Noordwest-Europese continentale plat bekeken; het AMM7, het DCSMv6-ZUNOv4 en het 3D DCSM-FM model. Bovendien hebben we de ruimtelijk-temporele prestaties van deze drie modellen beoordeeld bij het representeren van de zomergemiddelde kustwaterstand. De empirische ruismodellen zijn bepaald op basis van de verschillen tussen observatie- en modelgebaseerde zomergemiddelde kustwaterstanden bij getijdenmeters aan de kust. Alle drie de ruismodellen laten zien dat de modelruis inderdaad gecorreleerd is over zeeafstanden tot honderden kilometers. Tegelijkertijd vertonen ze allemaal een relatief grote discontinuïteit in de oorsprong (het nugget-effect); tussen $12,1$ cm^2 (3D DCSM-FM) en $16,3$ cm^2 (DCSMv6-ZUNOv4). De variantie (d.w.z. covariantie voor een zeeafstand van nul) voor deze twee modellen is respectievelijk $15,3$ cm^2 en $21,7$ cm^2. Door de waterstanden over drie zomers te middelen, werden de variantie en het nugget-effect voor 3D DCSM-FM verlaagd tot respectievelijk $12,7$ cm^2 en $10,0$ cm^2. Onze analyse toonde ook aan dat tussen de 30% en 50% van de variantie moet worden toegeschreven aan fouten in de verticale referentie van de getijdenmeters. Het ontbrak ons aan de informatie om te beoordelen welk deel van de waargenomen ruiscovarianties aan deze fouten moet worden toegeschreven. De beoordeling van de modelprestaties bracht significante variaties aan het licht in zowel ruimte als tijd, evenals tussen de drie hydrodynamische modellen. De resultaten suggereren dat er nog ruimte is voor modelverbetering. In de laatste experimenten hebben we het ruismodel van het best presterende model (het 3D DCSM-FM) gebruikt om de kwaliteitsimpact van het combineren van modelgebaseerde hydrodynamische waterpasdata en UELN data bij het realiseren van de EVRS opnieuw te beoordelen. De resultaten suggereren dat het niet meenemen van de ruiscorrelaties leidt tot een overschatting van de totale kwaliteitsimpact met 7% en 8%, wanneer we de waterstanden middelen over respectievelijk een en drie zomerperiodes.

Ten slotte presenteert Hoofdstuk 4 de eerste realisatie van het EVRS die geopotentiaalverschillen uit waterpassing/gravimetrie en modelgebaseerde hydrodynamische waterpassing combineert. De modelgebaseerde zomergemiddelde kustwaterstanden die werden gebruikt bij het berekenen van de hydrodynamische waterpasverbindingen, werden verkregen uit de hydrodynamische modellen Nemo-Nordic (Oostzee) en 3D DCSM-FM (Noordwest-Europees continentale plat). De impact van modelgebaseerde hydrodynamische waterpassing op het Europese verticale referentiestelsel (EVRF) is aanzienlijk, vooral voor Frankrijk en Groot-Brittannië. Vergeleken met een oplossing die alleen geopotentiaalverschillen uit waterpassing/gravimetrie gebruikt, lopen de verschillen in deze landen op tussen de tientallen en honderden kgal·mm. We zagen ook een verbeterde overeenkomst met normaalhoogten verkregen uit het verschil van GNSS hoogten en de hoogten van de Europese gravimetrische quasi-geoïde EGG2015. In Groot-Brittannië is de zuid-noordhelling van

48 mm deg^{-1}, aanwezig in de oplossing op basis van uitsluitend UELN data, teruggebracht tot 2,2 mm deg^{-1} . In Frankrijk beperkt de verbetering zich tot het zuidwesten. De keuze van de periode waarover de waterstanden gemiddeld worden, is van invloed op de resultaten, aangezien deze onder andere bepaalt welke set getijdenmeters beschikbaar zijn om de hydrodynamische waterpasverbindingen tot stand te brengen. Om dit te onderstrepen is een experiment uitgevoerd waarbij een middelingsperiode is gehanteerd die op basis van drie vastgestelde criteria als de minst geprefereerde keuze kan worden beschouwd. Met deze keuze is de positieve impact voor Frankrijk verdwenen. Voor Groot-Brittannië werd de geschatte zuid-noordhelling 12,6 mm deg^{-1}. Dit is groter dan de helling die wordt verkregen met de meest geprefereerde middelingsperiode, maar nog steeds aanzienlijk lager dan de helling die wordt verkregen wanneer alleen UELN data gebruikt wordt.

Het werk gepresenteerd in dit proefschrift zal naar verwachting bijdragen aan de verdere exploitatie van modelgebaseerde hydrodynamische waterpassing in de realisatie van verticale referentiesystemen (VRSs). Niet alleen het EVRS of een ander regionale VRS, de techniek heeft ook potentie als het gaat om de realisatie van het internationale hoogtereferentiesysteem (IHRS). Het ontsluiten van dit potentieel vereist wel een nauwe samenwerking tussen geodesie en oceanografie.

1

INTRODUCTION

1.1. BACKGROUND

The Dutch seaports are of vital importance to the national economy. As such, it is of crucial importance to keep them accessible. This is a real challenge because ships become bigger and ship drafts deeper while at the same time we have to maximize the nautical safety and limit prohibitively expensive dredging campaigns. *To face this challenge, a seamless forecasting system of the expected total water depths in the Dutch North Sea is needed.* Such a system must account for both the evolution of the seafloor topography (i.e., the bathymetry) and the expected water level. The Versatile Hydrodynamics project, of which this Ph.D. project is part of, addresses the water level component.

Forecasts of the expected water level are obtained from a hydrodynamic model. Given the large environmental, economic, and/or social consequences in case of accidents, it is crucial that these forecasts are *complete*, *accurate*, and *complemented by a realistic accuracy description* (i.e., one that is a function of space and lead time), and desirable that all available observations are exploited in the framework of data assimilation. Developing a model that provides such forecasts is the main overall goal of the Versatile Hydrodynamics project.

A *complete* model must resolve all 3D physical processes that contribute to the water level variations. In the southern North Sea, the main contributors are the astronomical tide, surge, and baroclinic forcing (i.e., forcing induced by variations in salinity and temperature). In forecasting storm surges, the primary application that drove hydrodynamic model development in the Netherlands, this contribution has so far been ignored. That is, barotropic (2D) models are used. The reason is that, overall, baroclinic processes mainly contribute to the low-frequency water level variations. At the same time, interaction between tides, surges, and the baroclinic water level are known to exist (e.g., Gräwe et al., 2014). Especially at the mouth of the Rotterdam Waterway. This region is very dynamic, with significant tidal currents modified by baroclinic effects due to the highly variable temperature and salinity. Missing the baroclinic forcing implies that also such interactions are ignored. In addition, there is another consequence, namely that the modeled water levels lack an absolute vertical reference (Slobbe et al., 2013). There are workarounds to account for baroclinic water level variations and to obtain the absolute water level, i.e., water levels in a 3D coordinate system (Slobbe et al., 2013). These are, however, not always valid and hard to operationalize. *Accurate* hydrodynamic models typically employ data assimilation.

The most obvious dataset to be assimilated into the model are tide gauge water levels. After adding the baroclinic forcing to the model, a straightforward implementation assimilates the *total* water levels. Doing so requires to adopt one and the same vertical datum for both the model and all water level observations. The latter includes tide gauge records acquired at islands and offshore platforms. The ability to transfer all water levels to a common vertical datum means is guaranteed once we have access to a regional vertical reference frame (VRF) covering the entire model domain. Given the fact that even small errors in the vertical referencing of the tide gauges imply large erroneous water fluxes and tilts in the water levels, and are a potential source for model instabilities, the VRF's required accuracy is expected to be in the order of 1 centimeter.

The designated VRF for the area of interest is the European Vertical Reference Frame 2019 (EVRF2019) (Sacher & Liebsch, 2019). The EVRF2019, however, lacks the required *coverage* and *accuracy*. This is mainly due to the fact that the EVRF2019 is solely based on geopotential differences obtained by spirit leveling/gravimetry. Spirit leveling (i) is prone to systematic errors (Vanicek et al., 1980), (ii) is expensive, and (iii) cannot be used to cross large water bodies. The latter implies that islands (including Ireland) and offshore platforms are not included and that the connection of coastal countries to the spirit leveling network used to realize the European Vertical Reference System (EVRS), i.e. the Unified European Leveling Network (UELN), is intrinsically weak. The latter was the reason why in computing the EVRF2019 Great Britain was excluded (only the datum offset between the EVRF and the British vertical datum was estimated); Great Britain is only connected to the UELN by two levelings through the Channel Tunnel.

GNSS/leveling (e.g., Catalão & Sevilla, 2008; Schwarz et al., 1987), an obvious alternative to express water levels relative to a common vertical reference surface, offers no solution in the short term. Applying this method requires (i) GNSS observations at all tide gauges of interest as well as the leveling data connecting the GNSS station with the tide gauge, and (ii) access to a high-accurate, high-resolution (quasi-)geoid covering the entire area of interest. Regarding the first, thanks to the efforts of among others the International GNSS Service, that initiated the Tide Gauge Benchmark Monitoring Pilot Project (Schöne et al., 2009), and the EuroGOOS Tide Gauge Task Team (EuroGOOS, 2023), many *coastal* tide gauges are equipped with GNSS (see https://www.sonel.org/-GPS-.html). However, this does not apply to the *offshore* tide gauges. On top, for many stations we lack the leveling data that allow to connect the GNSS station with the tide gauge. Concerning the accuracy of the (quasi-)geoid, to meet the 1 cm accuracy level it should be better than 0.87 cm (assuming that the GNSS heights can be obtained with 0.5 cm accuracy). The most accurate quasi-geoid model available covering the entire area of interest, the EGG2015 (Denker, H., 2015), only has an accuracy of ~ 7 cm (note that the accuracy varies significantly in different countries and areas). Computing a more accurate (quasi-)geoid mainly requires collecting gravity data with good coverage and quality. However, this is expensive and time consuming as it requires an joint effort of many countries.

The overall objective of this study is to realize a regional vertical reference system (VRS) meeting the accuracy and coverage requirements outlined above. Access to such a VRF allows to unify the vertical datum of all tide gauge water levels. The key ingredient in its realization is an approach that allows to connect height benchmarks or height datums *separated by large water bodies*. The next section provides a literature review of existing

approaches.

1.2. PREVIOUS WORK

The definition and realization of vertical reference systems (VRSs) is among the core tasks of geodesy. The underpinning theory is well-developed and extensively described in a number of textbooks (e.g., Hofmann-Wellenhof & Moritz, 2006; Torge & Müller, 2012) and will not be repeated here. This overview of previous work focuses on approaches described in literature that enable to connect height benchmarks or datums *separated by large water bodies*. Connecting height datums, also referred to as 'height datum unification' or 'height system unification', aims (Rummel, 2012) to detect, determine, and preferably, eliminate offsets between height systems, so that physical heights everywhere can be related to one and the same level surface. While addressing the same issue, connecting height datums does not go as far as realizing a regional/global VRS. The latter aims to connect local height *networks*. However, the challenge of crossing large bodies of water is similar in both. Three types of approaches are known from the literature, each of which has sub-variants.

1.2.1. THE GEODETIC BOUNDARY VALUE PROBLEM APPROACH

The development of the geodetic boundary value problem (GBVP) approach goes back to the work of Colombo (1980) and in particular Rummel and Teunissen (1988). At that time, it was just realized that national height datums defined by local mean sea level (MSL) do not refer to the same level surface. Early work by Balazs (1973), Fischer (1977), and Sturges (1974) focused on the definition of height datums, their relations to mean sea level (MSL), and their variations. Colombo (1980), Lelgemann (1977), and Mather (1973) defined all involved quantities and proposed solution strategies. Rapp (1983) outlined the need and prospects of a vertical datum with global coverage and discussed the role of space geodetic techniques (satellite positioning and altimetry). The latter has also been treated by Colombo (1985).

Rummel and Teunissen (1988) outlined a solution to the problem of connecting height datums, which relies on the combination of (i) precise geocentric positions of two or more points obtained using space geodetic techniques, (ii) their potential (or height) value in the respective height datum, and (iii) their geoid height difference. The latter was proposed to be obtained as the solution of a linear geodetic boundary value problem (GBVP) that assumes that observable potential (or height) differences and gravity anomalies refer to different height datums with unknown level differences among them. Provided the geocentric positions, leveled heights, and adequate gravity data are available for all datum zones, the unknown datum offsets appearing in the solution inside and outside the Stokes integral can be estimated using least squares. We refer to Sideris (2015), for an extensive and clear explanation of the GBVP approach.

Xu and Rummel (1991) and Xu (1992) investigated the quality (i.e., accuracy, reliability and detectability) of the proposed method using numerical simulations. It was shown that the quality of the vertical datum connection mainly depends on the quality of the geoid. Other factors include the relative accuracy of the leveling and positions of the 'space stations' (i.e., stations of which the geocentric positions are measured using space geodetic techniques), the number of space stations inside each datum zone, and their geographical

distribution. Only if the datum has to be determined in the absolute sense, the absolute accuracy of the space stations plays an important role.

Many have contributed to the further development and investigation of GBVP approach. They include, e.g., Balasubramania (1994), Heck (1989), Heck and Rummel (1990), Rapp, R. H. and Balasubramania, N. (1992), and Xu (1990). Sansò and Venuti (2002) gave an overview of various variants. Developments after 2000 include: (i) the use of other types of GBVPs to connect height datums instead of the scalar-free one (e.g., Ardalan et al., 2010; Ardalan & Safari, 2005; Grombein et al., 2016; Lehmann, 2000; L. Zhang et al., 2009), (ii) strategies to cope with biases in the gravity anomalies used to compute the local (quasi-)geoid models originating from the fact that they refer to local height datums (e.g., Amos & Featherstone, 2009; Gerlach & Rummel, 2013), and (iii) the exploitation of a satellite gravimetry derived geoid model (e.g., Rummel, 2012). Biases in the gravity anomaly data (Denker, 2001; Heck, 1990) significantly affect the estimated datum offsets; the error has about the same magnitude as the datum offsets, i.e. 1–2 m globally (Gerlach & Rummel, 2013). Amos and Featherstone (2009) tackled this issue by an iterative approach in which the gravity anomalies are updated at each iteration by additional gravity reductions corresponding with the datum offsets estimated at the previous step. By simulations, Gerlach and Rummel (2013) showed that the associated error becomes negligible if a satellite-only gravity model including GOCE data is employed up to spherical harmonic degree and order 200.

The GBVP approach has been applied in a number of studies (e.g., Rapp, 1994; Vu et al., 2021; P. Zhang et al., 2020). Amjadiparvar et al. (2016) applied the approach to connect the leveling-based Canadian height datum, CGVD28, with the leveling-based height datum NAVD88 used in the USA and Mexico. The datum offsets (with respect to a global equipotential surface defined by a GOCE-based geoid) were computed for eight smaller regions along the Canadian and US coastal areas to avoid contamination of the results by known systematic errors and distortions in the Canadian and US leveling networks. In the US coastal regions, the mean datum offset could be estimated with a 1 cm standard deviation. For the other regions the standard deviations range between 2.3 and 3.5 cm. At the same time, the authors noted that due to the low number of tide gauge stations in the Pacific regions, a decimeter standard deviation of the CGVD28 and NAVD88 offset seems plausible. Sánchez and Sideris (2017) applied the method to estimate the level differences between the South American height systems and the global level W_0. They demonstrated that the vertical datum parameters can be estimated with accuracy better than ± 5 cm in well-surveyed regions and ± 40 cm in sparsely surveyed regions.

With the advent of satellite gravimetry (in particular the GOCE mission (Drinkwater et al., 2003)), a number of studies have appeared in which the geoid is not explicitly estimated from gravity data. For example, Gruber et al. (2012) applied a procedure to connect local height datums by comparing the region-averaged 'local geoid height' obtained by differencing GNSS-derived ellipsoidal heights and leveled heights to the region-averaged 'global geoid height' obtained from a geoid computed from a GOCE-based global geopotential model. When accounting for the omission error using EGM2008, relative height datum offsets for the US, Australia, and Germany were obtained that agree at the 10 cm level to results from literature. For areas lacking dense and high quality ground observations, the accuracy was much lower. Rülke et al. (2012) applied a comparable approach to estimate

relative offsets between national height datums in Europe. Instead of averaging over a certain region, a plane model was fitted through the differences. They found an agreement of up to 5 cm for most countries with estimated offsets from spirit leveling.

1.2.2. OCEANOGRAPHIC APPROACH

The oceanographic approach connects tide gauges by means of hydrodynamic or ocean leveling (e.g., Cartwright & Crease, 1963; Proudman, 1953; Rummel & Ilk, 1995; Woodworth et al., 2013). It requires knowledge of the difference in mean dynamic topography (MDT) between them. Several variants to determine these differences have been developed. Cartwright and Crease (1963) determined them using a 1D hydrodynamic model, which was fed by local measurements of the atmospheric pressure, wind speed, and electrical potential across a submarine cable. They achieved an accuracy of 1.5 cm over a distance of 70 km. Wübbelmann (1992) achieved a comparable accuracy for a 20 km connection over the Fehmarn Belt. Woodworth et al. (2013) assessed the contribution of ocean leveling to height system unification by comparing the observation-derived MDTs at the coasts to estimates obtained from ocean circulation models and reported consistency at the subdecimeter level. Liibusk et al. (2013) used a satellite radar altimeter-derived MDT model and reported an accuracy of 2 cm over 65 km. Featherstone and Filmer (2012) successfully used the method to explain the north–south tilt in the Australian height datum. They exploited both 'geodetic' and 'oceanographic' MDT models, as well as models derived from a combination of geodetic and oceanographic data. The oceanographic MDT model, obtained by integration of temperature, pressure, and salinity fields on a water depth of 2 km, showed the best performance in explaining the tilt. A similar conclusion was obtained by Filmer and Featherstone (2012) who used five different models, as well as GNSS and two gravimetric quasi-geoid models, at tide gauges/tide gauge benchmarks to re-estimate the offset in the Australian height datum between mainland Australia and Tasmania. Filmer et al. (2018) also compared geodetic and oceanographic approaches to estimate the MDT. Their assessment included 13 numerical ocean models and 6 MDT models computed from observed geodetic and/or ocean data at 32 tide gauges around the Australasian coast. From their results, numerical ocean models appear to be a viable alternative method to geodetic methods. Their results indicate an error budget of ± 5 cm for some numerical ocean model MDTs compared to leveling for some tide gauge pairs.

Slobbe et al. (2018) proposed what they referred to as 'model-based hydrodynamic leveling' to connect islands and offshore tide gauges with the VRS on land. This variant exploits a regional, high-resolution hydrodynamic model that provides total water levels. From the model, differences in mean water level (MWL) are obtained between tide gauges at the mainland and at the islands or offshore platforms. Adding them to the MWL relative to the national VRSs at the mainland's tide gauges realizes a connection of the island and offshore platforms with the VRS on the mainland. Slobbe et al. (2018) applied the method using an extended version of the 2D Dutch operational storm surge model to transfer Amsterdam ordnance datum (NAP) from the Dutch mainland to the Wadden islands and nearby offshore platforms. Based on their results, the percentage of connections for which the absolute differences between the observation- and model-derived MWL differences are below 1 cm is about 34 % (46 out of 135 possible leveling connections). This result is based on a 19-year simulation period and by computing the MWL over the summer months only.

They found for each Wadden island several connections that allow transferring NAP with sub-centimeter accuracy.

1.2.3. CHRONOMETRIC LEVELING APPROACH

The most recent approach, still under development, is chronometric leveling. Chronometric leveling uses atomic clocks (Delva et al., 2019; Mehlstäubler et al., 2018; Müller et al., 2017; Wu & Müller, 2023). According to Einstein's theory of general relativity, the gravity potential difference between two locations can be derived from the comparison of clock frequencies measured at those locations. To achieve an accuracy of 1.0 cm in terms of height differences, this requires a relative frequency accuracy of 1.0×10^{-18}. Wu et al. (2018) assessed the potential of using clock networks for height datum unification (Bjerhammar, 1985) by simulations using the EUVN (European Unified Vertical Network) (Ihde et al., 2000) as a prior. In doing so, four local height systems were simulated from the EUVN by introducing individual offsets and tilts. Three or four clocks for each local region turned out to be sufficient to unify the height systems when considering offsets between the different height systems and tilts along the national leveling lines in both longitudinal and latitudinal directions. They highlighted that the clocks have to be interconnected and should be properly arranged to be able to sense the leveling tilts where necessary. Their results also indicated that even clocks with an accuracy of one order of magnitude lower than desired allow to unify the height systems to some extent.

1.3. MOTIVATION

A 1 cm accurate VRF covering the entire hydrodynamic model domain cannot be realized based on spirit leveling/gravity data only. Indeed, we need to cross large water bodies. Based on the overview of previous work, we conclude that hydrodynamic/ocean leveling is currently the only realizable technique that allows to do so with the required accuracy. The GBVP approach does not provide a solution for the same reasons stated in Sect. 1.1 regarding GNSS/leveling. The chronometric leveling approach seems very promising. However, no clock network is operational yet.

Regarding the variants of the oceanographic approach discussed in Sect. 1.2, the one proposed by Cartwright and Crease (1963) is not cost-effective and efficient as it requires particular infrastructure. Also, the variant which uses a radar altimeter-derived MDT is suboptimal due to a lack of reliable altimetry data in coastal areas (Vignudelli et al., 2011) and the lack of an approach to make the mean sea surface spectrally consistent with the geoid (e.g., Slobbe et al., 2012). The variant relying on an ocean model seems to be the most promising one. This Ph.D. project builds upon the preliminary work of Slobbe et al. (2018), who already stated that a rigorous implementation of their method requires access to a hydrodynamic model that resolves all 3D physical processes that contribute to the MWL rather than a 2D model in which the baroclinic forcing was added as a diagnostic forcing term.

Apart from using a hydrodynamic model that resolves all 3D physical processes that contribute to the MWL, there are a number of open questions regarding the implementation of model-based hydrodynamic leveling that are not addressed by Slobbe et al. (2018). In their study, they only assessed the number of connections between the mainland tide gauges

and the Wadden islands, for which the error is less than one centimeter. That is, they did not attempt to combine the model-based hydrodynamic leveling data with geopotential differences obtained using spirit leveling/gravimetry. Doing so requires (i) a careful selection of the time span in which the MWLs should be computed, (ii) a choice of the set of hydrodynamic leveling connections to be added, and (iii) a proper stochastic model for the model-based hydrodynamic leveling dataset. Regarding the first, the longer the time span, the more tide gauge records there are that are incomplete. As such, the distribution and number of available tide gauges are time-dependent. Moreover, between N tide gauges, only $N-1$ independent leveling connections can be established. Which connections to choose needs to be determined.

1.4. RESEARCH OBJECTIVES

The specific objective of this study is *to assess the potential of including model-based hydrodynamic leveling data in the realization of the EVRS*. In doing so, we will combine the model-based hydrodynamic leveling data with the geopotential differences from spirit leveling/gravimetry included in the UELN dataset. Contrary to what is done in realizing the EVRF2019 (see Sect. 1.1), we will include the British leveling data as this allows the British leveling network to be strongly tied to those on the European mainland. The reason to conduct this study for the EVRS stems from the fact that its domain overlaps with our region of interest.

From this objective, the following research questions are derived.

- ***What is the potential impact of adding model-based hydrodynamic leveling connections to the UELN dataset on the quality of the EVRF and which connections, and hence tide gauges, are most profitable in terms of impact?***

Before conducting a combination of UELN and *real* model-based hydrodynamic leveling data, we will first assess the impact by means of geodetic network analysis. As a result, different aspects can be studied without first acquiring and processing all the necessary data. This is, among others, relevant because we do not have access to state-of-the-art, high-resolution hydrodynamic models for all seas around Europe; the model being developed in the context of the Versatile Hydrodynamics project initially only included the Northeast Atlantic Ocean, including the North Sea and Wadden Sea (later versions included the Baltic Sea). As such, the impact may be greater than we can demonstrate from the hydrodynamic model at our disposal. In addition, the proposed analyses allow conclusions to be drawn without being limited by the possibilities offered by the current generation of hydrodynamic models. Indeed, future models may perform (much) better.

In our analysis, we will use variance information from the latest UELN adjustment. For the model-based hydrodynamic leveling data, we will assume that they are obtained from not-yet-existing hydrodynamic models covering either all European seas surrounding the European mainland or parts of it that provide the required mean water level with uniform precision. The quality aspects that will be studied include both precision and reliability. The tide gauges used in our analyzes are the ones included in the PSMSL database (Permanent Service for Mean Sea Level (PSMSL), n.d.). The second part of the research question provides guidance on where to focus in the development/calibration of the hydrodynamic model(s).

- ***What is the noise model for coastal summer mean water levels obtained by state-of-the-art hydrodynamic models covering the northwest European continental shelf?***

The combination of UELN and model-based hydrodynamic leveling data requires a proper stochastic model for both datasets that describes its noise characteristics. Such a model, referred to as a 'noise model', is not available for the latter dataset. The second part of this study aims to determine it empirically from the differences between observation- and model-derived MWLs. The adjectives 'coastal' and 'summer' for 'mean water levels' arise respectively from the fact that we will use coastal tide gauges (tide gauges must be connected to the UELN) in establishing the model-based hydrodynamic leveling connections, and that we will average the MWLs over the summer months May to September to avoid the storm surge periods. The latter is in line with the results obtained by Slobbe et al. (2018). The northwest European continental shelf comprises the Northeast Atlantic Ocean, North Sea and Wadden Sea. Empirical noise models will be obtained for three hydrodynamic models; the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (AMM7) (Tonani & Ascione, 2021), the DCSMv6-ZUNOv4 (Zijl et al., 2015; Zijl et al., 2013), and the 3D DCSM-FM (Zijl et al., 2020). The analysis of the obtained noise models will include an assessment of the spatiotemporal performance of the hydrodynamic models in representing the summer MWL. Moreover, to gain insight into the differences in model performance in the coastal waters versus the deep and shelf waters, empirical noise models will be developed for the latter two based on TOPEX/Jason satellite radar altimeter data and compared with the one for the coastal waters. This analysis will be conducted for the DCSMv6-ZUNOv4 model only.

- ***What is the impact of adding real model-based hydrodynamic leveling data between tide gauges in the Baltic Sea and the northwest European continental shelf on the EVRF?***

The model-derived SMWLs will be obtained from the Nemo-Nordic hydrodynamic model (Hordoir et al., 2019; Hordoir et al., 2015) and the 3D DCSM-FM (Zijl et al., 2020), which cover the Baltic Sea and the northwest European continental shelf, respectively. The impact will be quantified by comparing the obtained solutions to a solution that is solely based on UELN data. Where applicable, we will also compare the solutions to GNSS/leveling data.

1.5. OUTLINE OF THE THESIS

The content of this thesis is based on two published journal articles (Chapters 2 and 3) and one submitted article (Chapter 4). The Chapters 2 to 4 successively address the three research questions formulated in Sect. 1.4. They present the potential impact of hydrodynamic leveling on the quality of the EVRF (Chapter 2), the empirical noise model for the model-derived coastal summer mean water levels (Chapter 3), and the new realization of the EVRS obtained using model-based hydrodynamic leveling data (Chapter 4). Chapter. 5 provides conclusions and recommendations for future research.

BIBLIOGRAPHY

Amjadiparvar, B., Rangelova, E., & Sideris, M. G. (2016). The GBVP approach for vertical datum unification: recent results in North America. *Journal of Geodesy*, *90*(1), 45–63. https://doi.org/10.1007/s00190-015-0855-8

Amos, M. J., & Featherstone, W. E. (2009). Unification of New Zealand's local vertical datums: iterative gravimetric quasigeoid computations. *Journal of Geodesy*, *83*(1), 57–68. https://doi.org/10.1007/s00190-008-0232-y

Ardalan, A. A., Karimi, R., & Poutanen, M. (2010). A bias-free geodetic boundary value problem approach to height datum unification. *Journal of Geodesy*, *84*(2), 123–134. https://doi.org/10.1007/s00190-009-0348-8

Ardalan, A. A., & Safari, A. (2005). Global height datum unification: a new approach in gravity potential space. *Journal of Geodesy*, *79*(9), 512–523. https://doi.org/10.1007/s00190-005-0001-0

Balasubramania, N. (1994). *Definition and realization of a global vertical datum* (tech. rep.) [OSU report no. 427]. The Ohio State University, Columbus.

Balazs, E. I. (1973). Local Mean Sea Level in Relation to Geodetic Levelling Along the United States Coastlines. *National Geodetic Survey, Rockville, Md*.

Bjerhammar, A. (1985). On a relativistic geodesy. *Bulletin géodésique*, *59*(3), 207–220. https://doi.org/10.1007/BF02520327

Cartwright, D. E., & Crease, J. (1963). A comparison of the geodetic reference levels of England and France by means of the sea surface. *P. Roy. Soc. Lond. A. Mat.*, *273*(1355), 558–580.

Catalão, J., & Sevilla, M. (2008). The use of ICAGM07 geoid model for vertical datum unification on Iberia and Macaronesian islands. *American Geophysical Union, Fall Meeting 2008, abstract*.

Colombo, O. L. (1980). *A World Vertical Network* (tech. rep. No. 296). The Ohio State University, Columbus, Dept. Geodetic Science.

Colombo, O. L. (1985). Levelling with the help of space techniques [NOAA, Rockville, MD.]. *Proceedings of 3rd Int. Symp. on the North America Vertical Datum*.

Delva, P., Denker, H., & Lion, G. (2019). Chronometric geodesy: Methods and applications. In D. Puetzfeld & C. Lämmerzahl (Eds.), *Relativistic geodesy: Foundations and applications* (pp. 25–85). Springer International Publishing. https://doi.org/10.1007/978-3-030-11500-5_2

Denker, H. (2001). On the effect of datum inconsistencies of gravity and position on the European geoid computations [Budapest, Hungary, 2–7 Sep]. *Proceedings IAG scientific assembly*.

Denker, H. (2015). A new European gravimetric (quasi) geoid EGG2015. *26th IUGG General Assembly, June22–July2*.

Drinkwater, M. R., Floberghagen, R., Haagmans, R., Muzi, D., & Popescu, A. (2003). Goce: Esa's first earth explorer core mission. In G. Beutler, M. R. Drinkwater,

R. Rummel, & R. Von Steiger (Eds.), *Earth gravity field from space — from sensors to earth sciences: Proceedings of an issi workshop 11–15 march 2002, bern, switzerland* (pp. 419–432). Springer Netherlands. https://doi.org/10.1007/978-94-017-1333-7_36

EuroGOOS. (2023). EuroGOOS Tide Gauge Task Team [Accessed: June 2, 2023].

Featherstone, W. E., & Filmer, M. S. (2012). The north-south tilt in the Australian height datum is explained by the ocean's mean dynamic topography. *Journal of Geophysical Research: Oceans*, *117*(C8). https://doi.org/10.1029/2012JC007974

Filmer, M. S., & Featherstone, W. E. (2012). A Re-Evaluation of the Offset in the Australian Height Datum Between Mainland Australia and Tasmania. *Marine Geodesy*, *35*(1), 107–119. https://doi.org/10.1080/01490419.2011.634961

Filmer, M. S., Hughes, C. W., Woodworth, P. L., Featherstone, W. E., & Bingham, R. J. (2018). Comparison between geodetic and oceanographic approaches to estimate mean dynamic topography for vertical datum unification: evaluation at Australian tide gauges. *Journal of Geodesy*, *92*(12), 1413–1437. https://doi.org/10.1007/s00190-018-1131-5

Fischer, I. (1977). Mean sea level and the marine geoid – an analysis of concepts. *Marine Geodesy*, *1*(1), 37–59. https://doi.org/10.1080/01490417709387950

Gerlach, C., & Rummel, R. (2013). Global height system unification with goce: A simulation study on the indirect bias term in the gbvp approach. *Journal of Geodesy*, *87*(1), 57–67. https://doi.org/10.1007/s00190-012-0579-y

Gräwe, U., Burchard, H., Müller, M., & Schuttelaars, H. M. (2014). Seasonal variability in M2 and M4 tidal constituents and its implications for the coastal residual sediment transport. *Geophysical Research Letters*, *41*(15), 5563–5570. https://doi.org/10.1002/2014gl060517

Grombein, T., Seitz, K., & Heck, B. (2016). Height system unification based on the fixed gbvp approach. In C. Rizos & P. Willis (Eds.), *Iag 150 years* (pp. 305–311). Springer International Publishing.

Gruber, T. H., Gerlach, C., & Haagmans, R. (2012). Intercontinental height datum connection with GOCE and GPS-levelling data. *Journal of Geodetic Science*, *2*(4), 270–280. https://doi.org/10.2478/v10156-012-0001-y

Heck, B. (1989). A contribution to the scalar free boundary value problem of physical geodesy. *Manuscripta geodaetica*, *14*(2), 87–99.

Heck, B. (1990). An evaluation of some systematic error sources affecting terrestrial gravity anomalies. *J. Geod.*, *64*(1), 88–108. https://doi.org/10.1007/BF02530617

Heck, B., & Rummel, R. (1990). Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In H. Sünkel & T. Baker (Eds.), *Sea surface topography and the geoid* (pp. 116–128). Springer New York.

Hofmann-Wellenhof, B., & Moritz, H. (2006). *Physical geodesy*. Springer Science & Business Media.

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., Liu, Y., Pemberton, P., Schimanke, S., Andersson, H., et al. (2019). Nemo-Nordic 1.0: a NEMO-based ocean model for the Baltic and North seas–research and operational applications. *Geoscientific Model Development*, *12*(1), 363–386. https://doi.org/10.5194/gmd-12-363-2019

Hordoir, R., Axell, L., Löptien, U., Dietze, H., & Kuznetsov, I. (2015). Influence of sea level rise on the dynamics of salt inflows in the Baltic Sea. *Journal of Geophysical Research: Oceans*, *120*(10), 6653–6668. https://doi.org/https://doi.org/10.1002/2014JC010642

Ihde, J., Adam, J., Gurtner, W., Harsson, B. G., Sacher, M., Schlüter, W., & Wöppelmann, G. (2000). The height solution of the European vertical reference network (EUVN). *Veröff. Bayer. Komm. für die Internat. Erdmessung, Astronom. Geod. Arb*, *61*.

Lehmann, R. (2000). Altimetry–gravimetry problems with free vertical datum. *Journal of Geodesy*, *74*(3), 327–334. https://doi.org/10.1007/s001900050290

Lelgemann, D. (1977). On the Definition of the Listing-Geoid taking into Consideration Different Height Systems. *3rd Int. Symp. Geodesy and Physics of the Earth*, 419.

Liibusk, A., Ellmann, A., Kõuts, T., & Jürgenson, H. (2013). Precise Hydrodynamic Leveling by Using Pressure Gauges. *Mar. Geod.*, *36*(2), 138–163. https://doi.org/10.1080/01490419.2013.771594

Mather, R. S. (1973). The Influence of Sea Surface Topography on Geodetic Considerations. *Proceedings of The Earth's Gravitational Field and Secular Variations in Position*, 585–599.

Mehlstäubler, T. E., Grosche, G., Lisdat, C., Schmidt, P. O., & Denker, H. (2018). Atomic clocks for geodesy. *Reports on Progress in Physics*, *81*(6), 064401. https://doi.org/10.1088/1361-6633/aab409

Müller, J., Dirkx, D., Kopeikin, S. M., Lion, G., Panet, I., Petit, G., & Visser, P. N. A. M. (2017). High performance clocks and gravity field determination. *Space Science Reviews*, *214*(1), 5. https://doi.org/10.1007/s11214-017-0431-z

Permanent Service for Mean Sea Level (PSMSL). (n.d.). Tide Gauge Data [Retrieved 1 Jun 2020].

Proudman, J. (1953). *Dynamical oceanography*. Methuen.

Rapp, R. H. (1983). The need and prospects for a world vertical datum. *Proceedings of the International Association of Geodesy Symposia, International Union of Geodesy and Geophysics XVIII General Assembly*, *2*, 432–445.

Rapp, R. H. (1994). Separation between reference surfaces of selected vertical datums. *Bulletin Geodesique*, *69*(1), 26–31. https://doi.org/10.1007/BF00807989

Rapp, R. H. and Balasubramania, N. (1992). *A conceptual formulation of a world height system* (tech. rep.) [OSU report no. 421]. The Ohio State University, Columbus.

Rülke, A., Liebsch, G., Sacher, M., Schäfer, U., Schirmer, U., & Ihde, J. (2012). Unification of European height system realizations. *J. Geod. Sci.*, *2*(4), 343–354. https://doi.org/10.2478/v10156-011-0048-1

Rummel, R. (2012). Height unification using GOCE. *Journal of Geodetic Science*, *2*(4), 355–362. https://doi.org/doi:10.2478/v10156-011-0047-2

Rummel, R., & Ilk, K. H. (1995). Height datum connection-the ocean part. *AVN-Allgemeine Vermessungs-Nachrichten*, *102*(8-9), 321–330.

Rummel, R., & Teunissen, P. (1988). Height datum definition, height datum connection and the role of the geodetic boundary value problem. *Bulletin Géodésique*, *62*(4), 477–498. https://doi.org/10.1007/bf02520239

Sacher, M., & Liebsch, G. (2019). EVRF2019 as new realization of EVRS.

Sánchez, L., & Sideris, M. . (2017). Vertical datum unification for the International Height Reference System (IHRS). *Geophysical Journal International*, *209*(2), 570–586. https://doi.org/10.1093/gji/ggx025

Sansò, F., & Venuti, G. (2002). The height datum/geodetic datum problem. *Geophysical Journal International*, *149*(3), 768–775. https://doi.org/10.1093/gji/149.3.768

Schöne, T., Schön, N., & Thaller, D. (2009). IGS Tide Gauge Benchmark Monitoring Pilot Project (TIGA): scientific benefits. *Journal of Geodesy*, *83*(3), 249–261. https://doi.org/10.1007/s00190-008-0269-y

Schwarz, K. P., Sideris, M. G., & Forsberg, R. (1987). Orthometric heights without leveling. *J. Surv. Eng*, *113*(1), 28–40.

Sideris, M. (2015). Geodetic world height system unification. In W. Freeden, M. Z. Nashed, & T. Sonar (Eds.), *Handbook of geomathematics* (pp. 3067–3085). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-54551-1_83

Slobbe, D. C., Klees, R., Verlaan, M., Zijl, F., Alberts, B., & Farahani, H. H. (2018). Height system connection between island and mainland using a hydrodynamic model: a case study connecting the Dutch Wadden islands to the Amsterdam ordnance datum (NAP). *J. Geod.*, *92*(12), 1439–1456. https://doi.org/10.1007/s00190-018-1133-3

Slobbe, D. C., Simons, F. J., & Klees, R. (2012). The spherical Slepian basis as a means to obtain spectral consistency between mean sea level and the geoid. *Journal of Geodesy*, *86*(8), 609–628.

Slobbe, D. C., Verlaan, M., Klees, R., & Gerritsen, H. (2013). Obtaining instantaneous water levels relative to a geoid with a 2D storm surge model. *Cont. Shelf Res.*, *52*(0), 172–189. https://doi.org/10.1016/j.csr.2012.10.002

Sturges, W. (1974). Sea level slope along continental boundaries. *Journal of Geophysical Research (1896-1977)*, *79*(6), 825–830. https://doi.org/https://doi.org/10.1029/JC079i006p00825

Tonani, M., & Ascione, I. (2021). Product User Manual: Ocean Physical and Biogeochemical reanalysis NWSHELF_MULTIYEAR_PHY_004_009 NWSHELF_MULTIYEAR_BGC_004_011.

Torge, W., & Müller, J. (2012). *Geodesy*. De Gruyter. https://doi.org/doi:10.1515/9783110250008

Vanicek, P., Castle, R. O., & Balazs, E. I. (1980). Geodetic leveling and its applications. *Reviews of Geophysics*, *18*(2), 505–524. https://doi.org/https://doi.org/10.1029/RG018i002p00505

Vignudelli, S., Kostianoy, A. G., Cipollini, P., & Benveniste, J. (2011). *Coastal altimetry*. Springer Science & Business Media.

Vu, D. T., Bruinsma, S., Bonvalot, S., Bui, L. K., & Balmino, G. (2021). Determination of the geopotential value on the permanent GNSS stations in Vietnam based on the Geodetic Boundary Value Problem approach. *Geophysical Journal International*, *226*(2), 1206–1219. https://doi.org/10.1093/gji/ggab166

Woodworth, P. L., Hughes, C. W., Bingham, R. J., & Gruber, T. (2013). Towards worldwide height system unification using ocean information. *J. Geod. Sci.*, *2*(4), 302–318. https://doi.org/10.2478/v10156-012-0004-8

Wu, H., & Müller, J. (2023). Towards an international height reference frame using clock networks. In J. T. Freymueller & L. Sánchez (Eds.), *Beyond 100: The next century in geodesy* (pp. 3–10). Springer International Publishing.

Wu, H., Müller, J., & Lämmerzahl, C. (2018). Clock networks for height system unification: a simulation study. *Geophysical Journal International*, *216*(3), 1594–1607. https://doi.org/10.1093/gji/ggy508

Wübbelmann, H. (1992). Das hydrodynamische Nivellement am Beispiel eines Pegelnetzes am Fehmarn-Belt. *Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover*, (176), 1–147.

Xu, P. (1990). *Monitoring the sea level rise* [Delft Rep (New Ser) 90:1]. Delft University of Technology.

Xu, P. (1992). A quality investigation of global vertical datum connection. *Geophysical Journal International*, *110*(2), 361–370. https://doi.org/10.1111/j.1365-246X.1992.tb00880.x

Xu, P., & Rummel, R. (1991). *A quality investigation of global vertical datum connection* (tech. rep. No. 34). Netherlands Geodetic Commission. Delft.

Zhang, L., Li, F., Chen, W., & Zhang, C. (2009). Height datum unification between Shenzhen and Hong Kong using the solution of the linearized fixed-gravimetric boundary value problem. *Journal of Geodesy*, *83*(5), 411–417. https://doi.org/10.1007/s00190-008-0234-9

Zhang, P., Bao, L., Guo, D., Wu, L., Li, Q., Liu, H., Xue, Z., & Li, Z. (2020). Estimation of Vertical Datum Parameters Using the GBVP Approach Based on the Combined Global Geopotential Models. *Remote Sensing*, *12*(24). https://doi.org/10.3390/rs12244137

Zijl, F., Laan, S., & Groenenboom, J. (2020). Development of a 3D model for the NW European Shelf (3D DCSM-FM).

Zijl, F., Sumihar, J., & Verlaan, M. (2015). Application of data assimilation for improved operational water level forecasting on the northwest european shelf and north sea. *Ocean Dynam.*, *65*(12), 1699–1716. https://doi.org/10.1007/s10236-015-0898-7

Zijl, F., Verlaan, M., & Gerritsen, H. (2013). Improved water-level forecasting for the Northwest European Shelf and North Sea through direct modelling of tide, surge and non-linear interaction. *Ocean Dynam.*, *63*(7), 823–847. https://doi.org/10.1007/s10236-013-0624-2

2

THE POTENTIAL IMPACT OF HYDRODYNAMIC LEVELING ON THE QUALITY OF THE EUROPEAN VERTICAL REFERENCE FRAME

This chapter has been published in Journal of Geodesy Volume 95, Number 8, July 2021, Pages 1 – 18, https://doi.org/10.1007/s00190-021-01543-3

2.1. INTRODUCTION

The assimilation of *total* water levels measured by tide gauges into a hydrodynamic model requires that both the hydrodynamic model and the observed water levels refer to the same vertical datum. Total water level refers to the actual level of the water (with respect to a well-defined reference), which primarily varies due to tides, winds, and baroclinic effects (i.e., variations in water density). Since hydrodynamic model domains typically do not stop at national boundaries, the modelers are suddenly confronted with the need for a unified height datum. To fit their needs, this unified height datum should be (i) accessible at islands and offshore platforms inside the model domain where tide gauges are available and (ii) highly accurate; we expect the standard deviation to be in the order of 1 cm. The reason for the latter is that even small erroneous tilts in the vertical reference surface may induce large water fluxes, which are a potential source of model instabilities. These requirements pose even in well-surveyed areas with a good geodetic infrastructure a tremendous challenge for existing methods to realize a unified height datum.

Our area of interest serves in this respect as an illustrative example. The domain of the hydrodynamic model we are developing (i.e., the 3D DCSM-FM (Zijl et al., 2020)) with the aim to forecast total water levels in the southern North Sea covers the waters between 15°W to 13°E and 43°N to 64°N (see Fig. 2.1). The first-mentioned requirement shows immediately that the designated unified vertical reference frame for Europe, i.e., the European Vertical Reference Frame 2019 (EVRF2019) (Sacher & Liebsch, 2019), does not suffice. Indeed, Great Britain, Ireland, and other islands are not included. The reason for this is that the EVRF2019 is solely based on data of the Unified European Leveling Network (UELN) (Ihde et al., 2002), which are derived from spirit leveling and gravity data. It is well-known that spirit leveling cannot be used to cross large water bodies. Because of the second-mentioned requirement, GNSS/Leveling (e.g., Catalão & Sevilla, 2008; Schwarz et al., 1987) cannot currently be considered as an *operational* alternative to connect the islands and platforms. To meet the 1-cm accuracy level, a (quasi-)geoid model should be available with an accuracy of 8.7 mm (assuming the GNSS heights can be obtained with 5-mm accuracy). The most accurate quasi-geoid model available covering the entire area, the EGG2015 (Denker, H., 2015), only has an accuracy of ~ 7 cm (note that the accuracy varies significantly in different countries and areas). The same argument applies to the use of the geodetic boundary value problem (GBVP) approach (Amjadiparvar et al., 2016; Amos & Featherstone, 2009; Gerlach & Rummel, 2013; Heck & Rummel, 1990; Rummel & Teunissen, 1988; Sánchez & Sideris, 2017). Finally, despite having great potential, chronometric leveling (Mehlstäubler et al., 2018; Müller et al., 2017) is not yet operational.

The only remaining method for height datum connection known in literature is hydrodynamic/ocean leveling (e.g., Cartwright & Crease, 1963; Proudman, 1953; Woodworth et al., 2013). It is applied between tide gauges and requires knowledge of the differences in mean dynamic topography (MDT) between them. There are different approaches to derive MDT differences; we refer to Slobbe et al. (2018) for a concise review. Featherstone and Filmer (2012) successfully used the method to explain the north–south tilt in the Australian height datum. They exploited both 'geodetic' and 'oceanographic' MDT models, as well as models derived from a combination of geodetic and oceanographic data. The oceanographic MDT model, obtained by integration of temperature, pressure, and salinity fields on a water depth of 2 km, showed the best performance in explaining the tilt. A similar conclusion

was obtained by Filmer and Featherstone (2012) who used five different models, as well as GNSS and two gravimetric quasi-geoid models, at tide gauges/tide gauge benchmarks to re-estimate the offset in the Australian height datum between mainland Australia and Tasmania. In a later, more extended study, Filmer et al. (2018) compared 13 physics-based numerical ocean models and 6 MDTs computed from observed geodetic and/or oceanographic data at 32 tide gauges around the Australian coast to assess the suitability of different types of MDT for height datum unification. One of the main conclusions of the study is that numerical ocean models appear a viable alternative for height datum unification. Slobbe et al. (2018) proposed what they referred to as 'model-based hydrodynamic leveling'. Their method exploits a regional, high-resolution hydrodynamic model to derive mean water level (MWL) differences between tide gauges. The use of the term 'MWL' refers to the fact that the averaging period can be chosen freely; Slobbe et al. (2018) obtained the best results when averaging the water levels over the summer months of the 19-year simulation period. They applied the technique to transfer Amsterdam ordnance datum (Normaal Amsterdams Peil, NAP) from the Dutch mainland to the Dutch Wadden islands. Based on a high-resolution 2D hydrodynamic model, extended to account for depth-averaged water density variations, Slobbe et al. (2018) showed that for each Wadden island several connections are available that allow to transfer NAP with (sub-)centimeter accuracy.

In view of the above-formulated requirements for a unified height datum that meets the needs of the hydrodynamic modelers, we believe that for our area of interest hydrodynamic/ocean leveling has great potential. In particular, the implementation exploits a numerical model. The reasons are threefold. First, because the method indeed allows to transfer the height datum to all tide gauges inside the model domain. Second, because it is potentially accurate. Third, because a rigorous implementation (i.e., one that exploits a hydrodynamic model that resolves *all* relevant 3D physical processes) of the method is realizable in the short term. The second reason is suggested by the results obtained by Filmer et al. (2018), Slobbe et al. (2018), and Woodworth et al. (2013). Despite the fact that the numerical models used in these studies lacked spatial/temporal resolution and/or did not account for all relevant 3D physical processes [(Woodworth et al., 2013, Section 7.2) and (Slobbe et al., 2018, Section 5)], their performance was good in comparison with the results obtained with alternative methods. Moreover, the use of numerical models provides the freedom to choose the averaging period. This allows to avoid, for example, the storm periods. Regarding the last reason, indeed, many models have been developed (see https://eurogoos.eu/models/ for an overview) although for different applications. Many of these are incomplete in terms of physics and/or lack of resolution. As such, they are not suitable for our purpose. At the same time, however, we can highlight that all key building blocks to design a model that resolves *all* relevant 3D physical processes are available. This applies not only to our area of interest, but in fact to almost all European waters. These building blocks include parallel software packages that can handle unstructured meshes needed to run large, high-resolution models (e.g., Deltares, 2021), high-resolution meteorological forcing reanalysis datasets (e.g., Hersbach et al., 2020), river discharge data (e.g., Donnelly et al., 2015; Wilkinson et al., 2014), and a high-resolution bathymetry (e.g., EMODnet bathymetry consortium, 2018).

Indeed, to connect all tide gauges within the domain of our hydrodynamic model we could follow the approach by Slobbe et al. (2018). That is, we connect all tide gauges

to the NAP. This approach, however, requires a model that has a good performance at *all* tide gauge locations even though some are at the same mainland and can be connected by spirit leveling. Apart from that, all errors in the vertical reference of the involved Dutch tide gauge(s) propagate one to one to the vertical reference of the tide gauges of interest. Because the former involves great efforts to achieve it and the latter is not desirable, we propose combining 'hydrodynamic leveling data' with the UELN data and use the combined dataset to compute a new realization of the EVRS that covers our whole domain of interest. This proposal is, indeed, a bit similar to what is advocated by Filmer et al. (2014). To maximize the network strength, we advocate to establish hydrodynamic leveling connections in all European waters. This, of course, requires a model covering all European waters or a set of models that each cover a separate basin. With hydrodynamic leveling connections, we mean connections between tide gauge benchmarks that can be established by using the observation-derived MWLs relative with respect to the tide gauge benchmarks and the model-derived MWL differences between tide gauges. Note that here the model-derived MWL differences are obtained from models that do *not* assimilate geodetic information. The pursued strategy also benefits other users of the EVRF as we may expect that combining both datasets improves the quality of the leveling network and hence the derived VRF. In particular, adding hydrodynamic leveling data helps to detect/suppress systematic errors that spirit leveling is susceptible to (e.g., Penna et al., 2013).

The aim of this paper is twofold. First, to assess the impact of adding model-based hydrodynamic leveling connections to the UELN dataset on the quality of the EVRF. Second, to assess which connections, and hence tide gauges, are most profitable in terms of impact when realizing the EVRS. The second objective is motivated by the fact that in Europe there are many tide gauges. Not all of them can be used to establish hydrodynamic leveling connections. Indeed, a prerequisite is that the benchmarks of the tide gauges located on the European mainland are connected to the UELN. However, even if this requirement is met, the location might be unsuitable if the local water levels are not resolved by the hydrodynamic model. The question, however, can also be turned around: which hydrodynamic leveling connections do have the largest impact on the quality of the VRF? The answer to this question provides guidance where to focus in the development/calibration of the hydrodynamic model(s). To achieve our objectives, we conducted several geodetic network analyses using different scenarios. For the UELN data, we relied on variance information from the latest UELN adjustment. The required MWL differences between tide gauges are assumed to have a uniform precision and are assumed to be obtained from not-yet existing hydrodynamic models (see above) covering all European Seas surrounding the European mainland or parts of it. Indeed, the full potential of hydrodynamic leveling is exploited when we have one large model that allows to establish long-distance connections. In terms of model development, a more plausible scenario is to start with models covering separate sea basins (e.g., the Mediterranean Sea). This implies that we can only establish connections between tide gauges within the same sea basin.

The paper is organized as follows. Section. 2.2 describes the height network adjustment, the way the impact of adding hydrodynamic leveling data is assessed, and the method used to determine which hydrodynamic leveling connections are actually added. Section. 2.3 introduces the datasets used throughout this paper. Section. 2.4 introduces the setup of the experiments conducted in this study. The results of the experiments are presented and dis-

cussed in Sect. 2.5. Finally, we conclude by emphasizing the main findings and identifying topics for future research.

2.2. METHODOLOGY

2.2.1. HEIGHT NETWORK ADJUSTMENT

The height network adjustment is conducted using weighted least squares. For the two observation groups (i.e., the spirit and hydrodynamic leveling data), the Gauss–Markov model takes the form

$$y = Ax + e, \tag{2.1}$$

where

$$y = \begin{pmatrix} y_{\mathrm{sl}} \\ y_{\mathrm{hl}} \end{pmatrix}, A = \begin{pmatrix} A_{\mathrm{sl}} \\ A_{\mathrm{hl}} \end{pmatrix}, \text{ and } e = \begin{pmatrix} e_{\mathrm{sl}} \\ e_{\mathrm{hl}} \end{pmatrix}. \tag{2.2}$$

y is the observation vector, A is the design matrix, x is the vector of unknown parameters, e is the vector of residuals, and subscripts sl and hl stand for spirit leveling and hydrodynamic leveling, respectively. The stochastic properties of the residuals are described by

$$E\{e\} = 0,\ D\{e\} = Q_y = \begin{pmatrix} Q_{\mathrm{sl}} & \mathbf{0} \\ \mathbf{0} & Q_{\mathrm{hl}} \end{pmatrix}, \tag{2.3}$$

where $E\{.\}$ denotes the statistical expectation operator, $D\{.\}$ is the dispersion operator, Q_y is the combined variance-covariance matrix of the two observations groups, Q_{sl} is the variance-covariance matrix of the spirit leveling dataset, and Q_{hl} is the *full* variance-covariance matrix of the hydrodynamic leveling dataset. Q_{hl} is obtained by error propagation, assuming a uniform precision for the difference between the observation- and model-derived MWLs at a tide gauge location (where the observation-derived MWL is expressed relative to the tide gauge bench mark). That is,

$$Q_{\mathrm{hl}} = A_{\mathrm{hl}} Q_{\mathrm{dMWL}} A_{\mathrm{hl}}^{\mathrm{T}}, \tag{2.4}$$

where Q_{dMWL} is the diagonal variance-covariance matrix of the differences between the observation- and model-derived MWLs at the tide gauge locations and A_{hl} is the design matrix of the hydrodynamic leveling dataset. We assume that the contribution of the observation-derived MWLs expressed with respect to the tide gauge benchmarks to the error budget of the hydrodynamic leveling data is negligible. According to us, this is justified for the following reason. Today's instantaneous water levels are measured with a standard deviation of a few centimeters, which implies that the standard deviation of the mean is already at the sub-mm level for one month of data. Moreover, the connection between the tide gauge zero and the tide gauge benchmark can easily be determined using first-order leveling with sub-mm accuracy too. Regarding the contribution of the model-derived MWLs, we currently lack a proper stochastic model. It is expected that some degree of spatial correlation exists and that the accuracy will vary somewhat from location to location. The determination of a proper stochastic model will be the subject of a future study. Here we will use the most simple model possible, namely the model which assumes uniform and uncorrelated noise.

Note, anyway, that contrary to spirit leveling, the uncertainty of hydrodynamic leveling data is likely independent of the distance between the two involved tide gauges. In fact, the noise level mainly depends on the ability of the hydrodynamic model to represent the local MWL.

In realizing the EVRF2019, the datum defect is solved by adding the minimal constraint that for 12 datum points (see Fig. 2.1) the sum of the height changes is zero. The drawback of using this constraint is that the propagated standard deviations of the adjusted heights depend on the height marker (also referred to as "height benchmark" or "leveling benchmark") distance to the datum points (Sacher & Liebsch, 2019). To assess the impact of adding hydrodynamic leveling data on the quality of the leveling network, we conducted an experiment in which we used the constraint that the sum of height changes of all height markers is zero (Teunissen, 2006). This form of minimal constraint adjustment, known as inner constraint adjustment, provides similar results as using the pseudo-inverse in the least-squares adjustment (Ogundare, 2018). Indeed, to realize the EVRS the use of the inner constraint adjustment is not a proper alternative to solve the datum defect as also benchmarks in geodynamically unstable regions will affect the datum.

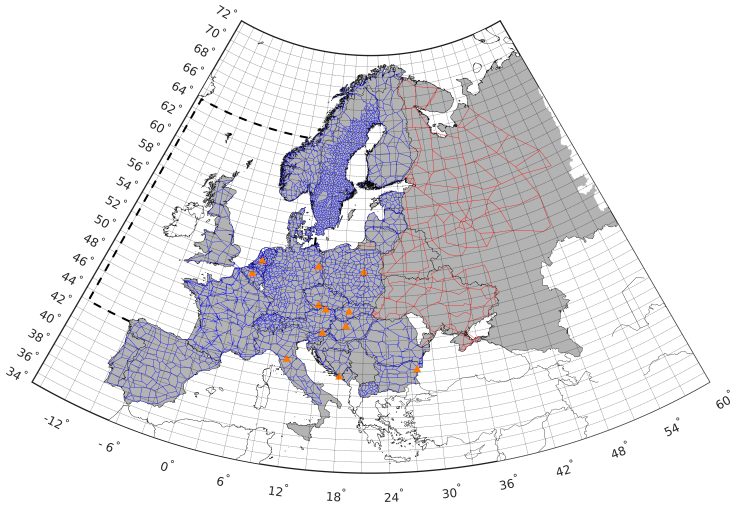


Figure 2.1: Domain of the EVRF2019 (gray areas), UELN leveling connections (blue lines indicate those received from the BKG while the red lines are the ones reconstructed from Sacher and Liebsch (2019, Figure 1), see Section 2.3.1 for more details), and locations of the 12 datum points (green triangle). The black dashed line indicates the domain of 3D DCSM-FM (see Experiment V).

2.2.2. ASSESSING THE IMPACT OF ADDING HYDRODYNAMIC LEVELING DATA

In general, the quality of a geodetic network can be characterized by i) precision, ii) reliability, and iii) cost (Amiri-Simkooei et al., 2012). In this paper, we focus on the first two criteria. The precision of a geodetic network is described by the variance-covariance matrix

of the estimated parameters $Q_{\hat{x}}$, with

$$Q_{\hat{x}} = \left(A^{\mathrm{T}} Q_y^{-1} A\right)^{-1}. \tag{2.5}$$

In reporting the precision, we focus on the propagated standard deviations (SDs) of the adjusted heights (i.e., the square root of the diagonal elements of $Q_{\hat{x}}$); the median of this value for all height markers of the network, as well as the median value per country. The median value for all height markers of the network is also used to determine which hydrodynamic leveling connections will be added to the network (see Sect. 2.2.3). Note that we use the median because the propagated height SDs for all height markers are not normally distributed.

The reliability of a geodetic network refers to its ability to detect and resist against outliers in the observations (Seemkooei, 2001a). We study the impact on the reliability by analyzing the redundancy numbers (see (Seemkooei, 2001b)). The redundancy numbers are the diagonal elements of the so-called redundancy matrix R defined as

$$R = I - A\left(A^{\mathrm{T}} Q_y^{-1} A\right)^{-1} A^{\mathrm{T}} Q_y^{-1}, \tag{2.6}$$

where I is the identity matrix. Redundancy numbers express how the redundancy is distributed over the observations. As such, they depend on the configuration of the network and how well the height markers are connected to each other. For uncorrelated measurements, their value ranges between 0 and 1. The smaller/larger its value, the larger/smaller the magnitude of the outlier that can be detected as well as its influence on the estimated parameters. It is desirable to have a network with relatively large and uniform redundancy numbers, so that the ability to detect outliers is the same in every part of the network (Baarda, 1968). Similar to the way we analyze the impact on precision, we will also report changes in the median value of the redundancy numbers for the entire network and per country.

2.2.3. CHOICE OF THE HYDRODYNAMIC LEVELING CONNECTIONS

Given N tide gauges, maximum $N-1$ *independent* hydrodynamic leveling connections can be established. By independent, we mean the connections that do not form any closed circuit. Indeed, the model-derived MWL differences between the tide gauges are obtained from the MWLs at the tide gauges. As such, adding a connection that closes a circuit does not add any new information and results the full variance-covariance matrix obtained using Eq. 2.4 to become singular. The number of possible connections can be extremely large. Considering N tide gauges, the number of possibilities to establish $N-1$ independent connections among them equals $K = N^{N-2}$ (Cayley's formula (Aigner & Ziegler, 1998)). Europe has a relatively dense network of tide gauges that contains hundreds of stations. Assuming 200 out of them can be used to establish hydrodynamic leveling connections, K equals 200^{198}. To evaluate which set of connections has the largest impact, i.e. results in the lowest median SD of the adjusted heights, K least-squares solutions have to be computed. Despite the fact that the computational load can be reduced significantly by exploiting the recursive least-squares method (Teunissen, 2006) and only computing the diagonal elements of the variance-covariance matrix of the estimated unknowns, still evaluating all K solutions is not feasible.

Therefore, we use a heuristic search method that identifies the connections one by one. In each step, we first identify all remaining possible connections (closed circuits are not

allowed) based on a depth-first search algorithm (Tarjan, 1972). Second, we identify which of these connections results in the lowest median SD of the adjusted heights. The identified connection is added to the list of found connections and removed from the list of remaining ones. The search process continues until no more connections are possible. The use of this heuristic search method indeed reduces the computational load significantly. To identify the first connection, $\binom{N}{2}$ least-squares solutions have to be computed. With every connection we add, this number decreases with the number of connections added in the previous step (always one) and the ones that form a closed circuit.

To further reduce the computational load, we (i) reduced the number of potential tide gauges (see Sect. 2.3.2) and (ii) did not allow connections among tide gauges (a) located within the same country and (b) located in neighboring countries for which the number of spirit leveling connections between the countries is larger than one.

2.3. DATA

2.3.1. SPIRIT LEVELING NETWORK AND DATA

From the Federal Agency for Cartography and Geodesy (BKG), we received (i) the locations of all UELN height markers, (ii) a list of leveling connections (only contains an overview of which height markers are connected; we did not receive the actual geopotential differences) in all countries except for Ukraine, Russia, and Belarus, iii) the a priori variances of the geopotential differences for the available connections, and iv) the variances obtained by variance component estimation (except for Great Britain, Ukraine, Russia, and Belarus). The reason why we did not receive the information for all countries is that either the BKG is not allowed to share the data (applies to the data of Ukraine, Russia, and Belarus), or the data have not been used in computing the EVRF2019 and as such are not available at all (applies to the variances obtained by variance component estimation for the data of Great Britain). Missing the data in Ukraine, Russia and Belarus makes the connection of central Europe to the Fennoscandia region to be based on just two leveling observations. This would artificially increase the impact of adding hydrodynamic leveling data. Therefore, we decided to reconstruct the missing leveling connections using Figure 1 in Sacher and Liebsch (2019). Figure. 2.1 shows both the part of the spirit leveling network obtained from the BKG and the reconstructed part. The data variances for the reconstructed connections are determined using the computed distances between the height markers and the reported standard deviations of unit weight per country (Sacher & Liebsch, 2019, Table 3). In all experiments conducted in this study, we used the variances that the BKG obtained by variance component estimation. For Great Britain, the a-priori variances were used.

To ease the interpretation of the results, all adjustments are conducted in terms of geometric quantities. That means that variances expressed in $kgal \cdot mm$ have been converted to meters, using the GRS80 (Moritz, 2000) normal gravity value.

2.3.2. CANDIDATE TIDE GAUGES AND LINK TO SPIRIT LEVELING NETWORK

Candidate tide gauges, i.e., tide gauges among which hydrodynamic leveling connections can be established, have to be located at the coast of one of the seas surrounding the UELN countries. Moreover, we only use tide gauges south of the TOPEX/Poseidon and Jason

maximum latitude of 66°N. The waters in higher latitudes have lower densities of high-quality satellite and in situ data for validation of hydrodynamic models. On top of that, these regions typically have a poor bathymetry (Stammer et al., 2014). Both will negatively affect the ability of hydrodynamic models to represent the MWL.

Tide gauges are selected from the ones included in the PSMSL database (Holgate et al., 2013; Permanent Service for Mean Sea Level (PSMSL), n.d.). In the area of interest, this database includes about 330 tide gauges. Indeed, more tide gauges are available (see, e.g., http://www.emodnet-physics.eu/Map/.). The PSMSL database is used, however, because the data are quality controlled and provided with extensive metadata. The metadata include, among others, descriptions of the tide gauge benchmarks and their locations. The latter information is indispensable when implementing hydrodynamic leveling.

To reduce the computational load (see Sect. 2.2.3), we only consider those tide gauges that are located within 10 km from the nearest UELN height marker. This results in a total number of 186 tide gauges. Figure. 2.2 shows the locations of the tide gauges. The number of tide gauges per country is presented in Table 2.1. To connect the tide gauges to the UELN, we added an artificial leveling connection between each tide gauge and the nearest UELN height marker. The variances for these added artificial leveling connections are determined assuming the leveling is conducted with a precision of 0.5 mm$/\sqrt{\text{km}}$ corresponding to the precision of first-order leveling (Bossler, 1984).

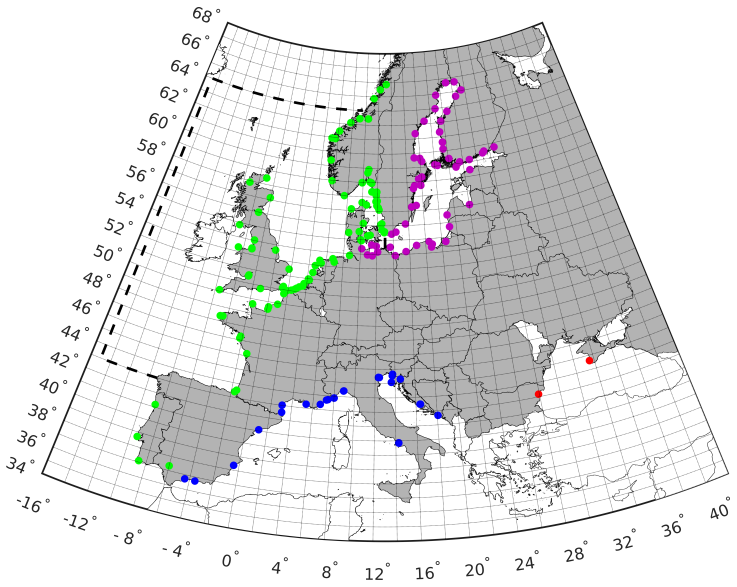


Figure 2.2: Location of all tide gauges used in this study. The different colors refer to the four sea basins used in Experiments I–III: the Mediterranean Sea (blue), Black Sea (red), Baltic Sea (magenta), and the North-East Atlantic region including the North Sea (green). The black dashed line indicates the domain of 3D DCSM-FM (see Experiment V).

2.4. EXPERIMENTAL SETUP

In this section, we describe and motivate the five experiments conducted in this study.

Experiment 0: using spirit leveling data only—The impact of adding hydrodynamic leveling is assessed by comparing the obtained realizations to the one obtained with spirit leveling data only. Since our spirit leveling dataset is not identical to the one used to obtain the EVRF2019 for reasons explained in Sect. 2.3.1, in Experiment 0, we assess the performance of what we refer to as the *spirit leveling-only solution*.

Experiment I: allowing for connections within basins only—Applying model-based hydrodynamic leveling between tide gauges requires the availability of a hydrodynamic model being capable to resolve the local MWL. As pointed out in Sect 2.1, in terms of model development a more plausible scenario for a European-wide implementation of hydrodynamic leveling is to start with a set of models each covering a separate sea basin. Assuming to have access to such a set of models that allow to derive the MWL differences with uniform precision, the objective of this experiment is to assess the impact of adding hydrodynamic leveling connections between tide gauges located in the same sea basin to the UELN. Four basins are considered: the Mediterranean Sea, Black Sea, Baltic Sea, and the North-East Atlantic region including the North Sea. Figure. 2.2 shows per basin the locations of all 186 PSMSL tide gauges that meet the criteria outlined in Sect. 2.3.2. We assume a uniform noise level of 3 cm for each connection. This means a variance of 4.5 cm^2 for the precision at which the model is able to reconstruct the MWL at each tide gauge location. Again, so far we lack a proper stochastic model for the hydrodynamic leveling dataset. The 3-cm accuracy level is a bit lower than the accuracy obtained by Slobbe et al. (2018) for the connection of the Dutch Wadden island tide gauges to the NAP. Woodworth et al. (2013) stated that ocean leveling is possible with a typical uncertainty of better than a decimeter. They also point out, however, that this statement "is subject to reservations concerning the limitations in the ocean models available for analysis, and to the fact that a global study remains to be made". Given the fact that we pursue an implementation based on models that resolve all relevant 3D physical processes plus the fact that we use MWL differences rather than MDT differences (i.e., we ignore the storm surge period in computing the MWL, see Sect 2.1), we believe 3 cm is challenging but not unrealistic.

Experiment II: varying the noise level—In Experiment I, we assumed a uniform variance of 4.5 cm^2 for the model-derived MWL at the tide gauge locations (corresponding to a precision of 3 cm for each hydrodynamic leveling connection). To assess how the assumed noise level impacts the results, in Experiment II we varied the noise level of the hydrodynamic leveling data from 1 to 5 cm in steps of 1 cm.

Experiment III: using the inner constraint adjustment—So far, we assessed the impact of hydrodynamic leveling connections on the quality of the EVRF. As explained in Sect. 2.2.1, the propagated SDs depend on the distance of the height markers to the locations of the datum points. As discussed, we can avoid this dependency by considering the so-called inner constraint adjustment. This is what we assess in this experiment. So, Experiment III differs from Experiment I in the use of the constraint added to solve the datum defect. Note that in this experiment, the improvement is quantified with respect to a spirit leveling-only solution obtained by applying the same constraint.

Experiment IV: adding hydrodynamic leveling connections among all European seas—This experiment aims to answer the question how the quality impact changes when

allowing connections among all European seas. Indeed, this requires a model covering all European seas. Note that in line with others (e.g., Lea et al., 2015) we treat the Black Sea as a closed basin. This means that any connection between a Black Sea tide gauge to one located at the coast of another sea is not allowed.

Experiment V: using the tide gauges within the 3D DCSM-FM domain only—In the project of which this study is part, we aim to develop a hydrodynamic model known as the 3D DCSM-FM (Zijl et al., 2020). One objective is to use this model to conduct hydrodynamic leveling. In Experiment V, we assess the quality impact in case we only have this model available. That is, in case we can only exploit the tide gauges available within the 3D DCSM-FM domain (see Fig. 2.2). Note that we only used the tide gauges that are located at a distance of at least one degree from the boundaries.

2.5. RESULTS AND DISCUSSION

In this section, we present and discuss the results of the experiments introduced in Sect. 2.4. In doing so, we first assess for all experiments the impact on the quality of the EVRF2019 (first research objective of this study). Thereafter, we assess which connections, and hence tide gauges, are most profitable in terms of impact. These are the ones we need to focus on in the development of the hydrodynamic model(s) (second research objective of this study).

2.5.1. THE IMPACT OF ADDING HYDRODYNAMIC LEVELING DATA ON THE QUALITY OF THE EVRF

EXPERIMENT 0: USING SPIRIT LEVELING DATA ONLY

Figure. 2.3 shows a map of the SDs of the adjusted heights for the spirit leveling-only solution (i.e., the solution which serves as a reference in experiments I, II, IV, and V). They cover a broad range of values between, ~ 5 and ~ 75 mm. The median value is 13.8 mm. Table 2.1 shows the median SD per country. Both Table 2.1 and Fig. 2.3 clearly show the large regional deviations; the values are lowest at the center (including Belgium, Germany, the Netherlands, Austria, Czech Republic, and Poland) and increase toward the margins. The mentioned countries all have high-quality leveling data (see Table 3 of Sacher and Liebsch (2019)). Moreover, most of the datum points are located in these countries (see Fig. 2.1). As mentioned before, the propagated SDs are not only affected by the precision of the observations, but also by the height marker distance to the datum points (Sacher & Liebsch, 2019). The absence of a datum point in the Scandinavian Peninsula explains why in Finland and Sweden (despite having high-quality leveling data) the adjusted height SDs are high compared to those in the center. Figure. 2.4 shows a map of the redundancy numbers and Table 2.1 presents the median redundancy number per country and for the entire network. Overall, the redundancy numbers, are small with values ranging from 0.002 (France) to 0.411 (Norway). The low value for France can be explained by the fact that the French leveling network contains many height markers that are only connected to one other height marker (see Fig. 2.1). The better the network is connected, the higher the redundancy numbers would be. By adding hydrodynamic leveling observations, we expect to increase the redundancy numbers in the coastal areas.

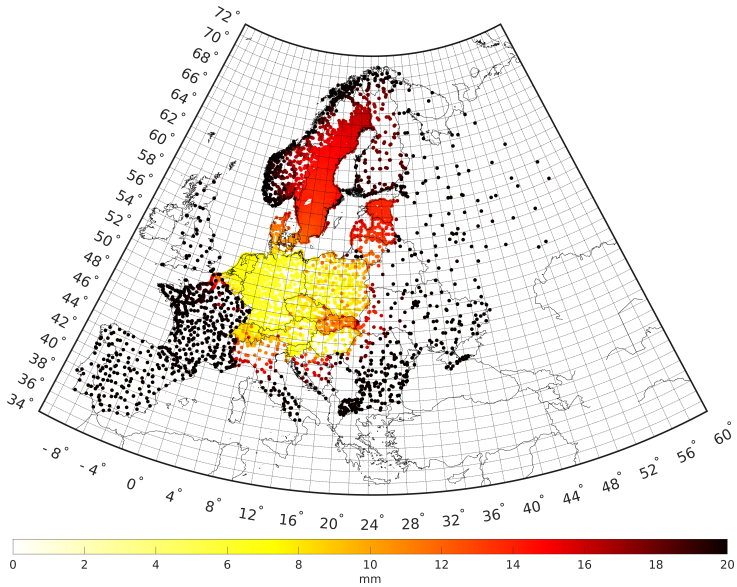


Figure 2.3: Propagated SDs of the adjusted heights in Experiment 0.

EXPERIMENT I: ALLOWING FOR CONNECTIONS WITHIN BASINS ONLY

Given the total number of 186 candidate tide gauges (Sect. 2.3.2) spread over 4 separate sea basins, the number of independent connections we can add is 182 (note that it is only allowed to establish connections within the same sea basins).

Adding all 182 connections reduces the median-propagated SD of all adjusted heights from 13.8 to 8.6 mm. This corresponds to an improvement of 38%. The improvement differs strongly per country as shown in Table 2.1; values range from 1% (Slovakia) to 60% (Great Britain). We observe larger improvements for coastal countries, except for the countries which are located along the Black Sea (see Fig. 2.5). We also notice a more significant improvement for coastal countries at the perimeter of the UELN network (i.e., Portugal, Spain, Great Britain, and the Scandinavian countries). The lower improvements for the Black Sea countries are explained by the fact that hydrodynamic leveling connections that link the Black Sea to the other European seas were not allowed. The reason why the impact is largest in Great Britain can be understood when we consider that the existing connection of the Great Britain leveling network to the remaining part of the UELN is extremely weak; it is connected by just two leveling campaigns through the channel tunnel. Since there are many tide gauges in Great Britain, hydrodynamic leveling allows to tie Great Britain much stronger to the rest of the UELN.

In terms of reliability, we observe increased redundancy numbers for the spirit leveling observations near the coastline (see Fig. 2.6). For most of these observations, the improvement is between 0.02 and 0.1. For about 1% of the observations the improvement is larger, the maximum being 0.9. In Great Britain, the numbers increase almost throughout the whole country. Here, they range between 0 and 0.5. The median redundancy number for Great Britain improves from 0.278 to 0.462 (note that in computing this value the redun-

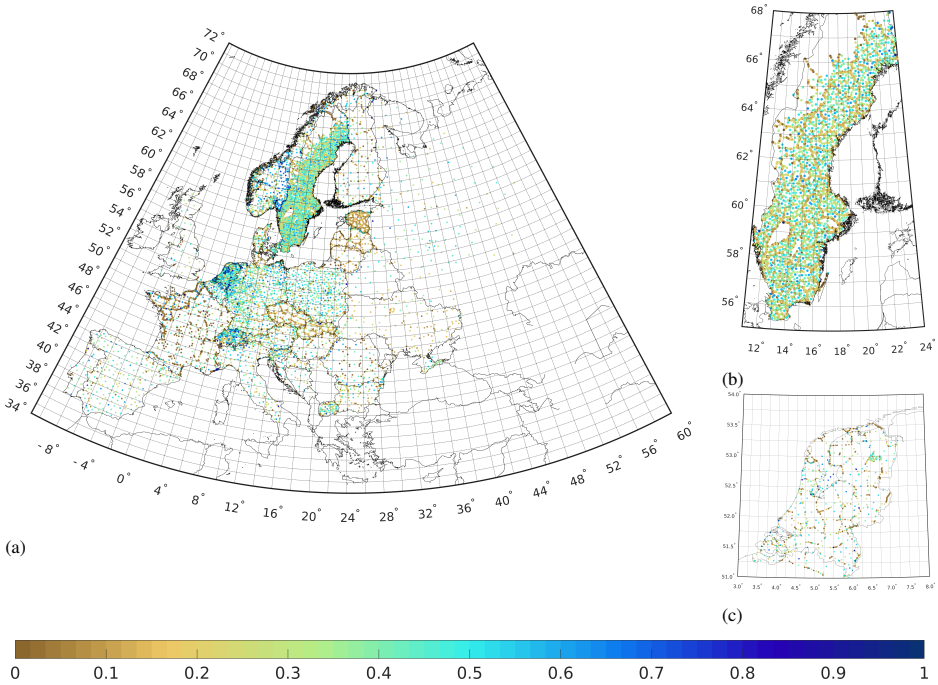


Figure 2.4: Redundancy numbers of the leveling observations in the spirit leveling-only solution (Experiment 0). Panels b and c show a zoom-in for Sweden and the Netherlands, respectively.

dancy numbers associated to the hydrodynamic leveling observations are excluded). For the other countries, we hardly observed any change in the median redundancy number. (For that reason, they are not included in Table 2.1.) This is reflected by the minor change in the median redundancy number for the whole network (0.161 versus 0.154 for Experiment 0).

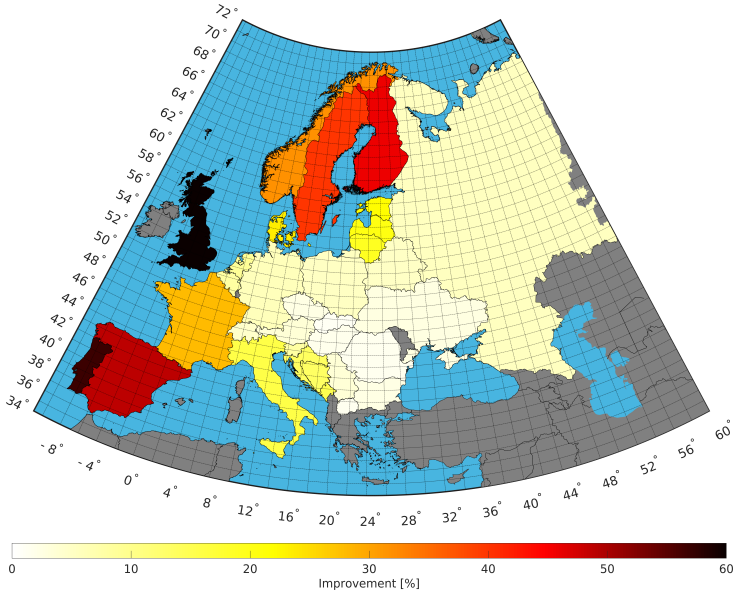


Figure 2.5: Improvement of the median-propagated SD of the adjusted heights per country obtained in Experiment I.

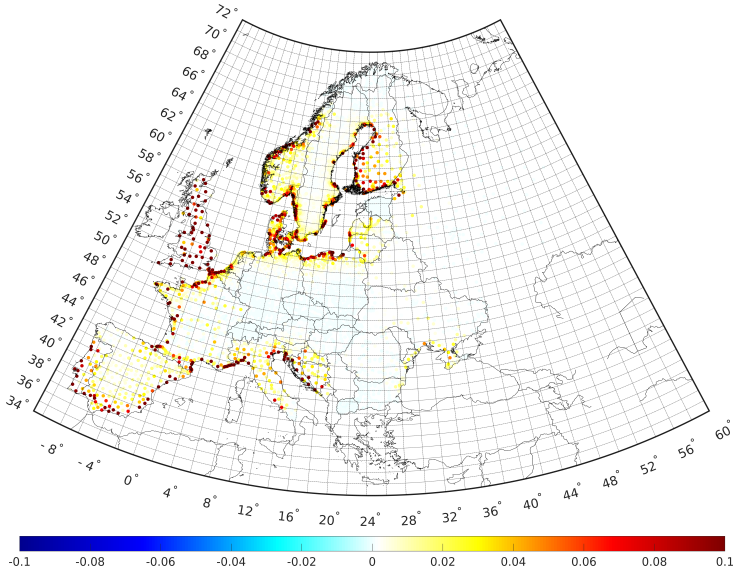


Figure 2.6: Difference between the redundancy numbers obtained in Experiments I and 0.

Table 2.1: Number of tide gauges per country and in total, and the median propagated standard deviation (SD) of the adjusted heights in millimeters (per country and in total).

Country	Number of tide gauges	Experiment 0		Experiment I		Experiment II				Experiment IV		Experiment V	
		SD	red. number	Noise 3cm		Noise 1cm		Noise 5cm		Noise 3cm			
Austria	0	6.8	0.220	6.6	(4)	6.4	(6)	6.7	(2)	6.6	(4)	6.8	(0)
Belgium	3	7.7	0.373	7.1	(8)	6.2	(19)	7.3	(5)	6.5	(15)	7.5	(2)
Bosnia and Herzegovina	0	16.3	0.279	14.1	(13)	12.9	(21)	15.1	(8)	13.7	(16)	16.3	(0)
Bulgaria	1	22.8	0.227	22.2	(2)	21.9	(4)	22.5	(1)	22.2	(2)	22.8	(0)
Belarus	0	19.2	0.374	18.3	(5)	17.8	(7)	18.6	(3)	18.1	(6)	19.2	(0)
Croatia	5	14.8	0.061	12.7	(14)	11.3	(23)	13.9	(6)	12.5	(16)	14.8	(0)
Czech Republic	0	7.4	0.076	7.2	(2)	7.1	(4)	7.3	(1)	7.2	(3)	7.4	(0)
Denmark	14	10.9	0.175	8.3	(24)	6.7	(38)	9.1	(17)	7.8	(29)	9.2	(16)
Estonia	1	13.8	0.083	10.7	(22)	7.2	(47)	12.2	(12)	10.2	(26)	13.7	(1)
Finland	19	18.3	0.158	9.9	(46)	7.4	(60)	11.9	(35)	9.0	(51)	16.5	(10)
France	20	20.7	0.002	14.9	(28)	13.2	(36)	16.4	(21)	14.6	(30)	15.7	(24)
Germany	7	6.5	0.316	6.1	(6)	5.8	(11)	6.2	(4)	5.8	(10)	6.5	(0)
Great Britain	20	31.1	0.278	12.5	(60)	8.3	(73)	15.0	(52)	11.8	(62)	13.1	(58)
Hungary	0	7.4	0.120	7.3	(2)	7.2	(3)	7.3	(1)	7.3	(2)	7.4	(0)
Italy	10	14.1	0.260	11.9	(16)	10.7	(24)	12.8	(10)	11.4	(19)	14.1	(0)
Latvia	2	13.7	0.117	10.9	(21)	8.2	(40)	12.2	(11)	10.4	(24)	13.6	(1)
Lithuania	1	12.7	0.067	10.2	(19)	8.4	(34)	11.3	(11)	9.8	(23)	12.6	(1)
Macedonia	0	28.1	0.297	27.6	(2)	27.4	(2)	27.8	(1)	27.6	(2)	28.1	(0)
Montenegro	0	23.4	0.279	19.9	(15)	17.6	(25)	21.3	(9)	19.5	(17)	23.4	(0)
Netherlands	9	6.8	0.167	6.2	(9)	5.6	(17)	6.5	(5)	5.7	(16)	6.7	(1)
Norway	18	19.1	0.411	13.0	(32)	10.7	(44)	14.6	(23)	12.5	(34)	14.0	(26)
Poland	6	7.8	0.355	7.4	(5)	7.0	(10)	7.6	(3)	7.2	(8)	7.8	(0)
Portugal	3	45.6	0.348	19.6	(57)	17.2	(62)	22.8	(50)	19.5	(57)	44.1	(3)
Romania	0	20.5	0.221	20.2	(2)	19.9	(3)	20.3	(1)	20.2	(2)	20.5	(0)
Russia	2	27.7	0.368	26.1	(6)	25.5	(8)	26.6	(4)	25.9	(6)	27.2	(2)
Serbia	0	19.7	0.231	19.0	(4)	18.7	(5)	19.3	(2)	18.8	(4)	19.7	(0)
Slovakia	0	10.3	0.104	10.2	(1)	10.1	(2)	10.2	(1)	10.1	(1)	10.3	(0)
Slovenia	2	7.0	0.244	6.4	(8)	5.9	(15)	6.6	(5)	6.2	(11)	7.0	(0)
Spain	7	40.6	0.344	20.8	(49)	19.4	(52)	22.7	(44)	20.6	(49)	39.1	(4)
Sweden	35	14.0	0.112	8.4	(40)	7.0	(50)	9.7	(31)	7.7	(45)	10.3	(27)
Switzerland	0	7.6	0.171	7.2	(5)	7.0	(7)	7.3	(3)	7.2	(5)	7.5	(0)
Ukraine	1	25.0	0.191	24.3	(3)	22.7	(9)	24.6	(2)	24.3	(3)	25.0	(0)
Total	**186**	**13.8**	**0.154**	**8.6**	**(38)**	**7.2**	**(48)**	**9.8**	**(29)**	**8.0**	**(42)**	**10.3**	**(25)**

For Experiments I, II, IV, and V the number between the brackets indicates the improvement in percentage relative to the value obtained for Experiment 0. For Experiment 0, the table also includes the median redundancy number per country and for all observations together.

EXPERIMENT II: VARYING THE NOISE LEVEL

The total number of added connections is the same as in Experiment I. Table 2.1 shows the median SDs when assuming a noise SD of 1 and 5 cm for each hydrodynamic leveling connection. The values corresponding to the other noise levels are in between these values (the ones corresponding to a noise level of 3 cm are those of Experiment I). As expected, the quality impact lowers with increasing noise level. For the most optimistic scenario, we found an improvement of 48%; however, for a noise level of 5 cm we still gain 29%. As expected, the values are quite different per country but show the same behavior for different noise levels. We always observe the largest improvement in Great Britain and Portugal, though the magnitude decreases with increasing noise level (see Table 2.1).

Figure. 2.7 shows for all considered noise levels the improvement of the overall median-propagated height SD (i.e., computed over all height markers) as a function of the number of added hydrodynamic leveling connections. We notice that the decrease of the level of improvement achieved by adding all connections is not linearly related to the increase of the noise level. Indeed, the distance between the curves for the number of added connections being equal to 182 gets smaller and smaller when the noise level goes up. We also observe that in order to achieve a certain level of improvement, more connections need to be added with increasing noise level. For example, a 25% improvement can be obtained with just 2 connections in case we have 1 cm accurate data, while 38 connections are needed when the accuracy level is 5 cm. Note, again, that the final level of improvement also depends on the available number of tide gauges as well as their location. In our experiments, both parameters are fixed. Still, we do not exploit all tide gauges available in Europe (see Sect. 2.3.2).

Varying the noise level did not change the median redundancy numbers for the entire network and per country.

EXPERIMENT III: USING THE INNER CONSTRAINT ADJUSTMENT

Experiment III uses the same number of hydrodynamic leveling connections as Experiment I does. The median-propagated SD of all adjusted heights is 7.0 mm, corresponding to an improvement of 28% (see Table 2.2). Note that the improvement is quantified compared to a spirit leveling-only solution obtained by applying the inner constraint adjustment. For this solution, the median-propagated SD was 9.8 mm. Compared to Experiment I, the improvement in the Netherlands, Germany, Belgium is more than 15% larger (cf. Figs. 2.8 and 2.5). For most other countries, we also observe improvements, though the magnitudes are smaller.

EXPERIMENT IV: ADDING HYDRODYNAMIC LEVELING CONNECTIONS AMONG ALL EUROPEAN SEAS

The total number of independent connections we can add in this experiment is 184; instead of 4, we have 2 sea basins (we treat the Black Sea as a closed basin).

The median-propagated SD of all adjusted heights is 8.0 mm, corresponding to an improvement of 42% compared to Experiment 0. This is, indeed, very close to the values obtained in Experiment I (8.6 mm and 38%, respectively). Also the improvements per country are quite similar (see Table 2.1); the largest increment in improvement is observed for the Netherlands where we go from 9% improvement to 16%. Keep in mind, however,

Table 2.2: Standard deviation (SD) of spirit leveling-only solution (in millimeter) where the inner constraint is used for the network adjustment and the standard deviation (SD) and percentage of improvement (values in bracket) per country for Experiment III.

Country	Reference spirit leveling solution	Experiment III
Austria	8.9	7.4 (17)
Belgium	7.9	5.8 (27)
Bosnia and Herzegovina	18.5	15.1 (18)
Bulgaria	24.7	23.4 (5)
Belarus	19.1	18.0 (6)
Croatia	16.6	13.5 (19)
Czech Republic	9.0	7.6 (15)
Denmark	8.4	6.7 (20)
Estonia	13.1	9.5 (27)
Finland	15.0	8.4 (44)
France	19.7	13.9 (30)
Germany	6.9	5.2 (25)
Great Britain	30.8	11.4 (63)
Hungary	9.9	8.6 (14)
Italy	15.0	12.2 (19)
Latvia	13.1	9.8 (25)
Lithuania	12.3	9.4 (23)
Macedonia	29.4	28.5 (3)
Montenegro	25.6	20.9 (18)
Netherlands	7.0	4.8 (31)
Norway	15.7	11.7 (25)
Poland	9.0	7.3 (18)
Portugal	44.6	19.1 (57)
Romania	22.2	21.2 (5)
Russia	26.5	25.6 (3)
Serbia	21.3	19.6 (8)
Slovakia	12.1	11.0 (9)
Slovenia	9.4	7.5 (20)
Spain	39.6	20.6 (48)
Sweden	9.4	6.2 (34)
Switzerland	8.3	6.9 (17)
Ukraine	25.9	25.0 (4)
Total	**9.8**	**7.0 (28)**

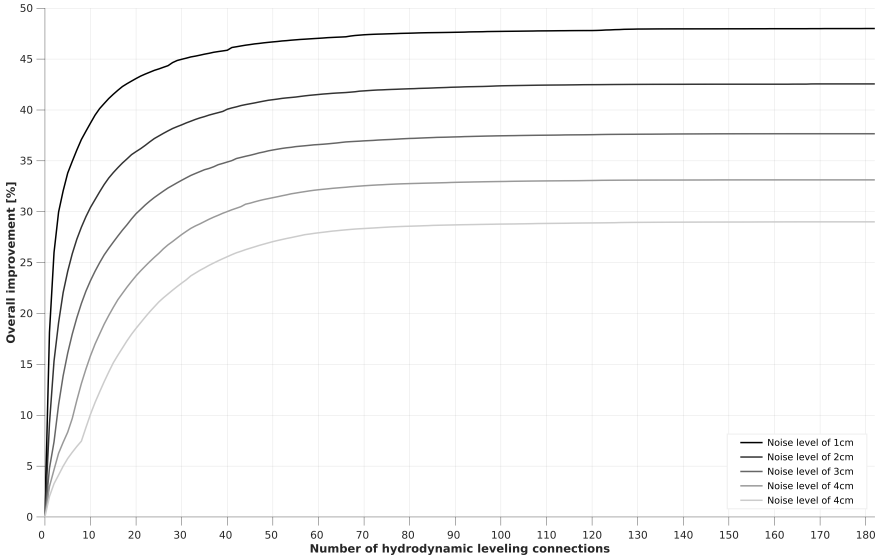


Figure 2.7: Overall improvement of the solutions obtained by adding hydrodynamic leveling data as a function of the number of added hydrodynamic leveling connections for different noise levels.

that in absolute numbers we go from 6.2 to 5.7 mm. So, the change is just at the sub-mm level.

In terms of reliability, we do not observe any significant differences compared to the results obtained in Experiment I. The median redundancy value for the entire network is 0.161 (also 0.161 in Experiment I). The median per country is almost the same as in Experiment I. From this, we conclude that the scenario in which we rely on a model that covers all European seas hardly changes the impact of adding hydrodynamic leveling on the quality of the EVRF. From a practical point of view, this is good news. First, there is no need to resolve the physics of water flow in narrow and complex waters connecting the basins, e.g., the Strait of Gibraltar. Second, the computational load is significantly reduced as the domain of the hydrodynamic model is much smaller.

EXPERIMENT V: USING THE TIDE GAUGES WITHIN THE 3D DCSM-FM DOMAIN ONLY

The total number of available PSMSL tide gauges within the 3D DCSM-FM domain is 77. The resulting number of added connections corresponds to the theoretical maximum of 76. The median value of the propagated SD after adding all 76 hydrodynamic leveling connections is 10.3 mm corresponding to an improvement of 25%. The median-propagated SDs per country are quite different compared with the values obtained in the other experiments (see Table 2.1). For Great Britain, the improvement is still large (58%). For the Scandinavian countries, the improvements are also substantial (27%, 26%, 16% and 10% for Sweden, Norway, Denmark, and Finland respectively) but lower than in Experiments I and IV. For all other countries, except France (24%), the improvements are equal to or less than 4%. Similar to what we observed in Experiment I, the redundancy numbers for the leveling connections near the coastline (here limited to the 3D DCSM-FM domain) are increased

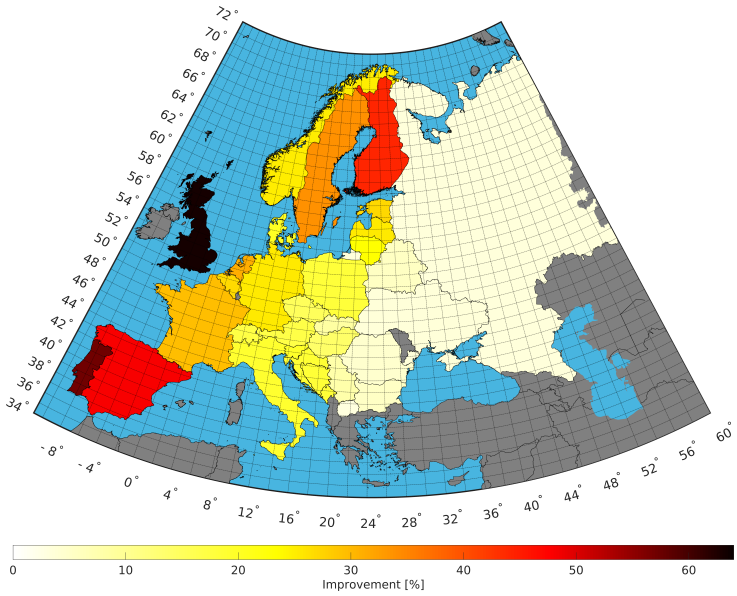


Figure 2.8: Improvement of the median propagated SD of the adjusted heights per country obtained in Experiment III.

(plot not shown). The median value, however, did not change significantly.

2.5.2. IDENTIFYING THE HIGH-IMPACT TIDE GAUGES

In this section, we want to identify the most profitable (in terms of quality impact on the realization of the EVRS) hydrodynamic leveling connections c.q. tide gauges. As motivated in the introduction, the answer to this question provides guidance where to focus in the development/calibration of the model. Note that in this section, we do not consider the results obtained from the inner constraint adjustment (Experiment III).

To start the analysis, we again point to Fig. 2.7 that shows for different noise levels the overall improvement as a function of the number of added hydrodynamic leveling connections. The third curve is associated with Experiment I. It shows a clear asymptotic behavior with a rapid increase of improvement in the beginning. Indeed, by just adding 20 connections the obtained improvement is already 30%. (The maximum improvement in Experiment I is 38%.) Also, in the other experiments (plots not shown in the paper), we observe this asymptotic behavior. To compare the results of the various experiments, we plotted in Fig. 2.9 the percentage of unique tide gauges being involved in establishing the hydrodynamic leveling connections as a function of the percentage of achieved improvement relative to the maximum improvement. From this plot, we observe that for all scenarios in which we used the 3 cm noise level (Experiments I, IV, and V) 75% of the maximum improvement can be achieved by using only 16–23% of the available tide gauges (which number is 186, 186, and 77 for Experiments I, IV, and V, respectively). Only when we increase/decrease the noise level, more/less tide gauges are needed to achieve a certain level of improvement (see the curves associated to Experiment II in Fig. 2.9). Still,

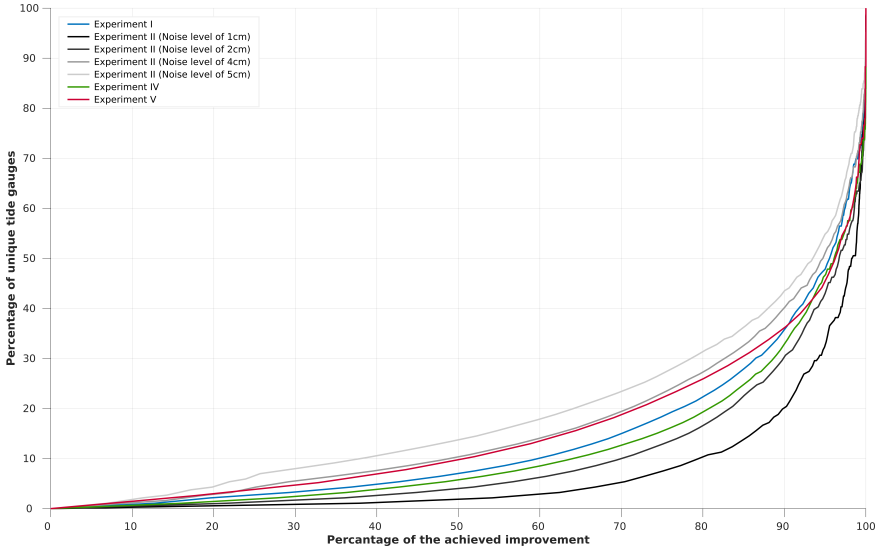


Figure 2.9: The percentage of unique tide gauges being involved in establishing the hydrodynamic leveling connections as a function of the percentage of achieved improvement relative to the total improvement for Experiments I, II, IV, and V.

even for a 5-cm noise level only 27% of the tide gauges are needed to achieve 75% of the maximum improvement. This is a positive result as it shows that when adding a limited number of hydrodynamic leveling connections between a small number of tide gauges a substantial improvement in the quality of the EVRF2019 can be achieved. Moreover, the analysis suggests that adding more tide gauges than the ones considered in this study will not significantly increase the overall level of improvement.

Next, we analyzed in more detail which tide gauges are involved in establishing the hydrodynamic leveling connections that resulted in 75% of the maximum improvement. For Experiments I, IV, and V, this involves 17, 15, and 9 connections, respectively. In Experiment I and IV, tide gauges in 8 and 7 countries are involved. Striking is that in both experiments, most connections involve tide gauges in Sweden (14 in Experiment I and 13 in Experiment IV). Note that these numbers do not represent the number of unique tide gauges involved. In some cases, tide gauges are involved in more connections. As seen earlier, increasing/decreasing the noise level (Experiment II) results in more/less connections being needed to achieve 75% of the maximum improvement. This basically means more tide gauges per country; the number of countries involved only slightly increased from 7 to 11 when we increase the noise level from 1 to 5 cm. Again, by far the Swedish tide gauges are favored most. They appear in 5, 9, 14, 17, and 17 connections by increasing the noise level from 1 to 5 cm. In Experiment V, the distribution over the countries is different. Only tide gauges in 5 countries are involved, namely France, the Netherlands, Norway, Sweden, and Belgium. Here, tide gauges from Netherlands, Sweden, and Norway are favored most; they are used in 7, 5, and 4 connections, respectively. In both Belgium and France, only one tide gauge is used.

The reason why in Experiments I, II, and IV the Swedish tide gauges are favored can be understood as follows. The criterion used to identify the best set of hydrodynamic leveling connections is based on the median SD computed over all height markers. In computing this median value, the country with the highest number of height markers will contribute the most. Sweden has the highest number of height markers; 32% of all height markers are located in Sweden. For all other countries, the percentages are below 13%. Of course, other metrics to identify the best set of hydrodynamic leveling connections are possible resulting in other tide gauges to be favored. For example, one might aim to minimize the median SD for a specific country or some countries.

The remaining question is whether indeed the identified tide gauges are important or whether it is sufficient to have tide gauges available in the specific countries. To answer this question, we repeated Experiment I after removing the 34 tide gauges involved in the 17 connections that allowed to achieve 75% of the maximum improvement. The resulting maximum improvement is 32%, which is still significant compared to the 38% obtained when including the 34 tide gauges. Again, the remaining Swedish tide gauges are favored. Hence, we conclude that in implementing model-based hydrodynamic leveling the tide gauges in the countries having the highest numbers of height markers are favored most. Moreover, the precise location of the involved tide gauges does not matter too much.

2.6. SUMMARY AND CONCLUSIONS

This work is part of the Versatile Hydrodynamics project that aims to develop a forecasting system for total water levels. To be able to assimilate observed total water levels from offshore tide gauges and tide gauges located at islands, an accurate unified VRF is needed. Obtaining such a VRF requires a technique to transfer heights over large water bodies. In this study, we exploit model-based hydrodynamic leveling proposed by Slobbe et al. (2018) to do so. The method requires information about the MWL difference between tide gauges. Two main objectives are addressed. First, we assessed the impact of adding model-based hydrodynamic leveling connections to the UELN dataset on the precision and reliability of the EVRF. Second, we assessed which connections, and hence tide gauges, are most profitable in terms of impact to focus on in realizing the EVRS. To achieve these objectives, geodetic network analyses were conducted based on different scenarios. In all experiments, it was assumed that we have access to hydrodynamic models covering either all European Seas surrounding the European mainland or parts of it. Moreover, it was assumed that these models provide the required MWL differences with uniform precision.

To identify which hydrodynamic leveling connections contribute most to the quality of the EVRF, a heuristic search algorithm was implemented. In each step, the algorithm identifies from all remaining possible connections the ones that result in the lowest median propagated SD of the adjusted heights. The algorithm stops when no more connections are remaining.

In total, five experiments were conducted. Experiment I assumes we have access to a set of hydrodynamic models, each covering a separate sea basin. Moreover, it assumes that each model provides the MWL differences with an SD of 3 cm. So, in the first experiment, we allowed for hydrodynamic leveling connections only among tide gauges located in the same basin. Four basins were considered, the Mediterranean Sea, Black Sea, Baltic Sea, and the North-East Atlantic region including the North Sea. Among 186 tide gauges, our

search algorithm identified 182 independent leveling connections that allowed to reduce the median SD of the adjusted heights from 13.8 mm for the spirit leveling-only solution to 8.6 mm, equivalent to a 38% improvement. Except for the countries around the Black Sea, coastal countries benefit the most with a maximum improvement of 60% for Great Britain. In terms of reliability, the impact was assessed by the redundancy numbers. We found increased redundancy numbers for observations close to the coast and over the entire Great Britain.

In Experiment II, we assessed the impact of the assumed precision for the added hydrodynamic leveling data. More particular, we varied the uniform precision from 1 to 5 cm in steps of 1 cm. As expected, lower precision resulted in lower impact. Though the results show that even in case the assumed precision is 5 cm, the overall improvement is still 29%. The increase of the improvement as a function of the number of added leveling connections, however, decreases. That means that more connections need to be added to achieve a certain level of improvement.

In Experiments III, we applied the inner-constraint adjustment in order to get rid of the impact the adopted datum points have on the results.

Experiment IV assumes we have access to a hydrodynamic model covering all European Seas (only the Black Sea is treated as a separate basin) surrounding the European mainland providing MWL differences with the same accuracy as in Experiment I. The overall improvement hardly changed compared to Experiment I; the median SD of the adjusted heights was 8.0 mm (8.6 mm for Experiment I).

In Experiment V, we assessed the impact in case our hydrodynamic model only covers the North Sea and part of the northeast Atlantic Ocean. In this scenario, the overall improvement was reduced to 25%. For Great Britain, however, the improvement is still about 60%.

Based on the results, we conclude that adding hydrodynamic leveling can highly impact the quality of the EVRF in terms of precision in case it is implemented involving tide gauges in entire Europe. Here, it is not required to rely on a model covering all European waters. That is, we can rely on models targeted for specific waters. Moreover, even in case the precision of the model-derived MWL is low (say, hydrodynamic leveling connections with an SD of 5 cm), the impact is substantial. In terms of reliability, the impact is confined to the coastal region. The exception is the leveling network of Great Britain. It might be interesting to study whether hydrodynamic leveling allows to identify the systematic errors in the British leveling network (Penna et al., 2013).

Another conclusion we can draw from these results is that the tide gauges involved in the connections that are most profitable in terms of impact are located in the country in which most height markers are located. The impact hardly depends on the geographic location of the tide gauges within a country; the impact did not drop dramatically after excluding the tide gauges involved in those connections that are responsible for 75% of the maximum improvement for Experiment I.

In identifying the hydrodynamic leveling connection to be added in each step as well as in reporting the improvements in the propagated SDs and redundancy numbers, we have used the median of the propagated height SDs rather than the mean as reported by Sacher and Liebsch (2019). This choice is motivated by the fact that the propagated SDs of all height markers together are not normally distributed (when considered per country, the

distribution is in most cases close to normal). The use of the mean instead of the median, however, hardly impacts the main findings of this study (experiments not included in this paper). These include the improvements observed per country [except for France for which the distribution of the propagated height SDs is skewed as a result of the low accuracy of one of the two available leveling datasets due to the presence of systematic errors (Sacher & Liebsch, 2019)], the behavior of the overall improvement as a function of the number of added hydrodynamic leveling connections, and the conclusion regarding the tide gauges involved in the connections that are most profitable in terms of impact. What did change for all experiments are the magnitudes of the overall improvements (i.e., the improvement of the propagated SDs computed over all height markers together). In terms of the mean propagated SD, the improvements are $> 10\%$ lower than in terms of the median.

In view of the requirement to have a VRF with a one centimeter accuracy (see Sect 2.1), we conclude that based on the formal errors, adding hydrodynamic leveling allows to achieve this target. Indeed, even in case we obtain the hydrodynamic leveling data with a SD of 5 cm the obtained median SD of the adjusted heights is < 1 cm (Experiment II). Regionally, though, the SDs vary significantly. More importantly, however, as stated above the numbers only represent the *formal precision*. Existing systematic errors in the spirit leveling data, vertical land motion, long-term sea level variations, etc., all reduce the ultimate accuracy. To what extent remains to be investigated.

A point of attention in the operationalization of the technique is the reduction of the observation- and model-derived tide gauge records to the reference epoch adopted in the EVRF (epoch 2000.0). This reduction is needed to correct for vertical land motion and/or long-term sea level variations. Applying such a reduction is, however, only feasible if the tide gauge records are sufficiently long and all needed metadata are available. An example of needed metadata is the epoch when the tide gauge benchmark is connected to the height system by means of leveling.

In a future work, we will derive a proper stochastic model for the hydrodynamic leveling dataset. Moreover, we will implement the method using the 3D DCSM-FM model currently under development. The upcoming release of the model will expand the model domain to the Baltic Sea. This would allow to establish hydrodynamic leveling connections among the North Sea and Baltic Sea tide gauges. At the same time, a more beneficial idea might be to launch a European project to develop regional hydrodynamic models to implement hydrodynamic leveling at the European scale.

BIBLIOGRAPHY

Aigner, M., & Ziegler, G. M. (1998). Cayley's formula for the number of trees. In *Proofs from THE BOOK* (pp. 141–146). Springer, Berlin. https://doi.org/https://doi.org/10.1142/9789814513388_0010

Amiri-Simkooei, A. R., Asgari, J., Zangeneh-Nejad, F., & Zaminpardaz, S. (2012). Basic concepts of optimization and design of geodetic networks. *Journal of Surveying Engineering*, *138*(4), 172–183. https://doi.org/10.1061/(asce)su.1943-5428.0000081

Amjadiparvar, B., Rangelova, E., & Sideris, M. G. (2016). The GBVP approach for vertical datum unification: recent results in North America. *Journal of Geodesy*, *90*(1), 45–63. https://doi.org/10.1007/s00190-015-0855-8

Amos, M. J., & Featherstone, W. E. (2009). Unification of New Zealand's local vertical datums: iterative gravimetric quasigeoid computations. *Journal of Geodesy*, *83*(1), 57–68. https://doi.org/10.1007/s00190-008-0232-y

Baarda, W. (1968). A testing procedure for use in geodetic networks. *Publication on Geodesy, New Series*, *2*.

Bossler, J. D. (1984). *Standards and specifications for geodetic control networks*. National Geodetic Information Branch, NOAA, FGCC.

Cartwright, D. E., & Crease, J. (1963). A comparison of the geodetic reference levels of England and France by means of the sea surface. *P. Roy. Soc. Lond. A. Mat.*, *273*(1355), 558–580.

Catalão, J., & Sevilla, M. (2008). The use of ICAGM07 geoid model for vertical datum unification on Iberia and Macaronesian islands. *American Geophysical Union, Fall Meeting 2008, abstract*.

Deltares. (2021). Delft3D flexible Mesh suite.

Denker, H. (2015). A new European gravimetric (quasi) geoid EGG2015. *26th IUGG General Assembly, June22–July2*.

Donnelly, C., Andersson, J. C. M., & Arheimer, B. (2015). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, *61*(2), 255–273. https://doi.org/10.1080/02626667.2015.1027710

EMODnet bathymetry consortium. (2018). EMODnet digital bathymetry (DTM 2018).

Featherstone, W. E., & Filmer, M. S. (2012). The north-south tilt in the Australian height datum is explained by the ocean's mean dynamic topography. *Journal of Geophysical Research: Oceans*, *117*(C8). https://doi.org/10.1029/2012JC007974

Filmer, M. S., & Featherstone, W. E. (2012). A Re-Evaluation of the Offset in the Australian Height Datum Between Mainland Australia and Tasmania. *Marine Geodesy*, *35*(1), 107–119. https://doi.org/10.1080/01490419.2011.634961

Filmer, M. S., Featherstone, W. E., & Claessens, S. J. (2014). Variance component estimation uncertainty for unbalanced data: application to a continent-wide vertical

datum. *Journal of Geodesy*, *88*(11), 1081–1093. https://doi.org/10.1007/s00190-014-0744-6

Filmer, M. S., Hughes, C. W., Woodworth, P. L., Featherstone, W. E., & Bingham, R. J. (2018). Comparison between geodetic and oceanographic approaches to estimate mean dynamic topography for vertical datum unification: evaluation at Australian tide gauges. *Journal of Geodesy*, *92*(12), 1413–1437. https://doi.org/10.1007/s00190-018-1131-5

Gerlach, C., & Rummel, R. (2013). Global height system unification with goce: A simulation study on the indirect bias term in the gbvp approach. *Journal of Geodesy*, *87*(1), 57–67. https://doi.org/10.1007/s00190-012-0579-y

Heck, B., & Rummel, R. (1990). Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In H. Sünkel & T. Baker (Eds.), *Sea surface topography and the geoid* (pp. 116–128). Springer New York.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holgate, S. J., Matthews, A., Woodworth, P. L., Rickards, L. J., Tamisiea, M. E., Bradshaw, E., Foden, P. R., Gordon, K. M., Jevrejeva, S., & Pugh, J. (2013). New data systems and products at the permanent service for mean sea level. *Journal of Coastal Research*, *29*(3), 493–504. https://doi.org/doi:10.2112/JCOASTRES-D-12-00175.1

Ihde, J., Augath, W., & Sacher, M. (2002). The Vertical Reference System for Europe. In *International association of geodesy symposia* (pp. 345–350). Springer, Berlin. https://doi.org/10.1007/978-3-662-04683-8_64

Lea, D. J., Mirouze, I., Martin, M. J., King, R. R., Hines, A., Walters, D., & Thurlow, M. (2015). Assessing a new coupled data assimilation system based on the met office coupled atmosphere–land–ocean–sea ice model. *Monthly Weather Review*, *143*(11), 4678–4694. https://doi.org/10.1175/mwr-d-15-0174.1

Mehlstäubler, T. E., Grosche, G., Lisdat, C., Schmidt, P. O., & Denker, H. (2018). Atomic clocks for geodesy. *Reports on Progress in Physics*, *81*(6), 064401. https://doi.org/10.1088/1361-6633/aab409

Moritz, H. (2000). Geodetic Reference System 1980 [10.1007/s001900050278]. *J. Geod.*, *74*(1), 128–162.

Müller, J., Dirkx, D., Kopeikin, S. M., Lion, G., Panet, I., Petit, G., & Visser, P. N. A. M. (2017). High performance clocks and gravity field determination. *Space Science Reviews*, *214*(1), 5. https://doi.org/10.1007/s11214-017-0431-z

Ogundare, J. O. (2018). *Understanding least squares estimation and geomatics data analysis*. Wiley, New York. https://doi.org/10.1002/9781119501459.ch14

Penna, N. T., Featherstone, W. E., Gazeaux, J., & Bingham, R. J. (2013). The apparent British sea slope is caused by systematic errors in the levelling-based vertical datum. *Geophysical Journal International*, *194*(2), 772–786. https://doi.org/10.1093/gji/ggt161

Permanent Service for Mean Sea Level (PSMSL). (n.d.). Tide Gauge Data [Retrieved 1 Jun 2020].

Proudman, J. (1953). *Dynamical oceanography*. Methuen.

Rummel, R., & Teunissen, P. (1988). Height datum definition, height datum connection and the role of the geodetic boundary value problem. *Bulletin Géodésique*, *62*(4), 477–498. https://doi.org/10.1007/bf02520239

Sacher, M., & Liebsch, G. (2019). EVRF2019 as new realization of EVRS.

Sánchez, L., & Sideris, M. . (2017). Vertical datum unification for the International Height Reference System (IHRS). *Geophysical Journal International*, *209*(2), 570–586. https://doi.org/10.1093/gji/ggx025

Schwarz, K. P., Sideris, M. G., & Forsberg, R. (1987). Orthometric heights without leveling. *J. Surv. Eng*, *113*(1), 28–40.

Seemkooei, A. A. (2001a). Comparison of reliability and geometrical strength criteria in geodetic networks. *Journal of Geodesy*, *75*(4), 227–233. https://doi.org/10.1007/s001900100170

Seemkooei, A. A. (2001b). Strategy for designing geodetic network with high reliability and geometrical strength. *Journal of Surveying Engineering*, *127*(3), 104–117. https://doi.org/10.1061/(asce)0733-9453(2001)127:3(104)

Slobbe, D. C., Klees, R., Verlaan, M., Zijl, F., Alberts, B., & Farahani, H. H. (2018). Height system connection between island and mainland using a hydrodynamic model: a case study connecting the Dutch Wadden islands to the Amsterdam ordnance datum (NAP). *J. Geod.*, *92*(12), 1439–1456. https://doi.org/10.1007/s00190-018-1133-3

Stammer, D., Ray, R. D., Andersen, O. B., Arbic, B. K., Bosch, W., Carrère, L., Cheng, Y., Chinn, D. S., Dushaw, B. D., Egbert, G. D., et al. (2014). Accuracy assessment of global barotropic ocean tide models. *Reviews of Geophysics*, *52*(3), 243–282. https://doi.org/10.1002/2014rg000450

Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM journal on computing*, *1*(2), 146–160.

Teunissen, P. J. G. (2006). *Network quality control*. VSSD.

Wilkinson, K., von Zabern, M., & Scherzer, J. (2014). Global Freshwater Fluxes into the World Oceans.

Woodworth, P. L., Hughes, C. W., Bingham, R. J., & Gruber, T. (2013). Towards worldwide height system unification using ocean information. *J. Geod. Sci.*, *2*(4), 302–318. https://doi.org/10.2478/v10156-012-0004-8

Zijl, F., Laan, S., & Groenenboom, J. (2020). Development of a 3D model for the NW European Shelf (3D DCSM-FM).

3

AN EMPIRICAL NOISE MODEL FOR THE BENEFIT OF MODEL-BASED HYDRODYNAMIC LEVELING

This chapter has been published in Journal of Geodesy Volume 97, Number 1, December 2022, Pages 1 – 21, https://doi.org/10.1007/s00190-022-01694-x

3.1. INTRODUCTION

Model-based hydrodynamic leveling is an efficient and flexible method to connect islands and offshore tide gauges with the height system on land (D. C. Slobbe et al., 2018). The technique uses mean water level (MWL) differences between tide gauges provided by a regional, high-resolution hydrodynamic model. The averaging period can be chosen freely. For instance, D. C. Slobbe et al. (2018) obtained the best results when averaging the water levels over the summer months of their 19-year simulation period. Afrasteh et al. (2021) showed that combining model-based hydrodynamic leveling data with data of the Unified European Leveling Network (UELN) may improve the quality of the European Vertical Reference Frame (EVRF). Assuming the model-based MWLs can be obtained with a uniform variance of 4.5 cm^2 (corresponding to a standard deviation of 3 cm for each hydrodynamic leveling connection), the median of the propagated height standard deviations improved by 38% compared to the spirit leveling-only solution. If the model(s) provide the MWLs with a uniform variance of 12.5 cm^2 (corresponding to a standard deviation of 5 cm for each hydrodynamic leveling connection), the reported improvement is still 29%. Although promising, a more realistic impact assessment is yet to be done as a proof is lacking that hydrodynamic models can indeed provide the MWL at the tide gauge locations with an accuracy of a few centimeters. Apart from that, Afrasteh et al. (2021) assumed that the noise variance-covariance (VC) matrix of the model-derived MWLs is a diagonal matrix. The goal of this paper is to develop a more realistic correlated error model and to confirm whether previous noise assumptions were sensible.

Obtaining the *full* noise VC matrix is not straightforward. Indeed, this would be the case if model output includes full noise VC matrices or when an ensemble of model outputs is available or can be generated. However, the first is typically not the case, and the latter is not feasible given the size of the model in terms of grid nodes and the intended simulation period (in the order of decades). Moreover, in both cases proper noise models need to be available for all forcing datasets as well as the open boundary conditions. Therefore, inspired by the successful approach of Ditmar et al. (2011) and Farahani et al. (2016), among others, our approach will be to develop an *empirical noise model for the model-derived MWLs* based on the differences between observation- and model-derived MWLs.

A successful implementation places at least four requirements on the observed water level data. First, the distribution of the observation sites should result in a representative sample of the coastal MWL model errors. Second, the observed time series must be long compared to the averaging period so that multiple noise realizations can be calculated over the water level averaging period we aim for (i.e., one or more summer periods). Third, the sampling interval must be sufficiently high to average high-frequency water level variations. Finally, the water levels must be expressed in the same height system. The adjective 'coastal' in the first requirement has been added as we want to establish the hydrodynamic leveling connections between tide gauges located in coastal waters (i.e., the waters up to a few km from the coast). Given the higher variability of the physical processes that play a role in these waters (Iglesias et al., 2020; Kantha & Clayson, 2000), the model performance in the coastal waters is expected to be poorer than in the deep and shelf waters. The third requirement relates to the possible use of satellite radar altimeter data, which have a much lower temporal resolution than tide gauge records.

Considering the first three criteria, tide gauge records are an appropriate data source to

develop an empirical noise model. Although compared to other parts of the Earth there are many tide gauges available in our target area (the European sea waters in the broad sense and the north-east Atlantic region including the North Sea in the narrow sense), we have to make compromises when deriving a noise model. The main limiting factor is that most tide gauges are not deployed to build long and stable time series. Therefore, in addition to gaps, long time series often show abrupt jumps. Moreover, it is not always clear when and with what accuracy a tide gauge is connected to the mainland height system, nor with what frequency the connections are verified. Therefore, beyond the usual and pragmatic assumptions of time stationarity and spatial isotropy, we need to shorten the timespan over which we average the water levels.

The main objective of this study is to develop and analyze an empirical noise model for model-derived coastal summer mean water levels (SMWLs) and use that to obtain a more realistic quality impact of combining hydrodynamic leveling and UELN data in realizing the European Vertical Reference System (EVRS). The motivation to use the *summer* MWLs rather than the MWLs averaged over the entire simulation period follows from the results of D. C. Slobbe et al. (2018); by only averaging over the summer months we ignore storm surge periods and get more accurate MWL differences. Note that we consider any SMWL signal not captured by the hydrodynamic models to be part of the noise. The analysis of the obtained noise model includes an assessment of the spatiotemporal performance level with which state-of-the-art hydrodynamic models are able to represent the SMWL. In doing so, we consider three hydrodynamic models for the Northeast Atlantic Ocean, including the North Sea and Wadden Sea; the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (AMM7) (Tonani & Ascione, 2021), the DCSMv6-ZUNOv4 (Zijl et al., 2015; Zijl et al., 2013), and the 3D DCSM-FM (Zijl et al., 2020). To gain insight into the differences in model performance in the coastal waters versus the deep and shelf waters, we will develop empirical noise models for the latter two based on TOPEX/Jason satellite radar altimeter data and compare them with the one for the coastal waters. This analysis will be conducted for the DCSMv6-ZUNOv4 model only. We will also assess the contribution of errors in the vertical referencing of the tide gauges to the observation-derived SMWLs. Note that throughout the paper, the term 'SMWL' refers to the average water level calculated over all May to September months of one or more years. Moreover, when we refer to a noise model for, e.g., 3D DCSM-FM, we refer to a specific reanalysis conducted with this model.

Indeed, comparing observation- and model-derived (mean) water levels is routinely done in assessing model performance and presented in many publications (e.g., Filmer et al., 2018; Woodworth et al., 2013). These include our area of interest (e.g., Hermans et al., 2020; Holt et al., 2005; D. C. Slobbe et al., 2018; D. C. Slobbe et al., 2013; Zijl et al., 2013). The combination of the following two aspects makes the present analysis different. First, based on the results of D. C. Slobbe et al. (2018) we are interested in the model performance in terms of *'summer'* MWL. Second, our analysis goes beyond typical comparison studies as we will particularly assess the magnitude of the noise correlations.

The paper is organized as follows. Sect. 3.2 describes the methodology used to generate the noise models, the metrics used to assess the model performance, and the experiment conducted to obtain a more realistic quality impact of combining hydrodynamic leveling and UELN data in realizing the EVRS. Sect. 3.3 introduces the data sets used in this pa-

per. In Sect. 3.4, we present, analyze, and discuss all results. Finally, we conclude by summarizing the main findings of the paper in Sect. 3.5.

3.2. METHODOLOGY

3.2.1. NOISE MODEL GENERATION

PRELIMINARIES

A hydrodynamic leveling connection comprises the geopotential difference (ΔW) between two UELN height benchmarks (HBMs), each connected to a tide gauge. It can be written as

$$\Delta W^{\mathrm{HBM}_y}_{\mathrm{HBM}_x} = \Delta W^{\mathrm{TGZ}_x}_{\mathrm{HBM}_x} + \Delta W^{\overline{\mathrm{SMWL}}_x}_{\mathrm{TGZ}_x} + \Delta W^{\widetilde{\mathrm{SMWL}}_y}_{\widetilde{\mathrm{SMWL}}_x} + \Delta W^{\mathrm{TGZ}_y}_{\overline{\mathrm{SMWL}}_y} + \Delta W^{\mathrm{HBM}_y}_{\mathrm{TGZ}_y}, \tag{3.1}$$

where x and y refer to a tide gauge, TGZ stands for tide gauge zero, and $\overline{\mathrm{SMWL}}$ and $\widetilde{\mathrm{SMWL}}$ are the observation- and model-derived SMWLs, respectively. We usually refer to the HBM to which the TGZ is connected as the 'tide gauge benchmark' (TGBM). However, not all TGBMs are part of the UELN. If not, $\Delta W^{\mathrm{TGZ}_\cdot}_{\mathrm{HBM}_\cdot} = \Delta W^{\mathrm{TGZ}_\cdot}_{\mathrm{TGBM}_\cdot} + \Delta W^{\mathrm{TGBM}_\cdot}_{\mathrm{HBM}_\cdot}$. Note further that in Eq. (3.1) any time variability is ignored. As pointed out in Afrasteh et al. (2021) both the observation- and model-derived tide gauge records need to be reduced to the reference epoch adopted in the EVRF (epoch 2000.0). The same applies to the leveling data acquired to connect the HBM to the TGZ. To simplify the present discussion, we assume this reduction has been done. Another premise is that in establishing the hydrodynamic leveling connections, the timespan over which we compute both the observation- and model-derived SMWLs is the same for *all* tide gauges. By doing this, we reduce potential time-dependent, large-scale errors in the modeled water levels, and it greatly simplifies the problem as the noise model is just 1-dimensional.

The uncertainties associated with the five independent terms in Eq. (3.1) together determine the precision of the hydrodynamic leveling data. By far the largest contributor is, however, the middle term $\Delta W^{\widetilde{\mathrm{SMWL}}_y}_{\widetilde{\mathrm{SMWL}}_x}$, which is expected to be at the cm level. Indeed, supposing the connections between the HBMs and TGZs are established by spirit leveling conducted with a precision of 0.5 mm/km (corresponding to the precision of first-order leveling (Bossler, 1984)) and that we only use tide gauges within 10 km from the UELN height benchmarks, the contributions of the first and last term are below 2 mm in terms of standard deviation. Similarly, the uncertainty of the MWL computed over one month of sea level observations is already < 1 mm in length units based on a 10-minute sampling and assuming white noise with a standard deviation of 5 cm.

Considering now a set of hydrodynamic leveling connections, the full noise VC matrix Q_{hl} is obtained by

$$Q_{\mathrm{hl}} = Q_1 + Q_2 + A_{\mathrm{hl}} Q_{\widetilde{\mathrm{SMWL}}} A^{\mathrm{T}}_{\mathrm{hl}} + Q_4 + Q_5, \tag{3.2}$$

where Q_1, Q_2, Q_4, and Q_5 are the *diagonal* noise VC matrices associated with $\Delta W^{\mathrm{TGZ}_x}_{\mathrm{HBM}_x}$, $\Delta W^{\overline{\mathrm{SMWL}}_x}_{\mathrm{TGZ}_x}$, $\Delta W^{\mathrm{TGZ}_y}_{\overline{\mathrm{SMWL}}_y}$, and $\Delta W^{\mathrm{HBM}_y}_{\mathrm{TGZ}_y}$, respectively; A_{hl} is the design matrix of the hydrodynamic leveling dataset; and $Q_{\widetilde{\mathrm{SMWL}}}$ the *full* noise VC matrix of the model-derived SMWLs. Estimating $Q_{\widetilde{\mathrm{SMWL}}}$ is one of the main goals of this paper and a critical step in the development of model-based hydrodynamic leveling.

OVERALL APPROACH TO GENERATE A NOISE MODEL OF THE MODEL-DERIVED SMWLS IN COASTAL WATERS

As motivated in Sect. 3.1, $Q_{\widetilde{\text{SMWL}}}$ will be computed from an empirical noise model derived from the differences between observation- and model-based SMWLs at a set of N coastal tide gauges within the model domain. The criteria applied to select the tide gauges are discussed in Sect. 3.3.2. The overall approach to compute the noise model from these data comprises four steps:

Step 1: Compute the T_{avg}-SMWL time series. We refer to the SMWL averaged over interval T_{avg} as the T_{avg}-SMWL. In dealing with tide gauge data, T_{avg} is one, two, or three summer periods. Given the time span of 22 years over which model-derived water levels are available (T_{sim}), a larger number of summer periods for T_{avg} will not provide sufficient realizations of the noise per tide gauge. The T_{avg}-SMWL time series are computed by averaging the monthly MWLs over all May to September months in each consecutive time interval of T_{avg} years. To increase the data availability we permit for one missing monthly MWL per 'summer' period. The model-derived monthly MWLs are in all cases calculated as the arithmetic mean of the modeled water levels. To account for 'small' data gaps in the observed water level time series (total time for which no data are available should be < 10 days per month), the observation-derived monthly MWLs are estimated along with the tides. For this we used the UTide Matlab functions (Codiga, 2020). The automated decision tree method, which is based on the equilibrium tide and the conventional Rayleigh criterion, has been applied to select the constituents to be included in the harmonic analysis (Codiga, 2011).

Step 2: Generate noise realizations. The next step in our approach is to generate a set of noise realizations. For each tide gauge, they are obtained by taking the difference between the observation- and model-derived T_{avg}-SMWL time series of step 1. If the simulation period is T_{sim}, the number of differences assuming no data gaps equals $\lfloor T_{\text{sim}}/T_{\text{avg}} \rfloor$. We consider these differences as realizations of the noise in the model-derived T_{avg}-SMWL. In case no difference could be computed due to data gaps in the observation-derived SMWL time series, the difference is set to `NaN`. Note that all monthly MWL differences are excluded from the analysis if the difference between the observation- and model-derived monthly MWLs exceeds the median plus/minus three times the standard deviation (estimated as $1.4826 \times$ the median absolute deviation (MAD) (Cook & Weisberg, 1982; Rousseeuw & Croux, 1993)). About 1% of the months were excluded this way. We also removed for each T_{avg} period the mean difference over all tide gauges. Based on a bootstrapping approach (Mooney & Duval, 1993), we then generate K sets of noise realizations at all N tide gauges.

Step 3: Compute empirical covariance functions. For each set, consisting of N differences, we compute an empirical covariance function $\hat{C}$ (e.g., Wackernagel, 2003) using:

$$\hat{C}(D_{\text{sea}}) = \frac{1}{M(D_{\text{sea}})} \sum_{M(D_{\text{sea}})} \left(Z(s_i) - \overline{Z}\right)\left(Z(s_j) - \overline{Z}\right), \tag{3.3}$$

where D_{sea} is the sea distance, $M(D_{\text{sea}})$ the set of tide gauges separated by sea distances within a given interval (see Sect. 3.2.1 for the way these intervals are defined), $Z(s)$ is the difference between the observation- and model-derived SMWLs at location s, and $\overline{Z}$ the average difference. The sea distance is, loosely spoken, the shortest distance to travel between two points over sea (see Sect. 3.2.1 for the method applied to compute these distances). It

is used as the distance metric to account for the presence of land. In computing $\hat{C}(D_{\text{sea}})$ it is assumed that the covariance function is isotropic.

Step 4: Fit an analytical noise model. The last step involves the fitting of an analytical model through the *average* empirical covariances. This is the subject of Sect. 3.2.1. Note that in computing the average empirical covariance function, the averaging is applied to the empirical correlograms obtained as $\hat{C}(D_{\text{sea}})/\hat{C}(0)$. To obtain the *average* empirical covariance function, we scale the average empirical correlogram by the variance computed as squared MAD of the differences between the observation- and model-derived SMWLs at all tide gauges scaled by 1.4826. To build $Q_{\widetilde{\text{SMWL}}}$, the analytical model has to be evaluated for the sea distances among the tide gauges involved in establishing the hydrodynamic leveling connections.

OVERALL APPROACH TO GENERATE A NOISE MODEL OF THE MODEL-DERIVED SMWLS IN DEEP AND SHELF WATERS

The noise models for the deep and shelf waters will be calculated from satellite radar altimeter data (see Sect. 3.3.3 for a description of the data). The limited temporal resolution (about 10 days) requires a different strategy than described in Sect. 3.2.1. In short, the period over which a reliable *average* water level can be calculated is much longer. In this study, we used the full data period of 1997–2019. The following procedure is applied using the instantaneous water levels as input.

Step 1: Compute difference time series. We interpolated the modeled instantaneous water levels at the altimeter data locations (see Fig. 3.1) and subtracted them from the observed ones. After that, we binned the differences resulting in one difference time series per 6.5 km in along-track direction. In the binning, we fitted a function through the data accounting for a slope in latitude and longitude directions as well as a linear trend over time and an annual and semi-annual cycle. To reduce the observations to the mean position of all data points in a bin, we used the estimated slope parameters.

Step 2: Compute the average difference over all May to September months and edit the dataset. From each time series, we selected the data points acquired in the months May to September and computed the average. All points for which the difference exceeds the median of all differences plus five times the standard deviation are excluded as an 'outlier'. Here again the standard deviation is estimated as $1.4826\times$ the MAD.

Step 3: Compute the 'sea distance matrix'. The sea distance matrix describes the sea distances among all bins. See Sect. 3.2.1 for further details.

Step 4: Split the dataset into deep and shelf waters. The separation between the deep and shelf waters is determined based on the 200 m depth contour line shown in Fig. 3.1.

Step 5: Compute the empirical covariance function. After determining the 'optimal' lag distances using the method described in Sect. 3.2.1, we compute the empirical covariance functions using Eq. (3.3).

COMPUTING THE SEA DISTANCE

To compute the sea distances among all observation locations (i.e., tide gauges or altimeter data points), we discretized the region using a uniform grid with a spacing of 0.05 degrees. Using the grid nodes classified as 'sea' based on GMT's `grdlandmask` (Wessel et al., 2019), we form a graph of which the vertices are connected in east-west and north-south

directions. Using Matlab's `graphshortestpath` function, we calculate the Dijkstra shortest path (Dijkstra, 1959) between nodes closest to the observation locations. Adding up the distance between the observation locations and the nearest vertices to the shortest path will end up in the sea distance. Note that we did not establish diagonal connections between vertices as all hydrodynamic models being considered in this study use rectangular grids.

DEFINING THE LAG DISTANCES

Usually, lags (also referred to as 'distance bins/classes') for which the empirical covariances are calculated have a uniform spacing. The maximum lag distance is half the distance between the two data points furthest from each other, and the lag tolerance (i.e., half the interval within which a distance must be to belong to a particular lag) is half the lag spacing. Given that in particular the number of tide gauges is limited, we use a non-uniform spacing and lag tolerance in which the lags are chosen so that approximately the same number of pairs are available for each lag distance. The maximum lag distance is set to 2100 km, which is approximately 65% of the sea distance between the two tide gauges furthest from each other. In using a non-uniform spacing and lag tolerance, we ensure that we have provided sufficient data points for computing the covariance value for each lag. The only exception to this is lag zero, for which the lag tolerance is set to zero. The latter means that the number of points available for this lag equals the number of tide gauges/altimeter data points.

THE ANALYTICAL COVARIANCE MODEL

Many models of covariance functions have been proposed. For some examples, we refer to e.g., Christakos (2012) and Hristopulos (2020). In this study, we use a composite model that is the superposition of the 'nugget effect' and I J-Bessel models (see Hristopulos, 2020, Sect. 4.2.2). The nugget effect describes a possible discontinuity at the origin. If present, it points to random errors in the observation- and model-derived SMWLs and signal at short spatial scales that cannot be resolved from the data or by the hydrodynamic model. The J-Bessel model is chosen because it admits negative values and provides a good fit of the empirical covariance functions. The composite model is defined as:

$$C(h) = \begin{cases} c_0 + \sigma_x^2 & \text{if } h = 0 \\ \frac{\sigma_x^2}{I} \sum_{i=1}^{I} 2^{\nu}\, \Gamma(\nu+1)\, h^{-\nu}\, J_{\nu}(h), \text{with } \nu \geq d/2 - 1 & \text{if } 0 < h \leq h_{\max}, \end{cases} \tag{3.4}$$

where h is the dimensionless lag defined as D_{sea}/ξ_i, ξ_i the characteristic length, c_0 is the nugget effect (i.e., the magnitude of the discontinuity at the origin, σ_x^2 is the variance by which the J-Bessel models are scaled, Γ denotes the Gamma function (Hristopulos, 2020, Eq. 4.17), $J_{\nu}(\cdot)$ is the Bessel function of the first kind of order ν (Watson, 1995), d denotes the number of spatial dimensions ($d = 2$ in our case), and $h_{\max}$ is the dimensionless maximum lag (i.e., 2100 km$/\xi_i$). Note that in this study the maximum value for I is equal to two. The parameters c_0, ξ_i, ν, and σ_x^2 are estimated by applying Matlab's global optimization solver `GlobalSearch`. The algorithm minimizes the sum of the squared differences between the empirical covariance function values and the analytical model such that $1E-4 \leq c_0 \leq 1E-2$, $1E2 \leq \xi_1 \leq 2E2$, $2E2 \leq \xi_2 \leq 5E2$, $0 \leq \nu \leq .25$, $1E-5 \leq \sigma_x^2 \leq 1E-3$, and $c_0 + \sigma_x^2$ equals the empirical covariance at $h = 0$. In case $I = 1$, we use $1E2 \leq \xi_1 \leq 5E2$. Moreover, we apply the constraint that the resulting noise

VC matrix (obtained by evaluating the model at the sea distances among the involved tide gauges) should be positive definite. In doing so, we set the objective function to infinite in case the minimum eigenvalue is smaller than machine epsilon (i.e., $2.2204E-16$). The last constraint is a technical solution to account for the fact that our distance metric is non-Euclidean. Without this constraint there would be no guarantee that the fitted model provides a positive definite noise VC matrix (Hristopulos, 2020, pp. 114).

3.2.2. METRICS USED TO ASSESS THE SPATIOTEMPORAL MODEL PERFORMANCE

Many metrics are available to assess a model's spatial and temporal performance (e.g., Bennett et al., 2013). Here, the temporal model performance is assessed for the subregions shown in Fig. 3.1 by the cumulative distribution functions of both the mean absolute error (MAE) and the Kling–Gupta efficiency (KGE) metric (Gupta et al., 2009) computed for all tide gauges within the sub-area. The spatial performance is assessed using both the MAE and the spatial efficiency (SPAEF) metric (Koch et al., 2018).

Part of the differences between observation- and model-derived SMWLs are explained by time-dependent but space-independent model biases (e.g., Jahanmard et al., 2021). Indeed, for the application at hand these errors do not matter as we use the differences in model-derived SMWLs between two tide gauges. To eliminate the contribution of these errors to the model performance, we apply a time-dependent bias correction calculated from a subset of tide gauges spanning the model domain for which a full time series is available (see Fig. 3.1). For the same reason, we use slightly modified formulations for the KGE and SPAEF metrics. The modified formulation for the Kling–Gupta efficiency metric, referred to as mKGE, is

$$\mathrm{mKGE} = 1 - \sqrt{(\alpha_Q - 1)^2 + (\beta_Q - 1)^2 + \gamma_Q^2}, \tag{3.5}$$

where α_Q is the Pearson correlation coefficient between the observation- and model-derived SMWL time series (only used in case it is significantly different from zero at the 0.05 significance level), β_Q is the relative variability based on the ratio of standard deviation in model-derived and observation-derived SMWL values and γ_Q is the difference between the averages of the model-derived and observation-derived SMWL time series normalized by the standard deviation of the observation-derived SMWL data. In the original formulation, γ_Q is the ratio of the averages. The modified spatial efficiency metric, referred to as mSPAEF, is

$$\mathrm{mSPAEF} = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{3.6}$$

where α is the Pearson correlation coefficient between the observation- and model-derived SMWL pattern, β is the ratio of standard deviation in model-derived and observation- derived SMWL values, and γ is the histogram intersection for the histogram of the observation-derived pattern and the histogram of the model-derived pattern. In the original formulation, β is the ratio of the standard deviation over the mean. To suppress the impact of outliers on the computed standard deviations we estimate these as $1.4826\times$ the MAD. For both the mKGE and mSPAEF metrics, the ideal value equals 1, while a value far below 0 indicates the model lacks any performance.

3.2.3. RE-ASSESSING THE EXPECTED QUALITY IMPACT OF COMBINING HYDRODYNAMIC LEVELING AND UELN DATA IN REALIZING THE EVRS

To obtain a more realistic expectation of the quality impact of combining hydrodynamic leveling and UELN data in realizing the EVRS, we conducted an experiment that has a similar setup as Experiments I and II of Afrasteh et al. (2021). The experiment assumes four hydrodynamic models are available, each covering part of the European sea waters. The four domains comprise the Mediterranean Sea, the Black Sea, the Baltic Sea, and the north-east Atlantic region including the North Sea. Connections between tide gauges are established only in case both are located in the same sea basin. It is further assumed that each of the four hydrodynamic models describes the coastal SMWLs with a precision consistent with the obtained noise model. That is, we assume our noise model derived for a hydrodynamic model covering the north-east Atlantic region including the North Sea also applies to the other basins. This assumption is considered to be reasonable in case i) the models have comparable resolutions, ii) the underlying bathymetries have similar quality, and iii) the models are forced using the same datasets. Note that in case the sea distance between two tide gauges is larger than the maximum lag distance for which we computed the empirical covariance functions, the covariance is set to zero. The full noise VC matrix $Q_{\widetilde{\text{SMWL}}}$ obtained this way is used to compute Q_{hl} (Eq. (3.2)). Here, we ignored the contributions of Q_1, Q_2, Q_4, and Q_5 (see Sect. 3.2.1). Using the heuristic search method described in Afrasteh et al. (2021, Sect. 2.3), we identified the set of hydrodynamic leveling connections that provide the lowest median of the propagated height standard deviations. The benchmark solution is the spirit-leveling only solution (see Afrasteh et al. (2021, Sect. 5.1.1)). To assess the impact of including noise correlations, we also repeat the experiment by setting the off-diagonal elements in $Q_{\widetilde{\text{SMWL}}}$ to zero.

3.3. DATA

3.3.1. MODEL-DERIVED WATER LEVEL TIME SERIES

In this study, three reanalysis products generated by three different hydrodynamic models have been used; the Atlantic – European North West Shelf – Ocean Physics Reanalysis (Tonani & Ascione, 2021), a reanalysis obtained using the coupled two-domain DCSMv6-ZUNOv4 model (D. C. Slobbe et al., 2018; Zijl et al., 2015; Zijl et al., 2013), and one generated using the 3D Dutch Continental Shelf Model – Flexible Mesh (3D DCSM-FM) (Zijl et al., 2020).

The first reanalysis is produced using the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (FOAM AMM7) which uses version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO) ocean model code (Madec et al., 2017) and the 3DVar NEMOVar system to assimilate observations (Mogensen et al., 2012). The data are publicly available at the Copernicus Marine Service (https://doi.org/10.48670/moi-00059). We refer to this dataset as the 'AMM7' dataset. Note that in this dataset all the grid points east of 10°E were masked (Renshaw et al., 2021).

The DCSMv6-ZUNOv4 dataset is obtained using the Dutch Continental Shelf Model version 6 (DCSMv6) and the Zuidelijk Noordzee model version 4 (ZUNOv4) in the outer and inner domain (Fig. 1 in D. C. Slobbe et al., 2018), respectively. The DCSMv6-ZUNOv4

is a 2D model; the baroclinic forcing has been added using the method described by D. C. Slobbe et al. (2013). This model has also been used by D. C. Slobbe et al. (2018). In this study, we redid the simulation to incorporate the updated salinity and temperature data from the earlier mentioned AMM7 dataset. Since the salinity and temperature data east of 10°E were masked in this dataset, we applied the same masking to the DCSMv6-ZUNOv4 derived water levels. The WAQUA software package, on which the DCSMv6-ZUNOv4 model is based, provides the opportunity to generate output at user-defined locations and epochs. In this study, we used this option to compute water levels at locations/epochs satellite radar altimeter data are available.

The 3D DCSM-FM reanalysis is obtained using the Delft3D Flexible Mesh software framework that allows for the use of unstructured grids. For this study we have used software version 2.17.05.72090. Note that the model is still under development. One known issue in the model is an apparent strong vertical circulation between the bottom and the pycnocline in the deep ocean originating from instabilities close to the open boundaries. This results in a less accurate representation of the MWL in deep ocean waters.

The three models/reanalyses differ in many ways. A detailed discussion of this goes beyond the scope of this paper. We refer to Table 3.1 for a summary of the main characteristics of the models as well as a brief overview of the setup applied when generating the reanalyses. Further details can be found in the cited references. For all products, we used/generated the water levels from 1997–2019.

3.3.2. TIDE GAUGE RECORDS

We collected about 200 high-resolution, coastal water level time series for the period 1997–2019. The time series were acquired by different national authorities in the countries Belgium, Denmark, France, Germany, Ireland, the Netherlands, Norway, Sweden, and Great Britain. All time series have been visually inspected for outliers and use the national height datum as the vertical reference for the water levels. For any meta-information regarding the national height datums we refer to http://crs.bkg.bund.de/crseu/crs/eu-national.php.

A tide gauge record is used to generate a noise model for a particular hydrodynamic model if: i) the tide gauge is located inside the model domain, ii) the tide gauge is outside the tidal flat areas, and iii) the tide gauge is east from 10° longitude (only in case of the AMM7 and DCSMv6-ZUNOv4 models). After excluding the tide gauge stations that do not meet these criteria, we applied a data editing step in which we excluded all records for which the median difference between the observation- and model-derived monthly MWL time series exceeded the median of the medians plus/minus three times the standard deviation (estimated as mentioned before). Fig. 3.1 shows the tide gauges available to generate a noise model per hydrodynamic model. For the AMM7, DCSMv6-ZUNOv4, and the 3D DCSM-FM models, there are 123, 150, and, 171 stations available, respectively. As can be inferred from Fig. 3.2, these numbers refer to the total number of tide gauges for which we have at least one summer of water levels available.

The most important preprocessing step for the tide gauge data is to unify the height datums. In this study, we will refer all observed water levels to the EGG2015 quasi-geoid model (Denker, H., 2015) in the mean-tide system (ζ_{EGG2015}). These water levels, referred to

Table 3.1: Summary of the main characteristics of the hydrodynamic models used in this study as well as a brief of the setup applied to generate the reanalyses. The abbreviations 'T', 'S', and 'SSH' denote temperature, salinity, and sea surface height, respectively.

	AMM7	DCSMv6-ZUNOv4	3D DCSM-FM
Model dimensions	3D	2D*[1]	3D
Domain	20°W–13°E 40°S–65°N	15°W–13°E 43°S–64°N	15°W–13°E 43°S–64°N
Horizontal grid	1/9°long × 1/15°lat	DCSMv6: 1/40°long × 1/60°lat ZUNOv4 (variable grid): 200–2000 m	variable grid 900–8000 m
Vertical grid	Hybrid S-σ-z, 51 layers (output interpolated to 24 geopotential levels)	N/A	z-σ, 50 layers
Forcing			
internal tides	included	included	included
atmospheric	ERA5	ERA-Interim	ERA5
Open boundary cond.			
	tides (15 constituents) T, S, SSH, barotropic u and v-velocities from reanalyses, river discharge daily time series	tides (26 constituents) surge (dynamic atmosph. corrections) baroclinic (AMM7)	tides (30 constituents) surge (inverse barometer correction) steric effect salinity and temperature monthly river discharge
Spin-up time	2 years	1 month	1 year
Data assimilation	yes (surface T and vertical profiles of T and S	no	no

*[1] Depth-averaged baroclinic pressure gradients added, computed from 4D salinity and temperature fields from the AMM7 model.

as χ_{EGG2015} are obtained as:

$$\begin{cases} \chi_{\text{EGG2015}} = \chi_{\text{NHD}} + (h^{\text{HBM}}_{\text{GNSS}} - H^{\text{HBM}}_{\text{NHD}}) - \zeta_{\text{EGG2015}} & \text{if GNSS is available} \\ \chi_{\text{EGG2015}} = \chi_{\text{NHD}} + h^{\text{HRS}} - \zeta_{\text{EGG2015}} & \text{otherwise,} \end{cases} \tag{3.7}$$

where χ_{NHD} is the water level expressed relative to the national height datum (NHD), $h^{\text{HBM}}_{\text{GNSS}}$ is the ellipsoidal height of a nearby HBM obtained using GNSS, $H^{\text{HBM}}_{\text{NHD}}$ is the physical height (e.g., orthometric or normal height) of the HBM, and h^{HRS} is the ellipsoidal height of the *national* height reference surface (HRS). Where applicable, all heights have been transformed to the mean-tide system. For the GNSS data, we used Petit and Luzum (2011, Eq. 7.14a). For the physical heights and EGG2015 height anomalies we used the equations provided by Mäkinen and Ihde (2008). Table 3.2 shows per country the information about the used HRS and the source of the GNSS data. Regarding the GNSS data, we always used the most recent GNSS solution available. To get an idea about the uncertainty of the second method (i.e., second row Eq. (3.7)), we will analyze in Sect. 3.4.3 the differences between $(h^{\text{HBM}}_{\text{GNSS}} - H^{\text{HBM}}_{\text{NHD}})$ and h^{HRS}.

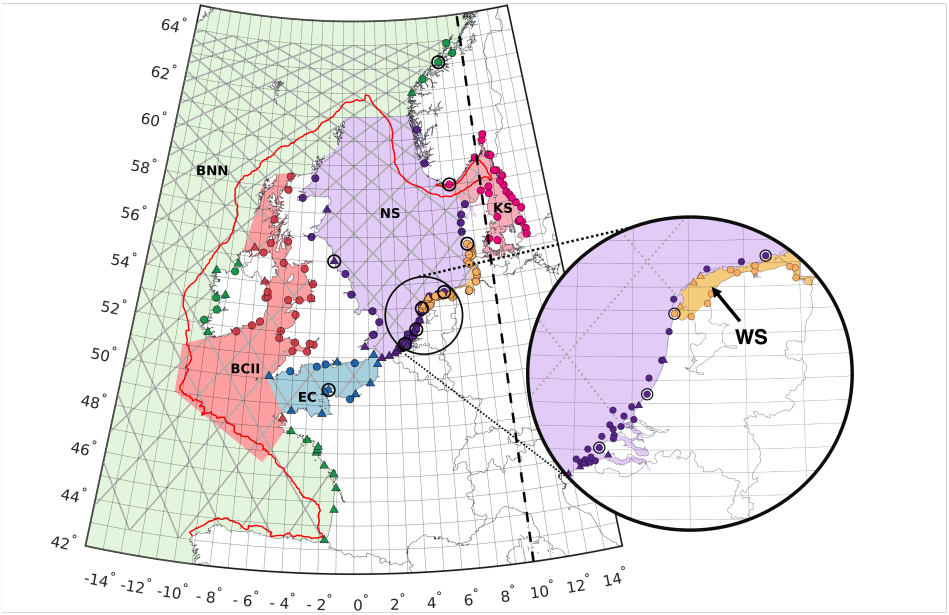


Figure 3.1: Map showing the tide gauge locations, altimeter data (grey dots), and the subregions used when assessing the temporal model performance. Here, BNN covers the Bay of Biscay, the North Atlantic Ocean and the Norwegian Sea; BCII covers the Bristol Channel, Celtic Sea, Irish Sea and St. George's Channel, and the Inner Seas off the West Coast of Scotland; EC stands for English Channel; KS is the Kattegat-Skagerrak Seas: NS stands for North Sea; and WS for Wadden Sea. Tide gauges plotted with a triangle symbol are measured using GNSS. The ones with a black circle around show the tide gauges used to correct the bias in the model-derived SMWL (See Sect. 3.2.2). All tide gauges east from the dashed black line are only available for the 3D DCSM-FM model. The red line is the 200 m depth contour separating the model domain in the deep and shelf waters (Sect. 3.2.1).

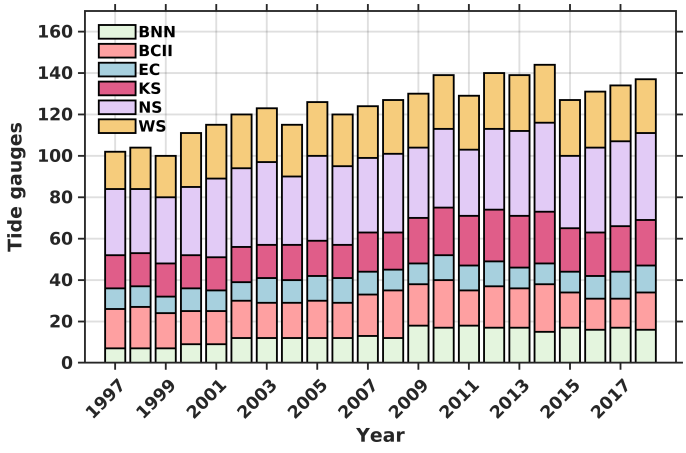


Figure 3.2: Number of available tide gauges per sub-area over time.

Table 3.2: Height reference surface and source of GNSS data per country.

Country	HRS	GNSS
Norway	HREF2018B (Lysaker & Vestøl, 2020)	KV
Sweden	SWEN17_RH2000 (Ågren et al., 2018)	—
Denmark	DKGEOID13B	—
Germany	GCG2016 Schwabe et al., 2016	—
Netherlands	NLGEO2018 (D. C. Slobbe et al., 2019)	RWS
Belgium	hBG18 (C. Slobbe et al., 2018)	NGI
France	RAF20 (IGN, 2021)	SHOM
Great Britain	OSGB36 (Ordnance Survey, 2015)	SONEL
Ireland	OSGB36 (Ordnance Survey, 2015)	MIFM

KV:Kartverket
RWS:Rijkswaterstaat
NGI:Nationaal Geografisch Instituut
SHOM: Service hydrographique et océanographique de la Marine
SONEL: Système d'Observation du Niveau des Eaux Littorales
MIFM: Marine Institute Foras na Mara

3.3.3. SATELLITE RADAR ALTIMETRY TIME SERIES

Satellite radar altimetry data, acquired by the TOPEX/Poseidon and Jason satellite altimeters, were obtained through the Radar Altimeter Database System (RADS, http://rads.tudelft.nl/rads/rads.shtml). All data from 1997–2019 were combined (except the data from the interleaved orbits). Note that RADS hardly contains data in the first five km from the coast. The following geophysical and range corrections were applied (Scharroo et al., 2016): ionosphere (smoothed dual-frequency altimeter observations), dry troposphere (ECMWF), wet troposphere (radiometer, ECMWF), solid tide (Cartwright & Edden, 1973; Cartwright, D. E. and Taylor, R. J., 1971), pole tide (Wahr, 1985), load tide (FES2014), and sea state bias (CLS). We also applied the reference frame offset and 'slope correction' discussed by Sandwell and Smith (2014). The latter is negligibly small, except over the shelf edge where it reaches a few centimeters. Finally, we transformed the data to the GRS80 ellipsoid and subtracted the EGG2015 quasi-geoid model in the mean-tide system. All observations for which the differences compared to the DCSMv6-ZUNOv4 derived water levels exceeds the median plus/minus five times the standard deviation (estimated as mentioned before) are flagged as outliers.

3.3.4. SPIRIT LEVELING DATA

The spirit leveling data used to re-assess the expected quality impact of combining hydrodynamic leveling and UELN data in realizing the EVRS are identical to the data used by Afrasteh et al. (2021). The data include i) the locations of all UELN height markers, ii) a list describing which height markers are connected (except for the countries Ukraine, Russia, and Belarus), iii) the a-priori variances of the geopotential differences for the available connections, and iv) the variances obtained by variance component estimation (except for Great Britain, Ukraine, Russia, and Belarus). As described by Afrasteh et al. (2021), we reconstructed the missing leveling connections in Ukraine, Russia, and Belarus. In all ex-

periments conducted in this study, we used the variances that the BKG obtained by variance component estimation. For Great Britain, the a-priori variances were used.

3.4. RESULTS AND DISCUSSION

3.4.1. NOISE MODELS FOR THE MODEL-DERIVED COASTAL SMWL

The average empirical noise covariance functions for the model-derived coastal MWLs computed over one summer period (i.e., the 1-SMWLs ($T_{\text{avg}} = 1$)) are shown in Fig. 3.3. Each of the empirical covariance functions is computed based on 10,000 bootstrap ensembles and for 31 lag distances, ranging between sea distances of 0 and 2100 km. Note that the maximum sea distance between two tide gauges is 3300 km. As also shown in Fig. 3.3, the number of pairs available per lag is except for the first and last lags typically larger than 210 (3D DCSM-FM), 165 (DCSMv6-ZUNOv4), and 100 (AMM7). For the 3D DCSM-FM model, we also computed the empirical noise covariance functions associated to the 2-SMWLs and 3-SMWLs (see Fig. 3.4).

The empirical noise covariance functions show the following:

- The model noise is spatially correlated. All functions show positive covariances for sea distances up to 250 km. For larger distances they fluctuate around zero for the 3D DCSM-FM. For the DCSMv6-ZUNOv4 and AMM7 models the covariances for larger distances are mostly negative.
- For all models there is a relatively large discontinuity in the empirical covariance function at zero sea distance (i.e., nugget effect). The estimated nugget effect is lowest for the 3D DCSM-FM (12.1 cm^2) and highest for the DCSMv6-ZUNOv4 (16.3 cm^2). The variance (i.e., covariance at zero sea distance) for these two models is 15.3 cm^2 and 21.7 cm^2, respectively. For the AMM7 model, the estimated nugget effect and variance are 15.3 cm^2 and 21.5 cm^2, respectively.
- Noise covariances have higher magnitudes for the AMM7 and DCSMv6-ZUNOv4 models compared to the 3D DCSM-FM model. Moreover, the first zero-crossing occurs for larger sea distances.
- Averaging the MWLs over two or three summer periods (i.e., the 2-SMWLs and 3-SMWLs), hardly changes the shape of the noise covariance functions. The largest change is in the variance, which decreases from 15.3 cm^2 to 14.1 cm^2 and 12.7 cm^2, and the estimated nugget effect that decreases from 12.1 cm^2 to 11.0 cm^2 and 10.0 cm^2, respectively.

Note that contrary to the empirical covariance functions for the DCSMv6-ZUNOv4 and AMM7 models, the one for the 3D DCSM-FM model is also based on tide gauges in the Kattegat-Skagerrak. A detailed analysis (not shown in this paper) indicates that the exclusion of the Kattegat-Skagerrak mainly affects the empirical covariances for sea distances > 1000 km.

Also, note that the computed error bars are likely too small. Indeed, in our approach any time correlation of the 'noise' is ignored. Since we also consider SMWL signals not captured by the model as noise, any such time-varying signal will result in noise realizations that are not independent. This, in turn, results in over-optimistic error bars.

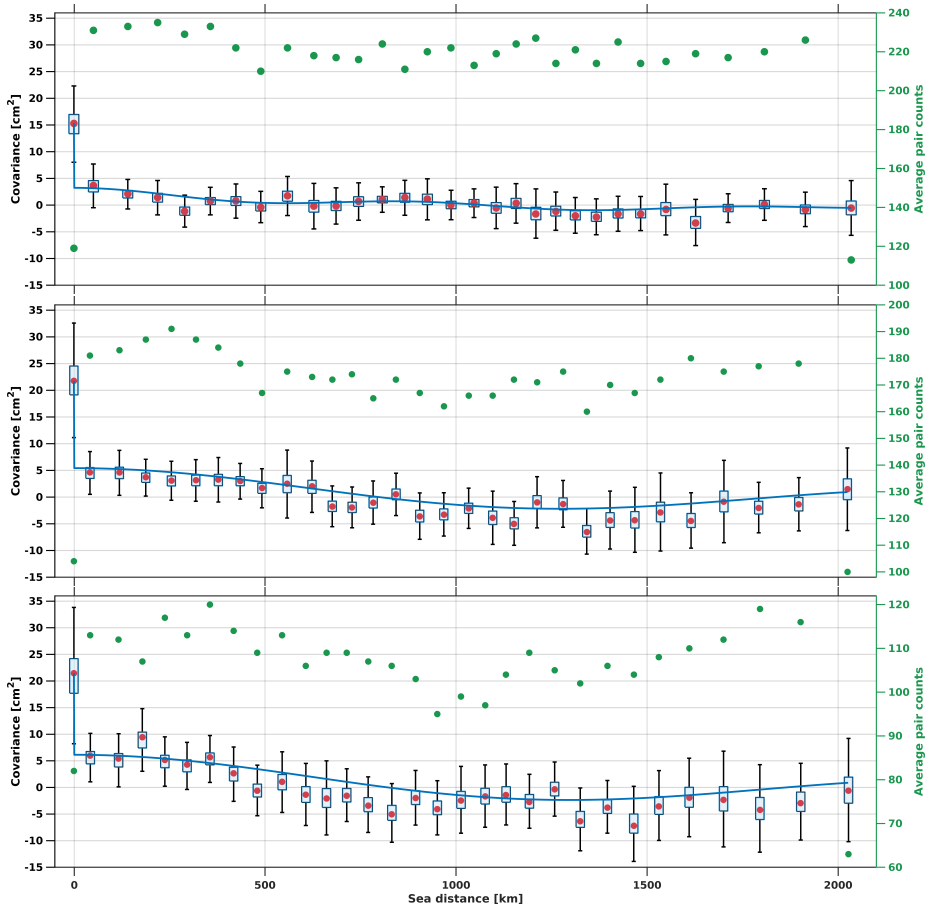


Figure 3.3: The empirical noise covariance functions for the 3D DCSM-FM (top panel), DCSMv6-ZUNOv4 (middle panel) and AMM7 (bottom panel) models. The red circles indicate the ensemble means. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme ensemble values not considered outliers, i.e., 99.3 percent coverage. The blue lines show the fitted analytical models presented in Eq. (3.4) and used in Sect. 3.4.5 (for 3D DCSM-FM only). Note that for 3D DCSM-FM we included two J-Bessel models in the composite model, while for the other hydrodynamic models we included one. The green dots for each plot indicates the number of pairs available per lag.

The empirical covariance functions provide, albeit limited, insight into the performance of the different models in representing the coastal 1-SMWL. At the same time, we should again realize that a contribution has been introduced by the uncertainty in the vertical referencing of the tide gauges and in the applied height datum unification. This uncertainty will partly contribute to the observed nugget effect (see further Sect. 3.4.3), but also partly explains the observed spatial correlations (quasi-geoid errors are spatially correlated). However, this uncertainty is *not* expected to explain the total nugget effect. As noted by Hristopulos (2020), the nugget effect describes: i) independent measurement errors that are due to the measurement process; ii) purely random fluctuations endogenous to the system, and iii)

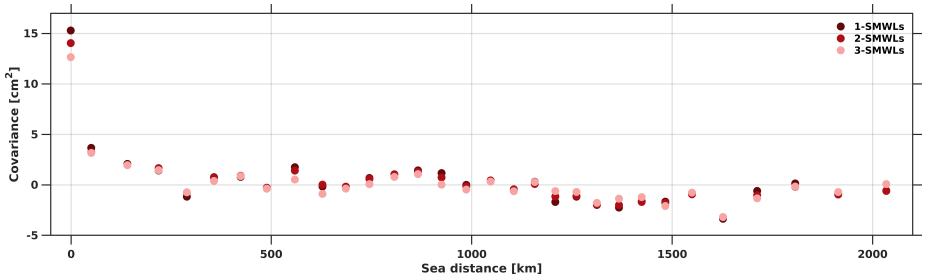


Figure 3.4: The averaged empirical noise covariance functions associated to the 3D DCSM-FM derived SMWLs computed over one, two or three summers.

sub-resolution variability, i.e., fluctuations with characteristic length scales that are below the resolution limit of the observations. In our case, the latter also points to the fact that hydrodynamic models are not always able to resolve the local SMWL variability at specific tide gauge locations. This is not surprising when we consider that many tide gauges are located in harbors, estuaries or other locations with complex bathymetry, or near flood defense structures. Solving the local hydrodynamics requires ultra-high resolutions, not only of the model, but also of the required forcing data. Moreover, where rivers flow into the sea, the models must be extended far into the rivers.

The comparison between the three models reflects the advances in modeling. Although the 3D DCSM-FM model should be considered as a preliminary model (see Sect. 3.3.1), we see significant improvements over the other two models. The variance and nugget effect are lower, implying that the model is better able to represent local processes. Moreover, the spatial error covariances are also lower. In comparison to the AMM7 model, the first could have been expected given the much higher resolution of the 3D DCSM-FM. With regard to the DCSMv6-ZUNOv4 model, the improved performance can partly be attributed to the resolution (for waters outside the North Sea and Wadden Sea), the fact that DCSMv6-ZUNOv4 is a 2D model (see Sect. 3.3.1) and differences in forcing data (see Table 3.1). In Sect. 3.4.2, we present the results of the spatiotemporal performance assessment to further analyze these differences.

3.4.2. SPATIOTEMPORAL MODEL PERFORMANCE IN REPRESENTING THE COASTAL 1-SMWL

For the analyses presented in this section, we used 120 tide gauges that are included in all models. The only exception is the temporal model performance assessment conducted for the KS region; the tide gauges being used there are only available for 3D DCSM-FM (see Sect. 3.3.1 and Fig. 3.1).

The metrics aimed to assess the temporal model performance are summarized in Fig. 3.5. Each panel shows the cumulative distribution functions for both the MAE and mKGE metrics for all tide gauges per subregion (see Fig. 3.1). Fig. 3.6 summarizes the assessment of the spatial model performance using both the MAE and mSPAEF metrics. The figures show the following:

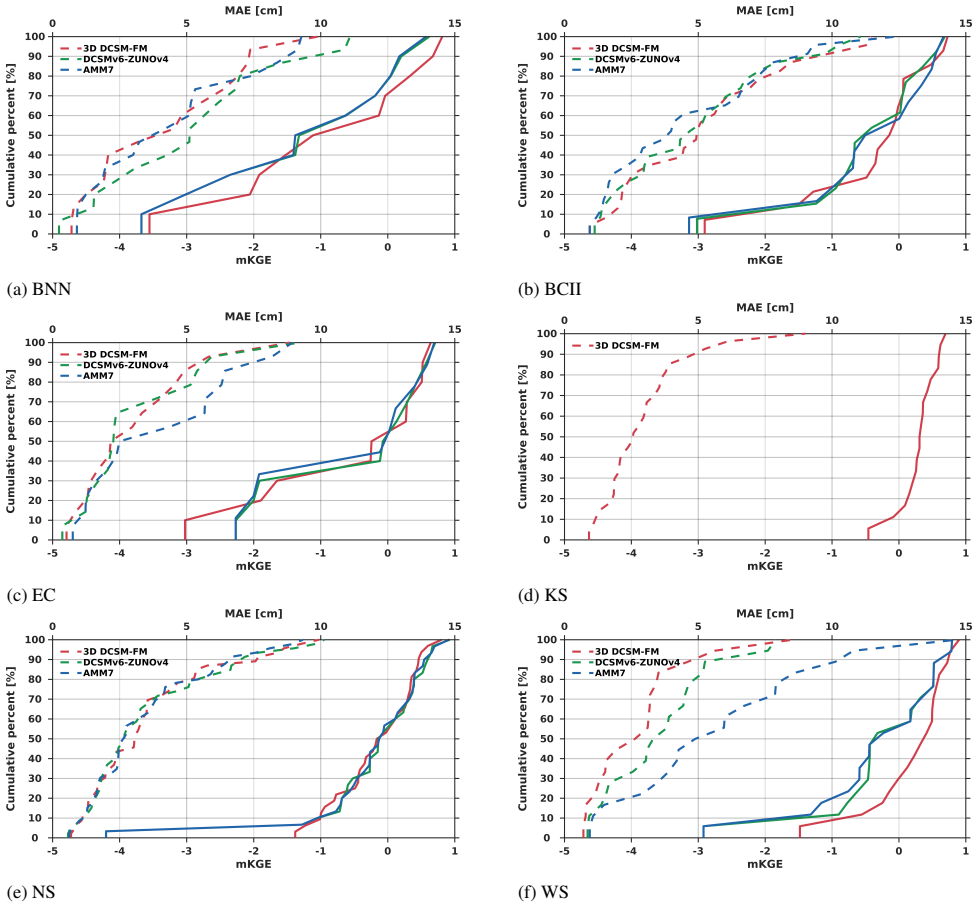
- The performance of each model depends on the subregion. This can be observed

by comparing both the MAE and mKGE cumulative distribution functions for the different subregions shown in Fig. 3.5.

- In all subregions, there are a number of tide gauges at which the models apparently lack the skills to represent the coastal 1-SMWLs (mKGE $<< 0$). Note that these tide gauges may differ per model. Expressed in percentages, we observe that in the worst case (subregion BNN) only 20% of all tide gauges have a mKGE > 0, while in the best case (subregion KS) this percentage is about 80%.
- In BNN and WS, the 3D DCSM-FM model clearly outperforms the other models. In BCII, the AMM7 model shows the best performance. In the NS and EC, the performance level is comparable for the different models.
- The best performance for the 3D DCSM-FM model is obtained in the KS region.
- The spatial performance assessment shows strong variability over time. In terms of MAE, we observe for all models an improved performance between 2004–2011. The largest improvement is observed for 3D DCSM-FM (from 4.25 cm in 1998 to ~ 3 cm in 2008). For all models, the performance degrades towards the end of the timespan considered in this study. In terms of the mSPAEF this behaviour is also observed, though it is less pronounced.
- Both in terms of MAE and mSPAEF, the 3D DCSM-FM model outperforms the other models over almost the entire timespan.

It should be noted that a poor temporal performance in terms of MAE and mKGE may be caused by biases in the observation- and/or model-derived 1-SMWL time series. Indeed, such a bias would show up one-to-one in the MAE values and would affect γ_Q (see Eq. (3.5)). A detailed analysis, not shown here, indeed shows that γ_Q has the largest variability and is mainly causing the lower mKGE values observed in Fig. 3.5.

Before interpreting the temporal performance, it is important to remember that although the size of the model domains suggests otherwise, the models may have been primarily developed for specific waters such as the Dutch waters in case of the 3D DCSM-FM and the DCSMv6-ZUNOv4 models. The importance of a good model performance elsewhere is proportional to the extent to which it influences the performance in these target waters. Both target application(s) and area largely explain the differences in the design of the models and the forcing data used in the reanalyses (see Table 3.1). These in turn explain the differences in performance in the different subregions. With this in mind, the higher performance of the 3D DCSM-FM in the WS is not a surprise; this model has a much higher spatial resolution (3D DCSM-FM vs. AMM7) and it is a 3D model (3D DCSM-FM vs. DCSMv6-ZUNOv4). Likewise for the higher performance of AMM7 in the BCII. Contrary to 3D DCSM-FM and DCSMv6-ZUNOv4, the AMM7 reanalysis is a British product. It goes beyond the scope of this paper to fully explain the observed performances. The take-home message from the results is that if one would have the resources to combine the good elements (such as the bathymetry in certain areas, schematization, and parametrization) of all models into a new model, further improvements are possible. At the same time, one has to keep in mind that for the application at hand it is not required to have a good performance at all tide gauge locations. Afrasteh et al. (2021) showed that adding already a small number of



(a) BNN (b) BCII

(c) EC (d) KS

(e) NS (f) WS

Figure 3.5: The cumulative distribution functions for both the MAE (dashed lines) and mKGE (solid lines) metrics per hydrodynamic model for all tide gauges located in the subregion.

hydrodynamic leveling connections to the UELN dataset has a significant impact on the quality of the EVRS. Our results show that in each subregions at least some tide gauges are available where the model has the necessary skills.

The spatial performance is somewhat difficult to explain. This mainly concerns the deterioration observed from the year 2011 onward. The improvement in the first period is in line with the improvement in the quality of the forcing data / open boundary conditions caused by the increased amount and quality of available observations to generate them (e.g., Hersbach et al., 2020). The deteriorated performance starting in 2011 does not fit into this picture. We would like to emphasize that time variations in the distribution of the number of tide gauges over the various subregions (Fig. 3.2) are not the explanation. When we repeat the analysis using only all tide gauges for which a full time series is available, we see the same pattern. Possible explanations for the deteriorated performance starting in 2011 include: i) errors in the forcing data/boundary conditions, and ii) model errors

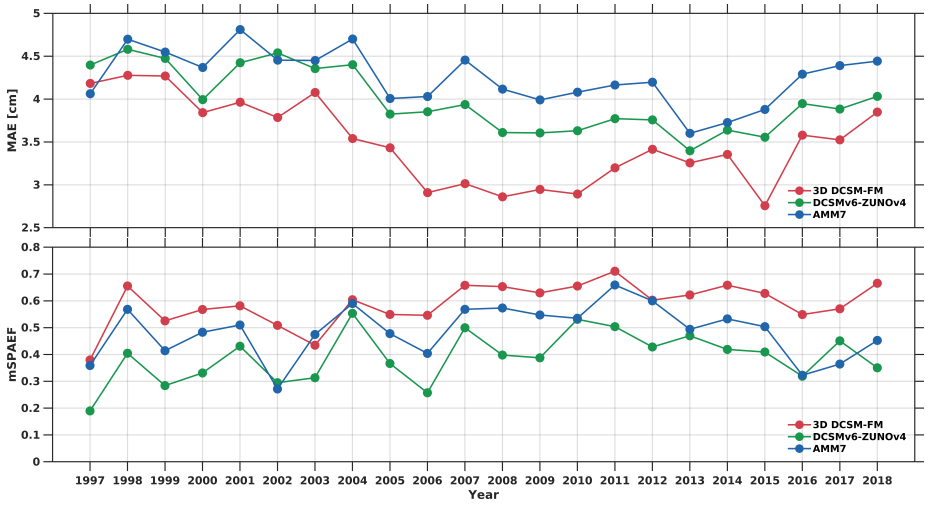


Figure 3.6: The MAE (top panel) and the mSPAEF (bottom panel) metrics used to assess the spatial model performance for each of the three considered hydrodynamic models.

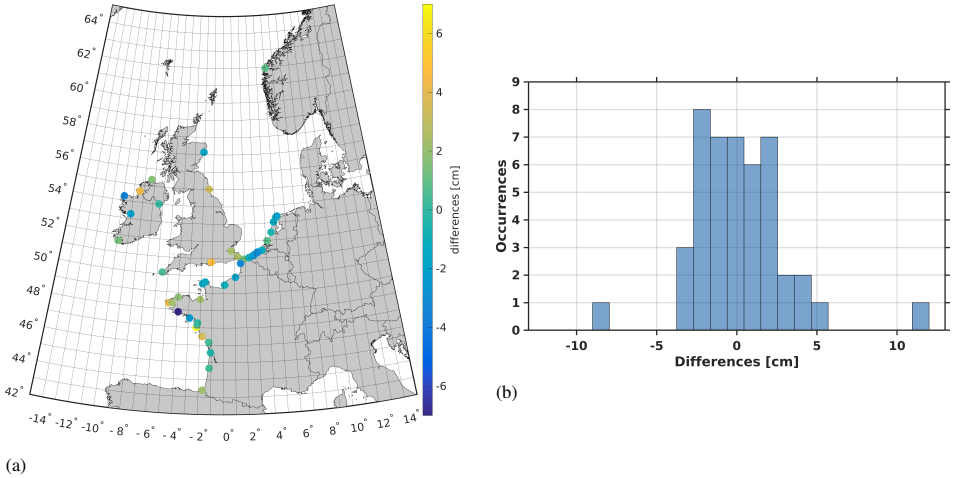
(e.g., missing large-scale, slow (multi-decadal) dynamic processes in the model physics). Regarding the latter, remember that all three models were originally developed to make short-term operational forecasts.

3.4.3. ASSESSMENT OF THE CONTRIBUTION OF ERRORS IN THE VERTICAL REFERENCING OF THE TIDE GAUGES TO THE OBSERVATION-DERIVED SMWLS

At 45 tide gauges (see Fig. 3.1), we computed the differences between $(h^{\mathrm{HBM}}_{\mathrm{GNSS}} - H^{\mathrm{HBM}}_{\mathrm{NHD}})$ and h^{HRS} (see Eq. (3.7)). A map and histogram of the differences are shown in Fig. 3.7. The results show the following:

- The median of the differences is 0.0 cm, while the standard deviation (estimated as before) is 2.7 cm.
- The French tide gauges Concarneau and L'Herbaudière behave as outliers (absolute difference > 5 cm). For Concarneau, the last leveling seems to be conducted in April 2003 SONEL (2007), while GNSS was installed in November 2007 SONEL (2003). The distance between the GNSS antenna and the tide gauge is about 1 km. We lack any details about how and when the GNSS has been connected to the TGBM. L'Herbaudière is quite a new station (data since June 2014). No further metadata for this tide gauge is available.
- We do neither observe much regional variability, nor strong spatial patterns.

Assuming that GNSS provides heights with an uncertainty between 0.5–1 cm and that we have connected both the tide gauge as well as the GNSS to the TGBM using precise leveling with an uncertainty of just a few mm, the expected uncertainty of the ellipsoidal



(a) (b)

Figure 3.7: Map and histogram of the differences between $(h^{\text{HBM}}_{\text{GNSS}} - H^{\text{HBM}}_{\text{NHD}})$ and h^{HRS} for tide gauges equipped with GNSS receivers.

height of the SMWL is about 1 cm. If this height is obtained by adding h^{HRS}, the uncertainty is between 1 and 2 cm. This gives an expected uncertainty of the differences between 1.4 and 2.2 cm. The observed standard deviation of 2.7 cm is slightly larger. Possible explanations include: i) vertical land motion between the time GNSS data were acquired and the reference epoch of h^{HRS} (typically, this enters via the corrector surface/innovation function added to the gravimetric (quasi-)geoid), and ii) a lower quality of the h^{HRS} along the coast (many countries including the Netherlands suffer from a gap in gravity data along the coast (Farahani et al., 2017)).

The estimated uncertainty (standard deviation) of the EGG2015 quasi-geoid is 1.9 cm (Denker et al., 2018). This makes the expected uncertainty of the SMWLs expressed relative to EGG2015 2.1 cm in case GNSS is exploited and between 2.1 and 2.8 cm otherwise. Consequently, a significant part of the observed variance (see Sect. 3.4.1) may be explained by uncertainties associated with the vertical referencing of the tide gauges/SMWLs. For example, in case of the 3D DCSM-FM the contribution is between approximately 30% and 50% for the 1-SMWLs. In terms of nugget effect, the contribution is between 35% and 65%. Indeed, part of these errors will be correlated. Denker et al. (2018) refer to a covariance function being derived for the computed height anomalies, which has a half-length of about 40 km and zeros at about 80, 220, 370 km, and so on. They stated that "over longer distances, e.g. beyond the second and third zero of the covariance function, the height anomalies are nearly uncorrelated". Since we lack the full VC matrix of the EGG2015, we cannot assess/remove the contribution of EGG2015 to the observed error correlations (remember that we use a different distance metric). In the remainder, we will remove a contribution of 2.15^2 $(1.9^2 + 1^2)$ from the observed variance. Indeed, the vertical referencing to EGG2015 is only required to obtain a noise model. It is not required when computing hydrodynamic leveling connections.

3.4.4. THE COASTAL WATER NOISE MODEL VERSUS THE ALTIMETER-DERIVED DEEP AND SHELF WATER NOISE MODELS

The empirical noise covariance functions for the DCSMv6-ZUNOv4 derived deep and shallow water 22-SMWLs are shown in Fig. 3.8. As a reference, we included the average empirical noise covariance function for the DCSMv6-ZUNOv4 derived coastal 1-SMWLs shown in Fig. 3.3 (middle panel). Note that 3856 points are in deep water, whereas 2294 points are in the shelf waters. Hence, for both empirical covariance functions the number of pairs available per lag is always larger than 1000. Also note that 95% of all altimeter time series comprise more than 345 data points. The empirical noise covariance functions show the following:

- The empirical noise covariance function for the deep waters is significantly different from the one for the shelf waters. The former shows larger covariance values for both short and long sea distances. The latter shows much less fluctuations. It does show, though, some small, large-scale pattern with positive covariances up to a lag distance of about 1000 km and negative covariances for larger distances.
- Both functions show a small jump at $\sim$ 190 km (shelf waters) and $\sim$ 220 km (deep waters). A detailed analysis revealed that up to this distance the covariances are mainly calculated in the along-track direction. From these distances, the covariances are partly determined by pairs in the across-track direction.
- Both functions also differ significantly from the empirical noise covariance function of the coastal 1-SMWL. Contrary to the latter, there seems to be no nugget effect. The pattern over short distances is also different.

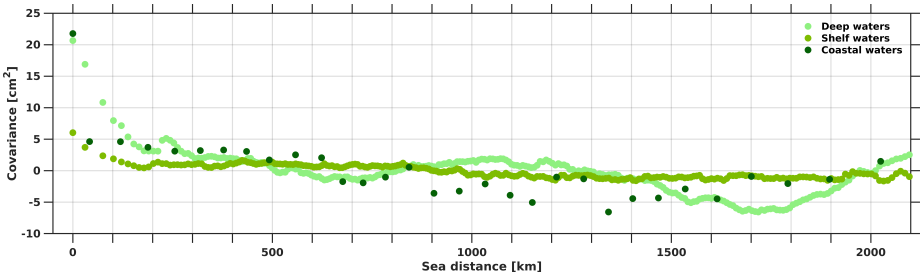


Figure 3.8: The empirical noise covariance function for deep, shelf waters computed using differences between satellite altimetry data and DCSMv6-ZUNOv4 averaged over period 1997–2019. As a reference, we included the empirical noise covariance function for the coastal 1-SMWL computed for DCSMv6-ZUNOv4 (Fig. 3.3, middle panel).

Some caution is required in interpreting the above results. The averaging period is different (1 to 3 summers for the coastal waters and 22 summers for the deep and shelf waters. Furthermore, this comparison has only been conducted for the DCSMv6-ZUNOv4 model. Since this is a 2D model, we cannot expect to get the best performance in deep waters where baroclinic processes dominate the SMWL variability (D. C. Slobbe et al., 2013). At the same time, the results do not contradict expectations based on oceanographic

arguments; along the coast, processes have a higher variability compared to the shelf and in deep water. This may have an impact on the empirical noise covariance function, notably at short spatial scales. The results of this experiment are a first confirmation that this is indeed the case. It shows that a noise model for the coastal SMWL cannot be calculated from altimeter data in the shelf and deep waters. Whether or not a dedicated coastal altimetry data product can help in this respect remains to be studied.

3.4.5. THE EXPECTED QUALITY IMPACT OF COMBINING HYDRODYNAMIC LEVELING AND UELN DATA ON THE EVRF REVISITED

The experiment described in Sect. 3.2.3 is conducted using the best-performing 3D DCSM-FM fitted noise model for the coastal 1-SMWLs and 3-SMWLs. Moreover, we quantified the effect of ignoring error covariances for the SMWLs, i.e., using the diagonal elements of the noise VC matrix only. The results, summarized in Table 3.3 and Fig. 3.9 show:

- Adding all 182 connections reduces the median propagated standard deviation of all adjusted heights from 13.8 to 10.6 mm (one summer averaging period) and to 10.3 mm (three summers averaging period). This corresponds to an improvement of 23% and 25%, respectively.

- Ignoring the noise covariances leads to an overestimation of the improvement by 7% (one summer averaging period) and 8% (three summers averaging period), i.e., by using only the diagonal elements of the full noise VC matrix the obtained improvements are 30% and 33%, respectively.

- The improvement differs strongly per country; values range from 1% (Slovakia) to $>$ 50% (Great Britain).

- Including noise correlations leads to a higher overall improvement for the first 9 (three summers averaging period) or 10 connections (one summer averaging period). For larger numbers of connections, the overall improvement is too optimistic if we ignore the noise correlations.

Ignoring the noise covariances, the precision with which we can derive the 1-SMWL differences is 4.6 cm in terms of standard deviation. This is close to the 5 cm that Afrasteh et al. (2021) used in Experiment II, which has the same setup as the experiment conducted in this paper. The results obtained here are therefore almost identical (30% vs. 29%) to the previously published results. What is new is the insight we get when comparing the results obtained with and without including the noise covariances. Indeed, ignoring the noise covariances results in a too optimistic quality impact (i.e., overall improvement is overestimated up to 8%). Only for the first few connections that we add, the overall improvement is larger. Apparently there are two conflicting effects. The first has to do with the fact that we use model-derived SMWL *differences* in hydrodynamic leveling. As noise in the SMWLs is correlated, some of the errors will be eliminated once we calculate the differences. This has a positive impact on the overall improvement. On the other hand, correlations among the hydrodynamic leveling connections result in less 'information' being added when the number of added connections increases. This is the second effect.

Table 3.3: The median propagated standard deviation (SD) of the adjusted heights in millimeters (per country and in total). Note that we only included the countries for which the improvement is larger than 10%. The number between the brackets indicates the improvement in percentage relative to the value obtained for the spirit-leveling only solution. 'Full VC' and 'diag VC' refers to the usage of either the full noise VC matrix of the SMWLs or the diagonal elements only.

Country	spirit leveling-only solution	1-SMWLs (full VC)	1-SMWLs (diag. VC)	3-SMWLs (full VC)	3-SMWLs (diag. VC)
Denmark	10.9	9.4 (14)	9.0 (18)	9.2 (15)	8.8 (19)
Estonia	13.8	12.2 (11)	12.0 (13)	12.0 (13)	11.6 (16)
Finland	18.3	13.5 (26)	11.6 (37)	13.0 (29)	11.0 (40)
France	20.7	16.8 (19)	16.2 (22)	16.3 (21)	15.6 (25)
Great Britain	31.1	16.2 (48)	14.5 (53)	15.2 (51)	14.0 (55)
Italy	14.1	13.2 (6)	12.6 (11)	13.0 (7)	12.2 (13)
Lithuania	12.7	11.3 (10)	11.2 (12)	11.1 (12)	10.9 (14)
Latvia	13.7	12.3 (11)	12.1 (12)	12.0 (13)	11.7 (15)
Montenegro	23.4	21.5 (8)	21.1 (10)	21.2 (10)	20.7 (12)
Norway	19.1	15.3 (20)	14.4 (24)	14.9 (22)	13.9 (27)
Portugal	45.6	23.1 (49)	22.1 (51)	22.1 (51)	21.2 (54)
Spain	40.6	23.4 (42)	22.3 (45)	22.6 (44)	21.8 (46)
Sweden	14.0	10.6 (25)	9.5 (32)	10.2 (27)	9.1 (35)
Total	**13.8**	**10.6 (23)**	**9.6 (30)**	**10.3 (25)**	**(9.2) (33)**

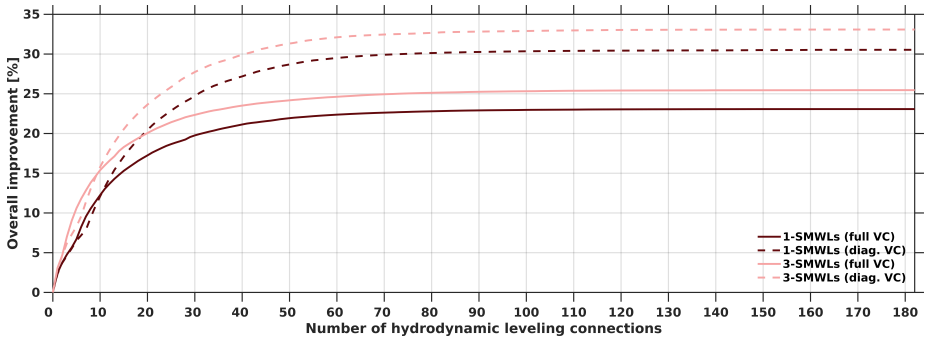


Figure 3.9: Overall quality improvement of the EVRF (expressed as the percentage with which the median propagated height standard deviation decreases compared to the spirit-leveling only solution) obtained by adding hydrodynamic leveling data as a function of the number of added connections. The solid lines correspond to scenarios in which the full noise VC matrix for the SMWL was used computed using the 3D DCSM-FM noise models associated to the coastal 1-SMWLs and 3-SMWLs. The dashed lines correspond to scenarios in which the covariances were ignored.

The choice of the composite analytical model fitted through the empirical covariance function has limited impact on the total quality impact. This follows from an experiment in which we used a composite analytical model defined by the superposition of the nugget effect and Cardinal Sine (see Hristopulos, 2020, Sect. 4.2.2) models (not shown). Similar to the J-Bessel model, the Cardinal Sine model admits negative covariances. Compared to the use of the J-Bessel model, the total quality impact increases from 23% to 25% for the coastal 1-SMWLs. The quality of the fit through the empirical covariance function,

however, is slightly lower.

Note that in presenting and discussing these results, we focused on the impact of our developed noise model. For a discussion of these results in a broader sense, we refer to Afrasteh et al. (2021).

3.5. SUMMARY AND CONCLUSION

A recent study by Afrasteh et al. (2021) showed that combining model-based hydrodynamic leveling data with data of the Unified European Leveling Network (UELN) may improve the quality of the European Vertical Reference Frame (EVRF) significantly. Assuming a variance of 4.5 cm^2 for the model-derived SMWL differences (corresponding to a 3 cm standard deviation for the hydrodynamic leveling connections), the observed reduction of the median standard deviation of the adjusted heights was 38%. In case the variance is 12.5 cm^2 (corresponding to a 5 cm standard deviation for the hydrodynamic leveling connections), this improvement is still 29%. Although promising, evidence so far has been lacking that hydrodynamic models can indeed represent the SMWL differences with this precision. In addition, the assumption of uncorrelated noise is not realistic. This study builds on our previous work by developing and analyzing a noise model for the model-derived coastal SMWLs and using it to obtain a more realistic quality impact of combining hydrodynamic leveling and UELN data in realizing the European Vertical Reference System (EVRS).

To develop the noise model, we used an empirical approach based on calculating an average empirical covariance function from the differences between tide gauge- and model-derived SMWLs. Three models have been used: the 3D DCSM-FM model, the DCSMv6-ZUNOv4 model and the AMM7 model. A reanalysis was performed for the first two models over the period 1997–2019. For AMM7, we used the output for the same timespan from a publicly available reanalysis. Given the differences in coverage, we had 171, 150, and 123 tide gauges available for 3D DCSM-FM, DCSMv6-ZUNOv4, and AMM7, respectively. In order to have multiple realizations of the noise, an averaging period of 1 summer was chosen. The impact of an averaging period of 2 and 3 summers was examined for the 3D DCSM-FM model.

First, we presented the empirical noise covariance functions for the different models. The functions show that the noise is indeed spatially correlated, although the correlations are different per model. Furthermore, they all show a relatively large discontinuity at the origin (i.e., nugget effect). This points to random errors in the observation- and model-derived SMWLs and signals at short spatial scales that cannot be resolved from the data or by the hydrodynamic model. The nugget effect is lowest for the 3D DCSM-FM (12.1 cm^2) and highest for the DCSMv6-ZUNOv4 (16.3 cm^2). The variance values for these models are 15.3 cm^2 and 21.7 cm^2, respectively. The empirical noise covariance functions obtained from an averaging period of two and three years do not really differ from those associated with an averaging period of one year. The biggest change is in the variance and nugget effect; the variance drops to 14.1 cm^2 and 12.7 cm^2, and the nugget effect to 11.0 cm^2 and 10.0 cm^2.

Second, we assessed both the spatial and temporal performance with which the considered hydrodynamic models are able to represent the coastal 1-SMWL. In both assessments, two different metrics have been used. The results show that for all three models, the perfor-

mance varies over space and time. Regarding temporal performance, there is no model that performs best everywhere; different models score better in different subregions. In some subregions we see no difference in performance. Also in all subregions there are more or less tide gauges where models do not show good agreement with the observations. In most cases the poor agreement is caused by biases in the observation- and/or model-derived 1-year SMWL time series. Regarding spatial performance, all models showed an improved performance in the period 2004–2011. Here we saw the best performance for 3D DCSM-FM over almost the entire period. It was beyond the scope of this study to explain the differences in performance. In any case, it is not important for hydrodynamic leveling that the models always and everywhere have a good performance; we can omit tide gauges/time periods where/in which the model has poorer performance. At the same time, the fact that the models have a different performance in different regions indicates that further improvements are possible. That is, one could combine the good elements of all models in one, new model.

Errors in the vertical referencing of the tide gauges contribute to the obtained noise covariance functions. In this study we looked at their contribution to the variance. An analysis of the differences between the ellipsoidal heights of the local height reference surface (HRS) obtained from GNSS/leveling and the HRS that is officially used results in an uncertainty of 2.7 cm in terms of standard deviation. This is slightly higher than the expected uncertainty (between 1.4 and 2.2 cm). Possible explanations are: i) vertical land motion between the time GNSS data were acquired and the reference epoch of the HRS, and ii) a lower quality of the HRS along the coast. Given the 1.9 cm uncertainty of EGG2015, we estimate that for 3D DCSM-FM ultimately between 30% and 50% (lower for the other models) of the variance is explained by the uncertainty in the vertical referencing of the tide gauge/SMWLs.

Next, we presented the empirical noise covariance functions for the deep and shelf waters in the target area calculated from TOPEX/Jason satellite altimetry data. As only the WAQUA software package, on which the DCSMv6-ZUNOv4 model is based, provided the opportunity to generate output at user-defined locations and epochs, the deep and shelf water empirical covariance functions are only calculated for this model. Both functions are not only significantly different from each other, but also from the function computed for coastal SMWLs. This is in line with oceanographic expectations, namely that the dynamics along the coast are more complex than in deep and shelf waters. Hence, altimeter data have limited value in obtaining a noise model for the coastal SMWLs. It remains to be studied, however, whether a dedicated coastal altimetry data product is useful in this respect.

Finally, we looked at the impact of the improved noise model on the quality of the EVRF. We used the noise model obtained for the 3D DCSM-FM assuming that a comparable performance can also be obtained in other European waters. This assumption is considered to be reasonable in case i) the models have comparable resolutions, ii) the underlying bathymetries have similar quality, and iii) the models are forced using the same datasets. The setup of the experiment was identical to Experiments I and II of Afrasteh et al. (2021), except for the assumption made in that study that the noise VC matrix of the model-derived MWLs was diagonal. The results show that using 1-SMWLs, the expected improvement in the median standard deviation of the adjusted heights is 23%. In the case of averaging over three summers, the improvement is 25%. Ignoring error correlations results

in an overestimation of the total quality impact by 7% (one summer averaging period) and 8% (three summers averaging period).

Developing a noise model for model-derived coastal SMWLs is indeed challenging. Compared to other parts of the world, European waters contain many tide gauges. At the same time, many of these tide gauges are not deployed to build long and stable time series; they include gaps and/or sudden jumps and sometimes exhibit spurious signals in the low frequencies that are probably best explained as measuring (or measurement correction) errors. In many cases, we also lack information about the vertical referencing and its control over time. This is crucial information, given that some tide gauges are located in areas with vertical ground movement. Ideally, one should at least correct for the latter. However, since in most cases the necessary information is missing, this correction is often neglected. In any case, we can expect that almost everywhere the model errors are significantly larger than the vertical land motion. A significant contributor to the computed empirical noise covariance functions are errors in the vertical referencing of the tide gauges/SMWLs. Removing this requires at least a full noise VC matrix for the (quasi-)geoid model used. If all the tide gauges are properly connected to a nearby UELN height marker an iterative approach is possible in which the vertical referencing is improved based on new realizations of the EVRS. However, given the small impact of taking the noise covariances into account as demonstrated in this study, the question is whether it is worth doing so.

In any case, the results of this research encourage further development of model-based hydrodynamic leveling. Indeed, we have shown that today's hydrodynamic models have the accuracy to improve the quality of the EVRF up to 25%. Our future work will be to demonstrate the quality impact of including model-based hydrodynamic leveling data in realizing the EVRS using real data.

BIBLIOGRAPHY

Afrasteh, Y., Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J., Snellen, M., & Zijl, F. (2021). The potential impact of hydrodynamic leveling on the quality of the European vertical reference frame. *Journal of Geodesy*, *95*(8). https://doi.org/10.1007/s00190-021-01543-3

Ågren, J., Kempe, C., & Jivall, L. (2018). Noggrann höjdbestämning med den nya nationella geoidmodellen SWEN17_RH2000 (in Swedish). Presented at the conference Kartdagarna, 20-22 March 2018, Linköping, Sweden.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., et al. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, *40*, 1–20. https://doi.org/https://doi.org/10.1016/j.envsoft.2012.09.011

Bossler, J. D. (1984). *Standards and specifications for geodetic control networks*. National Geodetic Information Branch, NOAA, FGCC.

Cartwright, D. E., & Edden, A. C. (1973). Corrected tables of tidal harmonics. *Geophys. J. Roy. Astron. Soc.*, *33*, 253–264. https://doi.org/doi:10.1111/j.1365-246X.1973.tb03420.x

Cartwright, D. E. and Taylor, R. J. (1971). New computations of the tide-generating potential. *Geophys. J. Roy. Astron. Soc.*, *23*, 45–74. https://doi.org/doi:10.1111/j.1365-246X.1971.tb01803.x

Christakos, G. (2012). *Random Field Models in Earth Sciences*. Dover Publictions Inc, USA.

Codiga, D. L. (2011). *Unified tidal analysis and prediction using the UTide Matlab functions* (tech. rep.) [GSO Technical Report 2011-01]. Graduate School of Oceanography, University of Rhode Island.

Codiga, D. L. (2020). UTide Unified Tidal Analysis and Prediction Functions.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman; Hall.

Denker, H., Timmen, L., Voigt, C., Weyers, S., Peik, E., Margolis, H. S., Delva, P., Wolf, P., & Petit, G. (2018). Geodetic methods to determine the relativistic redshift at the level of 10^{-18} in the context of international timescales: A review and practical results. *Journal of Geodesy*, *92*(5), 487–516.

Denker, H. (2015). A new European gravimetric (quasi) geoid EGG2015. *26th IUGG General Assembly, June22–July2*.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271. https://doi.org/10.1007/bf01386390

Ditmar, P., Teixeira da Encarnação, J., & Farahani, H. H. (2011). Understanding data noise in gravity field recovery on the basis of inter-satellite ranging measurements ac-

quired by the satellite gravimetry mission GRACE. *Journal of Geodesy*, *86*(6), 441–465. https://doi.org/10.1007/s00190-011-0531-6

Farahani, H. H., Klees, R., & Slobbe, D. C. (2017). Data requirements for a 5-mm quasi-geoid in the Netherlands. *Studia Geophysica et Geodaetica*, *61*(4), 675–702. https://doi.org/10.1007/s11200-016-0171-7

Farahani, H. H., Slobbe, D. C., Klees, R., & Seitz, K. (2016). Impact of accounting for coloured noise in radar altimetry data on a regional quasi-geoid model. *Journal of Geodesy*, *91*(1), 97–112. https://doi.org/10.1007/s00190-016-0941-6

Filmer, M. S., Hughes, C. W., Woodworth, P. L., Featherstone, W. E., & Bingham, R. J. (2018). Comparison between geodetic and oceanographic approaches to estimate mean dynamic topography for vertical datum unification: evaluation at Australian tide gauges. *Journal of Geodesy*, *92*(12), 1413–1437. https://doi.org/10.1007/s00190-018-1131-5

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Gupta, H. V. and Kling, H. and Yilmaz, K. K. and Martinez, G. F. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003

Hermans, T. H. J., Le Bars, D., Katsman, C. A., Camargo, C. M. L., Gerkema, T., Calafat, F. M., Tinker, J., & Slangen, A. B. A. (2020). Drivers of Interannual Sea Level Variability on the Northwestern European Shelf. *Journal of Geophysical Research: Oceans*, *125*(10). https://doi.org/10.1029/2020jc016325

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holt, J. T., Allen, J. I., Proctor, R., & Gilbert, F. (2005). Error quantification of a high-resolution coupled hydrodynamic–ecosystem coastal–ocean model: Part 1 model overview and assessment of the hydrodynamics. *Journal of Marine Systems*, *57*(1-2), 167–188. https://doi.org/10.1016/j.jmarsys.2005.04.008

Hristopulos, D. T. (2020). *Random Fields for Spatial Data Modeling – A Primer for Scientists and Engineers*. Springer Nature.

Iglesias, I., Avilez-Valente, P., Pinho, J., Bio, A., Vieira, J. M., Bastos, L., & Veloso-Gomes, F. (2020). Numerical Modeling Tools Applied to Estuarine and Coastal Hydrodynamics: A User Perspective. In J. S. A. D. Carmo (Ed.), *Coastal and marine environments*. IntechOpen. https://doi.org/10.5772/intechopen.85521

IGN. (2021). Grilles altimétriques | Géodésique [Accessed: 2022-03-22].

Jahanmard, V., Delpeche-Ellmann, N., & Ellmann, A. (2021). Realistic dynamic topography through coupling geoid and hydrodynamic models of the Baltic Sea. *Continental Shelf Research*, *222*, 104421. https://doi.org/https://doi.org/10.1016/j.csr.2021.104421

Kantha, L., & Clayson, C. (2000). Chapter 7 - Coastal Dynamics and Barotropic Models. In *Numerical models of oceans and oceanic processes* (pp. 493–528). Academic Press. https://doi.org/https://doi.org/10.1016/S0074-6142(00)80012-1

Koch, J., Demirel, M. C., & Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological

models. *Geoscientific Model Development*, *11*(5), 1873–1886. https://doi.org/10.5194/gmd-11-1873-2018

Lysaker, D. I., & Vestøl, O. (2020). The Norwegian vertical reference frame NN2000.

Madec, G., Bourdallé-Badie, R., Bouttier, P. A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., et al. (2017). Nemo ocean engine.

Mäkinen, J., & Ihde, J. (2008). The Permanent Tide In Height Systems. In *International Association of Geodesy Symposia* (pp. 81–87). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85426-5_10

Mogensen, K., Balmaseda, M. A., Weaver, A., et al. (2012). *The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4* (Vol. 668). ECMWF Reading, UK.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Newbury Park, CA, USA: Sage Publications.

Ordnance Survey. (2015). *A guide to coordinate systems in Great Britain* (tech. rep.). Ordnance Survey.

Petit, G., & Luzum, B. (2011). *IERS Conventions (2010)* (tech. rep.). Verlag des Bundesamts für Kartographie und Geodäsie. Frankfurt am Main.

Renshaw, R., Wakelin, S., Golbeck, I., & ODea, E. (2021). North West European Shelf Production Centre NWSHELF_MULTIYEAR_PHY_004_009.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. https://doi.org/10.1080/01621459.1993.10476408

Sandwell, D. T., & Smith, W. H. F. (2014). Slope correction for ocean radar altimetry. *J. Geod.*, *88*(8), 765–771. https://doi.org/10.1007/s00190-014-0720-1

Scharroo, R., Leuliette, E. W., Naeije, M. C., Martin-Puig, C., & Pires, N. (2016). RADS Version 4: An efficient way to analyse the multi-mission altimeter database. [ESA Special Publication SP-740]. *Proceedings of the ESA Living Planet Symposium*, 9–13.

Schwabe, J., Liebsch, G., & Schirmer, U. (2016). Refined computation strategies for the new German Combined Quasigeoid GCG2016. *Symposium on Geoid, Gravity and Height Systems (GGHS2016), Thessaloniki, Greece*, 19–23.

Slobbe, C., Klees, R., Farahani, H. H., Huisman, L., Alberts, B., Voet, P., & De Doncker, F. (2018). The belgian hybrid quasi-geoid: Hbg18. v. 1.0. https://doi.org/10.5880/isg.2018.003

Slobbe, D. C., Klees, R., Farahani, H. H., Huisman, L., Alberts, B., Voet, P., & De Doncker, F. (2019). The Impact of Noise in a GRACE/GOCE Global Gravity Model on a Local Quasi-Geoid. *Journal of Geophysical Research: Solid Earth*, *124*(3), 3219–3237. https://doi.org/10.1029/2018jb016470

Slobbe, D. C., Klees, R., Verlaan, M., Zijl, F., Alberts, B., & Farahani, H. H. (2018). Height system connection between island and mainland using a hydrodynamic model: a case study connecting the Dutch Wadden islands to the Amsterdam ordnance datum (NAP). *J. Geod.*, *92*(12), 1439–1456. https://doi.org/10.1007/s00190-018-1133-3

Slobbe, D. C., Verlaan, M., Klees, R., & Gerritsen, H. (2013). Obtaining instantaneous water levels relative to a geoid with a 2D storm surge model. *Cont. Shelf Res.*, *52*(0), 172–189. https://doi.org/10.1016/j.csr.2012.10.002

SONEL. (2003). Sonel GPS [Accessed: 2022-03-22].

SONEL. (2007). Sonel Leveling [Accessed: 2022-03-22].

Tonani, M., & Ascione, I. (2021). Product User Manual: Ocean Physical and Biogeochemical reanalysis NWSHELF_MULTIYEAR_PHY_004_009 NWSHELF_MULTIYEAR_BGC_004_011.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.

Wahr, J. M. (1985). Deformation induced by polar motion. *Journal of Geophysical Research*, *90*(B11), 9363–9368. https://doi.org/doi:10.1029/JB090iB11p09363

Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge university press.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, *20*(11), 5556–5564. https://doi.org/10.1029/2019GC008515

Woodworth, P. L., Hughes, C. W., Bingham, R. J., & Gruber, T. (2013). Towards worldwide height system unification using ocean information. *J. Geod. Sci.*, *2*(4), 302–318. https://doi.org/10.2478/v10156-012-0004-8

Zijl, F., Laan, S., & Groenenboom, J. (2020). Development of a 3D model for the NW European Shelf (3D DCSM-FM).

Zijl, F., Sumihar, J., & Verlaan, M. (2015). Application of data assimilation for improved operational water level forecasting on the northwest european shelf and north sea. *Ocean Dynam.*, *65*(12), 1699–1716. https://doi.org/10.1007/s10236-015-0898-7

Zijl, F., Verlaan, M., & Gerritsen, H. (2013). Improved water-level forecasting for the Northwest European Shelf and North Sea through direct modelling of tide, surge and non-linear interaction. *Ocean Dynam.*, *63*(7), 823–847. https://doi.org/10.1007/s10236-013-0624-2

4

Realizing the European Vertical Reference System using model-based hydrodynamic leveling data

This chapter has been published in Journal of Geodesy Volume 97, Number 10, September 2023, Pages 1 – 19, https://doi.org/10.1007/s00190-023-01778-2

4.1. INTRODUCTION

All realizations of the European Vertical Reference System (EVRS) (Ihde et al., 2002; Ihde et al., 2008) computed so far are solely based on data from the Unified European Leveling Network (UELN). The UELN dataset comprises geopotential differences between height benchmarks (HBMs) obtained by spirit leveling and gravimetry. The data are provided by all participating countries to the Federal Agency for Cartography and Geodesy (BKG) which is in charge of realizing the EVRS. As discussed in Afrasteh et al. (2021), the exclusive use of spirit leveling/gravimetry imposes limitations on the coverage of the European Vertical Reference Frame (EVRF). Indeed, spirit leveling cannot be used to cross large water bodies. Consequently, the EVRF does neither cover islands not connected to the European mainland leveling network through bridges or tunnels (e.g., Ireland), nor offshore platforms. In addition, the inability of spirit leveling to cross large water bodies reduces the strength of the leveling network in coastal countries. This, in turn, limits the precision and reliability of the computed realization. Apart from this, it should be noted that spirit leveling is prone to systematic errors (Vanicek et al., 1980). Leveling errors accumulate over long distances and may introduce slopes in the height system realizations (e.g., Wang et al., 2012). As for the countries that participate in the UELN project, systematic errors are known to be present in the Belgian (Slobbe et al., 2019), British (Penna et al., 2013), and IGN69 French leveling networks (Sacher & Liebsch, 2019).

One way to overcome the limitations in coverage and accuracy of the EVRF is offered by the so-called 'gravity field approach' (Rülke et al., 2012), which has often been solved as a geodetic boundary value problem (e.g., Amjadiparvar et al., 2016; Amos & Featherstone, 2009; Gerlach & Rummel, 2013; Heck & Rummel, 1990; Rummel & Teunissen, 1988; Sánchez & Sideris, 2017). Rülke et al. (2012) applied the approach to determine the height reference frame offsets among the European national height systems. In doing so, GNSS heights and different quasi-geoid models were exploited while the EVRF2007 was used as a reference. The offsets were estimated with an accuracy of about 5 cm for most countries. Wu et al. (2018) assessed the potential of using clock networks in height datum unification (Bjerhammar, 1985) by simulations using the EUVN (European Unified Vertical Network) (Ihde et al., 2000) as a prior. Although their results are very promising, no clock network is operational yet.

Afrasteh et al. (2021) proposed to overcome the limitations in coverage and accuracy of the EVRF imposed by the exclusive use of spirit leveling/gravimetry by combining the UELN dataset with model-based hydrodynamic leveling data. Model-based hydrodynamic leveling is introduced by Slobbe et al. (2018) as an efficient and flexible alternative method to connect islands and offshore platforms to the height system on land. The method, which is independent from any other method, uses a regional, high-resolution hydrodynamic model to derive mean water level (MWL) differences between tide gauges. By converting the MWL differences to geopotential differences and adding them to the geopotential differences between the HBMs and the MWLs at the tide gauges at both ends of the link, we establish a so-called 'hydrodynamic leveling connection'. Note that the period over which the water level is averaged is arbitrary. Slobbe et al. (2018) showed, however, that higher accuracy can be obtained by choosing a multi-year period and by only averaging over the summer months (the latter avoids the storm surges period). Therefore we use throughout the paper the term summer mean water levels (SMWLs), which refers to the average water

level calculated over all May to September months of one or more years.

Assuming the availability of a series of hydrodynamic models each covering a part of the European waters and providing uncorrelated SMWLs at the tide gauges with a uniform variance of 4.5 cm^2, Afrasteh et al. (2021) showed that combining UELN data with model-based hydrodynamic leveling data improves the median of the propagated height standard deviations by 38% compared to a solution that uses only spirit leveling/gravimetry. A variance of 4.5 cm^2 for the model-derived SMWL corresponds to a standard deviation of 3 cm for each hydrodynamic leveling connection (to establish a connection, we take the difference between two model-derived SMWLs). If the variance is 12.5 cm^2 (corresponding to a standard deviation of 5 cm for each hydrodynamic leveling connection), the reported improvement is still 29%. They also found increased redundancy numbers for leveling observations close to the coastlines. Afrasteh et al. (2023) reassessed the potential impact of hydrodynamic leveling data on the quality of the EVRF using an empirical noise model for the model-derived SMWLs allowing to include error correlations. Their results suggest an improvement up to 25% if the water levels are averaged over three summer periods.

Both Afrasteh et al. (2021) and Afrasteh et al. (2023) assessed the impact of hydrodynamic leveling data on the quality of the EVRF through geodetic network analysis, using only (assumed) stochastic information of the model-based hydrodynamic leveling data. The main objective of this study is to demonstrate the impact of adding *real* model-based hydrodynamic leveling data between tide gauges in the Baltic Sea and the northwest European continental shelf, including the North Sea and Wadden Sea, on the EVRF. The required model-derived SMWLs are obtained from the Nemo-Nordic hydrodynamic model (Hordoir et al., 2019; Hordoir et al., 2015) and the 3D DCSM-FM (Zijl et al., 2020), which cover the Baltic Sea and the northwest European continental shelf, respectively. The water levels are averaged over three successive summer periods that are the same per region. The reason for using three summer periods is that this is the maximum averaging period for which a noise model (needed to build the full noise variance-covariance matrix of the dataset) is available (Afrasteh et al., 2023).

We find it important to mention that the impact is a *provisional* impact and that *we are not presenting a new official realization of the EVRS that replaces the latest release EVRF2019*. The first follows from the fact that our study is subject to the following limitations. First, in establishing the hydrodynamic leveling connections no attempt was made to connect the tide gauges to the UELN by means of real levelings. Instead, we computed the potential differences from adjusted heights obtained during national height system realizations. Second, we made no attempt to reduce the potential differences to the reference epoch adopted for the EVRS (i.e., epoch 2000.0). Doing so requires information about the long-term sea level variability and vertical land motion at the locations of the involved tide gauges. For most tide gauges, these data are not available or there is a mismatch between the period for which the data are available and the period used in this study. Third, we took the used hydrodynamic models as is. That is, we made no attempt to improve their performance in representing the local MWL at the tide gauge locations (e.g., by improving the (local) spatial resolution of the hydrodynamic models). Finally, no validation is possible due to a lack of accurate control data. However, for those countries showing a large impact we assessed the agreement of the estimated heights with normal heights obtained by differencing GNSS heights and quasi-geoid heights obtained from the European gravimetric

quasi-geoid model EGG2015 (Denker, H., 2015).

The paper is organized as follows. Sect. 4.2 describes the methodology used to i) compute the hydrodynamic leveling data, ii) define the set of connections added to the adjustment, iii) conduct the adjustment and variance component estimation, and iv) assess the impact. Sect. 4.3 introduces the data sets used in this study. In Sect. 4.4, we present, analyze, and discuss all results. Finally, we conclude by summarizing the main findings of the paper in Sect. 4.5.

4.2. METHODOLOGY

4.2.1. FUNCTIONAL AND STOCHASTIC MODEL OF THE HYDRODYNAMIC LEVELING DATASET

In a *rigorous* implementation of model-based hydrodynamic leveling for the realization of the EVRS, the observation equation of the geopotential difference (ΔW) between two UELN height benchmarks (HBMs), one connected to tide gauge i and the other to tide gauge j, reads

$$\Delta W^{\mathrm{HBM}_j}_{\mathrm{HBM}_i} = \Delta W^{\mathrm{TGZ}_i}_{\mathrm{HBM}_i} + \Delta W^{\overline{\mathrm{SMWL}}_i}_{\mathrm{TGZ}_i} + \Delta W^{\widetilde{\mathrm{SMWL}}_j}_{\widetilde{\mathrm{SMWL}}_i} + \Delta W^{\mathrm{TGZ}_j}_{\overline{\mathrm{SMWL}}_j} + \Delta W^{\mathrm{HBM}_j}_{\mathrm{TGZ}_j}. \tag{4.1}$$

Here, TGZ stands for tide gauge zero, and $\overline{\mathrm{SMWL}}$ and $\widetilde{\mathrm{SMWL}}$ are the observation- and model-derived SMWLs, respectively. Note that it is assumed that all data have been reduced to the reference epoch adopted for the EVRS (epoch 2000.0). Also note that the hydrodynamic models assume a constant gravity acceleration, which we also used to compute $\Delta W^{\widetilde{\mathrm{SMWL}}_j}_{\widetilde{\mathrm{SMWL}}_i}$ from the model-derived SMWLs. The adjective 'rigorous' refers to an implementation in which the first and last terms in Eq. (4.1) are determined using spirit leveling and gravimetry. In the context of this project, acquiring these measurements is not realistic and feasible. In addition, none of the available tide gauge records uses TGZ as the vertical reference. The water levels are typically expressed relative to the national height datum (NHD) or chart datum. The metadata to transform the water levels back into water levels with respect to TGZ is missing. Hence, we cannot compute the second and fourth terms in Eq. (4.1).

Therefore, we establish $\Delta W^{\mathrm{HBM}_j}_{\mathrm{HBM}_i}$ using

$$\Delta W^{\mathrm{HBM}_j}_{\mathrm{HBM}_i} = \left(\gamma_{\mathrm{HBM}_i}\widehat{H}^{\mathrm{HBM}_i}_{\mathrm{NHD}_i} - \gamma_{\mathrm{TG}_i}\widehat{H}^{\overline{\mathrm{SMWL}}_i}_{\mathrm{NHD}_i}\right) + \Delta W^{\widetilde{\mathrm{SMWL}}_j}_{\widetilde{\mathrm{SMWL}}_i} + \left(\gamma_{\mathrm{TG}_j}\widehat{H}^{\overline{\mathrm{SMWL}}_j}_{\mathrm{NHD}_j} - \gamma_{\mathrm{HBM}_j}\widehat{H}^{\mathrm{HBM}_j}_{\mathrm{NHD}_j}\right), \tag{4.2}$$

where $\gamma_{\cdot}$ is the GRS80 normal gravity (Moritz, 2000), $\widehat{H}^{\mathrm{HBM}_{\cdot}}_{\mathrm{NHD}_{\cdot}}$ is the adjusted height of the nearest UELN benchmark with respect to the NHD, and $\widehat{H}^{\overline{\mathrm{SMWL}}_{\cdot}}_{\mathrm{NHD}_{\cdot}}$ is the height of the observation-derived SMWL expressed with respect to the NHD. Note that in computing the first and last term, we treat all adjusted heights as normal heights. However, Belgium and The Netherlands use an uncorrected leveled height system, Denmark an orthometric height system, and Great Britain an normal-orthometric height system (Federal Agency for Cartography and Geodesy, 2022b). The 'error' introduced is believed to be insignificant; the coastal areas in Belgium, The Netherlands, and Denmark have minimal topographic variations, whereas according to Filmer et al. (2010) the difference between normal-orthometric and normal heights does not exceed the 2–3 cm level for Australia where the highest mountain is almost 900 m higher than the one in Great Britain. Apart from that, for a small

distance between the HBM and tide gauge the error will partly cancel out in differencing $\widehat{H}^{\mathrm{HBM}_\cdot}_{\mathrm{NHD}_\cdot}$ and $\widehat{H}^{\overline{\mathrm{SMWL}}_\cdot}_{\mathrm{NHD}_\cdot}$ as the error is quite systematic. This also applies for any other systematic error present in the adjusted heights. Furthermore, note that $\widehat{H}^{\overline{\mathrm{SMWL}}_\cdot}_{\mathrm{NHD}_\cdot}$ is in fact the sum of three terms:

$$\widehat{H}^{\overline{\mathrm{SMWL}}_\cdot}_{\mathrm{NHD}_\cdot} = \widehat{H}^{\mathrm{TGBM}_\cdot}_{\mathrm{NHD}_\cdot} + \Delta H^{\mathrm{TGZ}_\cdot}_{\mathrm{TGBM}_\cdot} + H^{\overline{\mathrm{SMWL}}_\cdot}_{\mathrm{TGZ}_\cdot}, \tag{4.3}$$

where $\widehat{H}^{\mathrm{TGBM}_\cdot}_{\mathrm{NHD}_\cdot}$ is the adjusted height of the tide gauge benchmark (TGBM) with respect to the NHD, $\Delta H^{\mathrm{TGZ}_\cdot}_{\mathrm{TGBM}_\cdot}$ is the leveled height difference between the TGBM and TGZ, and $H^{\overline{\mathrm{SMWL}}_\cdot}_{\mathrm{TGZ}_\cdot}$ the tide gauge derived SMWL expressed relative to TGZ. The TGBM is the HBM to which the TGZ is connected. Most of them are not part of the UELN.

The stochastic model of the hydrodynamic leveling dataset needs to account for the contributions of all three terms in Eq. (4.2). The contribution of the first and last term is mainly determined by the uncertainty in $\widehat{H}^{\mathrm{HBM}_\cdot}_{\mathrm{NHD}_\cdot}$ and $\widehat{H}^{\mathrm{TGBM}_\cdot}_{\mathrm{NHD}_\cdot}$ (Eq. (4.3)). Indeed, supposing the connections between the TGBMs and TGZs are established by spirit leveling with a precision of 0.5 mm/km (corresponding to the precision of first-order leveling (Bossler, 1984)) and that the TGBM is typically close (within a few meters to a few kilometers) to the tide gauges, the contribution of $\Delta H^{\mathrm{TGZ}_\cdot}_{\mathrm{TGBM}_\cdot}$ is likely below 1 mm in terms of standard deviation. Similarly, the uncertainty of the MWL computed over one month of sea level observations is already < 1 mm based on a 10-minute sampling and assuming white noise with a standard deviation of 5 cm. As we lack the full variance-covariance matrices of the national height system realizations, we assume that the uncertainty in the first and last term of Eq. (4.2) is described by the precision of second-order leveling (Bossler, 1984), i.e., 1.3 mm/km in terms of standard deviation. The distance between the HBM and TGBM is approximated by the ellipsoidal distance between the nearest UELN benchmark and the tide gauge location.

The full noise variance-covariance matrix associated with the middle term has been obtained using the noise models developed by Afrasteh et al. (2023). For the specific model being used, see Sect. 4.3.1.

4.2.2. DEFINING THE SET OF HYDRODYNAMIC LEVELING CONNECTIONS

A key step in compiling the model-based hydrodynamic leveling dataset is to determine between which tide gauges connections should be established. Assuming N tide gauges are available, $N-1$ independent hydrodynamic leveling connections (i.e., connections that do not form any closed circuit) can be established. For this, there are at most N^{N-2} possibilities (Afrasteh et al., 2021). Similar to Afrasteh et al. (2021, 2023), we use a heuristic search method that identifies the connections one-by-one. Each connection added to the final set is the one that provided the lowest median standard deviation of the adjusted heights. No connections are allowed between tide gauges i) located in different hydrodynamic model domains (i.e., one in the 3D DCSM-FM and the other in the Nemo-Nordic model domain), ii) located within the same country, and iii) located in neighboring countries for which the number of spirit leveling connections between the countries is larger than one. The first criterion stems from the fact that there may be a bias between the water levels obtained from the different hydrodynamic models. The last two criteria are intended to reduce the computational load of the search method. Since we are using real data in this study, there is one additional preparatory step that will be described in the remainder of this section.

As noted in Sect. 4.1, the SMWL differences will be determined over three consecutive summer periods that are the same per region. The latter reduces potential time-dependent, large-scale errors in the modeled water levels (they cancel out in computing the difference) and simplifies the required noise model (Afrasteh et al., 2023). Model-derived water levels and tide gauge records are available from 1997–2019 and 2017–2021 for 3D DCSM-FM and Nemo-Nordic, respectively. Hence, we need to select which three successive years for each of these time spans will be used. In this choice, the following considerations were taken into account:

- The time difference between the center epoch of the three-year period and the reference epoch of the EVRS (i.e., 2000.0) — Ideally, there is no time difference. The greater the difference, the greater the impact of relative differences in long-term sea level variations and vertical land motion at the two involved tide gauge locations.
- Performance of the hydrodynamic model — Afrasteh et al. (2023) showed that for all three hydrodynamic models examined in their study, the performance to represent the SMWLs varies over space and time. As such, it makes sense to select a period in which the model shows a better performance.
- Distribution of the tide gauges — The tide gauges are not homogeneously distributed along the coastline. In addition, the distribution is time-dependent as in many cases the tide gauge records are not full (some tide gauges expired or became available later, and the records may contain gaps). There is no objective way to define the 'best' possible distribution. In this study, we use the maximum sea distance (Afrasteh et al., 2023, Sect. 2.1.4) between two adjacent tide gauges. More specifically, we minimize the weighted sum of the maximum 'sea distance' between two adjacent tide gauges per country, where the weighting is determined by the length of the country's coastline.

The criteria will be applied to the sets of tide gauges available per domain to identify what is referred to as the 'most preferred' and 'least preferred' time spans. These are subsequently used in Experiments I and II to assess the impact of model-based hydrodynamic leveling on the EVRF. The comparison of both experiments allows to evaluate the importance of the time span selection.

4.2.3. HEIGHT NETWORK ADJUSTMENT AND DATA WEIGHTING

The height network adjustment is conducted using weighted least-squares, see Afrasteh et al. (2021) for the equations. Similar to the approach followed in computing EVRF2019 (Sacher & Liebsch, 2019), the datum defect is solved by adding the minimal constraint that for 12 datum points the sum of the height changes is zero. The height of the 12 datum points are obtained from the EVRF2019 adjustment. Hence, the datum of the computed EVRF is the same as the one for the EVRF2019.

After determining which hydrodynamic leveling connections will be added, we conduct the height network adjustment and estimate variance factors for the UELN dataset and each model-based hydrodynamic leveling dataset. In doing so, we used the iterative minimum norm quadratic unbiased estimator (Rao, 1971). The iteration was terminated when the relative change of successive variance factors for all observation groups was smaller than

10^{-4}. Note that each of the observation groups that make up the UELN dataset are scaled with the variance factor estimated by the BKG in computing the EVRF2019 (Sacher & Liebsch, 2019, Table 3).

4.2.4. IMPACT ASSESSMENT

The impact of hydrodynamic leveling data on the EVRF is assessed by comparing the estimated geopotential numbers to the ones computed using UELN data only. If the differences for a country show a variability larger than 10 kgal·mm, we will assess the agreement of the solution with normal heights obtained by differencing GNSS and quasi-geoid heights. In doing so, we interpolate the differences between the adjusted EVRF heights and the heights expressed relative to the NHD to the GNSS data points. By adding the differences to the physical height of the GNSS data points, we obtain their EVRF heights. These are compared with the differences between the GNSS and quasi-geoid heights. Note that the comparison is conducted in the zero-tide system. As the UELN data were provided in the mean-tide system (Sect. 4.3.3), we apply the following transformation (Federal Agency for Cartography and Geodesy, 2022a):

$$C_{\text{zero}} = C_{\text{mean}} - 0.28841\sin^2(\phi) - 0.00195\sin^4(\phi) + 0.09722 + 0.08432, \quad (4.4)$$

where C_{zero} and C_{mean} are the geopotential numbers in respectively the zero- and mean-tide system, and ϕ is the geodetic latitude.

Apart from reporting statistics of the differences, including the median and standard deviation, we also assess the magnitude of trends in the differences in east-west and/or south-north directions. Here, the standard deviation was estimated as 1.4826 × the median absolute deviation (Cook & Weisberg, 1982; Rousseeuw & Croux, 1993). In case we estimate the trend in both directions, the magnitudes are estimated by fitting a plane through the differences. When estimating the south-north slope only, we fit a linear trend and intercept term to the differences as a function of latitude.

4.3. DATA

4.3.1. MODEL-DERIVED WATER LEVEL TIME SERIES

In this study, two reanalysis products generated by two different hydrodynamic models have been used; one for the northwest European continental shelf generated with the 3D Dutch Continental Shelf Model – Flexible Mesh (3D DCSM-FM) (Zijl et al., 2020), and one for the Baltic Sea generated with the Nemo-Nordic model (Nemo-Nordic NS01) (Hordoir et al., 2019), which is based on version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO) model code (Madec et al., 2017) and 3DEnVar data assimilation method (Axell & Liu, 2016). Some alternative products covering our area of interest are publicly available via https://data.marine.copernicus.eu/products. Key requirement is that the models used to generate them included all relevant physics contributing to the MWL variability.

The 3D DCSM-FM reanalysis is the same as used by Afrasteh et al. (2023). It is obtained using the Delft3D Flexible Mesh software framework that allows for the use of unstructured grids. For this study, we have used software version 2.17.05.72090. The 3D DCSM-FM is a three-dimensional hydrodynamic model with a grid resolution that varies between 0.5 and 4.0 nautical miles. It covers the area between 15° W to 13° E and 43° N

to 64° N. The water level time series are generated for the period 1997–2019. For more details about the model setup and forcing data, we refer to Zijl et al. (2020) and Afrasteh et al. (2023, Table 1). Note that the model is actively maintained and improved. One known issue in the version of the model used in this study is an apparent strong vertical circulation between the bottom and the pycnocline in the deep ocean originating from instabilities close to the open boundaries. This results in a less accurate representation of the MWL in deep ocean waters. Detailed investigations by the model developers showed that the issue has no impact on the shelf. A thorough assessment of 3D DCSM-FM's ability to represent the MWLs computed over one summer period (referred to as the '1-SMWLs') is given by Afrasteh et al. (2023). Their results show that the performance varies over space and time. With regard to the latter, they noticed improved performance in the period 2004–2011, although they had no explanation for it.

The Nemo-Nordic model, developed by the Swedish Meteorological and Hydrological Institutes, is a three-dimensional coupled ocean-sea ice model. The model domain covers both the Baltic and the North Sea (i.e., it ranges from 4.15278° W to 30.1802° E and 48.4917° N to 65.8914° N). In this study, however, we only use the water level time series for the Baltic area. The model is based on the NEMO ocean engine version 3.6. For a detailed description of the model setup and forcing data, as well as a validation of the water level time series, we refer to Jahanmard et al. (2021, 2022). Note that the hourly water levels are exported on grids with a horizontal resolution of 1 nautical mile. The reanalysis period ranges from 2017 to 2021.

The method applied to compute the SMWLs over three successive summer periods, referred to as 3-SMWLs, is described by Afrasteh et al. (2023). For the 3D DCSM-FM dataset, a noise model was computed by Afrasteh et al. (2023), which has been exploited in this study. For the Nemo-Nordic dataset, we used the noise model developed for the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (AMM7) (Tonani & Ascione, 2021). The only motivation we have for using this noise model is that the AMM7 hydrodynamic model, similar to the Nemo-Nordic model, relies on the Nucleus for European Modeling of the Ocean model code (Madec et al., 2017). Unfortunately, we lack the required long-time series of observation- and model-derived water levels to develop an empirical noise model specific to this dataset. The lack of a tailored noise model for the Nemo-Nordic dataset is among the reasons why variance component estimation is conducted.

4.3.2. OBSERVATION-DERIVED WATER LEVEL TIME SERIES

Tide gauge data were obtained from the national authorities in all coastal countries included in the two hydrodynamic model domains, except Spain and Ireland (3D DCSM-FM), and Lithuania and Russia (Nemo-Nordic). Spanish tide gauge data were not included in our database because we did not expect a good performance due to the model issue mentioned in Sect 4.3.1. For Ireland, we lack spirit leveling data. Regarding Lithuania and Russia, no data were available. The time span of the records is consistent with the time span of the reanalyses; 1997–2019 (3D DCSM-FM) and 2017–2021 (Nemo-Nordic). For the 3D DCSM-FM, the tide gauge dataset is similar to the one used and described by Afrasteh et al. (2023). It includes about 200 records. Note, however, that most of these records do not cover the entire reanalysis time span. For the Nemo-Nordic domain, about 50 tide gauge

records were available. The records were used to compute monthly MWL time series by means of a harmonic analysis using UTide (Codiga, 2020).

All water levels use the NHD as the vertical reference. For the tide gauges inside the Nemo-Nordic domain, a correction for the vertical land motion induced by glacial isostatic adjustment has been applied. The correction was computed using the regional land uplift model NKG2016LU_abs (Vestøl et al., 2019). It reduces all water levels to the reference epoch 2000.0. Note that this reference epoch is consistent with the one adopted for the EVRS.

A tide gauge is only considered as a potential candidate for a hydrodynamic leveling connection if it is i) outside the tidal flat areas, ii) connected to the NHD, and iii) within 40 km of the nearest UELN HBM. If multiple tide gauges were connected to the same UELN HBM, we used the one of which the median absolute deviation of the differences between the observation- and model-derived monthly MWLs was lowest. The 3-SMWLs were computed from the monthly MWL time series. Note that all monthly MWLs were excluded for which the value of the difference between the observation- and model-derived monthly MWLs exceeded the median of these differences plus/minus three times the standard deviation (estimated as before).

4.3.3. UELN DATA

The UELN data were provided by the BKG. The dataset comprises i) the geopotential differences between the UELN HBMs in the mean-tide system, ii) the a-priori variances for the geopotential differences, and iii) a variance factor for each observation group per country estimated by means of variance component estimation (see Table 4.1).

The dataset is, except for the following changes, the same as the one used to compute the EVRF2019 (release of September 2020). First, we included the data of the Third Geodetic Leveling in Great Britain acquired from 1951 to 1959. Second, we added a number of single connections (i.e., connections that do not form new loops) to HBMs for which GNSS data are available in Norway (61), Belgium (9), and Bulgaria (21). These were meant to facilitate a comparison with normal heights obtained by differencing GNSS heights and quasi-geoid heights obtained from the European gravimetric quasi-geoid model EGG2015. Note that the release of the EVRF2019 in September 2020 includes i) a sign correction in the tidal correction for the Polish dataset and ii) a number of minor updates to the data for Macedonia (error correction), Latvia, Lithuania, Italy (new nodal points), and Bulgaria (two border connections to Turkey have been included).

In computing the EVRF2019, only the datum shift between the Ordnance Datum Newlyn (ODN) and the EVRF was determined. This was done by adding four leveling connections between the HBMs ‘710137’ and ‘1300900’, ‘1300900’ and ‘1300385’, ‘1300385’ and ‘1300386’, and ‘1300386’ and ‘1300397’ to the French zero order leveling network (NIREF) dataset. With the exception of HBM ‘710137’, all mentioned HBMs are on British territory (see Fig. 4.1). The reason for determining only the datum shift stems from the fact that previous realizations of the EVRS showed a tilt relative to the ODN reference frame (Sacher & Liebsch, 2019). This tilt is caused by systematic errors in the British leveling data (see Penna et al. (2013) and cited references). The ODN has been computed based on an adjustment of data from the Third Geodetic Leveling while fixing the results from the Second Geodetic Leveling (1912–1952). On the other hand, the BKG only had (and

Table 4.1: Number of observations and variance factor for each observation group of the UELN dataset. Note that the datasets for Belgium, Bulgaria, and Norway include a number of single connections between a UELN HBM and a GNSS data point. Since no new loops are formed, these connections do not 'contribute' to the solution. The variance factors have been estimated by the BKG when computing the EVRF2019. As explained by Sacher and Liebsch, 2019, the estimated variance factors for the Belgian and old French datasets (referred to as 'France') were not used, but set manually. The variance factor for the British dataset has also been manually set as this observation group was not included in the computation of the official EVRF2019.

Country	**Number of observations**	**Variance factor**
Austria	179	0.832
Belgium	122	1
Bulgaria	118	2.678
Belarus	31	5.389
Croatia	81	1.97
Czech Republic new	185	0.687
Czech Republic old	83	1.41
Denmark	196	0.728
Estonia	418	0.052
Finland	272	0.545
France	344	100
France (NIREF)	1223	1.965
Germany	1112	0.431
Great Britain	61	14.54
Hungary	83	0.306
Italy	203	2.163
Latvia	151	0.722
Lithuania	65	0.566
Macedonia	66	0.905
Netherlands	1373	0.572
Norway new	550	1.805
Norway old	410	2.08
Poland	473	0.743
Portugal	30	4.046
Romania	133	3.044
Russia	176	4.908
Slovakia	196	2.212
Slovenia	89	0.383
Spain	227	5.649
Sweden	4206	0.994
Switzerland	719	0.825
Ukraine	211	2.894

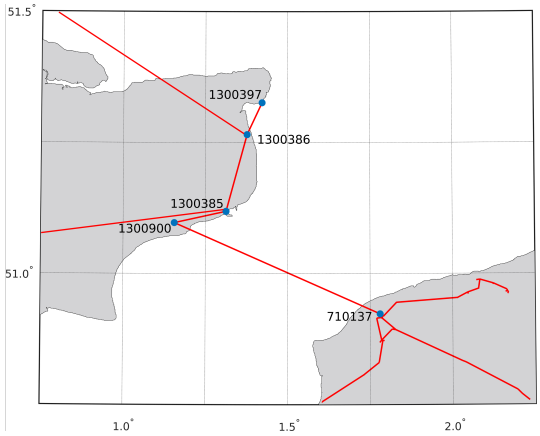


Figure 4.1: A zoom-in of the UELN, showing the connection between Great Britain and France through the Channel Tunnel.

still has) access to data from the Third Geodetic Leveling. In the dataset we used, the four connections mentioned above are part of the British dataset. The variance factor for this data set has been set equal to 9.867 by the BKG, which means that these four connections are given a lower weight than in the computation of the EVRF2019 (the variance factor for the French NIREF data is 1.965).

Fig. 4.2 shows the standard deviations of the UELN observations. Here, the variance factors presented in Table 4.1 are already applied.

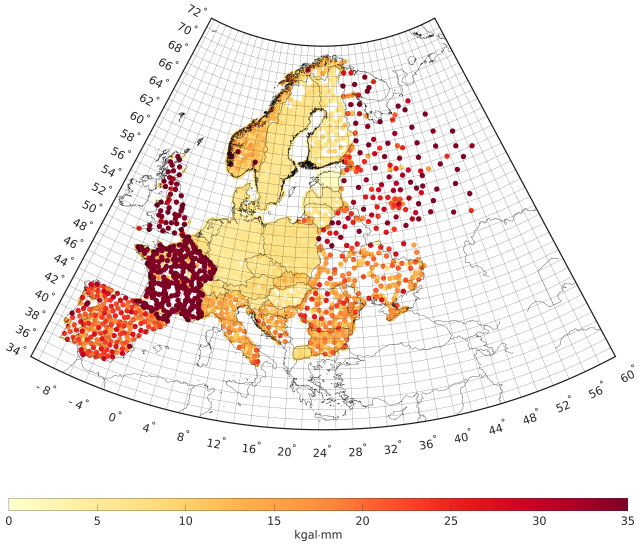


Figure 4.2: Standard deviations of the UELN observations. Note that the variance factors presented in Table 4.1 have been applied.

4.3.4. EUVN_DA DATA

The British and French GNSS data used in the impact assessment are part of the dataset assembled in the European Vertical Reference Network - Densification Action (EUVN_DA) project (Kenyeres et al., 2010). The British dataset comprises 181 data points and the French dataset 164 points. The dataset comprises ellipsoidal coordinates in the European Terrestrial Reference System 1989 (ETRS89) and leveled connections to UELN benchmarks. Kenyeres et al. (2010) did not provide details on when the specific datasets were acquired. It should be somewhere between 2003 and 2008. They also did not specify the length of the sessions over which GNSS data were acquired. They only stated that some countries, including France and Great Britain, "did not observe sessions of 24 hours, but submitted a denser database with a mean site separation close to 50 km." As such, we expect that the accuracy of the ellipsoidal heights is slightly lower than the 1 cm target accuracy.

4.3.5. THE EUROPEAN GRAVIMETRIC QUASI-GEOID MODEL EGG2015

The European gravimetric quasi-geoid model EGG2015 (Denker, H., 2015) is the latest of a series of European quasi-geoid models computed by the Institute of Geodesy at the Leibniz University of Hanover. The model was computed from surface gravity data in combination with topographic information, as well as the GOCO05S geopotential model (Mayer-Gürr, 2015). Here, the remove-compute-restore technique was exploited. The estimated uncertainty in terms of standard deviation is 1.9 cm (Denker et al., 2018). Further details about the datasets, computational method, and uncertainty are provided by Denker (2013) and Denker et al. (2018). Note that EGG2015 is in the zero-tide system. The normal heights obtained by differencing the GNSS and EGG2015 heights are referred to as the 'GNSS/EGG2015 normal heights'.

4.4. RESULTS AND DISCUSSION

4.4.1. EXPERIMENT 0: THE UELN-ONLY SOLUTION

To quantify the impact of model-based hydrodynamic leveling data on the EVRF, we used a solution computed using UELN data only as a reference. This solution is referred to as the 'UELN-only solution'. As we included the British leveling data as well as a few minor updates to some other observation groups of the UELN dataset (Sect. 4.3.3), in Experiment 0 we quantify the differences between this solution and the official EVRF2019 (release September 2020, available at https://evrs.bkg.bund.de/Subsites/EVRS/EN/EVRF2019/evrf2019.html).

The differences are significant at three HBMs. The largest difference is observed for HBM '1706226' (Poland); it reaches 489.5 kgal·mm. A detailed investigation shows that the geopotential number associated with the official EVRF2019 is incorrect. The other two HBMs, i.e., '1300386' and '1300397', are located in Great Britain. In the computation of the EVRF2019, the leveling connections to these HBMs were included in the France NIREF dataset. For both HBMs, the difference equals 6.527 kgal·mm.

A quality assessment of the UELN-only/EVRF2019 solutions is out of the scope of this study. The solutions are known to be contaminated by systematic errors in various observation groups, including but not limited to Belgium (Slobbe et al., 2019), France (Duquenne et al., 2015), and Great Britain (Hipkin et al., 2004; Penna et al., 2013; Ziebart et al., 2008).

Table 4.2: Variance factors estimated in Experiments I and II. Note that the UELN dataset is treated as one observation group. The relative weighting factors of all observation groups that make up the UELN dataset are provided in Table 4.1.

	Experiment I	Experiment II
UELN	0.973	0.964
3D DCSM-FM	2.156	1.121
Nemo-Nordic	2.273	2.174

When evaluating the impact of adding hydrodynamic leveling data, though, we included the UELN solution in the comparison with GNSS/EGG2015 normal heights.

4.4.2. EXPERIMENT I: ADDING HYDRODYNAMIC LEVELING DATA COMPUTED OVER THE MOST PREFERRED TIME SPAN

In Experiment I, we complemented the UELN dataset with hydrodynamic leveling connections between tide gauges in both the Baltic Sea (Nemo-Nordic dataset) and the northwest European continental shelf (3D DCSM-FM dataset). The 3-SMWLs were computed over the periods 2017–2019 and 2004–2006, respectively. Based on the criteria outlined in Sect. 4.2.2, these time spans are considered the most preferred ones. The number of available tide gauges is 50 (Nemo-Nordic) and 76 (3D DCSM-FM). We refer to Fig. 4.3, for a map showing the locations of the tide gauges. In total 124 hydrodynamic leveling connections were added; 49 in the Baltic Sea and 75 in the northwest European continental shelf. For an overview of which connections are being added we refer to Appendix A.

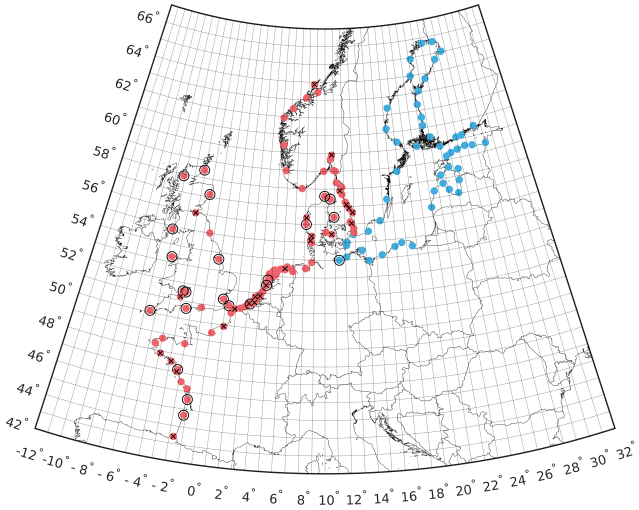


Figure 4.3: The locations of the available tide gauges inside the 3D DCSM-FM (red dots) and the Nemo-Nordic (blue dots) model domains. The tide gauges marked with a black circle are those that are only available for the most preferred time span, while those marked with a black cross are only available for the least preferred one. No black circle or cross means the tide gauge is available for both time spans.

Table 4.2 provides the estimated variance factors for the three observation groups. Fig. 4.4 shows the differences between the geopotential numbers of the UELN-only solution and the ones estimated in Experiment I, as well as the contribution to these differences from both hydrodynamic leveling datasets separately. These contributions were obtained by realizing the EVRS based on the UELN dataset and the identified hydrodynamic leveling connections in the specific sea basin. The applied variance factors are the same as those estimated when using both hydrodynamic leveling datasets. The results show the following:

- Both hydrodynamic leveling datasets are downweighted compared to the UELN dataset; for the latter dataset, the estimated variance factor is 0.973, whereas for the 3D DCSM-FM and Nemo-Nordic hydrodynamic leveling datasets the variance factors are 2.156 and 2.273, respectively. This means that the standard deviations of the added hydrodynamic leveling connections increase from 33.8–41.5 kgal·mm to 49.6–61.0 kgal·mm (3D DCSM-FM) and 38.4–59.9 kgal·mm to 58.0–90.3 kgal·mm (Nemo-Nordic).

- The differences between the geopotential numbers show a large-scale pattern, of negative (positive) values in the west (east). The median difference is -3.1 kgal·mm and the standard deviation (estimated as $1.4826 \times$ the median absolute deviation (MAD)) is 6.1 kgal·mm.

- Great Britain and France stand out in terms of the magnitude of the differences. In Great Britain, the values range between 23.8 and 444.2 kgal·mm. In France, the large differences are concentrated in the southwest. The values range between -96.7 and 10.5 kgal·mm. For all other countries, the values range between -15.2 (Spain) and 8.9 kgal·mm (Poland).

- Only in the countries France and Great Britain the range of the differences exceed 10 kgal·mm.

- The contribution of the Nemo-Nordic dataset (see Fig.4.4b) is largest in Norway, Sweden, and northern Finland (between -16.5 and -10.1 kgal·mm) as well as along the Polish coastline (up to 8.3 kgal·mm). Likewise, the contribution of the 3D DCSM-FM dataset (see Fig.4.4c) is the largest in Great Britain (between 27.9 and 450.8 kgal·mm), France (between -91.6 and 13.8 kgal·mm), and Spain (between -12.4 and -11.8 kgal·mm).

The observed impact of the hydrodynamic leveling datasets on the EVRF is significant, especially in Great Britain and France where differences reach tens to hundreds of kgal·mm. In most other countries the impact is substantially lower; we observe spatially correlated differences of low magnitude. This result is easily explained when considering i) the domain in which model-based hydrodynamic leveling connections have been established, and ii) the weighting (i.e., quality) of the (individual observation groups within the) UELN dataset relative to the two hydrodynamic leveling datasets.

Indeed, on the former, no major impact was expected for non-coastal countries and the countries around the Mediterranean and the Black Seas because the leveling networks of these countries were not strengthened with model-based hydrodynamic leveling connec-

tions. Note that the impact observed in Spain has been propagated from elsewhere; no Spanish tide gauges were used to establish model-based hydrodynamic leveling connections. The difference in impact between Great Britain and France on the one hand and the other North Sea countries on the other hand, as well as the small impact for the Baltic Sea countries, follows directly from the weighting of the different observation groups. Most northwestern European and Baltic Sea countries have high-quality leveling datasets (see Fig. 4.2). As such, it would have been more useful to include hydrodynamic leveling connections in the Mediterranean and Black Seas. The downweighting of the hydrodynamic leveling datasets relative to the UELN dataset also suppresses the impact. While the downweighting is significant (see the increase in the standard deviations of the added hydrodynamic leveling connections), it is within reasonable bounds. Further discussion of the obtained variance factors follows in the discussion of the results of Experiment II (Sect. 4.4.3).

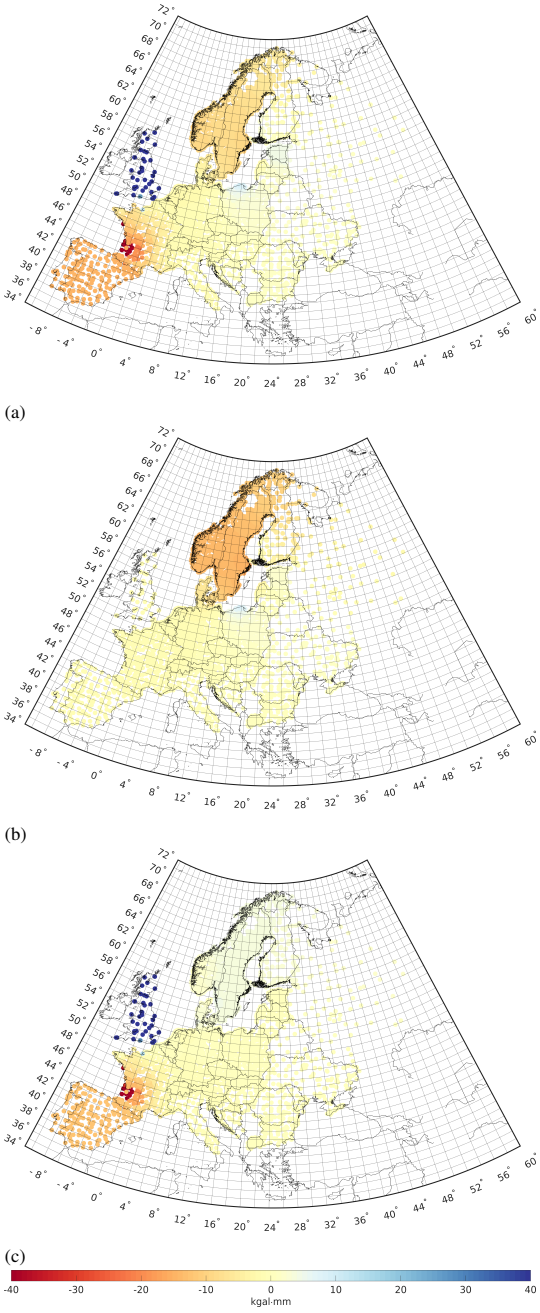


Figure 4.4: Differences between the geopotential numbers associated to the UELN-only solution (Experiment 0) and the ones estimated in Experiment I (a). The panels (b) and (c) show respectively the contributions of the 49 connections added in the Baltic Sea and the 75 in the northwest European continental shelf.

The fact that the differences in France are confined to the southwest is probably related to the difference in density of the old IGN69 leveling network versus that of the new NIREF network. As shown in Fig. 4.5, the two networks only overlap in the northwest. Outside the northwest, the IGN69 network is denser, i.e., contains smaller loops. Given the low weight assigned to this dataset in the adjustment (see Table 4.1), it is not surprising that the solution along the southwest coast is more impacted by the added hydrodynamic leveling connections.

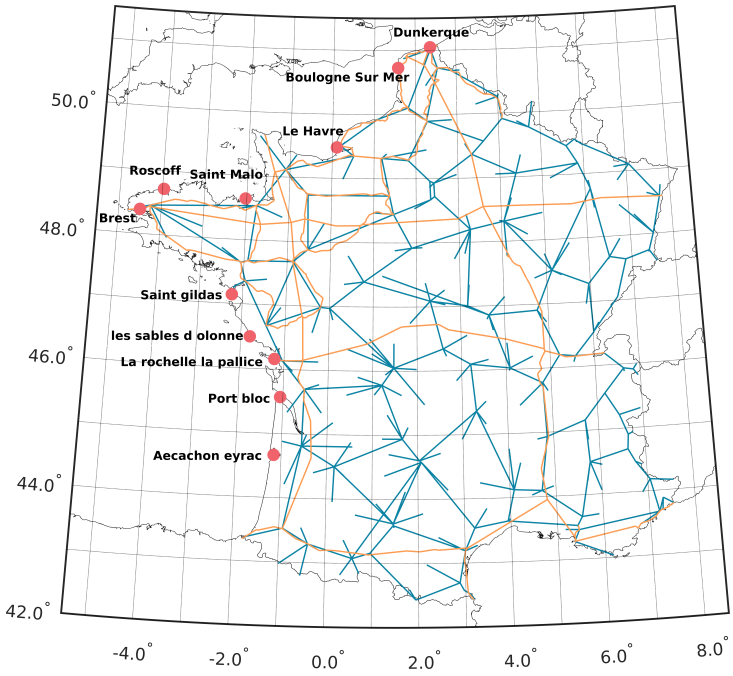


Figure 4.5: The old French leveling network IGN69 (blue) and the new zero-order leveling network named NIREF (orange). The old network is suspected for a tilt of 23 cm in North-South direction (Sacher & Liebsch, 2019).

It is striking to observe that the impact of adding hydrodynamic leveling connections among Baltic Sea tide gauges on the EVRF in Norway and Sweden is larger than when adding leveling connections involving tide gauges along the coasts of Norway and western Sweden (cf. Fig.4.4b and Fig.4.4c). In the latter case, there is hardly any impact. A conclusive explanation for this result is lacking. One explanation might be that the Norwegian and Swedish leveling networks fit well to the central part of the UELN (comprising The Netherlands, Germany, and Denmark) while the Nemo-Nordic dataset introduces some deformation of the network. Alternatively, the Nemo-Nordic dataset may correct some deformation of the network the 3D DCSM-FM dataset is not able to do.

Given the small ranges of the differences per country between the solutions obtained in Experiments 0 and I (they are < 10 kgal·mm, except for France and Great Britain), a comparison with GNSS/EGG2015 normal heights is only conducted for Great Britain and France. The results are summarized by difference maps (Fig. 4.6 and top panel of Fig. 4.8),

plots of the differences as a function of latitude (top panel in Fig. 4.7), plots showing the plane fitted to the differences (Fig. 4.9) and the histograms of the differences (bottom panels in Figs. 4.7 and 4.8). The comparison shows the following:

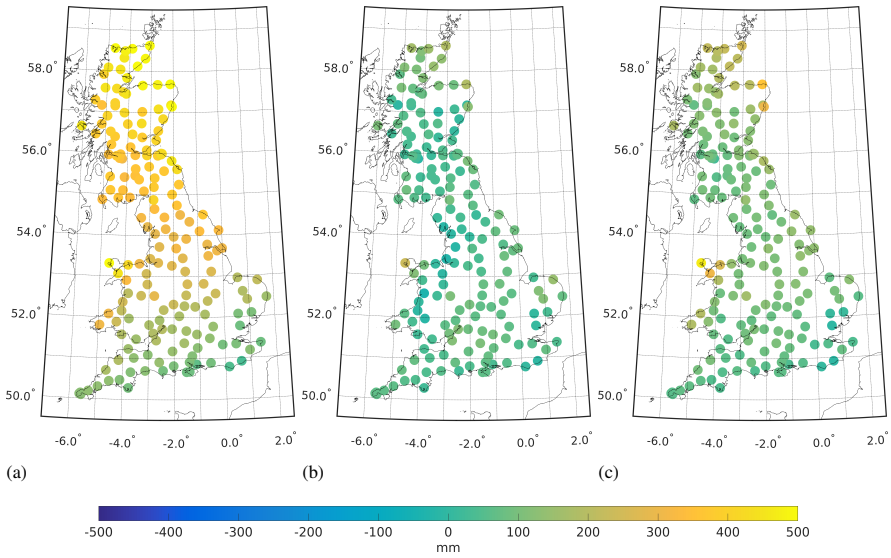


Figure 4.6: Differences between the EVRF heights estimated in Experiments 0 (a), I (b), and II (c) and the GNSS/EGG2015 normal heights at the EUVN data points in Great Britain.

- In Great Britain, the differences for the UELN-only solution show a south-north slope of 48 mm deg^{-1}. In the solution including model-based hydrodynamic leveling data (Experiment I) this trend almost disappeared; the estimated slope is 2.2 mm deg^{-1}.
- In Great Britain, the median of the differences reduces from 304 mm to 57 mm whereas the standard deviation (estimated as before) reduces from 156 mm to 35 mm.
- In France, the differences show a northwest-southeast slope in both solutions. In the southwest the differences are lower for Experiment I. The south-north slope in Experiment 1 reduces from 12.6 mm deg^{-1} to 10.1 mm deg^{-1}. However the east-west slope increases from 5.3 mm deg^{-1} to 7.6 mm deg^{-1}.
- In France, the median difference reduces from -31 mm to -21 mm. The standard deviation (estimated as before) does not change.

The first conclusion to be drawn from the comparison is that the EVRF heights obtained in Experiment I show better agreement with the GNSS/EGG2015 normal heights, both in France and Great Britain. In Great Britain, there is hardly any slope left in the differences associated with Experiment I. In France, the better agreement is concentrated in the southwest. Given the uncertainty of the GNSS data (Sect. 4.3.4) and the EGG2015 (Sect. 4.3.5), we can conclude that for these two countries the impact of model-based hydrodynamic leveling data on the quality of the EVRF is positive. The fact that the impact

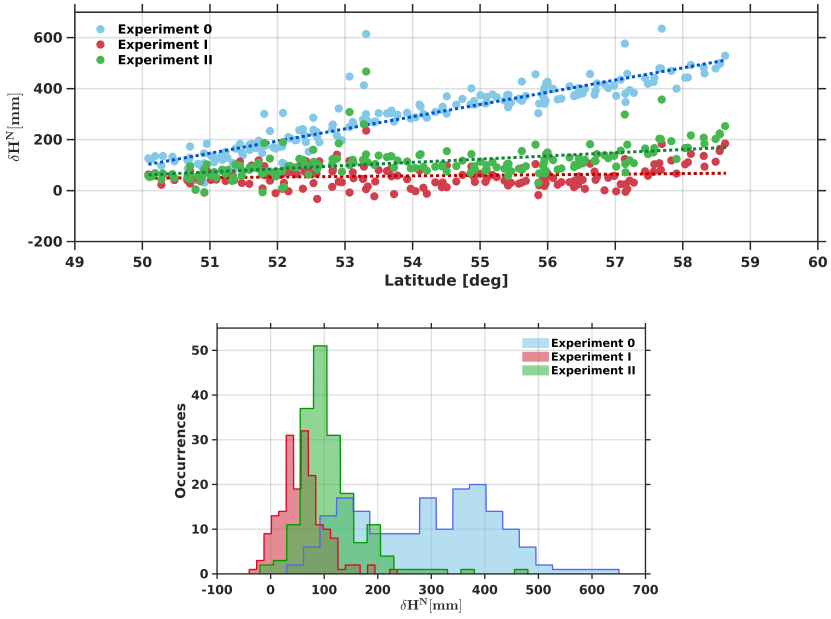


Figure 4.7: The top panel shows the differences between the EVRF heights estimated in the three Experiments and the GNSS/EGG2015 normal heights (δH^N) at the EUVN data points in Great Britain as a function of latitude as well as the fitted linear model. The bottom panel shows the histograms of the differences.

is greater for Great Britain can again be explained by the fact that the border of the British leveling network is completely adjacent to the sea, so that it could be strengthened over the entire perimeter. For France, the strengthening is only on the west side. However, there are possibilities for further reinforcement; in the southwest, France borders the Mediterranean Sea.

Note that the magnitude of the estimated slope associated with the UELN-only solution is larger than the -(20–25) mm deg^{-1} reported in (Penna et al., 2013). This difference originates from the use of both the Second and Third Geodetic Leveling datasets in the realization of the British vertical datum (Ordnance Datum Newlyn), whereas we only used the latter one (see also Sect. 4.3.3). To what extent model-based hydrodynamic leveling data can help to identify the exact source of the systematic leveling errors in the British dataset is outside the scope of this study.

4.4.3. EXPERIMENT II: ADDING HYDRODYNAMIC LEVELING DATA COMPUTED OVER THE LEAST PREFERRED TIME SPAN

In Experiment II, we repeat Experiment I while using time spans to compute the 3-SMWLs that can be considered as the least preferred choice based on the criteria outlined in Sect. 4.2.2. These are 2015–2017 for the 3D DCSM-FM dataset and 2018–2020 for the Nemo-Nordic dataset. Note that a different time span mainly affects the 3D DCSM-FM dataset; the time span over which the reanalysis is conducted is 22 years. For the Nemo-Nordic dataset, we

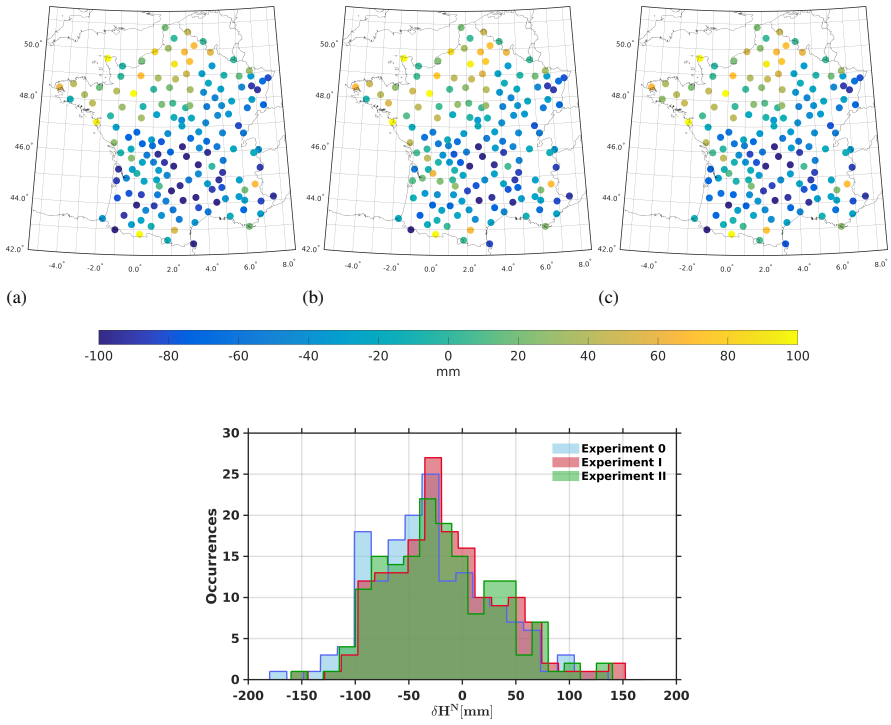


Figure 4.8: Top row shows the differences between the EVRF heights estimated in Experiments 0 (a), I (b), and II (c) and the GNSS/EGG2015 normal heights at the EUVN data points in France. The bottom panel shows the histogram of the differences.

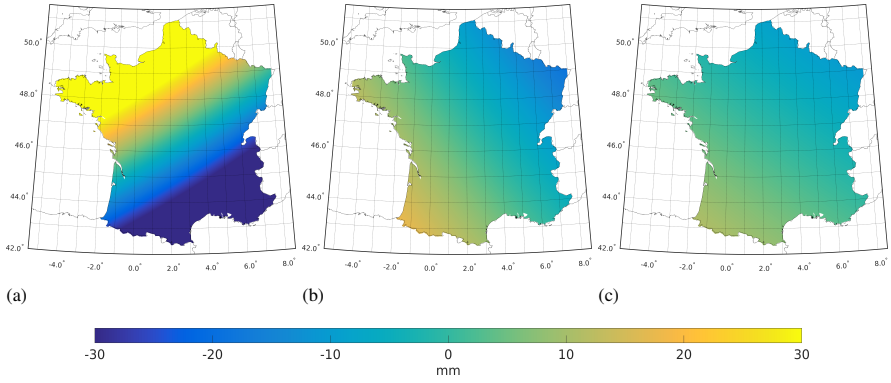


Figure 4.9: Panel (a) shows the plane fitted to the differences between the EVRF heights estimated in Experiments 0 and the GNSS/EGG2015 normal heights. Panels (b) and (c) show the change in the plane fitted in respectively Experiments I and II with respect to the one for Experiment 0. To enhance visibility, the mean values have been removed. These are 548.4, -110.8, and -88.1 in (a), (b), and (c), respectively.

only have 4 years of data available. Hence, there will be an overlap of two years. The choice of the time span impacts the quality of the hydrodynamic leveling connections (remember that the performance of the 3D DCSM-FM reanalysis varies over time (Afrasteh et al., 2023)) as well as the set of tide gauges available to establish connections. In Experiment II, we will assess to what extent the choice of the time span changes the impact of model-based hydrodynamic leveling on the EVRF observed in Experiment I.

For the Baltic Sea and northwest European continental shelf, we have 49 and 79 tide gauges available, respectively (see Fig. 4.3). Table 4.2 provides the estimated variance factors for the three observation groups. Fig. 4.10 shows the differences between the geopotential numbers of the UELN-only solution and the ones estimated in Experiment II, as well as the differences with respect to the solution obtained in Experiment I. The comparison with the GNSS/EGG2015 normal heights are summarized in Figs 4.6–4.8. Moreover, Fig. 4.9 shows the changes in the plane fitted to the differences with the GNSS/EGG2015 normal heights with respect to the one fitted in Experiment 0. The results show the following:

- The downweighting observed in Experiment I is lower. For the Nemo-Nordic dataset, the variance factor reduced from 2.273 to 2.174. For the 3D DCSM-FM, the value reduced from 2.156 to 1.210.
- The overall pattern in the differences compared to the UELN-only solution is the same.
- Changes compared to the solution obtained in Experiment I reach -0.4 kgal·mm and 2.2 kgal·mm in terms of the median and standard deviation (estimated as before), respectively. They are largest in Great Britain (between -235.8 and 16.5 kgal·mm), France (between -33.0 and 79.7 kgal·mm), and the Scandinavian countries (between -12.5 and 3.8 kgal·mm).
- The agreement with the GNSS/EGG2015 normal heights reduced compared to the solution in Experiment I for both France and Great Britain. For Great Britain, we observe an increased offset and slope. The slope in this case increases to 12.6 mm deg^{-1}. For France, we observe that the improved agreement in the southwest has gone. The tilt of the fitted plane in latitudinal direction does not show any significant change compared to Experiment I (i.e., it changed from -10.1 mm deg^{-1} to -10.6 mm deg^{-1}). In longitudinal direction, however, the tilt decreases from 7.6 mm deg^{-1} in Experiment I to 6.4 mm deg^{-1} in Experiment II.

The impact of the different time spans used to compute the 3-SMWLs is significant and to be expected considering the differences between the sets of tide gauges available to establish the hydrodynamic leveling connections (see Fig. 4.3). Compared to the set available in Experiment I, in Experiment II we miss 12 tide gauges in Great Britain while only 2 new ones were added. In southwest France, the set in Experiment I includes two more tide gauges. Also, the set of Danish tide gauges is much less homogeneously distributed in Experiment II. At the same time, the number of Swedish tide gauges on the Kattegat side has increased substantially. In fact, given the small number of tide gauges along the British coast in Experiment II (4), it is surprising that there is still such a level of agreement with the GNSS/EGG2015 normal heights compared to what is observed for the UELN-only solution.

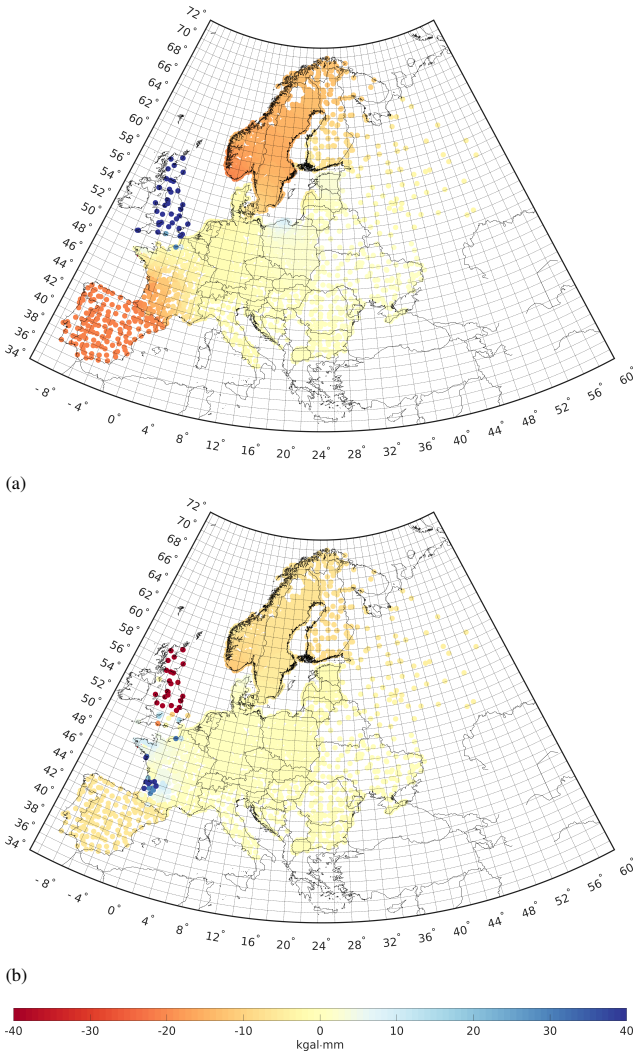


Figure 4.10: Differences between the geopotential numbers associated to the UELN-only solution (Experiment 0) and the ones estimated in Experiment II (a) and the ones associated to Experiments I and II (b).

A striking difference between the two Experiments is observed in the estimated variance factors for the 3D DCSM-FM dataset. In Experiment II the number is significantly smaller. Compared to the variance factor of the UELN dataset, the downweighting observed in Experiment II is lower. This result can be understood by considering the regional variability in performance of the 3D DCSM-FM reported by Afrasteh et al. (2023); the best performance was obtained in the Kattegat-Skaggerak region. In British waters (i.e., Bristol Channel, Celtic Sea, Irish Sea and St. George's Channel, and the Inner Seas off the West Coast of Scotland), the performance was less. Given the aforementioned differences between the two

sets of tide gauges available for establishing hydrodynamic leveling connections, Experiment II added more connections from/to tide gauges in the Kattegat-Skaggerak than to tide gauges in British waters. With regard to the Nemo-Nordic dataset, we would like to remark that a tailored noise model is missing; the length of the available time series is not sufficient to develop such a model. Instead, we relied on a noise model developed for the AMM7 hydrodynamic model covering the northwest European continental shelf (Sect. 4.3.1). The downweighting of the Nemo-Nordic dataset observed in both experiments suggests that the quality of the Nemo-Nordic model in representing the SMWLs is lower than that of the AMM7 model.

4.5. SUMMARY AND CONCLUSION

This study presents the first realization of a regional height reference system, namely the European Vertical Reference System (EVRS), based on the combination of geopotential differences determined with spirit leveling/gravimetry and model-based hydrodynamic leveling. The latter technique was introduced by Slobbe et al. (2018) as an efficient and flexible alternative method to connect islands and offshore platforms to the height system on land. The study built upon two previous studies on exploiting model-based hydrodynamic leveling data. In Afrasteh et al. (2021), we demonstrated the potential of the technique for the realization of the EVRS by means of geodetic network analyzes. That was, we assessed the potential impact on the quality of the European Vertical Reference Frame (EVRF). In Afrasteh et al. (2023), we presented empirical noise models for three different reanalysis products available for the northwest European continental shelf and reassessed the quality impact.

In the present study, we assessed the impact of model-based hydrodynamic leveling on the EVRF using real data. In doing so, we computed summer mean water levels (SMWLs) over three subsequent years (referred to as the 3-SMWLs) with the Nemo-Nordic and 3D DCSM-FM models covering the Baltic Sea and the northwest European continental shelf, respectively. About 250 coastal tide gauges, each located inside one of the two model domains but outside the tidal flat areas, were connected to the nearest Unified European Leveling Network (UELN) height benchmark (HBM). In establishing these connections, we relied on the adjusted heights obtained in the national height system realizations. Note that all tide gauges for which the distance to the nearest UELN HBM is > 40 km were excluded. Moreover, in case multiple tide gauges were connected to the same UELN HBM, we used the one in which the median absolute deviation of the differences between the observation- and model-derived monthly MWLs was lowest. For both reanalysis products, the 3-SMWLs were calculated over two different time spans. The first is considered the most preferred time span based on three criteria, while the second reflects the least preferred choice. The results obtained using these time spans are presented in Experiments I and II. Note that the choice of the time span is particularly relevant for the 3D DCSM-FM dataset; the reanalysis covers the period 1997–2019. The reanalysis based on the Nemo-Nordic model runs from 2017 to 2021. In determining the connections between the tide gauges, we made use of the heuristic search algorithm developed by Afrasteh et al. (2021). This algorithm identifies the set of hydrodynamic leveling connections that provide the lowest median of the propagated height standard deviations.

The impact of model-based hydrodynamic leveling on the EVRF was determined by

comparing the solutions with the so-called 'UELN-only solution'. Except for 3 HBMs, this solution is identical to the EVRF2019 (release September 2020). The differences include a mean difference of 6.527 kgal·mm at two British HBMs included in the EVRF2019.

Based on the comparison with the UELN-only solution, the impact of model-based hydrodynamic leveling in Experiment I has a long wave character and is significant for France and Great Britain where the differences range between -96.7 and 10.5 kgal·mm, and 23.8 and 444.2 kgal·mm, respectively. The range of differences for the other countries is lower than 10 kgal·mm. In Experiment II, the large-scale pattern is the same. However, the impact for both France and Great Britain is lower.

A comparison with normal heights obtained by differencing GNSS and EGG2015 quasi-geoid heights showed that in Experiment I for both France and Great Britain the systematic differences between both height data sets decreased. In Great Britain, the south-north slope disappeared almost completely after adding model-based hydrodynamic leveling data. In France, the decrease is mainly visible in the southwest.

Based on the results, we conclude that model-based hydrodynamic leveling has a positive impact on the EVRF. At the same time, and not unexpectedly, this impact does depend on the number and locations of tide gauges available for establishing the connections.

The results presented here show a provisional impact. Indeed, as indicated in Sect. 4.1, this study is subject to a number of limitations. First, in connecting the tide gauges to the UELN we mostly relied on potential differences computed from adjusted heights rather than real levelings. Second, we made no attempt to reduce the potential differences to the reference epoch adopted for the EVRS (i.e., epoch 2000.0). Third, we made no attempt to improve the performance of the hydrodynamic models in representing the local MWL at the tide gauge locations. In addition, to reduce the computational load no connections were allowed between tide gauges located i) within the same country, or ii) in neighboring countries for which the number of spirit leveling connections between the countries is larger than one. In particular for countries having a long coastline, it could be useful to relax these constraints. Finally, no connections have been added in the Mediterranean and Black Seas. Given the quality of the UELN data in the southern European countries, this would be a valuable addition.

All these aspects have to be addressed to achieve a rigorous implementation of model-based hydrodynamic leveling. However, this requires collaboration with experts who know which processes contribute to the (summer) mean water level variability at the tide gauge locations, and who have access to adequate models and datasets to model this variability. But also experts who can assess any vertical movement of the tide gauges and who can connect the tide gauges to the UELN. Undoubtedly, some things will not be realized overnight. At the same time, the prospect of having a technique that can connect islands and offshore platforms to the mainland height system and suppresses systematic errors in leveling networks makes the effort worthwhile.

Future research could also focus on using GNSS/leveling to establish (additional) connections between HBMs (separated by large water bodies). In addition to the required GNSS data, doing so requires the *full* noise variance-covariance matrix of the quasi-geoid model used. Obtaining the latter is probably the biggest challenge. Even if both are available, the question remains how to validate the results, as there is no dataset left to validate the solution.

BIBLIOGRAPHY

Afrasteh, Y., Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J., Snellen, M., & Zijl, F. (2021). The potential impact of hydrodynamic leveling on the quality of the European vertical reference frame. *Journal of Geodesy*, *95*(8). https://doi.org/10.1007/s00190-021-01543-3

Afrasteh, Y., Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J., Snellen, M., & Zijl, F. (2023). An empirical noise model for the benefit of model-based hydrodynamic leveling. *Journal of Geodesy*, *97*(1), 1–21.

Amjadiparvar, B., Rangelova, E., & Sideris, M. G. (2016). The GBVP approach for vertical datum unification: recent results in North America. *Journal of Geodesy*, *90*(1), 45–63. https://doi.org/10.1007/s00190-015-0855-8

Amos, M. J., & Featherstone, W. E. (2009). Unification of New Zealand's local vertical datums: iterative gravimetric quasigeoid computations. *Journal of Geodesy*, *83*(1), 57–68. https://doi.org/10.1007/s00190-008-0232-y

Axell, L., & Liu, Y. (2016). Application of 3-D ensemble variational data assimilation to a Baltic Sea reanalysis 1989–2013. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 24220. https://doi.org/10.3402/tellusa.v68.24220

Bjerhammar, A. (1985). On a relativistic geodesy. *Bulletin géodésique*, *59*(3), 207–220. https://doi.org/10.1007/BF02520327

Bossler, J. D. (1984). *Standards and specifications for geodetic control networks*. National Geodetic Information Branch, NOAA, FGCC.

Codiga, D. L. (2020). UTide Unified Tidal Analysis and Prediction Functions.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman; Hall.

Denker, H. (2013). Regional Gravity Field Modeling: Theory and Practical Results. In G. Xu (Ed.), *Sciences of Geodesy - II: Innovations and Future Developments* (pp. 185–291). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-28000-9_5

Denker, H., Timmen, L., Voigt, C., Weyers, S., Peik, E., Margolis, H. S., Delva, P., Wolf, P., & Petit, G. (2018). Geodetic methods to determine the relativistic redshift at the level of 10^{-18} in the context of international timescales: A review and practical results. *Journal of Geodesy*, *92*(5), 487–516.

Denker, H. (2015). A new European gravimetric (quasi) geoid EGG2015. *26th IUGG General Assembly, June22–July2*.

Duquenne, F., Coulomb, A., & LÉcu, F. (2015). *French Approach to Modernize Its Vertical Reference System* [FIG Working Week 2015, Sofia, Bulgaria, 17–21 May 2015, http://www.fig.net/resources/proceedings/fig_proceedings/fig2015/papers/ts06g/TS06G_duquenne_coulomb_et_al_7676.pdf].

Federal Agency for Cartography and Geodesy. (2022a). EVRF2019 [Accessed: November 16, 2022].

Federal Agency for Cartography and Geodesy. (2022b). Height datum relations [Accessed: November 16, 2022].

Filmer, M. S., Featherstone, W. E., & Kuhn, M. (2010). The effect of egm2008-based normal, normal-orthometric and helmert orthometric height systems on the australian levelling network. *Journal of Geodesy*, *84*(8), 501–513. https://doi.org/10.1007/s00190-010-0388-0

Gerlach, C., & Rummel, R. (2013). Global height system unification with goce: A simulation study on the indirect bias term in the gbvp approach. *Journal of Geodesy*, *87*(1), 57–67. https://doi.org/10.1007/s00190-012-0579-y

Heck, B., & Rummel, R. (1990). Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In H. Sünkel & T. Baker (Eds.), *Sea surface topography and the geoid* (pp. 116–128). Springer New York.

Hipkin, R., Haines, K., Beggan, C., Bingley, R., Hernandez, F., Holt, J., & Baker, T. (2004). The geoid EDIN2000 and mean sea surface topography around the British Isles. *Geophys. J. Int.*, *157*, 565–577. https://doi.org/10.1111/j.1365-246X.2004.01989.x

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., Liu, Y., Pemberton, P., Schimanke, S., Andersson, H., et al. (2019). Nemo-Nordic 1.0: a NEMO-based ocean model for the Baltic and North seas–research and operational applications. *Geoscientific Model Development*, *12*(1), 363–386. https://doi.org/10.5194/gmd-12-363-2019

Hordoir, R., Axell, L., Löptien, U., Dietze, H., & Kuznetsov, I. (2015). Influence of sea level rise on the dynamics of salt inflows in the Baltic Sea. *Journal of Geophysical Research: Oceans*, *120*(10), 6653–6668. https://doi.org/https://doi.org/10.1002/2014JC010642

Ihde, J., Adam, J., Gurtner, W., Harsson, B. G., Sacher, M., Schlüter, W., & Wöppelmann, G. (2000). The height solution of the European vertical reference network (EUVN). *Veröff. Bayer. Komm. für die Internat. Erdmessung, Astronom. Geod. Arb*, *61*.

Ihde, J., Augath, W., & Sacher, M. (2002). The Vertical Reference System for Europe. In *International association of geodesy symposia* (pp. 345–350). Springer, Berlin. https://doi.org/10.1007/978-3-662-04683-8_64

Ihde, J., Mäkinen, J., & Sacher, M. (2008). Conventions for the definition and realization of a European Vertical Reference System (EVRS) – EVRS Conventions 2007. *EVRS Conv*, 1–10.

Jahanmard, V., Delpeche-Ellmann, N., & Ellmann, A. (2021). Realistic dynamic topography through coupling geoid and hydrodynamic models of the Baltic Sea. *Continental Shelf Research*, *222*, 104421. https://doi.org/https://doi.org/10.1016/j.csr.2021.104421

Jahanmard, V., Delpeche-Ellmann, N., & Ellmann, A. (2022). Towards realistic dynamic topography from coast to offshore by incorporating hydrodynamic and geoid models. *Ocean Modelling*, *180*, 102124. https://doi.org/https://doi.org/10.1016/j.ocemod.2022.102124

Kenyeres, A., Sacher, M., Ihde, J., Denker, H., & Marti, U. (2010). *EUVN densification action – final report* (tech. rep.) [https://evrs.bkg.bund.de/SharedDocs/Downloads/

EVRS/EN/Publications/EUVN-DA_FinalReport.pdf?__blob=publicationFile&v=1, accessed October 6, 2022]. EUVN_DA Working Group.

Madec, G., Bourdallé-Badie, R., Bouttier, P. A., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., et al. (2017). Nemo ocean engine.

Mayer-Gürr, T. (2015). The combined satellite gravity field model GOCO05s. *EGU general assembly conference abstracts*, 12364.

Moritz, H. (2000). Geodetic Reference System 1980 [10.1007/s001900050278]. *J. Geod.*, *74*(1), 128–162.

Penna, N. T., Featherstone, W. E., Gazeaux, J., & Bingham, R. J. (2013). The apparent British sea slope is caused by systematic errors in the levelling-based vertical datum. *Geophysical Journal International*, *194*(2), 772–786. https://doi.org/10.1093/gji/ggt161

Rao, C. (1971). Estimation of variance and covariance components—minque theory. *Journal of Multivariate Analysis*, *1*(3), 257–275. https://doi.org/https://doi.org/10.1016/0047-259X(71)90001-7

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. https://doi.org/10.1080/01621459.1993.10476408

Rülke, A., Liebsch, G., Sacher, M., Schäfer, U., Schirmer, U., & Ihde, J. (2012). Unification of European height system realizations. *J. Geod. Sci.*, *2*(4), 343–354. https://doi.org/10.2478/v10156-011-0048-1

Rummel, R., & Teunissen, P. (1988). Height datum definition, height datum connection and the role of the geodetic boundary value problem. *Bulletin Géodésique*, *62*(4), 477–498. https://doi.org/10.1007/bf02520239

Sacher, M., & Liebsch, G. (2019). EVRF2019 as new realization of EVRS.

Sánchez, L., & Sideris, M. . (2017). Vertical datum unification for the International Height Reference System (IHRS). *Geophysical Journal International*, *209*(2), 570–586. https://doi.org/10.1093/gji/ggx025

Slobbe, D. C., Klees, R., Farahani, H. H., Huisman, L., Alberts, B., Voet, P., & De Doncker, F. (2019). The Impact of Noise in a GRACE/GOCE Global Gravity Model on a Local Quasi-Geoid. *Journal of Geophysical Research: Solid Earth*, *124*(3), 3219–3237. https://doi.org/10.1029/2018jb016470

Slobbe, D. C., Klees, R., Verlaan, M., Zijl, F., Alberts, B., & Farahani, H. H. (2018). Height system connection between island and mainland using a hydrodynamic model: a case study connecting the Dutch Wadden islands to the Amsterdam ordnance datum (NAP). *J. Geod.*, *92*(12), 1439–1456. https://doi.org/10.1007/s00190-018-1133-3

Tonani, M., & Ascione, I. (2021). Product User Manual: Ocean Physical and Biogeochemical reanalysis NWSHELF_MULTIYEAR_PHY_004_009 NWSHELF_MULTIYEAR_BGC_004_011.

Vanicek, P., Castle, R. O., & Balazs, E. I. (1980). Geodetic leveling and its applications. *Reviews of Geophysics*, *18*(2), 505–524. https://doi.org/https://doi.org/10.1029/RG018i002p00505

Vestøl, O., Ågren, J., Steffen, H., Kierulf, H., & Tarasov, L. (2019). NKG2016LU: a new land uplift model for Fennoscandia and the Baltic Region. *Journal of Geodesy*, *93*(9), 1759–1779. https://doi.org/10.1007/s00190-019-01280-8

Wang, Y. M., Saleh, J., Li, X., & Roman, D. R. (2012). The US Gravimetric Geoid of 2009 (USGG2009): model development and evaluation. *Journal of Geodesy*, *86*(3), 165–180. https://doi.org/10.1007/s00190-011-0506-7

Wu, H., Müller, J., & Lämmerzahl, C. (2018). Clock networks for height system unification: a simulation study. *Geophysical Journal International*, *216*(3), 1594–1607. https://doi.org/10.1093/gji/ggy508

Ziebart, M. K., Iliffe, J. C., Forsberg, R., & Strykowski, G. (2008). Convergence of the uk osgm05 grace-based geoid and the uk fundamental benchmark network. *Journal of Geophysical Research: Solid Earth*, *113*(B12).

Zijl, F., Laan, S., & Groenenboom, J. (2020). Development of a 3D model for the NW European Shelf (3D DCSM-FM).

5

CONCLUSIONS AND RECOMMENDATIONS

In this chapter, we answer the three research questions defined in Section 1.4, and evaluate whether the main research objective defined in Section 1.4 has been achieved. Thereafter, we provide some recommendations for future research.

5.1. CONCLUSIONS

The first research question defined in Section 1.4 was:

- ***What is the potential impact of adding model-based hydrodynamic leveling connections to the UELN dataset on the quality of the EVRF and which connections, and hence tide gauges, are most profitable in terms of impact?***

The answer to this research question is the main topic of Chapter 2. To assess the impact, a series of geodetic network analyses were conducted. It was assumed that hydrodynamic models are available covering either all European Seas surrounding the European mainland or parts of it. The impact was quantified in terms of precision and reliability. The variance information for the spirit leveling/gravimetry data was obtained from the latest UELN adjustment, while to get the variance-covariance matrix for the model-based hydrodynamic leveling data, each hydrodynamic model was assumed to provide the required coastal SMWLs with uniform precision. In Chapter 3 this latter assumption was abandoned and the influence of ignoring noise correlations on the estimated impact was examined.

To identify which hydrodynamic leveling connections contribute most to the quality of the EVRF, a heuristic search algorithm was implemented. In each step, the algorithm identifies from all remaining possible connections the ones that result in the lowest median propagated standard deviation (SD) of the adjusted heights. The algorithm stops when no more connections are remaining. Our motivation to use the median of the propagated height SDs instead of the mean is that the propagated SDs of all height markers are not normally distributed. Using the mean does, however, hardly impacts the main findings of the study (experiments not included in this thesis).

The results highlighted a significant improvement in the precision of the EVRF when adding model-based hydrodynamic leveling connections to the UELN data, especially if

tide gauges distributed over entire Europe are used. However, it is not necessary to use a model that covers all European waters; instead, we can use a combination of models, each designed specifically for a particular region. Using hydrodynamic leveling connections with a noise SD between 1 cm and 5 cm resulted in a 29% (noise SD of 5 cm) to 48% (noise SD of 1 cm) improvement in the median propagated SD of all adjusted heights. This indicates that even if the precision is low, the improvement is still significant. However, the higher the noise SD the more connections had to be added to achieve a certain level of improvement. The impact differs per country. Except for the countries around the Black Sea, the coastal countries benefit the most. In terms of reliability, the impact is mainly limited to coastal regions. Only in Great Britain the redundancy numbers increased almost throughout the whole country.

Restricting the use of tide gauges involved in establishing hydrodynamic leveling connections to specific sea basin(s) limits the impact. If hydrodynamic leveling connections are only added between tide gauges in the northwest European continental shelf, the overall impact reduces from 38% to 25% (assuming a noise SD of 3 cm). The change in impact per country differs strongly; for Great Britain the improvement remains about 60%.

The tide gauges that are involved in the most beneficial connections in terms of impact are located in the countries with the greatest number of height markers. However, the impact hardly depends on the geographic location of the tide gauges within a country. This was demonstrated in an experiment where the tide gauges responsible for 75% of the maximum improvement were excluded. The search algorithm was then run on the remaining tide gauges to identify a new set of optimal hydrodynamic leveling connections, and the impact did not decrease significantly.

Ignoring the noise correlations in the model-derived coastal SMWLs results in an overestimation of the total quality impact by 7% (one summer averaging period) and 8% (three summers averaging period). This aspect was studied in Chapter 3. It must be noted that impact in terms of precision reported in Chapters 2 and 3 only reflects the formal precision. Other factors, such as systematic errors in the spirit leveling data, vertical land motion, and long-term sea-level variations, can reduce the *real* impact.

The second research question defined in Section 1.4 was:

- ***What is the noise model for coastal summer mean water levels obtained by state-of-the-art hydrodynamic models covering the northwest European continental shelf?***

Combining geopotential differences from spirit leveling/gravimetry and model-based hydrodynamic leveling data requires a proper noise model for both datasets. For the latter dataset, no such model was available. Its development is the main topic of Chapter 3. Here, an empirical approach was used based on calculating an average empirical covariance function from the differences between tide gauge- and model-derived SMWLs. Three models have been used: the 3D DCSM-FM model, the DCSMv6-ZUNOv4 model and the Forecasting Ocean Assimilation Model 7 km Atlantic Margin model (AMM7) model. All three datasets covered the period 1997–2019. In order to have multiple realizations of the noise, an averaging period of one summer (defined as the period from May to September) was chosen. The impact of an averaging period of two and three summers was examined for the 3D DCSM-FM model.

The computed empirical covariance functions revealed that the noise in the model-derived coastal SMWLs is indeed spatially correlated. The level of correlation, however, varies among the models. Additionally, all models show a significant discontinuity at the origin, indicating a nugget effect. This points to random errors in the observation- and model-derived SMWLs and signals at short spatial scales that cannot be resolved from the data or by the hydrodynamic model. Of all three models used in this study, 3D DCSM-FM had the lowest nugget and variance values. The empirical noise covariance functions obtained for this model using an averaging period of two and three summers were not significantly different from those obtained using one summer. We observed a slight decrease in the variance values and nugget effects. A significant contributor to the obtained noise covariance functions are errors in the vertical referencing of the tide gauges. For 3D DCSM-FM, these errors explain between 30% and 50% of the variance. For the other models, the contribution is lower.

Chapter 3, also presents the spatiotemporal performance of the three hydrodynamic models in representing the coastal SMWLs computed over one summer period. The temporal performances were assessed for six subregions by the cumulative distribution functions of both the mean absolute error (MAE) and the modified Kling-Gupta efficiency (mKGE) metric. The spatial performances were assessed using both the MAE and the modified spatial efficiency (mSPAEF) metric. The modifications to the Kling-Gupta efficiency and spatial efficiency metrics aimed to eliminate the contribution of the time-dependent but space-independent model biases (these cancel out when computing the model-based hydrodynamic leveling data). For the same reason, we applied a time-dependent bias correction to the SMWLs computed from a subset of the tide gauges for which a full-time series is available.

The results indicate that the performance of the three models varies both spatially and temporally. Regarding the temporal performance, no single model consistently outperforms the others across all regions; different models score better in different subregions. Some subregions show no significant performance differences. Also in all subregions there are more or less tide gauges where models do not show good agreement with the observations. Biases in the derived SMWL time series were identified as the primary cause for the poor agreement between the models and observations, along with the models' inability to accurately represent coastal SMWLs at specific tide gauge locations. The variations in performance can be partially explained by considering the intended applications and the areas for which the hydrodynamic models were developed. Spatial performances display notable fluctuations over time, with all hydrodynamic models demonstrating improved performance between 2004 and 2011. Throughout most of the analyzed period, the 3D DCSM-FM model consistently outperformed the others in terms of MAE and mSPAEF. However, explaining the underlying reasons for these performance differences was outside the scope of this study.

The empirical noise covariance functions for the deep and shelf waters in the target area calculated from TOPEX/Jason satellite altimetry data, turned out to be significantly different from each other and from the function computed for coastal SMWLs. This is in line with oceanographic expectations, namely that the dynamics along the coast are more complex than in deep and shelf waters. Hence, altimeter data have limited value in obtaining a noise model for the coastal SMWLs.

The last research question defined in Section 1.4 was:

- ***What is the impact of adding real model-based hydrodynamic leveling data between tide gauges in the Baltic Sea and the northwest European continental shelf on the EVRF?***

The analysis in Chapter 2 presents the potential impact of model-based hydrodynamic leveling connections on quality of the EVRF in terms of precision and reliability, but does not include the impact on the estimated geopotential numbers / heights. This was the topic of Chapter 4. It presents the first attempt to compute a realization of the EVRS based on the combination of geopotential differences from spirit leveling/gravimetry and model-based hydrodynamic leveling. The latter are computed using coastal SMWLs averaged over three summers. For the Baltic Sea and the northwest European continental shelf, these are obtained from reanalyses with the Nemo-Nordic (covering the years 2017 to 2021) and 3D DCSM-FM (covering the years 1997 to 2019) models, respectively. All involved tide gauges were connected to the nearest UELN HBM using the adjusted heights obtained in the national height system realizations.

Based on the spatial distribution of the tide gauges along the coastlines, we identified the period 2004–2006 and 2017–2019 as the most preferred time span for the 3D DCSM-FM and Nemo-Nordic dataset, respectively. For this time span, we have respectively 76 and 50 tide gauges available allowing to establish a maximum of 124 hydrodynamic leveling connections. As we lacked the required long-time series of observation- and model-derived water levels to develop an empirical noise model specific to the Nemo-Nordic dataset, we used the noise model developed for the AMM7 model. The only motivation we have for using this noise model is that the AMM7 hydrodynamic model, similar to the Nemo-Nordic model, relies on the Nucleus for European Modeling of the Ocean (NEMO) model code. Moreover, we applied variance component estimation.

Based on the comparison with the 'UELN-only solution', i.e., a realization of the EVRS obtained using UELN data only, the impact of model-based hydrodynamic leveling is significant for France and Great Britain. The differences range between -96.7 kgal·mm and 10.5 kgal·mm, and 23.8 kgal·mm and 444.2 kgal·mm, for France and Great Britain, respectively. No significant change is observed in the Baltic countries. This may be explained by the high quality of the UELN data in the Baltic Sea countries.

A comparison with normal heights obtained by differencing GNSS and EGG2015 quasi-geoid heights showed that adding the hydrodynamic leveling connections lowers the systematic differences between both height data sets in France and Great Britain. In Great Britain, the south-north slope disappeared almost completely. In France, the decrease is mainly visible in the southwest.

5.2. REFLECTION ON THE MAIN RESEARCH OBJECTIVE

Motivated to achieve a 1 cm accurate VRF covering the entire 3D DCSM-FM hydrodynamic model domain, the main objective of this thesis was an assessment of the potential of including model-based hydrodynamic leveling data in the realization of the EVRS.

One of the main achievements of this thesis was enlarging the coverage of the EVRF using model-derived hydrodynamic leveling connections. In the latest realization of the

EVRS, the EVRF2019, only the datum shift between the Ordnance Datum Newlyn (ODN) and the EVRF was determined by levelings through the Channel Tunnel. By means of model-based hydrodynamic leveling we added multiple other connections between the British leveling network and the one on the European mainland. This allowed to include the full British leveling dataset in the adjustment, and resulted in a disappearance of the south-north tilt. This outcome suggests that the method can be applied to connect other islands (including Ireland) and offshore platforms to the EVRS.

Furthermore, the study demonstrated that combining geopotential differences from spirit leveling/gravimetry and model-based hydrodynamic leveling enhances the precision and reliability of the EVRF. In terms of precision, the coastal countries with poor-quality spirit leveling networks and those with a poor connection to the rest of the UELN network benefit the most. In terms of reliability, an improvement is observed for the network along the coastlines. To apply model-based hydrodynamic leveling, it is not required that the models always and everywhere have good performance; we can omit tide gauges/time periods where/in which the models have poorer performance. In Chapter 2, it was shown that adding already a small number of hydrodynamic leveling connections to the UELN dataset had a significant impact on the quality of the EVRS. The results in Chapter 3 showed that in each subregion, at least some tide gauges were available where the model had the necessary skills. Moreover, the results in Chapter 3 indicated that even using a short averaging period for computing the SMWLs (one year), the impact was significant. At the same time, the fact that the models have different performances in different regions suggested that further improvements are possible. That is, one could combine the good elements of all models in one, new model.

The study also demonstrated that adding model-based hydrodynamic leveling data has great potential in suppressing systematic errors in the leveling networks. A good example is Great Britain. As noted above, by adding model-based hydrodynamic leveling connections we got rid of the tilt.

Whether or not the addition of model-based hydrodynamic leveling to the UELN dataset will allow a 1-cm accurate realization of an EVRF remains to be seen. In any case, the results presented in this thesis encourage further development of the technique. In the next section we will make some recommendations for this.

5.3. RECOMMENDATIONS

While the work presented in this thesis revealed a significant potential of using model-based hydrodynamic leveling in height system realization, the results are provisional. The most important recommendations are listed below:

- As mentioned in Chapter 4, for a rigorous implementation of model-based hydrodynamic leveling we lacked (i) tide gauge records with respect to the tide gauge zero and (ii) the spirit leveling/gravimetry-derived potential difference between the tide gauge zeros and the UELN benchmarks. Regarding the first, such information must be available in the metadata of tide gauges, and it is just a matter of collecting it from the different authorities in the participating European countries. The second requires time and money and is best tackled in the framework of a European project.

- The model-based hydrodynamic leveling data have to be reduced to the reference

epoch adopted for the EVRS. Doing so requires information about the long-term sea level variability and vertical land motion at the locations of the involved tide gauges. To apply this reduction, future research should focus on tide gauges at which GNSS measurements are available.

- In the current study, hydrodynamic leveling connections were only established in the Baltic Sea and the northwest European continental shelf. Given the quality of the UELN data in some southern European countries, the addition of hydrodynamic leveling connections in the Mediterranean Sea and the Black Sea has definitely great value.

- Regarding the hydrodynamic leveling connections in the Baltic area, we lacked the long time series of the simulated and observed water levels to compute a proper noise model for the model-derived coastal SMWLs. Future research should focus on obtaining such a model by extending the time series.

- As explained in Chapter 3, this thesis uses tide gauges records for developing a noise model for the model-derived coastal SMWLs. While this dataset provides sufficient temporal resolution, it lacks a homogeneous spatial distribution throughout the coastal waters covered by the hydrodynamic model domain. It is suggested to investigate the possibility of using dedicated coastal altimetry products for deriving a noise model for coastal SMWLs.

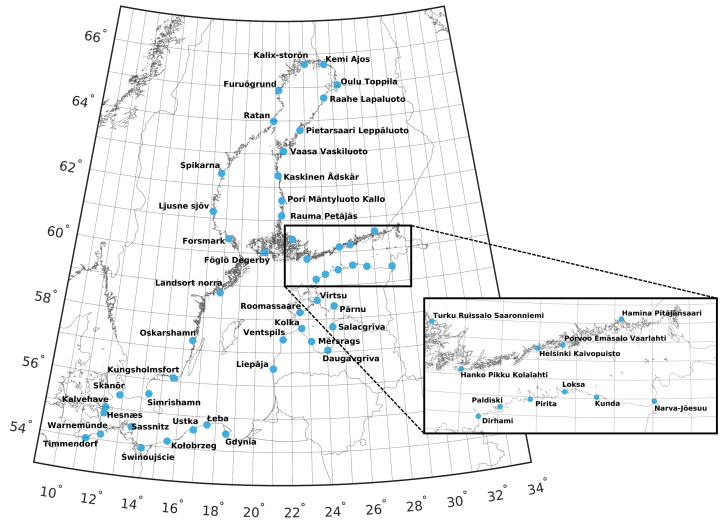
- The results in Chapter 3 showed that there is no hydrodynamic model that performs best everywhere; different models score better in different subregions. This indicates that there is room for improving the performance of hydrodynamic models. Tailoring the models with the specific aim to improve their performance in representing the coastal SMWLs is an interesting topic for future work.

- The results obtained in this thesis encourage to study the potential of model-based hydrodynamic leveling in realizing other regional VRSs. It would also be worth investigating to what extent it can contribute to the realization of the international height reference system (Sánchez et al., 2021). Note that the preliminary IHRF, completed in April 2019, comprises 170 stations of which 26 were co-located by tide gauges (https://ihrs.dgfi.tum.de/en/working-groups/012-ihrs-realization/). However, a prerequisite for a successful implementation is a close inter-disciplinary cooperation between geodesy and oceanography.

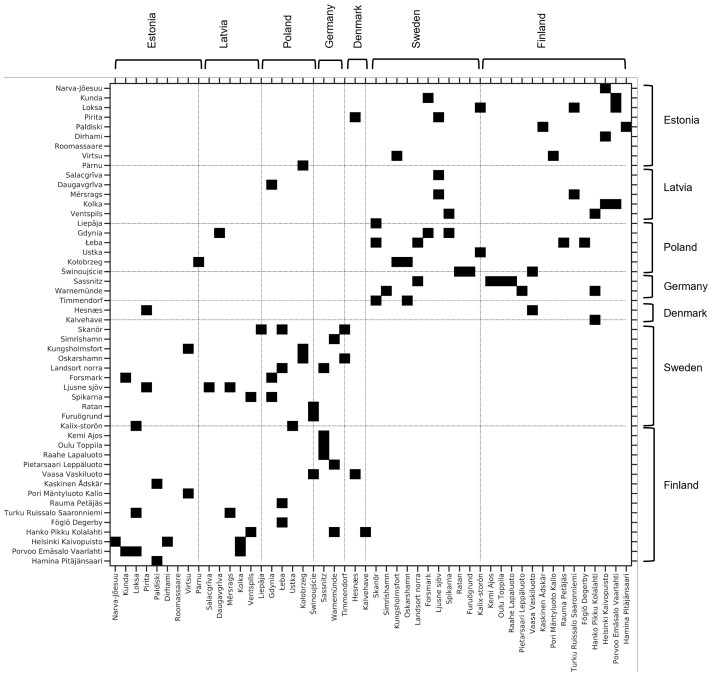
A

THE HYDRODYNAMIC LEVELING CONNECTIONS BEING ADDED

Figs. A.1b and A.2b show the model-based hydrodynamic leveling connections included in, respectively, the Nemo-Nordic and 3D DCSM-FM hydrodynamic leveling datasets. These connections are the ones being identified by our heuristic search algorithm described in Sect. 4.2.2. The maps showing the corresponding locations of the tide gauges are included in Figs. A.1a and A.2a.

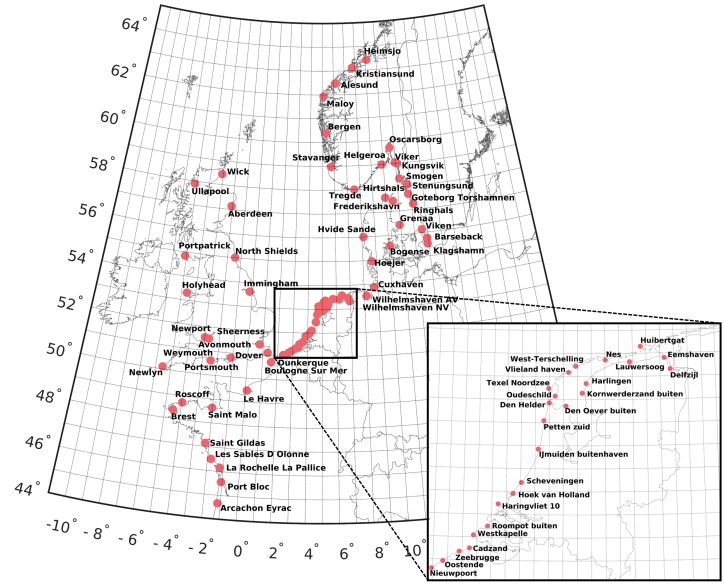


(a)

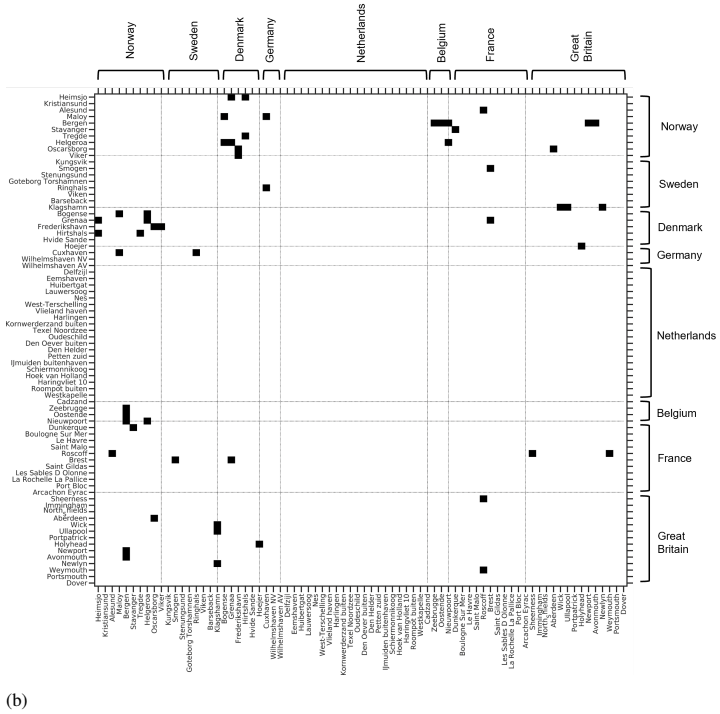


(b)

Figure A.1: Location of the tide gauges(a), and the connectivity matrix showing which model-based hydrodynamic leveling connections are included in the Nemo-Nordic dataset(b).



(a)



(b)

Figure A.2: Location of the tide gauges(a), and the connectivity matrix showing which model-based hydrodynamic leveling connections are included in the 3D DCSM-FM dataset(b).

ACKNOWLEDGEMENTS

Looking back on this long journey, I feel really lucky having people who supported, inspired, motivated, encouraged, and trusted me. I would like to take this opportunity to thank all of you, without whom I wouldn't have made it to the end.

First, I would like to thank my daily supervisor and copromoter. **Cornelis**, you were more than just a "boss" to me. Your positive energy, constant support, patience, and guidance were always strong motivators for me to continue my PhD journey. I have always admired your critical insight and attention to details. I have learned a lot from all our challenging discussions and will certainly miss (already miss!) working with you. **Martin**, it was a true pleasure to have you as a daily supervisor and promoter in my PhD project. Our weekly meetings were a source of inspiration, and I was always impressed by your ability to quickly grasp problems and always provide valuable ideas to enhance the work. I am deeply grateful to my promoter, **Roland**, who has never spared his support during my PhD. I appreciate your quick response and valuable input on my research and papers.

I would like to extend my gratitude to the committee members, **Ole Andersen**, **Ramon Hanssen**, **Laura Sanchez**, **George Vergos**, and **Peter Teunissen**. Many thanks for devoting your time to review my work.

Many thanks to all my colleagues from the Versatile Hydrodynamics project for all the collaborations and insightful discussions during our meetings. A big thank you to **Firmijn Zijl** who was always there for all my 3D DCSM-FM related questions. Great appreciation to **Martina Sacher** for her constant support related to UELN data and for being a warm and friendly host during my visit to BKG.

To all the people in the Geodesy and Remote Sensing department, it was a great pleasure to be part of the group. Having wonderful times with many friends and colleagues made my PhD life enjoyable. **Inger**, I really enjoyed sharing an office with you and all the fun activities we had outside of work. I believe I still owe you a white shirt since yours got ruined when you helped me change my bike chain in the very first days of my time in Delft. Thank you, **Henrique**, for introducing me to wakeboarding. It might not have been a success, but I definitely enjoyed it. **Frithjof**, it's nice to see you grow as a successful dad and scientist. I hope to receive your cheesy Christmas cards next year too. **Amey**, thanks for sharing the beautiful moments from your wedding.

I would like to thank all my Alten colleagues from the Basin Modeling project for their support and understanding while I was busy with the last steps to complete my PhD.

AmirMasoud, thank you for being a friend whom I could always count on for help and support. **Elahe**, when I moved to the Netherlands, knowing you are close by was heartwarming, thanks for all the support. **Khatereh**, thanks for your considerate and kind presence in all my important moments, from the master of ceremony to being my paranymph. **Parsa**, everyone needs a caring and fun friend like you. Thanks for always being there. **Stavros** and **Valentina**, the Greek trip and attending your wedding were among the highlights of my PhD years. Thanks for that and all the nice afterward gatherings. A big

appreciation to all my Delft friends, **Leila**, **Hoda**, **Zahra**, **Parviz**, **Farinaz** and **Mousa**. Thanks for all the wonderful moments we shared.

Finally, to my family and friends back home: **Arefeh**, thanks for your constant support and kindness even when we were thousands of kilometers away. **Sabah**, I'm not sure how you do it, but you are the one who, in tough moments, has convinced me that everything will be right. **Maman**, **Baba**, **Hossein**, **Somayeh**, and **Hassan**, you are my greatest treasures in life. Thanks for your encouragement, endless support, and unreserved love throughout this journey. Among all the achievements that the PhD life has brought for me, **Ramin**, you are definitely on top of them all. When I was struggling with a problem, you were the first one I would go to, and almost every time you inspired me with a nice idea. My warmest acknowledgment goes to you, for all the love and support even when you were dealing with your own PhD stuff.

CURRICULUM VITÆ

Yosra AFRASTEH

19-07-1991	Born in Tehran, Iran.
2009–2012	**Technical Associate Degree in Civil Engineering Surveying** Institute of Surveying and Mapping of National Geographical Organization, Iran Thesis title:
2012–2014	**Bachelor of Science in Surveying Engineering** Shahid Rajaee University, Iran
2015–2018	**Master of Science in Geodesy** University of Tehran, Iran Thesis title: Research internship at Univesity of New Brunswick (Canada)
2018–2022	**Doctoral Candidate** Delft University of Technology, The Netherlands

List of Publications

- **Afrasteh, Y.**, Slobbe, D. C., Verlaan, M., Sacher, V., Jahanmard, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J. Snellen, M. and Zijl, F. (2023). Realizing the European Vertical Reference System using model-based hydrodynamic leveling data. *Journal of Geodesy*, 97:1–19. https://doi.org/10.1007/s00190-023-01778-2.

- **Afrasteh, Y.**, Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J. Snellen, M. and Zijl, F. (2022). An empirical noise model for the benefit of model-based hydrodynamic leveling. *Journal of Geodesy*, 97:1–21. https://doi.org/10.1007%2Fs00190-022-01694-x.

- **Afrasteh, Y.**, Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J. Snellen, M. and Zijl, F. (2021). The potential impact of hydrodynamic leveling on the quality of the European vertical reference frame. *Journal of Geodesy*,95:1-18. https://doi.org/10.1007%2Fs00190-021-01543-3.

List of Presentations

Oral presentations

- Gravity, Geoid, and Height Systems (GGHS) symposium, September 2022, Austin, United States.
- European Geosciences Union (EGU) General Assembly, May 2022, Vienna, Austria.
- Hotine Marussi symposium, June 2022, Milan, Italy.
- Reference Frame Sub-Commission for Europe (EUREF) symposium, June 2022, Online.
- Nederlands Centrum voor Geodesie en Geo-informatica (NCG) symposium, April 2022, Wageningen, The Netherlands.

Poster presentations

- GSE/GRS Research Day, April 2019, Delft, The Netherlands.

Online displays

- American Geophysical Union (AGU) FallMeeting, December 2021.
- European Geosciences Union (EGU) General Assembly, April 2021.
- European Geosciences Union (EGU) General Assembly, May 2020.