Computing volumes and surface areas including party walls for the 3DBAG data set

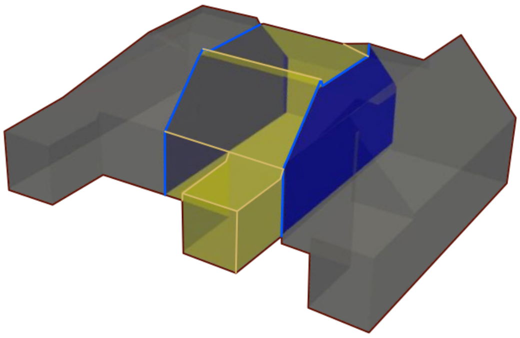
Agugiaro, Giorgio; Peters, Ravi; Stoter, Jantien; Dukai, Balázs

Publication date
2023

Document Version
Final published version

Citation (APA)
Agugiaro, G., Peters, R., Stoter, J., & Dukai, B. (2023). *Computing volumes and surface areas including party walls for the 3DBAG data set*. TU Delft/3DGI.

Computing volumes and surface areas including party walls for the 3DBAG data set



1 November 2023

Authors:
Dr. Giorgio Agugiaro
Email: g.agugiaro@tudelft.nl

Dr. Ravi Peters
Email: ravi.peters@3dgi.nl

Prof. dr Jantien Stoter
Email: j.stoter@tudelft.nl

MSc. Balázs Dukai
Email: balazs.dukai@3dgi.nl

Funded by Rijksdienst voor Ondernemend Nederland (RVO)

3D Geoinformation Group
Department of Urbanism
Faculty of Architecture and Built Environment
Delft University of Technology
Delft 2628BL - The Netherlands
Web: https://3d.bk.tudelft.nl/

3DGI
Van Boisotring 77
2722AB Zoetermeer
Nederland
Web: https://3dgi.nl

Table of contents

1 Introduction

This document provides a description of the project "Computing volumes and surface areas including party walls for the 3DBAG dataset" which has been carried out between the 3D Geoinformation group at TU Delft, 3DGI, and RVO in the timeframe between November 2022 and October 2023.
The goal of this project is to derive parameters from the 3DBAG that are relevant for energy consumption estimation, i.e. the enclosed volume of each building, as well as the party wall areas, the exterior wall areas, the ground floor areas and the roof areas.
As the detection of the party wall (i.e. the portion of the building shell that is shared between two buildings (BAG-panden)) is the most complex task to solve, specifically for a large data set, the main goal of the project is to define, evaluate and implement a methodology to compute the area extents of party walls between adjacent buildings from the 3DBAG data set. The 3DBAG dataset was first released in March 2021. A first revised version was released in September 2021 which we used for our analysis carried out during the first part of this project. The latest (5th) version has been released in October 2023 based on which we generated the final data for this project. In this last version AHN4 has been incorporated.
3DBAG is a country-wide dataset containing all buildings in the Netherlands, modelled in multiple LoDs, and based on the international standard CityGML. According to CityGML, a building can be modelled as a single-part unique object, or as an aggregation of building parts, each one having its own geometry. Additionally, each building is a geographical feature that can have several attributes (e.g. year of construction, number of storeys, etc.) and different geometries representing each one a specific Level of Detail (LoD). A graphical overview of the different LoDs according to CityGML v. 2.0 is given in Figure 1.

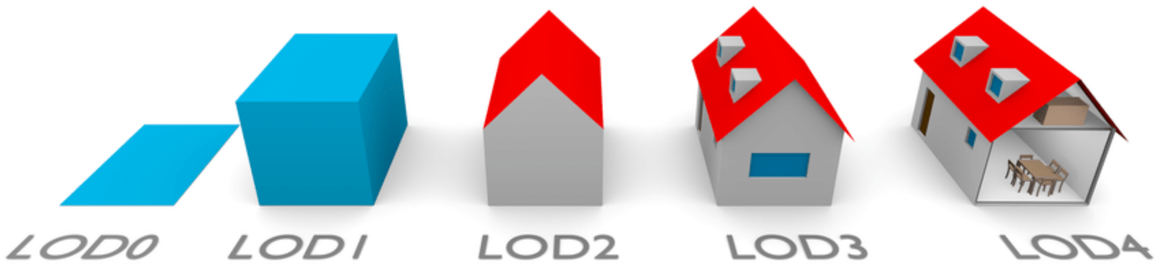


Figure 1. Levels of Detail for buildings according to CityGML v. 2.0. Image source: Biljecki et al. (2016)

In particular, the LoD2 allows differentiating between different thematic surfaces composing the building envelope. The geometries are semantically enriched and classified into GroundSurfaces, WallSurfaces and RoofSurfaces.
In the case of LoD2, RoofSurfaces represent the main planar surface(s) of the roofs. Smaller roof structures like chimneys and dormers are generally absent if their size is too small with regard to the surveyed data used for the 3D reconstruction process (e.g. the Lidar point cloud density).
The GroundSurfaces generally correspond to the planar extents of the roof surfaces projected onto the horizontal ground but they can also correspond to footprints. WallSurfaces connect vertically the Roof- and GroundSurfaces. This means that overhanging geometries (e.g. of roofs) lead to larger GroundSurfaces as in reality. But roof overhangs can also be represented. A graphical example can be seen comparing LoD2 and LoD3 in Figure 1. Finally, in LoD2 buildings there are no openings, i.e. neither doors nor windows.
More details about CityGML v. 2.0 and the modelling rules for the buildings can be found in the technical specifications of the standard published by the Open Geospatial Consortium.

1.1 The 3DBAG in a nutshell

The 3DBAG contains models of circa 9.5 million buildings in the Netherlands. All building models are generated by combining two country-wide open datasets: the BAG and AHN. The BAG (Basisregistratie Adressen en Gebouwen) contains, among the rest, a 2D polygon that represents the projected building extent as seen from above, the addresses and some other information. The height information in the 3DBAG comes from the AHN (Actueel Hoogtebestand Nederland), a Lidar-based point cloud dataset which is acquired on average every 4 years over the whole country. The 3DBAG dataset is available as open data and can be downloaded in different formats (e.g. CityJSON, GeoPackage and OBJ) from the web-based 3D viewer[1]. All buildings in LoD2.2 are modelled using thematic surfaces. The 3DBAG uses the refined LoD framework of Biljecki et al 2016, see Figure 2.

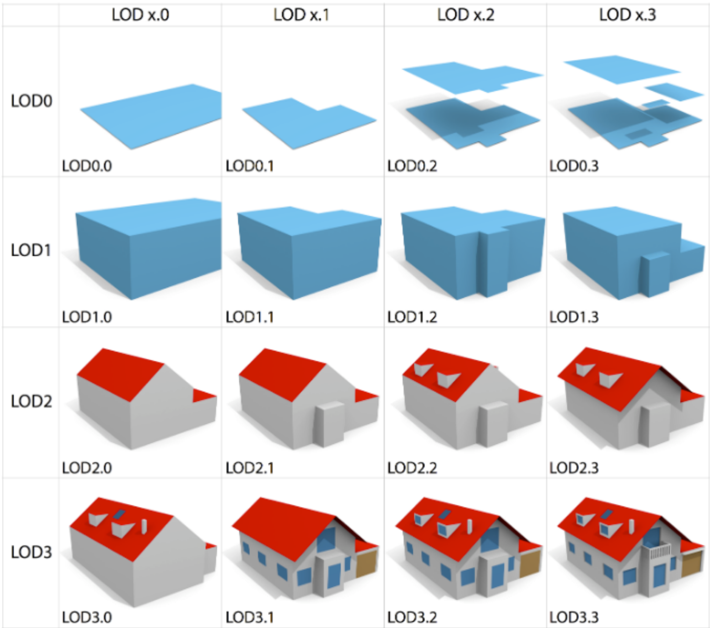


Figure 2. Example of improved LOD specification for 3D building model.[2]

When it comes to the accuracy of the reconstructed geometries, detailed information can be found in Dukai et al. (2021)[3].

Here only a summary of the main findings is presented. In general terms, only BAG geometries are used that temporally correspond to the acquisition time of the AHN. Additionally, BAG geometries representing underground features are filtered out, as they cannot be reconstructed due to the lack of AHN data. BAG geometries of above-ground buildings but spanning also over underground features (like an underground garage) are split into smaller polygons that correspond only to above-ground objects.

Several metrics are defined to assess the accuracy of the reconstructed geometries. When it comes to the RoofSurfaces, they primarily depend on the AHN data (point density and positional accuracy). In general terms, it can be stated that the RoofSurfaces polygons are no further away from the

[1] https://3dbag.nl/en/viewer

[2] F. Biljecki, H. Ledoux, J. Stoter. Computers, Environment and Urban Systems, 59: 25–37, 2016.

[3] Dukai, B., Peters, R., Vitalis, S., van Liempt, J., Stoter, J., 2021. Quality assessment of a nationwide data set containing automatically reconstructed 3D building models, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVI-4/W4-2021, 17–24, https://doi.org/10.5194/isprs-archives-XLVI-4-W4-2021-17-2021. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLVI-4-W4-2021/17/2021/isprs-archives-XLVI-4-W4-2021-17-2021.pdf

corresponding 3D Lidar points by a few cm (10 cm on average). Some roof structures (mostly dormers) are modelled, but this depends on their size and the Lidar point cloud density at that location. WallSurfaces connect the outer extents of the roof geometries with the respective GroundSurfaces, in this way overhangs are not modelled.
Further details on the generation of the 3DBAG can be found in Stoter et al. (2020)[4]. In this project, only LoD2.2 geometries will be used. All other geometries, from other LoDs, will not be considered.

1.2 Overview of this report

This report describes the results of the project which consisted of two main steps. In the first step we did several pre-tests (adjacency, volumes, area calculation for exterior and shared walls) using different methods on a test area to understand the impact of the required calculation methods when applied to the whole of The Netherlands, i.e. to see how robust the methods are and how sensitive they are for geometrical or topological errors or complexities. The results of the pre-analyses are reported in Chapter 2. Chapter 3 describes the second part of the project in which methods have been implemented to calculate for each 3DBAG building the enclosed volume, as well as the areas of the different types of surfaces for each building. Finally, chapter 4 contains conclusions of this project.

[4] Stoter, J., Peters, R., Commandeur, T., Dukai, B., Kumar, K., Ledoux, H., 2020. Automated reconstruction of 3D input data for noise simulation. Computers, Environment and Urban Systems, 80, 2020, pp. 101424.

2 Preliminary analyses

Starting from the Rijssen-Holten dataset, a set of preliminary analyses has been conducted in order to have a deeper understanding of the geometrical characteristics of the dataset, to investigate on potential sources of errors that may later lead to failure or additional errors in the computation of the party walls and the other envisioned parameters (volume, surface areas).
Therefore, a methodology has been set up. Several aspects have been investigated, and the procedure has been carried out at different scales, moving from the whole dataset, to the tile scale, and down to the single-building scale. The analysis has been carried out numerically, however criticalities have been further investigated also visually, in order to gather experience, collect ideas and provide suggestions for the successive steps.

Different sets of analyses have been carried out. The first one deals with the identification of issues when it comes to “adjacent” buildings. Two adjacent buildings are such if they have any portion of their respective WallSurfaces that share a certain amount of area. The geometry delimiting this area is called a party wall.
Due to the nature of the 3DBAG, and - more in general - the way LoD2 buildings are generated, the search for adjacent buildings can be reduced to the analysis of their footprint geometries. In other words, if the footprint polygons of two adjacent buildings are sharing a portion of at least one side, then the two buildings are considered adjacent.
The analysis for adjacency has covered two main existing methodologies developed by us. The results have been compared and criticalities collected. It must be noted that criticalities and errors are generally due to digitalization errors contained in the source data, namely the BAG dataset, which contains, among other things, the 2D geometries used as footprint during the generation of the 3DBAG.
A second set of analyses has covered the computation of the area of the building envelope elements (as a whole, or classified as RoofSurface, WallSurfaces, and GroundSurfaces). A third and last set of analyses has dealt with the computation of the gross volume enclosed by the building envelope surfaces.

For each analysis, results from different methods have been compared. The first method, called here “Method 1”, is the one experimentally applied as part of the 3DBAG project (at the time when this analysis was carried out). The second method, called “Method 2”, is based on algorithms and scripts developed in PostGIS, the database extension for the PostgreSQL database, which is used to store a copy of the 3DBAG. A detailed description of the two methods is beyond the scope of this report. Therefore, results will be presented just referring to the two methods. Nevertheless, further details can be found in Agugiaro et al. (2022)[5] where a preliminary analysis of the influence of party walls for urban energy modelling was carried out, using as test dataset containing the semantic 3D city model of the Dutch municipality of Rjissen-Holten[6] (León-Sánchez et al. 2022)[7] comparing it to yet another (simplified, 2D-based) method for the estimation of shared wall areas.

It is important to mention that due to the lack of ground truth data, no matter which analysis will be presented in the next sections, only a mutual comparison between the results of the two methods has been possible. There are no other (freely) available datasets containing information on adjacency, or surfaces areas, or enclosed gross volumes, that can be used as reference to validate the results from both methods.

[5] https://isprs-archives.copernicus.org/articles/XLVIII-4-W5-2022/9/2022/
[6] https://github.com/tudelft3d/Testbed4UBEM
[7] https://isprs-archives.copernicus.org/articles/XLVIII-4-W5-2022/97/2022/

2.1 Overview of the sample dataset

The sample dataset used for the preliminary tests refers to the municipality of Rijssen-Holten and was obtained by integrating, enriching and further processing data from the 3DBAG, the BAG, and additional datasets curated by the 3D Geoinformation group. From the 3DBAG (the version released in September 2021), 25 tiles were used. Their geographical extent is presented in Figure 3. They cover the whole area of the municipality of Rijssen-Holten and also include those buildings from the nearby municipalities falling within the selected tiles.

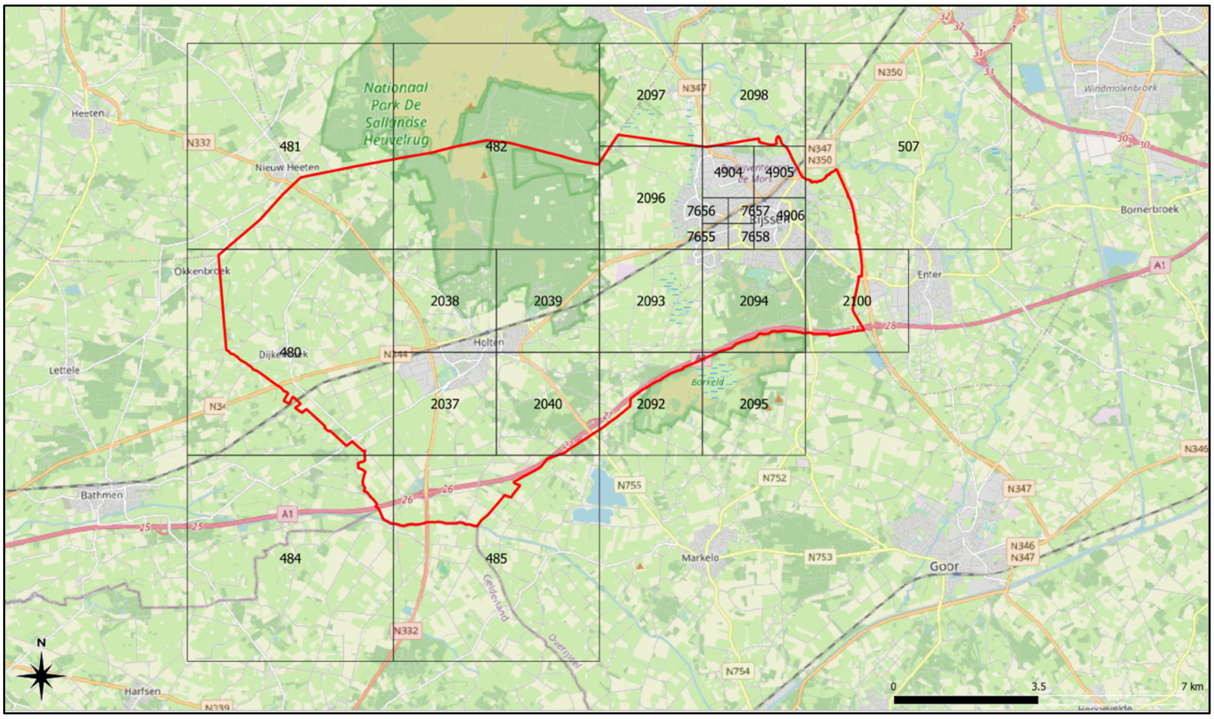


Figure 3. Extents of the sample dataset. The borders of the municipality of Rijssen-Holten are shown in red, while the 3DBAG tiles are overlaid in black.

The data sources have been harmonised and integrated to build a semantic 3D city model of the municipality of Rijssen-Holten. The 3D city model contains only building information, and, as mentioned before, is based on the international standard CityGML.

In terms of geometry, the modelling rules used for Rijssen-Holten are the following:

- A small number of buildings is modelled as multi-part buildings, i.e. they have hierarchically subordinated building-part objects. A multi-part building is associated with a unique Pand ID. Most of the time, a multi-part building is the result of the 3D reconstruction process where a rather large BAG polygon (e.g. spanning over underground features like a garage) is split into smaller polygons that correspond only to above-ground objects. An example of a multi-part building is presented in Figure 4
- The majority of the buildings (or building parts) are modelled in LoD2 via thematic surfaces, i.e. the building envelope is composed of WallSurfaces, RoofSurfaces and GroundSurfaces
- For a limited number of buildings, there are no LoD2 geometries in the used 3DBAG dataset, as the reconstruction process was not possible (e.g. due to lack or scarcity of Lidar points). These buildings are kept in the city model but have only a LoD0 geometry. Nevertheless, they are considered when computing the number of adjacent buildings. An example is presented in Figure 5.

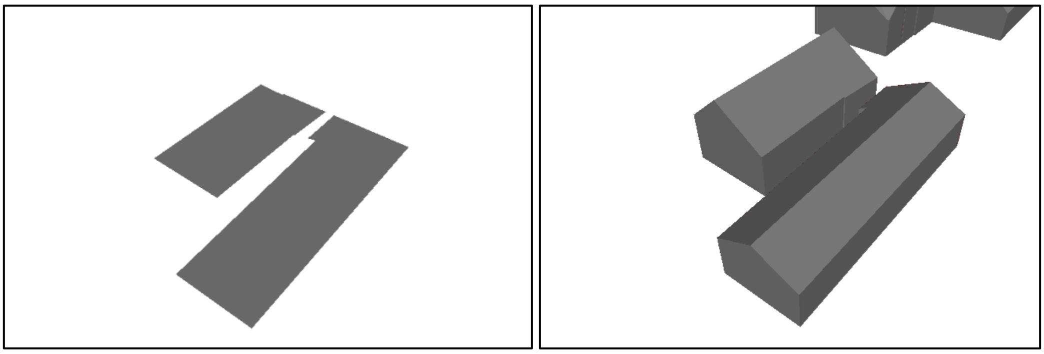


Figure 4. Example of multi-part building. Represented as LoD0 (2 footprints) [left], and as LoD2 (via thematic surfaces) [right].

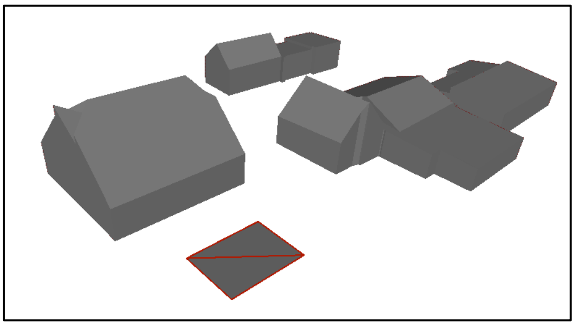


Figure 5. Example of building (highlighted in red) modelled only by means of LoD0 geometry.

The resulting dataset consists of circa 30500 buildings. Table 1 and Table 2 provide an overview of the test dataset in terms of buildings and thematic surfaces, respectively.

Table 1. Overview statistics of the Rijssen-Holten dataset in terms of buildings.

Buildings	Count	%
Number of buildings	30448	100.0%
of which:		
• only in LoD0	886	2.9%
• up to LoD2	29562	97.1%
• Single-part LoD2 building	29505	96.9%
• Multi-part LoD2 building	57	0.2%
• Free-standing building	17058	56.0%
• Non-free-standing building	13390	44.0%
• Residential building	14489	47.6%
• Mixed-use building	1235	4.1%
• Non-residential (single function)	9787	32.1%
• Non-residential (multi function)	200	0.7%
• Of unknown class	4737	15.6%

Table 2 Overview statistics of the Rijssen-Holten dataset in terms of thematic surfaces, **before** *the computation of the party walls ("Shared WallSurfaces").*

LoD2 thematic surfaces	Count		Area	
	n	%	m^2	%
Total	672129	100.0%	15834612.46	100.0%
of which				
• GroundSurface	29624	4.4%	4484810.56	28.3%
• RoofSurface	94237	14.0%	5118210.27	32.3%
• Ext. WallSurface	548268	81.6%	6231591.64	39.4%

• Shared WallSurface	N/A	N/A	N/A	N/A

2.2 Adjacency checks

In the following figures, the colour coding indicated in Table 1 has been used:

Table 3 Building footprint classification according to their adjacency

Colour	Description	Quantity
(white)	Non-adjacent buildings according to both methods	17199
	Adjacent buildings according to both methods	12820
	Adjacent buildings according to Method 1 only	291
	Adjacent buildings according to Method 2 only	138

A visual inspection was also carried out. The following Figures are screenshots taken from QGIS and represent excerpts of the Rijssen-Holten dataset allowing for a better visual perception of the results (and some of the identified issues).
Figure 6 gives a general overview of the results. In general, adjacent buildings are correctly identified (i.e. both methods identify them as being adjacent). However there are some that are missing. By further exploring the dataset, and zooming in at specific cases, one can see that it is not only small footprints (e.g. shed) being misclassified, but sometimes also proper and large buildings, as shown in Figure 7.
Only by performing a zoom, as shown in Figure 8, one of the reasons that lead to a wrong classification can be understood: the buildings are not adjacent due to small gaps caused by digitalization errors. This keeps the building polygons separated. Another typical digitalization error consists of drawing polygons that overlap. Examples will be given later in the report.

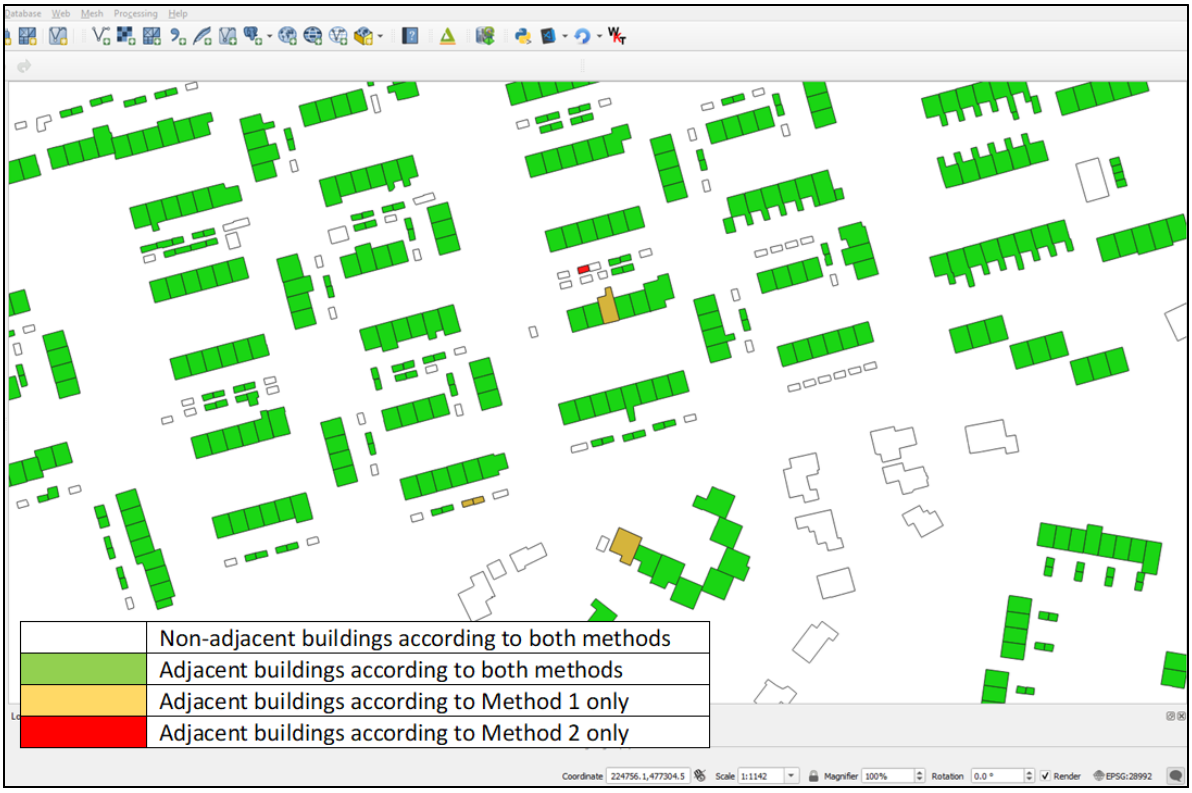


Figure 6. Example of classification of buildings according to their adjacency class.

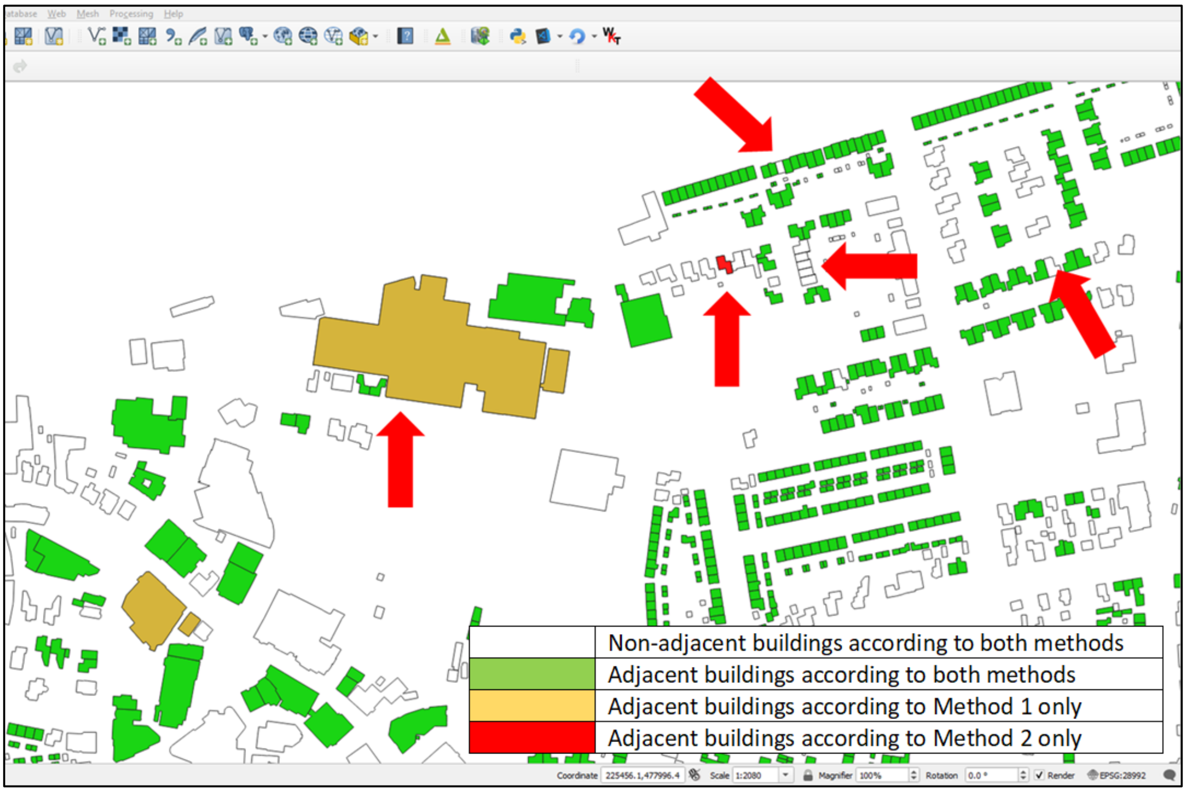


Figure 7. Example of issues in the classification of buildings according to their adjacency class.

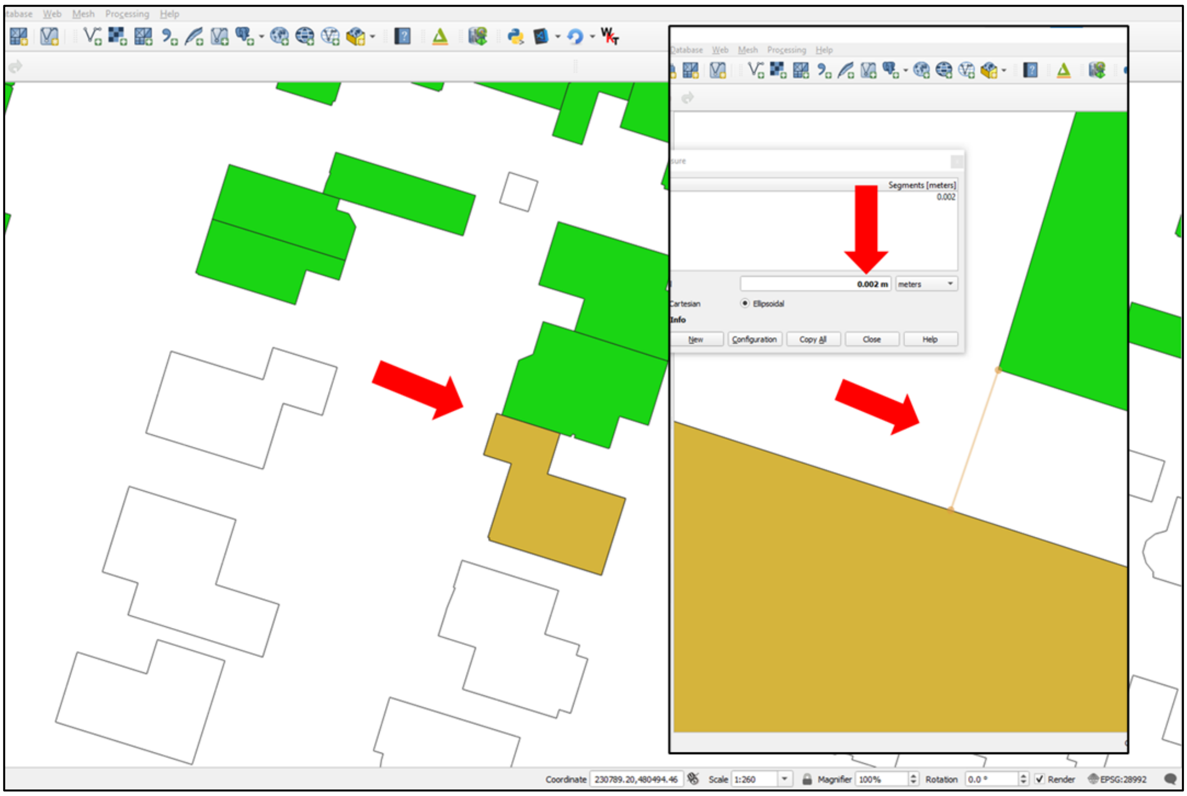


Figure 8. Example of disjoint buildings due to digitalisation errors.

The conclusions from the analysis on the identification of adjacent buildings can be summarised in the following points:

- The number of wrongly classified buildings as adjacent, in either method, is still not negligible. Both methods have a similar level of inaccuracy, despite the different strategies and algorithms employed;
- Given the imperfect nature of the source data (the BAG) used to generate the 3DBAG, it would be advisable to incorporate some tolerance values in the algorithms, in order to be able to tackle also those cases where two buildings are "slightly" disjoint or overlap. How this is done in the final results is explained om Section 3.1.
- From an analysis of similar experiences in other analogous projects in Europe, a recommended tolerance value for a search radius around the footprint polygon is 10 cm. Analogously, checks on parallelism deviation between adjacent polygon segments should be around 5° (decimal degrees);
- Another source of error is represented by duplicate buildings, or completely overlapping ones as they are present in the BAG. These should be removed beforehand, as they lead to further major issues when computing the party walls. More details about this specific issue will be given in the coming sections.

2.3 Computation of areas of the building envelope distinguishing between party and exterior walls

The purpose of this set of analysis is to compare the differences between the afore-mentioned two methods in terms of computed area of the party walls. The analysis has been carried out first at whole-dataset level, then at tile level, and finally manually, by visually inspecting four tiles with the largest differences.
Ideally, once the area of all 3D surfaces composing the building envelope is computed following either method, the difference between the total area should be zero. The same applies to the differences computed between the semantically aggregate values, e.g. the total area of RoofSurfaces, the total area of exterior WallSurfaces, the total area of party walls, etc. However, due to the different implemented set of algorithms, and the influence of the previously show adjacency issues, differences might be found in certain tiles, and – more specifically – with certain buildings. The purpose of this analysis is therefore to investigate whether and where major differences can be found.

Table 4 presents the results of this analysis. Sets of columns have been coloured for a better reading. The first column, in white, contains the tile code. All other values refer to aggregated areas for all buildings and are expressed in [m^2], while the "diff" columns contain the relative difference of the Method 2 compared to Method 1 in [%].

First and foremost, at whole-dataset level, there is a general accordance between the two methods, as can be seen from the last line of the table containing the totals. All "diff" values are below 1% – which is a rather good correspondence – with the only exception of the RoofSurfaces. This is due to some in-between simplifications of the roof geometries of Method 2.

If the analysis is then carried out at tile level, the following points can be highlighted:

- The second set of columns, in light green, presents the total area of the building envelopes. As can be observed from the "diff" column, the differences (in absolute value) are nearly all below 1%, i.e. the differences are negligible. Still, it can be observed that Method 1 delivers slightly higher aggregated area values compared to Method 2. This is, in general, due to differences in the computation of the roof areas (as mentioned before), and to Method 2 ignoring (e.g. dropping) the geometries of the party walls in case their area is smaller than a threshold of 0.0001 m^2. For further details, please refer to the previously mentioned paper by Agugiaro et al. (2022).
- The set of columns in light orange focuses on the total areas of the RoofSurfaces, and the next one, in light green, on total areas of the GroundSurfaces. In general, again, there is a rather good correspondence between the values computed with the two methods. For the GroundSurfaces,

small differences are due to the way a larger footprint is sometimes split into smaller parts in Method 1.

- The dark grey set of columns contain the comparison between the total area of the WallSurfaces, intended as a sum of exterior and party (i.e. shared between two buildings) walls. As can be seen from the “diff” column, there is a perfect correspondence between the two methods.
- In general, a rather good correspondence between the two methods can also be seen when considering the last two sets of columns, in light grey and light yellow, representing the area of the exterior walls and the area of the party walls, respectively. In general terms, Method 1 delivers aggregated area values for the party walls that are slightly larger than the corresponding ones from Method 2.
- Nevertheless, some tiles show remarkable differences that are larger (in absolute value) than 1%. This might be due to the previously seen differences in the way adjacent buildings are classified. Nevertheless, for the four tiles with the largest deviations (tiles: 485, 2092, 2095, 3096, highlighted in red in Table 4) a further, more detailed and manual (visual) inspection at single-building level was carried out. The results and the accompanying notes are presented and commented in the following Figures.

Table 4. Comparison of areas computed on building envelope elements using the two methods. Area values are aggregated at tile level.

Tile	Envelope	area [m^2]	Diff [%]	Roof	area [m^2]	Diff [%]	Ground	area [m^2]	Diff [%]	Walls_tot	area [m^2]	Diff [%]	Ext. Walls	area [m^2]	Diff [%]	Party Walls	area [m^2]	Diff [%]
	Method 1	Method 2		Method 1	Method 2	Diff	Method 1	Method 2	Diff	Method 1	Method 2	Diff	Method 1	Method 2	Diff	Method 1	Method 2	Diff
480	83587.40	83226.48	-0.43	27019.95	26659.07	-1.34	23546.12	23546.07	0.00	33021.33	33021.34	0.00	23883.31	24276.54	1.65	9138.02	8744.80	-4.30
481	142203.90	141569.03	-0.45	41671.70	41036.96	-1.52	33964.66	33964.66	0.00	66567.54	66567.40	0.00	48994.07	48939.28	-0.11	17573.47	17628.12	0.31
482	4632.93	4632.94	0.00	1748.19	1748.19	0.00	1596.39	1596.39	0.00	1288.35	1288.35	0.00	1167.03	1167.03	0.00	121.32	121.33	0.00
484	44953.78	44763.18	-0.42	16803.48	16612.90	-1.13	13843.85	13843.83	0.00	14306.45	14306.45	0.00	11710.31	11695.22	-0.13	2596.14	2611.23	0.58
485	117464.26	117053.57	-0.35	44206.30	43795.64	-0.93	37894.28	37894.27	0.00	35363.68	35363.66	0.00	**26796.96**	**27746.28**	**3.54**	**8566.72**	**7617.37**	**-11.08**
507	255252.25	254080.16	-0.46	73943.44	72843.05	-1.49	62773.76	62702.02	-0.11	118535.05	118535.09	0.00	80535.38	81521.28	1.22	37999.67	37013.81	-2.59
2037	372003.56	369936.56	-0.56	89981.72	87914.87	-2.30	73341.10	73341.03	0.00	208680.74	208680.66	0.00	129817.34	129987.94	0.13	78863.40	78692.72	-0.22
2038	207662.49	206043.64	-0.78	58174.64	56803.90	-2.36	50999.60	50751.56	-0.49	98488.25	98488.19	0.00	72395.83	72520.55	0.17	26092.42	25967.64	-0.48
2039	261085.50	260434.59	-0.25	59262.07	58611.03	-1.10	50307.46	50307.52	0.00	151515.97	151516.04	0.00	93647.91	93727.08	0.08	57868.06	57788.96	-0.14
2040	22216.61	22184.52	-0.14	5864.02	5831.95	-0.55	5074.40	5074.39	0.00	11278.19	11278.18	0.00	8377.38	8375.49	-0.02	2900.81	2902.69	0.06
2092	12893.37	12825.29	-0.53	4728.95	4660.93	-1.44	4053.42	4053.39	0.00	4111.00	4110.97	0.00	**3011.06**	**3452.09**	**14.65**	**1099.94**	**658.88**	**-40.10**
2093	85968.96	85588.42	-0.44	20597.12	20216.55	-1.85	17103.27	17103.24	0.00	48268.57	48268.63	0.00	30637.63	30906.48	0.88	17630.94	17362.16	-1.52
2094	718235.14	715090.38	-0.44	163880.17	160735.54	-1.92	137595.69	137595.60	0.00	416759.28	416759.24	0.00	256916.19	257893.25	0.38	159843.09	158865.99	-0.61
2095	34704.70	34430.94	-0.79	13302.98	13029.16	-2.06	11699.82	11699.84	0.00	9701.90	9701.94	0.00	**7067.01**	**7729.60**	**9.38**	**2634.89**	**1972.33**	**-25.15**
2096	417345.77	415589.42	-0.42	99380.45	97624.15	-1.77	81377.95	81378.04	0.00	236587.37	236587.23	0.00	148474.82	148634.93	0.11	88112.55	87952.30	-0.18
2097	6306.27	6301.13	-0.08	2188.62	2183.50	-0.23	1946.90	1946.90	0.00	2170.75	2170.73	0.00	1567.55	1566.74	-0.05	603.20	603.99	0.13
2098	42668.99	42180.46	-1.14	18079.70	17591.24	-2.70	16391.76	16391.76	0.00	8197.53	8197.47	0.00	**6747.64**	**7061.38**	**4.65**	**1449.89**	**1136.09**	**-21.64**
2100	91443.87	90983.81	-0.50	29519.95	29059.85	-1.56	27738.27	27738.30	0.00	34185.65	34185.65	0.00	23980.57	24483.63	2.10	10205.08	9702.02	-4.93
4904	132456.71	131340.77	-0.84	46497.06	45381.27	-2.40	43817.86	43817.79	0.00	42141.79	42141.71	0.00	32561.80	32637.55	0.23	9579.99	9504.16	-0.79
4905	104998.97	103980.46	-0.97	32037.86	31019.32	-3.18	27945.22	27945.20	0.00	45015.89	45015.94	0.00	35164.52	35189.78	0.07	9851.37	9826.16	-0.26
4906	695782.62	692112.73	-0.53	177020.03	173584.14	-1.94	149909.07	149676.93	-0.15	368853.52	368851.66	0.00	260168.13	260581.81	0.16	108685.39	108269.85	-0.38
7655	209843.33	208924.05	-0.44	50736.44	49818.87	-1.81	42396.21	42395.69	0.00	116710.68	116709.49	0.00	72588.02	73034.06	0.61	44122.66	43675.43	-1.01
7656	267939.87	266465.40	-0.55	64920.25	63445.65	-2.27	52223.58	52223.63	0.00	150796.04	150796.11	0.00	98354.89	98705.41	0.36	52441.15	52090.70	-0.67
7657	90284.55	89933.57	-0.39	23971.19	23629.46	-1.43	19427.06	19417.76	-0.05	46886.30	46886.35	0.00	34570.84	34649.07	0.23	12315.46	12237.28	-0.63
7658	259028.64	258332.18	-0.27	60694.72	59998.33	-1.15	51090.27	51090.19	0.00	147243.65	147243.65	0.00	94234.79	94787.53	0.59	53008.86	52456.12	-1.04
Total	**4680964.44**	**4658003.68**	**-0.49**	**1226231.00**	**1203835.54**	**-1.83**	**1038057.97**	**1037496.01**	**-0.05**	**2416675.47**	**2416672.13**	**0.00**	**1603370.98**	**1611269.98**	**0.49**	**813304.49**	**805402.15**	**-0.97**

Figure 9 provides an example of a first type of errors found in the source dataset BAG. From Figure 9a, one might expect that the building in yellow is clearly a free-standing building, as there are obviously no adjacent footprints. However, upon a better check, and as visible in Figure 9b, there is a smaller building which is overlapping and “contained” in the larger footprint. This issue might be due to a digitalization error, or to a wrong update operation in that one of the buildings has been replaced by the other, but not deleted from the BAG database. The effects on the computation of the shared walls are better visible in Figure 9c. Although there should be no party walls at all (it is a free-standing building), the overlapping building leads to the computation of “non-existing” party walls (in reality). Although this is not an error of either of the Methods, it does have consequences for the results.

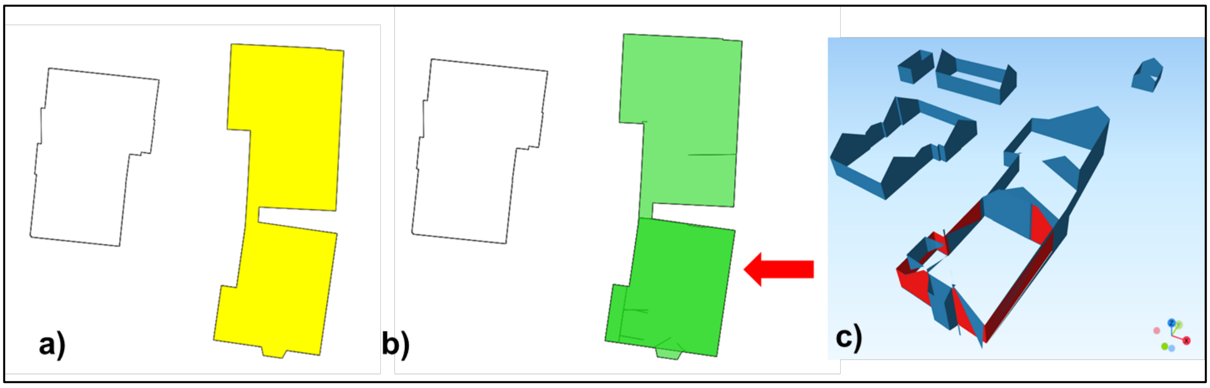


Figure 9. Example of errors due to overlapping buildings in the source dataset BAG. A free-standing building (a), is actually overlapping another one b), and this leads to the computation of party walls that should not exist in reality.

Figure 10 presents another example of problems related to the wrong classification of adjacent buildings. The building footprint in the upper part of the image (in white) is not classified as adjacent, due to digitalization errors. It can be clearly seen that there is a skewed gap between its footprint and the one in yellow, which is itself classified as adjacent due to the third footprint, in green. In terms of computation of the party walls, this case causes the algorithms to deliver different results:

- Method 1 does not compute the party walls between the white and the yellow building, most probably because it does not identify this as a case of adjacent buildings;
- Method 2 does compute these party walls (as shown in the right image in 3D). However, only the party wall of the yellow building is computed, and not the "reciprocal" one of the white building.

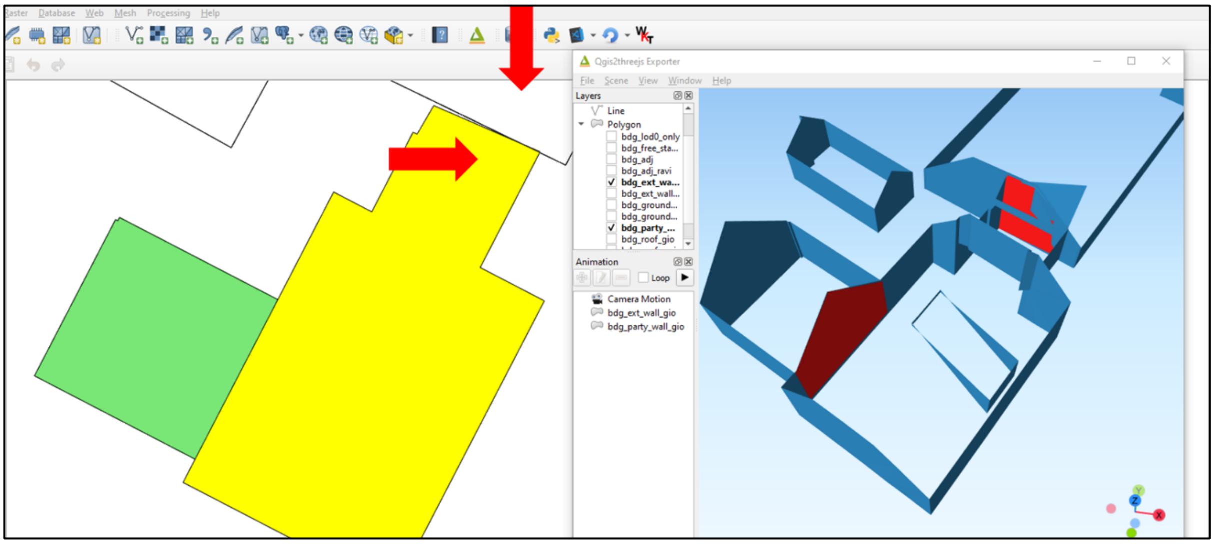


Figure 10. Example of issues due to wrong classification of buildings based on adjacency. The building in white, next to the building in yellow, is not classified as adjacent due to the small gap between the two footprints.

Figure 11 presents an example of issues caused by slightly overlapping building footprints. This case leads to the following results:

- Method 1 does not compute the party walls between the yellow and the green building;
- Method 2 does compute the party walls (as shown in the right image in 3D).

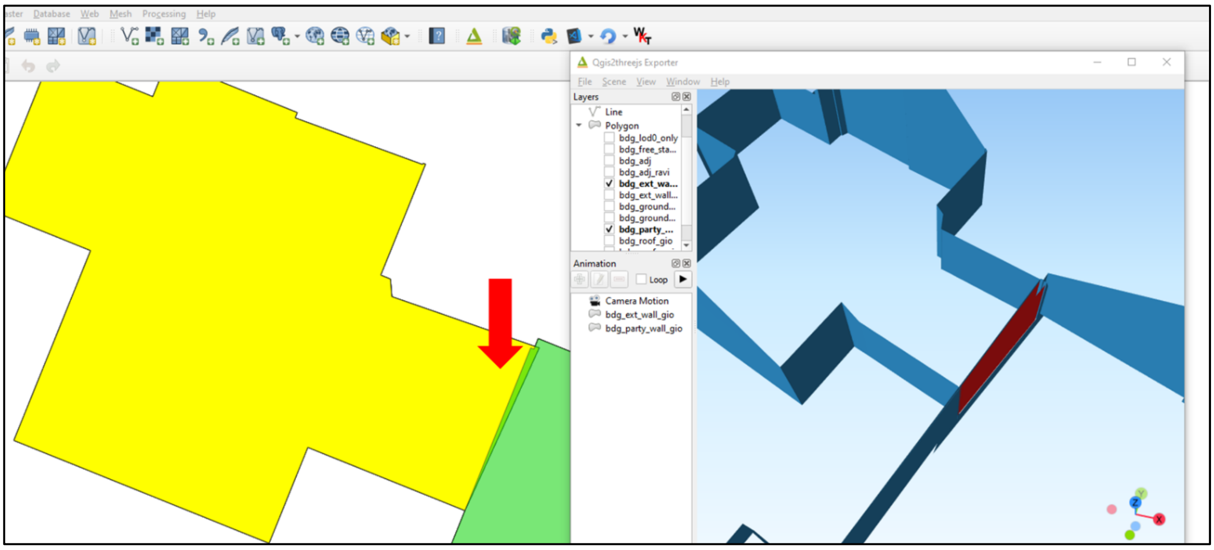


Figure 11. Example of buildings with slightly overlapping footprint polygons.

There may be cases also where multiple of the previously presented issues come together and are combined. Figure 12 presents an example thereof. The building in light brown is classified as adjacent by Method 1 and not by Method 2. A slight gap can be seen between the two footprint polygons. The third building, in green, is classified by both methods as adjacent. The configuration depicted here leads to the following results:

- Both methods compute the party walls between the yellow and the green buildings, as shown on the right side of Figure 12;
- Method 1 does not compute the party walls between the yellow and the light brown building;
- Method 2 does compute the party walls between the yellow and the light brown building, however only the one for the yellow building.

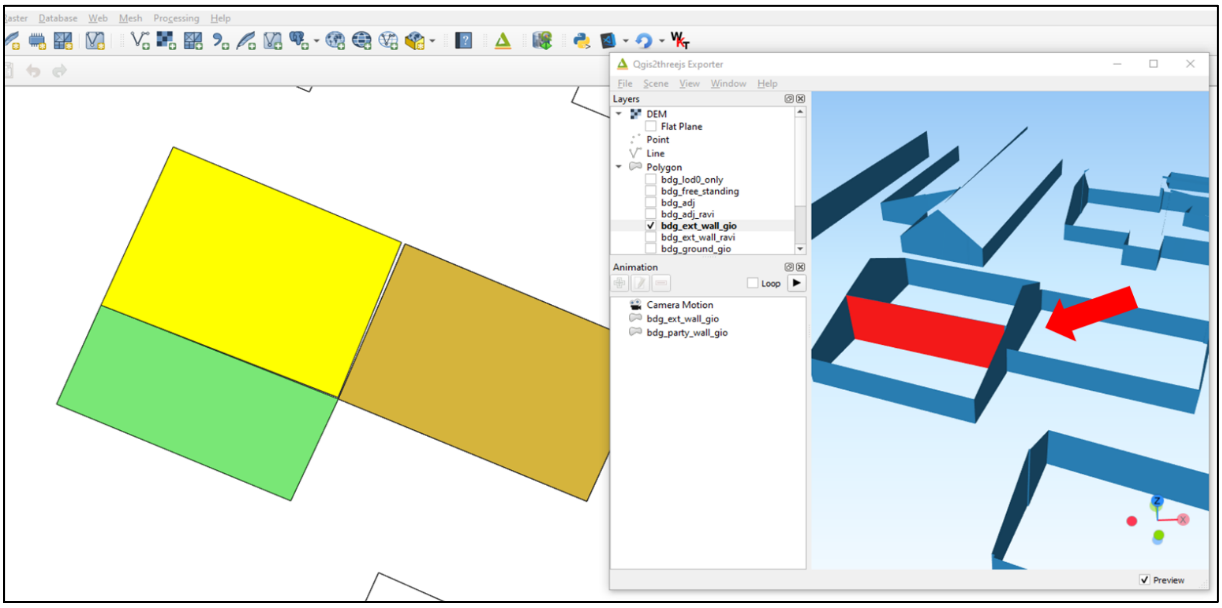


Figure 12. Example of combination of issues and their effects on the computation of part walls.

The adjacency issues notwithstanding, a further check has been carried out on the actual 3D geometries resulting from the computation of the party walls. In general, no major issues or differences have been encountered. Results from both methods are comparable. A simple visual example is presented in Figure 13.

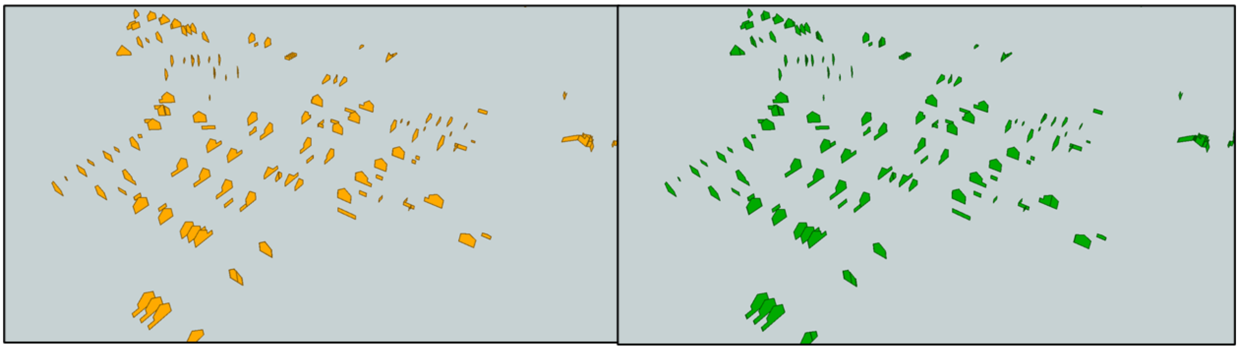


Figure 13. Visual comparison of geometries representing the party walls, computed using the Method 1 [left] and Method 2 [right].

In conclusion, when it comes to the analysis of the area computation, there are no major discrepancies between the two methods. However, with particularly "tricky" cases where multiple issues are present at the same time, Method 2 seems to perform slightly better in terms of robustness, although at the cost of a slower computation time. Therefore, the results from these preliminary analyses suggest that the renewed pipeline for the 3DBAG generation that will compute the surface areas should, as far as possible, merge (or include) aspects from both methods and, if possible, report about any potential adjacency issue.

2.4 Computation of the enclosed gross volume

Similar to the analysis carried out for the area computation, a set of checks has been carried out to investigate potential differences in the computation of the volume enclosed by the building envelope surfaces.
As previously described, the set of tests has been carried out at different scales to compare the two methods: first at the level of the whole dataset, then at tile level, and finally, manually by means of visual inspection at the single-building level, in case of particular tiles with significant deviations.
Again, it is important to mention here that, due to the lack of ground truth data, only a mutual comparison between the results of the two methods is possible.

Two different methods have been used to compute the gross volume of the buildings. The first method, called here "Method 1", is the one initially implemented (at the time when this analysis was carried out, this was later changed to a different method as described in Section 3.1) in the 3DBAG generation pipeline. It computes the volume by means of a voxelization algorithm which uses voxels of 0.5m size. The second method, called "Method 2", uses the volume computation tool of Safe Software's FME (version 2022). The only requirement is a valid solid geometry.
For this reason, before the actual computation of the volume, a preliminary analysis was carried out on the validity of the solid geometries. Method 1's validity check uses the val3dity[8] tool developed by the 3D Geoinformation group at TU Delft, while FME has its own set of modules to perform validity checks.
It should be noted that all statistics in the coming tables and graphs refer to volume computed with Method 2 as reference.

Table 5 presents the distribution of possible geometric errors in the solid geometries of the buildings according to val3dity. The first line, in green, highlights that 93,73% of buildings, corresponding to 91.58% in terms of volume, are valid (in the latest version of 3DBAG 99.15%% of all buildings are valid). The remaining lines contain different types of error codes. The explanation of such codes is available at the GitHub page of the project[9]. The sum of the volumes computed in both methods is

[8] https://github.com/tudelft3d/val3dity
[9] https://github.com/tudelft3d/val3dity/blob/main/docs/errors.rst

also available, as well as their difference. Looking at the last column, it can be observed that for the valid buildings, in general Method 1 delivers 3.17% higher values than Method 2, which corresponds to approximately 28.08 m^3 larger volume for each building.

Table 5. Distribution of valid and invalid geometries according to the results of val3dity.

Error labels	Count		Method 1		Method 2		Difference		Diff avg
		%	[m3]	%	[m3]	%	[m3]	%	[m3]
[no error]	28653	93.73%	26217447.57	91.58%	25412859.49	91.57%	804588.08	3.17%	28.08
[102,104,203]	4	0.01%	**4028.58**	0.01%	**3830.05**	0.01%	**198.53**	5.18%	49.63
[102,104]	1	0.00%	**833.25**	0.00%	**799.06**	0.00%	**34.19**	4.28%	34.19
[102,203]	7	0.02%	**9427.85**	0.03%	**9138.72**	0.03%	**289.13**	3.16%	41.30
[102]	3	0.01%	**3515.20**	0.01%	**3392.81**	0.01%	**122.39**	3.61%	40.80
[104,201,203]	12	0.04%	**16135.68**	0.06%	**15578.04**	0.06%	**557.65**	3.58%	46.47
[104,201]	18	0.06%	**25903.05**	0.09%	**25310.68**	0.09%	**592.36**	2.34%	32.91
[104,203]	147	0.48%	**190159.31**	0.66%	**184355.80**	0.66%	**5803.51**	3.15%	39.48
[104,302]	1	0.00%	**12127.58**	0.04%	**11943.29**	0.04%	**184.29**	1.54%	184.29
[104]	215	0.70%	**212732.95**	0.74%	**205299.24**	0.74%	**7433.71**	3.62%	34.58
[201,203]	29	0.09%	**29005.02**	0.10%	**28095.60**	0.10%	**909.42**	3.24%	31.36
[201]	85	0.28%	**68926.95**	0.24%	**66211.95**	0.24%	**2715.01**	4.10%	31.94
[203]	979	3.20%	**874851.66**	3.06%	**843963.59**	3.04%	**30888.07**	3.66%	31.55
[204]	155	0.51%	**171011.88**	0.60%	**165110.75**	0.59%	**5901.13**	3.57%	38.07
[302]	168	0.55%	**402807.88**	1.41%	**393220.01**	1.42%	**9587.86**	2.44%	57.07
[303,307]	9	0.03%	**13459.51**	0.05%	**13181.70**	0.05%	**277.81**	2.11%	30.87
[303]	82	0.27%	**375921.98**	1.31%	**368221.89**	1.33%	**7700.09**	2.09%	93.90
[306]	1	0.00%	**1160.62**	0.00%	**1122.27**	0.00%	**38.36**	3.42%	38.36
Grand Total	**30569**	**100.00%**	**28629456.52**	**100.00%**	**27751634.93**	**100.00%**	**877821.59**	**3.16%**	**28.72**

Table 6 presents the distribution of possible geometric errors in the solid geometries of the buildings according to the FME GeometryValidator module. Similar as before, the first line, in green, highlights that 85.5% of buildings, corresponding to 84.24% in volume, are valid. The remaining lines contain different types of error codes. The explanation of such codes is available at the GitHub page of the project[10]. The sum of the volumes computed in both methods is also available, as well as their difference. Looking at the last column, it can be said that, for the valid buildings, in general Method 1 delivers 3.17% higher values than Method 2, which corresponds to approximately 27.87 m^3 larger volume for each building. Please note that:

- The algorithms implemented to define the validity of a solid geometry may differ between the two methods, hence the different results.
- In the case of FME, only valid geometries that have not been modified by FME have been considered, for the sake of a rather conservative approach. Nevertheless, FME offers the possibility to repair certain types of common errors, therefore, considering also the corrected buildings, the percentage would grow to circa 98.7%.

Table 6. Distribution of valid and invalid geometries according to the results of the FME GeometryValidator module.

Error labels	Count		Method 1		Method 2		Difference		Diff avg
		%	[m3]	%	[m3]	%	[m3]	%	[m3]
passed	26138	85.50%	24117688.37	84.24%	23389188.74	84.28%	728499.628	3.11%	27.87
passed,repaired	1	0.00%	**433.55**	0.00%	**403.90**	0.00%	29.64	7.34%	29.64
repaired	4038	13.21%	**3417197.46**	11.94%	**3293144.84**	11.87%	124052.62	3.77%	30.72
repaired,passed	3	0.01%	**5346.32**	0.02%	**5157.20**	0.02%	189.12	3.67%	63.04
failed,passed	4	0.01%	**69805.10**	0.24%	**68164.48**	0.25%	1640.62	2.41%	410.15
failed	385	1.26%	**1018985.73**	3.56%	**995575.77**	3.59%	23409.95	2.35%	60.81
Grand Total	**30569**	**100.00%**	**28629456.52**	**100.00%**	**27751634.93**	**100.00%**	**877821.59**	**3.16%**	**28.72**

[10] https://github.com/tudelft3d/val3dity/blob/main/docs/errors.rst

Looking now at the gross volume of the buildings, the analysis follows the same approach as in the case of the computation of the surface area, going from a whole-dataset level down to the single-building one.

The gross volume values, aggregated for the whole dataset, are presented in Table 7. The RMSE can be considered as an indicator of accuracy associated with the average volume of a building computed with the methods. A value of circa 90 m^3 can be imagined as the average size of a room of 30 m^2 and 3 m high. In other words, when computing the value of a building, the value can be accurate to up to 1 room more (or less).

Table 7. Aggregated values of enclosed volume for the whole dataset.

N. of bdg	Method 1	Method 2	Difference		Diff. avg	RMSE
	[m3]	[m3]	[m3]	%	[m3]	[m3]
30569	**28629456.52**	**27751634.93**	**877821.59**	**3.16**	**28.72**	**90.21**

Further refining the analysis and aggregating the gross volume values at tile level allows to identify tiles which seem to be more problematic, see Table 8 and Figure 14 that present the results. In particular, the histogram plots the maximum and minimum volume differences, in [m^3], within each tile. From the histogram in Figure 14 it can be noted that, while negative differences tend to be relatively small, the positive volume differences are definitely non-negligible. If, on the one hand, this confirms the overall trend that Method 1 delivers larger gross volume values than Method 2, it can be also seen that the biggest issues are in two tiles specifically. More details for each tile can be read in Table 8.

Table 8. Aggregated values of enclosed volume at tile level.

Tile	Count		Method 1		Method 2		Difference		Diff. min	Diff. max	Diff. avg
		%	[m3]	%	[m3]	%	[m3]	%	[m3]	[m3]	[m3]
2037	1967	6.43%	**1807344.26**	6.31%	**1749880.51**	6.31%	57463.75	3.28%	-19.38	1044.06	29.21
2038	1320	4.32%	**1490918.78**	5.21%	**1450360.66**	5.23%	40558.12	2.80%	-64.16	1151.75	30.73
2039	1978	6.47%	**891622.56**	3.11%	**854544.54**	3.08%	37078.02	4.34%	-20.13	568.56	18.75
2040	1172	3.83%	**516095.53**	1.80%	**494864.24**	1.78%	21231.29	4.29%	-12.36	254.09	18.12
2092	285	0.93%	**241408.44**	0.84%	**234416.67**	0.84%	6991.77	2.98%	-10.45	311.48	24.53
2093	642	2.10%	**431235.05**	1.51%	**417412.18**	1.50%	13822.87	3.31%	-24.64	185.88	21.53
2094	3316	10.85%	**1374584.87**	4.80%	**1310704.76**	4.72%	63880.12	4.87%	-26.59	783.29	19.26
2095	245	0.80%	**255171.79**	0.89%	**248448.36**	0.90%	6723.44	2.71%	-12.38	354.49	27.44
2096	1947	6.37%	**1655654.34**	5.78%	**1598905.29**	5.76%	56749.05	3.55%	-35.36	883.94	29.15
2097	377	1.23%	**308574.16**	1.08%	**298060.02**	1.07%	10514.14	3.53%	-23.77	378.48	27.89
2098	487	1.59%	**516452.44**	1.80%	**503696.90**	1.82%	12755.54	2.53%	-27.92	517.73	26.19
2100	651	2.13%	**1618141.86**	5.65%	**1583179.04**	5.70%	34962.83	2.21%	-117.31	8856.68	53.71
480	1350	4.42%	**1568712.57**	5.48%	**1526630.92**	5.50%	42081.65	2.76%	-25.58	264.13	31.17
481	1340	4.38%	**1086615.20**	3.80%	**1047319.89**	3.77%	39295.31	3.75%	-10.50	311.20	29.32
482	102	0.33%	**112900.18**	0.39%	**110083.37**	0.40%	2816.82	2.56%	-15.56	193.04	27.62
484	1215	3.97%	**1342524.23**	4.69%	**1307394.29**	4.71%	35129.94	2.69%	-23.77	314.49	28.91
485	1254	4.10%	**1389939.42**	4.85%	**1353340.68**	4.88%	36598.74	2.70%	-25.12	604.16	29.19
4904	523	1.71%	**3532387.30**	12.34%	**3478551.13**	12.53%	53836.17	1.55%	-3.83	1225.25	102.94
4905	606	1.98%	**2164133.64**	7.56%	**2125536.59**	7.66%	38597.05	1.82%	-33.72	1182.54	63.69
4906	3258	10.66%	**2304140.10**	8.05%	**2214502.79**	7.98%	89637.30	4.05%	-30.76	1184.73	27.51
507	2255	7.38%	**1981251.89**	6.92%	**1908583.55**	6.88%	72668.34	3.81%	-35.06	8859.41	32.23
7655	1059	3.46%	**426312.99**	1.49%	**403382.33**	1.45%	22930.67	5.68%	-28.78	232.80	21.65
7656	1102	3.60%	**508712.49**	1.78%	**477913.07**	1.72%	30799.42	6.44%	-4.46	443.72	27.95
7657	737	2.41%	**595697.74**	2.08%	**573973.80**	2.07%	21723.95	3.78%	-6.38	550.47	29.48
7658	1381	4.52%	**508924.70**	1.78%	**479949.40**	1.73%	28975.29	6.04%	-19.43	995.43	20.98
Grand Total	**30569**	**100.00%**	**28629456.52**	**100.00%**	**27751634.93**	**100.00%**	**877821.59**	**3.07%**	**-117.31**	**8859.41**	**28.72**

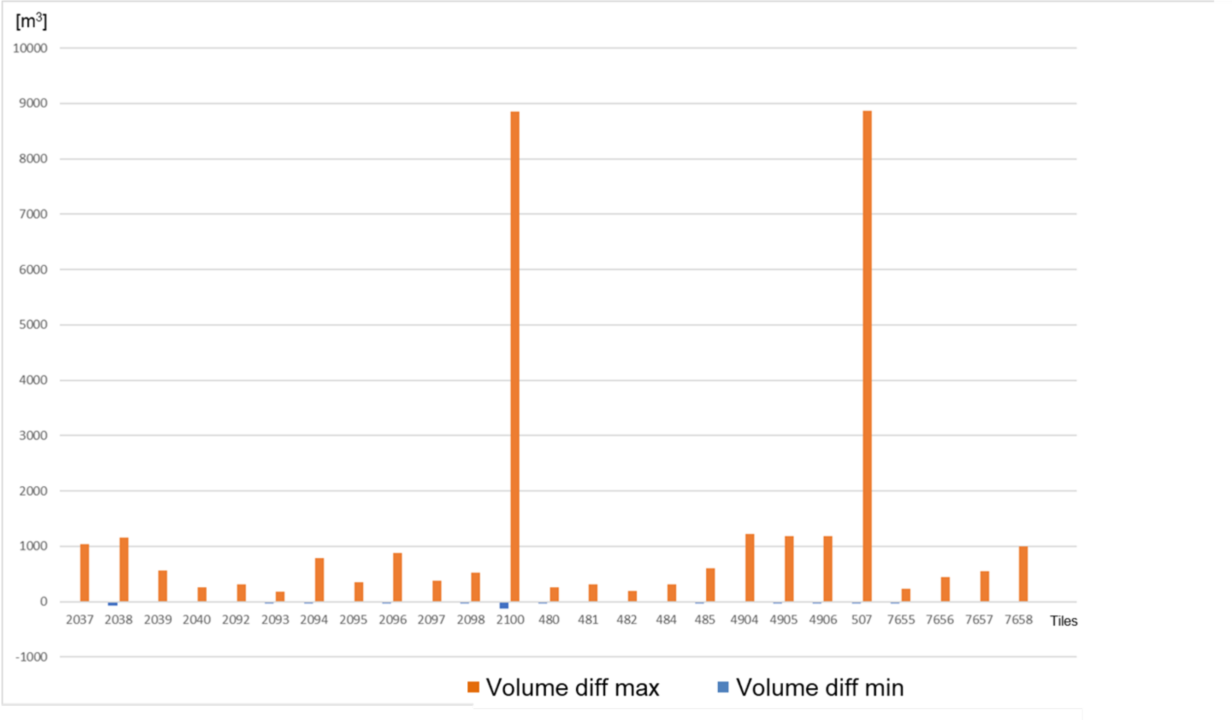


Figure 14. Histogram representing the maximum and minimum volume differences in [m3] within each tile of the dataset.

The analysis at tile level has helped in the further refinement of the analysis, in which the computed gross volume values have been investigated at single-building level, looking at particular problematic cases. Gross volume differences have been analysed in two ways: in terms of absolute volume differences, expressed in [m^3], and in terms of relative volume differences, expressed in [%] in order to normalise them according to the size of the building.

Table 9 represents an excerpt of the list of all buildings in the study area, ordered by ascending order of absolute differences of volume, in [m^3]. The table only shows the first 14 and the last 14 buildings, i.e. the 28 buildings that show the largest differences. The respective volume difference values are plotted in Figure 15 by ascending order of difference value. However, for better scaling (and therefore readability) of the graph, the maximum difference (+8856.68 m^3) has been omitted.

In a similar way, Table 10 represents an excerpt of the list of all buildings in the study area, ordered by ascending order of relative differences of volume, in [%]. The table only shows the first 14 and the last 14 buildings. The respective volume difference values are plotted in Figure 16 by ascending order of difference value. However, for better scaling (and therefore readability) of the graph, the maximum difference (+1732%) has been omitted.

In both tables, for better readability, the first two and the last two buildings in the list have been highlighted in red. It is interesting to observe that in 6 cases out of 8 (but the same trend can be observed also with the other buildings) their geometries have passed the validity test both in val3dity and FME and the validity check results therefore in errorless geometries.

Table 9. Excerpt of the list of buildings ordered by ascending order of absolute differences of volume, in [m3].

Identificatie	Tile	Method 1	Method 2	Difference		Val3dity	FME
		[m3]	[m3]	[m3]	%		
NL.IMBAG.Pand.0189100000015015	**2100**	**3596.30**	**3713.61**	**-117.31**	**-3.16**	**[]**	**passed**
NL.IMBAG.Pand.0189100000043705	**2100**	**15285.47**	**15396.10**	**-110.64**	**-0.72**	**[]**	**passed**
NL.IMBAG.Pand.1742100000008226	2038	2488.74	2552.90	**-64.16**	-2.51	[]	passed
NL.IMBAG.Pand.0189100000015016	2100	3421.38	3483.76	**-62.38**	-1.79	[]	passed
NL.IMBAG.Pand.1742100000001133	2096	529.45	564.81	**-35.36**	-6.26	[]	passed
NL.IMBAG.Pand.0189100000005162	507	1004.16	1039.22	**-35.06**	-3.37	[]	passed
NL.IMBAG.Pand.1742100000015671	4905	522.67	556.39	**-33.72**	-6.06	[]	passed
NL.IMBAG.Pand.1742100000010068	507	2093.27	2124.33	**-31.06**	-1.46	[]	passed
NL.IMBAG.Pand.1742100000005438	4906	3071.23	3101.99	**-30.76**	-0.99	[]	passed
NL.IMBAG.Pand.1742100000013572	7655	564.98	593.76	**-28.78**	-4.85	[]	passed
NL.IMBAG.Pand.1742100000096473	4906	645.88	673.92	**-28.04**	-4.16	[]	passed
NL.IMBAG.Pand.0189100000043069	2098	4460.79	4488.71	**-27.92**	-0.62	[]	passed
NL.IMBAG.Pand.1742100000006406	2094	364.59	391.17	**-26.59**	-6.80	[]	passed
NL.IMBAG.Pand.1742100000002806	4905	318.65	344.73	**-26.08**	-7.57	[]	passed
...	...	...	...	...	...	...	...
NL.IMBAG.Pand.1742100000007149	4904	62504.10	61513.06	**991.05**	1.61	[302]	failed
NL.IMBAG.Pand.1742100000007149	4905	62668.48	61675.85	**992.64**	1.61	[302]	failed
NL.IMBAG.Pand.1742100000001476	7658	38034.27	37038.84	**995.43**	2.69	[]	passed
NL.IMBAG.Pand.1742100000000155	4906	58204.85	57203.74	**1001.11**	1.75	[303]	failed
NL.IMBAG.Pand.1742100000001821	4905	141477.92	140450.61	**1027.31**	0.73	[]	passed
NL.IMBAG.Pand.1742100000001746	2038	65592.60	64550.02	**1042.58**	1.62	[]	passed
NL.IMBAG.Pand.1742100000001746	2037	64445.34	63401.28	**1044.06**	1.65	[]	passed
NL.IMBAG.Pand.1742100000004019	4905	112442.37	111386.46	**1055.91**	0.95	[]	passed
NL.IMBAG.Pand.1742100000009205	2038	156296.30	155144.55	**1151.75**	0.74	[]	passed
NL.IMBAG.Pand.1742100000001822	4905	110858.48	109675.93	**1182.54**	1.08	[]	passed
NL.IMBAG.Pand.1742100000014713	4906	49442.31	48257.58	**1184.73**	2.46	[303]	failed,passed
NL.IMBAG.Pand.1742100000001822	4904	111150.86	109962.92	**1187.94**	1.08	[]	passed
NL.IMBAG.Pand.1742100000001813	**4904**	**216072.56**	**214847.31**	**1225.25**	**0.57**	**[]**	**passed**
NL.IMBAG.Pand.1742100000005324	**2100**	**80772.23**	**71915.56**	**8856.68**	**12.32**	**[]**	**passed**

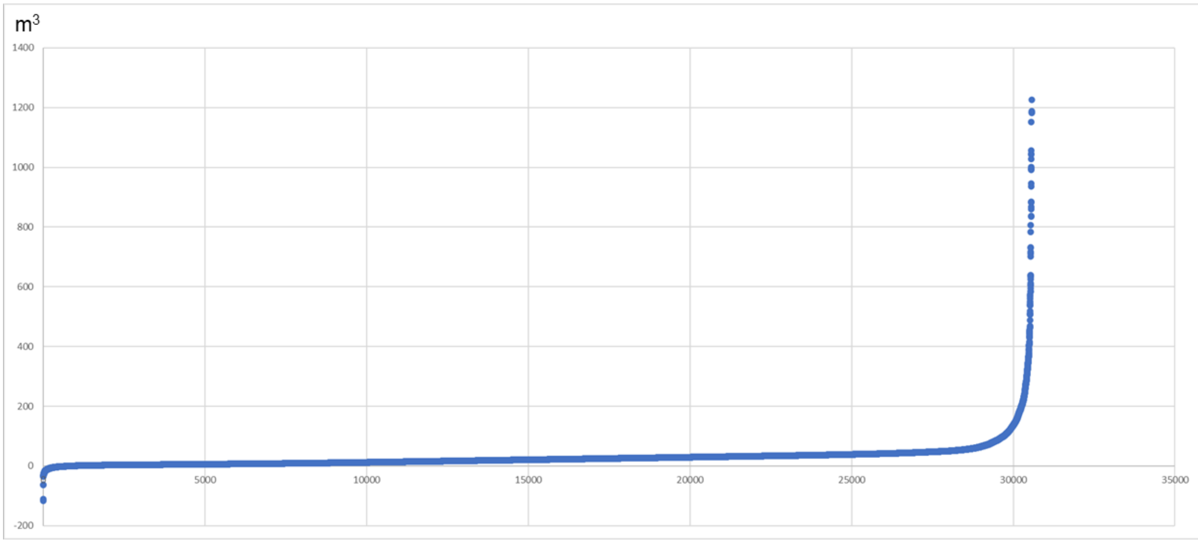


Figure 15. Plot of ordered absolute differences in volume, in [m3], of all buildings belonging to the dataset. For scaling and better readability reasons, the last value of difference (+8856.68 m3) has been omitted.

Table 10. Excerpt of the list of buildings ordered by ascending order of relative differences of volume, in [%].

Identificatie	Tile	Method 1	Method 2	Difference		Val3dity	FME
		[m3]	[m3]	[m3]	%		
NL.IMBAG.Pand.1742100000021308	**2094**	**35.79**	**48.10**	**-12.30**	**-25.58%**	**[]**	**passed**
NL.IMBAG.Pand.0262100000037246	**484**	**12.68**	**16.59**	**-3.91**	**-23.58%**	**[]**	**passed**
NL.IMBAG.Pand.0189100000045509	2098	32.11	41.49	-9.39	**-22.62%**	[]	passed
NL.IMBAG.Pand.1742100000017645	2096	21.91	27.57	-5.66	**-20.53%**	[]	passed
NL.IMBAG.Pand.1742100000005418	4906	11.07	13.78	-2.71	**-19.70%**	[]	passed
NL.IMBAG.Pand.0262100000034332	485	9.27	11.21	-1.94	**-17.31%**	[]	passed
NL.IMBAG.Pand.1742100000095818	2096	50.95	60.95	-10.00	**-16.40%**	[]	passed
NL.IMBAG.Pand.1742100000099467	2096	11.29	13.48	-2.19	**-16.25%**	[]	passed
NL.IMBAG.Pand.1742100000020050	2040	12.81	15.28	-2.47	**-16.18%**	[]	passed
NL.IMBAG.Pand.1742100000004905	2094	41.60	49.17	-7.57	**-15.39%**	[]	passed
NL.IMBAG.Pand.1742100000098546	2094	26.72	31.57	-4.85	**-15.36%**	[]	passed
NL.IMBAG.Pand.0189100000005069	2098	5.97	7.03	-1.07	**-15.21%**	[]	passed
NL.IMBAG.Pand.1742100000017664	2096	22.34	26.29	-3.95	**-15.01%**	[]	passed
NL.IMBAG.Pand.0189100000015185	507	4.88	5.73	-0.85	**-14.79%**	[]	passed
...	...	...	...	...	...	...	...
NL.IMBAG.Pand.1742100000095838	4906	20.72	15.11	5.61	**37.11%**	[]	passed
NL.IMBAG.Pand.0189100000047697	507	11.34	8.26	3.09	**37.36%**	[]	passed
NL.IMBAG.Pand.1742100000095326	2096	58.37	42.35	16.02	**37.81%**	[]	passed
NL.IMBAG.Pand.0189100000009401	507	15.16	10.98	4.18	**38.09%**	[]	passed
NL.IMBAG.Pand.0189100000009106	507	12.80	9.26	3.54	**38.25%**	[]	passed
NL.IMBAG.Pand.0189100000008106	507	13.57	9.69	3.88	**40.00%**	[]	passed
NL.IMBAG.Pand.1742100000099309	4906	12.72	9.03	3.69	**40.82%**	[]	passed
NL.IMBAG.Pand.1742100000000688	2094	446.82	312.67	134.16	**42.91%**	[302]	failed
NL.IMBAG.Pand.1742100000099312	2096	23.34	16.33	7.01	**42.93%**	[]	passed
NL.IMBAG.Pand.0189100000044687	4905	51.24	34.49	16.75	**48.57%**	[]	passed
NL.IMBAG.Pand.1742100000015009	2040	17.62	11.52	6.10	**52.92%**	[]	passed
NL.IMBAG.Pand.1742100000099313	2096	17.68	10.88	6.79	**62.41%**	[]	passed
NL.IMBAG.Pand.1742100000006269	**2094**	**61.81**	**32.61**	**29.20**	**89.55%**	**[302]**	**failed**
NL.IMBAG.Pand.1742100000016155	**2094**	**33.52**	**1.83**	**31.69**	**1732.48%**	**[302]**	**failed**

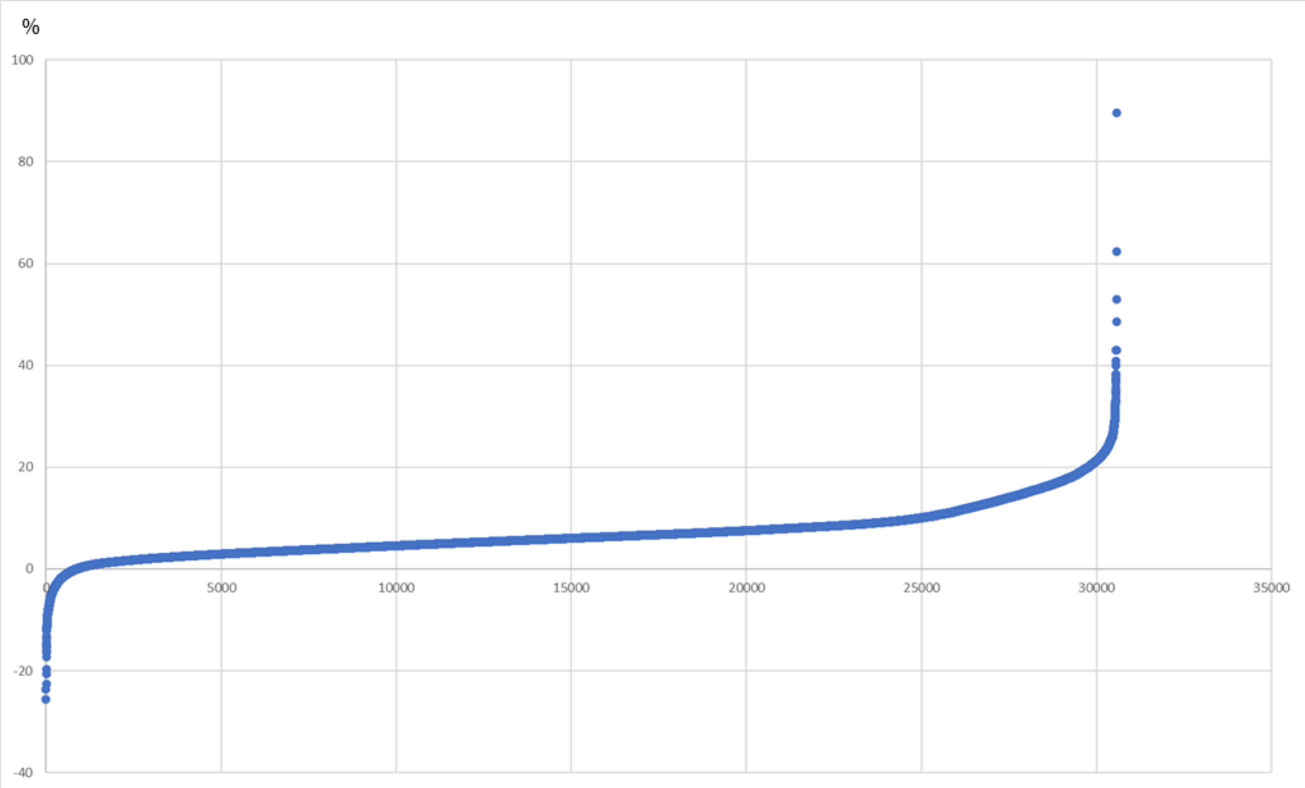


Figure 16. Plot of ordered relative differences in volume, in [%], of all buildings belonging to the dataset. For scaling and better readability reasons, the last value of difference (+1732%) has been omitted.

Both Table 9 and Table 10 have been used to pick those buildings with the largest differences (either in [m^3] or in [%]) and to check them one by one manually by visually inspecting the associated 3D geometries.

Figure 17, Figure 18, Figure 19 and Figure 20 represent the first two and the last two buildings of Table 9. Considering that the geometries are valid, also from the visual inspection no particular errors have become evident. It seems however that the volume differences are somehow proportional to the size and the (geometrical) complexity of the buildings. The latter one can be intended as proportional to the number of polygons needed to define the building envelope.
Nevertheless, the visual inspection of a building next to the one depicted in Figure 20 has allowed us to identify an occasional error in the orientation of the geometries. As highlighted by the red arrow, the RoofSurface of the building in Figure 21 is flipped, i.e. its normal is not pointing "outside" (towards the sky) but towards the inside of the building. This is represented by means of a semi-transparent surface, unlike the solid grey ones of all other buildings. Notice that rare geometric errors like these can be automatically detected and haven been flagged in the final results (see Section 3.2).

Figure 22 and Figure 23 represent the first and the last buildings of Table 10, i.e. those buildings with the minimum and maximum volume differences in [%]. Again, from the simple visual inspection of the building in Figure 22, no particular anomalies in the geometry can be detected visually (and the building has valid geometries, according to Table 10). However, in Figure 23 an error due to the 3D reconstruction pipeline becomes visible. A portion of the building is composed by a set of "flipped" surfaces that therefore contribute to the wrong computation of the volume. This is visible in two buildings. They are highlighted by the red arrows.

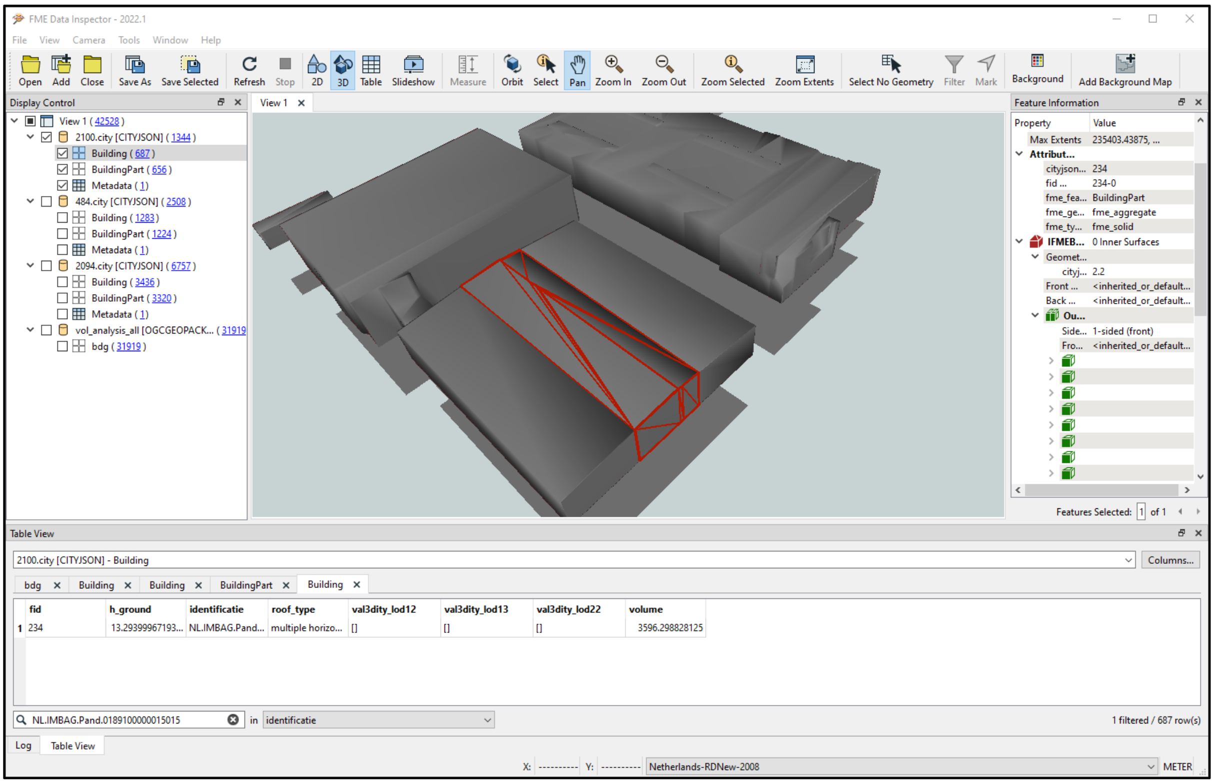


Figure 17. 3D visualisation of the building having id NL.IMBAG.Pand.0189100000015015 (please refer to Table 9 for details).

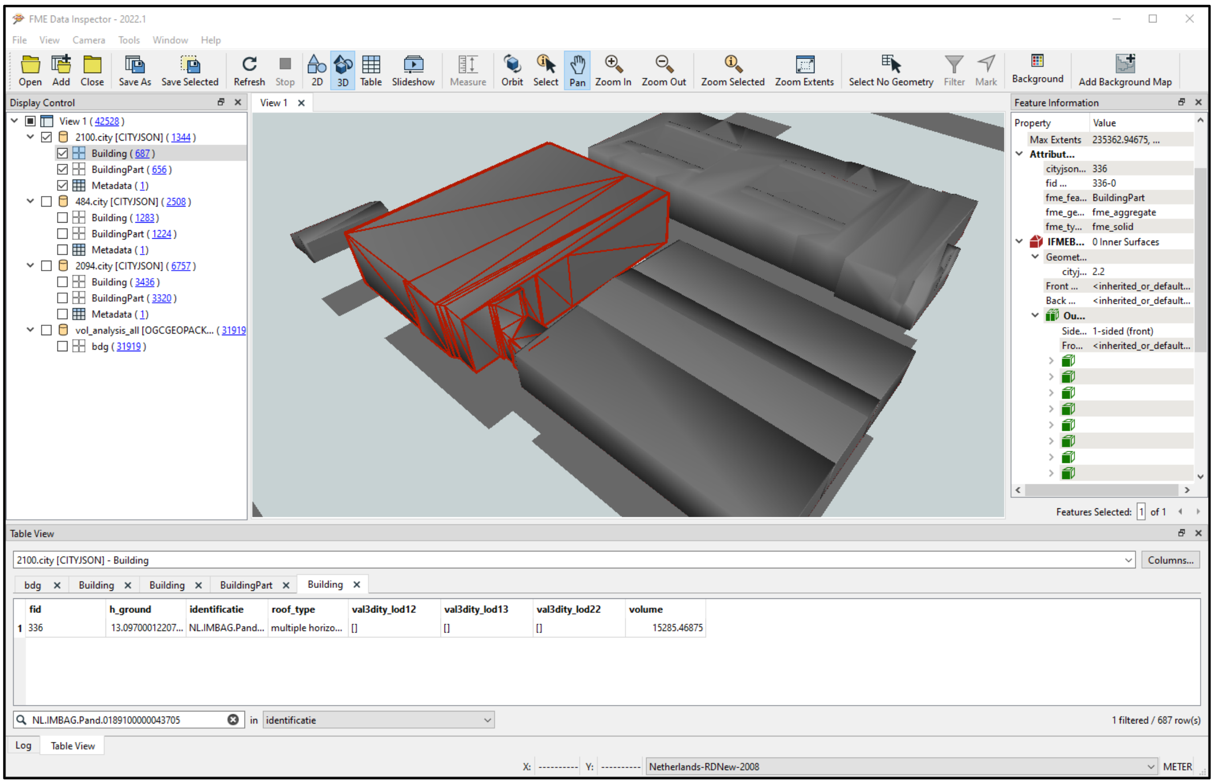


Figure 18. 3D visualisation of the building having id NL.IMBAG.Pand.0189100000043705 (please refer to Table 9 for details).

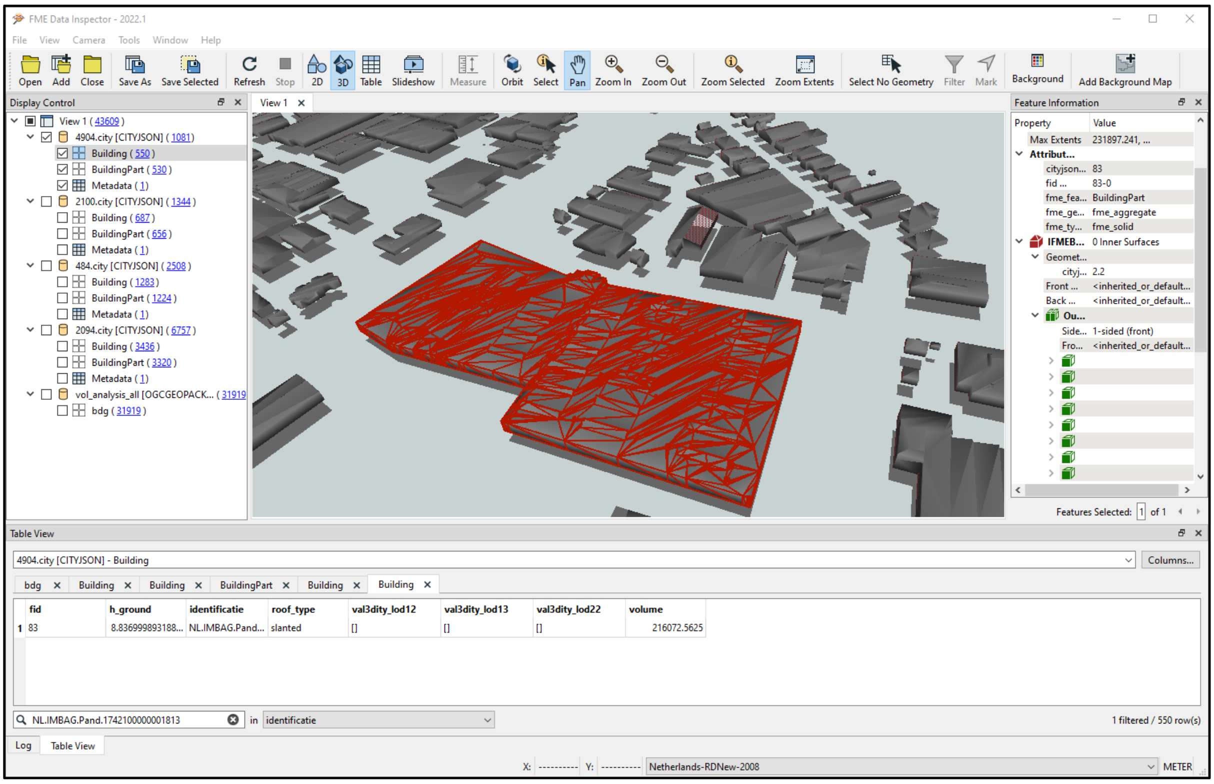


Figure 19. 3D visualisation of the building having id NL.IMBAG.Pand.1742100000001813 (please refer to Table 9 for details).

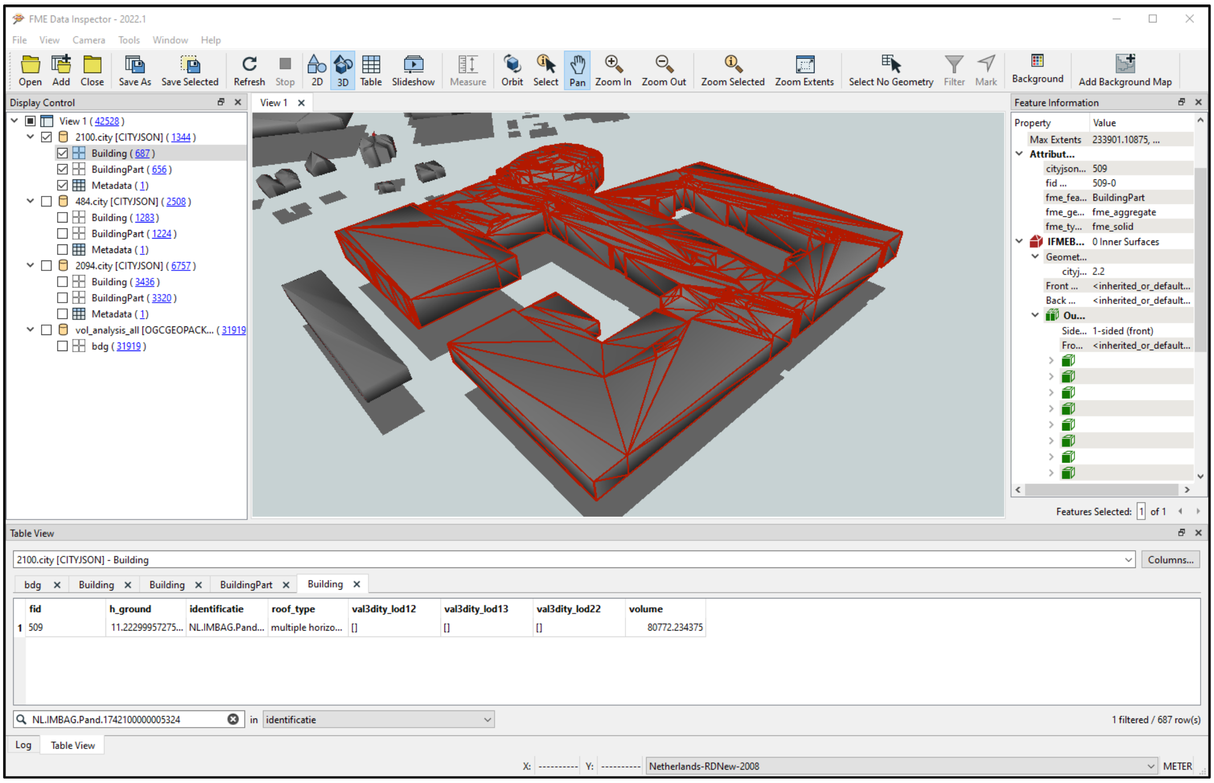


Figure 20. 3D visualisation of the building having id NL.IMBAG.Pand.1742100000005324 (please refer to Table 9 for details).

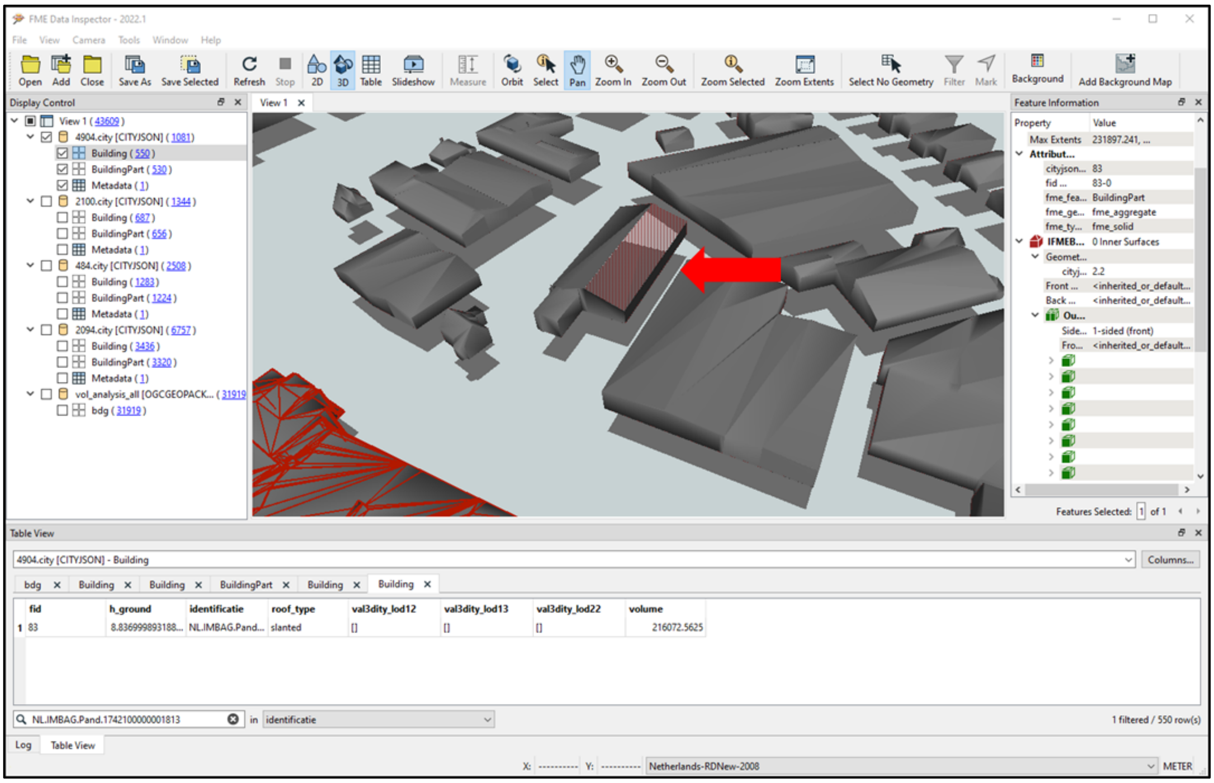


Figure 21. Example of wrongly oriented RoofSurface.

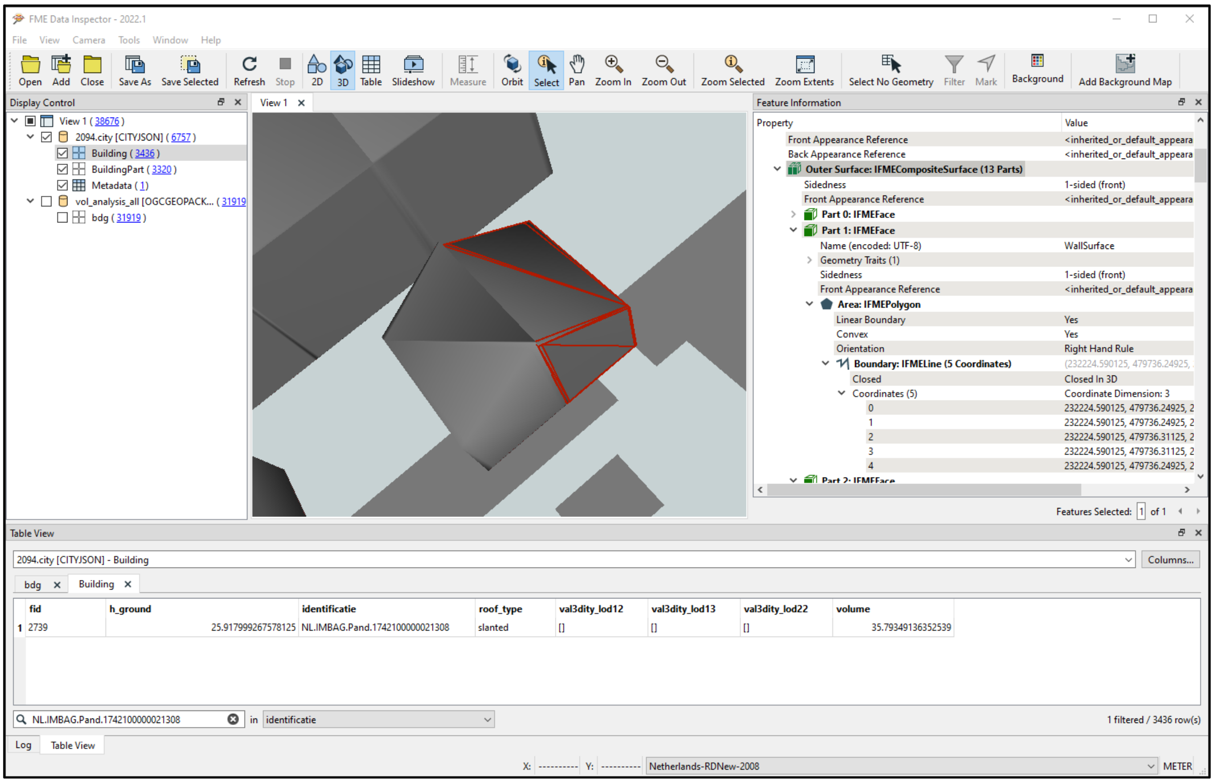


Figure 22. 3D visualisation of the building having id NL.IMBAG.Pand.1742100000021308 (please refer to Table 10 for details).

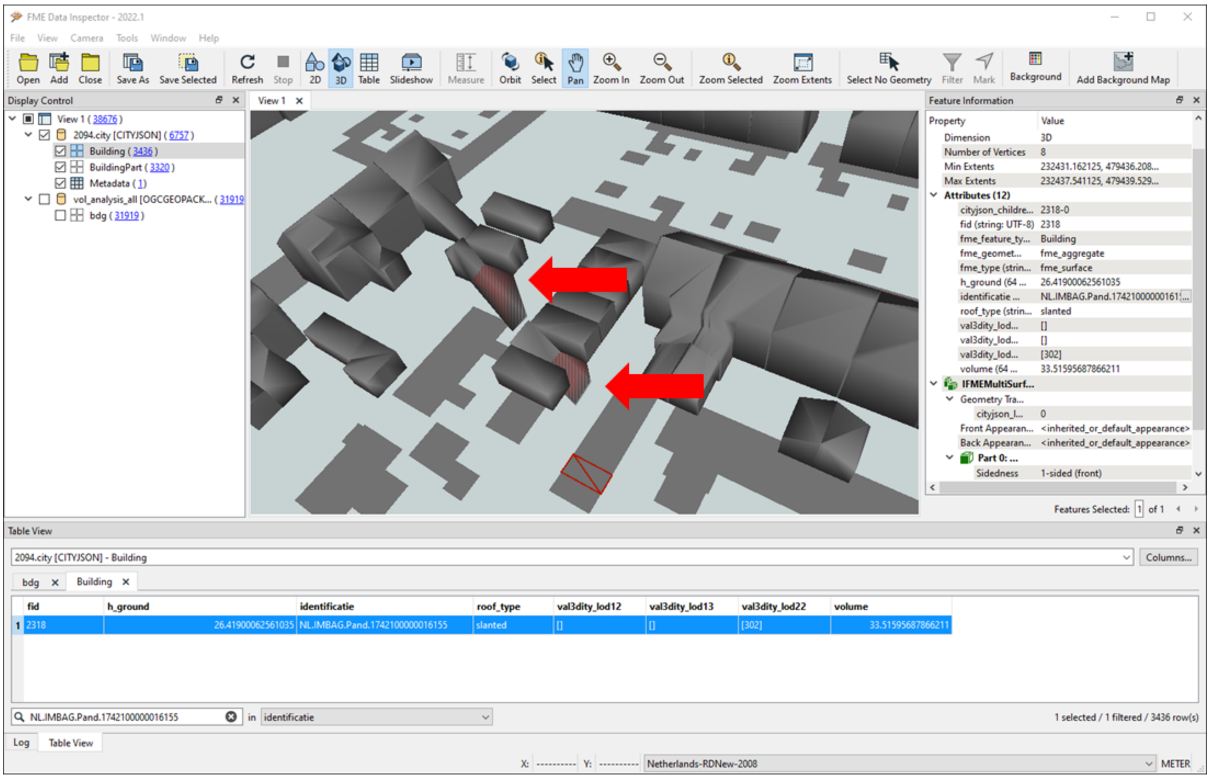


Figure 23. 3D visualisation of the building having id NL.IMBAG.Pand.1742100000016155 (please refer to Table 10 for details).

When it comes to the analysis of the gross volumes enclosed by the building envelopes, it can be summarised from the results that:

- The geometrical validity of the building envelopes "as is" is > 90% using val3dity and >80% using FME GeometryValidator. If, however, one considers also the geometries automatically repaired by FME, the latter figure reaches nearly 99%;
- Method 1 delivers slightly higher results than Method 2 (circa +3% per building), or, alternatively, a RMSE of circa 90 m^3, which might be considered as the "size of a room";
- Nevertheless, when disaggregating the results at tile- and single-building level, major differences are found. It seems that the differences are proportional to the size and geometrical complexity of the building.
- One of the reasons why Method 1 delivers higher values could be dependent on the strategy used to compute the volumes, which uses a discretisation approach relying on voxels of 0.5m size. A reduction of this value (e.g. down to 0.25 cm) might lead to better results, as the approximating volume would be closer to the real one. Alternatively a different volume computation method could be used (and has been used at the end) that does not do any discretisation (see Section 3.1.1);
- A by-product of the visual inspection of the buildings has been the identification of other issues, namely some wrongly oriented (i.e. "flipped") geometries, and some wrongly reconstructed buildings with portions that are completely delimited by flipped geometries. It is important to remind here that for this latter case, the identification of this problem cannot be carried out automatically by means of simple volume comparison, therefore the 3D reconstruction algorithm must be improved to perform additional checks (which can be flagged) and avoid these errors in the future.

Still, it is crucial to remind again that, due to the lack of ground truth data, only a mutual comparison between two methods is possible. This comparison of the results allows therefore to determine how the two methods relate to each other, but it is *per se* not indicative of which method is better in absolute terms. The lack of ground truth data is a limiting factor in this kind of analysis.

3 Generation of volumes and surfaces areas (including party and exterior walls) for all 3DBAG buildings

The aim of the project is to calculate for each 3DBAG building, the volumes as well as the different areas of the envelope discriminating between roof, ground, party wall, and exterior wall. How this is implemented for the whole 3DBAG as the final results of this project, is described in this section.

The analyses described in Chapter 2, have been carried out on the 3DBAG published in 2021. In June 2023 and October 2023 new versions of the 3DBAG have been reconstructed and published. The nationwide calculation of the required parameters has been done on the latest release. Therefore, the main changes are described in Section 3.1. Section 3.2 and Section 3.3 describe the implementation of the party wall area and the volume calculation methods. The generation of the output file that contains per 3DBAG building the volume and areas of surfaces, as well as a quality tag, is described in Section 3.3.

3.1 Calculation method for volumes and surface areas

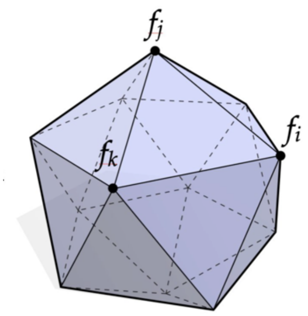
For the majority of the buildings, the methods that were compared in Chapter 2 to detect party walls and calculate their areas behaved comparably. The main differences were caused by the source data, ie. buildings that have geometrical or topological errors or that have complex (real world) geometries. Therefore, for the inclusion in the 3DBAG pipeline, we implemented methods to calculate volumes and discriminate between party and exterior walls that appeared to be most robust and computationally efficient. To deal with the problematic buildings as identified in the previous Chapter, we calculate a 'reliability' attribute as described in Section 3.2.

3.1.1 Calculation of volume

Based on the findings of Chapter 2, we have implemented another calculation method[11] than the one that was evaluated in Chapter 2. This new method is vector based and does not use any kind of discretisation. The volume computation therefore yields exact results, without discretisation error. It is also robust to invalid geometries, such as small gaps in the 3D mesh.

We calculate the volume from the triangulated 3D building model using the formula:

$$\text{volume}(f) = \frac{1}{6} \sum_{ijk \in F} f_i \cdot (f_j \times f_k)$$



We thus calculate for each triangle the dot product of one vertex with the cross-product of the two other vertices. We then sum these and divide the result by 6. This method was implemented in the Geoflow software[12] that is used to perform the 3D building reconstruction for the 3DBAG.

[11] Discrete differential geometry: an applied introduction, Keenan Crane

[12] https://github.com/geoflow3d/geoflow-bundle

3.1.2 Calculation of surface areas

The 3DBAG models already contain the necessary semantic information to distinguish between ground surfaces, wall surfaces and roof surfaces (see Figure 24). The roof surfaces are split further into flat roofs (slope <5 degrees) and sloped roofs (slope > 5 degrees) simply based on their inclination. These calculations can be directly performed on the 3DBAG models. However, the wall surfaces must be further divided into party walls and exterior walls and doing this takes significantly more effort in terms of implementation work and computation time.

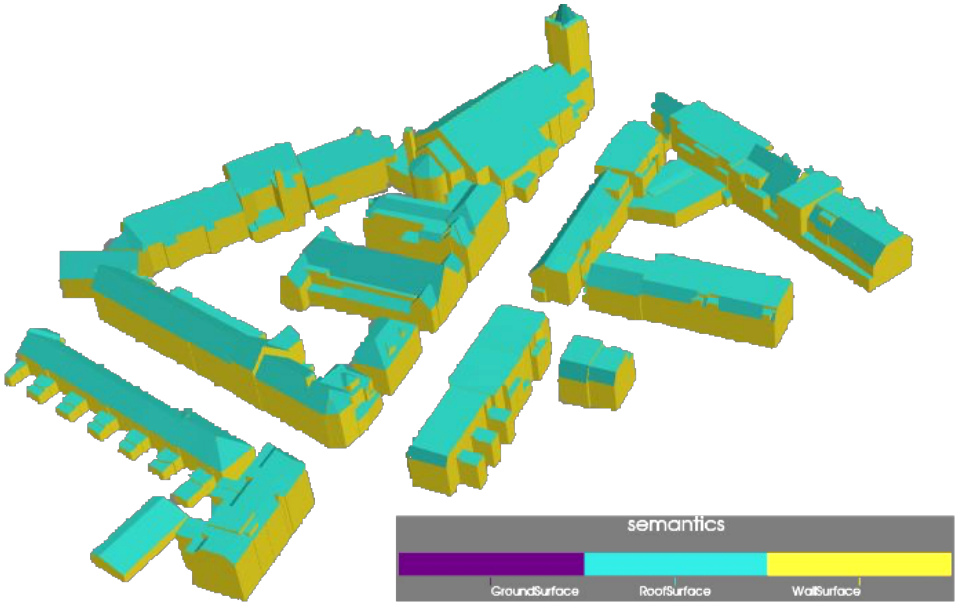


Figure 24. Semantic surface labels that are available in the 3DBAG

To divide the wall surfaces into exterior walls and party walls for a given building we use the following steps.

1. Find the neighbouring buildings using a 2D bounding box buffer of the current building.
2. Select all surfaces that are labelled as wall for all buildings in the buffer.
3. Cluster all wall surfaces based on their supporting 3D planes. During this operation co-planar wall surfaces from different buildings will end up in the same cluster. To do this we employ agglomerative hierarchical clustering. The planes are parametrisised using the general form $ax+by+cz+d=0$ and a Euclidean distance threshold of 0.1 is used during the clustering on the parameters *a*, *b*, *c*, and *d*.
4. In each cluster: intersect the wall surfaces from the current building with the wall surfaces from the adjacent buildings resulting in an intersection surface. If such an intersection surface is found, we have found a party wall. The parts of the walls without intersections are exterior walls.
5. Calculate the area of the exterior and party wall geometries that we found in step 4.

Figure 25 illustrates the result of this method for one building. This method was originally implemented by Labetski et al. (2023)[13], and further improved for this project. The source code can be found online[14].

13 Anna Labetski, Stelios Vitalis, Filip Biljecki, Ken Arroyo Ohori & Jantien Stoter (2023): 3D building metrics for urban morphology. International Journal of Geographical Information Science, 37(1): 36-67. DOI: 10.1080/13658816.2022.2103818. Code: https://github.com/tudelft3d/3d-building-metrics

14 https://github.com/3DGI/urban-morphology-3d

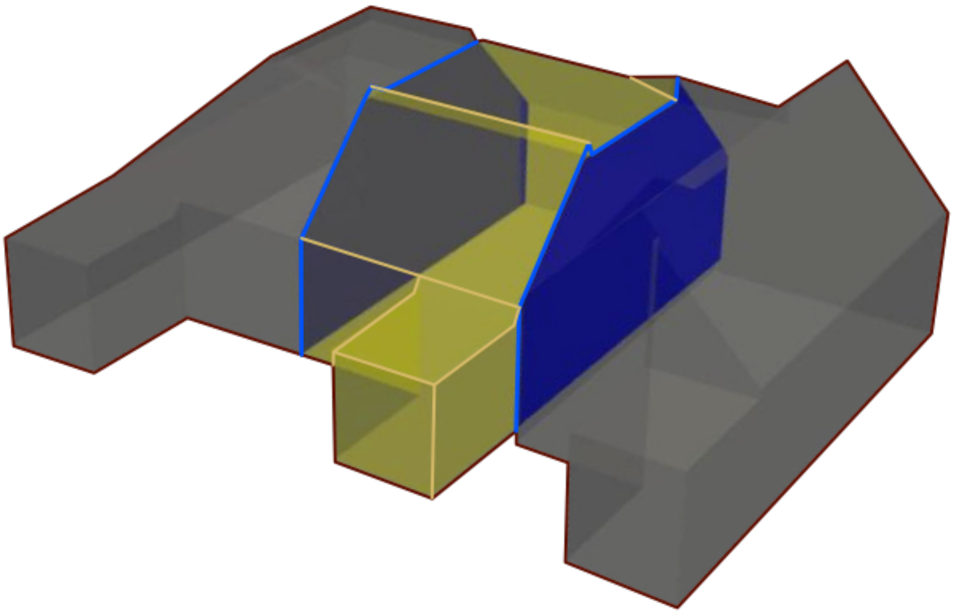


Figure 25. 3D visualisation of a BAG building (pand) and its two adjacent buildings (gray). The blue surfaces indicate the surfaces for which an intersection was found with the adjacent buildings, ie. the party walls.

3.2 Reliability indicators

As demonstrated in Chapter 2, there are some errors in the 3DBAG data that affect the volume and surface calculations. These errors can be caused by faults in the input datasets, eg. slivers and overlapping polygons from the BAG dataset, missing data from the AHN point cloud or by limitations of the 3D building reconstruction method. We cannot fix all these errors automatically, but we can provide an indication of which buildings are likely to contain an error. To do this we provide a number of 'reliability indicators'. Each of these quantifies a certain property that can help us determine how reliable the volume and area calculations are. We provide the following reliability indicators for each building:

1. Area of overlap between the BAG polygon of this building and other BAG polygons (`attribute b3_bag_bag_overlap`). Such overlaps could indicate an error in the BAG dataset such as duplicate geometries or point to complex buildings (see eg. Figure 26).
2. Geometric errors found in the 3D model using the val3dity library (attribute `b3_val3dity_lod22`). For instance, surfaces with an incorrect normal orientation will be detected with val3dity.
3. The difference in area between the input BAG polygon and the ground floor surface of the reconstructed 3D model (`_opp_grond_verlies`). This value is high for cases where the BAG polygon contains large areas of ground ('maaiveld') points in the elevation data, often indicative of underground structures that the 3DBAG does not reconstruct. These underground parts are filtered out during 3D reconstruction resulting in a difference between the original BAG polygon area and the area of the ground surface in 3D BAG.

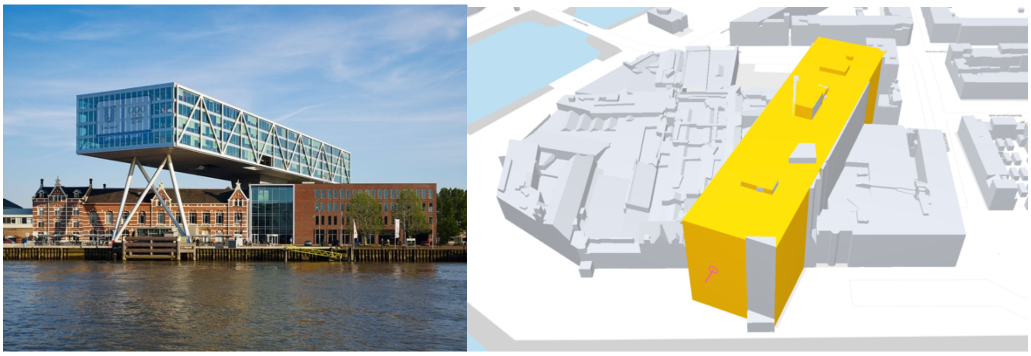


Figure 26. Example of a 'complex' building, photo (left) and 3DBAG (right). The roof of the box shaped Unilever building is is also used for the buildings below it, this is because we only have elevation data for the top most roofs. Because the BAG polygons of the Unilever building and the buildings below it will overlap, this case can be detected.

We assume that the computed surface areas and volumes are reliable if

1. `b3_bag_bag_overlap` is zero
2. `b3_val3dity_lod22` is empty, which means no geometric errors were detected
3. `_opp_grond_verlies` is relatively small, ie. less than 20% of the input bag polygon area. This criterium is conservatively chosen and could be relaxed further by choosing a higher percentage than 20%.

We combine these criteria into the following formula to decide if a building has reliable values for the surface areas and the volume:

```
_betrouwbaar = b3_bag_bag_overlap == 0 AND b3_val3dity_lod22 == "[]"
AND _opp_grond_verlies < ( _opp_bag_polygoon * 0.2 )
```

The resulting boolean attribute `_betrouwbaar` is **True** if the surface area and volume values are probably reliable and **False** if not.

Notice that this attribute is not 100% accurate, and it cannot be since we have no ground truth data. Nevertheless, we believe this is still a very useful and pragmatic attribute to quickly filter out buildings for which the computed wall and surfaces areas may be incorrect.

3.3 Integration with 3DBAG generation pipeline

Most of the implementation work for this project was integrated into the 3DBAG generation pipeline[15]. This means that the volumes and the surface areas for the different surface types are now automatically computed for every new 3DBAG release. Volumes are included since version 2023.06.22 and the surface areas are included since version 2023.10.08.

Notice that the 2021 release of the 3DBAG, that was also used for the analysis described in Chapter 2, was based on the now outdated AHN3 elevation data. Since the 2023.06.22 version the 3DBAG release is based on most recent AHN4 elevation data. In addition, the reconstruction process has been improved, based on experiences and feedback of the 2021 version. For more information about the improvements of the recent 3DBAG releases one can read the release notes[16].

[15] https://github.com/3DGI/3dbag-pipeline
[16] https://docs.3dbag.nl/en/overview/release_notes/

3.4 Specification of the delivered CSV file

The main result of this project is a data file that contains the surfaces areas of different types and the volumes for all buildings in the national 3DBAG dataset. This data file is delivered as a CSV (comma separated values) file that was extracted from the 3DBAG dataset version 2023.10.08. Table 11 lists the attributes that are included in the CSV file.

It should be noted that glazing areas are not included, since these are not present as objects in the 3DBAG and can therefore be estimated as a certain percentage of the exterior walls.

Table 11 Attributes that are provided in the delivered CSV file. The second column indicates where the attribute originated from. The label RVO means the attribute was specifically computed for this project and is currently not part of the 3DBAG dataset.

Attribute	**Dataset**	**Unit**	**Description**
`identificatie`	BAG	-	BAG identificatie code
`oorspronkelijkbouwjaar`	BAG	-	Original year of construction
`status`	BAG	-	BAG status
`b3_bag_bag_overlap`	3DBAG	m^2	Sum of overlap area of BAG polygon with other BAG polygons.
`b3_opp_grond`	3DBAG	m^2	Area of ground surface 3D model
`b3_opp_dak_plat`	3DBAG	m^2	Total area of flat roof surfaces of 3D model
`b3_opp_dak_schuin`	3DBAG	m^2	Total area of slanted roof surfaces of 3D model
`b3_opp_scheidingsmuur`	3DBAG	m^2	Total area of party wall surfaces of 3D model
`b3_opp_buitenmuur`	3DBAG	m^2	Total area of exterior wall surfaces of 3D model
`b3_pw_datum`	3DBAG		Acquisition date of AHN elevation data
`b3_volume_lod22`	3DBAG	m^3	Volume of 3D model
`b3_val3dity_lod22`	3DBAG		Error codes for geometric invalidities of 3D model
`_opp_bag_polygoon`	RVO	m^2	Area of BAG polygon
`_opp_grond_verlies`	RVO	m^2	`_opp_bag_polygoon` – `b3_opp_grond`
`_ratio_grond_tot_volume`	RVO		`_opp_bag_polygoon` / `b3_volume_lod22`
`_ratio_dak_tot_volume`	RVO		`_opp_bag_polygoon` / `b3_volume_lod22`
`_ratio_buitenmuur_tot_volume`	RVO		`_opp_bag_polygoon`

			/ b3_volume_lod22
_betrouwbaar[17]	RVO		Boolean attribute that gives an indication of the reliability of the b3_opp_* and b3_volume_lod22 attributes. 'True' means they are probably reliable. 'False' means there might be a problem with the source data or 3D model that causes incorrect values. See section 3.2.

4 Conclusions

In this report we investigated how to compute on a national level building volumes , surface areas (for roofs (flat and slanted), party walls, exterior walls and ground floors), and derived parameters in a way that is useful in practice. This information is relevant for building energy consumption estimation. The basis for these computations is our national 3DBAG dataset.

We performed a detailed analysis on the output of two methods to calculate the surface areas as well as several methods for volume calculation.
For the surface areas we find that both methods gave very similar results. Any discrepancies that we discovered, can be explained by faults in the input datasets, eg. slivers and overlapping polygons from the BAG dataset, missing data from the AHN point cloud, or by limitations of the 3D building reconstruction method. We cannot fix these problems automatically, but we can detect the building that are likely to have them. To do this we calculate reliability indicators, that we use to create the boolean attribute _betrouwbaar that can be used to filter out these potentially problematic buildings. Using this attribute one can easily filter out buildings that may have incorrect volume and surface area values in any analyses.
For the volume calculation, we discovered that our initial discretisation-based method proved to be inaccurate especially for buildings with a larger footprint area. We therefore implemented a different method that computes exact volumes and is still reasonably robust to possible errors in building geometry.

It should be noted that we did not have any ground truth data available to verify our results. We do believe that our results for the 3DBAG models to be accurate based on our comparison of the different methods. However, the 3DBAG models themselves are merely approximations of the buildings in reality and may therefore contain approximation errors which will have an impact on the calculated volume and surface areas. We expect these approximation errors to be small given the high accuracy of the input datasets, but we could not quantify them.

The calculation methods for the volumes and surfaces areas have been implemented in the 3DBAG generation pipeline. We have selected the methods based on their computational efficiency and ease of integration into the 3DBAG pipeline. With this integration, the data that we computed for this project will stay up to date with future versions of the open 3DBAG dataset and it will also be available to the general public. For this project specifically, we have exported the data into a CSV file that also contains a number of additional derived attributes.

[17] _betrouwbaar = b3_bag_bag_overlap == 0 AND b3_val3dity_lod22 == "[]" AND _opp_grond_verlies < (_opp_bag_polygoon * 0.2)