

A MULTI-HEIGHT LOD1 MODEL OF ALL BUILDINGS IN THE NETHERLANDS

B. Dukai¹, H. Ledoux¹, J. E. Stoter¹

¹ Delft University of Technology, Netherlands - (b.dukai, h.ledoux, j.e.stoter)@tudelft.nl

KEY WORDS: 3D building models, 3D building reconstruction, quality of 3D data, open data, roof classification

ABSTRACT:

The 3D representation of buildings with roof shapes (also called LoD2) is popular in the 3D city modelling domain since it provides a realistic view of 3D city models. However, for many application block models of buildings are sufficient or even more suitable. These so called LoD1 models can be reconstructed relatively easily from building footprints and point clouds. But LoD1 representations for the same building can be rather different because of differences in height references used to reconstruct the block models and differences in underlying statistical calculation methods. Users are often not aware of these differences, while these differences may have an impact on the outcome of spatial analyses. To standardise possible variances of LoD1 models and let the users choose the best one for their application, we have developed a LoD1 reconstruction service that generates several heights per building (both for the ground surface and the extrusion height). The building models are generated for all ~10 million buildings in The Netherlands based on footprints of buildings and LiDAR point clouds. The 3D dataset is updated every month automatically. In addition, for each building quality parameters are calculated and made available. This article describes the development of the LoD1 building service and we report on the spatial analysis that we performed on the generated height values.

1. INTRODUCTION

3D city models are increasingly used in urban applications in order to make cities cleaner, sustainable, better accessible, greener, CO₂-neutral, etc (Biljecki et al., 2016). Models of buildings are prominent objects in these models. The building models can be created at different Levels of Detail (LoD). If we take the terminology of CityGML (see Figure 1), a building can be modelled at four different levels of detail for the outer shell of the building: LoD0, LoD1, LoD2, and LoD3; and at LoD4, for the interior of the building.



Figure 1. The levels of detail for buildings in CityGML (Biljecki et al., 2016)

Representing buildings with roof shapes (i.e. LoD2) is popular within the 3D city modelling domain and many researches have carried out studies to automatically generate LoD2 models.

LoD2 provides a realistic experience and is therefore preferred in visualisations. In addition, due to the more realistic representation, LoD2 models are often thought to be more accurate than block models (i.e. LoD1). However, it is good to realise that LoD2 models are still an abstraction of the real world objects. In addition, simple block models of buildings are more suitable for many 3D data applications for example volume estimations or shadow analysis (Biljecki et al., 2015).

For many roofs the shapes can be generated fully automatically, but these are still approximations of the real shapes and objects such as chimneys, ventilation systems, or furniture on a roof terrace are not modelled. In addition, for more complex roof shapes or in case of insufficient number of height points, the automated reconstruction of roof shapes can be difficult. Moreover, in case of complex roof shapes, their abstraction is

often ambiguous. Especially in historic city centres where roofs of all shapes and sizes occur. See for example the case of “what is a flat roof” in Figure 7.

In contrast to LoD2 (and LoD3 models), LoD1 models for every building can be generated fully automatically from point clouds and 2D building polygons (i.e. footprints), which are increasingly available as open data. Therefore, LoD1 models are already frequently generated by various organisations based on building footprints and height points and are used in applications such as wind flow simulations, prediction of energy consumption and loss, and noise simulations. However, there are still some challenges in the reconstruction and use of LoD1 models since automatically generated LoD1 models for the same area can differ in their reference heights for various reasons.

Firstly, differences between independently reconstructed LoD1 models can be caused by differences in the reference heights used. For example, the upper surface of the block can represent the height of the roof edge, but also the ridge height or the maximum height (for example, chimney). Furthermore, the underlying statistical calculations might differ, e.g. is the height of the building block calculated as the average or the median of height points that fall in the building polygon? Finally, the used height points per building may be different, e.g. the number of points available for the statistical calculation can vary from one to hundreds of points (also depending on the acquisition method) and with this the accuracy of the calculated heights of the building models.

Many users are not aware that there are a large number of options for modelling buildings as blocks based on their 2D polygon, while these options do influence the outcome of analyses for which the LoD1 models are used (Biljecki et al., 2018). In addition, usually little is known about the quality of the automatically reconstructed LoD1 models because no metadata is kept.

Therefore, there is a need to standardise LoD1 models and provide insight into the method and quality of each generated building. In order to standardise possible LoD1 variations and to let the user choose which one to use, we developed a 3D building service that generates multiple reference heights per building based on statistical calculations, and that updates the 3D models monthly for all ~10 million buildings in the Netherlands (see Figure 2). In addition, the service provides insight into the quality of the generated building models.

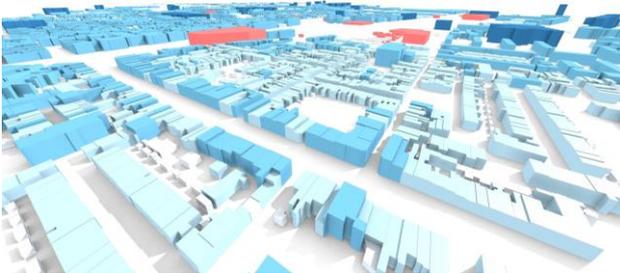


Figure 2. Visualisation of the generated 3D BAG data. Blue represents different height classes. The point heights for the red buildings are not representative, because the buildings are newer than the points (in this case, old buildings have been replaced by new ones)

2. OUR METHOD

We automatically generate LoD1 models from 2D footprints as registered in the building registration and point clouds obtained from airborne laser scanning. The source data for both types of data as well as the underlying software used are presented in Section 2.1.

Section 2.2 presents our method to calculate reference heights for all buildings; Section 2.3 presents our method to scale up to the whole country and keeping the models up-to-date. The quality parameters that we calculate per (and assign to each) building model are presented in Section 2.4. Finally, our method to identify flat roofs (which is important information to check the quality of reconstructed models) is presented in Section 2.5

What we add to existing LoD1 solutions is that we calculate various reference heights for extrusion (monthly) and additionally calculate and maintain metadata on how the data was created. We finally add parameters that describe the quality of the calculated reference heights of each building. This is especially important in an automatic process that is applied to 10 million buildings, whereby it is not possible to visually check all generated building models.

2.1 Source data and software used

2.1.1 BAG: The Building and Address register of The Netherlands (BAG) contains all buildings and addresses in the Netherlands. The geometry of addresses is collected as points and those of buildings as polygons (i.e. outline as seen from above). Municipalities are responsible for collecting the BAG data and keeping the data up-to-date. The BAG data is provided via the national geo portal PDOK both in a viewer and as download service. BAG also contains the history of buildings, i.e. buildings that are planned and buildings that have existed in the past but now have been removed.

We use the monthly BAG extracts provided by the NLEextract service (NLEextract, 2019). From the extract we only use a selection of the BAG buildings that have been realised and have not (yet) been demolished, so the data set represents the current situation.

2.1.2 AHN: The national height model of the Netherlands (AHN, 2019) is a point cloud acquired by airborne lidar systems. The first version of AHN (with a density of at least one point per 16 square meters, and in forests one point per 36 square meters) was completed in 2003. In the period of 2009 to 2012, the second version of the data set was acquired with an average point density of 10 points per square meter. Currently data for the third version (enriched with pulse count information) has been collected and is becoming available in chunks, after being validated and corrected. The AHN3 coverage for the complete Netherlands is expected in 2019. The resolution of AHN3 is similar to the one of AHN2. In addition, it contains a classification of the point cloud. For our research, we use AHN3 for the areas for which the data is available (about two third of the country). For the other areas, we use AHN2.

AHN data is automatically downloaded from the governmental data server in our solution. We use AHN3 (acquired between 2014-2019) where already available and otherwise AHN2 (acquired between 2007-2012). The points are classified in both versions, albeit in different ways. For AHN2 we use "filtered out" to determine the building heights and "filtered" points for the height at ground level. For AHN3 we use the classes "building" respectively "ground points" to determine building heights and heights at ground level.

2.1.3 Software 3dfier: For the extrusion of building we use the open source software 3dfier (Commandeur et al., 2019). This open source software is being developed in collaboration between the 3D Geoinformation research group of TU Delft and the Kadaster. The software automatically generates 3D city models based on 2D topography and point clouds, also for other objects than buildings and ensures that generated 3D building models connect to the terrain.

2.2 Method to calculate reference heights

To address the need for different reference heights as well as providing metadata on these heights, we have developed a solution whereby different reference heights are calculated as well as a set of relevant quality values. For the height of blocks, we calculate six percentiles with the 3dfier software based on all "building" (AHN3) or "filtered" (AHN2) points that fall in each building polygon in order to capture six different reference heights of each building, see Figure 3.

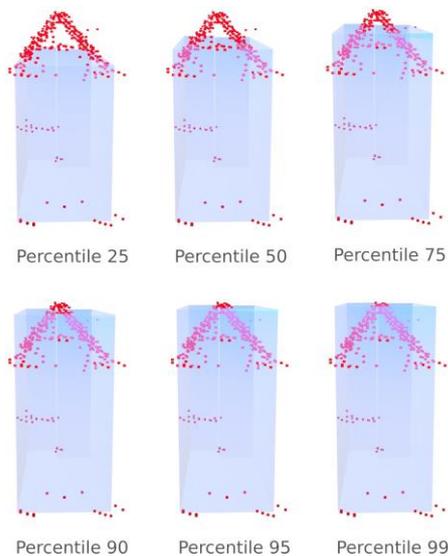


Figure 3. Visualization of the calculated reference heights based on percentiles and the height points that were used in the calculation for each building

For ground level height references (the minimum height of a 3D BAG building from which a building is extruded), i.e. 0, 10, 20, 30, 40 and 50 percentile heights, are calculated based on “filtered out” and “ground points” in a 0.5m buffer around the building.

The calculated ground level and reference heights are all added as attributes to the 2D BAG geometries. The user can then extrude the 2D building to a height that best suits her needs. The AHN version (2 or 3) used for each building is also added as an attribute. The data set and the LoD1 service is available at Dukai et al. 2018. Users can view, query and download 3D BAG data via this website. The generated dataset for the whole of the Netherlands is available as a WFS and WMS service, GeoPackage (7 GB), PostgreSQL backup (1.5 GB) and a CSV file without geometry (2 GB). The source code of the software “bag3d” with documentation is also open and freely available (Dukai, 2019).

2.3 Scale up and keep up to date

The BAG contains around 10 million buildings and the AHN2 639 billion height points (AHN3 is expected to contain around 700 billion points). Together these data sets cover more than one terabyte of storage space. We use a so-called “tiling” approach to scale up our process for the whole of the Netherlands. We use the AHN’s 1377 original tiles for this. On our server, 30 tiles can run simultaneously and we can generate the LoD1 BAG in about three days. Buildings that are overlapping more than one tile is assigned to the tile in where its centroid lies. The adjacent AHN data is used as well for determining the height.

Every month we check whether there is an update available from NLEextract and AHN, and we generate the entire LoD1 BAG in this way again. The updates ensure that new buildings are added, demolished buildings are removed and AHN3 is used if available. Buildings that have not changed in reality remain exactly the same in our data set. Older versions of the LoD1 remain available (and downloadable) on the site.

2.4 Quality parameters

In addition to the different reference heights from which a user can choose, we calculate and assign quality-related information. Firstly, we add an attribute per building that states when the used AHN tile was generated. This attribute indicates, together with the construction date of a building (available from the BAG), when a building is newer than the used height points which makes the calculated height reference invalid.

In addition, we determine on the basis of a spatial analysis per building how many height points are available for the calculation for that building and assign this as attribute. Here we make a distinction between available points at ground level and points available for building roofs.

In some cases, no points are available. This can have different reasons. An obvious explanation is when there is a new building. But missing height points can also be caused by an incorrect classification of the AHN points, see Figure 4.

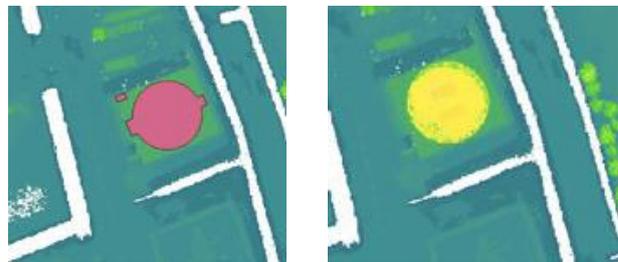


Figure 4. The BAG building on the left has no height, because the relevant point heights are classified in AHN as “ground points” (right)

In addition, missing points for a building can be due to a laser beam that cannot reach the ground, such as the ground level at a building that is completely occluded by other buildings or buildings under bridges, see Figure 5.



Figure 5. Buildings under bridges have no corresponding heights in AHN. Picture from © Google

A few models have missing ground height. The main reason for this is an occasional, slight misalignment of the footprints and the point cloud. In these cases the small search radius (0,5m) is not sufficient to bridge the gaps of missing AHN points (usually due to occlusion at the foot of the building), and the buffer of building points, thus no ground points are found for the model (Figure 6).

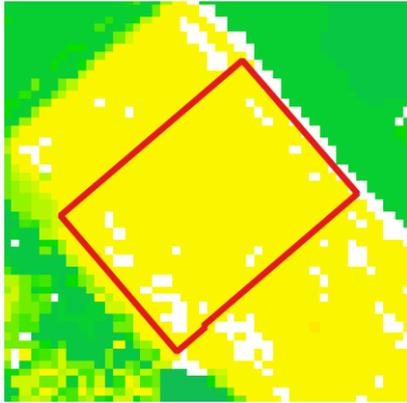


Figure 6. A building model (red) overlaid on the rasterized AHN3 (cellsize 0,5m, colour gradient by height where green is low, yellow is high elevation, missing data is white). In case of this particular model, the yellow cells are classified as “building”, the green cells are classified as “ground” in AHN3.

In an earlier version, also points were missing since AHN3 tiles were not yet completely covered (which was not documented). In order to use the AHN3 (i.e. newer) heights for these tiles where they are available, we generate the 3D BAG here twice: once with AHN2 and once with AHN3, after which the most recent heights are selected for each building. This operation is less expensive than making an area selection in the relevant AHN2 and AHN3 files.

Another quality parameter that we calculate is the Root Mean Square Error to show to what extent a block model actually approaches the building, i.e. to show the height difference between the AHN points and calculated reference heights. The RMSE of the geometric difference between the point cloud and the constructed 3D building model is calculated for each building and for each percentile and we add this quality data as attributes. Finally, we identify and assign to each building whether the roof is flat or not (see for more details in next section)

All quality information, i.e. information about the “too old” height data, missing input points, RMSE per percentile, flat roofs and the number of points that were included in the statistical calculations is recalculated and saved for each update. Table 1 provides an overview of the attributes that we create as part of the 3D building service.

Attribute	Description
ground-00, ground-10, ground-20, ground-30, ground-40, ground-50	The height of the ground surface of the building at the given percentile.
roof-25, roof-50, roof-75, roof-90, roof-95, roof-99	The height of the roof surface of the building at the given percentile. For example, roof-99 is the height of the

95, roof-99	building when the roof surface is set at the 99th percentile of the z-coordinates of the point cloud of the building.
rmse-25, rmse-50, rmse-75, rmse-90, rmse-95, rmse-99	The Root Mean Square Error or the geometric difference between the 3D building model and the point cloud that was used for generating the model. Or, the average discrepancy in meters between the 3D building model and the real-world building. This measure also accounts for the whole building, not only the roof.
roof_flat	Indicates that roof of the real-world building is flat or not. The value 1 means a flat roof, the value 0 means a not flat roof.
nr_ground_pts	The number of points in the point cloud that were used for determining the ground-height of the building model. If this value is 0, that means that the ground-points are missing from the point cloud at given model.
nr_roof_pts	The number of points in the point cloud that were used for determining the roof-height of the building model. If this value is 0, that means that the roof-points are missing from the point cloud at given model.
ahn_version	The version of the AHN that was used to obtain the height information.
ahn_file_date	The creation date of the AHN file that was used to obtain the height information. This is not the same as the timestamp of the LiDAR points of a particular building.
height_valid	0 – invalid height, because the building was built after the point cloud was acquired; 1 – valid height, because the building was built before the point cloud was acquired
tile_id	The ID of the tile where the building belongs to.

Table 1. The attributes that the 3D dataset contains additional to the attributes from the original building registration.

2.5 Identifying flat roofs

An important first step in modelling roof shapes and identifying how well an extruded building block approaches its real-world counterpart, is the detection of flat roofs. Buildings with flat roofs allow to determine the accuracy for LoD1 buildings because, in theory, the geometry of LoD1 and LoD2 should coincide (Section 3.1).

It seems obvious that a human can identify “a flat roof”. But looking at the built reality, this is not always true. Take for example the roofs in Figure 7. In a (small and non-representative) survey we asked 28 people to indicate for these buildings whether or not it has a flat roof in their opinion. The respondents were unanimous in case of buildings with a single, horizontal roof surface (4, 5, reading from left to right). But for three buildings (1, 3, and 7, reading from left to right) the opinions were divided. If it is difficult to unambiguously categorise reality through human interpretation, how do we formalise this into an algorithm?

Also interesting is that respondents indicated that the definition of a "flat roof" can have different meanings for different applications. For example, a waste water engineer is interested in knowing when the actual flat surface is at least 50% of the total surface area, significantly reducing water flow, while a real estate developer is interested in those roofs that fit a (small) terrace, and a climate expert from the municipality wants to know on how many roofs it is possible to realize a green roof.



Figure 7. Are these roofs flat or not? Images from © Google

Thus "flat roof" is in itself a misleading term, and in our method we opted for the strictest definition, in which only those buildings have a "flat roof" that have a single, horizontal roof surface (4, 5, reading from left to right in Figure 7).

In the first iteration of the 3D BAG we use a rudimentary but computationally cheap method for identifying flat roofs. It exploits the fact that AHN was collected through aerial laser scanning, thus most points fall on the roof of the buildings. The distribution of height values of the point cloud of a building is in most cases multimodal, where the modes describe a roof part, the tails usually contain points on the walls. We assume that for each building the variance of the distribution with the largest mean gives the variance of the roof. In a manually labelled set of 470 buildings we used a Gaussian Mixture model to identify the clusters of flat- and not-flat roofs based on their variance (Figure 8). In the 3D BAG we use these clusters to label each building. In our sample data set we estimated an 84% accuracy. According to this method, we determined that about 34% of the buildings have a flat roof in the Netherlands.

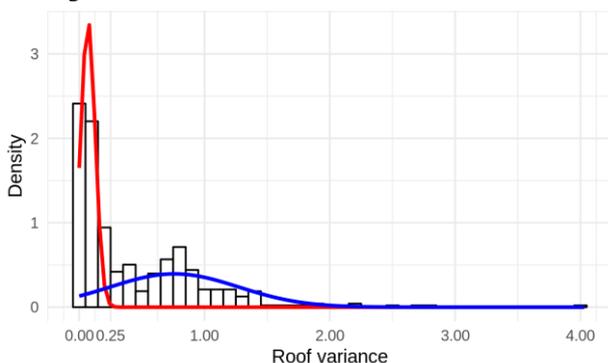


Figure 8. Gaussian Mixture model and histogram of roof variances in the sample data set. Flat roofs (red) and not-flat roofs (blue). Sample size 470 buildings.

3. RESULTS & DISCUSSION

In this section we present and analyse the results in order to provide insights into the quality of reconstructed LoD1 building models. In the subsections we present respectively a statistical summary of the whole data set and some of its quality parameters (Section 3.1), and analyses of the RMSE of different percentiles in Section 3.2.

3.1 Quality of the extruded building models

The quality parameters provide the possibility to calculate the quality for the whole LoD1 data set and monitor this through updates. For each update, these statistics are summarised and published on the 3D building service website to provide insight into the quality of the entire 3D building dataset.

We calculate the percentage of buildings that do not have corresponding point cloud in the AHN. This value indicates the timeliness of AHN in relation to the building stock on the country level. Many consider the AHN severely outdated due to its update cycle that spans over about 5 years. In the March 2019 version, 3.9% of the approximately 10 million were found to be newer than the height data used. Additionally, we determine the number of buildings for which no AHN height points are available. For the March 2019 version, 2% of the buildings have no roof height and 2.3% have no ground height.

The information about the timeliness of the AHN and coverage of the point cloud per building is comprised in the "invalid height" attribute. Thus the height of a building is "invalid" when the building is newer than the AHN or the AHN does not have points in the appropriate class that cover the building. The "invalid height" is a boolean value that helps to quickly filter unreliable building models. Table 2 shows that both in case of AHN2 and AHN3, 5% of the models are invalid, although AHN2 covers only about 30% of the buildings.

AHN version	Invalid height	Amount of buildings
2	5%	29%
3	5%	71%

Table 2. Amount of models with invalid height per AHN version. As of March 2019.

3.2 RMSE of LoD1 models

In order to gain insight into the extent to which the geometry for the different percentiles approximates the actual building, we have reviewed the RMSE for all the buildings in the Netherlands. The RMSE calculation includes both the points on the walls and on the roof.

Figure 9 shows the distribution of the RMSE for all the building models in the Netherlands. The clearly bimodal distribution indicates that the variance of the distances between point cloud and model, can be exploited for identifying flat roofs, at least in large numbers. We used this property in our method for

classifying the roof types (see Section 2.5).

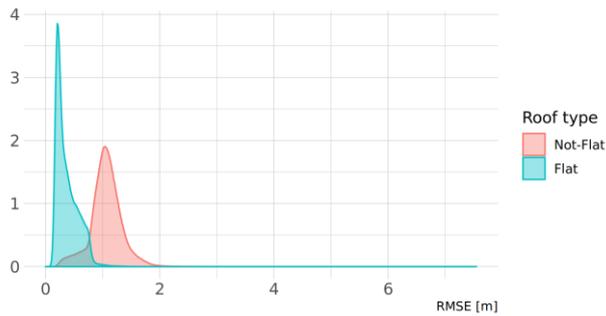


Figure 9. Distribution of RMSE of the geometric difference between the point cloud and building model for all the buildings in the Netherlands.

A close look at the RMSE in Figure 10 reveals that the median RMSE is 1.06 meters for not-flat roofs and 0.31 meters for flat roofs. Also, there is minimal variation in RMSE across all percentiles in buildings with flat roofs, while in case of not-flat roofs the RMSE is inversely proportional to the percentile. For buildings with not-flat roofs, the RMSE remains below 1 m for percentiles 75-99, from which we can conclude that automatically generated LoD1 buildings are indeed suitable for most GIS analyses.

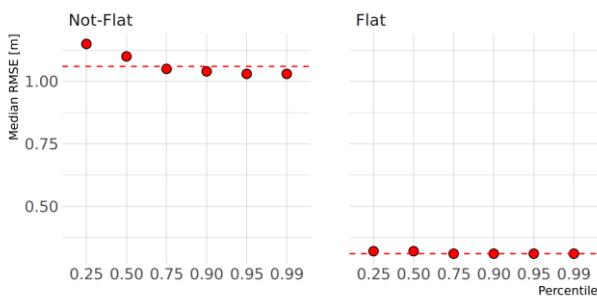


Figure 10. The median RMSE of the geometric difference between the point cloud and extruded 3D building models, analysed for both non-flat roofs (right) and flat roofs (left). The median per roof type is indicated with the red dashed line, which is 1.06 m for non-flat roofs and 0.31 m for flat roofs.

Looking at the calculated building reference heights we can observe a clear distinction between buildings with flat and not-flat roofs. In Figure 11 we plotted the median building height in the municipalities of Netherlands per roof type. The building height is calculated as the distance from the ground and the model height at 95th percentile. The median height of buildings with flat roofs is around 2.5 meters, with relatively low variation. From this we infer that our method for classifying roof types mainly identifies small structures with a single horizontal roof (eg. garages and alike). This is accordance with our strict definition of a “flat roof”.

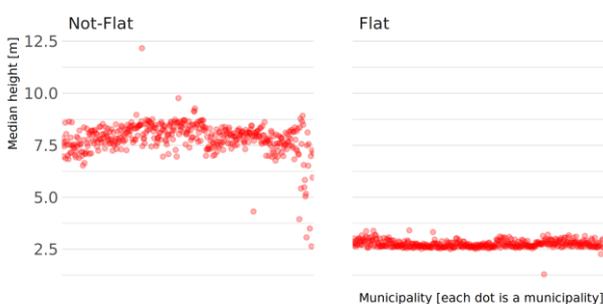


Figure 11. Median building height in municipalities of the Netherlands: Left for non-flat roofs and right for flat roofs. Each dot represents a municipality. The building height is calculated as the distance between the ground and the model height at 95th percentile.

4. CONCLUSION

In this article we have described our 3D building service, which generates six different reference heights per building for the whole of the Netherlands in order to extrude a building footprint and obtain a LoD1 building model. The service provides insight into the quality of the data generated, i.e. for each building quality parameters are calculated and these parameters are also summarized for the whole dataset for each (monthly updated) version. The quality information per building allows a user to choose which reference height she uses in her application and possibly to interactively correct 3D buildings that are less reliable.

The buildings are built on the basis of the airborne LiDAR dataset of the Netherlands. For more up-to-date height data, future work will study the use of point clouds obtained from matching aerial images which are acquired every year. Future work will also focus on automatically generating more detailed 3D building geometries such as LoD2 and LoD1.3 (i.e. LoD1 building with different reference height for one building to model height jumps in buildings such as a church with a tower or a garage attached to a building, see Biljecki et al. 2016), and improve the classification of roof types.

Finally, we will explore how our 3D building service can be integrated with 3D buildings models as acquired by municipalities, i.e. use our automatically generated height information as initial values which can be replaced by municipalities with values acquired by for example terrestrial measurements.

5. ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 677312 UMnD).

6. REFERENCES

AHN, 2019. www.ahn.nl (14 May 2019)

Balázs Dukai. 2019. tudelft3d/bag3d: Zenodo DOI (Version v1.0.1). Zenodo. <http://doi.org/10.5281/zenodo.2806867>

Balázs Dukai, Jantien Stoter, Hugo Ledoux. 2018 3D Registration of Buildings and Addresses (BAG). 4TU.Centre for Research Data. Dataset. <https://doi.org/10.4121/uuid:f1f9759d-024a-492a-b821-07014dd6131c>

Filip Biljecki, Gerard B. M. Heuvelink, Hugo Ledoux and Jantien Stoter. 2018. The effect of acquisition error and level of detail on the accuracy of spatial analyses.. Cartography and Geographic Information Science 45(2), 2018, pp. 156–176.

Filip Biljecki, Hugo Ledoux, Jantien Stoter, 2016, An improved LOD specification for 3D building models, Computers, Environment and Urban Systems 59, 25-37

Filip Biljecki, Jantien Stoter, Hugo Ledoux, Sisi Zlatanova and Arzu Çöltekin. 2015 Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information 4(4), December 2015, pp. 2842–2889.

NLEExtract, 2019. <https://nlextract.nl> (14 May 2019)

Tom Commandeur, Hugo Ledoux, Ravi Peters, Maarten Pronk, Balázs Dukai, Filip Biljecki, & Anna Labetski. 2019. tudelft3d/3dfier: 3D BAG version 1.0.0 (Version 3dbag-v1.0.0). Zenodo. <http://doi.org/10.5281/zenodo.2807056>