Rebuilding the Cadastral Map of The Netherlands, the Geodetic Concept

Frank van den HEUVEL, Gerbrand VESTJENS, Gerard VERKUIJL, Mark van den BROEK, The Netherlands

Key words: cadastral map, geometric quality, network adjustment, statistical testing

SUMMARY

The Dutch cadastral map is a nation-wide, topologically correct index for the cadastral registration. It is available as open data. However, its positional quality of about half a meter is not considered to be sufficient in a future where people want to derive the exact location of their legal boundaries from the digital map themselves. A research program aiming at renewal of the cadastral map was started and we succeeded in building a prototype software that is able to largely automatically read and vectorise the more than 5 million historical field sketches containing the original survey information. The prototype facilitates connecting the resulting local networks to gather the information for large-scale network adjustments that result in the new geometry of the cadastral map stored in the so-called reconstruction map. Currently, we are in the middle of a pilot project in which we will produce the reconstruction map from measurements extracted from about 10 thousand field sketches.

The geodetic concept adopted for the production of the reconstruction map is based on the Delft method of testing where quality control is performed in all steps of the process. This starts with the adjustment and testing of the measurements of the many small survey projects individually, of which the measurements are stored in the field sketches. After georeferencing the survey projects are connected using corresponding points in the overlap between the projects. All measurements are weighted and the so-called idealisation precision is accounted for in relation to the type of point. With every newly added project the redundancy improves, the network is re-adjusted, and the measurements are tested for errors. In this way the geometric base for the new cadastral map is being built while at the same time errors in the measurements are eliminated.

In this paper intermediate results of our research on the geodetic challenges in building the reconstruction map are presented. The main challenges are: the large number of field sketches, the variability of their content, and the related number of errors in combination with a limited redundancy. Furthermore, we are investigating how to cope with the limitations in network size, as a nation-wide integral adjustment is not feasible. A fast and large-scale adjustment software is under development, as well as a procedure for renewal of the cadastral map based on the results of the large-scale adjustments.
1. INTRODUCTION

An introduction and motivation to the research project “Rebuilding the Cadastral Map of The Netherlands” is given in (Hagemans et al., 2020). There it is outlined that the goal of the project is to build the so-called “reconstruction map” from the survey measurements that have been gathered since the start of the Dutch Cadastre in 1832 in millions of field sketches. The reconstruction map is the improved version of the cadastral map, and will replace the official cadastral map in the future. An overview of the process from field sketches to reconstruction map is presented in Figure 1. Steps 2, 3, and 4 will be discussed in this paper in more detail.

That it is deemed to be feasible to extract and process the many millions of historic survey documents or field sketches is credit to the developments in artificial intelligence in recent years. Artificial intelligence is essential in automating the digitization of the field sketches that hold the original survey measurements. An example of a field sketch is depicted in Figure 2. More details on this topic can be found in (Franken et al., 2020).

In this paper we focus on the geodetic aspects of building the reconstruction map: the adjustment and validation of the historic survey measurements, the large geodetic network adjustments involved, and finally the integration and improvement of the cadastral map. In the next section we outline the approach adopted for this map renewal process.

![Figure 1: Overview of the four-step approach to renewal of the cadastral map.](image-url)
2. THE APPROACH FOR REBUILDING THE CADASTRAL MAP

An overview of our approach for rebuilding the cadastral map is depicted in Figure 1. There are two main sources of information available as a starting point for rebuilding the cadastral map of The Netherlands. Firstly, there is the current cadastral map, and secondly, there are around 5.5 million so-called field sketches available. The purpose of the cadastral map is to give an overview of the parcels shape and location; the geometry is known to be relatively poor, and errors up to about 1 meter can occur. The boundaries of parcels are more accurately registered in the field sketches. In a field sketch (see an example in Figure 2) the cadastral surveyor has noted the field measurements that connect the boundaries to reference points such as corners of buildings or benchmarks in the terrain. As described in (Franken et al., 2020) the digitization of the field sketches is at the basis of the improvement of the cadastral map. An example of the digitization is shown in Figure 3. That this first step in our procedure is a challenging project in itself is not only related to the number of field sketches: the field sketches can be up to two centuries old, and as a result they are quite diverse in their content and quality. Furthermore, the field sketches contain only a subset of the current boundaries as a result of splitting and merging of parcels. In fact, it is possible to reconstruct the historical (time-dependent) cadastral map from the field sketches. Currently, we focus on the improvement of the present cadastral map.

Figure 2: An example of a field sketch. The Dutch Cadastre has scanned all 5.5 million field sketches of which some date back to the early 19th century.
In the second step of our procedure field sketches are positioned on the map and connected by the identification of common points. We will not violate the fundamental geodetic rule of “working from large scale to small scale”, however, the process that aims at improvement of the cadastral map starts with the analysis of the measurements of the individual field sketches. In this way the measurements are adjusted and tested in an early stage and errors can be corrected and the approach known as DIA (Detection, Identification, Adaptation) is implemented (Teunissen, 2000). Specifically the method for statistical testing as applied in DIA is often referred to as the 'Delft School' of Mathematical Geodesy, and was developed more than 50 years ago by prof. Baarda (Baarda, 1968). Field sketches that overlap, i.e. have points in common, are connected and larger networks are built. Networks constructed using up to about 50 field sketches are analysed with the geodetic network adjustment software MOVE3 (MOVE3, 2020), our choice for the implementation of DIA. We refer to these adjustment as small-scale adjustments and these are the topic of section 3.

![Figure 3: The above field sketch, overlayed with the result of the vectorization. Buildings have been detected, parcel numbers (red boxes) and measurements (yellow boxes) read.](image)

For larger networks we have developed adjustment software in-house that is limited in functionality but significantly faster. The process of vectorizing field sketches and small-scale adjustments of up to 50 connected field sketches is outsourced to external parties. Large-scale adjustment and improvement of the cadastral map are performed at the Cadastre. These third and fourth step of the procedure are currently under development. Large-scale adjustment
aims at the integral adjustment and testing of the observations of hundreds of field sketches. First results and an assessment of the improvement in speed of the newly developed adjustment software is presented in section 4.

We anticipate a total of more than hundred million network points for the whole country. Therefore it is not possible to adjust all observations in a single adjustment computation. As all the measurements are between points that are relatively close together, say up to 1 kilometre, it is easy to split the adjustment in several (or many) parts. However, this implies that the large-scale adjustments will have to be connected, because these networks will have common points that will obtain different coordinates from different adjustments. This topic is discussed in section 4.1.

In section 5 the final step of the procedure is presented: the improvement of the cadastral map. This step is implemented as an adjustment for connecting two point fields. The coordinates resulting from the connected large-scale adjustments constitute the primary point field. Its approximated covariance matrix is derived by error propagation of the precision adopted for the original measurements such as distances and angles, and verified in the small-scale adjustments. The cadastral map is regarded as a point field with a quality description in the form of an artificial covariance matrix that reflects the correlation between the coordinates. The connection between the two point fields is established by the identification of identical points as well as geometric relations. Geometric relations are needed because many points of the cadastral map are not directly surveyed but constructed using the survey measurements.

That the approach outlined above is “working from large scale to small scale”, is guaranteed by the large set of more than a million reference points distributed over the country. These points have known coordinates with high precision (centimetre level) in the national RD coordinate system and are used to connect to in the large-scale adjustments. The small-scale adjustments have the detection and correction of measurement and vectorization errors (DIA-approach) as their primary goal, and can in principle be performed in a local coordinate system.

3. NETWORK ADJUSTMENT FOR FIELD SKETCHES

3.1 Validation and positioning of individual field sketches

Once the field sketches have been vectorized, the measurements from the individual field sketches can be adjusted to find errors in the vectorization. Typical errors are errors in the automatic recognition of written tape measures or starting points of measurement lines. Since the field sketches are built up in a way that the boundaries can be reconstructed in the field based on measurements to monumented points or buildings, the field sketches do not contain sufficient information to solve all points in the adjustment. To solve the rank deficiency in the adjustment the strategy used is adding all points as weighted constrained points with their initially computed value and with a standard deviation of 20 m. Thus a solution can always be computed and there where there is redundancy, the measurements can be tested for outliers. Typically, when using the cadastral settings for the standard deviations
for measurements and idealization precision, the marginally detectable bias is at the decimeter level. Thus errors of less than 0.10 m will not be detectable at this stage.

For the field sketch in Figure 2 and Figure 3, testing results and error detection shows an error of about 0.20 m in the tape distance between point p171 an p22, as shown in the output of the network adjustment software MOVE3:

<table>
<thead>
<tr>
<th>Record</th>
<th>Station</th>
<th>Target</th>
<th>Test</th>
<th>Factor</th>
<th>Red</th>
<th>Est err</th>
</tr>
</thead>
<tbody>
<tr>
<td>83</td>
<td>Tape distance</td>
<td>p171</td>
<td>p22</td>
<td>W-test</td>
<td>1.6</td>
<td>23</td>
</tr>
</tbody>
</table>

Tape Distance was vectorized as 9.76 m, but a visual check showed that 9.96 m is more obvious. The 0.20 m difference is also confirmed by the 0.19 m error estimate. Because of the use of 20 m standard deviation for all points the standard deviations after adjustment will also show large values, representing the inaccuracy in absolute positioning. Fixing 2 arbitrary points to zero standard deviation will give a better representation of the relative precision of the points in the field sketch. The typical average precision will be a few decimeters. The redundancy numbers of the measurements show that about 65% of the distance measurements have reasonable to good control. About 35% of the measurements have weak or no control (Figure 5).

Initially, the adjustment is done in an arbitrary local system to focus on error detection and improvement of relative geometry. Once the field sketch adjustment has an accepted overall F-test and no rejected w-test for measurements are identified, the field sketch can be located in the Dutch RD system. An automated algorithm is used to find an approximate location on the cadastral map, with the aid of parcel numbers and comparing distances between points on the field sketch and map. This automated process has a success rate of about 87% in urban areas and 52% in rural areas where less building information is available. The remaining field sketches must be manually located using at least 2 points identified in the map. Usually this is done via buildings that are visible on both the field sketch and the cadastral map. After a successful localization of individual field sketches they are linked in the next step.
3.2 Adjustment of clusters of field sketches

Linking field sketches will improve redundancy, and thus allow testing of measurements that could not be tested when adjusting individual field sketches. Testing of the links ensure that the proper points are linked together. We performed a test in which 65 field sketches were linked. These field sketches cover an area of about 2 by 2 km. Figure 6 shows a screenshot of the specially designed tooling called VeCToR (Franken et al., 2020), that is used to specify the links between points of overlapping field sketches and facilitates the adjustment of the complete network.

The adjustment contains 4747 unique points and thus 9494 unknown coordinates to be adjusted using 10822 measurements and 2412 points linked. After adjustment some points...
will show large standard deviations, indicating that the precision of these points is weak (Figure 7). This is due to a lack of observations for these points. Adjusting the 65 field sketches takes approximately 1 minute on an average laptop (3 iterations). Tests have shown that a solution with 3 iterations is practically identical to a solution with more iterations. If sufficient GPS measured reference points with a fixed standard deviation of 2 cm are available, the average a posteriori standard deviation is about 4 cm. Here we linked to 62 reference points. The reliability as expressed in the Minimal Detectable Bias (MDB) is shown in Figure 8. MDB is the size of the error that can be detected in the W-test with a probability of 80%. The figure shows that the average size of the error that can be detected in testing is about 0.15 m, smaller errors will be considered as measurement noise.

Figure 9 shows the network in MOVE3 after adjustment. Before the adjustment the coordinates are based on the preliminary coordinates resulting from the positioning step: see Figure 6. After the adjustment the linked points coincide.

Figure 8: Marginal Detectable Bias Tape measurements (red) and Chainage and Offsets (blue).

Figure 9: MOVE3 screenshot after adjustment of 65 field sketches (southern part).
4. LARGE-SCALE NETWORK ADJUSTMENT

4.1 Speeding-up the adjustment
A single field sketch contains ca. 100 points on average, which means 200 variables (x,y-coordinates) to be estimated by solving the adjustment problem. The number of observations is typically a factor two higher. In our specific use case however, ultimately we are not dealing with single field sketches but with much larger networks. The Netherlands has geographically been divided in municipalities, which themselves are again divided in sections. A section easily corresponds to 10k field sketches resulting in an adjustment problem with in the order of a million variables. The extreme case is given by combining all measurements of the whole of the Netherlands together in one adjustment problem. We estimate the size of such a problem at around 500 million variables and in the order of a billion observations.

The software currently used at the Dutch Cadastre for performing adjustments, geodetic network adjustment software MOVE3 (MOVE3, 2020), was not developed with such large networks in mind. After a short feasibility study on extending MOVE3 to be able to handle such networks, it was concluded that a separate stand-alone solver for solving large networks was to be developed. The main requirement for this solver was solving large cases at speed, at a high enough level of accuracy. As such, the choice was made to only consider a 2D-representation of the coordinates (as opposed to the 3D capabilities of MOVE3). Only the features deemed essential were chosen to be implemented for this solver.

On a high-level, the solver works as follows:

1. **A weighted least-squares adjustment problem is defined from the given variables and observations.** Various observation types are supported, the most important ones being tape distances, collinear points, and perpendicular angles.
2. **The resulting non-linear problem is solved using an iterative method called Levenberg-Marquardt (Marquardt, 1963).** This method effectively linearizes the problem at each iteration and, at each iteration, a sparse direct solver is used to solve the linearized problem.
3. **Once the solution is deemed converged, optionally some statistics are calculated to get more information on the reliability of the solution.** These include the so-called redundancy numbers, marginal detectable errors, w-tests (for each observation) and the precision ellipses (for each point).

**Ad 2. On the Levenberg-Marquardt algorithm:** this algorithm is not guaranteed to find the global minimum, but only a local minimum. The Levenberg-Marquardt algorithm interpolates between the Gauss–Newton algorithm and the method of gradient descent. The Levenberg-Marquardt algorithm is more robust than the Gauss–Newton algorithm, which means that in many cases it finds a solution even if it starts very far off the final minimum. It tends to be a bit slower than the Gauss–Newton algorithm but we still prefer it because of the more robust behaviour. Finally, the Levenberg-Marquardt algorithm can be regarded as a Gauss–Newton
algorithm using a so-called trust region approach, which means that the new solution from one iteration to the other is constrained to be close to the current solution.

*Optimizing the solver for speed*

As it turns out, the overall bottleneck (in terms of computational time) is not in step 2 where the pure adjustment problem is iteratively solved, but in step 3 where some statistics of the solution need to be calculated. However, we looked into optimizing both steps with respect to speed since we do not always need the extra statistics to be calculated. We will now discuss the two steps in more detail with respect to optimizing the (wall clock) time of these steps.

To solve the adjustment problem fast, two things are essential:

1. Fast calculation of the residual vector and the Jacobian matrix at each iteration of the Levenberg-Marquardt algorithm. This was achieved by *vectorizing* the calculation of these elements (per observation type). This was all done in Python, using the Numpy library.
2. Having a fast solver for the linearized problem. We have experimented with open source alternatives and found PyPardiso to be the fastest. This is effectively a Python interface to the Intel MKL Pardiso library to solve large sparse linear systems of equations.

We tested the performance of our solver on a set of self-constructed test cases of increasing size. We report our timings in Table 1. Since the timings are driven by the stop criterion used, we also report the number of iterations carried out. The calculations were done on a single laptop (32GB Ram, Intel Xeon 2.7 GHz CPU, 6 cores). We also report the timing of the case detailed in section 3.2 of this paper (with 9494 variables): this case was solved in 0.17 seconds for one iteration. We did not explicitly compare performance with MOVE3 on all test cases but observed a speedup of approximately 2 orders of magnitude on a medium sized problem (10k variables) and also achieved a better scalability with respect to problem size.

Of all the statistics we would like to have, the covariance matrix of the solution, which is effectively the inverse of the normal matrix, would be the bottleneck in terms of calculation time. After that, the calculation of the so-called redundancy numbers (for each observation)
would form the bottleneck. To calculate these redundancy numbers, one would need only selected elements of the inverse of the normal matrix (and not all of them as in the case of the full covariance matrix).

At the moment, we are investigating the state-of-the-art in so-called sparse selected inversion that would enable us to calculate selected entries of the covariance matrix more efficiently. There is recent research on exactly this topic and a good starting point is given by (Verbosio & Schenk, 2019). We have narrowed down our alternatives to MUMPS (open source) and PARDISO (commercial solver). As a reference, naively calculating the full inverse (using the PyPardiso solver and reusing the factorization from step 2 leads to the timings in Table 2.

4.2 Connecting overlapping large-scale networks
It is not feasible to perform a single large-scale adjustment of all observations of the 5.5 million field sketches. This implies that results of multiple large-scale adjustments have to be combined as these adjustments will have points in overlap. We are currently researching how to deal with overlapping large-scale adjustments. There are two options for dealing with the discrepancies between overlapping adjustments:

- Adjusting the differences: this involves an adjustment in which the coordinate differences in the coinciding points are “translated” into updates of the other points. This adjustment involves (an approximation of) the full covariance matrices of both point fields, as well as updating these covariance matrices (Teunissen, 2006).

- Not adjusting the differences: in that case we will start the renewal of the cadastral map with possibly multiple sets of coordinates for the same point.

In both solutions it is important to avoid using the same measurement in more than one adjustment, because then that measurement will be used multiple times and its weight in the map renewal process is multiplied as well. Thus overlap between adjustments in terms of common points does not seem an issue, while overlap in observations does.

We prefer the second option because the extensive computations of the first option are avoided. This implies that the updating process of the cadastral map is to be done for each large-scale adjustment separately. Alternatively, the discrepancies between overlapping parts of the large-scale adjustments could be avoided by using only the centre (non-overlapping) part of the large-scale adjustments. However, the overlap between the large-scale adjustments needed in this approach will increase the computational burden.

5. RENEWAL OF THE CADASTRAL MAP

5.1 Connecting the cadastral map to the large-scale networks
The large-scale adjustments described in the previous paragraph result in a point field with a significantly higher quality than the current cadastral map. However, there are several reasons why this point field often does not correspond with the cadastral map:

- The point field results from the adjustment of historic measurements: many parcel boundaries have disappeared due to parcel merging. Furthermore, many buildings have been modified, extended, or demolished since the survey took place.
- Not all boundaries have been registered in the field sketches, only those that were altered after the initial registration in the first half of the 19th century.
- The point field resulting from the large-scale adjustment contains many auxiliary points on measurement lines that do not correspond to features of the cadastral map.
- Last but not least: the positional accuracy of especially the cadastral map is limited, which leads to mismatches between the two.

An example of the limited correspondence between the cadastral map and the result of the adjusted survey measurements is shown in Figure 10.

![Figure 10: Cadastral map (legal boundaries in black, buildings in red) and on the right in overlay (purple) with the result of the adjusted survey measurements of the field sketch of Figure 2. Note that the map is oriented to the north while the field sketch of Figure 2 is not.](image)

A major challenge in this step of our approach to cadastral map renewal is finding the correspondences between the cadastral map and the point field derived from the field sketches. Two types of correspondences have to be established:

1. Points of the field sketches can correspond to points of the cadastral map: point – point correspondence.
2. Points of the field sketches can be on parcel boundaries: point – line correspondence.

The method for matching the results of the large-scale adjustments and the cadastral map is under development. We are investigating several approaches.

5.2 The adjustments for renewal of the cadastral map and its quality description
The actual renewal of the cadastral map is an iterative adjustment process in which the cadastral map is adjusted to the point field resulting from each of the large-scale adjustments based on the geometric relations, as discussed in the previous paragraph. Figure 11 gives a schematic overview: subsets of the point fields of the large-scale adjustments (denoted \(Z_i\)) are related to the parcel corners and boundaries in the cadastral map (subset of point field \(W\)).

Different approaches for renewal of a map are presented in (Polman & Salzmann, 1996). The procedure we adopted for renewal of the cadastral map with large-scale adjustment point field \(Z_i\) can be summarised as follows:

1. Establishment of geometric relations between (subsets of) point fields \(Z_i\) and \(W\): see paragraph 5.1.
2. Adjustment of the geometric relations: this step affects only the points featuring in the geometric relations. This adjustment does not involve a transformation, as both point fields are in the national coordinate system.

3. Adjustment of the points of the cadastral map not affected by the geometric relations: the so-called “free” points. There is no need to adjust the free points of the large-scale adjustment point field: they can be regarded as auxiliary points.

![Diagram](image_url)

*Figure 11: Point fields and their relations for updating the cadastral map with the results of the large-scale adjustments.*

Of course, both (weighted) adjustments are greatly affected by the quality of the point fields involved. The quality description of the point fields in the form of covariance matrices is critical and needs further investigation:

- The quality of the present cadastral map is not well-known and currently investigated in a separate project (Hagemans et al., 2020). The results are planned to be used to establish a sparse artificial covariance matrix that will also represent the correlation between neighbouring points.

- Updating the cadastral map will require updating its quality description, and consequently its covariance matrix. With rigorous updating fill-in of the covariance matrix will take place and as a result it will need more storage space. It is likely that this covariance matrix is to be simplified in order to limit the storage space burden.

- The size of the cadastral map in number of points is too large to store a full covariance matrix, however, a partial or sparse representation implies an approximation. The trade-off between required storage space and quality has to be made. It is important to keep in mind that the cadastral map will have to be updated frequently in the future as well.

- The covariance matrices of the point fields resulting from the large-scale adjustments are full in principle. Currently we investigate the feasibility of the computation of this matrix and its usage in the map renewal step.
6. CONCLUSIONS

In this paper, results have been presented of the research program of the Dutch Cadastre for renewal of the cadastral map. We focussed on the geodetic aspects of the map renewal process that is based on the survey measurements of millions of historic field sketches. The used procedure for error detection and elimination is based on the ‘Delft School of Mathematical Geodesy’. We show the first results of this approach in a network that contains data of 65 field sketches. Measurement and vectorisation errors were eliminated, and coordinates of the nearly 5000 points computed.

Furthermore, performance of newly developed large-scale adjustment software is evaluated, and an approach for renewal of the cadastral map based on the results of the large-scale adjustments is presented. The quality description of the current and updated cadastral map plays an important role in this approach. It is concluded that rigorous nation-wide adjustment of all historic measurements and storage of a full covariance matrix of the coordinates is not feasible, and thus a trade-off between quality and computational burden is to be made.

REFERENCES

- Franken, J., Florijn, W., Hagemans, E., Hoekstra, M., 2020, "Rebuilding the cadastral map of The Netherlands, the artificial intelligence solution”, proceedings FIG-2020, Amsterdam.
- Hagemans, E., Busink, R., Grift, J., Schouten, F., 2020, "Rebuilding the cadastral map of The Netherlands, the overall concept”, proceedings FIG-2020, Amsterdam.
BIOGRAPHICAL NOTES

Frank van den Heuvel is working as a geodetic specialist at the Dutch Cadastre. He worked as an assistant professor at the Delft University of Technology and obtained a PhD-degree in 2003. He specialized in photogrammetry and worked for several companies before joining the Cadastre in 2018. He is working on quality assurance in large-scale photogrammetric projects and research programs, specifically on renewal of cadastral map of The Netherlands.

Gerbrand Vestjens is working as a geodetic specialist at the Dutch Cadastre. After obtaining his M.Sc. degree in Geodesy from the Delft University of Technology he worked at Ingenieursbureau Geodelta until 2016. He is experienced in drafting technical specifications for nationwide geodetic data collection. He is currently working on large-scale photogrammetric projects and the research program for renewal of the cadastral map of The Netherlands.

Gerard Verkuijl
Gerard Verkuijl is Geodetic Consultant at Sweco Nederland B.V. He has a Geodesy degree from the Delft University of Technology and is one of the developers of the MOVE3 adjustment software package. From this expertise he joined the research program for renewal of the cadastral map of The Netherlands.

Mark van den Broek
Mark van den Broek has been involved in the research program for renewal of the cadastral map of The Netherlands with key contributions in machine learning and optimization algorithm development. Mark has over 15 years of experience in analytics and algorithm development in various application domains. His current focus is on machine learning. Mark studied in parallel Mathematics at Eindhoven University and Econometrics at Tilburg University.

CONTACTS

Dr. Frank van den Heuvel, Ir. Gerbrand Vestjens
Dutch Cadastre, Land Registry and Mapping Agency
Postbus 9046, 7300 GH Apeldoorn, The Netherlands
Tel: +31 88 183 2200
Email: frank.vandenheuvel@kadaster.nl, Gerbrand.vestjens@kadaster.nl
Website: www.kadaster.nl