Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Afrasteh, Y., Slobbe, D. C., Verlaan, M., Sacher, M., Klees, R., Guarneri, H., Keyzer, L., Pietrzak, J., Snellen, M., & Zijl, F. (2021). The potential impact of hydrodynamic leveling on the quality of the European vertical reference frame. *Journal of Geodesy*, *95*(8), Article 90. https://doi.org/10.1007/s00190-021-01543-3

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

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Journal of Geodesy (2021) 95:90
https://doi.org/10.1007/s00190-021-01543-3

ORIGINAL ARTICLE

The potential impact of hydrodynamic leveling on the quality of the European vertical reference frame

Y. Afrasteh[1] · D. C. Slobbe[1] · M. Verlaan[2] · M. Sacher[3] · R. Klees[1] · H. Guarneri[1] · L. Keyzer[1] · J. Pietrzak[1] · M. Snellen[1] · F. Zijl[2]

Received: 29 October 2020 / Accepted: 9 July 2021
© The Author(s) 2021

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

Keywords Tide gauge · Quality · Hydrodynamic leveling · Network

1 Introduction

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

Our area of interest serves in this respect as an illustrative example. The domain of the hydrodynamic model we are developing (i.e., the 3D DCSM-FM (Zijl et al. 2020)) with the aim to forecast total water levels in the southern

✉ Y. Afrasteh
y.afrasteh@tudelft.nl

1 Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

2 Deltares, Boussinesqweg 1, 2629 HV Delft, The Netherlands

3 Federal Agency for Cartography and Geodesy, Richard-Strauss-Allee 11, 60598 Frankfurt am Main, Germany

Published online: 24 July 2021

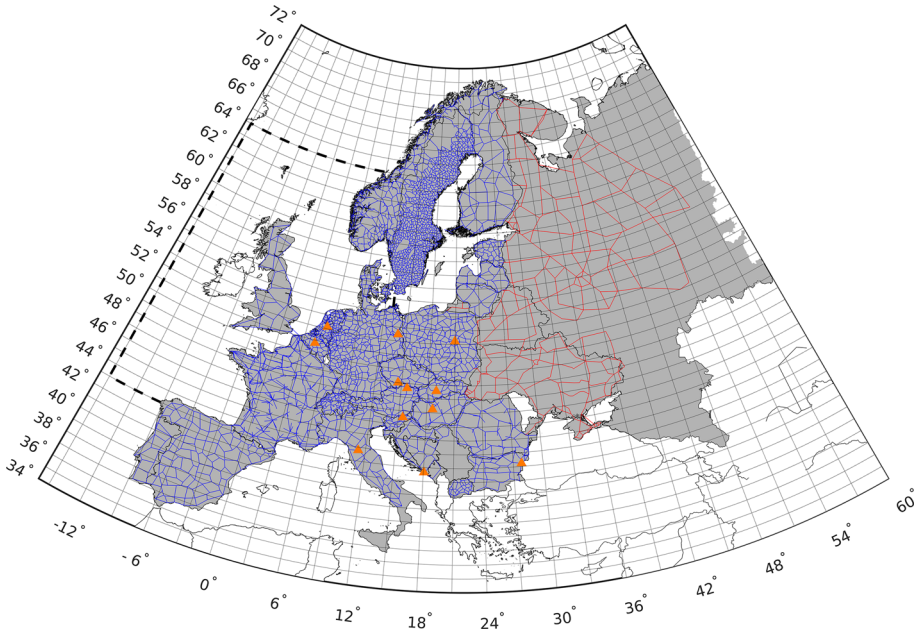


Fig. 1 Domain of the EVRF2019 (gray areas), UELN leveling connections (blue lines indicate those received from the BKG, while the red lines are the ones reconstructed from (Sacher and Liebsch 2019, Figure 1), see Sect. 3.1 for more details), and locations of the 12 datum points (green triangle). The black dashed line indicates the domain of 3D DCSM-FM (see Experiment V)

North Sea covers the waters between 15°W to 13°E and 43°N to 64°N (see Fig. 1). The first-mentioned requirement shows immediately that the designated unified vertical reference frame for Europe, i.e., the European Vertical Reference Frame 2019 (EVRF2019) (Sacher and Liebsch 2019), does not suffice. Indeed, Great Britain, Ireland, and other islands are not included. The reason for this is that the EVRF2019 is solely based on data of the Unified European Leveling Network (UELN) (Ihde et al. 2002), which are derived from spirit leveling and gravity data. It is well known that spirit leveling cannot be used to cross large water bodies. Because of the second-mentioned requirement, GNSS/Leveling (e.g., Schwarz et al. 1987; Catalão and Sevilla 2008) cannot currently be considered as an *operational* alternative to connect the islands and platforms. To meet the 1-cm accuracy level, a (quasi-)geoid model should be available with an accuracy of 8.7 mm (assuming the GNSS heights can be obtained with 5-mm accuracy). The most accurate quasi-geoid model available covering the entire area, the EGG2015 (Denker 2015), only has an accuracy of $\sim$ 7 cm (note that the accuracy varies significantly in different countries and areas). The same argument applies to the use of the geodetic boundary value problem (GBVP) approach (Rummel and Teunissen 1988; Heck and Rummel 1990; Amos and Featherstone 2008; Gerlach and Rummel 2012; Amjadiparvar et al. 2015; Sánchez and Sideris 2017). Finally, despite having great potential, chronometric leveling (Müller et al. 2017; Mehlstäubler et al. 2018) is not yet operational.

The only remaining method for height datum connection known in literature is hydrodynamic/ocean leveling (e.g., Proudman 1953; Cartwright and Crease 1963; Woodworth et al. 2013). It is applied between tide gauges and requires knowledge of the differences in mean dynamic topography (MDT) between them. There are different approaches to derive MDT differences; we refer to Slobbe et al. (2018) for a concise review. Featherstone and Filmer (2012) successfully used the method to explain the north–south tilt in the Australian height datum. They exploited both 'geodetic' and 'oceanographic' MDT models, as well as models derived from a combination of geodetic and oceanographic data. The oceanographic MDT model, obtained by integration of temperature, pressure, and salinity fields on a water depth of 2 km, showed the best performance in explaining the tilt. A similar conclusion was obtained by Filmer and Featherstone (2012) who used five different models, as well as GNSS and two gravimetric quasi-geoid models, at tide gauges/tide gauge benchmarks to re-estimate the offset in the Australian height datum between mainland Australia and Tasmania. In a later, more extended study, Filmer et al. (2018) compared 13 physics-based numerical ocean models and 6 MDTs computed from observed geodetic and/or oceanographic data at 32 tide gauges around the Australian coast to assess the suitability of different types of MDT for height datum unification. One of the main conclusions of the study is that numerical ocean models appear a viable alternative for height datum unification. Slobbe et al. (2018) proposed what they referred to as 'model-based hydrodynamic leveling'. Their method exploits a regional, high-resolution hydrodynamic model to derive mean water level (MWL) differences between tide gauges. The use of the term 'MWL' refers to the fact that the averaging period can be chosen freely; Slobbe et al. (2018) obtained the best results when averaging the water levels over the summer months of the 19-year simulation period. They applied the technique to

transfer Amsterdam ordnance datum (Normaal Amsterdams Peil, NAP) from the Dutch mainland to the Dutch Wadden islands. Based on a high-resolution 2D hydrodynamic model, extended to account for depth-averaged water density variations, Slobbe et al. (2018) showed that for each Wadden island several connections are available that allow to transfer NAP with (sub-)centimeter accuracy.

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

Indeed, to connect all tide gauges within the domain of our hydrodynamic model we could follow the approach by Slobbe et al. (2018). That is, we connect all tide gauges to the NAP. This approach, however, requires a model that has a good performance at *all* tide gauge locations even though some are at the same mainland and can be connected by spirit leveling. Apart from that, all errors in the vertical reference of the involved Dutch tide gauge(s) propagate one to one to the vertical reference of the tide gauges of interest. Because the former involves great efforts to achieve it and the latter is not desirable, we propose combining 'hydrodynamic leveling data' with the UELN data and use the combined dataset to compute a new realization of the EVRS that covers our whole domain of interest. This proposal is, indeed, a bit similar to what is advocated by Filmer et al. (2014). To maximize the network strength, we advocate to establish hydrodynamic leveling connections in all European waters. This, of course, requires a model covering all European waters or a set of models that each cover a separate basin. With hydrodynamic leveling connections, we mean connections between tide gauge benchmarks that can be established by using the observation-derived MWLs relative with respect to the tide gauge benchmarks and the model-derived MWL differences between tide gauges. Note that here the model-derived MWL differences are obtained from models that do *not* assimilate geodetic information. The pursued strategy also benefits other users of the EVRF as we may expect that combining both datasets improves the quality of the leveling network and hence the derived VRF. In particular, adding hydrodynamic leveling data helps to detect/suppress systematic errors that spirit leveling is susceptible to (e.g., Penna et al. 2013).

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

establish connections between tide gauges within the same sea basin.

The paper is organized as follows: Section 2 describes the height network adjustment, the way the impact of adding hydrodynamic leveling data is assessed, and the method used to determine which hydrodynamic leveling connections are actually added. Section 3 introduces the datasets used throughout this paper. Section 4 introduces the setup of the experiments conducted in this study. The results of the experiments are presented and discussed in Sect. 5. Finally, we conclude by emphasizing the main findings and identifying topics for future research.

2 Methodology

2.1 Height network adjustment

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

$$y = Ax + e, \tag{1}$$

where

$$y = \begin{pmatrix} y_{\mathrm{sl}} \\ y_{\mathrm{hl}} \end{pmatrix}, \; A = \begin{pmatrix} A_{\mathrm{sl}} \\ A_{\mathrm{hl}} \end{pmatrix}, \text{ and } e = \begin{pmatrix} e_{\mathrm{sl}} \\ e_{\mathrm{hl}} \end{pmatrix}. \tag{2}$$

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

$$E\{e\} = 0, \; D\{e\} = Q_y = \begin{pmatrix} Q_{\mathrm{sl}} & \mathbf{0} \\ \mathbf{0} & Q_{\mathrm{hl}} \end{pmatrix}, \tag{3}$$

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

$$Q_{\mathrm{hl}} = A_{\mathrm{hl}} Q_{\mathrm{dMWL}} A_{\mathrm{hl}}^{\mathrm{T}}, \tag{4}$$

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

In realizing the EVRF2019, the datum defect is solved by adding the minimal constraint that for 12 datum points (see Fig. 1) the sum of the height changes is zero. The drawback of using this constraint is that the propagated standard deviations of the adjusted heights depend on the height marker (also referred to as "height benchmark" or "leveling benchmark") distance to the datum points (Sacher and Liebsch 2019). To assess the impact of adding hydrodynamic leveling data on the quality of the leveling network, we conducted an experiment in which we used the constraint that the sum of height changes of all height markers is zero (Teunissen 2006). This form of minimal constraint adjustment, known as inner constraint adjustment, provides similar results as using the pseudo-inverse in the least-squares adjustment (Ogundare 2018). Indeed, to realize the EVRS the use of the inner constraint adjustment is not a proper alternative to solve the datum defect as also benchmarks in geodynamically unstable regions will affect the datum.

2.2 Assessing the impact of adding hydrodynamic leveling data

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

parameters $Q_{\hat{x}}$, with

$$Q_{\hat{x}} = \left(A^{\mathrm{T}} Q_y^{-1} A\right)^{-1}. \tag{5}$$

In reporting the precision, we focus on the propagated standard deviations (SDs) of the adjusted heights (i.e., the square root of the diagonal elements of $Q_{\hat{x}}$); the median of this value for all height markers of the network, as well as the median value per country. The median value for all height markers of the network is also used to determine which hydrodynamic leveling connections will be added to the network (see Sect. 2.3). Note that we use the median because the propagated height SDs for all height markers are not normally distributed.

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

$$R = I - A\left(A^{\mathrm{T}} Q_y^{-1} A\right)^{-1} A^{\mathrm{T}} Q_y^{-1}, \tag{6}$$

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

2.3 Choice of the hydrodynamic leveling connections

Given N tide gauges, maximum $N-1$ *independent* hydrodynamic leveling connections can be established. By independent, we mean the connections that do not form any closed circuit. Indeed, the model-derived MWL differences between the tide gauges are obtained from the MWLs at the tide gauges. As such, adding a connection that closes a circuit does not add any new information and results the full variance-covariance matrix obtained using Eq. 4 to become singular. The number of possible connections can be extremely large. Considering N tide gauges, the number of possibilities to establish $N-1$ independent connections among them equals $K = N^{N-2}$ (Cayley's formula (Aigner and Ziegler 1998)). Europe has a relatively dense network of tide gauges that contains hundreds of stations. Assuming 200 out of them can be used to establish hydrodynamic leveling connections, K equals 200^{198}. To evaluate which set of connections has the largest impact, i.e. results in the lowest median SD of the adjusted heights, K least-squares solutions have to be computed. Despite the fact that the computational load can be reduced significantly by exploiting the recursive least-squares method (Teunissen 2006) and only computing the diagonal elements of the variance-covariance matrix of the estimated unknowns, still evaluating all K solutions is not feasible.

Therefore, we use a heuristic search method that identifies the connections one by one. In each step, we first identify all remaining possible connections (closed circuits are not allowed) based on a depth-first search algorithm (Tarjan 1972). Second, we identify which of these connections results in the lowest median SD of the adjusted heights. The identified connection is added to the list of found connections and removed from the list of remaining ones. The search process continues until no more connections are possible. The use of this heuristic search method indeed reduces the computational load significantly. To identify the first connection, $\binom{N}{2}$ least-squares solutions have to be computed. With every connection we add, this number decreases with the number of connections added in the previous step (always one) and the ones that form a closed circuit.

To further reduce the computational load, we (i) reduced the number of potential tide gauges (see Sect. 3.2) and (ii) did not allow connections among tide gauges (a) located within the same country and (b) located in neighboring countries for which the number of spirit leveling connections between the countries is larger than one.

3 Data

3.1 Spirit leveling network and data

From the Federal Agency for Cartography and Geodesy (BKG), we received (i) the locations of all UELN height markers, (ii) a list of leveling connections (only contains an overview of which height markers are connected; we did not receive the actual geopotential differences) in all countries except for Ukraine, Russia, and Belarus, (iii) the a priori variances of the geopotential differences for the available connections, and (iv) the variances obtained by variance component estimation (except for Great Britain, Ukraine, Russia, and Belarus). The reason why we did not receive the information for all countries is that either the BKG is not allowed to share the data (applies to the data of Ukraine, Russia, and Belarus), or the data have not been used in computing

the EVRF2019 and as such are not available at all (applies to the variances obtained by variance component estimation for the data of Great Britain). Missing the data in Ukraine, Russia and Belarus makes the connection of central Europe to the Fennoscandia region to be based on just two leveling observations. This would artificially increase the impact of adding hydrodynamic leveling data. Therefore, we decided to reconstruct the missing leveling connections using Figure 1 in Sacher and Liebsch (2019). Figure 1 shows both the part of the spirit leveling network obtained from the BKG and the reconstructed part. The data variances for the reconstructed connections are determined using the computed distances between the height markers and the reported standard deviations of unit weight per country (Sacher and Liebsch 2019, Table 3). In all experiments conducted in this study, we used the variances that the BKG obtained by variance component estimation. For Great Britain, the a priori variances were used.

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

3.2 Candidate tide gauges and link to spirit leveling network

Candidate tide gauges, i.e., tide gauges among which hydrodynamic leveling connections can be established, have to be located at the coast of one of the seas surrounding the UELN countries. Moreover, we only use tide gauges south of the TOPEX/Poseidon and Jason maximum latitude of 66°N. The waters in higher latitudes have lower densities of high-quality satellite and in situ data for validation of hydrodynamic models. On top of that, these regions typically have a poor bathymetry (Stammer et al. 2014). Both will negatively affect the ability of hydrodynamic models to represent the MWL.

Tide gauges are selected from the ones included in the PSMSL database (Holgate et al. 2013; PSMSL 2020). In the area of interest, this database includes about 330 tide gauges. Indeed, more tide gauges are available (see, e.g., http://www.emodnet-physics.eu/Map/.). The PSMSL database is used, however, because the data are quality-controlled and provided with extensive metadata. The metadata include, among others, descriptions of the tide gauge benchmarks and their locations. The latter information is indispensable when implementing hydrodynamic leveling.

To reduce the computational load (see Sect. 2.3), we only consider those tide gauges that are located within 10 km from the nearest UELN height marker. This results in a total number of 186 tide gauges. Figure 2 shows the locations of the tide gauges. The number of tide gauges per country is presented in Table 1. To connect the tide gauges to the UELN, we added an artificial leveling connection between each tide gauge and the nearest UELN height marker. The variances for these added artificial leveling connections are determined assuming the leveling is conducted with a precision of 0.5 mm/$\sqrt{\text{km}}$ corresponding to the precision of first-order leveling (Federal Geodetic Control Committee et al. 1984).

4 Experimental setup

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

Experiment 0: using spirit leveling data only—The impact of adding hydrodynamic leveling is assessed by comparing the obtained realizations to the one obtained with spirit leveling data only. Since our spirit leveling dataset is not identical to the one used to obtain the EVRF2019 for reasons explained in Sect. 3.1, in Experiment 0, we assess the performance of what we refer to as the *spirit leveling-only solution*.

Experiment I: allowing for connections within basins only—Applying model-based hydrodynamic leveling between tide gauges requires the availability of a hydrodynamic model being capable to resolve the local MWL. As pointed out in Sect. 1, in terms of model development a more plausible scenario for a European-wide implementation of hydrodynamic leveling is to start with a set of models each covering a separate sea basin. Assuming to have access to such a set of models that allow to derive the MWL differences with uniform precision, the objective of this experiment is to assess the impact of adding hydrodynamic leveling connections between tide gauges located in the same sea basin to the UELN. Four basins are considered: the Mediterranean Sea, Black Sea, Baltic Sea, and the North-East Atlantic region including the North Sea. Figure 2 shows per basin the locations of all 186 PSMSL tide gauges that meet the criteria outlined in Sect. 3.2. We assume a uniform noise level of 3 cm for each connection. This means a variance of 4.5 cm^2 for the precision at which the model is able to reconstruct the MWL at each tide gauge location. Again, so far we lack a proper stochastic model for the hydrodynamic leveling dataset. The 3-cm accuracy level is a bit lower than the accuracy obtained by Slobbe et al. (2018) for the connection of the Dutch Wadden island tide gauges to the NAP. Woodworth et al. (2013) stated that ocean leveling is possible with a typical uncertainty of better than a decimeter. They also point out, however, that this statement "is subject to reservations concerning the limitations in the ocean models available for analysis, and to the fact that a global study remains to be made". Given the fact that we pursue an implementation based on models that resolve all relevant 3D physical processes plus the fact that we use MWL differences

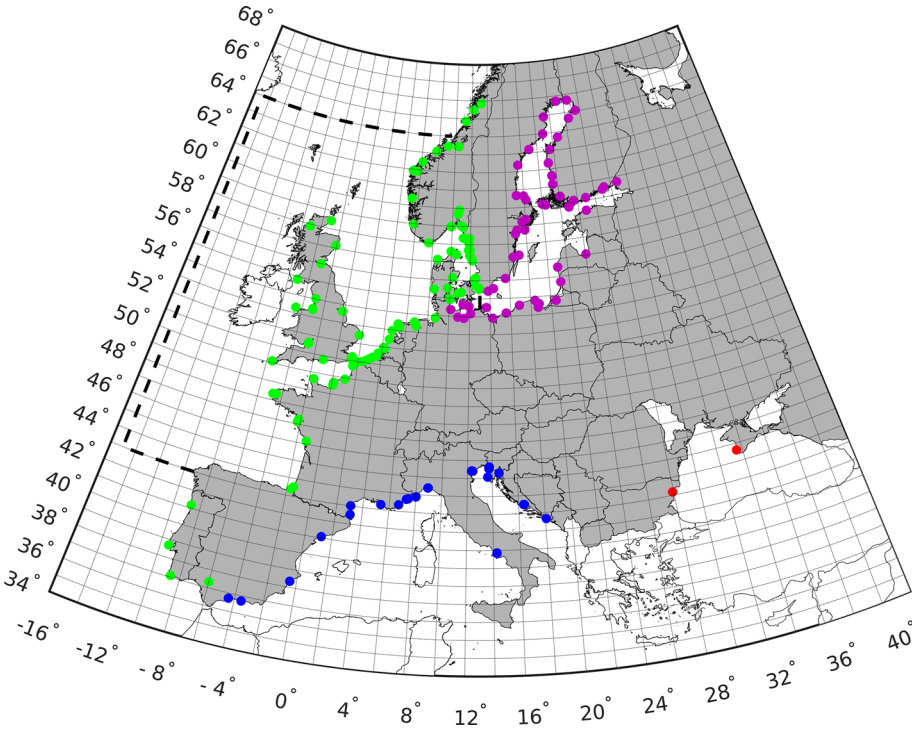


Fig. 2 Location of all tide gauges used in this study. The different colors refer to the four sea basins used in Experiments I–III: the Mediterranean Sea (blue), Black Sea (red), Baltic Sea (magenta), and the North-East Atlantic region including the North Sea (green). The black dashed line indicates the domain of 3D DCSM-FM (see Experiment V)

rather than MDT differences (i.e., we ignore the storm surge period in computing the MWL, see Sect. 1), we believe 3 cm is challenging but not unrealistic.

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

Experiment III: using the inner constraint adjustment—So far, we assessed the impact of hydrodynamic leveling connections on the quality of the EVRF. As explained in Sect. 2.1, the propagated SDs depend on the distance of the height markers to the locations of the datum points. As discussed, we can avoid this dependency by considering the so-called inner constraint adjustment. This is what we assess in this experiment. So, Experiment III differs from Experiment I in the use of the constraint added to solve the datum defect. Note that in this experiment, the improvement is quantified with respect to a spirit leveling-only solution obtained by applying the same constraint.

Experiment IV: adding hydrodynamic leveling connections among all European seas—This experiment aims to answer the question how the quality impact changes when allowing connections among all European seas. Indeed, this requires a model covering all European seas. Note that in line with others (e.g., Lea et al. 2015) we treat the Black Sea as a closed basin. This means that any connection between a Black Sea tide gauge to one located at the coast of another sea is not allowed.

Experiment V: using the tide gauges within the 3D DCSM-FM domain only—In the project of which this study is part, we aim to develop a hydrodynamic model known as the 3D DCSM-FM (Zijl et al. 2020). One objective is to use this model to conduct hydrodynamic leveling. In Experiment V, we assess the quality impact in case we only have this model available. That is, in case we can only exploit the tide gauges available within the 3D DCSM-FM domain (see Fig. 2). Note that we only used the tide gauges that are located at a distance of at least one degree from the boundaries.

5 Results and discussion

In this section, we present and discuss the results of the experiments introduced in Sect. 4. In doing so, we first assess for all experiments the impact on the quality of the EVRF2019 (first research objective of this study). Thereafter, we assess which connections, and hence tide gauges, are most profitable in terms of impact. These are the ones we need to focus on in the development of the hydrodynamic model(s) (second research objective of this study).

5.1 The impact of adding hydrodynamic leveling data on the quality of the EVRF

5.1.1 Experiment 0: using spirit leveling data only

Figure 3 shows a map of the SDs of the adjusted heights for the spirit leveling-only solution (i.e., the solution which serves as a reference in experiments I, II, IV, and V). They cover a broad range of values, between $\sim$ 5 and $\sim$ 75 mm. The median value is 13.8 mm. Table 1 shows the median SD per country. Both Table 1 and Fig. 3 clearly show the large regional deviations; the values are lowest at the center (including Belgium, Germany, the Netherlands, Austria, Czech Republic, and Poland) and increase toward the margins. The mentioned countries all have high-quality leveling data (see Table 3 of Sacher and Liebsch (2019)). Moreover, most of the datum points are located in these countries (see Fig. 1). As mentioned before, the propagated SDs are not only affected by the precision of the observations, but also by the height marker distance to the datum points (Sacher and Liebsch 2019). The absence of a datum point in the Scandinavian Peninsula explains why in Finland and Sweden (despite having high-quality leveling data) the adjusted height SDs are high compared to those in the center. Figure 4 shows a map of the redundancy numbers, and Table 1 presents the median redundancy number per country and for the entire network. Overall, the redundancy numbers are small with values ranging from 0.002 (France) to 0.411 (Norway). The low value for France can be explained by the fact that the French leveling network contains many height markers that are only connected to one other height marker (see Fig. 1). The better the network is connected, the higher the redundancy numbers would be. By adding hydrodynamic leveling observations, we expect to increase the redundancy numbers in the coastal areas.

5.1.2 Experiment I: allowing for connections within basins only

Given the total number of 186 candidate tide gauges (Sect. 3.2) spread over 4 separate sea basins, the number of independent connections we can add is 182 (note that it is only allowed to establish connections within the same sea basins).

Adding all 182 connections reduces the median-propagated SD of all adjusted heights from 13.8 to 8.6 mm. This corresponds to an improvement of 38%. The improvement differs strongly per country as shown in Table 1; values range from 1% (Slovakia) to 60% (Great Britain). We observe larger improvements for coastal countries, except for the countries which are located along the Black Sea (see Fig. 5). We also notice a more significant improvement for coastal countries at the perimeter of the UELN network (i.e., Portugal, Spain, Great Britain, and the Scandinavian countries). The lower improvements for the Black Sea countries are explained by the fact that hydrodynamic leveling connections that link the Black Sea to the other European seas were not allowed. The reason why the impact is largest in Great Britain can be understood when we consider that the existing connection of the Great Britain leveling network to the remaining part of the UELN is extremely weak; it is connected by just two leveling campaigns through the channel tunnel. Since there are many tide gauges in Great Britain, hydrodynamic leveling allows to tie Great Britain much stronger to the rest of the UELN.

In terms of reliability, we observe increased redundancy numbers for the spirit leveling observations near the coastline (see Fig. 6). For most of these observations, the improvement is between 0.02 and 0.1. For about 1% of the observations the improvement is larger, the maximum being 0.9. In Great Britain, the numbers increase almost throughout the whole country. Here, they range between 0 and 0.5. The median redundancy number for Great Britain improves from 0.278 to 0.462 (note that in computing this value the redundancy numbers associated to the hydrodynamic leveling observations are excluded). For the other countries, we hardly observed any change in the median redundancy number. (For that reason, they are not included in Table 1.) This is reflected by the minor change in the median redundancy number for the whole network (0.161 versus 0.154 for Experiment 0).

5.1.3 Experiment II: varying the noise level

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

Figure 7 shows for all considered noise levels the improvement of the overall median-propagated height SD (i.e., computed over all height markers) as a function of the number of added hydrodynamic leveling connections. We notice that the decrease of the level of improvement achieved by adding all connections is not linearly related to the increase of the noise level. Indeed, the distance between the curves for the number of added connections being equal to 182 gets smaller and smaller when the noise level goes up. We also

Table 1 Number of tide gauges per country and in total, and the median-propagated standard deviation (SD) of the adjusted heights in millimeters (per country and in total)

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

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

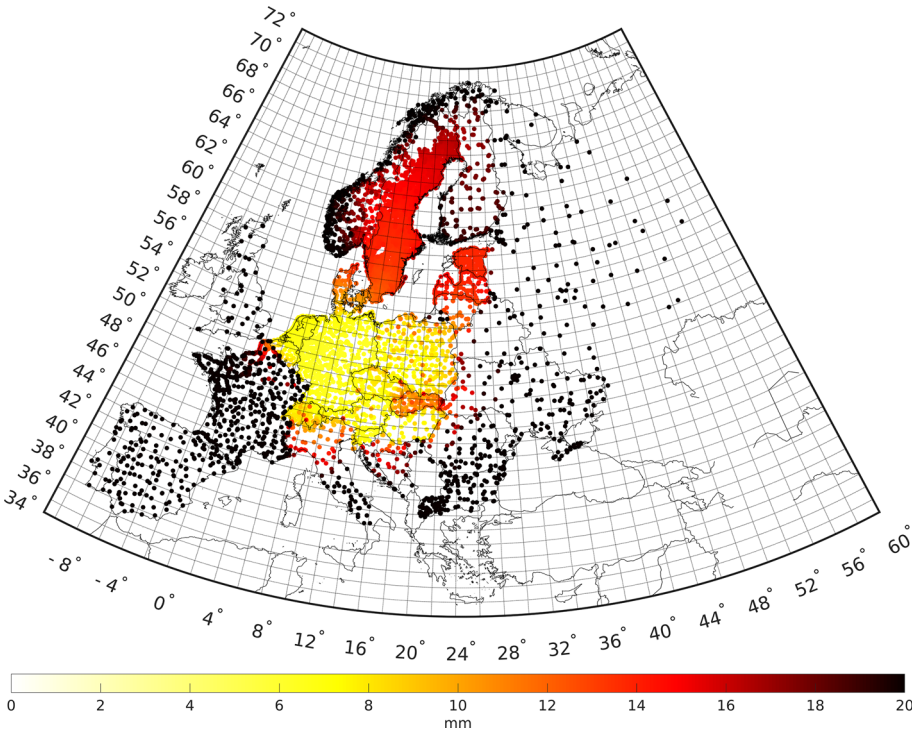


Fig. 3 Propagated SDs of the adjusted heights in Experiment 0

observe that in order to achieve a certain level of improvement, more connections need to be added with increasing noise level. For example, a 25% improvement can be obtained with just 2 connections in case we have 1 cm accurate data, while 38 connections are needed when the accuracy level is 5 cm. Note, again, that the final level of improvement also depends on the available number of tide gauges as well as their location. In our experiments, both parameters are fixed. Still, we do not exploit all tide gauges available in Europe (see Sect. 3.2).

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

5.1.4 Experiment III: using the inner constraint adjustment

Experiment III uses the same number of hydrodynamic leveling connections as Experiment I does. The median-propagated SD of all adjusted heights is 7.0 mm, corresponding to an improvement of 28% (see Table 2). Note that the improvement is quantified compared to a spirit leveling-only solution obtained by applying the inner constraint adjustment. For this solution, the median-propagated SD was 9.8 mm. Compared to Experiment I, the improvement in the Netherlands, Germany, Belgium is more than 15% larger (cf. Figs. 8 and 5). For most other countries, we also observe improvements, though the magnitudes are smaller.

5.1.5 Experiment IV: adding hydrodynamic leveling connections among all European seas

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

The median-propagated SD of all adjusted heights is 8.0 mm, corresponding to an improvement of 42% compared to Experiment 0. This is, indeed, very close to the values obtained in Experiment I (8.6 mm and 38%, respectively). Also the improvements per country are quite similar (see Table 1); the largest increment in improvement is observed for the Netherlands where we go from 9% improvement to 16%. Keep in mind, however, that in absolute numbers we go from 6.2 to 5.7 mm. So, the change is just at the sub-mm level.

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

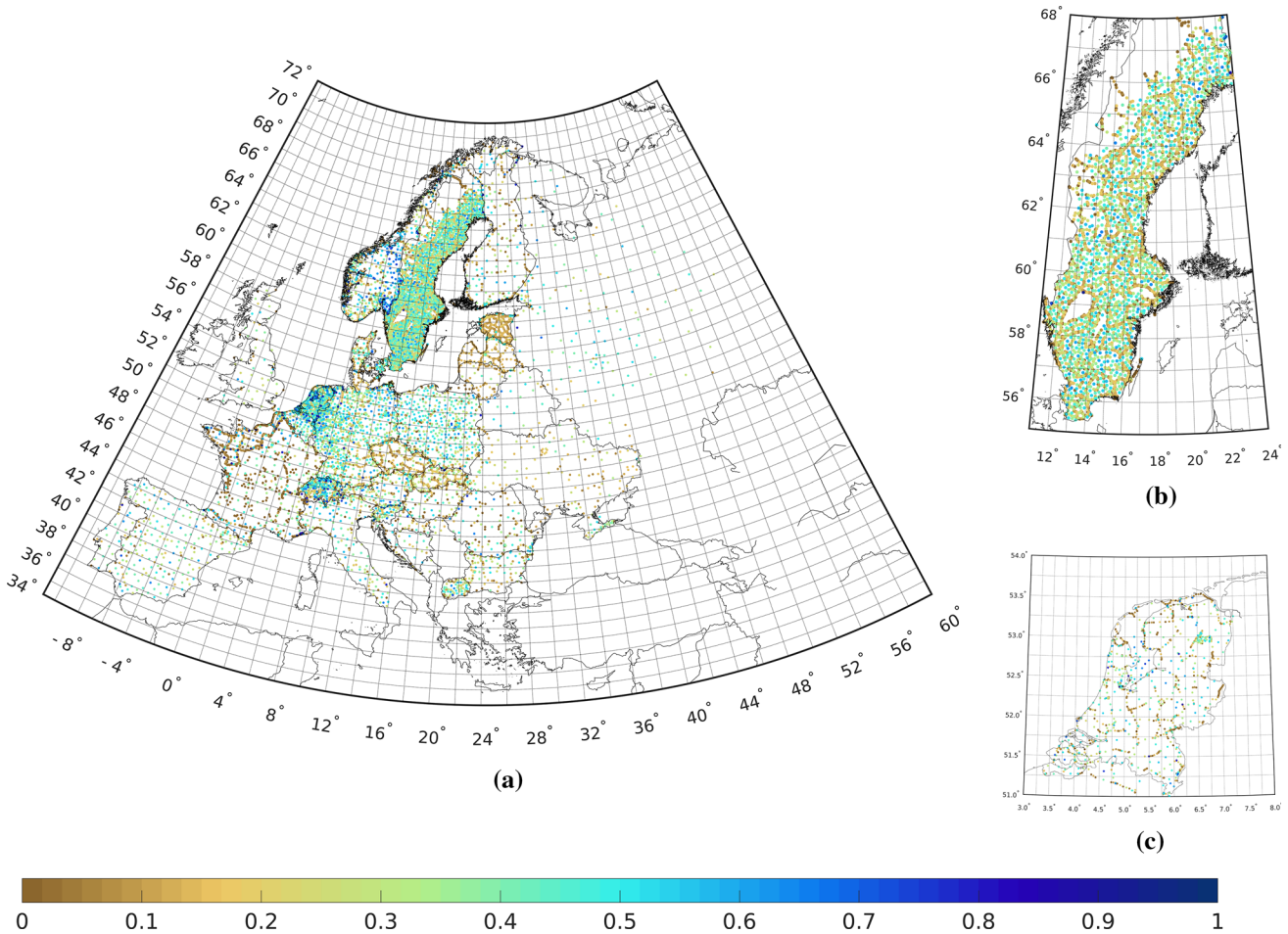


Fig. 4 Redundancy numbers of the leveling observations in the spirit leveling-only solution (Experiment 0) (**a**). A zoom-in for Sweden (**b**) and the Netherlands (**c**)

5.1.6 Experiment V: using the tide gauges within the 3D DCSM-FM domain only

The total number of available PSMSL tide gauges within the 3D DCSM-FM domain is 77. The resulting number of added connections corresponds to the theoretical maximum of 76. The median value of the propagated SD after adding all 76 hydrodynamic leveling connections is 10.3 mm corresponding to an improvement of 25%. The median-propagated SDs per country are quite different compared with the values obtained in the other experiments (see Table 1). For Great Britain, the improvement is still large (58%). For the Scandinavian countries, the improvements are also substantial (27%, 26%, 16%, and 10% for Sweden, Norway, Denmark, and Finland, respectively) but lower than in Experiments I and IV. For all other countries, except France (24%), the improvements are equal to or less than 4%. Similar to what we observed in Experiment I, the redundancy numbers for the leveling connections near the coastline (here limited to the 3D DCSM-FM domain) are increased (plot not shown). The median value, however, did not change significantly.

5.2 Identifying the high-impact tide gauges

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

To start the analysis, we again point to Fig. 7 that shows for different noise levels the overall improvement as a function of the number of added hydrodynamic leveling connections. The third curve is associated with Experiment I. It shows a clear asymptotic behavior with a rapid increase of improvement in the beginning. Indeed, by just adding 20 connections the obtained improvement is already 30%. (The maximum

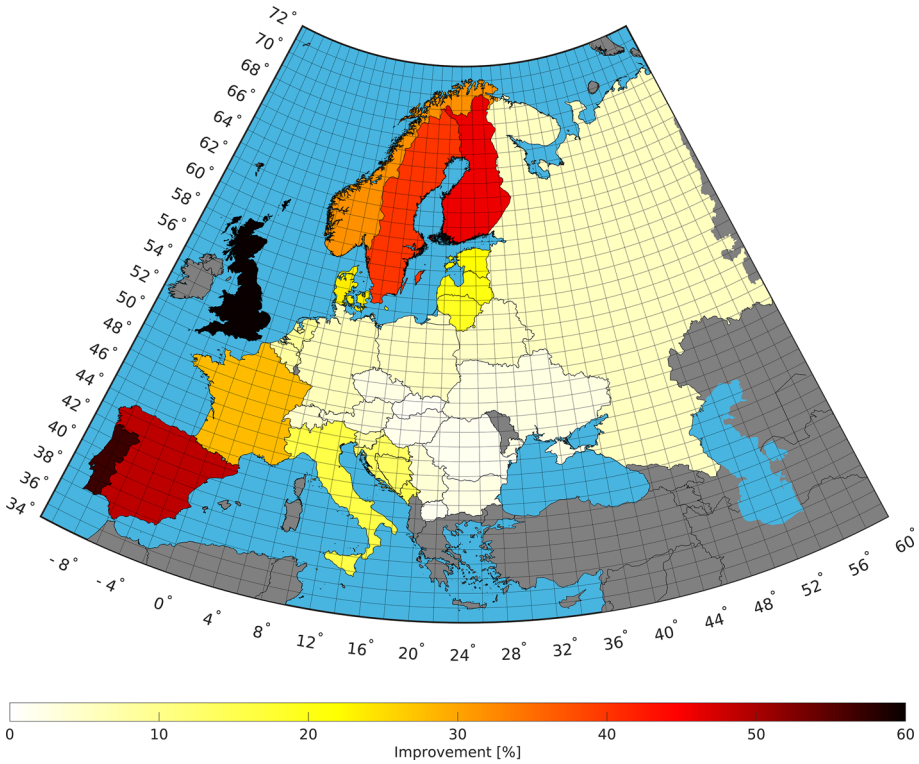


Fig. 5 Improvement of the median-propagated SD of the adjusted heights per country obtained in Experiment I

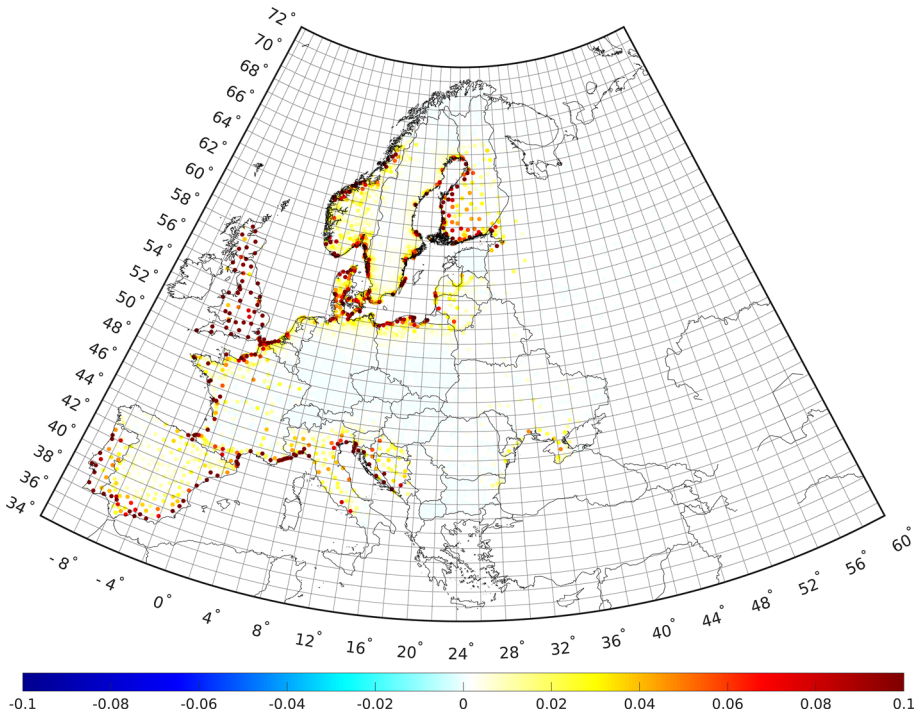


Fig. 6 Difference between the redundancy numbers obtained in Experiments I and 0

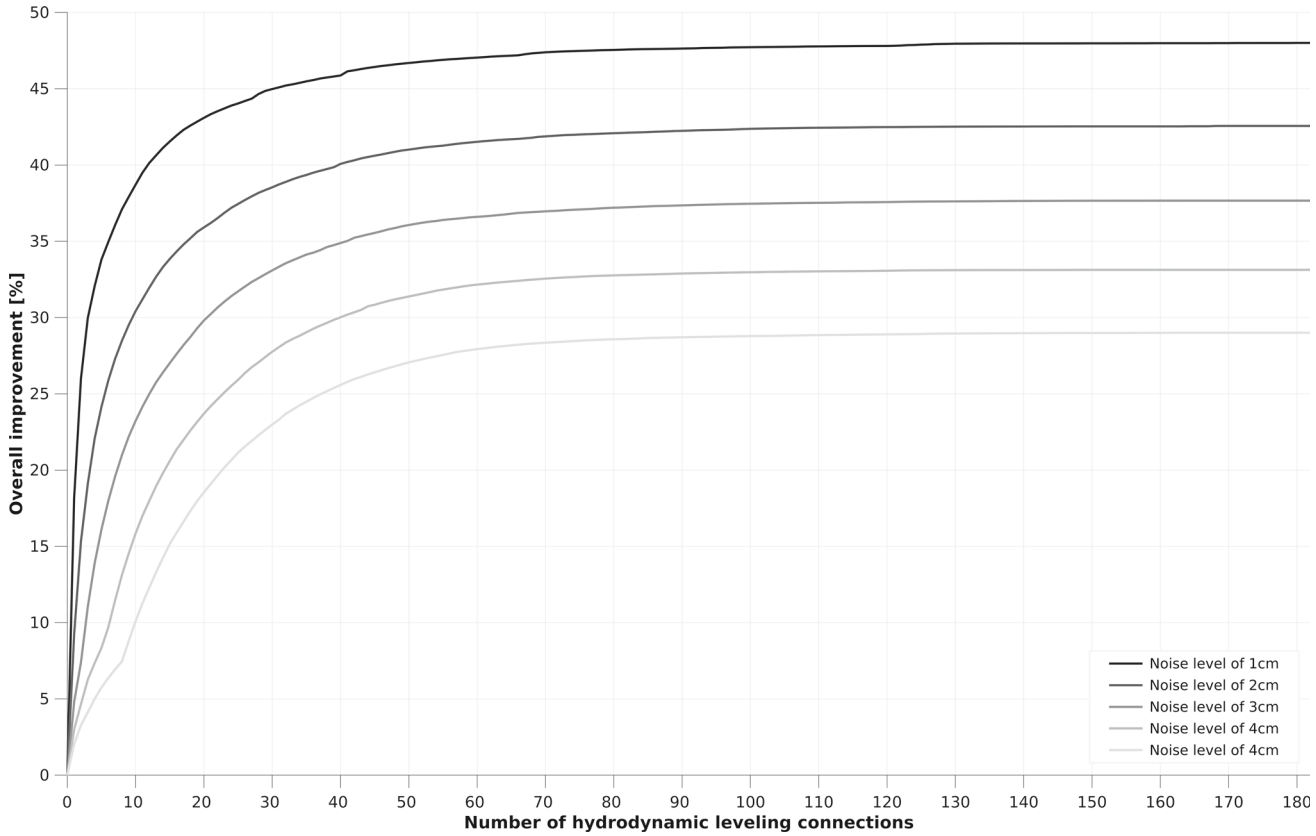


Fig. 7 Overall improvement of the solutions obtained by adding hydrodynamic leveling data as a function of the number of added hydrodynamic leveling connections for different noise levels

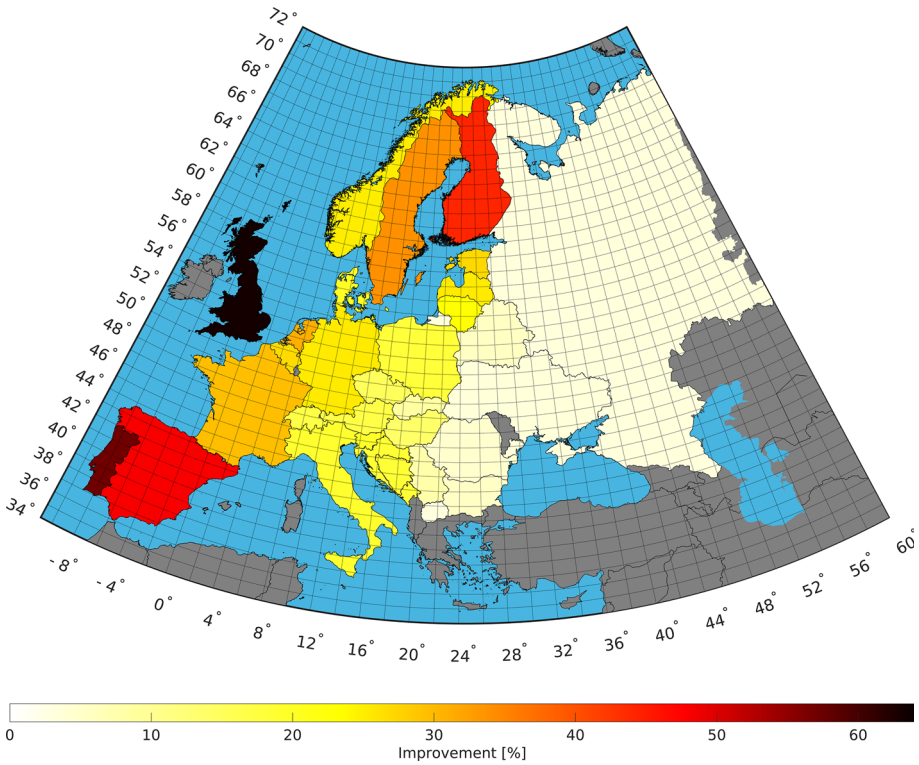


Fig. 8 Improvement of the median-propagated SD of the adjusted heights per country obtained in Experiment III

Table 2 Standard deviation (SD) of spirit leveling-only solution (in millimeter) where the inner constraint is used for the network adjustment and the standard deviation (SD) and percentage of improvement (values in bracket) per country for Experiment III

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

improvement in Experiment I is 38%.) Also, in the other experiments (plots not shown in the paper), we observe this asymptotic behavior. To compare the results of the various experiments, we plotted in Fig. 9 the percentage of unique tide gauges being involved in establishing the hydrodynamic leveling connections as a function of the percentage of achieved improvement relative to the maximum improvement. From this plot, we observe that for all scenarios in which we used the 3 cm noise level (Experiments I, IV, and V) 75% of the maximum improvement can be achieved by using only 16–23% of the available tide gauges (which number is 186, 186, and 77 for Experiments I, IV, and V, respectively). Only when we increase/decrease the noise level, more/less tide gauges are needed to achieve a certain level of improvement (see the curves associated to Experiment II in Fig. 9). Still, even for a 5-cm noise level only 27% of the tide gauges are needed to achieve 75% of the maximum improvement. This is a positive result as it shows that when adding a limited number of hydrodynamic leveling connections between a small number of tide gauges a substantial improvement in the quality of the EVRF2019 can be achieved. Moreover, the analysis suggests that adding more tide gauges than the ones considered in this study will not significantly increase the overall level of improvement.

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

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

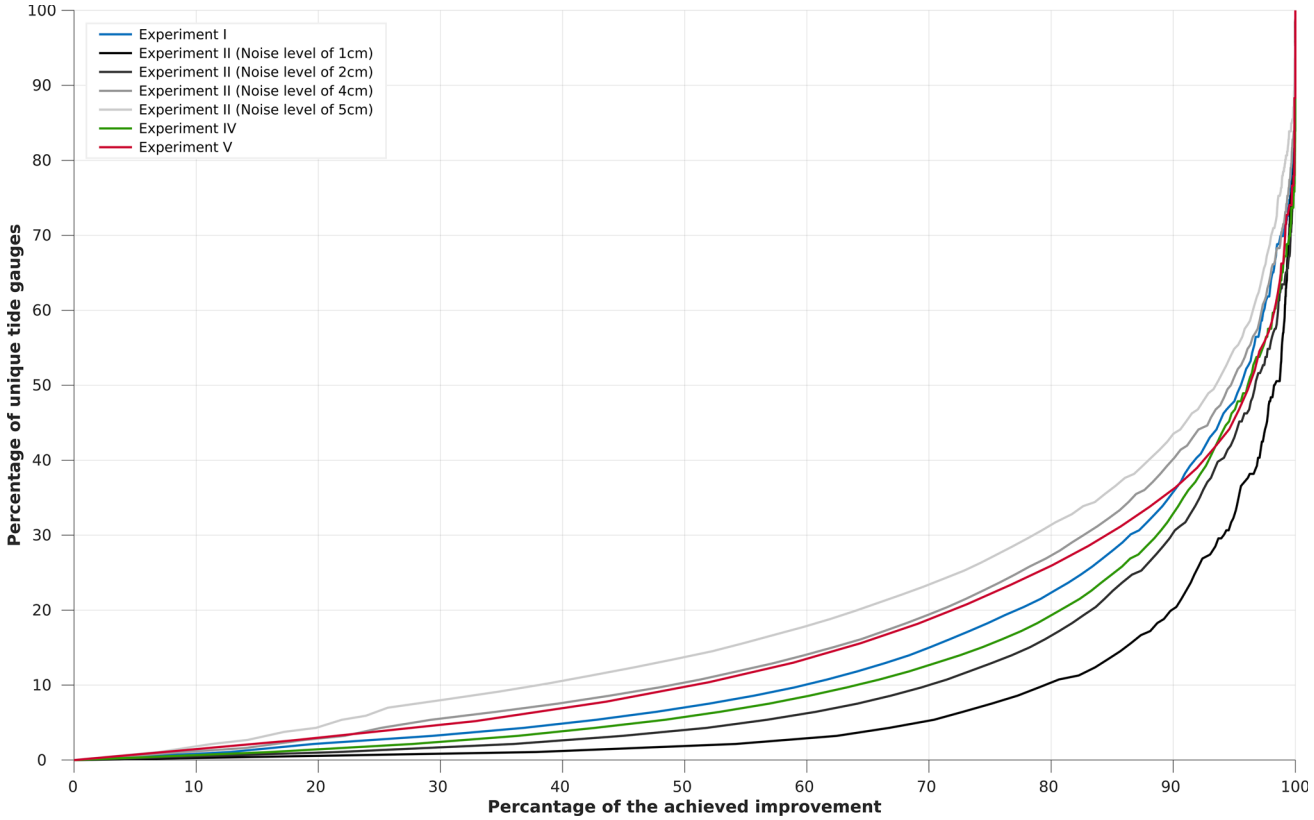


Fig. 9 The percentage of unique tide gauges being involved in establishing the hydrodynamic leveling connections as a function of the percentage of achieved improvement relative to the total improvement for Experiments I, II, IV, and V

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

6 Summary and conclusions

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

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

In total, five experiments were conducted. Experiment I assumes we have access to a set of hydrodynamic models,

each covering a separate sea basin. Moreover, it assumes that each model provides the MWL differences with an SD of 3 cm. So, in the first experiment, we allowed for hydrodynamic leveling connections only among tide gauges located in the same basin. Four basins were considered, the Mediterranean Sea, Black Sea, Baltic Sea, and the North-East Atlantic region including the North Sea. Among 186 tide gauges, our search algorithm identified 182 independent leveling connections that allowed to reduce the median SD of the adjusted heights from 13.8 mm for the spirit leveling-only solution to 8.6 mm, equivalent to a 38% improvement. Except for the countries around the Black Sea, coastal countries benefit the most with a maximum improvement of 60% for Great Britain. In terms of reliability, the impact was assessed by the redundancy numbers. We found increased redundancy numbers for observations close to the coast and over the entire Great Britain.

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

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

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

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

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

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

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

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

A point of attention in the operationalization of the technique is the reduction of the observation- and model-derived tide gauge records to the reference epoch adopted in the EVRF (epoch 2000.0). This reduction is needed to correct for vertical land motion and/or long-term sea-level variations. Applying such a reduction is, however, only feasible

if the tide gauge records are sufficiently long and all needed metadata are available. An example of needed metadata is the epoch when the tide gauge benchmark is connected to the height system by means of leveling.

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

Acknowledgements This study was performed in the framework of the Versatile Hydrodynamics project, funded by the Netherlands Organization for Research (NWO). This support is gratefully acknowledged. We thank the editor and three anonymous reviewers for their constructive remarks and corrections, which helped us to improve the quality of the manuscript.

Declarations

Author contributions CS proposed the initial idea. YA, CS, MV, and RK designed the algorithm and experiments. YA and CS developed the software and wrote the manuscript. All analyzed the results of the study and reviewed the manuscript.

Data availability The UELN data are provided by the Federal Agency for Cartography and Geodesy (BKG). These data are not publicly available. The tide gauge locations are obtained from the publicly available PSMSL database (https://www.psmsl.org/data/obtaining/).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aigner M, Ziegler GM (1998) Cayley's formula for the number of trees. In: Proofs from THE BOOK. Springer, Berlin, pp 141–146. https://doi.org/10.1007/978-3-662-22343-7_22

Amiri-Simkooei AR, Asgari J, Zangeneh-Nejad F, Zaminpardaz S (2012) Basic concepts of optimization and design of geodetic networks. J Surv Eng 138(4):172–183. https://doi.org/10.1061/(asce)su.1943-5428.0000081

Amjadiparvar B, Rangelova E, Sideris MG (2015) The GBVP approach for vertical datum unification: recent results in North America. J Geod 90(1):45–63. https://doi.org/10.1007/s00190-015-0855-8

Amos MJ, Featherstone WE (2008) Unification of New Zealand's local vertical datums: iterative gravimetric quasigeoid computations. J Geod 83(1):57–68. https://doi.org/10.1007/s00190-008-0232-y

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

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

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

Deltares (2021) Delft3D flexible Mesh suite. https://content.oss.deltares.nl/delft3d/manuals/D-Flow_FM_User_Manual.pdf

Denker H (2015) A new European Gravimetric (Quasi)Geoid EGG2015. 26th IUGG General Assembly, June 22–July 2, Prague, Czech Republic

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

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

Featherstone W, Filmer M (2012) The north-south tilt in the Australian height datum is explained by the ocean's mean dynamic topography. J Geophys Res Oceans. https://doi.org/10.1029/2012JC007974

Federal Geodetic Control Committee, et al (1984) Standards and specifications for geodetic control networks. NOAA, National Geodetic Information Branch, N/CG17X2, Sect 2:1

Filmer MS, Featherstone WE (2012) A re-evaluation of the offset in the Australian height datum between Mainland Australia and Tasmania. Mar Geod 35(1):107–119. https://doi.org/10.1080/01490419.2011.634961

Filmer MS, Featherstone WE, Claessens SJ (2014) Variance component estimation uncertainty for unbalanced data: application to a continent-wide vertical datum. J Geod 88(11):1081–1093. https://doi.org/10.1007/s00190-014-0744-6

Filmer MS, Hughes CW, Woodworth PL, Featherstone WE, Bingham RJ (2018) Comparison between geodetic and oceanographic approaches to estimate mean dynamic topography for vertical datum unification: evaluation at Australian tide gauges. J Geod 92(12):1413–1437. https://doi.org/10.1007/s00190-018-1131-5

Gerlach C, Rummel R (2012) Global height system unification with GOCE: a simulation study on the indirect bias term in the GBVP approach. J Geod 87(1):57–67. https://doi.org/10.1007/s00190-012-0579-y

Heck B, Rummel R (1990) Strategies for solving the vertical datum problem using terrestrial and satellite geodetic data. In: Sea surface topography and the Geoid. Springer, New York, pp 116–128. https://doi.org/10.1007/978-1-4684-7098-7_14

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

Holgate SJ, Matthews A, Woodworth PL, Rickards LJ, Tamisiea ME, Bradshaw E, Foden PR, Gordon KM, Jevrejeva S, Pugh J (2013) New data systems and products at the permanent service for mean sea level. J Coast Res 29(3):493–504. https://doi.org/10.2112/JCOASTRES-D-12-00175.1

Ihde J, Augath W, Sacher M (2002) The Vertical Reference System for Europe. In: International association of Geodesy symposia.

Springer, Berlin, pp 345–350. https://doi.org/10.1007/978-3-662-04683-8_64
Lea DJ, Mirouze I, Martin MJ, King RR, Hines A, Walters D, Thurlow M (2015) Assessing a new coupled data assimilation system based on the met office coupled atmosphere-land-ocean-sea ice model. Mon Weather Rev 143(11):4678–4694. https://doi.org/10.1175/mwr-d-15-0174.1
Mehlstäubler TE, Grosche G, Lisdat C, Schmidt PO, Denker H (2018) Atomic clocks for geodesy. Rep Progress Phys 81(6):064 401. https://doi.org/10.1088/1361-6633/aab409
Moritz H (2000) Geodetic Reference System 1980. J Geod 74(1):128–162. https://doi.org/10.1007/s001900050278
Müller J, Dirkx D, Kopeikin SM, Lion G, Panet I, Petit G, Visser PNAM (2017) High performance clocks and gravity field determination. Space Sci Rev 214(1):5. https://doi.org/10.1007/s11214-017-0431-z
Ogundare JO (2018) Understanding least squares estimation and geomatics data analysis. Wiley, New York. https://doi.org/10.1002/9781119501459.ch14
Penna NT, Featherstone WE, Gazeaux J, Bingham RJ (2013) The apparent British sea slope is caused by systematic errors in the levelling-based vertical datum. Geophys J Int 194(2):772–786. https://doi.org/10.1093/gji/ggt161
Permanent Service for Mean Sea Level (PSMSL) (Retrieved 01 Jun 2020) Tide Gauge Data. http://www.psmsl.org/data/obtaining/
Proudman J (1953) Dynamical oceanography. Methuen, New York
Rummel R, Teunissen P (1988) Height datum definition, height datum connection and the role of the geodetic boundary value problem. Bull Géodésique 62(4):477–498. https://doi.org/10.1007/bf02520239
Sacher M, Liebsch G (2019) EVRF2019 as new realization of EVRS. https://evrs.bkg.bund.de/SharedDocs/Downloads/EVRS/EN/Publications/EVRF2019_FinalReport.pdf?__blob=publicationFile&v=2
Sánchez L, Sideris MG (2017) Vertical datum unification for the International Height Reference System (IHRS). Geophysical Journal International p ggx025, https://doi.org/10.1093/gji/ggx025
Schwarz KP, Sideris MG, Forsberg R (1987) Orthometric heights without leveling. J Surv Eng 113(1):28–40
Seemkooei AA (2001a) Comparison of reliability and geometrical strength criteria in geodetic networks. J Geod 75(4):227–233. https://doi.org/10.1007/s001900100170
Seemkooei AA (2001b) Strategy for designing geodetic network with high reliability and geometrical strength. J Surv Eng 127(3):104–117. https://doi.org/10.1061/(asce)0733-9453(2001)127:3(104)
Slobbe DC, Klees R, Verlaan M, Zijl F, Alberts B, Farahani HH (2018) Height system connection between island and mainland using a hydrodynamic model: a case study connecting the Dutch Wadden islands to the Amsterdam ordnance datum (NAP). J Geod 92(12):1439–1456. https://doi.org/10.1007/s00190-018-1133-3
Stammer D, Ray RD, Andersen OB, Arbic BK, Bosch W, Carrère L, Cheng Y, Chinn DS, Dushaw BD, Egbert GD et al (2014) Accuracy assessment of global barotropic ocean tide models. Rev Geophys 52(3):243–282. https://doi.org/10.1002/2014rg000450
Tarjan R (1972) Depth-first search and linear graph algorithms. SIAM J Comput 1(2):146–160
Teunissen PJG (2006) Network quality control. VSSD
Wilkinson K, von Zabern M, Scherzer J (2014) Global Freshwater Fluxes into the World Oceans. http://doi.bafg.de/BfG/2014/GRDC_Report_44.pdf
Woodworth PL, Hughes CW, Bingham RJ, Gruber T (2013) Towards worldwide height system unification using ocean information. J Geod Sci 2(4):302–318. https://doi.org/10.2478/v10156-012-0004-8
Zijl F, Veenstra J, Groenenboom J (2020) The 3D Dutch continental shelf model-flexible mesh (3D DCSM-FM). https://www.deltares.nl/app/uploads/2020/12/Development-of-a-3D-model-for-the-NW-European-Shelf-3D-DCSM-FM.pdf