NOAA Technical Report NOS NGS 72

GEOID 18

Kevin Ahlgren
Galen Scott
David Zilkoski
Brian Shaw
Nagendra Paudel

Silver Spring, MD
May 2020

Acknowledgements

GEOID18 would not be what it is without the many hours that thousands of volunteer surveyors, state and local agencies, and federal agency partners put into collecting GPS data on bench marks. NGS sincerely thanks all the people who participated in data collection over the past few years and who have helped create GEOID18.

The authors would like to thank Joe Evjen and Yan-Ming Wang for providing comments and reviewing this document and Lina Dutky for editing. Additional thanks goes to members of the GEOID18 Advisory Team which includes Michael Dennis, John Ellingson, Joe Evjen, Dan Roman, and Yan-Ming Wang for providing guidance throughout the GEOID18 project.

Table of Contents

1 Executive Summary

NOAA Technical Report NOS NGS 72

GEOID18

GEOID18 is the National Geodetic Survey's (NGS) latest hybrid geoid model, which enables GPS users to access orthometric heights that are consistent with official vertical datums of the National Spatial Reference System (NSRS). GEOID18 covers the 48 CONtiguous United States (CONUS), Puerto Rico, and the U.S. Virgin Islands, includes 1 arcminute grids for the geoid height, the geoid height estimated uncertainty, and deflections of the vertical (DEFLEC18), and is accessible at geodesy.noaa.gov/GEOID/GEOID18.

GEOID18 shows significant improvement over its predecessor, demonstrating an 18% smaller overall standard deviation (1.39 cm versus 1.7 cm) for the CONUS region. The overall performance of GEOID18 is further remarkable considering the model is based on 29% more GPS on bench mark observations (32,000+ vs. 24,900+). These benefits are widespread and not restricted to just a handful of states with 41 out of 51 states/regions having more GPS on bench marks used in the modeling process, and 47 out of 51 states/regions experiencing smaller standard deviations compared with GEOID12B. Unlike prior NGS hybrid geoids, GEOID18 does not cover Alaska or Pacific islands.

There are two main reasons for the improvement. First, significant advancements in gravimetric geoid modeling theory and gravimetric data quality have occurred since 2012. Overall, this improvement causes approximately 1.2 mm of improvement in GEOID18, and occurs in 39 out of 51 states/regions. Secondly and most importantly, NGS has a much larger and more refined GPS on bench mark dataset, which provides the most significant impact on GEOID18's performance. Overall, the GPS on bench marks provide 4.6 mm of improvement with a large number of states (20 out of 51) experiencing 5+ mm improvements from GEOID12B to GEOID18.

DEFLEC18 provides hybrid deflections of the vertical on the Earth's surface. GEOID18 heights are used to compute deflections on the geoid surface, which are then corrected for the plumbline curvature based on Helmert's definition (Torge, 1991; Jekeli, 1999) and a digital elevation model (DEM). Comparisons with historical deflection observations shows accuracies at the 1.1 to 1.2 arc-second level (in root mean squared, RMS) over all of CONUS. Additional comparisons with more recent deflection observations are even better, showing accuracies at the 0.15" to 0.30" RMS level along NGS's geoid slope validation surveys (GSVS11 and GSVS14) survey lines.

Find this entire report here:

https://geodesy.noaa.gov/library/pdfs/NOAA_TR_NOS_NGS_0072.pdf

2 Purpose

The purpose of the GEOID18 hybrid geoid model is to provide geoid heights that can be combined with GPS-derived NAD 83 (2011) epoch 2010.00 ellipsoid heights to produce values consistent with the official vertical datum published orthometric heights on NGS Data Sheets. Throughout this document, the official vertical datums (NAVD 88 for CONUS, PRVD02 for Puerto Rico, and VIVD09 for the U.S. Virgin Islands) are referred to as vertical datum, or VD. Hybrid geoid models are created by constraining a gravimetric geoid model to published heights using GPS observations on leveled bench marks. GEOID18 is intended to be the last hybrid geoid model that NGS creates before the current vertical datums are replaced by the North American-Pacific Geopotential Datum of 2022 (NAPGD2022) .

A hybrid geoid model requires two datasets:

1. A purely gravimetric geoid model, xGEOID19B, (Li, et al., 2019), which is created from a number of terrestrial, airborne, and space-based gravimetric datasets
2. A network of passive bench marks, GPSBM18, (Ahlgren, et al., 2020), where both the ellipsoid height from GPS and the orthometric height from geodetic leveling are observed

Combining these two datasets gives the hybrid geoid model the positive attributes from each individual dataset. The gravimetric geoid model is very accurate over long distances (or wavelengths) like the 2,500 miles from Florida to the state of Washington. The gravimetric geoid is also a continuous surface with no gaps, allowing areas where no leveling lines are present to be accurately modeled. The GPS on bench mark data, where they exist, accurately reflect the shape of the VD and provide higher resolutions over small geographic regions. The fusion of these datasets results in a hybrid that is both seamless and accurate at all distances.

The following sections focus on the methodology, input datasets, analysis, and performance of GEOID18 and associated products:

- Methodology
- Input Datasets
    - Gravimetric Geoid Model, xGEOID19B
    - GPS on Bench Marks dataset (GPSBM18)
- Results
- Performance Analysis
    - Omission Error Analysis of GPS on bench marks and GEOID18 model
    - Commission Error Analysis of GPS on bench marks and GEOID18 model
- Geoid uncertainty model
- Surface deflection of the vertical model (DEFLEC18)

The following sections provide technical information, but avoid detailed background or explanations of the mathematical or geophysical concepts and do not cover all possible scenarios for usage of GEOID18, the geoid uncertainty grid, and DEFLEC18. For definitions of terms used in this document, please refer to the NGS Geodetic Glossary. For technical assistance in using these products, contact an NGS Regional

Geodetic Advisor. For questions related to the technical details of the data or methods used in the modeling, contact the NGS GEOID Team.

3 Technical Specifications

3.1 Spatial Resolution

The GEOID18 model and associated gridded products are provided as 1 arc-minute grids. The GPS on bench mark dataset is provided at discrete, heterogeneous locations, which are limited by the availability of bench marks and GPS observations.

3.2 Grid Cell Values

Geoid heights are reported as 4-byte binary numbers. This implies 10^{-38} precision, but practically, the precision is limited to approximately 0.1 mm. For more discussion on how NGS implemented the number of digits and the implied accuracy in the past and plans to do so in the future, see NGS (2019).

3.3 Units

Geoid heights are in meters. The estimated uncertainty is provided in meters at 1-sigma (1σ) standard in the grid, and are scaled by 1.96 to get the 95% confidence interval for users of the NGS Online Computation Tool (https://geodesy.noaa.gov/GEOID/GEOID18/computation.html).

3.4 Datums

GEOID18 is intended for use with coordinates in the North American Datum of 1983 (2011) [NAD 83 (2011) epoch 2010.00]. It provides orthometric heights consistent with the North American Vertical Datum of 1988 (NAVD 88), the Puerto Rico Vertical Datum of 2002 (PRVD02), or the Virgin Islands Vertical Datum of 2009 (VIVD09), depending on location.

GEOID18 does not incorporate any time dependency in the model, which is consistent with the static nature of the VD. In areas with significant vertical land motion, discrepancies in GEOID18 have been mitigated with updated leveling adjustments, outlier detection schemes, and redundant observations, as permissible.

3.5 Coverage Area

GEOID18 is developed specifically for the 48 CONtiguous United States (CONUS) and the U.S. territories of Puerto Rico and the U.S. Virgin Islands, as defined in Table 1. It is not recommended for use outside of the land covered portions of these areas due to insufficient GPS on bench mark constraints and/or lack of jurisdiction. The coverage area does include some foreign territories including Canada, Mexico, Bermuda, and the Bahamas but has no basis for use in these territories.

Table 1: Geographic Areas for GEOID18 and other related products:

1’ resolution	Min. Latitude	Max. Latitude	Min. East Longitude	Max. East Longitude
CONUS	24.0	58.0	230.0	300.0
Puerto Rico + U.S. Virgin Islands	15.0	21.0	291.0	296.0

Table 2: Geographic Areas NOT INCLUDED in GEOID18:

Users should continue to use GEOID12B in these areas.				
Alaska	*49.0*	*72.0*	*172.0*	*234.0*
Hawaii	*18.0*	*24.0*	*199.0*	*206.0*
Guam / Northern Mariana Islands	*11.0*	*18.0*	*143.0*	*146.0*
The ASVD 02 vertical datum has been superseded by Federal Register Notice and previous hybrid geoids are not supported for further use in this region.				
American Samoa	*-17.0*	*-11.0*	*186.0*	*192.0*

4 Methodology

The following section describes the methodology used to create GEOID18 and associated products. GEOID18 consists of two regions: CONUS and Puerto Rico/U.S. Virgin Islands. The CONUS geoid grid and the Puerto Rico/U.S. Virgin Islands geoid grid are both computed independently from one another, so equations that refer to NAVD 88 in the following section for CONUS are also applied separately to PRVD02/VIVD09 results.

4.1 Least-Squares Collocation, Residuals, and Covariance Function

The hybrid geoid methodology makes use of a residual as defined by the following equation, which is computed at every bench mark in the GPSBM18 file:

$$r_{raw} = N_{xGEOID19B} - (h_{NAD83(2011)} - H_{NAVD\ 88}) \quad (1)$$

where:
$N_{xGEOID19B}$ is the gravimetric **geoid height** determined by the NGS xGEOID19B.
$h_{NAD83(2011)}$ is the GPS derived **ellipsoid height** with respect to NAD83(2011).
$H_{NAVD\ 88}$ is the **orthometric height** derived from leveling with respect to the vertical datum.

Like many leveling-based vertical datums (Zilkowski, et al, 1992, Featherstone and Filmer, 2012), the residuals obtained in (1) are contaminated with a continental tilt and bias that is estimated and removed with a simple two-dimensional planar surface of the form shown in (2) using all valid bench marks and least squares to estimate the 3 unknown parameters (*A*, *B*, and *C*). It is outside the scope of GEOID18 to determine exactly why a bias and tilt exist in (1), and the goal with GEOID18 is to remove any systematic effects in (1) in order to perform the least squares collocation.

$$\lambda * A + \varphi * B + C = r_{raw} \quad (2)$$

where:
λ is the longitude
φ is the latitude
A, B, C are the unknown parameters of the plane that are solved for with least squares.

Removing the tilt and bias from the residuals in (1) results in bias-free and tilt-free residuals, which will be referred to as pre-model residuals throughout the remainder of this document and is shown in (3):

$$r_{pre-model} = N_{xGEOID19B} - \{h_{NAD83(2011)} - H_{NAVD\ 88} - (\lambda * \hat{A} + \varphi * \hat{B} + \hat{C})\} \quad (3)$$

where:
$\hat{A}, \hat{B}, and\ \hat{C}$ are the estimated parameters of the 2D plane obtained from least squares.

The pre-model residuals are then used to determine a mathematical model using least squares collocation (LSC) (Moritz, 1980) and multiple Gaussian functions to describe the behavior seen at the bench marks in the form of a covariance function (Roman, et al. 2004). The general form of the covariance function used to create the analytical signal for GEOID18 is shown in (4).

$$C_{ll} = A_0 e^{-(\frac{D_{ll}}{L})^2} \quad (4)$$

where:

C_{ll} is the covariance between two points
A_0 is the amplitude at auto-correlation
D_{ll} is the distance between two points
L is the correlation length

To more accurately reflect the varying correlation at different distances, GEOID18 makes use of (4) with six differing $A_0$and L parameters that are added together to obtain the final multi-matrix variance-covariance function for CONUS as shown in (5). For the PRVI computations, only two correlation lengths are used.

$$C_{tt} = C_{l_1 l_1} + C_{l_2 l_2} + C_{l_3 l_3} + C_{l_4 l_4} + C_{l_5 l_5} + C_{l_6 l_6} \quad (5)$$

Once the relationship between the points is modeled, a regular 1 arc-minute grid over the entire domain is interpolated using the following equation from LSC where all GPS on bench marks are included in the prediction:

$$\hat{s} = C_{st} * (C_{tt} + C_{nn})^{-1} * r_{pre-model} \quad (6)$$

where:
$\hat{s}$ is the vector of predicted values at the 1 arc-minute grid cells. This vector contains the local 'warping' component due to the GPS on bench marks that is one component of the predicted $N_{GEOID18}$ values.
C_{st} is the variance-covariance matrix between the prediction location and each observation, $r_{pre-model,i}$
C_{nn} is the variance-covariance matrix of random noise for the observations in $r_{pre-model}$

Finally, the bias/tilt component and the xGEOID19B component that are removed at the point level in (1) and (3) must be restored as shown in (7). With all components now in 1 arc-minute grids, the warped component found in (6) is combined with the continental tilt and bias surface found from (2) and the gravimetric geoid surface from xGEOID19B. This results in the final hybrid geoid model, GEOID18.

$$N_{GEOID18} = N_{\hat{s}} + N_{\lambda * \hat{A} + \varphi * \hat{B} + \hat{C}} + N_{xGEOID19B} \quad (7)$$

where:
$N_{\hat{s}}$ is a 1 arc-minute gridded output that is obtained from the vector found in (6)
$N_{\lambda * \hat{A} + \varphi * \hat{B} + \hat{C}}$ is a 1 arc-minute grid obtained from the predicted parameters ($\hat{A}, \hat{B}, and\ \hat{C}$) from (2)
$N_{xGEOID19B}$ is a 1 arc-minute grid from the gravimetric geoid, xGEOID19B

An additional step can be taken to evaluate the performance of the hybrid geoid model by computing a *post-model residual* at every GPS on bench mark. The post-model residual will be investigated in a number of situations later in this document.

$$r_{post-model} = N_{GEOID18} - (h_{NAD83(2011)} - H_{NAVD\ 88}) \quad (8)$$

In additional to the actual geoid values that are obtained at any given location from (7), the formal estimated uncertainty can be determined at any location. The estimated uncertainty of the predicted values in $\hat{s}$ from (6) can also be estimated from the following (Moritz, 1980):

$$\sigma^2{}_{correlated} = C_{ss} - C_{st} * (C_{tt} + C_{nn})^{-1} * C_{ts} \quad (9)$$

where:

C_{ss} is the autocorrelation or $C(0)$

$$C_{ts} = C_{st}{}^T$$

Since the estimated uncertainty from (9) is only that component which is from the correlation, an additional uncorrelated error component is added to the GEOID18 total estimated uncertainty as shown in (10). The rationale behind this uncorrelated error is that only the leveling and gravimetric geoid portion of (1) can be fully captured by the correlated component. The GPS-derived ellipsoid height, which might be partially corrected to neighboring marks, is much more uncorrected. This is especially true as we incorporate OPUS Share solutions that haven't been processed in a combined network adjustment, processed with different arrangements of the NOAA CORS Network, etc. The uncorrelated amount is 1.4 cm and 1.7 cm for CONUS and Puerto Rico/U.S. Virgin Islands, respectively. Under ideal circumstances, all three components in (8) would have their own associated error estimates. NGS currently does not have error estimates associated with xGEOID19B, so the uncorrelated term in (10) attempts to absorb that and effectively acts as a 'floor' or minimum estimated uncertainty for GEOID18.

$$\sigma_{N_{GEOID18}} = \sqrt{(\sigma^2{}_{uncorrelated} + \sigma^2{}_{correlated})} \quad (10)$$

5 Input Data

The following section will highlight the two input datasets that were used in the construction of GEOID18: xGEOID19B and GPSBM18. Additionally, the covariance function model that blends these two datasets together will also be shown. The majority of the discussion will focus on the GPSBM18 dataset and how it was assembled, since a considerable amount of effort was invested in that process.

5.1 Gravimetric Geoid Model

GEOID18 is based on NGS's most recent gravimetric geoid model, xGEOID19B (Li, et al., 2019). xGEOID19B is the latest in a series of Experimental Geoid Models (xGEOIDs) that show what National Spatial Reference System users can expect with the forthcoming, modernized NAPGD2022 datum. xGEOID19B uses the latest gravity data from surface, airborne, altimetry, and satellite models along with an updated DEM. The result is that xGEOID19B, and therefore GEOID18, make use of all of the airborne gravity data available from the NGS GRAV-D project as of July 2018. There are 53 GRAV-D blocks throughout CONUS, Alaska, and Puerto Rico/U.S. Virgin Islands that are included in the xGEOID19B model, and their impact on GEOID18 is illustrated in Figure 1. The greatest GRAV-D contribution to GEOID18 is that it provides a consistent surface from land to water bodies, which helps improve the quality of the geoid in littoral areas and across international borders where data is often lacking. A handful of states that exhibit significant improvement (2 cm+) over modest areas include Arizona, California, Maine, Louisiana, Texas, and Washington. Since the gravimetric model is based on IGS08, a transformation to NAD83 (2011) epoch 2010.0 must be done using the 14 parameter Helmert transformation (https://geodesy.noaa.gov/CORS/coords_alt.shtml, Soler and Snay (2004).

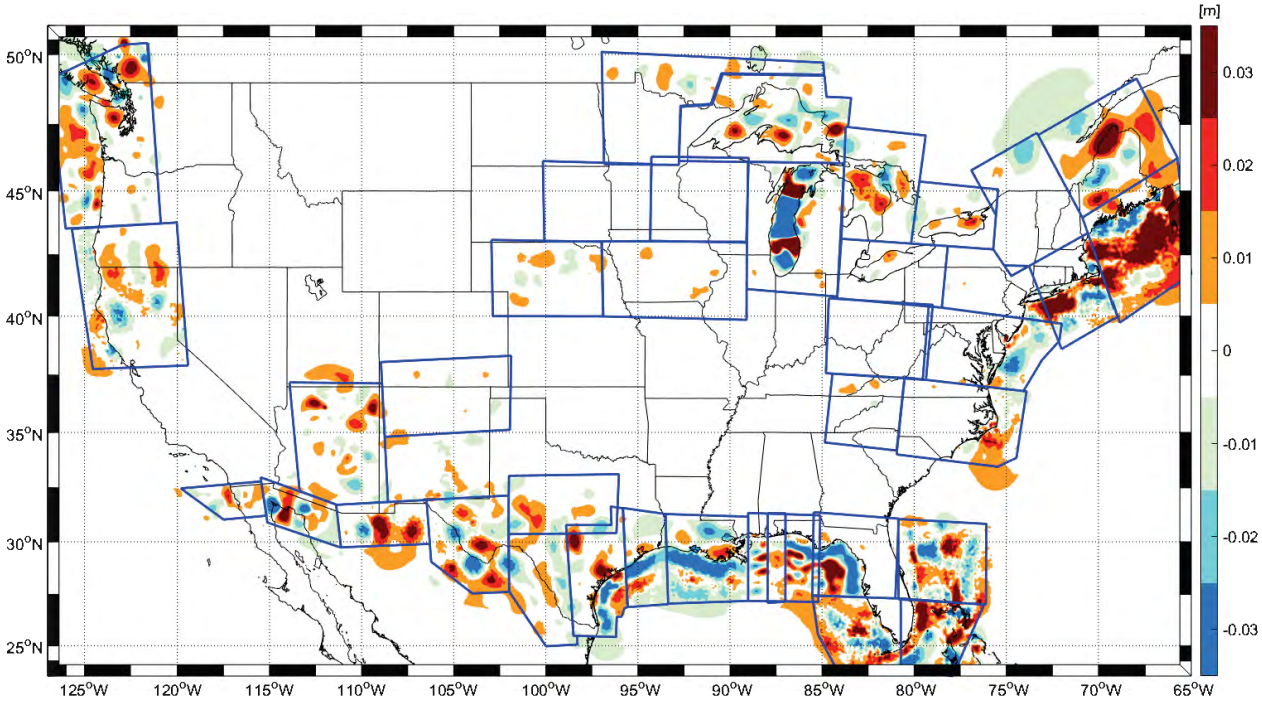

Figure 1: Impact of GRAV-D on GEOID18 — Difference between models using xGEOID19B and xGEOID19A. Blue outlines show the general GRAV-D Block boundaries used in xGEOID19B and GEOID18.

5.2 GPS on Bench Marks

Since 2014, NGS has sponsored annual crowdsourced data collection campaigns called GPS on Bench Marks (GPSonBM) to help improve the accuracy and geographic coverage of GEOID18 and other NGS products. For many of these years, NGS has worked with the National Society of Professional Surveyors (NSPS) to promote participation during National Surveyors Week each March. In 2018 alone, nearly 600 people and agencies from across CONUS and Puerto Rico submitted over 3,800 four-hour GNSS observations on about 2,500 bench marks. This additional data has significantly improved the model by closing data gaps and resolving conflicts in older data.

5.2.1 GPS on Bench Marks Overview

For GEOID18, a total of 32,357 bench marks are used in the modeling in CONUS and 127 bench marks are used in Puerto Rico and the U.S. Virgin Islands as illustrated in Figure 2 and Figure 3, respectively. The ellipsoid height component of this dataset was used from one of two NGS databases: 1) the Integrated Database (IDB) through "Bluebooking" and 2) the OPUS Database (also called OPUS Share or OPUS Shared Database). A great deal of analysis and quality control work was undertaken by NGS to arrive at the above mentioned list of high-quality bench marks and is presented in the section below.

Overall, 4,224 bench marks in CONUS are not used in the geoid model construction as these were deemed to be outliers or are derived from single GPS occupations (see Figure 4). In Puerto Rico and the U.S. Virgin Islands, there were 18 total outliers or single GPS occupations (see Figure 5). The entire GPS on Bench Mark dataset (Ahlgren, et al., 2020) is available on the NGS website and should be considered a companion to this technical report. State-by-state statistics can be found in Appendix I.

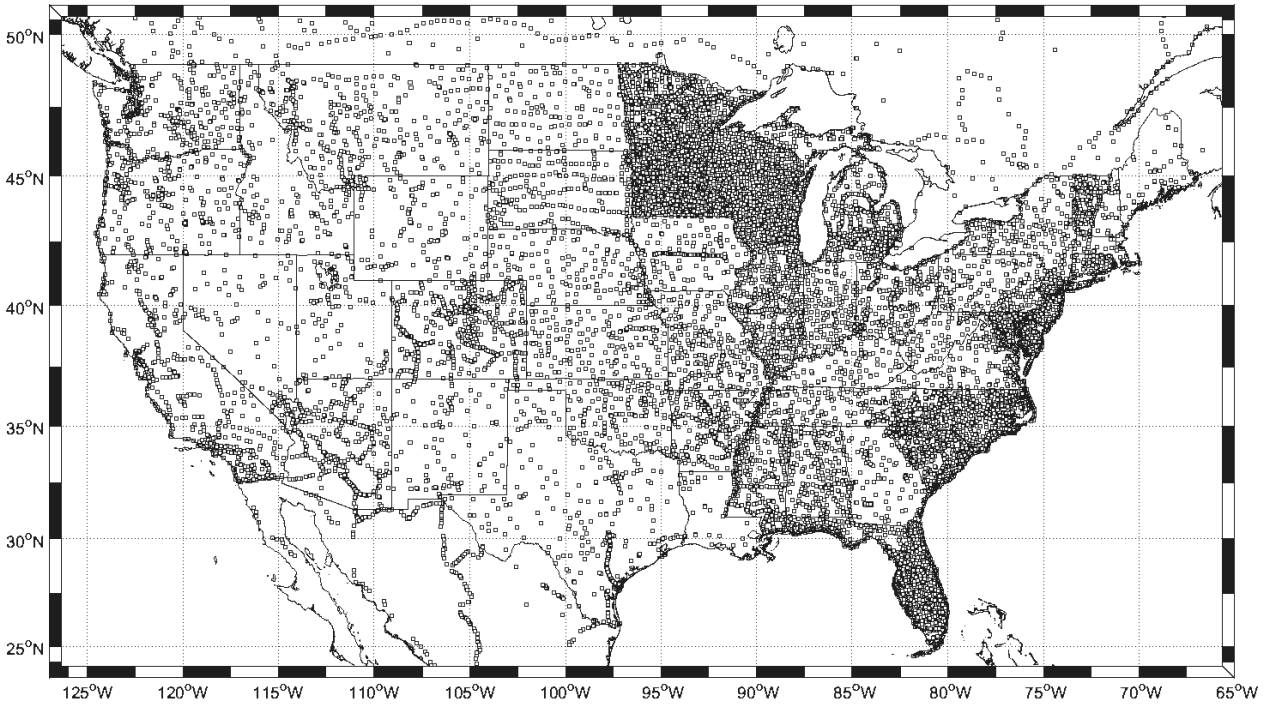

Figure 2: GPS on BMs dataset used in GEOID18 for CONUS

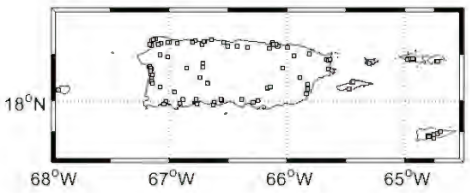

Figure 3: GPS on BMs dataset used in GEOID18 for Puerto Rico/U.S. Virgin Islands

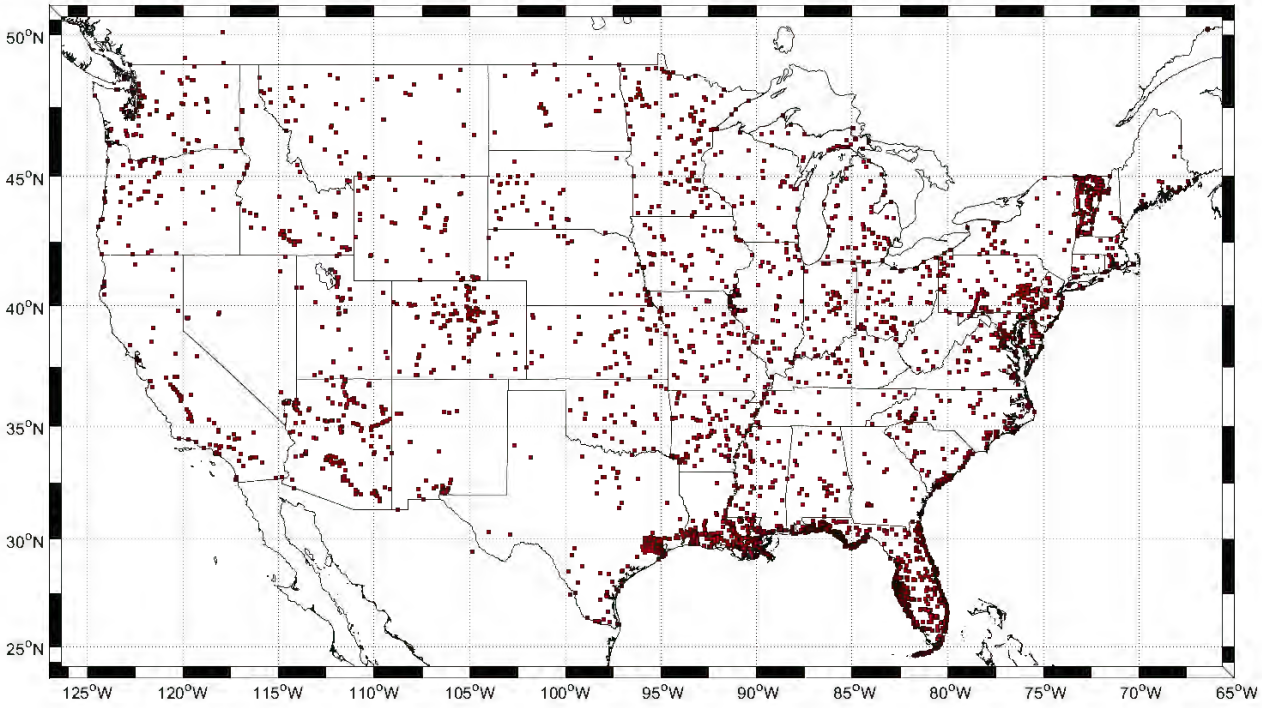

Figure 4: GPS on BMs not used in GEOID18 but available in the supplemental dataset

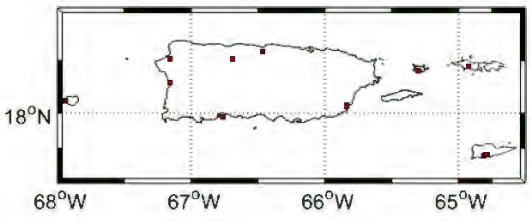

Figure 5: GPS on BMs not used in GEIOD18 but available in the supplemental dataset

Additional context is provided in the rest of this section to highlight some of the nuances between the two ellipsoid height data sources (IDB and OPUS Share). In Figure 6 and Figure 7, the bench marks USED in GEOID18 with ellipsoid heights coming from the OPUS Share database are shown for CONUS and Puerto Rico/U.S. Virgin Islands, respectively. The majority of these used bench marks have 2+ satisfactory GPS observations; however, there are a few bench marks used in GEOID18 that are derived from a single OPUS Share observation in regions with limited GPS on Bench Mark coverage. An even smaller number of bench marks utilize an OPUS Share-derived ellipsoid height that was found to be

superior to the corresponding IDB ellipsoid height. These OPUS Share observations that supersede the IDB values were only accepted on bench marks with 2+ OPUS Share observations and the approximate improvement was at the 10 to 20 cm-level. Inevitably, there are additional OPUS Share heights that outperform the IDB height on an individual bench mark at a few cm-level, but identifying those differences were not within the scope of this project.

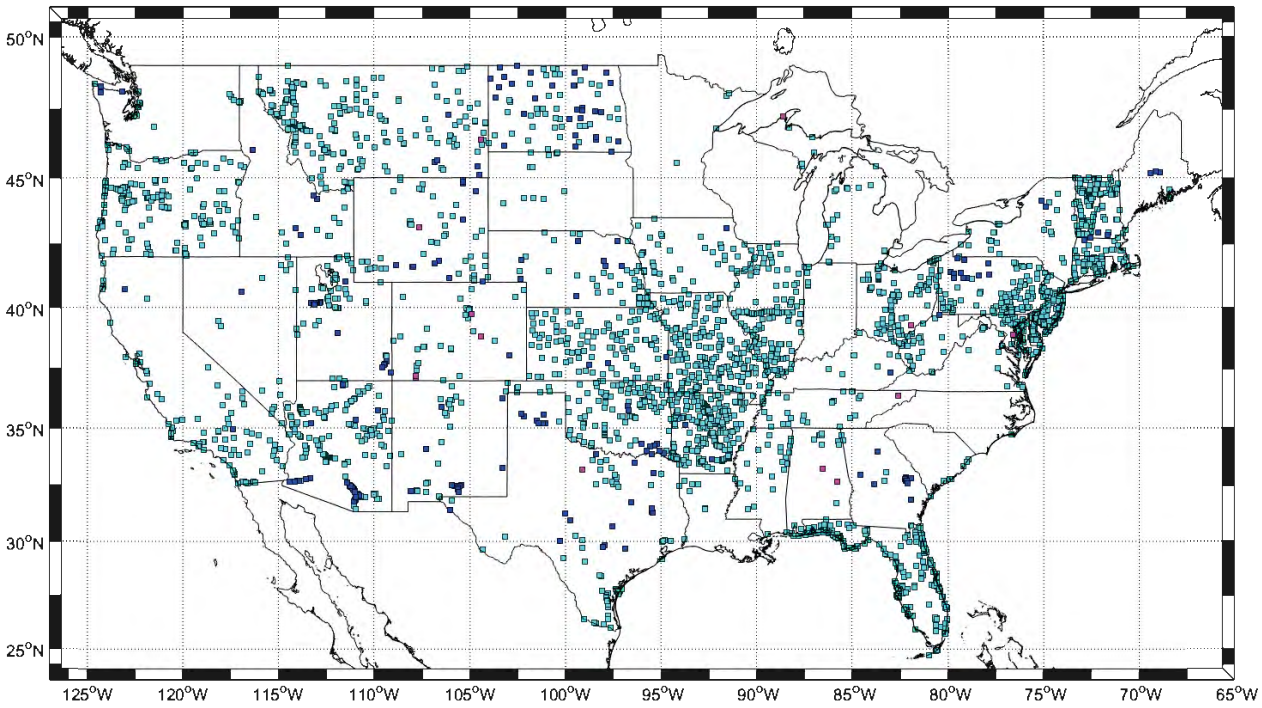

Figure 6: OPUS Share used bench marks — Cyan: OPUS 2+ observations; Blue: Single observations; Magenta: OPUS observation supersedes the IDB.

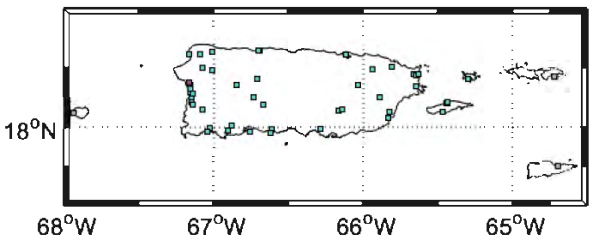

Figure 7: OPUS Share used bench marks in PRVI — Cyan: OPUS 2+ observations; Magenta: OPUS observation supersedes the IDB.

In a very limited number of cases, a bench mark with a 'No Check' ellipsoid height originating in the NGS IDB is used (see Figure 8). The 'No Check' category simply means that the coordinates are based on a single GPS occupation.

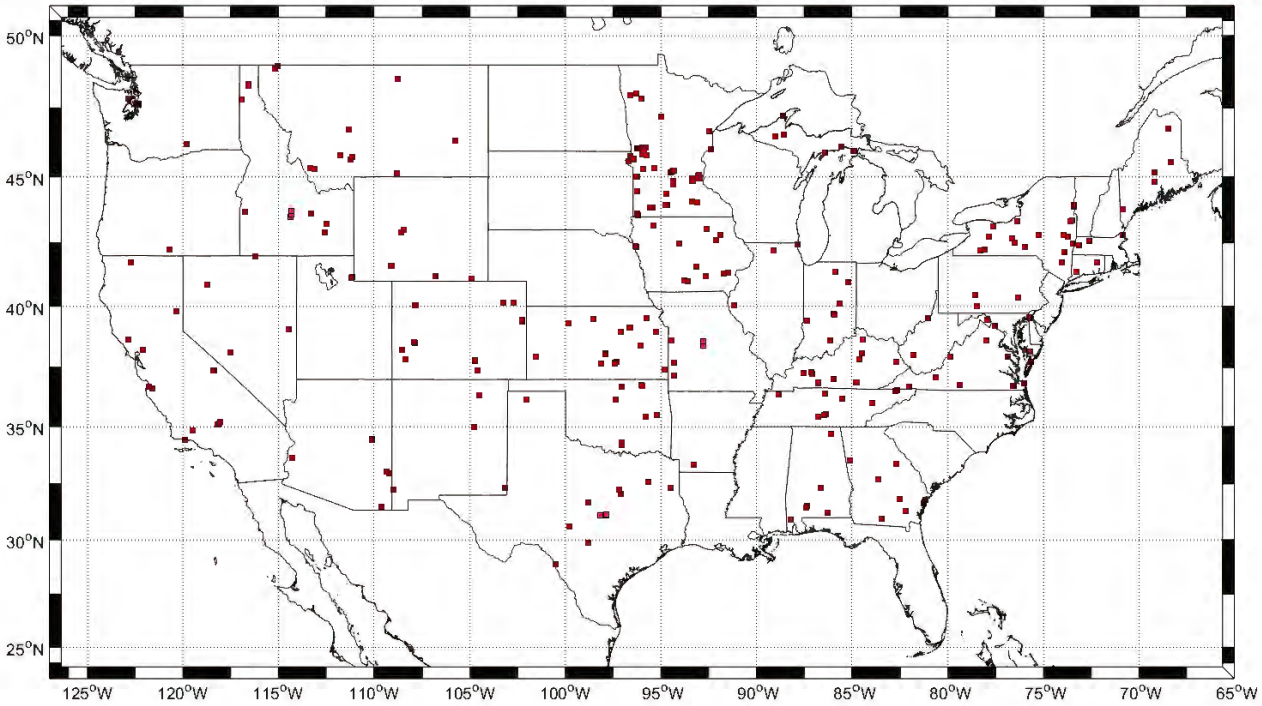

Figure 8: 'No Check' GPS solutions from the NGS IDB used in GEOID18

In addition to the overall GPS on Bench Marks distribution shown in the preceding figures, many individual states have dramatically increased the number of bench marks available. Figure 9 shows a state-by-state summary of the percent increase in the number of GPS on Bench Marks used from GEOID12B to GEOID18. The positive values signify an increase in the number of used GPS on Bench Marks while negative values signify a decrease. Additional state-by-state summary statistics can be found in Appendix I.

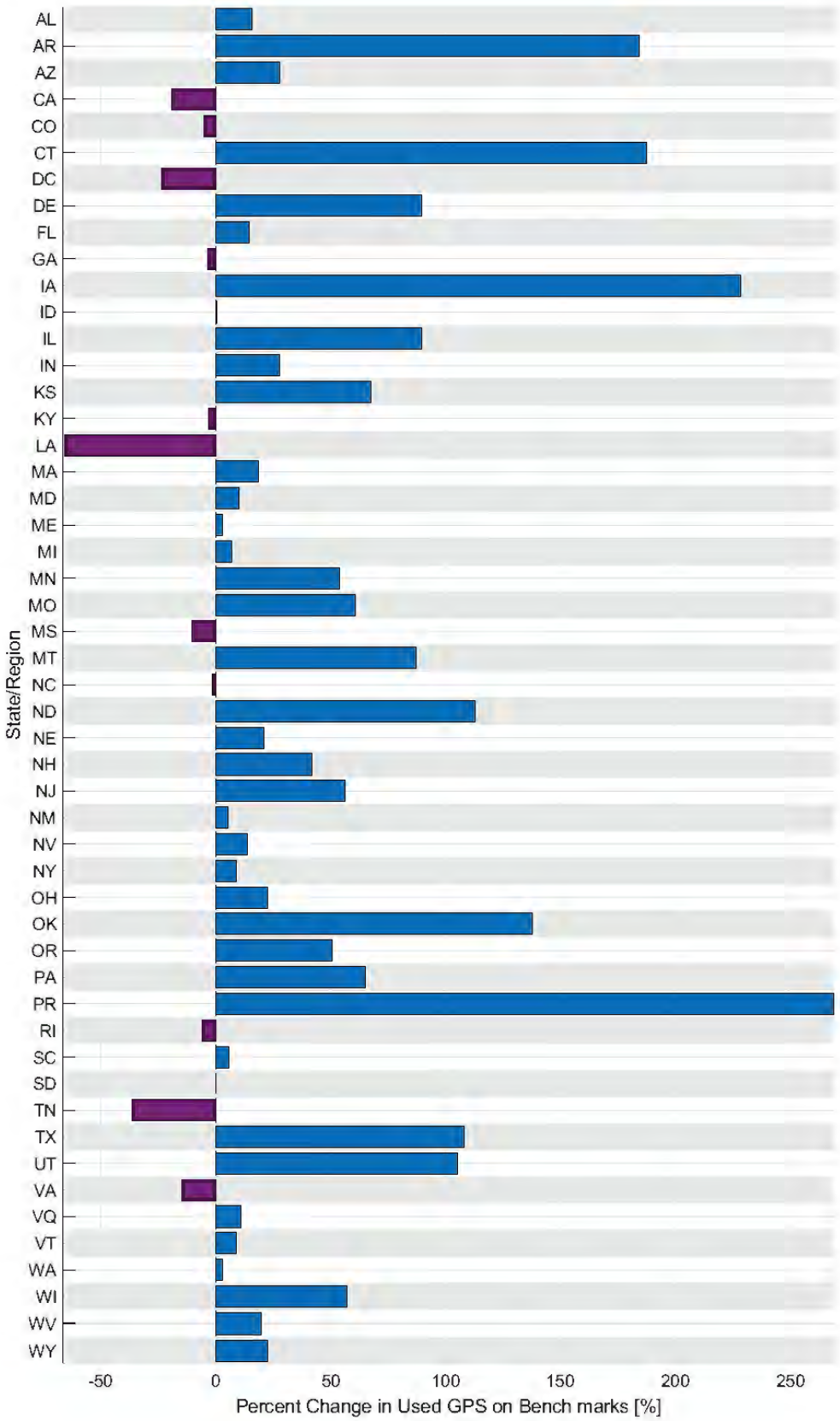

Figure 9: Change in GPS on Bench Marks used in GEOID18 compared with GEOID12B. Blue/positive changes signify increased number of bench marks.

5.2.2 GPS on Bench Marks Detailed Analysis

GPS on Bench Marks are a fundamental component in creating a hybrid geoid, but each bench mark should be valid; errors are not uncommon due to mark movement in the years between leveling and GPS, or height mistakes in some GPS surveys. Each hybrid geoid has rejected some GPS on Bench Marks, and the process for isolating and rejecting them becomes easier as the density and precision of the measurements improves.

NGS evaluated all available GPS on Bench Marks data to create GEOID18. Some stations with anomalous residuals were flagged and requested to be re-observed with GPS. New observations were then used to help determine if the vertical datum heights are no longer valid or if there are undetected errors in the published ellipsoid heights. Users supported the development of GEOID18 by strategically occupying those stations identified as outliers and/or located in areas devoid of GPS on Bench Marks stations. Users provided new GPS on Bench Marks data to NGS by 'Sharing' it through NGS' OPUS web tool and/or by submitting GPS projects to NGS for incorporation into the NSRS and publication by NGS (Bluebooking). To have a high level of confidence that the OPUS Share results were accurate, at least two matching, independent GPS observations were typically required for each mark. It should be noted that when a station was only occupied once, it was still useful for validating the hybrid geoid model.

The following questions needed to be addressed when analyzing the GPS on Bench Marks (GPS on BM) residual values:

1) Is the large GPS on BM residual due to an issue with the NAVD 88 orthometric height or the NAD 83 (2011) ellipsoid height?
2) Should larger residuals be tolerated in areas with fewer GPS on Bench Marks stations?

Many of the large GPS on BM residuals could be due to an invalid vertical datum height because the bench mark has moved since the last time the height was adjusted and published, and/or an undetected error in an ellipsoid height due to a weak GPS project design. Most of these stations with large GPS on Bench Mark residuals don't accurately represent the current vertical datum. Following the appropriate federal geodetic survey guidelines, procedures, and specifications when performing a geodetic survey, a user would identify these stations as bench marks with invalid heights. Therefore, these bench marks were not used in the hybrid geoid model just like they would not be used in controlling geodetic surveys. The goal of GEOID18 is to create a hybrid geoid model that is consistent with valid published vertical datum values.

The basic procedures and criteria used to identify and remove data outliers in the GPS on Bench Marks dataset are shown in Figure 10 and Table 3.

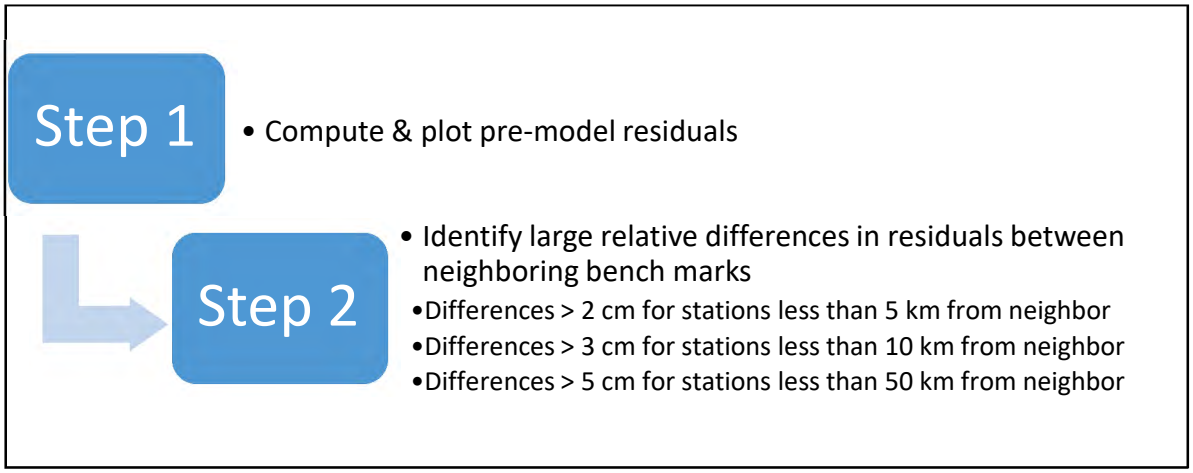

Figure 10: Procedure for GPS on Bench Marks analysis, which was repeated a number of times as iterations occurred.

Table 3: Rejection Criteria — Attributes considered when rejecting GPS on Bench Marks

	Attributes of leveling heights	Attributes of GPS ellipsoid heights:
1	Age of the leveling line used to establish the orthometric height	Station's local and network accuracy value
2	"New minus Old" values in areas where there was enough repeat leveling to estimate movement of bench marks	Station's involvement in the NA2011 Network adjustment — primary or secondary (age of GNSS data)
3	Bench marks heights based on leveling spurs	Network design used to establish the GPS-derived ellipsoid height
4	NGSIDB datasheet indicates that the bench mark's NAVD 88 height value was not estimated from an adjustment	GPS network design where there wasn't a direct connection between closely spaced stations that have large relative differences in outliers
5	Bench mark was only leveled once using single-run leveling procedure	OPUS Share vs IDB ellipsoid height source
6	Incorrect orthometric height due to poor modeled NAVD 88 gravity value	

5.2.3 Iterating Residual Analysis and New Model Creation

All residuals were evaluated using the criteria provided in the section above. This process of analysis and removal of GPS on Bench Marks was done iteratively, starting with residuals greater than +/- 7.5 cm. New versions of the hybrid geoid model were then created without the rejected stations and the review process was repeated. The team looked at both the pre-model residual (as described above and in (3)) and post-model residual from (8). Some of the rejections which had small post-model residuals were put back into the geoid model (i.e. un-rejected), and new stations with large pre-model and/or post-model residuals were further analyzed resulting in new recommendations about including or removing them from the geoid model.

Table 4 is an example of the documentation prepared and used to evaluate the GPS on Bench Marks. As previously mentioned, the pre-model residuals are used in the first round of analysis to determine which marks should be removed from the model, and in the second round of analysis, both the pre-model and post-model residuals are used as final confirmation that all used GPS on Bench Marks are appropriate and any removed GPS on Bench Marks should continue to be not used in the model. The table available for the second round of analysis is shown with an additional column (in yellow) showing the post-model residuals.

Table 4: Geodetic information for GPS on Bench Marks used in the analysis

PID	Pre-Modeled Residual v8.2.1 (cm)	Post-Modeled Residual v8.2.1(cm)	Designation	State	NAD 83 (2011) Ellipsoid Height (m)	NAVD 88 Orthometric Height (m)	Use (blank) or Do_Not_Use (X)	GNSS Information	Leveling Information	Comments
DK0537	18.8	9.6	X 240	MS	15.596	42.298	X	GNSS Obs in 2001	1993 - 1/2	Do not use - nearby stations with smaller residuals
DL9709	13.7	5.0	CHE 26	NC	467.202	495.927	X	GNSS Obs in 2009	2008 -2/2 - on a spur line	Do not use - nearby stations with smaller residuals
EH0774	10.4	3.4	J 223	AR	28.42	55.002		GNSS Obs in 2002	1976 - 1/2	Okay to use - large pre-modeled residual but consistent with neighbors
BG5044	9.4	4.4	P019	FL	-25.06	2.275		GNSS Obs in 1991	2015 - 1/2	Okay to use - recent leveling data; large pre-modeled residual but consistent with neighbors

5.2.4 Analysis of GPS on Bench Marks Residuals on OPUS-Share Stations

As previously mentioned, users have supported the development of GEOID18 by strategically occupying stations that were identified as outliers and/or were located in areas void of GPS on Bench Marks stations. Users provided their results using the NGS OPUS Share web tool. In most situations, at least two OPUS solutions on an individual bench mark were needed to be used in constraining GEOID18. As shown in Figure 6 and Figure 8, there are a small number of exceptions to this rule in areas devoid of GPS on Bench Mark coverage. Additionally, on bench marks with three or more OPUS solutions, an automatic outlier detection is performed that flags and removes any extreme outliers with respect to the ellipsoid height. This outlier detection is based on a scaled Median Absolute Deviation (MAD) technique where any observation outside the following range centered on the 'median of h' is considered an outlier and removed from the set:

$$\pm 3 * c * MAD = \pm 3 * 1.4826 * median(|h_i - median(h)|) \quad (11)$$

All of the remaining OPUS Share solutions' ellipsoid heights are then averaged into the final NAD83 (2011) ellipsoid height used in (1). To get a sense of how this outlier detection impacts the results, a handful of real situations encountered for GEOID18 are shown in Table 5. This outlier detection impacts 193 stations in GEOID18 at magnitudes of about 2 cm. The change in the mean ellipsoid height for these 193 stations is illustrated in Figure 11.

Table 5: Examples of outlier detection on OPUS Share results

	PID: AB0937			PID: AB4080			PID: HD0371		
Shared Solution	Observation date:	Ellipsoid height:	Outlier:	Observation date:	Ellipsoid height:	Outlier:	Observation date:	Ellipsoid height:	Outlier:
1	11/28/2006	-15.676		11/29/2017	-7.090	X	03/23/2011	262.737	X
2	07/23/2018	-15.765	X	03/22/2018	-7.005		06/08/2016	262.909	
3	07/25/2018	-15.682		03/26/2018	-7.008		08/01/2018	262.915	
4	08/22/2018	-15.685							
5	08/23/2018	-15.700							
Mean h:		-15.7016	-15.6857		-7.0343	-7.0065		262.8537	292.9120

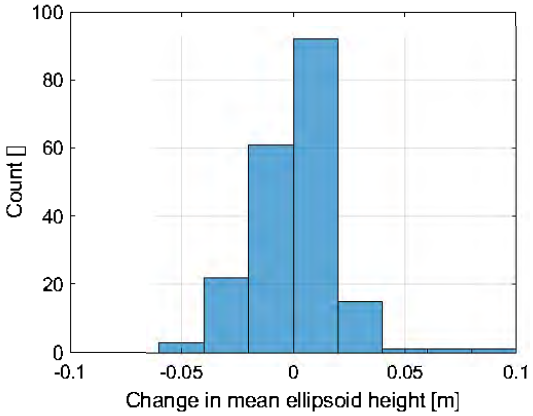

Figure 11: Change in the mean ellipsoid height caused by outlier detection with the OPUS Share dataset

Table 6 shows an example of the documentation that was utilized to evaluate the OPUS Shared GPS on Bench Mark residuals. Notice that both the pre-model residuals and post-model residuals (in yellow) are shown in this table. In the initial evaluation, only the pre-model residuals would be available to assist in the evaluation, then after modeling, the post-model residuals would provide final confirmation on how well the model performs at any particular bench mark.

Table 6: OPUS Share information used in analysis

PID	Pre-Modeled Residual v8.2.1 (cm)	Post-Modeled Residual v 8.2.1 (cm)	Designation	State	NAD 83 (2011) Ellipsoid Height (m)	NAVD 88 Orthometric Height (m)	Use (blank) or Do_Not_Use (X)	GNSS Information - OPUS Share Ellipsoid Heights (m)	Leveling Information	Comments
DO4461	1.1	-0.4	OPUS:0790029	IL	86.669	115.483		(1) 86.646 (2) 86.692	2012 - 2/1	Okay to use - Good repeat OPUS Shared solutions (< 5cm), small pre-modeled residuals, and recent leveling data
PA0772	-5.0	4.5	OPUS:1 LRC	OR	1434.089	1453.943	X	(1) 1434.134 (2) 1434.108 (3) 1434.035	1970 - 2/0	Do not use - There are three OPUS shared solutions. Two OPUS Shared solutions agree within 3 cm but there's a large varation in third OPUS Shared Solution (approximately 10 cm).
CP0032	9.5	1.1	OPUS:1 V 13 MSHD	MS	4.959	31.397		(1) 4.963 (2) 4.955	1978 - 2/0	Okay to use - Good repeat OPUS Shared solutions (< 1cm); large post-modeled residual but consistent with neighbors
DG9172	8.2	0.5	OPUS:14416	MI	195.322	228.945		(1) 195.316 (2) 195.327	2003 - 1/2	Okay to use - Good repeat OPUS Shared solutions (< 1cm) and recent leveling; large pre-modeled residual but consistent with neighbors

5.2.5 Example of an Accepted GPS on Bench Mark from OPUS Share

The following situation includes an accepted GPS on Bench Mark from OPUS Share for station 0790029 (PID: DO4461). The relevant attributes for this station are shown in Table 5 and the OPUS Share solutions can be obtained online here: DO4461 OPUS Share Solution from 2015 and DO4461 OPUS Share Solution from 2018. This station was included in the development of the GEOID18 due to the consistency in pre-model residuals between this mark and surrounding marks as shown in Figure 12. The geographic consistency together with two OPUS Share solutions that agree to within 5 cm (see Table 7) and recent leveling surveys to determine the NAVD 88 orthometric height all provide evidence and confidence that the observed height values associated with this bench mark are correct and can be used in the development of GEOID18.

Table 7: OPUS Shared ellipsoid heights for PID: DO4461

Shared solution:	Observation date:	NAD83(2011) Ellipsoid height [m]:
1	05/20/2015	86.692
2	07/09/2018	86.646

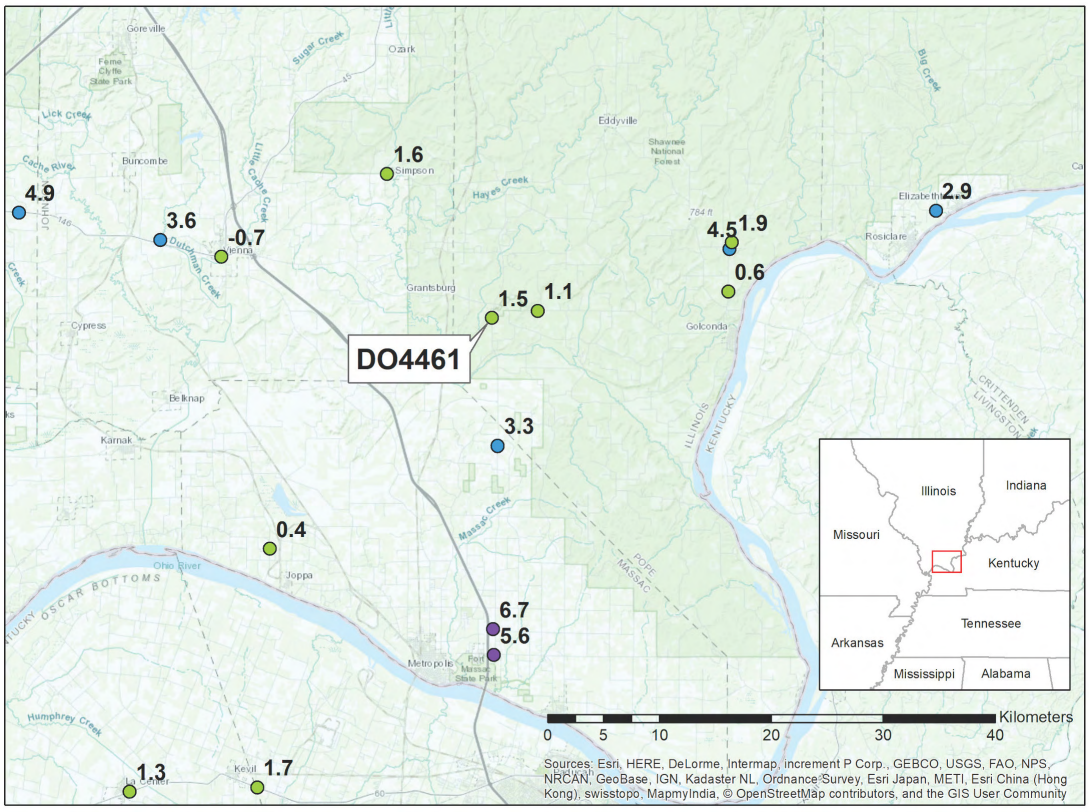

Figure 12: Pre-model residuals in [cm] in the vicinity of PID = DO4461. This mark was used in development of GEOID18.

5.2.6 Example of a Rejected GPS on Bench Mark – PID: PA0772

The following section presents an example of a rejected GPS on bench mark station: 1 LRC (PID: PA0772) in Oregon. This station was not included in the development of the hybrid geoid model because the pre-model residual was inconsistent with neighboring bench marks as illustrated in Figure 13. PA0772 has a pre-model residual of -4.1 cm while the surrounding marks are in the -8 to -10 cm range. There are three OPUS Share solutions on PA0772 (OPUS Share solution 1, OPUS Share solution 2, and OPUS Share solution 3) that are all used to arrive at the mean ellipsoid height equal to 1434.089 m as shown in Table 8.

In this case, the outlier filter does not remove any of the solutions; however, upon meticulous inspection, it might appear that the solution from 2014 is questionable as it is 8+ cm below the two solutions from 2018. Just south of this mark are two other bench marks (PIDs = PA0773 and PA0774) that show almost the same signature in their OPUS Share derived ellipsoid heights with solutions from 2014 being 8 to 10 cm lower than solutions from 2018 (see Table 8). However, on both of these two marks, the outlier filter removes the 2014 solutions and computes the mean ellipsoid height based only on the 2018 solutions. Additionally, the next two bench marks (PA0775 and PA0776) to the south also have solutions from 2014 and 2018, but these are mutually consistent with each other at ~cm level. It is beyond the current scope of this document to uncover the root cause of this 8 to 10 cm discrepancy in the marks from 2014 to 2018, but likely caused by either natural uplifting processes or errors in the 2014 OPUS Share solutions.

The overall result in this area is that GEOID18 is consistent with published NAVD 88 elevations and the 2018 ellipsoid heights at the 1 to 2 cm level. This type of meticulous analysis and understanding would not be possible without the numerous surveys and submissions done by the Oregon Department of Transportation in this area.

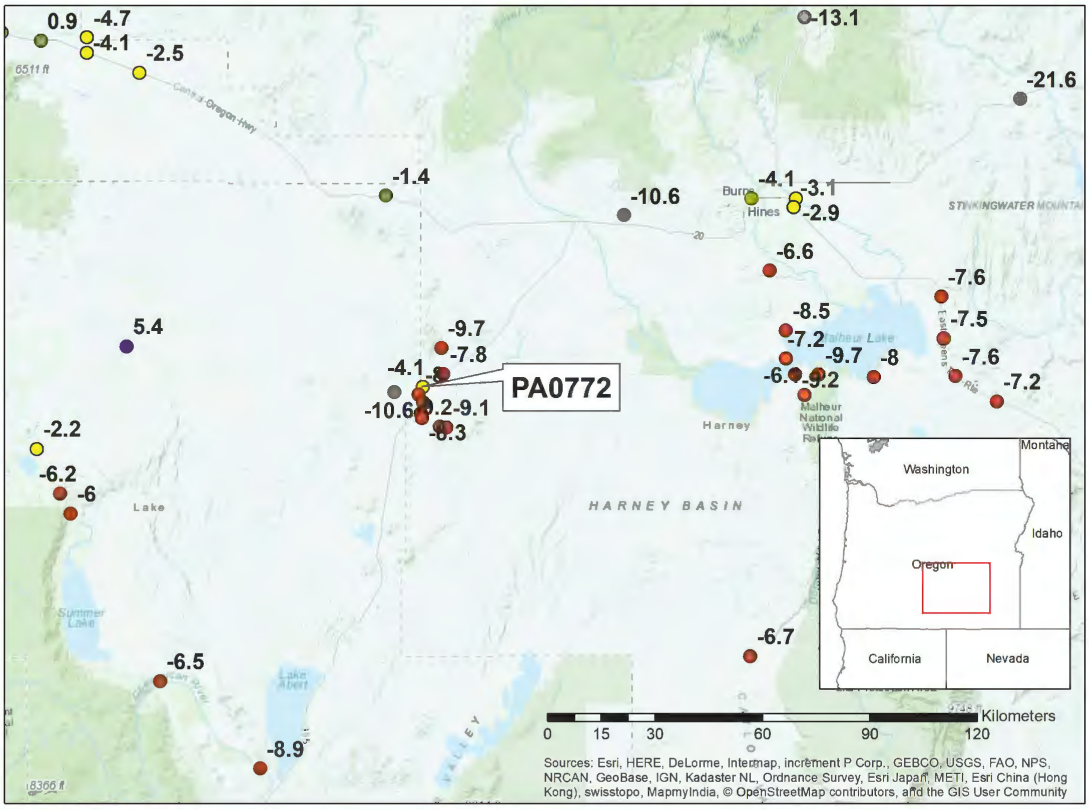

Figure 13: Pre-model residuals in [cm] in the vicinity of PID: PA0772. PA0772 has a pre-model residual of -4.1 cm while the surrounding marks are in the -8 to -10 cm range. This mark was not used in the development of GEOID18 due to this inconsistency.

Table 8: OPUS Share Solution ellipsoid heights on PID: PA0772 and surrounding bench marks

	PA0772		PA0773		PA0774		PA0775		PA0776	
Shared Solution:	Observation date:	Ellipsoid height:	Observation date:	Ellipsoid height:	Observation date:	Ellipsoid height:	Observation date:	Ellipsoid height:	Observation date:	Ellipsoid height:
1	05/20/2014	1434.025	05/20/2014	*1421.256**	05/15/2014	*1422.319**	05/15/2014	1412.752	05/19/204	1464.178
2	07/25/2018	1434.108	07/25/2018	1421.360	07/25/2018	1422.409	07/24/2018	1412.740	07/24/2018	1464.188
3	09/17/2018	1434.134	09/17/2018	1421.365	09/17/2018	1422.423				
Raw Mean Ellipsoid Height		1434.089		1421.327		1422.384		1412.746		1464.183
Final Ellipsoid Height		1434.089		1421.363		1422.416		1412.746		1464.183

* Flagged as outlier by (11) and removed from solution set.

5.2.7 Example of a Rejected GPS on Bench Mark – PID: AF9779

The following is an example of two closely spaced stations in the state of Washington that were rejected in GEOID18. The stations have very large pre-model (and post-model) residuals, and there were a number of nearby stations with consistent residuals, which is illustrated in Figure 14. Additional metadata about this mark can be found in Appendix III. AF9779 has a pre-model residual of 28+ cm and AF9780's pre-model residual is 14+ cm. The ellipsoid heights of the stations were from the IDB and determined using GPS data from 1998. These stations are both located within 2 to 4 km from several stations with small residuals, which allowed AF9770 and AF9780 to be removed.

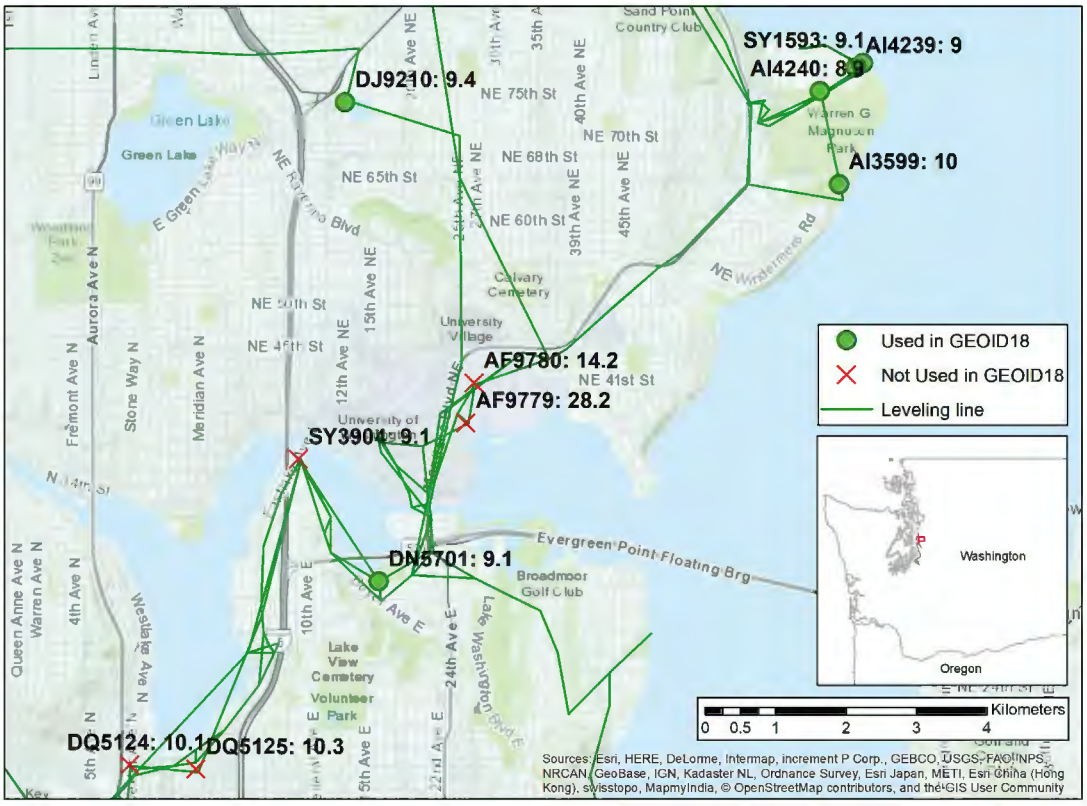

Figure 14: Pre-model residuals around PID: AF9779 in the state of Washington. Both AF9779 and AF9780 were not used in GEOID18 due to large pre-model residuals inconsistent with the neighboring marks. The other three PIDS (SY3904, DQ5124, and DQ5125) that are not used have consistent residuals but only have a single OPUS Share solution.

5.2.8 Example of a rejected GPS on Bench Mark – PID: HF0299

The following is an example of a station (PID: HF0299) in the state of Kansas that was rejected in GEOID18 for several reasons and is illustrated in Figure 15 (with additional metadata in Appendix III). The station has a large pre-model residual of -8.4 cm, which is 6 to 10 cm different than surrounding bench mark residuals. The likely cause of this discrepancy is that the orthometric height was established on a spur section of a very short, 2nd Order-Class 0 leveling line performed in 1962. Furthermore, this spur was tied to an even older 2nd Order-Class 0 line from 1934. The inconsistent pre-model residual along with the spatial distribution of surrounding marks caused this PID to be not used in GEOID18.

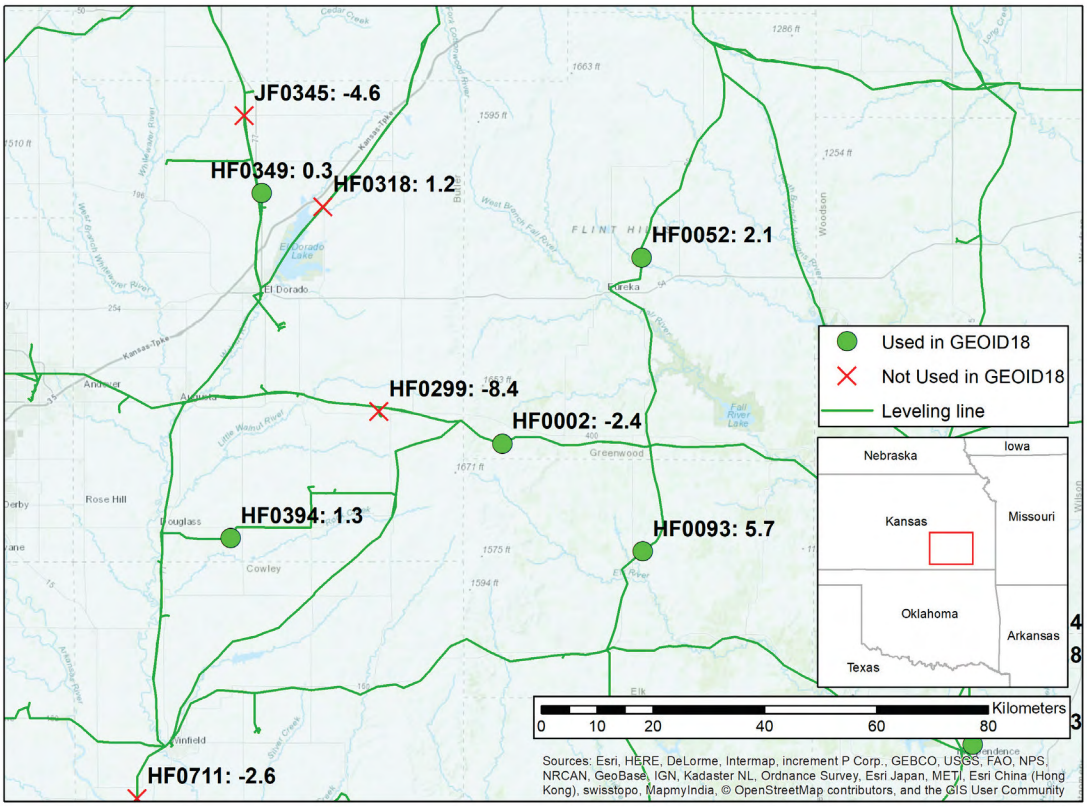

Figure 15: Pre-model residuals around PID: HF0299 in the state of Kansas. This bench mark is not used in GEOID18 due to its residual being inconsistent with neighboring values.

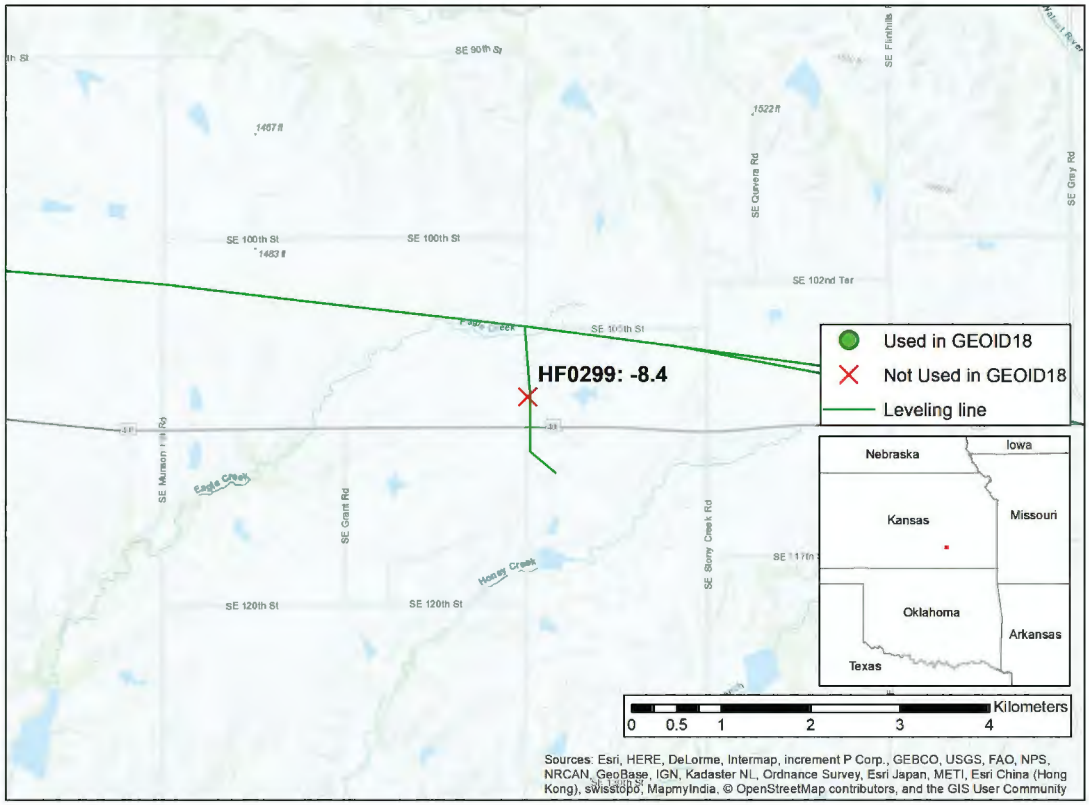

Figure 16: Pre-model residual for PID: HF0299 with spur leveling line. Units: [cm]

5.2.9 Example of a Rejected GPS on Bench Mark – PID: EK0599

The following is an example of a station (PID: EK0599) in the state of Oklahoma that was rejected in GEOID18 and illustrated in Figure 17 (with additional metadata in Appendix III). Most importantly, the pre-model residual is -25 cm, which is inconsistent with neighboring marks at the 20 cm level. First, the station's orthometric height was established in 1934 using 2nd Order-Class 0 leveling data. Secondly, the ellipsoid height was established using older GPS from 2001. Finally, the station is labeled as a "No Check" station, because its coordinates were estimated based on single GPS vector (see excerpt from the NGS Datasheet in Figure 18 below). It should be noted that a large void area (approximately 50 km) was created by rejecting this station, but the residual was simply too large to be used in the model.

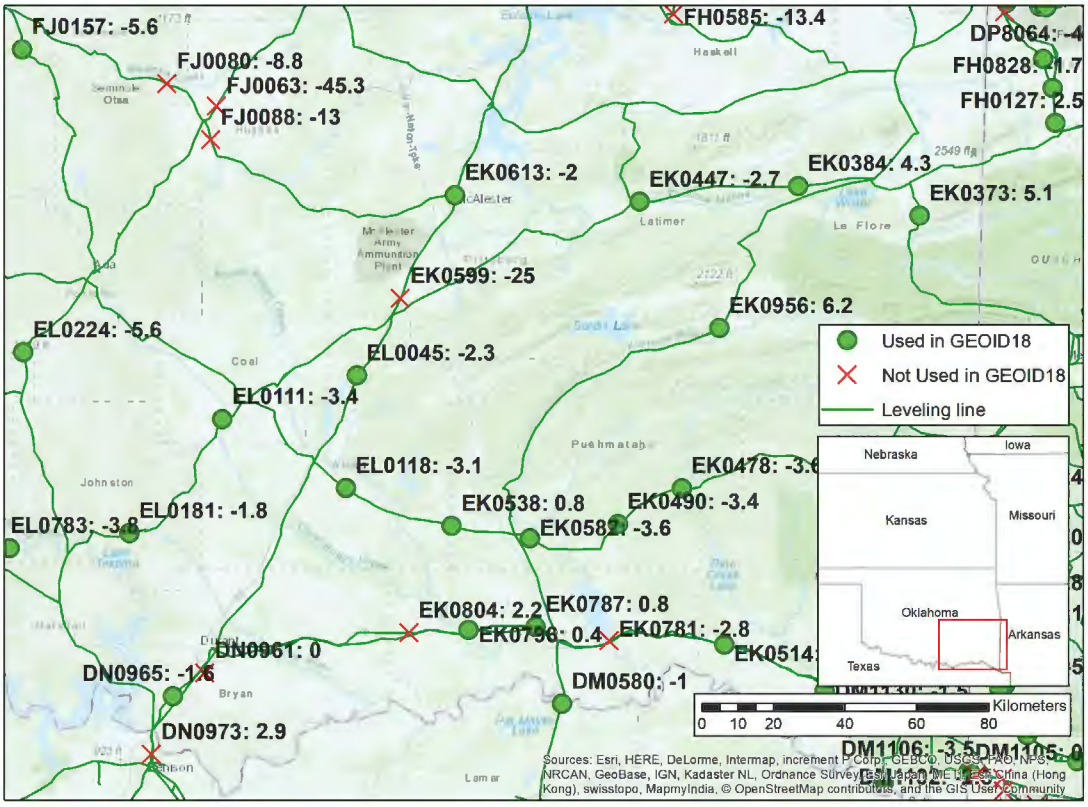

Figure 17: Pre-model residuals around PID: EK0599. Units = [cm]

The NGS Data Sheet

See file dsdata.pdf for more information about the datasheet.

```
PROGRAM = datasheet95, VERSION = 8.12.5.4
1        National Geodetic Survey,    Retrieval Date = OCTOBER 20, 2019
 EK0599 ***********************************************************************
 EK0599  DESIGNATION -  F 16
 EK0599  PID         -  EK0599
 EK0599  STATE/COUNTY-  OK/PITTSBURG
 EK0599  COUNTRY     -  US
 EK0599  USGS QUAD   -  KIOWA (1978)
 EK0599
 EK0599                         *CURRENT SURVEY CONTROL
 EK0599
 EK0599* NAD 83(2011) POSITION- 34 43 09.57817(N) 095 54 27.02957(W)   NO CHECK
 EK0599* NAD 83(2011) ELLIP HT-   197.577 (meters)        (06/27/12)   NO CHECK
 EK0599* NAD 83(2011) EPOCH    -  2010.00
 EK0599* NAVD 88 ORTHO HEIGHT -   226.860 (meters)      744.29  (feet) ADJUSTED
 EK0599
 EK0599  GEOID HEIGHT    -        -29.515 (meters)                     GEOID18
```

Figure 18: NGS Datasheet for PID: EK0599, which highlights the 'NO CHECK' status of the geometric coordinates of this mark.

5.2.10 Example of a GPS on Bench Mark That Uses OPUS Share to Supersede the IDB Value – PID: DH0882

The following section illustrates an example of a bench mark that has an ellipsoid height from both database sources, and the OPUS Share results were ultimately used in GEOID18. This is an overall extremely rare occurrence in GEOID18 with only 14 bench marks having OPUS Share ellipsoid heights that effectively supersede their IDB corresponding heights. This was only done when there was overwhelming evidence that the IDB ellipsoid height was questionable with respect to other bench

marks in GEOID18 and the OPUS Share solution(s) provided a clear, more consistent fit with neighboring bench marks.

In Figure 19a, it is pretty evident that the pre-model residual derived from the NGS IDB ellipsoid height does not fit well with neighboring bench marks. It is likely in disagreement at the 8 to 10 cm level (8 cm to the northerly and westerly bench marks and 10 cm to the southeasterly bench mark). Upon substitution of the OPUS Share ellipsoid height, the residual is much more consistent at the 1 to 3 cm-level with neighboring bench marks (see Figure 19b). The OPUS Share derived ellipsoid height (an average of the two solutions shown in Table 9) was ultimately used in GEOID18.

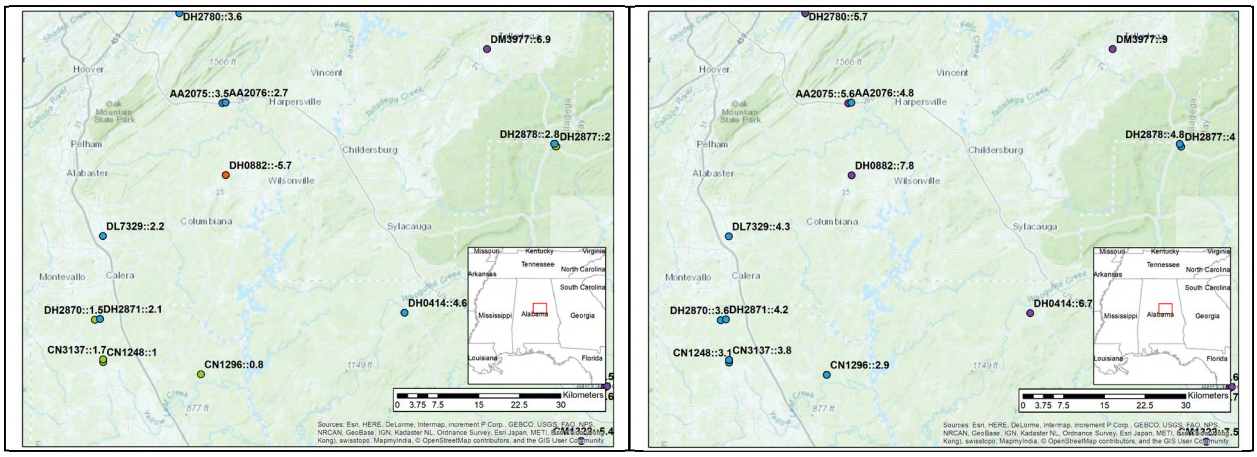

Figure 19: Surrounding pre-model residuals of PID: DH0882. Units: [cm]. **a** (left): DH0882 ellipsoid height from NGS IDB. **b** (right): DH0882 ellipsoid height from OPUS Share.

Table 9: Heights from NGS IDB and OPUS Share for PID: DH0882

Data Source:	NAD83(2011) Ellipsoid height:	Date:	NAVD 88 Orthometric height:	Comments:
IDB	112.734	07/11/1990	141.786	Derived from 2 occupations, each with 2 hours of duration
OPUS Share Solution 1	112.640	04/09/2018		OPUS Shared Solution
OPUS Share Solution 2	112.598	09/17/2018		OPUS Shared Solution

5.2.11 Selection of GPS on Bench Marks along the Louisiana and Texas Gulf Coast

There are areas of apparent crustal movement in the Gulf Coast region of the United States such as from the southeastern Texas region eastward through Louisiana to the Mississippi border. Due to apparent crustal movement, many control station elevations in the region are obsolete. The selection of GPS on Bench Marks stations used in the development of GEOID18 in this region was kept to a minimum to limit the influence of crustal movement in the hybrid geoid model. The GPS on Bench Marks residuals were analyzed based on a number of different preliminary hybrid geoid models using various scenarios of

constraints to determine the best set of GPS on Bench Marks stations to be used in GEOID18 along the Texas/Louisiana Gulf Coast.

The selection of stations used and not used in GEOID18 in the southeastern Texas region eastward to the Mississippi border is shown in Figure 20, which highlights the sparseness of the data used in GEOID18 in this region.

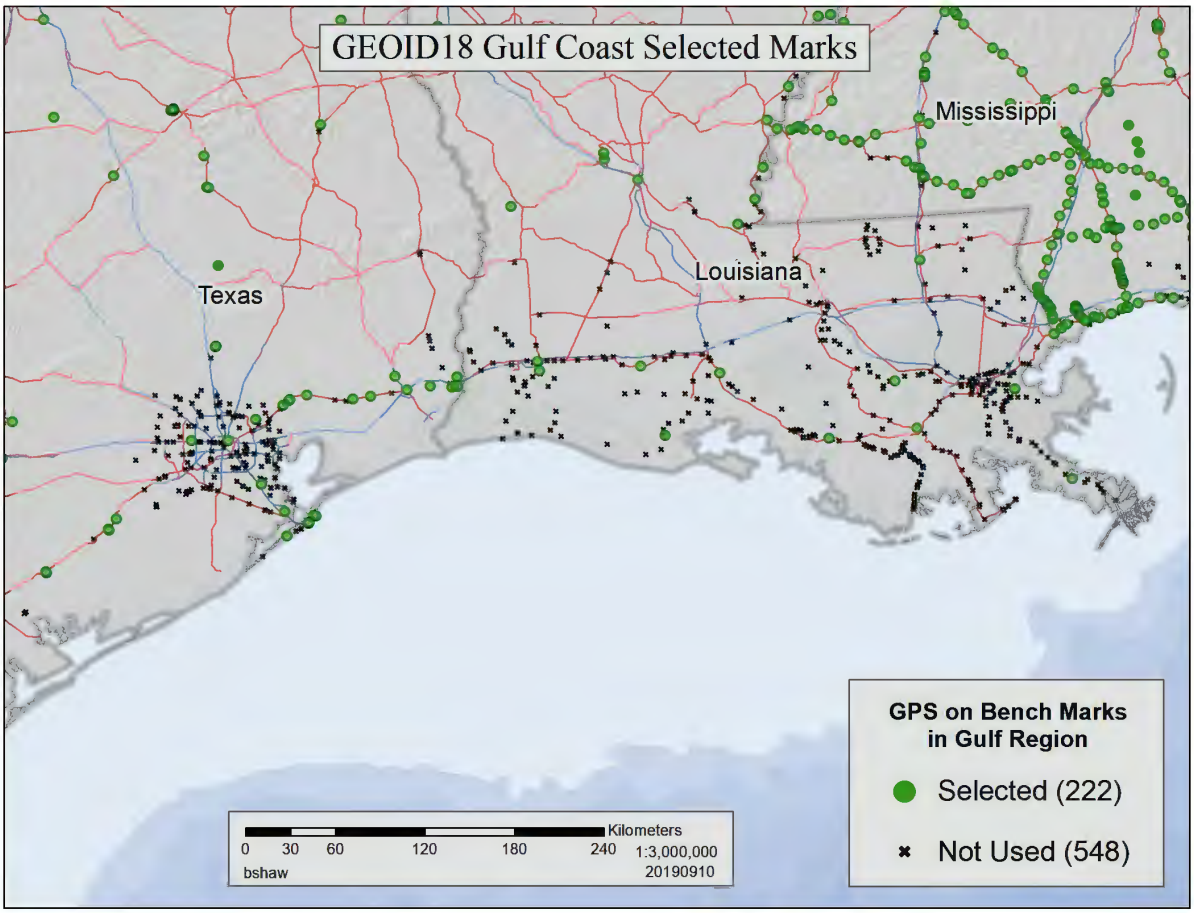

Figure 20: GEOID18 Gulf Coast Selected Marks. Selected marks are those used in the development of GEOID18.

5.3 Least Squares Collocation

The following section describes how the two previously discussed input datasets (xGEOID19B and the GPS on Bench Marks) are combined in a prescribed method, which extracts the positive attributes from each input and blends them into a hybrid geoid model. This method is essentially a 'warping' of the gravimetric geoid model to the GPS on Bench Marks, which is performed using multi-matrix least squares collocation (Roman, et al. 2004). The covariance function from (4) and (5) is altered slightly in this model from that used in GEOID12B due to the sheer number of new bench marks and the overall better alignment with the gravimetric geoid model due to a very thorough and meticulous analysis of the residuals on a mark-by-mark basis. The covariance function is illustrated in Figure 21 where individual parameters are also defined in Table 10. Additionally, the covariance function parameters for PRVI are shown in Table 11. The modeling for PRVI is much more sensitive to changes in the parameters

compared to the CONUS modeling. This is primarily due to the number of GPS on Bench Marks and their correlation that support the model. In PRVI, there are 127 GPS on Bench Marks used that result in 8001 combinations or correlations. This causes the PRVI model to be based on a much smaller number of empirical covariances (i.e. observations) compared with the CONUS model. For example, PRVI has only 231 combinations over the 0 to 5 km interval compared with 97,000+ combinations for CONUS.

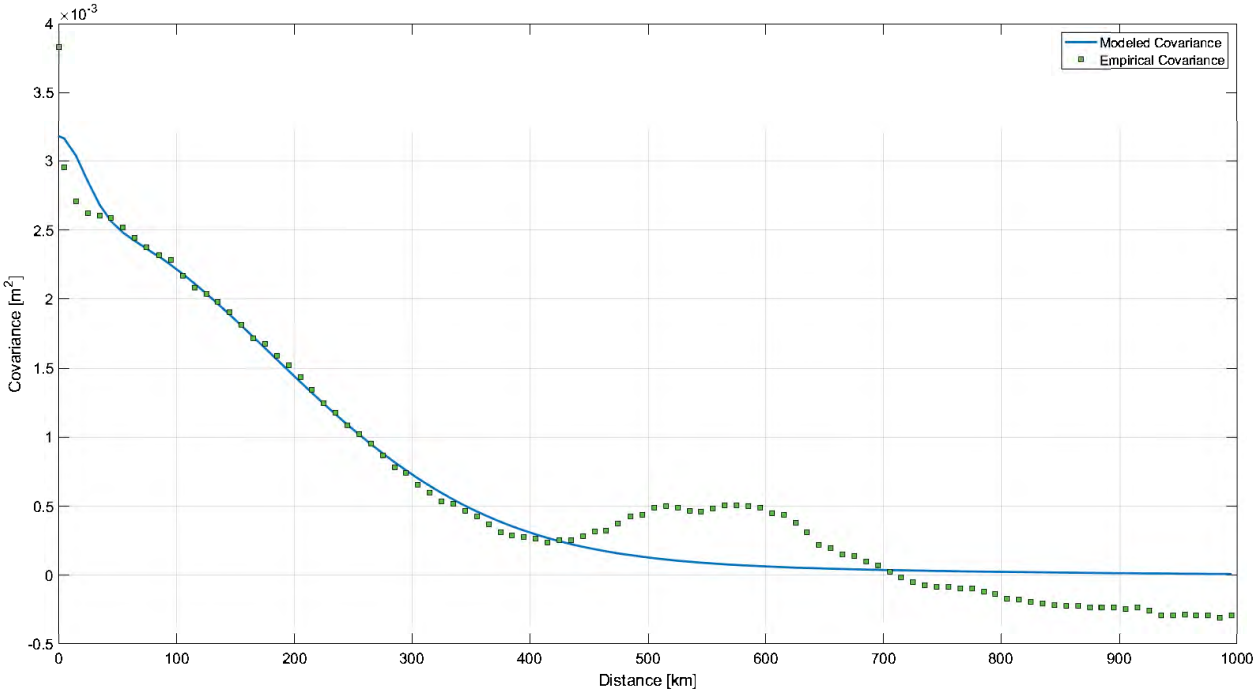

Figure 21: Covariance function and empirical residual data at 10 km intervals from 0 to 1000 km.

Table 10: Covariance function parameters used for CONUS

	$\sqrt{A_0}$ [m]	L [km]
$C_{l_1 l_1}$	0.01193	600
$C_{l_2 l_2}$	0.04814	260
$C_{l_3 l_3}$	0.01023	180
$C_{l_4 l_4}$	0.00100	90
$C_{l_5 l_5}$	0.00423	60
$C_{l_6 l_6}$	0.02446	30

Table 11: Covariance function parameters used for PRVI:

	$\sqrt{A_0}$ [m]	L [km]
$C_{l_1 l_1}$	0.0010	60
$C_{l_2 l_2}$	0.0345	10

Additional terms are not necessary as the maximum bench mark spacing in this region is much smaller, and there are not an adequate number of bench marks to support more terms.

For the final component needed in (6), GEOID18 uses a noise value of 2 cm for all the GPS on Bench Marks in C_{nn}. While this is probably too scientifically simple, it actually is quite appropriate in an operational sense. For example, preliminary tests show that when the noise value was increased to 5 cm on all the bench marks within a particular state, the fit is less constrained and the model does not fit the

bench marks as well. While this is expected and scientifically more appropriate, it is not what NGS is trying to achieve with a hybrid geoid model. The goal of GEOID18 is to fit high-quality GPS on Bench Marks as closely as possible, which is achieved most practically with 2 cm noise values on all the residuals.

5.3.1 Error Model Using Covariance Function and 'Jackknifing':

One technique that has proven very useful for improving the overall geoid model quality is a resampling technique known as jackknifing (Quenouille, 1949) or sometimes referred to as "leave-one-out." In this technique, each residual in the pre-model residuals ($r_{pre\text{-}model,i}$) is removed from the model, a model is built using all residuals except for $r_{pre\text{-}model,i}$, and then the predicted value can be compared to the observed residual. This is then repeated for all residuals resulting in n = 32,357 individual geoid models for CONUS and a set of misfit values (jackknife prediction — observed) that are very useful in practice. These misfits are illustrated for CONUS in Figure 22. The statistics associated with these misfits are also shown in Table 12. While this type of statistic is often too optimistic (Brown, et al. 2018), we find that it provides a lot of value to users of a hybrid geoid model to understand how good (or how poorly) an individual bench mark fits with the model.

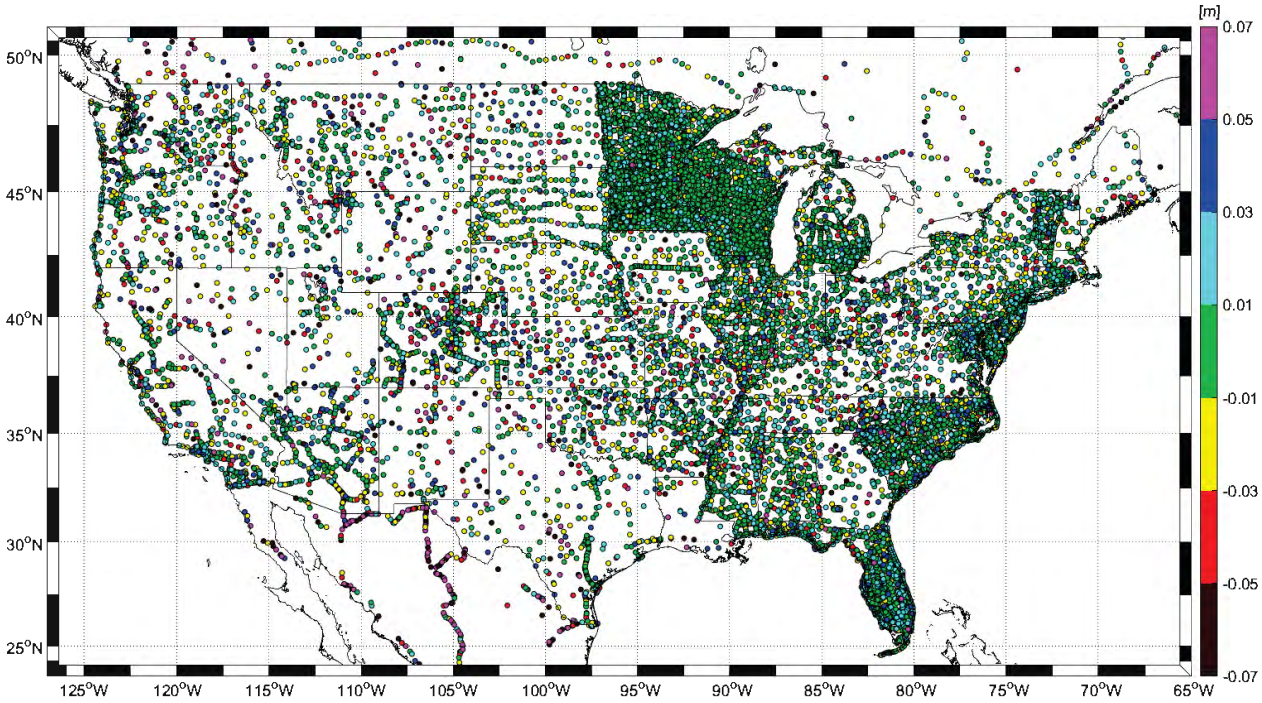

Figure 22: Results from jackknifing over CONUS where each GPS on Bench Mark is evaluated by a temporary geoid model which does not contain that bench mark. This shows the misfit between the jackknife prediction and the observed residual.

The distribution of the misfit residuals for CONUS is shown in Figure 23. In addition to the actual distribution, a normal distribution is also shown, which highlights the optimistic tendency of this method. In Figure 23a, it is evident that the misfits are much more centered than a normal distribution and exhibit longer tails. In Figure 23b, a normal probability is shown, which compares the misfit distribution to a normal distribution. If the data is normally distributed, it should align with the red curve. It is evident that this dataset does not follow the normal distribution, with only the interior

approximately 80% (10% to 90%) aligning with a normal distribution. This is confirmed by hypothesis testing where a Lillefors test (Lillefors, 1967) was performed, which tests the null hypothesis that the misfit distribution comes from a normal distribution. The result of this hypothesis test is rejection of the null hypothesis at the 5% significance level, signifying the data does not follow a normal distribution.

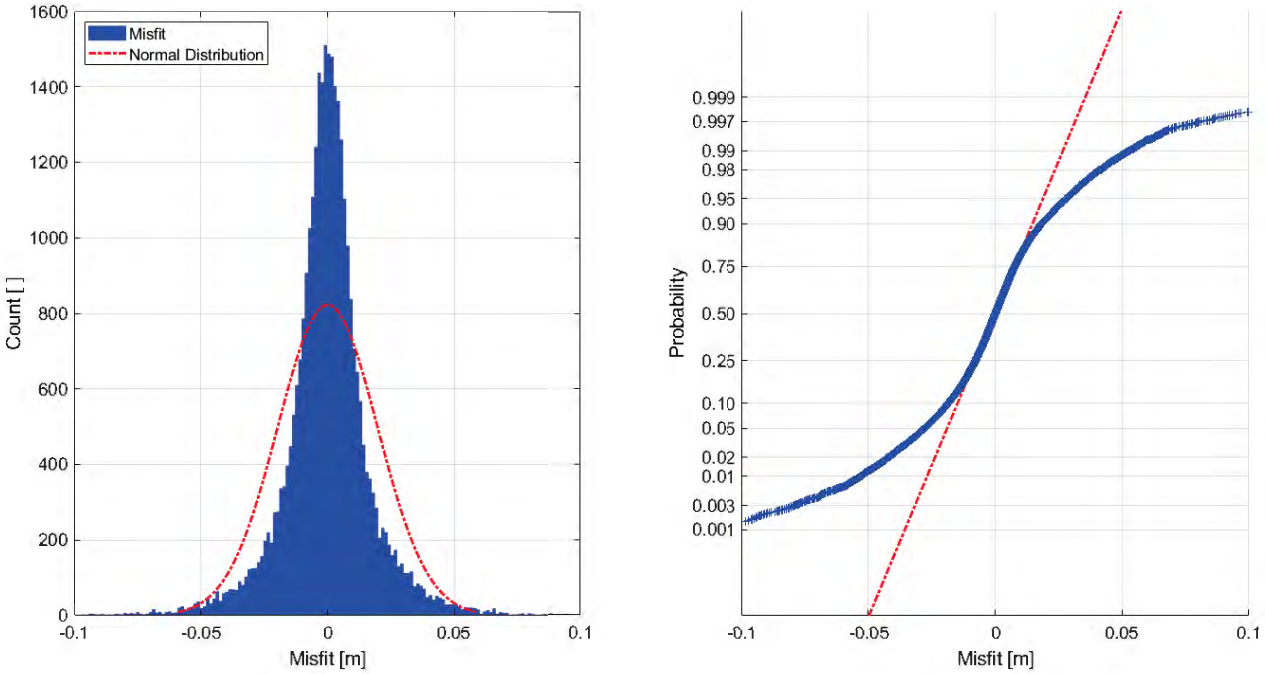

Figure 23: Misfits in CONUS compared to a normal distribution. **a** (left): histogram of misfit residuals with a normal distribution. **b** (right): normal probability of the misfit residuals where the empirical data will align with the red curve if it follows a normal distribution. In this situation, the misfit data do not follow the normal distribution.

A similar process was done for Puerto Rico and the U.S. Virgin Islands; however, due to the very small number of points (127), it is difficult to draw meaningful conclusions for this region. The geographical distribution of the misfits is shown in Figure 24 along with the data distribution compared with a normal distribution in Figure 25. The Lillefors hypothesis test fails to reject the null hypothesis at the 5% significance level, signifying the data follows a normal distribution. This is confirmed visually with Figure 25b.

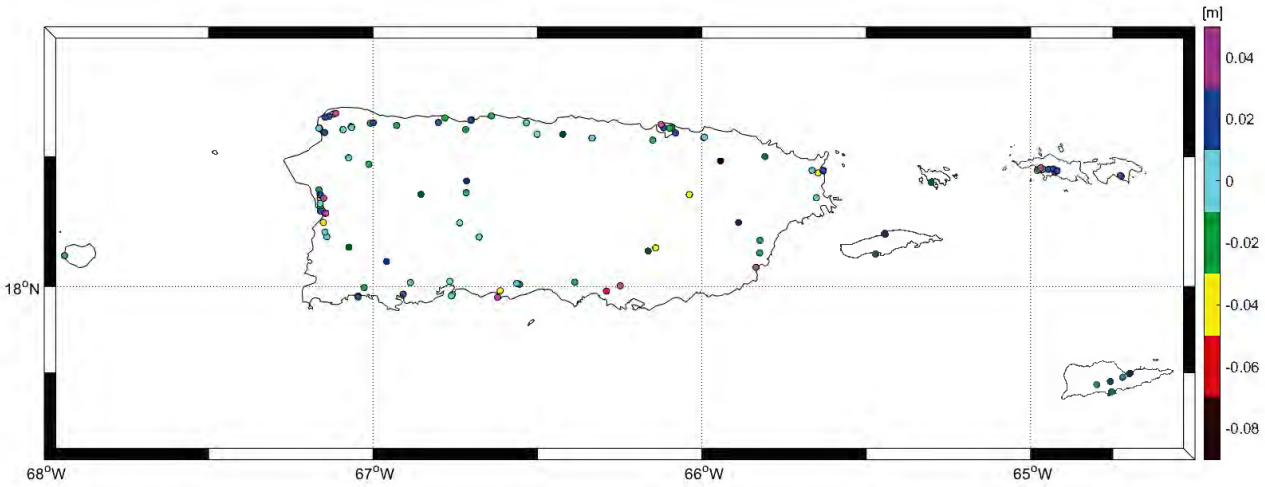

Figure 24: Results from jackknifing over PRVI where each GPS on Bench Mark is evaluated by a temporary geoid model which does not contain that bench mark. This shows the misfit between the jackknife prediction and the observed residual.

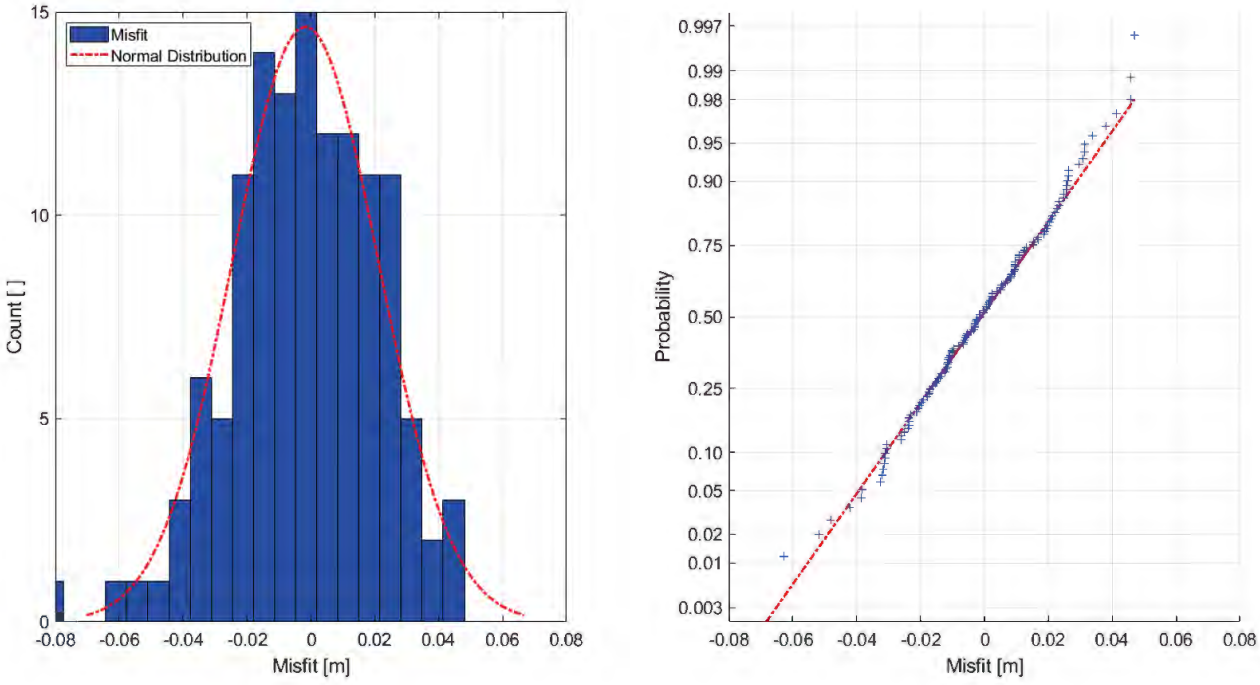

Figure 25: Actual distribution of misfits compared to a normal distribution for PRVI. **a** (left): histogram of misfit residuals with a normal distribution. **b** (right): normal probability of the misfit residuals where the empirical data will align with the red curve if it follows a normal distribution. These results appear to be normally distributed, though only 127 points are used.

Table 12: Statistics of the misfits from jackknifing for CONUS and PRVI. Units: [cm]

	Min.	Max.	Mean	StdDev.	16%	25%	75%	84%
CONUS	-32.82	29.84	0.00	1.97	-1.27	-0.78	0.77	1.25
PRVI	-8.28	4.69	-0.17	2.29	-2.37	-1.68	1.47	2.22

The misfit statistic from the jackknifing procedure is quite useful in a practical sense but also has limitations. The practicality of these misfits can be illustrated in the following example. The misfit for a station (PID: MG0388) in Muscatine County, Iowa is quite large at 8.4 cm compared to the surrounding marks as shown in Figure 26. Since the jackknife misfits can only be computed at bench marks used in the model, this signifies that the geoid model possibly has some unknown level of error in this region. This also signifies that something about this particular bench mark might be suspect (ellipsoid height, orthometric height, or the monument itself). However, this mark and geoid model could be completely fine. The jackknife misfits simply give the user some evidence that there may be a concern in a particular localized area and addition survey care and redundancy should be exercised while using that particular bench mark or the hybrid geoid model in that particular area.

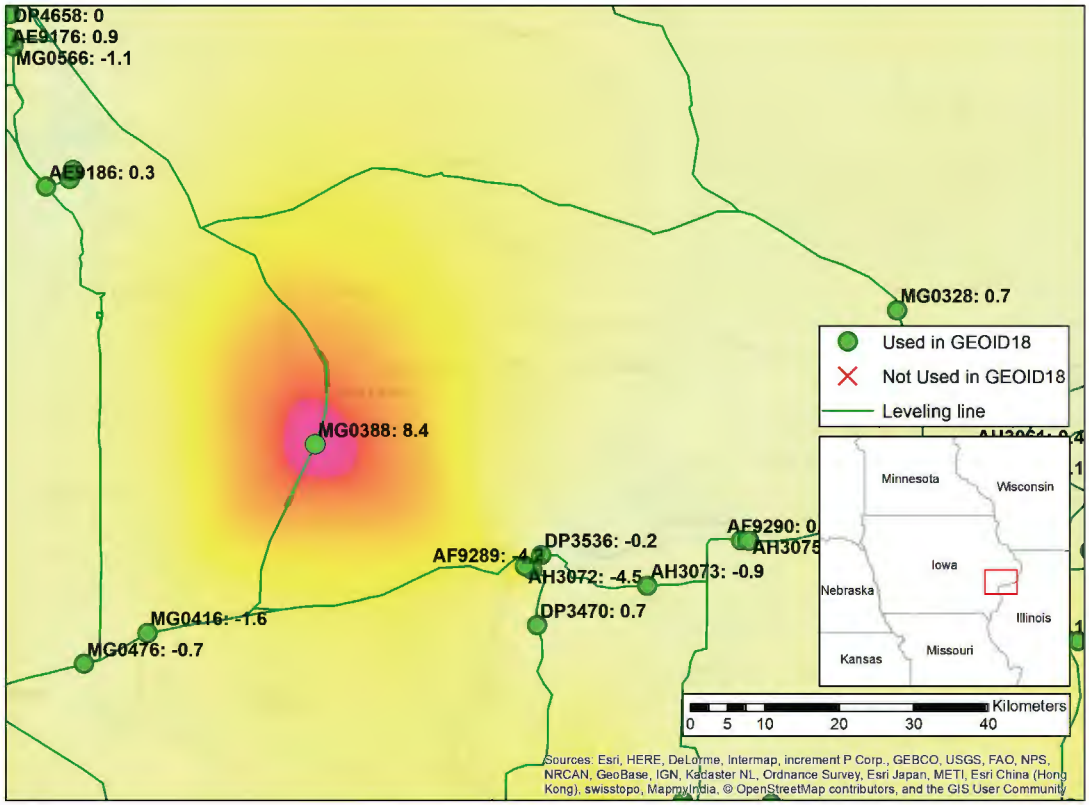

Figure 26: Misfit in [cm] computed at each used GPS on Bench Mark around MG0388 in eastern Iowa. The misfit represents how much discrepancy would exist if any individual mark was not used in the hybrid geoid model.

One might question why these marks are still included in the hybrid geoid model, if they have some questionable attributes. These questionable marks are in that middle ground where the residuals are not large enough to warrant removal from the model but they are still a bit inconsistent with the surrounding values. Additionally, they are typically in areas where very few bench marks exist, so providing something is better than nothing. In the previous example for MG0388, the surrounding pre-model residuals are shown in Figure 27 with the misfit highlighted in the background. The pre-model residual is 14.6 cm, which is consistent with a number of bench marks approximately 50 km to the northwest (approximately 1.3 cm different). Additional benchmarks approximately 30 km to the southwest and southeast are in the 3.7 to 7.3 range, which is the root cause of this inconsistency. This

mark is ultimately kept in the geoid modeling though since there are no additional bench marks within 30+ km.

This demonstrates that the model's quality is dependent on the quality AND density of GPS on BM observations; the density being critical in isolating and removing bad observations.

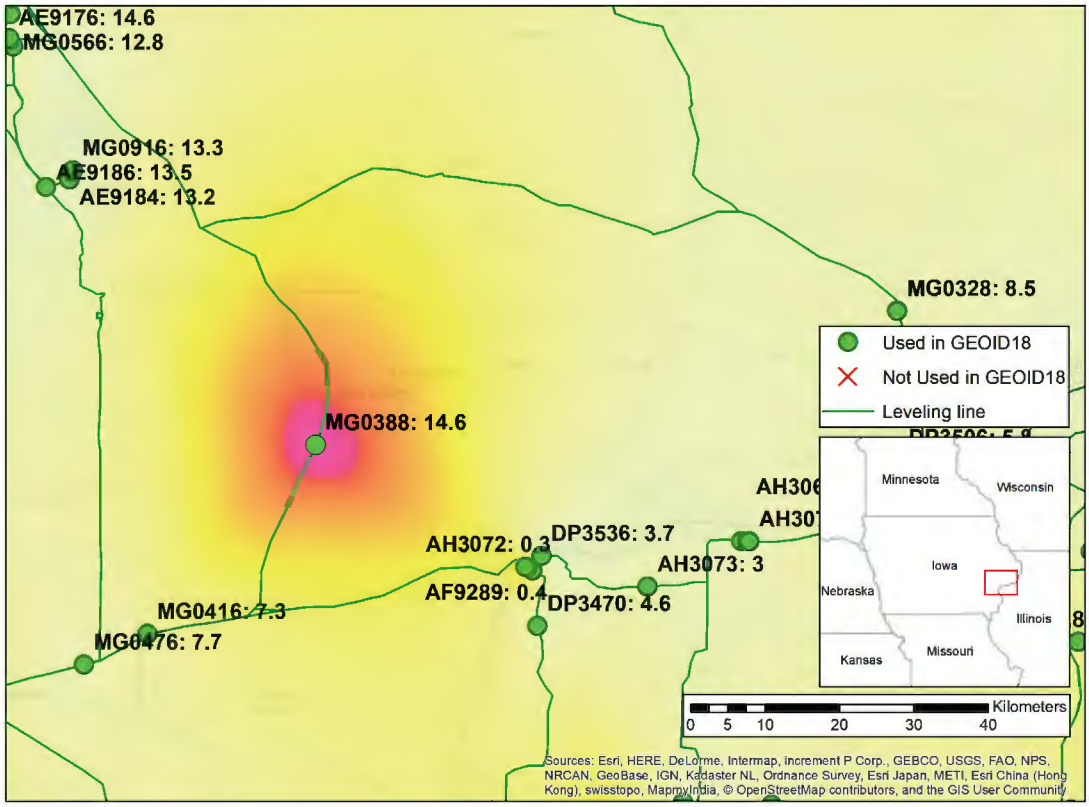

Figure 27: Pre-model residual in the area surrounding MG0388 in [cm]. This mark is used in GEOID18 due to the sparse surrounding bench marks and general agreement with marks 50 km to the northwest. The transparent surface reflects the interpolated misfits for easier visualization.

6 Results

6.1 Geoid Results

In this section, the results for GEOID18 will be shown from a broad perspective. This includes overall results for the raw residuals, pre-model residuals, and post-model residuals; the least squares collocation resultant grids; and the final GEOID18 model.

6.1.1 Raw Residuals

The initial set of raw residuals as computed in (1) for the full CONUS-area (including Canada and Mexico bench marks) and PRVI are illustrated in Figure 28 and Figure 29, respectively, along with their associated statistics in Table 13. The raw residuals are most significantly influenced by the continental tilt present in the vertical datum. For the CONUS-area, the raw residuals have what first appears to be a very large standard deviation of 29.4 cm; however, this is almost completely an artifact of the continental tilt in NAVD 88[1]. There is also a 56.2 cm bias present in the raw residuals, which is a mostly driven by the offset between the NAVD 88 datum and the W_0 value for xGEOID19B. This agrees with one estimate for the separation, which puts NAVD 88 55 ± 2 cm below the W_0 = 62 636 856.0 m^2/s^2 surface (Bursa, et al 2004).

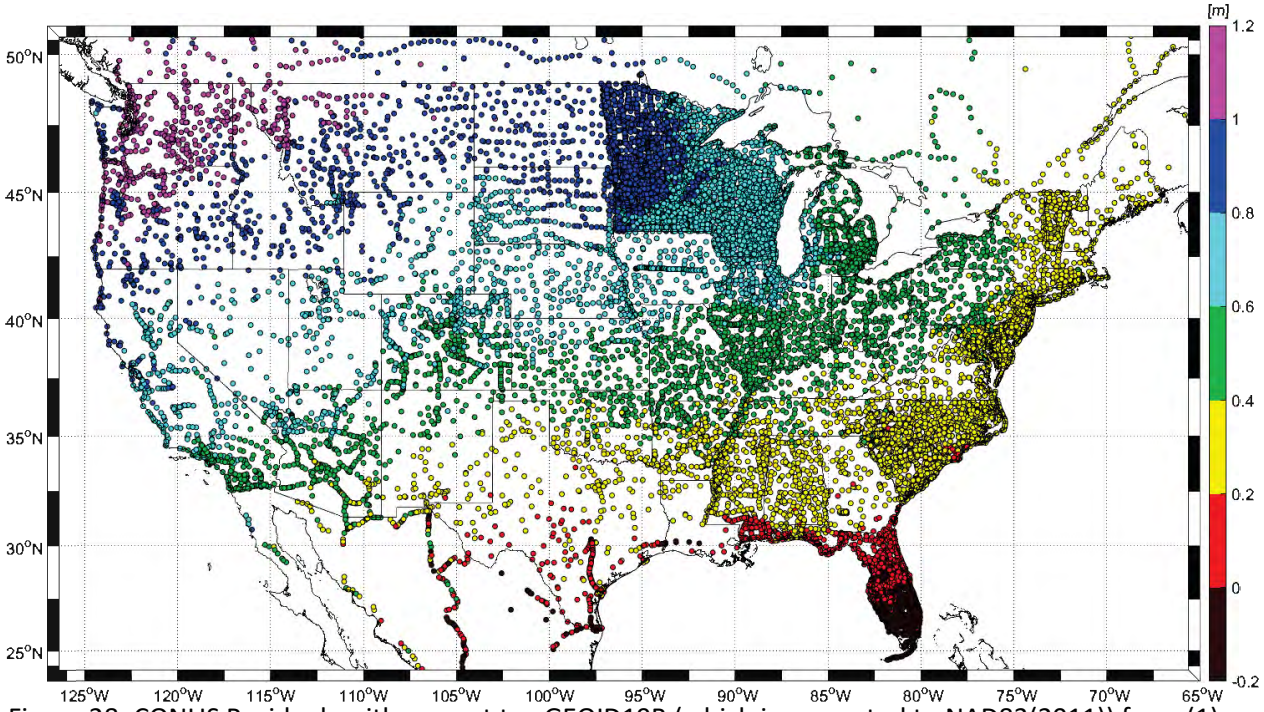

Figure 28: CONUS Residuals with respect to xGEOID19B (which is converted to NAD83(2011)) from (1)

[1] There is also a continental tilt in the ellipsoid heights between IGS08 and NAD83(2011) that is similar to the NAVD 88 tilt; however, this geometric tilt has been removed with the transformation of xGEOID19B to a NAD83(2011) version. If not removed, the combined tilt is even more severe and approximately 2.4 m across CONUS.

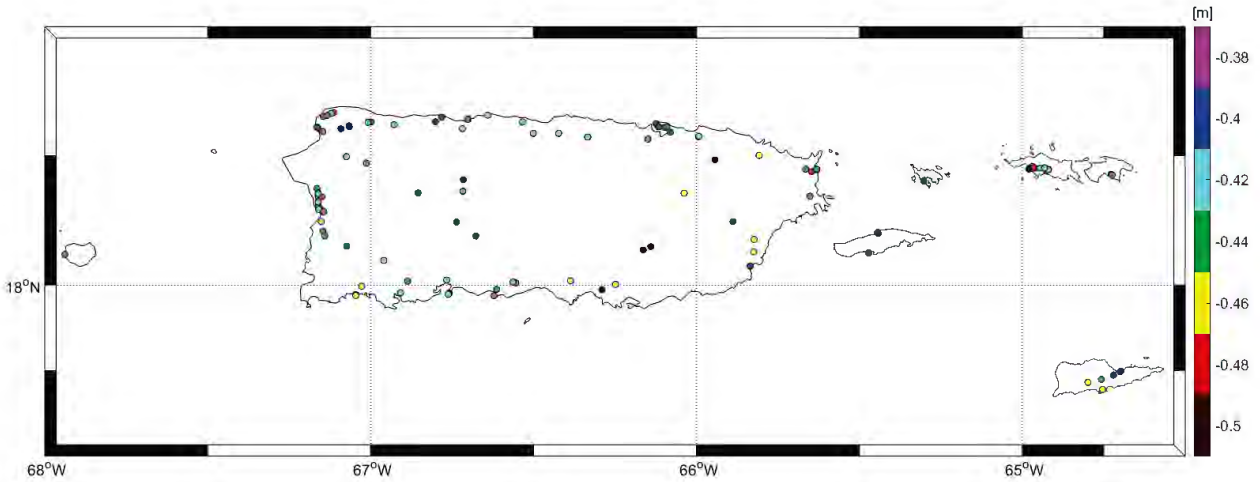

Figure 29: PRVI Residuals with respect to xGEOID19B (NAD83(2011) version) from (1)

Table 13: Statistics of residuals for CONUS and PRVI with xGEOID19B

	Number Residuals	Min. [cm]	Max. [cm]	Mean [cm]	Std. Dev. [cm]
CONUS, Canada, and Mexico	32,357	-26.37	127.50	56.22	29.43
CONUS-only	31,580	-18.05	123.72	56.05	29.13
PRVI	127	-51.90	-37.14	-42.74	2.99

6.1.2 Pre-model Residuals

In the following section, the residuals are still compared with xGEOID19B but the bias and tilt parameters are estimated from (2) and removed resulting in the pre-model results, which are the input needed for the LSC. The pre-model residuals still have some minor systematic effects present in a number of states along the periphery of the removed planar surface. States like Minnesota, Wisconsin, Michigan, New York and the New England region still have a systematic trend where the planar fit is not supported well enough. This is visually less evident in California, Arizona, and New Mexico but still present. The exact reason for this systematic boundary effect is not well understood at NGS presently.

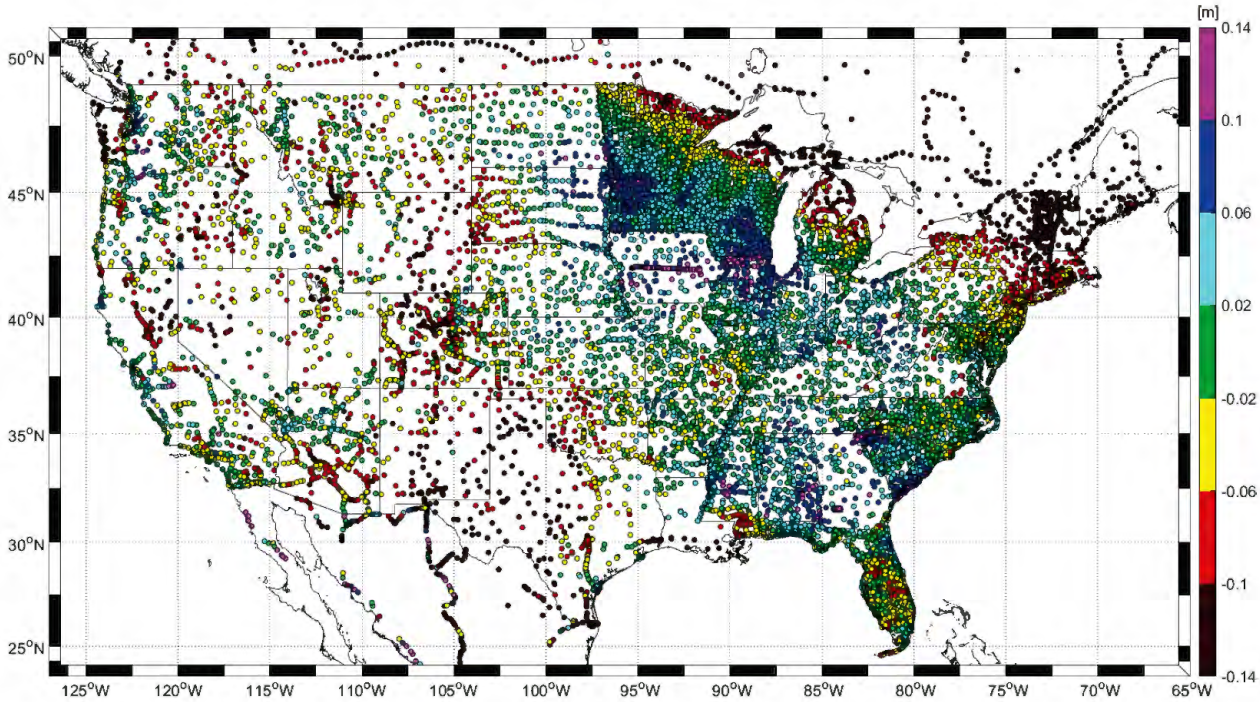

Figure 30: CONUS Pre-model Residuals — bias-free/tilt-free with respect to xGEOID19B from (3)

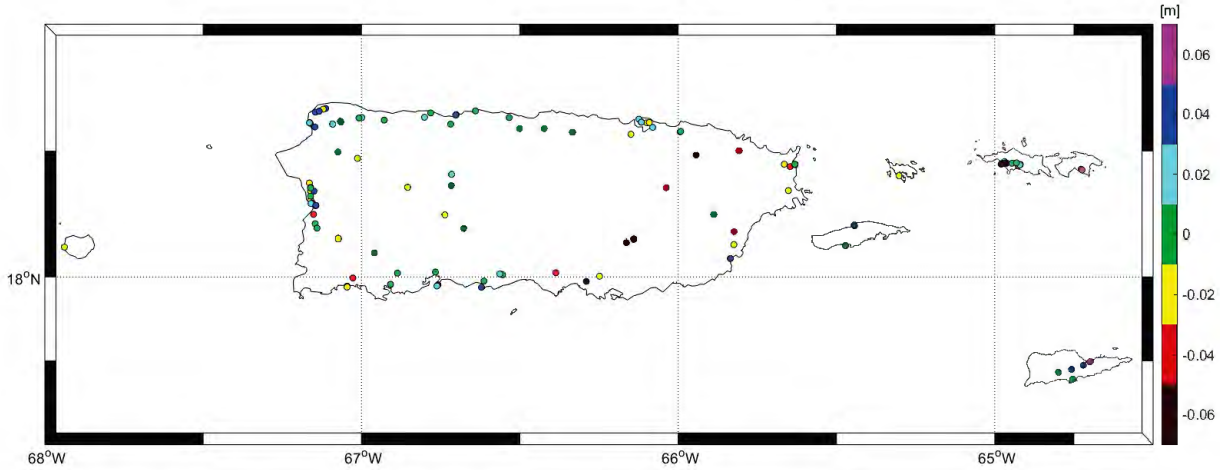

Figure 31: PRVI Pre-model Residuals — bias-free/tilt-free with respect to xGEOID19B from (3)

Table 14: Statistics of pre-model residuals for North America, CONUS, and PRVI with xGEOID19B

	Number Residuals	Min. [cm]	Max. [cm]	Mean [cm]	Std. Dev. [cm]
CONUS, Canada, and Mexico	32,357	-66.89	33.64	0.00	7.12
CONUS-only	31,580	-39.82	21.70	5.03	5.98
PRVI	127	-8.74	6.07	-0.00	2.86

6.1.3 Least Squares Collocation

In the LSC prediction phase, the pre-model residuals are used in (6) to estimate the warped component of the hybrid geoid model. The estimated warped surface is shown in Figure 32 and Figure 33 for CONUS and PRVI, respectively. This surface illustrates where there are localized differences in the vertical datum compared with the gravimetric geoid surface. In the next step, the tilted surface and bias are added back to the warped surface resulting in the conversion surface from the NAD83 (2011) morphed xGEOID19B[2] as shown in Figure 34 and Figure 35. Finally, the conversion surface is combined with xGEOID19B resulting in the GEOID18 model as shown in Figure 36 and Figure 37.

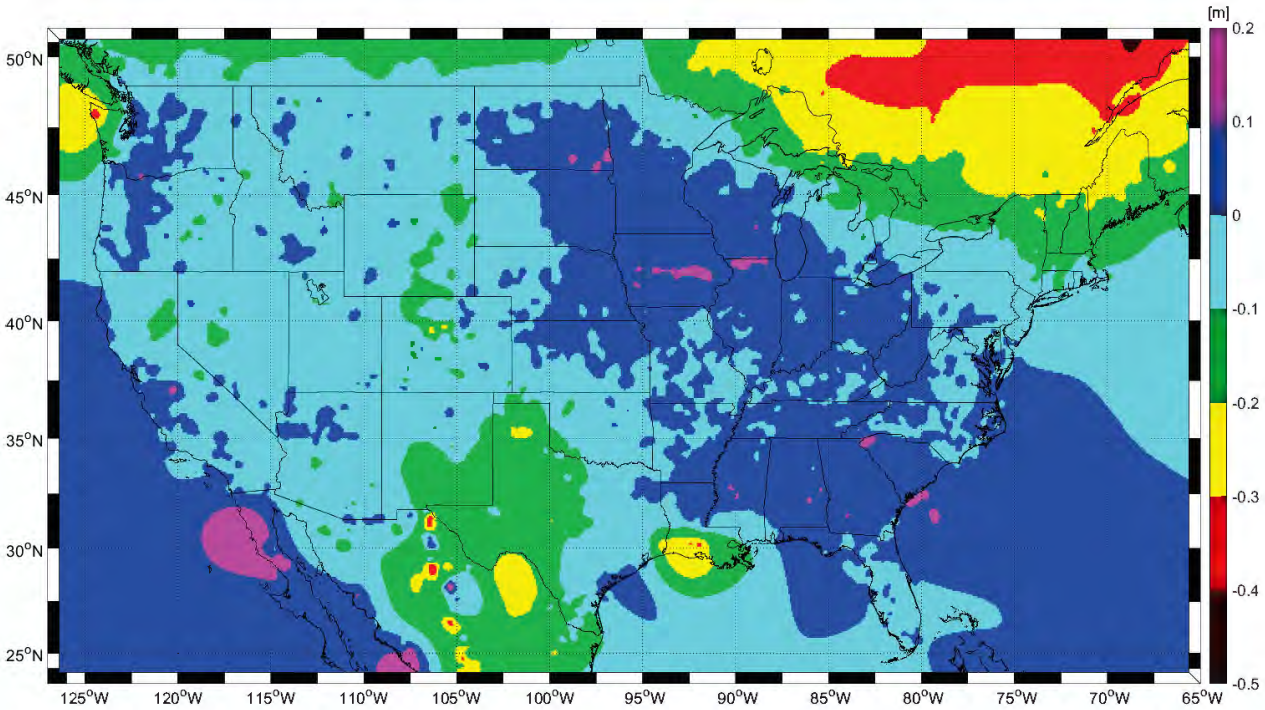

Figure 32: Warped surface component from LSC over CONUS from (6)

[2] xGEOID19B is provided with respect to IGS08, which requires a transformation to NAD83 (2011) as specified previously. Consequently, Figure 34 and Figure 35 are very close to the difference between orthometric heights in NAVD88 and NAPGD2022 except for this missing transformation element from NAD83 (2011) to IGS08.

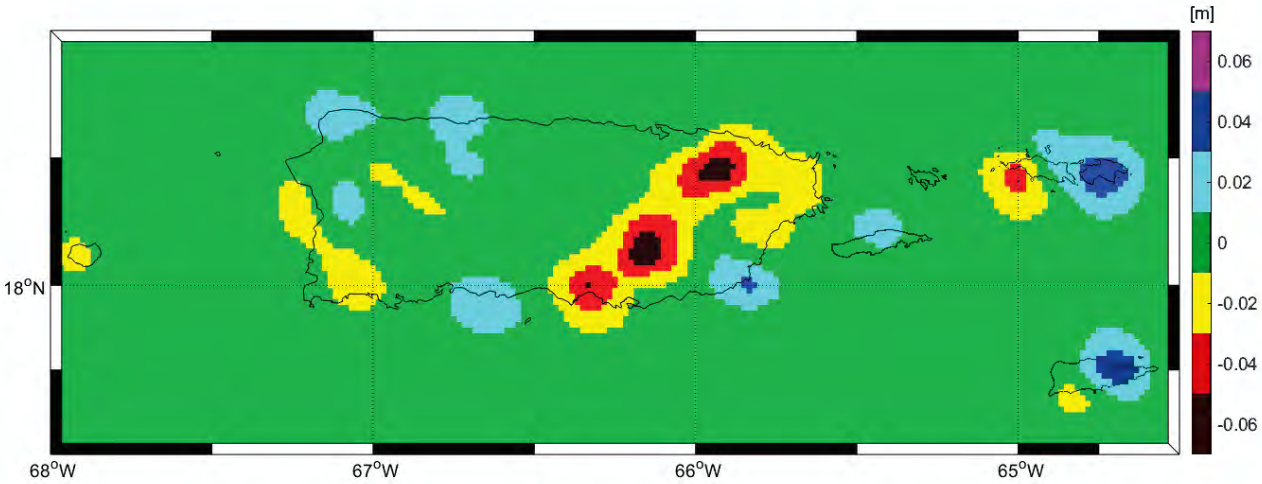

Figure 33: Warped surface component from LSC over PRVI from (6)

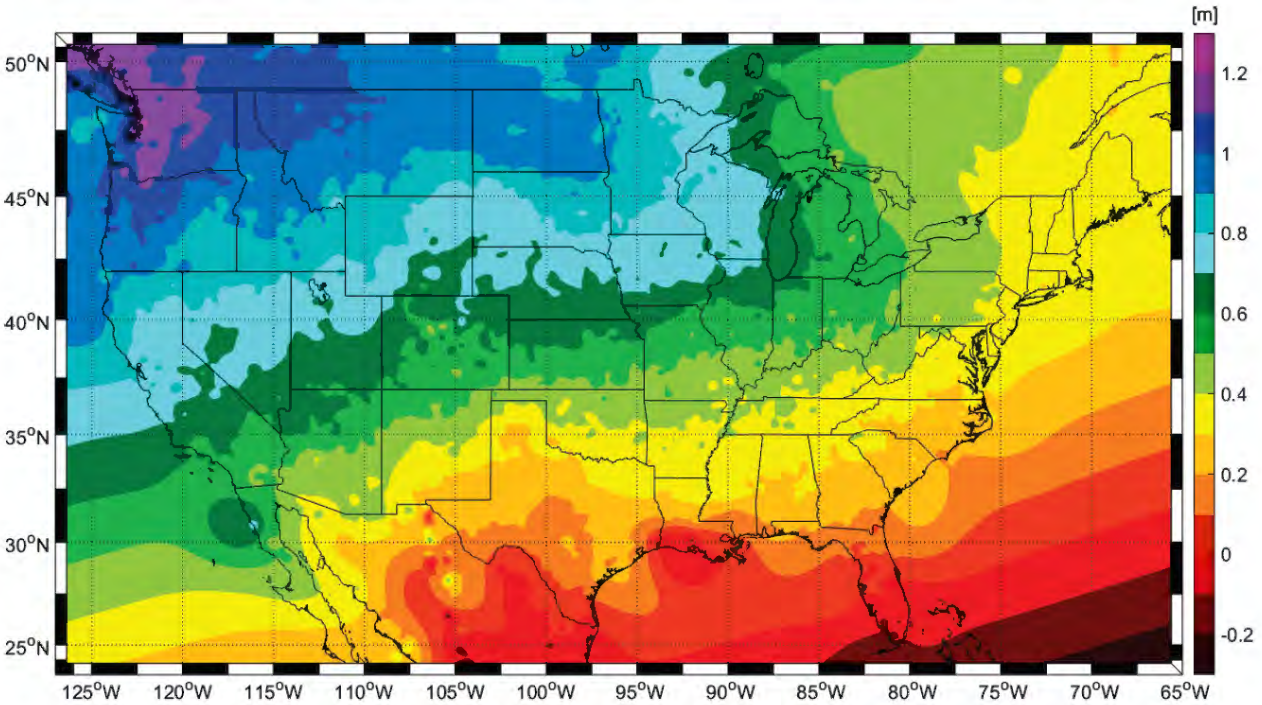

Figure 34: Warped + Bias + Tilt correction surface for CONUS. This surface gets added to the NAD83 (2011) morphed version of xGEOID19B, which results in GEOID18. Equivalent to the first two terms in (7).

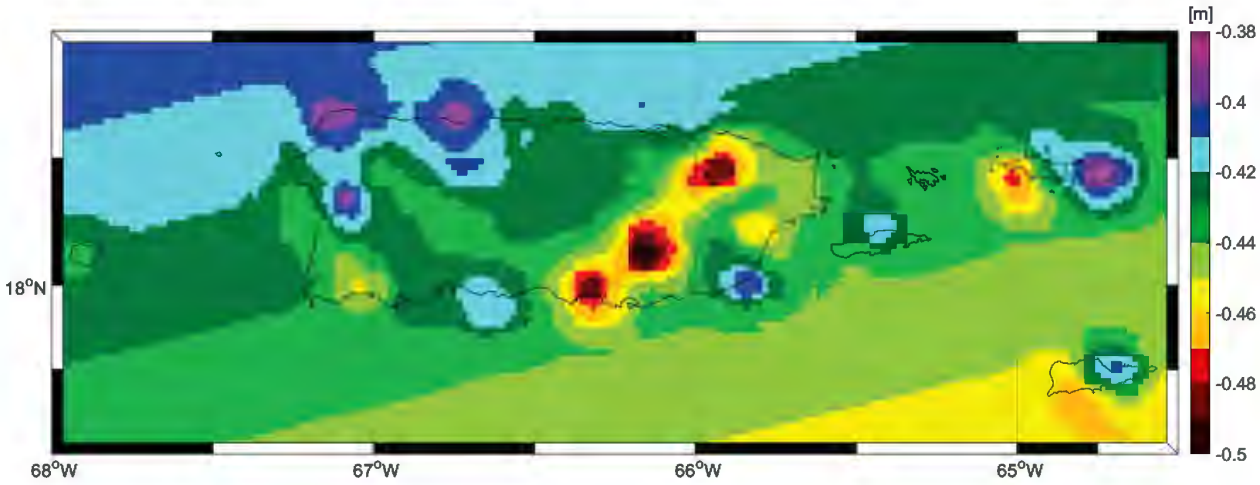

Figure 35: Warped + Bias + Tilt correction surface for PRVI. This surface gets added to the NAD83 (2011) morphed version of xGEOID19B, which results in GEOID18. Equivalent to the first two terms in (7).

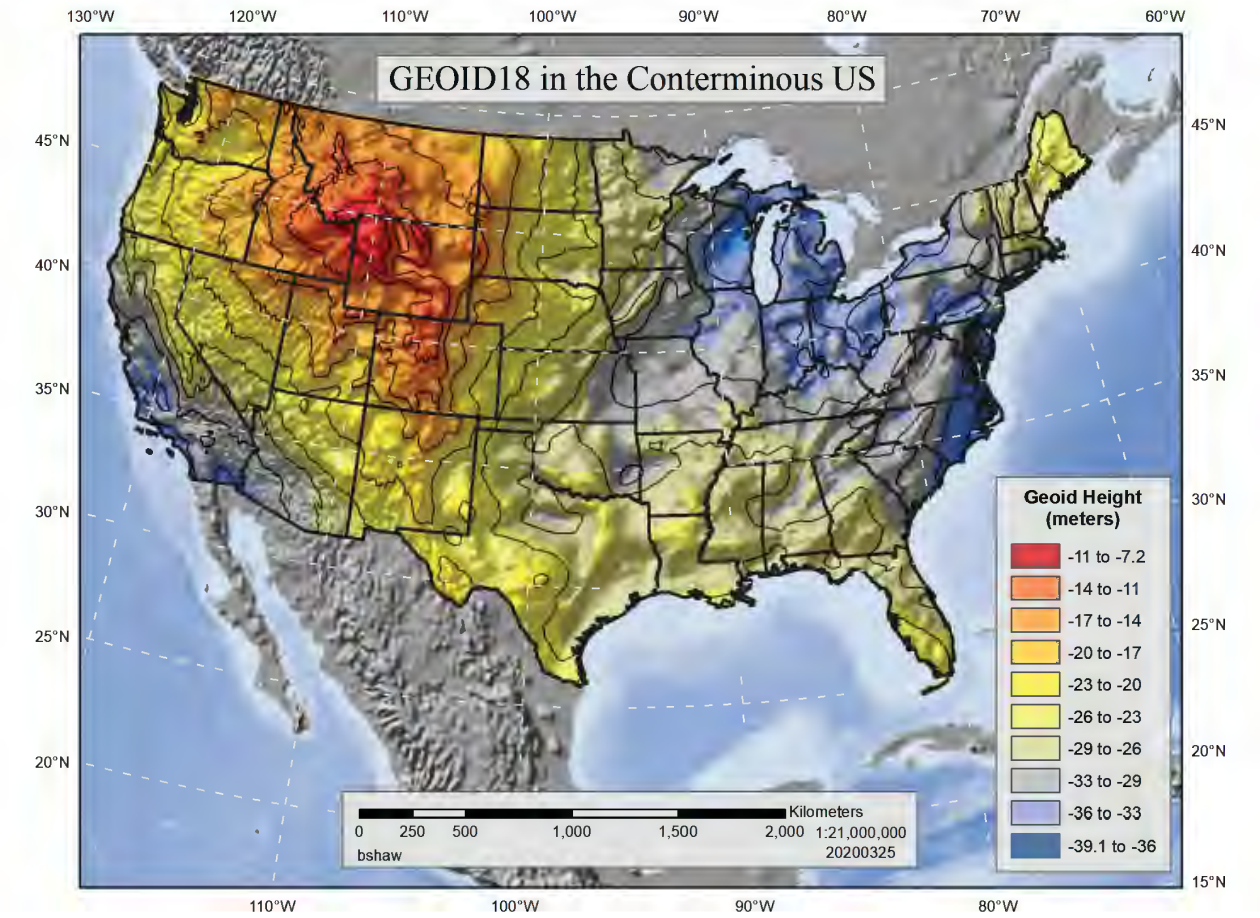

Figure 36: GEOID18 in the Conterminous US.

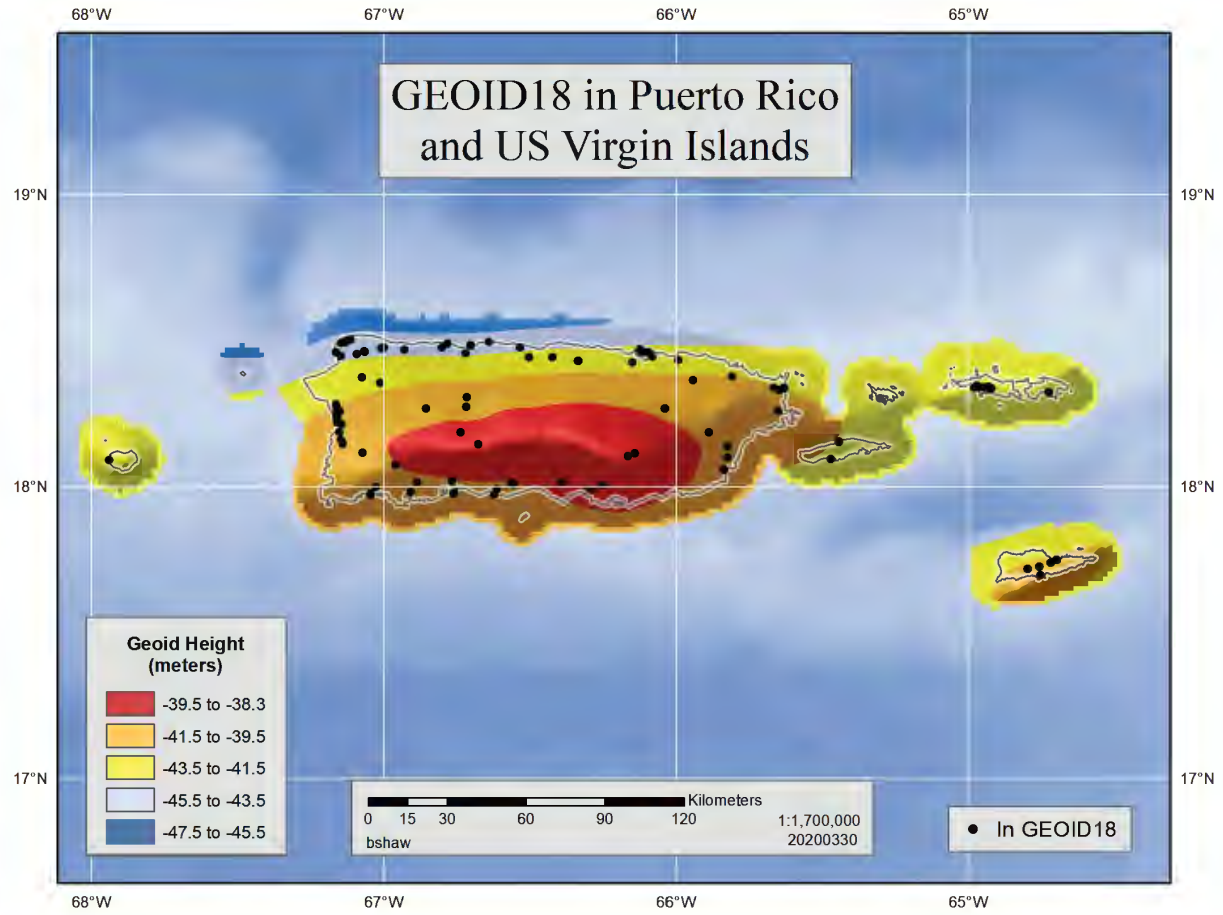

Figure 37: GEOID18 in Puerto Rico and U.S. Virgin Islands.

6.2 Uncertainty Estimates

The estimated uncertainty associated with GEOID18 is calculated from (9) and (10) and provided in a 1 arc-minute grid. The estimated uncertainty grids for CONUS and PRVI are illustrated in Figure 38 and Figure 39, respectively. Additional discussion and analysis of the uncertainty estimates are provided in Section 6.2.2.

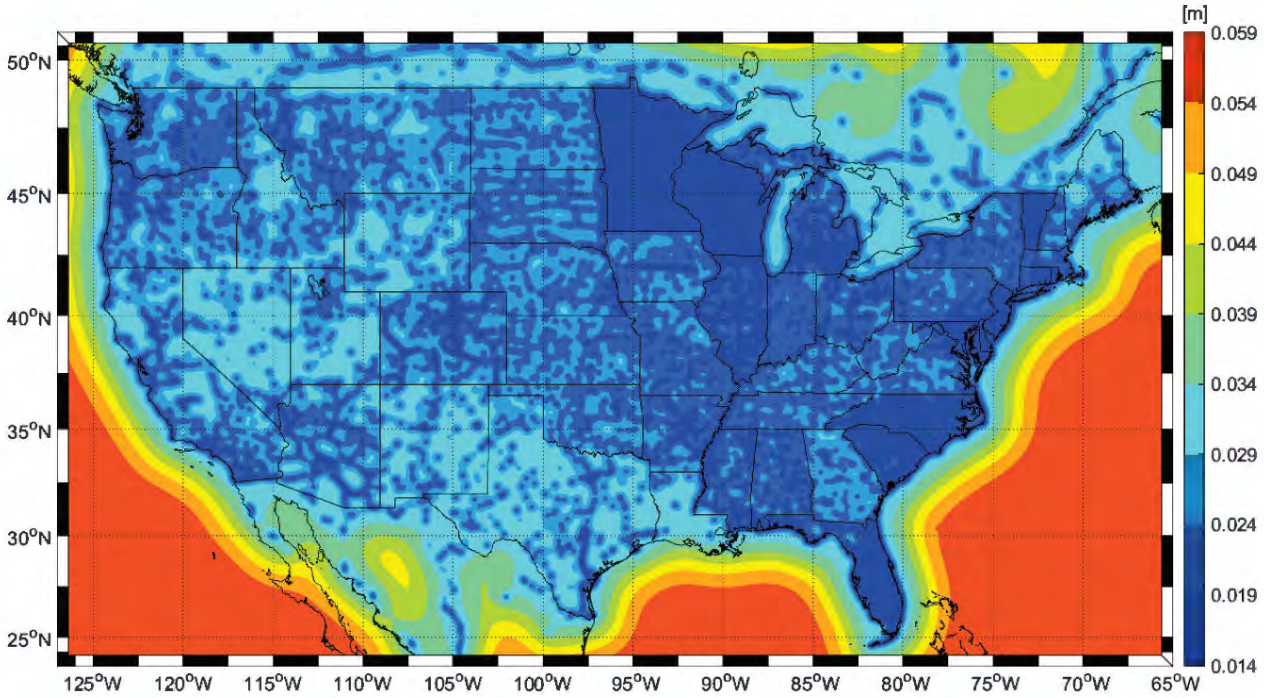

Figure 38: GEOID18 estimated uncertainty (1-sigma) over CONUS.

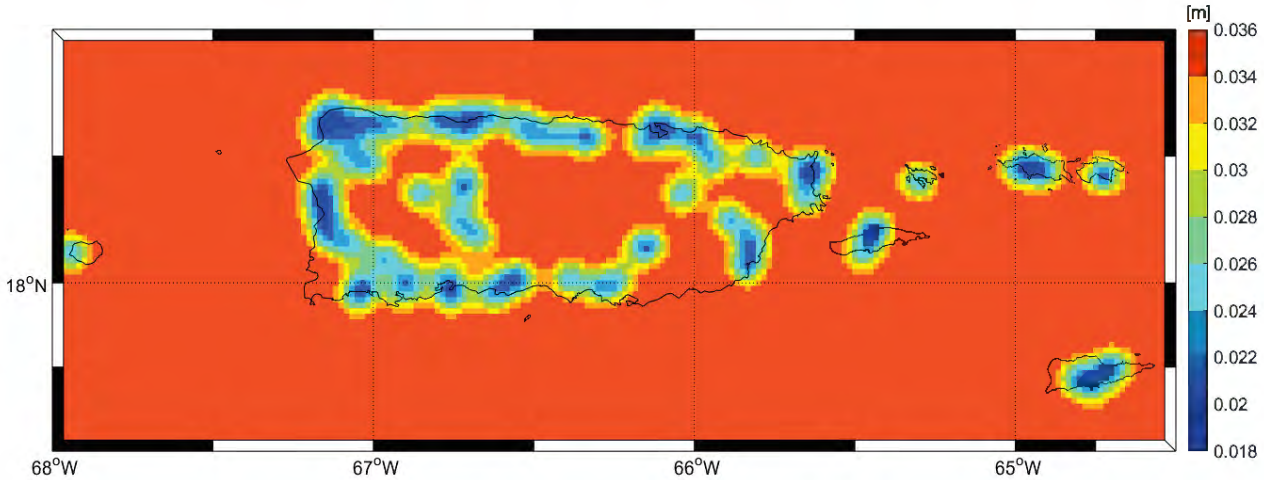

Figure 39: GEOID18 estimated uncertainty (1-sigma) over PRVI. Note: color scale is different than CONUS figure above.

7 GEOID18 Model Performance and Evaluation

In the following section, a number of additional metrics and statistics are presented to evaluate the overall performance of the GEOID18 model. This is meant to illustrate the overall high quality of GEOID18 by investigating its performance on a state-by-state basis. The rationale behind this more regional investigation is that while the overall model performance is definitely important, GEOID18 can be quite variable from state-to-state due to a number of factors including the GPS on Bench Marks quality and coverage and the gravimetric geoid quality within a particular state.

Evaluating the performance of a hybrid geoid model is not a straight-forward task with no perfect method to utilize. To overcome this deficiency, we make use of a number of statistics that will consider two classes of errors:

1. Commission-type error — A commission-type error is caused by inexact numerical observations. Example of a commission-type error is a higher residual on an individual mark or group of marks (i.e. noisier data).
2. Omission-type error — An omission-type error is caused by lack of appropriate sampling. An example of an omission-type error would be a situation where very few GPS on Bench Marks are available within a particular region.

Commission can only be evaluated on the existing data whereas omission evaluates the lack of data. As can be seen by the examples, these error classes are quite different to evaluate but both affect the performance of GEOID18. Additionally, they can act in opposite directions with the commission error decreasing and the omission error increasing at a particular location or within a specified region.

An additional way to illustrate GEOID18's performance is through comparison to GEOID12B where an improvement or degradation represents the ***relative*** performance increase or decrease from GEOID12B to GEOID18. An example of an improvement would be the lowering of the overall residual standard deviation from GEOID12B to GEOID18. This metric still must be separated into commission-type improvements and omission-type improvements. In this document, these relative improvements will be presented either as percentages (20% improvement) or in absolute changes (25 more GPS on Bench Marks) depending on the situation.

7.1 Commission-Type Errors / Improvements

To assess the overall commission-type performance of GEOID18, the residuals with respect to the final GEOID18 hybrid model are computed resulting in the post-model residuals that are illustrated in Figure 40 and Figure 41. The statistics for the post-model residuals are shown in Table 15. Since individual states often behave quite differently, the state-by-state post-model residuals are shown for GEOID18 and GEOID12B in Figure 42 and Figure 43, in an absolute sense and a relative sense, respectively. The CONUS-wide standard deviation of this residual is 1.39 cm, which is considerably lower than GEOID12B. For PRVI, there is a small increase in the standard deviation from 1.36 cm in GEOID12B to 1.66 cm for GEOID18. The 3 mm increase in standard deviation is caused by a slightly different covariance function used in GEOID18 that has a 10 km wavelength as the shortest component compared with GEOID12B which used a 5 km wavelength (see Section 8 Relative Accuracy and Figure 62 for the general discussion and examples on how the covariance function parameters can alter the fit of the model). The 10 km

wavelength parameter was used to avoid overfitting in PRVI and relies on the addition of almost 3x the number of bench marks used in GEOID18 compared with GEOID12B.

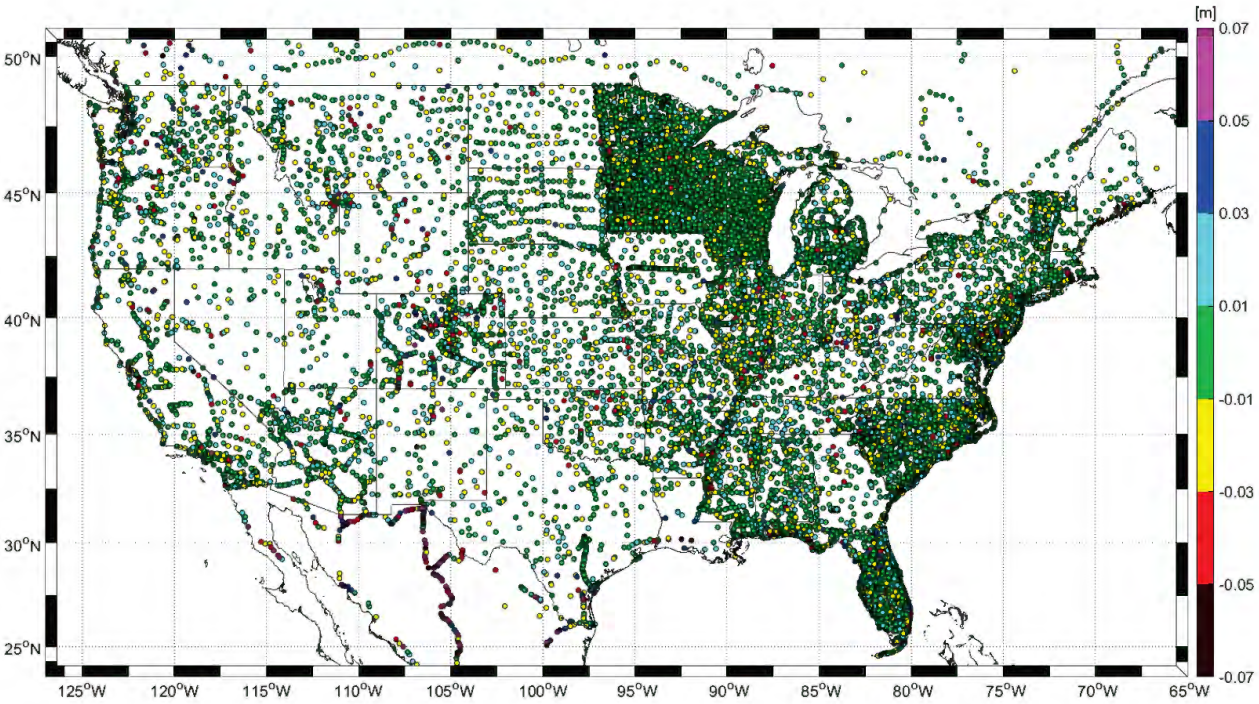

Figure 40: Post-model residuals with GEOID18

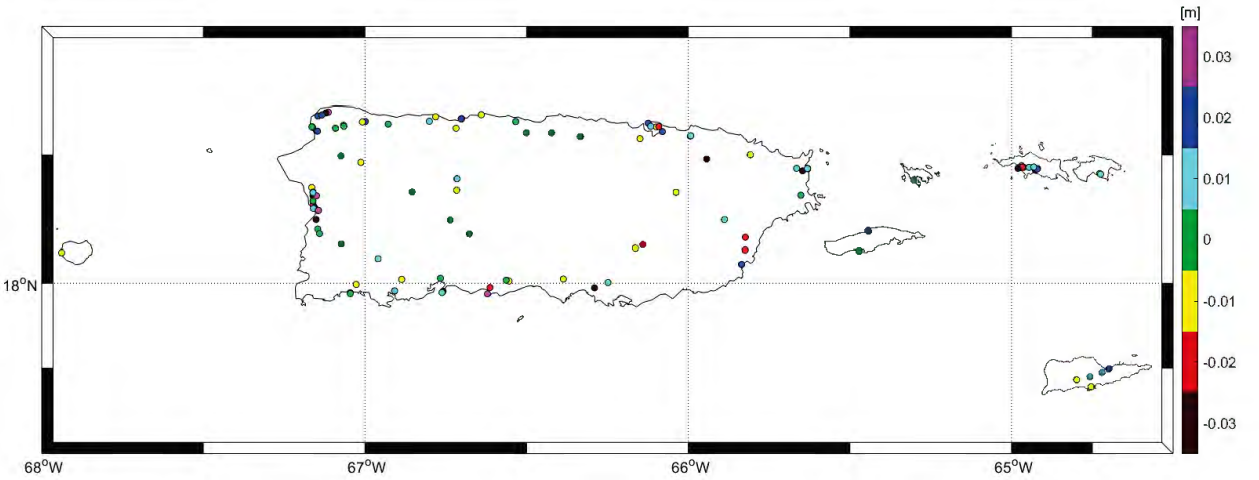

Figure 41: PRVI post-model residuals with GEOID18

Table 15: Statistics of post-model residuals for CONUS and PRVI with GEOID18

	Number Residuals	Min. [cm]	Max. [cm]	Mean [cm]	Std. Dev. [cm]
CONUS, Canada, and Mexico	32,357	-19.43	20.60	-0.00	1.39
CONUS	31,580	-10.12	8.17	0.00	1.27
PRVI	127	-4.37	4.52	-0.00	1.66

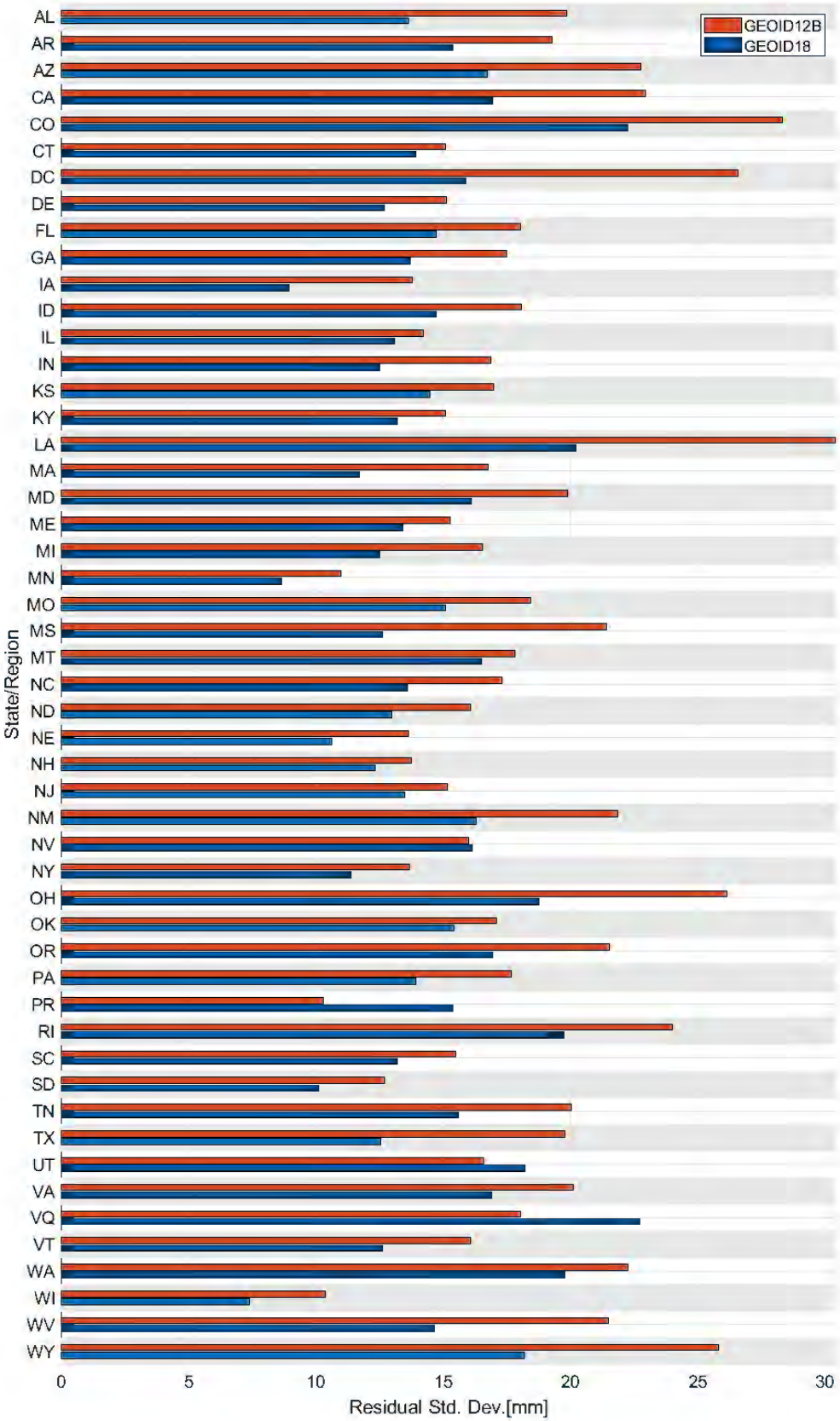

Figure 42: State-by-state standard deviation of post-model residual (N_{Hybrid} – h + H)) for GEOID12B and GEOID18

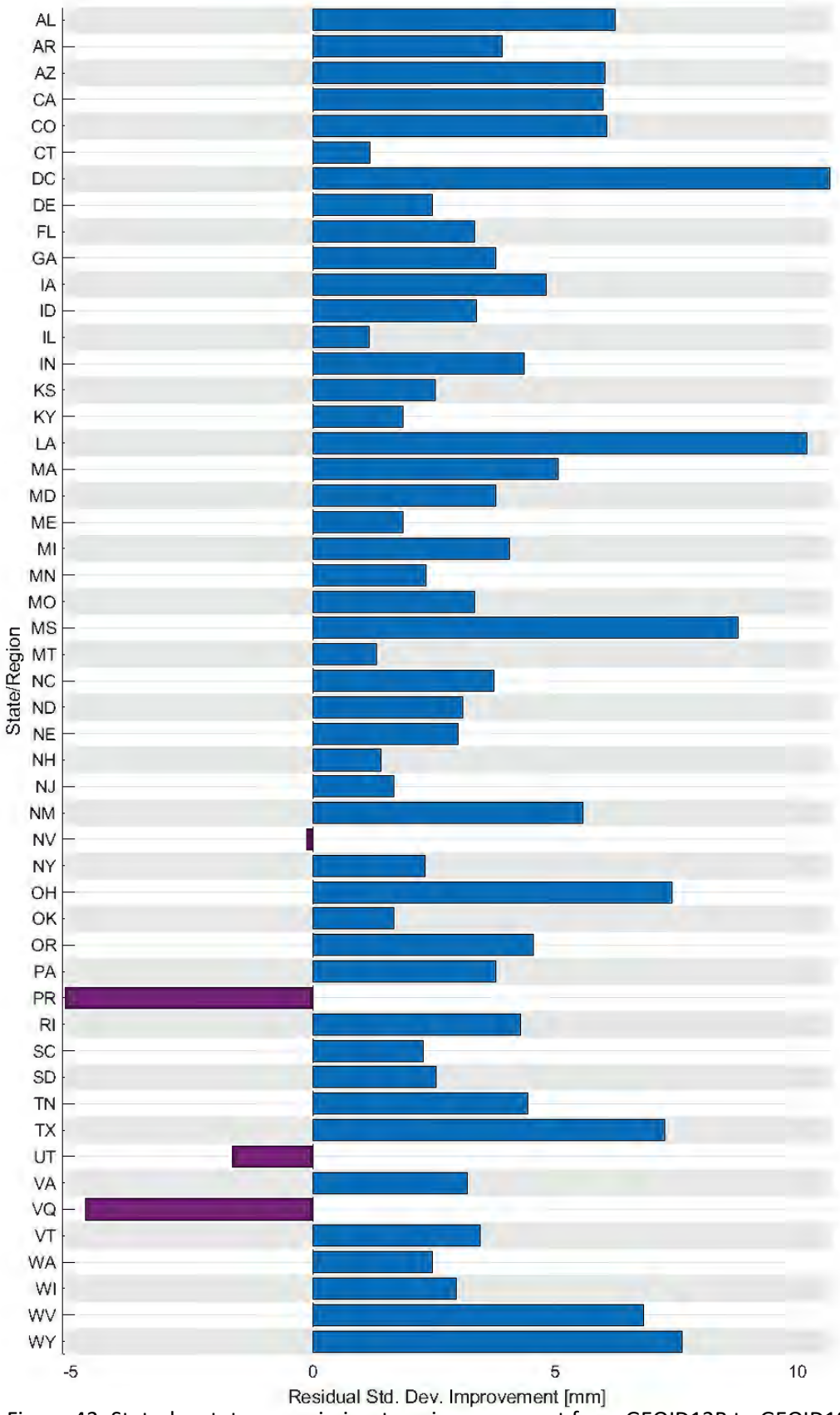

Figure 43: State-by-state commission-type improvement from GEOID12B to GEOID18

7.1.1.1 Causes of the Commission-Type Improvements in GEOID18 Compared to GEOID12B

There are two main reasons for the improvement in GEOID18 compared with GEOID12B:
1) xGEOID19B and
2) removal of outliers in the GPS on bench marks.

The improvement that came from the use of xGEOID19B is a combination of a number of factors including (in no particular order) a better Digital Elevation Model, better surface gravity data, better satellite gravity from the Gravity field and steady-state Ocean Circulation Explorer (GOCE) mission and Gravity Recovery and Climate Experiment (GRACE) mission, inclusion of the Gravity for the Redefinition of the Vertical Datum (GRAV-D) airborne gravity data, and refined geoid modeling theory. It is nearly impossible to isolate the impact of any single factor within xGEOID19B that is causing the improvement. The lone exception is the GRAV-D data which can be investigated through comparison with the xGEOID19A model.

To get a sense of the level of improvement in the hybrid geoid model from xGEOID19B, we compute residuals with respect to the gravimetric geoid model used in GEOID12B, USGG2012, and compare these with the xGEOID19B residuals on a state-by-state basis. Over all of CONUS, the improvement from USGG2012 to xGEOID19B is very minor with standard deviation improvement of just over 1 mm. On a state-by-state basis, this improvement is slightly more impactful but still only a modest improvement in the overall GEOID18 model. In Figure 44, the improvement due to xGEOID19B is shown for the majority of states/regions covered by GEOID18 (39 out of 51). Maine has the largest improvement at 19 mm while Mississippi has the most degradation of any state at just 4 mm.

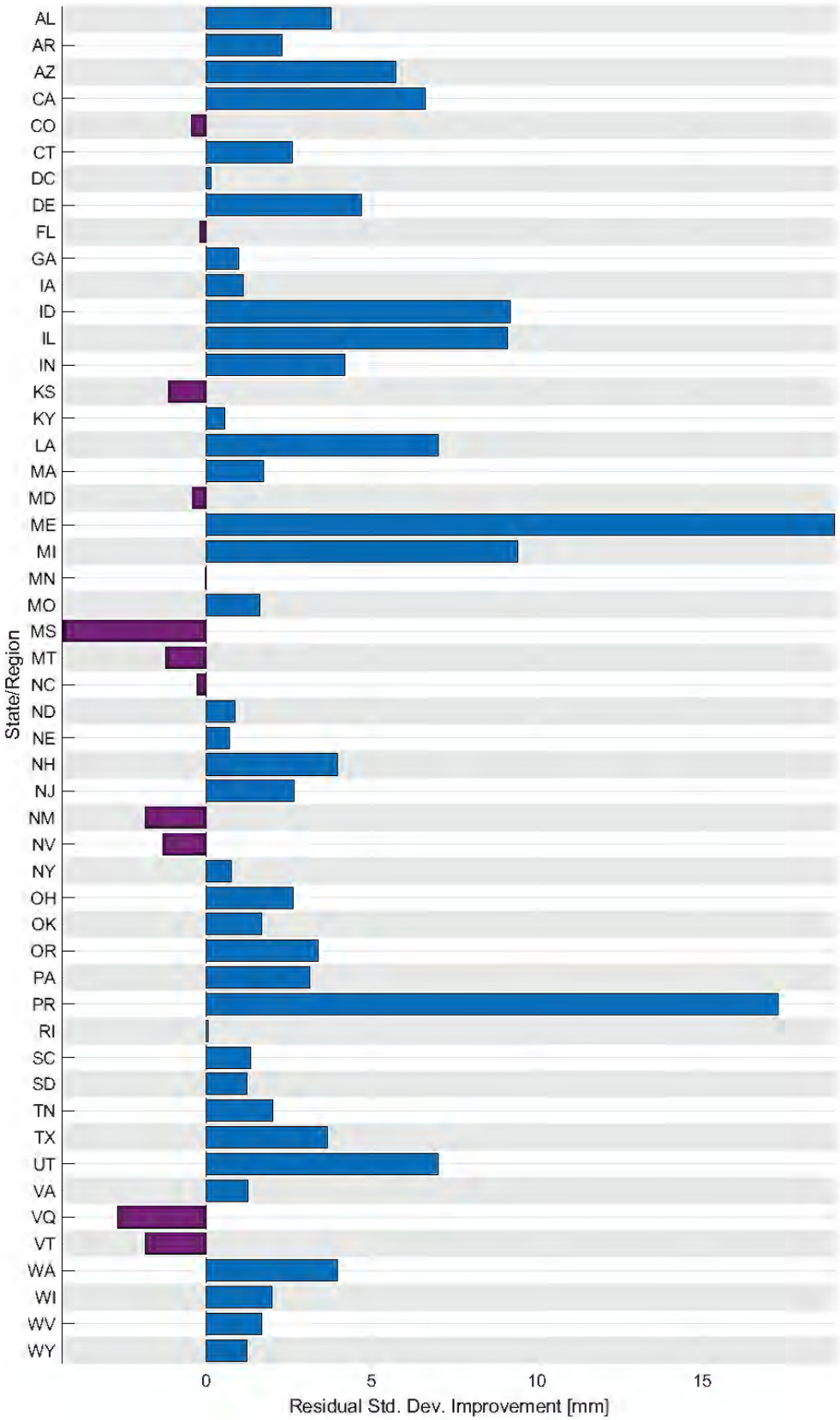

Figure 44: Improvement in xGEOID19B compared to USGG2012. Bench marks from GEOID18 compared with xGEOID19B and USGG2012.

Attempting to pull out the improvement caused solely from GRAV-D, the commission-type improvement from xGEOID19A to xGEOID19B is shown in Figure 45.

Two major themes from the GRAV-D inclusion are:
1) most regions see improvement with GRAV-D and
2) this improvement is at the 2-3 mm level.

The vast majority of states/regions (42 out of 51) show improvement or no change[3]. Figure 45 can be thought of as one contribution to the overall improvement in Figure 44. For example, Maine has an overall improvement of 19 mm with about 12.5 mm coming from GRAV-D. Generally, the GRAV-D contribution is quite minor though with most states/regions only being impacted at the 2–3 mm level and only 4 states have a 5+ mm improvement.

[3] There are still a handful of states that have no GRAV-D coverage so no change is expected over those states.

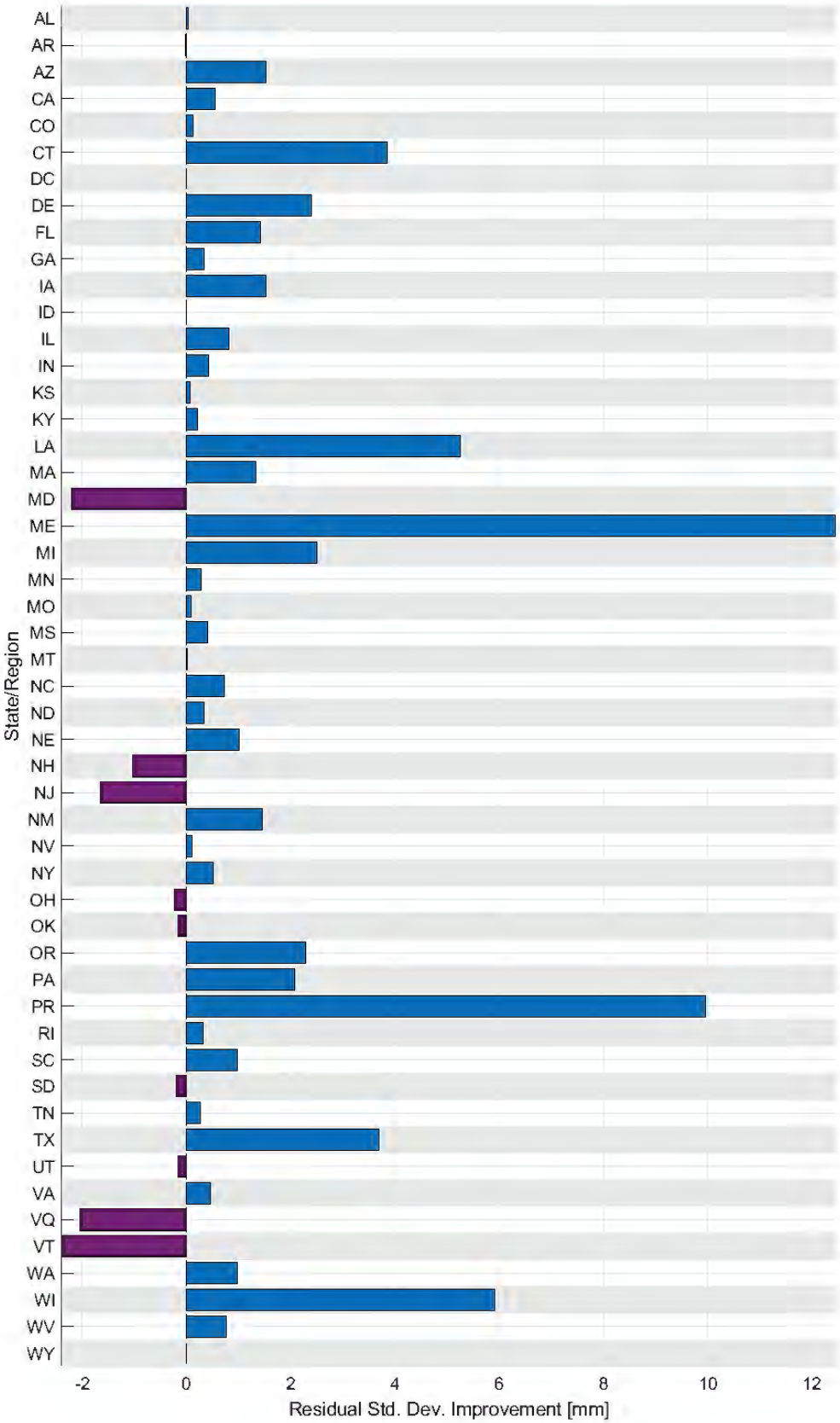

Figure 45: Improvement in GEOID18 caused by using xGEOID19B compared to xGEOID19A. Bench marks from GEOID18 compared with xGEOID19B and xGEOID19A.

Considering the very minor improvement in GEOID18 caused by xGEOID19B, it is evident that the primary improvement in GEOID18 is driven by the use and removal of the GPS on Bench Marks. To illustrate this, the GPS on Bench Marks dataset from GEOID12B is compared with GEOID18's GPS on Bench Marks. Both GPS on Bench Marks sets are computed with respect to xGEOID19B and illustrated in Figure 46. The CONUS-wide standard deviation improvement from GPSBM12 to GPSBM18 is 4.6 mm. This is more an indicator of how well the gravimetric geoid fits the data and degradations in this metric are not necessarily indicative of a degradation in GEOID18. This is especially true for states near the periphery that have significant new GPS on Bench Marks.

These new bench marks are negatively impacted by the poorly fitting tilt of the surface, which causes a 'false positive' in some of the negatively impacted states including Nevada and Wisconsin. If we replace the systematic correction surface from a 2D plane to a four-parameter model surface as described in Kotsakis and Sideris, (1999), we see that a standard deviation improvement actually exists for Nevada and Wisconsin as illustrated in Figure 47. Two other states (Louisiana and New Hampshire) are still both negative but show improvement with the use of the four-parameter model compared with the 2D plane. The states that still have a decrease in performance that is not improved with the four-parameter model include Illinois and New York. The likely cause for Illinois is the enormous amount of new bench marks included in GEOID18 (796 compared to 420), and these new residuals are just a little bit noisier (+2.1 mm standard deviation) than previous data as illustrated in Figure 48. For the New York residuals, there is just a slight degradation in the fit between the bench marks and xGEOID19B when going from GPSBM12B to GPSBM18. This is illustrated in Figure 49 where the off-centered bins have slightly more power for the GPSBM18, which causes the standard deviation to increase slightly for GEOID18. In both of these states, the decrease is so minor that it is not a cause of major concern. Again, this is a commission-type error and not necessarily indicative of the overall quality of GEOID18 in these states that show degradation.

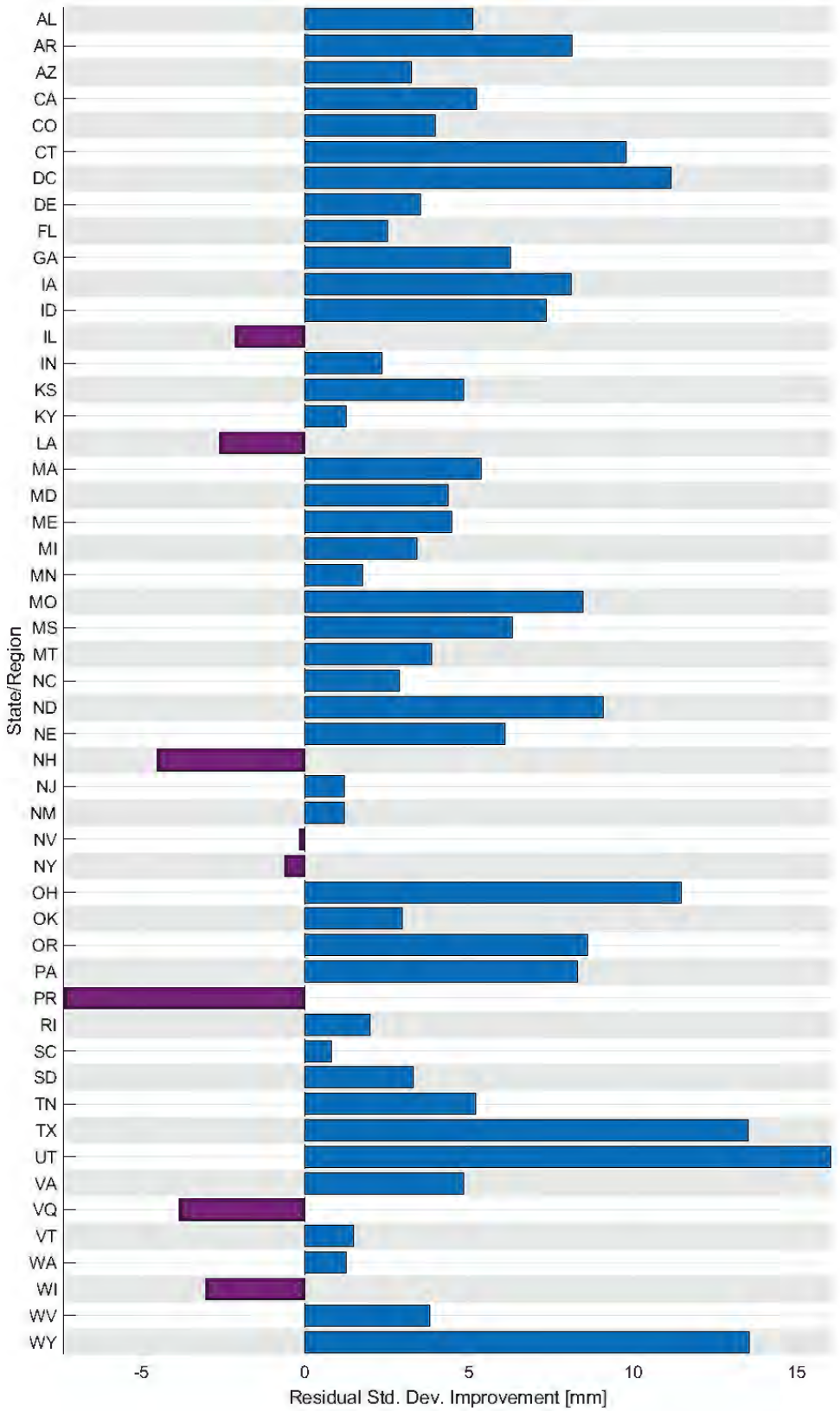

Figure 46: Improvement in GEOID18 caused by GPS on Bench Marks. Improvement is the decrease in the standard deviation between GPSBM18 and GPSBM12 both with respect to xGEOID19B.

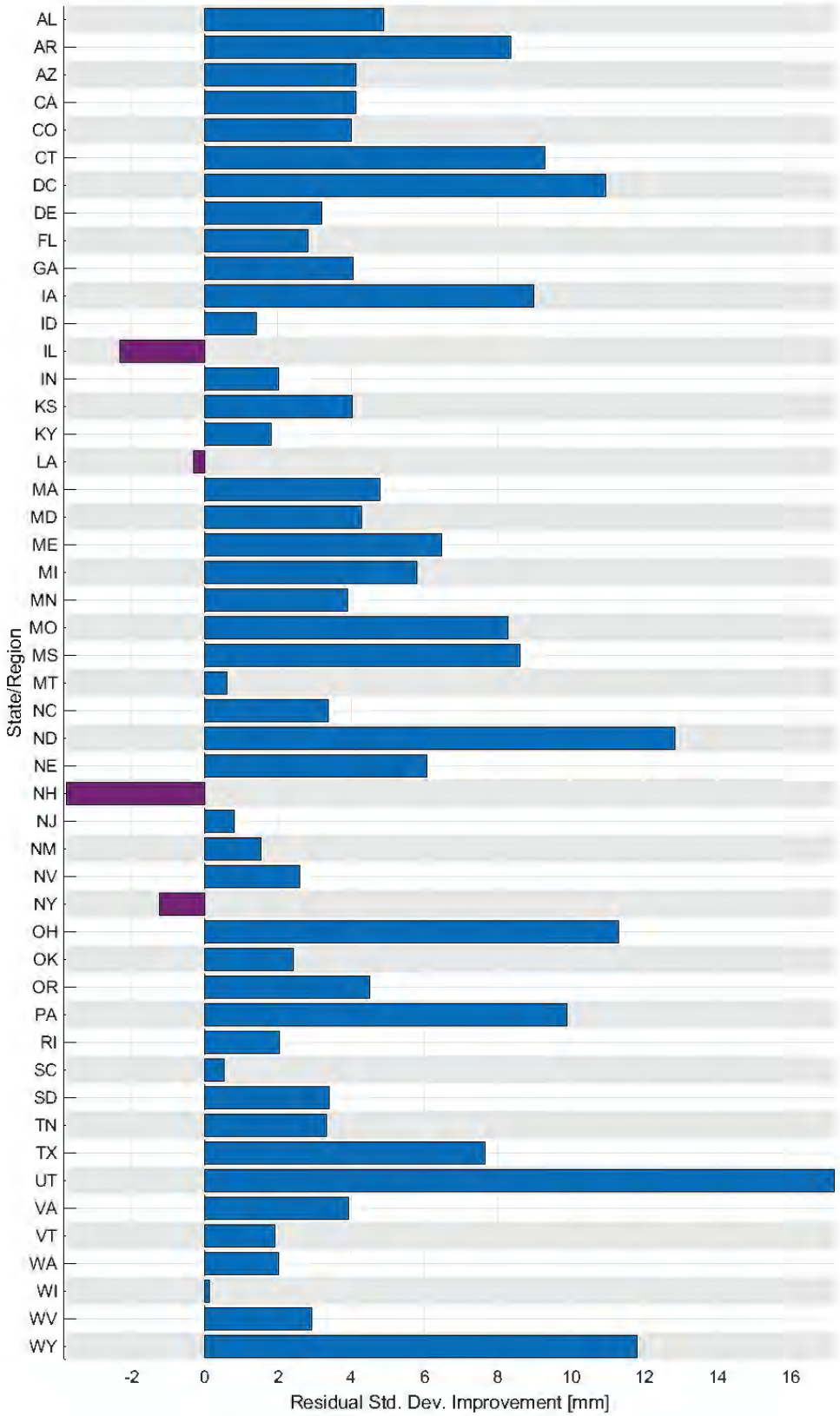

Figure 47: Improvement from GPS on BMs with a 4 parameter surface removed after Kotsakis and Sideris, (1999). Improvement is the decrease in the standard deviation between GPSBM18 and

GPSBM12 both with respect to xGEOID19B. Puerto Rico/U.S. Virgin Islands are not shown here as that region is not affected by the systematic effects that the 4 parameter surface would address.

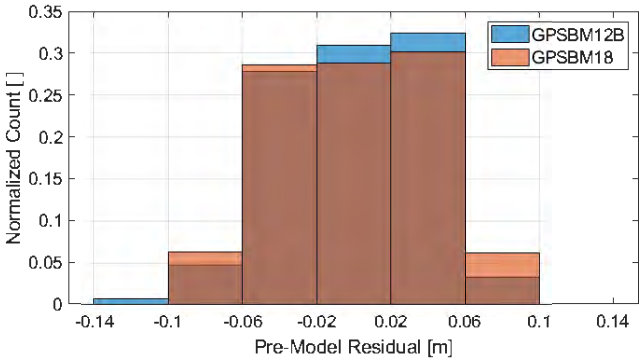

Figure 48: Pre-model residuals for the state of Illinois from GPSBM12B and GPSBMS18. The additional residuals in the ±(0.06 to 0.10 cm) bins causes the overall standard deviation to worsen from the GPSBM12B data to the GPSBM18 data.

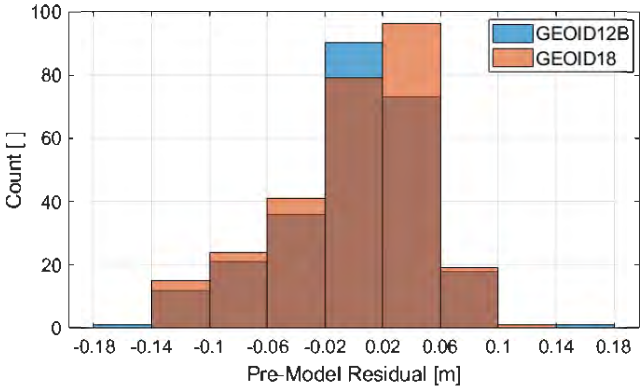

Figure 49: Pre-model residuals for the state of New York from GPSBM12B and GPSBMS18. The slight increase in the off-centered bins causes the overall standard deviation to worsen from the GPSBM12B data to the GPSBM18 data.

7.2 Omission-Type Errors

In the following section, a number of metrics are investigated to evaluate the omission-type improvement in GEOID18. In general, this is difficult to assess because any available high-quality bench mark with a GPS-derived ellipsoid height was used in the GEOID18 modeling, leaving little to check the omission-type error. A number of factors that will be presented are the changes between GEOID12B and GEOID18 with respect to the distance to the nearest used GPS on bench mark, the estimated uncertainties, and the jackknifing results that were previously presented.

7.2.1 Minimum Distance to Bench Mark:

A key factor in the overall quality in a hybrid geoid model is the GPS on Bench Marks data and how well distributed they are. In this omission-type context, how close any given location is to a used bench mark is critical to how well the geoid model fits that area. This is completely driven by the covariance function used in the LSC prediction, where the closer one is to a used bench mark, the lower the estimated uncertainty and higher confidence in the geoid model performance. This distance improvement is shown in Figure 50 where the minimum distance to a used bench mark is compared between GEOID18

and GEOID12B. Overall, 25.4% of land areas within CONUS exhibits an improvement of 1 km or more in the minimum distance to a used bench mark compared to GEOID12B. This is in comparison to approximately 8.4% or land areas that exhibit a degradation of 1 km or more. The results for PRVI are illustrated in Figure 51, where the minimum distance has been vastly improved. The percentage of land areas for CONUS and PRVI that are within a variable distance to a used GPS on Bench Mark are shown in Table 16 and Table 17, respectively.

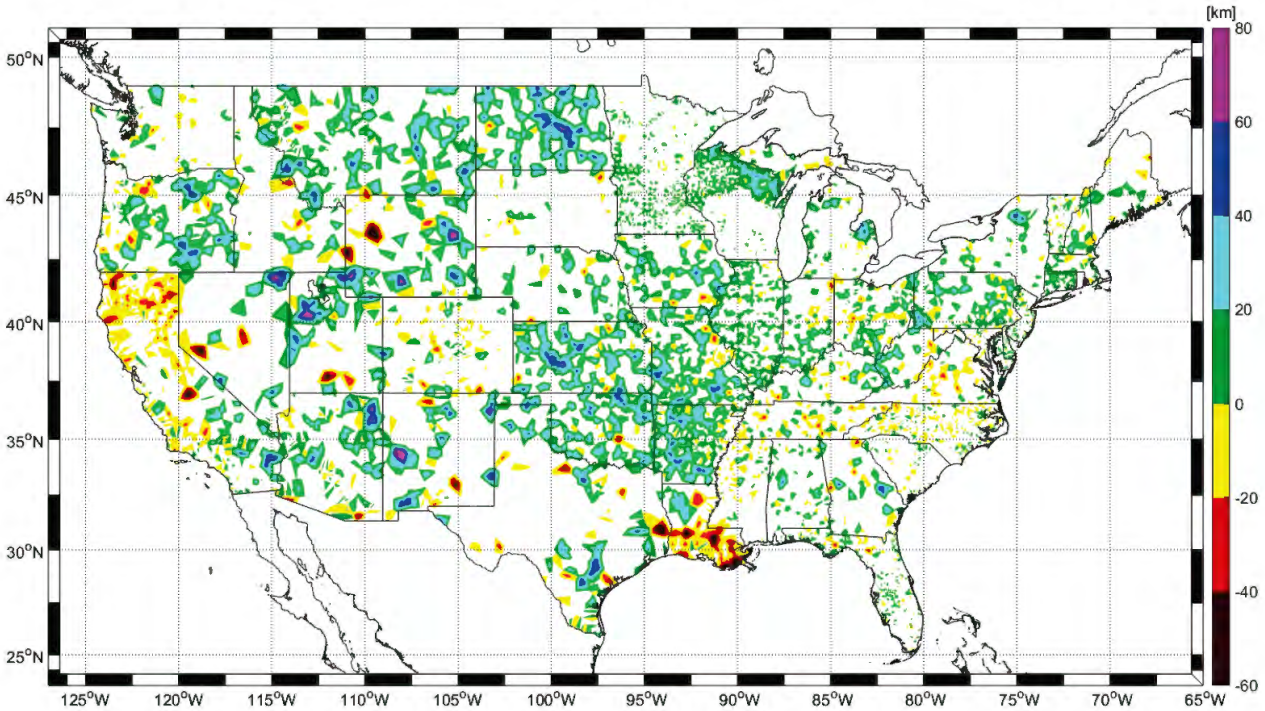

Figure 50: Distance improvement between GEOID12B and GEOID18. Positive numbers indicate a location is closer to a GPS on Bench Mark in GEOID18 compared to GEOID12B. Areas that show essentially no difference (within 1 km) are masked to highlight those areas that show change.

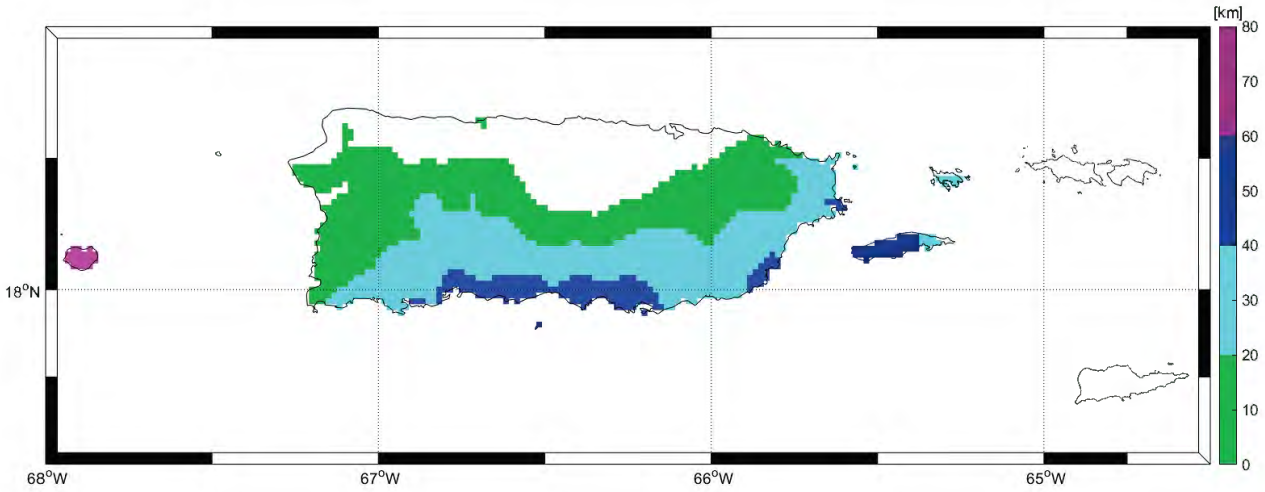

Figure 51: PRVI Distance improvement between GEOID12B and GEOID18. Positive numbers indicate a location is closer to a GPS on Bench Mark in GEOID18 compared to GEOID12B. Areas that show essentially no difference (within 1 km) are masked to highlight those areas that show change.

Table 16: Percentage of CONUS below a minimum distance to a GPS on Bench Mark

	10 km	20 km	30 km	45 km	60 km
GEOID12B	27.13%	56.37%	76.50%	92.19%	97.84%
GEOID18	31.88%	65.36%	83.92%	94.82%	98.56%

Table 17: Percentage of PRVI below a minimum distance to a GPS on Bench Mark

	10 km	20 km	30 km	45 km	60 km
GEOID12B	25.00%	44.79%	63.68%	89.90%	99.10%
GEOID18	75.07%	98.19%	99.97%	100%	100%

While the previous figures illustrate where the spatial improvement is occurring, it is a bit difficult a get a general sense of this omission improvement. In Figure 52 and Figure 53, the percentage of the total land area within a used GPS on Bench Mark at all distances is shown for CONUS and PRVI, respectively. For CONUS, we see that the largest improvement of roughly 9% occurred at the 20 km level (peak in Figure 52b), which corresponds very well with NGS's recommendation to observe bench marks at 30 km spacings, which would result in any location to be 15 km away from a used bench mark. Results in PRVI as shown in Figure 53 are even more impressive where the largest impact has over 50% more of the land area within 10 km of a used GPS on Bench Mark in GEOID18 compared to GEOID12B! Overall, over 98% of PRVI land areas are within 20 km of a used bench mark.

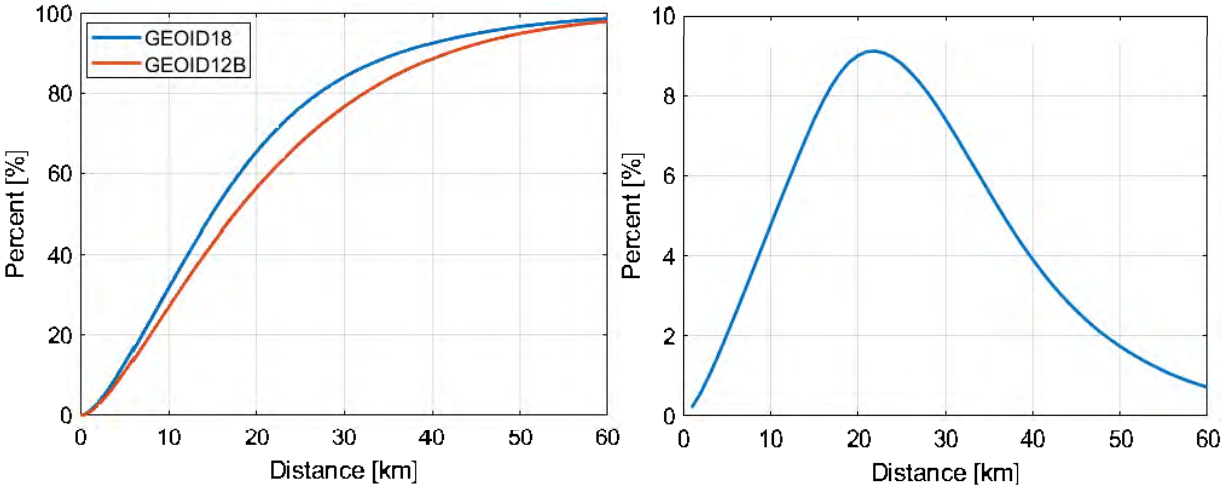

Figure 52: **a** (at left) Percent of CONUS land areas within a minimum distance to a used GPS on Bench Mark for both GEOID12B and GEOID18. The improvement is represented by the shifting to the left of the curve from GEOID12B to GEOID18. **b** (at right) Difference in GEOID12B and GEOID18 curves in left figure. This is the distance improvement percent from GEOID12B to GEOID18.

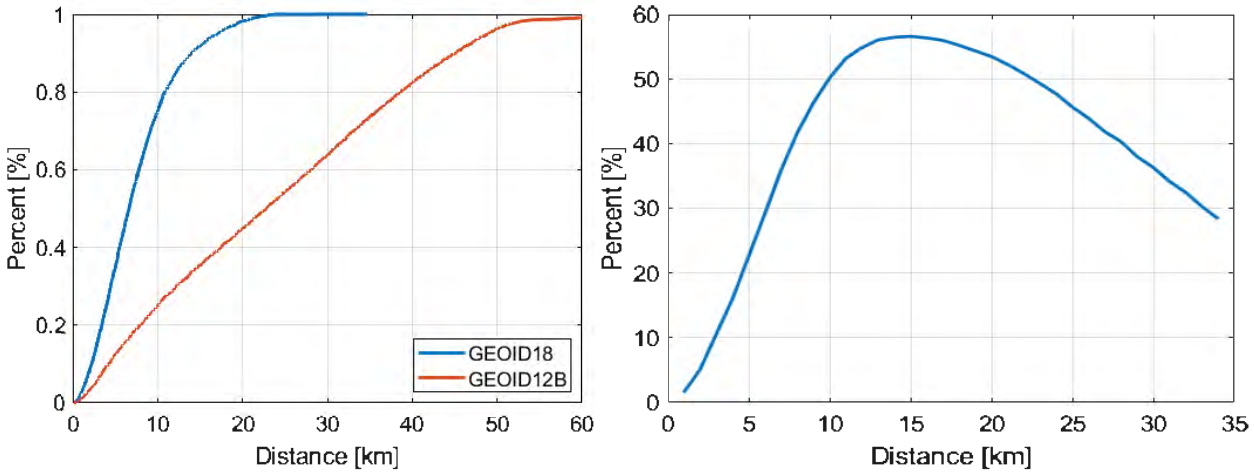

Figure 53: **a** (at left) Percent of PRVI land areas within a minimum distance to a used GPS on Bench Mark for both GEOID12B and GEOID18. The improvement is represented by the shifting to the left of the curve from GEOID12B to GEOID18. **b** (at right) Difference in GEOID12B and GEOID18 curves in left figure. This is the distance improvement percent from GEOID12B to GEOID18.

Each individual state can have quite a difference in the GPS on Bench Marks distribution and minimum distances compared to the overall CONUS results. The improvement (change in percent of land area) from GEOID12B to GEOID18 is shown in Figure 54. The majority of states exhibit a noticeable improvement at the 30 km level. Overall, only seven states show a decrease in the percent of land area within 30 km of a GPS on Bench Mark with the only significant decrease in Louisiana. The reason that some states have a very small decrease is that GEOID18 used a slightly more restrictive threshold for NAVD 88 leveling observations within the GPS on Bench Marks than was done in GEOID12B. For GEOID18, marks that were codified as 'LEVELING' were not used whereas they were in GEOID12B. The reason for this decision is that this class of marks has not been rigorously adjusted within the NAVD 88 network. In certain locations like Northern California and Tennessee, these LEVELING-based marks make up a large portion of the bench marks.

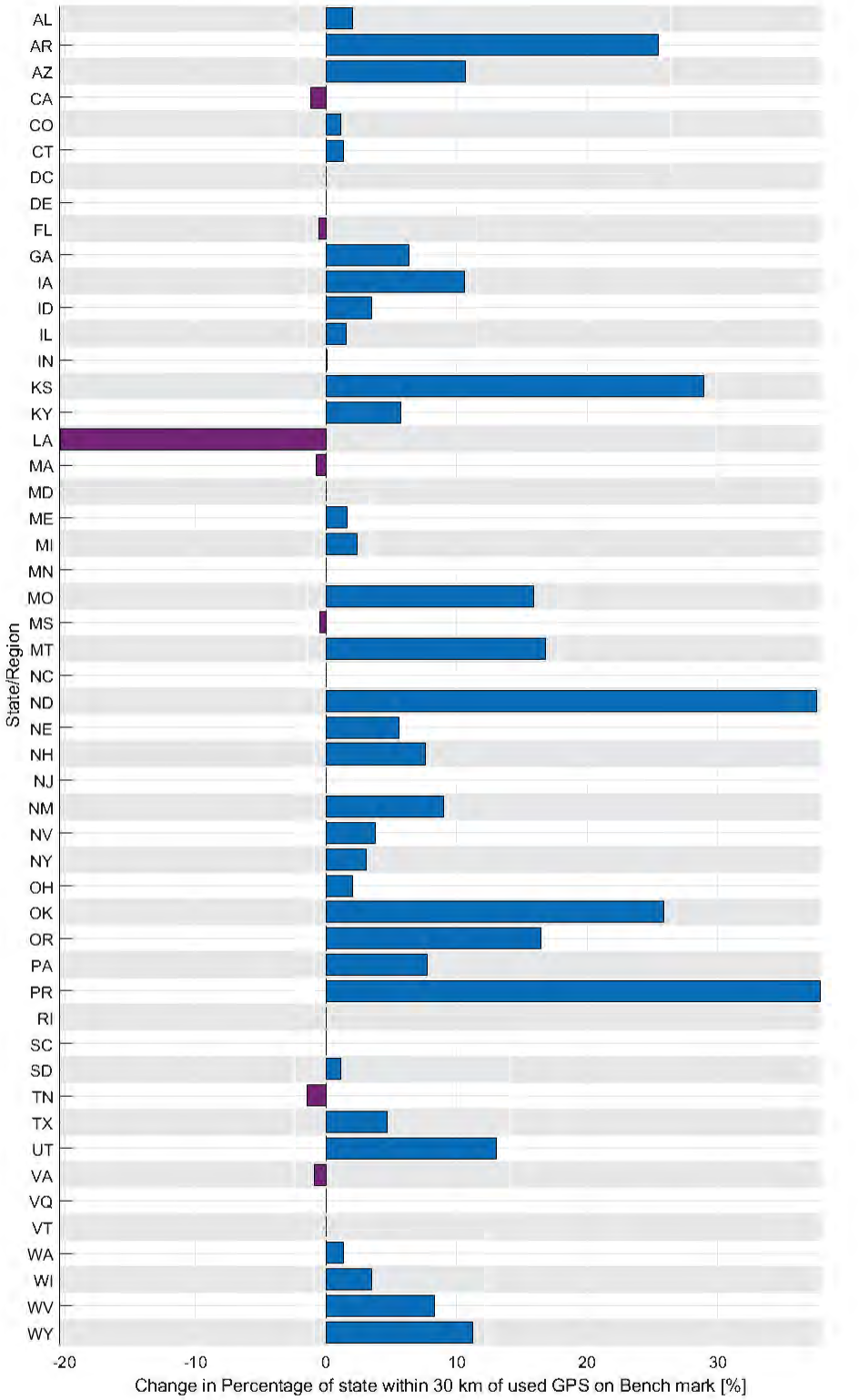

Figure 54: Change in percentage of a state that is within 30 km of a used GPS on Bench Mark from GEOID12B and GEOID18. States with blue or positive percentage show improvement – i.e. more of the state is within 30 km of used GPS on Bench Mark.

7.2.2 Uncertainty Estimates

The following section illustrates some essential factors related to the uncertainty estimates (see Figure 38 and Figure 39). Two essential factors to remember throughout the remainder of this section: 1) the overall quality of any residual has NO influence on the uncertainty estimate and 2) the geographic distribution is irrelevant. With regards to 1), when computing the uncertainty estimate from (9), the actual residual amount is not present — only the distance and the *a priori* uncertainty. Secondly, the LSC prediction used for GEOID18 relies on the following assumptions (see Moritz, 1980; Sanso, 1986 for more information):

1. Stationarity: the residuals reflect a stationary process, which means that the statistics are assumed to be the same everywhere.
2. Isotropy: An isotropic covariance function based on empirical data is used, which means that the covariance function is independent of direction. Practically, this means that our geographic distribution of residuals does not factor into the prediction, which is counterintuitive to many geodetic networks and problems.

In many geodetic applications and problems, it is known to be mathematically advantageous to have a strong, well-distributed geographic arrangement or spatial homogeneity as shown in Figure 55a, where the estimated uncertainty is desired at the green mark and the magenta marks represent GPS on Bench Mark locations that are used to compute the uncertainty. The four magenta marks are all at roughly a 30 km distance from the desired mark. In Figure 55a, the magenta marks are all well distributed surrounding the point of interest. However, when computing the uncertainty estimate with LSC, this support is irrelevant and the situation shown in Figure 55b provides the same level of impact to the uncertainty estimate as all marks shown in purple are at roughly 30 km from the central green location. Others have used anisotropic covariance functions in LSC (Darbeheshti and Featherstone, 2009), but to provide consistency with previous NGS hybrid geoid models, GEOID18 continues to use stationary and isotropic assumptions. To reiterate, these two geographic distributions would definitely cause different impacts on the computed geoid model (especially when one considers the effect of the tilt), but they do not cause different impacts to the uncertainty estimate.

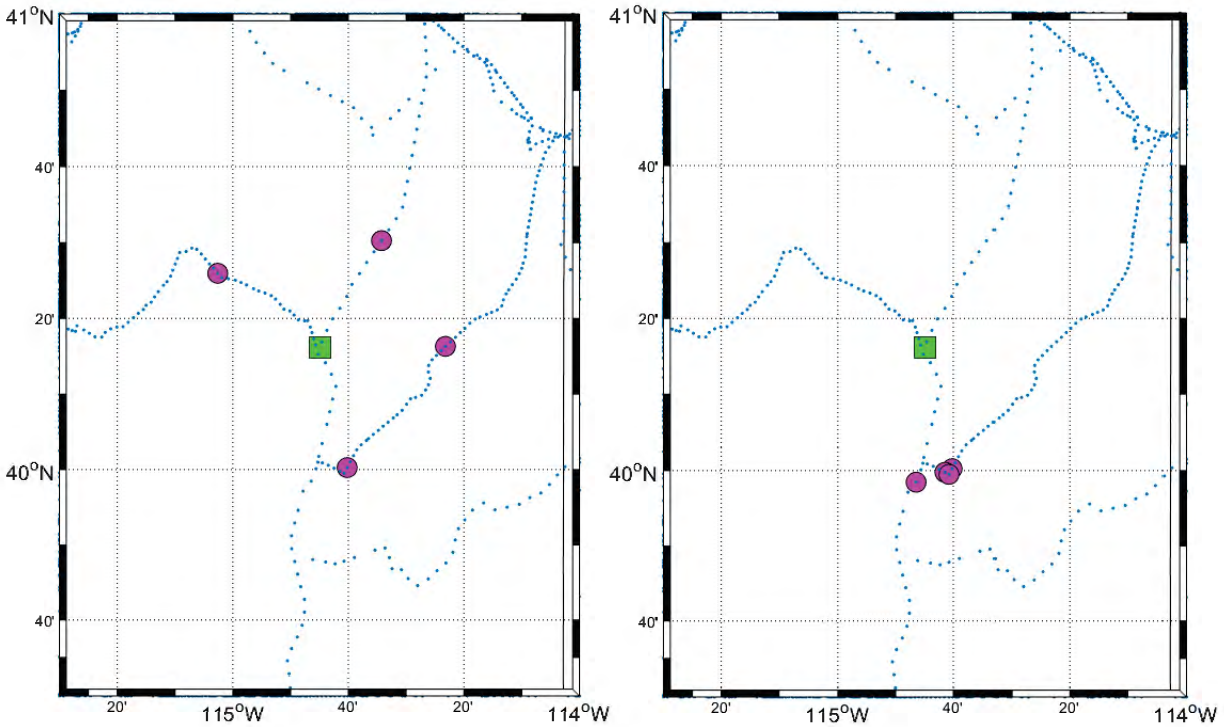

Figure 55: **a** (Left): Actual geographic distribution of bench marks around focal mark in green — all purple marks are at ~30 km with good geographic distribution. **b** (Right): Poor geographic distribution of purple marks at ~30 km distances. Both situations produce the same estimated uncertainty at focal mark. Blue marks are the actual locations of bench marks in this area.

The following situation describes how the estimated uncertainty at a particular location, ***P***, changes by artificially adding GPS on Bench Marks to the LSC estimation. Two scenarios are illustrated and described to assess their impact on the estimated uncertainty at ***P*** (σ_P): 1) A single bench mark is added at variable distances from 1200 m to 100 km and 2) a number of bench marks are added with the first additional GPS on Bench Mark being at a predetermined distance. While it is possible to highlight this with truly synthetic data, the scenarios are generated using actual NGS bench mark locations and their distribution. The goal of this is to illustrate how the uncertainty estimate can be lowered or improved by adding new observations.

7.2.2.1 Estimated Uncertainty Scenario 1

In the first scenario, the initial estimated uncertainty at ***P*** (σ_P) has a standard deviation of 2.8 cm (represented as the horizontal line in Figure 56 and Figure 57 below). This location is almost the furthest land location within CONUS from a GPS on Bench Mark at approximately 74 km, which is why the estimated uncertainty is so large. The resulting uncertainty estimate when a hypothetical observation is added to the LSC estimation at variable distances is shown in Figure 56. It is clearly evident how the uncertainty is impacted by the added observation's distance to ***P***. Adding the closest bench mark at

approximately 1.2 km as an observation causes the uncertainty to drop to 1.6 cm, whereas an added bench mark at 30 km causes an uncertainty of 2.55 cm.

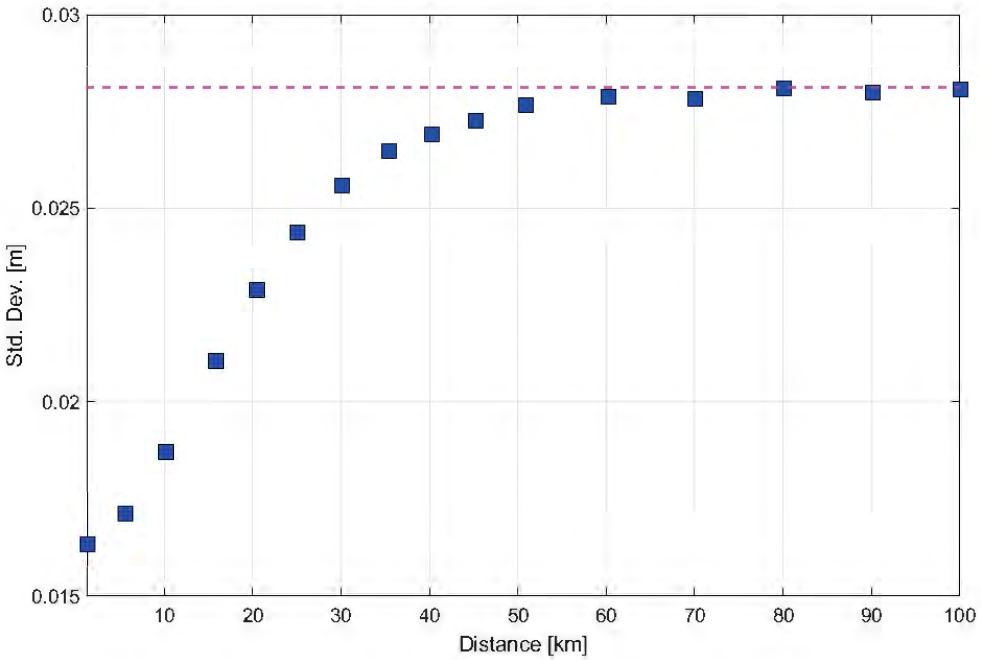

Figure 56: Estimated uncertainty as single GPS on Bench Marks are added at different distances.

7.2.2.2 Estimated Uncertainty Scenario 2

In the second scenario, the initial estimated uncertainty at ***P*** is still 2.8 cm, and multiple bench marks are added at variable initial distances sequentially to illustrate the impact of adding 2, 3, 4, … bench marks at a particular distance from ***P***. For example, if we sequentially add bench marks at an initial distance of 45 km from ***P***, we see that there is almost no added benefit (reflected by a lower standard deviation) with more than 2 additional bench marks (all the data points at 45 km are clustered at 2.6 – 2.7 cm). However, when we go to shorter initial distances like 20 km and 10 km, there is clearly a greater improvement by adding more bench marks. At distances shorter than 10 km, there is only very incremental improvement in the estimated uncertainty which can be visualized by how the curves are all very flat at distances below 10 km.

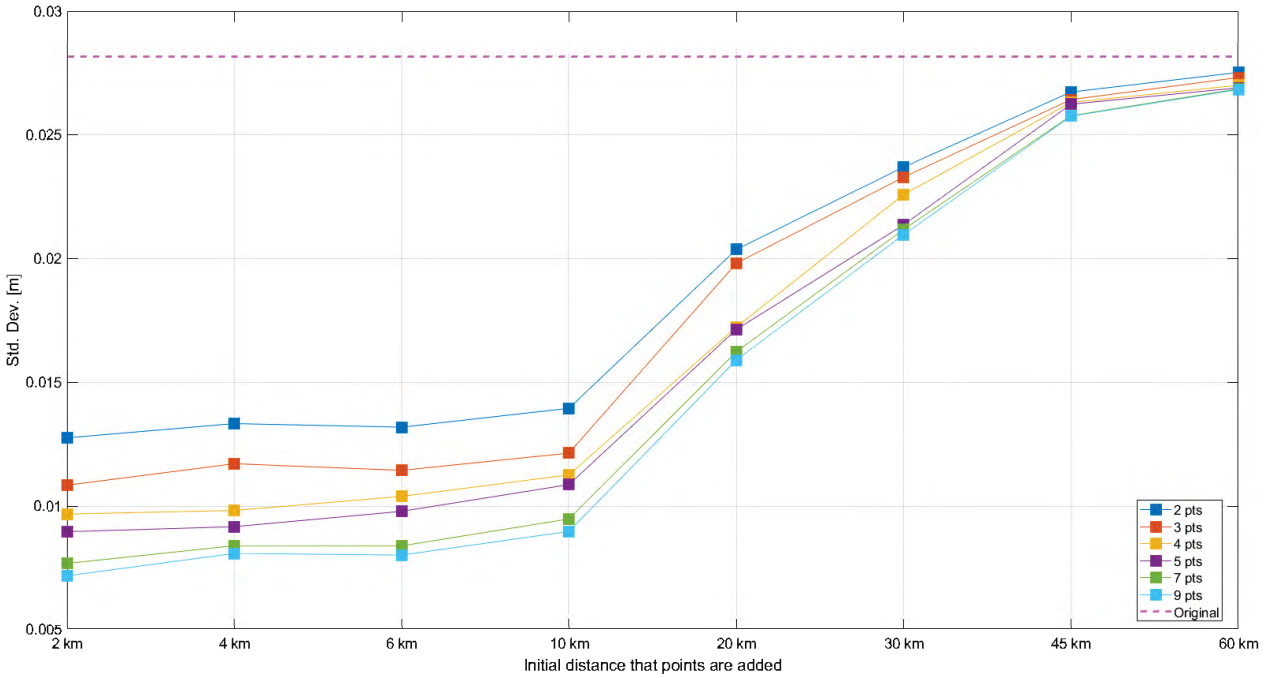

Figure 57: Improvement in the estimated uncertainty at P by adding multiple observations (bench marks) at variable initial distances. Blue curve shows how the uncertainty changes by adding 2 points, red curve = 3 points, …, up to 9 points.

8 Relative Accuracy

Many applications that utilize GEOID18 are more concerned with how well the model performs in a relative sense over different distances. This is the situation if we want to compare results based on GPS heights and a geoid model to geodetic leveling. This type of analysis has been applied to both formal error estimates and empirically derived height differences by Smith and Roman, 2001 and Brown, et. al. 2018, respectively. The methodology is also very similar to analysis performed for the NGS Geoid Slope Validation Surveys (Smith, et al, 2013; Wang, et al, 2017; and van Westrum, in press). For this analysis, all used GPS on BMs are compared against one other in terms of residual *differences*. For bench mark *i* and bench mark *j* that are separated by some distance (d_{ij}), the residual difference can be determined from (12):

$$\Delta r = r_j - r_i = (h_j - H_j - N_j) - (h_i - H_i - N_i) \;=\; (H_i - H_j) - [(h_i - N_i) - (h_j - N_j)] \qquad (12)$$

A residual difference is computed for every possible pair of used GPS on Bench Marks, which is 523,471,546 combinations for the CONUS GEOID18 dataset. The set of residual differences are then combined into 1 km distance bins based on the bench mark distance (d_{ij}). Statistics (e.g. mean, standard deviation, RMS) for residual differences (10) can then be computed for each 1 km bin. This type of relative accuracy can be performed with either the gravimetric geoid model or the hybrid model (i.e. pre-model vs. post-model residuals) and with different GPS on Bench Mark datasets (e.g. GEOID12B bench marks vs. GEOID18 bench marks). In the section below, the comparison for xGEOID19B and GEOID18 will be initially shown, then some comparisons with combinations from GEOID12B.

The number of combinations, mean, and RMS for each 1 km bin using GEOID18 is shown in Figure 58. The pre-model residual differences are with respect to the xGEOID19B and the post-model are with respect to GEOID18. For the RMS statistics, the lower the curve is — the better. Figure 58c specifically shows the overall and significant benefit of utilizing a hybrid geoid, which is warped to bench marks as it has an RMS < 2 cm over all distances compared with the gravimetric model that has > 5 cm RMS at 200+ km. However, at the very shortest distances (< 5 km), the gravimetric and hybrid geoids show very similar performance as illustrated in Figure 59.

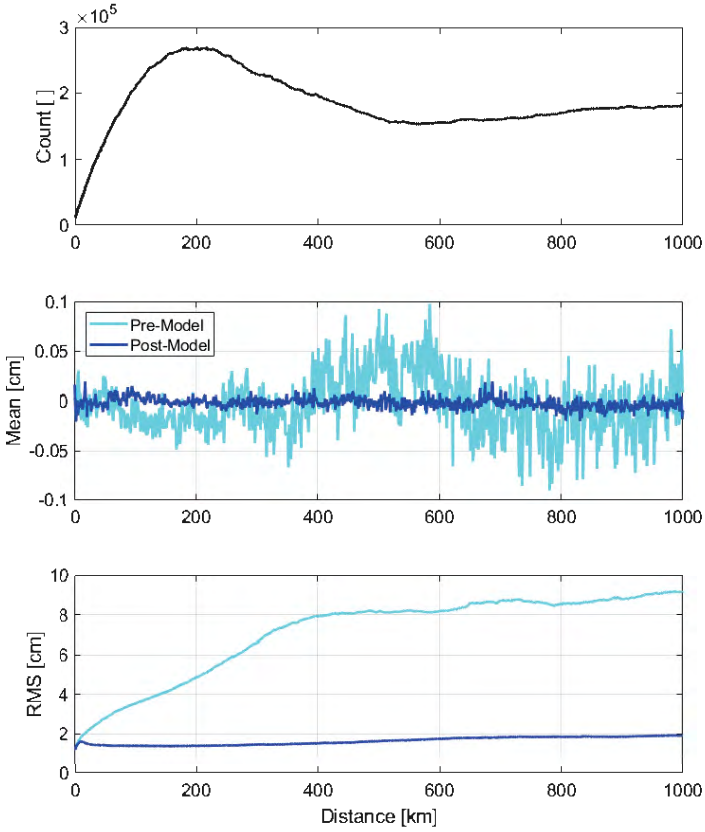

Figure 58: Relative accuracy for pre-model and post-model residuals for GEOID18. **a** (Top): number of combinations per 1 km bin; **b** (Middle): mean of relative differences; **c** (Bottom): RMS of relative differences

To get a sense of how the relative accuracy compares with geodetic leveling standards, the RMS is shown in Figure 59 compared to 3rd Order leveling standards (FGCC, 1984). This is meant to reiterate that GEOID18 is not a substitute for geodetic leveling as even 3rd Order leveling is more accurate at distances less than 50 km compared to GEOID18.

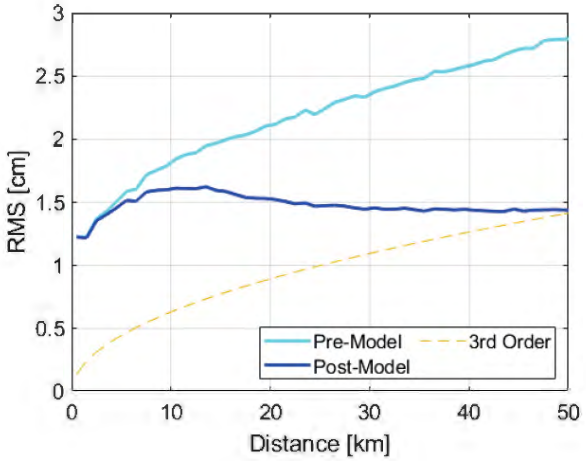

Figure 59: RMS for GEOID18 pre-model and post-model residuals at distances from 0 to 50 km. The 3rd order leveling standard is shown as $2.0\sqrt{k}$ in mm with k being the distance in km (FGCC, 1984).

The relative accuracy for both GEOID12B and GEOID18 using the used bench marks from each respective model is shown in Figure 60. Figure 60a shows the significant increase in combinations in GEOID18 driven by additional benchmarks. Since GEOID18 uses approximately 7,100 more bench marks in the model, the number of combinations is massively increased. The total number of combinations almost doubles from 317,331,028 in GEOID12B to 523,471,546 in GEOID18. The general shape of the number of combinations at different distances is fairly similar, but this inherent difference is likely driving some of the differences in the relative accuracy shown throughout this section. Figure 60b illustrates the mean of (12) for both GEOID12B and GEOID18 with both models showing almost 0 mean. Figure 60c illustrates the RMS for both models, which highlights the improvement with GEOID18 in a relative sense over all distances. In Figure 61, the relative RMS is shown for distances up to 50 km to highlight the

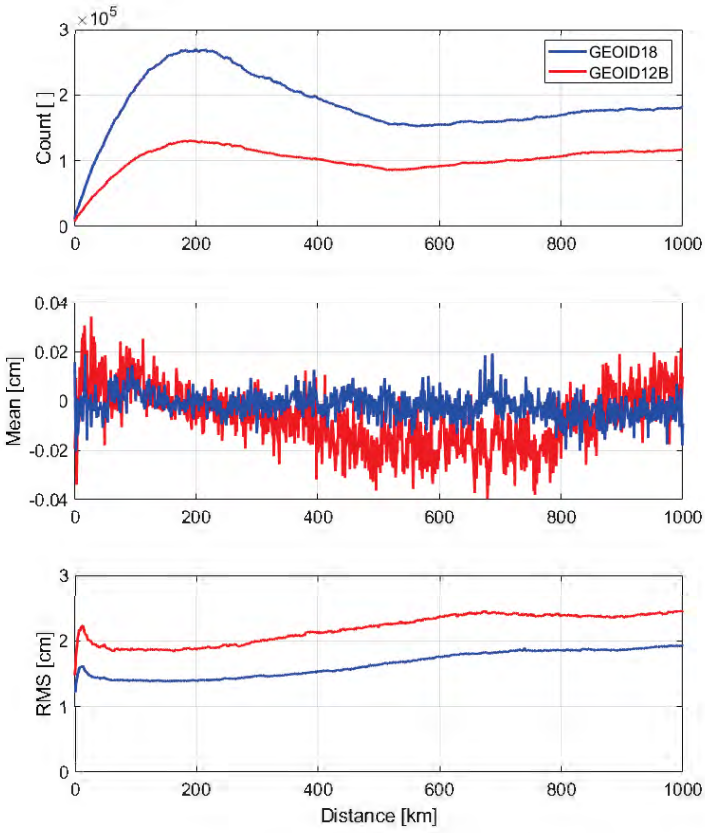

Figure 60: Comparison of GEOID12B and GEOID18 post-model residuals. **a** (Top): Count in each 1 km bin; **b** (Middle): Mean; **c** (Bottom): RMS.

details at these wavelengths along with geodetic leveling specifications for 3rd Order ($2.0\sqrt{k}$) and other leveling-like error propagation (FGCC, 1984).

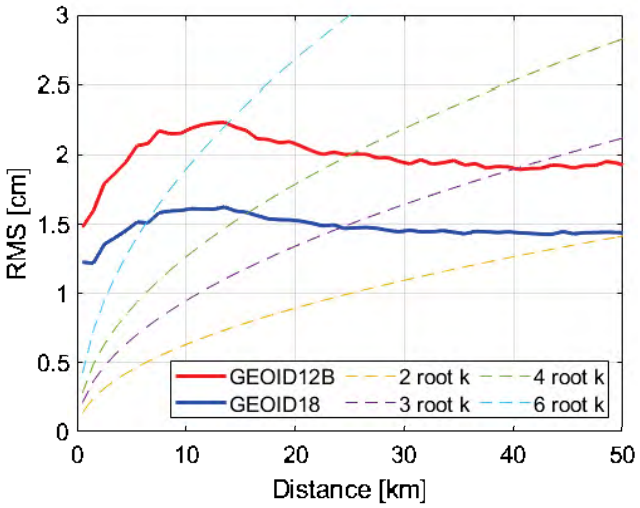

Figure 61: RMS of relative differences for GEOID12B and GEOID18 with leveling specifications: 2 root k is Third Order, 3 root k, 4 root k, and 6 root k are present to simply illustrate what magnitude of relative accuracy is likely at various distances.

From Figure 60 and Figure 61, it is pretty evident that a localized peak is present in the relative RMS in both the GEOID12B and GEOID18 results at approximately 15 km. This is caused by the choice of wavelengths used in the covariance model. GEOID18 and GEOID12B both use 30 km as the shortest wavelength, which implies that high frequency information below the half wavelength (i.e. 15 km) is not adequately captured. To illustrate this behavior at the shortest wavelengths, a prototype hybrid geoid model was constructed exactly the same as GEOID18 except it used a 15 km term in its covariance function. The relative RMS for this model and GEOID18 are shown in Figure 62, which clearly shows the shift in the localized peak from 15 km in GEOID18 to approximately 7.5 km in the prototype model.

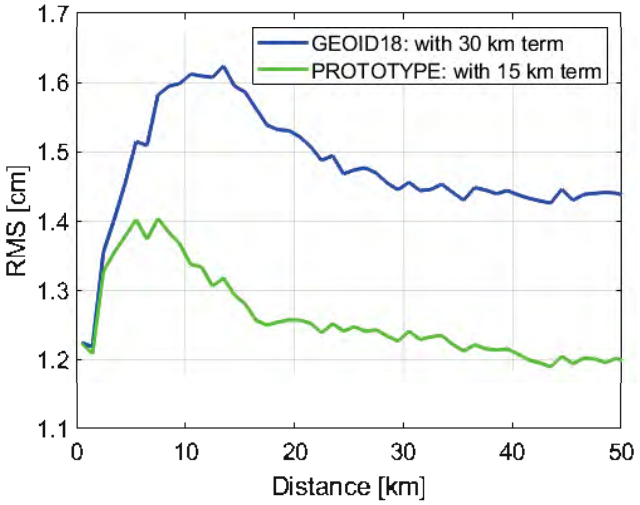

Figure 62: RMS of relative differences for GEOID18 (with 30 km shortest wavelength) and a prototype GEOID18-like model (with 15 km shortest wavelength). The local peak shifts from 15 km to 7.5 km when the higher frequency component is included.

Additionally, the pre-model residual is investigated in the following section to determine where improvements are originating from and at what distances. The relative accuracy for the pre-model residual for both GEOID12B and GEOID18 using the used bench marks from each respective model is shown in the first column of Figure 63 (top is to 1000 km and bottom is to 200 km). Overall all distances, GEOID18 shows a relative improvement compared to GEOID12B. In the second column of Figure 63, both sets of bench marks (GPSBMS12B and GPSBM18) are taken with respect to xGEOID19B and analyzed relatively. This isolates all differences to be only caused by differences in the bench mark datasets. In the third column of Figure 63, the GPSBM18 dataset is taken with respect to USGG2012 and xGEOID19B to isolate any differences caused by the gravimetric geoid models. From these figures, a number of inferences can be made: 1) the GPS on Bench Marks data in GEOID18 provide improvement over all distances; 2) below 350 km, both USGG2012 and xGEOID19B are fairly consistent with USGG2012 being very slightly better than xGEOID19B between 50 and 150 km; 3) at distances over 400 km, xGEOID19B performs better than USGG2012 likely due to improvements in satellite gravity from GRACE and GOCE; and 4) the most impact is driven by the GPS on Bench Marks dataset improvement.

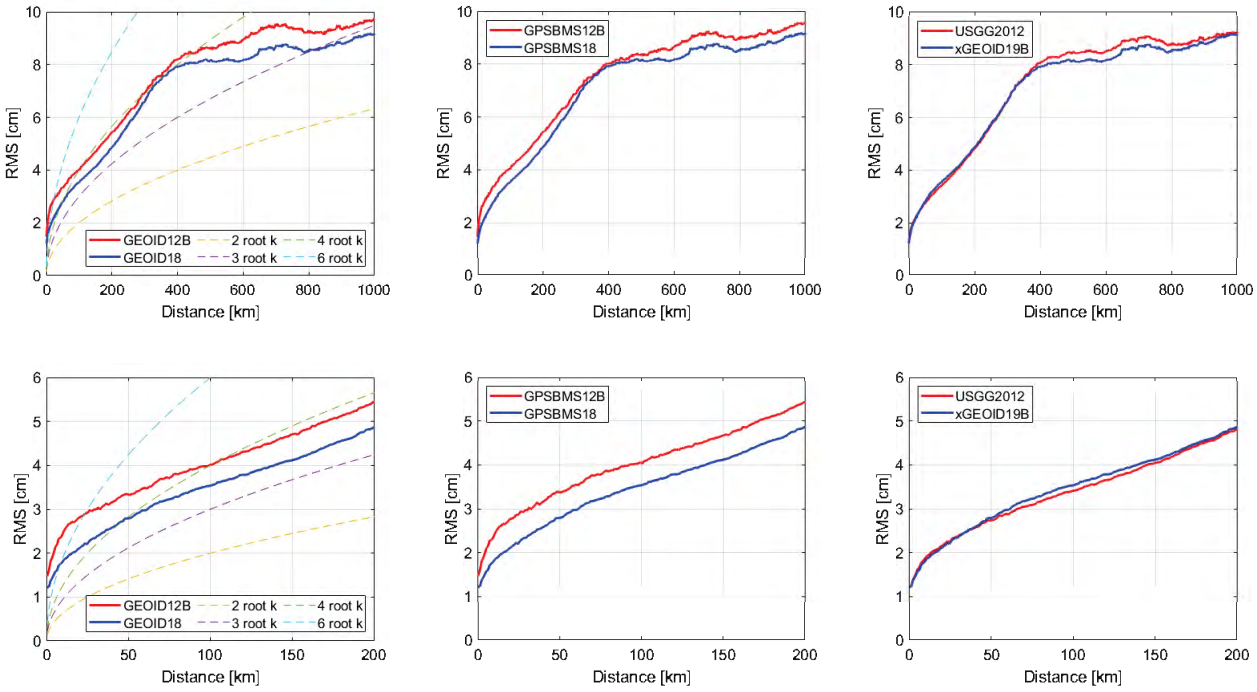

Figure 63: Relative accuracy for GEOID12B and GEOID18. First column is the overall difference. Middle column is the difference caused by the GPS on Bench Marks data. Right column is the difference caused by the gravimetric geoid. Top row has distance out to 1000 km. Bottom row out to 200 km. The blue curve is the same on all figures.

9 DEFLEC18

A derived product from the hybrid geoid model is a deflection of the vertical (DOV) model for both the north-south component (ξ or xi) and the east-west component (η or eta). The DEFLEC18 model provides angular differences between the normal to the hybrid geoid surface and the normal to the ellipsoid at the Earth's surface. These are based on the Helmert definition for deflections of the vertical with the important caveat that the hybrid geoid surface is not a "true" equipotential surface (i.e. a true gravimetric geoid like xGEOID19B). These can be used in the following equations to convert astronomic observations and geometric observations on the Earth's surface.

$$\begin{aligned} \xi &= \Phi - \varphi \\ \eta &= (\Lambda - \lambda) * cos\varphi \end{aligned} \tag{13}$$

where: Φ= astronomic latitude, Λ= astronomic longitude, φ= geometric latitude, λ= geometric longitude.

The creation of this hybrid deflection of the vertical model requires three 1' grids: GEOID18, Bouguer gravity, and a DEM. The Bouguer gravity grid is based on the simple plate-based correction (or Helmert definition) (for more information, see Hinze, et al. 2005) and derived from the xGEOID19B gravity grid for consistency with the gravimetric model. Additionally, the DEM is that which is used in xGEOID19B for consistency and based on SRTMv4 (Jarvis, et al. 2008). The Bouguer gravity grid and the DEM are shown in Figure 64 and Figure 65 for CONUS and Figure 66 and Figure 67 for PRVI.

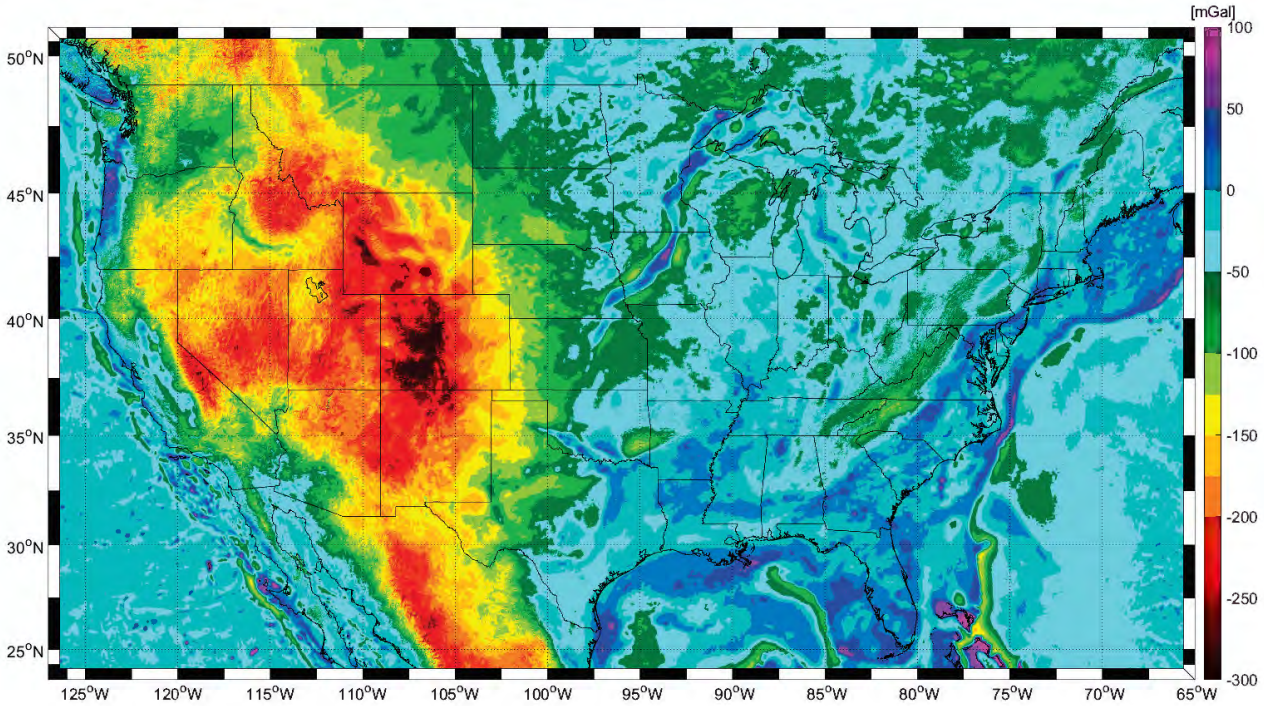

Figure 64: Bouguer gravity grid used in DEFLEC18.

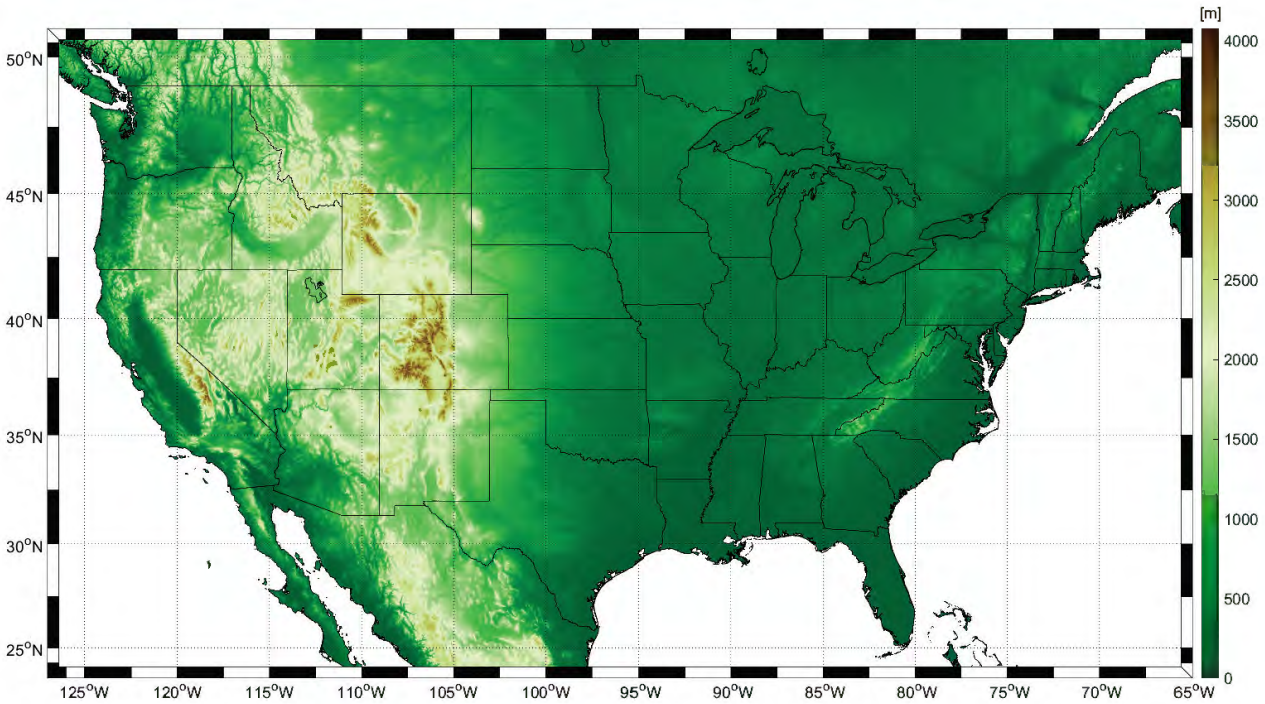

Figure 65: DEM grid used in DEFLEC18 based on xGEOID19B DEM model.

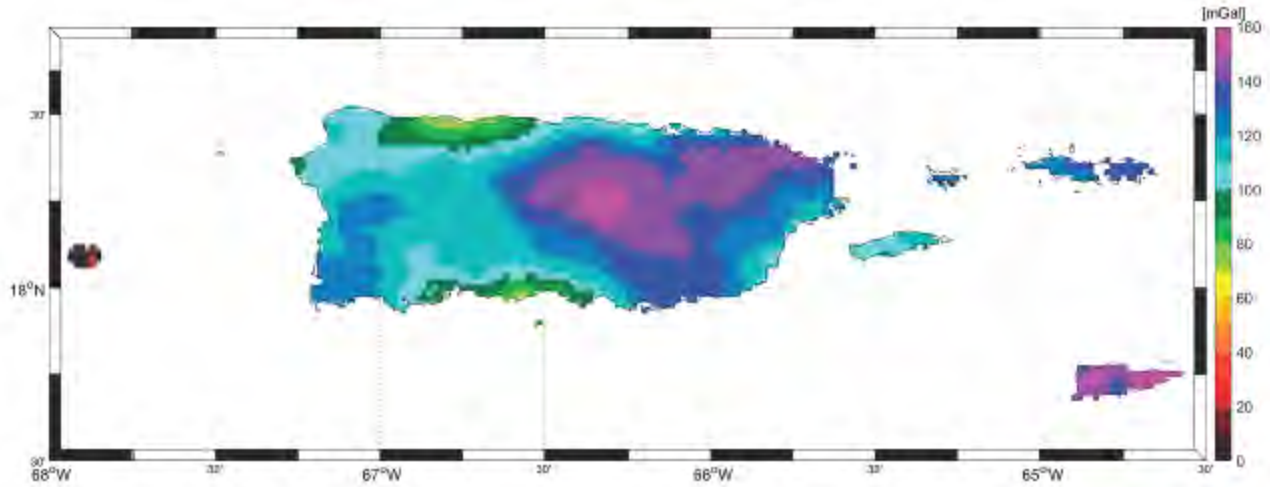

Figure 66: Bouguer gravity grid used in DEFLEC18 for PRVI.

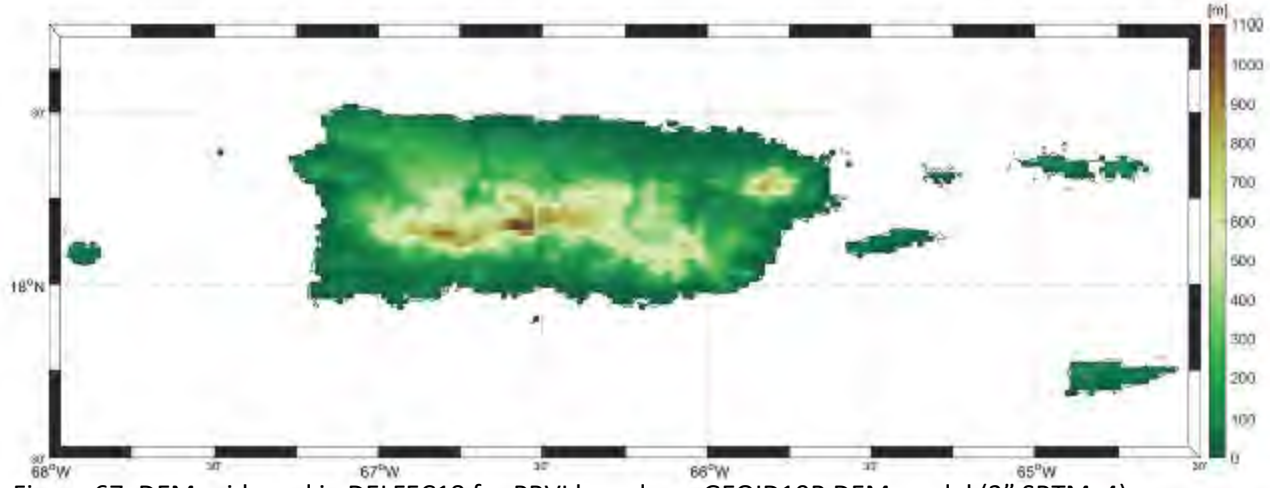

Figure 67: DEM grid used in DELFEC18 for PRVI based on xGEOID19B DEM model (3" SRTMv4).

9.1 Methodology:

The deflections are computed in each direction (north-south and east-west) by taking the derivative of the geoid surface along that direction as shown in (14) (Heiskanen and Moritz, 1967, eq. 2-204).

$$\xi_{geoid} = -\frac{1}{R}\frac{\partial N}{\partial \varphi}$$
$$\eta_{geoid} = -\frac{1}{R\cos\varphi}\frac{\partial N}{\partial \lambda} \quad (14)$$

where:
R = mean earth radius

This derivative is computed using a 5-point cubic spline over the 1 arc-minute resolution GEOID18 grid after Smith & Roman, (2001). These deflections are associated with the geoid though and NOT the Earth's surface so a correction term specified in (15) is used to correct for the curvature of the plumbline from the geoid to the Earth's surface (Heiskanen and Moritz, 1967, eq. 5-32) where the x-axis points positive northwards and the y-axis points positive eastward.

$$\delta\xi = \frac{H}{\bar{g}}\left(\frac{d\bar{g}}{dx} + 0.0424\frac{dH}{dx}\right)$$
$$\delta\eta = \frac{H}{\bar{g}}\left(\frac{d\bar{g}}{dy} + 0.0424\frac{dH}{dy}\right) \quad (15)$$

where:
$\bar{g} = g + 0.0424 * H$

These two correction terms are similar to the Helmert orthometric correction used in leveling and are illustrated in Figure 68 and Figure 69 for ξ and η in CONUS and Figure 70 and Figure 71 for PRVI.

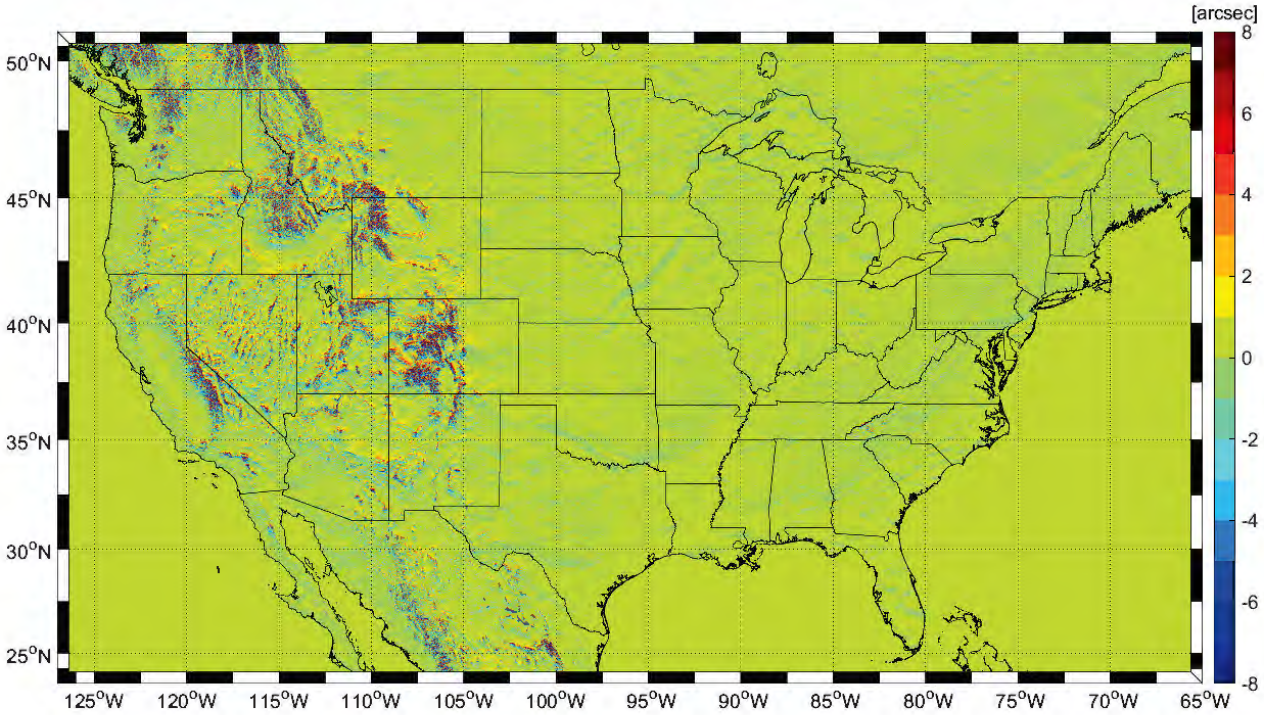

Figure 68: Plumbline correction component in ξ(N-S) direction.

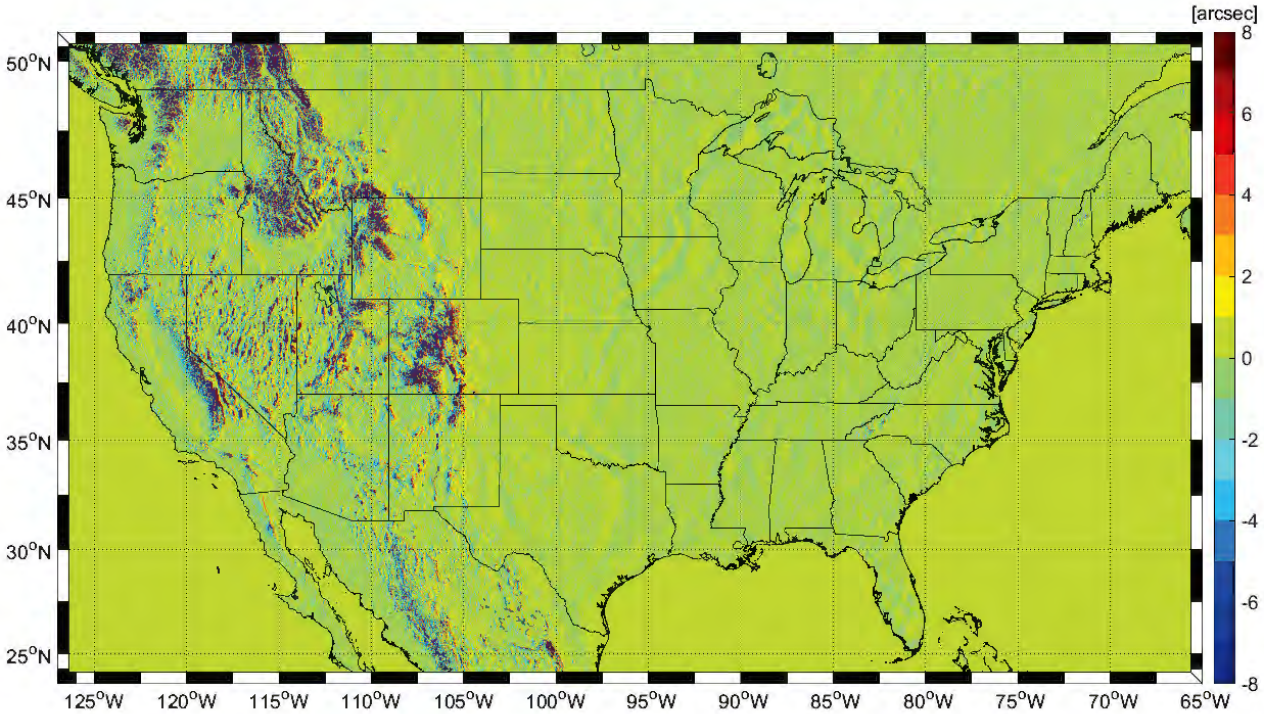

Figure 69: Plumbline correction component in η(E-W) direction.

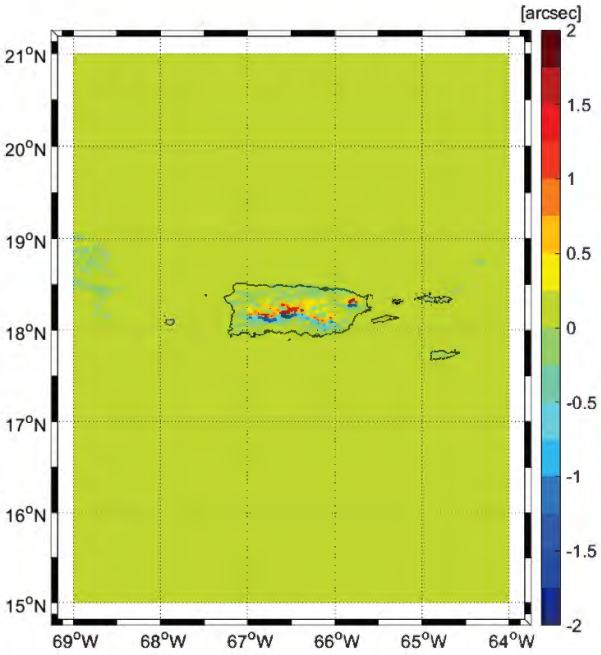

Figure 70: PRVI Plumbline correction component in ξ(N-S) direction.

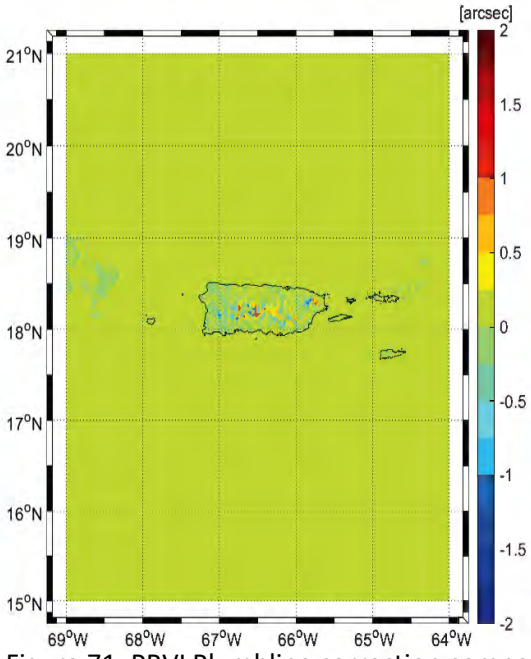

Figure 71: PRVI Plumbline correction component in η(E-W) direction.

The plumbline correction terms are then added to ξ_{geoid} and η_{geoid} resulting in the deflection of the vertical components on the Earth's surface and the ultimate DEFLEC18 model values.

$$\begin{aligned} \xi &= \xi_{geoid} + \delta\xi \\ \eta &= \eta_{geoid} + \delta\eta \end{aligned} \tag{16}$$

9.2 Results

The DEFLEC18 model for CONUS is shown in Figure 72 and Figure 73 for ξ and η, respectively. PRVI models are shown in Figure 74 and Figure 75. Statistics for the deflection of the vertical components are also shown in Table 18.

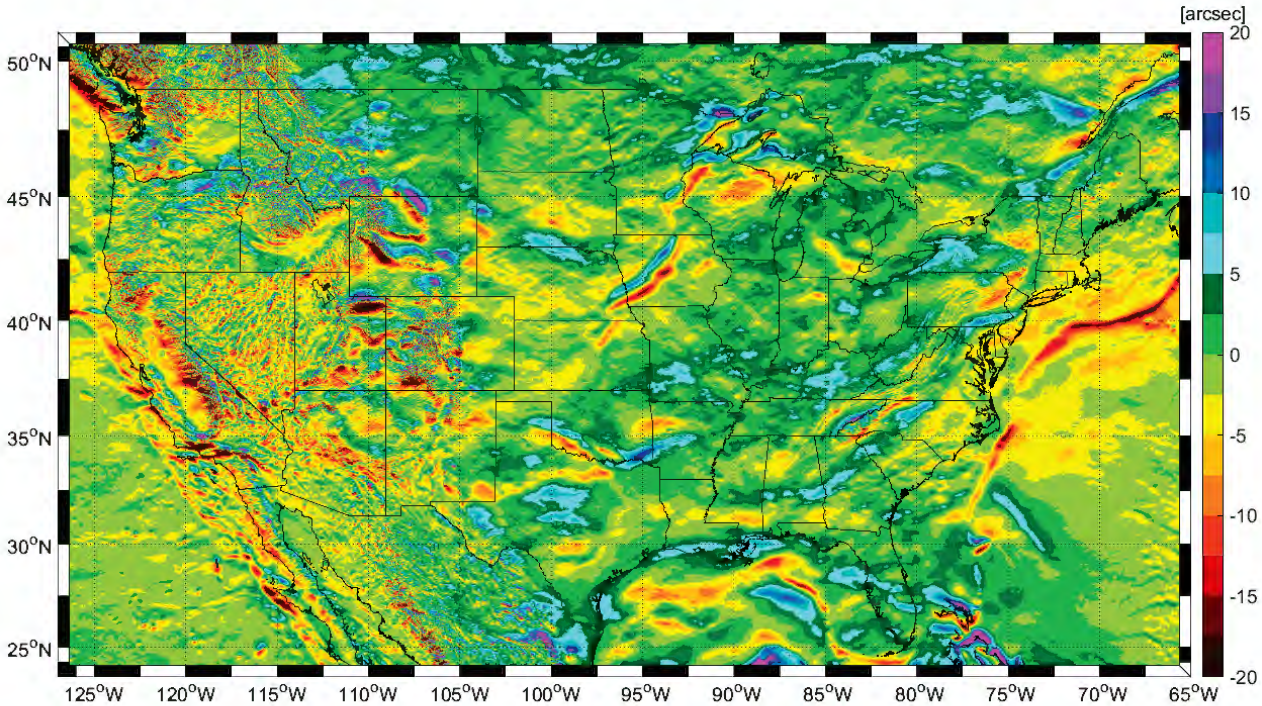

Figure 72: North-south component of DEFLEC18. Note: Colorbar is not the full range of values.

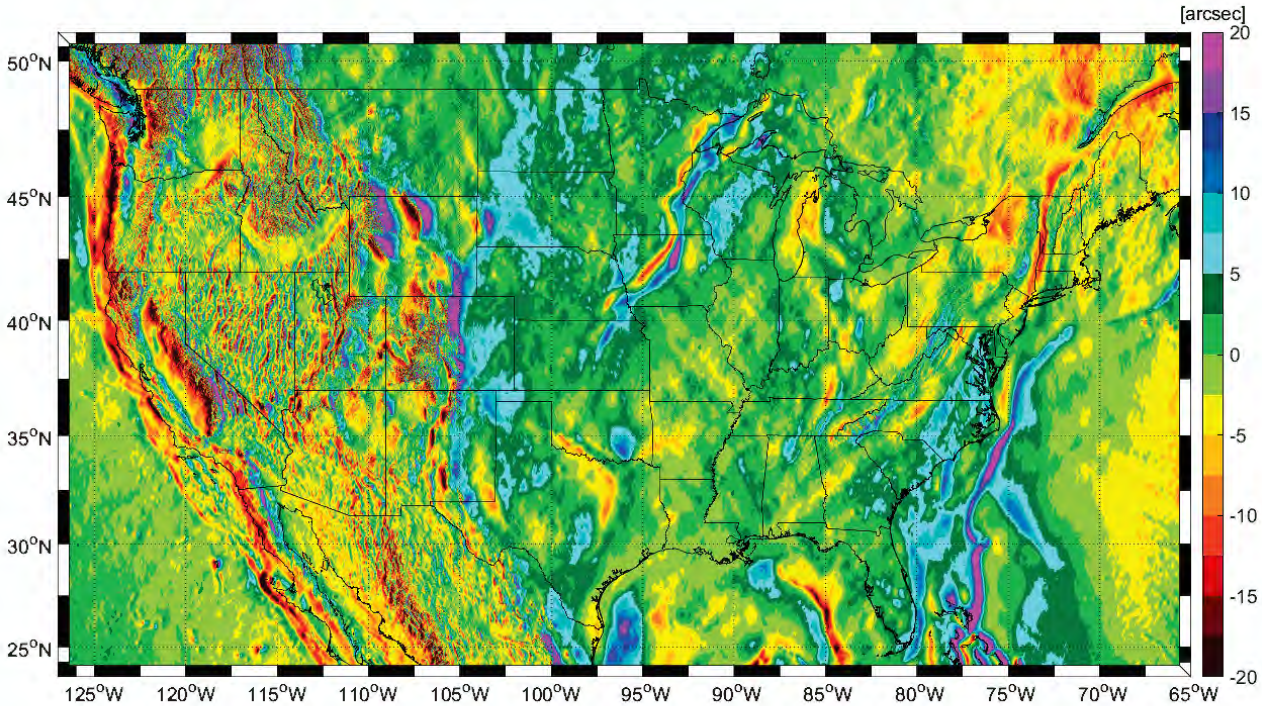

Figure 73: East-west component of DEFLEC18. Note: Colorbar is not the full range of values.

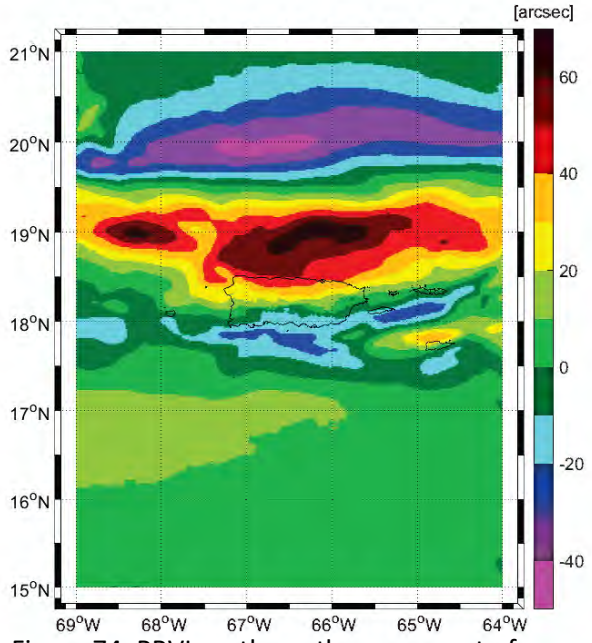

Figure 74: PRVI north-south component of DEFLEC18.

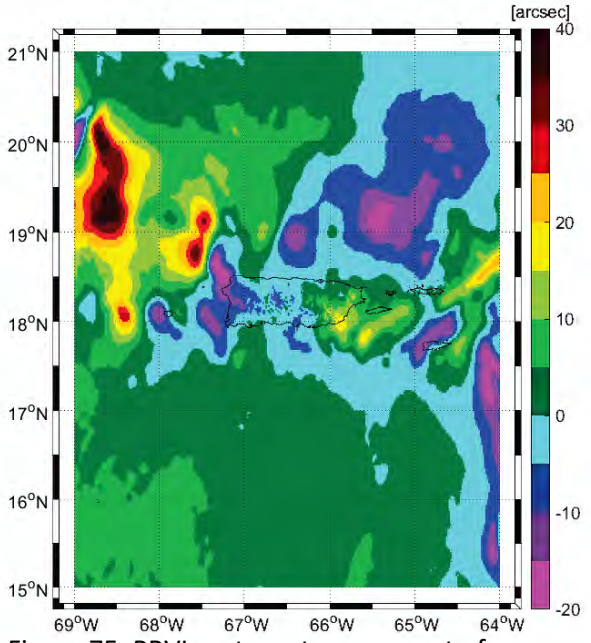

Figure 75: PRVI east-west component of DEFLEC18.

Table 18: Statistics for DEFLEC18 1' grids. Units = arc-seconds.

CONUS	Min.	Max.	Mean	Std. Dev.
ξ (XI)	-52.0827	76.8717	-0.16585	4.47812
η (ETA)	-74.27598	68.42022	-0.238515	5.45124
ξ Topo Corr.	-49.2779	69.9094	0.0880142	1.214090
η Topo. Corr.	-66.6992	52.0124	-0.0131939	1.61085
PRVI	Min.	Max.	Mean	Std. Dev.
ξ (XI)	-44.08111	66.1245	4.5866	20.57322
η (ETA)	-20.30841	40.535671	2.28550	7.54032
ξ Topo Corr.	-2.75147	2.80977	0.0010360	0.066477
η Topo Corr.	-1.63891	2.02519	0.000277	0.042397

9.3 Comparison with Observed Deflections of the Vertical

NGS has a very small number of observed deflections of the vertical that have been used for evaluation purposes for geoid models and deflection of the vertical models. Additionally, the three GSVS lines in Texas, Iowa, and Colorado have observed high-quality deflections that can also be used to evaluate DEFLEC18. At the time of publication, the GSVS17 deflection of the vertical observations have not been publically released, so only the GSVS11 and GSVS14 lines are shown for comparison purposes. Due to the very small number of observed deflections of the vertical in Puerto Rico and U.S. Virgin Islands, those areas are not considered in this comparison.

NGS also has additional deflection of the vertical models which are provided as BETA versions that are designed to be consistent with the xGEOID model series. These BETA deflections or xDEFLEC models can

be found at the following link: https://beta.ngs.noaa.gov/GEOID/xDEFLEC19/index.shtml. The most recent xDEFLEC19B will be used as a comparison throughout this section.

The historical deflection of the vertical dataset is shown in Figure 76 and Figure 77 as residuals between the modeled DEFLEC18 value and the observed value. There are approximately 3,400 point observations of astronomic coordinates (Smith and Roman, 2001) that can be combined with geometric coordinates to determine the deflection of the vertical components. These observations are all many decades old so very little additional information is known about them. The DEFLEC18 model values agree at 1.14" RMS in ξ and 1.18" RMS in η with respect to the observed deflections. Additional statistics can be found in Table 19. DEFLEC18 exhibits slightly worse performance with respect to this historical DOV dataset than DEFLEC12B with the only truly significant difference being a minor increase of 0.14" in the η RMS. Both DEFLEC12B and DEFLEC18 have similar means in both components and similar ξ RMS terms.

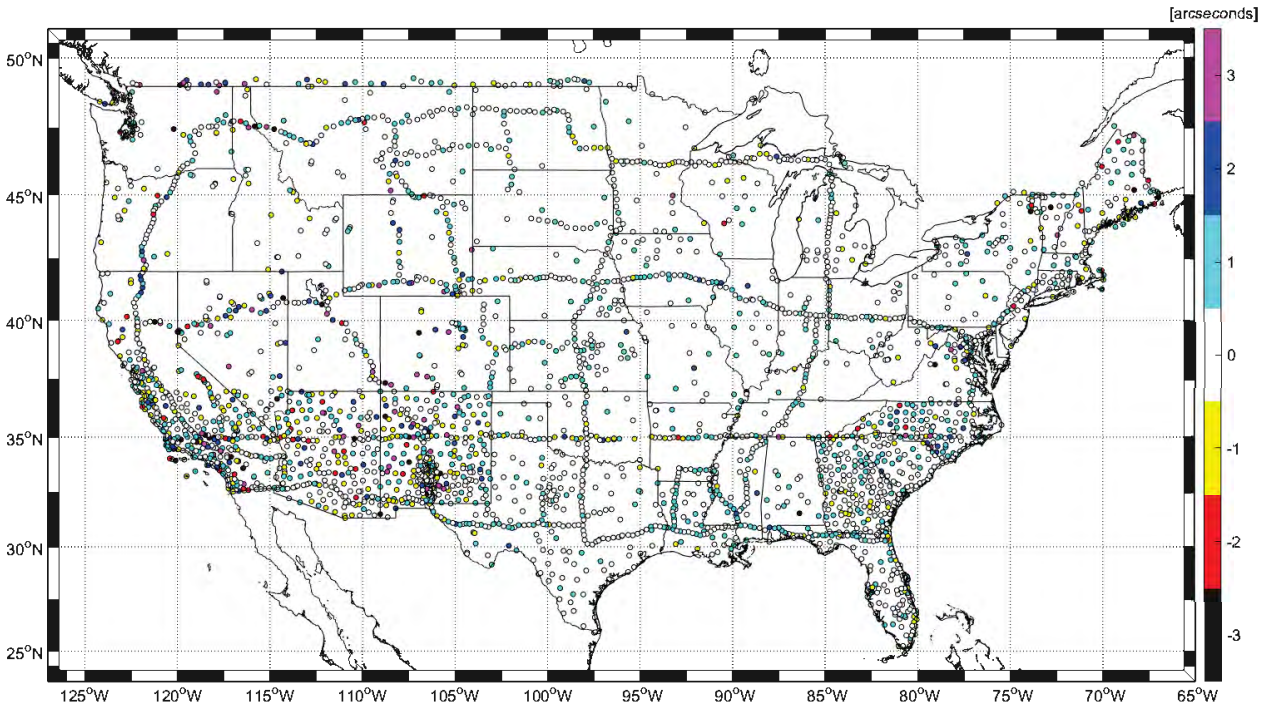

Figure 76: ξ component residual with respect to DEFLEC18.

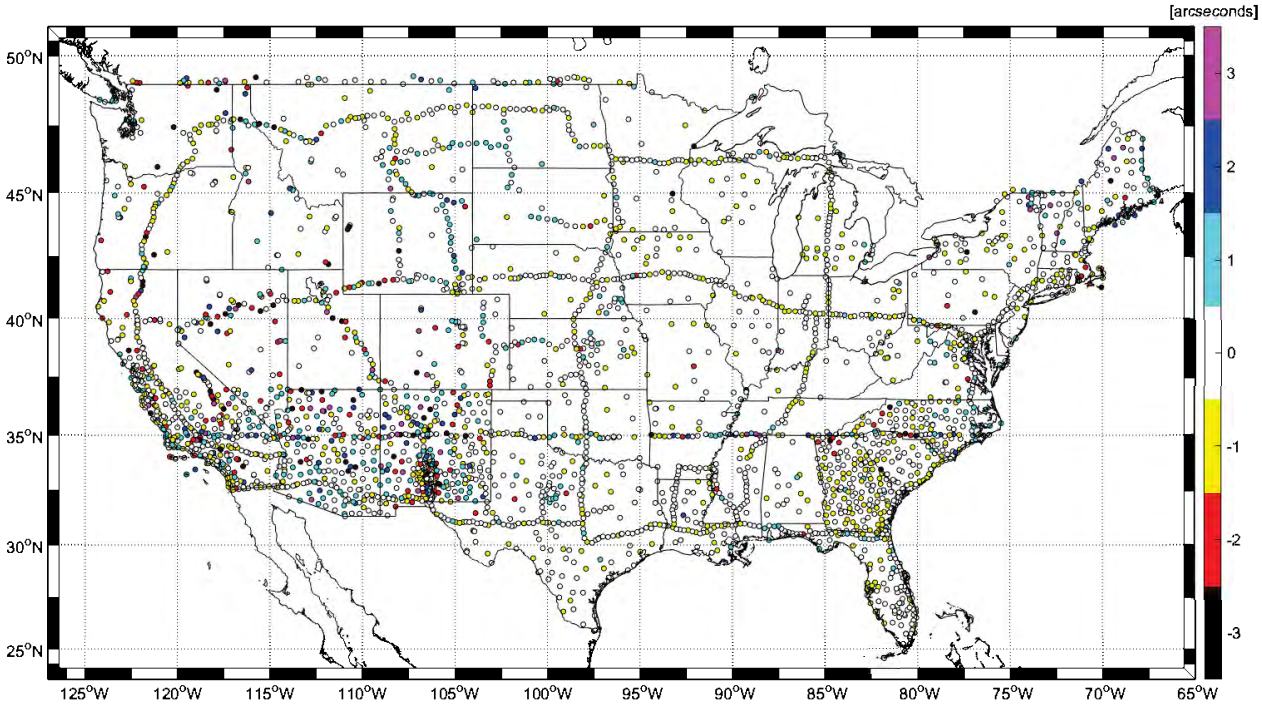

Figure 77: η component residual with respect to DEFLEC18.

Table 19: Statistics of residuals for deflection of the vertical components with respect to DEFLEC12B, DEFLEC18, and xDEFLEC19B. Units in arc-seconds.

Dataset	Model	# pts.	ξ (XI)					η (ETA)				
			Min	Max	Mean	StdDev	RMS	Min	Max	Mean	StdDev	RMS
Historical DoVs	DEFLEC12B	3398	-20.892	12.172	0.114	1.095	1.101	-11.929	8.277	-0.174	1.042	1.056
	DEFLEC18		-20.861	13.653	0.178	1.128	1.141	-19.665	8.867	-0.203	1.166	1.183
	xDEFLEC19B		-20.929	13.199	0.013	1.053	1.053	-13.949	9.12	-0.094	1.011	1.016
GSVS11	DEFLEC12B	218	-0.706	1.276	0.046	0.246	0.249	-0.356	0.614	0.141	0.166	0.217
	DEFLEC18		-0.649	1.119	0.028	0.189	0.191	-0.35	0.657	0.055	0.150	0.160
	xDEFLEC19B		-0.593	1.027	-0.056	0.168	0.177	-0.406	0.690	0.053	0.156	0.165
GSVS14	DEFLEC12B	204	-0.578	1.019	0.034	0.326	0.327	-0.723	0.82	0.018	0.289	0.289
	DEFLEC18		-0.528	0.907	0.033	0.303	0.304	-0.638	0.805	-0.008	0.253	0.253
	xDEFLEC19B		-0.598	0.882	-0.035	0.310	0.311	-0.574	0.858	0.053	0.258	0.263

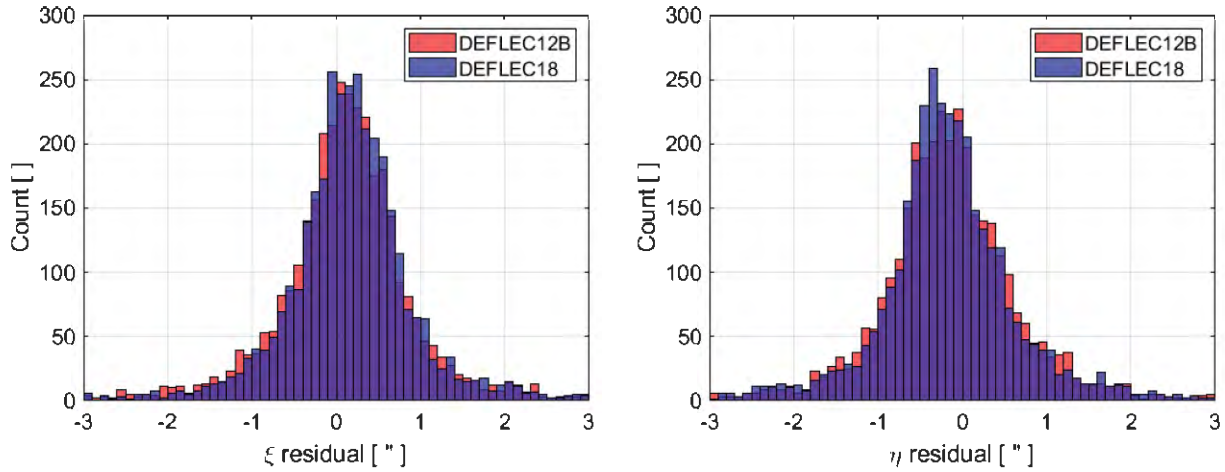

Figure 78: Deflection of the vertical component residuals for DEFLEC12B and DEFLEC18. (model – observed). **a** (left) ξ (north-south) component. **b** (right) η (east-west) component. Units: [arcseconds].

Due to the unknown nature of the observed deflection of the vertical data, it is very likely that some of these observations are subject to systematic errors and/or blunders. For this reason, a very soft outlier detection is performed with respect to each of the residual components using Grubbs method (Grubbs, 1950). The lightest of detection is performed here by employing Grubbs test at the 1% significance level (compared to 5% or 10%, which would both detect more outliers) for each DOV component as shown in Figure 79. After employing the outlier tests which flag approximately 1.5% of the observation residuals as outliers, the remaining residuals have an RMS of 0.8” in ξ and 0.9” in η. Similarly, the xDEFLEC19B *sans-outlier* residuals have an RMS of 0.75” in ξ and 0.8” in η.

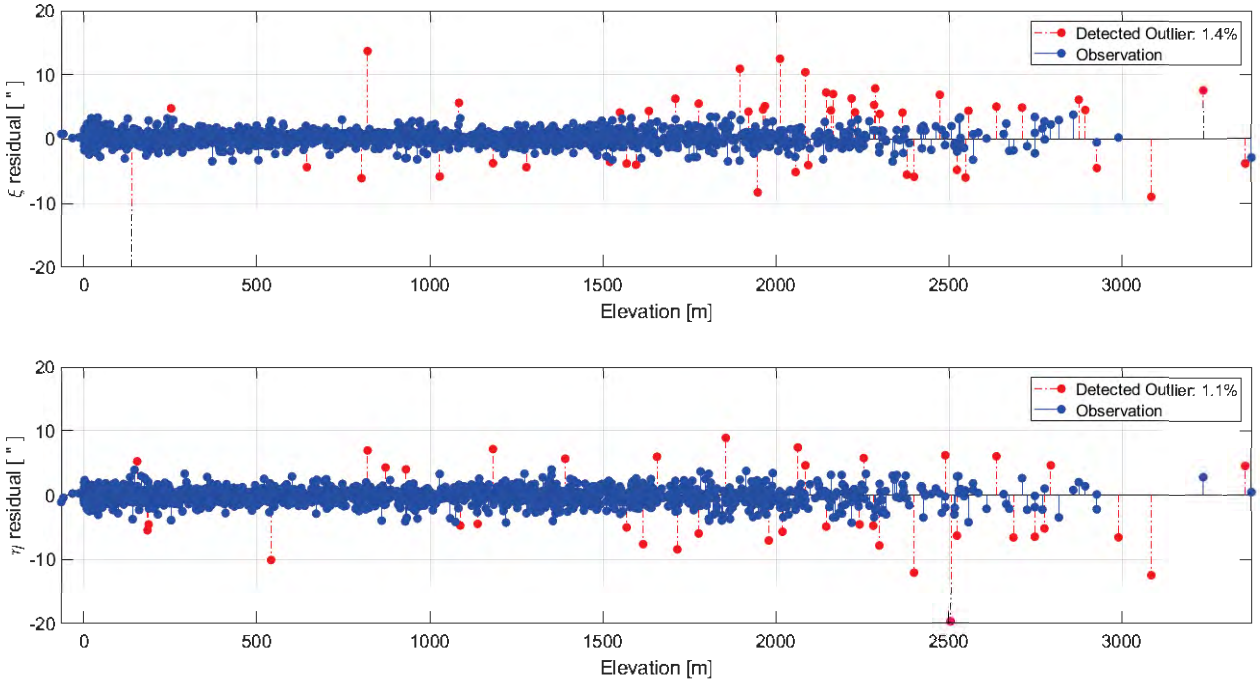

Figure 79: Deflection of the vertical residual (DEFLEC18 model – observed) components with respect to elevation. Flagged outliers are shown in red based on Grubbs test at 99% significance level. **a** (top) ξ

component which flags 48 out of 3398 (1.4%) residuals as outliers. **b** (bottom) η component which flags 38 out of 3398 (1.1%) residuals as outliers.

There is clear relationship with elevation in the DOV residuals with observations at higher elevations exhibiting more noise compared with those at lower elevations. This is reinforced in Figure 80, which shows the percentage of observations that are flagged as outliers over different elevation ranges. There are definitely fewer observations at higher elevations; however, there is clearly more discrepancy as the elevation increases. This is likely due to two factors: 1) the lack of resolution in a 1 arc-minute geoid model to capture high frequency DOV content that is likely present in mountainous terrain; and 2) the inexactness in the plumbline correction term which is based on Helmert's mean gravity along the plumbline, which neglects any residual terrain content.

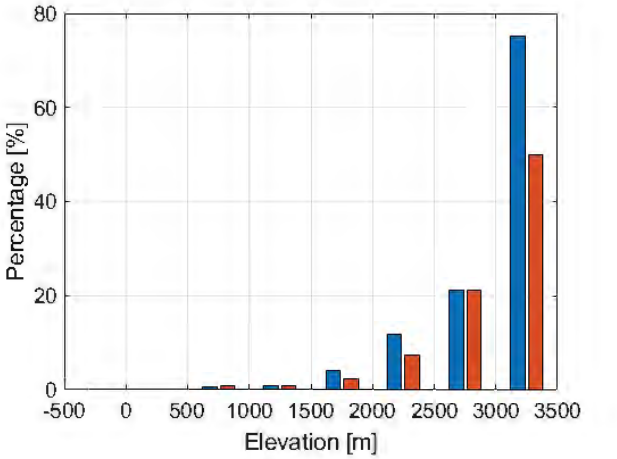

Figure 80: Percentage of observations flagged as outliers with respect to elevation. Blue: ξ component. Red: η component.

The GSVS11 deflection component residuals are shown in Figure 81. Both components are within approximately 0.2 arc-seconds RMS of the DEFLEC18 model, which is just marginally better than the DEFLEC12B model over this flat region. In Figure 82, the GSVS14 deflection component residuals are shown and are slightly larger than GSVS11, but they still show very good agreement at 0.3 arc-seconds RMS with respect to DEFLEC18. Again, a very minor improvement occurs from DEFLEC12B to DEFLEC18. The statistics for these residuals are shown in Table 19.

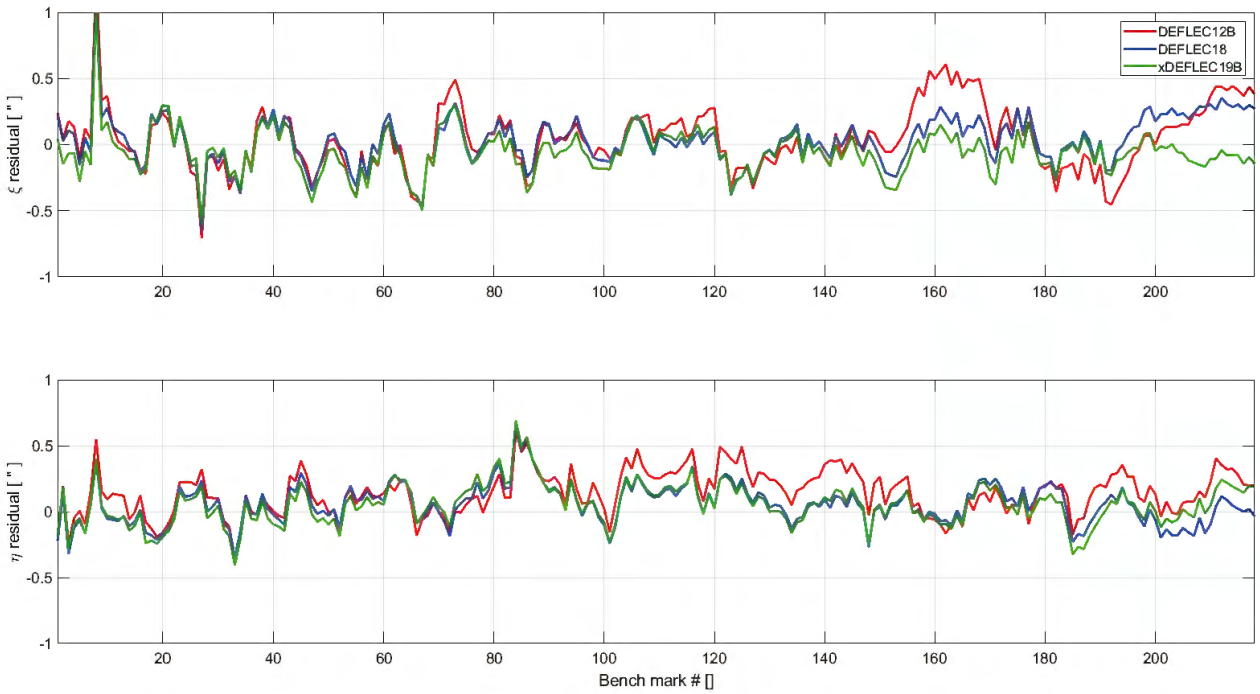

Figure 81: GSVS11 deflection of the vertical component residuals (modeled — observed). Line runs north to south from Austin, Texas, to Corpus Christi, Texas.

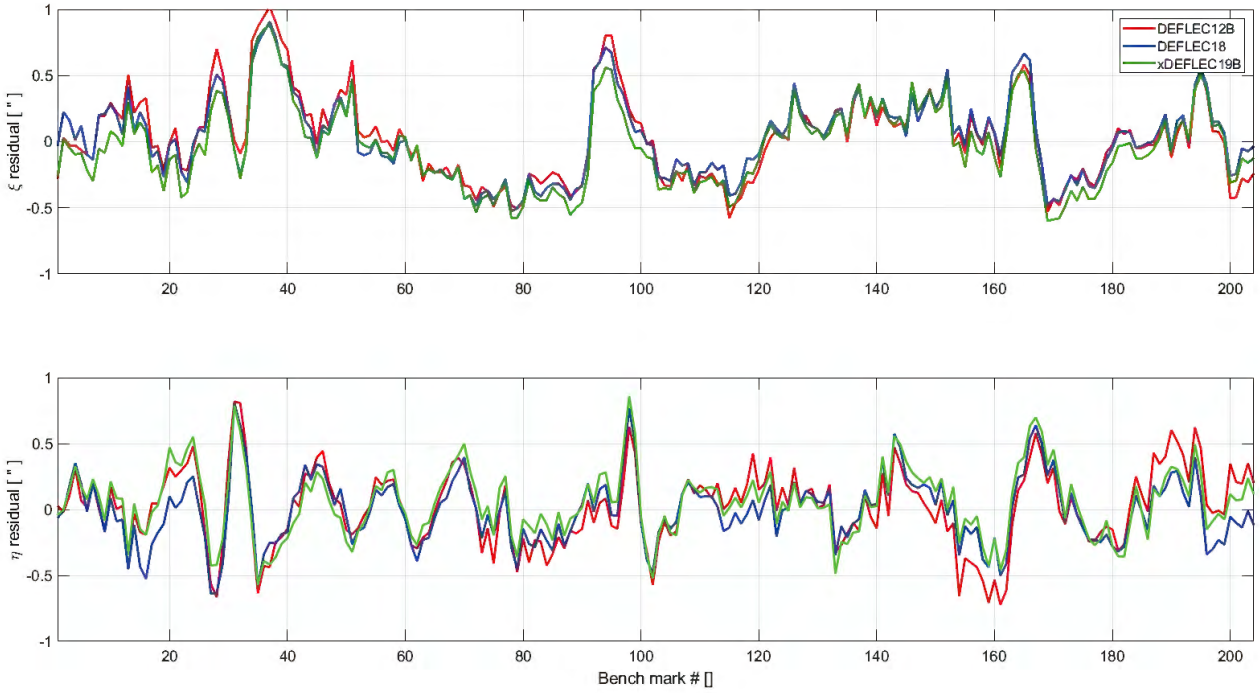

Figure 82: GSVS14 deflection of the vertical component residuals (modeled — observed). Line runs west to east from Carroll, Iowa to Cedar Rapids, Iowa.

The xDEFLEC19B model provides a more consistent agreement with the observed datasets used in this section than DEFLEC18. This is somewhat expected when one considers what warping a gravimetric geoid like xGEOID19B with GPS on Bench Marks data does to the deflections of the vertical. A deflection of the vertical observation taken on the Earth's surface measures the slope of the geoid surface based

completely on the gravitational field. If we now warp that geoid surface with GPS on Bench Marks data, the deflection of the vertical observation does not sense that artificial warped component whatsoever.

This is very much reflected in the historical deflections of the vertical (see Table 19) where the standard deviation is slightly worse for DEFLEC18 compared to xDEFLEC19B for ξ (1.128 versus 1.053) and η (1.166 versus 1.011). This slight degradation is also illustrated in Figure 83, which shows the overall worsening in both ξ and η with respect to elevation. The DEFLEC18 ξ and η components have slightly larger standard deviations as the elevation increases compared to the xDEFLEC19B components.

Over the GSVS11 and GSVS14 lines, both the DEFLEC18 and xDEFLEC19B models perform at almost the same level of accuracy with both models having approximately 0.2 to 0.3 arc-seconds RMS for both DOV components. This is somewhat surprising since one would expect the xDEFLEC19B model to more accurately reflect the true gravitational field and observed deflections.

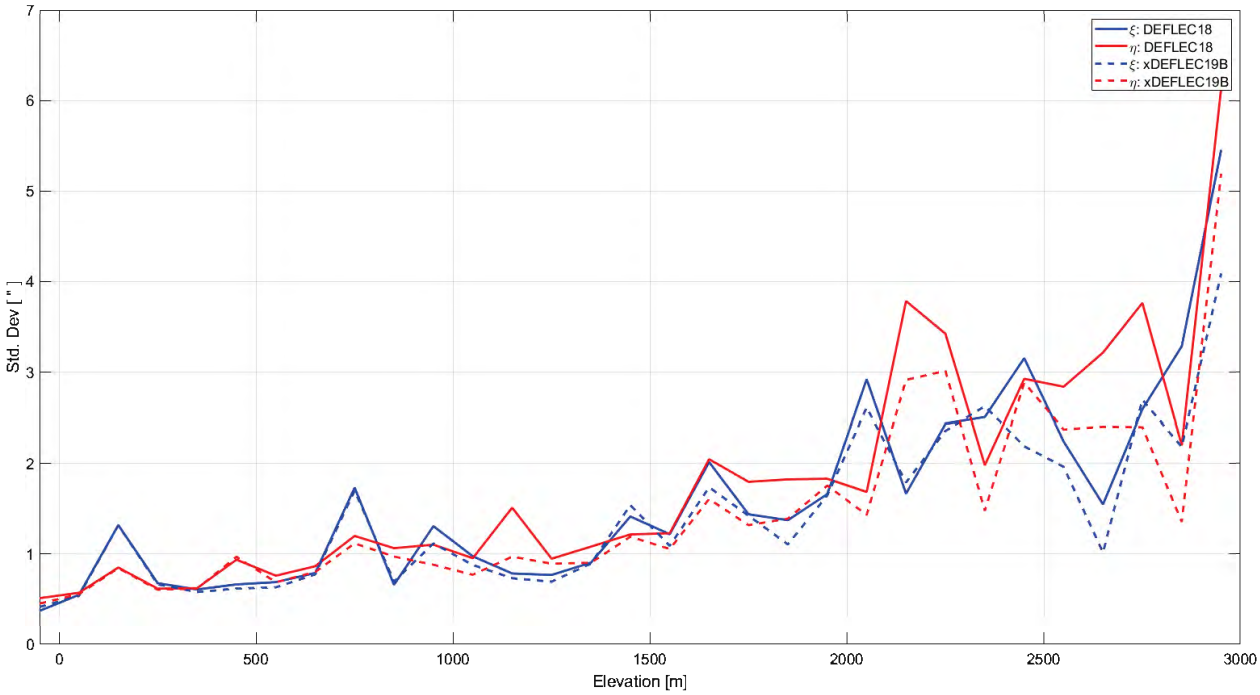

Figure 83: Standard deviation of the ξ and η residuals at different elevations for DEFLEC18 and xDEFLEC19B

10 Conclusions

For many years, NGS has been developing gravimetric and hybrid geoid models that continue to show progressive improvement. GEOID18 is the latest in the NGS hybrid geoid model series that effectively provides a conversion surface between NAD83 (2011) ellipsoid heights obtained with GPS positioning and the official vertical datum in a particular region. GEOID18 is specifically developed for use and applications within CONUS, Puerto Rico, and the U.S. Virgin Islands. This geoid model is provided in the form of a 1 arc-minute grid over these areas. Additional 1 arc-minute gridded products derived from GEOID18 include an estimated uncertainty grid and a deflection of the vertical grid (DEFLEC18).

Due to a number of factors, GEOID18 shows significant improvement compared with GEOID12B. While very few perfect validation schemes exist to absolutely confirm this improvement, all evidence points to significant improvement. Overall performance of the model when compared to GPS on Bench Marks is 1.39 cm standard deviation, which is a 3 mm improvement from GEOID12B. This is quite a remarkable amount considering there are 32,000+ observations and 7,100+ additional new observations. This 1.39 cm standard deviation reflects how well the model fits to GPS on Bench Marks and what users can expect to see. Most importantly, this improvement is not driven by any one state – 47 out of 51 states/regions experience this type of improvement. Additionally, from the omission-type improvement perspective, the vast majority of states/regions show an improvement in the percentage of the state that is within 30 km of a bench mark. Furthermore, a number of states/regions have significantly added coverage with five states over 20% improvement and 12 states with over 10% improvement compared to GEOID12B.

There are two main reasons for the improvement in GEOID18 compared with GEOID12B. First, there have been significant advancements in gravimetric geoid modeling theory and data quality since 2012. This includes improvements in surface gravity data, GRAV-D data, satellite gravity from GRACE and GOCE, DEM data, and the modeling theory. Overall, this improvement is only about 1.2 mm; however, improvement can be observed in 39 out of 51 states/regions as illustrated in Figure 44 and nine states have 5+ mm of improvement. The GRAV-D contribution is approximately 0.4 mm overall but various states experience much higher levels with five states having 5+ mm of improvement due to GRAV-D. Secondly and most importantly, the refined GPS on Bench Marks dataset provides the most significant impact on GEOID18's performance. Overall, this contribution is 4.6 mm and is illustrated state-by-state in Figure 46 where 20 out of 51 states/regions have 5+ mm improvements from GEOID12B to GEOID18.

The improvement can also be illustrated from a relative accuracy perspective where GEOID18 shows less than 2 cm RMS over all distances up to 1000 km. Over distances less than 10 km, GEOID18 exhibits relative accuracy at the 1.6 cm RMS level. In a relative sense, GEOID18 shows improvement over all distances when compared against both the gravimetric model, xGEOID19B, and the previous hybrid geoid model, GEOID12B.

The deflection of the vertical model (DEFLEC18) is constructed using the following 1' grids: GEOID18, a DEM, and a Bouguer gravity model. DEFLEC18 is a hybrid deflection of the vertical model on the Earth's surface, which makes use of GEOID18 to compute the deflections on the geoid. The deflections on the geoid are then corrected for the plumbline curvature based on Helmert's definition and using the DEM and Bouguer gravity model, which results in the surface deflections. Very limited validation of this hybrid deflection model is possible; however, comparisons between CONUS-wide historical

observations, GSVS11, and GSVS14 provide some external quality assessment. Over CONUS, DEFLEC18 has 1.14 arc-seconds RMS in ξ and 1.18 arc-seconds RMS in η based on comparisons with 3,400 historical deflection observations. The perceived accuracy is even better when validating against the newer deflection observations performed in GSVS11 and GSVS14 where the accuracies are 0.19 arc-seconds RMS in ξ and 0.16 arc-seconds RMS in η for GSVS11 and 0.30" RMS in ξ and 0.25" RMS in η for GSVS14.

11 Bibliography

Ahlgren, Kevin; Scott, Galen; Shaw, Brian; Zilkoski, David; Paudel, Nagendra (2020). GEOID18 Height and Estimated Uncertainty for GPS on Bench Marks dataset used to create GEOID18 (NCEI Accession 0209231). NOAA National Centers for Environmental Information. Dataset. https://doi.org/10.25921/d2rg-hv41. Accessed 04/09/2020.

Brown, N. J., McCubbine, J. C., Featherstone, W. E., Gowans, N., Woods, A., & Baran, I. (2018). "AUSGeoid2020 combined gravimetric–geometric model: location-specific uncertainties and baseline-length-dependent error decorrelation." *Journal of Geodesy*, 92(12), 1457–1465.

Burša, M., Kenyon, S., Kouba, J., Šíma, Z., Vatrt, V., & Vojtíšková, M. (2004). "A global vertical reference frame based on four regional vertical datums." *Studia geophysica et geodaetica*, 48(3), 493–502.

Darbeheshti, N., & Featherstone, W. E. (2009). "Non-stationary covariance function modelling in 2D least-squares collocation." *Journal of Geodesy*, 83(6), 495–508.

Featherstone, W. E., & Filmer, M. S. (2012). "The north-south tilt in the Australian Height Datum is explained by the ocean's mean dynamic topography." *Journal of Geophysical Research*: Oceans, 117(C8).

Federal Geodetic Control Committee (1984). *Standards and Specifications for Geodetic Control Networks*. John D. Bossler, Chairman, Rockville, MD.

Grubbs F.E. (1950). "Sample criteria for testing outlying observations." *Ann Math Statist* 21:27–58.

Heiskanen, W.A. and H. Moritz (1967). *Physical Geodesy*. W.H. Freeman and Co, San Francisco.

Hinze, W. J., Aiken, C., Brozena, J., Coakley, B., Dater, D., Flanagan, G., ... & Kucks, R. (2005). "New standards for reducing gravity data: The North American gravity database." *Geophysics*, 70(4), J25–J32.

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org), 15, 25–54.

Jekeli C (1999). An analysis of vertical deflections derived from high-degree spherical harmonic models. J Geod 73(1): 10–22. doi:10.1007/s001900050213.

Li, X., Crowley, J. W., Holmes, S. A., & Wang, Y. M. (2016). "The contribution of the GRAV-D airborne gravity to geoid determination in the Great Lakes region." *Geophysical Research Letters*, 43(9), 4358–4365.

Li, X., K. Ahlgren, R. Hardy, J. Krcmaric, Y.M. Wang (2019). *The development and evaluation of the experimental gravimetric geoid model 2019.* https://beta.ngs.noaa.gov/GEOID/xGEOID19/xGeoid19_tech_details.v10.pdf.

Lilliefors, H. W. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." *Journal of the American Statistical Association*. Vol. 62, pp. 399–402.

Moritz, H (1980). *Advanced physical geodesy*. Herbert Wichmann, Karlsruhe.

National Geodetic Survey. (2017a). *Blueprint for 2022, part 1: geometric coordinates*, NOAA Technical Report NOS NGS 62, available at: https://www.ngs.noaa.gov/PUBS_LIB/NOAA_TR_NOS_NGS_0062.pdf, downloaded 2019 09 18.

National Geodetic Survey. (2017b). *Blueprint for 2022, part 2: geopotential coordinates*, NOAA Technical Report NOS NGS 64, available at: https://www.ngs.noaa.gov/PUBS_LIB/NOAA_TR_NOS_NGS_0064.pdf, downloaded 2019 09 18.

National Geodetic Survey (2019). *Blueprint for 2022, part 3: working in the modernized NSRS*. NOAA Technical Report NOS NGS 67. https://www.ngs.noaa.gov/PUBS_LIB/NOAA_TR_NOS_NGS_0067.pdf. Downloaded 2019 12 18.

Quenouille, M. H. (1949). "Problems in plane sampling." *The Annals of Mathematical Statistics*, *20*(3), 355–375.

Roman, D. R., Wang, Y. M., Henning, W., & Hamilton, J. (2004). "Assessment of the new national geoid height model, GEOID03." *Surveying and Land Information Science*, 64(3), 153–162.

Sanso, F. (1986). "Statistical methods in physical geodesy." In: Sunkel H (ed) *Mathematical and numerical techniques in physical geodesy*. Springer, Berlin Heidelberg New York, pp 49–156.

Smith, D. A., & Roman, D. R. (2001). "GEOID99 and G99SSS: 1-arc-minute geoid models for the United States." *Journal of Geodesy*, *75*(9-10), 469–490.

Smith, D. A., Holmes, S. A., Li, X., Guillaume, S., Wang, Y. M., Bürki, B., ... & Damiani, T. M. (2013). "Confirming regional 1 cm differential geoid accuracy from airborne gravimetry: the Geoid Slope Validation Survey of 2011." *Journal of Geodesy*, 87(10-12), 885-907.

Soler, T., & Snay, R. A. (2004). "Transforming positions and velocities between the International Terrestrial Reference Frame of 2000 and North American Datum of 1983." *Journal of Surveying Engineering*, 130(2), 49–55.

Torge W (1991) *Geodesy*, 2nd edn. Walter de Gruyter, Berlin.

van Westrum et al. (in press). GSVS17.

Wang, Y. M., Becker, C., Mader, G., Martin, D., Li, X., Jiang, T., ... & Bürki, B. (2017). "The Geoid Slope Validation Survey 2014 and GRAV-D airborne gravity enhanced geoid comparison results in Iowa." Journal of Geodesy, 91(10), 1261–1276.

Zilkoski, D. B., J.H. Richards, and G.M. Young. (1992). "Results of the general adjustment of the North American Vertical Datum of 1988." *Surveying and Land Information Systems*, 52(3), 133–149, http://www.ngs.noaa.gov/PUBS_LIB/NAVD88/navd88report.htm, downloaded 2019 09 05.

12 Appendix I

Table 20: State-by-state statistics for GEOID18 and GPS on Bench Marks. Also included is CD = Canada, MX=Mexico, and ON = Ontario from NGS IDB.

GEOID18				Pre-model residual Used Marks [cm]				Post-model residuals Used Marks [cm]			
State	Total Marks	Marks Used	Marks Not Used	Min.	Max.	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.
AL	496	452	44	-7.9	15.9	4.9	3.5	-4.9	5.4	0.0	1.4
AR	508	409	99	-8.2	10.0	0.7	3.0	-4.9	5.2	0.0	1.5
AZ	624	431	193	-18.8	6.4	-4.0	4.7	-7.0	6.0	0.0	1.7
CA	723	641	82	-21.4	21.7	-2.2	5.0	-6.8	7.1	0.0	1.7
CD	574	565	9	-66.9	6.1	-24.4	12.8	-5.9	8.0	-0.2	1.8
CO	688	571	117	-27.5	9.9	-6.4	6.6	-7.3	8.2	0.0	2.2
CT	78	69	9	-14.4	-1.7	-9.3	2.5	-3.8	3.4	0.0	1.4
DC	17	13	4	-6.2	-0.2	-1.6	1.6	-5.1	1.0	-0.4	1.6
DE	152	129	23	-4.3	3.3	-0.7	1.8	-4.2	4.1	0.1	1.3
FL	3892	2894	998	-11.6	10.8	-2.3	3.5	-6.2	6.1	0.0	1.5
GA	162	145	17	-3.6	13.4	4.7	3.9	-3.8	3.2	0.1	1.4
IA	393	338	55	-1.4	13.5	9.0	3.1	-3.1	4.5	0.0	0.9
ID	191	138	53	-17.0	7.4	-3.0	4.1	-3.7	5.6	0.1	1.5
IL	922	796	126	-4.9	14.5	4.8	4.0	-4.6	4.5	0.0	1.3
IN	210	170	40	-5.2	10.2	3.7	2.8	-3.7	3.7	0.0	1.3
KS	242	196	46	-11.9	9.2	0.4	3.9	-4.6	3.2	0.1	1.4
KY	214	178	36	-6.4	6.6	0.8	2.2	-6.5	3.5	-0.1	1.3
LA	381	51	330	-37.7	7.5	-1.8	11.6	-6.7	4.5	-0.1	2.0
MA	63	51	12	-16.2	-4.8	-10.3	2.4	-3.8	3.1	0.0	1.2
MD	667	614	53	-7.6	7.9	-1.2	2.4	-5.1	6.4	0.0	1.6
ME	89	71	18	-24.9	-10.0	-17.2	3.4	-3.7	3.1	0.0	1.3
MI	939	862	77	-25.2	10.6	-4.0	6.4	-4.7	4.1	0.0	1.3
MN	11011	10891	120	-14.3	13.4	3.0	4.4	-5.5	6.6	0.0	0.9
MO	372	310	62	-8.4	10.7	0.6	3.3	-4.0	4.2	0.0	1.5
MS	492	419	73	-9.7	12.2	1.1	5.3	-3.8	4.7	0.0	1.3
MT	475	421	54	-19.2	5.8	-3.4	4.4	-5.2	5.3	0.0	1.7
MX	210	206	4	-48.4	33.6	-10.0	16.1	-19.4	20.6	0.0	7.1
NC	2050	1957	93	-10.3	11.0	0.7	3.0	-6.0	5.8	0.0	1.4
ND	157	134	23	-8.8	13.3	2.0	4.7	-4.0	3.9	0.0	1.3
NE	219	191	28	-9.1	7.4	0.5	3.4	-4.3	2.9	-0.1	1.1
NH	43	34	9	-24.4	-11.7	-15.8	3.0	-3.4	1.9	-0.1	1.2
NJ	633	599	34	-12.3	2.9	-4.0	2.7	-4.4	4.9	0.0	1.4
NM	166	137	29	-25.4	11.6	-10.2	5.4	-5.6	6.8	-0.1	1.6
NV	90	82	8	-20.2	6.8	-4.2	4.7	-4.4	5.0	0.1	1.6
NY	316	275	41	-22.9	0.6	-9.5	5.0	-4.6	3.1	0.0	1.1

OH	457	382	75	-8.8	12.0	3.2	3.4	-5.9	6.2	0.0	1.9
OK	228	195	33	-19.4	5.2	-4.7	4.3	-3.9	4.5	0.0	1.5
ON	7	6	1	-18.5	2.6	-6.2	7.5	-0.4	1.4	0.2	0.7
OR	439	368	71	-15.2	13.1	-2.4	5.0	-5.9	5.3	0.1	1.7
PA	311	200	111	-10.1	6.1	-1.5	2.8	-3.4	4.3	0.0	1.4
PR	121	107	14	-8.7	4.6	-0.1	2.6	-3.4	3.9	0.0	1.5
RI	40	31	9	-11.8	-2.5	-7.6	2.5	-3.7	4.1	0.1	2.0
SC	1696	1627	69	-3.7	15.2	5.4	3.5	-6.0	6.8	0.0	1.3
SD	283	250	33	-11.6	11.5	0.7	6.0	-2.6	3.0	0.0	1.0
TN	220	199	21	-6.0	10.0	2.2	2.8	-4.4	3.2	-0.1	1.6
TX	776	539	237	-30.4	6.6	-6.3	6.2	-3.7	5.7	0.0	1.3
UT	154	119	35	-13.1	4.7	-4.0	4.0	-7.7	4.7	0.0	1.8
VA	465	408	57	-9.8	6.3	-0.8	2.6	-5.4	5.6	0.0	1.7
VQ	24	20	4	-6.3	6.1	0.6	3.9	-4.4	4.5	0.2	2.3
VT	762	489	273	-24.2	-8.3	-17.0	2.9	-4.3	4.9	0.0	1.3
WA	420	333	87	-39.8	13.1	-3.0	7.6	-10.1	4.5	-0.1	2.0
WI	1593	1557	36	-7.2	11.6	3.5	3.6	-3.0	3.7	0.0	0.7
WV	81	73	8	-6.8	9.2	3.6	3.1	-4.4	3.2	0.2	1.5
WY	190	141	49	-20.5	6.1	-6.0	5.4	-4.4	4.8	0.0	1.8

13 Appendix II

Table 21: State-by-state statistics for minimum distance to a used GPS on Bench Marks

	GEOID12B	GEOID18	GEOID12B					GEOID18				
State	Median Distance [km]	Median Distance [km]	10km	20km	30km	45km	60km	10km	20km	30km	45km	60km
AL	12.0	11.0	38.8%	84.3%	97.8%	100.0%	100.0%	43.7%	90.2%	99.8%	100.0%	100.0%
AZ	22.2	17.5	18.1%	44.5%	67.6%	88.5%	97.0%	24.4%	57.3%	78.2%	94.4%	98.7%
AR	21.3	11.5	16.6%	46.0%	71.9%	93.6%	100.0%	42.0%	84.2%	97.3%	100.0%	100.0%
CA	16.9	18.6	27.8%	58.1%	77.5%	92.8%	98.9%	22.4%	54.0%	76.3%	92.6%	98.4%
CO	13.8	13.7	34.3%	71.5%	91.8%	99.4%	100.0%	34.1%	72.8%	93.0%	99.9%	100.0%
CT	11.6	6.9	40.2%	85.7%	98.7%	100.0%	100.0%	77.2%	99.2%	100.0%	100.0%	100.0%
DE	5.2	3.7	90.7%	100.0%	100.0%	100.0%	100.0%	98.7%	100.0%	100.0%	100.0%	100.0%
DC	2.1	2.7	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
FL	5.9	5.4	77.2%	98.0%	99.9%	100.0%	100.0%	80.0%	96.9%	99.4%	100.0%	100.0%
GA	19.9	18.7	17.6%	50.3%	80.0%	98.4%	99.9%	18.4%	55.0%	86.3%	100.0%	100.0%
ID	22.6	21.4	14.7%	42.9%	67.5%	87.3%	95.1%	15.4%	45.7%	71.0%	89.6%	96.0%
IL	10.6	7.4	46.3%	87.7%	98.4%	99.9%	100.0%	68.6%	98.1%	100.0%	100.0%	100.0%
IN	12.6	10.5	35.5%	82.3%	98.6%	100.0%	100.0%	46.7%	93.6%	98.6%	100.0%	100.0%
IA	18.0	14.6	19.3%	57.7%	85.2%	99.3%	100.0%	28.2%	73.8%	95.8%	100.0%	100.0%
KS	23.9	15.5	11.6%	38.1%	67.0%	93.0%	99.2%	23.3%	71.6%	96.0%	99.9%	100.0%
KY	15.8	14.7	25.3%	66.2%	90.3%	99.3%	100.0%	27.6%	73.1%	96.1%	100.0%	100.0%
LA	20.2	28.7	21.5%	49.4%	73.1%	89.3%	95.6%	9.7%	30.1%	52.7%	80.0%	95.6%
ME	21.8	21.5	16.0%	45.5%	66.0%	83.4%	93.9%	15.9%	46.0%	67.6%	84.5%	94.5%
MD	4.6	4.5	96.4%	100.0%	100.0%	100.0%	100.0%	94.8%	100.0%	100.0%	100.0%	100.0%
MA	12.4	11.1	35.7%	82.5%	97.2%	99.6%	100.0%	42.8%	87.9%	96.4%	99.6%	100.0%
MI	9.4	8.6	53.4%	87.2%	96.9%	99.4%	99.8%	58.1%	92.9%	99.3%	99.7%	99.8%
MN	4.8	4.1	88.0%	99.6%	100.0%	100.0%	100.0%	90.5%	99.7%	100.0%	100.0%	100.0%
MS	10.0	10.2	50.0%	87.8%	98.7%	100.0%	100.0%	48.8%	88.3%	98.2%	99.7%	100.0%
MO	18.2	12.4	20.4%	56.2%	82.6%	97.1%	99.9%	36.2%	85.6%	98.5%	100.0%	100.0%
MT	22.2	16.7	14.0%	43.4%	69.8%	92.1%	99.1%	22.6%	62.6%	86.6%	97.5%	99.9%
NE	18.1	16.2	18.2%	57.5%	86.7%	99.5%	100.0%	22.3%	65.6%	92.3%	99.8%	100.0%

NV	33.5	31.6	6.8%	22.6%	43.0%	71.0%	89.7%	7.7%	25.5%	46.7%	73.7%	90.9%
NH	17.1	13.8	24.0%	60.0%	86.8%	100.0%	100.0%	32.9%	73.1%	94.3%	100.0%	100.0%
NJ	5.1	4.5	87.4%	100.0%	100.0%	100.0%	100.0%	91.5%	100.0%	100.0%	100.0%	100.0%
NM	31.2	27.4	8.2%	25.5%	47.3%	77.5%	92.6%	9.5%	31.3%	56.2%	85.7%	97.7%
NY	11.9	11.2	38.2%	84.0%	95.4%	98.9%	100.0%	42.0%	88.6%	98.5%	100.0%	100.0%
NC	5.7	6.3	81.6%	99.6%	100.0%	100.0%	100.0%	75.9%	98.9%	100.0%	100.0%	100.0%
ND	26.6	16.4	9.4%	32.1%	58.7%	87.8%	97.3%	20.8%	67.7%	96.3%	99.9%	100.0%
OH	12.2	10.6	38.8%	80.8%	97.7%	100.0%	100.0%	46.6%	89.5%	99.7%	100.0%	100.0%
OK	23.5	15.0	11.5%	38.9%	68.9%	94.7%	99.8%	26.1%	72.2%	94.8%	99.7%	100.0%
OR	20.8	16.5	18.2%	47.6%	72.0%	92.7%	98.6%	24.6%	62.6%	88.4%	99.0%	100.0%
PA	15.5	11.7	25.1%	68.5%	91.2%	99.8%	100.0%	39.6%	87.2%	99.0%	100.0%	100.0%
PR	23.9	6.8	21.9%	42.4%	62.1%	89.5%	99.1%	74.1%	98.1%	100.0%	100.0%	100.0%
RI	5.5	5.4	91.3%	97.8%	100.0%	100.0%	100.0%	91.0%	97.8%	100.0%	100.0%	100.0%
SC	5.4	5.3	80.9%	99.2%	100.0%	100.0%	100.0%	80.7%	99.0%	100.0%	100.0%	100.0%
SD	14.3	14.1	29.2%	70.7%	92.4%	99.8%	100.0%	29.8%	72.7%	93.6%	100.0%	100.0%
TN	12.2	12.5	37.8%	86.9%	99.7%	100.0%	100.0%	36.0%	83.8%	98.3%	100.0%	100.0%
TX	29.2	27.3	8.0%	27.4%	51.9%	81.5%	95.3%	10.0%	31.6%	56.5%	83.9%	95.8%
UT	31.9	25.7	6.7%	24.0%	46.0%	72.4%	87.8%	11.6%	36.3%	59.0%	80.8%	91.5%
VT	6.5	6.0	75.5%	99.8%	100.0%	100.0%	100.0%	80.1%	99.8%	100.0%	100.0%	100.0%
VA	12.6	13.4	36.1%	80.2%	97.4%	100.0%	100.0%	33.1%	76.7%	96.5%	100.0%	100.0%
VQ	4.0	4.0	96.7%	100.0%	100.0%	100.0%	100.0%	96.7%	100.0%	100.0%	100.0%	100.0%
WA	13.5	13.1	32.9%	74.7%	91.3%	98.9%	100.0%	33.9%	76.4%	92.7%	99.4%	100.0%
WV	16.1	13.9	22.6%	66.1%	90.5%	99.9%	100.0%	28.7%	79.7%	98.9%	100.0%	100.0%
WI	6.6	4.8	70.6%	89.2%	96.5%	100.0%	100.0%	92.8%	100.0%	100.0%	100.0%	100.0%
WY	26.8	22.6	10.8%	32.7%	58.0%	84.5%	95.9%	14.1%	42.3%	69.3%	91.1%	97.9%

14 Appendix III

Table 22: Additional metadata used to evaluate GPS on Bench Marks examples in Section 4.2.

PID	Pre-Modeled Residual v14.1.1 (cm)	Post-Modeled Residual (cm)	Designation	State	NAD 83 (2011) Ellipsoid Height (m)	NAVD 88 Orthometric Height (m)	Use (blank) or Do_Not_Use (X)	GNSS Information	Leveling Information	Comments
AF9779	29.7	18.7	SEAT B	WA	-16.76	7.189	X	GNSS Obs in 1998	2007 - 2/1	Do not use - Very large pre- and post-modeled residuals and there are nearby stations with smaller residuals
AF9780	15.7	4.7	SEAT A	WA	-11.888	11.92	X	GNSS Obs in 1998	2007 - 2/1	Do not use - Very large pre-modeled residual and large post-modeled residual relative to its neigbors that are within a few kms
HF0394	0.8	-0.3	SMILEY	KS	370.462	399.67		GNSS Obs in 2002	1962 - 2/0	Okay to use - small pre- and post-modeled residuals
HF0299	-8.7	-8.5	POLE	KS	426.644	456.14	X	GNSS Obs in 1997 and 2002	1962 - 2/0 (On a spur section of a short leveling line tied to a 1934 leveling line)	Do not use - nearby stations with smaller residuals (PID HF0394)
EK0599	-26.0	-23.4	F 16	OK	197.577	226.86	X	GNSS Obs in 2001 (Based on Single GNSS Vector - No Check Station)	1933- 1/2	Do not use - Very large pre- and post-modeled residuals, and GNSS ellipsoid height based on a single vector
EL0045	-3.8	0.0	A 16 Reset	OK	175.296	204.226		GNSS Obs in 1993	1933- 1/2	Okay to use - small pre- and post-modeled residuals
EK0613	-3.0	-0.3	P 16 Reset	OK	189.445	219.462		GNSS Obs in 1993 and 2001	1934 - 2/0	Okay to use - small pre- and post-modeled residuals