NOAA Technical Report NOS NGS 78

Technical Details of the Experimental GEOID 2020

Yan Ming Wang
Xiaopeng Li
Kevin Ahlgren
Jordan Krcmaric
Ryan Hardy

National Geodetic Survey

Marc Véronneau
Jianliang Huang

Canadian Geodetic Survey

David Avalos

National Institute of Statistics and Geography, Mexico

June 2022

Technical Details of the Experimental GEOID 2020

NOAA Technical Report NOS NGS 78

Yan Ming Wang, Xiaopeng Li, Kevin Ahlgren, Jordan Krcmaric and Ryan Hardy

The National Geodetic Survey, NOAA, U.S.A.

Marc Véronneau and Jianliang Huang

The Canadian Geodetic Survey, Surveyor General Branch, Natural Resources Canada

David Avalos

The National Institute of Statistics and Geography, Mexico

June, 2022

Abstract

For the upcoming North American-Pacific Geopotential Datum of 2022 (NAPGD2022), the United States National Geodetic Survey (NGS), the Canadian Geodetic Survey (CGS), and Mexico's National Institute of Statistics and Geography (INEGI) have computed the first joint experimental gravimetric geoid model for North America. xGEOID20A/B are 1'×1' grids for the region bordered by latitude N0° and N85°, and longitude E180° and E350°. For this joint effort, NGS and CGS computed geoid models using a common dataset, but each agency used their own computational method. This report records this joint effort and shows the model evaluation results.

Contents

1. Introduction

Since 2014, NGS has been computing two types (A and B) of experimental geoid models (xGEOID) using the latest techniques and data. On the one hand, the A model is computed without using the GRAV-D airborne gravity data. On the other hand, the B model includes all GRAV-D data available. The GRAV-D contribution to the geoid model can be measured by differencing models A and B.

Section 2 describes the fundamental parameters and gravity data used in the computation. It summarizes the gravity data and the newly constructed digital elevation model (xDEM20). Section 3 summarizes the computational methods used by NGS and CGS. The combination of NGS/CGS models is also given in Section 3. The evaluation of the geoid models using independent data sets is performed in Section 4. The computation of xDEFLEC20 and its evaluation is given in Section 5. Section 6 is devoted to the uncertainty estimation for xGEOID20 and xDEFLEC20. Summary and conclusions are discussed in Section 7.

2. Fundamental parameters and data used

2.1 Fundamental parameters

The xGEOID20 (A and B) models are in a tide-free system, and their geopotential (W_0) is 62,636,856.0 m^2s^{-2} (NGS 2017). The product of the gravitational constant and the total mass of the Earth used for the geoid models is GM=3.986044415×10^{14} m^3s^{-2}, the same as the International Height Reference Frame (IHRF, Sanchez et al. 2021). The geoid undulations are calculated in ITRF2014 at epoch 2010.0 using the GRS80 reference ellipsoid (Moritz 2000). The geopotential of xGEOID20 is 4.85 m^2s^{-2} smaller than the geopotential of the GRS80 ellipsoid (U_0=62,636,860.850 m^2s^{-2}) and 2.6 m^2s^{-2} larger than the geopotential W_0 = 62636853.4 m^2s^{-2}, adopted by the IHRF. It should be noted that the IHRF geopotential is in the mean tide system.

2.2 Gravity data used

The terrestrial and altimetry-derived gravity data are the same as for the xGEOID19 computation (Li et al. 2019). The terrestrial gravity data consists of 1,633,376 gravity points provided by the National Geospatial–Intelligence Agency (NGA) and 135,290 gravity points collected by NGS. Over the oceans, the DTU15 altimetry-derived gravity anomalies (Andersen and Knudsen 2016), available in a 1'×1' grid, are used in the computation. The main update to the gravity data is 9 new GRAV-D survey blocks over portions of Alaska, Arizona, California, Georgia, Oklahoma, South Carolina, Texas, and Washington. A total of 63 GRAV-D airborne gravity survey blocks are now available for xGEOID20 computation. Another major update is the use of the latest satellite-only global gravity model GOCO06s (Kvas et al. 2021). It replaces GOCO05s (Mayer-Gürr et al. 2015), which was used for xGEOID19 calculation.

2.3 xG20DEM

A new digital elevation model, xG20DEM, was developed for the xGEOID20 computation (Krcmaric et al. in prep). The model was composed TanDEM-X (Wessel et al. 2018), MERIT (Yamzaki et al. 2017) and the USGS 3D Elevation Program (3DEP) dataset (USGS 2019). All datasets were converted to GRS80 ellipsoidal heights and combined at a $3''$ spatial resolution. In comparisons with the ellipsoidal heights of GPS on Benchmarks (see section 4.1) and the geoid slope validation surveys 2011, 2014 and 2017 (GSVS11, GSVS14, and GSVS17, see section 4.2), the accuracy (1σ) of xG20DEM was estimated to be about 1 meter in flat and low elevation regions. However, the accuracy decreased to about 2 to 3 meters in high mountains.

3. Geoid computation and combination

This section summarizes the geoid computation methods used by NGS and CGS. Computation details are referred to the cited papers. The combination methods are also discussed.

3.1 CGS geoid computation

CGS continues to make use of the Stokes-Helmert's scheme for the calculation of geoid models for Canada and North America. This approach has been the focus point in Canada since the mid-1980s with early work at the University of New Brunswick toward the theoretical formulation of the geoid to achieve the mm accuracy. The research project saw collaboration with the University of Calgary, CGS, NGS and international academic institutions over the years. This effort culminated in 2013 with the publication of the CGG2013 model, which now serves as the vertical datum for Canada (Véronneau and Huang 2016).

CGS developed two geoid models for the xGEOID20 project: one without GRAV-D (CGSA) and one with GRAV-D data (CGSB). The CGS calculation uses the same datasets (gravity and DEM) as NGS to minimize discrepancies associated with data.

For CGSA, the long wavelength components of the geoid heights rest entirely on the Earth Gravity Model xG20RefA (see section 3.2) up to degree 150. The degree-banded Stokes kernel contributes to the calculation of the middle and short wavelength components by integrating terrestrial gravity data, which are introduced gradually between degrees 150 and 220 using a cosine function. The Ref20A model and terrestrial gravity data are previously transformed into the Helmert space, i.e., an Earth where topographical masses above the geoid are removed and restored following Helmert's second condensation method. This scheme generates a co-geoid that is transformed to the geoid by adding the primary indirect terrain effect. Finally, the geoid is corrected by the zero-term (N_0) to represent equipotential surface 62,636,856 ($\mathrm{m}^2/\mathrm{s}^2$).

The Helmert terrestrial gravity grid is produced by:
1) calculating a refined Bouguer anomaly at each gravity point,
2) interpolating the refined Bouguer anomalies into a $30"\times30"$ grid using least-squares collocation,
3) adding a Bouguer plate with a height given by the Digital Elevation Model,
4) applying systematic corrections to obtain the Helmert anomalies, and finally
5) averaging the grid to 1'×1' spacing.

The grid is a composite of land gravity data, marine gravity from DTU15, ArcGP v2020 for Greenland, and EGM2008 for data gaps over land (outside Canada, and USA). The Ref20A Helmert gravity anomalies (degree 2190) are subtracted from the Helmert terrestrial gravity anomalies prior to the Stoke integration.

CGSB is produced by adding a correction to CGSA by using xG20RefB (see section 3.2) between spherical harmonic degrees 150 and 2159. This approach is to include GRAV-D signals, which is part of the xG20RefB model, but it also means that xG20RefB replaces the terrestrial gravity in the regions where GRAV-D is not available. This deteriorates the solution over Canada because the xG20RefB gravity field is not as accurate as the terrestrial gravity grid produced by CGS. The highest frequencies (above 2159) of CGSA are still calculated from the terrestrial gravity data. This approximate approach was implemented because a pure GRAV-D grid was not available. Future models will combine spectrally an Earth Gravity model, a GRAV-D gravity grid and a terrestrial gravity grid by using a remove-compute-restore scheme with a kernel modification. This approach will also require a tapering technique when transitioning between regions with and without GRAV-D data (e.g., between U.S. and Canada).

3.2 NGS geoid computation

The NGS method is based on Molodensky's problem (Moritz 1980). The height anomaly is computed first using the analytical downward continuation solution, then the geoid is computed by adding the geoid-quasigeoid separation term. Detailed computation formulas can be found in Wang et al. (2012, 2020).

The computation started with the spherical harmonic coefficient model Ref16A, which was developed using the method following closely that of the EGM2008 (https://beta.ngs.noaa.gov/GEOID/xGEOID16/xGEOID16_technical_details.shtml). The model was then merged with the latest satellite gravity model GOCO06s (Kvas et al. 2021) from degree 2 to 230. To have a smooth combination, a cosine taper function at degree 120 was applied from degrees 150 to 220. This GOCO06s updated model was call xG20RefA.

Using this coefficient model as the reference model, the residual gravity disturbances were computed at the altitude of each flight, then each individual flight line was de-biased using its median. The residual gravity disturbances were resampled and edited for outliers, then reduced to the mean flight altitude h. The residual gravity disturbances were gridded into $5' \times 5'$ cells, and modeled through the harmonic analysis on a larger ellipsoidal with semimajor axis $a + h$ (Smith et al. 2013), where a is the semimajor axis of the reference ellipsoid GRS80. The coefficients were then scaled to the reference ellipsoid, and this model was called xG20RefB.

NGS computed two quasigeoid models. One used the residual terrain model (RTM) to reduce the terrain effect, and then interpolate the residual gravity anomalies into $1' \times 1'$ cells. Another model used the complete Bouguer anomaly for gravity interpolation and data gridding.

The type A and type B quasigeoid models were computed following exactly the same procedure, but use the coefficient models xG20RefA and xG20RefB, respectively. In addition to the RTM method (Wang et al. 2020), the refined Bouguer anomalies were used to interpolate the data points into 1'×1' grids. After computing quasigeoid models using RTM and Bouguer approaches, the two models (RTM and Boug) were converted into geoid models by applying the geoid-

quasigeoid separation (GQS) that considers higher order terms (complete separation) than the Bouguer anomaly-only term (simple separation). The higher terms are negligible in lower elevation regions, but they can reach decimeters in high mountains. For instance, in Colorado where the 1-cm geoid computation experiment took place (Wang et al. 2021a), NGS and CGS models became more similar after applying the complete separation instead of the simple separation. Table 1 gives the statistics of the differences between the NGS and CGS geoid models when NGS models are converted using the simple (sGQS) and complete separation (cGQS).

Table 1. Geoid height differences (CGS-NGS) in Colorado (latitude N35° to N40°, longitude E250° to E258°), units in cm.

	RTMA-CGSA		**BougA-CGSA**		**RTMB-CGSB**		**BougB-CGSB**	
	sGQS	**cGQS**	**sGQS**	**cGQS**	**sGQS**	**cGQS**	**sGQS**	**cGQS**
Mean	0.1	-0.3	-1.2	-1.7	-0.0	-0.5	-1.3	-1.8
STD	3.7	2.7	3.1	2.1	3.3	2.7	2.1	1.3
RMS	3.7	2.7	3.3	2.7	3.3	2.7	2.5	2.2
Min.	-20.0	-18.0	-12.2	-11.2	-26.4	-21.3	-17.0	-16.4
Max.	37.1	16.0	26.8	13.4	35.3	12.2	24.1	8.1
Range	57.1	34.0	39.0	24.6	61.7	33.7	41.1	24.5

Table 1 shows clearly that the complete separation term reduces not only the range of geoid differences between the NGS and CGS models noticeably in decimeters, but also the RMS of the differences in Colorado, where the heights range from 932 m to 4,385 m (Wang et al. 2021a). For this reason, the complete geoid-quasigeoid separation term (Wang et al. 2021b) was applied to the quasigeoid models using the RTM and Boug approaches and the simple geoid-quasigeoid separation term was removed from consideration. From this point forward in the text, the two NGS geoid models (RTM20 and Boug20) are the geoid models converted from the quasigeoid models by adding the complete GQS term.

3.3 Model combination

The NGS and CGS geoid models have a 1′×1′ spatial resolution, but the node points are registered differently. The NGS models are “upper-left corner-registered” while the CGS models are “cell-centered-registered.” The CGS models were interpolated into cell-cornered registration using the bi-spline method before the combination. The bi-spline method is preferred over bi-quadratic to maintain smoothness. The NGS models cover the area between latitude N0° and N85°, and longitude E170° and E350°. The CGS models cover the region bordered by latitude N0° and N90°, and longitude E180° and E360°. The combined models include the common area of the NGS and CGS models (N0°/N85°/E180°/E350°).

Because the NGS and CGS computation methods differ, we expect some small numerical differences. The following figures show the geoid differences between the NGS and CGS geoid models.

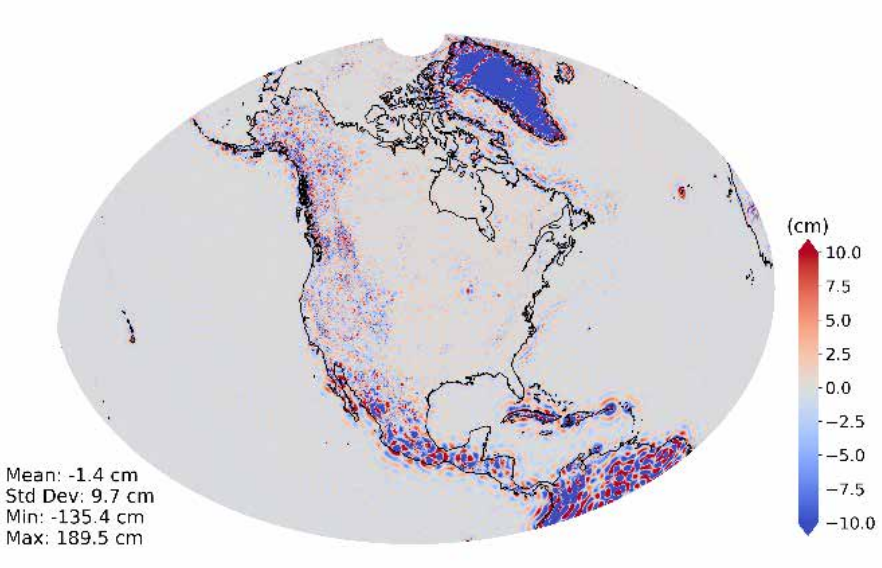

Figure 1a. Geoid height differences between NGS and CGS models: CGSA – RTMA.

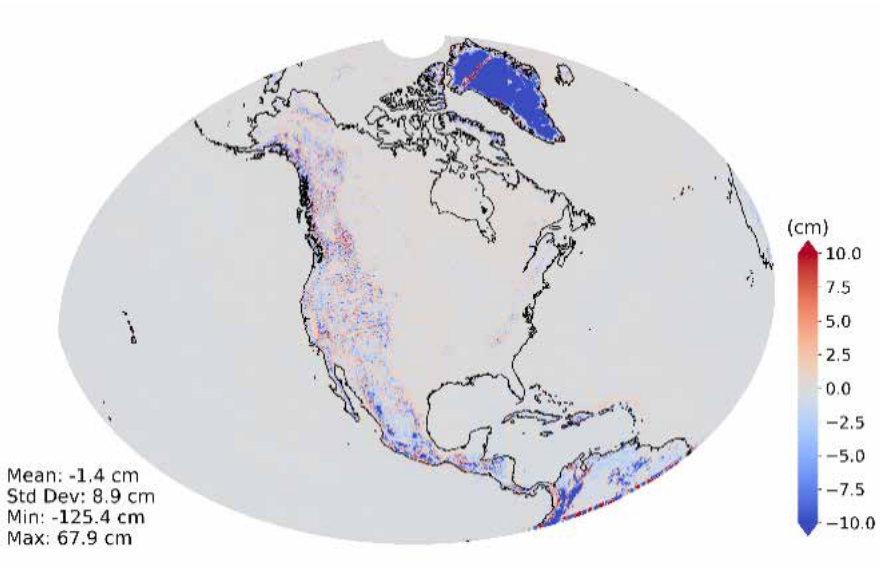

Figure 1b. Geoid height differences between NGS and CGS models: CGSB – RTMB.

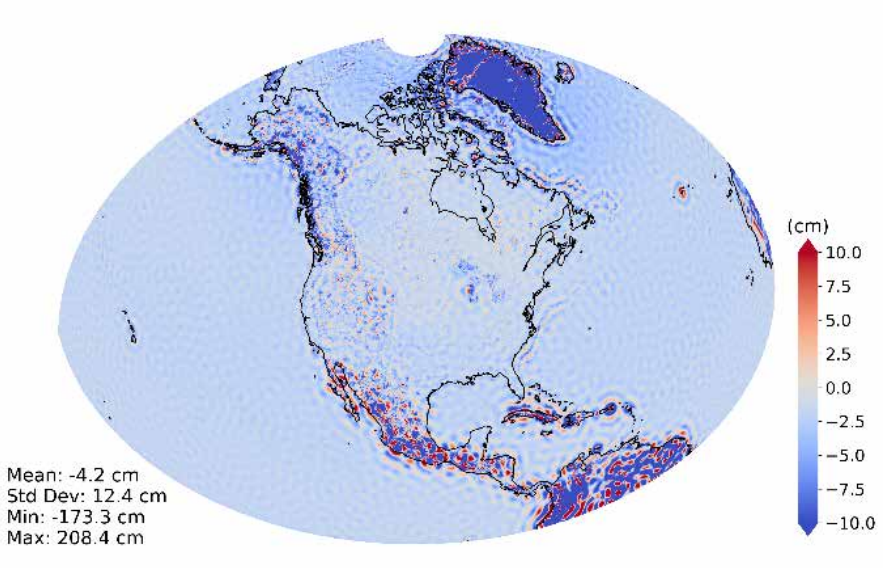

Figure 1c. Geoid height differences between NGS and CGS models: CGSA – BougA.

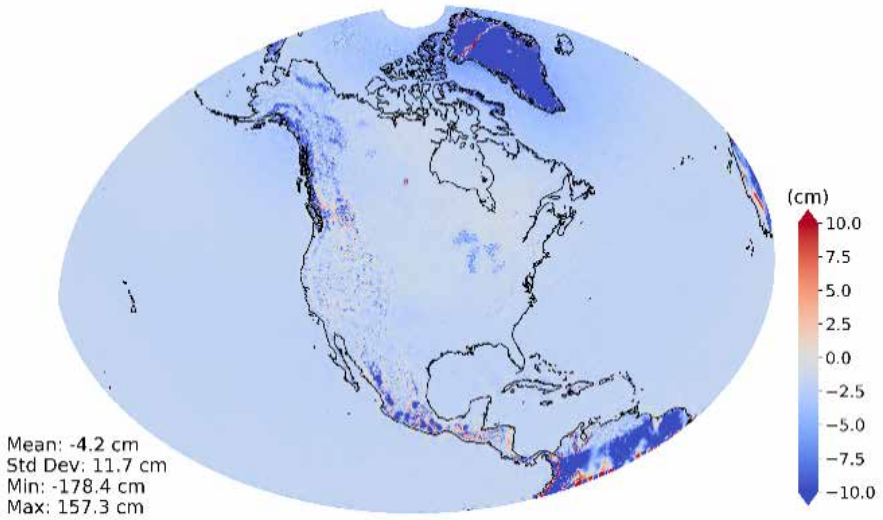

Figure 1d. Geoid height differences between NGS and CGS models: CGSB – BougB.

Figures 1a–1d show large differences in areas such as Greenland and South America. These noticeable large differences are associated with areas where terrestrial gravity data are missing or sparse. These large geoid differences are consistent with the predicted geoid errors, plotted in Figure 7. In other words, the areas with large differences are having large geoid errors, too. Another observation in Figures 1a–1d is the geoid differences between NGS BougA/B and CGS models, which reach 10 cm in the Great Lakes region. In contrast, the RTM solutions agree well with CGS models. The larger differences in NGS Boug models may be caused by improper use of the gravity data over the lakes (e.g., mixing incorrectly ship and lake-bottom observations). Another observation is that the mean biases between NGS RTM and CGS models are 1.4 cm while they are 4.2 cm between NGS Boug and CGS.

Two weighting schemes were evaluated in the model combination: one consisted in using the simple arithmetic mean, the other used weights based on the GPS/leveling differences. The weights were determined as the inverse of the variances of the geoid differences between the GPS/leveling data and geoid models on a state-by-state basis.

Out of the four NGS and two CGS geoid models, four combined models were computed: RTMA&CGSA, BougA&CGSA, RTMB&CGSB, and BougB&CGSB. In the next section, these models are compared against the historical GPS on Benchmarks and the GSVS datasets.

4. xGEOID20 validation

4.1 GPS on benchmarks

The historical GPS/leveling data consists of the published ellipsoidal and (Helmert) orthometric heights on 31,514 benchmarks This data set was used in the GEOID18 computation, and its data selection is documented in (Ahlgren et al. 2020). The published ellipsoidal heights were in the North American Datum 1983 (NAD83), and they were converted into the geocentric ITRF08 at epoch 2010.00. The published (Helmert) orthometric heights were in the North American Vertical Datum 1988 (NAVD88), which was stablished in 1991 by the minimum-constraint adjustment of the Canadian-Mexican-United States leveling observations. It held fixed the height (6.271 meters) of the primary tidal bench mark at Father Point/Rimouski, Quebec, Canada (Zilkoski et al. 1992).

To avoid using the Helmert orthometric heights, the geopotential numbers *C* on each mark was retrieved from NGS IDB database. By using the complete geoid-quasigeoid separation term δ (Wang et al. 2021b), the geoid height on the mark is computed as

$$N_{GPSL} = h - H^N + \delta \tag{1}$$

where N_{GPSL} is the geoid height computed from the GPS/leveling data, h is the ellipsoidal height, and H^N is the normal height computed by (Hofmann-Wellenhof and Moritz 2006):

$$H^N = C/\bar{\gamma} \tag{2}$$

where $\bar{\gamma}$ is the mean normal gravity at the benchmark:

$$\begin{aligned} &\bar{\gamma} = \gamma_0[1 - \tfrac{1}{a}(1 + f + m - 2f sin^2\varphi)(h - \zeta)], \\ &\gamma_0 = 9.780\ 327\ (1 + 0.005\ 3024\ sin^2\varphi - 0.000\ 0058 sin^2 2\varphi), \\ &a = 6378137, f = 0.003\ 352\ 810\ 681\ 18, m = 0.003\ 449\ 786\ 003\ 08. \end{aligned} \tag{3}$$

It is well-known that the GPS/leveling data implied geoid undulations have a continental tilt in the north-west to the south-east direction of about 1 m (Wang et al. 2012) with respect to satellite gravity models. Therefore, the comparisons with the historical GPS/leveling data are done in a state by state manner. Table 2 shows the means and standard deviations of the geoid undulations differences between those from GPS/leveling (N_{GPSL}) and those from the models (N_{model}).

Table 2. Statistics of the geoid differences ($N_{GPSL} - N_{model}$) at 31,514 GPS/leveling data state by state (mean/STD), units in cm. The states with the standard deviations larger than 10 cm are in red color.

Equal Weights					
	#	CGSA&RTMA	CGSB&RTMB	CGSA&BougA	CGSB&BougB
AL	452	-28.8/ 5.98	-28.8/ 6.05	-29.9/ 6.02	-29.8/ 6.08
AZ	431	-49.6/ 9.64	-49.9/ 9.26	-50.8/ 9.18	-51.2/ 8.86
AR	408	-39.1/ 3.77	-39.0/ 3.81	-40.1/ 3.73	-40.0/ 3.76
CA	640	-66.5/12.52	-67.0/12.21	-68.0/12.37	-68.5/12.10
CO	568	-61.9/ 6.04	-62.0/ 6.25	-62.7/ 6.15	-62.8/ 6.49
CT	69	-33.4/ 2.42	-33.1/ 2.38	-34.2/ 2.36	-33.9/ 2.33
DE	129	-35.3/ 3.09	-35.8/ 2.67	-236.1/ 3.13	-36.6/ 2.71
DC	13	-36.4/ 1.55	-35.7/ 1.55	-37.0/ 1.55	-36.4/ 1.54
FL	2894	0.3/ 8.32	0.3/ 8.30	-0.6/ 8.36	-0.6/ 8.34
GA	145	-25.0/ 6.07	-25.1/ 6.04	-26.0/ 6.07	-26.1/ 6.04
ID	138	-92.5/ 8.10	-92.9/ 8.40	-93.2/ 8.21	-93.6/ 8.49
IL	796	-59.2/ 8.84	-59.1/ 8.83	-60.0/ 8.82	-59.9/ 8.82
IN	160	-53.6/ 4.93	-53.7/ 5.18	-54.4/ 4.94	-54.5/ 5.18
IA	338	-74.0/ 3.46	-73.9/ 3.50	-74.8/ 3.47	-74.7/ 3.49
KS	196	-56.8/ 5.64	-56.8/ 5.62	-57.7/ 5.63	-57.7/ 5.62
KY	178	-42.3/ 3.21	-42.2/ 3.14	-43.2/ 3.21	-43.1/ 3.14
LA	51	-25.9/14.23	-26.1/14.01	-27.0/14.25	-27.1/14.02
ME	71	-34.2/ 3.36	-33.2/ 2.52	-35.0/ 3.35	-33.9/ 2.50
MD	614	-36.5/ 2.58	-36.4/ 2.63	-37.3/ 2.55	-37.2/ 2.59
MA	51	-33.6/ 1.75	-33.5/ 2.11	-34.3/ 1.73	-34.2/ 2.09
MI	861	-58.1/ 4.78	-58.0/ 4.60	-58.9/ 4.78	-58.8/ 4.61
MN	10883	-81.9/ 4.35	-81.9/ 4.40	-82.7/ 4.34	-82.6/ 4.39
MS	419	-28.0/ 9.16	-28.1/ 9.05	-29.0/ 9.16	-29.1/ 9.05
MO	310	-50.7/ 6.94	-50.6/ 6.96	-51.6/ 7.00	-51.5/ 7.02
MT	420	-97.1/ 7.82	-97.6/ 8.19	-98.0/ 7.85	-98.5/ 8.29
NE	189	-68.6/ 4.03	-68.7/ 3.99	-69.5/ 4.03	-69.5/ 3.99
NV	82	-70.8/ 8.19	-71.0/ 8.11	-72.0/ 7.84	-72.2/ 7.78
NH	34	-33.1/ 2.11	-33.4/ 2.15	-34.1/ 2.12	-34.4/ 2.24
NJ	599	-35.7/ 2.21	-35.7/ 2.12	-36.5/ 2.22	-36.5/ 2.14
NM	137	-38.8/10.15	-39.1/10.20	-40.1/ 9.98	-40.4/10.04
NY	275	-40.1/ 5.30	-40.0/ 5.34	-40.9/ 5.32	-40.9/ 5.36
NC	1950	-27.4/ 4.78	-27.5/ 4.78	-28.3/ 4.77	-28.4/ 4.77
ND	134	-93.3/ 3.14	-93.3/ 3.23	-94.0/ 3.19	-94.0/ 3.28
OH	381	-51.7/ 3.75	-51.6/ 3.80	-52.6/ 3.83	-52.6/ 3.87
OK	195	-41.0/ 5.39	-40.9/ 5.43	-42.1/ 5.38	-42.0/ 5.42
OR	367	-98.8/ 7.88	-98.9/ 7.99	-99.9/ 8.01	-100.0/ 8.12
PA	199	-42.8/ 3.99	-42.7/ 3.93	-43.6/ 3.99	-43.5/ 3.93

RI	31	-34.5/ 2.44	-34.5/ 2.64	-35.1/ 2.41	-35.1/ 2.57
SC	1626	-28.0/ 6.46	-28.1/ 6.01	-29.0/ 6.53	-29.0/ 6.09
SD	250	-80.6/ 5.60	-80.6/ 5.63	-81.4/ 5.56	-81.4/ 5.59
TN	199	-38.4/ 3.19	-38.4/ 3.27	-39.4/ 3.12	-39.4/ 3.19
TX	516	-20.2/ 8.14	-19.5/ 8.16	-21.3/ 8.11	-20.7/ 8.14
UT	119	-70.7/ 6.94	-71.2/ 7.17	-72.0/ 6.93	-72.4/ 7.21
VT	488	-35.4/ 2.05	-35.3/ 2.03	-36.0/ 2.20	-35.9/ 2.21
VA	407	-35.4/ 3.09	-35.1/ 3.04	-36.2/ 3.08	-35.9/ 3.04
WA	332	-108.3/ 6.96	-108.7/ 7.02	-109.4/ 6.77	-109.8/ 6.89
WV	73	-44.6/ 4.09	-44.7/ 4.34	-45.4/ 4.17	-45.5/ 4.40
WI	1555	-73.7/ 3.86	-73.8/ 3.63	-74.4/ 3.83	-74.5/ 3.59
WY	141	-76.7/ 7.64	-77.1/ 8.48	-77.4/ 7.51	-77.8/ 8.39
Total	31514	-56.3/ 5.86	-56.3/5.84	-57.1/5.85	-57.1/5.83

State by state, the STD values of the geoid differences vary from 2 cm to 14 cm, and the nationwide average is slightly better than 6 cm. Three states have standard deviations larger than 10 cm. California is known as having significant crustal motions due to the tectonic motion and subsidence due to groundwater removal. There is significant subsidence around the Louisiana's Gulf coast. However, there is no clear reason for New Mexico. The mean differences vary widely from 0 cm to 109 cm, indicating large long wavelength errors in the historical GPS/leveling dataset.

Although statistics are not shown in this paper, the same analysis is also conducted on the combined geoid models weighted by the inverse variances of the GPS/leveling differences. The statistics show improvement at sub-mm level in only a few states. The improvement is insignificant and the models computed using the equal weights are used in the following evaluations.

4.2 Geoid slope validation survey 2011, 2014, and 2107

The comparison with the historical GPS/leveling data set gives a rough estimation of the geoid models performance because the accuracies of the geoid undulations derived from the historical GPS/leveling dataset may be worse than the geoid models. To obtain a reliable assessment, NGS conducted three surveys in areas with different types of terrain. These surveys are known as Geoid Slope Validation Surveys and were conducted in 2011 (Texas), 2014 (Iowa) and 2017 (Colorado). Each surveyed line has a length exceeding 300 km. On the three traverses, NGS collected simultaneously with high standards GPS, leveling, gravity and astrogeodetic deflections of the vertical (DoV) to assure epoch consistency between the datasets. The accuracies of the GPS/leveling-derived geoid undulations between any marks of the GSVS lines are estimated to be 1 to 1.5 cm (Smith et al. 2013; Wang et al. 2017; van Westrum 2021). These datasets provide a reliable assessment of the geoid models in a relative manner. This is evidenced by the small STD of the geoid differences, which are listed in Table 3.

Table 3. Statistics of the geoid differences ($N_{GPSL} - N_{model}$) on the marks of three GSVS lines, units in cm.

GSVS11					
Model	**Mean**	**STD**	**Min**	**Max**	**Range**
RTMA	32.1	1.6	28.0	36.6	8.6
RTMB	30.5	1.1	27.6	34.0	6.4
BougA	35.4	1.6	31.3	40.1	8.8
BougB	33.9	1.1	31.1	37.6	6.4
CGSA	32.1	1.3	28.8	36.8	8.0
CGSB	30.7	1.1	27.9	34.3	6.4
xG20RefA	32.5	1.7	28.9	20.0	8.9
xG20REfB	31.3	1.1	28.4	34.7	6.3
RTMA&CGSA	**32.1**	**1.4**	**28.4**	**36.7**	**8.3**
RTMB&CGSB	**30.6**	**1.1**	**27.7**	**34.2**	**6.4**
BougA&CGSA	**33.4**	**1.4**	**29.6**	**38.1**	**8.4**
BougB&CGSB	**31.9**	**1.1**	**29.0**	**35.6**	**6.5**

GSVS14					
Model	**Mean**	**STD**	**Min**	**Max**	**Range**
RTMA	75.2	1.4	72.6	79.0	6.4
RTMB	75.2	1.6	72.2	79.1	7.0
BougA	75.8	1.3	73.4	79.7	6.4
BougB	75.9	1.5	73.0	79.0	6.0
CGSA	75.7	1.5	72.4	79.2	6.8
CGSB	75.5	1.6	72.5	78.9	6.4
xG20RefA	75.3	1.4	72.6	78.8	6.2
xG20REfB	75.1	1.6	72.6	79.4	6.8
RTMA&CGSA	**75.4**	**1.3**	**72.6**	**78.8**	**6.2**
RTMB&CGSB	**75.4**	**1.6**	**72.3**	**79.0**	**6.6**
BougA&CGSA	**76.2**	**1.3**	**73.3**	**79.8**	**6.5**
BougB&CGSB	**76.2**	**1.5**	**73.2**	**79.4**	**6.2**

GSVS17					
Model	**Mean**	**STD**	**Min**	**Max**	**Range**
RTMA	-26.5	2.8	-32.4	-17.3	15.1
RTMB	-26.8	2.8	-32.1	-17.3	14.9
BougA	-24.8	1.8	-29.4	-21.0	8.4
BougB	-25.1	1.8	-30.2	-21.0	9.2

CGSA	-28.1	1.6	-31.9	-23.3	8.6
CGSB	-27.0	2.3	-32.7	-22.4	10.4
xG20RefA	-27.1	2.9	-35.7	-21.9	13.8
xG20REfB	-27.4	3.0	-37.2	-22.5	14.7
RTMA&CGSA	**-27.3**	**1.9**	**-30.7**	**-21.1**	**9.6**
RTMB&CGSB	**-26.9**	**2.1**	**-31.1**	**-22.6**	**8.5**
BougA&CGSA	**-26.3**	**1.4**	**-30.0**	**-23.7**	**6.4**
BougB&CGSB	**-25.9**	**2.0**	**-30.8**	**-21.6**	**9.3**

Each GSVS leveling line is tied to a single NAVD 88 mark. Therefore, the biases for each GSVS line are mainly the systematic errors in the NAVD 88 heights. Excluding the biases, the combined four geoid models agree with the GPS/leveling-derived geoid heights between 1 and 2 cm (STD) for the GSVS11, GSVS14 and GSVS17. This agreement is more than three times better than with the historical GPS/leveling dataset. The improved agreement is very encouraging in terms of the quality of the models and validation data. Further improvement in agreement is possible, but will be difficult. The computation of geoid models with cm accuracy is challenging given current data accuracy and distribution, and collecting cm-accurate independent validation data (e.g., GSVS) is also a demanding effort and its improvement is limited by instruments and observations techniques.

Figures 2a – 2c show the agreement between the geoid models in Texas (low and flat), Iowa (medium elevation), and Colorado (high mountains). The reference models xG20RefA and xG20RefB are included as control.

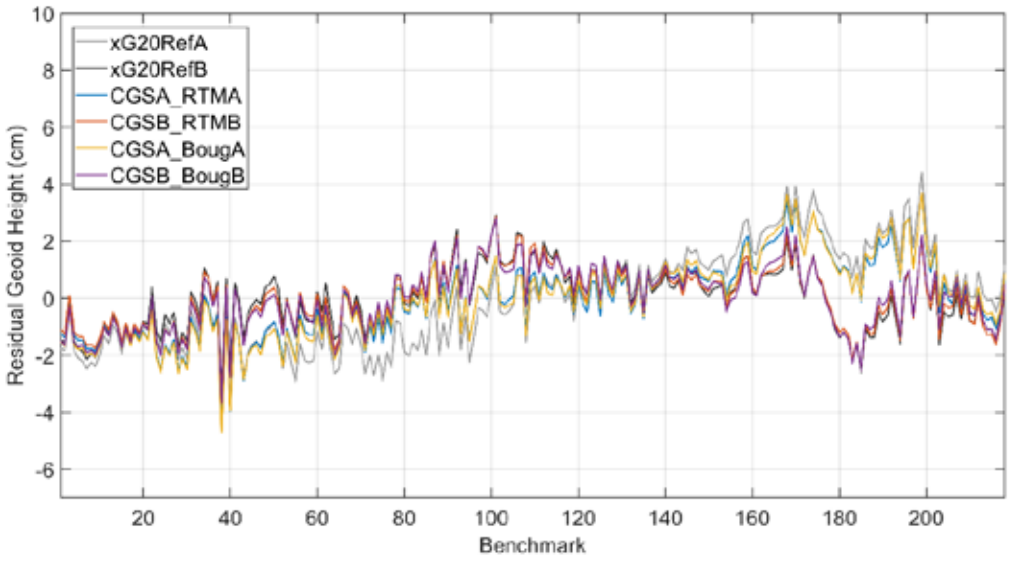

Figure 2a. Geoid height differences: Model - GSVS11.

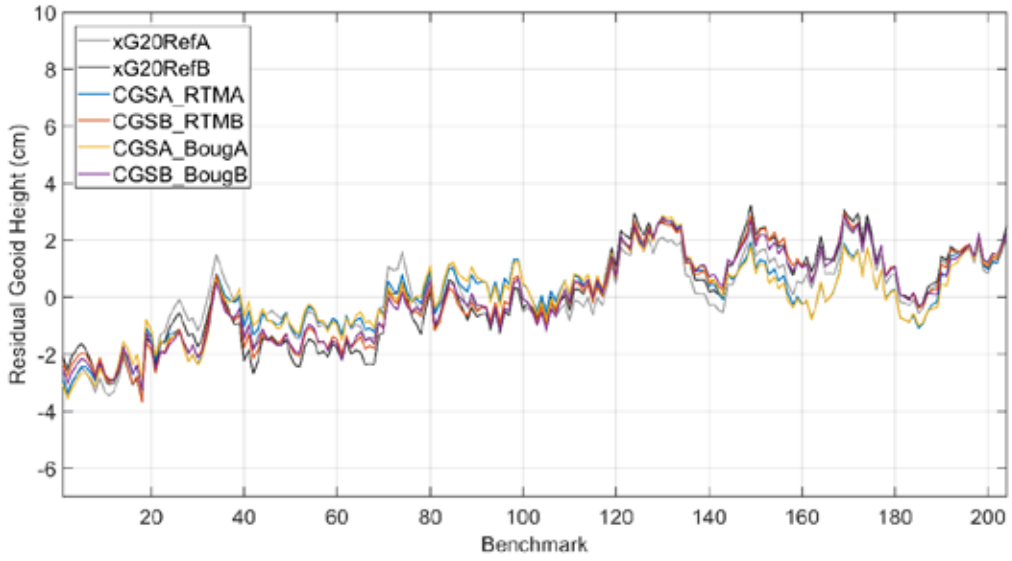

Figure 2b. Geoid height differences: Model - GSVS14.

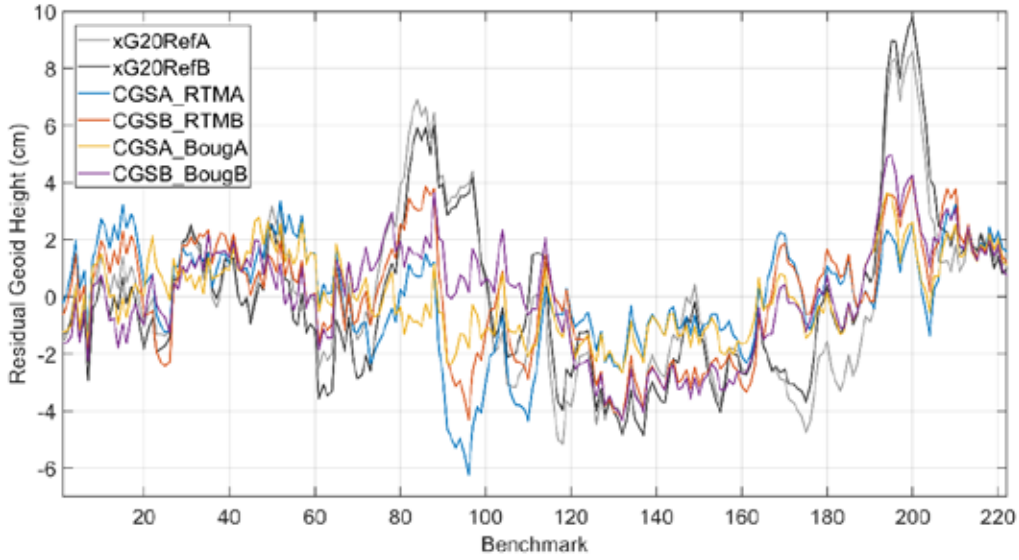

Figure 2c. Geoid height differences: Model - GSVS17.

Figures 2a, 2b, and 2c show the geoid models agree with the GSVS data sets impressively well, not only in low and flat areas (GSVS11), but also in high mountains (GSVS17). In both cases, there are no significant slopes observed in the differences. However, there is about 4 cm tilt on the Iowa traverse (GSVS14). It is unknown whether the slope is in the GSVS data, geoid models or in both. Further analysis is needed to identify the cause of the tilt. Another observation is that the reference models xGEOID20RefA/B perform as well in flat area and area with medium elevations (Figure 2a and 2b), but they have larger differences at the two peaks in Fig 2c.

Another comparison is the geoid slope comparisons (Smith et al. 2013; Wang et al. 2017; van Westrum 2021). The following figures show the RMS of the geoid slopes at different baseline lengths.

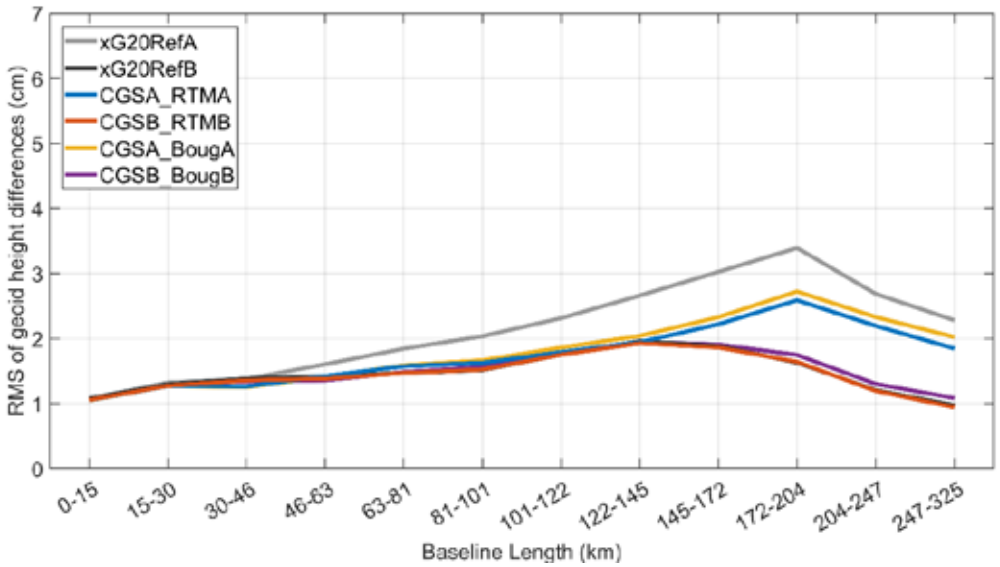

Figure 3a. RMS of geoid differences at different baseline lengths on the GSVS11 line.

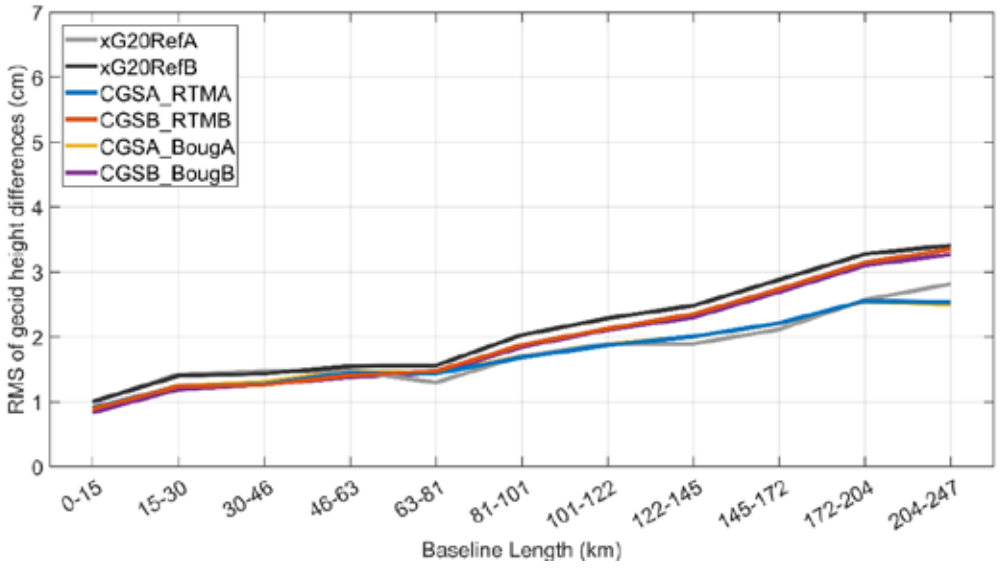

Figure 3b. RMS of geoid differences at different baseline lengths on the GSVS14 line.

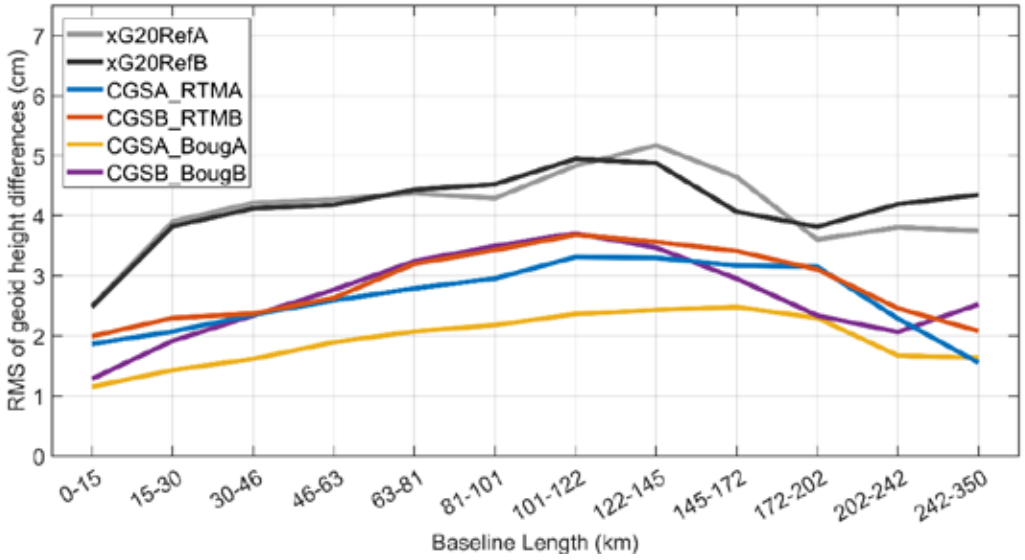

Figure 3c. RMS of geoid differences at different baseline lengths on the GSVS17 line.

Figures 3a, 3b, and 3c show that the RMS of geoid differences are around 1 to 5 cm at every baseline length for the three GSVS lines. The RMS of the geoid differences are listed in Table 4.

Table 4. RMS of geoid differences at different baseline lengths. All units in cm.

GSVS11

Baseline (km)	CGSA& RTMA	CGSB& RTMB	CGSA& BougA	CGSB& BougB	xG20RefA	xG20RefB
0-15	1.1	1.1	1.0	1.1	1.1	1.1
15-30	1.3	1.3	1.3	1.3	1.3	1.3
30-46	1.3	1.4	1.3	1.3	1.4	1.4
46-63	1.4	1.4	1.4	1.4	1.6	1.4
63-81	1.6	1.5	1.6	1.5	1.8	1.5
81-101	1.6	1.5	1.7	1.6	2.0	1.5
101-122	1.8	1.8	1.9	1.8	2.3	1.8
122-145	1.9	1.9	2.0	1.9	2.7	2.0
145-172	2.2	1.9	2.3	1.9	3.0	1.9
172-204	2.6	1.6	2.7	1.7	3.4	1.6
204-247	2.2	1.2	2.3	1.3	2.7	1.2
247-325	1.8	0.9	2.0	1.1	2.3	1.0

GSVS14

Baseline (km)	CGSA& RTMA	CGSB& RTMB	CGSA& BougA	CGSB& BougB	xG20RefA	xG20RefB
0-15	0.9	0.9	0.9	0.8	1.0	1.0
15-30	1.2	1.2	1.2	1.2	1.4	1.4
30-46	1.3	1.3	1.3	1.3	1.5	1.4
46-63	1.4	1.4	1.5	1.4	1.5	1.5
63-81	1.4	1.5	1.4	1.4	1.3	1.6
81-101	1.7	1.9	1.7	1.8	1.7	2.0
101-122	1.9	2.1	1.9	2.1	1.9	2.3
122-145	2.0	2.4	2.0	2.3	1.9	2.5
145-172	2.2	2.7	2.2	2.7	2.1	2.9
172-204	2.6	3.1	2.5	3.1	2.6	3.3
204-247	2.5	3.4	2.5	3.3	2.8	3.4

GSVS17

Baseline (km)	CGSA& RTMA	CGSB& RTMB	CGSA& BougA	CGSB& BougB	xG20RefA	xG20RefB
0-15	1.9	2.0	1.2	1.3	2.5	2.5
15-30	2.1	2.3	1.4	1.9	3.9	3.8
30-46	2.4	2.4	1.6	2.3	4.2	4.1
46-63	2.6	2.6	1.9	2.8	4.3	4.2
63-81	2.8	3.2	2.1	3.2	4.4	4.4
81-101	3.0	3.4	2.2	3.5	4.3	4.5
101-122	3.3	3.7	2.4	3.7	4.8	4.9
122-145	3.3	3.6	2.4	3.5	5.2	4.9
145-172	3.2	3.4	2.5	3.0	4.6	4.1
172-202	3.2	3.1	2.3	2.3	3.6	3.8
202-242	2.3	2.5	1.7	2.1	3.8	4.2
242-350	1.6	2.1	1.6	2.5	3.8	4.3

The combination models with the weights based on the GPS/leveling differences are also compared with the GSVS data, but the results are not presented as the differences are at sub-mm level to those in Table 4.

4.3 Mean lake surface heights of the Great Lakes from the satellite altimetry

Satellite altimeters have been measuring the instantaneous lake surface height for decades. Merged TOPEX/Poseidon/Jason1/OSTM (Jason2) (TPJO) altimetric data (Beckley et al. 2013), from NASA's MEaSUREs project (https://podaac.jpl.nasa.gov), at 1 Hz rate (~5.8 km), were used to compute the mean lake surface height along the ground tracks. The TPJO time series has 822 ten-day repeat cycles from 1992 until the present time. The first 10 cycles are known to have had an off-nadir attitude problem (B. Beckley, personal communication, 2015). Only the

remaining cycles from 11 to 822 were used to form the mean lake surface height along the tracks (Li et al. 2016) and used for evaluation. After careful data editing and averaging, the mean lake surfaces along each altimetry track are good approximations of equipotential surfaces. Thus, the dynamic heights along these tracks should be close to a constant. There are 16 TOPEX/Poseidon tracks that cross the Great Lakes (Figure 4a). The standard deviations of the altimetry-derived dynamic heights are plotted in Figure 4b. As the control, the comparison results of xG19RefA/B are included too.

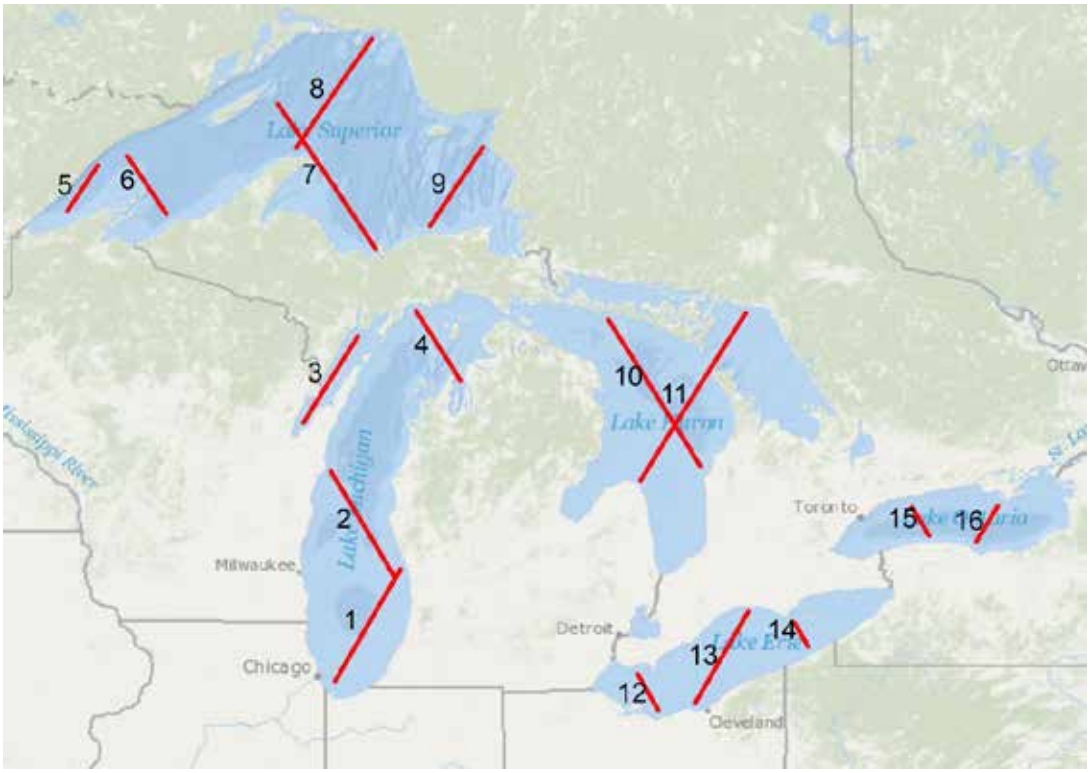

Figure 4a. TOPEX/Poseidon Ground track over the Great Lakes.

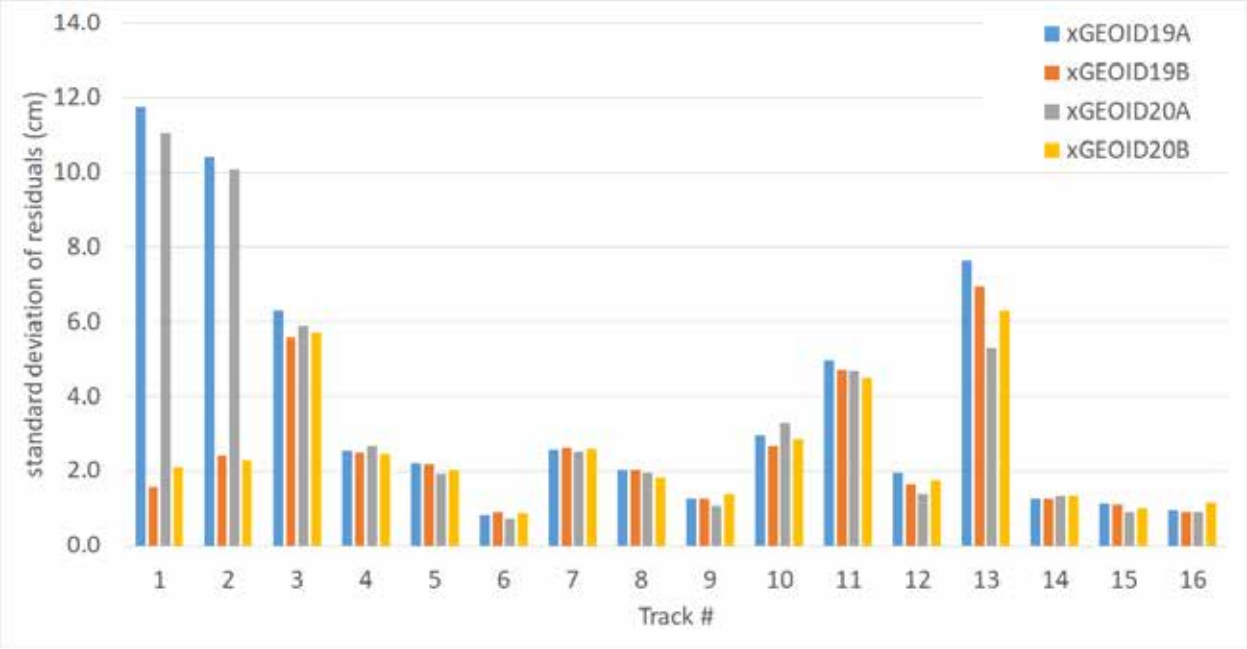

Figure 4b. Standard deviation of dynamic height along 16 TOPEX/Poseidon altimetry tracks. Track 1, 2 and 3 are over Lake Michigan, where the largest improvement in the GRAV-D data enhanced models is seen.

The interpretation of Figure 4b must be done carefully because of the varying quality of the mean lake surface heights from the altimetry data. Some tracks are mostly long segments over deep water, but some segments are short over shallow water. Nevertheless, a few-cm agreement between the altimetry data and the xG20RefB imply overall good quality of altimetry data and geoid models. The significant improvement of xG20RefB over xG20RefA over Lake Michigan (track 1, 2 and 3) can be explained by the contribution of the GRAV-D data. This shows well the usefulness of airborne gravity in areas with problematic shipborne data and lake bottom gravity measurements.

There are other satellite altimeter missions available and it can improve the coverage of the TOPEX/Poseidon missions over the lakes. A more complete dataset of satellite altimetry data with estimated accuracy would be useful for future validation of geoid models.

4.4 The Canadian Great Lakes water gauge data

The water level has been monitored at gauges around the Great Lakes for decades. There is a total of 21 Canadian permanent gauges on lakes Superior (4), Huron (6), Erie (6) and Ontario (5). These gauges can be used in the evaluation of geoid models. Seven years (2007–2013) of data for the summer months (June–September) are used to determine the mean water level above the chart datum at each gauge. The water level can change significantly at the meter level due to flooding, drought and water management. However, the changes are for the whole lake and should affect all gauges in the same way. Thus, the mean lake surface can be approximated as an equipotential surface.

GPS campaigns have been conducted at all the permanent gauges on the lakes. The GPS survey is either directly on a gauge mark or a near-by benchmark tied by levelling to a gauge mark. The

GPS coordinates ($\phi, \lambda. h$) are in ITRF2008 epoch 2010.0. The epoch difference (~0.5 years) between mean water level and GPS coordinates is neglected. First, the ellipsoidal heights (h) are transformed to orthometric heights (H) by subtracting the geoid undulations (N) from the model to be validated. If the GPS station is not directly on a gauge mark, the orthometric height is transferred to the gauge mark by adding the levelling height difference (ΔH). Second, the orthometric height of the mean water level with respect to the geoid (H^{MWL}) is calculated by subtracting the height of the gauge mark above the chart datum (H_{CD}) and adding the height of the mean water level above the chart datum (Z_0). Finally, the orthometric height of the mean water level is converted to dynamic heights (H^D). In an ideal case, the dynamic heights at each gauge would be the same on a lake if a lake is an equipotential surface.

The calculation of the dynamic height can be summarized as

$$H^D = [(h - N) + \Delta H - H_{CD} + Z_0]\bar{g}/\gamma_{45} \quad (4)$$

where γ_{45} is the normal gravity (GRS80) at 45 degrees latitude, $\bar{g}$ is the mean gravity along the plumbline between the geoid and the surface and it is approximated by (Heiskanen and Moritz 1967)

$$\bar{g} = g + 0.0424H, \; g \text{ in gal and } H \text{ in km.} \quad (5)$$

The validation of the geoid models is done by analyzing the difference in heights at all gauges on a same lake. Table 5 gives the statistics of the geoid differences between the model and the water gauge for Lakes Superior, Huron, Erie and Ontario. Results highlighted in bold show best combination solutions for each lake. The height discrepancies between the gauges on a lake come from error in the GPS ellipsoidal heights, water measurements and geoid models, and could include lake surface topography too (i.e., permanent topography deviating from an equipotential surface).

Table 5. Statistics of the residual geoid heights (model—water gauge). Units in meters.

Superior: 4 stations, 72 samples					
Model	**Mean**	**STD**	**Min.**	**Max.**	**Range**
RTMA	182.624	0.019	182.595	182.667	0.072
RTMB	182.628	0.021	182.600	182.677	0.077
BougA	182.578	0.044	182.525	182.670	0.145
BougB	182.581	0.047	182.527	182.680	0.153
CGSA	182.597	0.020	182.569	182.638	0.069
CGSB	182.596	0.024	182.552	182.637	0.085
xG20RefA	182.581	0.021	182.544	182.623	0.079
xG20REfB	182.584	0.023	182.549	182.633	0.084
RTMA&CGSA	**182.600**	**0.019**	**182.572**	**182.635**	**0.063**
RTMB&CGSB	**182.601**	**0.022**	**182.566**	**182.648**	**0.082**
BougA&CGSA	**182.578**	**0.030**	**182.540**	**182.637**	**0.097**
BougB&CGSB	**182.579**	**0.031**	**182.546**	**182.649**	**0.103**

Huron: 6 stations, 115 samples					
Model	**Mean**	**STD**	**Min.**	**Max.**	**Range**
RTMA	175.667	0.025	175.594	175.708	0.114
RTMB	175.673	0.022	175.618	175.710	0.092
BougA	175.640	0.021	175.599	175.682	0.083
BougB	175.645	0.019	175.612	175.684	0.072
CGSA	175.646	0.018	175.603	175.680	0.077
CGSB	175.655	0.019	175.607	175.691	0.084
xG20RefA	175.632	0.024	175.568	175.670	0.102
xG20REfB	175.638	0.020	175.593	175.672	0.079
RTMA&CGSA	**175.649**	**0.021**	**175.591**	**175.687**	**0.096**
RTMB&CGSB	**175.657**	**0.020**	**175.606**	**175.692**	**0.086**
BougA&CGSA	**175.636**	**0.019**	**175.594**	**175.674**	**0.080**
BougB&CGSB	**175.643**	**0.019**	**175.608**	**175.679**	**0.071**

Erie: 6 stations, 58 samples					
Model	**Mean**	**STD**	**Min.**	**Max.**	**Range**
RTMA	173.731	0.018	173.694	173.762	0.068
RTMB	173.730	0.018	173.693	173.760	0.067
BougA	173.699	0.030	173.646	173.746	0.100
BougB	173.698	0.031	173.647	173.743	0.096
CGSA	173.717	0.012	173.696	173.753	0.057
CGSB	173.717	0.019	173.677	173.747	0.070
xG20RefA	173.697	0.019	173.659	173.728	0.069
xG20REfB	173.696	0.018	173.658	173.725	0.067
RTMA&CGSA	**173.718**	**0.015**	**173.691**	**173.752**	**0.061**
RTMB&CGSB	**173.717**	**0.018**	**173.679**	**173.747**	**0.068**
BougA&CGSA	**173.703**	**0.021**	**173.669**	**173.744**	**0.075**
BougB&CGSB	**173.702**	**0.024**	**173.660**	**173.739**	**0.079**

Ontario: 5 stations, 74 samples					
Model	**Mean**	**STD**	**Min.**	**Max.**	**Range**
RTMA	74.498	0.021	74.458	74.526	0.068
RTMB	74.499	0.023	74.452	74.531	0.082
BougA	74.477	0.025	74.431	74.513	0.082
BougB	74.478	0.028	74.427	74.520	0.093
CGSA	74.477	0.025	74.428	74.511	0.083
CGSB	74.477	0.027	74.423	74.515	0.092
xG20RefA	74.463	0.027	74.415	74.502	0.087
xG20REfB	74.464	0.030	74.409	74.509	0.100
RTMA&CGSA	**74.481**	**0.023**	**74.437**	**74.512**	**0.075**
RTMB&CGSB	**74.482**	**0.025**	**74.431**	**74.516**	**0.085**
BougA&CGSA	**74.471**	**0.025**	**74.424**	**74.505**	**0.081**
BougB&CGSB	**74.472**	**0.027**	**74.419**	**74.512**	**0.093**

The GPS/water level data comparison shows again good agreement over all lakes. The STD of geoid differences range between 1 and 3 cm, showing high accuracies of the geoid models and the tide gauge datasets.

4.5 Conclusions of xGEOID20 validation

Based on the validation of the geoid models against independent GSVS, satellite altimeter and lake tide gauge data, the combination models CGSA&RTMA and CGSB&RTMB are selected as the xGEOID20A/B models. The combination models that include the NGS Boug type models are not selected largely due to their problems in the Great Lakes region. Geoid model accuracy in the Great Lakes region is of primary significance and, in selecting the optimal model, Great Lakes altimetry and tide gauge agreement outweighs agreement of the NGS Boug type models with CGS models in the Rocky Mountains and GSVS17 validation data set.

5. Computation and evaluation of xDEFLEC20

xGEOID20 is accompanied by xDEFLEC20. It is two grids ($1' \times 1'$) for the two components of the deflections of the vertical (DoV). They are calculated at the topographic surface from the xGEOID20B model. A deflection of the vertical describes the deviation of the plumb line at the topographic surface with respect to the ellipsoid normal. The computation of xDEFLEC20 follows the methods described in Ahlgren et al. (2020).

The deflection of the vertical is defined as the difference between astronomic and geodetic coordinates (Heiskanen and Moritz 1967):

$$\eta = (\Lambda - \lambda)\cos\phi, \tag{6}$$

$$\xi = \Phi - \phi, \tag{7}$$

where η and ξ are the east-west and north-south components of DoV; Λ and Φ are astronomic longitude and latitude (which describe the direction of the plumb line); and λ and ϕ are geodetic longitude and latitude (which describe the direction of the ellipsoid normal).

The deflection components are computed from the $1'$ geoid grid, Bouguer gravity anomaly, and topographic elevation. The first step in the computation is the numerical derivative of the xGEOID20B grid:

$$\eta_N = -\frac{1}{R\cos\phi}\frac{\partial N}{\partial \lambda}, \tag{8}$$

$$\xi_N = -\frac{1}{R}\frac{\partial N}{\partial \phi}. \tag{9}$$

where R is the mean Earth radius. The gradients of the geoid grid are complemented with the Bouguer plumb line curvature term to obtain the surface deflection of the vertical (cf. Heiskanen and Moritz 1967, Eq. (5-32)):

$$\eta = \eta_N + \frac{1}{R\cos\phi}\frac{H}{\bar{g}}\left(\frac{\partial \bar{g}}{\partial \lambda} + k_H\frac{\partial H}{\partial \lambda}\right) \tag{10}$$

$$\xi = \xi_N + \frac{1}{R}\frac{H}{\bar{g}}\left(\frac{\partial \bar{g}}{\partial \phi} + k_H\frac{\partial H}{\partial \phi}\right) \tag{11}$$

where $k_H$is the Helmert constant, 0.0424 mGal m^{-1}.

The Bouguer gravity grid used in the calculation is the same as for xGEOID19. This calculation uses 1-arcminute ETOPO1 for its DEM which differs from the DEM used for xGEOID20. As deflections of the vertical are not observably sensitive to elevation changes of less than 100 meters, these differences are not expected to introduce much error. The input grids for this product will be harmonized in future iterations.

xDEFLEC20 is evaluated using observed deflections of the vertical, including historical astronomical deflections of the vertical, astronomical deflections of the vertical measured on GSVS lines with the CODIAC zenith camera, and lake-surface deflections observed with ICESat-2's laser altimeter.

The most precise and accurate DoV observations come from the GSVS lines in Texas (2011), Iowa (2014) and Colorado (2017). These deflections of the vertical are observed with a precision of 0.05″. On the GSVS11 and GSVS14 lines, xDEFLEC20 performs comparably with xDEFLEC19. Along GSVS11, xDEFLEC20 has a residual standard deviation of 0.18″ in η and 0.15″ in ξ. Similarly, xDEFLEC20's GSVS14 residuals have standard deviations of 0.24″ in η and 0.28″ in ξ. Along both lines, the overall bias is less than 0.1″ in any direction. These results indicate that xDEFLEC20 can distinguish slopes on the order of 1 mm per km (0.21″) in the topography of the eastern United States.

The rugged terrain of Colorado captured by GSVS17 proves more challenging for xDEFLEC20, with standard deviations of the residuals of more than 0.8″ in either direction. However, these results are a dramatic improvement over xDEFLEC19, which have residual standard deviations of 2.2″ in η and 1.9″ in ξ. These improvements are thought to come from changes in the derivative technique, averaging the NGS geoid model with the CGS geoid model and the improved xG20DEM.

Table 6. Statistics for residual deflections of the vertical on GSVS lines (GSVS - Model, arcseconds).

	xDEFLEC19		xDEFLEC20	
GSVS11	η	ξ	η	ξ
Mean	0.05	-0.06	0.08	-0.05
STD	0.16	0.17	0.18	0.15
Min.	-0.38	-0.62	-0.44	-0.59
Max.	0.69	1.03	0.48	0.73
GSVS14	η	ξ	η	ξ
Mean	0.05	-0.03	0.04	0.01
STD	0.26	0.31	0.24	0.28
Min.	-0.57	-0.60	-0.51	-0.51
Max.	0.86	0.88	0.78	0.77
GSVS17	η	ξ	η	ξ
Mean	-0.38	0.06	-0.10	0.25
STD	2.21	1.93	0.82	0.87
Min.	-10.06	-9.65	-2.40	-2.13
Max.	9.29	6.70	3.21	4.80

The deflection product is also compared with 3,991 historical astronomical measurements. Within the conterminous United States (CONUS), the overall RMS weighted by inverse squared observational error (wRMS) of the residuals is 0.8″ in η and 0.7″ in ξ. Dividing CONUS according to time zone into eastern (Eastern and Central) and western (Mountain and Pacific) states, xDEFLEC20 predicts the historical deflection of the vertical in the eastern zones with a residual wRMS of 0.5″ in any direction, but 1.2″ and 0.9″ in η and ξ, respectively, in the more rugged western zones. These wRMS are comparable to those from xDEFLEC19 within 0.1″. In Alaska, the wRMS for the 47 stations improve substantially over xDEFLEC19's: 1.3″ to 1.1″ in η and and 2.9″ to 2.0″. The 97 astronomical sites in mainland Canada have a wRMS of 1.1″ in η and 0.8″ in ξ. The full regional breakdown of regional residual wRMS is given in Table 7.

xDEFLEC20 is also compared with lake surface slopes measured with ICESat-2's six parallel laser altimeters. This unique altimetry setup enables ICESat-2 to measure lake surface slopes in both the along-track and cross-track directions across a 6 km baseline with precision comparable to astronomical measurements. This technique is limited by the cross-track error in ICESat-2's attitude and the possibility of hydraulic and wind-driven deviations of the water surface from an equipotential surface. The uncertainty in η tends to be larger than the uncertainty in ξ because the slope measurements are made based on range measurements from three different pairs of laser beams. While measurements of ξ mostly depend on comparing along-track range measurements from the same beams across time, the east-west measurements come from comparing cross-track measurements from different beams with distinct biases. More details of the technique are given in the xGEOID19 technical details (Li et al. 2019). The most notable differences in this year's comparison are that the ICESat-2 dataset has been expanded to include 19,722 lake surface slopes globally. Of these, 13,770 in North America are used to evaluate xDEFLEC20. In CONUS, 6,204 surface slopes are used, with 6,096 in eastern states with abundant surface water.

These eastern sites had wRMS residuals of 1.3″ in η and 0.9″ in ξ. The more arid west has only 108 sites with extensive enough surface water to yield a valid measurement, with degraded agreement in η.

ICESat-2 offers little additional benefit over CONUS, where astronomical observations are abundant. However, in Alaska and Canada, where few deflections of the vertical are available to NGS, the ICESat-2 dataset fills extensive gaps. 5,770 observations are available over the many lakes of the Canadian Shield. These observations have a wRMS agreement with xDEFLEC20 of 1.2″ in η and 0.9″ in ξ, which is comparable with observations in CONUS. In Alaska, the 142 lake surface slopes reveal wRMS agreement of 1.5″ in η and 1.2″ in ξ. Furthermore, ICESat-2 provides critical validation over the Great Lakes with 1,654 deflection measurements. These comparisons are shown alongside astronomic comparisons in Table 7.

Table 7. Weighted RMS residuals of ICESat-2 and historical deflections-of-the-vertical comparisons with xDEFLEC20, units in arcsecond.

wRMS	**ICESat-2**			**Astronomical**		
	Points	η	ξ	Points	η	ξ
CONUS	6,204	1.27	0.86	3,991	0.82	0.69
CONUS East	6,096	1.26	0.85	2,164	0.46	0.52
CONUS West	108	1.61	0.89	1,827	1.15	0.86
Alaska	142	1.54	1.17	47	1.06	1.96
Great Lakes	1,654	1.23	0.65	0	n/a	
Canada	5,770	1.24	0.86	97	1.10	0.83

6. Uncertainty estimation of xGEOID20 and xDEFLEC20

xGEOID20 is the first NGS experimental geoid model to include a grid of uncertainty estimates for the geoid undulations and both components of the deflections of the vertical. Uncertainty estimates indicate the confidence in the modeled geoid undulations at a given location. From the perspective of the user, they illustrate the degree of caution that should be applied when using GPS-derived orthometric heights to predict geopotential differences. For modelers, this grid helps tune expectations for performance in the absence of validation data and highlights the impacts of region-specific data sparsity.

6.1 Forward error grids

The 5′ forward error grid accompanies the results of the xGEOID20 geoid undulations. This uncertainty estimate is generated with a spectral combination of linear error propagation of the Bouguer anomaly error grid and expanded GOCO06s spherical-harmonic errors. Following Pavlis et al. (2012), the errors are estimated as the sum of a high-frequency component and a low-frequency component. The high-frequency component $\sigma_H^2(\mathrm{N})$ is developed by propagating gridded gravity errors from satellite-altimetry, terrestrial, and other dense data sources.

The low-frequency component $\sigma_L^2(\mathrm{N})$ is derived from the formal errors of GOCO06s.

$$\sigma^2(\mathrm{N}) = \sigma_L^2(\mathrm{N}) + \sigma_H^2(\mathrm{N}) \quad (12)$$

The high- and low-frequency components of the error model are tapered in the spectral domain with a complementary weighting function that matches the function used to combine GOCO06s and the surface input data in the development of xGEOID20.

The first step in the error estimation process is to assemble a 5′ grid of Bouguer gravity anomaly errors from the available input data. It contains error contributions from the available terrestrial gravity point data and satellite altimetry grids. Where neither altimetry nor terrestrial data is given in a particular grid cell, the formal free-air-anomaly errors from EGM2008 are used instead.

For a 5′ grid cell i with either terrestrial gravity or gridded altimetry-derived gravity data with respective errors $\sigma^2(\Delta g_{\mathrm{Terrestrial}})_{i,j}$ and $\sigma^2\left(\Delta g_{\mathrm{Altimetry}}\right)_{i,j}$ for the jth data point in the cell, the total error associated with that cell is given as the harmonic mean of the formal errors of each contributing data point.

$$\frac{1}{\sigma^2(\Delta g)_i} = \sum_{\mathrm{j=1}}^{\mathrm{n_{Terrestrial,\,i}}} \frac{1}{\sigma^2(\Delta g_{\mathrm{Terrestrial}})_{i,j}} + \sum_{j=1}^{n_{\mathrm{Altimetry,i}}} \frac{1}{\sigma^2(\Delta g_{\mathrm{Altimetry}})_{i,j}} \quad (13)$$

This approach approximates the standard error of the mean of data within that cell. The grid is illustrated in Figure 5.

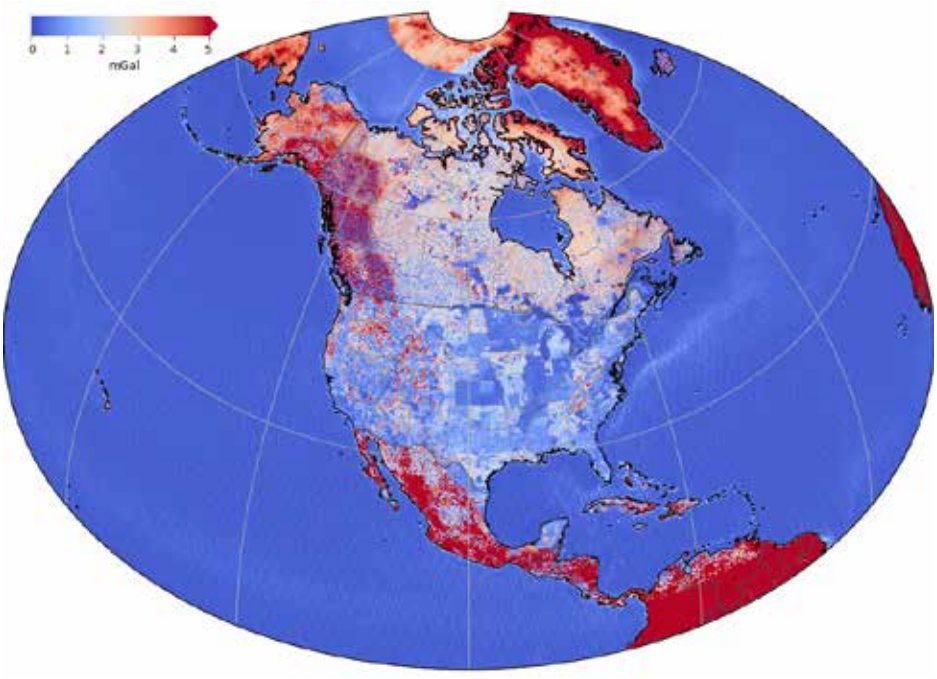

Figure 5. xGEOID20 Bouguer gravity error grid used for the high-frequency component of the geoid error.

The next step is computing the high-frequency component of the geoid error. Following Pavlis et al. (2012), this is given by

$$\sigma_H^2(\mathrm{N}) = \left(\frac{R}{4\pi\gamma}\right)^2 \iint_\Omega \sigma^2(\Delta g) S_H^2(\psi) d\Omega^2, \quad (14)$$

where γ is the mean surface gravity, ψ is the spherical distance between convolution points, and Ω is a solid angle over the spherical domain of Earth's surface.

The modified high-degree Stokes kernel is precomputed as follows:

$$S_H(\psi) = \sum_{l=2}^{l_{max}} w_l \frac{2l+1}{l-1} P_l(\psi) \quad (15)$$

The weight coefficients w_l describe the relative spectral weight of the reference model vs. the satellite model. These coefficients should approach 0 at low degrees and 1 at high degrees. For this operation, these coefficients will be identical to the weights determined for combining the terrestrial gravity data with GOCO06s. In this instance, the weights are defined by a cosine taper from $l = 150$ to $l = 220$. Weights at degrees above 220 equal 1 and weights below degree 150 equal 0. A 1-D FFT method is used to expedite the convolutions for each band of latitude, where the Stokes kernel is precomputed and interpolated.

The low-frequency component of the error grid is computed directly from the formal variance-covariance matrix of GOCO06s. This covariance matrix is projected into the spatial domain according to the following equation.

$$\sigma_L^2(\mathrm{N}) = R^2 \sum_{i,j} Y_{l(i)m(i)} Y_{l(j)m(j)} \sigma^2_{l(i)m(i),l(j)m(j)} (1 - w_{l(i)})(1 - w_{l(j)}) \quad (16)$$

The indices i and j signify the indices of the covariance matrix, which are uniquely tied to combinations of degree and order l and m, according to the arrangement of the covariance matrix. The spherical harmonic functions Y_{lm} are defined conventionally. The weighting coefficients w_l are the same as used to compute the high-frequency error coefficients. GOCO06s is available to degree and order 300, which means its variance-covariance matrix is too large for direct computation on personal computers. Its covariance matrix is given in approximately 1,000 2048x2048 blocks for tractable computation.

The resulting error grid is illustrated in Figure 6. It is predominantly zonal in nature, with higher precision near the North Pole due an abundance of tracking stations and crossovers for GRACE and GOCE. The error generally increases toward the equator, where the spacecraft ground tracks are farther apart. Within North America, typical values of the 1σ uncertainty are between ±0.5 and ±1.0 cm.

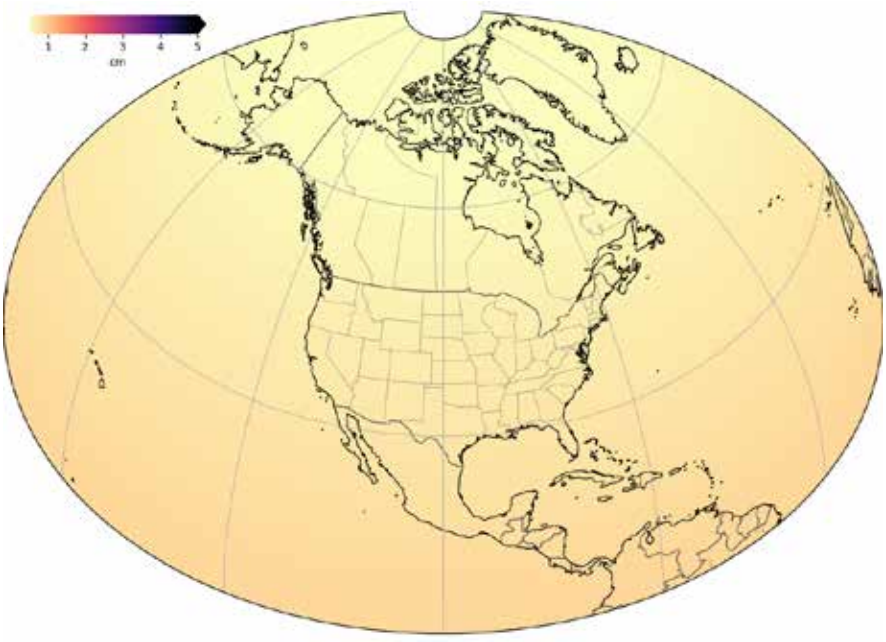

Figure 6. Low-frequency component of xGEOID20 1σ errors derived from the GOCO06s formal covariance matrix, which a cosine taper from degrees 150 to 220.

The final error grid, representing the sum of spectrally complementary 1-sigma errors from GOCO06s errors and high-density terrestrial gravity, is shown in Figure 7. The error map shows lower errors (1-2 cm) in the eastern the United States and higher errors (1-4 cm) in the western contiguous United States and southern Alaska. The errors largely reflect differences in the density of available terrestrial gravity and the impacts of topography.

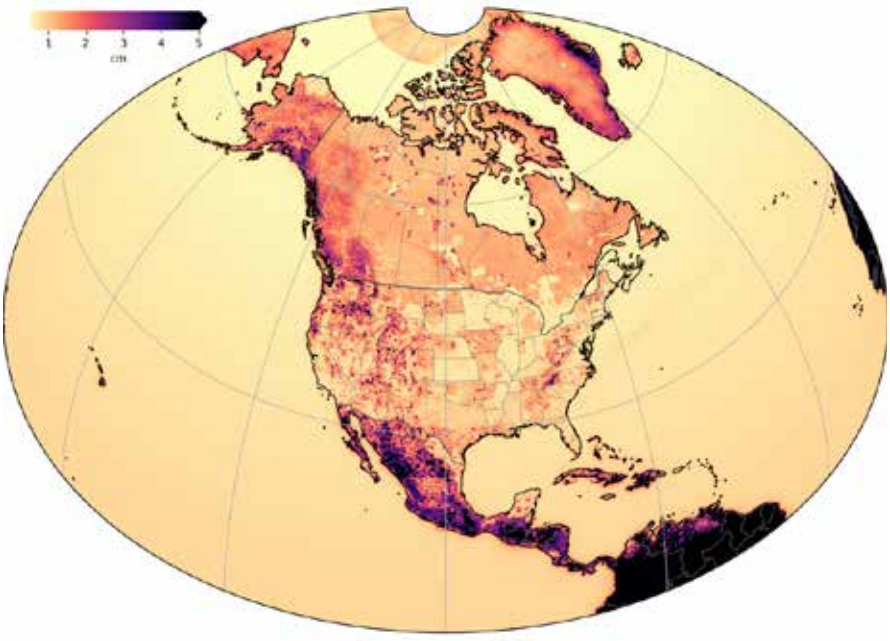

Figure 7. xGEOID20 1σ error grid, representing the total contributions of high-frequency and low-frequency geoid error terms.

Errors in the defections of the vertical are also computed using the high-frequency gravity error grid with a modified Stokes integration approach as before. The low-frequency component of the DoV uncertainty from GOCO06s is treated as a constant scalar variance estimated from the sum of the spectrally weighted degree variance. This scalar, equal to 0.066 arcseconds, is added to the high-frequency error variance grid to get the total error variance. RMS DoV errors are shown in Figure 8. In CONUS, they range from ±0.1-0.3″ in smooth terrain, but exceed ±0.5″ in mountainous terrain. They are especially high outside of CONUS, particularly in southern Alaska, where uncertainties exceed ±1″.

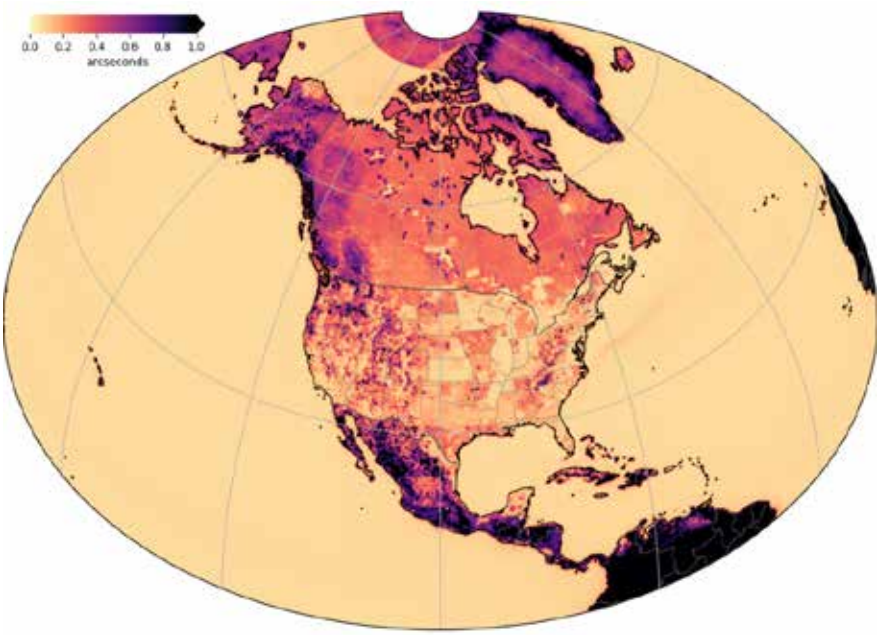

Figure 8. RMS of the east-west and north-south components of the xDEFLEC20 1σ deflection of the vertical error grid, indicating the typical error in deflection for a random azimuth.

These error grids were verified against the GSVS lines. If they are well estimated, their RMS errors along each of the lines should match the standard deviation of GPS/leveling and DoV residuals in Table 6. A comparison of the predicted GSVS line errors and the actual residuals is shown in Table 8. This comparison reveals that the predicted errors are meaningfully consistent in scale with the observed residuals, particularly along GSVS11 and GSVS14. The error grid may, however, be too optimistic along the rugged terrain of GSVS17. This may indicate the need for additional work to capture topographic contributions to these uncertainties.

Table 8. Comparison of predicted RMS xGEODI20 and xDEFLEC20 errors along GSVS lines with standard deviations of observed xGEOID20 and xDEFLEC20 residuals.

	GSVS11		GSVS14		GSVS17	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
Geoid (cm)	1.4	1.7	1.4	1.5	2.0	1.5
η (arcseconds)	0.18	0.30	0.24	0.25	0.82	0.26
ξ (arcseconds)	0.15	0.34	0.28	0.29	0.87	0.31

6.2 Empirical error grids

The 5′ forward error grid is only as accurate as its data inputs. Given the size and diversity of the data sources, there's no guarantee that the formal errors assigned to certain gravity datasets accurately reflect the true uncertainty of the data. For this reason, the geoid model uncertainty grids must be checked against the available validation data.

The most spatially abundant dataset for validating geoid models in North America is the GPS/leveling dataset. One of the major problems with working with the GPS/leveling dataset is that leveling is a relative measurement. Leveling errors tend to increase with the distance between benchmarks. This presents challenges for comparing geoid model performance between different regions. The standard deviation of GPS/leveling residuals for California will naturally be larger than residuals in Massachusetts because the difference in size between the states results in differences in the average baseline between stations.

A remedy for this is computing geoid model performance statistics within equal-area cells. This is functionally equivalent to high-pass filtering the GPS/leveling residuals. A convenient gridding scheme for this purpose is a 1-degree reduced Gaussian grid developed for the Goddard Space Flight Center mascon solutions (Loomis et al. 2019). This gridding scheme is chosen for expediency as we already have software written to manipulate and visualize this grid—other gridding schemes may be considered. The differential error contribution from leveling at this scale approaches 1 cm. As such, these error grids unavoidably overestimate the true geoid model error, but keep instrumental contributions to the residuals consistent across regions.

With GPS/leveling residuals aggregated into 1-degree cells, the performance of the geoid model may be compared between regions and states. A map of the standard deviation of residuals within cells is given in Figure 9. This grid may be compared with the forward error grid discussed earlier. A down-sampled version of the forward error grid, computed from the RMS of predicted 1-sigma errors within the 1-degree cells, is shown alongside the empirical grid in Figure 9.

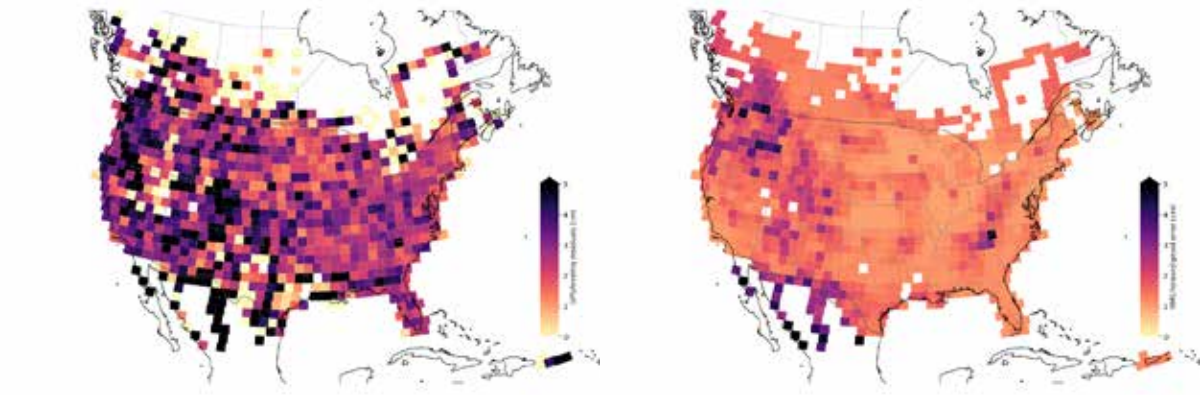

Figure 9. xGEOID20 errors resampled to 1-degree equal-area bins. Left: Empirical error estimates derived from binned GPS/leveling residuals from xGEOID20B. Right: Down sampled forward error grid.

The two error maps show similar structure, but different scale. The salient point of resemblance between these grids is higher errors in western CONUS than the east. Accounting for degrees of freedom and the number of data points in each cell, the RMS standard deviation of GPS/leveling

residual variation within cells is 2.8 cm. The corresponding RMS of cell-averaged forward errors across the same spatial extent is 1.8 cm. The difference between these uncertainties is largely attributable to combined GPS and leveling uncertainties of ±2 cm.

The empirical errors are also aggregated by state as shown in Figure 10. In contrast with the results in Table 3, pre-aggregating the residuals in 1-degree cells removed much of the scale-dependent effects on leveling errors, ensuring that these errors primarily reflect geoid model uncertainty.

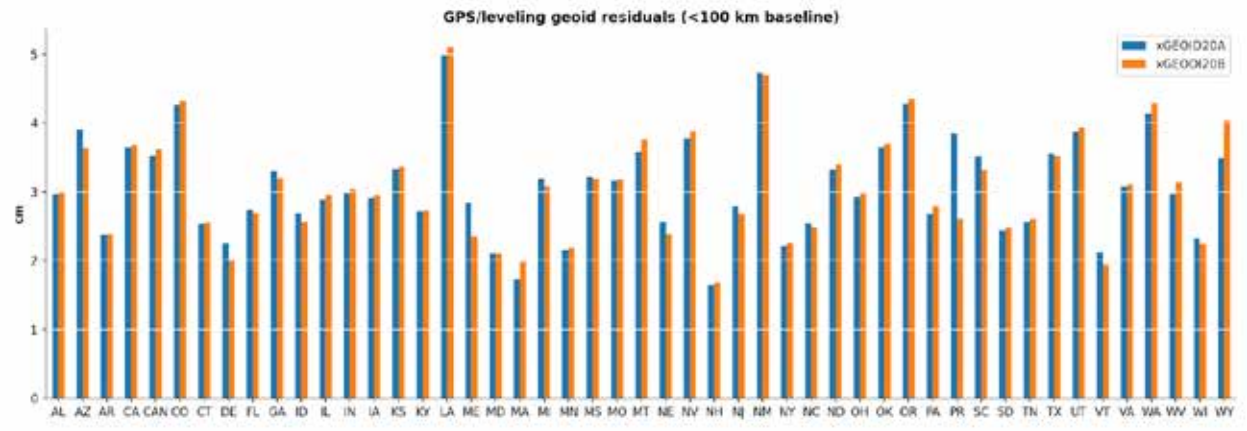

Figure 10. RMS of GPS/leveling residual standard deviations with respect to xGEOID20A and xGEOID20B aggregated within 1-degree cells and then aggregated by state.

6.3 Discussion of error grids

The forward and empirical error grids represent a first step toward quantifying the uncertainty of NGS's geoid products. Comparison against available validation data suggests that the current grid is largely consistent with observed errors, but may need additional work to capture the effects of rugged terrain. The GSVS GPS/leveling and deflection of the vertical datasets provide the most locally precise check of these grids, while NGS's GPS/leveling and astrogeodetic datasets provide the most spatially abundant checks. Error grids estimated from GPS/leveling should broadly resemble the error grids estimated through the forward error propagation approach. Their convergence is the best indicator of an accurate error grid.

A stochastic approach, such as bootstrapping (Mooney et al. 1993), where an array of quasigeoids are computed from resampled input gravity data, may also serve to check the forward error estimates for errors inherent in technique. The contributions from GRAV-D have been ignored in these uncertainty estimates, but the value of their contribution should be considered in the future. Future versions of this error grid should also consider alternatives to the nominal formal errors as input datasets for error calculations.

Finally, the error estimates presented here represent uncertainties in the components of the Bouguer quasigeoid up to degree and order 2160. These data-driven uncertainties are assumed to be dominant. However, as we refine this component of the error estimate, we should revisit other

contributions to the geoid error, such as geoid-quasigeoid separation and the indirect effect of the topography.

7. Summary and Conclusions

The xGEOID20 models are computed jointly by NGS and CGS. The final models are the simple average of the NGS and CGS models. The combined models are evaluated by using the GPS/leveling data sets (the historical and GSVS 11/14/17), the mean lake surface heights from nearly three decades TOPEX/Poseidon altimetry data over the Great Lakes, and the water gauges over Lake Superior, Huron, Erie and Ontario. The xGEOID20-derived deflections of the vertical are also compared with the ICESat-2 over inland water bodies and the historical astronomical observations. The validation data sets have different accuracies, which are reflected in the magnitude of the geoid height differences. For instance, the STDs of the differences between the xGEOID20 and historical GPS/leveling data range from 2 to 14 cm state by state, about 6 cm nationwide. However, the STDs of the differences are much smaller (around 2 cm) on marks of three GSVS lines over 300 km of length. The small discrepancies indicate high quality of the geoid models and the GSVS GPS/leveling data. The mean lake surface heights from the altimetry data over the Great Lakes have different accuracies due to their locations, thus the STDs of discrepancies are different from track to track, but still in the range of a few cm. The seven-year average of water gauge data shows about 2-3 cm agreement (STD) with the xGEOID20 models over the four lakes, a similar level to the GSVS comparison results.

The geoid slope comparison is also performed for the three GSVS lines. The RMS differences range from 1 cm to 4 cm over baseline lengths from 0 to 300 km. The xGEOID20 deflections of the vertical are also validated using those observed on the GSVS 11/14/17 marks. The agreement is within several tenths of arcsecond.

Accompanied with the geoid grids, geoid errors are estimated using error variances of GOCO06s and gravity errors. The estimated errors were compared against the historical and GSVS GPS/leveling geoid differences. The scale of the predicted errors is consistent with the residuals, but further research is needed.

The first joint geoid model xGEOID20 passes the validation process and shows a relative accuracy around 2 cm in areas with good gravity coverage. However, there are areas that need to improve for the future models.

a. Reduction of the geoid differences originating from the computation methods. The large geoid differences are caused to a great extent by the correctional and computational procedure for the gravity anomalies over the Great Lakes, the use of shipboard data along coastal oceans, RTM and terrain corrections over mountainous areas, data interpolation and the Stokes kernel modification, and to a lesser extent by the methodological approximations (Stokes-Helmert vs. Molodensky), numerical integration methods, radius of the integration cap, and estimation of the far-zone contribution.

b. Polar geoid computation (latitude 60° to 90°, longitude 170° to 350°) needs special numerical treatment because of the meridian convergence. The ARCGP gravity file in 3'x3' provided by NGA should be combined consistently and optimally with CGS's

and NGS's terrestrial gravity data, and shipboard gravity data and altimetry-derived marine gravity data.

c. The Greenland geoid model needs topographical correction for the effect of density differences between ice and the topographical mass. A key task is compiling an ice thickness grid. CGS has developed the method for estimating the total ice correction to the geoid model, and will contribute an ice correction grid to the combined geoid model over Greenland.

d. Both CGS and NGS need to study the height anomaly and geoid computation over lakes. Water density needs to be considered for the geoid-quasigeoid separation computation.

In addition, the dynamic geoid model will be developed mainly accounting for the glacial isostatic adjustment around Hudson Bay, and glacial melt over Greenland, Queen Elizabeth Islands, and several major ice caps in the north part of the Western Cordillera, including Alaska. CGS has investigated the data requirement for the dynamic model, and NGS has carried out various studies including the geoid change due to glacial melt over Alaska. There is a need to develop a strategy and plan for the dynamic geoid model.

References

Ahlgren, K.M., G. Scott, D. Zilkoski, B. Shaw, and N. Paudel (2020) *GEOID18*. NOAA Technical Report NOS NGS, 72. Silver Spring, Maryland.

Andersen, O.B. and P. Knudsen (2016). *Deriving the DTU15 Global high resolution marine gravity field from satellite altimetry*. Abstract from ESA Living Planet Symposium 2016, Prague, Czech Republic. http://lps16.esa.int/page_session189.php#1558p .

Beckley, B., R. Ray, S. Holmes , N. Zelensky, F. Lemoine, X. Yang, S. Brown, S. Desai, G. Mitchum G, and J. Hausman (2013) *Integrated multi-mission ocean altimeter data for climate research complete time series version 2*. Available from ftp://podaac.jpl.nasa.gov/allData/merged_alt/preview/L2/docs/multi_alt_handbook_v2.pdf.

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

Hofmann-Wellenhof, B. and H. Moritz (2006) *Physical Geodesy*, 2nd edition. Springer, Wien. https://doi.org/10.1007/978-3-211-33545-1.

Krcmaric, J. et al. (2021) xG20DEM, in preparation.

Kvas, Andreas, Jan Martin Brockmann, Sandro Krauss, Till Schubert, Thomas Gruber, Ulrich Meyer, Torsten Mayer-Gürr, Wolf-Dieter Schuh, Adrian Jäggi, and Roland Pail (2021) "GOCO06s — a satellite-only global gravity field model," *Earth System Science Data*, 13(1), 99–118. https://doi.org/10.5194/essd-13-99-2021.

Li, Xiaopeng, Kevin Ahlgren, Ryan Hardy, Jordan Krcmaric, and Yan Ming 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.

Li, X., J.W. Crowley, S.A. Holmes and Y.M. Wang (2016) "The contribution of the GRAV-D airborne gravity to geoid determination in the Great Lakes region." *Geophys. Res. Lett*. 43, 4358–4365, https://doi.org/10.1002/2016GL068374.

Loomis, B.D., S.B. Luthcke, and T.J. Sabaka TJ (2019) "Regularization and error characterization of GRACE mascons." *J Geod* 93, 1381–1398. https://doi.org/10.1007/s00190-019-01252-y.

Mayer-Gürr, T., A. Kvas, B. Klinger, D. Rieser, N. Zehentner, R. Pail, et al. (2015). The new combined satellite only model GOCO05s. https://doi.org/10.13140/RG.2.1.4688.6807.

Mooney, Christopher Z., et al. "Bootstrapping: A nonparametric approach to statistical inference." No. 95. sage, 1993.

Moritz, H. (2000) "Geodetic Reference System 1980." *J Geod* 74, 128–133. https://doi.org/10.1007/s001900050278.

Moritz, H. (1980) *Advanced Physical Geodesy*, Herbert Wichmann Verlag, Karlsrihe.

National Geodetic Survey (2017) *NOAA Blueprint for 2022, Part 2: Geopotential Coordinates*. NOAA Technical Report, NOS NGS 64.

Pavlis, N.K., S.A. Holmes, S.C. Kenyon, and J.K. Factor (2012) "The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)," *J. Geophys. Res*., 117, B04406, https://doi.org/10.1029/2011JB008916.

Sánchez, L., J. Ågren, J. Huang, Y.M. Wang, J. Mäkinen, R. Pail, R. Barzaghi, G. Vergos, K. Ahlgren, and Q. Liu (2021). *J Geod* 95, 33. https://doi.org/10.1007/s00190-021-01481-0.

Smith, D.A. et al. (2019) *Blueprint for 2022, Part 3: Working in the Modernized NSRS*, NOAA Technical Report NOS NGS 67.

Smith, D.A., S.A. Holmes, X. Li, S. Guillaume, Y.M. Wang, B. Bürki, D.R. Roman, and T.M. Damiani (2013) "Confirming regional 1 cm differential geoid accuracy from airborne gravimetry: the Geoid Slope Validation Survey of 2011." *J Geod* 87:885–907. https://doi.org/10.1007/s00190-013-0653-0.

van Westrum, D., K. Ahlgren, C. Hirt et al. (2021) "A Geoid Slope Validation Survey (2017) in the rugged terrain of Colorado, USA." *J Geod* 95, 9. https://doi.org/10.1007/s00190-020-01463-8.

Véronneau, M. and J. Huang (2016) "The Canadian Geodetic Vertical Datum of 2013 (CGVD2013)." *Geomatica*. 70(1): 9-19.

Wang, Y.M. et al. (2021a) "Colorado geoid computation experiment—Overview and summary," *J Geod* 95.12 (2021): 1–21.

Wang, Y.M. et al. (2021b) "On the accurate computation of the geoid-quasigeoid separation in a mountainous region—a case study in Colorado with a full extension to the experimental geoid region," in preparation.

Wang, Y.M., X. Li, K. Ahlgren et al. (2020) "Colorado geoid modeling at the US National Geodetic Survey." *J Geod* 94, 106. https://doi.org/10.1007/s00190-020-01429-w.

Wang, Y.M., C. Becker, G. Mader et al. (2017) "The Geoid Slope Validation Survey 2014 and GRAV-D airborne gravity enhanced geoid comparison results in Iowa." *J Geod* 91, 1261–1276. https://doi.org/10.1007/s00190-017-1022-1.

Wang, Y.M., J. Saleh, X. Li, and D.A. Roman (2012) "The US Gravimetric Geoid of 2009 (USGG2009): model development and evaluation." *J Geod* 86, 165–180.

Wessel, B., M. Huber, C. Wohlfart, U. Marschalk, D. Kosmann, and A. Roth (2018): "Accuracy Assessment of the Global TanDEM-X Digital Elevation Model with GPS Data." ISPRS *Journal of Photogrammetry and Remote Sensing*. Vol. 139, pp. 171-182.

Yamazaki, D., D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J.C. Neal, C.C. Sampson, S. Kanae, and P.D. Bates (2017) "A high accuracy map of global terrain elevations," *Geophysical Research Letters*, vol.44, pp.5844–5853.

Zilkoski, D.B., J.H. Richards, and G.M. Young (1992) "Results of the general adjustment of the North American Vertical Datum of 1988." *Surv Land Inf Syst* 52(3):133–149.