A nation of volunteered data: Analysis of the NGS's GPS on Bench Marks project

Kevin AHLGREN, Jacob HECK, Galen SCOTT, Brian SHAW, United States

Key words: GPS, datums

SUMMARY

The National Oceanic and Atmospheric Administration's (NOAA) National Geodetic Survey (NGS) has a crowd-sourced data collection effort called GPS on Bench Marks. This project is a collaborative effort with thousands of volunteer surveyors across the United States at the local, regional, state and federal level. Volunteer surveyors provide their static survey data and metadata, along with a few photographs to NGS. In turn, NGS processes the data, provides survey coordinates in the National Spatial Reference System (NSRS), and stores the data for future NSRS modernization and development. The vast majority of this data is open access and can be easily retrieved online.

The GPS on Bench Marks project has been ongoing since 2014 with slightly different goals from year to year. Recently, the project has primarily focused on collecting data for hybrid geoid modeling and building a transformation between the current vertical datums and a future geoid-based vertical datum. User participation has exploded in the last 36 months as NGS gets closer to a modernized NSRS. This paper will highlight the overall project; workflows and web maps that NGS has established to make it easy for volunteer surveyors to find, collect, and submit their surveys; and describe how NGS uses this data for NSRS modernization.

A nation of volunteered data: Analysis of the NGS's GPS on Bench Marks project

Kevin AHLGREN, Jacob HECK, Galen SCOTT, Brian SHAW, United States

1. INTRODUCTION

GPS on Bench Marks is a program within NGS that is helping the geodetic and broader geospatial communities prepare for a modernized National Spatial Reference System (NSRS). It is a collaborative, crowd sourced data collection effort between NGS and the community with the mutual goal to produce more accurate geodetic products and ease the transition to the modernized system. The program provides a mechanism for NGS to highlight regions throughout the country where additional static GNSS observations would improve the quality of a particular NGS product. Local surveyors have the opportunity to invest their time and resources to observe these priority stations and ultimately benefit from the locally improved geodetic product/model.

Currently, there are two primary inter-related priorities for GPS on Bench Marks along with a number of secondary benefits. The first priority is to collect a robust dataset for modeling of a transformation surface between the current vertical datums (NAVD 88 and others) and the modernized North American-Pacific Geopotential Datum of 2022 (NAPGD2022). The second primary priority is to obtain new observations that will be utilized to generate reference epoch coordinates (RECs) (National Geodetic Survey, 2021) in the modernized NSRS, which are a very popular product in the current datums. Some of the secondary benefits derived from the GPS on Bench Marks data include updated passive control status, use of a mark for RTN field checks, and geoid model validation.

Over the past decades, the utility to define and maintain the National Spatial Reference System (NSRS) through continuous GNSS stations has changed the role of passive control. As a result of this shift, and an associated reduction in congressional funding for mark maintenance, responsibility for maintenance of passive control has reverted to the States. Passive control remains important locally for surveyors and there are a number of ways data can be submitted to NGS for ties to the NSRS and to be used to improve the local accuracy of NGS models and tools. The most rigorous process continues to be through "Bluebooking", which allows for a network adjustment to be performed with careful vetting of the observations. Once approved, these observations are then stored in the NGS Integrated Database (IDB) and authoritative coordinates are provided by NGS on Datasheets. Coordinates on datasheets provide the authoritative source of geodetic control for land surveying and engineering projects. The second method is through OPUS Share, which provides an easy way for local surveyors to send in data collected on passive marks to NGS. Users can submit their static GPS surveys to NGS, designate that they want them to be shared, and then NGS will process the data and share the results with the geodetic community. There are minimal requirements to 'share': 4+ hour observations on a re-observable monument, a dual frequency receiver, an accurate antenna height, and 2 photographs.

Both of these two processes will be discussed in a bit more detail in this paper in Section 3 along with some examples of what is possible with this data from a research perspective. Additionally, we present the current status of GPS on Bench Marks in Section 2. This paper has three objectives aimed at different user groups: 1) for the local surveying community, consider submitting data to NGS for inclusion in geodetic products; 2) for the research community, NGS has a wealth of accurate, completely public geodetic data that can be utilized for numerous investigations; 3) for other international geodetic agencies, we highlight best practices for crowd-sourced geodetic data and our experiment interacting with geodetic partners throughout the community.

2. GPS ON BENCH MARKS

GPS on Bench Marks is a program within NGS to provide outreach, coordination, and support the efforts of the geodetic community with the mutual goal of producing more accurate geodetic products. The process is rather simple as illustrated in Figure 1 with cooperation between NGS and geodetic surveyors that ultimately results in improved geodetic products and models, which benefits NGS, geodetic surveyors, and the general public.

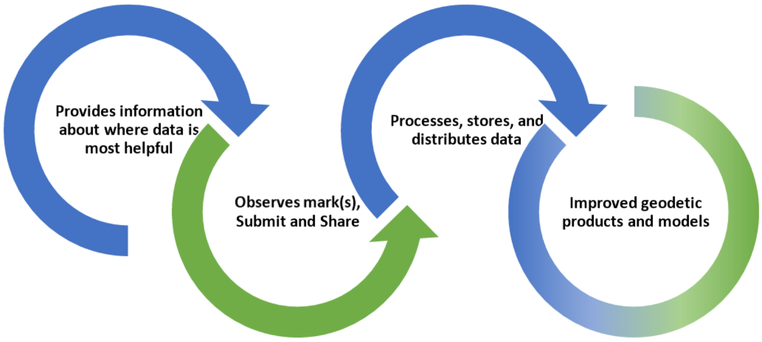

Figure 1: GPS on Bench mark general process and relationship between NGS (in blue) and the geodetic community (in green).

GPS on Bench Marks has evolved over the years with changing needs at NGS. Most recently, the goal was to support data collection efforts for GEOID18 (Ahlgren et al., 2020), the latest hybrid geoid to convert GNSS ellipsoid heights to orthometric heights in the appropriate NSRS datum for different regions (e.g. NAD83(2011) to NAVD 88). Currently, the primary focus of GPS on Bench Marks is to support data collection for a future transformation surface between NAVD 88 and NAPGD2022 with a secondary focus on updating passive control coordinates in a new regional frame at a reference epoch. For more information about this, please see National Geodetic Survey (2021).

2.1 CURRENT PRACTICES WITH GPS ON BENCHMARKS

This section will provide a brief overview of how and why NGS is requesting data, currently. As our primary focus is to support a future transformation between old and new datums, we have a spatial resolution that would ideally be met, which was determined to be 10 km. Based on this resolution, we divide the entire U.S. into a 10 km hexagon tiling resulting in 52,000+ hexagons (see Figure 2). This number would be larger but many areas of the country have no passive geodetic control in them at this fine resolution. The 10 km hexagons are our primary method of tracking the project status as will be shown in the next section. For regions that have the capability and desire to have a denser coverage, we also provide a 2 km hexagon tiling – which only is generated when the underlying 10 km hexagon is first completed. This has proved successful for cities and counties that have an active geodetic community, city/county surveyor, water district, department of transportation, etc. This increased spatial resolution will be most beneficial in *complex* areas where the relationship between the old and new datums can vary over shorter distances.

Based on previous campaigns that were extremely successful, we determined that providing information about individual and specific passive control marks eases the burden for partners and allows us to target the necessary data resolution in specific areas. This results in NGS designating a particular passive mark as a 'Priority' mark. This priority designation is arrived at based on an algorithm that looks at age of previous GPS data, dates the mark was last recovered, if the mark is observable, and a number of esoteric codes from the NGS database. However, local knowledge and use of a passive geodetic control mark is always preferred so partners can observe on other marks (not the designated 'Priority' mark) that still meet the NGS requirements and are within the same hexagon.

Lastly, there is quite a multilayered messaging about GPS on Bench Marks. This encompasses a publicly available web map with multiple layers to sort, query, and visualize both what is/is not completed. This is shown in Figure 2 and can be found online (GPS on Bench Marks for the Transformation Tool Web map). This map is regularly updated (currently at weekly intervals) as more data is submitted to NGS. The data for this web map is available on ESRI's Living Atlas of the World as a Feature Service and can be incorporated into other web maps. Some state agencies have adapted this data set for use in their own web maps to encourage regional collaboration and keep track of GPS on Bench Mark contributions at a local level.

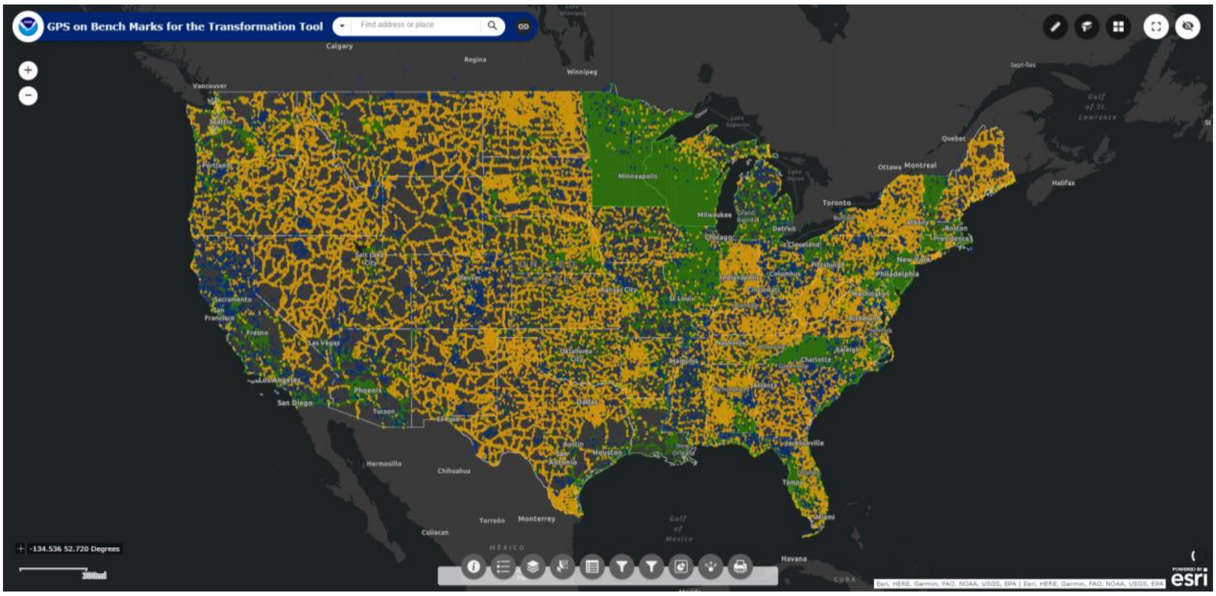

Figure 2: GPS on Bench Marks web map. Green regions are completed 10 km hexagons. Blue and Yellow regions are awaiting additional GPS observations.

Additionally, we have a regular newsletter that is delivered via email where we provide updates, highlight the work of a partner, etc. We also host a webinar at least once per year to describe the ongoing effort and any improvements. Finally, we have a number of tutorials, instructions, technical details, and other content on the GPS on Bench Marks webpage (https://geodesy.noaa.gov/GPSonBM/).

2.2 REPORTING FOR GPS ON BENCH MARKS

In addition to the operational side of the program described in the previous section, we also have a number of ways to visualize and monitor the status of the campaign that will be presented in this section. Some of these are real-time tools available to the public while others are metrics that are monitored internally at NGS.

The primary reporting tool that is used by NGS and available to the public is the GPS on Bench Marks Progress Dashboard, shown in Figure 3. This tool is generated by the same data layers as used in the web map but allows for a more detailed look at the completion status with options to analyze completion by state, by year/month, and by spatial resolution.

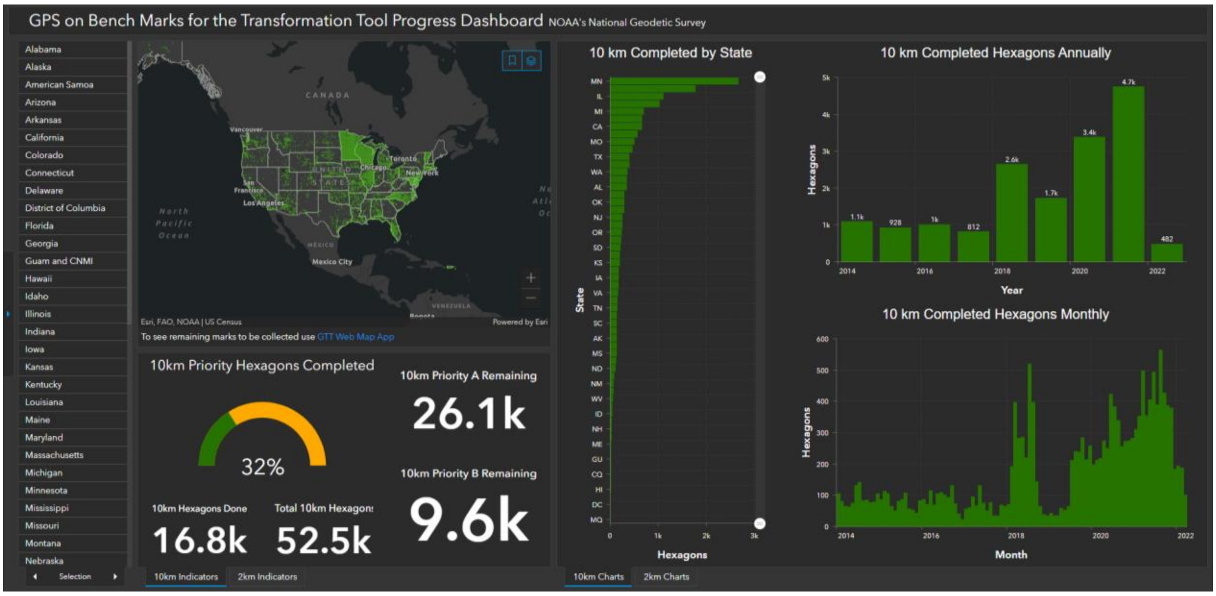

Figure 3: GPS on Bench Marks Progress Dashboard

A number of other metrics are monitored at NGS that are not provided in the Progress Dashboard. These include various time-based, cumulative statistics. For example, on an annual basis we monitor the growth of the campaign based on a number of statistics including total observations, total unique PIDs, total hours observed, and total unique contributors, which is illustrated in Figure 4. A couple of key takeaways from these statistics include:

1. Over 200,000 hours of observation time!
2. Nearly 29,800 total occupations
3. Steady growth in number of contributors – 100% increase from 204 contributors in 2019 to 424 in 2021.
4. Growth in the number of submissions per contributor from a median of 4 per contributor in previous years to 6 per contributor in 2021.

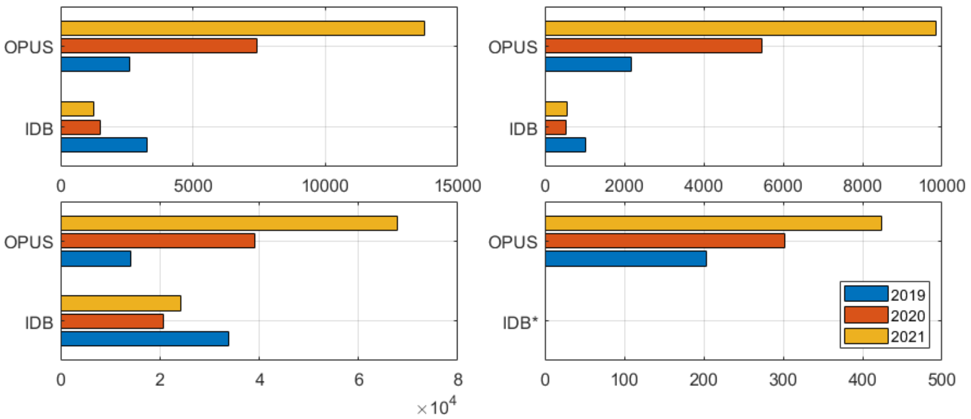

Figure 4: Annual statistics for GPS on Bench Marks from both data sources (OPUS and IDB). Upper-left: total occupations; Upper-right: total unique PIDs; Lower-left: total hours; lower-right: total unique contributors (IDB is not available presently)*

Additionally, the week-by-week 10 km hexagon completion statistics are monitored as illustrated in Figure 5. This gives a more detailed look at the campaign over time with the usual ebbs and flows based on *survey seasons* (e.g. lower totals in winter – higher totals in spring/summer) but can highlight the existence of a more underlying problem (e.g. IT related issues, supporting datasets being unavailable, etc.).

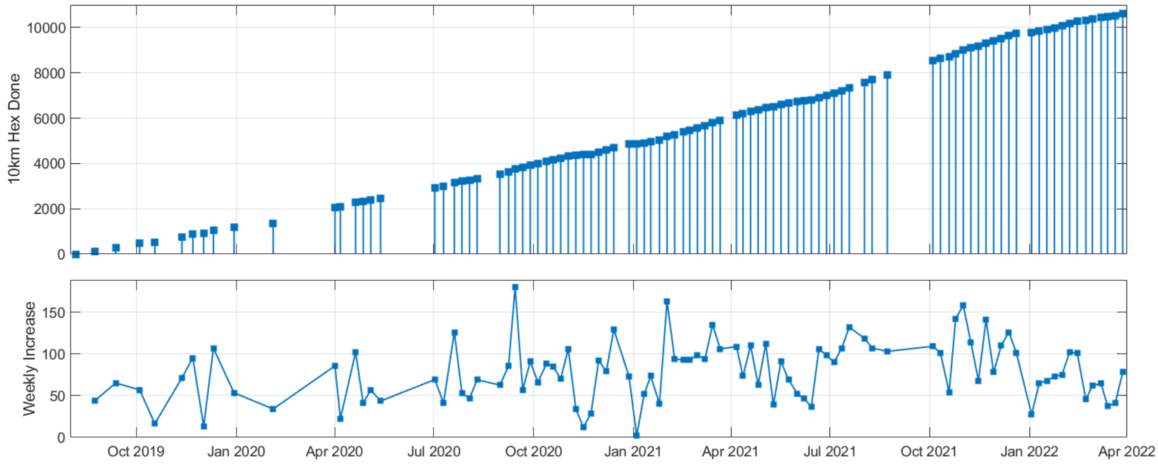

Figure 5: GPS on Bench Marks weekly 10 km hexagons completed. Upper: total 10 km hexagons completed since the start of the campaign in August 2019. Lower: weekly increase in 10 km hexagons completed.

3. OPUS SHARE DATABASE

The following section provides a look into the OPUS Share database, which is the most popular route for submission to GPS on Bench Marks. We focus on this source of data for a number of reasons including: it is the main data submission route for individual observations, it has enormous potential for external applicability for academic and applied research, and it is very easily accessible. The first part of this section consists of a broad overview of the data present in OPUS Share while the final sections present examples of what is possible with the OPUS Share data.

3.1 OVERVIEW

The OPUS Share Database consists of nearly 58,000 individual submissions as of 15 March 2022. The general geographic distribution of the database for CONUS and Alaska, Hawaii, and Puerto Rico is illustrated in Figure 6 and Figure 7, respectively, classified by the number of submissions an individual mark.

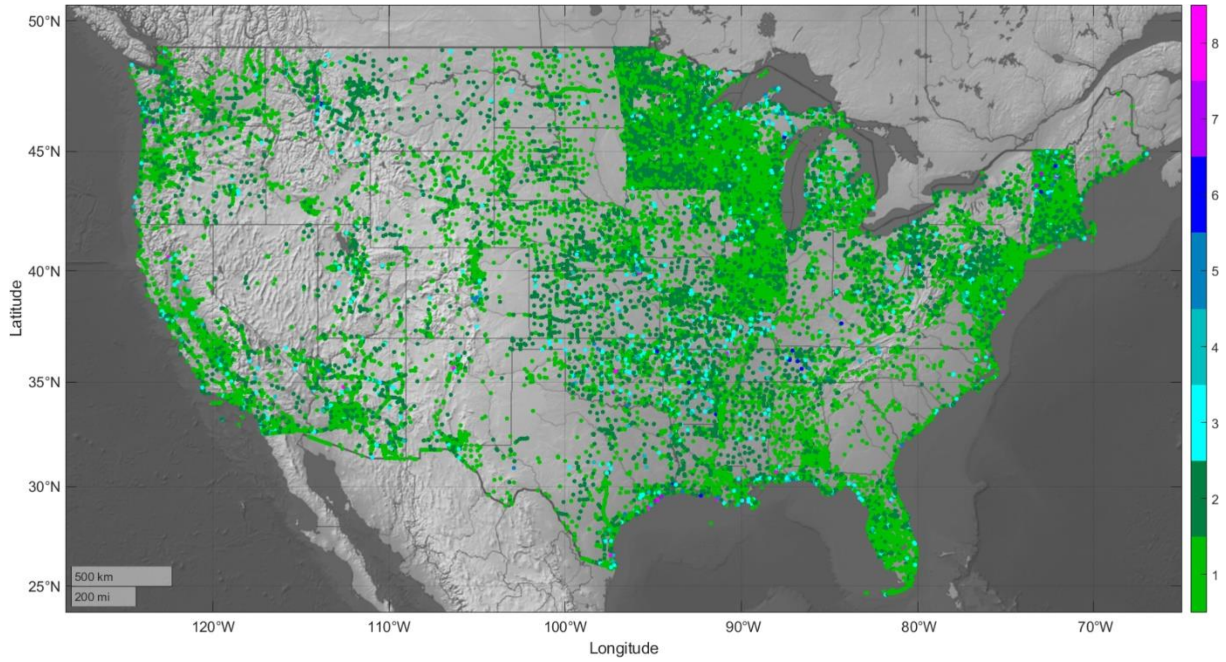

Figure 6: OPUS Share Database coverage throughout CONUS (as of 15 March 2022).

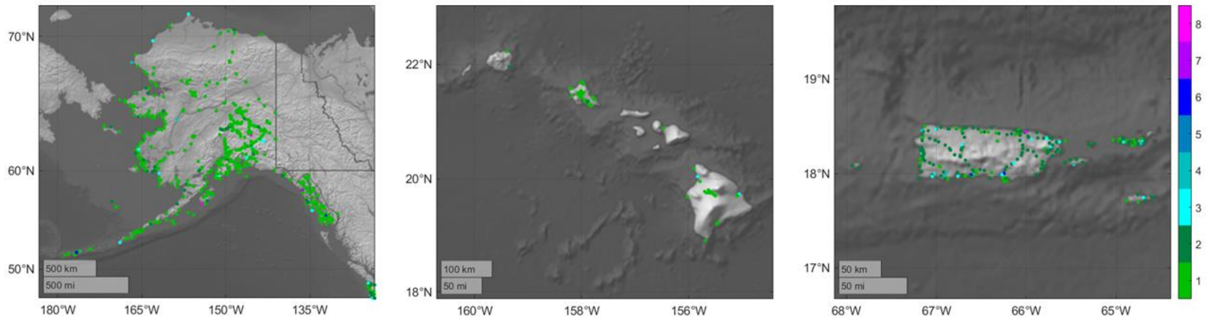

Figure 7: OPUS Share Database coverage for Alaska, Hawaii, and Puerto Rico/U.S. Virgin Islands (left to right, as of 15 March 2022).

It is evident from the figures that the majority of marks have been observed just one time, but this is a bit misleading as many of these marks have GPS observations present in the IDB (see Section 3.2). Additionally, there are 15,079 marks that have multiple observations as illustrated by in Figure 8.

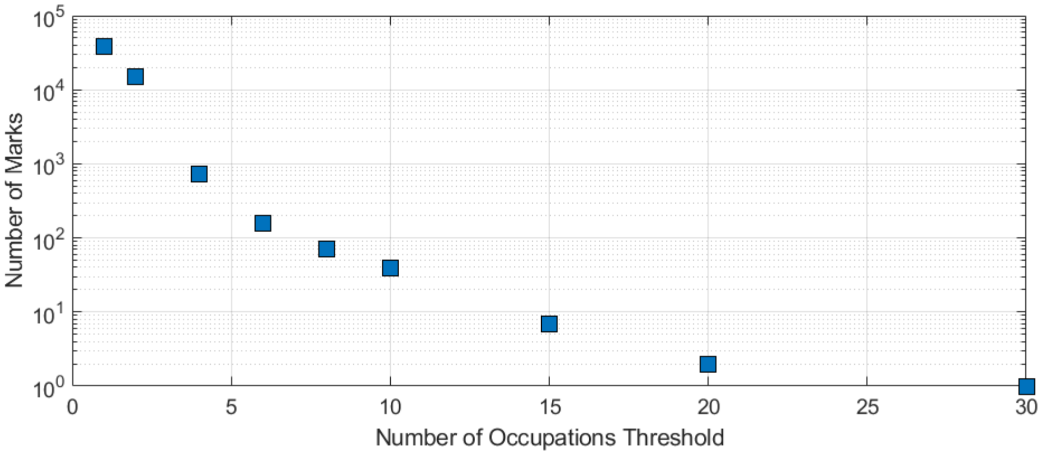

Figure 8: Number of marks that have a minimum number of occupations (log scale). 15,079 marks have 2+ occupations; 734 marks have 4+ occupations; 39 marks have 10+ occupations.

Another critical element for the overall accuracy of a GPS solution is the observational duration, which has a minimum of 4 hours for possible inclusion in OPUS Share. As one would expect, the vast majority of submissions are just above that 4-hour minimum as illustrated in Figure 9. There are a very small number of submissions (~900) that have observation durations less than 4 hours that are not shown in Figure 9. These submissions are exceptions to the rule and usually just a few minutes less than 4 hours (almost 600 out of 900 are 3:50+).

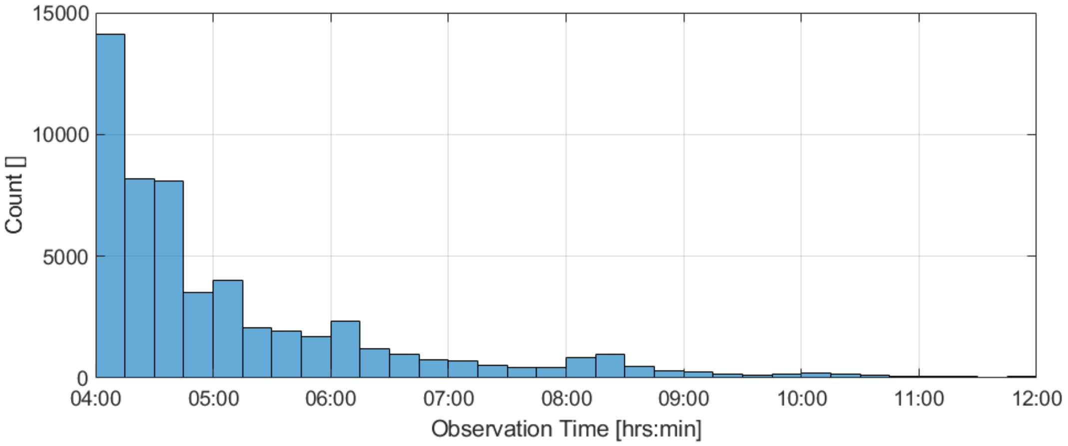

Figure 9: Number of marks with respect to observation duration.

3.2 NAD 83(2011) COMPARISON

One aspect that is investigated here is how the OPUS Share results compare with the currently published geodetic control values on any given passive mark. The published NAD 83(2011) epoch 2010.0 ellipsoid height can be found on a given mark's Datasheet and is the authoritative geodetic control from NGS. The difference between this value and the OPUS Share derived ellipsoid height is illustrated in Figure 10. In many instances, this difference can be interpreted as real vertical land motion captured between the original GPS survey(s) on a given mark and the more recent OPUS Share observation(s). For example, there is stronger evidence and more confidence that the change is real when we observe a geographic region where there is persistent and consistent observations showing that change. Alternatively, a single observation without any surrounding observations is much less convincing and could be caused by a number of other issues (e.g. the individual mark, antenna height, observational setup, etc.).

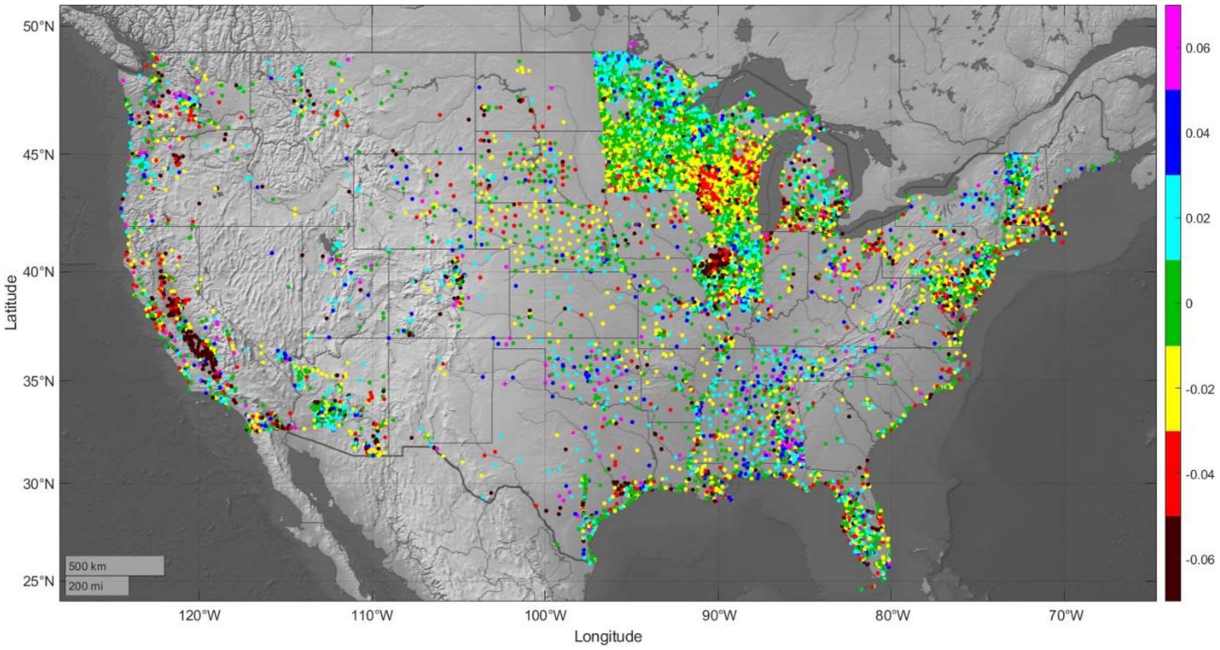

Figure 10: Difference between ellipsoid height in NAD 83(2011) epoch 2010.0 as published on NGS Datasheets and the OPUS Share determined value.

The original GPS survey(s) go back as far as the late 1980s/early 1990s and when compared with OPUS Share data observed over the last decade, can provide a longer time series than almost all CORS and provide a more densified network coverage of vertical land motion. In Figure 10, there are a number of regions with consistent ellipsoid height changes over a more widespread area that strongly support the theory that what is represented is actual land motion rather than some other artifact or error. If we look at the Central Valley of California, suburban Houston, Texas, and the western Great Lakes states, we find a consistent pattern of change that is being observed on numerous if not hundreds of marks. The detailed, localized research necessary to investigate these phenomena is beyond the scope of this paper but highlights the power of augmenting these NGS data sources with potentially other geodetic data to better understand what is occurring at local and regional levels.

3.3 INDIVIDUAL MARK TIME SERIES

The following section highlights the unique ability to utilize OPUS Share solutions to construct a vertical land motion, time-series analysis. We illustrate this on just a single mark and acknowledge that no major, general conclusions should be drawn from such a limited scenario. This is simply meant to spark conversations about what the possibilities are with this type of dataset. However, this example provides a benefit of using OPUS Share in that the submissions are from multiple observers representing a more complete geodetic history of a mark compared to the situation where only a single surveyor is observing a mark.

The mark highlighted is a tidal bench mark located on Grand Isle in Louisiana with NGS Permanent ID = AT0685. This region has been undergoing vertical land motion for many years and has been the subject of numerous geodetic and geophysical investigations (Shinkle and Dokka, 2004; Dokka et al. 2006). The mark has been observed and shared 14 times between 2007 and 2019 with the NAD 83(2011) ellipsoid height time series illustrated in Figure 11. This is the derived/estimated NAD 83(2011) ellipsoid height if one would simply grab the value from the OPUS Share solution's webpage (e.g. OPUS Share solution). Additionally, we include an estimated linear vertical velocity based on a simple, equal-weight least squares adjustment (see Table 1 for all estimated linear velocity solutions). However, the concern with this is that the OPUS Share solutions utilize a mixture of reference frames and constraint coordinates depending on when the solution was originally submitted to NGS.

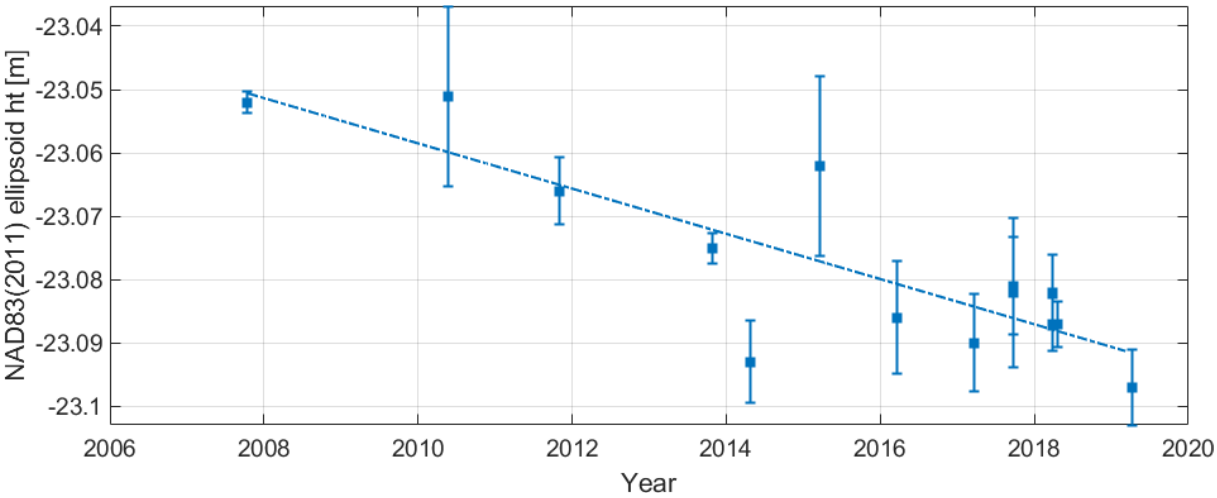

Figure 11: NAD 83(2011) ellipsoid height available on OPUS Share solution for tidal bench mark in Grand Isle, Louisiana. The estimated velocity is -3.6 mm/yr. Error bars shown are based on the ellipsoid height peak-to-peak value / 1.6929 (Schwarz, 2006) but are not used in the velocity estimation.

An enormous benefit of the OPUS Share solutions is that the geometric vector information (dX, dY, dZ) between the campaign station and the three CORSs used in the solution are included in an xml formatted file. This file also has the full variance-covariance information for the vector baselines. Using this information allows us to carry out a number of more rigorous estimates of the vertical velocity including a consistent reference frame, consistent and reprocessed CORS coordinates and velocities, and full variance-covariance information for the vectors. This analysis is carried using two different sources for the CORS coordinates: 1) the published ITRF2014 coordinates and velocities from NGS as part of MYCS2 (Saleh et al. 2021) and 2) an estimated ITRF2014 coordinate from the University of Nevada-Reno (Blewitt et al 2018). The reason we use both types of solutions is simply to show an independent comparison and highlight the variability in velocity and uncertainty estimates. The variance-covariance information is provided by the OPUS processing so is identical for both solution types with the only difference being the CORS coordinates at survey epoch. The time series with the estimated vertical velocity for both solution types is shown in Figure 12. The estimated standard deviation for each solution is also illustrated with an error bar, which is hardly visible due to the extremely low magnitude.

It is not too surprising to see that the estimated error is much too optimistic as this was seen most recently in the National Adjustment of 2011 (Dennis, 2021), which relies on the same underlying GPS processing software. One aspect that is surprising is the degree to which the UNR solution and the MYCS2 solution differ from one another This isn't reflected in Figure 12 but can be assessed when we compute the estimated variance of unit weight from (1) (or estimated standard deviation if one takes the square root) as illustrated in Figure 13. This value is expected to be approximately 1 (the *a priori*, variance of unit weight).

$$S_0^2 = \frac{\tilde{e}^T P \tilde{e}}{n - m} \tag{1}$$

where $\tilde{e}$ is the vector of residuals from the adjustment, *P* is the weight matrix, *n* is the number of observations, and *m* is the number of unknowns.

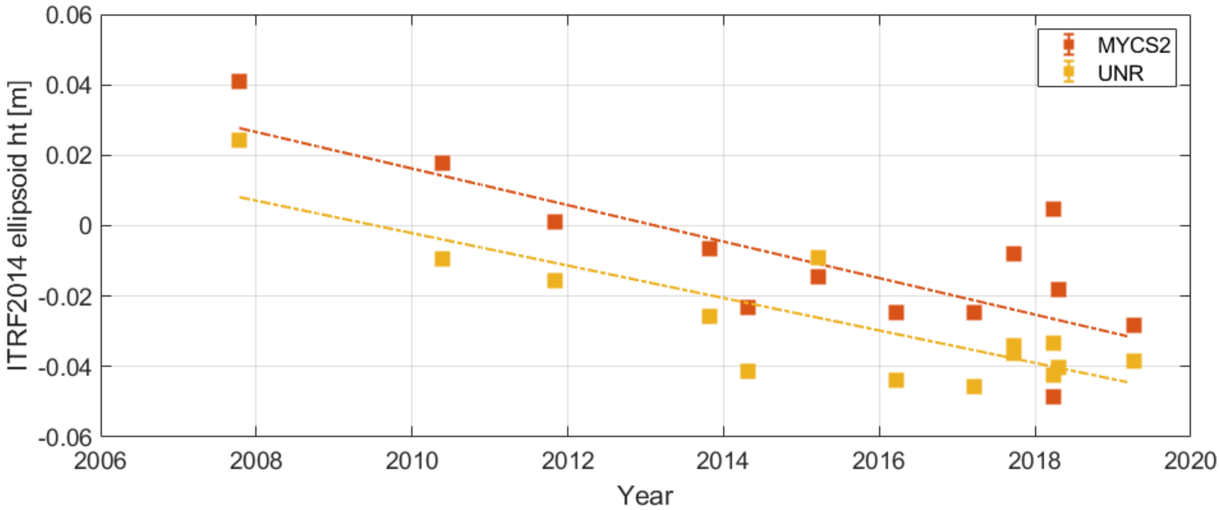

Figure 12: ITRF2014 ellipsoid height (with constant of -24.45 m removed) time series. Full unscaled, variance/covariance used to estimate the individual solution coordinates at survey epoch.

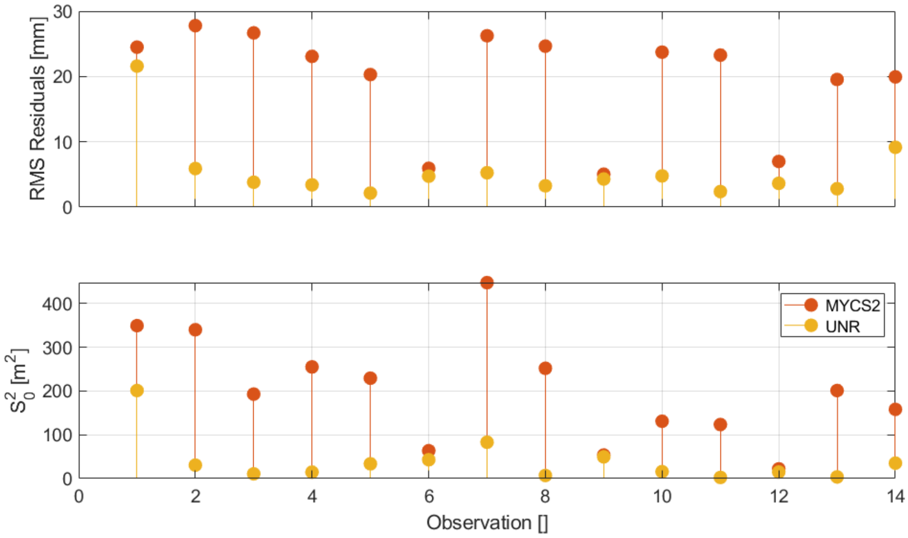

Figure 13: (top) RMS of the residuals [mm] and (bottom) estimated variance of unit weight [m^2] for each of the 14 solutions.

In this example, the MYCS2 solutions consistently have a much larger S_0^2 than the UNR solutions (100 to 300 compared with 10 to 40). The only difference between these solutions is the CORS coordinates used as constraints, which signifies that the UNR based CORS coordinates are much more self-consistent with the associated weights provided by OPUS. While the exact cause of this is beyond the scope of this paper and the subject of future work, we hypothesis that one of the MYCS2 CORS coordinates is inconsistent with the others. The ultimate goal though is to estimate an appropriate uncertainty for each solution (and propagate that to the vertical velocity and uncertainty estimate).

To achieve that goal, we perform a very crude scaling of the entire weight matrix (inverse of the variance/covariance matrix) based on approximately 1 / S_0^2 for each observation and solution type. Since the MYCS2 solutions have a much larger S_0^2, their corresponding weight matrices are scaled by a much smaller amount compared with the UNR solutions. After this type of scaled, variance-covariance adjustment, new coordinates are estimated along with uncertainties. Additionally, the new S_0^2 values can be computed, which are now nearly 1 for all solutions. Using these coordinates and uncertainties in the linear velocity estimation, this new time-series is illustrated in Figure 14 where the individual ellipsoid heights are not changed much from the original adjustment but the estimated uncertainties are much more appropriate. The impact of the more self-consistent UNR solutions is represented here quite clearly as this solution has much smaller error bars, which is reassuring that the crude scaling of the weights is being performed appropriately.

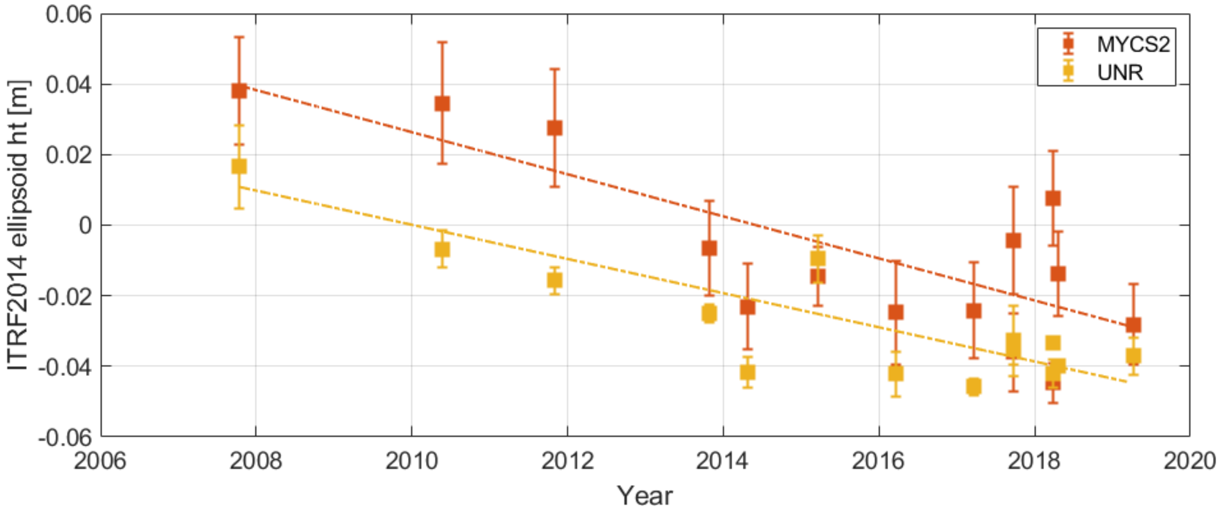

Figure 14: ITRF2014 ellipsoid height (with constant of -24.45 m removed) time series. Full, scaled variance/covariance used to estimate the individual solution coordinates at survey epoch.

Table 1: Estimated linear vertical velocities with uncertainty for all solutions

Estimated vertical velocity: [*mm/yr*]	Estimated uncertainty, 1-sigma: [*mm/yr*]	Weighting scheme:	Constraints Solution:	Reference Frame:
-3.6	+/- 0.7	Equally weighted	Original OPUS Share	NAD 83(2011)
-5.2	+/- 1.3	Original variance/covariance	MYCS2	ITRF2014
-4.6	+/- 0.8	Original variance/covariance	UNR	ITRF2014
-6.0	+/- 1.2	Scaled variance/covariance	MYCS2	ITRF2014
-4.8	+/- 0.7	Scaled variance/covariance	UNR	ITRF2014

4. CONCLUSION

The GPS on Bench Marks campaign has been an extremely successful endeavor the past few years and highlights the collaborative work between NGS and the local geodetic community. While the priority and efforts have changed slightly over the years, the overall goal remains the same – better national level geodetic products for local and regional scales. The campaign has seen tremendous growth the past few years when looking at total submissions, number of stations observations, overall observation times, and number of participants. Hopefully, this can serve as an example for other national geodetic agencies to follow and build upon with the overall goal of an improved spatial reference system.

However, we argue that there is still untapped potential with this type of fully transparent and public dataset. To just provide the most basic example of the applied research possible, we present two very small examples highlight the type of information contained within the dataset. While we acknowledge that no general conclusions should be drawn from this limited experiment, it is evident that a number of experiments can be performed that have meaningful impact for the geodetic community.

Finally, we encourage users throughout the United States to continue their support of this effort and consider participating, if they haven't done so as of yet. There are a number of locally realized benefits and efficiencies that can be obtained for users of the Modernized NSRS.

REFERENCES

Ahlgren, K., G. Scott, D. Zilkoski, B. Shaw, & N. Paudel (2020). GEOID 18. NOAA Technical Report NOS NGS 72. https://geodesy.noaa.gov/library/pdfs/NOAA_TR_NOS_NGS_0072.pdf

Blewitt, G., W. C. Hammond, and C. Kreemer (2018), Harnessing the GPS data explosion for interdisciplinary science, Eos, 99, https://doi.org/10.1029/2018EO104623.

Dennis, M.L. (2021). The National Adjustment of 2011. NOAA Technical Report NOS NGS 65.

Dokka, R. K., Sella, G. F., & Dixon, T. H. (2006). Tectonic control of subsidence and southward displacement of southeast Louisiana with respect to stable North America. Geophysical Research Letters, 33(23).

National Geodetic Survey (2021). Blueprint for the Modernized NSRS, Part 3: Working in the Modernized NSRS. NOAA Technical Report NOS NGS 67 https://geodesy.noaa.gov/library/pdfs/NOAA_TR_NOS_NGS_0067.pdf

Saleh, J., Yoon, S., Choi, K., Sun, L., Snay, R., McFarland, P., ... & Coloma, F. (2021). 1996–2017 GPS position time series, velocities and quality measures for the CORS network. Journal of Applied Geodesy, 15(2), 105-115.

Schwarz, C. R. (2006). Statistics of range of a set of normally distributed numbers. Journal of surveying engineering, 132(4), 155-159.

Shinkle, K., and Dokka, R. (2004). Rate of Vertical Displacement at Benchmarks in the lower Mississippi Valley and the Northern Gulf Coast. NOAA Technical Report NOS/NGS 50.

BIOGRAPHICAL NOTES

Kevin Ahlgren is a geodesist and geoid modeler within NGS's Observations and Analysis Division. He has a Ph.D. in Geodetic Science and Surveying from the Ohio State University.

CONTACTS

Dr. Kevin Ahlgren
NOAA / National Geodetic Survey
1315 E. West Highway
Silver Spring, MD
UNITED STATES
Tel. +01 240-533-9894
Email: kevin.ahlgren@noaa.gov
Web site: https://geodesy.noaa.gov/GPSonBM