Decadal GPS Time Series and Velocity Fields Spanning the North American Continent and Beyond: New Data Products, Cyberinfrastructure and Case Studies from the EarthScope Plate Boundary Observatory (PBO) and Other Regional Networks

G43B-0946
Fall AGU 2017

David A. Phillips[1] (phillips@unavco.org), T.A. Herring[2], T.I. Melbourne[3], M.H. Murray[4], W.M. Szeliga[3], M. Floyd[2], C.M. Puskas[1], R.W. King[2], F.M. Boler[1], C.M. Meertens[1], G.S. Mattioli[1]
[1]UNAVCO (www.unavco.org), [2]Massachusetts Institute of Technology, [3]Central Washington University, [4]New Mexico Tech

Abstract

The Geodesy Advancing Geosciences and EarthScope (GAGE) Facility, operated by UNAVCO, provides a diverse suite of geodetic data, derived products and cyberinfrastructure services to support community Earth science research and education. GPS data and products including decadal station position time series and velocities are provided for 2000+ continuous GPS stations from the Plate Boundary Observatory (PBO) and other networks distributed throughout the high Arctic, North America, and Caribbean regions. The position time series contain a multitude of signals in addition to the secular motions, including coseismic and postseismic displacements, interseismic strain accumulation, and transient signals associated with hydrologic and other processes. We present our latest velocity field solutions, new time series offset estimate products, and new time series examples associated with various phenomena.

Position time series, and the signals they contain, are inherently dependent upon analysis parameters such as network scaling and reference frame realization. The estimation of scale changes for example, a common practice, has large impacts on vertical motion estimates. GAGE/PBO velocities and time series are currently provided in IGS (IGb08) and North America (NAM08, IGb08 rotated to a fixed North America Plate) reference frames. We are reprocessing all data (1996 to present) as part of the transition from IGb08 to IGS14 that began in 2017. New NAM14 and IGS14 data products are discussed.

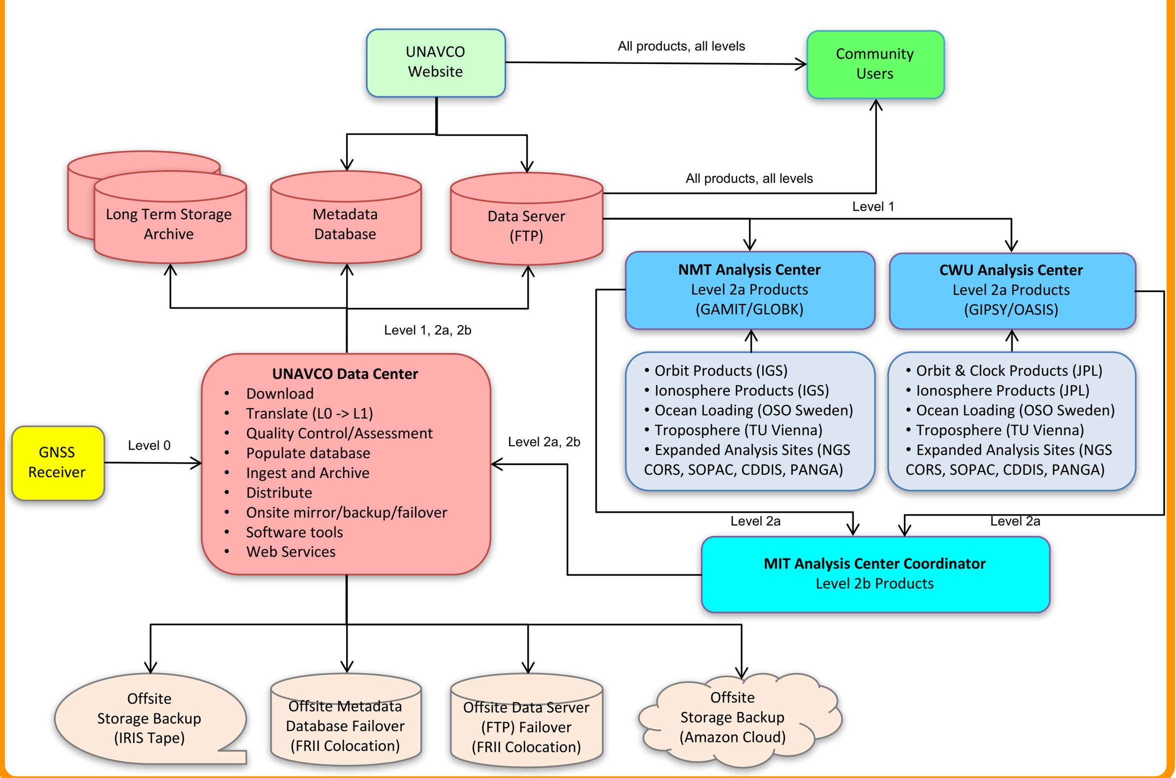
GAGE/PBO GPS data products are currently generated using onsite computing clusters. As part of an NSF funded EarthCube Building Blocks project called "Deploying MultiFacility Cyberinfrastructure in Commercial and Private Cloud-based Systems (GeoSciCloud)", we are investigating performance, cost, and efficiency differences between local computing resources and cloud based resources. Test environments include a commercial cloud provider (Amazon/AWS), NSF cloud-like infrastructures within XSEDE (TACC, the Texas Advanced Computing Center), and in-house cyberinfrastructures. Preliminary findings from this effort are presented. Web services developed by UNAVCO to facilitate the discovery, customization and dissemination of GPS data and products are also presented.

GAGE Facility GPS Data Flow, Analysis and Products

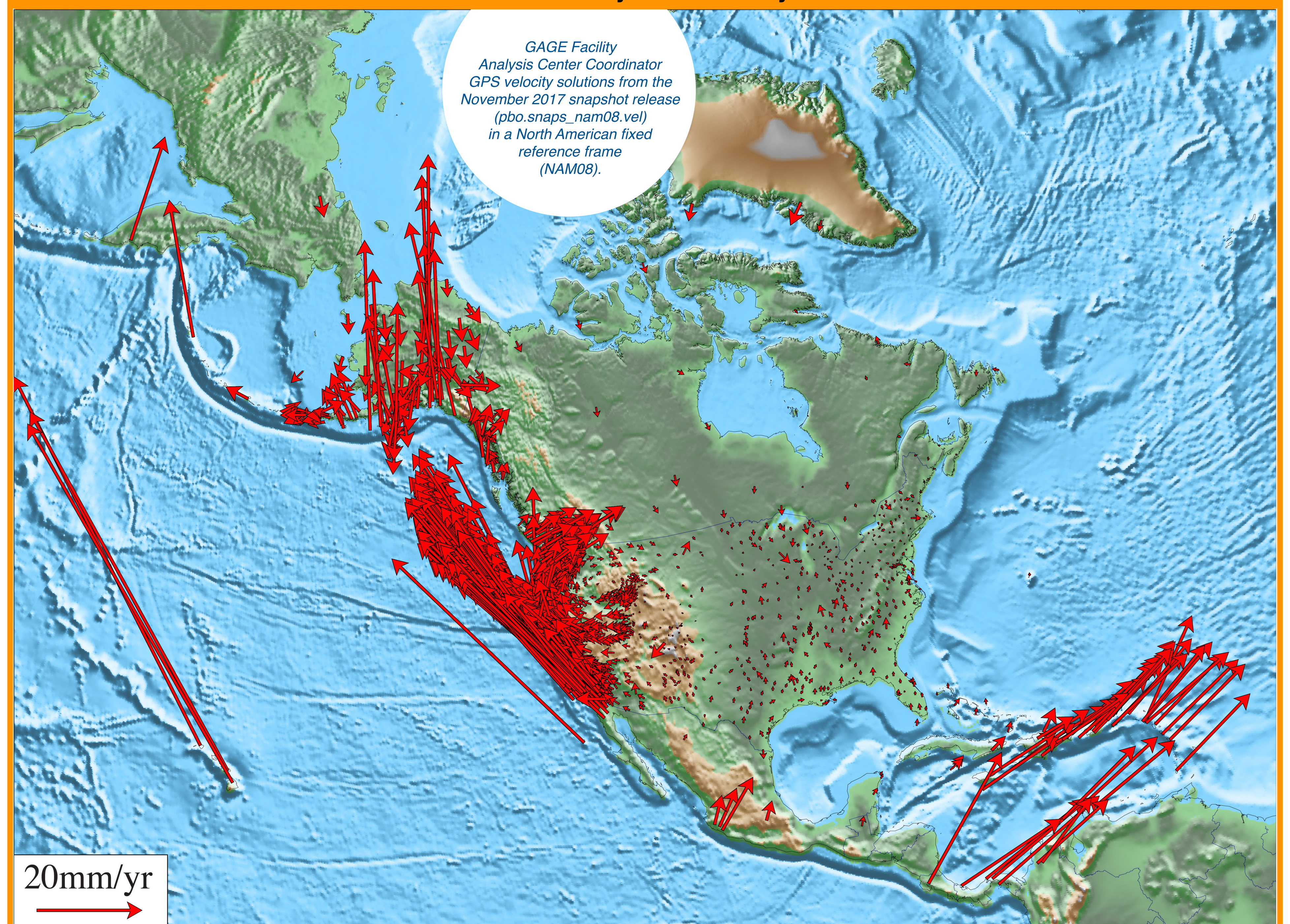
GAGE Facility Level 2 GPS data products are generated by two independent analysis centers (AC's), at Central Washington University (CWU) and New Mexico Tech (NMT), and an Analysis Center Coordinator (ACC) at the Massachusetts Institute of Technology (MIT). The CWU and NMT AC's process GPS phase and pseudorange data to generate unconstrained position solutions of the GPS sites averaged over 24 hour periods. The ACC then constrains and combines the AC solutions to generate final data products including position solution time series, position offset estimates, and velocity solutions.

All products are available from the UNAVCO FTP site: ftp://data-out.unavco.org/pub/products.

Data Type	Data Level	Data Product	Format	Generation Frequency	Generated By
GPS	0	Standard rate data (15-sec)	Raw, BINEX	Hourly, Daily	UNAVCO
		High rate data (1-, 2-, 5-sps)	Raw, BINEX	Hourly, Daily	UNAVCO
		Real-time, high rate data stream	BINEX, RTCM	Real-time	UNAVCO
		Community continuous data	Raw, RINEX	Hourly, Daily	Community
		Survey-mode (campaign) data	Raw, RINEX	Varies	Community
		Metadata	Database	Varies	UNAVCO
	1	Standard rate data (15-sec)	RINEX	Daily, varies	UNAVCO
		High rate data (1-, 2-, 5-sps)	RINEX	Varies	UNAVCO
		Community continuous data	RINEX	Daily, varies	UNAVCO
		Survey-mode (campaign) data	RINEX	Daily, varies	UNAVCO
	2	Position solutions (unconstrained)	SINEX	Daily	CWU, NMT
		Position solutions (constrained, combined)	SINEX	Daily	MIT
		Position offsets (e.g. coseismic)	ASCII	Varies	MIT
		Time series (constrained, combined)	ASCII, CSV	Daily	MIT
		Velocity solutions (constrained)	ASCII	Monthly	MIT
		Tropospheric parameter estimates	ASCII	Daily	CWU, NMT
		Position solution QA parameters	ASCII	Daily, varies	UNR
		Hydrologic loading models	ASCII	Quarterly	UNAVCO



GAGE Facility GPS Velocity Solutions



GAGE GPS network velocity solutions are generated by two different methods on two different time scales: "final" velocity solutions are generated annually and "snapshot" velocities are generated monthly. The annual final velocity solutions are derived from a computationally intensive full SINEX combination, and use discontinuity files and process noise models to account for temporal correlations in the time series. Snapshot velocity solutions are generated from direct fitting to time series as opposed to the much more time consuming and rigorous combination of SINEX files. Final velocity solutions are provided in both ASCII and SINEX formats in NAM08 and IGS08 reference frames (see note regarding transition to IGS14 below), while snapshot velocity solutions are provided in ASCII format. The most recent final velocity solutions here released in January 2017. The next final solution is currently in preparation and is scheduled for release in January 2018. Digital Object Identifiers (DOI's) are now minted for each annual release of final products (see References).

All velocity field solutions incorporate estimates of offsets from antenna changes, equipment problems (such as broken radomes), and earthquakes. For earthquakes with large post-seismic deformations, time dependent models are estimated, in the form of log functions with earthquake dependent time constants, so that the velocity estimates will not be greatly affected by the earthquake post-seismic motions. Since the post-seismic signals take time to develop, post-seismic log terms may be found to be needed many months after an earthquake and so these parameterizations evolve with time.

ITRF 2014 Transition

GAGE GPS time series and velocity solutions are currently provided in the ITRF 2008 no-net-rotation reference frame (IGS08) as well as in a North American fixed reference frame (NAM08). NAM08 is based on rotating IGS08 into the North America frame using the Euler pole from Altamimi et al. [2012]. Following the official release of ITRF 2014 earlier this year, we will begin to generate products in equivalent NAM14/IGS14 frames in the near future.

On January 29, 2017 (GPS week 1934 day 0), the IGS switched its operational products to use the IGS14 system, replacing the current IGb08 system. There are two major changes associated with this switch: (1) A new set of coordinates, velocities and, for some stations, post-seismic deformation models are being adopted for the core set of IGS sites that define the IGS realization of the ITRF2014 system; and (2) the antenna phase center models are being changed for the GPS satellites and for some GPS ground antennas. Details of expected effects are discussed in IGSMAIL-7399. The change in the reference frame realization should be small but the change in the antenna model will change the coordinates of some stations by up to 19 mm in height and 5 mm in horizontal coordinates.

The immediate effect on GAGE solutions is small because the GAGE AC's will not adopt the new IGS14 ANTEX file until a later date when we have reprocessed the data prior to week 1934 with the new IGS14 ANTEX file. IGS combined orbits based on the IGS14 system have already been generated, and the GAGE network processing AC (NMT) is using these files now. The GAGE PPP AC (CWU) will start reprocessing when JPL orbits and clocks become available in the IGS14 system. The current plan is to reprocess all data back to 1996 and then release the full set of time series in the NAM14 and IGS14 reference frames when all reprocessing is complete. The operational processing will also switch at the time of the release so that there will be no artificial offsets in the time series. The reprocessing is expected to take several months and an announcement will be made when the time series products are to be updated. Interim updates are available on the UNAVCO website.

http://www.unavco.org/data/gps-gnss/derived-products/derived-products.html

GAGE Facility GPS Time Series and Offset Estimates

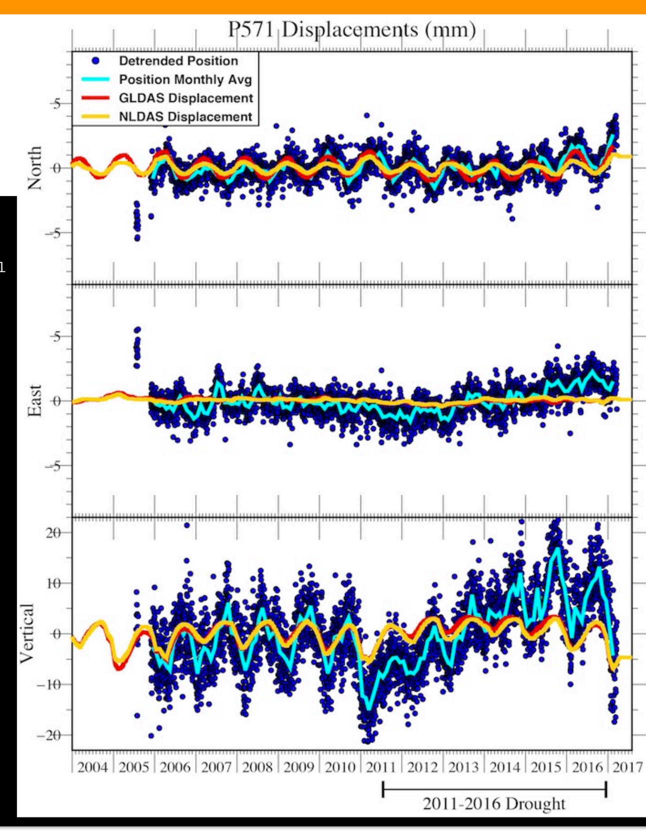
In generating velocity solutions, offsets due to earthquakes and equipment changes are estimated and low-quality outliers due to snow, for example, are removed from the velocity estimate solution so as to not corrupt it. However, these offsets are not removed from the position time series products, so that users may analyze all displacements and make their own decisions as to whether or not certain displacements should be "corrected". (In the spirit of "one researcher's noise is another researcher's data.") Nevertheless, we have developed and introduced several new data products to help users identify known and/or estimated time series displacements.

Some displacements can appear as sharply constrained breaks in time and space. These static offsets are often associated with geophysical events such as earthquakes, or maintenance events such as antenna swaps. The time series for TLALOCNet/UNAM-Geofisica site OXTH, shown here, includes offsets related to both an equipment change and an earthquake. In order to help users identify and adjust for such offsets, a new product file was developed and released in 2017 that lists all static offsets identified by the GAGE ACC using a Kalman filter, including N/E/U offset estimates and cause, if known.

Other offsets can be more subtle and variable. These transient displacements are often associated with hydrologic, atmospheric, or postseimic phenomena, and can be more challenging to identify, constrain and interpret. Moreover, such signals may be amplified, obscured, or even induced by different data processing methods, such as the estimation (or not) of scale changes as discussed in Herring et al., 2016. The time series for PBO site P571, shown here, includes time varying displacements due to hydrologic loading. In order to help users identify and adjust for such displacements, UNAVCO has developed new data products based on hydrologic surface loading. Using estimates of surface water mass from NASA Global and National Land Data Assimilation Systems (GLDAS and NLDAS, respectively, types of environmental models used in climate and weather models), UNAVCO models the expected displacements from these loads at GAGE-processed GPS stations.

p571_noah125_nldas2.hyd

```
Hydrological Surface Loading Time Series based on NLDAS NOAH Land Surface Model
Format Version:  1.0.0
Author:  UNAVCO
Land Surface Model:  NOAH
Land Surface Model grid size (degrees):  0.125
Hydro model of station (4-character id):  P571
Station lat:  36.231400
Station lon:  -118.767000
Begin Date:  2000-01-15
End Date:  2017-09-14
Release Date:  2017-10-19
Start Field Description
Date        Date for given epoch
YYYY-DOY    Year, day-of-year for given epoch
MJD         Modified Julian day for given epoch
DispN       Modeled north displacement from surface loading (mm)
DispE       Modeled east displacement from surface loading (mm)
DispU       Modeled vertical displacement from surface loading (mm)
End Field Description
Date, YYYY-DOY, MJD, DispN, DispE, DispU
. . .
2017-09-01,2017-244,57997.0,-0.48,0.00,1.68
2017-09-02,2017-245,57998.0,-0.48,0.00,1.69
2017-09-03,2017-246,57999.0,-0.48,-0.00,1.71
2017-09-04,2017-247,58000.0,-0.49,-0.00,1.73
2017-09-05,2017-248,58001.0,-0.49,-0.00,1.75
. . .
```



Above right: Detrended GPS position time series from PBO site P571 near Springville, California, on the western slope of the Sierra Nevada Mountains, compared to displacement time series from GLDAS and NLDAS. The 2011-2016 drought is marked. During the drought, P571 experienced uplift not represented in the models. This was attributed to depletion of stored groundwater and permanent snowpack, which was not included in the surface loading models based on the GLDAS and NLDAS environmental parameters. However, the surface loading displacement models do have reduced peak-to-peak seasonal amplitudes during the drought, reflecting the reduced soil moisture. Above left: header and example entries from the UNAVCO surface loading time series file (p571_noah125_nldas2.hyd) for this site.

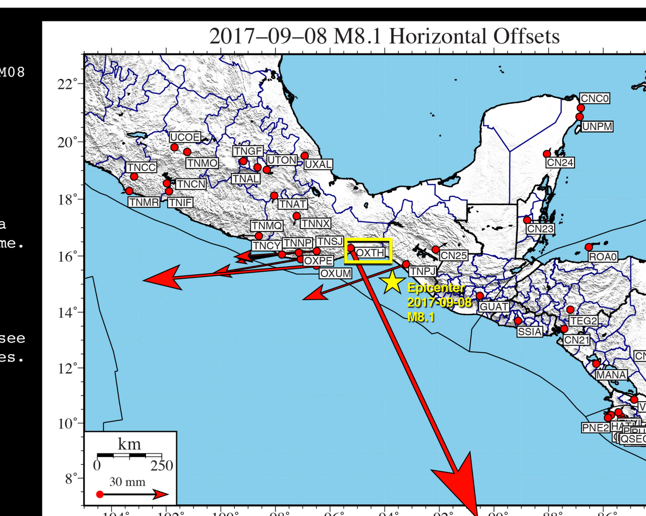
pbo.kalts_nam08.off

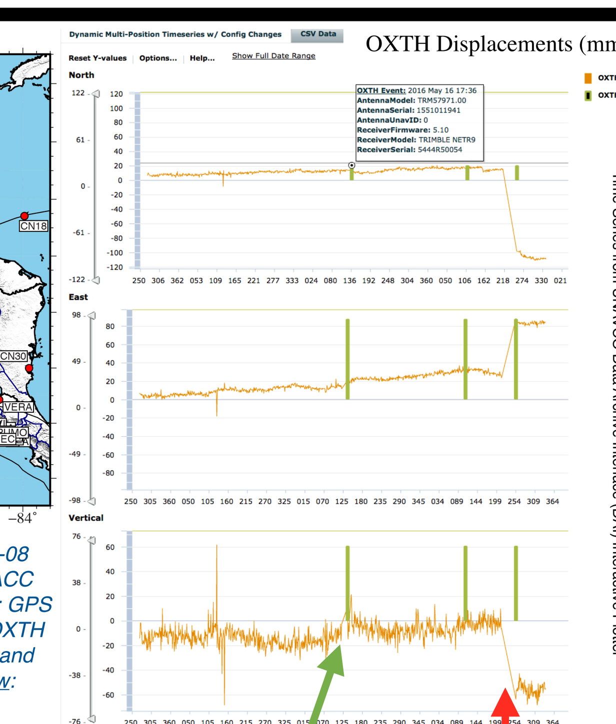
```
PBO Offset estimates from SNAPKF/kf_<site>.pbo.final.det Reference Frame : NAM08
Format Version: 1.0.0
Release Date  : 201710161735
Start Field Description and Notes
NOTES
Extracted using sh_extOffsets Date Mon Oct 16 18:01:32 EDT 2017
This file is created from Kalman Filter fits to the GAGE time-series. The
process noise values used in the Kalman filters are computed for each site as
described in Herring et al., 2016. Offsets can be of Break type due to antenna
change or due to unknown reasons or of type OffEq due an earthquake at the time.
If the standard deviation of the offset estimates (sN, sE, sU) are large, the
offset can not be accurately determined.  In some cases this is due to there
being no usable data prior to the offset or, in some cases, multiple offsets
with no usable data between the offsets. Earthquakes with large non-linear
standard deviations are not reported. This product is not quality checked to see
if the estimated offsets remove the apparent discontinuities in the time series.
For sites with very systematic residuals, the estimates might not remove the
discontinuities.
The columns are
Site             : GAGE 4-character site ID
YYYY MM DD  HR MN : Year, month, day, hour, minute of the time of the offset
dN (mm)  sN (mm)  : Change in North,  1-sigma standard deviation (mm)
dE (mm)  sE (mm)  : Change in East,   1-sigma standard deviation (mm)
dU (mm)  sU (mm)  : Change in Height, 1-sigma standard deviation (mm)
TYPE              : Either Break (equipment) or OffEq (earthquake)
Description       : Free format comments specific to offset type
                    Types are
                    EQ -- Earthquake with GAGE code (from All_PBO_eqs.eq)
                    AN -- Antenna or radome change (from All_PBO_ants.eq)
                    UN -- Unknown reason or from damage (from All_PBO_unkn.eq)
Reference:  Herring, T. A., T. I. Melbourne, M. H. Murray, M. A. Floyd, W. M.
Szeliga, R. W. King, D. A. Phillips, C. M. Puskas, M. Santillan, and L. Wang
(2016), Plate Boundary Observatory and related networks: GPS data analysis
methods and geodetic products, Rev. Geophys., 54, 759-808, doi:
10.1002/2016RG000529.
End Field Description and Notes

.Site  YYYY MM DD  HR MN     dN (mm)  sN (mm)    dE (mm)  sE (mm)    dU (mm)  sU (mm)  TYPE  ! Description
  . . .
 OXTH  2016 05 16  17 36      -2.87     0.52       6.97     0.81      10.90     1.52  Break ! AN Antenna swap TRM41249.00 to TRM57971.00 Receiver TRIMBLE NETRS to TRIMBLE NETR9
 OXTH  2017 09 08  04 50    -115.55     0.65      56.95     0.97     -46.95     2.38  OffEq ! EQ 40 Location   15.06780  -93.71500
  . . .
```

Offset from Equipment Change on 2016-05-16

Offset from Earthquake "EQ 40" on 2017-09-08





Above: horizontal coseismic displacements from the 2017-09-08 M8.1 Pijijiapan, Mexico earthquake estimated by the GAGE ACC (event file: pbo_170908_0450_eq40_coseis_kalts.evt). Right: GPS position time series from TLALOCNet/UNAM-Geofisica site OXTH showing offsets due to equipment changes in 2016 (antenna and receiver swap) as well as the Pijijiapan earthquake. Left/below: header and OXTH entries from the new offset file (pbo.kalts_nam08.off).

NSF EarthCube GeoSciCloud Project: Experimental GPS Processing in the Cloud

EarthCube Building Blocks: Collaborative Proposal: Deploying Multi-Facility Cyberinfrastructure in Commercial and Private Cloud-based Systems. (GeoSciCloud)

It is common to hear that it would be optimal to perform computations "in the cloud".

And it is indeed true that many commercial companies are moving their information technology into a cloud environment. However, NSF funded scientific data centers including UNAVCO and IRIS have unique requirements and constraints. Funding is limited and costs of managing data centers using cloud technology can be quite costly. Additionally, government funded research organizations typically have much smaller IT staffs than do corporations. The impact of managing IT operations in the cloud is not identical between large corporations and NSF funded data centers.

As part of the NSF funded GeoSciCloud project, IRIS, UNAVCO, and the GAGE Facility GPS Analysis Centers at CWU and NMT are deploying data collections and operational processing services in different computing environments in order to assess the feasibility and impact of cloud computing versus in-house infrastructure. Specifically, we are comparing operations in a commercial cloud environment (Amazon Web Services (AWS)), an NSF supported large computing facility with cloud computing characteristics (XSEDE), and current in-house environments. This project will thereby help NSF/EarthCube identify the most suitable IT environment in which EarthCube should deploy and support shared infrastructure. The potential reliability and cost-savings are excellent motivating factors.

One of the most computationally intensive parts of UNAVCO GAGE Facility operations is reprocessing of decades worth of 15 second-sampled GPS data from 2,000+ globally distributed stations. This involves reprocessing of literally millions of data files and takes several weeks to months to complete. While it is not practical to perform an entire reprocessing run in the cloud for this project, it was proposed that the GAGE GPS Analysis Centers at CWU and NMT would provide the basis for assessing the benefit of the elasticity and co-located cloud storage by reprocessing two months worth of data.

As of December 2017, the AWS testing phase is mostly complete and preliminary findings are summarized here. The next phase of the project will be to repeat the experiment in the XSEDE environment.

AWS Resource Costs: GAGE Facility GPS Data Analysis Centers

GPS Analysis Center (Processing Software)	Billed AWS Cost for Processing 1 Network Day	Billed AWS Cost for Storage for Processing 1 Network Day	Projected AWS Cost* for Processing of 3,652 Network Days (10 Network years)	Projected AWS Cost* for Storage for Processing of 3,652 Network Days (10 Network years)	Time to Reprocess All Existing Data 1996-present (Local cluster)	Time to Reprocess All Existing Data 1996-present (AWS)
Central Washington University (GIPSY)	$3.86	$1.95	$14,097	$7,121	~24 days	~1,000 days (~19 months)
	$5.81		$21,218			
New Mexico Tech (GAMIT)	$9.69	$3.50	$35,388	$12,782	~4 months	A few days to weeks (depending on automation)
	$13.19		$48,170			
Notes	Proposal was to process 60 days but only 1 day processable to date. CWU processed day 2016-275 w/ 1137 sites. NMT processed day 2016-304 w/ 1852 sites.		These projections may be very inaccurate due to the small sample size on which they are based. Also see caveats below.		AWS vs. local cluster reprocessing times are reversed for the two analysis centers/software.	

*Caveats that significantly affect these (re)processing cost projections:
- There may be free or discounted rates/tiers for certain time periods and/or job sizes.
- It is not clear based on work to date how data storage costs would scale.

Summary of overall user experience with Amazon Web Services (AWS)
Overall rating scale: 1 (It was great) - 10 (Would not choose to go through this again).

Task	Rating	Summary Comments	AWS Advantages	AWS Challenges
UNAVCO FTP Service	5	Very straightforward to set up and get running.	None so far, but when storage expansion is needed AWS should be simpler.	Minor; need to learn AWS terminology.
UNAVCO GeoWS (Geoscience Web Services)	2	Setup straightforward and less time after initial learning curve. AWS architecture used reduces complexity for development/support of web services.	Infinite scaling, theoretically zero downtime, less monitoring support required	Learning features of different products to improve response times and help reduce costs. Cost to optimize software to take advantage of lower cost offerings
CWU AC Processing (GIPSY)	10	Difficult and frustrating overall for CWU; local optimization methods don't port well to AWS; (re)processing time much slower than local system.	No clear advantage.	Understanding how to map CWU GPS processing workflow topology to AWS. Significantly slower. Lack of effective job scheduler on AWS.
NMT AC Processing (GAMIT)	5	NMT/GAMIT scripts seem to map relatively well to AWS and enable a mostly automated environment; (re)processing time potentially much faster than local system.	Main advantage is access to large numbers of CPUs and storage space that would enable processing of multiple days/weeks simultaneously.	Need a better understanding about how to fully automate launching of instances to facilitate large-scale processing operations.

References

ARTICLES:

Altamimi, Z., L. Métivier, and X. Collilieux (2012), ITRF2008 plate motion model, J. Geophys. Res., 117, B07402, doi:10.1029/2011JB008930.

Herring, T. A., T. I. Melbourne, M. H. Murray, M. A. Floyd, W. M. Szeliga, R. W. King, D. A. Phillips, C. M. Puskas, M. Santillan, and L. Wang (2016), Plate Boundary Observatory and related networks: GPS data analysis methods and geodetic products, Rev. Geophys., 54, doi:10.1002/2016RG000529.

DOI's:

Herring, T.A., T.I. Melbourne, M.H. Murray, M.A. Floyd, W.M. Szeliga, R.W. King, D.A. Phillips, C.M. Puskas, M. Santillan, 2016, GAGE Processing GPS Plate Boundary Observatory Expanded Analysis Product for 2016: Final Position Time Series; Combination in NAM08 and IGS08 Reference Frames Produced by the Massachusetts Institute of Technology (Analysis Center Coordinator), UNAVCO, Geodetic Data Product, doi:10.7283/P2PH25.

Herring, T.A., M.A. Floyd, and R.W. King, 2017, GAGE Processing GPS Plate Boundary Observatory Expanded Analysis Product for 2016: SINEX Analysis Based Final (Annual) Velocity Field; Combination in NAM08 and IGS08 Reference Frames Produced by the Massachusetts Institute of Technology (Analysis Center Coordinator), UNAVCO, Geodetic Data Product, doi:10.7283/P29665.

Herring, T.A., M.A. Floyd, and R.W. King, 2017, GAGE Processing GPS Plate Boundary Observatory Expanded Analysis Product for 2016: Combination in NAM08 and IGS08 Reference Frames Using Weighted Least Squares Analysis of Time Series from CWU (NAM08 only), NMT (NAM08 only) and PBO Analyses Produced by the Massachusetts Institute of Technology (Analysis Center Coordinator), UNAVCO, Geodetic Data Product, doi: 10.7283/P21T0M.

UNAVCO, a non-profit university-governed consortium, facilitates geoscience research and education using geodesy.
UNAVCO operates the U.S. National Earth Science Geodetic Facility, known as the Geodesy Advancing Geosciences and EarthScope (GAGE) Facility.
Core funding is provided by the U.S. National Science Foundation (NSF) and the U.S. National Aeronautics and Space Administration (NASA).