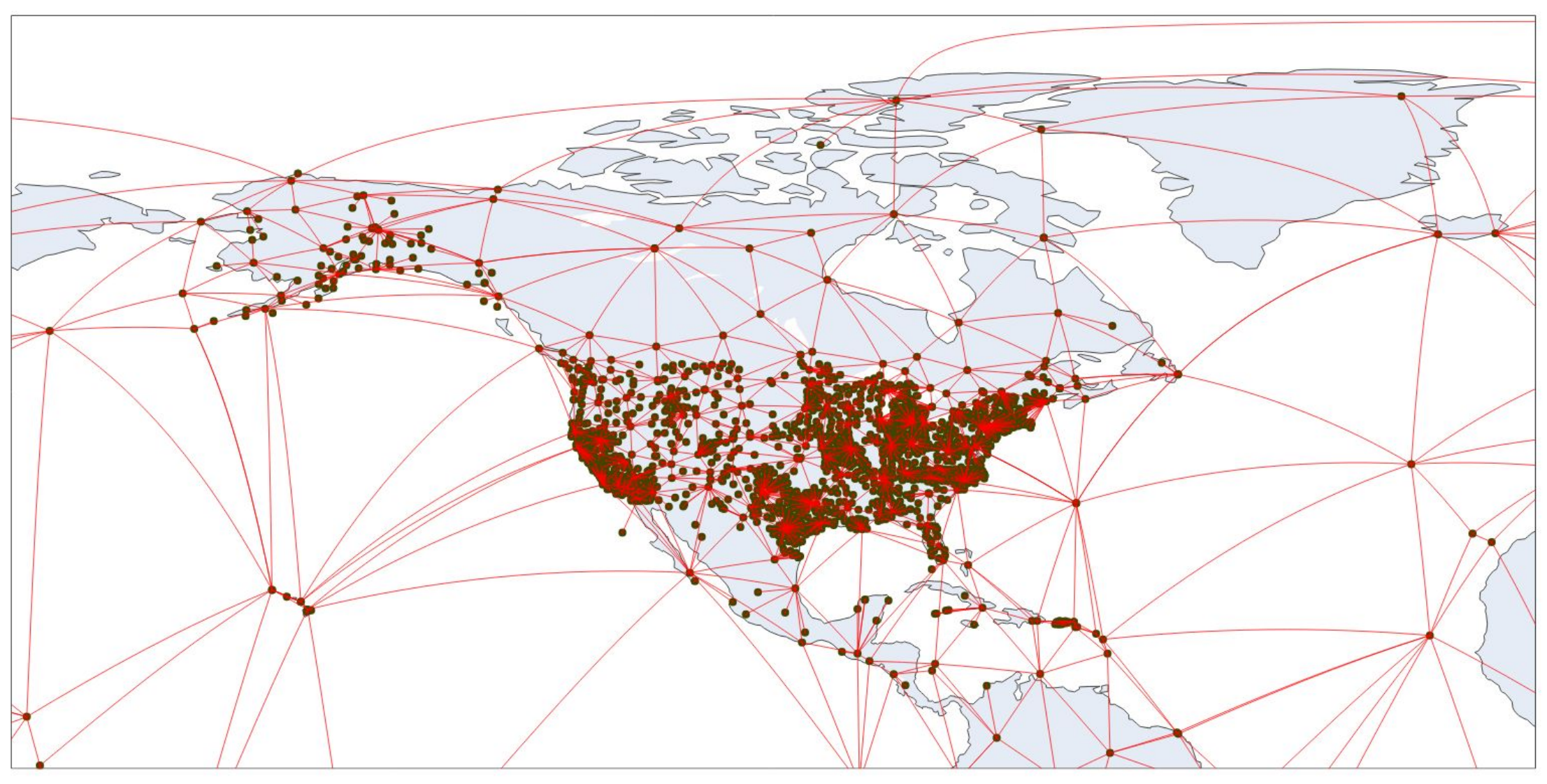
Automated analysis of GNSS time series for large networks

Rick Bennett (rick.bennett@noaa.gov), Phillip McFarland, and Amy Whetter
NOAA/National Geodetic Survey, Silver Spring, Maryland, USA

IGS Workshop 2024 Poster P2: 018

1. Summary

With the explosion of continuous GNSS stations around the globe over the past three decades, there is an increasing need for algorithms to objectively process coordinate time series data with minimal human input.



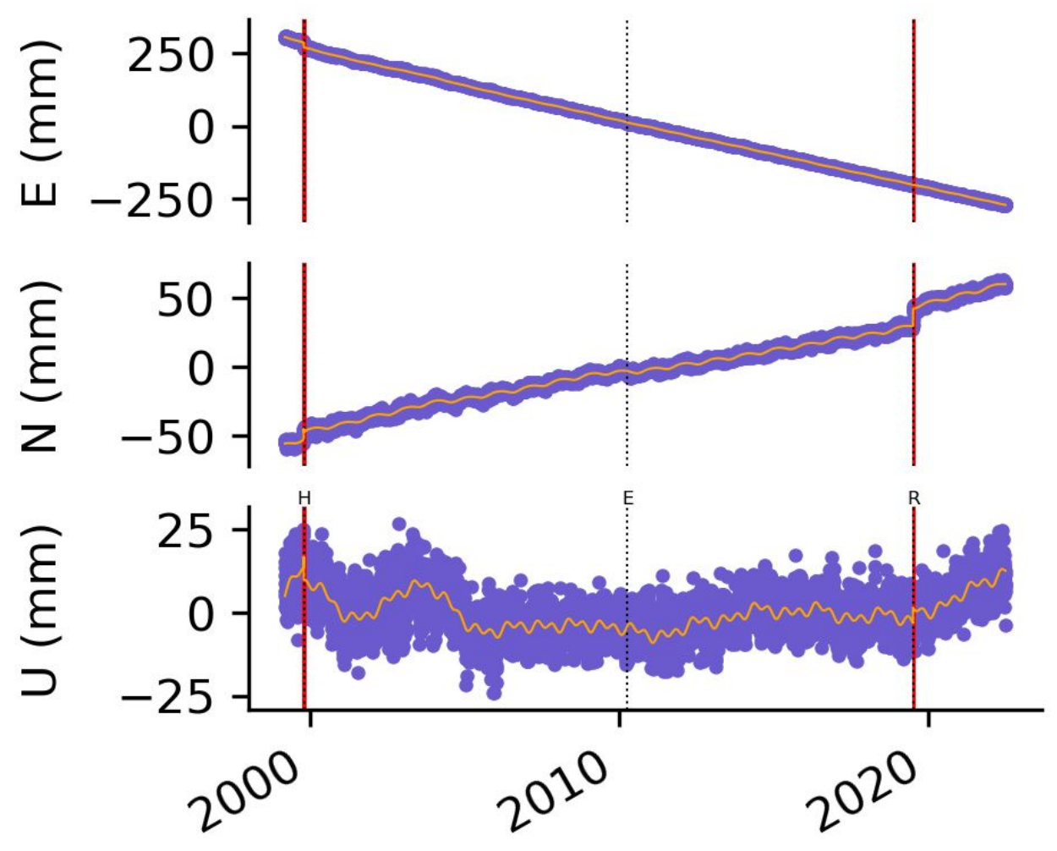
The NOAA CORS Network (NCN) comprises ~1900 operational GNSS stations, operated by > 230 organizations, public and private, and serves a wide range of applications. The NCN is heterogeneous in several respects, including instrumentation type, monumentation style and site characteristics, and tectonic setting. There is no one size fits all routine for automatically analyzing data from the NCN. Crustal deformation and equipment changes cause some site coordinates to evolve appreciably, and these motions must be accounted for to precisely realize the National Spatial Reference System (NSRS). Incomplete site motion models can cause velocity solutions to crash, or become distorted, degrading the accuracy of frame realizations. Anomalous displacements unaccounted for by the coordinate functions compromise access to the NSRS.

2. Analysis methods

To address the problem of determining coordinate functions for the NCN in an automated and objective way, we developed a simple algorithm that derives its utility primarily through an iterative application of the LASSO (Least Absolute Shrinkage and Selection Operator) regression (Tibshirani, 1997).

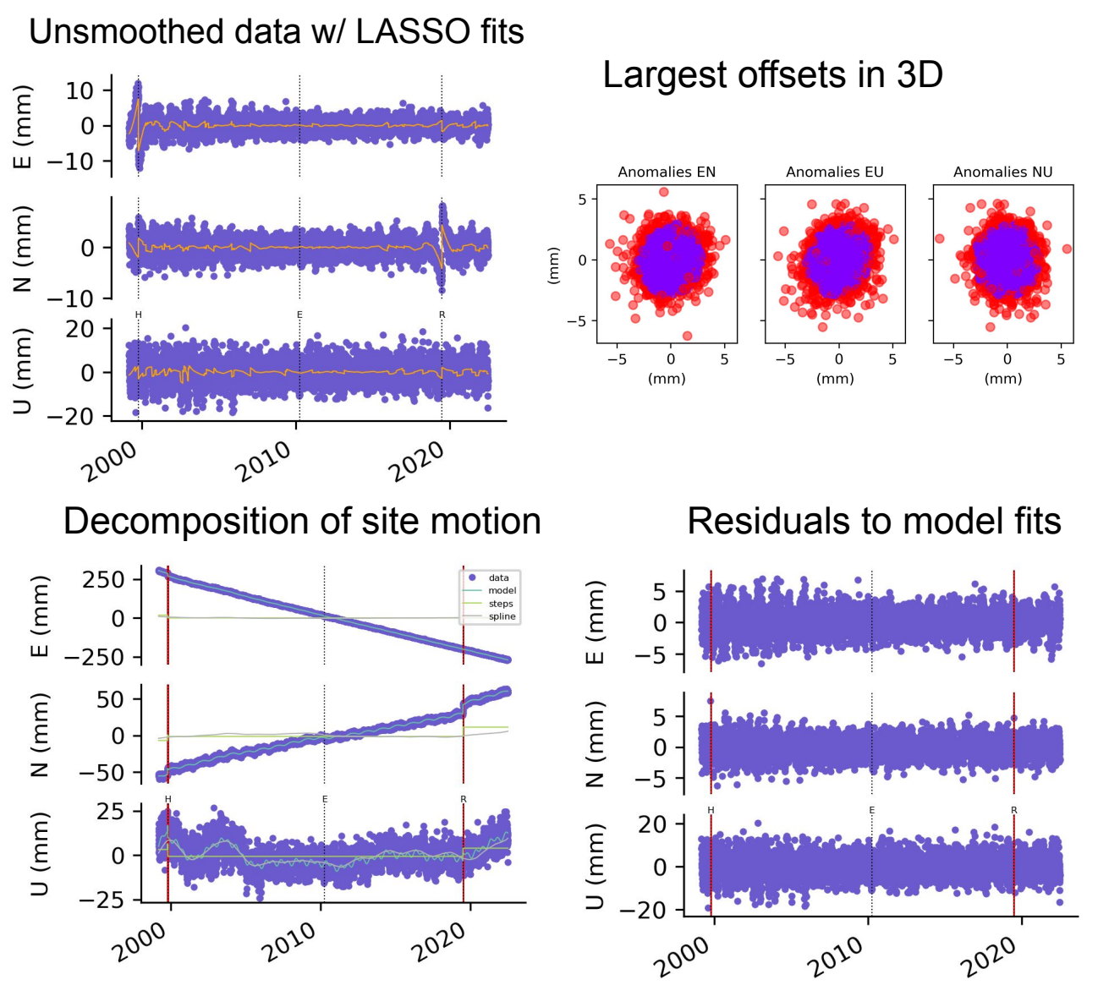
I. Pre-processing for outlier detection – gross outliers are detected using a modified z-score based on median absolute differences (MAD) after fitting the data with a loose Gaussian process model. Outliers are removed before proceeding to the next step.

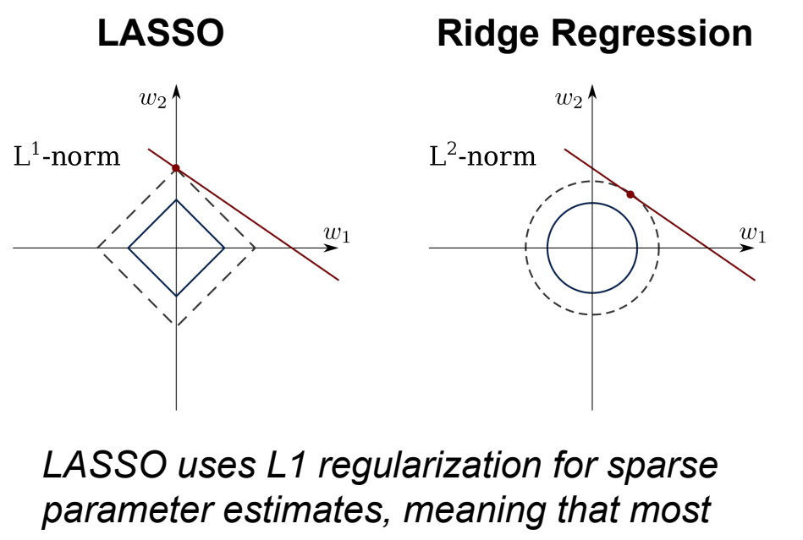
II. Iterative application of the LASSO regression followed by ridge regression (McDonald, 2009) to determine the epochs of offsets, with the LASSO tightened on each successive iteration until no statistically significant jumps remain undetected.



Example ENU time series for site SCIA in southern California. The site lies within 500 kilometers of three major earthquakes: 1999 M7.1 Hector Mine (**H**), 2010 M7.2 El Mayor-Cucapah (**E**), and 2019 M7.1 Ridgecrest (**R**). The **vertical black dotted lines** mark the epochs of the earthquakes. **Orange curves** shows the trajectory models fit to the SCIA time series. The epochs of the earthquakes were not used by the algorithm. Instead, the algorithm identified significant offsets at two of the three events (**red**) (Hector Mine and Ridgecrest). These were the two earthquakes closest to the site.

Application of the LASSO





*LASSO uses L1 regularization for sparse parameter estimates, meaning that most parameters are set equal to zero, e.g., the 2010 El Mayor-Cucapah (**E**) earthquake was not selected. The non-zero parameter estimates are biased, however.*

(**top left**) "Unsmoothed" data are scanned for large offsets (**red points**) in 3D (**top right**). With candidate jump epochs selected, the LASSO is run to asses their importance in fitting the data. LASSO picks are not adopted unless they are found to be statistically significant in a subsequent ridge regression solution for the coordinate function model.

(**bottom left**) The model consists of a traditional linear regression model (**Fα''**) with annual and semi-annual components, plus a smooth spline line function (**G'β'**) to capture nonlinear site motions and step functions (**Hγ**) at offset epochs.

(**bottom right**) Residuals to the model.

Model decomposition by cascading orthogonal projections

$$\tilde{y} = F\alpha + (Q_F + P_F)G\beta + H\gamma + \tilde{e}$$
$$= F\alpha' + G'\beta + H\gamma + \tilde{e}$$
$$= F\alpha' + G'\beta + (Q_A + P_A)H\gamma + \tilde{e}$$
$$= F\alpha'' + G'\beta' + H'\gamma + \tilde{e},$$

Matrices P and Q are orthogonal projectors, such that P + Q = I. The subscripts on these matrices indicate the associated subspaces (e.g., Q_F is a projector onto the subspace spanned by the columns of F, and P_F is a projector onto the nullspace of F).

The matrix A = [F G].

3. Performance evaluation

All algorithms have their shortcomings, but so do people. How good does an algorithm need to be before it may be suitable for operational research?

To begin to address this question, we performed three experiments described below and to the right.

3a. Experiment 1

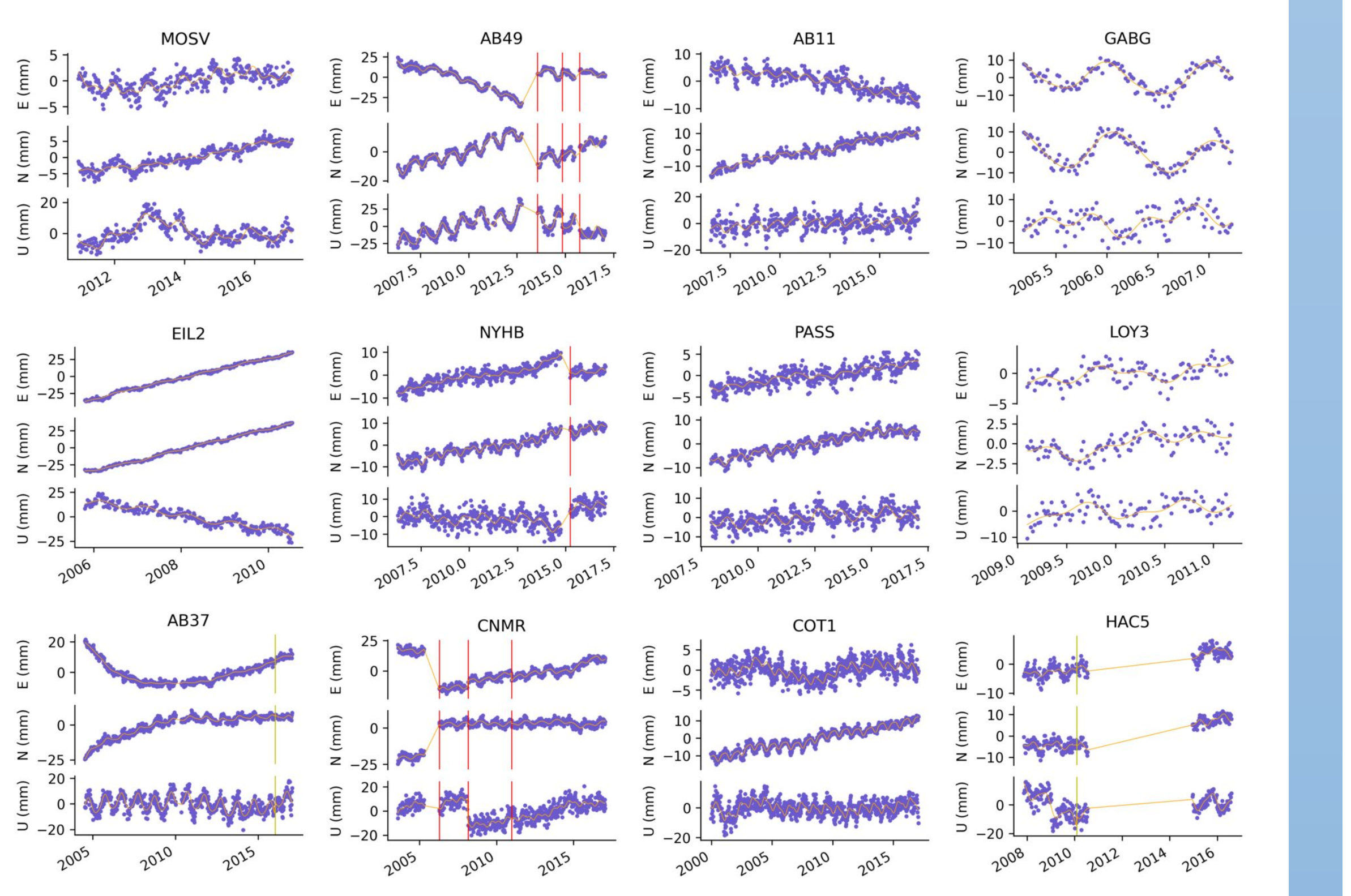
We used coordinate time series, velocity, and offset data products derived by two independent analysis centers (MEASURES, and the Geodesy Laboratory at the University of Nevada, Reno) from 59 continuously operating GNSS stations in the vicinity of the 2019 Ridgecrest earthquake sequence. We compared velocity estimates derived by our algorithm using GPS coordinate time series data from SOPAC (SOP) and University of Nevada, Reno (UNR), with velocity estimates from UNR's MIDAS solution (Blewitt et al., 2016), and MEASURES (Jamason et. al., 2004). A comparison of velocities determined by SOPAC and UNR with those obtained using our algorithm is shown in the table below.

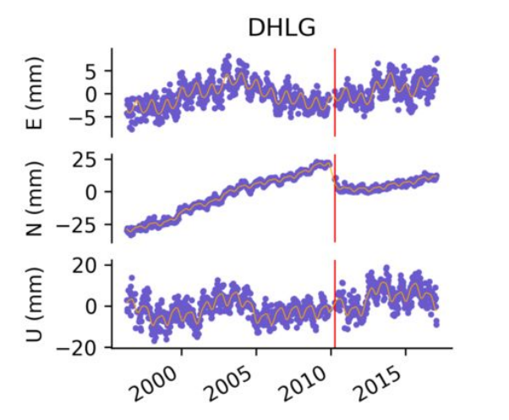
RMS velocity differences (mm/yr)
Summary statistics for 59 sites

Comparison - method (dataset)	RMS
This method (SOP) vs MEASURES (SOP)	0.7
This method (SOP) vs This method (UNR)	0.5
This method (UNR) vs MIDAS (UNR)	0.4
MEASURES (SOP) vs MIDAS (UNR)	0.4

3b. Experiment 2

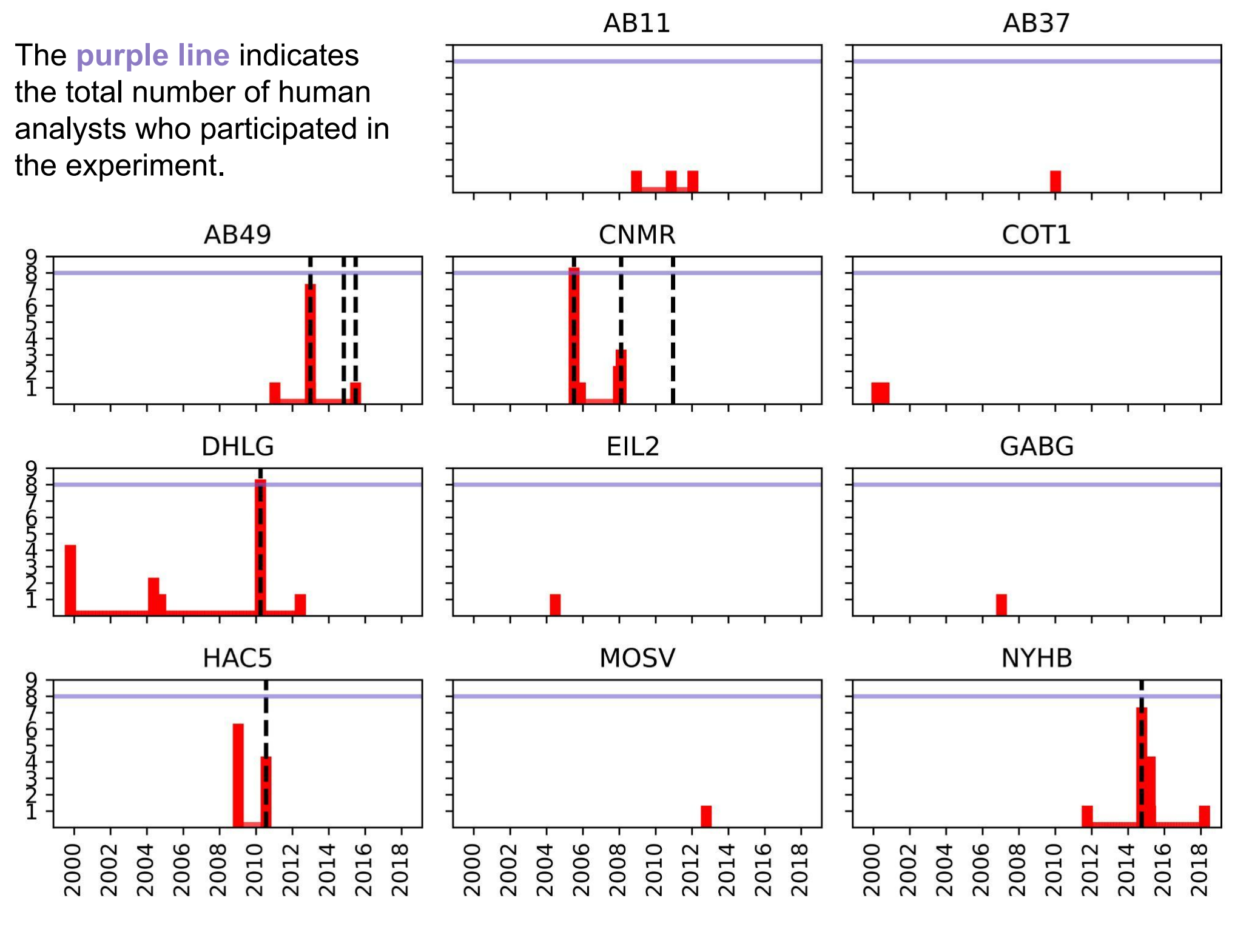
We used a set of 13 coordinate time series from a previous multi-year combination solution (MYCS2; Saleh et al., 2021) (selected by Jarir Saleh) to assess the algorithm's ability to accurately pick coordinate offsets. Eight human analysts independently picked the epochs of offsets by visual inspection of the time series. The automated picks are shown below.





(**above**) Automatically picked coordinate functions for 13 CORS (**orange curves**). Offsets are indicated by **vertical red lines**. **Yellow vertical lines** are possible offsets suggested by the code for human inspection. The straight line segments that span gaps in the data is a plotting artifact. The model was not computed for epochs within the gaps, although it could be with a little more coding.

(**below**) Histogram of offset picks by 8 human analysts (**red**) compared with offsets picked automatically by the algorithm (**vertical dashed black lines**). Variation among human analysts is apparent, with several offset epochs being picked by only one of the 8 analysts. Interestingly, only two of the offsets were picked by all 8 of the analysts. The automated algorithm picked two offsets (sites AB49 and CNMR) that was not picked by any human, but all other automated picks were corroborated by 3 or more human analysts. There are two epochs selected by 4 or more human analysts that the algorithm did not pick (sites HAC5 and DHLG). The offset in HAC5 is apparent in the vertical component. The algorithm did not identify this offset, but did flag the time series for human inspection. The "missed" pick for site DHLG was apparent to only 4 of the 8 analysts; who is right in this situation?

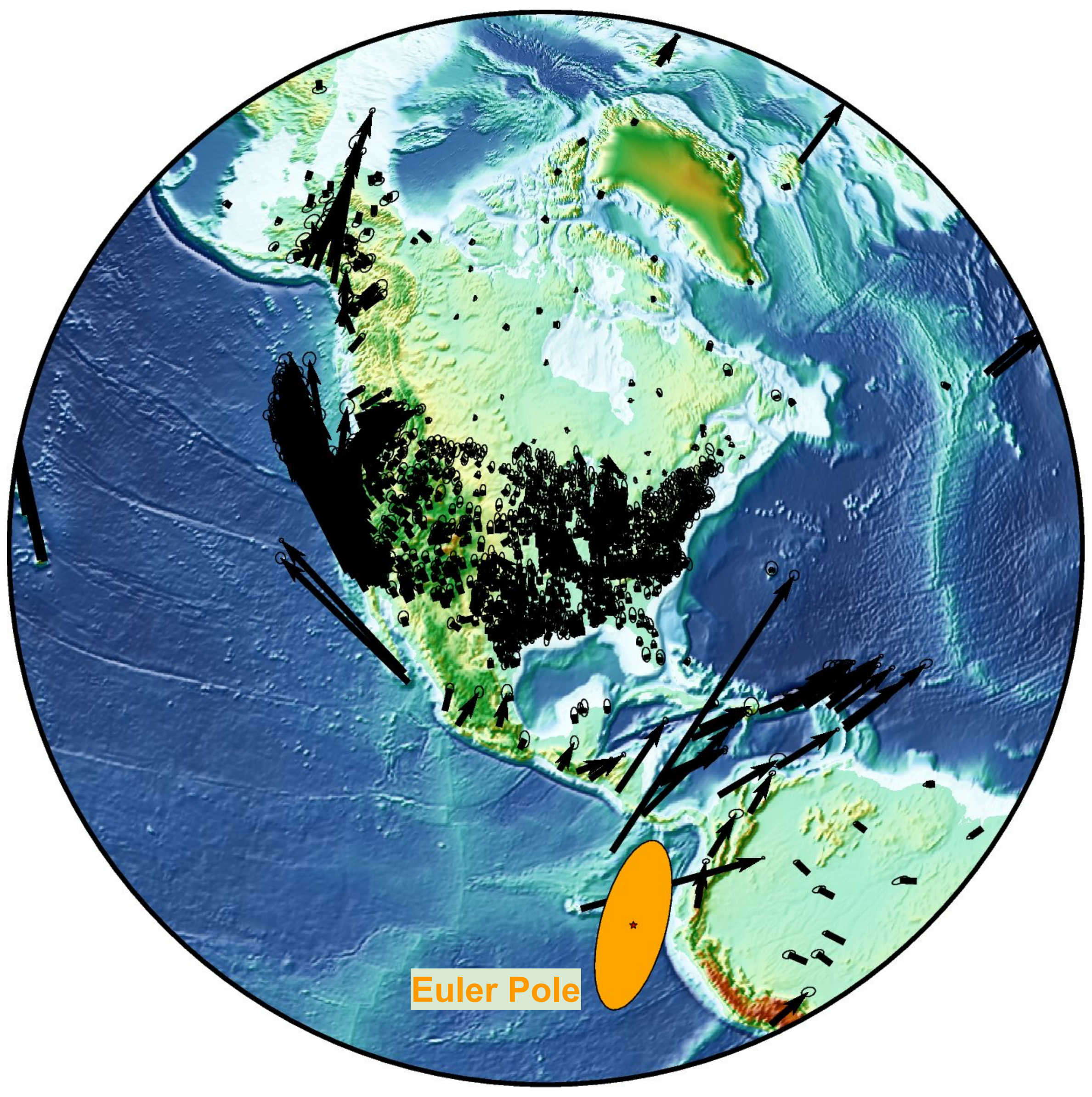


The **purple line** indicates the total number of human analysts who participated in the experiment.

3c. Experiment 3

We estimated trajectories models for all IGS stations using the LASSO algorithm and compared the resulting coordinate functions with official IGS20 products. Our analysis involved four main steps:

1. Analysis of 3222 continuously operating GNSS stations concentrated mostly in North America with the PAGES (double-differencing) software.
2. Stacking of the resulting weekly SINEX files through time using GLOBK to form precise coordinate time series aligned to the IGS20 frame.
3. Automated analysis of the coordinate time series data using the LASSO algorithm.
4. Using the offset and outliers identified by LASSO to estimate site coordinate functions, including velocities and offsets, using CATREF (Altamimi et al., 2024).

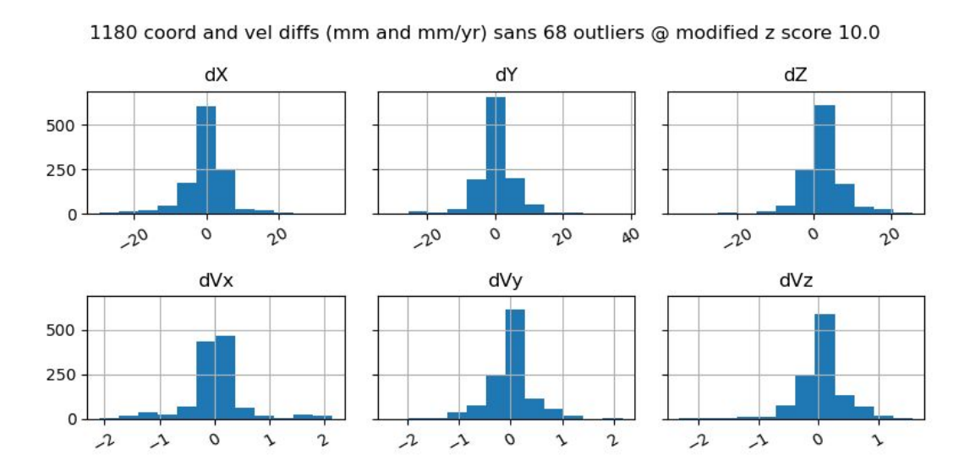


NOAA CORS Network IGS20 frame (North America plate fixed)

IGS20 auto-determined trajectories vs IGS20 products

STATISTICS

```
Removed 85 outliers after 9 iterations
Number of comparison points used: 1061
RMS East pos                :   15.154 mm
RMS North pos               :   14.633 mm
RMS Up pos                  :   31.560 mm
RMS East vel                :    1.433 mm
RMS North vel               :    1.379 mm
RMS Up vel                  :    3.075 mm
```



TRANSFORMATION PARAMETERS

```
       |     TX      TY      TZ      SC      RX      RY      RZ   |
-------|----------------------------------------------------------|
  pos  |     0.0     0.1     2.3    0.00  -0.024   0.035  -0.011  |
  +/-  |     0.4     0.4     0.4    0.00   0.015   0.015   0.015  |
  vel  |    -0.0    -0.1     0.0    0.00  -0.003   0.002   0.000  |
  +/-  |     0.1     0.0     0.0    0.00   0.002   0.002   0.002  |
-------|----------------------------------------------------------|
Units: mm, mm/yr, ppb, ppb/yr, mas, mas/yr
```

Trajectory models fit to IGS stations with the exact parameterization in the soln_IGS0.snx file also exhibit the same Z-bias, indicating that the bias is a deficiency in our realization of the IGS20 frame using GLOBK, rather than a consequence of missing an offset or other flaw of the automated algorithm. This does not affect the performance of the automated time series analysis in any way.

IGS20 CATREF w/ auto picks vs IGS20 products

STATISTICS

```
Removed 67 outliers after 4 iterations
Number of comparison points used: 798
RMS East pos                :    4.062 mm
RMS North pos               :    3.754 mm
RMS Up pos                  :   11.528 mm
RMS East vel                :    0.544 mm
RMS North vel               :    0.465 mm
RMS Up vel                  :    1.505 mm
```

TRANSFORMATION PARAMETERS

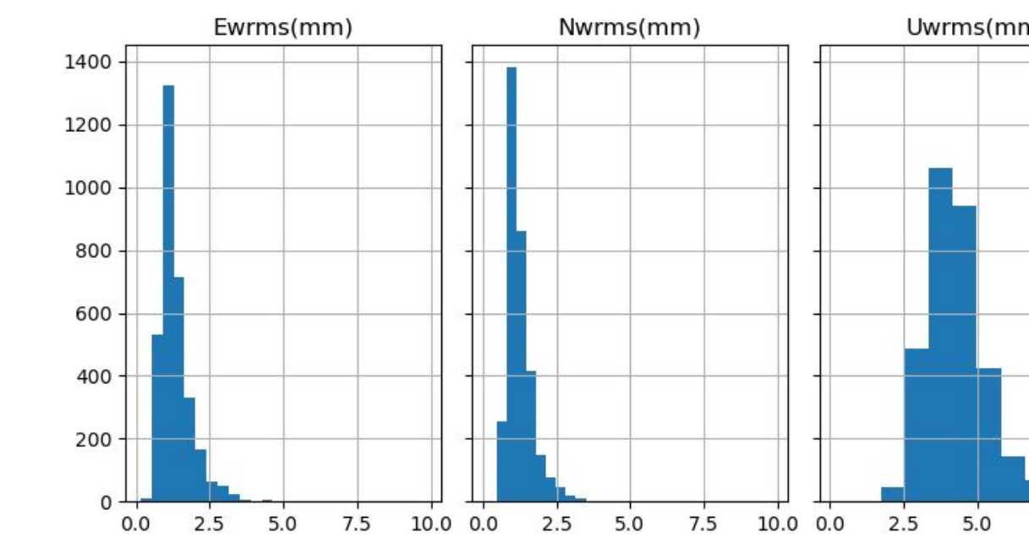
```
       |     TX      TY      TZ      SC      RX      RY      RZ   |
-------|----------------------------------------------------------|
  pos  |    -0.4    -0.0    -0.9    0.00  -0.001   0.014   0.039  |
  +/-  |     0.1     0.1     0.1    0.00   0.004   0.004   0.004  |
  vel  |    -0.0     0.0    -0.0    0.00   0.001   0.002   0.001  |
  +/-  |     0.0     0.0     0.0    0.00   0.001   0.001   0.001  |
-------|----------------------------------------------------------|
Units: mm, mm/yr, ppb, ppb/yr, mas, mas/yr
```

Summary statistics

ALL DATA SUMMARY (count 3222 minus 0 fails)

	Used	Omits	Brks	Ewrms(mm)	Emae(mm)	Nwrms(mm)	Nmae(mm)	Uwrms(mm)	Umae(mm)
mean	579.1	43.0	2.5	1.32	1.01	1.26	0.97	4.32	3.33
std	272.1	44.5	2.4	0.52	0.39	0.44	0.33	1.05	0.78
min	10	0	0	0.16	0.13	0.15	0.12	0.14	0.11
25%	391	16	1	0.98	0.76	0.96	0.75	3.61	2.81
50%	**583**	**29**	**2**	**1.19**	0.93	**1.15**	0.89	**4.19**	3.26
75%	764	52	3	1.53	1.18	1.43	1.11	4.84	3.76
max	**1385**	464	41	4.63	3.71	4.16	3.53	9.84	7.50

wrms = weighted root-mean-square, mae = mean absolute error



The distribution of WRMS values is slightly skewed due to a few larger WRMS values. The median values are **1.2 mm** and **4.2 mm** in the horizontal and vertical components, respectively. The mean duration of the time series is 11 years (**583 weeks**), and the maximum number of years is 26 (**1385 weeks**). On average (median) there were **2 breakpoints** inserted and **29 outliers** assigned by the automated time series analysis algorithm, showing that the low WRMS values are not in general a consequence of excessive breakpoints additions or outlier rejections.

4. Conclusions

Automated GNSS coordinate time series modeling appears to be a competitive alternative to visual inspection by human analysts.

5. References

- Altamimi, Z., P. Sillard, X. Collolieux, P. Rebischung, CATREF Software: Combination and Analysis of Terrestrial Reference Frames, April 29, 2024.
- Blewitt, G., Kreemer, C., Hammond, W. C., & Gazeaux, J. (2016). MIDAS robust trend estimator for accurate GPS station velocities without step detection. *Journal of Geophysical Research: Solid Earth*, *121*(3), 2054-2068.
- Jamason, P., Y. Bock, P. Fang, B. Gilmore, D. Malveaux, L. Prawirodirdjo, and M. Scharber (2004). SOPAC Web site (http://sopac. ucsd. edu). *GPS solutions*, *8*, 272-277.
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93-100.
- Saleh, J., S. Yoon, K. Choi, L. Sun, R. Snay, P. McFarland, S. Williams, D. Haw, F. and Coloma (2021). 1996–2017 GPS position time series, velocities and quality measures for the CORS Network. *Journal of Applied Geodesy*, *15*(2), 105-115.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.