Addendum to: Methods for rapidly estimating velocity precision from GNSS time series in the presence of temporal correlation: a new method and comparison of existing methods, https://doi.org/10.1029/2019JB019132

John Langbein

April 22, 2020

Guidance for assessing the reliability of rate uncertainties calculated by the modified version of *midas*.

The primary goal of the *midas* program is to provide a fast and robust estimates of secular rates from GNSS time series without having to physically look at the time series. This lends itself to analyzing large networks that encompasses many GNSS sites. With the additional variance of rate (*midas_vr*) described in *Langbein* [2020], one should also question the reliability of the estimated rate uncertainty provided by the modified version of *midas*. In this addendum, I'll provide some guidance, but, in the end, one needs to be skeptical of methods that skip the necessary step of physically examining each time series.

Although the original *midas* program can handle time series with sparse observations that are associated with campaign style GNSS measurements, the VR algorithm assumes that there are frequent observations such that the time series can be broken into short intervals in order to estimate rates spanning a variety of intervals. Consequently *midas_vr* assumes that the data are from continuously operating GNSS sites with roughly one measurement per day. Built into the VR algorithm is a test applied to each interval under analysis to determine whether that interval is atleast 75% complete. If the completeness falls under 75%, that interval is not included in the statistics for the variance in rate computed for that interval. The specification of 75% was an arbitrary choice by me. None the less, although *midas_vr* will provide a rate uncertainty for time series with less than 75% completeness, it will base its results on just a few intervals that meet the 75% threshold. So, I suggest not using *midas_vr* to analyze GNSS time series with less than 75% completeness. The output of *midas_vr* provide information on the completeness, specifically

columns 5 through 7.

The other assessment that *midas_vr* provides is a measure of fit of the observed variance of rate to the modeled variance in rate, which as the form of $a\tau^{\nu-3}$. That is, repeating equation 6 from *Langbein* [2020],

$$\sigma^2_{vr}(\tau) = 12\sigma^2_{wh}\tau^{-3} + A\tau^{(\nu-3)} + B_{365}F(365,\tau) + B_{182.5}F(182.5,\tau) \quad (1)$$

where τ is the length of the interval for the calculation of the variance of rate, and $F(T,\tau)$ is equation 10 of *Hackl et al.* [2011] Corresponding to each interval is a weighting term that is proportional to $\sigma^2_{vr}(\tau)/n^{0.5}$, where n is of intervals used to compute the variance in rate for each interval spanning τ days. Therefore, the quality of fit is

$$RMS_{fit} = \frac{1}{m}[\sum_{i=1}^{m}(\sigma^2_{vri} - \hat{\sigma^2_{vri}})^2/\sigma^2_{wght_i}]^{0.5} \quad (2)$$

where m is the number of intervals analyzed and $\hat{\sigma^2_{vri}}$ is the predicted value of the variance in rate using equation 1.

To get a sense of the ranges of RMS_{fit} for a well behaved examples, I made a histogram of RMS_{fit} for all of the simulations (4200) that I did in the published paper [*Langbein*, 2020] which is shown in Figure 1. RMS_{fit} ranges between 1 and 5, with a median value of 2.8 and with 95%, 99% and 99.5% of RMS_{fit} less than 3.6, 4.1, and 4.8 respectively.

For the 569 time series of real data analyzed by *Langbein* [2020] for which the completeness is greater 75%, those histograms of the distribution of RMS_{fit} are shown in Figure 2. Recall that the VR2 option relied upon using the program *est_noise* to aid in the removal of outliers and to remove known offset. Since that option is not in the spirit of *midas* where any hands-on editing of the data is not allowed, I will only discuss the VR1 and VR3 histograms as VR1 uses *midas* where known offset are acknowledged, and VR3 where known offsets are ignored. Both of those histograms show a larger spread than those from the ideal cases derived from simulations. Where with the simulations, the 95% and 99% levels were 3.6 and 4.1, those levels for RMS_{fit} increased to 5 and 6, respectively for real data. I examined plots of four of the time series with the largest RMS_{fit} (MONB east, P215 up, LUTZ n, and P267 up) and in two of the cases, P215 up and LUTZ n, I found some transient changes in the time series which may have impacted the analysis. For the other two time series, I found no obvious causes for a high RMS_{fit}.

A further test of the efficacy of the RMS_{fit} statistic is to reproduce part of Figure 3 from *Langbein*, [2020] and labeled Figure 3 here. This compares the rate uncertainties computed by *est_noise* with those computed with the modified version of *midas* using real data. Only those

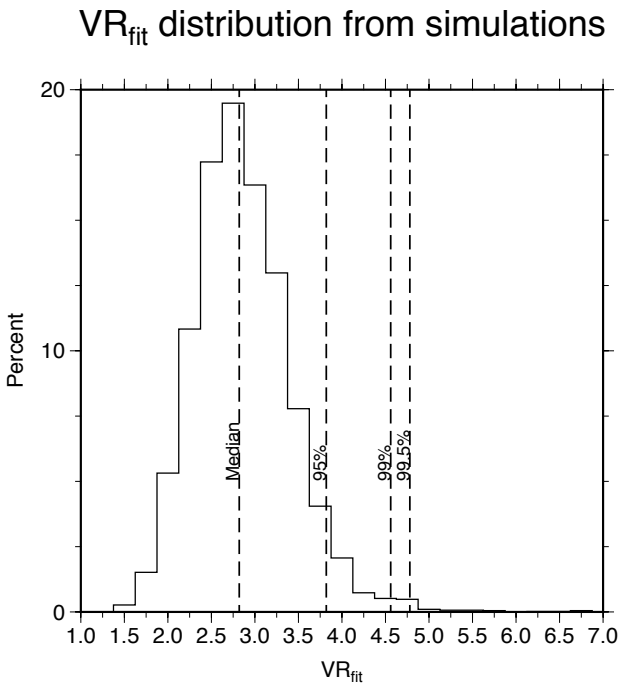


Figure 1: Histogram of the goodness of fit of the regression of observed variance of rate to the modeled variance of rate as measured using equation 2. Data are from 4200 simulations from *Langbein*, [2020]

time series that are more than 75% complete are shown. In addition, the points with $RMS_{fit} > 6$ are plotted with larger symbols. Visually, the scatter about the trend of the larger symbols is about the same as the other, smaller symbols. Consequently, if RMS_{fit} is large, it does not necessarily mean that the calculated rate uncertainty from the VR algorithm is incorrect.

Another potential tool is to compare the rate uncertainties between those calculated by *midas* and that calculated by the VR algorithm. Those results are shown in Figure 4. Two of the large symbols shown for VR1 which deviate most from a linear trend are two time series LUTZ n (0.1 vs 0.9) and MONB east (0.1 vs 0.9). These were previously identified with excessively large RMS_{fit}.

Finally, I downloaded and analyzed approximately 7000 time series from the Nevada

Geodetic Laboratory (NGL) with an American reference frame with *midas_vr*. On a two core, MacIntosh laptop, it took a little over 2.25 hours to complete the download and analysis using a residential, cable modem. I selected 5900 sites that were at least 75% complete and had more than 5 years of data. For this set of examples, I did not use the list of known offsets provided by NGL; had I included the known offsets, the results discussed below would be slightly different. A scatter plot of rate uncertainties provided by *midas* and the VR algorithm is shown in Figure 5a. Like the above figures, those points with $RMS_{fit} > 6$ are plotted with large circles. They are rather scarce; only 0.5% of those points had RMS_{fit} that exceeded 6. Perhaps more diagnostic are points that deviate greatly from their nominal trends established by fitting a function, $\sigma_{vr} = A * \sigma^{B}{}_{midas}$ where σs are the rate uncertainties computed by *midas* and the VR algorithm. They can help identify data that one or the other method, or both, failed to provide a reliable estimate of the rate uncertainty. To highlight the points that deviate most from the trends, I have identified them with large, black squares.

A slightly refined version of the same analysis was done by only looking at sites processed by NGL located in the western US and southwestern Canada; that removed about 4500 sites. Figure 5b shows a scatter plot of the rate uncertainties calculated by both methods The range of uncertainties is two orders of magnitude less than that from the entire data set. However, there are outliers from the cloud of points.

I studied a few of the time series of the subset of data shown in Figure 5b that had both a large misfit to the least squares trends and large RMS_{fit}. These are listed in the first half of Table 1. The second half of Table 1 are time series with only a large misfit to the least squares trends. The notes specify what I observed by looking at the time series, and in a few cases, using *est_noise* to calculated the rate uncertainty.

In many cases, especially those time series with both large σ_{vr} and large misfit to the least squares trend in Figure 5b, inspection of the time series reveals significant problems with the data; thee are listed under the notes column in Table 1.

In other cases, the time series had very low noise; the VR algorithm indicated that the index of power law noise was less than that for flicker noise. For the case both P738 and P535, *est_noise* validated the power law index and likewise, the rate uncertainty. But, in other case, like IDIR, although the VR algorithm indicated a power law index of 0.6, *est_noise* could detect some random walk component which increases the estimate of rate uncertainty to 0.41 mm/yr

In some cases, it is worth examining the optional output of the *midas_vr* by using the *-o* option. For each component, the program outputs three files: one being the calculated values of the variance of rate over different periods up to one-quarter of the length of the time series; a second file consisting of the predicted values variance of rate extrapolated to span the length of

Table 1: Data with potentially suspicious rate uncertainties

Site, comp.	rate uncertainty, mm/yr		RMS_{fit}	misfit to	location	notes
	σ_{midas}	σ_{vr}		LSQ fit		
HVWY u	1.6760	0.0560	5.68	3.97	Yellowstone	multiple inflation deflation events; 90mm
AVRY e	0.1290	1.4900	21.92	2.13	S. Calif.	half of time series noisy
P667 u	0.9610	23.4020	12.46	2.97	Mt. Lassen	winter snow
P667 n	0.3280	7.9720	9.61	2.47	Mt. Lassen	winter snow
P162 e	0.1330	1.6440	11.25	2.18	N. Calif.	multiple offsets; midas rejected
					C. Mendocino	several intervals est_noise 0.17mm/yr
RMRK u	0.9570	10.7090	8.65	2.20	Idaho	Numerous, unknown offsets est_noise 0.51mm/yr
SOMT n	0.0930	0.6600	6.81	2.03	S. Calif	est_noise 0.17 mm/yr
WDCB n	0.5180	0.1340	3.31	2.36	N. Calif.	Large white noise est_noise 0.31mm/yr
IDIR n	0.3120	0.0610	2.66	2.33	Idaho	VR; low index 0.6 PL est_noise 0.41mm/yr
COND n	0.1590	1.9860	2.39	2.26	Central OR	Nothing obvious est_noise 0.097mm/yr
P738 e	0.1670	0.0300	3.28	2.20	Central OR	VR; low index 0.5 PL est_noise 0.035mm/yr
P535 e	0.1200	0.0180	3.13	2.17	S. Calif.	VR; low index 0.4 PL est_noise 0.024mm/yr

the time series; and a third file consisting of the various trials of functional models used to fit the calculated variance in rate. The first two files can be used to make plots.

One example, shown in Figure 6, is a of the variance in rate data is shown for the site RLAP (located in the Mississippi valley). Note that in Figure 6a, there is a gross outlier from the trend at 2000 days. Consequently, the fit to the data is poor ($VR_{fit} = 123$). In this case, I chose not to use the "step option" in *midas*. However, when the step option was used (Figure 6b), the outlier was eliminated and the fit of the trend is satisfactory. With the better fit, the calculated uncertainty in rate from *midas_vr* is 0.17 mm/yr, consistent with that from *est_noise* of 0.26 mm/yr. Note, to achieve reasonable results from *est_noise*, I needed to look at the data from that site (http://geodesy.unr.edu/NGLStationPages/stations/RLAP.sta) and remove several years of noisy data.

In another example, the vertical component at the site HVWY (located in Yellowstone), shows the impact of using different weightings when the a function of variance in rate is fit to the observed variance in rate. With the default weighting, 0.5, discussed briefly in the JGR paper, the fit is unsatisfactory. By changing the weight (the -W option in *midas_vr*, to 0.25, it gives the variance in rate at the longer periods more "weight". Note that the long period trend is roughly $\tau^{2.5-3}$, characteristic of a power law with a high, 2.5, index. The listing in Table 1 indicates that this site under goes significant uplift and subsidence from Yellowstone caldera and that the apparent "noise", is due to this long period, transient behavior.

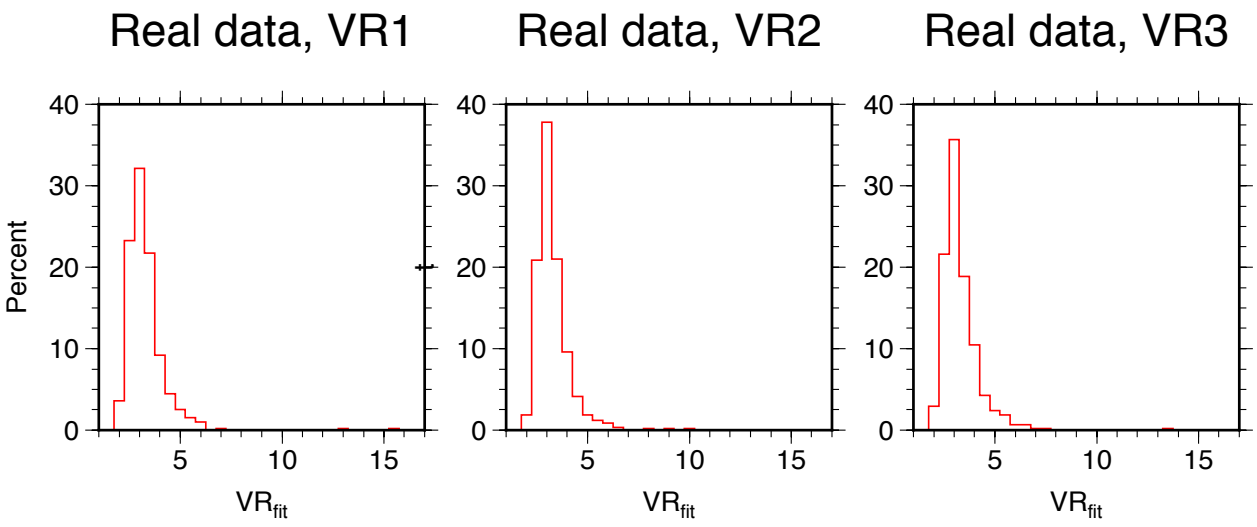


Figure 2: Histogram of the goodness of fit of the regression of observed variance of rate to the modeled variance of rate as measured using equation 2. Data are from 560 time series of real data analyzed by *Langbein*, [2020]

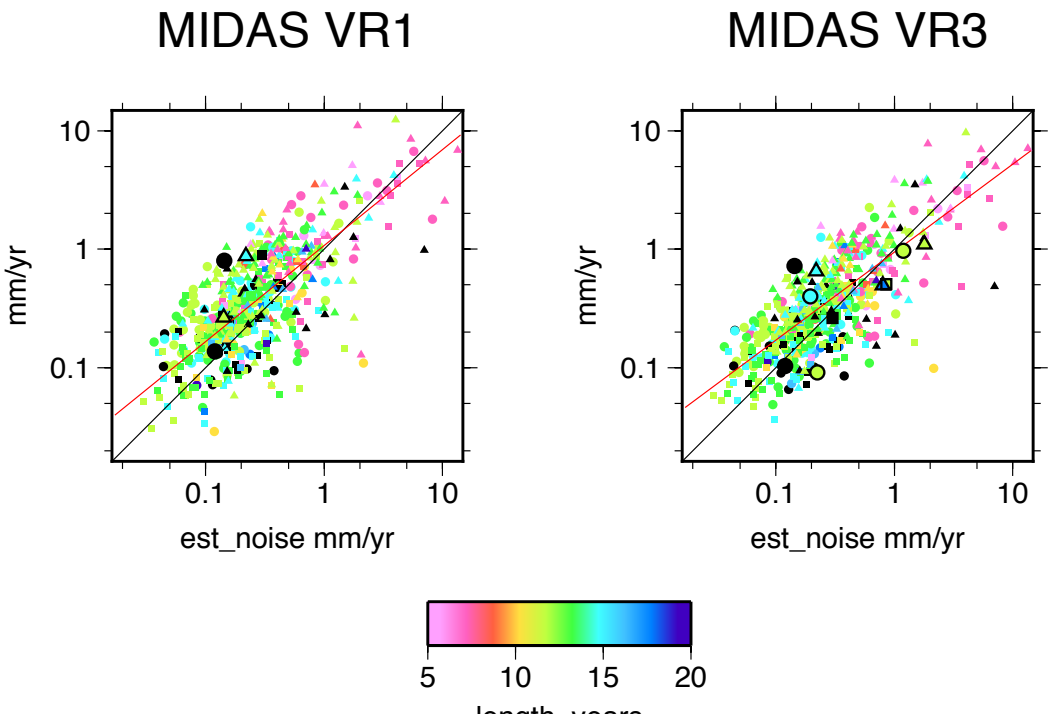


Figure 3: Plots showing the rate uncertainties computed from real data using the modified version of *midas* with the VR algorithm in comparison with rate uncertainties estimated using *est_noise*. Data are from time series for the sites in the greater San Francisco Bay area. The results are color coded as to the length of their time series. Horizontal and vertical uncertainties are shown. Circles are east component, squares are north component, and triangles are vertical component. The black line represents equivalence of estimated rate uncertainties between the methods. The red line is a best fit to power-law trend, $y = \alpha x^{\beta}$, between those estimated by *midas* and *est_noise* Unlike the original figure 3 in *Langbein* [2020], only the results from time series with better than 75% completeness. **In addition, the large symbols outlined in black are for the rate uncertainties from calculated by the modified *midas* with** $RMS_{fit} > 6$.

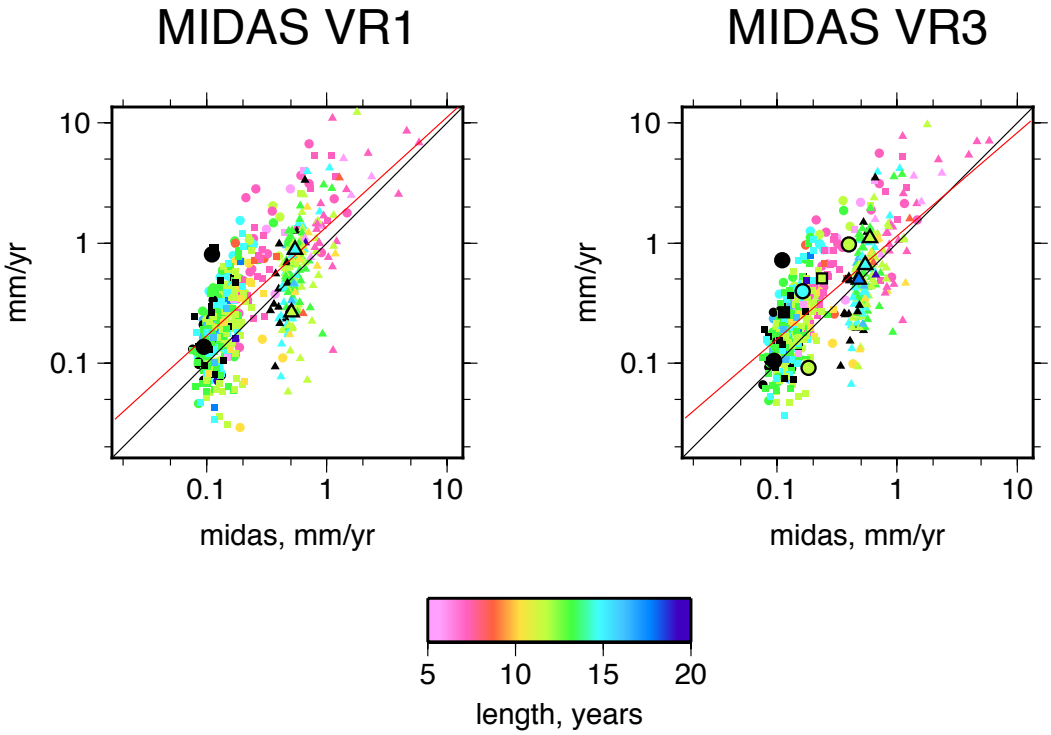
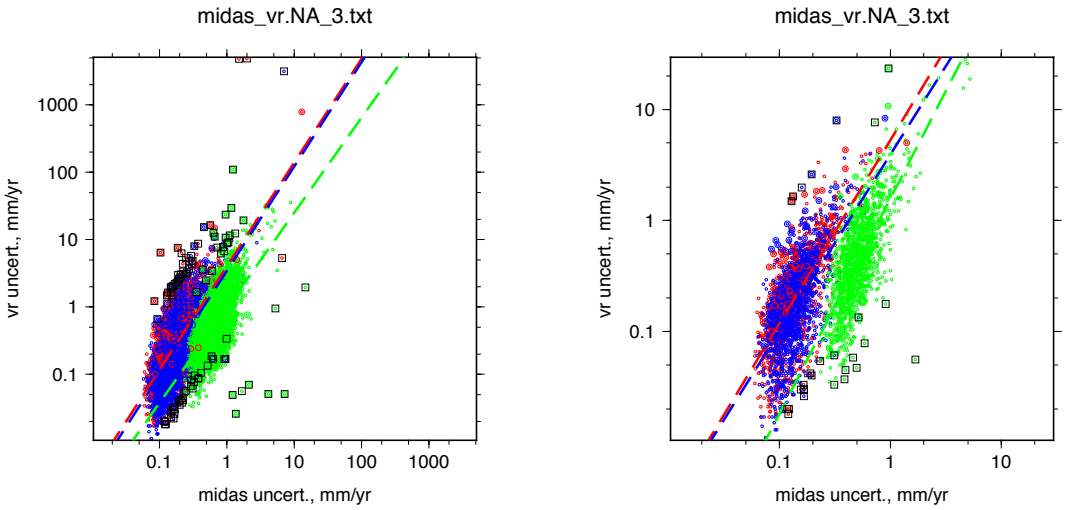
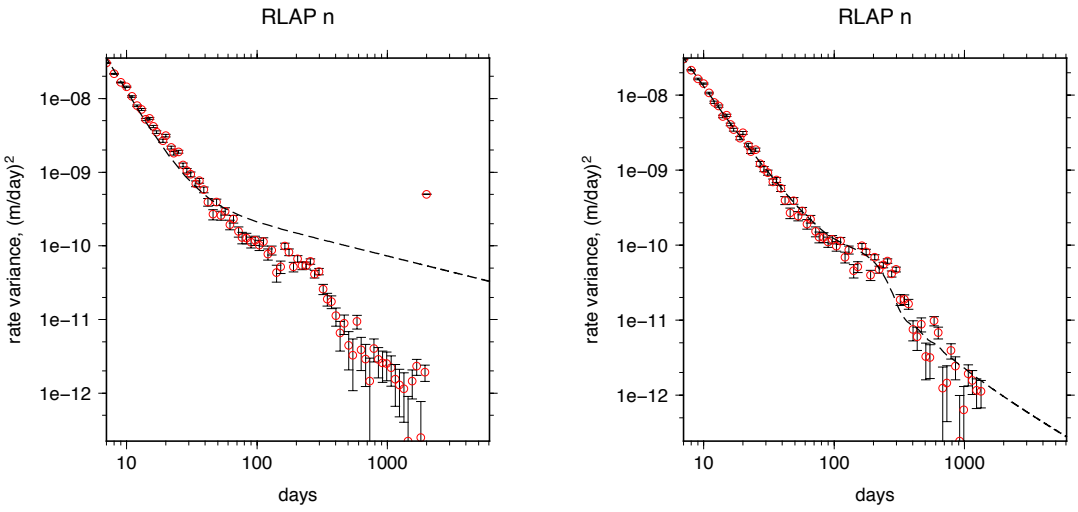


Figure 4: Similar plot as Figure 3 but rather than comparing the VR uncertainties with those from *est_noise*, the comparison is with those calculated by *midas*. The large symbols outlined in black are for the rate uncertainties from calculated by the modified *midas* with $RMS_{fit} > 6$



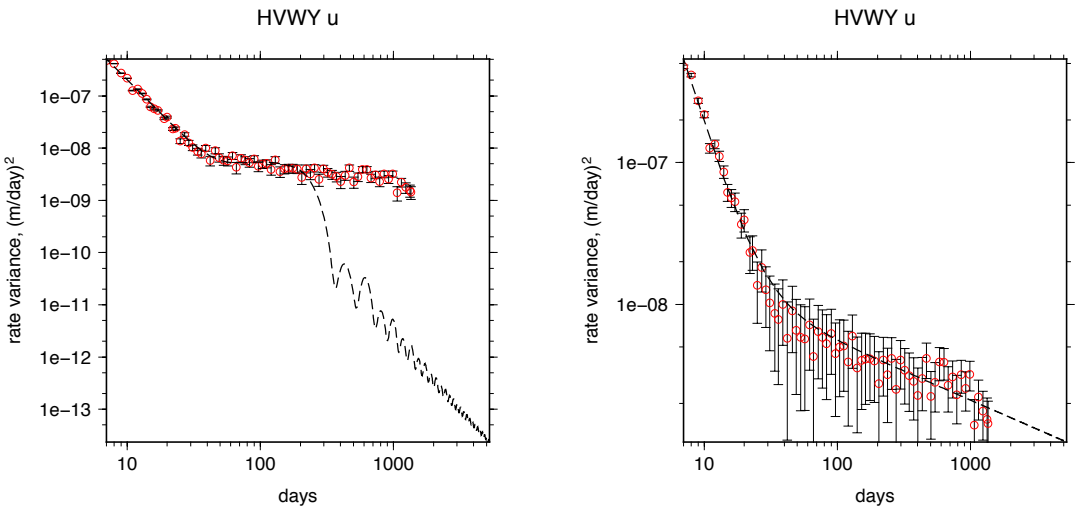
(a) (b)

Figure 5: A scatter plot of rate uncertainties calculated by *midas* and the VR algorithm using data processed in the North American reference frame (NA12) by the NGL. Dots with red are east, blue is north, and green is up. The large colored symbols are for the rate uncertainties from calculated by the modified *midas* with $RMS_{fit} > 6$. The dash lines are a least squares fit to the rate uncertainties calculated by both methods can color coded to their component. The left hand plot (a) consists of results from 5900 sites on the North American plate analyzed by NGL. The right hand plot (b) is a subset of sites, 1400, that are also analyzed by UNAVCO and located in the western US and southwest Canada. (Note that I used the NGL solutions.



(a) (b)

Figure 6: Example for the site RLAP showing the calculated variance in rate, along with its "weight", and the predicted values of variance in rate. In A), the known offsets were not factored into the calculations, while in B) the offsets were factored into the calculations. Note the *midas* does not calculate the size of the offsets, but, instead, deletes intervals where offsets are known to exist.



(a) (b)

Figure 7: Example for the site HVWY showing the calculated variance in rate, along with its "weight", and the predicted values of variance in rate. In A), the default weighting, 0.5, for the calculated variance in rate was used with fitting a trend. In B) the weighting was changed to 0.25 to provide more influence of the longer intervals on fitting the data to the trend.