

Interactive comment on “Precision of continuous GPS velocities from statistical analysis of synthetic time series” by Christine Masson et al.

Hammond (Referee)

whammond@unr.edu

Received and published: 31 December 2018

General Comments:

Masson et al. present an analysis of constructed position time series that were made using analytical forms of signal and noise that are typically observed in geodetic GPS data. They then use these synthetic time series to identify the factors that contribute the most to uncertainty in the estimated trend. Once the most important factors are identified (time series length, spectral content and amplitude of colored noise, etc.), they offer a few rules of thumb that can be applied to categorize time series according to how precise they are expected to be. One of their conclusions is that the time series duration is invariably the most influential factor in maintaining low uncertainty in velocity

C1

estimation, which is very important when considering how GPS networks are funded and maintained.

Aside from the focus on synthetic time series, two elements in the paper stand out as looking new to me. First, they employ a regression-tree approach that rank orders parameters in terms of their overall impact on the velocity uncertainty. This is useful since these factors are sometimes known for real data in advance and can be used to generate expectations for which time series provide the lowest uncertainties, before any more detailed analysis is undertaken. Second they introduce a new method for scanning the time series for discontinuities that are undocumented, i.e., their existence and time of occurrence are unknown beforehand. Their method is interesting because they flip the problem by determining which epochs contain *no* step to within data uncertainty, thereby narrowing the set of epochs that could have steps.

I had a number of suggestions, mostly minor which I placed in the technical comments below. The introduction could benefit from some short additional text, possibly in the last paragraph, on the general value of looking at synthetic time series as opposed to 1) real ones when so many are available, or 2) simple formulas that mathematically represent the content of signals+noise in them (e.g., Williams, 2003). The answer might be e.g., the ability to know the true answer in order to evaluate the validity of false and true detections, which might be obvious at the onset to expert readers but not everyone. That paragraph would be a good place to also mention the limitations of an analysis like this, since many real GPS time series contain signals of types not included in their synthetic tests. They mention a few examples in the paper but do not discuss the impact of the potential presence of these signals in detail.

Some detailed/technical comments:

Page 2 line 26. Could replace "low deformation" with "low rate of deformation"

Page 2 line 31. In equation 1 they may have meant to use $H(t)$ rather than the Kronecker delta function to indicate the occurrence of a step in the time series. $H(t)$, the Heaviside

C2

function, is zero before t and one after t , and is also the time integral of the Dirac delta function. The Kronecker delta function is a discrete version of the Dirac delta function, in physics literature. Here Masson et al., define $\delta(t)$ in a way that works for their paper so it is probably all OK and self-consistent here, but might cause some minor confusion to call it the “Kronecker delta function”.

Figure 2. It seems odd to me at first to lump all the horizontal and vertical data together in the analysis, and in this one plot. I guess it all works out in the end. But I wondered if including a new binary parameter in the regression tree, horizontal vs. vertical time series, would have a strong predictive ability in the tree.

Figure 3. The caption lists values of k as positive, but on page 3 they are said to be always negative.

Page 5 line 26. It is a little confusing sometimes that they interchange the terms “accuracy” and “uncertainty”. For example, Figure 4 is a nice plot, but “accuracy” should be changed to “uncertainty” since accuracy should improve (increase) with time series duration, but the quantity shown decreases with time series duration. Also in Figure 4, they should state in the caption what is the meaning of the vertical extent of the vertical black bar, and also what is indicated by the extent of the blue box. Then in 5, 6, 7 it can be said they are as in Figure 4. These plots may be standard in some literature but probably not everyone will already know the details of construction.

Page 5 line 25. “possible asymptotic value ca. $v_{95} = 0.05 \text{ mm yr}^{-1}$ ”. Possibly? In Figure 4 it looks like the uncertainty is still decreasing, though more slowly, at duration 20 years. I would have thought that theoretically the asymptote would be $v_{95}=0$ and maybe we will do at least a little better if we run a GPS station for 100 or 1000 years. I don’t see evidence of an asymptote at 0.05 mm yr^{-1} .

Figure 8. I think in the caption v should be v_{95} ? The “95” is dropped in several places when it should be included.

C3

Page 6 line 18. Probably meant “if the series is short.”?

Page 8 line 11. it would be better to use a lower case t , rather than T for the step time so as not to confuse with time series duration.

page 8 line 14. “consists in” should be “consist of”

Page 8 line 16. Instead of “amp_off”, notation for non-offsets might be better stated in same class as true offsets, e.g., $C_{\text{notanoffset}}$ or something shorter.

Page 8 line 25. I have a few questions about their interesting new offset detection method. First, it seemed that a part of the explanation may be missing. It is stated that it is repeated “until a significant offset is found”. But once an offset is found there could be others and, since only evenly spaced arrays of offsets are tested in each iteration, there may be epochs that have not yet been tested. So how does the algorithm guarantee the completeness of the scan for a step at every epoch? Secondly, if a large true step exists and the adjacent epoch is tested, it will likely be evaluated as a significant step. Is there a mechanism to replace the adjacent epoch with the correct one once it has been tested? When the process is repeated are steps and non-steps identified in previous iterations excluded from being considered as steps? If so I expect that would improve efficiency and reduce ambiguity in the algorithm. Finally, could this method be applied to real data? It seems that the calibration method determining b and δ_t in Appendix B relies on the quantity of false and true identifications, so might not be available for real data(?).

Page 9 line 21. “real” need not be in quotes.

Page 9 line 27. “is no the case” should be “is not the case”

Page 10 line 14-22. These classifications may be useful but possibly a bit dismissive of the utility of some of the categories when signals are large enough to stand out from the noise. For example the Oregon coast rises $>4 \text{ mm/yr}$ owing to elastic strain accumulation on the subduction zone. Even short time series may be useful there.

C4

Page 10 line 31. “These results may indicate a lower limit in velocity accuracy ca. 0.1 mm yr⁻¹”. But it said in the previous sentence that some were 0.05 mm yr⁻¹...

Page 11 line 13. “v” should be “v₉₅”? Also “A ratio of 1 corresponds to a standard error equal to its velocity”. In a Gaussian distribution +/- one standard deviation contains 68% of the samples, whereas the definition of v₉₅ in this paper is the limit that contains 95%. So would not a ratio of 2 indicate that the standard error and velocity accuracy are similar?

Page 12 line 14. I did not see Masson et al., 2018 in the reference list.

Page 13 line 28. “A significant outcome of our analysis is the fact that very long series durations (over 15 – 20 years) do not ensure a better accuracy compare to series with 8 – 10 years of measurements”. However, Figure 4 says they are still getting more precise even at 20 years (though apparently at a decreasing rate of improvement) so I’m not sure if this statement is strictly true. It may be true that if a specific requirement for uncertainty is 0.1 mm/yr then there is no need to collect longer time series, but that requirement standard depends on the application and we may not yet know all future standards that are needed from the data.

Page 14 line 5. Acknowledgements sections often now contain proper attribution to those who collected (in this case the RENAG network), archived, processed the data, and from where the processed time series were downloaded, i.e ftp server, web site, etc., and on what date. In this case the authors may have had prior access to the data (?), i.e. processed it themselves, but it would improve repeatability of this work if others could be guided to where they could access the data. Separate questions: Are the synthetic time series developed here openly available?

Page 25, line 5. Why not show b=10, discussed in the text, on the plot?

Interactive comment on Solid Earth Discuss., <https://doi.org/10.5194/se-2018-77>, 2018.