Answers to the Interactive comment on "Noise characteristics in Zenith Total Delay from homogeneously reprocessed GPS time series" by Anna Klos et al. written by Anonymous Referee #3

Review of manuscript amt-2016-385: "Noise characteristics in Zenith Total Delay from homogeneously reprocessed GPS time series" by A. Klos and co-authors.

This work investigates deterministic and stochastic modelling of reprocessed GPS ZTD time series from a global GPS network composed of 120 stations. The deterministic component of ZTD series has been quite thoroughly analysed in past studies, namely their seasonal and diurnal variations, and this study doesn't bring new light on these features. In contrast, no clear model for the stochastic component of ZTD time series is established yet. A proper modelling of ZTD noise characteristics is however known to be of special importance when subsequent parameters are estimated from the time series such as secular trends. This work is thus timely since long GPS ZTD time series start being used for climate applications. To my opinion, the relevance of this study is to analyse a global GPS network for nearly the full GPS history (1995-2015). However, in its present form, this work suffers several major methodological deficiencies which make the results highly uncertain and incomplete. The manuscript also lacks clarity and thorough interpretation of the results. A major revision and substantial rewriting are necessary. General comments for the revision are given below while an annotated manuscript is provided with specific comments.

Major comments

1. Inadequate homogenization method

I think there are several problems with the method adopted to detect and correct the offsets in the ZTD time series. First, it is assumed that discontinuities (offsets) detected in coordinate time series are also valid for ZTD time series. Though equipment and processing changes can induce discontinuities in both coordinates and ZTD series, it is not established that they impact them simultaneously in a systematic way. Moreover, the authors also included earthquakes as sources of discontinuity for ZTD. Though this is valid for position, earthquakes have no impact on ZTD estimates unless the station coordinates are tightly constrained during the processing, but this is not the case here. For a proper and unambiguous detection offsets in the ZTD series, it is necessary to analyse the ZTD time series, or more efficiently ZTD difference time series, using e.g. nearby stations or a NWP model as a reference (Vey et al., 2009; Ning et al., 2016).

Second, the estimation of a series of offset parameters simultaneously with an overall trend parameter (eq. 2) is an ill-posed problem. This is easy to check from the parameter covariance matrix of the deterministic model. Moreover, the estimation of offsets from the absolute ZTD series and not from differences is very likely to amplify the instability and produce unrealistic offset values and trend estimates. The presence of offset values as large as -69 or +50 mm (given in Table S1) or significant changes in the trend estimates when the noise model is changed (as reported in Table 2) confirms this deficiency.

Both major deficiencies need to be corrected before the trend estimates can be given any geophysical credit and noise characteristics can be determined accurately. Note that from their benchmarking of homogenization algorithms for monthly climate data, Venema et al. (2012) concluded on the superiority of relative homogenization algorithms over absolute (without using neighboring stations). Worse, the absolute homogenization algorithm they tested decreased the homogeneity of their simulated data.

Venema et al., Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89-115, doi:10.5194/cp-8-89-2012, 2012.

Thank you very much for this valuable comment. Indeed, the values of trends changed much when we switched the noise model from WH into AR(1)+WH.

For now, we analysed the ZWD series instead of ZTD data. We performed new homogenisation for ZWD data. We employed the epochs we found in GPS position time series, excluded earthquakes, as suggested and estimated the amplitudes of offsets with a least-squares estimation. We validated all epochs manually by checking the amplitude of offset and its significance in terms of estimated error and also by checking if the offsets which are close to each other do not introduce any numerical bias. Having validated all offsets manually, we reduced the number of offsets from 530 offsets which were reported in a set of 120 stations we analysed into 333 offsets. On the basis on these epochs, we corrected the ZWD series and these data were examined further.

As you can see in the newest version of the paper, we do not implement offsets together with deterministic model anymore. The only series we analyse are the ones which were homogenised (corrected by offsets).

2. Incomplete deterministic and stochastic models

The authors decide to include 4 periodic components for the seasonal cycle as well as 2 components for the diurnal cycle. This choice is quite arbitrary and poorly justified. Contrary to past studies which focussed only on the interpretation of the periodic parameters or on the overall trends, it is especially important here to include all relevant periodic components that might mix with the noise models. Inspection of the PSD plots provided in the manuscript and in the supplement suggests that higher diurnal harmonics should be included. A more rigorous spectral analysis should be presented to derive the general deterministic model. I would also like to see the RMS residual as an indicator of the amount of signal (or noise) remaining in the series after the deterministic model is fitted. The same remark holds for a preliminary choice of stochastic models. Inspection of the PSD plots would help pre-selecting the proper classes of noise models. For example, one might argue that, contrary to the coordinate time series PSDs, the ZTD time series do not resemble a power law since the spectrum is nearly flat at both high and low frequency ends. On the other hand, although the authors introduce various general stochastic models, such as white noise, power law and ARFIMA, they don't explore sufficient options and combinations from these models. Though the power law model could be excluded a priori, a more exhaustive number of combinations of ARFIMA(p,d,q) and white noise models should be tested. The decision of where to stop can be inferred from the inspection of AIC, or BIC, or another criterion. In regard to the objective of determining the optimal noise model, the present work considered too few variants to achieve this goal.

In the corrected version of our manuscript, we still consider 4 harmonics of annual signal: f=365.25, 182.62, 121.75 and 91.31 days. However, now, the choice of the harmonics is commented. First and second harmonics of annual frequency were included in the deterministic model for the entire set of stations, as they were proved to be significant for all stations using the stacked PSD. The periods of 121.75 and 91.31 days were assumed only for those stations, for which their amplitudes were higher than the errors of their estimates. The diurnal and semi-diurnal frequencies were assumed for the entire set of analysed stations, following previous studies by e.g. Jin et al. (2009). Other harmonics up to 12th were also examined, but they were not significant. The amplitudes of these frequencies ranged between 0.05 and 0.3 mm, whereas the standard deviation of noise is equal to approx. 30 mm. Having taken this fact into consideration, we decided not to model higher harmonics of a day. The peaks you might have seen on the individual PSDs we included in the previous version of our manuscript were the result of the length of the window we employed to compute the Welch periodogram. Having modelled the exact frequencies we have clearly seen at the spectrum, the power of the peaks did not decrease, which confirmed that they were artificially computed. So as not to misled the reader with those peaks, we removed the individual PSD and replaced it with stacked PSD. We also added a comment on that.

We estimated the RMS values of the differences between ZWD values and the model we employed (we refer to these differences as residuals). The values of RMSs of residuals differ between 1.76 and 50.57 mm, depending on the station. The higher the annual signal is, the larger is the RMS of the difference. The largest RMS computed for residuals was found for TSKB station, for which the amplitude of annual signal was equal to 122 mm. Naturally enough, the smallest RMSs were found for highest and lowest latitude stations, especially those with lowest amplitudes of annual signal. Other harmonics of annual signal and daily and sub-daily period are not so important when RMS value of residuals is being considered, as their amplitudes are relatively small comparing to the annual frequency. A comment about the RMS values was added to the text.

For now, we have only focused on two noise models: WH only and a combination of AR(1)+WH, which we prove to be preferred over a pure white noise assumption to describe the stochastic properties of ZWD data. We characterize the stochastic part of ZWD data as AR(1)+WH and show that adding a simple AR(1) to a pure white noise provides much more reliable estimates of a trend and its error than a simple white noise itself. We comment that employing a more sophisticated noise model and going further into higher orders of autoregressive process until X will cause that your model starts to be dependent from X previous values. However, higher orders of AR process provide the errors of ZWD trends which only differ of about 0.05 mm/yr from the estimates provided with AR(1)+WH.
We characterize the AR(1)+WH noise model by a fraction of AR(1) which means how much AR(1) contributes into the employed noise model. We estimate the ratio of uncertainties delivered with WH and AR(1)+WH noise models. We show that the trend error might be underestimated by a factor of 14 at maximum. We end with a recommendation that at least a simple AR(1)+WH noise model have to be assumed to obtain more reliable uncertainties for ZWD trends.

3. Relevance and interpretation of results.
Section 3.1 present results about seasonal and diurnal cycle of ZTD. Though these are significant components in the temporal variation of ZTD series, the results presented in this study don't add new information compared to past studies using ZTD, ZWD, or IWV (e.g. several publications by Jin et al., Bock et al., Nilsson and Elgered, or Ning et al.). I suggest removing this part from the manuscript but mention the main results in regard to past studies. An option would be to add the fitted parameters of the seasonal and diurnal cycle in the form a text file or a Table in the supplement.

Indeed, the seasonal cycles of ZWD and ZTD have already been described in numerous studies. Accordingly, the part about seasonal signals has now been shortened. We focus more on the noise analysis than we did in the previous version of the paper.

Using supplemental data is a valuable means for offering the interested reader the possibility to cross-check results and performing complementary analyses of the data. However, the discussion of the main manuscript results should not rely on figures and tables in the supplement. Figures necessary to the discussion (e.g. Figure S1 or a stacked PSD) should be presented in the main text.

All necessary figures and those cited in the text were moved into the main part of the paper.

Section 3.2 should be reorganised and follow a more logical and sequential description of results. First, the performance of a comprehensive set of different noise models should be compared on the basis of BIC and/or other model selection criteria. Global results should be presented (mean over all sites) as well as some typical examples (Table 2). Then, the trend estimates and their uncertainties can be analysed and results from different noise models inter-compared. I think it is necessary to add statistical significance tests on the estimated parameters (trends, offsets, stochastic model parameters). This would help assessing the relevance of the parameters, especially for the trends, and give an additional criterion for selecting the order of noise models. It would also help clarifying the discussion and interpretation.

For now, the global results are being discussed in the paper. We focus more on the stochastic properties of the applied noise model rather than deterministic part itself. Section 3.2. was re-written.

The paper is lacking an a priori discussion of the origin and nature of noise in the ZTD series. I think at least measurement and processing errors as well as atmospheric variability should be mentioned. As for the nature, it would be interesting to make some assumptions based on the cited literature, e.g. measurement as white noise, atmospheric variability as AR(1), and discuss them later when interpreting the results.

A discussion on the origin and nature of noise in the ZWD time series was added. We mentioned the possible sources of noise, and also gave the references. Basing on that, the noise model we recommended was well-justified.

On the question of whether analysing ZTD, ZWD, or IWV time series, it seems to me that this work could stay at the ZTD level since the main goal is describing a methodology for determining an optimal noise model in ZTD time series and assessing the impact on trend estimates and uncertainties. Using ZWD or IWV data would certainly allow more insight into the water vapour origin of variability, but this would pose the problem of using accurate ZHD and Tm data in order not to introduce spurious trends and noise. Though the main conclusions from the ZTD analysis would very likely also apply to ZWD, this cannot be affirmed without being verified. If only ZTD data are analysed, the authors should thus be cautious in their interpretations. The results cannot be interpreted unambiguously in terms of water vapour variability related to atmospheric processes.

In the revised version of the manuscript, we focus on the ZWD series, as was suggested by Reviewer #1. We agree that the ZWD series tell us more about the water vapour content of the atmosphere than the ZTD. According to our previous studies, ZWD do show similar stochastic characteristics to ZTD.

Specific to this work is the presentation of results in climate zones following the Köppen-Geiger classification. Regarding the annual and diurnal components, it seems to me that this classification is not relevant (no coherence emerges in the maps or latitude plots). Maybe, this is due to an excessive reduction to only 5 classes? Did the authors consider using the full classification proposed by Peel et al., 2007?

The classification into different climate zones was removed from the paper, as it was stated not to be appealing.

4. Lack of clarity of presentation

In general I found the manuscript rather difficult to read for the following reasons. The model description uses inconsistent notations among equations (2) to (12). All equations should be introduced in the methodology section and not later, e.g. in the results sections (eq. 10 to 13). Some equations are probably not useful (e.g. 3, 7, 8, 10, 11, 12, 13) and some are incomplete (eq. 5, 9).

In the revised version of the paper, the notation is consistent throughout the paper. We moved equations from Results section into Methodology and deleted those, which were found to be unnecessary. Thank you very much for this comment. The changes we made will hopefully help all readers to get the point of our paper.

A lot of descriptions of properties of GPS time series and results from past studies concern coordinate time series and are not relevant to ZTD time series (in the introduction, P2L21-25, P3L6-914, in the homogenization section, P5). Conclusions relevant only to position time series should generally not be included. I also noticed that many of the references to past studies of station position series are improperly cited when discussing properties of ZTD series (e.g. P2L10, P3L11, P5L3) or erroneous results/values are cited (e.g. from Jin et al., 2007, 2008, or from Bock et al., 2014).

The references concerning GPS position time series were deleted. We also re-wrote the part describing the position noise in the Introduction. We checked the values from Jin et al. and Bock et al. and changed them into the proper ones.

The same topics/discussions are repeated in several places (e.g. descriptions of GPS processing, homogenization, climate zone classification in the introduction and in the methodology and results sections; results from past studies on noise models P6, P9 and P12) or are interrupted and restarted (mainly the comparison of results for different noise models P10).

We re-wrote the entire text, hopefully making it more consistent and clear. Now, the Introduction part is followed by data description, methodology including all equations we used throughout the text, results we obtained and discussion of our findings.

Improper wording is used in some places (e.g. when describing atmospheric propagation in the introduction and GPS processing in section 2).

The revised version of our manuscript was carefully checked, corrected and re-written. We hope that this would provide the clear message we try to send.

More specific comments on all these points are given in the annotated PDF.

Thank you very much. All comments included in the annotated PDF were put into the text.