Review of egusphere-2025-6473

While reading the paper for review I ran across the description of the calculation of the Confidence Intervals of the drifts in the ground-based/satellite differences (lines 178-195 and in particular lines 192-195).

> “The drift is estimated as the median of the slopes all the lines connecting possible pairs of points”.

This is the basis of the Theil-Sen/Kendall-based method, as referred to in line 189.

> “The confidence interval for the drift-estimate is given by the interval containing the middle 95% (to obtain an estimate of the 2σ error).”

If I have my statistics right, what is described above is incorrect (i.e. using all pair-point-slopes for the 95% CI), this is not the way to calculate the CI’s in the Theil-Sen method. Details are outlined below, the Theil-Sen 95% CI’s are much narrower than the values reported here, a quick back-of-the-envelope calculation suggests easily an order of magnitude.

Which brings me to the second issue: the Theil-Sen method does not give accurate CI’s if the dataset has significant autocorrelation. From Figure 2 this appears to be the case. If so, another method or a modification of Theil-Sen should be used to derive CI’s.

There are several methods for calculation of CI’s, more on that also below. I often use a bootstrap approach because of its methodological simplicity and because I find it rather intuitive to understand (uncertainty is in the data so use the data for deriving uncertainty). But that is just a personal preference.

As a consequence, there is little added value in continuing the review before these issues are addressed. Once addressed and incorporated in a revision, I will review the paper

Hence why I label this review as “major revision”.

Note that I do not necessarily expect results to materially change as there are two aspects here that work in two directions with regard to the CI’s: the appropriate Theil-Sen CI is much narrower but autocorrelation will widen CI’s.

Detailed comment

The following should be done:

> **[1]** check if indeed an inappropriate/incorrect method is used to calculate the Theil-Sen CI's (for example check the original papers). As stated, I think I have my statistics correct but that is no guarantee either.
>
> **[2]** check if the data has autocorrelation (lag-1) larger than 0.2
>
> **[3A]** if both are the case, select method that accounts for autocorrelation, for example:
>
> - block bootstrap of the Theil–Sen slope
> - GLS/ARIMA trend estimation
> - state-space trend models (e.g. Kalman filter)
> - Generalized least squares (GLS) with autocorrelation structure
> - LOESS + derivative uncertainty
>
> **[3B]** if **[1]** is correct but **[2]** is not, apply the correct Theil-Sen CI calculation
>
> **[3C]** if **[1]** is not the case but **[2]** is, select a different method that accounts for autocorrelation

After that, the results should be rechecked and if needed adjustments to the paper should be made.

The Theil-Sen CI will likely give a much smaller CI – order of magnitude, even more - but if there is autocorrelation the Theil-Sen CI is inaccurate as well. Applying an approach that accounts for autocorrelation may widen the CI value again, possible even much wider.

This is why I noted that in the end this all may not materially affect the results and findings of the paper, why it is important to get this done properly and why I propose a withdraw-and-resubmission.

Details about Theil-Sen & CI's

Using the middle 95% of all pairwise slopes (Theil-Sen approach) as a proxy for the 2-sigma error of the median slope is not theoretically justified and should only be viewed, at best, as a crude heuristic. It is thus not "wrong" but the better approach would be to use the appropriate Theil–Sen confidence intervals.-sigma error of the median slope is not theoretically justified and should only be viewed, at best, as a crude heuristic.

A bootstrap for the median slope and corresponding CI would also be a possibility but a regular bootstrap does not account for autocorrelation, for that a block-bootstrap would be a more appropriate approach.

1. what is done here:

- take all pairwise slopes between two time series (or one series vs time),
- compute the median slope (this is essentially the Theil–Sen estimator), and
- then take the middle 95% of all slopes as an "error band",

which means treating the empirical distribution of pairwise slopes as if:

1. each slope were an independent draw, and
2. the central 95% of that distribution corresponded to a 95% interval for the median slope.

Neither is really true.

2. Why the middle 95% of slopes is not an accurate 2σ proxy

- Strong dependence: Each slope uses two data points, so slopes share points and are heavily correlated. There are not $n(n-1)/2$ independent pieces of information.
- Wrong target: The distribution of all pairwise slopes is not the sampling distribution of the median slope. The median slope is a functional of that set, but its uncertainty is much smaller than the spread of all slopes.
- Coverage is not 95% for the median: In the Theil–Sen framework, the exact 95% confidence interval for the true slope is constructed by taking slopes between specific rank indices, derived from the binomial distribution of the median—not simply the 2.5% and 97.5% quantiles of all slopes.

So the "middle 95% of slopes" will generally not correspond to a 95% (≈2σ) interval for the median slope; it will usually be too wide.

3. The Theil-Sen Confidence Interval

Suppose there are n n data points. Then the number of pairwise slopes becomes:

$$N = \frac{n(n-1)}{2}$$

But these slopes are not independent. Meaning he 2.5% and 97.5% quantiles cannot be used as the 95% CI.

Sen (1968) showed that:

- The median slope is the (N+1)/2-th slope in the sorted list.
- A 95% CI corresponds to slopes ranked between:

$$L = \frac{N - z_{0.975}\sqrt{V}}{2} \text{ and } U = \frac{N + z_{0.975}\sqrt{V}}{2}$$

where V is the variance of Kendall's tau under the null.

This is better than "middle 95% of slopes" because:

- The slopes are strongly dependent
- The distribution of slopes is not the sampling distribution of the median slope
- The width of the slope cloud reflects noise + leverage + spacing, not estimator uncertainty
- The correct CI is much narrower than the middle 95% of slopes

Theil–Sen/Kendall gives a statistically valid interval with correct coverage, but only if the lag-1 autocorrelation of the time series is sufficiently small (using a threshold of < 0.2 – in absolute value - is common). Otherwise an approach should be selected that accounts for autocorrelation.