Citizen rain gauges improve hourly radar rainfall bias correction using a two-step Kalman filter - Review

Anonymous Referee

November 5, 2021

1 General comments

Thank you for this manuscript.

Using citizen rain gauge data is an interesting approach and the authors analyzed the gathered data carefully and with a lot of different techniques. There seems to be a benefit in using these Citizen rain gauge data, so from my point of view, this study should be published.

I had some difficulties to understand where all the errors are coming from and how these errors are calculated. There are often some unreferenced assumptions on the error characteristics. Sometimes I would have wished to have a bit more insight into the error distributions of the different measurements.

There is also nothing mentioned on radar calibration errors. Most weather radars are not well calibrated and this obviously influences the estimation of R via the $Z - R$ relationship. Is this measurement bias automatically canceled out by calculating an individual $Z - R$ relationship?

Regarding the KF and its mathematical formulation: I guess that this has been done properly, at least from my understanding. However, Kalman filters are especially great for multi-dimensional data sets with observation vectors being larger than 1. So I'm therefore not sure if averaging everything together and calculating one bias value is the best approach. But at least it is an approach that shows some nice results.

There are still quite some language issues and after a couple of page I got tired to correct them all. So there should be some polishing done on that.

Overall I think that the manuscript can be published after some corrections, which are detailed below.

2 Detailed comments

- Line 12: *of Sattahip radar station and gauge rainfall* → of Sattahip radar station, gauge rainfall...
- Line 27: *accuracy of flash flood estimates and warning.* → accuracy of flash flood estimates and warnings.
- Line 28: *However, weather radar provides indirect measurement of backscattered electromagnetic waves called radar reflectivity data (Z).* $rightarrow$ However, a weather radar provides an indirect measurement of backscattered electromagnetic waves called radar reflectivity ~~data~~ (Z).
- Line 30: *ground-truthing by rain gauge data is required to calibrate the Z-R relationship (*$Z = AR^b$*)* → I'm not happy with this sentence. It's not a priori clear that just the Z-R relation needs to be calibrated. Biases between the radar rainfall intensity can also stem from radar calibration errors, i.e., from a bias in Z. Most radar operators do not alter the Z-R relationship at all, but just adjust the radar rainfall vs gauge rainfall bias (Sideris, 2014).
- Line 60: *variances affecting the mean field bias estimate.*
- Line 76: *However, these methods are not usually designed for real-time* → However, these methods are usually not designed for real-time
- Figure 1: I can't see the location of the radar in this map. It's indicated in the legend but not on the map (or hardly visible).

- Line 107: *The Tubma basin is covered within the range of Sattahip radar station.* → The Tubma basin is located within the coverage of the? Sattahip radar station.
- Line 107: *a beam width of 1.0°* → a half power beam width of 1.0°
- Line 109: *The radar reflectivity product is in a Cartesian* → The radar reflectivity product is provided in a Cartesian
- Line 110: *The Sattahip radar provides the CAPPI reflectivity data derived from the 2.5-km constant altitude plan position indicator (CAPPI).* → CAPPI derived from CAPPI: that's not a meaningful sentence.
- Line 114: *Additionally, the noise and hail effects were eliminated by setting reflectivity values below 15 dBZ to zero, and reflectivity values greater than 53 dBZ to 53 dBZ.* → I'm not sure if this is allowed. Could you give some reasoning on this method or a reference? If I use a Marhsall-Palmer relation ($Z = 200\,R^{1.6}$ then 100 mm of rain already give 55 dBZ. 100 mm are not unrealistic in the tropics, even higher values are possible. But probably Marshall-Palmer is not suitable in the tropics? And clipping below 15 dBZ might be ok, but reflectivity values below this value or not necessarily due to noise, especially not receiver noise. It might need a specification of what you mean by 'noise'.
- Line 137: *radar rainfall accumulation (mm/h)* → Accumulation is not in mm/h
- Line 141: *was validated against a second, independent dataset. Results found that a locally calibrated Z-R relationship that was used in this* → I don't understand this sentence. What is the $Z - R$ relation of the climatological dataset? What is the $Z - R$ relation of the locally calibrated dataset?
- Line 146: *These 15-min rain gauges*: → These rain gauge data have a temporal resolution of 15 minutes
- Line 148: *double mass curves method*: → What is this?
- Line 155: *based on spatial decorrelation analysis in the process.* → based on spatial decorrelation analysis for this? process.
- Line 180: RMSE does not tell you much about the bias between the TMD and the citizen gauge. Bias needs to be given as well.
- Line 188: *First, daily citizen...* → From line 126 onward you were using 'Firstly, Secondly...'.
- Line 206: I'm not sure if 'noise' is the right word here. KF accounts for measurement errors or uncertainties, but not for noise.
- Line 208: *different uncertainty characteristics, i.e., hourly...*
- Line 210: *Since the MFB (G/R ratio) is assumed to follow a log-normal distribution.* → Has this sentence a relationship to the previous sentence? And what is an ordinary KF scheme
- Line 300: *downscaled hourly citizen rain gauge data were used to back-calculate the hourly citizen rain gauges data* → I don't understand this.
- Figure 5: This is Kalman filter bias corrected rain accumulation data from gauges?
- Line 420: *%-ile* → percentile
- Line 508: *obviously appear steady light rainfall accumulation* → ?
- Figure 9: *Rainfall Depth* → probably not the right annotation.

3 Kalman filter

I guess I have understood your KF approach, but there are some questions left:

- y_t and z_t are your log-transformed bias observations from the TMD and the citizen gauge network, respectively, corresponding to O_t in Eq. 10?
- Figure 4: In the second KF step you take the variance estimates P_t from the previous KF step as a priori variances?
- You calculate the Kalman gain K_t in the second step with these P_t variances. From my point of view, this Kalman gain should be named differently, since K_t is already used for the first step.
- Likewise, the equation (2) in the 2nd KF step in Figure 4 should be rewritten, since the value on the left side of the equation ($\hat{\beta}_t$) is an update of the bias value on the right side.
- Same is true for Eq. 3 in Figure 4, 2nd step: P_t is an update of P_t on the right side of the equation.
- I'm somehow irritated by the usage of these capital letters K and P, which, at least for me, would represent matrices. In KF theory, P then represents the error covariance matrix, but since you are using variances only it is somehow strange to go from $\sigma^2_{M_{z,t}}$ to P_t. Probably just an unimportant detail.
- Line 229 onwards: Error estimation of $\sigma^2_{M_{z,t}}$ and $\sigma^2_{M_{y,t}}$: I did not fully understand this. $\sigma^2_{O_t}$ is the (spatial?) variance of all stations? And why is Equation 19 necessary if you calculate $\sigma^2_{M_{z,t}}$ and $\sigma^2_{M_{y,t}}$ by the spatial variance of the individual stations?