# Detailed response to Davide Zoccatelli's comment

Joost Buitink, Remko Uijlenhoet, and Adriaan J. Teuling

## 1 Density Weighted Distance

One the main comments by Davide Zoccatelli was related to the use of the new metric that was introduced in the manuscript, the Density Weighted Distance (DWD). Davide Zoccatelli made several valid comments, which we will address below.

- Davide Zoccatelli correctly mentioned that high DWD values can mean two things: either that the low_res model result is outside of the range of values simulated with the high_res model, or that the high_res model has high internal variability. This is indeed a property of this method, which we will explicitly mention when we introduce this method. If the low_res model results is different than the entire range of values from the high_res model, we recommend not to use DWD since this is not the intended situation to apply DWD. To prevent this, all conclusions related to DWD values are not solely based on the DWD results, but are interpreted in combination with the violin plots, which proved the necessary additional information. We agree that this is currently not clearly presented as such in the manuscript, and we will make sure this is clarified in the new version of the manuscript.

- DWD can give high values, even if the low_res model is perfectly representing the accumulated high_res results. This intended behavior of the method, since we want to quantify how well the low_res model represents the entire high_res model results. However, we do agree that a comparison between the low_res model and the aggregated catchment response from the high_res model is currently missing from the manuscript. In Figure 1 we present a comparison between the aggregated high resolution model response and the response from the low resolution model. All pixels within the high resolution model are averaged and compared with the anomaly calculated for the low resolution model. Ideally, the low resolution model should match the aggregated high resolution model response. This figure shows that generally both models simulate the same trend, yet the order of magnitude of the anomaly does not always match. The presented average difference represents the mean absolute difference between the high and low resolution model results. This value shows that the resolution difference generally causes a bigger disagreement in the Alpine basins than in the pre-Alpine basins.

- Davide Zoccatelli correctly noted the the $P_{\text{low\_res}} - 0.05$ and $0.95 - P_{\text{low\_res}}$ can be interpreted as weights, and that the sum of those is different from 1. Therefore, we propose a new slight modification of the DWD, based on the following

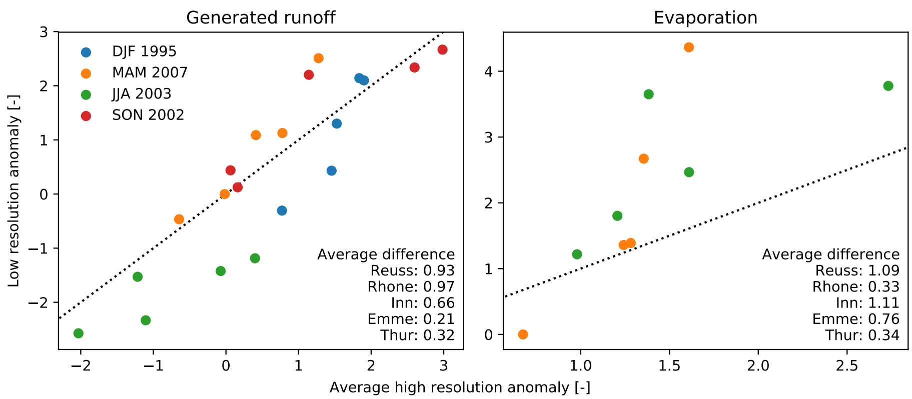


**Figure 1.** Comparison between the average high resolution model response and the low resolution model response, for the generated runoff. Colors indicate the different extreme seasons, and the dotted line represents the 1:1 line.

equations:

$$DWD = W_{\text{lower}} \cdot d_{\text{lower}} + W_{\text{lower}} \cdot d_{\text{lower}}, \tag{1}$$

$$W_{\text{lower}} = \max\left(0, \min\left(1, \frac{P_{\text{low\_res}} - P_{\text{lower}}}{P_{\text{upper}} - P_{\text{lower}}}\right)\right), \tag{2}$$

$$W_{\text{upper}} = \max\left(0, \min\left(1, \frac{P_{\text{upper}} - P_{\text{low\_res}}}{P_{\text{upper}} - P_{\text{lower}}}\right)\right), \tag{3}$$

$$d_{\text{lower}} = Z_{\text{low\_res}} - Z^{5\%}_{\text{high\_res}}, \tag{4}$$

$$d_{\text{upper}} = Z^{95\%}_{\text{high\_res}} - Z_{\text{low\_res}}, \tag{5}$$

where the $W_{\text{lower}}$ and $W_{\text{upper}}$ replace the original $P_{\text{low}} - 0.05$ and $0.95 - P_{\text{low}}$ terms, respectively. Both weights are corrected between 0 and 1, which is only necessary when $P_{\text{low\_res}}$ is outside the $P_{\text{lower}} - P_{\text{upper}}$ range. This has some minor implications for the DWD results, since the distances are now weighted with a total weight of 1, instead of 0.9 in the previous version. This has the effect that the DWD values are slightly higher. The absolute DWD values will change, but it has no implications for the overall conclusions, see Figure 2 and Table 1. These results are a function of the chosen threshold, yet we would recommend to choose $P_{\text{upper}}$ and $P_{\text{lower}}$ to include most of the high resolution data, since every pixel can be considered equally important. We are happy to learn if this modified definition solves the original concerns raised.

## 2 Hydrological model validation

– Davide Zoccatelli also questioned the validation of the model. Originally, our validation was only shown in Figure 5b (original manuscript). We think this might have easily been overlooked. Therefore, we have updated our calibration/validation figure to better represent the validation of the model, and added KGE values as model validation. In

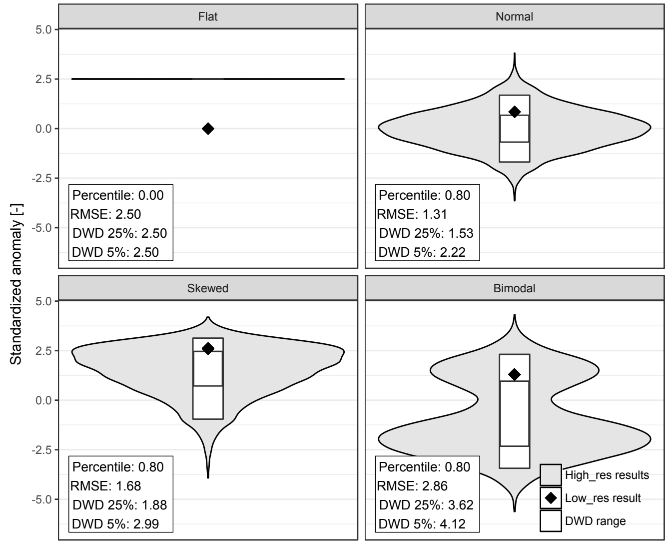


**Figure 2.** DWD examples with the updated DWD equations.

**Table 1.** Scale mismatch between the high and low resolution models as measured by DWD, for both hydrological fluxes during the four extreme seasons (based on the updated DWD equations).

| Basin | **Total generated runoff** | | | | **Actual ET** | |
|---|---|---|---|---|---|---|
| | DJF 1995 | MAM 2007 | JJA 2003 | SON 2002 | MAM 2007 | JJA 2003 |
| Reuss | 2.63 | 2.31 | 4.81 | 2.58 | 2.44 | 3.70 |
| Rhone | 2.86 | 2.82 | 4.29 | 1.88 | 0.85 | 1.52 |
| Inn | 2.40 | 1.82 | 4.56 | 1.84 | 4.36 | 2.19 |
| Emme | 1.25 | 0.33 | 0.79 | 1.01 | 0.46 | 5.96 |
| Thur | 0.97 | 0.73 | 1.48 | 0.66 | 0.62 | 4.33 |

original figure, calibration and validation data for the low resolution model was missing, so we added this to the new version of this graph, see Figure 3. We hope this figure gives a better overview of the model performance.

– Davide Zoccatelli also noted that there is no internal validation of the model. To get an idea of the quality of the internal flux representation, we compared the simulated fluxes with observed data from the research catchment Rietholzbach, situated within the Thur basin (Seneviratne et al., 2012). We obtained the evaporation data as measured with a long-term research lysimeter, and discharge data from this catchment. Both discharge and evaporation from the corresponding pixel were extracted from SPHY, to compare with the observations. We calculated the anomalies over the entire simulation period. The comparison between the observed and simulated anomalies can be found in Table 2. This table shows that the simulated anomalies agree well with the direction and magnitude of the observed anomalies. Winter and autumn

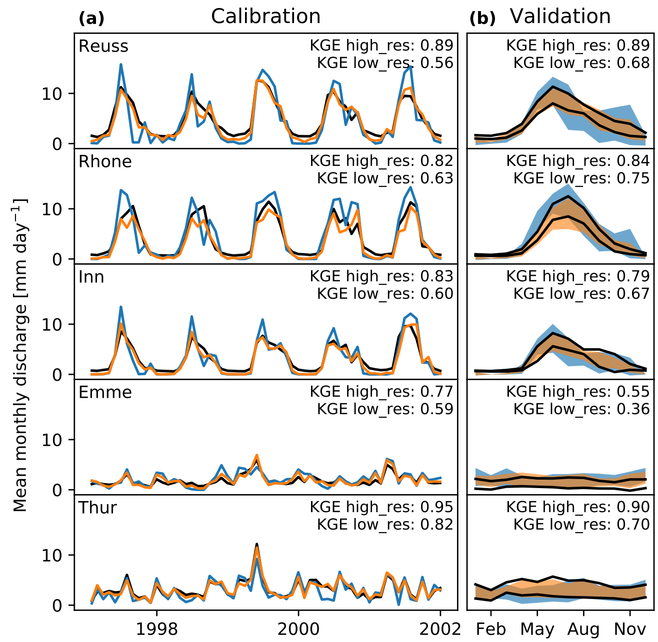


**Figure 3.** Calibration and validation of SHPY. The model is validated over the entire period, excluding the calibration period. The two black lines, and the shaded areas present the average discharge plus/minus the standard deviation. For Inn, all data after 2003 was excluded, since the observed discharge pattern changed after this period.

values for evaporation are gray, since they are not the focus of this study due to the fact that SPHY does not allow for evaporation during snow covered periods. We believe that this table is a valuable addition to the validation of the model.

**Table 2.** Comparison between anomalies simulated with SPHY and observed in the Rietholzbach, anomalies are based of the entire simulation period.

| Event | **Runoff** | | **Evaporation** | |
|---|---|---|---|---|
| | Observed | Simulated | Observed | Simulated |
| DJF 1995 | 1.68 | 2.34 | 0.86 | 0.53 |
| MAM 2007 | -0.52 | -0.27 | 1.61 | 0.98 |
| JJA 2003 | -2.15 | -2.17 | 1.66 | 3.61 |
| SON 2002 | 2.65 | 2.62 | -1.76 | -0.26 |

There is a slight mismatch between the evaporation anomalies during the summer of 2003, yet both describe unusually high values. To further investigate this, we also plotted the evaporation time series in Figure 4. This figure shows that

SPHY is able to accurately simulate the evaporation, yet there are some differences between the two time series. These can be attributed to the scale difference between the lysimeter and a single high resolution SPHY pixel, and the fact that SPHY does not account for all factors influencing evaporation since it uses the temperature-based Hargreaves method. The difference between the two time series influences the mean and standard deviation, and therefore the resulting anomaly values.

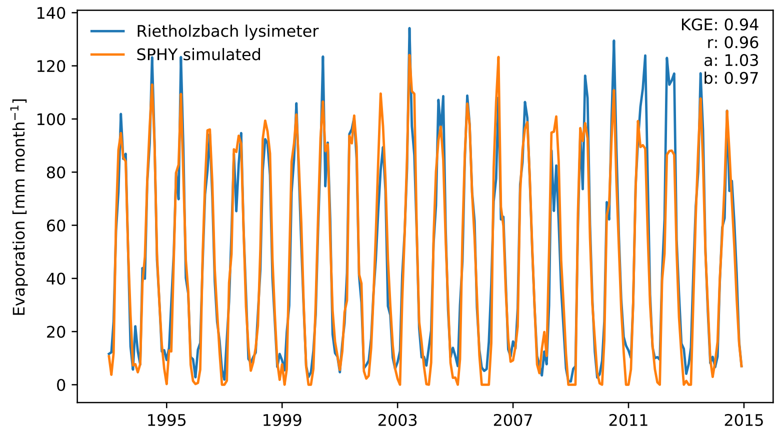


**Figure 4.** Comparison between observed evaporation in a lysimeter, and the simulated evaporation of the corresponding pixel.

## 3 Comments and other changes

- We will add average yearly precipitation to Table 1 in the original manuscript.

- We will add more references and background information to the manuscript regarding scaling issues in hydrology/hydrological modeling.

- SPHY assumes that each model pixel is independent, meaning that there is no communication between the individual pixels. This is a limitation of the model, however most (conceptual) hydrological models are programmed this way. We believe the impact on the results to be small.

- SPHY does include runoff propagation, where runoff is transported to the downstream cells using a recession coefficient. For the exact details for the routing conceptualization, we refer to the paper of Terink et al. (2015).

- We calibrated the model based on the sum of squares, since this is the more common approach to model optimization However, KGE is easier to interpret than the square of residuals.

## References

Seneviratne, S. I., Lehner, I., Gurtz, J., Teuling, A. J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schroff, K., Vitvar, T., and Zappa, M. (2012). Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resources Research*, 48(6):W06526.

Terink, W., Lutz, A. F., Simons, G. W. H., Immerzeel, W. W., and Droogers, P. (2015). SPHY v2.0: Spatial Processes in HYdrology. *Geosci. Model Dev.*, 8(7):2009–2034.