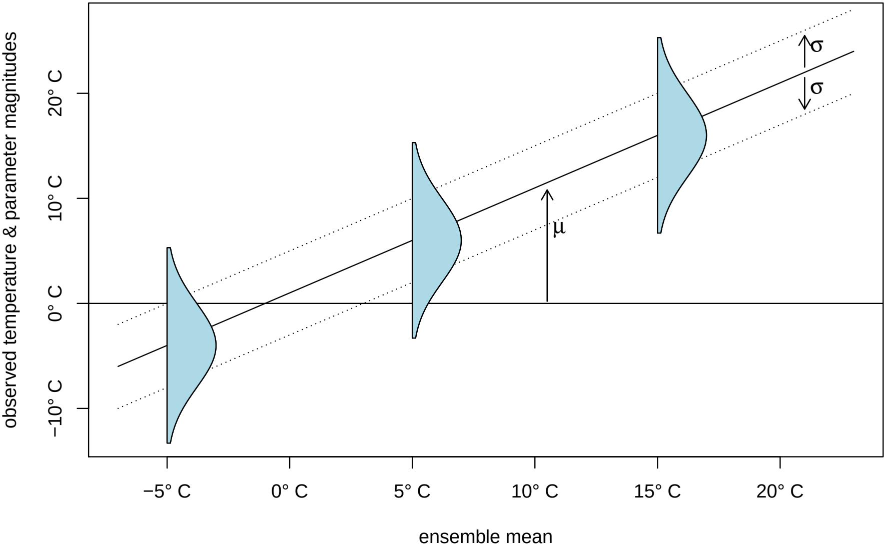
Multivariate Statistical Post-Processing of Ensemble Forecasts of Precipitation and Temperature

Michael Scheuerer

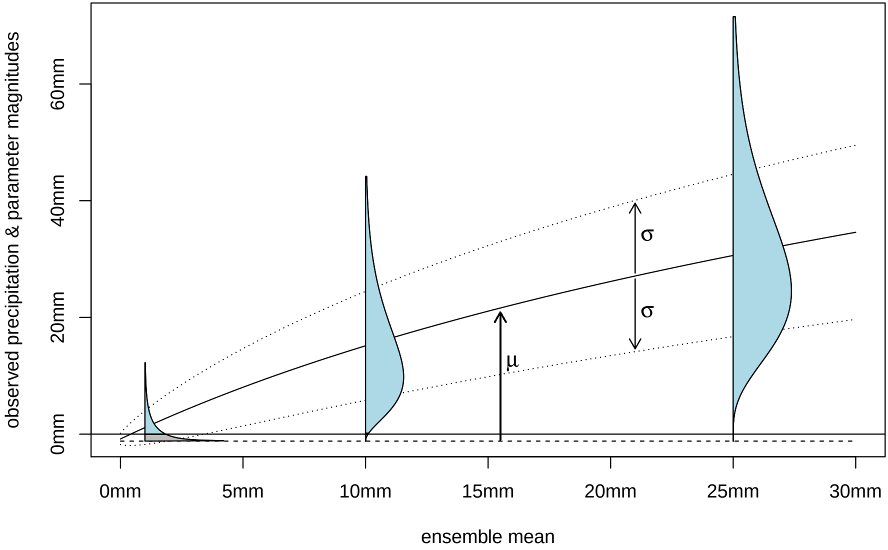
NOAA/ESRL, PSD

June 2016

Univariate post-processing: temperature



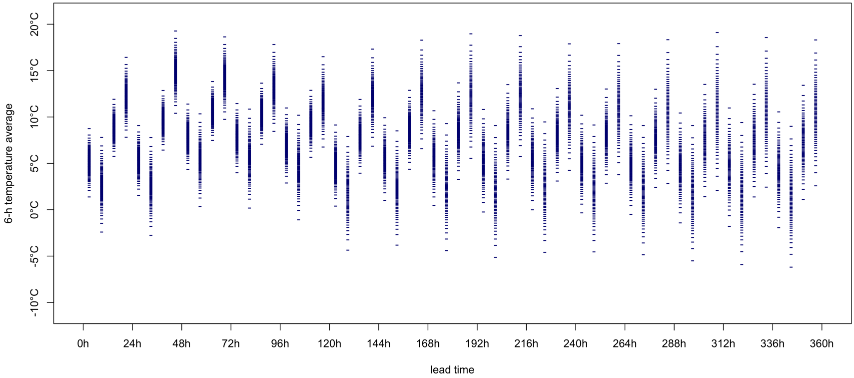
Univariate post-processing: precipitation accumulations



Serial dependence of temperature forecast trajectories

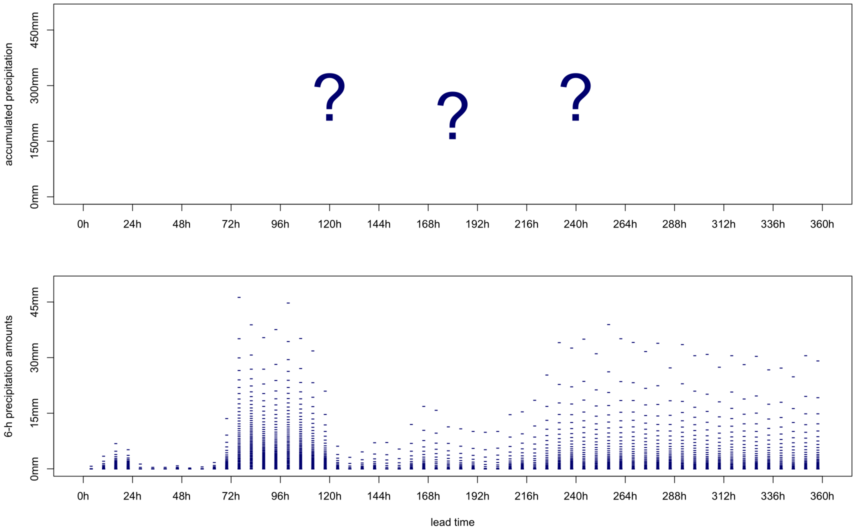
By calculating certain quantiles, the predictive forecast distributions can be turned back into an ensemble (of any desired size).

Univariate post-processing, however, does not provide any information about serial dependence, i.e. we don't know how to connect the ensemble forecasts at different lead times.



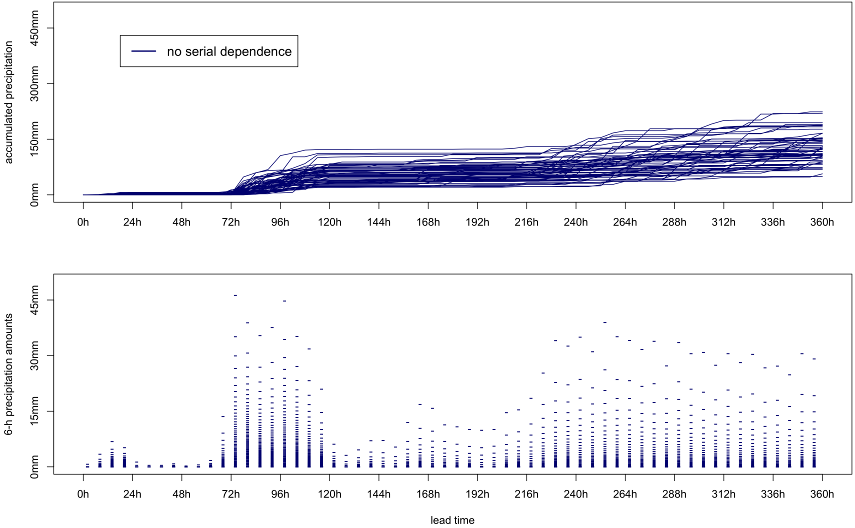
Serial dependence of precipitation forecast trajectories

That serial dependence information, however, is critical e.g. for calculating cumulative precipitation amounts over the forecast period.



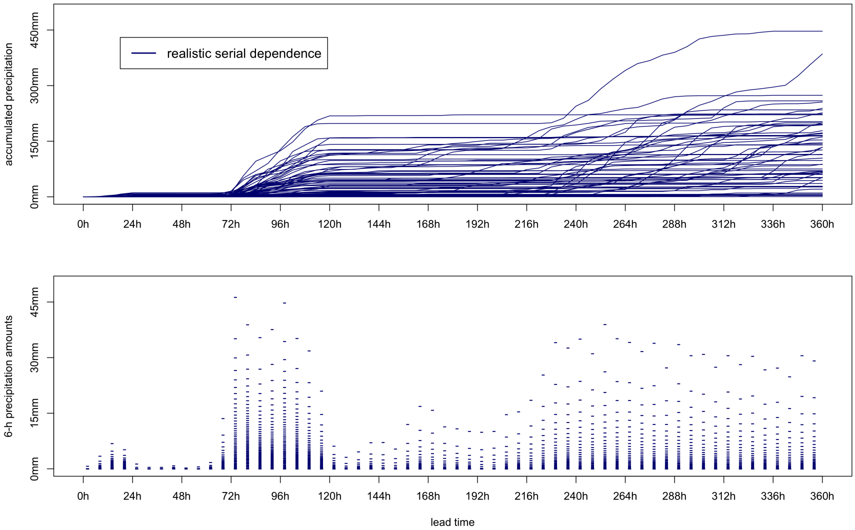
Serial dependence of precipitation forecast trajectories

Naively assuming that the individual 6-h accumulations are independent results in underdispersive predictions of longer-term accumulations.

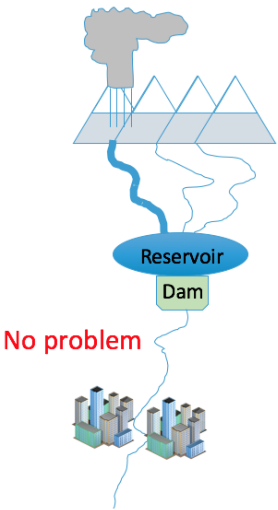


Serial dependence of precipitation forecast trajectories

The exact same marginal distributions in conjunction with a more realistic serial dependence structure yield much more dispersed predictions.



Spatial dependence of precipitation forecast trajectories



Hydrologists need to know not only the intensity of rainfall, but whether or not that intense rainfall is expected at several locations simultaneously.

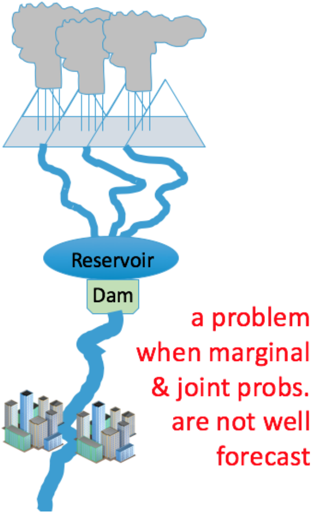
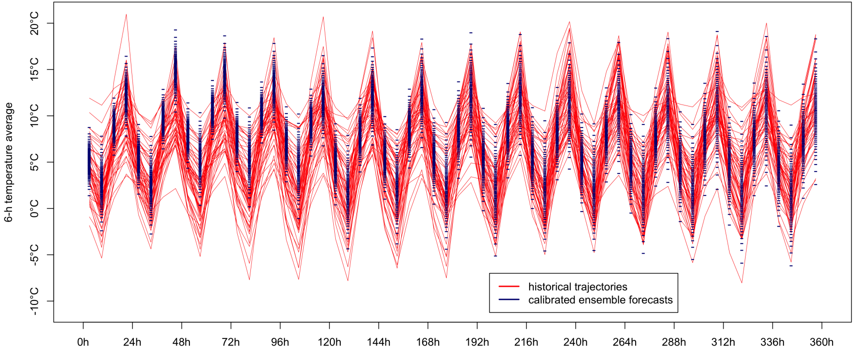


image courtesy: Tom Hamill

Modeling spatio-temporal dependence: the Schaake Shuffle

Idea of the Schaake Shuffle:

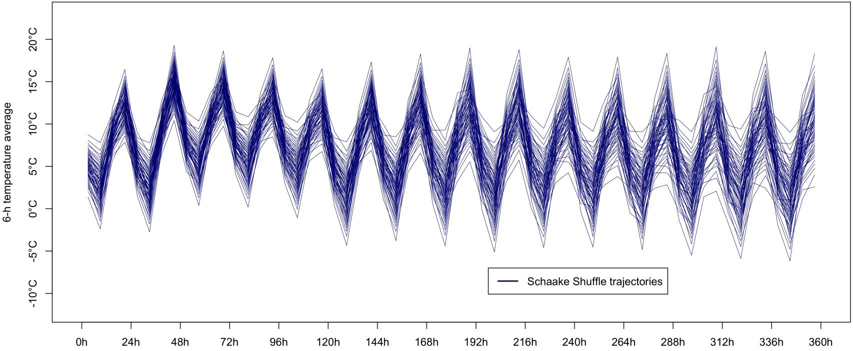
1. use historical observation trajectories as 'dependence template'
2. re-map the values of this historical ensemble to the values of the (univariate) forecast ensemble while retaining its rank order



Modeling spatio-temporal dependence: the Schaake Shuffle

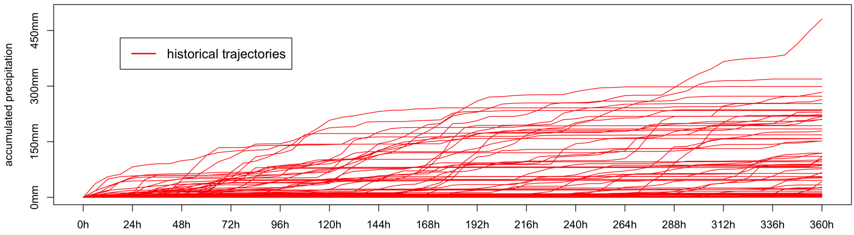
Idea of the Schaake Shuffle:

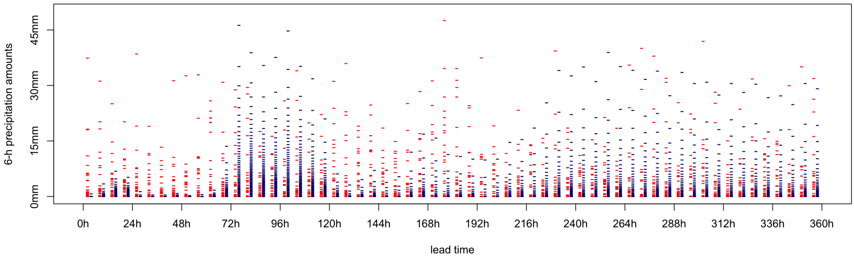
1. use historical observation trajectories as 'dependence template'
2. re-map the values of this historical ensemble to the values of the (univariate) forecast ensemble while retaining its rank order



Modeling spatio-temporal dependence: the Schaake Shuffle

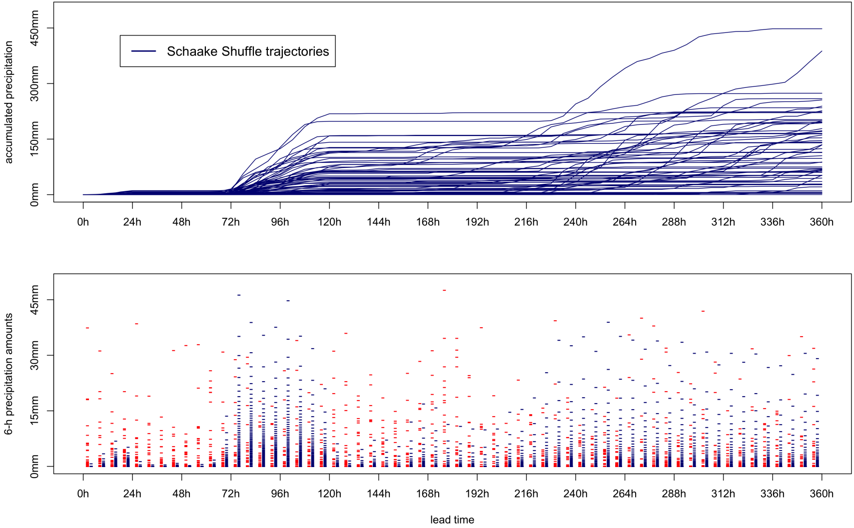
Spatio-temporal precipitation trajectories are obtained in the same way:





Modeling spatio-temporal dependence: the Schaake Shuffle

Spatio-temporal precipitation trajectories are obtained in the same way:



Main idea for improving the Schaake Shuffle

The Schaake Shuffle (Clark et al., 2004) has been used very successfully since it has been proposed. Yet, there are a few concerns:

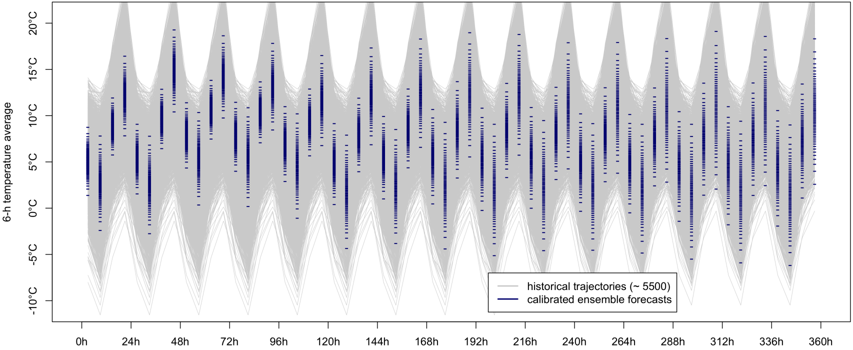
- the re-mapping of ensemble values can be quite substantial. Are the re-mapped trajectories still realistic?
- in particular: is it adequate to assume that spatial/ temporal correlation is state-independent, i.e. for example the same for low and high levels of precipitation?

To address these concerns we propose an algorithm that selects historical trajectories whose marginal distributions already resemble the distributions of the forecast ensemble, thus reducing the required amount of re-mapping.

Temperature and precipitation are considered simultaneously, but since precipitation is more important and more complex it will be treated in a more sophisticated way.

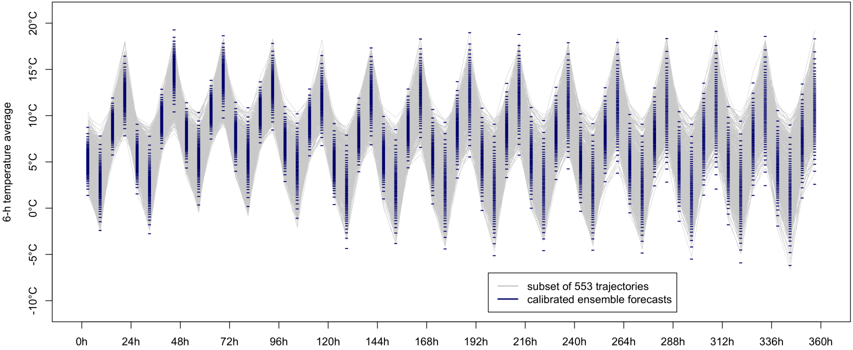
Improving the Schaake Shuffle: step 1

Starting from a much increased set of historical dates/trajectories, we apply a simple subsetting criterion based on temperatures and discard all trajectories with too many values outside the 99.9% predictions intervals of the forecast distributions.



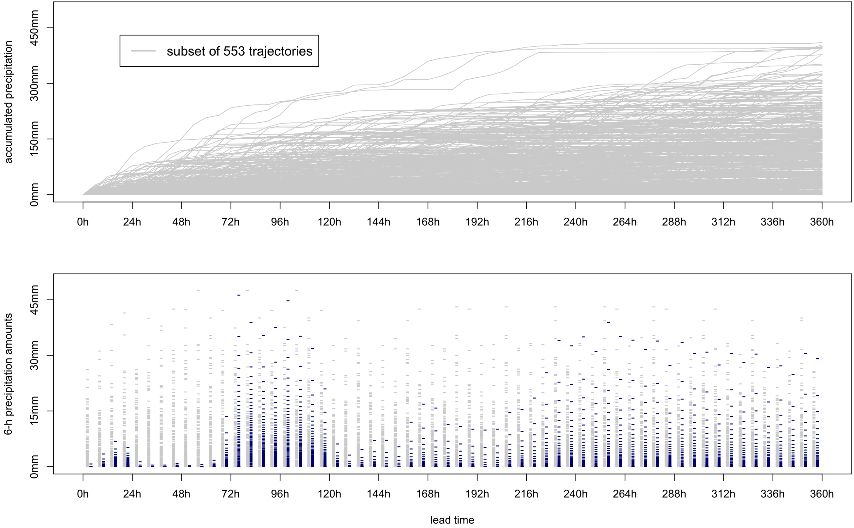
Improving the Schaake Shuffle: step 1

Starting from a much increased set of historical dates/trajectories, we apply a simple subsetting criterion based on temperatures and discard all trajectories with too many values outside the 99.9% predictions intervals of the forecast distributions.



Improving the Schaake Shuffle: step 2

Further reduction of the number of historical trajectories will be based on the similarity of the marginal distributions to those of the forecasts.



Improving the Schaake Shuffle: step 2

Similarity will be measured by the divergence

$$d(F_m, G_n) = \int \left(F_m(x) - G_n(x)\right)^2 dx$$

where F_m is the CDF of the m-member predictive ensemble and G_n is the CDF of the n historical trajectories at a particular lead time and location.

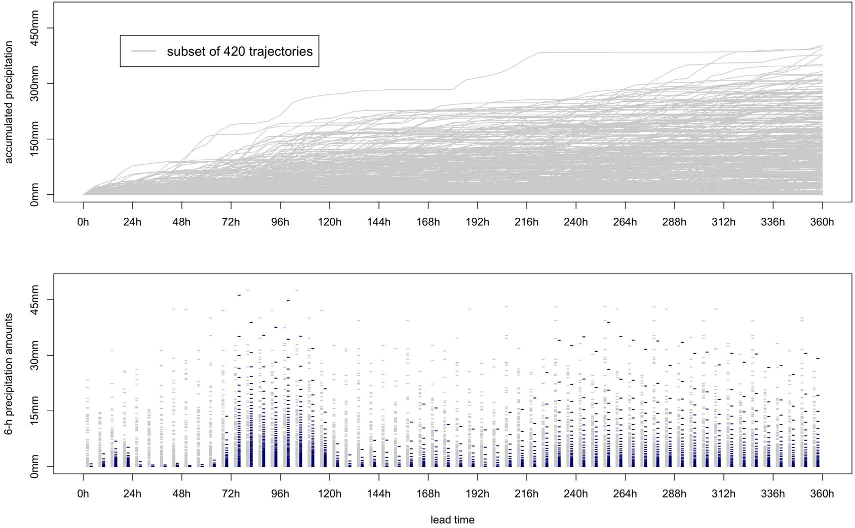
A subset of those n trajectories will be chosen iteratively based on the mean divergence over all lead times and all locations in the basin:

- for each trajectory i, we calculate the divergence $d(F_m, G_{n,-i})$ between F_m and the CDF $G_{n,-i}$ obtained after omitting that trajectory.
- trajectories whose omission reduces the divergence from F_m most will be discarded, all others retained for the next iteration

This process will be repeated until only m trajectories remain.

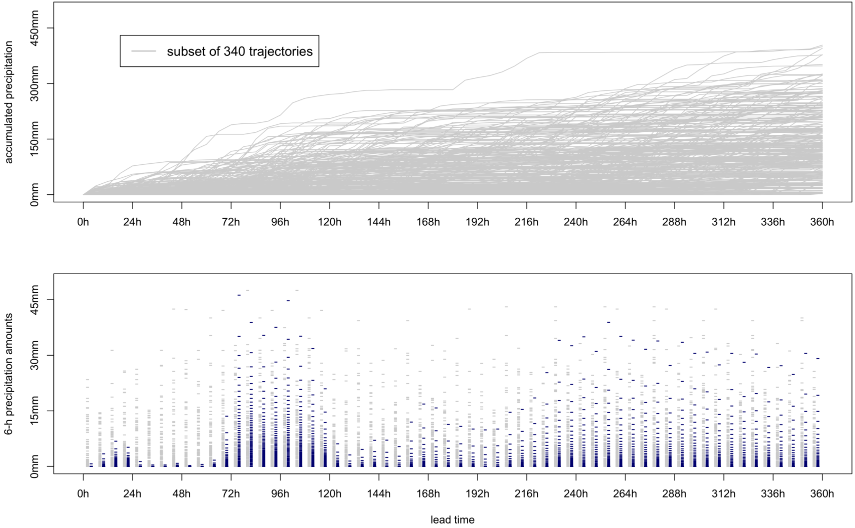
Improving the Schaake Shuffle: step 2

In the first iteration step the original 553 trajectories are reduced to 420:



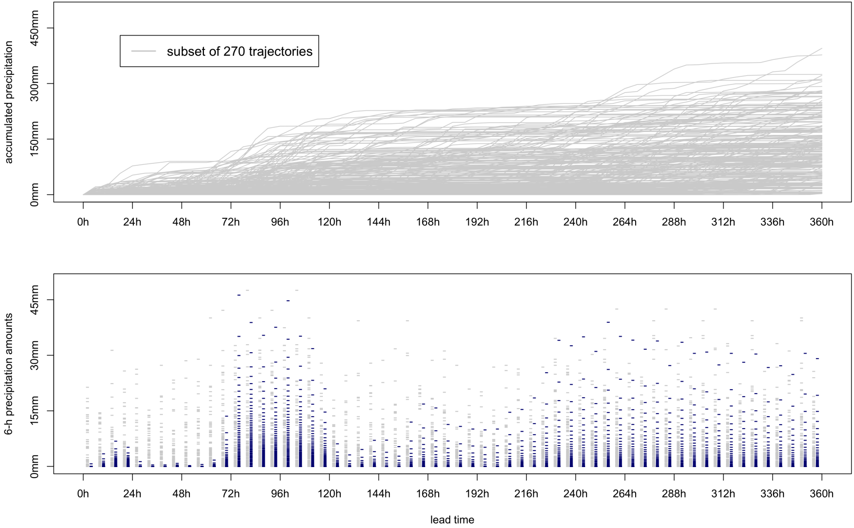
Improving the Schaake Shuffle: step 2

In the second iteration step they are further reduced to 340:



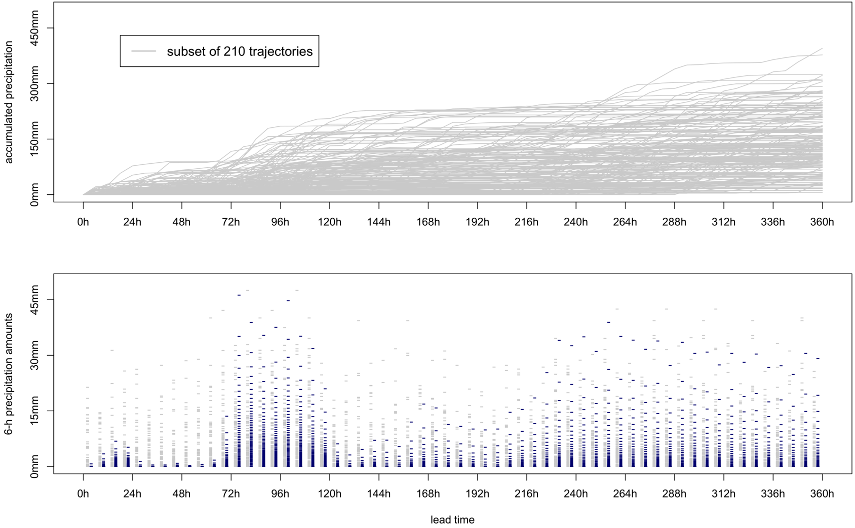
Improving the Schaake Shuffle: step 2

In the third iteration step they are further reduced to 270:



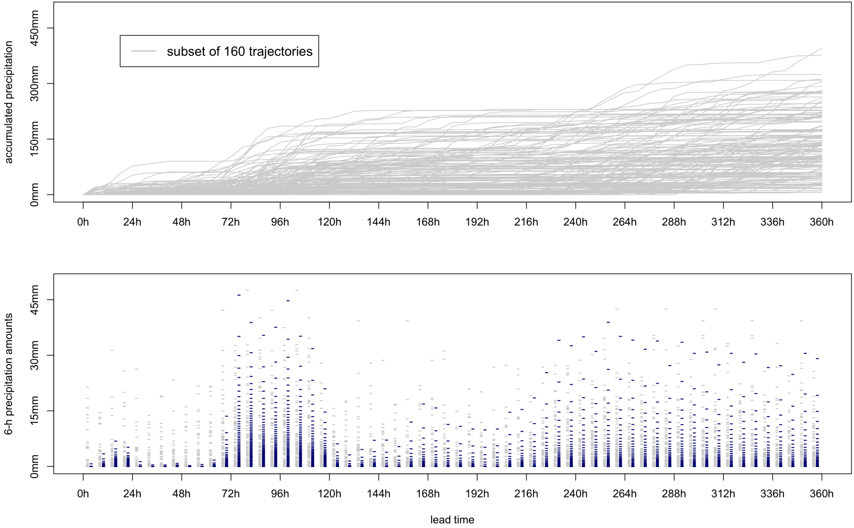
Improving the Schaake Shuffle: step 2

In the 4th iteration step they are further reduced to 210:



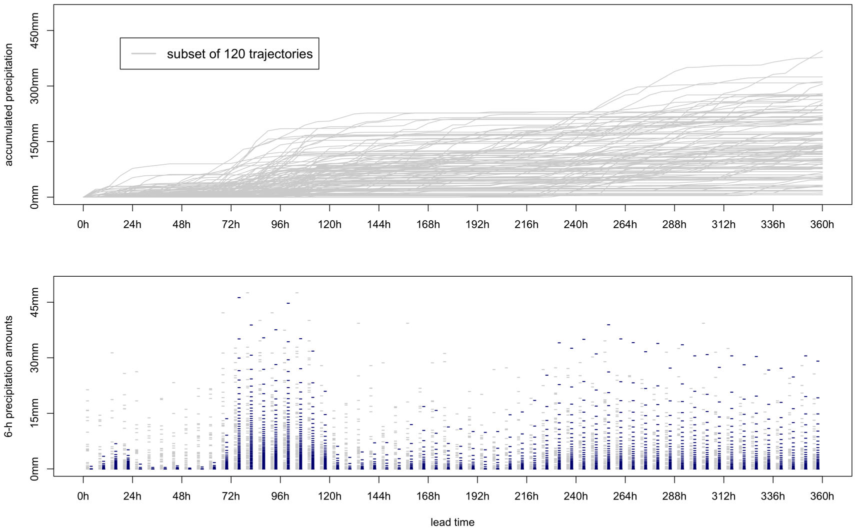
Improving the Schaake Shuffle: step 2

In the 5th iteration step they are further reduced to 160:



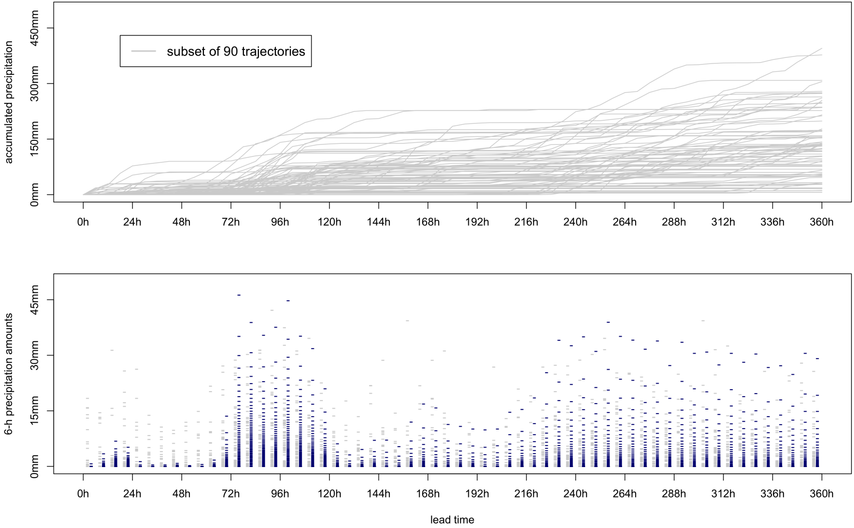
Improving the Schaake Shuffle: step 2

In the 6th iteration step they are further reduced to 120:



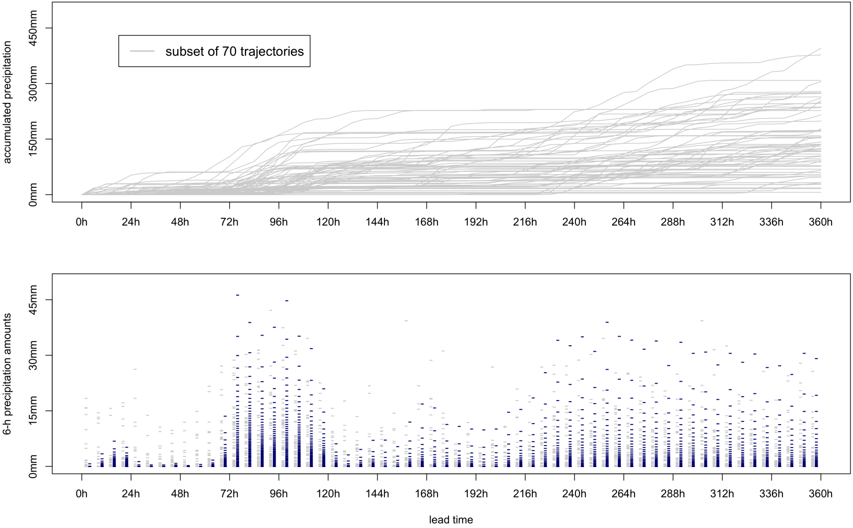
Improving the Schaake Shuffle: step 2

In the 7th iteration step they are further reduced to 90:



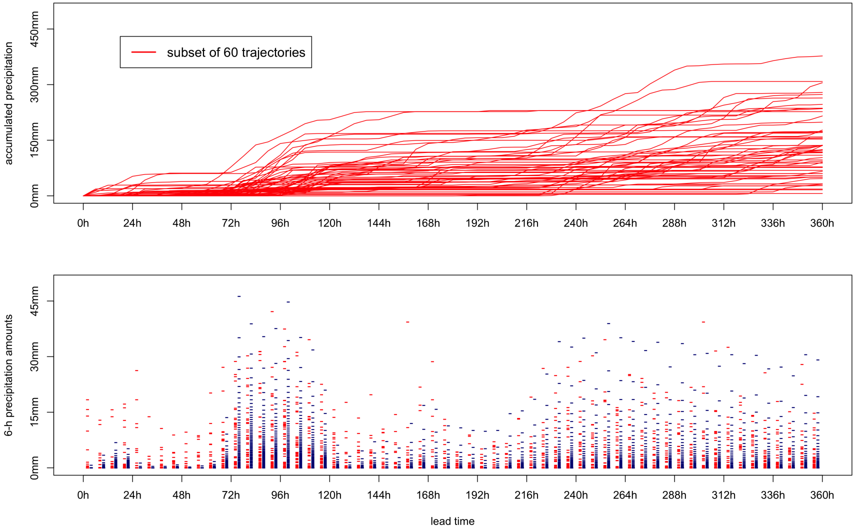
Improving the Schaake Shuffle: step 2

In the 8th iteration step they are further reduced to 70:



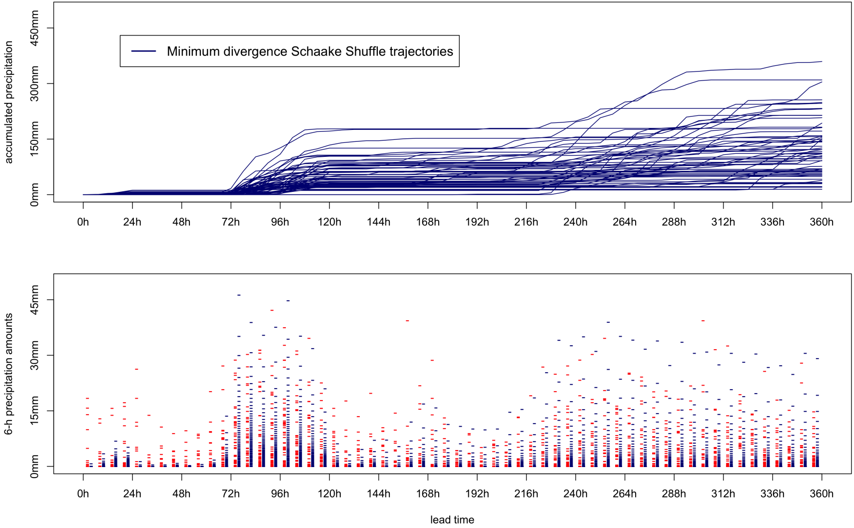
Improving the Schaake Shuffle: step 2

Finally, only m (here: 60) of the original 553 trajectories are retained:



Improving the Schaake Shuffle: step 2

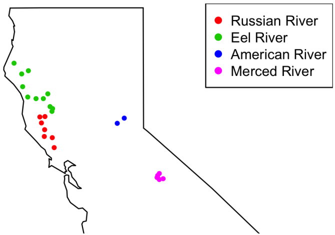
Now, less re-mapping is required to match the predictive ensemble:



Application to four river basins in Northern California

In the following we consider

- 25 stations in 4 different basins
- verification period: 1985 to 2010
- 6-h precipitation accumulations & 6-h average surface temperatures
- forecasts lead times up to 15 days



First, the GEFS ensemble (11 members) forecasts are calibrated with the univariate techniques (Gneiting et al., 2005; Scheuerer and Hamill, 2015) sketched above, using NOAA's 2nd generation reforecast data set.

Then, we apply the standard Schaake Shuffle (StSS) and the minimum divergence Schaake Shuffle (MDSS) described above to generate calibrated 60-member forecast trajectories for the two weather variables.

Application to four river basins in Northern California

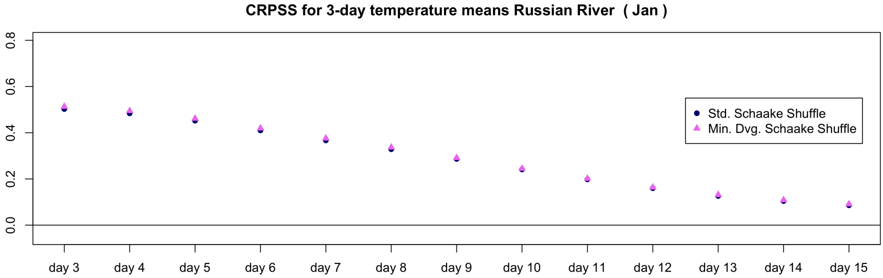
To compare the performance of StSS and MDSS we

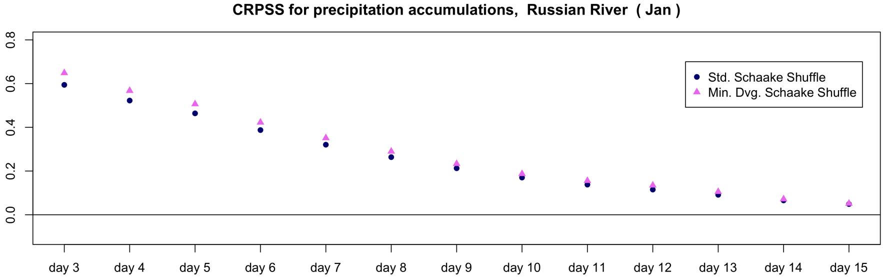
- use leave-one-year-out cross validation
- aggregate the temperature trajectories to a univariate quantity by considering 3-day mean temperatures, averaged over all stations within each basin
- aggregate the precipitation trajectories to a univariate quantity by considering 3-day accumulated precipitation, averaged over all stations within each basin
- calculate the continuous ranked probability skill score (CRPSS)

$$crps(F, y) = \int \big(F(x) - \mathbf{1}(x \geq y)\big)^2 dx$$

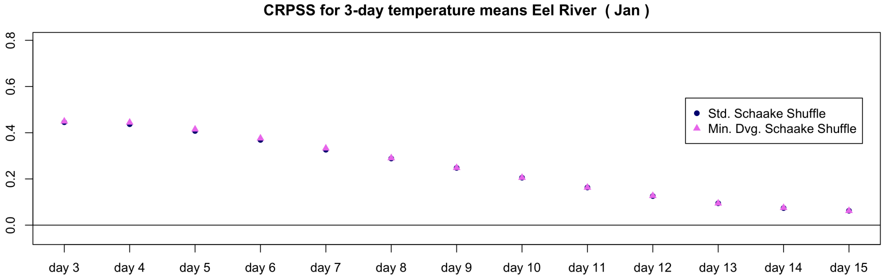
of the StSS and MDSS forecasts for these aggregated quantities.

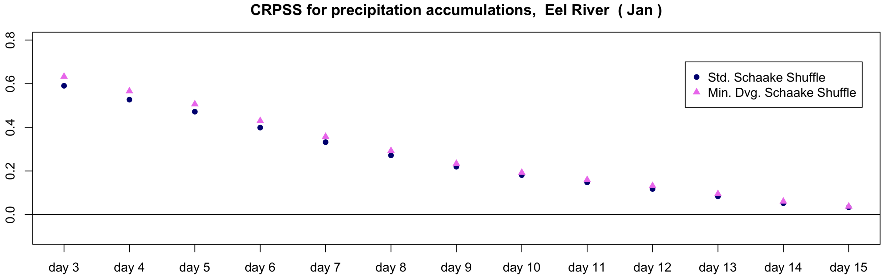
CRPSS for Russian River Basin (January)



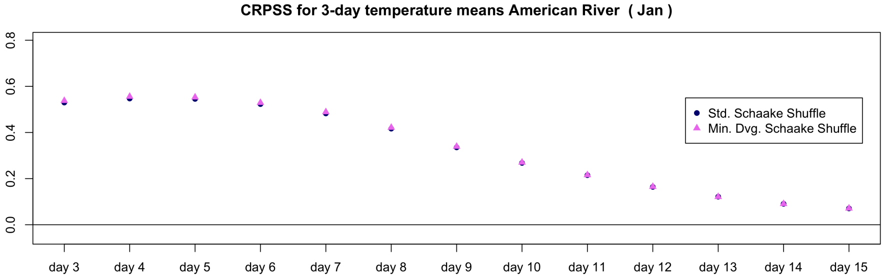


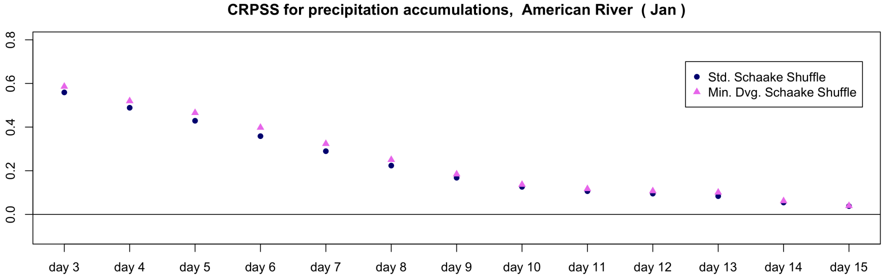
CRPSS for Eel River Basin (January)



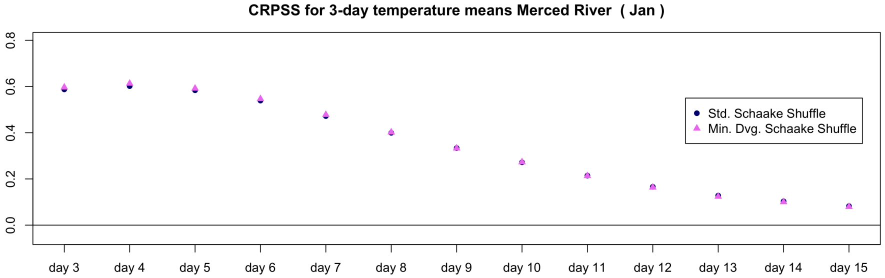


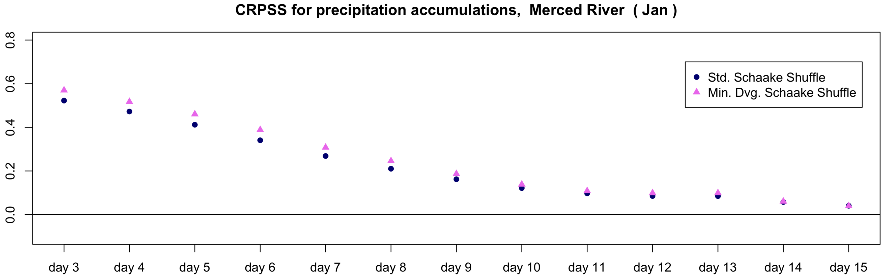
CRPSS for American River Basin (January)



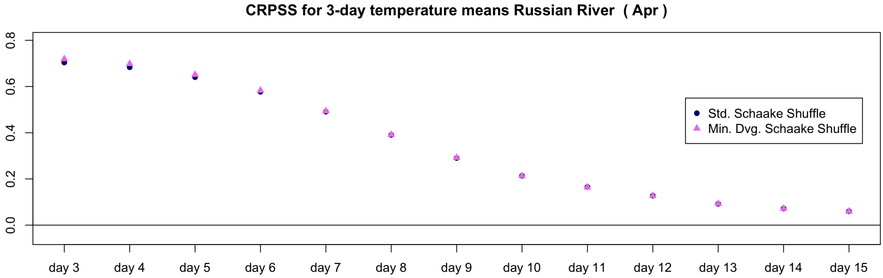


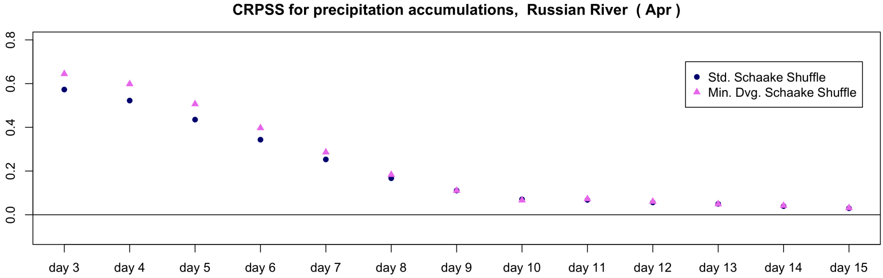
CRPSS for Merced River Basin (January)



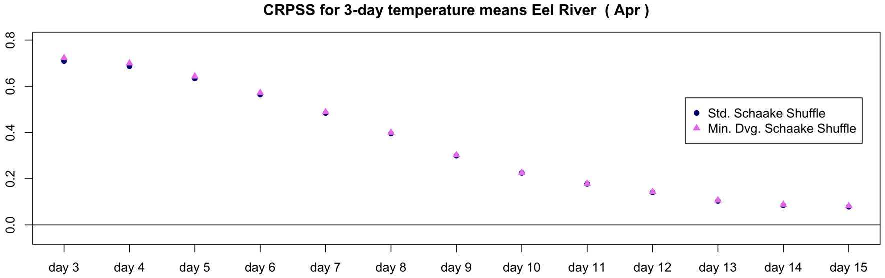


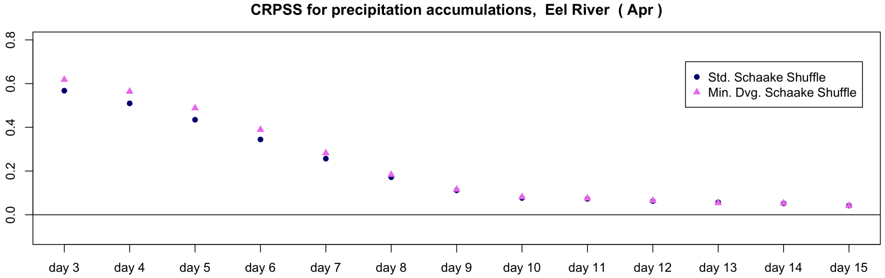
CRPSS for Russian River Basin (April)



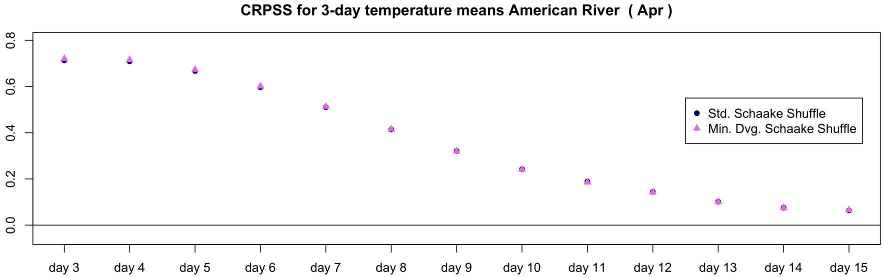


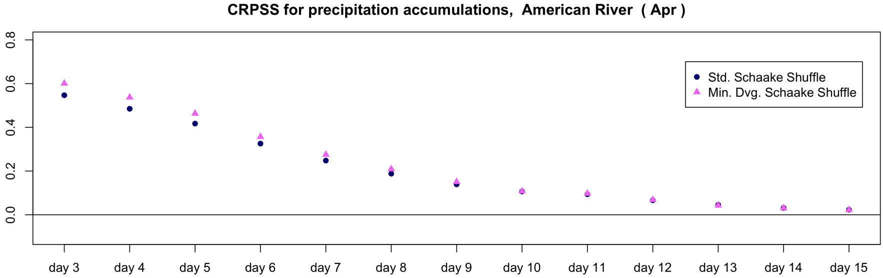
CRPSS for Eel River Basin (April)



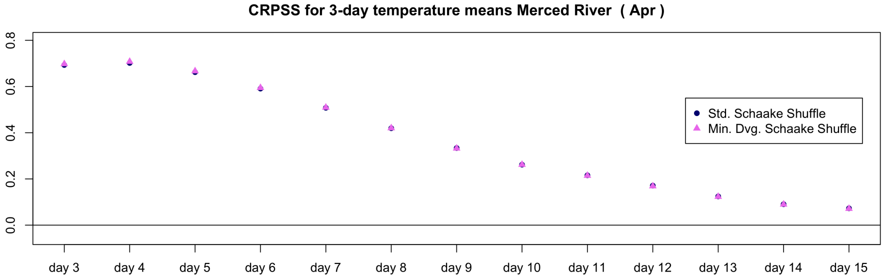


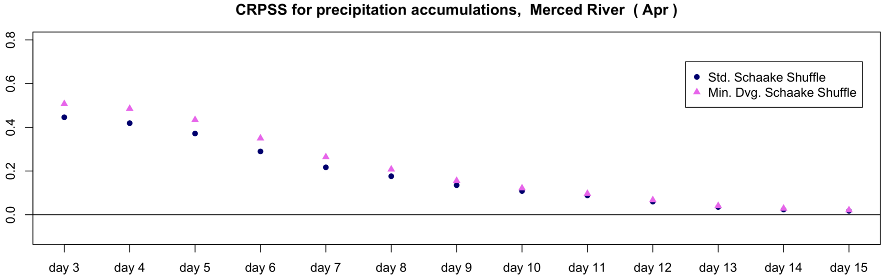
CRPSS for American River Basin (April)



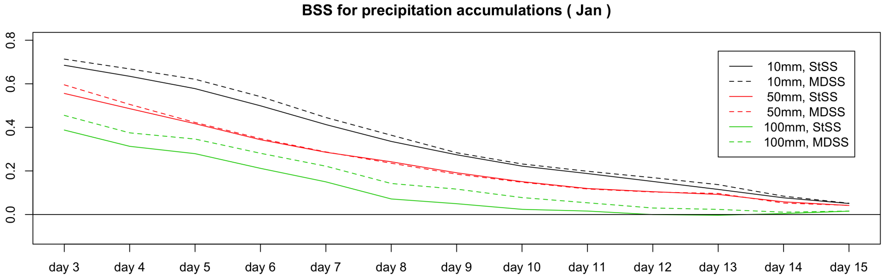


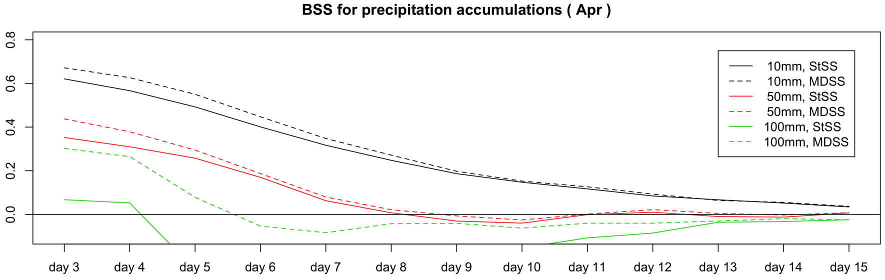
CRPSS for Merced River Basin (April)





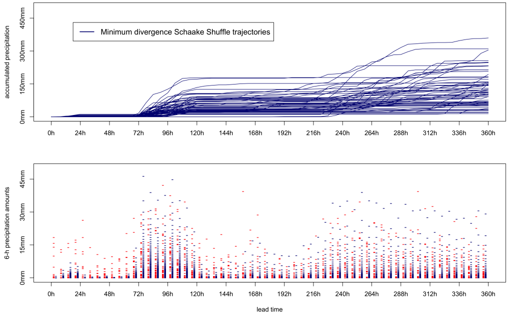
Brier skill scores (all basins)





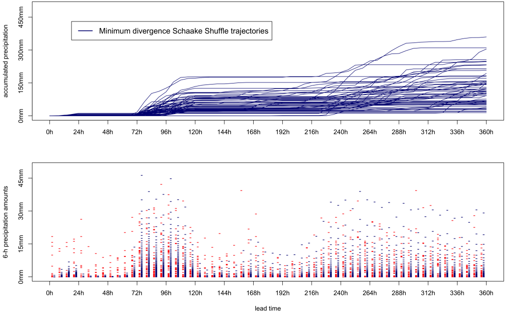
Summary

- Hydrological streamflow forecast models strongly rely on predictions of temperature and precipitation amounts as inputs ('forcings')
- Statistical post-processing must account for spatial and temporal dependence and dependence between the two weather variables in order to yield realistic forecast trajectories
- The Schaake Shuffle is an excellent way to achieve that; we presented a variant of it (MDSS) that makes it even better



Summary

- Hydrological streamflow forecast models strongly rely on predictions of temperature and precipitation amounts as inputs ('forcings')
- Statistical post-processing must account for spatial and temporal dependence and dependence between the two weather variables in order to yield realistic forecast trajectories
- The Schaake Shuffle is an excellent way to achieve that; we presented a variant of it (MDSS) that makes it even better



Thanks for listening!

References I

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.
The Schaake shuffle: A method for reconstructing spaceâtime variability in forecasted precipitation and temperature fields.
J. Hydrometeor., 5:243–262, 2004.

Gneiting, T., Raftery, A.E., Westveld, A.H., and Goldman, T.
Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation.
Mon. Wea. Rev.,133:1098–1118, 2005.

Scheuerer, M. and T. M. Hamill
Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted Gamma distributions.
Mon. Wea. Rev., 143:4578–4596, 2015.