

Lab assignment 5

ATM S 544

Dominik Stiller

November 20, 2023

1 Description of code

The script uses the serial ensemble square root filter (EnSRF) from Whitaker & Hamill (2002). Instead of perturbing observations for the deviation update of each ensemble member, this algorithm calculates a scaling factor α for the Kalman gain to force the correct covariance for the deviations.

Cycling data assimilation in the script is implemented as follows:

1. Initialize the ensemble from staggered ERA5 analyses. (A real ensemble forecasting system would use carefully chosen perturbations here, e.g., based on Bred vectors.)
2. For each day between 2022-11-15 and 2022-12-31:
 - a) Forecast each ensemble member from the previous analysis using PanguWeather with a 24-h lead time.
 - b) Get the "observation" y from ERA5.
 - c) Calculate the observation estimate $H\mathbf{x}^P$ for each ensemble member.
 - d) Perform the EnSRF update for each field. (This could be parallelized.)
 - e) Set the analysis as initial condition for the next forecast.

2 Sample covariance

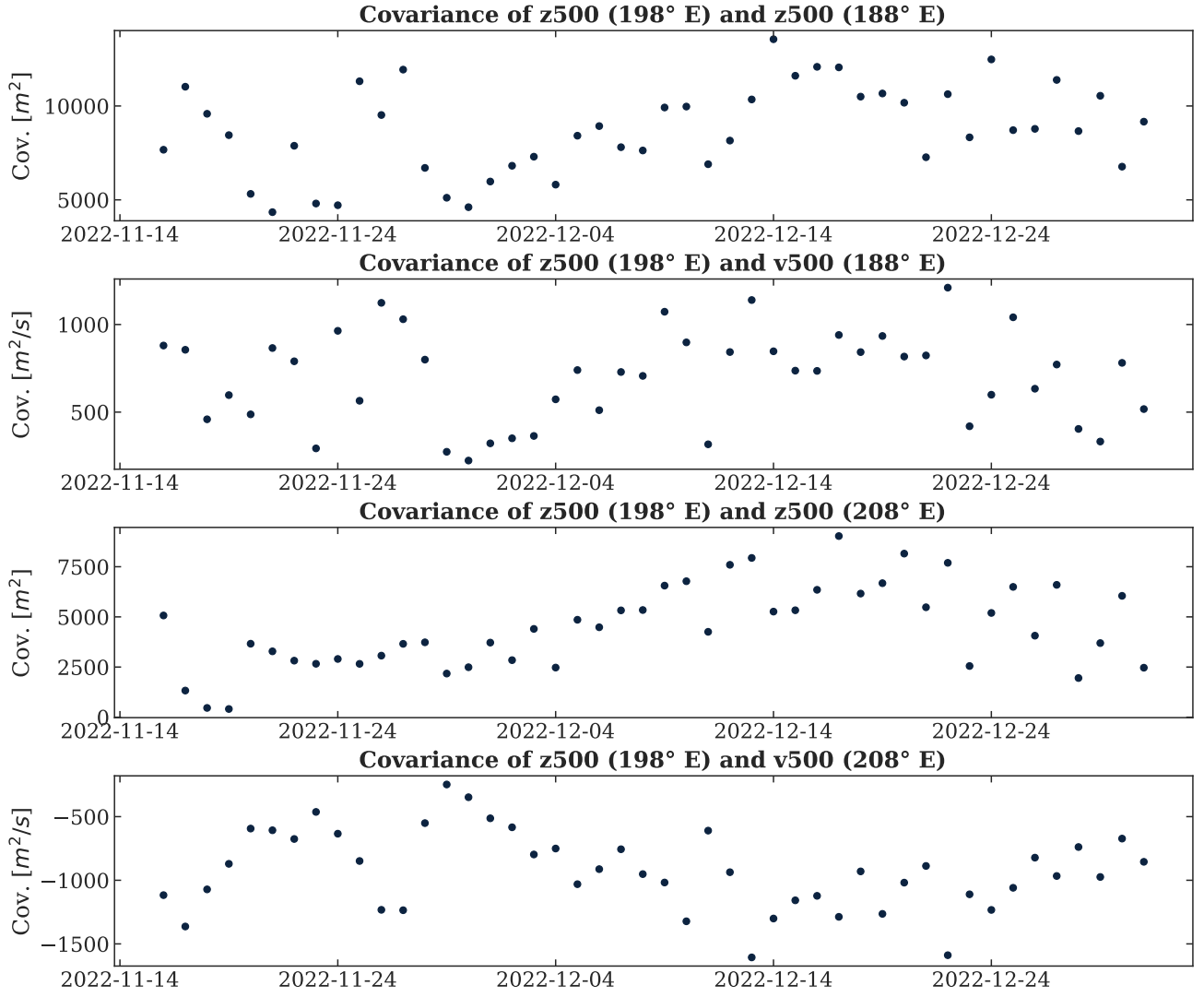


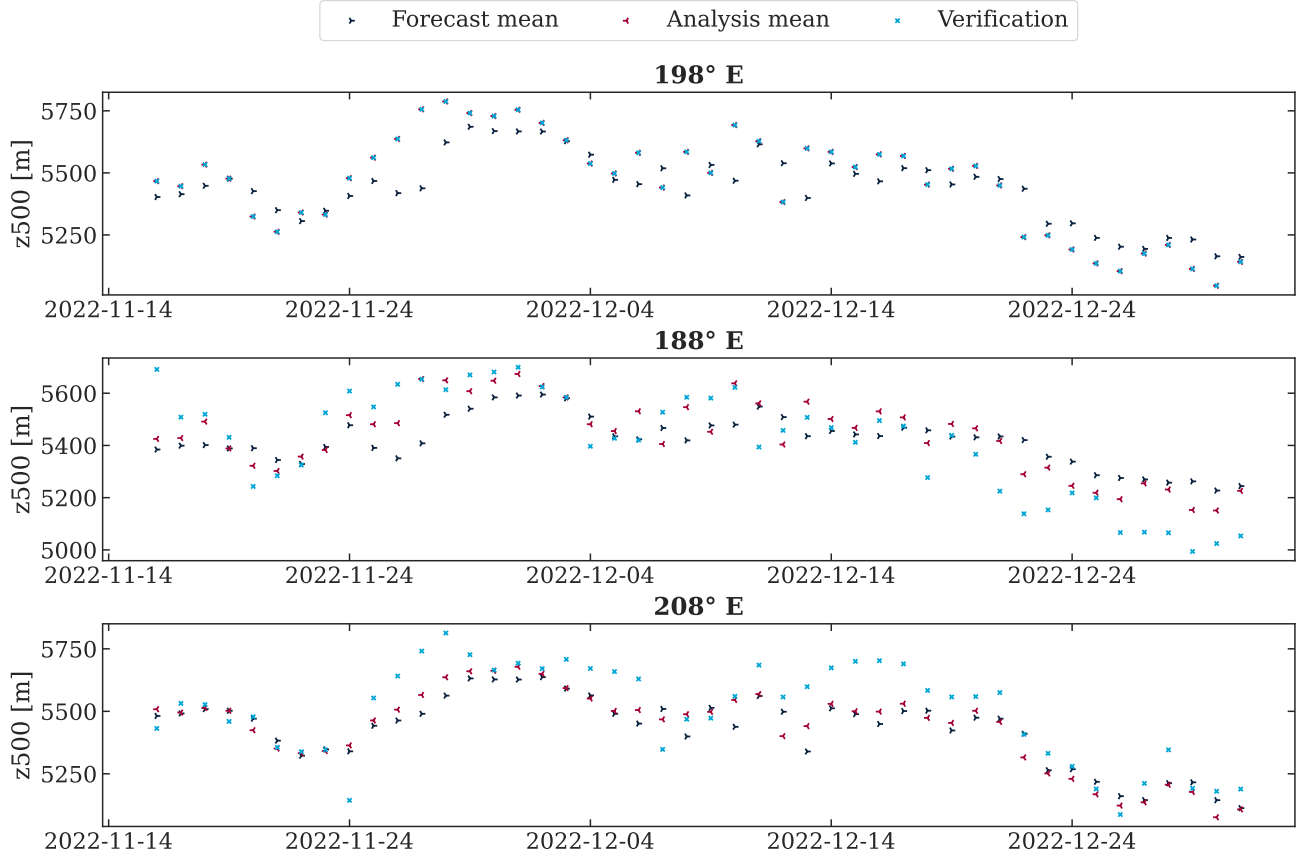
Figure 1: Sample covariance from the prior between fields and locations.

Time series of the sample covariances at the three locations are shown in Figure 1. The covariances change with time, but it appears to be a random walk rather than a physically meaningful change. There is a minimum in most of the absolute covariances around December 1. The meridional wind speeds have covariances of opposite signs with the geopotential height in between, as postulated by geostrophic flow.

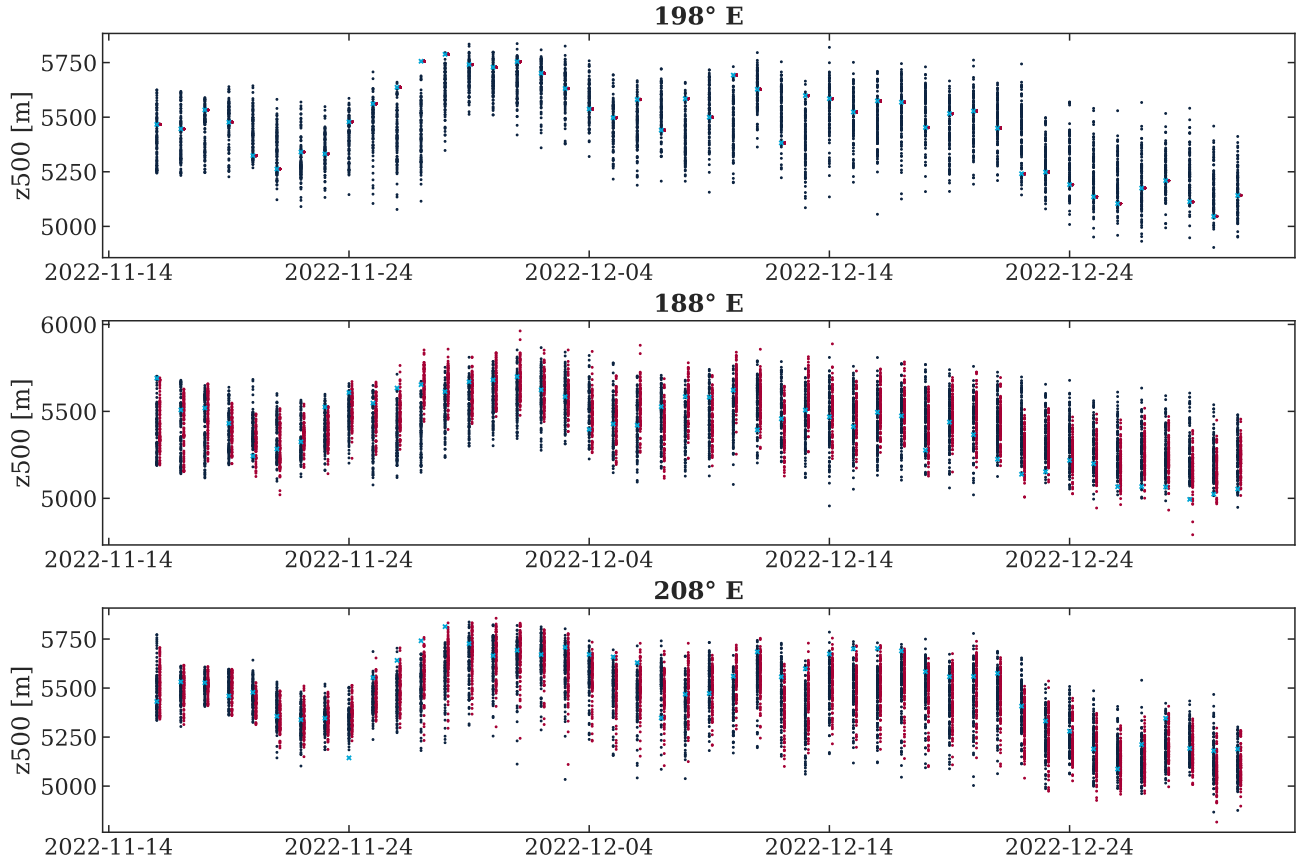
3 Geopotential height

Time series of the geopotential height at the three locations are shown in Figure 2. At the observation location, the analysis always coincides with the verification (which is identical to the observation for this location) since the state–observation covariance here is, obviously, maximal (correlation of 1). Therefore, the analysis spread is almost zero (exactly zero is prevented by the finite observation variance). At the other locations, the analysis is generally closer to the observation than the prior, although this is not always the case. The analysis spread is generally slightly less than the prior spread, but not by much. In some cases (e.g., 188° E on the third last day), the minimum–to–maximum range is even larger in the analysis, although the variance has to be smaller than that of the forecast by virtue of the EnSRF equations. The ensemble spread remains relatively constant over time because it is reduced in the analysis but bumped in the forecast (emulating the $+Q$ term of the non-ensemble Kalman filter).

The ensemble variance and mean squared error (both for the forecast) are shown in Figure 3. The ensemble variance appears to be a lower bound for the true error, but usually underestimates it slightly, and in some cases dramatically.



(a) Ensemble mean



(b) Ensemble samples

Figure 2: Geopotential height.

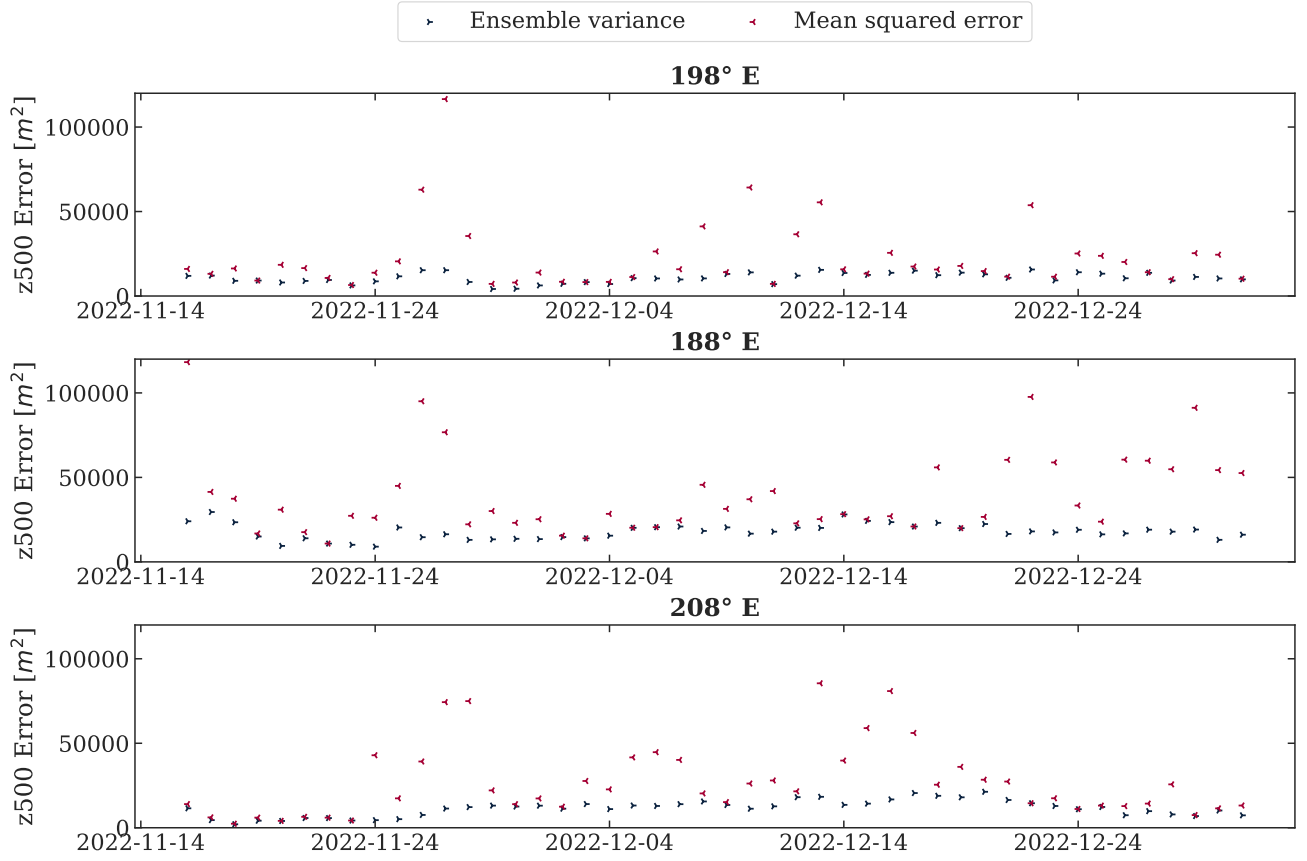


Figure 3: Ensemble variance and mean squared error for geopotential height forecast.