Correlated observation errors in the perturbed observation ensemble data assimilation

Anna Shlyaeva
NOAA/ESRL, Boulder

Mark Buehner
Environment Canada, Montreal

Chris Snyder
NCAR, Boulder

Motivation

- Moving towards global high-resolution systems that resolve wide range of scales. Background may have errors on scales from global to convective; for ensemble systems simple one-scale spatial localization approaches might not be optimal
- Observation error statistics unknown and hard to estimate, but there's evidence that observations may have spatially correlated errors
- Both background and observation error covariances influence how different scales in the analysis are resolved

Simple 1-D problem

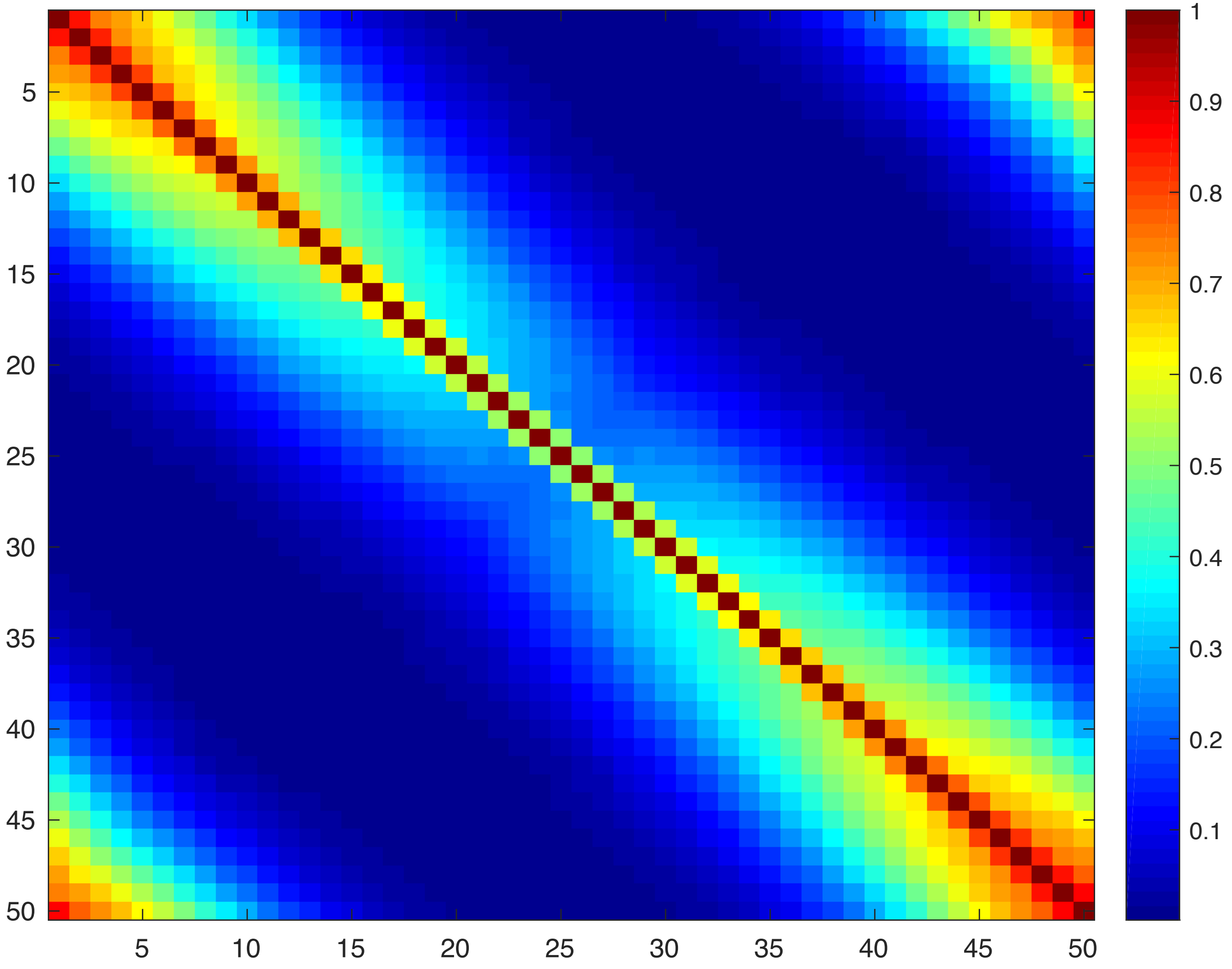
- 1D periodic domain, 50 points

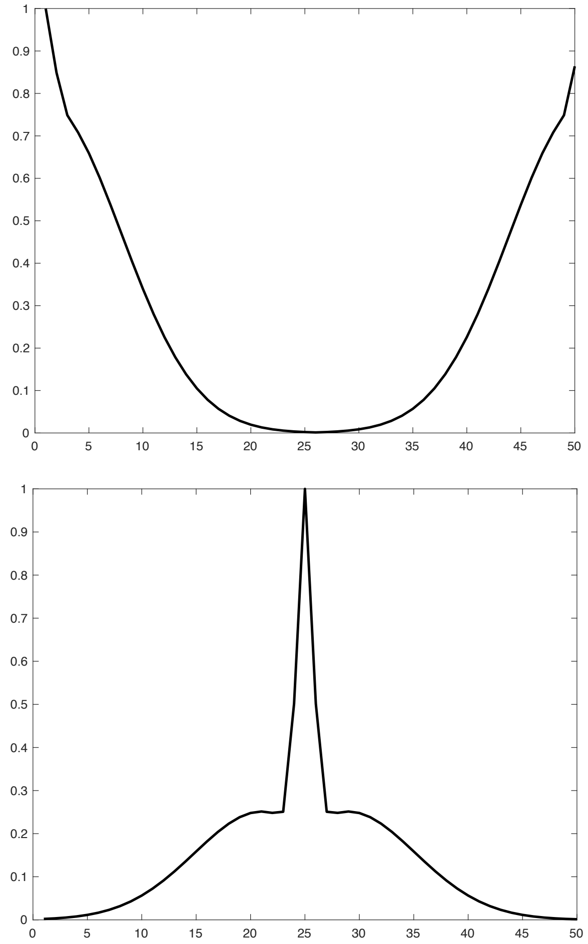
- Know true B (blend of simple correlation functions with large and small scales)
- Know true R
- Fully observed network (H is identity)

- Look at A (analysis spread and the error of analysis ensemble mean) depending on choices in ensemble assimilation (choosing localization for B and treatment of R)

True B:

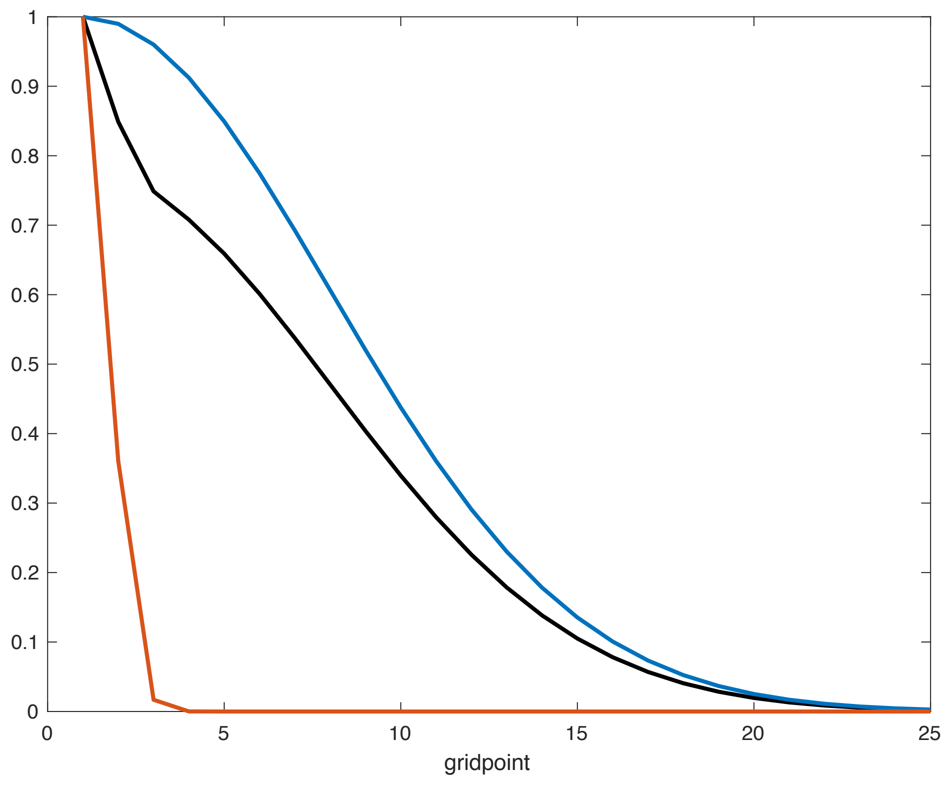
weighted sum of small (0.7) and large (7) scale Gaussian covariances



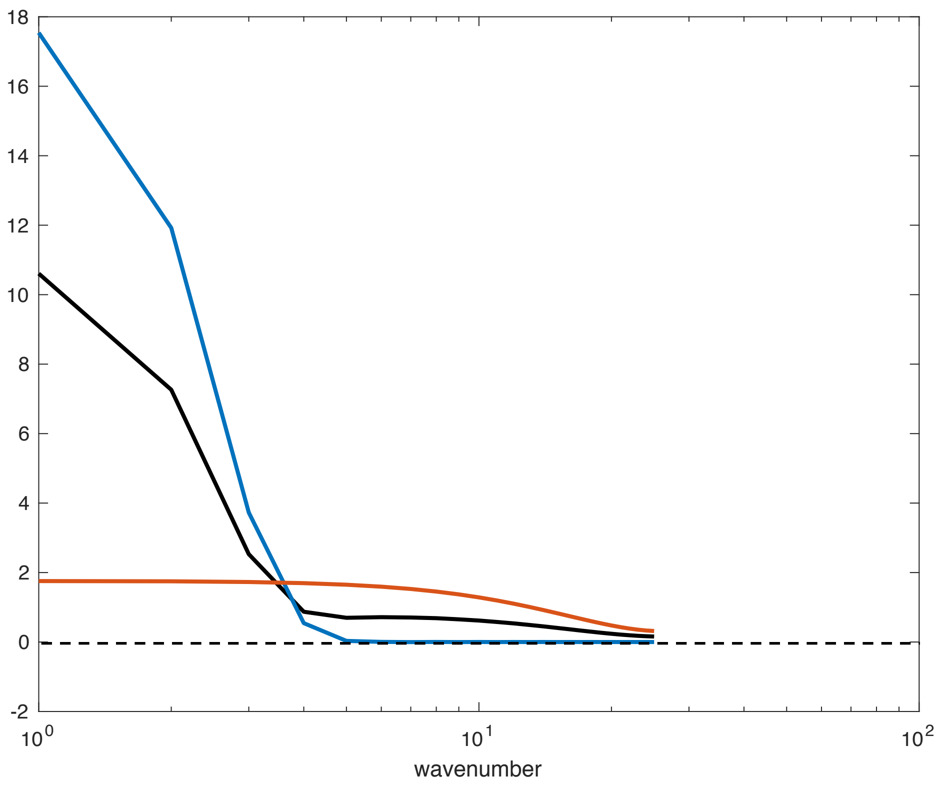


Covariances and spectral variances

Covariance in grid space



Variance in spectral space



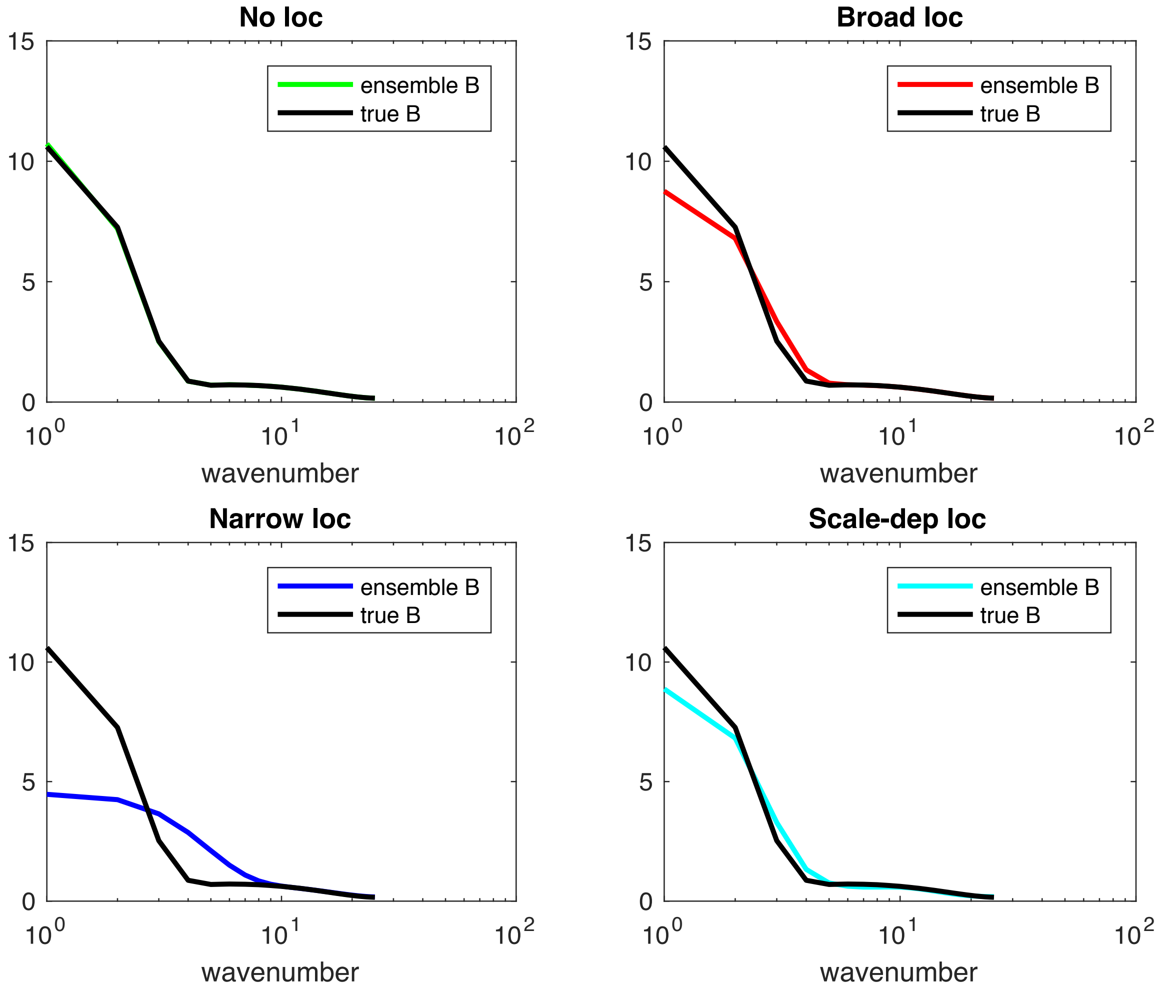
Ensemble estimates of B

- B is sampled from true B by limited size ensemble (40 members) x^b; $X^b = {}^1\!/_{\sqrt{Nens-1}}\left(x^b - \overline{x^b}\right)$
- Localization:
  - No localization: $B = X^b X^{bT}$
  - Broad localization: $B = L_{10} \circ X^b X^{bT}$
  - Narrow localization: $B = L_3 \circ X^b X^{bT}$
  - Scale-dependent localization: $B = \sum_{j_1=1,3} \sum_{j_2=1,3} X_{j_1}^b X_{j_2}^{bT} \circ L_{j_1,j_2}$ where X_j^b are scale-separated background perturbations (large, medium, small scales), $L_{j_1,j_2} = L_{j_1}^{1/2} L_{j_2}^{T/2}$ and L_j are localization functions for different scales (broad for large, narrow for small)

    (Buehner, Shlyaeva 2015; see J-F Caron talk)

Ensemble estimates of B (spectral variances)

Spectral space background error variances



Based on 5000 realizations. Color lines for the mean estimate

- No localization gives unbiased estimate of B, but the standard deviation of the estimate (not shown) is high
- Localization introduces bias in the mean estimate of B but reduces standard deviation of the B estimate
- Narrow localization is damping large scales strongly
- Scale-dependent localization gives best results compared to single-scale localizations

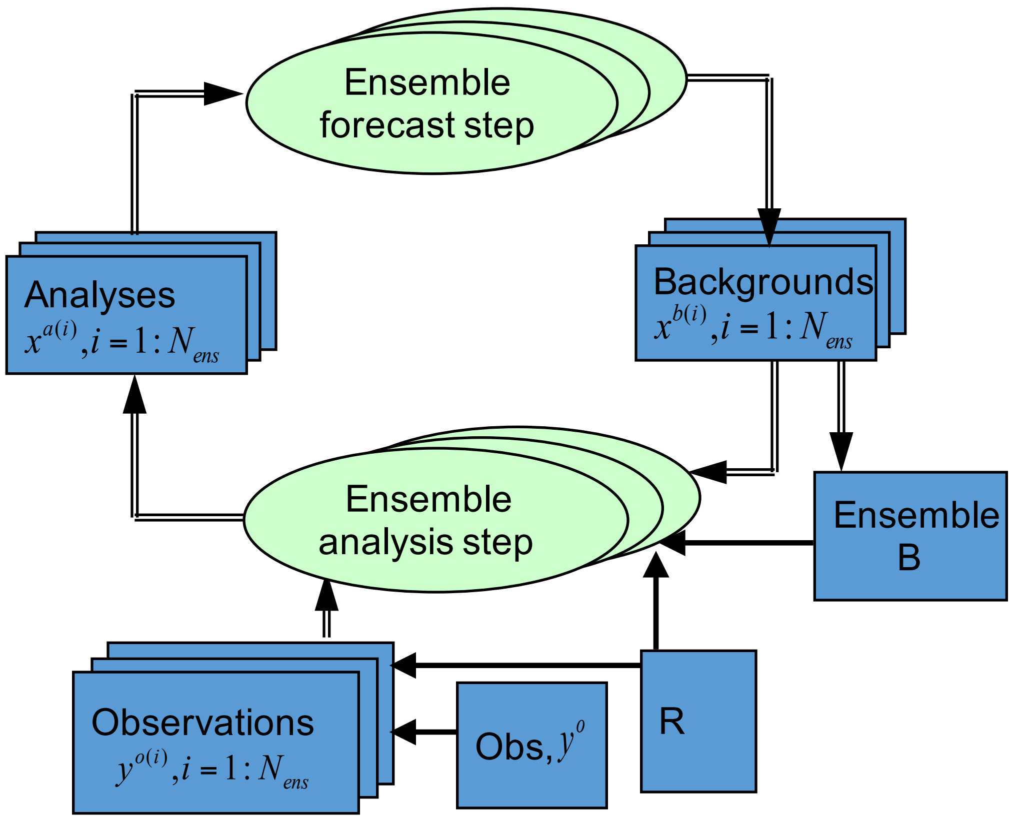
Assimilation

- Perturbed observations EnKF
  - Observations always perturbed with true R
  - Avoiding inbreeding (self-exclusion):

$$x_i^a = x_i^b + K_i\left(y_i - Hx_i^b\right)$$
$$K_i = B_{(i)}H^T\left(HB_{(i)}H^T + R\right)^{-1}$$

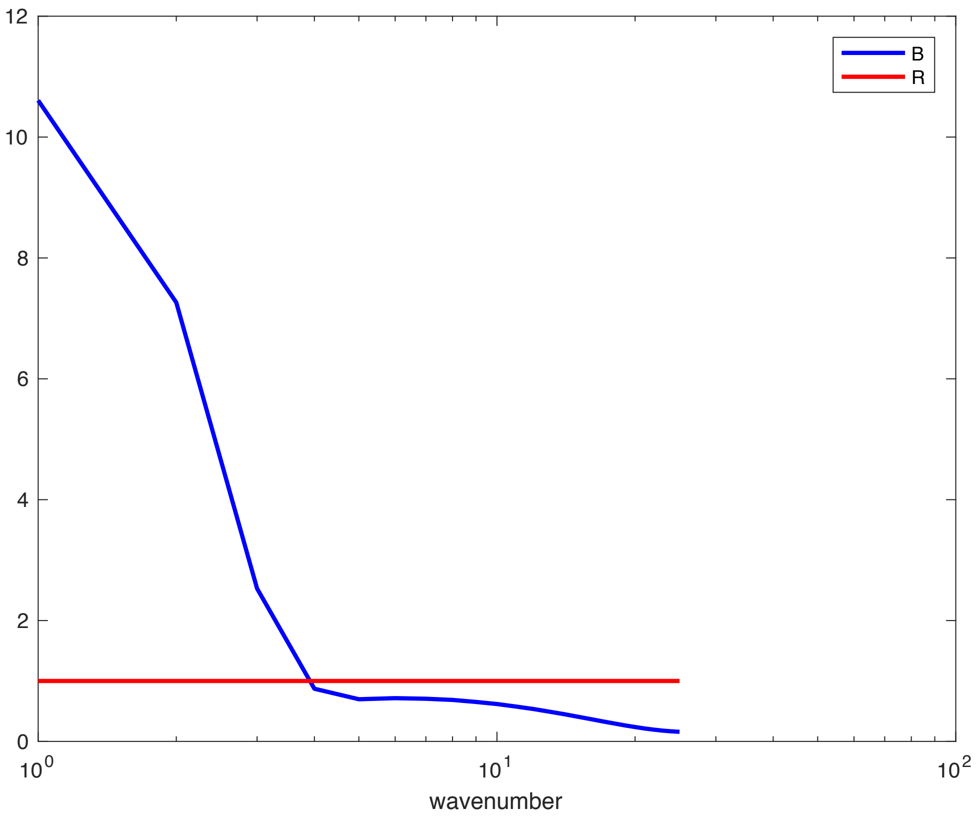
$B_{(i)}$ is estimated on all members but i^th

- Analysis spread $A = X^aX^{aT}$ very similar to analysis error covariance of ensemble mean $E = (x^a - x^t)(x^a - x^t)^T$ for this ensemble size
- For the following experiments only analysis error covariance is shown

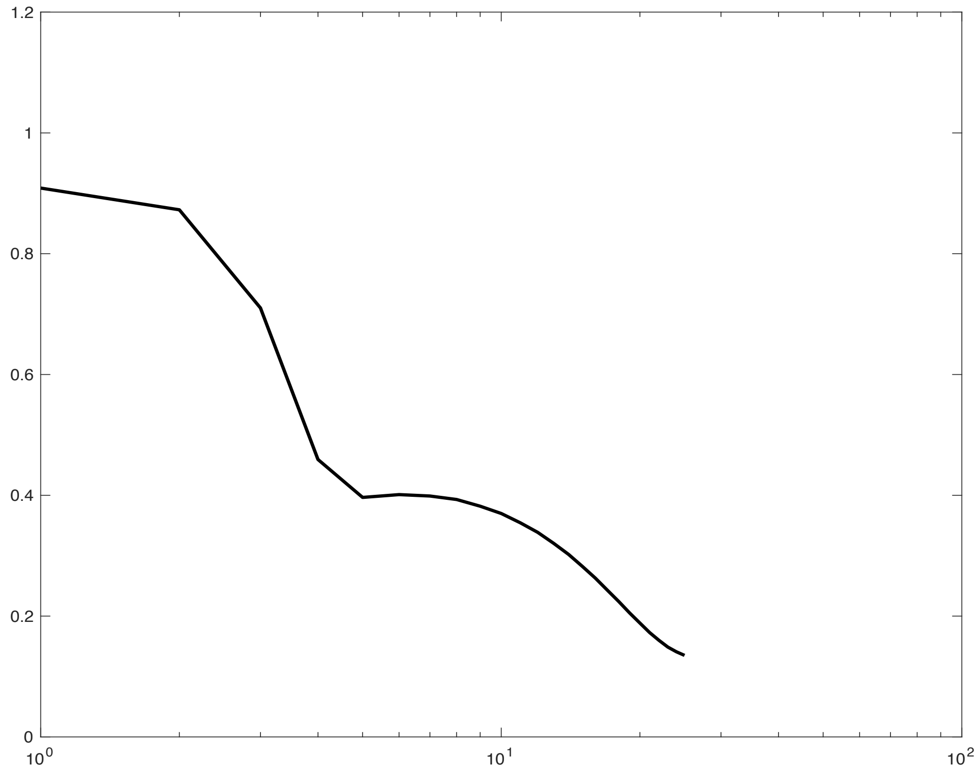


Uncorrelated observation errors & true B

Background and observation error spectral variances



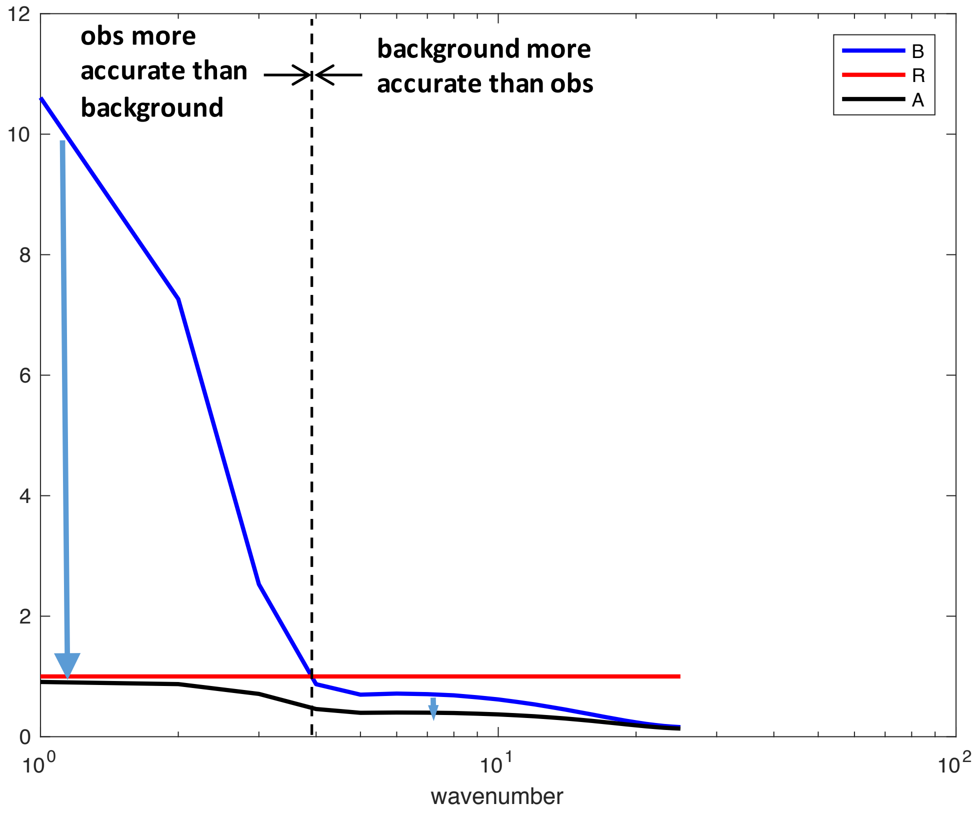
Diagonal of Kalman gain in spectral space



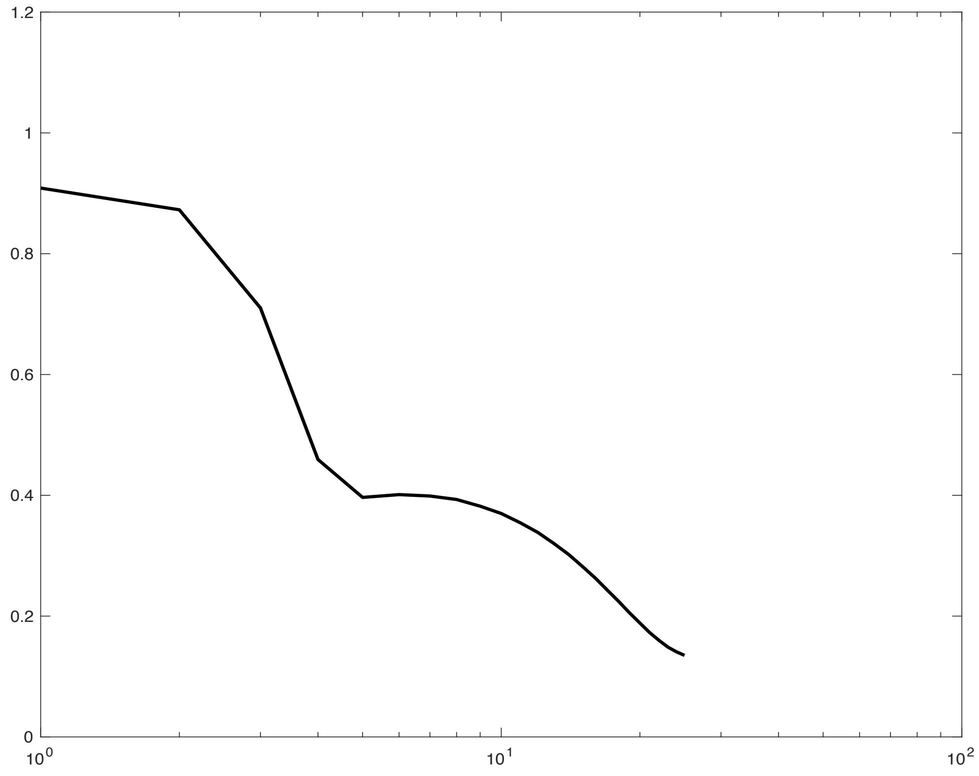
Observations have uncorrelated errors with the same variance as background errors

Uncorrelated observation errors & true B

Background, observation error and error in the analysis mean spectral variances



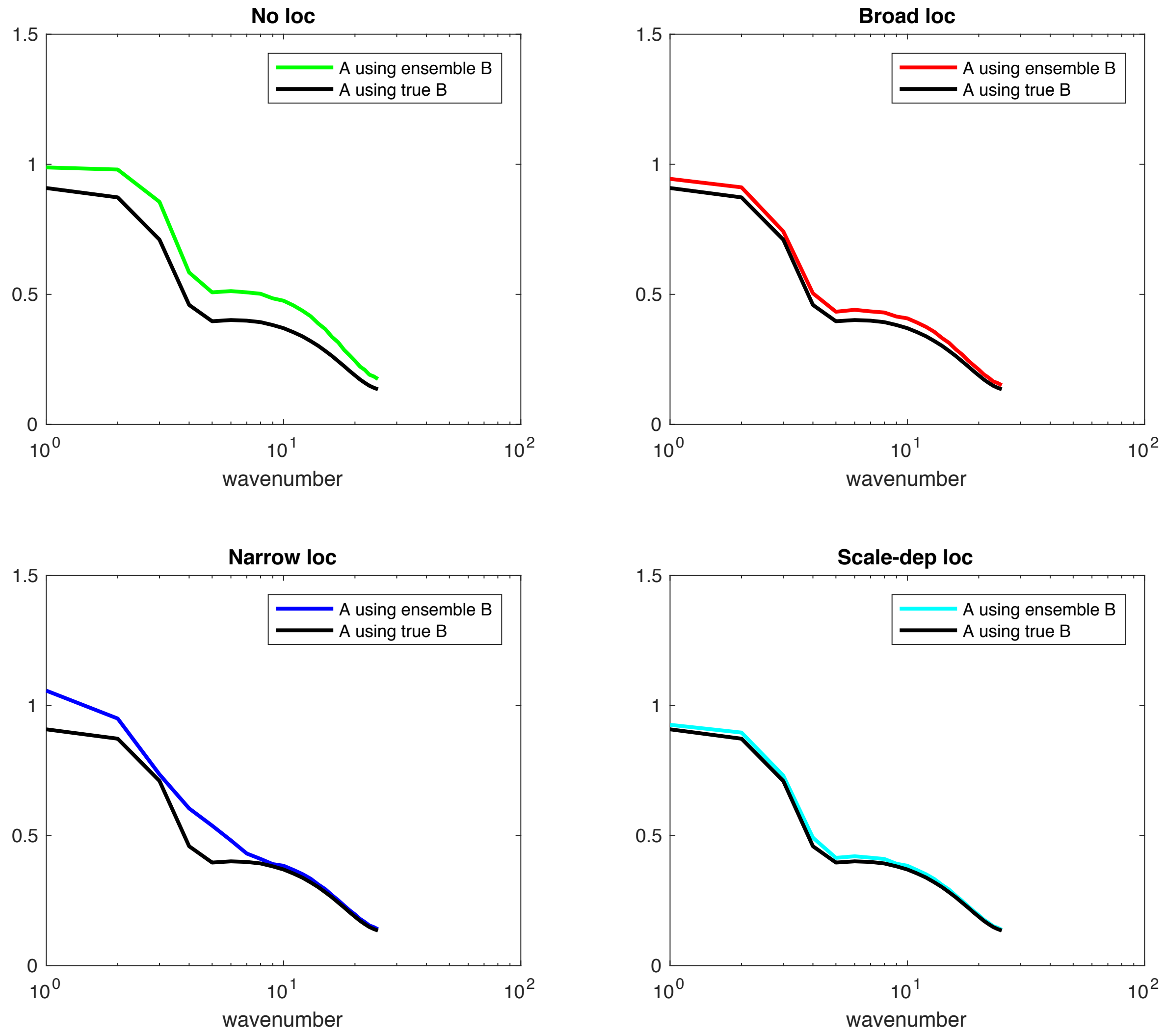
Diagonal of Kalman gain in spectral space



With diagonal R mostly large scales can be corrected by observations

Uncorrelated observation errors & ensemble B

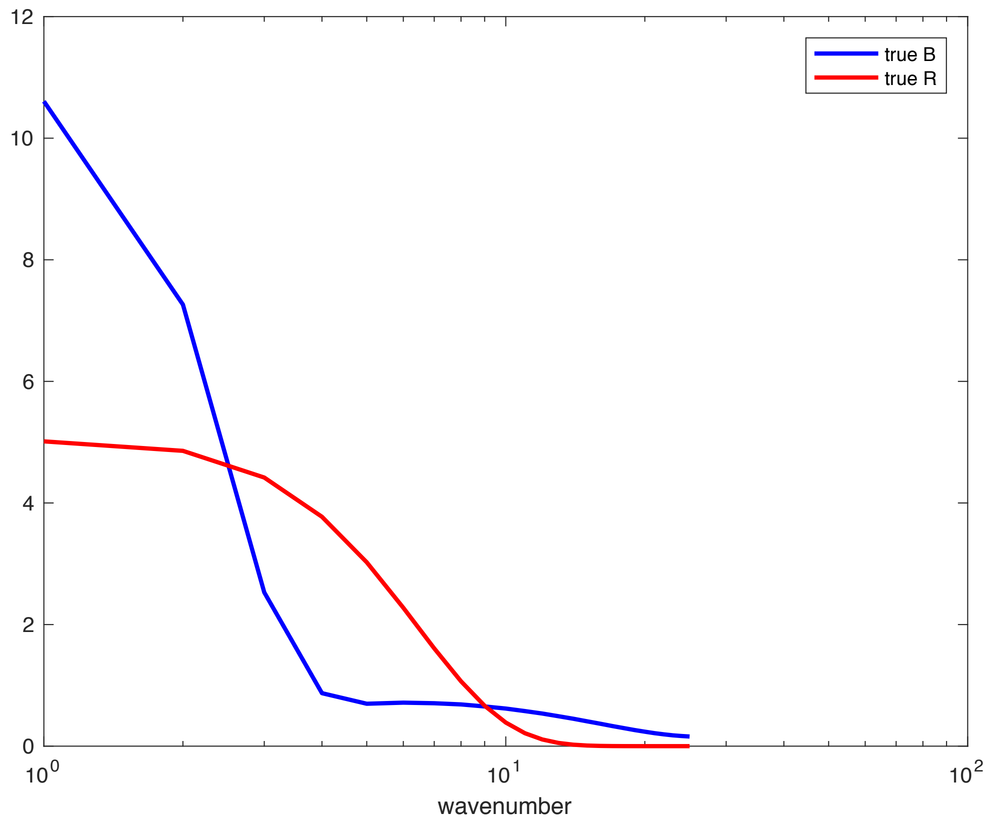
Spectral space analysis error variances



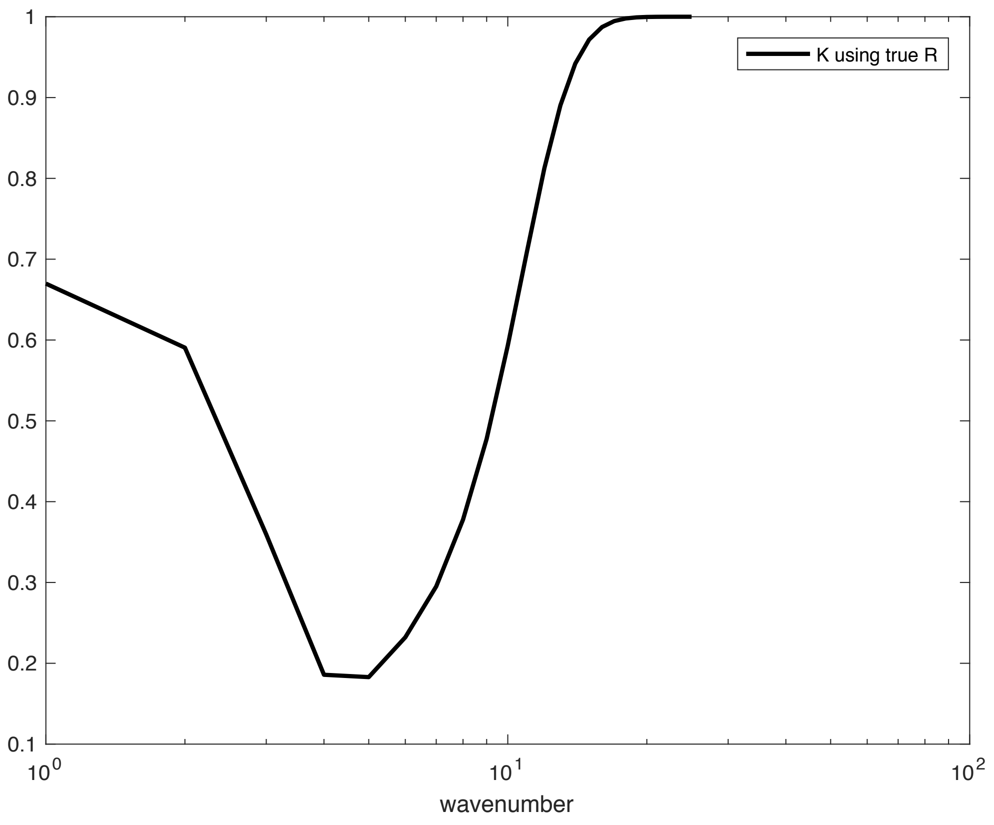
- No localization: even though B estimate is unbiased, A estimate is biased
- Narrow localization (in presence of large scale error) leads to higher errors in large scales than other types of localization
- Best results obtained with scale-dependent localization
- The benefit of using scale-dependent localization over broad localization increases with smaller ensemble size

Correlated observation errors & true B

Background and observation error spectral variances



Diagonal of Kalman gain in spectral space

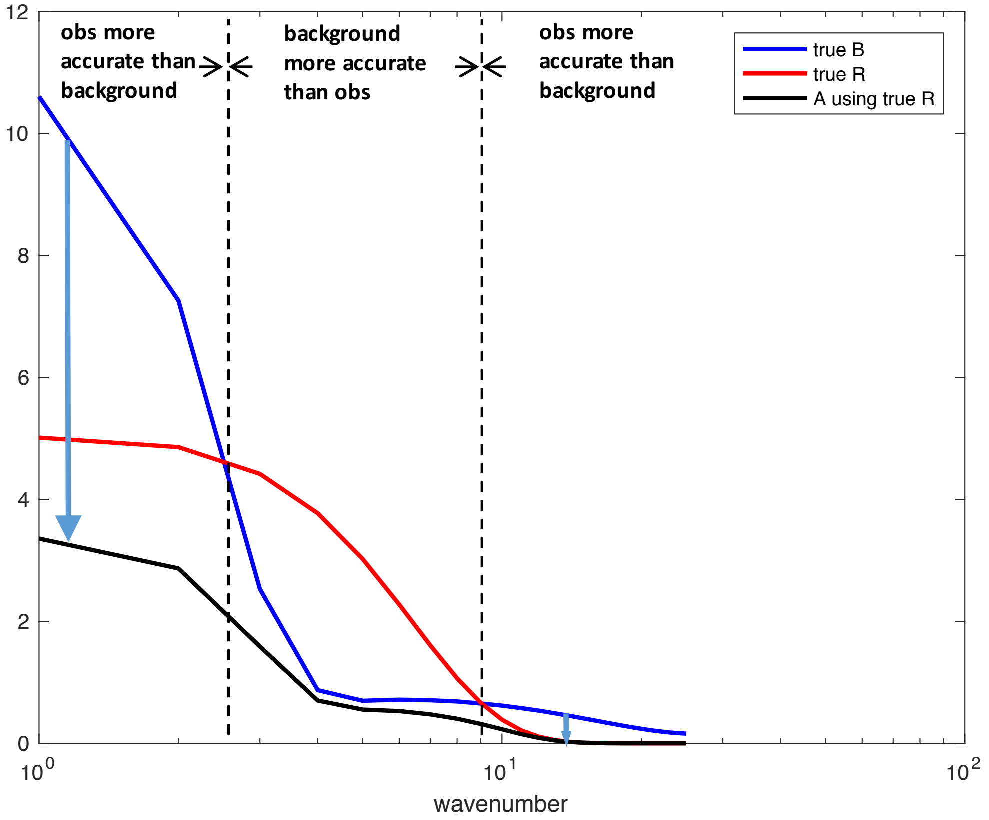


Using true correlated R in assimilation

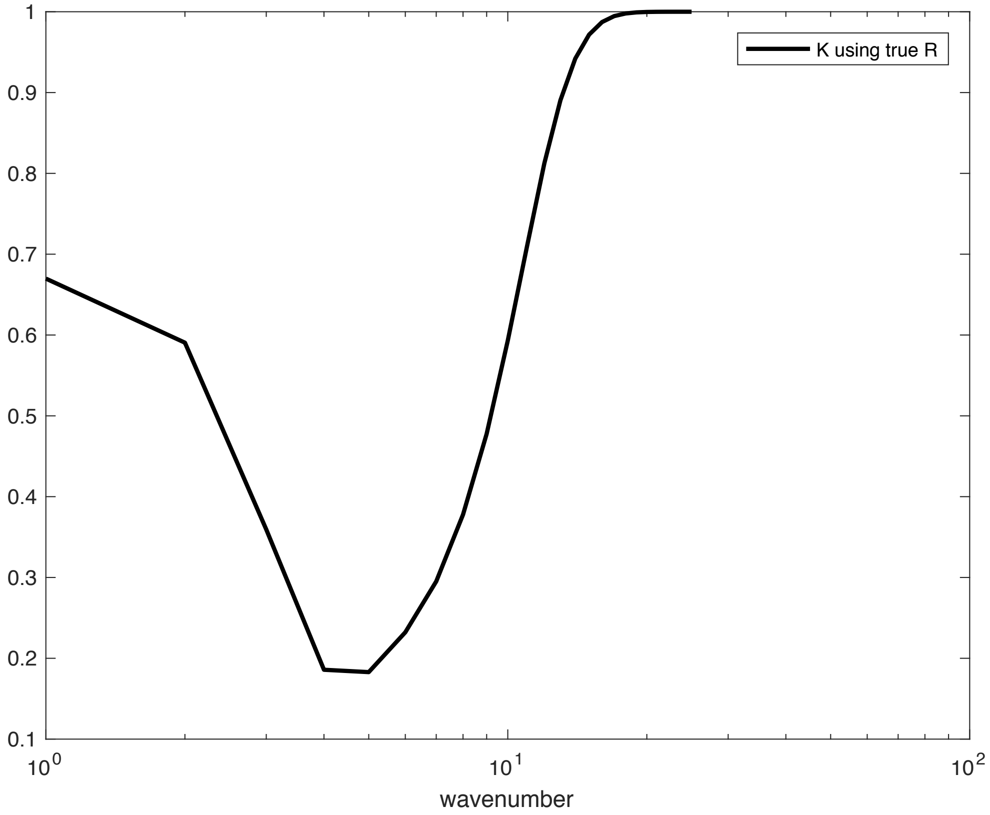
Observations have correlated errors with lengthscale (Ls=2) in between bgnd error lengthscales and same variances

Correlated observation errors & true B

Background, observation error and error in the analysis mean spectral variances



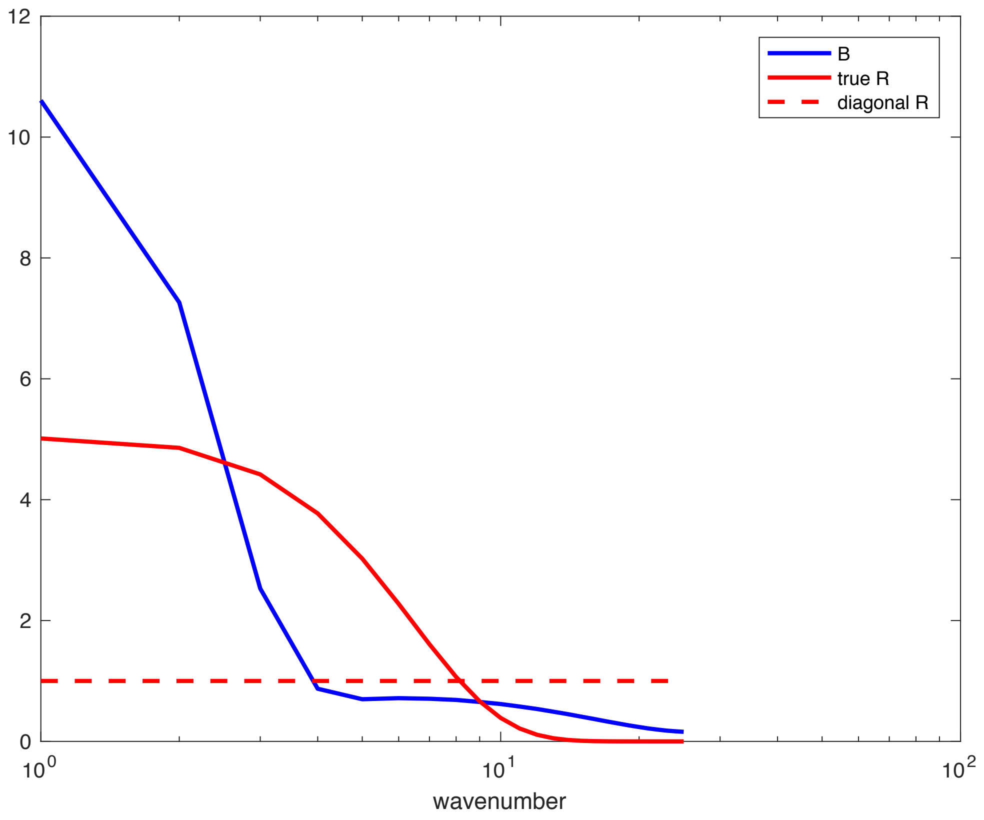
Diagonal of Kalman gain in spectral space



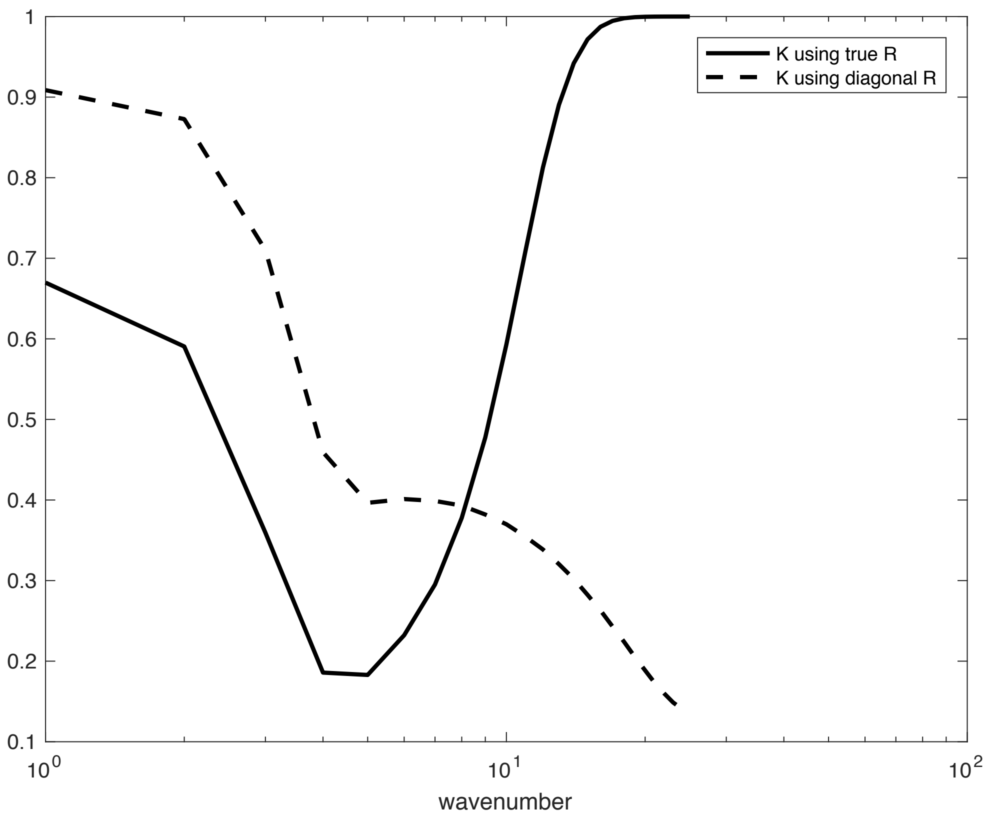
With non-diagonal R small scales can be significantly corrected by observations

Correlated observation errors & true B

Background and observation error spectral variances



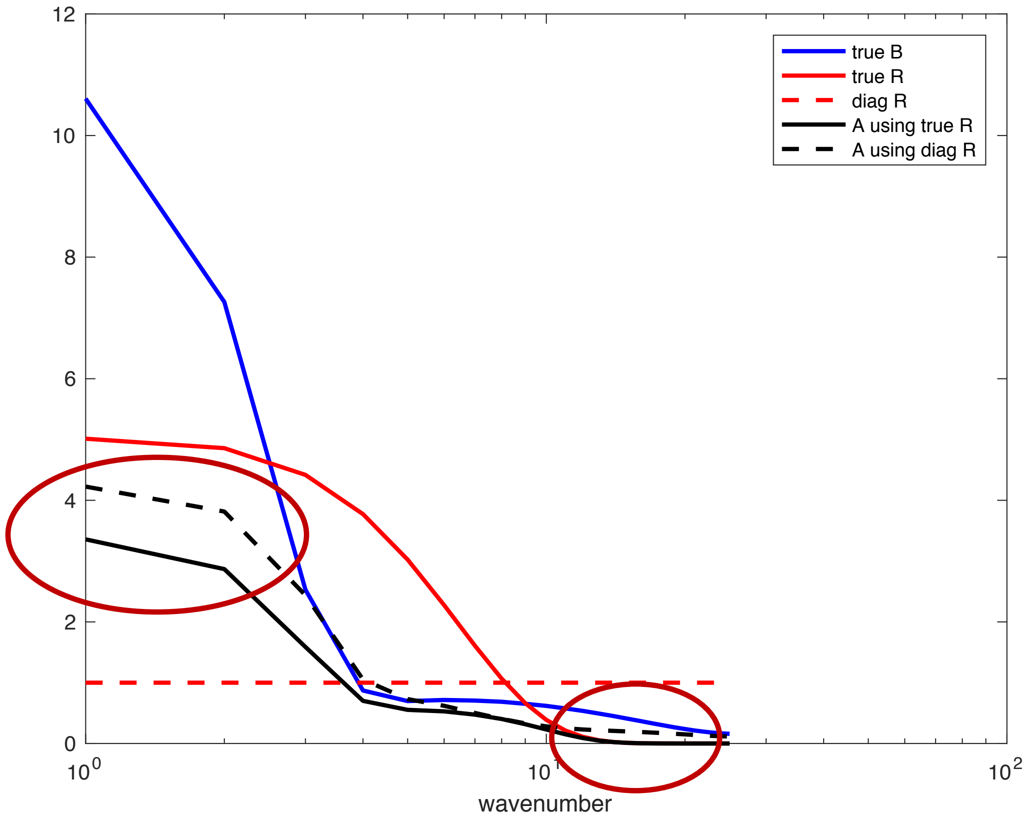
Diagonal of Kalman gain in spectral space



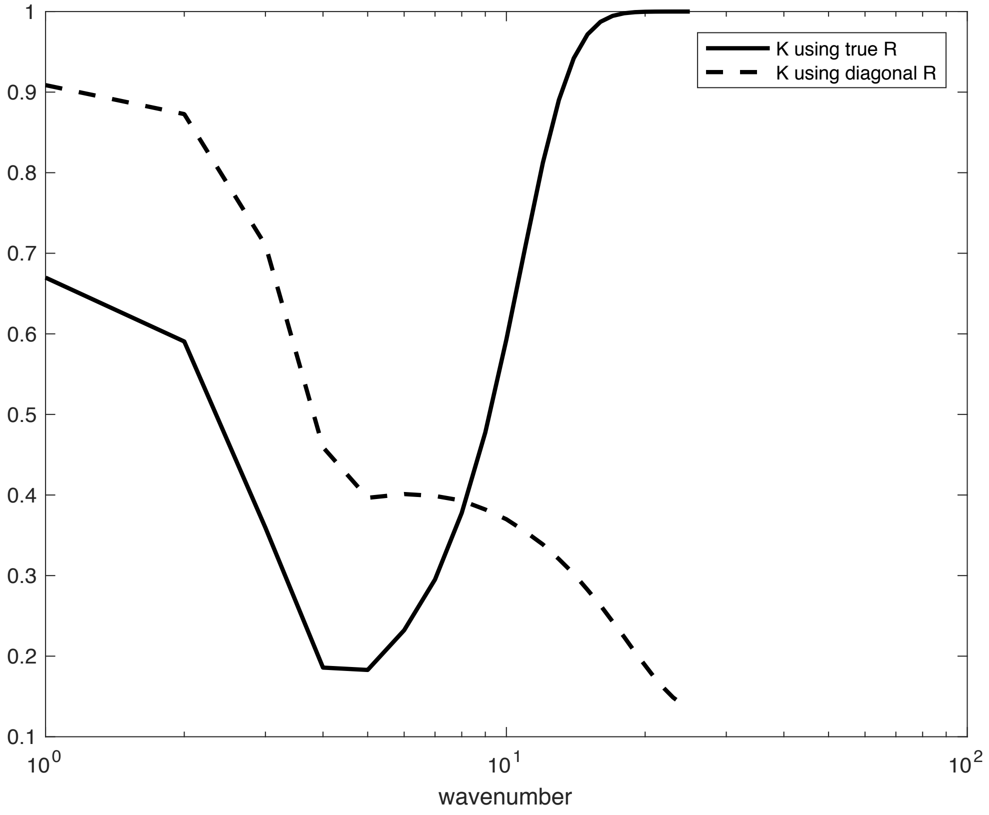
Using diagonal R (ignoring off-diagonal elements) in assimilation

Correlated observation errors & true B

Background, observation error and error in the analysis mean spectral variances



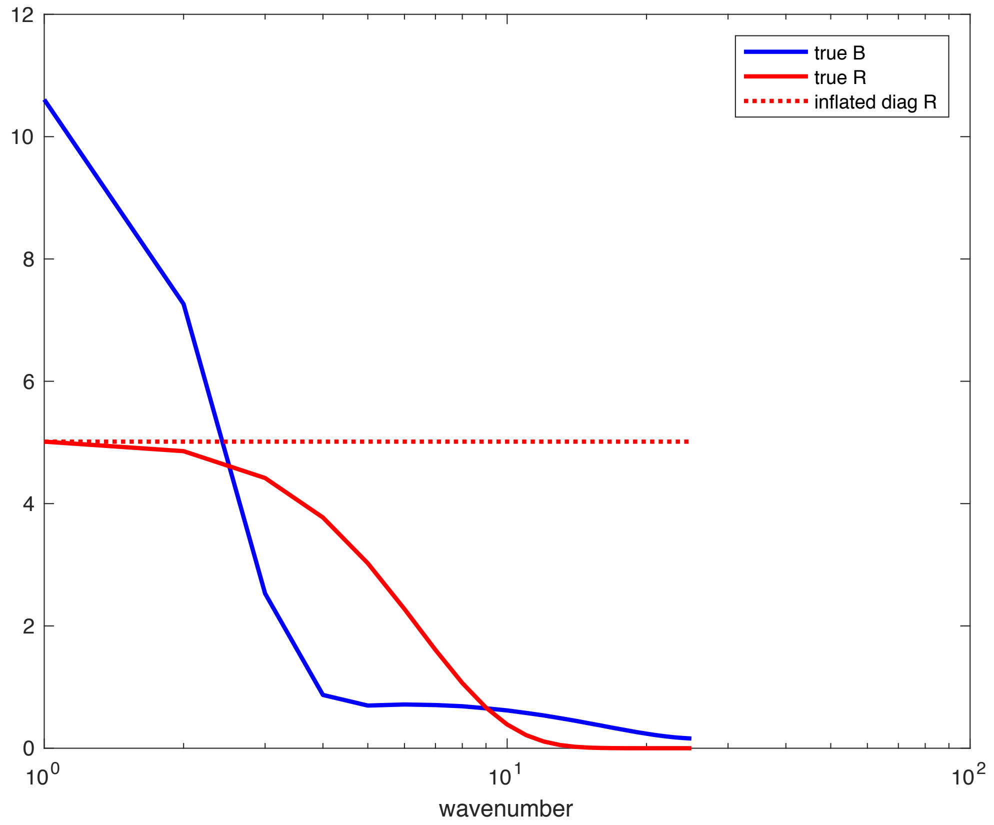
Diagonal of Kalman gain in spectral space



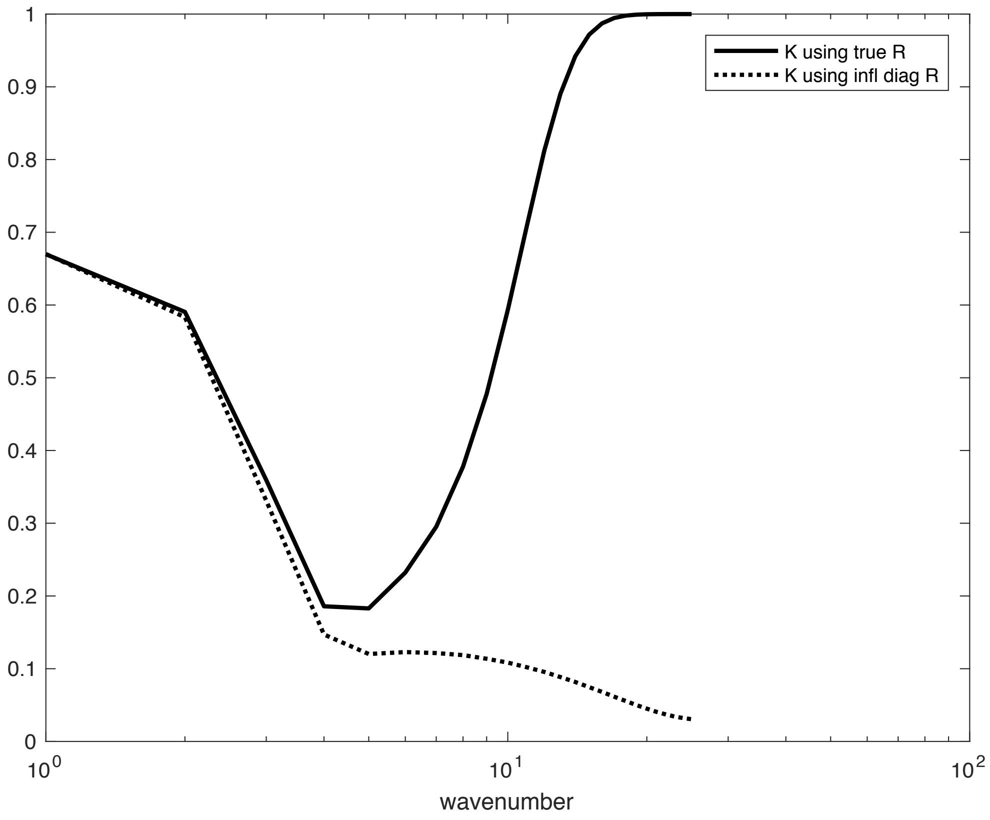
Using diagonal R in assimilation (ignoring off-diagonal elements): errors are suboptimal both for large scales (overfitting obs) and small scales (not using small-scale information in obs)

Correlated observation errors & true B

Background and observation error spectral variances



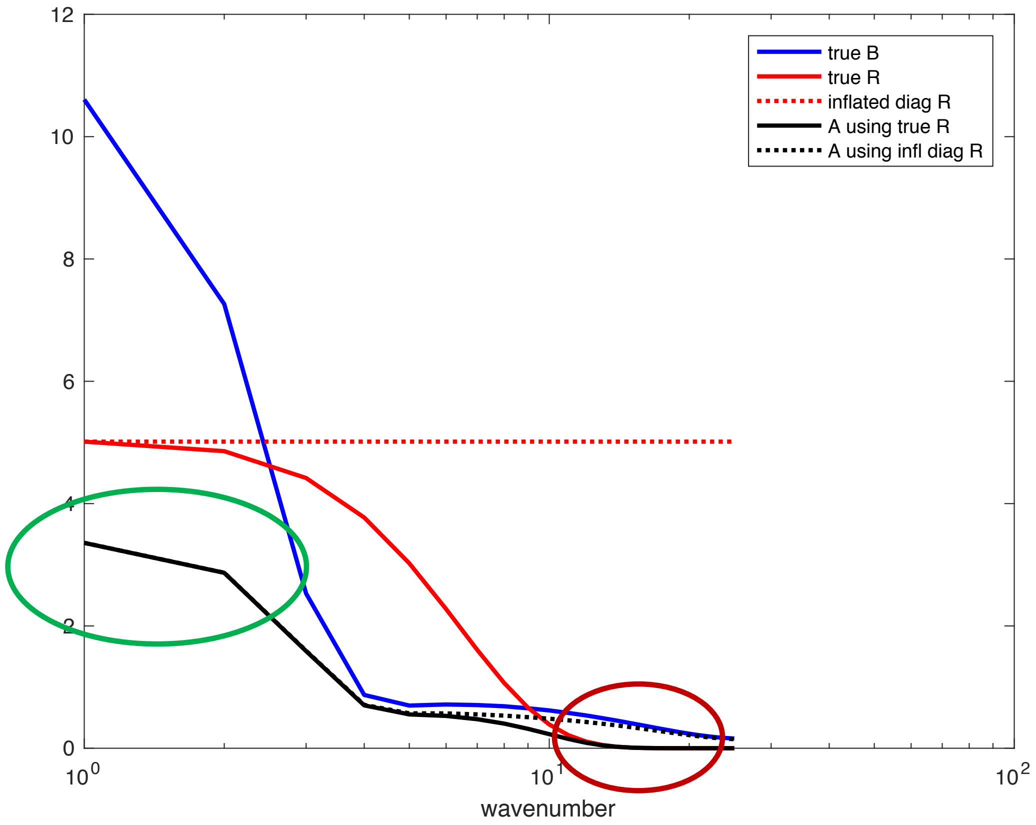
Diagonal of Kalman gain in spectral space



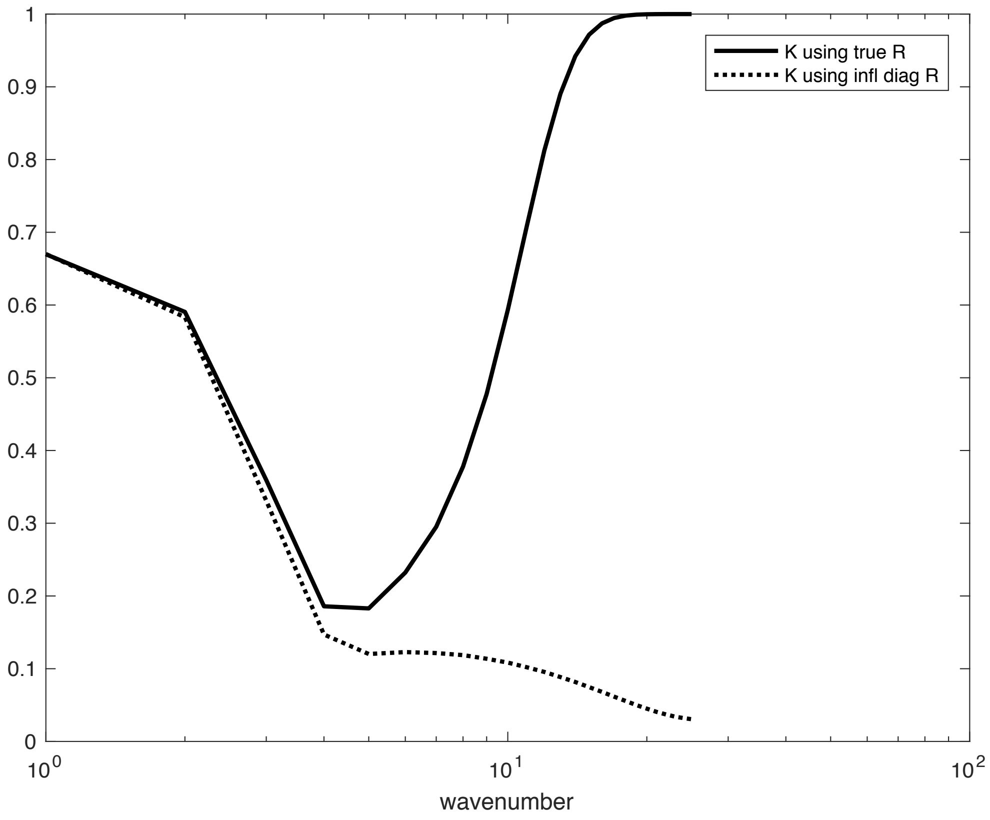
Using inflated diagonal R in assimilation (inflation factor chosen to fit largest scale perfectly)

Correlated observation errors & true B

Background, observation error and error in the analysis mean spectral variances



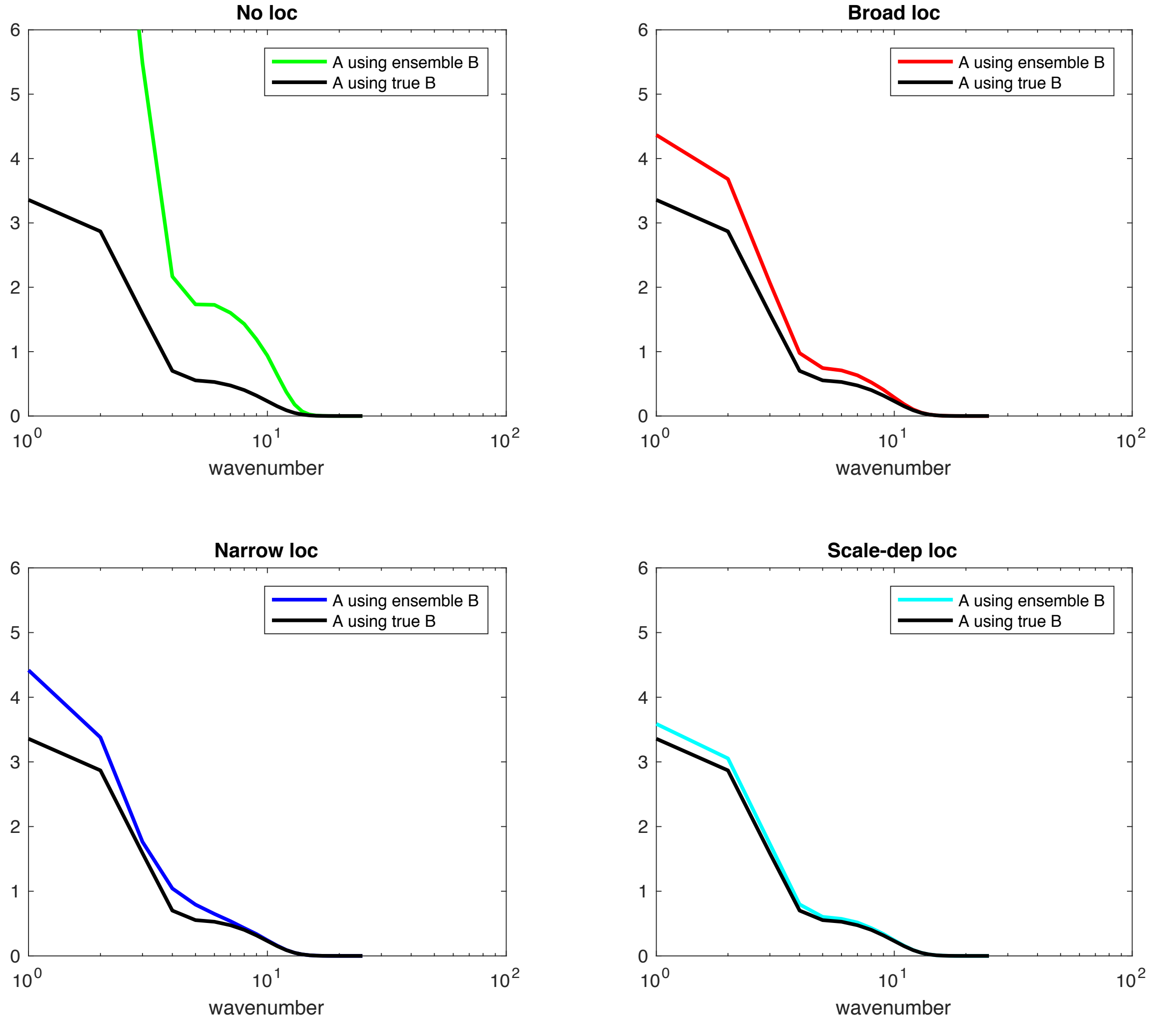
Diagonal of Kalman gain in spectral space



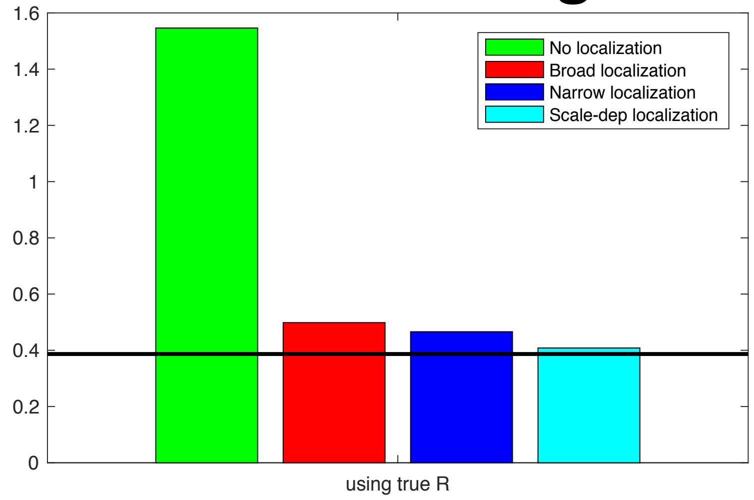
Using inflated diagonal R in assimilation:
good results for large scales, but small scales are suboptimal, worse than without inflation

Correlated observation errors & ensemble B (using true R in assimilation)

Spectral space analysis error variances



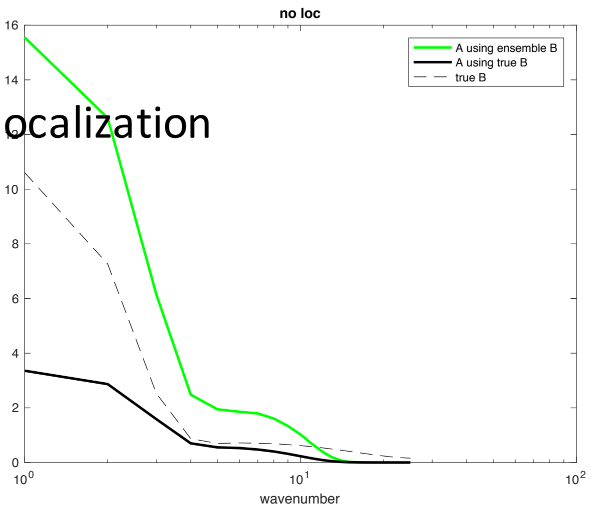
- No localization: very high errors
- Single-scale localization: high errors for large scales
- Best results with scale-dependent localization. Mean-square analysis errors (mean of A variances in grid space):



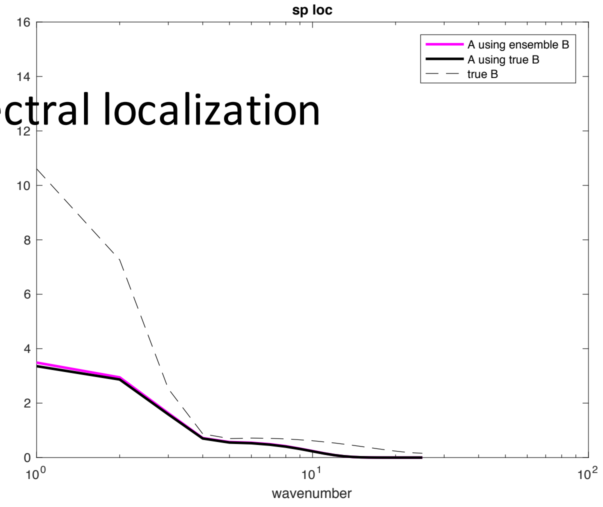
Why is analysis error so high when using correlated R and non-localized B?

Spectral space analysis error variances

No localization



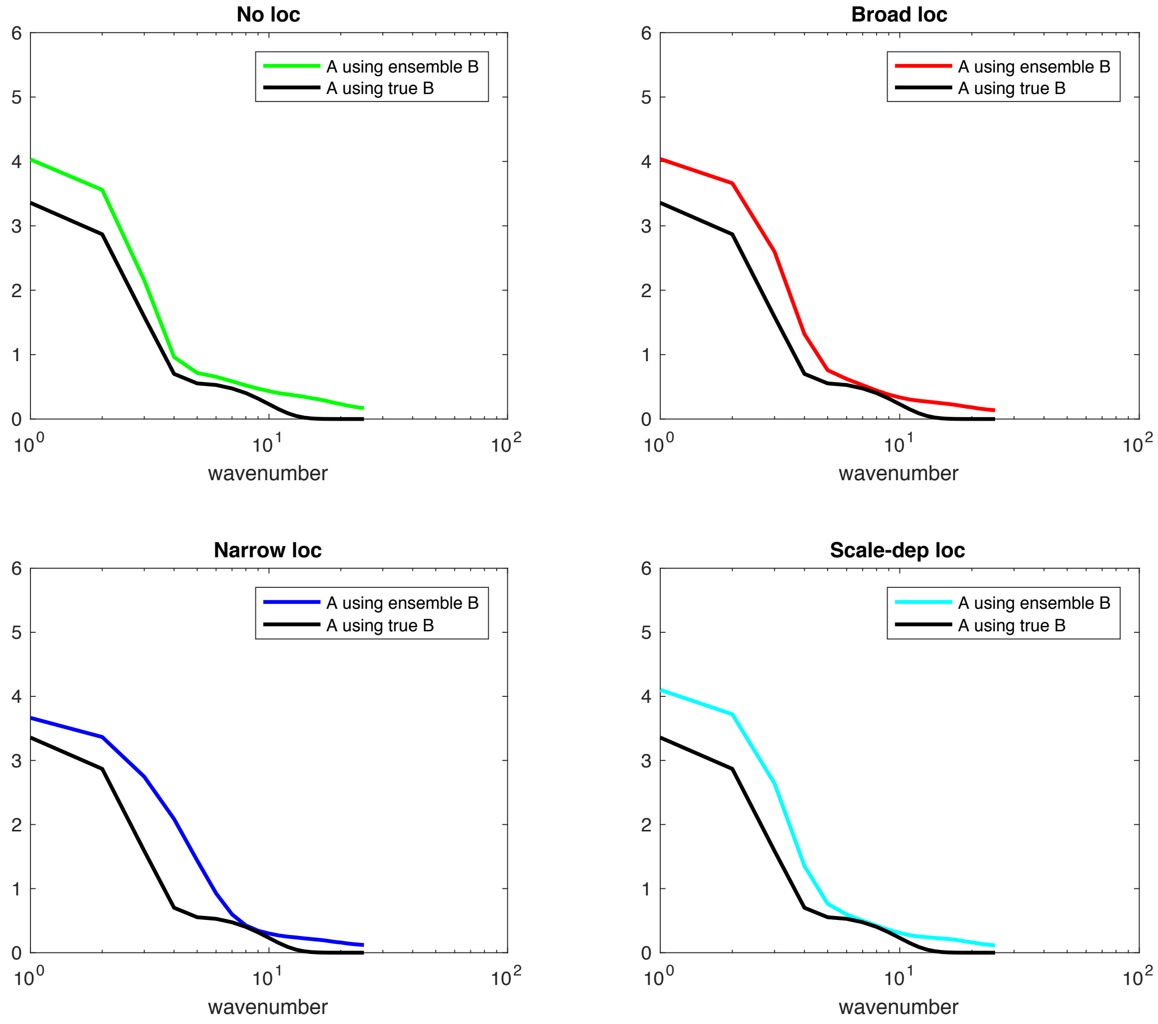
Spectral localization



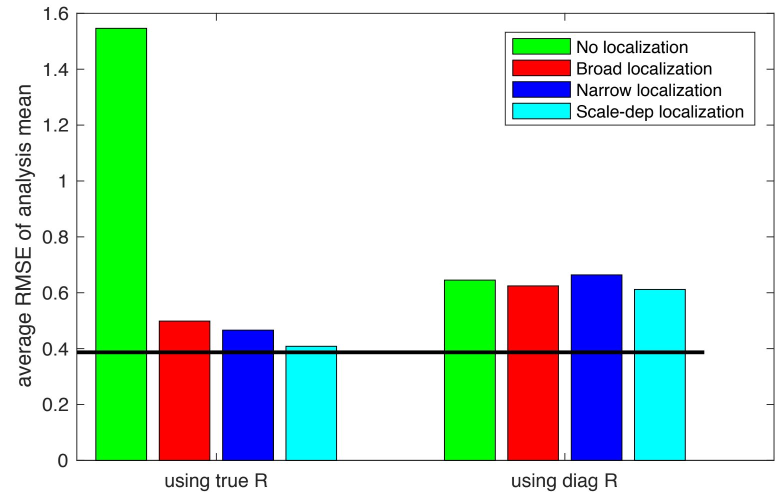
- Mostly because of the spurious cross-scale correlations (the error is close to optimal when removing all cross-scale correlations with spectral localization)
- Using correlated R leads to strong update of small scales
- Spurious cross-scale correlations lead to spurious update of large scales, but observations have significant errors in large scales

Correlated observation errors & ensemble B (using diagonal R in assimilation)

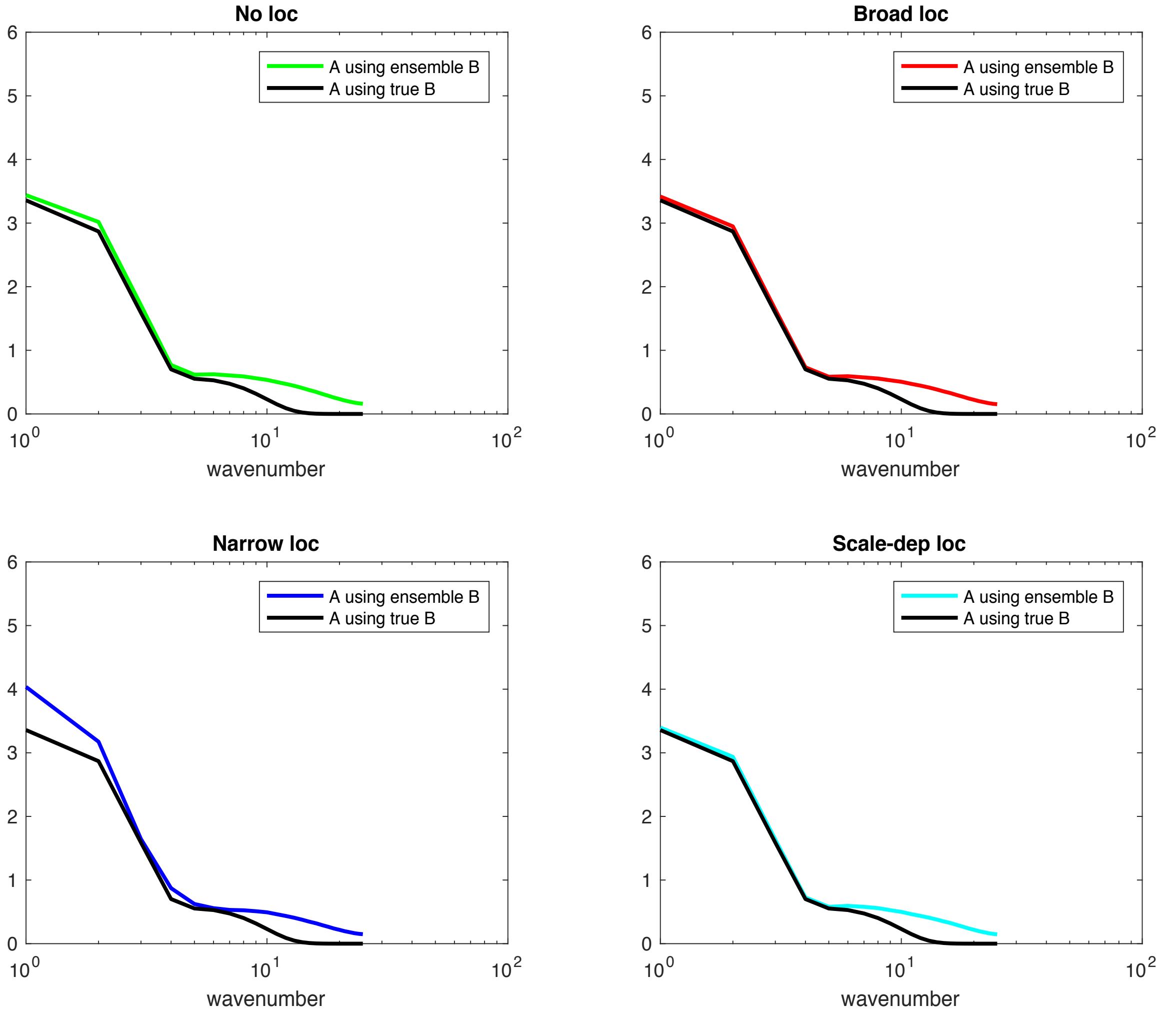
Spectral space analysis error variances



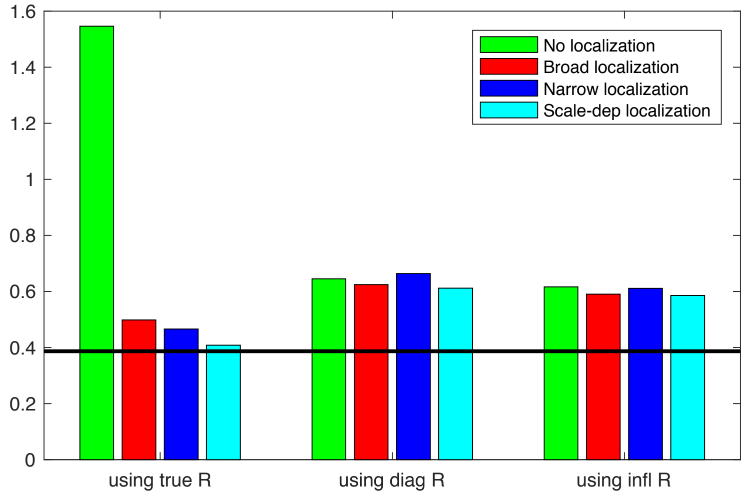
- All localizations perform similar: suboptimal both for large and small scales
- Higher mean square analysis errors than when using true R (for localized B):



Correlated observation errors & ensemble B (using inflated diagonal R in assimilation)



- Narrow localization: high error for large scales; all the rest: better errors at large scales than when using non-inflated errors
- Mean square analysis errors better at locations dominated by large and worse at locations dominated by small scales:



Conclusions

- Observation errors:
  - When observation errors are correlated, this means they have relatively less uncertainty at small scales and therefore small scales can be corrected by assimilation
  - Ignoring correlations leads to higher analysis errors both for large and small scales
  - Using inflated variances helps to improve analysis for large scales, but is even more suboptimal for small scales
- Background error localization:
  - Narrow localization in presence of some large scale errors in the background generally leads to high analysis errors for large scales
  - Scale-dependent localization generally gives better results than single-scale localization, especially when assimilating observations with correlated errors
  - Localization seems to be even more important when assimilating observations with correlated errors due to uneven distribution of uncertainty across scales
- These conclusions may depend on assumption of a fully observed system with H=I