Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2018-132-RC3, 2019

Ocean Science
Discussions

Interactive comment on "A multiscale ocean data assimilation approach combining spatial and spectral localisation" *by* Ann-Sophie Tissier et al.

Lemieux-Dudon (Referee)

benedicte.lemieux@cmcc.it

Received and published: 16 January 2019

This article focuses on the ensemble data assimilation systems which use the domain localization technique to prevent the contamination of the analysis with spurious long-range correlations due to ensemble error covariance matrix sampling errors. The authors propose a new data assimilation algorithm to solve the well-known drawbacks associated to the traditional localization techniques. Localization requires to define a decorrelation length (or radius of influence) which is known to be scale dependent. As a result a single decorrelation length cannot be suited for the wide range of scales represented in models and observations. Localization can also create noise especially when local analyses are superimposed. It also discards the true long-distance correlations (creating imbalance and loss of relevant information). This paper specifically

address this later issue. The objective of the authors is to improve the analysis of the large scales without worsening the small scale components of the signal in order to use the full constraint of the observing system at all scales. To separate the small and large scales, the authors apply a spectral transformation based on a spherical harmonics decomposition which enables to carry out two successive analyses : i) a first analysis is performed in the spectral domain with a spectral localization of the large scales components of the signal, ii) a second analysis is performed in the spatial domain with the traditional domain localization applied over the remaining smaller scales of the residual components of the signal. This approach called "multiscale observational update algorithm" is said to be computationally affordable because it avoids the small scale spectral transformation by using the multi-scale filtering technique.

This technique is tested in the framework of a twin experiment using synthetic satellite altimetry observations over a realistic regional configuration (ie, North Atlantic and Nordic Seas) at resolution of a quarter of degree. A prior ensemble of 70 members is simulated with NEMO v3.6 and the multi-scale analysis is implemented in the SAM2 data assimilation system (Mercator Ocean).

This article deserves to be published with minor changes. The state of art is properly introduced, the methodology is relevant and precisely described (eg, twin experiment), the results are commented and illustrated with numerous figures. Several verification methods are used to validate the proposed data assimilation algorithm (rank histograms, RMSE, ensemble spread before and after the analysis,..). Figure 7 is especially convincing in showing that the multi-scale analysis algorithm can improve the update of the large scale components without downgrading the small scales.

The objective of the authors appears to be reached when using synthetic SSH observations in a realistic regional configuration. I have few comments and questions which are listed below. The particular points for which I would recommend clarifications concern questions related to the transformation of the observations and the consistency analysis in the spectral space presented on Figure 10.

* p.2 line 28 : typography.

"compute removal of the between-scale covariances" : complete ?

* p.2 line 35 : methodology.

Can you explain the effect of neglecting the in between scale covariance ? (see related questions p.9 line 8).

* p.3 line 20 : suggestion.

This information was given several times and could be removed here : "This combines spectral localization for the large scales with spatial localization for the small scales"

* p.4 line 14 : typography.

uncertainties repeated twice

* p.4 line 29 : clarification appreciated.

The ensemble spread in the Gulf stream and Siberian Sea are too small/too large, respectively : "these characteristics do not affect the evaluation of the multiscale algorithm that is performed in this study". Can you explain why ?

* p.5 line 15 : language ?

"which is very close to whose used by the Ensemble Transform Kalman Filter (ETKF)"

* p.6 line 25 : suggested clarification.

"For a multivariate three-dimensional variable, this transformation can be applied to each vertical level of each model variable". Which physical parameter do you control and update? This question could be clarified when introducing the SEEK p.5.

* p.6 line 28 : minor comment.

"The reversible spectral transformation preserves the information for all degrees $l \leq l$ max" Information about lmax unclear at that stage : the spherical harmonics decompo-

sition (ST) is not fully introduced yet.

* p.6 line 30 and p.7 line 1 : suggestion when referring to Figure 3a, 3b

You could explicitly mention that you are showing the f_lm coefficients of the spherical harmonics decomposition.

* p.7 line 7 : correction.

In Equation (3) (the inverse ST) since part of the original signal f(theta,phy) is very likely filtered out by applying the cut-off degrees lmin and lmax, I would suggest that you replace the name of the function f.

* p.7 line 16 : clarification.

Could you explain further what is the "scale separator" your refer to?

* p.8 line 1 : clarification / correction

The least square problem of equations (4) and (5) is set to identify the best spherical harmonic coefficients (ie, f_lm) so that the reconstructed observation field f(theta, phi) (here ssh) taken at the observation locations (ie, theta_k, phi_k) minimizes the distance to the observations (ie, fo_k) with possibly an additional regularization term (minimal norm) and/or some bogus observations. You mention that the least square problem is only used for observations which are not on a regular grid (else you would apply the ST directly). It is said that "p is the number of grid points of the domain" in equation (4). It seems to me that p is rather the observation space dimension.

* p.9 line 8 : clarification.

This comment is connected to my previous question about the impact of neglecting the in between scale covariance "p.2 line 35". You are mentioning here that the "correlations between very different scales are weak and should be neglected by the data assimilation scheme and reduced to zero". I am not fully convinced about this statement when looking at Figure 5a and 5b. You illustrate your statement showing the cross-scales correlations for two "small" degree parameters (ie, large spatial scales) with the other spectral scales of the system (l in [0,60]). The cross-scales correlations "between very different scales" do not seem to be all negligible. Could you explain if these figures are meant to justify : i) the spectral scale separation algorithm (correlations between large and smaller spatial scales are said to be weak), ii) the need of a spectral localization for the large scales. What is the link with the next statement "For the same reasons, each localization window will contain a number of degrees of freedom sufficiently low to be controlled with an ensemble of moderate size"?

* p.11 line 28 : suggestion for Figure 6a, 6b and 6c

You mention that "Figures 6a and 6b show respectively the large scale of the mean ensemble of analysis increments obtained respectively with spatial localization or spectral localization (see Sect. 4.1.1). Hence, they have to be as similar as possible to the large scale part of the true anomaly showed on Fig. 4c". Comparison might be easier if you would directly plot the difference between the true anomaly and mean ensemble of analysis increments.

* p.13 line 29 : clarification

Regarding the results of Figure 8, did you apply zero bogus observations ?

* p.13 line 29: suggestion for Figure 8a and 8b.

Comparison might be easier for the reader if you directly plot the difference between the increment of analyses for the full spectrum and the True anomaly.

* p.15 line 8 : clarification needed for the rank histogram in the spectral space.

You mention that "Ranks maps in the spectral space provide additional indication that all algorithms provides reliable updated ensembles". You also say that "Figure 10 shows the maps of ranks in the spectral space for" the "ensemble". Rank histograms usually present the following information (similarly to Figure 9) along the x and y-axis: 1) the N+1 ranks extracted from the N ensemble members (ie, the N+1 probability

bins), 2) the counts of the verifying observation in each bins. On the "ranks map" of Figure 10, I cannot understand where are the counts of the synthetic observation used for verification (ie, the "true state") ? On such plot, it seems to me that one dimension is missing to measure the statistical reliability of the ensemble with respect to the verifying observation (ie, the counts). Is Figure 10 equivalent in a way to plotting several rank histograms for several spectral range ? What type of features you expect to see to confirm the statistical reliability of the ensemble (prior and updates) with such map of ranks ? Where are the counts ?

Finally, as stated in the other reviews there are small language issues which you should try to correct (I cannot help on that topic).

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-132, 2018.