Ocean Science

Discussions

EGU

Open Access

Interactive comment on "Data assimilation of sea surface temperature and salinity using basin-scale EOF reconstruction: a feasibility study in the NE Baltic Sea" *by* Mihhail Zujev et al.

Anonymous Referee #2

Received and published: 15 July 2020

General comments:

The manuscript describes an unusual data assimilation (DA) method which employs Empirical Orthogonal Functions (EOFs) to correct only the large-scale patterns of an ocean model for the northeastern Baltic proper. A training dataset of five years of model data was used to calculate the EOF modes. Only sea surface temperature (SST) and sea surface salinity (SSS) are considered, and the method relies on observations from a time window of up to 30 days centred around the analysis time. The authors found that the DA method is feasible to use for assimilation of SST and SSS and that it is computationally efficient.

I think the authors have made an interesting investigation of the current setup using the so-called HBM ocean model and the proposed DA technique, and I recommend the paper be published after some corrections.

General remarks:

The manuscript is well structured but not always so easy to read. I recommend a language check by a native speaker, if possible.

It is not clear whether the validation dataset was independent from the observations used in the DA process. On lines 239-240 it is stated that all observational data were used in the DA, but on lines 427 etc. it seems half of the gridded observations were reserved for validation. Please explain this more carefully. Also, even if every second gridded cell of observations were kept for validation, it is not clear to me that these observations are truly independent from the ones used in the DA, as they all originate from the same lines of FerryBox data. Is it possible to use data from certain ships for data assimilation, and use data from other ships for validation? Or is it possible to reserve the ICES data for validation, and just use the FerryBox data in the DA? Finally, the satellite-derived SST data in Figure 4 (d) is used to discuss the results in a qualitative way only; why not use it (together with similar satellite data) also for a statistical validation? In short, I would like to see a more careful validation using truly independent observations.

The DA method relies on the use of EOF reconstructions of SST and SST. Please show some examples of the reconstructions that were used in the DA.

It has been shown that the DA method works in a "reanalysis mode", in which observations "from the future" can be used in a time window up to 30 days wide, centred on the analysis date. I can see a problem when this method is used in forecast mode, where observations mainly from the past several days can be used. Is it enough to have a time window of, say, six hours? Please discuss this more.

Detailed, technical comments:

(I omit page numbers as the line numbers are unique)

l.8-9: "...based on covariance estimates from long..." (too many "the")

l.10; "...on a regular grid."

l.35: "...do not presently include DA..." (word order)

l.66: "Baltic proper" (should not be spelled with capital P; see also other occurrences in the manuscript)

l.130: "Two sets of compressed (averaged) FerryBox..." (for clarity); also on line 133.

l.134: "...it was chosen not to..." (word order)

l.142: "...data was too irregular..."

l. 146: "...time fixed..." How do you mean? Are they time-independent? Or just interpolated to pre-determined, fixed times, e.g. 00 UTC?

l.170: "observation operator Hi..."

l.179-181: "In practice, ..." I don't quite understand this sentence; please rewrite...

l.183: "...are not made at the same time." (simpler)

l.183: "...to cover a larger sea area..."

l.184: "...observation operator Hp..."

l.185: remove one instant of "that"

l.204: What happens if there is only one observation available?

l.215: "...without DA..." This puzzles me, as the analysis field depends on DA..?

l.224: "...can easily be included..." (word order)

C3

l.271: "...frequently became unrealistic."

l.274: "The time windows for experiments (a) and (b) were selected to be 10..."

l.317: "...revealed a deep... ...and a shallower..."

l.344: "...strips of lower salinity..."

l.395: "...when DA had decreased the FR temperature..." FR = free run; so DA cannot affect the temperature... I think you mean "decreased the temperature in the SST01 and SST02 datasets"...?

l.411: Is it the centred RMSD that is being used? Why not the usual RMSD? The centred RMSD is calculated after removing the bias; which do you mean?

l.418: "...so many observations."