Review of paper GMD-2022-241 by M. Beauchamp et al.

This paper presents an application of the 4DVarNet framework to SSH interpolation from nadir and wide-swath satellite altimetry. While the methodology and the results are clearly of interest, the paper can not be considered for publication under its present form. In particular, the explanation of the methodology is very difficult to follow because several of the mathematical quantities involved are not even defined. The quality of the english can be largely improved. Several typos are also present. Comments are detailed below:

Main Comments

- The background definition is not clearly put. It begins with "the state analysis x^as results in (from) a gradient-based minimization of the defined variational cost $J(x) = J_\phi(x, y, \Omega)$. None of these mathematical quantities is properly defined. Φ is said to be "a time-stepping operator associated with the dynamical model", which is far too vaue. It can be understood that x and y are temporal vectors of respectively the state of the system and the observations but this should be stated explicitly. Please be more rigorous in the description of the objectives at hand and the mathematical objects and tools you use.

- The paragraph linking the 4DVar formulation with the 3DVar and optimal interpolation is confusing. You should develop to make it clearer.

- Equation (4) that is supposed to be at the core of the 4DVarNet approach is not understandable under this form. The quantities g, h and c are not even defined. The reader can guess that $\mathcal{L}$ is the "LSTM" in the equation but otherwise, the explanation is not clear. You should take the time to explain properly what is done at each iteration of the algorithm while defining all the mathematical quantities involved. In the following paragraph, a solver Γ comes out of nowhere. How is it linked with equation (4)? The quantities of interest that seem to be the NN Φ and Γ are learned by minimizing a cost function that is given 3 pages later. The augmented framework that is described in section

3.2 involves two anomalies dx_1 and dx_2, neither of which are defined. Consequently, I think that section 3 should be completely rewritten to make it much more pedagogical as it is far from being understandable in its current form.

- Section 3.4 is also difficult to follow, especially how the patch are built. In can be read in the caption that "The spatial size of the patches is chosen to match the maximal distance with spatial autocorrelation of the SSH". Please develop and justify properly.

- Section 4.2 seems to explain how to build the training and test set but this should be stated explicitly.

- In the beginning of section 4.3, the training and evaluation settings are described. It is not clear to me why the test set is built upon older data with respect to the training set. Please justify.

- Concerning the evaluation metrics, the models obtaining lower values for λ_t and λ_x seem to be favored. You should report what these criteria are instead of just giving a reference. Also, it is not clear what the σ is: as the approach provides only a state estimate given the observations, I do not know where this σ comes from.

- Section 5.4 proposes a way to estimate uncertainties from several learnings of the 4DVarNet. It may be seen as a bootstrap procedure and therefore provides an estimate of the uncertainty of the prediction offered by 4DVarNet but not of the phenomenon itself. You should make a clear distinction between both.

Minor comments

There are lots of typos. Here I list those I have found:

- in the abstract: "SWOT (Surface Ocean and Water Topography)" reads as SOWT

- p.1 l.22: SSH is not defined

- p.2: the last sentence of the first paragraph is not understandable

- p.2 l.36: "4DVarNet" instead of "4VarNet"

- p.2 l.47: "We believe these contributions to contribute...", please rephrase

- p.2 l.59: the observation operator is "potentially trainable". Trainable means that a parametric representation of this operator exists whose parameters can be estimated. Otherwise, you should use "learnable"

- p.2 l.75: "optimal formulation (OI)" -> interpolation ?
- p.2 l.78: "framework", without the "s"
- p.2 l.81: "smoothing" instead of "smooting"
- p.5 l.118: "with" instead of "which"
- p.5 l.119: "libraries" rather than "framework"
- p.7 l.169: "...the their gradients..."
- p.13 l.259: "interpolaion"