Manuscript title: Deep Learning for Super-Resolution of Mediterranean Sea Surface Temperature Fields
Authored by: Claudia Fanelli, Daniele Ciani, Andrea Pisano, and Bruno Buongiorno Nardelli
Manuscript ID: https://doi.org/10.5194/egusphere-2024-455

REVIEWER #3

Reviewer: In their work the authors are tackling an important problem in geophysics, namely the reconstruction of high resolution fields of Sea Surface Temperature from partial high-resolution observations. Super-resolution is an ill posed problem, given that identical low-resolution fields can correspond to different high-resolution fields. The authors adapt their previous work on ADT to SST field reconstructions, as well as their well established knowledge of SST field reconstruction, showing improvements over the Mediterranean basin, within the confines of their experiment.

I really appreciated many parts of the article, notably the varying metrics and case studies to evaluate the quality of the reconstructions.

As it stands, I have some major and minor criticisms for the article that, should the authors address, would make for a significant contribution to the community.

Response: The authors would like to thank the anonymous reviewer for their interest in reading our manuscript. We think that the manuscript has been significantly improved by taking into account the reviewer's feedback. Please find the detailed responses below with the reference to the revisions appearing in the re-submitted files (highlighted in yellow).

Reviewer: The validation process is prone to data leakage. 4 days out of a year of data were omitted, but there is no mention of removing some days before or some days after in order to prevent data leakage. The physical reasoning of this is absent. Are the structures that decorelated after one day given the removal of the 200km smoothed field?

In general the L119 statement: "The test dataset is finally selected separating the 15% of the tiles available after 120 the preprocessing, chosen in order to be able to reconstruct the full geographical coverage of four days which are representative of different seasons." requires clarification. I read it as 4 individual days, one in each season. It could be understood as patches covering the whole area, spread out over each season. I would expect to have more of a cross-validation approach given the one year dataset limitation.

Response: Thank you for pointing out the lack of clarity and previous limitations regarding the choices made to construct the test dataset. Indeed, the initial test was carried out selecting all the tiles covering four individual days (one in each season), as clarified in line 132 of the revised version of the manuscript. However, we do agree about the limitations of that initial test dataset and thus set up a much more extended test including one full year of totally independent data. The new results are presented in the new Section 3.2.

Reviewer: Another major concern is the input. The input, presently, is only the first guess (removing a sliding window), and the information coming from the L3 satellite product is not used as a complimentary input. Why did the authors deprived themselves from potential additional input such as multiple time steps and L3 products? Other works (such as Archambault et al, Martin et al) in SSH fields have training procedures where some of the satellite information is omitted from the target in order to validate the approach.

Response: Thanks to the reviewer's r comment, we realized that we needed to better introduce the objectives and processing steps involved in our workflow . In fact, we are not dealing here with an algorithm designed for gappy-fields interpolation, but rather with an algorithm designed to improve one specific step in our processing chain, namely improving the resolution of the low-resolution gap-free field used as the background for the 1 km optimal interpolation implemented in our chain. As such, unlike the classical interpolation methods, in which it is important and useful to use ground-truth data and multiple time steps as input when available, the goal here is effectively only to learn how to reconstruct a single image at very high resolution starting from a low resolution one. L3 data are indeed e used as input in the final interpolation step, though, and are of course also used as target during the training phase. We have clarified all this in the Introduction of the revised version of the manuscript (lines 87-91).

Reviewer: There is no mention of how the total field reconstruction over the whole Mediterranean sea is output. If the image was made by recomposing a sliding window reconstruction that should be mentioned. Given that the network learns filters, one could conceivably apply them on the whole image, though I expect the attention layers to pose an issue.

Response: Following from previous comment/response, we hope to have clarified this point adding a more detailed explanation at the end of Section 2, in lines 208-212 of the revised version of the manuscript.

Reviewer: The choice of architecture, while documented, is not justified. Were the hyper-parameters optimized? Other architectures evaluated? The authors mention a lot of competing methods, but do not compare their architecture to them. (DINEOF, DINCAE, to cite but two) Are the computational and expertise cost justified versus other methods? The results seem to indicate a 0.02°C improvement; is the architecture stable through different initializations?

Response: As clarified in the revised text, we are not dealing here with a novel interpolation algorithm, but rather on a single-image end-to-end algorithm, that is applied to improve our background field used in a two-step optimal interpolation algorithm. As such, comparisons with different interpolation algorithms as DINCAE/DINEOF would be out of scope. On the other hand, we realized that we did not explain why we have chosen a specific configuration of the network, and we also understood that at least one additional test was needed. In fact, we did not develop a new architecture but we exploited the one developed by Buongiorno Nardelli et al. (2022) (https://doi.org/10.3390/rs14051159), and also aimed to reduce computational costs. As such, we originally relied on the choices made in that specific work. In the revised paper, however, we now describe one additional test carried out by modifying the network depth, i.e. reducing the number of Multiscale Adaptive Residual Blocks (called dADRSR/2 in the revised version of the manuscript) and

compare that also with the results obtained by applying other deep learning methods (i.e., the EDSR and the ADR networks) The assessment of the various network configurations considered are now shown in Table 1.

Reviewer: The method section (2.2) seems to assume unfamiliarity with neural networks, providing intuitive explanations for basic architectural blocks, but then very quickly skims over important details of the more complicated blocks of the architecture. This part would benefit reducing the initial explanation of activation functions and CNNs (such as the interpretation of lines 59 to 61 which is intuitive but could easily not correspond to the exact explanations provided given the non-linear activations) and expanding on the reasoning of the architectural choices (the adaptive part of the ARB is not discussed, implying the rest of the architecture is non adaptive).

Response: This has now been revised considering the reviewer suggestion. We clarified the adaptive strategy used in the network in lines 72-75 and lines 161-164 of the revised version of the manuscript.

Reviewer: No mention is made to VIT and diffusion-based super resolution techniques that have become state of the art in computer vision. I can understand the daunting nature of these, but should they not be mentioned as potential further steps, at least? The latter is especially significant: the field reconstructions obtained through optimizing RMSE favor smoothness, and often do not represent physically feasible oceanic states. Graphcast for example has been abandoned in favor of Gencast for that very reason. Given that the model is in NRT, and therefore would be used for constraining operational models, it might be interesting to at least think about this. It is even more important given the non-bijective nature of the problem.

Response: We agree with the reviewer that it is worth mentioning also other techniques that have gained a lot of interest lately, which we would like to explore in the future. Following their suggestion, we added a brief paragraph in the conclusion on the topic (lines 321-320 of the revised version of the manuscript).

Reviewer: Fig.1 would benefit from locating with a bounding box or three the patches on the right hand side.

Response: Thank you for your suggestion. We modified Figure 2 of the revised version of the manuscript with an example of a pair of tiles and the corresponding position over the Mediterranean Sea.

Reviewer: 182: max(I) in denormalized space i.e. K°? Or in the space where the large field is removed 200km? Is it computed over the patches, or the whole Mediterranean?

Response: All the errors (including the max(I) used to calculate the PSNR) are computed over the final reconstructed image over the whole Mediterranean Sea, as explained in lines 208-212 of the revised version of the manuscript.