Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-128-AC4, 2020

Geoscientific
Model Development
Discussions

Interactive comment on "DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations" *by* Alexander Barth et al.

Alexander Barth et al.

a.barth@uliege.be

Received and published: 25 February 2020

Reply to the review of DINCAE 1.0: a convolutional neural network with error estimates to recon- struct sea surface temperature satellite observation

We would like to thank the reviewer for carefully reading the manuscript. Her/His complete review is copied below while our answers are inserted below every comment and written in boldface.

This study presents a novel approach of reconstructing sea surface temperatures from cloudy satellite data by making good use of modern deep learning techniques. While I

believe that the study has been carefully designed and executed, I have severe problems with the paper in terms of its presentation and accuracy in writing. This study could become a high-impact publication if it were better structured and methods and outcomes were described clearer. I advise the authors to apply major revisions and seek the help of a native speaker to avoid erroneous or ambiguous statements. Below, please find my detailed comments.

All coauthors and I, tried our best to improve the manuscript to avoid erroneous or ambiguous statements. We hope that the manuscript is now clear.

Abstract: the sentence "However, it is unclear how to handle missing data (or data with variable accuracy) in a neural network when using incomplete satellite data in the training phase." is not very clear. Perhaps rephrase as "Contrary to standard image reconstruction with neural networks, this application requires a method to handle missing data (or data with variable accuracy)."

OK, we changed this sentence to:

> **Contrary to standard image reconstruction with neural networks, this application requires a method to handle missing data (or data with variable accuracy) already in the training phase.**

L7: suggest to remove "essentially"

Ok, done.

L9: what is "relatively long"? Provide a number, please.

We agree, and we changed the sentence to:

> **The approach, called DINCAE (Data-Interpolating Convolutional Auto-Encoder) is applied to a 25-year time-series of Advanced Very High Resolution Radiometer (AVHRR) sea surface temperature data**

L11: "a method to reconstruct missing data": suggest to rephrase "a previously published method", "the current standard method", "the state-of-the-art DINEOF method", or similar.

Thank you for the suggestion. In the revised manuscript we refer to the methods as "state-of-the-art".

L16: what is meant by "the ocean current signal"? A signal always refers to a measurement or sensing process. Here you want to refer to a physical process in the ocean.

We used the term signal indeed quite broadly in the original manuscript and replaced it in the revised manuscript with a more precise term. In the revised manuscript, this was changed to "the ocean velocity variability depends thus partially on ocean temperature".

L17: replace "like" with "e.g."

OK, done.

L20: replace "sensor" with "measurement". This sentence refers to the measurement principle, not the technical instrument, which performs the measurement.

The submitted manuscript reads:

> **However, as for any sensor working in the infrared or visible bands, clouds often obscure large parts of the field-of-view.**

In the revised manuscript we changed sensor by "measuring technique".

L22: "but often small scale information is filtered out because of the transient and stochastic nature of these structures." The transient and stochastic nature produces variability and is clearly not the reason why small scale information is filtered out. This is rather a result of the averaging procedures which are applied in practically all known techniques to interpolate. As this is a critical sentence for setting the stage of this study, please rethink the phrasing and provide a more precise description of the issue, which you are trying to solve.

We agree. The following has been added to the manuscript:

> **A truncated EOF decomposition will focus primarily in spatial structures with a "strong" signature (or more formally defined with a significant L2 norm compared to the total variance). Small scale structures can be included in a truncated EOF decomposition as long as their related variance is large enough to be present in the retained EOF modes. But small scale structures tend to be transient (short-lived) and therefore are often not retained in the dominant EOF modes. It should be noted that there is no explicit spatial filtering scale in DINEOF removing small scales (unlike other methods like optimal interpolation, kriging, spline interpolation). But in practice a similar smoothing effect is noticed because of the EOF truncation.**

We removed the terms "transient" and "stochastic" in the revised manuscript and we clarified that there is no explicit filtering in DINEOF by using a predefined spatial length-scale.

L23: DINEOF falls from the sky here. For non-experts in the field of sea surface temperature reconstructions, it is completely unclear what this is. Also, as DINEOF appears here for the first time, it is a must that you provide a reference. The reference comes two sentences further down, which is too late. A brief description of the method would be appropriate here.

The original manuscript was:

> **DINEOF (Data Interpolating Empirical Orthogonal Functions), provides an accurate way of retrieving missing data and reducing noise in satellite datasets using a set of optimal EOFs. The optimal number of EOF is determined by cross-validation. More information on the DINEOF approach is documented in (Beckers and Rixen, 2003;Alvera-Azcárate et al., 2005).**

The reference was now put in the first sentence. The following information was added when describing DINEOF. More information has been added later in DINEOF sections, because it would be too technical for the introduction.

> **DINEOF (Data Interpolating Empirical Orthogonal Functions, Beckers and Rixen, 2003; Alvera-Azcárate et al., 2005), is an iterative method to reconstruct missing observations reducing noise in satellite datasets using empirical orthogonal functions (EOF). A truncated EOF decomposition using the leading EOFs is performed and the initially missing data are reconstructed using this EOF decomposition. The EOF decomposition and reconstruction is repeated until convergence.**

page 2: L9: "to detect the presence of non-linear, stochastic features" - I disagree that neural networks "detect" these features. Rather they are able to "learn" such features and thus potentially produce more detailed reconstructions of them.

OK, this is changed the revised manuscript:

> **Neural networks are therefore specially well positioned to learn non-linear, stochastic features measured at the sea surface by satellite sensors, and their use might prove efficient in retaining these structures when analysing satellite data, for example for reconstructing missing data.**

L11: This statement is too general. I strongly suggest to first explain briefly what types of neural networks exist and how they can/could be used for the problem you want to solve (including references to the most important deep learning papers). Then you can make the argument that these networks (and in particular CNN derivates) are generally trained with complete data, whereas in your application you need to find a method which can train with scenes containing missing data, because there are no complete satellite scenes available (or only very few).

We added the following information to the manuscript to give a general overview of the types of neural networks. Later in the manuscript we will focus on some examples in oceanography.

> **Neural networks can be composed of a wide variety of building blocks, such as fully connected layers (Rosenblatt, 1958; Widrow and Hoff, 1962) recurrent networks (e.g. Long Short-term Memory (Hochreiter and Schmidhuber, 1997), Gated recurrent unit (Cho et al., 2014)), convolutional layers (LeCun et al., 1998; Krizhevsky et al., 2012). Recurrent networks work typically with a one dimensional list of inputs of a variable length (such as a text sentence). Fully connected layers and convolutional layers require to have a full dataset without missing**

> **data, at least for the training phase. For a review on neural networks the reader is referred to Schmidhuber (2015) and references therein.**

L14: this paragraph contains some of the literature review I am asking for in my previous comment. However, here it is on the one hand too specific (only ocean data applications), but on the other hand too superficial as it doesn't become clear why you need to develop a new approach and cannot simply apply for example the method of Krasnopolsky et al.

In the revised manuscript we first referenced the general idea in artificial neural networks and then give some examples in oceanography. The field is too wide to give a comprehensive overview of all applications in geoscience and we limit therefore the applications to oceanography. The updated manuscript goes into more details on the limitations of the method of Krasnopolsky et al. (2016):

> **The neural network by Krasnopolsky et al. (2016) uses as input satellite sea surface elevation, sea surface salinity, sea surface temperature and *in situ* Argo salinity and temperature vertical profiles with some auxiliary information (like longitude, latitude and time) to estimate the Chlorophyll-a concentration. The network does not use measured Chlorophyll-a concentration at a given location as input during inference (the reconstruction phase), nor the information from nearby grid points to infer Chlorophyll-a concentration. The network is exposed to the chlorophyll measurements only during the training phase.**

L33: I storngly suggest to re-organize the paper so that it follows the classical structure and describes the method before the data, and in particular before another reconstruction (DINEOF) is reported.

For us it was important to describe the data first because the ubiquity of clouds and the strong seasonal cycle in the data sets are an important constraint for the method. Therefore we believe that for most readers it will be easier to understand the method once its input (i.e. the satellite data) is presented. In fact, the type of data motivates the choices that were made during the design of the method. Also, we believe it helps the reader if we give concrete size of the arrays and matrices involved when the method is described. However, these matrix sizes depend on the data. Since the method section depends heavily on the data section, and the data section does not depend on the method section, we choose to present the data first. We actually tried to reverse the order as suggested by the reviewer but we ended up with too many forward references which would harm the readability.

We however presented the DINEOF method after the DINCAE method as suggested in the revised manuscript.

Page 3: L4: Delete the first sentence. You don't need a motivation within your "Data" section. Such an argument belongs in the introduction, if you wish to explain, for example, why you designed your study based on this dataset and not another one. In section 2 you should only describe the dataset, without any "discussion".

We deleted the first sentence.

L19: The cross-validation deserves more explanation, because you are publishing in a journal which is primarily read by non-experts in the field of machine learning. It is important to note that (contrary to standard image analysis) subsequent scenes from the AVHRR data are not independent. Therefore you cannot use a random sample to construct your test dataset (or "validation dataset", whichever terminology you prefer). I am wondering if 50 scenes are indeed sufficient to thoroughly test the generalization of the network. Not being an ocean scientist, I can only assume that typical transport time scales in your study region are on the order of a week(?). This would imply that the first 7 scenes of your "independent" test data are still "polluted" and thus not fully independent. Have you tried retaining a larger test sample?

In the revised manuscript we also added a reference to a classical textbook for the cross-validation method outside of the realm of machine learning.

> **To assess the accuracy of the reconstruction method, cross-validation is used wilks1995. For cross-validation a subset of the data is withheld from the analysis and the final reconstruction is compared to the withheld dataset to access its accuracy. Since clouds have a spatial extent, we wanted to withhold data with a similar spatial structure. In the last 50 images we removed data according to the cloud mask of the first 50 images of the SST time series. The last 50 images represent the data from 2009-09-25 to 2009-12-27 (since some scenes with too few data have been dropped as mentioned before). These data are not used at all during either the training or the reconstruction phases, and can therefore be considered independent. In total, 106 816 measurements (i.e. individual pixels) have been withheld this way.**

We did not try to have a temporal gap between the training data and test data or a larger test sample. It is true that there is some correlation between the training data and test data (the last few scenes used for learning might correlate with the first few scenes of the validation set). But we also performed a validation with in situ data in the manuscript. Both validation methods lead to the same outcome. We realize by reading the other comments from the reviewer that the reference to the table with in situ validation was not quite clear in the original manuscript.

For the "best" DINCAE experiment, we also recomputed the cross-validation

RMS error using only the last 43 scences and the RMS error is with 0.3754 °C very similar (and even slightly lower) that the RMS error using the last 50 images 0.3834°C. If there would be a significant "pollution" effect, then one would expect that the RMS error with 43 scence to be larger than with the 50 scences. But this was not observed. Given the large pool of training data (5216 scenes) the effect of a handful potentially correlated scenes does not have any significant effect.

L21: How can you retain > 100,000 measurements from 50 scenes? Are these individual pixels, or did you actually apply cross-validation with random sampling, thus ignoring the argument I made above?

Yes, this is the count of individual pixels of the 50 images used for cross-validation as measured by the AVHRR sensor. We clarified this in the revised manuscript.

Page 4: Figure 1: I cannot see any arrows, which are referenced in the figure caption.

We are sorry about this problem. It has been corrected in the revised manuscript.

L3: Again, some explanation of DINEOF is warranted in this paper. It should be clear what this method does, without having to access the referenced papers. For details you can refer to them, but not for the fundamental "explanation".

We expand this section and included more information about how DINEOF reconstructed the missing data:

> **A truncated EOF decomposition using the leading N EOFs is performed and the initially missing data are reconstructed by combining the retained EOF modes and their corresponding amplitudes. The**

EOF decomposition and reconstruction of missing data is repeated until convergence. The optimal number of EOFs N is determined by cross-validation.

Page 5: Table 1: instead of "fewer layers" or "more layers", the number of layers should be given.

We agree, and this has been added to the table ("fewer layers": 3 convolutional layers and "more layers": 5 convolutional layers)

L5: I don't think this is an appropriate citation here. It is the principle of EOFs to detect relations between variables and construct an orthogonal set of linear functions to model these relations. Due to the cutoff after N EOFs, there is always smoothing applied. A proper citation here would be some standard statistics book.

We added a reference to Wilks (1995).

L7: I am not at all surprised by this result: if you extend the timeseries, you will be more likely to sample patterns, which have not been observed before and which don't fit well to the already "learned" EOFs. Hence, there is less structure that can be described by the EOFs and more noise

Original manuscript reads:

> **As only 13 modes are retrained by DINEOF for the reconstruction, some small scale structures are smoothed-out, which is a well known property of a truncated EOF decomposition (e.g. Alvera-Azcárate et al., 2009). This smoothing effect results in an RMS error of 0.3864°C**

> **when comparing the reconstructed dataset to all the initially present SST (i.e. used for the reconstruction). A somewhat surprising result is that when using less data (only from the last two years, i.e. 2008 to 2009), 19 EOFs modes are retained, leading to a reconstruction with richer structures.**

Here we did not extend the time series but just used a subsample of the time series. For the full time series (1985-2009, 25 years), 13 modes have been retained as optimal. For a subset (2008-2009, 2 years), more EOFs (19 modes) have been retained. The number of time instances is an upper bound for the number of EOFs with non-zero singular values. For a shorter time series, this upper bound is thus lower, yet more EOFs have been retained with the shorter time series. This result was unexpected for us.

L14: I disagree that deep neural networks are "extensively" used in Earth sciences. This field is developing rapdily, but the applications are so far far from "extensive".

We agree and the sentence was revised:

> **Convolutional and other deep neural networks are extensively used in computer vision and find an increasing number of applications in Earth sciences [...]**

Page 6: L6: what does "different errors" mean? Different to what?

The error can be different from one pixel to another. We changed this in the revised manuscript as "error varying in space and/or time" to be more clear.

Also: as mentioned before, for an article in this journal there needs to be a brief description of CAEs in the method section, which should contain enough information that an uninitiated reader (e.g. an ocean modeller) understands why the approach might actually work. Clearly, before the discussion taking place here, the reader must know how the network is constructed, which activation functions are used, which optimizer is used, whether regularization techniques are applied, etc. And the basis for the network built here is probably coming from some deep learning paper, which then needs to be cited. Such a description follows on page 7. Please restructure.

This is the structure of the original manuscript:

On page 6:

- **We explain how missing data is handled in data assimilation which motivates the present work**
- **Handling of missing data in the input is done in analogy here**
- **Describe the input of the neural network**

On page 7-8:

- **General structure of the network**
- **Skip connections**
- **activation function**

On page 9:

- **Cost function**

On page 10:

- **Optimizer**
- **regularization techniques**

To us, it seems logical and didactical to make a description of the method step-by-step: first the input of the network, then how the input is transformed by the network (general structure, activation function, skip connections), how to assess the accuracy of the output (cost function), how to optimize the accuracy and finally how to prevent overfitting (regularization techniques). In the revised manuscript we make it clear that the description of the convolutional autoencoder will come in the following.

In the revised manuscript, we put the reference to Hinton and Salakhutdinov (2006); Ronneberger et al. (2015) more prominently as the neural network structure proposed in those papers is quite similar to one used here. We also updated the overview of the paper in the introduction to make the links between sections clearer.

Page 7: L7: the references refer to convolutional layers only. However, you are applying an autoencoder approach, so the appropriate references should be made. For example: G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504, 2006.

Note that we cite Hinton and Salakhutdinov (2006) on page 2 in the introduction of the original manuscript:

> **An auto-encoder is a particular type of network which can compress and decompress the information in an input dataset (Hinton and**

Salakhutdinov, 2006), effectively reducing the dimensionality in the input data.

To put the reference of Hinton and Salakhutdinov (2006) more prominently in the manuscript we changed on page 7, line 7 of the revised manuscript:

> **The main building blocks of the neural network (Table 2) are convolutional layers (LeCun et al., 1998; Krizhevsky et al., 2012).**

was changed to:

> **The overall structure of the neural network (Table 2) is a convolutional autoencoder (Hinton and Salakhutdinov, 2006; Ronneberger et al., 2015). Its main building blocks are convolutional layers (LeCun et al., 1998; Krizhevsky et al., 2012).**

L9: "different number of filters" - not filter sizes. Your filter size is always 3x3 as you state-of-the-art just below.

Thank you for pointing this out. We corrected this and changed it throughout the manuscript.

L16: composed of

Thank you, we corrected this (and we found a similar error which is corrected too).

Page 8: Please also write down the loss function.

The cost function is "Equation 9" (page 9) from the original manuscript:

> **[...] The cost function has finally the following form:** $\mathbf{J}(\hat{y}_{ij}, \hat{\sigma}_{ij}) = \frac{1}{2N} \sum_{ij} \left[\left(\frac{y_{ij} - \hat{y}_{ij}}{\hat{\sigma}_{ij}} \right)^2 + \log(\hat{\sigma}_{ij}^2) + 2 \log(\sqrt{2\pi}) \right]$

We do not make a distinction between "cost function" and "loss function" and use it as synonyms as in Goodfellow et al. (2016):

> **The function we want to minimize or maximize is called the objective function, or criterion. When we are minimizing it, we may also call it the cost function, loss function, or error function. (page 80, of Goodfellow et al. (2016))**

The loss function per individual scalar sample is the term in brackets of equations 9. This information has been added to the manuscript.

Page 9: L2: So here it appears that indeed your "independent" validation data are not independent.

This is line 2, page 9 for the original manuscript:

> **The input data set is randomly shuffled (over the time dimension) and partitioned into so-called mini-batches of 50 images, as an array of the size 8 x 112 x 112 x 50.**

It is unclear to us why the reviewer thinks that the data is not independent. The data marked for cross-validation is not used during the training. It is just a coincidence that the mini-batch size is equal to the number of images used for cross-validation. These numbers are not related.

L9: I don't understand the random masking: is one random mask applied to each image and then the same image used in each epoch? Or do you apply different random masks to the same image as a means to augment your data and increase generalization?

The later is the case, the paragraph has been revised to make this clear:

> **For every input image, more data points were masked (in addition to the cross-validation) by using a randomly chosen cloud mask during training. The cloud mask of a training image would thus be the union of the cloud mask of the input dataset and a randomly chosen cloud mask. This allows us to assess the capability of the network to recover missing data under clouds. Without the additional clouds, the neural network would simply learn to reproduce the SST values that are already received as input. At every epoch a different mask is applied to a given image to mitigate overfitting and aid generalization.**

L24: Remove the statement about other variables, because you focus exclusively on SST. This can go in the conclusions section, but is confusing here.

OK, done.

Page 10: L10: I am confused here as to how the reconstruction and the training works. Normally, you first train your network and then you reconstruct (in particular on unseen data). Then you cannot take any average between epochs 200 and 1000.

The revised paragraph now reads:

> **The neural network is updated using the gradient for every mini-batch during training and after every 10 epochs the current state of the neu-**

> **ral network is used to infer the missing data over the whole time series, and in particular reconstructing the missing data is the cross-validation dataset. But importantly, the network is not updated using the cross-validation data.**

So effectively, we temporarily suspend the training after every 10 epochs and reconstruct the missing data but then continue the training.

Page 13: L5: "[..] underestimate the actual error by 15underestimation of the expected error of this magnitude should be acceptable for most purposes." This sentence deserves further explanation. What is the generally accepted accuracy of the error estimate?

In the revised manuscript we avoided the term "acceptable" and added the following information:

> **An interpolation technique which is commonly used in operational context, is optimal interpolation. This technique is able to provide an expected error variance of the interpolated fields based on a series of assumptions, in particular that the errors are Gaussian distributed with a known covariance and zero mean. Given these assumptions, the error variance of the optimal interpolation algorithm is only found to be weakly related to the observed RMSE in a study of Pisano et al. (2016) using satellite sea-surface temperature in the Mediterranean Sea. In this context, the fact that DINCAE underestimates the actual error only by 15% on average can be seen as an improvement.**

Page 14: L4: again: not "filter sizes" but "number of filters"

Thank you, this is corrected. We corrected all 5 occurrences of this issue.

Page 15: Figure 5: the figure titles are misleading. Apparently you are always showing results for one specific day. This day should then be mentioned in the caption and nt as title on the first panel. The way the panels are labelled now suggests that you compare apples with oranges (which I don't believe you do). Also here and in Figure 6 it is not quite clear to me if the DINEOF reconstruction also had to deal with the "added clouds" or not. Higher up, when you mention the addition of random clouds it would be good to see a typical fraction of image size which is obstructed by these random clouds. From figures 5 and 6, this obstruction seems to be quite large.

We removed the dates from the first panel. It was already mentioned in the caption and the reviewer is right that the date is common to all panels of figure 5.

Initially, the averaged cloud coverage of the dataset is 46% (over all 25 years). The cloud coverage for the 50 last scenes is increased to 77% when the cross-validation points are excluded. It is true that a significant part of the scene is obscured (after marking the data for cross-validation), but in the Mediteranan Sea the cloud coverage is relatively low compared to the globally average cloud coverage which is 75% (Wylie et al., 2005). Removing some data for cross-validation makes the cloud coverage thus more similar to the global average.

Page 17: L4: where can the reader see the comparison between in-situ obs and the reconstructions? This paragraph remains qualitative and doesn't contribute anything meaningful.

The RMS errors can be seen in table 3, but the reference in the original manuscript was unfortunately not clear:

> **As expected, biases play now a more important role when comparing**

> **in situ observations with reconstructed satellite data (3).**

"(3)" has been changed to "(Table 3)". We apologize for this issue which could easily cause a reader to overlook this table.

L15: indeed - if the deep learning method was applied correctly (which is somewhat difficult to judge from this paper due to all the issues described in this review), then this is a very nice and important result, which shows the superiority of deep learning approaches with their ability to learn non-linear functions compared to standard statistical methods. This key result could probably be brought out even clearer.

Thank you for your encouragement! We hope that the revised manuscript based on the comments of all reviewers is now clearer.

Page 18: L2: why is this method "practical"? I assume it is, but this is only because of my background knowledge. This point needs to be made explicit somewhere in the paper. You list computation times for training, but you don't say how long it takes to reconstruct a scene once the network has been trained.

Reconstructing the data of all 25 years takes only 8 seconds on the GeForce GTX 1080 for a trained network, but training the network can take several hours as mentioned in the manuscript. The manuscript has been updated with the reconstruction time for a trained network.

In the revised manuscript we removed the term "practical" because it was not possible for us to give it a precise meaning.

Page 19: After rewriting other parts of the manuscript I suggest to re-read the final part of the conclusions to see if the paper ends with the highest impact message.

We agree that the ending of the conclusion was quite dull in the original manuscript. We revised the conclusions accordingly:

> **The tests conducted in this paper show that DINCAE is able to provide a good reconstruction of missing data in satellite SST observations and retaining more variability than the DINEOF method. In addition, the expected error variance of the reconstruction is estimated avoiding several assumptions (difficult to justify in practice) of other methods like optimal interpolation.**

References

Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J.-M.: Reconstruction of incomplete oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea Surface Temperature., Ocean Modelling, 9, 325–346, https://doi.org/10.1016/j.ocemod.2004.08.001, http://hdl.handle.net/2268/4296, 2005.

Beckers, J.-M. and Rixen, M.: EOF calculation and data filling from incomplete oceanographic datasets, Journal of Atmospheric and Oceanic Technology, 20, 1839–1856, https://doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2, 2003.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, CoRR, abs/1406.1078, http://arxiv.org/abs/1406.1078, 2014.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Hinton, G. E. and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks, Science, 313, 504–507, https://doi.org/10.1126/science.1127647, https://science.sciencemag.org/content/313/5786/504, 2006.

Hochreiter, S. and Schmidhuber, J.: Long Short-term Memory, Neural computation, 9, 1735–80, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Krasnopolsky, V., Nadiga, S., Mehra, A., Bayler, E., and Behringer, D.: Neural Networks Technique for Filling Gaps in Satellite Measurements: Application to Ocean Color Observations, Computational Intelligence and Neuroscience, 2016, https://doi.org/10.1155/2016/6156513, 2016.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems 25, edited by Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., pp. 1097–1105, Curran Associates, Inc., http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf, 2012.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86, 2278–2324, 1998.

Pisano, A., Nardelli, B. B., Tronconi, C., and Santoleri, R.: The new Mediterranean optimally interpolated pathfinder AVHRR SST Dataset (1982-2012), Remote Sensing of Environment, 176, 107 – 116, https://doi.org/10.1016/j.rse.2016.01.019, 2016.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., pp. 234–241, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain, Psychological Review, 65, 386–408, https://doi.org/10.1037/h0042519, 1958.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85 – 117, https://doi.org/10.1016/j.neunet.2014.09.003, 2015.

Widrow, B. and Hoff, M. E.: Associative Storage and Retrieval of Digital Information in Networks of Adaptive "Neurons", pp. 160–160, Springer US, Boston, MA, https://doi.org/10.1007/978-1-4684-1716-6_25, 1962.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 1995.

Wylie, D., Jackson, D. L., Menzel, W. P., and Bates, J. J.: Trends in Global Cloud Cover in Two Decades of HIRS Observations, Journal of Climate, 18, 3021–3031, https://doi.org/10.1175/JCLI3461.1, 2005.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-128, 2019.