

Title: Neural networks for data assimilation of surface and upper-air data in Rio de Janeiro
Author(s): Vinícius Albuquerque de Almeida et al.
MS No.: gmd-2022-50
MS type: Development and technical paper

Minor Reviews:

Line 40: should be “and medicine”.

Line 54: “encompass researches” should be “encompasses research”

Line 61: “in the latter 1990s” sounds a bit weird

Line 80: “the most of operational centers for weather prediction has” -> “most operational centers for weather prediction have”

Line 85-95: My issue with this is equation 1 doesn't match up with the notation you are describing. You describe x_{N+1}^a , x_{N+1}^b , and y_{N+1}^o but these are not shown in equation 1. You should describe where n and a appear if you use them, or don't mention them if you don't use them.

Line 89 and 94: co-variance vs covariance.

Line 108: “is” should be “are”

Line 110: “can or not be” sounds awkward

Line 111: “or Gaussian one.” The word “one” seems weird here.

Line 112: You say there are two types (supervised and unsupervised) but you only describe one (the supervised). It would be good to add what the unsupervised method is.

Line 124: “Therefore.... Sparse matrix” needs to be reworked. Since you cite it in the line before, you could even just remove this line. If you want to keep it in, it needs a little fixing.

Line 184: This sentence needs to be improved.

Line 186: “regularly” to “regular” OR just remove “regularly” since it's not needed.

Line 190: “1-week data assimilation” what does this mean? This paragraph reads like you are doing updates every 6 (or 12) hours, with forecasts of 6 hours, and this is done over many years, but I am confused what is happening for 1 week?

Line 236: “also on to determine” I am not sure what you mean here.

Line 314: “lets to looking” should probably be “let's look”

Major Reviews:

1. Why train using the 3D-Var solution instead of real observations as the NN target variable? Was 3D-Var solution the only target variable? How does using the 3D-Var solution as the target variable compare to using the real observations? By using the 3D-Var solution as the target variable for the Neural Networks, you are estimating the 3D-Var solution and not the observations. I am a little confused at why you would do that and if it's beneficial.
2. Only one date (1 day or 1 week) as results, what if those dates are good but the rest are worse? This needs some sort of overall result from all dates. You have the whole of 2019 for testing data, does this change if you have different months? Maybe this method worked well for February 1st but not so well on other dates. I like the figures for examples that you have, they

really do show what happens at one time, but I think an overall 2019 average for each method would be nice. Maybe a nice year long line graph showing table 3 results for each day would be nice. That would really show that the results are not cherry picked (I am not saying they are, but people could come to that conclusion here without an overall year plot).

3. This one is related to number 1. If you use the 3D-Var as a target, then your neural networks are not estimating the system, they are estimating the 3D-Var solution. In figure 4, your temperature error maps. 3D-Var will be estimating the system, so your error would be 3D-Var solution minus the real observation. Are you NNs set out like:
 - 1) NN estimating 3D-Var minus real observations, or
 - 2) NN estimating 3D-Var minus 3D-Var solution.