1. Lines 13 & 14. Rephrase the sentences, possibly as follows: “Tandem neural network architecture (TNNA) is a machine learning algorithm which as been recently proposed for estimating uncertain parameters with inverse mappings”.
2. Lines 22 & 25. Here a percentage of noise is mentioned (1% or 10%), but it is not clearly stated which are the measured quantities and what is used as reference value.
3. Lines 31 to 36. Recent publications only have been considered. However, these concepts are well-established since a long time and can be considered text-book material. Moreover, I wonder whether the papers referenced for inverse modeling are the most relevant ones. Many other review papers on inverse problems in hydrology are available and should be considered (e.g., 10.1016/0022-1694(87)90207-1, 10.2136/sssabookser5.4.c40, 10.1016/S0167-5648(04)80146-1, 10.2166/nh.2007.024, 10.1029/96WR00160, 10.1007/BF01547729, 10.1016/0309-1708(91)90039-Q, 10.3390/hydrology11110189, 10.1029/WR022i002p00095, 10.1002/hyp.3360060305, and many others).
4. Line 39. I would not use “deterministic” to characterize Bayesian methods, which are based on the theory of stochastic processes.
5. Line 62. Substitute “CNN” with “convolutional neural network (CNN)”.
6. Line 70. Substitute “DNN” with “deep neural network (DNN)”.
7. Line 80. Methods based on the minimization of an objective function, can be improved from the point of view of the computational effort, through the use of the adjoint equation for the computation of the gradient of the objective function. This should be considered by the authors and possibly mentioned or discussed in the manuscript.
8. Lines 83 & 360. Is “designs” the best word? May be, “is based on”, “considers” or “proposes”? Similarly for “designed” at line 360.
9. Line 85. Substitute “was” with “is”, because the present tense is used in the following sentences.
10. Line 91. Expression “parameter values transition smoothly across space” could be rephrased, possibly as “the spatial variation of parameter values is quite smooth”.
11. Line 95. Is “curse” the best word?
12. Lines 95 to 102. These sentences could be improved to explain why different methods have been used for the different scenarios and to motivate the specific choice of each method. This should help to improve the description of what is novel in this work, otherwise the comment by one of the reviewers remains crucial (“The manuscript presents a thorough comparison, but it fails to identify which is the clear innovation brought forward.”)

13. Line 136 to 138. These sentences could be rephrased, possibly as "These four methods were proposed at different stages of the development of machine learning, but the application for constructing surrogate models in most groundwater modeling scenarios is still relevant." Did I interpret correctly your thoughts? If so, this sentence remain rather nevertheless rather apodictic and I wonder whether it can be supported in a better way from physical arguments or is it necessary.
14. Line 140. Sentence "The surrogate model for inversion will be constructed using the most accurate among them" remains vague.
15. Line 141. Expression "the values for different simulation components" is not fully clear to me. All the data sets used for the training are normalized with the formula $X_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$, where x_i is the i-th value of the data set, x_{min} and x_{max} are respectively the minimum and maximum value of the data set, and X_i is the normalized value. Is this right?
16. Equation (4). how is this equation related to the parameters of equations (1) to (3)? Are x and y scalar or vector quantities?
17. Line 150. Substitute "Eq.(5)~(6)", possibly with "equations (5) and (6)".
18. Lines 151 to 155. The notation has to be modified. What is w^j? In the second line of equation (6) it could be better to use $(u - \varepsilon)^2$. Remark "ε... insensitive tube" can be erased.
19. Lines 164 & 165. Erase "the penalty parameter" and "the kernel function parameter", the name of the variable is sufficient. However, σ is not defined, is it?
20. Equation (7). Do W and B have the same meaning as the same quantities in (4)? σ was defined to be a parameter at line 165, here is a function: this is confusing for the Readers who are not familiar with the applied methods. Erase × from the formula.
21. Lines 177 to 204. The notation is unclear, it does not correspond with the notation introduced in the previous part of the manuscript. For instance, symbols F and G have already been used for different quantities. H is not defined is it? The loss function has the same symbol as an hyperparameter of MSVR. ω_i in equation (12) is not defined, is it?
22. Lines 205 to 215. Is the information about the number of neurons in each hidden layer relevant here, namely, in the description of the methodology? It should be stated later and the motivation for the choice of this value should be given. The same comment applies for the type of activation functions. The whole paragraph could be moved to another point, i.e., after the description of the data sets and where the method is applied.
23. Lines 217 & 218. I partially disagree with statement "the purpose of a surrogate model is to minimize the difference between the predicted outputs and the

numerical modeling outputs": the purpose of a surrogate model is to substitute a high-dimensional model with a low-dimensional model. So the surrogate model must

24. Line 217. Why an L1 norm? L2 norms have been used so far in the work!
25. Line 221. Statement "a widely used machine learning framework" can be erased.
26. Line 226. Symbol G has already been used to denote other quantities, functions, etc.
27. Line 238 & 239. Sentence "For example,... the reduced-dimensional parameters" can be erased, the citation could be sufficient. However, I wonder whether it is the optimal one.
28. Section 2.2.2. Once again the notation is confusing: symbols that have been used previously for some quantities are used here to denote different quantities. Formula $\boldsymbol{z} \sim \boldsymbol{q}(\boldsymbol{z})$ is given without an explanation.
29. Section 2.3.1. Once again the notation is confusing and sometimes not rigorous. These parts could be moved to the appendix, or, even better in the supplementary material.
30. Section 2.3.2. This section requires a thorough revision, with a clear definition of individual quantities.
31. Line 362 & 366. The measurement unit is a different concept from relevant temporal scales.
32. Line 366. Could "plain" be substituted with "alluvial"?
33. Line 374. Which observation data have been simulated? Hydraulic head? Solute concentrations? This is stated much later only.
34. Lines 374 to 379. The added noise is proportional to the value of the "measured" value. Therefore, this means that the error on hydraulic head is assumed to be very small close to the boundaries where the prescribed head is 0 m and to be the highest at the opposite border of the domain, where the prescribed head attains high values. Unfortunately, hydraulic head represent a potential and as such it could be changed by adding a constant value, without changing the hydraulic gradient, which is the "engine" of groundwater flow. Therefore, if one used a different reference height for hydraulic head, the absolute value of errors on hydraulic head and the errors on hydraulic gradients would change a lot.
35. Line 376. Once again, ε is used to denote a different quantity.
36. Lines 383, 515, 649. Symbol "~" should be substituted with "to".
37. Sections 3.1 to 3.3. Which method is used for the simulation of flow and transport? Finite differences, finite elements, finite volumes,...? Eulerian or Lagrangian methods for solute transport? Which time spacing is used? Is the transport model purely convective?

38. Lines 390, 402, 423. Words “meshes” or “grids” should be substituted, possibly with “cells” or “elements”.
39. Figure 2, Lines 565ff. Here upper case K is used for permeability, whereas lower case k is used in the text. I prefer the latter choice, but a uniform symbol should be used throughout the whole manuscript.
40. Line 394. Word “uncertain” can be erased.
41. Line 406. Expression “are as:” should be corrected.
42. Line 409. Word “stable” should be substituted with “stationary” or “steady-state”.
43. Line 410. Add a reference for “equifinality”. Indeed, in this way an important prior information and regularization is introduced, without proper discussion.
44. Line 428. Expression “*t*=2~24 years” could be substituted as “from 2 years to 24 year”.
45. Line 450. Expression “Figure S3~Figure S6” should be substituted, possibly with “Figures S3 to S6 in the supplementary material”.
46. Lines 458ff, Figures 5, 6 & 13, Tables 1 & 3. Measurement units for RMSE are missing. How is RMSE computed for all the model outputs? Head and concentration errors cannot be simply summed up, as they bear different measurement units.
47. Figures 5 and 6. Expression “(a) ~(c) are” should be substituted, possibly with “Plots (a) to (c) show”.
48. Section 4.2.1. What is the “logarithmic average convergence” represented in Figure 7? Is it the RMSE?
49. Figure 7. Why the initial value is different among different algorithms? The caption does not specify what is the difference between the four plots. It would be important to recall that the TNNA curve is the same for all the plots. Why is the curve of DE so “noisy”? I have not recognized such an irregular behavior in my experience with that algorithm. The TNNA curve is quite smooth, but it shows very small bumps, in particular slightly after 150 iterations. Is there any explanation for that behavior?
50. Line 505. Is the noise additive or multiplicative? It seems to be additive, now. So there is a difference with respect to what has been described at lines 374 to 379. Why?
51. Lines 516ff. Validation should refer to the use of data sets corresponding to different physical situations from those considered during calibration. So this is not a standard and thorough “model validation”.
52. Line 572. Numbers in “K4 and K6” should be subscripts.
53. Figures 8 & 9. The captions do not provide full descriptions of the figure content.
54. Lines 584ff. Once again, “deterministic” is used in a context where the Bayesian, stochastic approach is mentioned.

55. Figure 10. The figure caption should be rewritten. Six rows are mentioned, but the figure has 4 rows and 6 columns. No explanation is given for (a) to (d).
56. Figure 11. The figure caption must be completed with the description of what is represented in the four images.
57. Figure 13. Upper case letter should not be used for measurements units: “days”, not “Days”.
58. Figure 14. The second row of plot (a) shows a “wavy” behavior. Can it be explained?
59. Section 4.2.3. It is not clear if the values of permeability of the two hydrofacies have been estimated or have been fixed. In other words, which are the parameters to be identified in this tests?
60. Line 669. Expression “Figure 15-16” should be substituted with “Figures 15 and 16”.
61. Lines 716 & 717. Sentence “three key aspects should be considered to extended for real-world applications” should be rephrased.
62. Lines 724ff. The statement “heterogeneity exhibits ambiguous statistical features” is not clear to me and this makes it unclear also the following remarks.
63. Line 732 & 733. Expression “such designs are also to eliminate” should be rephrased.
64. Line 763 to 774. These sentences discuss potential future developments, which are not based on the results of this work: therefore, they can be erased.
65. Lines 787, 795, 804, 814, 843, 855, 859, 874, 886, 915, 918, 921, 926, 954, 973, 981. The page numbers or the paper numbers of these scientific articles are missing.
66. Line 835. Volume number and page or paper numbers are missing.
67. Lines 839, 848, 913. Several details are missing for these references.
68. Line 847. Details of this reference should be checked.
69. Lines 850, 904. DOI is missing for these references.
70. Line 908. Details of the reference should be corrected.
71. Line 911. “npj Digital Medicine” should be checked.