The authors followed the reviewers' suggestions and answered most of my comments.

Overall, I appreciate their effort, but some comments have not been addressed in a fully satisfactory way and some residual language or typo mistakes should be fixed before accepting the manuscript for publication.

1. The answer to comment "b. There is no discussion about parameter sensitivity and hyperparameter optimization of the TNNA algorithm" by referee #1 is not satisfactory. I warmly ask the authors to carefully reconsider this issue.
2. The answer to comment "d. There is insufficient detail about how the surrogate models are trained and on which parameters they are trained on" by Referee #1 is not fully satisfactory. Some information has been added, but I am afraid that the information is not sufficient for a reader who would like to apply the same procedure.
3. Lines 370ff. the description of the noise added to the data is now very clear. Unfortunately, my comment remains still valid and basically not answered. In fact, if x is the "true" physical quantity, then the normalized value is $X = (x - x_{min})(x_{max} - x_{min})^{-1}$ and $x = x_{min} + X(x_{max} - x_{min})$.
   The noisy normalized value is $X' = (1 + \delta)X$, where δ follows a normal distribution, with zero mean and unit standard deviation. The corresponding noisy physical quantity is given by $x' = x_{min} + (1 + \delta)X(x_{max} - x_{min})$.
   The absolute error on the physical quantity is then $x' - x = \delta X(x_{max} - x_{min})$. It is clear from this formula, that the absolute error is proportional to X, so that it is negligible for values of x close to x_{min}. On the other hand it is maximum, when x is close to x_{max}. I do not see any physical reason for such a choice.
   Moreover, the relative error is given by
   $$\frac{x' - x}{x} = \frac{\delta X(x_{max} - x_{min})}{x_{min} + X(x_{max} - x_{min})}$$
   If $x_{min} = 0$, then the relative error is equal to δ, otherwise it depends also on the value of X.
   I think that a proper discussion of this issue is very important.
4. The answer to the comment on Figure 14, namely on the "wavy" behavior shown in the second row of plot (a), opens a relevant question. If I understood properly, the hydraulic head has been simulated with a threshold of 0.01 m on the iterative method applied to solve the flow equation: if this is the case, the simulation error is greater than 0.01 m. However, let us assume that this is the order of magnitude of the simulation error on "noise-free" heads. Than, this is the same order of magnitude as the error added in the low-noise tests, when a 1% standard deviation is considered. In other words, the noise-free data share an error with the same order of magnitude

as the noise in the low-level tests. However, the basic difference is that the simulation error could have the same absolute error everywhere, whereas the added noise is proportional to X, as demonstrated above. Am I wrong?

5. Line 43 & 44. Sentence "Among available algorithms, methods based on objective functions established from maximum a posteriori estimation and solved by optimization techniques represent a significant category" remains quite ambiguous. May be, it could be substituted with "Methods based on the minimization of objective functions or the maximization of posterior distributions require the application of optimization techniques".
6. Lines 67 to 70. Indeed, my comment intended to stress that the use of the adjoint equation limits the number of runs of the forward problem for the application of gradient-based optimization algorithms.
7. Line 115. The sentence should be rephrased. It might be ambiguous to what word "respectively" refers.
8. Line 117. Specify which is the benchmark with respect to which ML methods give an advantage.
9. Lines 119 & 120. Sentence "With advancements… for future studies" could be erased.
10. Line 141. $\boldsymbol{z}$ is not defined, is it?
11. Line 145, equation (3). If I understand correctly the notation, the operator for parameter dimensionality reduction $\boldsymbol{G}$ computes the high-dimensional parameter vector $\boldsymbol{m}$ starting from a low-dimensional vector $\boldsymbol{z}$. Then the computed vector $\boldsymbol{m}$ is used in the forward (high-fidelity?) model. I think this is not the proper description of surrogate models.
12. Line 150. Expression "calculating their responses" should be rephrased.
13. Line 155. Expression "convolutional neural network" can be erased, because the acronym CNN has already been defined.
14. Line 162. Is the reference Chen et al. (2021) significant? The min-max normalization was introduced long time before that paper.
15. Lines 173ff. $\boldsymbol{m}_i$ is not defined, is it?
16. Line 237. Other choices, e.g., hyperbolic tangent, could provide output in the range from 0 to 1. So, why the Sigmoid was chosen?
17. Lines 253 & 254. The motivation for the use of L1 norm for the loss function is not very informative. Moreover, is "constraints" the right word here?
18. Line 286. G has already been used at line 141, even if here it is a scalar and there it was a vector.
19. Line 333. "After obtain" should be corrected as "After obtaining".
20. Line 338. Word "vector" should be corrected as "vector".

21. Line 355. Expression “measured in days” should be substituted, possibly with “of 60 days”.
22. Line 358. Expression “measured in year” should be substitute, possibly with “of several years (up to 30 years)”.
23. Line 409. The concept of equifinality has been introduced in hydrology at least in 1992 by Beven & Binley (DOI: 10.1002/hyp.3360060305).
24. Line 429. Word “focus” should be corrected as “focuses”.
25. Line 524. Word “during” should be substituted, possibly with “as a function of the”.
26. Lines 555ff, Figure 7. The remark about the “noisy” curve of DE that the authors included in their “Response to comments” should be added to the text of the manuscript.
27. Lines 740ff. Punctuation should be checked.
28. Line 743. Words “also” and “more” could be erased.
29. Lines 743 & 744. Expression “representative parameter field datasets” should be clarified.
30. Line 761. Meteorological factors can affect head measurements, but other factors could be much more relevant.