

1 **REVIEW 1 (our reply in blue)**

2 **Summary**

3 Aqel et al. present a neural network (NN) based approach to predict matric potential from soil water
4 content observations. Using an autoencoder, they extract the most relevant features of the soil water
5 retention dynamics. They input their results into a deep neural network (DNN), which increases
6 the transferability of the DNN.

7 **Assessment**

8 The approach presented in this paper is convincing. The manuscript is well-written. Prediction of
9 hysteresis in soil water retention is of interest to the soil hydrology and soil physics community.

10 I don't have major comments. Thus, I recommend accepting the manuscript after minor revisions.
11 I have some minor comments below.

12 **We thank the reviewer for the positive feed-back and the specific comments (they are addressed**
13 **below).**

14 **Nash-Sutcliffe efficiency**

15 The Nash-Sutcliffe coefficient tends to emphasise maxima in a time series, which might bias the
16 results. An additional interesting metric would be the Kling-Gupta efficiency (Knoben et al., 2019;
17 doi: 10.5194/hess-23-4323-2019).

18 **Thank you for the input. We computed the Kling-Gupta Efficiency (KGE) for the nine sites (see Table**
19 **1 on the next page).**

20 **KGE was developed to address some limitations of the Nash-Sutcliffe Efficiency (NSE) by**
21 **incorporating three components: correlation, bias, and variability (Liu, 2020;**
22 **<https://doi.org/10.1016/j.jhydrol.2020.125488>). The KGE provides a more comprehensive assessment**
23 **of model performance by balancing these aspects. The value of KGE ranges from negative infinity to**
24 **1, with a value of 1 indicating perfect agreement between observed and modeled data.**

25 **In a study by Gupta et al. (2009; <https://doi.org/10.1016/j.jhydrol.2009.08.003>), a KGE of 0.6 was**
26 **considered acceptable for streamflow simulations. In our analysis, we follow Towner et al. (2019,**
27 **<https://hess.copernicus.org/articles/23/3057/2019/>) that used KGE > 0.75 as threshold for "good" model**

28 performance. This threshold suggests that the model accurately captures the dynamics of the observed
 29 data, including the mean, variability, and correlation structure.

30 **KGE values for deep neural network modeling without autoencoder**

31 After running the deep neural network model (section 3.1), the KGE values were as shown in Table 1.
 32 Only two sites, Matzendorf (site #7) and Etziken (site #5), from the nine sites had a KGE value of less
 33 than 0.75 in the validation (KGE > 0.75 for training). These two sites were mentioned in section 3.1 as
 34 sites needing more training data, which follows the expected scenario by NSE.

35 In conclusion, the KGE-analysis defines good model performance for seven out of nine sites (four sites
 36 according to NSE-criterion of $NSE \geq 0.80$).

37 *Table 1: Statistical assessment of calibration (1825 days, until year 2019/2020) and validation results (years 2018/2019/2020*
 38 *until years 2020/2021/2022) for nine sites. The holdout dataset was part of the training period and includes 548 days (30 %*
 39 *of calibration).*

Location	AUV (-)	Training (holdout)			Validation		
		NSE (-)	RMSE (cm)	KGE (-)	NSE (-)	RMSE (cm)	KGE (-)
1 Aetigkofen	1.95	0.92	48	0.91	0.89	60	0.87
2 Bellach	7.00	0.70	98	0.84	0.62	125	0.77
3 Breitenbach ^{a, b}	3.56	0.86	82	0.78	0.83	96	0.84
4 Dulliken ^a	2.19	0.82	55	0.86	0.73	103	0.76
5 Etziken ^a	1.90	0.88	56	0.90	0.75	70	0.65
6 Hofstetten -Flüh ^b	5.59	0.76	90	0.79	0.63	123	0.81
7 Matzendorf	6.39	0.76	83	0.79	0.59	133	0.63
8 Stüsslingen	4.49	0.80	71	0.84	0.80	98	0.85
9 Zunzgen	6.44	0.87	62	0.82	0.83	73	0.77

40 ^a forest sites.

41 ^b Sites with limited available data. For those sites, only 1200 days were used for training; Within this training period, a subset
 42 of 360 randomly selected days was designated as a holdout dataset; the validation period for those specific sites was from
 43 2018/2019 to 2022.

44 **KGE values for deep neural network using the autoencoder value (AUC-DNN)**

45 The results of the Kling-Gupta Efficiency (KGE) for the Deep Neural Network Autoencoder (AUC-
 46 DNN) model in section 3.3 show that three out of the six validation sites have a KGE value less than
 47 0.75 (Table 2, next page). The two sites Hofstetten-Flüh (site #6) and Matzendorf (site #7) have the
 48 lowest NSE values, indicating that the model captures the general dynamics of these site rather than the
 49 exact values. This is consistent with our conclusion in section 3.3. The third site, Breitenbach (site #3),
 50 was identified in section 3.3 as a site where underestimation is expected (see Figure 7 in manuscript),
 51 which explains why its KGE value is below the threshold of 0.75.

52 *Table 2: AUC-DNN Model performance for the period 2012-2022. Three training sites were used to build the AUC-DNN*
53 *model that was then applied for the other six sites. The sites are listed according to the corresponding autoencoder value*
54 *(AUV). The asterisks mark the sites with forest; The AUV was scaled from 1.9 to 7.0 to simplify input. Alternatively, scaled*
55 *values ranging from 0 to 1 could also be utilized.*

Location	AUV	AUV (type)	used as	NSE (-)	RMSE (cm)	KGE (-)
5 Etziken*	1.90	Type 1	Training site	0.82	70	0.81
1 Aetgikofen	1.95	Type 1	Validating	0.76	88	0.76
4 Dulliken*	2.19	Type 1	Validating	0.65	100	0.77
3 Breitenbach*	3.56	Transitional	Validating	0.71	73	0.68
8 Stüsslingen	4.49	Transitional	Training site	0.85	116	0.91
6 Hofstetten-Flüh	5.59	Transitional	Validating	0.60	113	0.72
7 Matzendorf	6.39	Type 2	Validating	0.58	123	0.56
9 Zunzgen	6.44	Type 2	Validating	0.69	104	0.81
2 Bellach	7.00	Type 2	Training site	0.71	104	0.80

56

57 Similar to the analysis without autoencoder discussed in the previous paragraph, the KGE-analysis
58 defines good model performance for more sites than according to the NSE. Accordingly, the NSE-
59 thresholds for good model performance are more challenging compared to KGE and we focus on NSE
60 (and not KGE) in the paper.

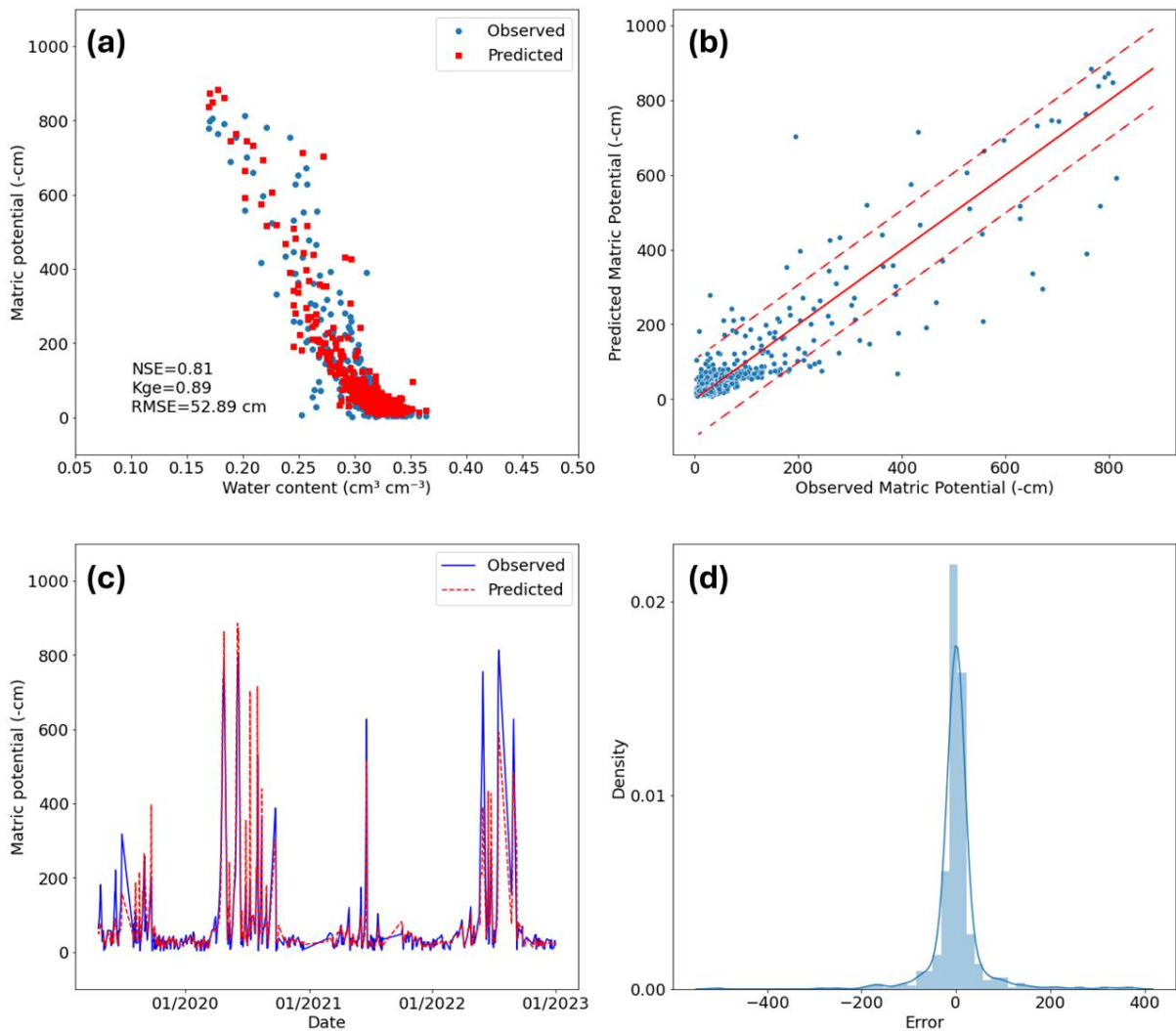
61 **Local nature of the model**

62 The results obtained are from sites that share similar climate and topography. I wonder if this
63 workflow would work as well in different regions of the world, or if further adjustments must be
64 made.

65 To show what is needed to apply the workflow for other regions, we used the Deep Neural Network
66 Autoencoder (AUV-DNN) model described in section 3.3 to predict the matric potential for a site called
67 ‘Wasen’ in the hilly region around the Napf mountain in Switzerland. Compared to the canton of
68 Solothurn (the sites presented in the paper), the Napf region has a different geology and is known to be
69 colder in winter and having more rainfall.

70 The AUV value for Wasen was 2.54 and falls within the range of the training sites (1.90 to 7.00).
71 Accordingly, we could expect that the soil moisture dynamics was similar to the sites in the canton of
72 Solothurn. The model was thus able to predict the matric potential with high quality, as shown in the
73 figure on the next page. The NSE for the model was 0.81, and the KGE was 0.89, indicating that the
74 model performs well in predicting unseen sites in other locations.

75 However, Wasen is still in Switzerland, with soil texture similar to the training sites. It is expected that
 76 the model may not perform well for sites with AUV values outside the training range. We commented
 77 on this in section 4.1 in the revised manuscript.



78
 79 *Figure 1. Evaluation of the Deep Neural Network with Autoencoder (AUC-DNN) model performance at the Wasen site for the*
 80 *period 2019-2023. (a) Comparison between the expected Soil Water characteristics curve (SWC) and the observed SWC. (b)*
 81 *Scatter plot that compares observed data points with their corresponding simulated values, providing a visual representation*
 82 *of the level of conformity to the identity line. The two dashed lines represent the 95% confidence interval around the identity*
 83 *line, providing a visual assessment of the level of agreement. (c) Time series comparison showing the observed and predicted*
 84 *matric potential for the entire period. (d) Analysis of the distribution of prediction errors (observed minus modelled value)*
 85 *using positively mild skewed distribution.*

86 **Physically-based modelling**

87 I partially agree that the reductionist mechanistic models might be unable to account for the full
 88 complexity inherent in the soil water retention process. Input-agnostic approaches such as neural
 89 networks surely have an advantage when it comes to predicting matric potential. However,

90 physically-based modelling is also a tool for process understanding that could potentially help us
91 disentangling the effects of all the interacting processes that control soil water retention. I know
92 that there are efforts to make machine learning a tool for process understanding as well. Perhaps
93 the authors could comment briefly on this and place their work in this discussion?

94 We agree with the reviewer that we need physically-based modelling for process understanding. Due
95 to the complexity of the involved physical processes at the field scale (hysteresis, non-equilibrium,
96 seasonal dynamics of soil structure), we don't have a yet a physical model to predict these processes.
97 Machine learning could help to disentangle these effects, for example by classifying periods that are
98 affected by structural changes and periods that are dominated by non-equilibrium effects. For the
99 different periods, specific amendments in the description of the physical process and properties could
100 be developed (i.e., the application of season-dependent and rate-dependent soil hydraulic properties).

101 Alternatively, physically-induced machine learning (PIML) should be applied in the future, to link the
102 knowledge we have on the physical processes with the data-driven machine learning approaches. There
103 are recent applications of PIML in hydrology: Degen et al. (2023;
104 <https://gmd.copernicus.org/articles/16/7375/2023/>) replaced the complex numerical simulations of the
105 Richards equation with a surrogate model using a set of physically-based basis functions; Bhasme et al.
106 (2022; <https://doi.org/10.1016/j.jhydrol.2022.128618>) combined a set of simple physically-based mass
107 balance equations with machine-learning to predict successfully evapotranspiration and streamflow
108 from a catchment. A similar approach is possible for the problem addressed in our paper: we could
109 combine the physically-based description of the Richards equation with machine-learning based
110 hydraulic functions that change continuously with season or with the drainage rate. Such an approach
111 would provide insight in the changing hydraulic functions and test the validity of the Richards equation
112 (to see if other processes like macropore flow must be included). We will test this in the future but did
113 not address it in this paper.