REVIEW 2 (our reply in blue)

General Comments

This manuscript presents an approach for predicting soil water potential and its hysteresis under natural field conditions by combining deep neural networks (DNN) with autoencoder neural networks. This integration leverages the strengths of both methods, with the autoencoder effectively compressing and capturing site-specific features of soil moisture dynamics, and the DNN utilizing these features to enhance prediction accuracy.

Overall, the method is promising and convincing, and the manuscript is well-organized and clearly written. I have only a few concerns and suggestions, primarily regarding the model's generalization capability to clay soils and regions with significantly different climatic conditions, and the model's interpretability.

We acknowledge the detailed comments of the reviewer. The suggestions and concerns are addressed below.

Specific Comments

Lines 106-109: Generalization Capability: Autoencoders are highly dependent on the quality and diversity of the training data. As shown in Figure 1, the selected region has relatively similar climatic conditions and soil types, mainly loams with a clay fraction less than 50%. I am curious about the model's generalization capability to different regions with varying climatic conditions and soil types, especially for clayey soils. Suggest expanding Section 4.1 to discuss this point and potential approaches to address this issue. Additionally, suggest discussing the possibility of using other autoencoders, such as variational autoencoder (VAE).

We agree with the reviewer that the sites presented in the paper are very similar with respect to climate but also with respect to geology due to their vicinity to the Jura. The “transfer” of the model to sites with different climate or soil properties will be tested in a following study. But motivated by the comment of the reviewer, we checked the application of the model for a site called ‘Wasen’ in the hilly region of the Napf mountain in the Prealps of Switzerland. In that region the geology is different, and the climate is wetter and temperatures in winter are lower compared to the sites of the paper. The model was able to predict the matric potential with high quality, as shown in the figure below. The NSE for the model was 0.81 indicating that the model performs well in predicting unseen sites in other locations. We relate the good model performance to the Autoencoder value (AUV=2.54) that was within the range of the sites presented in the paper (1.9 to 7.0) and we hypothesize that if the autoencoder value deduced from the water content time series is within the range of the training data, the model performs well. A paragraph has been added to Section 4.1 to clarify the model limitations described here.

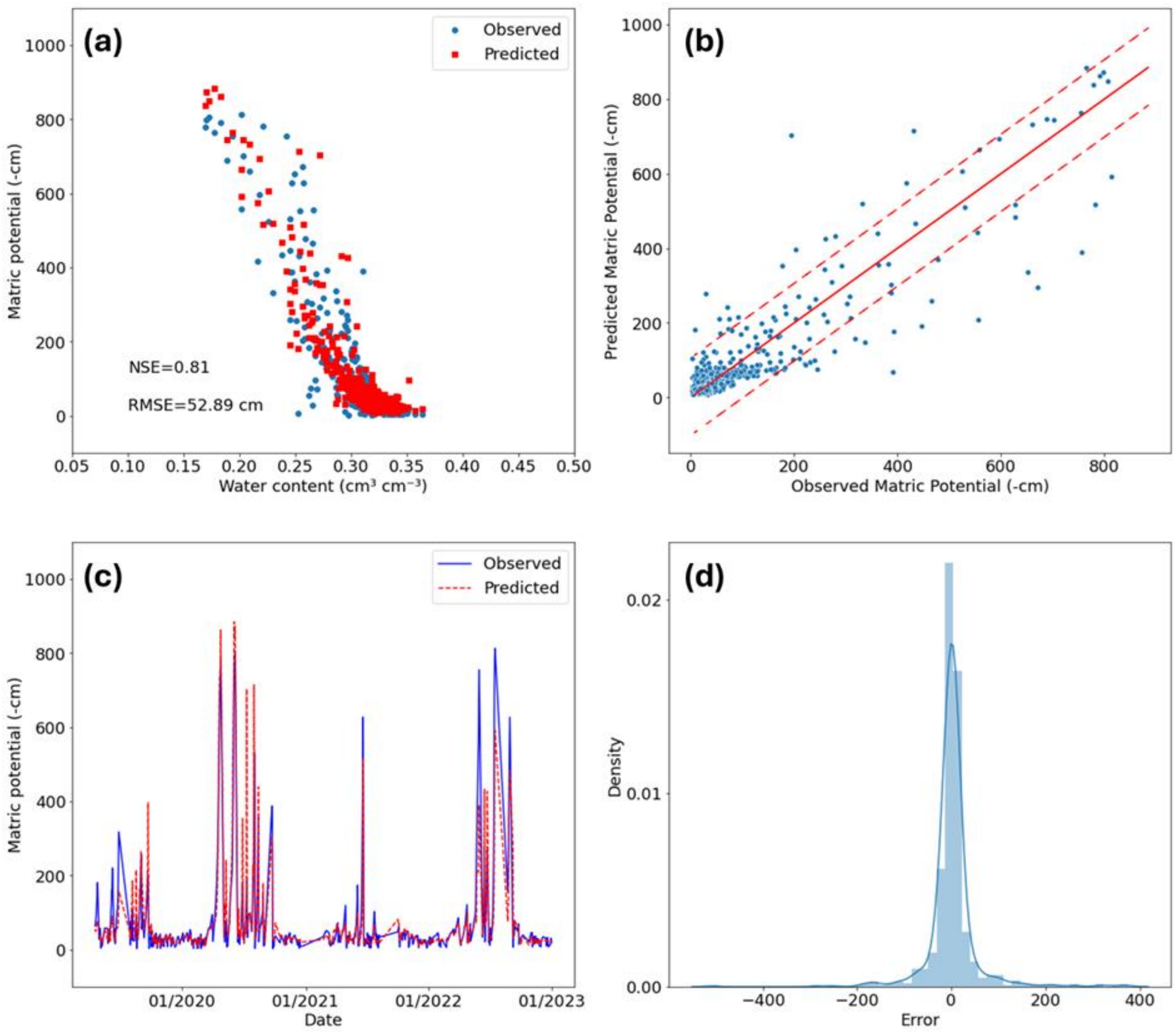

Figure 1. Evaluation of the Deep Neural Network with Autoencoder (AUC-DNN) model performance at the Wasen site for the period 2019-2023. (a) Comparison between the expected Soil Water characteristics curve (SWC) and the observed SWC. (b) Scatter plot that compares observed data points with their corresponding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

We are grateful for the comment on the variational autoencoder. This could be especially helpful for soils with high variations in water content for the same matric potential value as can be expected for clay soils (the autoencoder values were higher for the four sites with clay content ≥ 30%). We expect that using a variational autoencoder instead of a deterministic autoencoder would improve the prediction of matric potential because it leverages regularization in the latent space that explicitly considers the variance (second moment) of the data distribution, leading to a more robust and accurate representation of the water content timeseries (Xu & Liang, 2021; https://doi.org/10.1002/wat2.1533). A paragraph has been added to Section 4.2 to explain why it is recommended to use the variational autoencoder in case of clay soil.

Lines 218: Model Interpretability: The interpretability of the autoencoder's hidden layer representations is typically challenging. Suggest Including a discussion in the results analysis or discussion section on potential techniques to visualize the features learned by the autoencoder's hidden layers, which can help readers understand the model's internal workings

The interpretation of the autoencoder representation was shortly discussed in section 4.2 and is now expanded to clarify the link between the autoencoder value and the water. The analysis indicates that the value of AUC in the model is not equal to the average water content but is highly affected by it (the higher the average water content, the higher the autoencoder value). Deterministic autoencoders, which map inputs deterministically to a lower-dimensional space, tend to capture prominent statistical properties of the input data. The first moment, or the average, is a primary statistical property. Therefore, the hidden layer representations (AUV) in a deterministic autoencoder will indeed be influenced by the average (first moment) of the water content.

However, the average water content alone cannot explain the distribution of autoencoder values found for the nine sites, but the variations of water content must be included as well. This was shown in section 4.2, revealing that the shape of the envelope embracing all variations in water content must be included to explain the autoencoder value of the nine sites. To confirm that the average water content is not sufficient to classify the dynamics at different sites. First, we run the autoencoder model to analyze the yearly changes. Second, we scale the values from 0 to 1, assigning a value of 0 to the lowest yearly AUV across all sites and a value of 1 to the highest. Finally, we calculate the average yearly water content and scale it similarly to AUV. To highlight the results here, we quantified the annual changes of AUV and average water content for two sites that have Type 2 category (see type description in Figure 6 in the manuscript). AUV tracked changes in average water content but exhibited a different magnitude of variation. The site with higher sand content (Bellach) showed a higher variation of AUV compared to the other site (Zunzgen). This observation supports our conclusion in section 4.2 that the hidden layer is capturing more than just the average water content.

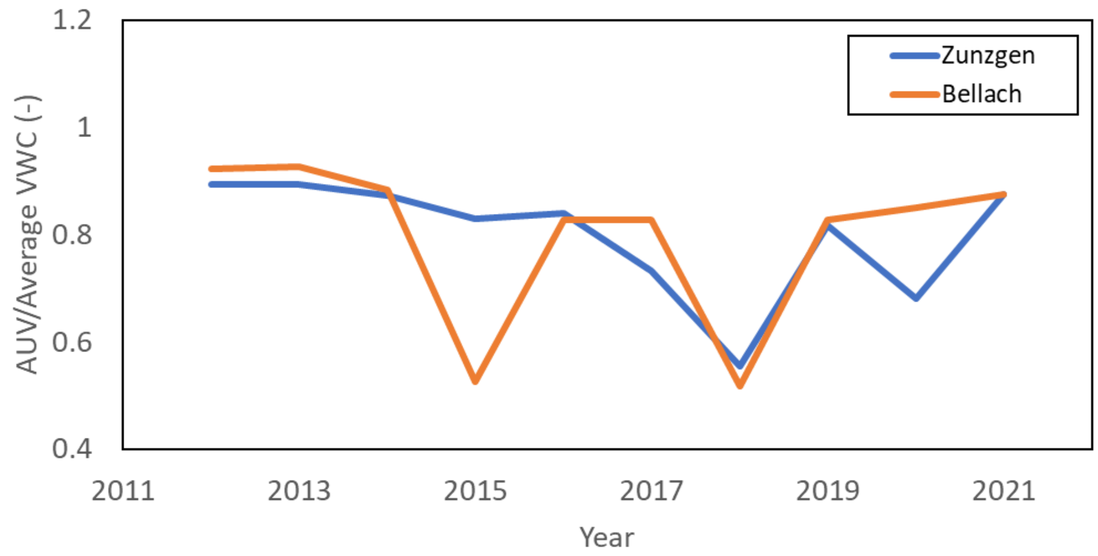

Figure 2: Annual Variations in Autoencoder Values and Average Water Content for two Sites*. The x-axis represents the years, while the y-axis shows the ratio of the scaled AUV to the scaled average water content for the same year. The plot demonstrates how the autoencoder's hidden layer representations track changes in average water content, reflecting variations and additional properties derived from the water content time series.*

To conclude, we consider average water content as a central parameter for visualizing AUV. However, the variability and other higher-order statistical moments (e.g., variance, skewness) significantly influence the precise value of these hidden representations. These additional properties could include:

- Variability (second moment): Reflecting how much the water content fluctuates around the mean.
- Trend: Long-term increase or decrease in water content over time.
- Periodic Components: Seasonal or cyclical patterns in water content.

Therefore, AUV primarily reflects the average water content, combined with other properties derived from the variation in the water content time series. This is discussed in section 4.2.

Lines 289-290: Why adopt an NSE value > 0.80 as the criterion for an optimal model? Please provide the rationale for selecting this value.

Several studies have shown that the performance of hydrological modeling is good when NSE values are around 0.75 or higher (Lin et al., 2017; https://doi.org/10.1061/(asce)he.1943-5584.0001580). Other studies suggest categorizing NSE results into levels to evaluate model simulation outcomes, where an NSE > 0.75 indicates a very good model, while an NSE value < 0.5 signifies unsatisfactory results (Moriasi et al., 2007; https://doi.org/10.13031/2013.23153). In Gupta et al. (1999;

https://ascelibrary.org/doi/abs/10.1061/(ASCE)1084-0699(1999)4:2(135)), an NSE value of > 0.80 was considered as good ('efficient') and NSE < 0.50 as poor. These references are now added in the paper. Here we use NSE > 0.80 as well as criterion for good model performance. For this study, the chosen sites are mainly part of a network designed to provide real-time matric potential information for mitigating soil compaction. We found that when the NSE value is over 0.80, the confidence intervals for matric potential predictions are as follows: around 70 cm for 68% confidence interval, around 120 cm for 90% confidence interval, and around 150 cm for 95% confidence interval. This indicates that despite high predictive accuracy, different confidence intervals provide varying levels of precision and certainty, which can be strategically used for effective soil compaction management:

- 68% Confidence Interval (around ±70 cm): This high precision interval is useful for routine monitoring and precise irrigation adjustments, ensuring that matric potential levels are optimal to prevent over-compaction or drying.
- 90% Confidence Interval (around ±120 cm): This balanced interval offers a reliable estimate for planning soil management practices and designing traffic patterns to minimize soil compaction, providing a good compromise between precision and confidence.
- 95% Confidence Interval (around ±180 cm): This interval, offering the highest confidence, is essential for high-risk scenarios and long-term planning. It ensures that comprehensive measures are in place to prevent severe compaction and maintain soil stability, considering the widest range of potential matric potential variations.

By linking these confidence intervals with high NSE values, we can optimize soil compaction mitigation strategies, tailoring interventions to match the precision and risk tolerance required for various applications, from routine monitoring to high-stakes infrastructure planning.

Figure 2: The common unit for matric potential is -kPa. Please explain the relationship between the -kPa and -cm used in this manuscript.

The data downloaded from the soil moisture network were given in centibars (cbar) with 1 cbar = 1 kPa units of pressure, i.e., energy per volume). In the paper we expressed it as a head (length; energy per weight) considering water density of 1000 kg m^3 and gravity acceleration of 10 m s^{-2}, resulting in units of cm that are 1/10 of kPa. This is now stated in section 2.

Equation 1: Please ensure that all parameters are clearly defined after the equation, and that their mathematical notation (bold, italic) is consistent throughout the manuscript.

We rearranged the text to define the parameters after the equations and checked the notation throughout the manuscript.