Ref.: Ms. No. COMPAG-D-23-04782

Title: Using a Transformer Network to Determine Soil Water Content along a Time Domain Reflectometry Sensor

Dear Editor and Anonymous Reviewers

**The authors appreciate the comments from the editors and the reviewers.**

Some of the comments are related to the basic TDR concepts, some of the comments are related to technical details and basic neural network concepts, and other comments (such as why EC are not included or why EC is assumed to be constant) are related to philosophical concerns for TDR (and may be related to future hardware development).

Many of the comments are not specific to the “TDR-transformer” model structure. Because we are submitting this “TDR-transformer” work as a method/technical paper, it is **not** appropriate to incorporate our responses to the comments in the main paper. **Instead, we include many of our responses in a supplemental material section for the TDR sensor, waveform, and transformer network.** The supplemental material section is written topic by topic. Some heuristic ideas and discussion are provided to help readers understand TDR equipment and TDR-Transformer. We hope readers are able to follow the supplemental material concepts step-by-step.

Sincerely

Zhuangji Wang, Dennis Timlin and Robert Horton

|  |  |  |
| --- | --- | --- |
| **Editorial Remarks:** | | |
| **1** | I have completed the review of your manuscript, and a summary is appended below. The reviewers recommend reconsideration of your paper following major revision. I invite you to resubmit your manuscript after addressing all reviewer comments. | Point-by-point responses are provided by the authors in this file. The authors appreciate the time and work from the reviewers and the editors. |
| **Reviewer ONE** | | |
| **2** | The title has an appeal for a novel approach to interpret TDR waveforms. The use of "Transformer Network" probably has less appeal (less awareness) than the term, "Neural Network" if this doesn't make much difference to the authors. | **The authors appreciate the comments from the reviewer.**  We agree in part with this review comment.   1. Essentially, “Transformer” is a family of models in the field of deep learning, but it may not be well known to natural resource scientists. For example, in the name “chatGPT”, that last letter stands for transformer “T = Transformer”.   In deep learning-related publications, in general, people use “Transformer” (without “network”). We want to be clear to our readers, so we add “Network” following “Transformer”.   1. “Transformer” indicates the architecture of the neural network. Different types of neural networks may be very different from each other (like two animal species). So, we have to say “transformer” to clarify our neural network model architecture. 2. **In the revised title, we include “Transformer Neural Network”** to emphasize it is a “Neural Network”. But in the body of the article, we still use “Transformer Network” and just “Transformer”. |
| **3** | This paper is a follow-on paper from Wang et al., (2023; 10.1029/2022WR033895), though this paper is surprisingly sketchy compared to this previous publication in WRR. For example, the WRR paper contains six complex figures with many depictions, while here there are just 3 figures. | 1. In our earlier paper [Wang et al. (2023), https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022WR033895], the model is a stack of three networks, like an “OREO cookie”. So, we present the three networks one by one and then assemble them together.   While for Transformer, the network structure is very integrated. For example, the “encoder” sends information to the “decoder”. If a reader notes “” in the decoder, it means the decoder receives and processes information (from encoder) 4 times with different parameters. Moreover, the Transformer Network includes iterative computations marked on the right side of Fig. 1.  The presentation style follows a commonly used Transformer structure diagram. If the reviewer searches for “Transformer AI”, they will observe a similar format for other applications.     1. This paper follows a similar writing style as Wang et al., (2023), i.e., the data transformations follow the model diagrams.   The writing style is appropriate (i.e., not superficial). We present step-by-step descriptions that guide readers through the paper. To assist the readers to go through the model, there are multiple places in the original paper where we marked the shapes of the input-output data, i.e., data size. For example,     1. To help readers better grasp larger grids, **an illustrative version with smaller data size is provided in the supplemental material section.**   Please note:   1. Our description in the original paper is detailed compared to typical deep learning papers. 2. **Hopefully, the illustrative version (with a “toy example”) provided** in the supplemental material section is easy to follow. |
| **4** | The authors cite several key figures in other papers, e.g., Timlin and Pachepsky (2002), which I am unable to access due to a lack of subscription to Soil Science. Why wouldn't the authors reproduce those key figures, readers are not interested in chasing figures from unavailable publications. | **This is a very good comment, and we appreciate it.**   1. Fig. 3 is “reproduced” from the earlier paper, because the TDR waveforms there are digitalized from Timlin and Pachepsky (2002). 2. Reproducing existing research results would be a “digression” to cause the manuscript to not focus on the main objective/method of this study. 3. Therefore, details in Timlin and Pachepsky (2002) will be summarized in a supplemental material section, and we also provide a copy of that paper in the supplement material. |
| **5** | There is also an overabundance of jargon used, which in many cases could be better translated to the specific terminology of the application of TDR waveform analysis. Readers will struggle to follow much of the narrative here in my opinion. | **We appreciate this reviewer comment.**  The comment does not clearly say which terminology is acceptable and which terminology is unacceptable for a TDR paper. Therefore, we make some assumptions in our response.  We identify some terms that may be considered as “jargon” (as mentioned by the reviewers), such as “Transformer”, “Convolutional Embedding”, “Attention Mechanism”. **However, those terms should remain for the following reasons.**   1. Although the terms may not seem to be commonly used in TDR descriptions, they are standard terms widely used by many scientists and engineers. Thus, it is not necessary or appropriate to create new terms for our TDR analysis. Even though our application of transformer to TDR is new, we should keep using standard terms that already exist. 2. To the best of our knowledge, this manuscript is one of the first applications (if not the first one) that adopts a transformer model architecture for TDR waveform analysis. So, for the machine learning terms, there is no corresponding “specific terminology of the application of TDR waveform analysis (mentioned by the reviewer)”. 3. When the terms first appear in the text, we present explanations, sometimes with demonstration on the change of data size (from A-by-B data frame to M-by-N data frame), sometimes with citations. |
| **6** | Also, please be clear when referring to the TDR probe, versus the TDR rods. These are not the same thing. | **Respectfully, we appreciate the reviewer comment, but we CLAIM that we do not use ambiguous terms.**   1. Respectfully, in the main text of this paper, we never use or refer to “TDR probe”, while the word “TDR probe” or “probe” appears three times but all of them are in References as parts of titles of existing papers. 2. Throughout this manuscript, we use “TDR sensor” and “sensor rod/waveguide”.   “TDR Sensor” means **the overall equipment**. Traditionally, it does not include the cable tester, while as the development of Acclima TDR, some signal generation and cable tester functions are integrated into the sensor.  “TDR sensor rod” means the parallel metal ‘needles’ or rods of the sensor. A group of parallel rods (commonly 2 or 3 rods) becomes the “waveguide”. For example, in the manuscript, we say “*To perform a TDR measurement, a waveguide with two (or three)* ***parallel metal rods*** *is first inserted into a soil……* *When the signal enters the waveguides, reaches the ends of* ***the sensor rods****……*”  “Waveguide” means a group of “TDR sensor rods”, for a TDR sensor, a waveguide often has 2 or 3 TDR sensor rods (based on the common design).  **WE DO NOT USE “TDR PROBE”**. In general, “TDR probe” should be equivalent to “TDR Sensor”, but we believe “TDR Sensor” is clearer. Because in some designs, “TDR probe” is not one “sensor”, and the TDR sensor could look like an “edge”. **Therefore, “TDR probe” is too general and may cause ambiguity.**  For example, the Campbell SoilVUE-10, referred to as the “big screw,” is one TDR probe (as named by Campbell), but the metal parts are TDR sensors (green circles) with parallel metal rods along the edges. |
| **7** | There is no mention of the actual TDR device used in the creation of the TDR waveforms used in the example. It is not appropriate to simply cite a paper and expect readers to look elsewhere for these critical details. | **This is a very important comment, and we appreciate it.**   1. It is a leverage between “tightly focus on the critical target or contribution from this paper (i.e., TDR)” and “description of data or measurements that are already published elsewhere”. 2. Based on this review comment, as well as a previous comment regarding the availability of Timlin and Pachepsky (2002) paper, we decided to include a supplemental material section to address these issues.   By doing that, the manuscript still tightly focuses on the main objective of this study, while some technical details are included in the supplemental material. |
| **8** | Title: Suggest being more explicit in the title since the sensor head is not part of the waveform analysis, "along the rods of a TDR sensor" | The title is reworded. |
| **9** | One of the Core Ideas states, "Variations of and along TDR sensors are determined with the new model", however, the physical measurement and waveform interpretation is only for travel-time along the rods. | Although the reviewer does not request a specific revision, we want to provide some comments and clarifications.   1. The direct physical measurement from TDR should be voltage. The information from TDR measurements is embedded within the variations of the electrical signal voltage. 2. Although travel time analysis is commonly used in direct models, by saying “*the physical measurement and waveform interpretation is only for travel-time along the rods*” in the review comment, it seems TDR reports the travel time to the users. We know the goal of travel time analysis is to determine εr. It is not appropriate to just report a travel time to soil scientists, because relative permittivity or water content values are the desired results. 3. By saying “*however, the physical measurement and waveform interpretation is only for travel-time along the rods*” in the review comment, **that implicitly assumes the physical travel distance is known**. This assumption is adopted in traditional travel time analysis, which provides average signal speed (and hence one average εr along the whole waveguide).   **In other words, travel time analysis works (and is used as the core step for direct models), because the travel distance is known.** For example, in direct models, such as Tangent-Line, AWIGF, BMO, the physical travel distance is the length of the TDR rods (or the length if both forward propagation and backward reflection are considered). Then, using travel time, one can estimate signal speed (= travel distance / travel time) and then determine relative permittivity by comparing the signal speed with the speed of light.  **However, if the water content or relative permittivity varies along the rod, then using travel time analysis alone is not sufficient.** For example, if the TDR rods are in a 3-layered soil with three different water contents (or relative permittivity values), even if you can obtain a travel time for each layer, you still do not know the thickness of each layer (or say the physical distance along the rod occupied by each layer). Then you cannot compute the signal propagation speed and relative permittivity for each soil layer.  In other words, although you know the whole TDR rod length, you do not know, and you cannot assume, the length portion occupied by each of the three soil layers, along the TDR rod. Thus, the traditional travel time analysis cannot be simply applied in this general case.   1. Thus, travel time analysis may only work for uniform soil relative permittivity. Otherwise, you must assume or know external information. For example,   *If the TDR rod is partially exposed to air, you will have to know that layer is air so you can assume a relative permittivity of air (about 1), or you need to know the length of the rod exposed to air.*  *Therefore, you must know something in addition to travel time analysis when the soil* *relative permittivity is not uniform.*  These are some of the basic concepts that TDR users should be aware of, and this type of discussion is incorporated **in Supplemental Material Topic 1**. Travel time implicitly adopts assumptions, but the assumptions may be built into existing software without user awareness.  Some existing inverse methods (directly inverse fitting the waveform via the telegraph equation) use TDR measured voltages, and not travel time analysis either. |
| **10** | The interpretation of is subject to a slew of other factors and in that sense, should be left out of the statement here and elsewhere in the manuscript since waveform interpretation does not yield water content per se, i.e., from the neural network. | **This is a valuable comment, and we appreciate it.**   1. In this manuscript (or in some other TDR modeling papers), the Topp et al. (1980) equation is implicitly assumed. So, we treat and as equivalent quantities. Once is obtained, we just need a (deterministic) Topp et al. (1980) equation to compute 2. Indeed, there exist other calibration equations (or working curve) to relate and .   So, in the revised version, we clarify that we are determining by the TDR waveform interpretation, and then we clearly mentioned that can be used to compute via the Topp et al. (1980) equation or another equation depending on soil properties.  That will not affect the TDR-transformer because the TDR-transformer and most of our analyses are based on . |
| **11** | This is especially the case if those factors including porous medium-surface area, -bound water, -constituent phase configuration, -texture, -electrical conductivity (e.g., assumed constant), etc. are not involved in the Neural Network analysis. Although the goal is determination of it is misleading to simply suggest that determination directly yields (see Line 64 for example). | **This is a very valuable and an important comment. We really appreciate it.**   1. We agree that soil properties may exert a variety of influence. However, is soil electrical permittivity, which is a quantity determined via TDR waveform interpretation. This is independent of the factors that the reviewer mentioned.   To make an extreme example, given a soil , we can use water, ethylene glycol or other liquids to make up a solution with the same value of , and theoretically, the TDR sensor cannot distinguish between the soil and the solution (while measurements in soil may have larger noise).   1. However, converting to soil water content includes assumptions regarding the soil properties, as mentioned by the reviewer in this comment. Those are also the assumptions that should be considered when using Topp et al. (1980) model. 2. **In the revised paper, we emphasize that** “ is the directly measured soil electrical properties reported from TDR measurements. To convert to , soil minerals, texture, etc. may exert various affects. Therefore, the output from TDR-transformer is and the conversion from to relies on existing calibration curves, such as the Topp et al. (1980) model.” |
| **12** | L102: The word "Transformer" by itself has a vague meaning here and does not match the reference used in the abstract of "TDR-Transformer" with citation (Abadi et al., 2015). Is this a generic version of the same thing? Provide a reference and context along with the word. | 1. “Transformer” is a generic neural network architecture, and “Transformer” can be used in a range of applications. So “Transformer” in the sentence L102 should not be "TDR-Transformer".   More importantly, in L102, we write “*Transformer is a* ***versatile neural network*** *based on attention mechanism, and it has been used for both image and time series analyses.*” It should be clear that this sentence gives a background of “Transformer” to provide clarity and to remove vagueness. We believe this sentence gives a clear description of Transformer from the following three aspects,   1. Transformer is a neural network; 2. It is based on attention mechanism; 3. It has some applications.   Then in L111, we write “*Since transformer can handle both geometrical and time domain information, it is promising in TDR waveform analysis.*” This should be a clear indicator that we change from the general review of “Transformer” to our target “TDR-Transformer”.   1. In the abstract, we write “*TDR-Transformer is developed using Python and Google TensorFlow with Keras API (Abadi et al., 2015).*”   “*(Abadi et al., 2015)*” is a citation for “*Google TensorFlow with Keras API*”. This is a tool for neural network development. Just like we are using “MICROSOFT WORD software” to write this response. Abadi et al. (2015) presents a tool, rather than Transformer or TDR-Transformer.  In order to avoid such a misunderstanding, and “*(Abadi et al., 2015)*” is not that necessary in the abstract and already cited elsewhere, **in the revised version, we removed the citation from the abstract.**  In the original abstract, we write “*We present “TDR-Transformer” as a new waveform interpretation model. Equipped with a transformer network backbone, TDR-Transformer can determine ……*”  Based on this sentence, it should be clear that   1. TDR-Transformer is the new model proposed in this study and there is no other reference indicating that TDR-Transformer already exists. 2. TDR-Transformer uses Transformer network as the backbone, which gives a clear relationship between “TDR-transformer” and general “transformer network”. 3. **We agree with this comment** “*Provide a reference and context along with the word.*”   In the revised paper, “Vaswani et al., 2017” is cited as the first paper of attention mechanism and Transformer, in L102. |
| **13** | L137-139: What is meant by sampling at a constant frequency here? Is this the algorithm, a hypothetical waveform or a TDR-measured waveform? For the TDR-measured waveform, the frequency content is broadband and is often a function of the porous medium properties through which it passes. It may contain reduced frequency content, for example, when passing through lossy porous media such as clayey soils. You should provide context behind this scanning index and how it relates to the travel time or travel distance (TDR units).  The point is that you are suggesting you can analyze waveform geometry but the geometry of waveforms is tied to layers with permittivities and depths so why not explain things in terms of time or depth, at least when you get to the analysis. | 1. In the manuscript, we wrote “*Since the reflection coefficients are sampled at a constant frequency, the horizontal axis is labeled using the “scanning index” (from 1 to 1000) rather than elapsed time.*”   Therefore, “*sampled at a constant frequency*” means the cable tester reports the reflection coefficient value at a constant frequency, where “reflection coefficient” is the vertical axis of a TDR waveform.  Recall that direct physical measurement processes for TDR are that (1) the cable tester (or signal generator) injects a voltage signal into the TDR sensor and (2) the cable tester samples voltage. Therefore, one can understand that the cable tester records the voltage values from the TDR sensor and then reports it as the “reflection coefficient” at a constant frequency.    **We revised the text to make it clearer, but the revision is not in this place but in L58-L60 where related information first appeared.**   1. Sampling frequency is not an algorithm. For example, I measure my height every day, then the sampling frequency for my height is “once per day”; I measure my weight twice per day, the “twice per day” is the sample frequency for my weight. 2. A TDR voltage signal contains a wide range of frequency, because ideally the voltage signal should be a “Heaviside Step Function”. A step function contains a wide band of frequency and can be demonstrated via Fourier Transformation. This is just a fact, and we agree with the reviewer on that fact.   However, the sampling frequency is the frequency for the cable tester to measure the TDR voltage and it is not related to the frequency components in the TDR voltage signal.   1. The reviewer asked this question “*Why not presenting TDR waveform with respect to time or layer distance*?” 2. Both time (usually several ns) and scanning index (often 251 but can be up to 5000) are commonly used to present TDR waveforms. They are essentially equivalent, because the cable tester measures voltage at a constant sampling frequency.   For example, in the simulated TDR waveform, when we say the “sampling frequency” is constant, that means for every 2e-11 s, the cable tester reads a voltage datum from the sensor. Therefore, one scanning index is associated with a time advance of 2e-11 s.  In TDR papers or documents such as TDR100/200, papers for TACQ, the scanning index (251 data point) is presented, and in papers for AWIGF, time is usually used.  “Presenting in time” or “scanning index” has nothing to do with permittivity and depth. The cable tester always makes voltage measurements, and the measured voltage is always there independent to permittivity and depth, no matter if we put the sensor in air or soil or even in water.   1. Then the question is “*since time and index are the same, why use index rather than time?*”   That is because the model we developed treats a TDR waveform as an array of numbers. So, an index is preferred rather than time.  For example, when I take a photograph. The direct outputs are pixels. Although I know each pixel corresponds to a physical distance, that does not matter when I process (smoothing or sharpening) the images. Operation on the image is done by treating it as a matrix of integers, and operations are on each integer by the index rather than physical distance.  (Note that B-scan ultrasonography may provide functions in measuring physical length but that belongs to post-processing steps)  Similarly for TDR-Transformer, we treat a TDR waveform as a “1D array” and TDR-Transformer operator the signal by index. That is the reason we prefer using index.  In the revised version, we clearly marked how each index corresponds to a specific time interval, so the reader can convert index to elapsed time back and forth freely.   1. *“Why not distance?”*   The reviewer suggested using the depth of soil layer thickness, while we consider “the depth” is the same as the distance from the head of the TDR sensor.  First, we do not know where the water content changes. If we can plot or explain a TDR waveform with respect to distance (as the horizontal axis of a TDR waveform), that means we should know where the water content changes. However, that is the quantity we want to determine. Therefore, plotting or explaining TDR waveforms with respect to distance falls into a circular-argument logic trap, because we wrongly assume we know the value of a desired quantity as a pre-condition.  Second, distance is a secondary quantity and should not be an independent variable. Because the TDR sensor is connected to the cable tester only at one end (may be the sensor head or the coaxial cable depending on the sensor design). So, physically, we do not take measurement on the voltage/reflection coefficient at an arbitrary position along TDR rods.  **This response is heuristic, and some of it is included in the paper or the supplemental material section.** |
| **14** | Line 141-149: This is quite confusing to follow, perhaps since you are not describing things explicitly in terms of the waveform geometry. Most TDR users familiar with the geometrical features of the waveform would be expecting to have the rising portion of the second reflection (end of rods) to be included in the analysis, which your Figure 1A does not seemingly include. This should be corrected as the rising portion just outside of your shaded pink region | **We understand the reviewer’s concern, so we provide one more sentence to define the geometrical features. The following discussion incorporated into the revised paper.**   1. The “Geometrical Features” is everything in the TDR waveform rather than just the second reflection position.   More importantly, the reviewer may notice that throughout this paper, for the TDR-Transformer model construction, we never mentioned “reflection positions”. Reflection Positions in this manuscript are for other methods or used for comparison proposes.  Essentially, TDR-Transformer, as well as some other inverse models, directly analyzes the TDR waveforms without explicitly pointing out the reflection positions. That is because when soil water content varies, “reflection positions” could be complicated, and a waveform can jump up and down.   1. The shaded areas in Figure 1A are hand-drawn and for demonstration proposes only, so we can extend the last section a little bit more. However, the shaded area will only cover a small portion of the rising part.   To simply understand this, “rising part” means the TDR is responding to a change in , so it does not necessarily mean the time when the voltage signal front hit the boundary of changes (or the end of the sensor rods) during the “rising part”.  The actual time when the voltage signal front hit the boundary of changes should be at the beginning (a small interval) of the rising portion.  Although Tangent line and TACQ use information at the middle of the rising portion for computations, that is not very related to this study. |
| **15** | L151-161: It would be very helpful for readers if you put the analysis in the context of traditional waveforms in soil as well as your disciplinary jargon. For example, where is the first reflection and the second reflection in homogeneous soils. Other reflections would appear where a disparity in permittivity was present. Typically, derivatives of the waveform are used for relating the peak derivative to the point of reflection for travel time analysis. Is any of this the same? | Respectfully, we cannot revise the manuscript based on this comment, even though we understand some familiar languages could be good for TDR users.   1. TDR-transformer does not explicitly rely on the reflection positions. **In other words, TDR-transfer model structure does not explicitly force the model to trace the reflection position, and then use the reflection positions to do the following computations**.   This is one of the big differences compared to direct models, such as tangent line, AWIGF, BMO. In direct models, the reflection positions are explicitly determined and then used in the following computations.   1. Not focusing on reflection is not unusual in TDR analysis either. For example, in inverse analysis models, the whole waveform can be directly fitted without knowing if a specific data point is a reflection position or not.   **Our revisions:**   1. From a computation process perspective, this paragraph is clear. But some terms such as “geometrical features” will be redefined in the previous paragraph.   In the revised manuscript, we say “*TDR geometrical feature is a summary of the functional patterns, including the increasing, decreasing, and horizontal patterns in the TDR waveform. Each local geometrical feature (1-by-32 array) focuses on an interval of 40 TDR data points, and each entry in the 1-by-32 vector has the ability to provide one representation of the (geometrical) functional patterns for that 40-by-1 TDR data interval, including but not limited to the increasing trend, decreasing trend, horizontal trend, slope or curvature.*”  Thus, we provide a description on the geometrical feature. For TDR waveforms, as well as for arbitrary functional curves, the geometrical feature is not limited to reflection (a.k.a., jumps).   1. For some terminologies like “embedding”, “convolution”, “position embedding layer”, they are standard names and there are no corresponding terms in TDR, unless we create new terms.   The terms are commonly used in transform neural network, and to the authors’ knowledge, this is one of the first studies (if not the first one) to bring transformer to TDR. There are no traditional TDR terminologies that fully address the transform neural network terms.   1. Finally, we emphasize that “determining reflection positions” is not 100% necessary for all TDR data interpretation models. |
| **16** | Line 255-256: This is a great example where you explicitly relate the waveform to water content when it is inappropriate since you haven't and shouldn't define a relationship between water content and permittivity in this exercise as it depends on other factors not part of this algorithm. Please remove the water content relationship in this manuscript since it isn't important other than in the introduction as stating an example of the application. | **This is a very valuable comment. We really appreciate it. But we have to make the following clarifications.**   1. Relating water content or not does not affect the TDR waveform analysis. Because after reporting relative permittivity, TDR analysis is performed.   Since calibration equations (e.g., Topp et al., 1980) are commonly used to convert relative permittivity to soil water content. In the original manuscript, we assume that relative permittivity and soil water content are “equivalent”.   1. In the revised version, we are more careful when relating relative permittivity to soil water content. However, we think that mentioning soil water content is good to include for most TDR users.   Therefore, instead of totally removing water content, **we put assumptions that Topp et al. (1980) is assumed in this study, and we also clearly mention that calibration curves between relative permittivity and soil water content may vary due to soil properties.** |
| **17** | Line 261-262: How were waveforms simulated, using what frequency band, what rod length, etc.? | We added a supplemental material section that includes technical details, please see supplemental material Topic 3. |
| **18** | Line 367-369: Your statement, "When spatial variations occur in both soil electrical conductivity and soil , it is difficult to determine whether the geometrical fluctuations in the waveform are due to the variations in electrical conductivity or .  However, one of the strengths of TDR is that losses from electrical conductivity (lossy) soil conditions have little impact on the determination of water content (permittivity). This is because the EC impacts are generally in reducing the Reflection coefficient or waveform magnitude, while the travel time analysis for permittivity determination is simply a function of the reflection positions along the time or distance axis, i.e., vertical versus horizontal scales. How then is your algorithm sensitive to permittivity determination if traditional TDR waveform analysis software doesn't have a problem with determination of permittivity until the second reflection is completely attenuated? Please expand on this explanation. | The statement from the reviewer “*This is because the EC impacts are generally in reducing the reflection coefficient or waveform magnitude, while the travel time analysis for permittivity determination is simply a function of the reflection positions along the time or distance axis,*” is conceptually correct, while EC reduces the energy of the TDR so we say EC causes continuous decreases in the waveform (voltage or reflection coefficient) not only near the reflection position.  One of the problems is, when soil water content varies along the rods, multiple reflection positions occur. Then, the rising period of the previous reflection positions can superpose on the following reflection position. If the water content varies layer by layer, the magnitude of waveform variations will be large enough to “shade” relatively slow EC changes.  In this study, we do not explicitly mention EC because the determination of coupled and EC variations has not yet been solved (i.e., and EC that vary arbitrarily and randomly, while TDR waveform is the only input). However, when the model is trained, we randomly impose a uniform EC value up to 0.5 mS/cm as the background condition.  The EC values are small but able to cover tap water EC ranges in infiltration measurements. This is also mentioned in the supplemental material. Therefore, the results in Fig. 2 already include some EC effects.  However, determining EC is not the goal of this study and EC must be assumed to be constant along the sensor rods. And determining EC under varying soil via TDR waveform is not yet possible. This is the reason we do not actively mention EC in this study. **Also see Review Comment 26. Topic 4.1 in the supplemental material.**  ----------------------------------------------------  All in all, statements in the review comment are true and easy to implement when soil and EC are uniform. However, if soil or EC have spatial variations, some concepts we assume for in uniform and EC may not be true anymore. |
| **19** | Section 2.1 Encoder Structure: There are many undefined terms here, e.g., convolutional embedding layer, zero padding, stride, feature vectors, data frame output, mask size, etc. Please provide context and definition for readers who are not likely familiar with the discipline where these are familiar. | We added a supplemental material section that includes technical details.  We understand the concern from the reviewer, but we have to say that the concepts can be found in related textbooks. Therefore, the terms are somewhat elementary, so we include a citation “Fig. 14-4, Géron, 2019”.  “data frame output” is not a terminology, in the manuscript, we said “…*which results in a 50-by-32 data frame output after the convolutional embedding operation*”. That sentence should be understood as “…*which results in [a 50-by-32 data frame] + [output after the convolutional embedding operation]*”.  We change “after” to “from”, which seems to make this sentence clearer, i.e., the data frame is “output from” the convolutional embedding operation.  We added a supplemental material section to provide a more easy-to-follow version of the transformer. |
| **20** | Figure 1: you cite Figure 1 on page 6, but it doesn't appear until page 11. It should appear on page 7. | Done. |
| **Reviewer TWO** | | |
| **21** | The major accomplishment of this study is to construct a "TDR-transformer" using neural network based on attention mechanism for determining dielectric permittivity (or soil water content) "distribution." While there is an innovative component in this study, it falls short in several key issues. The major comments are: | **The authors appreciate the comments from the reviewer.** |
| **22** | - Clarification 1: Be more specific on how TDR waveforms are simulated. Of course, the governing equation is the transmission line equation (Line 254-255). To generate synthetic TDR waveforms, the source function, impedance profile, and dielectric model should be given. | Because this paper is designed to be a concentrated technical paper, that information was omitted from the original paper. In the revised version, we include a supplemental material section to provide technical details. |
| **23** | - Clarification 2: From Fig. 1, it's not clear how training is carried out. In addition to encoder structure and decoder structure, how the transformer is trained should be clearly elaborated. | Due to the existence of TensorFlow, training a network is quite straightforward. We just need to specify the input, output and the target function (minimize the L2 distance between the predicted and ground-truth ).  Please see Topic 2.4 in the newly provided supplemental material section. |
| **24** | In Section 3.2, how is the transformer network trained? Is it trained by the simulated waveforms and applied to both synthetic and experimental data? If the pulse rise time and impedance mismatches are not modeled correctly, would that affect the TDR-transformer prediction? | This network is trained and tested by simulated waveforms (because a large number of waveforms are needed) in Section 3.1, then applied to the measured waveforms in Section 3.2.  The waveform simulation is run twice, each time generating 80K waveforms. So, we have two simulated waveform datasets. One for training and one for testing. Then the trained TDR-transformer is applied directly in Section 3.2 to measured waveforms (under a different TDR sensor configuration).  We have to use simulated waveforms because training a neural network requires a large amount of data, which is not supported by existing databases.  -------------------------------------------------------------------  We are not 100% sure about the meaning of “impedance mismatches”. Because the training uses simulated waveforms, we know the ground truth values for each waveform. Therefore, we interpret the “impedance mismatches” as that “the change in soil that occur in the middle of one 40-by-1 data interval, then what value for this interval can be predicted?”  If our interpretation is correct, then “impedance mismatches” indicated by the reviewer is a “resolution problem”. If that happens, TDR-transformer reports an intermediate value.  Based on Fig. 2 in the original paper, we say TDR-transformer is stable against these “impedance mismatches”. That is because based on the strategy of how a TDR waveform is sliced and the overlapping between two slices, the change of soil should always occur within one 40-by-1 data interval, or say the “impedance mismatches” always happens, but the results shown in Fig. 2 are still stable.  Regarding to “how a TDR waveform is sliced and the overlapping between two slices”, please see Fig. 1 for the relation between data intervals and strides, also see Fig. S4 for a more illustrative example. |
| **25** | - Other clarifications: It's not clear how the causality (restricting the attention to the left) is enforced in Eq. (1). It's not clear how the Linear Regression is conducted in Fig. 1B. | Those technical details are commonly used for neural networks, which are independent of TDR-transformer, so we put this in the supplemental material section. |
| **26** | - Assumption of electrical conductivity: Like dielectric permittivity, EC is a function of water content. Depending on soil type, EC may be more sensitive to water content than dielectric permittivity. Assuming EC is constant is not a reasonable assumption. When the transformer is calibrated, EC is assumed constant. Besides the EC variation, what about the value of EC? If the value of EC in application is different from that used for training, would that affect the result? Since simulation tool is available, the effect of EC variation and EC value should be investigated and discussed. | Determining coupled arbitrary variations of soil relative permittivity and soil EC is not yet possible, and existing studies either restrict one quantity to solve for the other one, or apply external assumptions. For example, in existing inverse studies, without extra assumptions, they mainly report soil water profiles (e.g., <https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2002WR001890>) and seldomly mention EC profiles. In our previous study, we mention EC profiles, but then our uniform assumption will be in the water content part (e.g., <https://www.sciencedirect.com/science/article/pii/S0168169921000302>).  Some inverse models also use similar waveform simulation tools. Therefore, even though we have simulation tools, the waveform simulation tools are independent in order to resolve arbitrary variations in both soil relative permittivity and soil EC (i.e., they are two different things).  In this paper, we only made small perturbations on the EC (0-0.5 mS cm^-1), because that EC is just a background condition and determining EC is not the goal of this study.  We understand that in practice, variations in both soil relative permittivity and soil EC can occur, but we did not yet find a way to resolve both variations, so any discussion on EC effects at this point will be not strict.  Since this is a method paper and “constant EC assumption” is the main reason for the reviewer to raise this question, we provide a discussion from a technical perspective. This assumption (constant EC) is based on how to use the TDR information. Since we are developing a TDR data analysis model, we must analyze how much information we obtain, and how much information we can extract.  The detailed discussion is placed in the supplemental material section topic 4.1. We do not put that discussion in the paper because that discussion may relate to a broad topic in TDR, and which needs future research advancements. |
| **27** | - Generalization:　What is the ability of the TDR-transformer to determine dielectric profile that changes gradually instead of sharply with an interface? | In Section 3.2, one of the wetting fronts is relatively sharp, and one is relatively smooth, due to the soil conductivity. Therefore, for the infiltration with a smooth wetting front, there is a gradual soil variation, and our TDR-transformer model can capture it.  We did not test our model with a “linear” change in soil water content. That could be a “gap” for a lot of similar studies that determine soil water content variations. We will address it as a future research topic.  Essentially, this question/comment provides a reason why TDR-transformer, and a range of inverse models, do not put the “reflection position” as the first priority (as the direct models do). |
| **28** | - Inverse analysis: The results are compared with inverse analysis. The estimated dielectric permittivity from inverse analysis is apparently not correct. If the inverse analysis is carried out properly, it should provide model-based reasonable results. | That value is taken from the paper, which is also included in the supplementary material since the “soil science” journal is no longer being published. Those waveforms are measured results.  The target of this paper is not to judge one existing model. The goal of this testing is to demonstrate that the TDR-transformer can produce appropriate results. Therefore, we only provide discussion on TDR-Transformer.  The old model predicted well the position where soil water content changed. Therefore, it is still worth comparing against that study. That leads to the following discussion points:  One digression: positions where soil changes v.s. soil values:  When we compare TDR-transformer and TDR-CNN, we present discussion on one individual model,   1. if the model provides good determination on the position where soil changes, the values may not be that good (like TDR-CNN); 2. the model’s resolution for changing position may not be that good, but then values could be accurate (like TDR-transformer);   It looks like “ changing position” and “ value” is a tradeoff. Therefore, although the value of the inverse model is not that good, the wetting front position is better than TDR-Transformer. |
| **29** | - Title: What is directly estimated is dielectric permittivity, and it should focus on the "profile" obtained. So it's better to change it to "…. Determine Dielectric Permittivity Profile…" | Done. We prefer to use “relative permittivity” to substitute for “Soil water content”.  Thanks for this suggestion. |
| **30** | - Please discuss how the structure parameters (e.g., stride number, mask size, etc) were decided and their effects on the results. | Based on the previous study (<https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022WR033895>), we wrote this sentence in the manuscript “*The number of geometrical features (32) and the mask size (40-by-1) follows the TDR-CNN design, which are sufficient to represent TDR waveforms (Wang et al., 2023).*”  Therefore, the convolution embedding is designed based on existing studies. We revise this sentence a little bit to include “stride number, padding, etc.” |
| **31** | - In Section 3.1, the number of training waveforms and testing waveforms are not specified. | This question is the same as some of the questions in “Review Comment 24”. After that comment is addressed, this comment is resolved. |
| **32** | - Some references should be given, for example, Eq. (1) and when describing the feeding forward box and residual step. | Done. The study [Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., Polosukhin, I. 2017. Attention is all you need. Advances in Neural Information Processing Systems 30: 5998–6008 (NIPS 2017)] is the **BREAKTHROUGH** study, so it is cited. |