

Subject: Final author comments (ACs)

Title: Integrated approach for characterizing aquifer heterogeneity in alluvial plains

Author(s): Igor Karlović et al.

MS No.: egusphere-2025-327

MS type: Research article

We sincerely thank the reviewers for their thorough evaluation and valuable suggestions to improve our manuscript. Below, we provide point-by-point responses to all comments (reviewer's text underlined), followed by the corresponding revisions made to the text. We begin by responding to the comments from Reviewer 1 (Thomas Hermans, RC1), followed by Reviewer 2 (Anonymous Referee, RC2), and finally the community comments from Lee Slater (CC1).

**RC1 (<https://doi.org/10.5194/egusphere-2025-327-RC1>)**

Dear authors,

I read with interests this paper entitled: “Integrated approach for characterizing aquifer heterogeneity in alluvial plains”. In this article, a methodology is proposed to integrate geophysical data as a constraint for geostatistical simulations meant for generating realistic realizations of alluvial aquifer heterogeneity. The topic is relevant, as these aquifers are amongst the most complex to characterize, while there are often exploited for drinking water production and highly contaminated in and around cities due to industrial activity. I support any effort related to a better characterization of these complex systems. In that sense, the paper is interesting, but from my point of view, the reader is left with a feeling of unfulfilled expectations. My major concerns are described below:

1. Geophysical data are not really used as a soft constraint in this study, but merely to estimate correlation lengths. This is a strong limitation of this work since correlation lengths from tomographic methods are known to be biased (Day-Lewis and Lane, 2004). This likely

explains the limited added value of ERT for the evaluation criteria compared to other studies that used ERT in a similar context (e.g. Hermans et al., 2015). Actually, looking at your objectives, I see many similarities with the study by Hermans et al. (2015). They used ERT to constrain hydrofacies simulated by MPS. They also used falsification to deduce the most realistic training image. They further constrained their simulations by hydrogeological data collected during a pumping test. Also see the work by Barfod et al. (2018) for a similar approach. Therefore, I think it is important to better describe the global context of using geophysical data to constrain geological models, and discuss the proposed methodology in that perspective. Ideally, it would be good to also include simulations where the geophysical data are actually used as soft constraints for the detailed area where data is available.

We acknowledge that geophysical data are not used as a "soft constraint" in exact same context as in the works cited by the reviewer. In our study, these data serve to estimate the length and geometry of characteristic sediments with distinct hydrogeological properties, i.e. hydrofacies units. As Turner (2021) explains, soft data consist of "indirect observations of geological properties, as well as qualitative and interpretative information from geophysical surveys or conceptualizations of the depositional system." We maintain that our approach using ERT data still represents soft data, though applied in a different context. To ensure clarity for readers, we have modified the text to provide a better description of our methodology and the broader context of using geophysical data to constrain geological models, as suggested by the reviewer (also in line with another reviewer, CC1). Throughout our revised paper, we have consistently accompanied the term "soft data" with its specific description as "ERT-derived lens lengths." We emphasize that our ERT-derived soft constraints (both vertical and lateral) are grounded in the actual vertical positions of hydrofacies units observed in wells, addressing the reviewer's concerns about potential bias. We also note that this does not contradict or is in conflict with the work of Day-Lewis and Lane (2004).

Regarding the limited added value of ERT data, the observed similarities between ERT-derived lens lengths and T-PROGS default values in Model Area 2 offer a potential explanation for the similar prediction outcomes. We outline this interpretation in Section 3.2.2 and our response to CC1.

Based on the clarifications made about the methodology and general use of geophysical data, we consider unnecessary to perform MPS simulations to estimate the facies probability linked to a dc-resistivity value (or range of values) as suggested by the reviewer (similar to cited works). However, we recognize the value of this perspective and will certainly consider applying such methods for estimating key hydrogeological parameters in future research.

Reference: Turner, K. A.: Discretization and Stochastic Modeling, in: Applied Multidimensional Geological Modeling: Informing sustainable human interactions with the shallow subsurface, edited by: Turner, K. A., Kessler, H., and van der Meulen, M. J., John Wiley & Sons, Ltd, <https://doi.org/10.1002/9781119163091.ch13>, 2021.

2. The introduction should focus more on studies which investigated heterogeneity characterization, which is also the topic of this paper. Here are a few references that are relevant, there are many more (and more recent): Baines et al. 2002; Bowling et al. 2005, 2007; Bersezio, Giudici, and Mele 2007; Mastrocicco et al. 2010; Doetsch et al. 2010, 2012a. Gottshalk et al., 2017.

We have carefully reviewed all references suggested by the reviewer and have incorporated those most relevant to support our research.

3. The methodology uses thresholds on resistivity to classify the deposits in hydrofacies (L120-122). But with such a low number of validation wells, the uncertainty cannot be captured, does it? See Hermans and Irving (2017) for a study dedicated to uncertainty analysis in a similar context. This paper shows that such a threshold actually does not exist. For any resistivity value, every hydrofacies has a specific probability of occurrence. This comes mostly from the limitation of geophysical inversion which smooths interfaces, but can also results from the heterogeneity within the sediments. This has also been demonstrated by Isunza-Manrique et al. (2023) in another context. To me, this aspect is essential for any study aiming at a robust integration of geophysical data in a stochastic framework. I would also extract 1D resistivity distribution at the location of borehole and show in parallel the hydrofacies (L182). Because of the smoothing effect of inversion, I really doubt that fixed boundaries can be used. It could reveal a lack of co-located data to derive these ranges.

We recognize that the number of wells used to define the resistivity thresholds is limited to two wells located near four ERT profiles. However, the resistivity anomalies (defining hydrofacies units up to 20 m depth in each profile) are very well defined, with clear resistivity contrasts between them that align with vertical lithological contacts observed in the wells. This result is very important to consider, because it allows us to confidently establish resistivity thresholds for each hydrofacies within the first 20 m depth, without uncertainty. Thus, we differ with reviewer's point of view of implicitly suggesting that uncertainty analysis would show each hydrofacies having a probability of occurrence for any resistivity value, implying no distinct resistivity thresholds exist. From a mathematical perspective, this approach appears unsustainable given the physical and chemical properties of porous media under an electric field. For instance, between clays and gravel, resistivity thresholds can be clearly defined based on well-understood physical characteristics (grain size, pore geometry, grain density, tortuosity) and chemical properties (mineral composition, CEC), along with saturation conditions and pore fluid salinity. The clear differentiation of materials in the Varaždin ERT data (through resistivity thresholds) arises from their distinct physico-chemical properties, producing strong resistivity contrasts. This interpretation is validated by its consistency with vertical lithological contacts observed in wells.

Regarding the limitations of geophysical inversion using the R2 code (ResIPy), we have used synthetic modeling to address potential artifacts associated with the geometries and resistivity contrasts among hydrofacies units in the upper part (above 20 m depth), as well as for the conductive materials at 20 m depth or below (Sgcs, CSs) with unknown geometries (as explained in the text). We also used it to address the limited depth of investigation caused by conductive materials at greater depths (Sgcs, CSs), along with the typical loss of resolution with increasing depth. The smoothed interfaces observed in all synthetic models were not associated with artifacts when using reasonable theoretical (true) resistivity values. We obtained very similar results to those for the field data, validating the quality of field measurements and inversion protocols. Also, synthetic modeling suggests that each unit in the field is in its interior not heterogeneous (at least for the resolution and sensitivity of used electrode array), and that the smoothed interface in the inverted model are reliable. Overall, these synthetic modeling results demonstrate that the smoothness-based regularization approach (Occam's inversion) used by R2 (ResIPy) does not produce any misinterpretation

linked to the smooth appearance of electrical resistivity distribution. In fact, it effectively resolves the geometric characteristics of hydrofacies with minimal residual error, meaning it fits the synthetic resistivity data very well. Therefore, for the field data-based model, we conclude the inversion settings used in R2 were the most suitable (relative to a blocky inversion approach for example) for characterizing the expected smooth-shape structures of alluvial deposits at the study site, resulting in an excellent fit, as indicated by the low misfit error.

While the study by Isunza-Manrique et al. (2023) provides valuable insights, its focus on artificial environments such as municipal landfills makes it fundamentally incompatible with our investigation of natural deposits. These artificial settings involve complexities that extend far beyond simple heterogeneities, which is why we consider their methodology inappropriate as a reference for our work.

In our revised manuscript, we have clarified that our approach does not involve robust integration of geophysical data within a stochastic framework, but is rather related to estimating lens dimensions. In our response to the first comment, we have explained why determining probability of occurrence for each unit is unnecessary in our study context. Following the reviewer's suggestion, Fig. 4 displays the resistivity values extracted from the ERT profile (1D resistivity distribution) alongside borehole SPV-5. To avoid confusion and maintain clarity, there is no need to show the boundaries of  $\rho_{nf}$  values or resistivity lines contours next to the borehole (SPV-5) that correspond to different lithologies in Fig. 4, as suggested by RC1. To address this point, we have provided a detailed explanation in the text describing how these values were determined (Section 3.1), which are also shown above the resistivity color scale in Fig. 4 for direct comparison. The fixed resistivity boundaries for each unit using a smooth resistivity distribution are completely validated from synthetic modeling results. Given the type of deposits in the alluvial system and the horizontal dimensions of hydrofacies of several tens of meter, we would expect to find rounded units with flat surfaces as those in synthetic models, which can be perfectly delineated with Wenner or Wenner-Schlumberger array configurations, as we have successfully achieved in our work.

4. I find  $n=10$  realizations a very low number to look at uncertainties. In particular, the objective of a stochastic studies should be to deduce posterior probabilities, and not only deduce the most likely model (see line 162). I think more realizations are needed given that 4 facies are considered.

During hydrofacies model development, four alternative modeling approaches were tested alongside the maximum entropy method (Carle, 1999), each with 10 stochastic realizations. The results showed no statistical deviation from the maximum entropy method's performance range. Since these alternatives offered no substantial advantage while increasing computational demands, they were excluded from further analysis. Given this, we proceeded with the maximum entropy approach for subsequent analysis, as the method provides intuitive way to observe vertical transition patterns between hydrofacies.

While we acknowledge that additional realizations could benefit uncertainty quantification studies, this work focused primarily on methodological development for subsurface heterogeneity characterization and identification of the most probable hydrofacies model, as it forms the basis for defining hydraulic conductivity fields. Future work will expand on utilizing the hydrofacies model as direct input for groundwater flow and contaminant transport simulations to support groundwater management in the study area.

5. Some important elements of the methodology are unclear. For example, it is not explained how the correlation lengths are extracted from the geophysical inversion (L169-170). It is crucial to the methodology, and any subjective element or involved parameters should be identified, also considering my other remarks (see point 3 above).

In the revised manuscript, we have provided a detailed step-by-step explanation of our approach for estimating lens dimensions in Sections 2.2 and 3.1. Briefly, we used a joint interpretation of ERT data, borehole records, and synthetic ERT modeling to systematically characterize all identifiable hydrofacies lenses and estimate their lengths across the 10 ERT profiles. Based on these measurements, which are now presented in the updated Table 1, we calculated mean length values for each hydrofacies type.

6. L197-203. This part is presented as one of the novelty of the paper. However, as explained above, other studies using TI for hydrofacies simulations proposed a more thorough analysis of this (see Hermans et al., 2015 for MPS, and Hermans and Irving, 2017 for synthetic studies and Gottshalk et al. 2017 for indicator simulations). Here, a deterministic approach is first used to derive some correlation lengths that are then integrated in geostatistical simulations, this is not really an integration of ERT into stochastic modelling, which would imply some consideration of the related uncertainty (i.e. probability distributions). Note also that the approach of using synthetic models to validate interpretation is not new. See Caterina et al. 2013 for example. The modelled lenses are very thin. Can they really explain the low resistivity observed, or is it simply a loss of resolution with depth ? How can you be sure only the background profile is found at depth?

We revised the sentence related to the novelty of the approach used in our work, since we did not perform full integration of ERT into stochastic modelling (i.e. probability distributions). Regarding synthetic modeling, we have revised the text to remove any implication of novelty, as the reviewer correctly pointed out that this technique is not new. While we do not cite the work of Caterina et al. (2013), the exercise of carrying out synthetic modeling dates back to the early days of the computer era, initially applied to VES, or to modeling lateral changes and near-surface inhomogeneities prior to ERT. However, we would like to emphasize that, based on the synthetic modelling results, the conductive anomaly at 20–22 m depth plays a key role in defining the depth of investigation for ERT measurements in the Varaždin area, and consequently affects the resolution at greater depths. These two aspects are clearly and easily observed in both the pseudo-3D model shown in Fig.3 and the refined ERT interpretation of each measured line, such as VIN-1 shown in Fig. 4. We are aware, and fully agree with RC2, that synthetic models are equally affected by the conductive layer (CSs-Sgcs lenses or layer), i.e., simulated ERT measurements will exhibit loss of resolution and limited depth of investigation at around 20 m depth. As Reviewer 2 (RC2) points out, our synthetic models explicitly incorporate prior geological knowledge, which cannot be neglected in synthetic modelling. Borehole data confirms the presence of a conductive layer formed by CSs and Sgcs materials with a thickness of approximately 3 m at this depth. The extent of low resistivity values below the depth of the CSs-Sgcs material (whether as a thin layer or lenses) is explained as an artifact due to the loss of ERT resolution, limited by current

penetration depth, rather than representing a true geological feature. Furthermore, borehole data reveal that beneath the conductive CSs/Sgcs layer lies the Gsc hydrofacies, which should exhibit significantly higher resistivity relative to CSs/Sgcs. Refined deterministic analysis of ERT data from iso-resistivity contours maps was very effective in confirming that the apparent extensive conductive zone below 22 m depth in VIN-1 is caused by two distinct conductive anomalies whose depths are underestimated by the inversion approach (i.e., their iso-resistivity contour lines are observed deeper, about 35–40 m). This is observed in both the synthetic model and field data, demonstrating that the CSs-Sgcs materials are highly conductive and significantly affect the ERT data (resolution and depth of investigation). Overall, our expectation from synthetic modeling was to define a plausible geometry of the upper boundary of these materials (as a thin layer or lenses), then constrain their thickness based on borehole data for final interpretation and estimate horizontal lengths used in statistical modeling. This exercise forms the foundation of synthetic modeling work and represents best practice in any geophysical investigation.

#### Specific comments:

We have carefully addressed all of the specific comments to the best of our ability. For those suggestions that did not include specific line references, we have made reasonable interpretations about their intended locations in the manuscript. Should any of these require adjustment, we would greatly appreciate further clarification to ensure we implement the suggested improvements precisely.

1. There are too many technical details for an abstract. It is not understandable what the grid refers to (ERT inversion grid, hydrofacies simulation grid?). I also miss some context. Typically an abstract should be structured as: 1) Global context 2) Specific research gap 3) Proposed methodology 4) Main results 5) Conclusion.

We have restructured the abstract to follow the recommended format, avoiding excessive technical details while providing clearer context and research approach. We believe that the revised version more effectively communicates our findings to a broader audience.

2. The resolution of ERT always decreases with depth.



We fully agree with the reviewer's observation about ERT resolution decreasing with depth. While our results show the characteristic resistivity decrease with depth (Fig. 4), borehole data demonstrate that coarser, more resistive units underlie the fine-grained Sgcs-CSs hydrofacies lenses. This highlights that thin conductive lenses mask deeper resistive formations. The current tends to flow through conductive materials, which fundamentally limits depth of investigation and affects resolution (i.e. its capability to demarcate the boundaries of such conductive materials). The text has been revised to clarify that borehole data support this interpretation.

3. Not clear what is meant by boundary conditions in this context.

The text has been revised to eliminate potential confusion.

4. MPS can also be pixel-based. The original SNESIM algorithm was a pixel-based MPS algorithm (Strebelle, 2002), so is the direct sampling algorithm (Mariethoz et al., 2010).

We appreciate the insight and have revised the text to improve clarity, avoiding overgeneralization of simulation methods.

5. “Depending on” rather than “Given”.

We have implemented the suggested modification in Section 2.1.

6. Reference to specific Excel functions is not necessary.

We have implemented the suggested change by removing the Excel function reference in Section 2.3.1.

7. Reference to an automated python script is not necessary. It is expected that you made the process automatic.

All Python references have been removed from the validation methodology sections for both model areas (2.3.1 and 2.3.2) as suggested.

8. A resistivity of 4600 Ohm.m seems very high for alluvial sediments. Is this realistic?

The 4,600 ohm·m reflects an inversion artifact resulting from using 10,000 ohm·m as an upper boundary constraint during ERT inversion. Synthetic modeling demonstrates that hydrofacies G exhibits resistivities of 500-1,500 ohm·m, which aligns with expected values for coarse alluvial sediments (as seen in Fig. 4).

9. I don't see any sharp boundaries in the figure, which is a result of the smoothness constrained used for inversion. Have you considered other inversion approaches (blocky inversion, minimum gradient support, etc.)?

We have revised the text to describe boundaries as gradual transitions rather than sharp contrasts, consistent with our use of smoothness-constrained inversion. This method is more adequate to represent the deposition environment, reflecting the expected smoothed, rounded geometries of different materials in the study area.

10. Figure 4. You select different values for the different lenses, why? Why is the CSs layer discontinuous? Wouldn't a continuous layer also explain the data?

Our approach integrates borehole data, ERT measurements, and synthetic model simulations. The geometries of the materials and their resistivity values in these models were constrained by ranges from the field ERT data, as outlined in Section 3.1. A fundamental step in synthetic modeling involves evaluating different initial (true) resistivity values within the possible range for geologic materials, which are primarily influenced by moisture content and, consequently, salinity variations. It is well-known and standard practice to correlate bulk resistivity with electrolyte (pore water) resistivity as a function of moisture and/or salinity variations at laboratory scale. However, in the Varaždin study area, field measurements surprisingly revealed minimal variation, i.e., each material or hydrofacies exhibited well-defined, nearly constant resistivity values, with only slight differences depending on their spatial position (see Fig. 3 for example). Based on this observation, it is reasonable to infer that moisture and/or salinity variations may occur within the same material at different locations. Accordingly, we assigned slightly different true resistivity values to represent natural variability. For the Sgcs-CSs materials, we used values of 150 and 180 ohm·m near the surface and 20 ohm·m at greater depths, reflecting the expected decrease in resistivity with depth due to increasing moisture content and salinity. For the G materials, we applied

true values of 1500 ohm·m and 1000 ohm·m, with the 500 ohm·m difference representing plausible spatial variations in moisture or salinity.

Revisiting RC2's remark, prior geological knowledge constrains our synthetic modeling. While a continuous layer CSs could also explain the data, borehole records clearly demonstrate its discontinuous nature across the study area, as several boreholes did not intersect this layer. Additionally, the resistivity contrast between Gsc and CSs-Sgcs in the first 0-3 m depth is clear, showing the discontinuity (i.e., lenses) of these materials. Note again that the moisture content between these two types of textures plays an important role in generating such resistivity contrasts. We have clarified these points in the revised manuscript to improve understanding.

11. What is “the virtual position of the borehole”? Do you mean a projection of the borehole on the profile?

Thank you for this observation. We have revised the text to use more precise terminology in the Section 3.1.

12. Table 1 gives mean values and ranges, but it is not mentioned how many lenses are detected to calculate them.

As suggested, we have added the number of lenses (n) used to calculate the mean lengths in Table 1.

13. Figure 8. The number of correct predictions is quite low. What would be the score if the most abundant facies (background?) would be predicted everywhere?

The results in Figure 8 are based on 1,200 stochastic realizations, showing a range of correct predictions, with mean values used to compare performance across different grid resolutions. While the number of correct predictions may seem low, the key focus was to evaluate the relative improvement in predictive capability between models developed using only borehole data and those incorporating ERT-derived lens lengths (in addition to testing grid resolutions). The choice of Gcs as the background hydrofacies is the most logical from a sedimentological point of view, representing transitional deposits between coarser and finer

hydrofacies and reflecting shifts between high- and low-energy depositional environments. This interpretation is supported by the ERT profile in Fig. 4, which shows Gcs located between other hydrofacies lenses. Predicting the most abundant facies everywhere would oversimplify the system, ignoring the spatial heterogeneity our method aims to resolve. Such an approach would produce accuracy scores matching the hydrofacies proportions in borehole data, but fail to capture realistic geological structures. As our goal is to reproduce plausible heterogeneity, this hypothetical scenario is not directly applicable. However, if assessed, the accuracy would align with the relative hydrofacies proportions observed in boreholes, as outlined in Section 3.2.2.

14. It is not clear to me what “oversegmentation of the lenses” is. Aren’t you overinterpreting distributions that are not significantly different given the low number of samples (n=10)? Wouldn’t you need many more validation data to analyze the risk of oversegmentation?

We evaluated grid resolution effects using 1,200 simulations in Model Area 2. The results demonstrated that optimal cell size corresponds to estimated lens lengths. Increasing model detail beyond these dimensions did not improve accuracy, making finer grids computationally inefficient. We acknowledge that “oversegmentation” was an imprecise term and have modified it throughout the text.

15. L335-340. Maybe refer to the work of Danish colleagues who developed a scale of “reliability” when building their geological models using hard and soft data (e.g., Enemark et al., 2024 and references therein).

The suggested paper, while valuable, addresses propagating interpretation uncertainties from 3D hydrostratigraphic models to groundwater models, representing a different research phase from our current borehole-based hydrofacies identification in section 4 (1). We maintain our qualitative reliability framework (defined in Fig. 1) as the appropriate approach for this phase of research, though we will consider the suggested methodology in future research.

16. L342-359. See my main comments related to other studies which proposed more advanced methodologies.

We reviewed the suggested literature and adopted relevant elements that we believe added value to our manuscript.

## **RC2 (<https://doi.org/10.5194/egusphere-2025-327-RC2>)**

### General comments

Thank you for submitting this manuscript. Personally, I am curious about applying stochastic modelling for the purposes of electrical resistivity response of the subsurface, so I found it an interesting and engaging read. The reasoning, or wider context, behind the research is well formed (in that the aquifer which is the subject matter of the research is critical to the local water supply). The methodology in this manuscript seems competent, though I'm not familiar with the T-PROGS software that underpins the stochastic modelling efforts. I find the results insightful, however, I do have concerns about the statistical significance the stochastic outputs with and without the benefit of geoelectrical information.

Regardless I wish the authors well, and hope my comments are useful to them. Furthermore, I would like to recommend that this manuscript be accepted on the basis of moderate revisions.

### Specific comments

Lines 93 – 95: Is it known how K is determined in these boreholes? Are they determined in situ (e.g. falling/rising head tests) or via laboratory investigations on samples.

The reliability classes presented in Figure 1 were developed to reflect the range and quality of available hydrogeological data in the boreholes, where parts related to information for K determinations refer to all possible estimation methods, including grain-size distribution analyses, permeameter tests from borehole samples, field investigations such as slug tests, pumping tests etc.

Lines 198 – 202: As the ERT is sensitive to 20 m below the ground level, I agree that the synthetic modelling should also be limited to 20 m. Although, I'm unsure if the inclusion of synthetic modelling improves the sensitivity of the electrical measurements to various hydrofacies below 20

m depth. If one proposes a tomography model (be it through a smoothness regularised gauss newton approach, as with R2, or via stochastic approaches) the raw data is still fundamentally only sensitive to the upper 20 m in this case. It's a limitation of where the electrical current flows, as the authors point out (Lines 198-199). Nevertheless, I think the authors could argue here that the synthetic modelling benefits ground model development by incorporating prior knowledge of the geology.

We agree with the reviewer that both field data-based and synthetic models are reliable down to a depth of approximately 20 meters. More precisely, this sensitivity limit corresponds to the depth where the conductive CSs/Sgcs layer is encountered, e.g. in borehole SPV-5 these two materials form thin conductive layer ranging from 20.7 to 23 m depth. We fully agree with the reviewer's observation that ERT measurements are fundamentally sensitive to the upper 20 meters, regardless of the approach used to estimate a possible geological model. This is the primary reason why Model Area 2 was limited to the upper 20 meters, allowing us to assess how ERT-derived lens lengths and varying grid resolutions influenced model prediction accuracy, while avoiding potential uncertainties arising from synthetic modeling at deeper levels. The purpose of employing synthetic modeling was not to improve sensitivity of ERT measurements, but rather to assess or test different possible geometries, improving the ERT imaging interpretation. The synthetic modeling was beneficial because it provided reliable and realistic hydrofacies dimension estimates that were used as critical input parameters for the T-PROGS model, which in turn improved the procedure for constructing the hydrofacies model in the subsurface. We acknowledge that our original text may have created some confusion about ERT sensitivity limitation, and we have carefully corrected this in the revised manuscript in Section 2.2.

Figure 4: How did the authors build the lens shapes in the mesh? The results are quite exciting. What is the data misfit of the traditional tomography section and synthetic model versus the real data?

In the revised manuscript, we include a step-by-step description of our methodology for estimating lens dimensions in Sections 2.2 and 3.1. The synthetic model development in ResIPy was informed by both ERT field data interpretations and borehole records. Particularly valuable were the borehole data, which provided information about depth intervals and thicknesses of individual

hydrofacies units. Some boreholes did not intersect the CSs and Sgcs layer, indicating its discontinuous, lens-shaped nature. In the revised manuscript, we included the final RMS misfit values for both the synthetic model and real data inversions in the caption of Fig. 4.

Figure 5 b: Referring to Figure 2 (which shows a 3D visualisation of the boreholes) the “Clay to silt, sandy” (CSs) hydrofacies appears to dominate the near surface, furthermore, previous studies (Karlović et al., 2021) suggest that the geology is layered. Yet in this figure (5 b) the CSs hydrofacies distribution has little lateral continuity. Can the authors comment on why that might be?

The main purpose of developing the hydrofacies model was to more accurately characterize subsurface heterogeneity. Previous aquifer characterizations relied on simplified, layer-based conceptual models that often overlooked the covering layer. As discussed in the text, this layer is typically thin or absent, suggesting high infiltration potential and increased groundwater vulnerability to surface contamination. Consequently, the layer likely exhibits limited lateral continuity near the surface. Fig. 5b was constructed using 80% of the boreholes from Fig. 2 as hard constraints, ensuring the presence of the CSs hydrofacies at borehole locations. Another important aspect is the model scale, the entire Model area 1 domain ( $27,984 \text{ m} \times 16,142 \text{ m} \times 100 \text{ m}$ ) was discretized into 1,000,000 cells ( $100 \times 100 \times 100$  in the x, y, and z directions), resulting in a horizontal resolution of approximately  $280 \text{ m} \times 161 \text{ m}$  per cell.

Line 254: What is not clear to me is which information from the electrical resistivity tomography is included or how it is included into the T-PROGS software. Is it just the lens lengths as stated on line 167?

This observation aligns with Reviewer 1's comments (RC1). In response, we have revised the text to more clearly define the role of ERT data as soft constraints within our methodological framework, specifically as ERT-derived hydrofacies lens lengths.

Figure 7 b: See above comment about Figure 5 b.

Similar to our response regarding Fig. 5b, the lateral continuity of the CSs hydrofacies is limited at the surface of Model Area 2 (Fig. 7b), but remains well-constrained at borehole locations. Note that the hydrofacies model in this figure has a horizontal cell resolution of  $20 \text{ m} \times 20 \text{ m}$ . Moreover,

Fig. 3 effectively illustrates this pattern, as high-resistivity materials, consistent with the gravel and sand observed during field ERT measurements, are frequently present at the surface of the Vinokovščak wellfield.

Lines 308 – 309: Is an improvement of 0.3 to 5.0 % statistically significant enough to argue that the inclusion of ERT has improved the outcome of stochastic modelling? My experience with Markov chain Monte Carlo methods is that the answer can differ for repeated runs by a few percentage points. If the T-PROGS software is utilising Markov chains as the authors state (Line 136) then the question is whether the results are repeatable.

Both approaches (combining borehole data with either ERT-derived lens lengths or default lens lengths) produce a range of prediction accuracies. The mean values for all tested horizontal grid resolutions ( $10 \times 10$  m,  $20 \times 20$  m,  $40 \times 40$  m,  $60 \times 60$  m,  $80 \times 80$  m, and  $100 \times 100$  m) consistently demonstrate better performance with ERT-derived lengths across 1,200 realizations (Fig. 8). While these results demonstrate repeatability, the observed improvement remains slight, as noted in the Abstract, Sections 3.2.2 and 4 (4). Other reviewers (RC1 and CC1) similarly observed the modest improvement in the prediction of hydrofacies distribution when ERT is added. In our discussion in Section 3.2.2, we propose a possible explanation to comparable results, related to similarities between ERT-derived lens lengths and T-PROGS default values (set at  $10 \times$  hydrofacies thickness). Additional details are presented in our response to CC1.

**CC1 (<https://doi.org/10.5194/egusphere-2025-327-CC1>)**

This is an interesting contribution to studies that aim to improve hydrofacies distribution by combining electrical geophysical datasets with discrete borehole logging datasets. The approach is applied to an important study area and addresses issues with nitrate contamination of critical aquifers used for water supply.

The comments posted to date identify some key areas for improvement. In particular, I agree with the need to [1] better define this contribution within the scope of previous similar (foundational?) work [2] better recognize the limitations of using smooth electrical resistivity tomography (ERT) inversions for defining discrete hydrofacies boundaries, and [3] reassess the use of the synthetic



models to assess the reliability of the estimated hydrofacies models. With respect to the third point, I am not entirely sure what was done. The approach described by the following statement is hard to follow: “This issue has been addressed by introducing synthetic models into the ERT interpretation, based on the assumption that if the ERT imaging from synthetic modeling matches or closely resembles to ERT imaging obtained from field measurements, the interpretations and resulting hydrofacies models are reliable”. As already noted, it is not reasonable to synthetically infer structures at depth beyond the resolution of the acquired field measurements. Perhaps I am misunderstanding the strategy applied. I suggest that the revisions include a better illustration/justification of the approach with a concrete example of the value added.

In response to point [1], we have reviewed the literature and incorporated relevant prior research, including studies on heterogeneity characterization and previous investigations that utilized geophysical data to constrain geological models.

We have addressed point [2] in our response to comment 3 by RC1. Although we acknowledge that the regularization parameter helps determine a stable solution, we must note that it can also introduce spurious heterogeneity in the solution. However, as demonstrated in our synthetic modeling, the Occam's inversion approach employed by R2 is very effective for determining the regularization parameter to achieve the smoothest possible model that fits the data within acceptable misfit errors, while avoiding both over- and under-fitting.

Regarding point [3], we recognize that the original text may have caused some confusion. We used synthetic modeling as a well-established and reliable tool to assess the sensitivity of our field electrode arrays, depth of investigation, and capability to electrically resolve lithological contrast and geometry between different subsurface units. Beyond these applications, the results of synthetic modeling were very useful to strengthen and validate the final hydrofacies model developed for the Varaždin study area. We have revised the manuscript to describe how synthetic modeling contributed to different aspects of our analysis. Below, we summarize the specific purposes and advantages of our synthetic modeling approach:

1) To validate the reliability of field ERT measurements to depths of 20 m with a very high quality/noise ratio. More precisely, this depth corresponds to the appearance of conductive CSs/Sgcs layer. This was confirmed by synthetic models, resolving the lithology and geometry

within this depth range, suggesting a range of resistivity values consistent with those from field measurements.

2) To detect and validate the influence of the Sgcs and CSs hydrofacies at 20-23 m depth. With synthetic models, we were able to estimate the conductivity/resistivity magnitude of these materials, confirming their role in limiting the depth of investigation of our field ERT measurements. Synthetic modeling results were very useful, demonstrating that the Sgcs/CSs materials exhibit high conductivity at these depths, greater than initially anticipated. This explains the conductive anomaly visible in our ERT imaging (Figs. 3 and 4).

3) To determine whether the Sgcs/CSs materials at 20-23 m depth form lens-shaped features or constitute a single layer. The Sgcs/CSs units are not at depths beyond resolution, but are rather responsible to limit the depth of investigation and resolution. This was the main challenge concerning geophysical analysis. Using synthetic models, we wanted to evaluate the effects of these two possible scenarios (lens-shaped and single layer) to establish the most probable geometry and resistivity magnitudes of these materials. The lens-shaped geometry shown in Fig. 4 emerged as one plausible configuration. Our acceptance criteria for this interpretation rely on strong similarities in the shape of the conductive anomaly appearing at 20 m depth in the synthetic model and the equivalent anomaly in the field-data inversion model. We note that the thickness of the layer is consistent with well-constrained measurements from borehole data. Moreover, since this layer is absent in certain boreholes, we infer that it is not a continuous across the study area.

We acknowledge that the original text lacked clarity regarding this analysis. In response, we have substantially revised Sections 2.2 and 3.1 to provide a better explanation of our strategy applied and justification for the modeling approach. With these textual improvements clarifying the purposes and results of our approach, we believe the illustrated example in Fig. 4 now appropriately demonstrates our findings.

One additional point to consider in the revisions is that the paper implicitly assumes that variations in grain size primarily control variations in electrical resistivity. Of course, resistivity also depends strongly on porosity, fluid conductivity, and the degree of saturation. Although variations in porosity sensed with resistivity could contribute to improved hydrofacies discrimination, variations in fluid conductivity and degree of saturation would likely complicate the delineation

of hydrofacies. Is there any field data available to constrain the magnitude of these variations? Presumably, water tables are available in the wells. Significant variations in fluid conductivity might be expected given the nitrate contamination problem. Perhaps these complicating factors can partly explain why the improvement in the prediction of hydrofacies distribution is modest when ERT is added? Some further discussion of this issue is required.

This paper will make a valuable contribution to the hydrogeophysics literature once the comments provided in the posted discussions are addressed.

While we acknowledge that electrical resistivity depends on multiple factors including porosity, fluid conductivity, and degree of saturation, our study considers grain size variations as the primary control for hydrofacies discrimination. Although variations in fluid conductivity and saturation could potentially complicate facies delineation, our observations in Model Area 2 demonstrate limited saturation effects for hydrofacies G. Long-term water table records (2005–2019) from observation wells (prefix SPV; Fig. 1) reveal water table fluctuating between 3.68 and 10.95 m, depending on the hydrological season, wellfield pumping regime, proximity to pumping wells, and Drava River levels. More recent field measurements from monthly campaigns in the Vinokovščak wellfield area show a stable water table, ranging from 4.49 m to 7.65 m (median: 6.74 m). Field measurements from 2018-2022 revealed stable groundwater electrolytic conductivity (479-501  $\mu\text{S}/\text{cm}$ , median = 492  $\mu\text{S}/\text{cm}$ ), indicating consistent total dissolved solids content. Due to its negative charge, nitrate is mobile, resulting in conservative transport behavior. We expect that nitrate mobility follows the hydrofacies sequence:  $G > G_{sc} > S_{gcs} > CSs$ . In the Varaždin aquifer, nitrate concentrations are typically higher in shallow zones due to proximity to contamination sources. Unfortunately, multi-depth screened wells at the Vinokovščak wellfield yield mixed samples during pumping, preventing depth-specific parameter analysis and any direct correlations with resistivity measurements. As shown in the Fig. 4, the high-resistivity lens associated with hydrofacies G at the beginning of the profile exhibits comparable resistivity values in both dry (surface) and saturated (deeper) zones. Variations between lenses along the profile (500  $\text{ohm}\cdot\text{m}$  in the field data model, and 700  $\text{ohm}\cdot\text{m}$  in the synthetic model computed using average resistivity values) suggest these differences are primarily spatial rather than related to saturation state.

Degree of saturation in the vadose zone is expected to be a function of lithology, i.e. it decreases with increasing the grain size and pore radius. At shallow depths, the resistivity of Sgcs-CSs material exhibits resistivity values twice as high as those at 20-25 m depth in the saturated zone (low resistivity, 60 ohm·m from inverted field data, 140 ohm·m estimated from synthetic model and 20 ohm·m suggested as the true resistivity value from synthetic model). The hydraulic conductivity (K) of the four hydrofacies units is the subject of a separate paper currently under review. While these results are not yet published and we cannot discuss them in detail, our methodology involves estimating K values from grain-size distribution curves using empirical methods, followed by validation against pumping test data. The derived K ranges for individual hydrofacies show clear separation, with only minor overlaps occurring in transitional deposits (Sgcs and Gsc) that vary in their sand and gravel ratios, as extensively documented in the literature.

However, our analysis revealed that similarities between ERT-derived lens lengths and default T-PROGS values (set at  $10 \times$  hydrofacies thickness) may account for the comparable prediction outcomes between the two approaches, a point we have now included in our discussion in Section 3.2.2. The ratios of ERT-derived to default lens lengths vary across different hydrofacies, with hydrofacies G showing ratios of 2.9–3.1, Sgcs 1.8, and CSs 1.2–1.5 (derived from the data in Table 1). While hydrofacies G exhibits the greatest dimensional discrepancy, its volumetric predominance within the system compensates for these variations, making it the dominant control on the overall hydrofacies model.