This paper reports on several challenges addressed by the authors while pursuing this study. The paper is interesting to read, primarily because it raises several interesting issues and challenges.

The main concern I have is that the approach pursued here, although referred to as stochastic, does not present a complete and sound statistical theory. This shows up in different ways:

- The approach presented in this paper combines statistical concepts (such as transition probabilities) with multiple decisions that are based on aesthetics and subjective judgment calls. The judgment calls made by the authors are not translated into statistical rules. As a result, the ensemble of realizations that are generated does not constitute a statistically-meaningful ensemble. Rather, it represents a bunch of (presumably and subjectively) reasonably-looking realizations. This creates a confusing mix of statistical models and art. What constitutes "reasonable" and less "reasonable" is not clear. It is a judgment call, and hence it crosses the boundary between science and art. Without a statistically-meaningful ensemble, meaning, without a complete representation of high and low-probability realizations, it becomes impossible to quantifying predictions. Quantifying predictions is the main goal of such studies, and when that is not made possible, what is the point?
    - Along this line, I should refer to the authors' comment on p. 15221 (line 20) concerning ".. equally likely probable realizations..". First, if the authors believe the realizations are equally-probable, they need to show how these probabilities are computed. Next, I believe that in the absence of a multi-point statistical model, as is the case here, one cannot really use the term "equally-likely". At best these are realizations that were produced following a similar procedure (on p. 15221 line 21, the authors refer to the realizations they produce as "plausible", which I think is more appropriate) . Lastly, the authors should explain why they see merit in producing equally-probable realizations. I believe that in application we are interested in a variety of probabilities, high-and low-probabilities: we know how to prepare

for high-probability events. It is the low-probability (less plausible?) events that lead to disasters.

- No attempt is reported in this study on computing or employing statistics beyond single- and two-point correlations. This opens up the question of the implications of neglecting higher-order statistics.
- On p. 15220 (line 20), the authors mention five criteria used for validation. This statement and the strategy it represent raise several challenges:
  - First, validation is not possible in groundwater applications. This was pointed out in a paper by Naomi Oreskes (http://www.likbez.com/AV/CS/Pre01-oreskes.pdf) and echoed by many scientists ever since. The best the authors could say on this context is that they examined their realizations from five different perspectives.
  - The five criteria used for validation represent information that was known a-priori. One could argue that as such, these criteria could have affected the judgment calls made by the authors along the way, and hence they do not represent independent and unbiased evaluation criteria.
  - In my opinion, the information represented by the 5 evaluation criteria should be used to construct statistical priors in a Bayesian sense. In this context, there are several papers I should mention, including cf., Woodbury and Rubin, 2000, Hou and Rubin, 2005.

A few more comments in other directions:

1. On p. 15222, line 24, the authors state that "Until now there are no published studies on the incorporation of a comprehensive and continuous soft conditioning datasets..". To my knowledge this is not accurate. As an example, I should mention the concept of anchors, discussed in Rubin et al., (2010) which can be used to condition on so-called "soft" data. Anchors can be used to represent data of all sorts using statistical distributions.
2. On p. 15223 line 4 (and on multiple other locations) the authors refer to "overconditioning" [sic]. The authors do not define what they mean with this term, and my interpretation of it as that it means some sort of

challenge related to highly-dense data used for conditioning. This is supported by a statement made on p. 15234 line 13 that "The observed problem of overconditioning is caused by spatially correlated data which are incorporated into the modeling process". The relationship with spatially-correlated data is correct, in my opinion, only that this is an avoidable problem, because it is an outcome of the authors' decision to use kriged data for conditioning. This decision needs to be revisited. Kriging produces point estimates that are optimal in some sense. It does not create, and is not intended to create, fields that are defined the geostatistical models that are used for kriging. Kriging is a smooth interpolator, not a random field generator. Kriging eliminates important variability, and cannot be used for conditioning (see Rubin, 2003, p. 60 and p.71, discussion on estimation vs. simulation). Kriging produces unrealistic and inflated correlation lengths. These correlation lengths do not represent spatial variability of the geophysical variables, because they are obtained from a graphic representation of kriging estimates. This is possibly the reason for the effect referred to by the authors as overconditioning. As an alternative, I would suggest to the authors to generate realizations of the geophysical data for conditioning. I would possibly represent the geophysical data using a series of anchors (each defining a statistical distribution (Rubin et al., 2010).Then, for simulation, I would suggest using a nested structure approach (see Maxwell et al., 2000) which involves (a) generating random skytem field realizations, followed by (b) using each of these realizations as a starting point instead of the kriged estimates. An alternative would be to convert the geophysical data into anchor representation of the facies at selected location, and use these anchors as a starting point for simulation. (Rubin et al., 2010; Murakami, 2010)

A few additional comments:

3. Single point cross-correlations: show examples, explain how done. Explain how the discrepancy between the scale of the borehole measurements on one hand and the scale of the geophysical data as accounted for. In MAD (Rubin et al.,

2010) a case is made that anchors could be used to account for that (a scaling model is needed). Additionally:

- The use of statistical correlations to relate between the geophysical and geological attributes pursued in this study is reported very scantly. It is not clear how good or bad these correlations are, and this needs to be discussed.

- Please discuss and demonstrate the implications of using single- rather than multi-point statistics.

- There is an extensive body of work on the use of petrophysical models for relating the geophysical and geological attributes (Rubin et al., 1992; Mavko et al., 2009). It would be interesting to know if the statistical correlations provided better results compared to physically-based, statistical models.

4. On several occasions in this paper the authors point out that conditioning is producing a trends (e.g., p. 15220 lines 19-20). Stated differently, trends are identified in application. This is a problem because the existence of a deterministic trend indicates that the trend was not removed prior to computing the two-point correlations, which violates the requirement for stationarity (Rubin, 2003, p. 58). When a trend exists, it must be accounted for a-priori, and not as an outcome.

5. On page 15220, line 13, the authors identify “the incorporation of two distinct datasources [sic] into the stochastic modeling…..sparse borehole data and abundant SkyTEM data” as the “novelty of this study”. In making this statement, the authors should recognize the large body of published work that did precisely the same, including: Rubin et al., 1992, Copty and Rubin, 1995, Hubbard and Rubin, 2000, Hubbard et al., 2005, Hou and Rubin, 2005, Kolwaksy et al., (2001, 2004).

References:

Copty, N., and Rubin, Y., Stochastic approach to the characterization of lithofacies from surface seismic and well data, Water Resour. Research, 31 (7), 1673 - 1686 1995.

Dai, Z., R.W. Ritzi., C. Huang, Y. Rubin, D. Dominic, Transport in heterogeneous sediments with multi-modal conductivity and hierarchical organization across scales, J. of Hydrology, 294, 68-86, DOI: 10.1016, j.jhydrol, 2003.10.024, 2004.

Hou, Z., and Y. Rubin, On minimum relative entropy concepts and prior compatibility issues in vadose zone inverse and forward modeling, Water Resour. Res., 41(12), W12425, doi:10.1029/2005WR004082, 2005

Hubbard, S., Chen, J., Peterson, J., and Y. Rubin, Hydrogeological characterization of the South Oyster Bacterial transport site using geophysical data, Water Resour. Res., 37(10), 2431-2456, 2001.

Hubbard, S., Rubin, Y., Hydrogeological parameter estimation using geophysical data: A review of selected techniques, J. Contam. Hydrology, 45, 3-34, 2000.

Kowalsky, M.B., Dietrich, P., Teutsch, G., and Y. Rubin, Forward modeling of GPR data using digitized outcrop images and multiple scenarios of water saturation, Water Resour. Res., 37(6), 1615-1625, 2001.

Kowalsky, M., S. Finsterle, Y. Rubin, Estimating flow parameters distributions using ground-penetrating radar and hydrologic measurements during transient flow in the vadose zone, Adv. Water Resour., 27(6), 585-599, 2004.

Mavko, G., T. Mukerji, J. Dvorkin, The Rock Physics Handbook, Cambridge University Press, 2009

Maxwell, R., Kastenberg, W., and Rubin, Y., Hydrogeological site characterization and its implication on human exposure risk assessment, Water Resour. Res., 35(9), 2841-2855, 1999

Murakami, H., X. Chen, M. S. Hahn, Y. Liu, M. L. Rockhold, V. R. Vermeul, J. M. Zachara, and Y. Rubin, Bayesian approach for three-dimensional aquifer characterization at the Hanford 300 Area, Hydrol. Earth Syst. Sci., 14, 1989-2001, doi:10.5194/hess-14-, 1989-2010, 2010.

Ritzi, R.W., Z. Dai, D. Dominic, and Y. Rubin, Spatial correlation of permeability in cross-stratified sediment with hierarchical architecture, Water Res. Research, 40(3), 2004.

Rubin, Y., Mavko, G. and Harris, J., Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, Water Resour. Research, 28(7), 1992.

Rubin, Y., Applied Stochastic Hydrogeology, Oxford University Press, 2003

Rubin, Y., X. Chen, H. Murakami, and M. Hahn, A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, *Water Resour. Res., 46*, W10523, doi:10.1029/2009WR008799, 2010.

Seong, K. and Rubin, Y, Field investigation of the WIPP Site (New Mexico) using a non stationary stochastic model with a trending hydraulic conductivity, Water Resour. Res., 35(4), 1999.

Woodbury, A., and Rubin, Y., A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod Experimental site, Water Resour. Res., 36(1), 159-171, 2000