Accounting for Correlated Data Errors during Inversion of Cross-Borehole Ground Penetrating Radar Data

Cordua, Knud Skou; Looms, Majken C.; Nielsen, Lars

Published in:
Vadose Zone Journal

Link to article, DOI:
10.2136/vzj2007.0008

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Accounting for Correlated Data Errors during Inversion of Cross-Borehole Ground Penetrating Radar Data

Knud S. Cordua, Majken C. Looms, and Lars Nielsen*

Least-squares tomographic inversion of cross-borehole ground penetrating radar (GPR) data was used to estimate the radar wave velocity distribution in the upper ~10 m of the vadose zone at a field site in northern Zealand, Denmark. The radar wave velocities were transformed to values of water saturation and formed a basis for hydrologic studies of flow characteristics of the vadose zone. Cross-borehole GPR data are likely to be contaminated by correlated data errors that may give rise to significant artifacts in the inverse estimate of the velocity distribution if they are not properly accounted for during the inversion process. We analyzed two sources of correlated data errors (unknown cavities and small-scale clay-enriched zones close to the borehole walls), which we assumed to play a significant role in the cross-borehole GPR data sets collected at our field site. The correlated errors may be accounted for by specification of data error covariance matrices, which are included as a priori knowledge in the inverse operator used in the tomographic algorithm. The study indicates that proper accounting for correlated data errors significantly suppresses the effects of these data errors and results in trustworthy inverse estimates of the radar wave velocities between the boreholes. Suppression of the influence of the correlated data errors is a necessity for obtaining realistic hydrologic models of the vadose zone. Using popular inverse methods that disregard the data error correlation properties can lead to severe artifacts in the inverse model estimate and, therefore, can lead to incorrect hydrologic interpretations. Our findings are based on both synthetic tests and a real data example.

ABBREVIATIONS: GPR, ground penetrating radar.
of the standard deviation and correlation properties of the data errors. These data error properties can be described in the data error covariance matrix of the least-squares inverse operator, and thereby the data errors may be accounted for during inversion (Jackson, 1979; Tarantola, 1987; Nielsen and Jacobsen, 1996).

We used cross-borehole GPR investigations for studying water saturation and flow characteristics of the vadose zone in a sandy environment in northern Zealand, Denmark. In our study area, we expect unknown cavities in the borehole walls and small-scale low-velocity anomalies close to the source and receiver positions to give rise to strongly correlated data errors. These types of data errors are often referred to as static errors and they may have significant influence on the quality of cross-borehole tomographic inverse estimates (e.g., Squires et al., 1992; Maurer and Green, 1997; Vasco et al., 1997). In this study, we investigated the influence of such correlated data errors on our GPR data, and we assessed the standard deviation and correlation properties of the errors. We outline below how data error covariance matrices, which are included in the inverse operator, may be specified for the expected correlated data errors. In a synthetic study, we demonstrate how the data errors propagate into the inverse estimate and affect model resolution. Further, we applied the method to a field data example from our study area. We show that the assumptions made about data errors and their correlation during inversion are crucial for the quality of the resulting inverse estimate.

Static-like errors in cross-borehole seismic or GPR data sets are traditionally treated as source- or receiver-dependent extra model parameters, which are estimated from the travel time data during the inversion process (Maurer and Green, 1997; Vasco et al., 1997). We discuss the main differences between our approach and the traditional approaches used for suppressing static-like errors.

**Site and Data Acquisition Description**

Our field site is located on the peninsula of Arrenæs, northern Zealand, Denmark (Fig. 1). Here, the vadose zone is approximately 25 m thick. The sediments of this zone are loosely packed glacial deposits, which consist mainly of sandy or gravelly material. Minor intercalations of clay-rich sediments exist in the area. These sand- and gravel-dominated glacial deposit sediments have an average radar wave velocity of 0.13 to 0.14 m/ns (Looms et al., 2008). The boreholes used for GPR investigations have been placed in a regular geometry (Fig. 1b). The boreholes were drilled with a 0.10-m-wide auger drill stem, and 0.05-m-wide polyvinyl chloride (PVC) tubes for GPR antenna access were placed in these boreholes shortly after drilling. The antennae, which were provided by Sensors & Software Inc., Mississauga, ON, Canada, have a nominal frequency of 100 MHz, and are 0.026 m in diameter. The boreholes were drilled to a depth of 14 m. When the PVC tubes were lowered into the boreholes, it was discovered that the lowermost 1 to 2 m of the boreholes were filled with sediments. Thus, a significant amount of the loosely packed sediments must have fallen from the borehole walls to the bottom of the borehole shortly after drilling. The void between the borehole walls and the PVC tubes was backfilled with the material brought to the surface during the drilling process. It is very likely, however, that some significant cavities are still present close to the borehole walls.

We used a ray geometry consisting of a total of 702 rays for our cross-borehole tomographic GPR setup in this study (Fig. 2). The spacing between neighboring transmitter and receiver positions in the two boreholes is 1.0 and 0.25 m, respectively. All rays dipping >45° were omitted from the data set. The two-dimensional model section between the two boreholes was parameterized using a regular grid of cells, where each cell has constant slowness (inverse velocity). The grid cell length is 0.50 m. For each transmitter–receiver pair, we picked the first arrival travel time to be able to estimate the radar wave velocity by tomographic inversion. In general, the signal/noise ratio of the recorded waveforms was high, and the picking uncertainties were very low and mainly determined by the temporal sampling interval of the data set.

**Correlated Data Errors and Covariance Specification**

Several sources of data error exist in cross-borehole GPR tomographic problems. Some examples of important error sources are (see Peterson [2001] for an exhaustive analysis of different error types that may influence cross-borehole GPR data sets):
(i) unknown cavities in the borehole walls close to the receiver or transmitter antenna, (ii) unknown small-scale anomalies close to the antenna positions, (iii) incorrect positioning of the transmitter or receiver antennae during data acquisition, (iv) time jumps or time drift due to problems with transmitter calibration, (v) incorrect information about the geometry of the two boreholes, (vi) mispicking of the travel time of the first arrival, (vii) electrical background noise from the surroundings or the GPR instrumentation.

Data error sources (i) through (v) will result in correlated data errors. Only error sources (vi) and (vii) may be expected to result in uncorrelated data errors. The unknown cavities and small-scale anomalies close to the borehole walls are treated as sources of data error because they are believed to be smaller than the grid cell size, and, therefore, they cannot be captured by the model grid. Furthermore, they are positioned in areas where the spatial resolution provided by the cross-borehole data is low due to a limited number of crossing ray paths. The influence of the different error types will vary from data set to data set, since they strongly depend on the local geology, instrumentation, and background noise level.

In the following, we discuss why we expect errors sources (i) and (ii) to significantly contaminate our cross-borehole GPR data sets, and we show how these error types may be described through data error covariance specification. As described above, observations made during drilling of the boreholes strongly indicated that material had fallen from the borehole walls to the bottom of the boreholes shortly after the drilling process had ended, and small-scale clay-rich anomalies are known from other investigations to exist in the study area. The cavities in the borehole walls will generate correlated errors for data collected using receiver or transmitter positions close to the cavities. In the vadose zone, the cavities will primarily be filled with air, which has a radar wave velocity of 0.3 m/ns, i.e., approximately twice the average velocity of the sediment that is studied here. If the radius of the unknown cavities is on the order of 0.05 to 0.25 m, then the unknown cavities may cause all travel times collected close to the cavities to be ~0.5 to 1.5 ns early. Furthermore, we know from wells drilled in the area that minor intercalations of clay-rich material may be found in the overall sandy–gravelly sediments of the study area (well information from the area is available in the Jupiter borehole database at www.geus.dk/jupiter/index-dk.htm [in Danish, verified 13 Sept. 2007]). The clay-enriched material is expected to have a radar wave velocity of about 0.06 m/ns, which is about half the velocity of the overall sedimentary material under study. Thus, even small (0.05–0.25 m in diameter), unknown anomalies enriched in clay may cause all the data collected using a specific receiver or transmitter position to be about 0.5 to 2 ns late if these anomalies are located close to the receiver or transmitter antenna position. These anomalies are too small to be captured by the model parameterization. Therefore, we choose to regard the influence of such small-scale anomalies as errors of the data set.

A priori information about data errors and data error correlation is specified in the data error covariance matrix, \( C_D \), which is part of the least-squares inverse estimate of the model parameters (e.g., Tarantola, 1987):

\[
\begin{align*}
\langle m \rangle &= m_{\text{prior}} + \\
(G^T C_D^{-1} G + C_M^{-1})^{-1} G^T C_D^{-1} (d_{\text{obs}} - G m_{\text{prior}})
\end{align*}
\]

where \( m_{\text{prior}} \) contains the a priori values of the slowness distribution of the grid of cells describing the two-dimensional section between the two boreholes, \( C_M \) is the model covariance matrix, \( G^T \) is the linear mapping matrix expressing the relation between model parameters and observed data. Matrix \( G \) contains the distances that the different rays have traveled in the individual cells.

The data error covariance matrix is a square and invertible matrix consisting of \( N \) by \( N \) elements, where \( N \) is the total number of picked travel times. For more details regarding the characteristics of data error covariance matrices, we refer to Tarantola (1987). The choice of \( C_D \) has significant impact on the inverse estimate. The diagonal elements of \( C_D \) contain the variance of the individual data errors, whereas the off-diagonal elements describe the correlation between the individual data errors. In the special case where all data errors are assumed to be uncorrelated and have the same standard deviation, the data error covariance matrix takes on the familiar form

\[
C_D = \sigma^2 I
\]

where \( I \) is the identity matrix, and \( \sigma \) is the standard deviation of the data errors.

Error types I and II are spatially correlated, and the spatial correlation properties must be specified in the off-diagonal elements of the data error covariance matrix. Clearly, the unknown cavities and low-velocity anomalies strongly affect data collected with a transmitter or receiver position close to the location of a cavity or a low-velocity anomaly. We specify separate data error covariance matrices related to the transmitter and receiver positions (\( C_{T_r} \), and \( C_{R_e} \) respectively). For simplicity, we refer to the corresponding correlated data errors as the transmitter static errors and the receiver static errors, respectively, in the following. The term static is borrowed from the reflection seismic literature, where this term is used for describing seismic travel time data errors, which are (almost) the same for all seismic data recorded at a specific location. We choose a covariance function to describe the different elements of \( C_{T_r} \) and \( C_{R_e} \) (e.g., Walden and Hosken, 1985). For \( C_{T_r} \) we have

\[
C_{T_r}(i, j) = \sigma^2 \exp[-s(i, j)/L]
\]
where \( s(i,j) \) is the distance between the \( i \)th and the \( j \)th transmitter position, and \( L \) is the spatial correlation length. Matrix \( C_{Re} \) is specified in a similar way.

The covariance matrix for uncorrelated data errors related to mispicking of the first arrival may be set up using Eq. [2]. As mentioned above, the picking uncertainty is low for our data set, which offers a high signal/noise ratio for the recorded waveforms. In the synthetic study below, where we focused on the effects of expected correlated error types, we chose a low standard deviation of 0.1 ns for the uncorrelated data errors.

**Simulation of Correlated Data Errors**

In a synthetic test study, we used a spatial correlation length, \( L \), of 5 m and a standard deviation, \( \sigma \), of the correlated data errors of 0.5 ns for setting up \( C_{Tr} \) and \( C_{Re} \) using Eq. [2]. Matrices \( C_{Tr} \) and \( C_{Re} \) are illustrated in Fig. 3a and 3c, respectively. For simplicity, only 200 measurements were used to create these figures. Note the blocky appearance of \( C_{Tr} \). Each of the white blocks following the diagonal of \( C_{Tr} \) describe the perfect correlation of the data errors of all the travel times coming from a specific transmitter position. The gray tones of the other blocks show the amount of correlation of the data errors originating from different transmitter points. Matrix \( C_{Re} \) also has a blocky appearance but shows different correlation properties within the individual blocks due to the way the travel-time data set is organized: The transmitter antenna is lowered to a certain position in the borehole. From that position, radar signals are transmitted to all the different receiver antenna locations in the other borehole. Then the transmitter is lowered 1.00 m to the next transmitter position, and signals are again transmitted to all receiver locations in the other borehole. The receiver positions are only 0.25 m apart. This procedure is continued until all transmitter locations in both boreholes have been visited (Fig. 2). The data error covariance matrix related to the uncorrelated data errors (standard deviation of 0.1 ns) is specified using Eq. [2] and illustrated in Fig. 3e.

Based on the specified data error covariance matrices, we may simulate the corresponding data noise following the scheme outlined by Jacobsen (1993). Figure 3 shows simulations of transmitter static errors, receiver static errors, and the uncorrelated data errors. The strongly correlated nature of the static errors is evident. The transmitter static errors have a step-like appearance. In contrast, neighboring receiver static errors are typically different; however, the receiver static errors are repeated in a cyclic manner through the data set. These appearances of the static error types are a consequence of the data sampling scheme outlined above.

**Inversion of Synthetic Data with Correlated Errors**

In a synthetic test, we assessed the influence of the correlated errors on the least-squares tomographic inverse estimate of the slowness distribution between a set of boreholes placed 5 m apart. Using the ray geometry of Fig. 2 and an anomaly test pattern with a known slowness perturbation in the central part of the model grid (Fig. 4d, 4h, and 4l), we generated a noise-free first arrival travel-time data set. We then set up the data error covariance matrices for the transmitter and receiver static errors as well as the uncorrelated mispicking errors using Fig. 3. Subsequently, we added the three data error covariance matrices and created the total error budget, which we then added to the noise-free synthetic travel-time data set to obtain a synthetic data set including correlated data errors.

The synthetic data set contaminated by the correlated transmitter and receiver static errors as well as the uncorrelated mispicking errors was inverted using Eq. [1]. We used an a priori model in which all cells were assumed to have a slowness equal...
to the background slowness of the true anomaly test pattern. We performed nine different inversions that differed only in the choice of a priori data error covariance matrix.

First, we performed the inversion disregarding the two correlated data error types, i.e., we used an a priori data error covariance matrix that was a diagonal matrix containing only the variance of the uncorrelated data error (Fig. 3c). The inversion result is shown in Fig. 4a. The test anomaly in the central part of the model grid is indicated in the inverse estimate, although it is smeared and its shape is distorted (compare Fig. 4a and 4d). The correlated data errors have a significant effect on the inverse estimate, and they give rise to pronounced artifacts in the estimated velocity model, especially, but not only, close to the boreholes. Such effects of static-like errors are expected and have also been noted for seismic cross-borehole investigations (e.g., Squires et al., 1992). When severe errors are assumed to be present in a given data set used for tomographic inversion, the standard procedure is often to simply increase the standard deviation of the a priori data uncertainty above the level of the expected picking uncertainty to damp the least-squares solution. Therefore, we again performed the inversion just accounting for uncorrelated data errors, but this time we increased the standard deviation of the a priori data uncertainties from 0.1 to 0.7 ns to account for the total amplitude of the composite error budget (Fig. 4b). The artifacts related to the correlated data errors are damped but still evident. The estimation of the central test anomaly has suffered from additional smoothing. The approach where we only account for uncorrelated data errors can partly suppress the unfortunate effects of the correlated data errors if the a priori standard deviation is increased to 1.4 ns, which is twice the standard deviation of the total error budget of the synthetic data set (Fig. 4c). In this case, however, the resolution of the central test anomaly is significantly improved and the amplitude of the artifacts is reduced compared with the previous inversion result (compare Fig. 4f to Fig. 4e). The main artifacts are now observed along the top and bottom of the model grid where the resolution of the data set is poor due to the limited ray coverage in these areas (cf. Fig. 2). For the estimate of Fig. 4g, we again made the correct assumption that \( L = 5 \) m, but we

FIG. 4. Synthetic test study: (a–c), (e–g), and (i–k) inversion results; (d, h, and l) true slowness test pattern that should ideally be fully recovered by inversion. In (a–c), data error correlation was not taken into account during the inversion process. In (e–g) and (i–k), the inversion was performed assuming that the data errors were correlated; different assumptions about the correlation properties and the standard deviations of the correlated data errors were made. Only in (j) was the inversion performed using the correct data error covariance matrix of the total error budget.
overestimated the standard deviations of the correlated static errors by a factor of 4 (2 ns). The central test anomaly is well resolved, and the main artifacts observed in the inverse estimate are limited to the areas where the ray coverage was poor. For the inverse estimates in Fig. 4i to 4k, we made the correct assumptions regarding the standard deviation of the different error types. These three estimates only differ in the choice of $L$. From Fig. 4i to 4k, $L$ increases in steps: 0, 5, and 10 m. All three inversion results provide good resolution of the central test anomaly, although the estimate shown in Fig. 4k ($L = 10$ m) exhibits slightly more smearing than the estimates of Fig. 4i and 4j; however, the differences are minor. Significant artifacts of the inverse estimates are, again, limited to the regions of the model with the poorest ray coverage (top and bottom).

Taking the correlation of the data errors into account resulted in significantly better recovery of the central test anomaly than the popular inversion approach in which correlation of the data errors is disregarded. The synthetic test examples indicate that the approach of taking data correlation into account is robust. Even if the standard deviation of the correlated data errors is underestimated by a factor of 2 or overestimated by a factor of 4, the inversion scheme provides good resolution of the central test anomaly and the influence of artifacts is limited. Over- or underestimating the value of the correlation length, $L$, by 100% results in essentially the same inversion results as the case where the correct choice of $L$ is made.

**Tomographic Inversion of Field Data**

Tomographic data sets were obtained using the GPR boreholes at the Arrenes field site (Fig. 1b). We inverted and interpreted first arrival travel-time data acquired between Boreholes 1 and 3 using the source–receiver geometry illustrated in Fig. 2. Looms et al. (2007) described several data sets obtained at the Arrenes field site and discussed the hydrogeologic implications of the geophysical interpretations. We inverted the field data, assuming different standard deviations and correlation properties of the data errors (Fig. 5). The inverse estimates that we found and discuss below only differ in the choices we made regarding a priori data error correlation and standard deviation.

First, we inverted the tomographic data using the popular approach where data error correlation is not taken into account (Fig. 5a–5c). We used an a priori data error standard deviation of 0.8 ns to obtain the result of Fig. 5a. Overall, the velocity model of Fig. 5a is characterized by relatively high GPR wave velocities above $\sim7.5$-m depth and lower velocities below this depth. In the uppermost 7.5 m, a high-velocity anomaly is present in the central part of the model grid. Marked high-velocity anomalies are found at 2- to 3-m depth near both borehole walls. Below $\sim3$- to 4-m depth, low-velocity anomalies are found along both boreholes. In the area of lower velocities below $\sim7.5$-m depth, a pronounced low-velocity anomaly is found just below 8-m depth about 1 m to the right of the leftmost borehole. Thus, marked two-dimensional variations are observed in the velocity field. The high-velocity areas could be interpreted as sandy or gravelly sediments with low water saturation, whereas the low-velocity areas could be interpreted as areas with an increased content of fine-grained material or increased water content (e.g., Reynolds, 1997). The strong lateral variations in the velocity field are inconsistent, however, with results from interpretation of GPR reflection data sections, which show that, overall, the geologic layering of the subsurface is horizontal to subhorizontal, with average dips of $\sim6^\circ$ and maximum dips of $\sim10^\circ$ (see Hansen et al. [2008] for illustration of the GPR reflection data). We suspect that the observed strong lateral velocity variations may be caused by unknown cavities in the borehole walls or local, small-scale, low-velocity anomalies near the boreholes, as discussed above. We sought inverse estimates where these unrealistic anomalies are suppressed. The inverse estimate illustrated in Fig. 5b was found assuming uncorrelated data errors with a standard deviation of 1.9 ns. The two high-velocity anomalies near the boreholes are now partly suppressed and the central high-velocity anomaly above 7-m depth is smeared. Relatively strong lateral variations are still present in the model, however, although all velocity structures appear more smoothly smeared than the model of Fig. 5a. Increasing the standard deviation of the assumed uncorrelated data errors to 3.8 ns resulted in the inverse image of the velocity field shown in Fig. 5c. Here, all local anomalies are significantly suppressed, and, in general, the velocity field is smooth, with more limited lateral velocity variations.

To suppress the unrealistic velocity anomalies in the inverse estimates while keeping a high degree of model resolution with limited smearing in the central parts of the model grid, we inverted the travel-time data while accounting for correlated, static-like errors through a priori data error covariance specification as outlined in the synthetic studies above. Inversion results obtained assuming different characteristics of the static-like data errors are shown in Fig. 5d to 5h. In these five inversions, we also assumed uncorrelated data errors with a standard deviation of 0.8 ns. Thus, we assumed all three data error types illustrated in Fig. 3 to be present in the field data. For the three inverse estimates of Fig. 5d to 5f, we used $L = 5$ m, whereas we made different choices regarding the standard deviation of the correlated data errors. In Fig. 5d, we assumed a standard deviation of 0.5 ns for the two correlated data error types. Although partly suppressed, the unrealistic anomalies near the boreholes and the high-velocity anomaly in the central part of the model grid are still evident in the tomographic image. Increasing the a priori standard deviation of the assumed correlated data errors to 1.2 ns (Fig. 5e) and 2.0 ns (Fig. 5f) resulted in velocity models showing less overall lateral variation. Note that in Fig. 5f the anomalies near the boreholes have been effectively suppressed. Furthermore, the high velocities above $\sim7.5$-m depth are not centered around the middle of the model grid. Instead, we observe a more laterally continuous high-velocity body terminating at about 7.5-m depth, consistent with the observations made in the reflection GPR data. Note that the low-velocity anomaly just below 8-m depth seems to be well constrained by independent crossing ray paths and has not suffered smearing although the standard deviation of the total error budget has been significantly increased. This anomaly could represent a small anomaly enriched in fine-grained material (clay and silt). Setting $L$ to 0 m while keeping the standard deviation of the correlated data errors at 2.0 ns resulted in the inverse estimate of Fig. 5g, whereas increasing $L$ to 10 m (still keeping the correlated data error standard deviation at 2.0 ns) resulted in the velocity distribution shown in Fig. 5h. The results of Fig. 5g and 5h show much resemblance to the results of Fig. 5e and 5f, although slightly more lateral variation is observed in the velocity field of Fig. 5g. Thus, in this real data example also, our approach seems to be robust: a broad selection of a priori standard deviations and correlation lengths of the
correlated data errors resulted in inverse estimates in which the general features are almost identical.

**Discussion**

Several sources of data error exist in cross-borehole GPR studies, and some error types are strongly correlated. Based on studies related to our field site in northern Zealand, Denmark, we investigated the influence of selected sources of correlated data errors (unknown irregularities of the borehole walls and small-scale anomalies close to the borehole walls), which are believed to have a significant effect on our cross-borehole GPR data. We found that these correlated error types had a significant negative effect on the inverse estimate of the radar wave velocity distribution if not properly accounted for by careful a priori data error covariance specification. They introduced pronounced positive and negative velocity anomalies not only close to the boreholes but also in other parts of the model grid if a standard least-squares operator, which does not take data error correlation into account, was used. The synthetic studies show that if we account for data error correlation, significantly improved inverse estimates may be obtained even if the correlation properties are only approximately known. The inversion approach used to suppress the effects of the correlated data errors is robust: a wide range of a priori standard deviations and correlation lengths of the correlated data errors resulted in inverse estimates in which the recovered test anomaly patterns are similar. In the field data example, strong isolated high-velocity anomalies are present in the inverse estimates if data error correlation is not accounted for. These high-velocity anomalies could well represent cavities in the borehole walls, especially since they occur in the uppermost 2 to 3 m of the boreholes, where we expect the sediments to be most loosely packed due to lack of compaction. The undesired high-velocity anomalies can only be suppressed by the popular approach where error correlation is disregarded if very high a priori data uncertainties are assumed, in which case the entire model is significantly smoothed and the velocity structures appear to be poorly resolved. Furthermore, the popular approach results in relatively strong lateral variations in the main velocity anomalies of the inverse estimate. This observation is in conflict with observations made in reflection GPR data sections from the field site (Hansen et al., 2008). Consistent with our findings, Squires et al. (1992) showed that static-like errors introduce unrealistic, strong lateral variations in the inverse estimates of seismic wave velocities in a seismic cross-borehole study. Accounting for data error correlation suppresses the effects of the anomalies close to the borehole walls and reduces the unrealistically strong lateral variation of the main velocity anomalies, consistent with observations made in the reflection GPR data sections from the field site. As for the synthetic test examples, the inversion scheme accounting for data error correlation appears to be robust.

Other inversion schemes have been used for suppression of static-like errors. A common feature of these schemes is that the sources of the correlated data errors are included as extra model parameters that have to be estimated during the inversion process. For example, Crosson (1976) used a source-relocation technique to account for incorrect information about the position of earthquake sources in large-scale seismic investigations. Maurer and Green (1997) used a similar technique to suppress the systematic data errors resulting from incorrect knowledge about the position and receiver positions in a high-resolution cross-well seismic tomographic experiment. Vasco et al. (1997) used an inversion scheme where source and receiver static terms were estimated as extra model parameters in a cross-borehole tomographic GPR study. Although GPR tomographic problems normally involve more data points than model parameters, the inverse problem cannot be solved without using some kind of regularization technique. Vasco et al. (1997) minimized the roughness of the velocity model during inversion.
using a finite-difference approximation of the spatial derivatives of the velocity field, and they did not allow for extremely large-amplitude velocity variations and unrealistically large source and receiver static terms. Thus, also in the approach of Vasco et al. (1997), prior assumptions were made regarding the amplitude of the static-like errors. In our study, we chose to account for the expected data errors through specification of data error covariance matrices, which are included in the inverse operator. With this approach, we do not need to increase the number of model parameters that must be estimated from the same number of travel times, and we can specify geologically reasonable standard deviations and correlation properties of the data errors. Furthermore, our approach is not limited to accounting for static-like features in the data sets. The scheme is flexible and any error type for which a covariance matrix may be specified can be accounted for using this approach.

We have focused on the correlated data errors generated by unknown cavities or small-scale anomalies near the boreholes, which we assume may play an important role at our field site. As discussed above, several other sources of correlated data errors occur. We have not identified any systematic data errors that may unequivocally be linked to erroneous positioning of the source and receiver antennae. The effects generated by such errors may, however, at least partly be accounted for by the covariance matrices that we set up in this study. Before, during, and after each cross-borehole survey, we performed transmitter and receiver calibration (adjustment of time zero). We did not note any drift or time jumps, which could cause significant miscalibration. Incorrect knowledge about borehole geometry will also produce significant systematic errors (Peterson, 2001). We have measured the distance and topographical differences between the boreholes using a high-precision theodolite instrument, resulting in uncertainties of <0.001 m. In this study, we assumed that our uncertainties linked to borehole geometry were insignificant.

The error types that we have studied in detail here may not have a significant impact on cross-borehole GPR data sets collected in other environments. For example, unknown cavities around the borehole walls may be less likely to arise in more compact sediments such as limestone or sandstone. Moreover, the sources of error should be carefully analyzed for each separate data set.

Peterson (2001) discussed a number of systematic (or correlated) errors likely to contaminate cross-borehole GPR data sets and outlined how they may be corrected. Correction of known errors should, of course, always be made before inversion. Sometimes purely static errors can be identified by visual inspection of travel times or travel-time residuals (e.g., Squires et al., 1992). Manual inspection and correction of all travel-time picks may be not be a feasible approach, however, especially for large data sets containing 700 travel-time picks or more, and the risk of missing significant errors during the manual correction process is large. We have demonstrated that if the variance and correlation properties of correlated data errors can be approximately estimated, then the effect on the inverse estimate of these correlated data errors may be greatly damped through careful specification of a priori data error covariance matrices.

In the synthetic test examples shown in this study, it is evident that the inversion results obtained accounting for data error correlation provide better resolution of the test anomaly pattern. In general, tomographic algorithms are known to produce severe smearing of anomalies (e.g., Hole, 1992). In particular, in areas with uneven ray coverage (few rays in one direction and many rays in a different direction) standard inversion algorithms, which disregard spatial correlation, tend to cause smearing of anomalies in the direction of the many rays. Nielsen and Jacobsen (1996) demonstrated that taking spatial correlation into account suppresses the smearing effects and increases model resolution.

Most program packages used for cross-borehole GPR tomographic inversion do not offer the possibility of taking correlated data errors into account by flexible specification of the data error correlation properties in the a priori data error covariance matrix, which is part of the least-squares inverse operator. Here we have demonstrated how correlated data errors can be taken into account through flexible data error covariance specification. The approach is simple in the way that it only requires basic reasoning regarding data errors likely to occur in a given data set. As soon as the interpreter has got an estimate of the standard deviation of the data errors and their spatial (or temporal) correlation properties, the data error covariance matrix can be set up. The incorporation of the data error covariance matrix in the least squares operator is straightforward (Eq. [3]; Tarantola, 1987).

Conclusions

Cross-borehole GPR data used for studying water saturation and hydrologic properties of the vadose zone are likely to be contaminated by correlated data errors. Some of the correlated data errors may be difficult or practically unfeasible to correct before inversion of the GPR data. Standard software for interpretation of cross-borehole GPR data do not normally offer the possibility of taking correlated data errors into account during the tomographic least-squares inversion process. Disregarding the correlation properties during inversion may lead to severe artifacts in the inverse model estimate and, therefore, lead to incorrect results regarding water saturation and hydrologic properties.

At our field site in northern Zealand, Denmark, we expected that unknown cavities and unknown small-scale low-velocity anomalies existed at or close to the walls of the boreholes used for the GPR data collection. These two sources of error may significantly affect the recorded first arrival travel times of the cross-borehole GPR data sets, and the resulting travel-time errors have a strongly correlated nature. Through analysis of the correlation properties of these error types, we have specified data error covariance matrices that describe the main characteristics of the errors. In synthetic inversion examples, we have shown that disregarding these correlated error types during inversion leads to significant contamination of the inverse estimate by undesired anomalies close to the borehole walls. The influence of the correlated data errors was effectively damped and the resolution properties of the inverse estimate were significantly improved if data error correlation was accounted for during the inversion process. The proposed method is robust in the sense that even significant overestimation of the data error standard deviation or incorrect assumptions about the spatial correlation length of the data errors results in inverse estimates that are similar to the ideal case where the exact correct standard deviation and spatial correlation length are known.

The proposed inversion scheme was applied to a tomographic GPR data set acquired at the field site at Arrenæs. The inverse estimates so obtained show more geologically plausible
velocity structures and significant suppression of spurious high-amplitude anomalies near the borehole walls than inversion results calculated using standard inversion schemes disregarding data error correlation.

Acknowledgments
We wish to thank the Water Department, Section for Water Quality, at Copenhagen Energy for all their help and for the use of their field location at Arrenæs infiltration plant. The cross-borehole instrumentation was purchased at Sensors & Software Inc., Mississauga, ON, Canada.

References