

TERI project, MALM and current source density inversion

Discrete constrained linear inversion

1 Introduction

1.1 MALM

MALM acquisitions are performed with one current electrode connected to the plant stem and the second to soil. The other electrodes are used to measure the distribution of the electric potential, see figure 1.

The passive-capillary water uptake directed from the roots to the leaves bases on the continuity of the water path inside the plant. Then, the injected current is expected to flow from the stem to roots through this water continuum. The active roots becomes the connection between the plant electrode and the soil electrode, they behave as current sources.

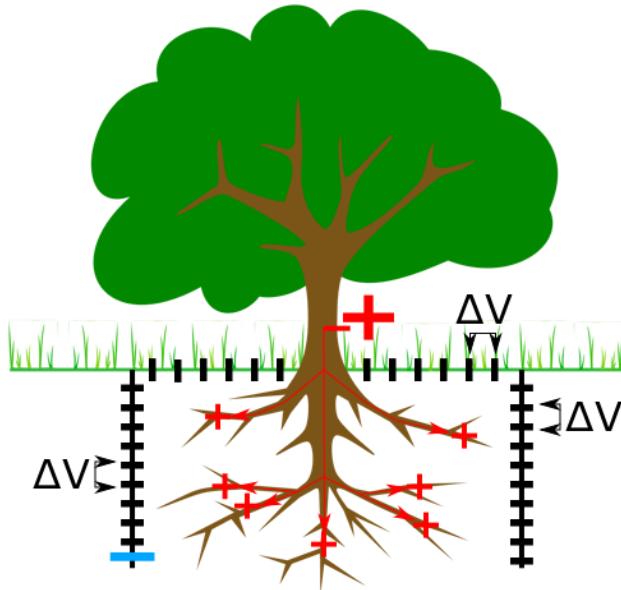


Figure 1: Possible 2D MALM configuration with boreholes. One current electrode is connected to the plant trunk (red plus), the second current electrode is chosen among the electrodes in the soil (blue minus). All the other electrodes are used to acquire potential differences and characterize the distribution of the electric potential about the rhizosphere. Soil electrodes can be placed on the surface and/or in the soil through the use of boreholes. The current electrode in the soil can be changed in order to acquire multiple sets and improve the information on the position of the roots.

The distribution of the so generated electric voltage depends on:

- Soil conductivity distribution
- Position of the soil electrode
- Boundaries of the system
- Position of the roots

The soil conductivity is derived from a normal ERT acquisition and inversion. The position of the soil electrode is known. The system boundaries are also known, e.g., lab 3D geometry or field semi-infinite conditions. It is then possible to infer the last factor, the unknown position of the roots.

1.2 Current Source Density estimation

The estimation of the current source distribution is performed on a grid of virtual electrodes (VRTe). The VRTe are the possible sources of current that the inversion process is going to test, the result of the inversion represents the fraction of current associated with each VRTe.

All the VRTe are added as nodes to the mesh used by the ERT inversion code (BERT). For linearity, we can sum the effects of different sources; this means we can calculate the response for each tested VRTe and define how much it is contributing to the distribution of potential measured in the LAB data. Physically this is a projection of a *linear* system components.

Single or all in once?? It is technically and numerically possible to account for multiple sets of MALM data, each one acquired with a different soil electrode. This would strongly improve the resolution of the iCSD. For example, having a set of soil electrodes around the expected position of the roots. However, if the distribution of the current among the roots depends on the soil return electrode position it is will generate different current source distributions for each soil electrode used. Clearly, this would make more difficult the combination of different data sets.

Abbreviations:

VRTe = virtual electrode, tested source; R = resistance; SOILe = real-lab electrode fixed current electrode in the soil

Our goal is to estimate the VRTe current fractions from the VRTe simulated ΔP and the measured resistances.

2 BERT calculation of the VRTe responses

2.1 Mesh

Add virtual (VRTe) electrodes to the BERT mesh. Then, the mesh contains both the real electrodes, which are used for the potential differences and for the return electrode, and the VRTe, which are used for the current injection.

2.2 Sequence

VRTe1 SOILe1 M1 N1

VRTe1 SOILe1 M2 N2

...

VRTe1 SOILe1 Mlast Nlast

"""

VRTe2 SOILe1 M1 N1

VRTe2 SOILe1 M2 N2

...

VRTe2 SOILe1 Mlast Nlast

"""

VRTe3 SOILe1 M1 N1

VRTe3 SOILe1 M2 N2

...

VRTe3 SOILe1 Mlast Nlast

...etc... VRTe1 SOILe2 M1 N1

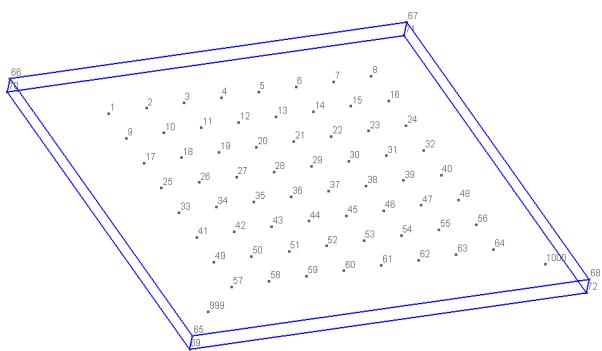
VRTe1 SOILe2 M2 N2

...

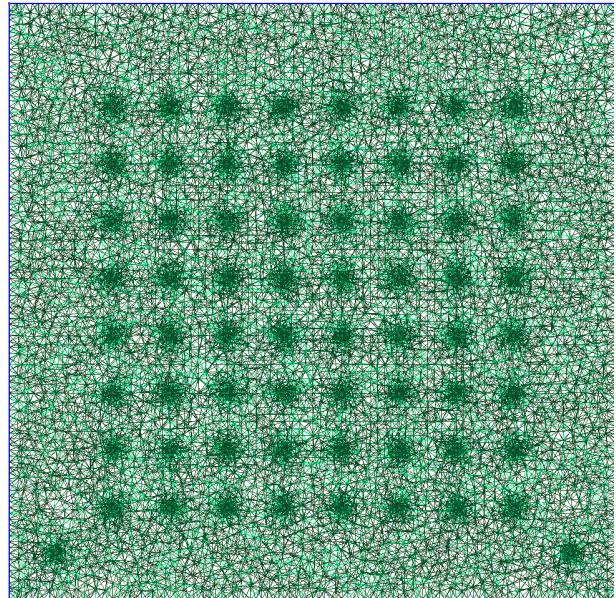
VRTe1 SOILe2 Mlast Nlast

"""

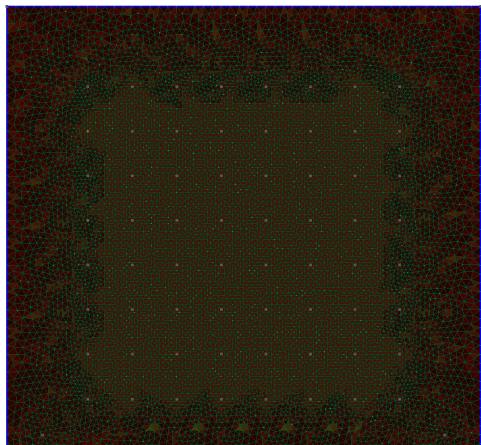
VRTe2 SOILe2 M1 N1



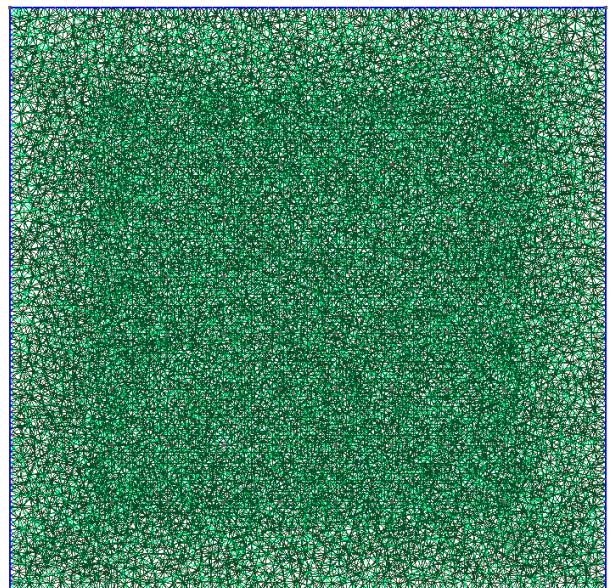
(a) Rhizotron geometry and electrodes (1:64).



(b) Local refinement.



(c) Check mesh quality.



(d) Distributed refinement.

Figure 2: Mesh preparation.

```

VRTe2 SOILe2 M2 N2
...
VRTe2 SOILe2 Mlast Nlast
"""
VRTe3 SOILe2 M1 N1
VRTe3 SOILe2 M2 N2
...
VRTe3 SOILe2 Mlast Nlast
...etc...

```

Run forward model on the inverted resistivity model, one single sequence can contain all the possible sources.

As the forward simulation run only once, it is possible to use a very fine mesh. This can be done using *secMesh.bms* and a rho.map extracted from the *fop-model.vtk*.

3 Least squares inversion

The least squares method is adopted to estimates the current fraction associated with each VRTe.

Spatial regularization, current conservation and parameters boundaries are added to the least squares problem. A specific and adjustable weight is given to each of these constrains to control the trade-off between data fitting and regularization.

3.1 Linear system

Because of its linearity, the inverse problem can be expressed as $\mathbf{Ax} = \mathbf{b}$. Where \mathbf{A} is the model, a matrix containing the VRTe simulated responses and the regularization; \mathbf{x} is a vector containing the model parameters to be estimated, the VRTe current fractions; and \mathbf{b} is a vector containing the observed data and regularization targets.

The simulated VRTe responses are organized into the \mathbf{A} matrix, each VRTe response being a column. Therefore, the number of columns of \mathbf{A} corresponds to the number of VRTEs.

Each row in \mathbf{A} expresses the contributions of the VRTEs to the relative data in \mathbf{b} . Therefore, the number of rows in \mathbf{A} corresponds to the number of observed data in \mathbf{b} .

3.2 Current conservation constrain

Because of the current conservation, the condition that the sum of all the current fractions is equal to 1 has to be respected. In this way, the total injected current will be consistently returned by the return electrode, whose intensity is normalized to -1.

This condition can be included into least squares problem by adding a new all-ones row to the \mathbf{A} matrix. The new row represents the sum of all the elements in the column \mathbf{x} , which is set to be equal to 1 by adding this value as new element to the column \mathbf{b} .

3.3 Spatial regularization

A 2D spatial regularization is implemented to stabilize the numerical inversion and control the trade-off between solution roughness and data fit. As for the previous constrain, this condition is included into the least squares problem by adding new rows to the \mathbf{A} matrix.

Each new row expresses a first order finite difference between two consecutive VRTEs, -1 and 1 are inserted in the columns of the 2 VRTEs considered. The difference is set to be equal to 0 by adding this new entry into the column \mathbf{b} .

$$\begin{array}{c}
W \\
\hline
\left(\begin{array}{cccc} W_{obs} & 0 & \cdots & 0 \\ 0 & w_{obs} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{reg} \end{array} \right) \quad \mathbf{A} = \left(\begin{array}{cccc} S_1 R_1 & S_2 R_1 & \cdots & S_n R_1 \\ S_1 R_2 & S_2 R_2 & \cdots & S_n R_2 \\ \vdots & \vdots & \ddots & \vdots \\ S_1 R_m & S_2 R_m & \cdots & S_n R_m \\ 1 & 1 & \cdots & 1 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array} \right) \quad \mathbf{x} = \left(\begin{array}{c} S_{1,cf} \\ S_{2,cf} \\ \vdots \\ S_{n,cf} \end{array} \right) \quad \mathbf{b} = \left(\begin{array}{c} R_{1,meas} \\ R_{2,meas} \\ \vdots \\ R_{m,meas} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{array} \right) \quad W = \left(\begin{array}{cccc} W_{obs} & 0 & \cdots & 0 \\ 0 & w_{obs} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{reg} \end{array} \right)
\end{array}$$

3.4 Constrain and regularization weights

The weight of the different observations and constraints can be modified by multiplying both \mathbf{A} and \mathbf{b} a weighting diagonal matrix. In this way, a small misfit relative to the target value is forced, where desired, by increasing its absolute value compared to the other observations and/or constraints.

After the inversion, the residuals of the unweighted problem can still be obtained. The residuals of the weighted least squares solution are divided by the weights, the so obtained unweighted residuals can then be better analyzed and compared. For example to construct the L-curve.

3.5 L-curve and optimal regularization

In order to identify the optimum value for the regularization weights, the model-to-data fit can be compared with the solution roughness. This is done by performing several least squares inversion with different regularization weights and plotting the solution roughness against the data fit. The solution roughness is defined as the sum of the unweighted regularization residuals, that is how the model deviates from the regularized target. The obtain curve, called L-curved, illustrates how the model-to-data fit decreases while increasing the regularization.

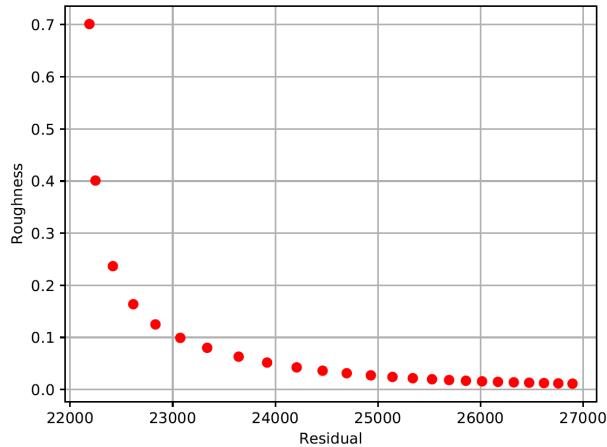


Figure 3: Trade-off between model-to-data fit and regularization.

The L-curve allows to understand at which point the model-to-fit improvement is minimal and does not justify the associated increment of the solution roughness. The optimum level of regularization is that at which the curve deflects sharply upwards, that is the solution deviates from the regularization conditions without a significant improvement of the data fit.

3.6 Bounding values

The weighted linear problem is solved with the use of a python library that offers two different numerical algorithms. Both algorithms solves the linear system in the least-squares sense and allows the definition of variable boundaries. An implementation of a subspace trusted region method proposed by Branch et al., 1999, has been used for the current inversion.

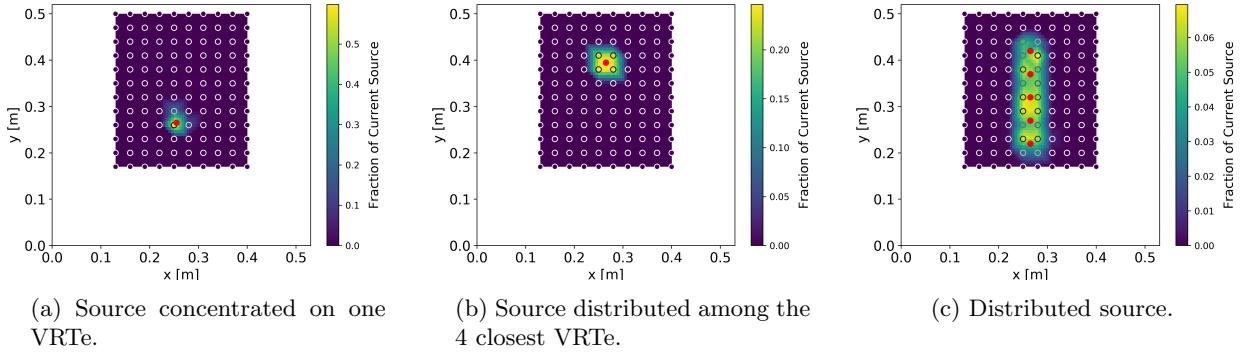


Figure 4: Synthetic models, linear interpolation of the current fraction associated with the VRTe.

3.7 Comparison with PEST

PEST is a standard software package for parameter estimation and numerical uncertainty. PEST was initially used for the current source density inversion, but the present and specific code offers two main advantages:

- Take advantage of the problem linearity. PEST is designed to solve nonlinear problems, such as hydraulic or resistivity distribution. As consequence, PEST does not account for the problem linearity and calculates the Jacobian matrix at each iteration. The calculation of the Jacobian matrix requires, without linearity assumption, PEST to externally call the physical forward model at least as many times as there are adjustable parameters (VRTe). The specific number of calls depends on the type of derivative used, and may need to be increased to obtain higher resolution and reduce rounding errors. For each model call PEST writes the modified parameters and reads the model results through text files. As consequence of the many model calls and write/read time, the inversion time significantly increases.
- In order to facilitate the problem definition PEST takes care of the matrices preparation. The user defines the problem to solve through a list of commands and inputs in a series of configuration files. While this allows the user to work on a higher level, it necessarily introduce an overhead of commands make it difficult to automatize the inversion procedure compared to the direct definition of the problem in matrix form.

Because of its reliability, user-friendliness, and extensive documentation, PEST represented an optimal starting point and provided the necessary initial numerical validation.

4 Numerical and laboratory tests

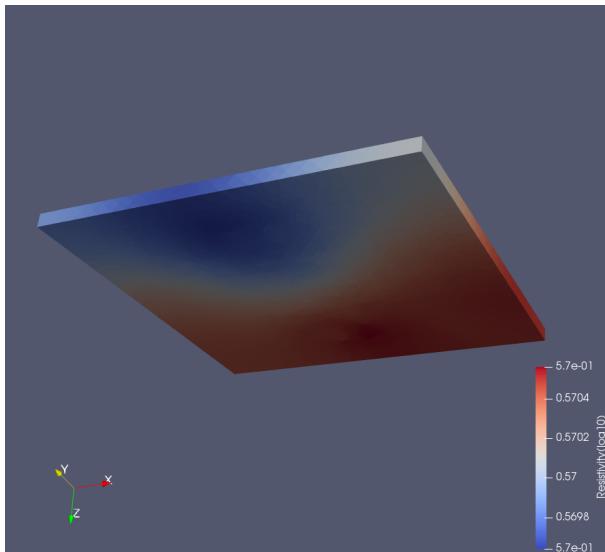
The ERT inversion quality was evaluated first. Different type of meshes were tested because of the particular rhizotron geometry (we used *gmsh*).

Three initial MALM synthetic tests were performed on a homogeneous distribution of conductivity, i.e. running the forward model on the rhizotron geometry with a homogeneous resistivity of 1 Ohm/m. In the first test the current source was placed close to a VRTe to evaluate if the inversion could properly concentrate on that VRTe the source of current. In the second test the current source was placed in a position that forced the inversion to more equally distribute the current. In the third test a vertical diffuse source of current was simulated by combining 5 sources.

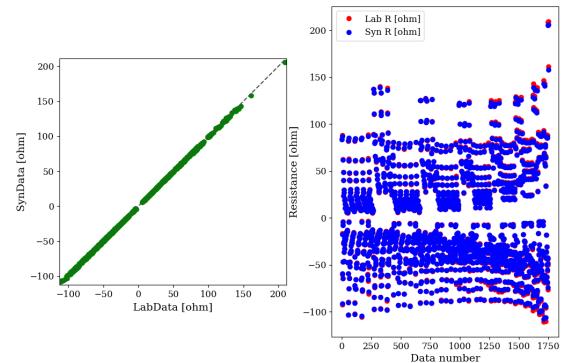
Following the synthetic tests, a first laboratory experiment was performed by filling the rhizotron with only water and dropping a wire representing the roots from the top of the rhizotron. A single electrode was connected at the end of the wire. The ERT inversion produced an homogeneous distribution of resistivity (3.7 Ohm/m). Before the MALM acquisition and inversion, we tested the capability of the inverted model to reproduce the LAB data, this is important in order to evaluate if the sensitivity of the data could support the investigation of a current source.



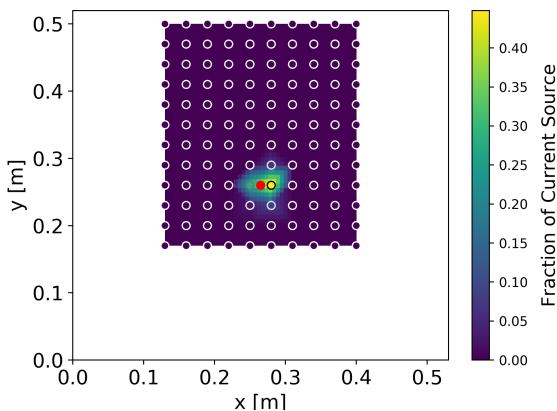
Figure 5: MALM configuration for the lab test, the red cable supplies a small injection electrode positioned in the center of the rhizotron.



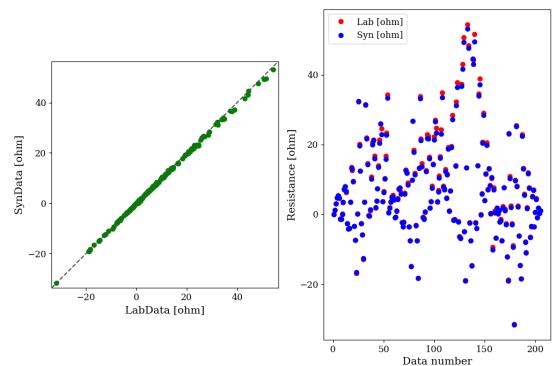
(a) Inverted resistivity distribution.



(b) ERT data check, comparison between laboratory ERT data and resistances modeled on the inverted resistivity model.



(c) Estimated current source distribution, the red point represents the position of the electrode inside the rhizotron in the laboratory.



(d) Comparison between laboratory and simulated MALM resistances. The simulation of the resistances bases on the inverted resistivity distribution (BERT) and estimated current source distribution.

Figure 6: Lab experiment with single source.

In order to test the overall resolution of the current source density inversion we performed 6 laboratory tests in which two point sources where placed at different distances. We could then verify if, and for which minimum distance, the two sources were correctly inverted. For this tests a non-homogeneous water conductivity was forced by the addition of salt, the values of resistivity agreed with the solution conductivity values measured with a conductivity meter. Furthermore, the overall current source density inversion appears to properly take into account the conductivity distribution.

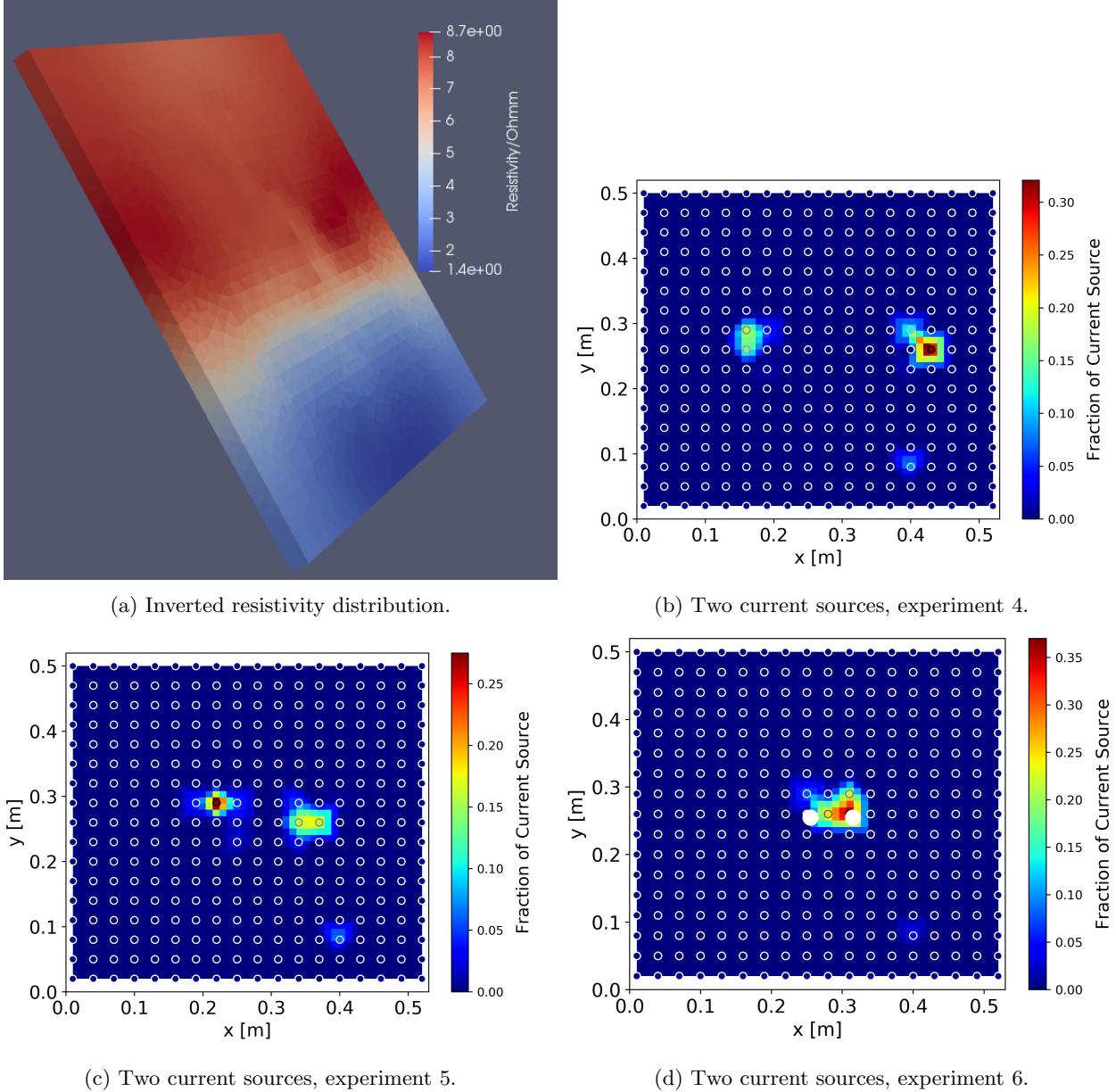


Figure 7: Lab experiment with two sources, the sources are correctly located and distinguished in experiments 4 and 5. In experiment 6 the two sources are not distinguished, see white dots, and a unique elongated source results from the inversion process.

5 Plant in water

After the numerical and laboratory methodological tests, we moved to explore the current source densities associated with MALM measurements on maize plants. In two first experiments the roots we positioned in the rhizotron filed with water.

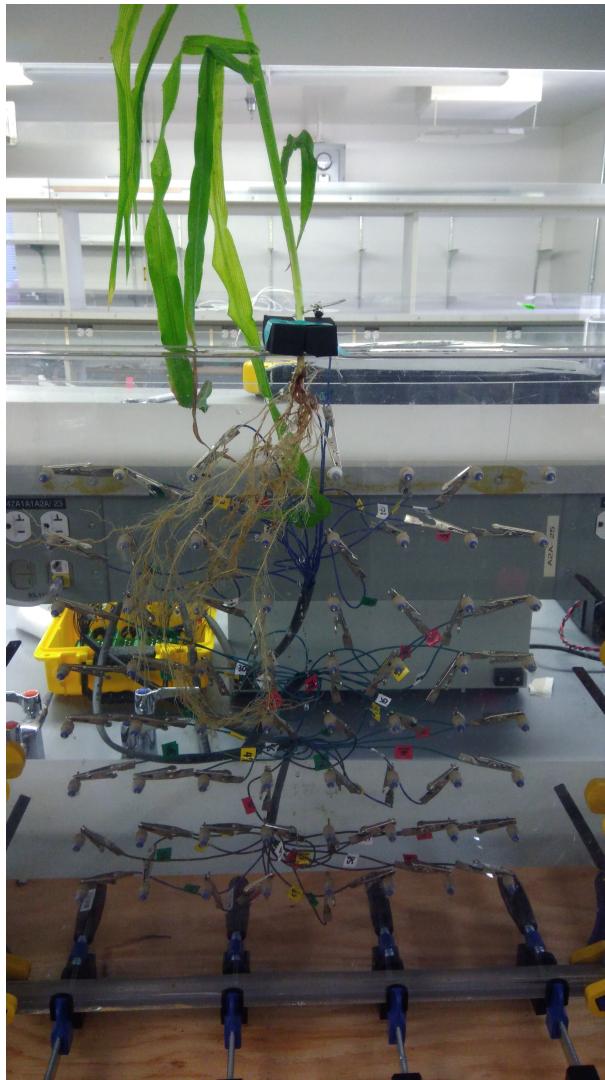


Figure 8: Plant experiment with water.

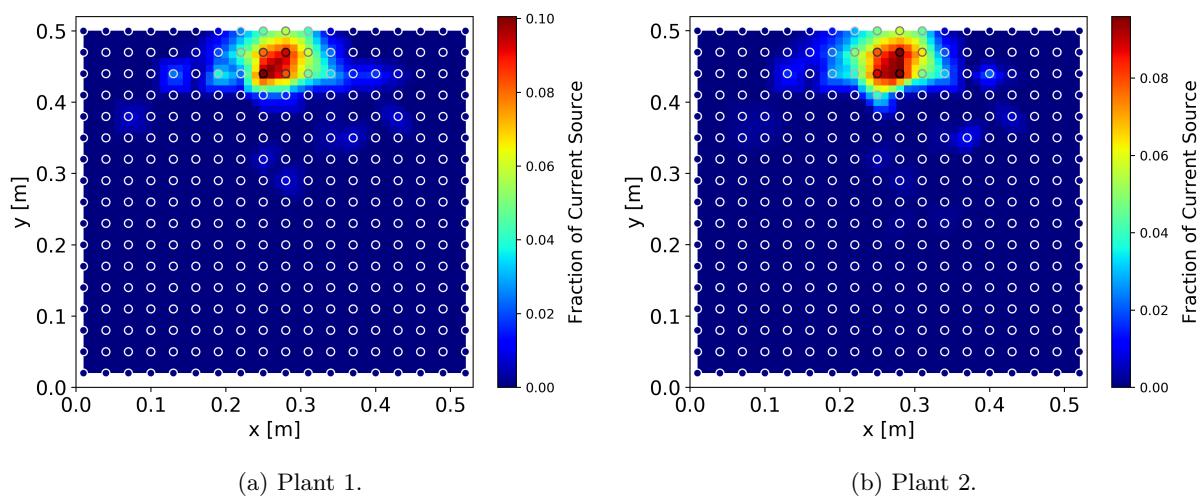


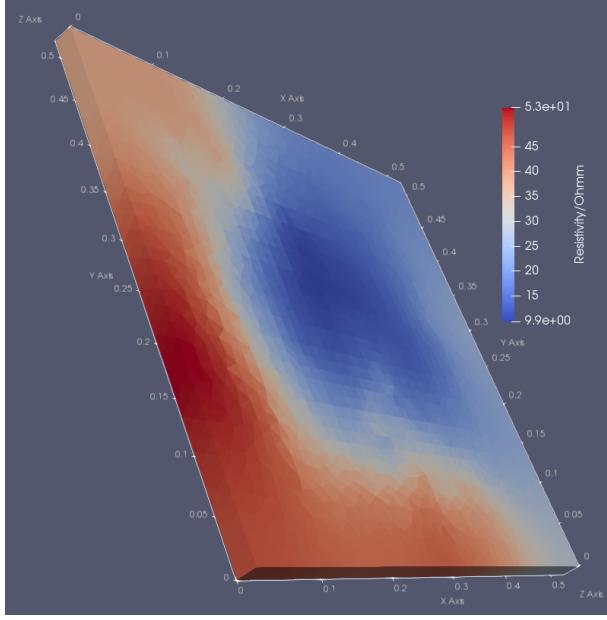
Figure 9: Current source densities obtained from the two laboratory experiments with maize plants in water.

6 Plant in Soil

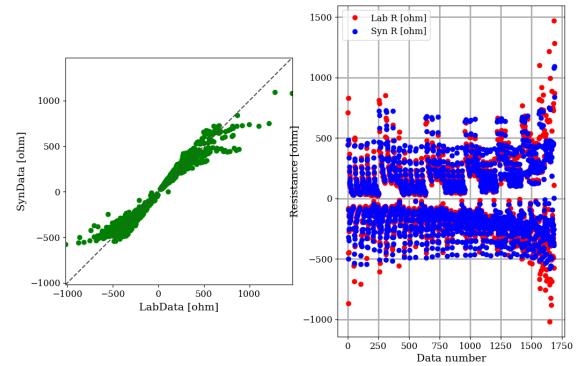
On the basis of the positive initial tests, we performed an experiment with a maize plant in natural soil. An homogeneous distribution of the current source was used as initial value for the estimation procedure.



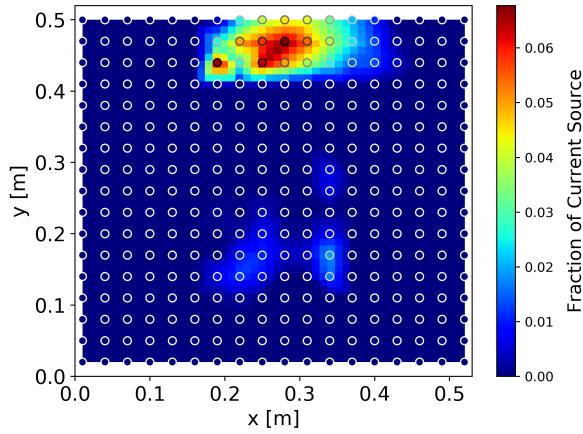
Figure 10: Distribution of the roots at the moment of the MALM acquisition.



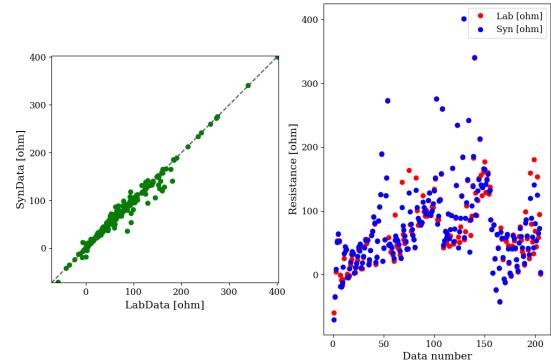
(a) Inverted resistivity distribution.



(b) ERT data check, comparison between laboratory ERT data and resistances modeled on the inverted resistivity model. The particular distribution of the misfit is likely due to the Occam regularization implemented by the ERT code.



(c) Estimated current source distribution, the red point represents the position of the return electrode.



(d) Comparison between laboratory and simulated MALM resistances. The simulation of the resistances bases on the inverted resistivity distribution (BERT) and estimated current source distribution.

Figure 11: Experiment with maize plant in soil.