Technical Report: Soil moisture dynamics identification in a ecological plot of zucchini, beets and lettuce

Erid Pacheco*, Gabriela Cáceres*, and Luis Orihuela,†

October 11, 2023

1 Introduction

The main purpose of this report is to present detailed information regarding the identification of linear and non-linear models for the moisture dynamics in agricultural soil. These models have been identified for their later use for model-based controllers, such as Model Predictive Control.

The document is structured as follows: Section 2 describes the experimental site, from both the agricultural and data acquisition perspectives. Section 3 details all the data preparation techniques. Finally, Section 4 introduces the identified models.

2 Experimental site description

The implementation took place in a farm belonging to Bioalverde. This company grows ecological crops in its fields located in Dos Hermanas, Seville.

The plot had a dimension of 50x17 meters and was divided longitudinally into 10 irrigation lines. Additionally, was subdivided into three distinct subplots, each dedicated to cultivating a different crop. The first subplot featured zucchini with two lines, the second accommodated beets with four lines, and the third was designated for lettuce, also with four lines.

The monitoring system consisted of ground nodes equipped with soil moisture sensors to measure the volumetric water content (VWC). Each subplot was served by three ground nodes, resulting in a total of nine ground nodes per plot. These ground nodes were equipped with a total of six soil moisture sensors, evenly distributed between the first and second soil layers.

The data collected from a ground node can be seen in table 1.

Data	Data name	Data type
s11	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$
s12	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$
s13	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$
s21	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$
s22	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$
s23	h_{gnd},t_{gnd}	$h_{gnd}, t_{gnd} \in \mathbb{R}$

Table 1: Data collected from the ground node

One solenoid valve and one flow meter were installed for each subplot and connected to a LoRaWAN I/O controller device.

Data from the sensors and I/O devices is transmitted wirelessly (using the LoRa protocol) to a gateway that sends the data to the cloud, this data is collected and stored in a database on a server with a timestamp of when the measurement was saved.

*Departamento de Ingeniería, Universidad Loyola Andalucía, Avda. de las Universidades s/n 41071 Dos Hermanas, Sevilla, Spain

†Departamento de Ingeniería Electrónica, Sistemas Informáticos, y Automática, Universidad de Huelva, Avda. de las Fuerzas Armadas s/n, 21007, Huelva, Spain

For the system identification, the authors use the data acquired from the ground node and I/O device controller extracted from the database from 00:00:00 (UTC) on June 15th, 2023 until 00:00:00 June 21st, 2023. The sampling time of the ground nodes is 15min and for the I/O devices 5min. The variables acquired are time, valve state $\{1, 0\}$, pulse count and VWC of each sensor.

The data collected from an I/O device can be seen in table 2.

Data	Data name	Data type
valve status	v_{ctr}	$v_{ctr} \in \{0, 1\}$
pulse count	p_{ctr}	$p_{ctr} \in \mathbb{N}$

Table 2: Data collected from I/O device

3 Data preparation

After collecting the data from ground nodes (h_{gnd} and t_{gnd}) and control nodes (v_{ctr}, p_{ctr} and t_{ctr}) used in the experiment (as explained in Section 2), several preparatory steps were taken before performing the identification process.

For the control nodes, the following steps were carried out:

1. Data Filtering: Some of the data obtained from the control nodes, such as the valve state (v_{ctr}) and pulse count (p_{ctr}), underwent a filtering process. This filtering took into consideration the relationship between changes in the valve state and the corresponding pulse count in order to have the correct v_{ctr}.

2. Data Resampling: Control nodes data v_{ctr} and t_{ctr} were resampled using a linear interpolation method, resulting in a uniform sampling time t of 15 minutes.

3. Data Storage: Finally, the prepared data v_{sca} and t was organized and stored in tables, with each table corresponding to a specific subplot of the experiment. [1]

In particular, when it comes to data treatment within the control node, it closely examines variables such as time, valve state, and pulse count. The flowchart illustrating this treatment process can be found in Figure 1. In this context, n represents the time of actual valve state measurements.

Please note that, even though the valve state v_{ctr} is a boolean variable, the prepared v_{sca} is a real variable, belonging to the interval $[0, 1]$, trying to accommodate possible inter-sampling periods where the valve was partially open and closed.

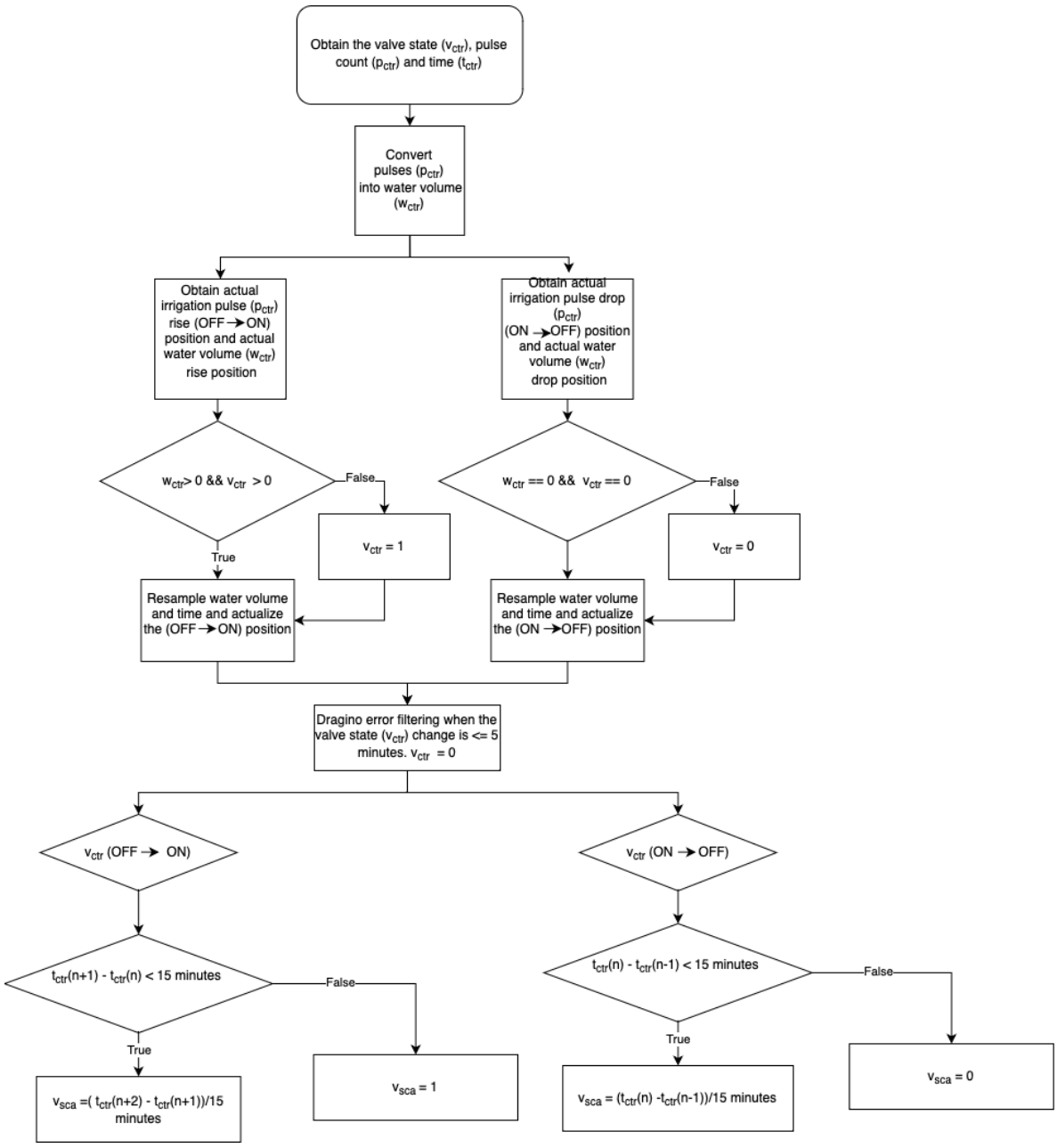

Figure 1: Flowchart of the valve states (v_{ctr}) treatment and resampling.

The ground node treatment and storage steps are outlined as follows:

1. Data Resampling: Ground nodes data h_{gnd} and t_{gnd} were resampled using a linear interpolation method, resulting in a uniform sampling time t of 15 minutes.

2. Interpolation and Extrapolation: For sensor data, a specific process was applied, involving interpolation and extrapolation techniques. This step was carried out by considering the relationship between the h_{gnd} data and the timing of changes in the valve state t_{ctr}. Preserving causality was crucial in this step.

3. Data Scaling: Since the data for each h_{gnd} component may vary significantly, it is advisable to standardize or scale them h_{sca}. This simplifies the subsequent identification process.

4. Data Storage: Finally, the prepared data h_{sca} and t was organized and stored in tables, with each table corresponding to a specific subplot of the experiment. [1]

In the case of ground node data treatment, the analysis considered both the timestamped VWC (Volumetric Water Content) values and the resampled valve state data, as previously performed. It is applied to three conditions for either interpolating, extrapolating, or retaining the actual value for each valve state change, as outlined in the flowchart presented in Figure 2. Subsequently, the processed data, collected and treated, is made available in the repository [1].

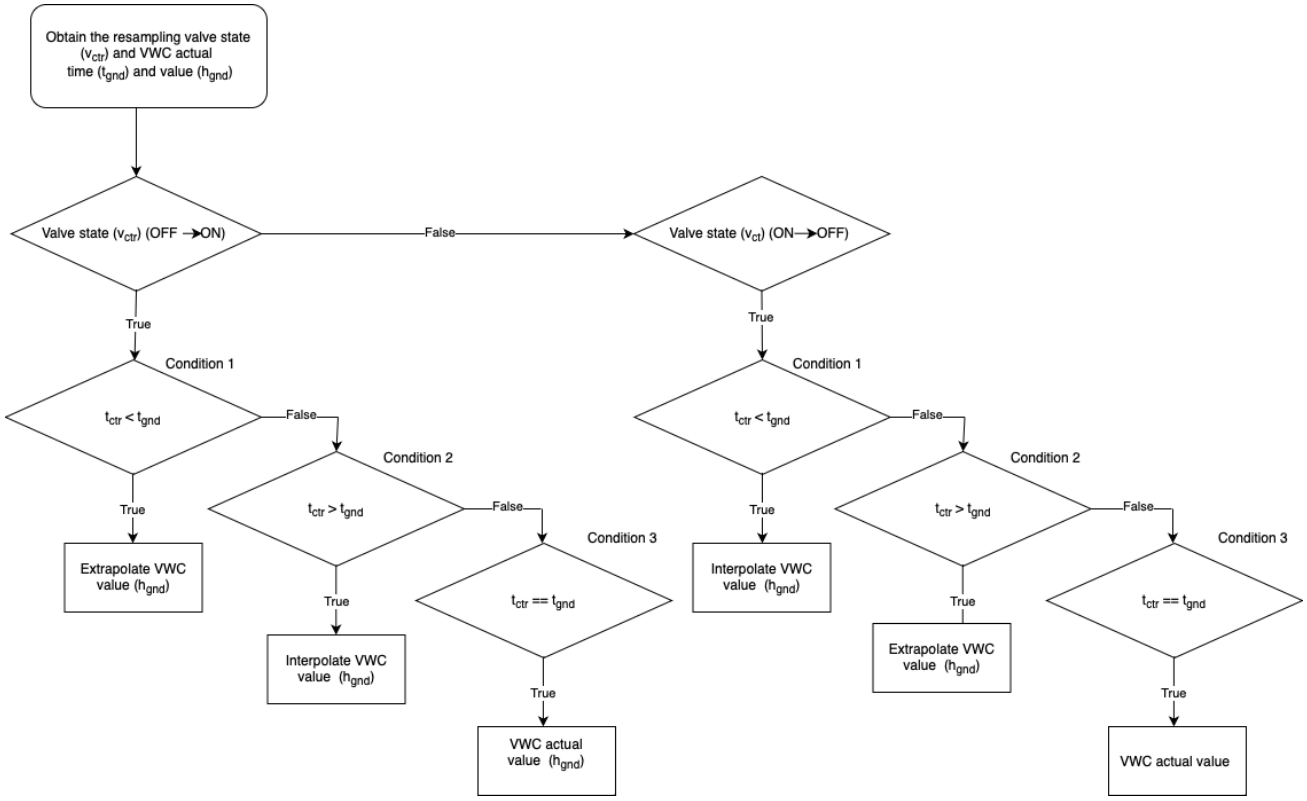

Figure 2: Flowchart of the h_{gnd} interpolation and extrapolation.

4 Soil dynamics identification

As mentioned before, each ground node have two pairs of three sensors, installed in two different depths. The model for the soil moisture dynamics at the location of each ground node is obtained using the data of only two sensors, one of each soil layer. Then, for each ground node, a visual inspection of the acquired data from the sensors of each measurement node was conducted, followed by the selection of one sensor from each layer for the identification of both nonlinear and linear models. The selected sensors are detailed in Table 3, where s_{ij} stands for the layer $i \in 1, 2$ and sensor $j \in 1, 2, 3$. The labels used for the ground nodes are the same than those in the shared database.

Node	Layer 1	Layer 2
gnd004	s_{11}	s_{23}
gnd006	s_{12}	s_{21}
gnd003	s_{12}	s_{23}
gnd007	s_{12}	s_{21}
gnd011	s_{12}	s_{23}
gnd005	s_{13}	s_{23}

Table 3: Sensor selected for each ground node

In the next subsections, and for the location of each ground node, input-output models will be obtained to model the discrete-time dynamics between the input v_{sca} and the outputs h_{sca} (moisture

in each layer) for a sampling time of 15 minutes (such as the common time t). Aiming at using a more standard notation, the inputs will be termed $u(k)$ and the outputs $x_i(k),\ i \in 1, 2$.

4.1 Non-Linear Model

The soil moisture dynamics is assumed to be model as follows:

$$x_1(k+1) = x_1(k) + f_1(x_1(k)) + c_1 x_1(k) u(k) + b_1 u(k) \tag{1}$$

$$x_2(k+1) = x_2(k) + f_2(x_1(k), x_2(k)) + c_2 x_2(k) u(k) + b_2 u(k), \tag{2}$$

where $x_i(k)$ stands for the soil moisture on layer i (layer 2 is deeper than layer 1), $u(k) \in [0, 1]$ is the control action, $f_i(\cdot)$ are nonlinear functions to be identified, and c_i, b_i are constants to be identified.

Please note that, when the control action is set to zero, model (1)-(2) simplifies to:

$$x_1(k+1) = x_1(k) + f_1(x_1(k))$$
$$x_2(k+1) = x_2(k) + f_2(x_1(k), x_2(k)),$$

where it is easy to see that functions $f_i(\cdot)$ intend to capture the difference between $x_i(k+1)$ and $x_i(k)$ as a function of the actual state.

It is assumed that these functions are polynomial functions. Therefore, the definition of function $f_1(\cdot)$ is:

$$f_1(x_1(k)) = \alpha_0 + \alpha_1 x_1(k) + \alpha_2 x_1(k)^2 + \ldots + \alpha_n x_1(k)^n \tag{3}$$

On the other hand, function $f_2(\cdot)$ can take different definitions:

$$f_2(x_1(k), x_2(k)) = \begin{cases} \beta_0 + \beta_{11} x_2(k) + \beta_{12} x_2(k)^2 + \ldots + \beta_{1m} x_2(k)^m, & \text{method 1} \\ \beta_0 + \beta_{11} x_1(k) + \beta_{21} x_2(k) + \ldots + \beta_{1m} x_1(k)^m + \beta_{2m} x_2(k)^m, & \text{method 2} \\ \beta_0 + \beta_{11} x_2(k) + \beta_{21}(x_1(k) - x_2(k)) + \ldots + \beta_{1m} x_2(k)^m + \beta_{2m}(x_1(k) - x_2(k))^m, & \text{method 3} \\ \beta_0 + \beta_{11} x_1(k) + \beta_{21}(x_1(k) - x_2(k)) + \ldots + \beta_{1m} x_1(k)^m + \beta_{2m}(x_1(k) - x_2(k))^m, & \text{method 4} \\ \beta_0 + \beta_{11}(x_1(k) - x_2(k)) + \ldots + \beta_{1m}(x_1(k) - x_2(k))^m, & \text{method 5} \end{cases} \tag{4}$$

Constants n, m are to be chosen.

4.1.1 Identification steps

These models have been identified using the next procedure:

1. Data is collected from the nodes deployed for the experiment, as explained in Section 2.
2. Some data preparation techniques have been applied to the data (see Section 3).
3. Previous dataset is split into a training dataset and a validation dataset.
4. For a given method $i \in \{1, 2, 3, 4, 5\}$ and maximum orders n, m, the model is identified using least squares.
5. The chosen model is validated to make predictions in open loop for 24 hours using the validation dataset
6. If the results are not adequate, change the method or the orders and go back to step 4.

4.1.2 Identified parameters

Tables 4 includes the identified method for each of the node.

Ground node	Method
gnd004	2
gnd006	1
gnd003	3
gnd007	1
gnd011	3
gnd005	3

Table 4: Identified methods

Tables 5-6-7 includes the identified constants for each of the plots and priorities.

Ground node	α_0	α_1	α_2	α_3	α_4	α_5
gnd004	-0.0012	-0.0004	-0.0221	0.0024	-0.0915	0
gnd006	0.0004	-0.0148	0.0161	-0.1288	0	0
gnd003	-0.0003	-0.0418	0.2467	-0.5415	0.3242	0
gnd007	-0.0066	0.1719	-1.7098	6.3428	-9.5919	4.8365
gnd011	0.0008	-0.1160	0.8312	-2.3557	2.9620	-1.4978
gnd005	0.0038	-0.1060	0.5696	-1.0600	0.4596	0

Table 5: Identified constants for $f_1(\cdot)$

Ground node	β_0	β_{11}	β_{12}	β_{13}	β_{14}	β_{21}	β_{22}	β_{23}	β_{24}
gnd004	0.0062	0.1662	-0.3343	0.0118	0.1926	-0.3542	1.4179	-2.2136	1.0721
gnd006	0.0011	-0.0428	0.1094	-0.0971	0	0	0	0	0
gnd003	0.0007	-0.0368	0.0756	-0.0416	0	0.0023	0.0069	0.0251	0
gnd007	0.0171	-0.1344	0.1609	0.2154	-0.3311	0	0	0	0
gnd011	0.0014	-0.0581	0.3091	-0.8070	0.4016	0.0156	0.1048	0.7304	-5.2301
gnd005	-0.0004	-0.0640	-0.1184	0.4106	-0.6389	0.0183	0.0165	-0.0058	0.9228

Table 6: Identified constants for $f_2(\cdot)$

Ground node	c_1	c_2	b_1	b_2
gnd004	-0.6837	-0.0977	0.7609	0.2278
gnd006	-0.3503	-0.0847	0.4642	0.2083
gnd003	0	0	0.1340	0.0146
gnd007	0	0	0.0569	0.0688
gnd011	-0.2986	-0.1952	0.4303	0.3283
gnd005	-0.4110	-0.2172	0.4440	0.3466

Table 7: Identified constants for c_i, b_i

4.2 Linear Model

We assume that the irrigation system model is linear and described in discrete time by the following equations:

$$x_1(k+1) = A_{11}x_1(k) + A_{12}x_2(k) + B_{11}u(k), \tag{5}$$
$$x_2(k+1) = A_{21}x_1(k) + A_{22}x_2(k) + B_{21}u(k), \tag{6}$$

where $x_1(k)$ and $x_2(k)$ represent the soil moisture levels in layers 1 and 2, respectively, and $u(k)$ represents the binary control action ($\{0,1\}$). Coefficients A_{ij} and B_{ij} are unknown parameters that we aim to identify.

The matrix representation of this system is:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} u(k) \rightarrow x(k+1) = Ax(k) + Bu(k). \tag{7}$$

To identify the elements of matrices A and B, the Kalman filter is employed [2]. The state vector will contain the components of matrices A and B. Therefore, the dynamic system used in the Kalman filter equations will be:

$$\phi(k+1) = \theta_k\phi(k) + \delta_k m(k) + w_k, \quad w_k \sim \mathcal{N}(0, Q_k) \tag{8}$$
$$x(k+1) = H_k\phi(k) + v_k, \quad v_k \sim \mathcal{N}(0, R_k) \tag{9}$$

where $\phi(k)$ is the state vector to be estimated. It should be noted that the state vector $\phi(k)$ does not merely contain the system parameters, but rather approaches them as $k \to \infty$. This means that the identified parameters will be the steady-state values of the state vector.

w_k and v_k are the process and measurement noise, respectively, which are assumed Gaussian with zero mean. Matrices Q_k and R_k are the noise covariance matrices.

The matrix R has been assigned a constant value, which is the same for all sensors and equals the value of the sensor in a steady-state condition, $\sigma^2 I_{6\times6}$. This value has been experimentally adjusted to reach a value of $R = 0.0004 I_{6\times6}$.

Considering that the elements of the state vector are constants, matrix θ is a identity matrix, this is, $\theta = I_{6\times6}$ and $\delta_k = 0$, and there is no model uncertainty, which implies $Q_k = 0, \forall k$.

The Kalman filter operates in two stages: prediction and correction. The general equations for these stages are described below:

Prediction:

$$\hat{\phi}_k^- = \theta \hat{\phi}_{k-1}^+ \tag{10}$$

$$P_k^- = \theta P_{k-1}^+ \theta^T + Q_{k-1} \tag{11}$$

Correction:

$$\overline{K}_k = P_k^- H_k^T (H_k P_k^- H_k^T + R)^{-1} \tag{12}$$

$$\hat{\phi}_k^+ = \hat{\phi}_k^- + \overline{K}_k (z_k - H_k \hat{\phi}_k^-) \tag{13}$$

$$P_k^+ = P_k^- - \overline{K}_k H_k P_k^- \tag{14}$$

The matrices used for identification are defined as follows:

$$x_k = [x_1(k), x_2(k)]^T$$

$$H_k = \begin{bmatrix} x_1(k-1) & x_2(k-1) & 0 & 0 & u_{k-1} & 0 \\ 0 & 0 & x_1(k-1) & x_2(k-1) & 0 & u_{k-1} \end{bmatrix}$$

4.2.1 Identification Steps

The models have been identified using a similar procedure than for the nonlinear model:

1. Data is collected from the deployed sensors and actuators.
2. Data preparation techniques are applied.
3. The dataset is split into training and validation sets.
4. The model is identified using the Kalman Filter.
5. The identified model is validated for 24-hour open-loop predictions using the validation dataset.

4.2.2 Identified Parameters

Table 8 lists the coefficients of the identified models for each measurement node.

Ground node	A_{11}	A_{12}	A_{21}	A_{22}	B_{11}	B_{22}
gnd004	0.7912	0.1608	0.0017	0.9688	0.2611	0.1742
gnd006	0.7641	0.1051	0.0271	0.9708	0.2780	0.1534
gnd003	0.9852	0.0015	0.0117	0.9889	0.1333	0.0138
gnd007	0.9365	0.0359	-0.0134	1.0030	0.0607	0.0583
gnd011	1.1532	-0.2490	0.2128	0.6724	0.1959	0.1996
gnd005	1.0846	-0.2407	0.1109	0.6937	0.1984	0.2427

Table 8: Identified constants for linear models

The mean squared errors and the correlation of the validation data with respect to the data used in identification can be observed in Table 9.

Ground node	MSE (x_1)	MSE (x_2)	Correlation r^2 (x_1)	Correlation r^2 (x_2)
gnd004	0.0236	0.0080	0.7366	0.7841
gnd006	0.0044	0.0176	0.8971	0.7864
gnd003	0.0122	0.0355	0.8191	0.0105
gnd007	0.0122	0.0197	0.5915	0.6171
gnd011	0.0312	0.0366	0.0897	0.4022
gnd005	0.0088	0.0044	0.7288	0.7994

Table 9: Errors respect to validation data.

References

[1] E. Pacheco, G. Cáceres, and L. Orihuela, "Sensors and Actuators Data for Model Identification of an Irrigation System," Universidad Loyola, Tech. Rep., 10 2023.

[2] A. Nevado R., P. Cabrera C., and M. Martín S., *Conceptos básicos de filtrado, estimación e identificación.* Editorial UNED, 2015.