

Soil moisture forecasting for irrigation recommendation

James Brinkhoff* John Hornbuckle**
Carlos Ballester Lurbe***

* *University of New England, Armidale NSW 2351 Australia, (email: james.brinkhoff@une.edu.au)*

** *Deakin University, Geelong VIC 3220 Australia, email: j.hornbuckle@deakin.edu.au*

*** *Deakin University, Geelong VIC 3220 Australia, email: carlos.ballesterlurbe@deakin.edu.au*

Abstract: This study integrates measured soil moisture sensor data, a remotely sensed crop vegetation index, and weather data to train models, in order to predict future soil moisture. The study was carried out on a cotton farm, with wireless soil moisture monitoring equipment deployed across five plots. Lasso, Decision Tree, Random Forest and Support Vector Machine modeling methods were trialled. Random Forest models gave consistently good results (mean 7-day prediction error from 8.0 to 16.9 kPA except in one plot with malfunctioning sensors). Linear regression with two of the most important predictor variables was not as accurate, but allowed extraction of an interpretable model. The system was implemented in Google Cloud Platform and a model was trained continuously through the season. An online irrigation dashboard was created showing previous and forecast soil moisture conditions, along with weather and normalized difference vegetation index (NDVI). This was used to guide operators in advance of irrigation water needs. The methodology developed in this study could be used as part of a closed-loop sensing and irrigation automation system.

© 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Soil moisture; irrigation automation; forecasting; time-series modeling.

1. INTRODUCTION

Irrigated agriculture is important for the food and fibre needs of the world, consuming around 20% of global arable land, but generating 40% of produce, Sauer et al. (2010). However, there is increasing pressure on the world water resources, with agriculture accounting for 85% of consumptive water use, Foley et al. (2005). This has resulted in reduced river flows, increasing salinity, fertilizer pollution and declining water tables. With a growing population, the world's irrigated area is increasing. Considering water scarcity there is a need to improve water productivity, Harrison (2002).

This motivates the development of techniques and technologies to optimize crop water use. Deficit irrigation techniques are used to improve water productivity (crop production per unit of water consumption) and profit, Fereres et al. (2003). These techniques trade off water supply with crop yields to achieve the desired economic or environmental outcomes. Fereres et al. (2003) showed that while reducing water may reduce yield, it can result in increased revenue in horticultural crops. Ballester et al. (2019) reported a study of the effects of various irrigation deficits on cotton quality and yield.

Scheduling irrigation is an important aspect of water management. Continuous soil moisture monitoring can help, by giving growers an indication of when the soil moisture is reaching plant-limiting levels, Brinkhoff et al. (2017).

However, such sensors only indicate the current conditions. As daily evapotranspiration can change dramatically depending on weather, it can be hard to precisely predict when irrigation will be needed. There are delay and flow limitations inherent in irrigation water supply systems, both at the network (Mareels et al. (2005)) and field scales (North et al. (2010)). There is thus a need to predict soil moisture levels, based on current levels, crop growth stage (which affects evapotranspiration) and forecast weather conditions. This will enable irrigation water requirements to be forecast, and ensure timely delivery.

Many previous works have reported on soil moisture monitoring, focusing on current conditions. Thompson et al. (2007) suggested the use of soil matric potential (otherwise known as soil tension) is more reliable than water content based sensing for scheduling irrigation. Navarro-Helln et al. (2016) trained an irrigation decision system based on current weather and soil moisture conditions, and agronomist recommendations for watering in the coming week, but didn't consider forecast weather conditions. Adeyemi et al. (2018) predicted soil moisture one day ahead using a dynamic neural network model, trained on soil moisture and meteorological measurements. McCarthy et al. (2014) described a model predictive control strategy to optimize cotton irrigation to achieve a variety of outcomes. The strategy utilized a crop growth model (OZCOT) and measured and forecast weather data. This is a powerful methodology, provided an accurate crop model, the expertise to train it and implement the system

are available. Delgoda et al. (2016) fit linear models to simulated and measured soil moisture deficits as a function of evapotranspiration, rainfall and irrigation amounts and showed how it could be used to calculate daily irrigation amounts given a required moisture deficit. Goldstein et al. (2018) focused on evaluating a range of machine learning algorithms to provide recommended irrigation amount in the coming week based on different combinations of soil moisture and weather data. Montgomery et al. (2015) developed an online tool using evapotranspiration data from weather stations and forecasts together with NDVI from satellite remote sensing to predict soil moisture deficits, but it didn't integrate actual local soil moisture measurements.

We aim to use (i) measured soil moisture data, (ii) measured weather conditions, (iii) measured crop coefficient (using remotely sensed NDVI) to predict future soil moisture using weather forecasts. This can then be used to decide on irrigation water requirements in the coming seven days, and potentially opens the way to automating water orders and automating on-farm irrigation controllers in the future.

2. MATERIALS AND METHODS

The location of the study site was at Whitton in New South Wales, Australia. An NDVI image of the site from Sentinel-2 is shown in Fig. 1. The location of the main soil moisture sensors and weather station are indicated.

2.1 Soil moisture and weather monitoring using the WiField platform

A WiFi network was installed at the site. This consisted of a cellular data modem connected by ethernet to a Ubiquiti Nanostation M2 directional WiFi access point. This was able to provide internet connectivity to the edges of all plots, a maximum distance of about 800 metres from the access point.

Ten WiField agricultural data loggers (Brinkhoff et al. (2017)) were installed, with a variety of sensors connected. Each of these used an Electric Imp WiFi microcontroller to gather sensor data and send it to various online cloud databases. One of the loggers was configured as a weather station, with temperature, humidity, wind, rain and solar radiation sensors connected. This gathered local meteorological data every 10 minutes.

Each plot had a WiField logger with two Watermark 200SS soil tension sensors at 25 cm below the surface and an Enviropro multisensor capacitance probe. These gathered data hourly. The temperature from the Enviropro sensor at 25 cm and the resistance from the Watermark sensors was used to calculate the soil tension in kPA using a standard formula (Shock et al. (1998)).

These WiField loggers uploaded data to Google Cloud Platform. The data was ingested and processed in real time as described in the following section.

2.2 Data processing

The data from the probes (hourly), weather station (10 minutes) and NDVI from Sentinel-2 processed in Google



Fig. 1. Sentinel-2 NDVI image of the site from 2019-01-13, showing the location of the soil moisture loggers and weather station (WS). Green indicates NDVI>0.9, red indicates NDVI<0.7.

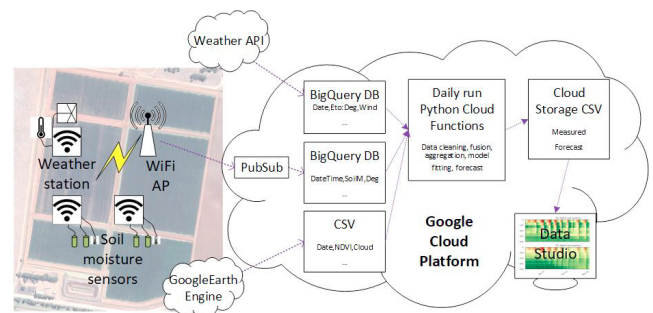


Fig. 2. Field and online data sources, processing and visualization flow.

Earth Engine (≈ 5 days), was ingested and analyzed in Google Cloud Platform as shown in Fig. 2. Daily aggregate data was produced using a Python Cloud Function with the following steps:

- The last daily value of the soil tension and soil moisture readings was selected (in the hour before midnight), as this will be affected by the total daily evapotranspiration, which is calculated from midnight to midnight.
- Irrigation dates were detected by checking the delta between the previous and current day's soil tension and rainfall readings. If the current soil tension is near 0kPA, and the previous day's is sufficiently low (<-

20kPa), and there was limited rainfall, then there was an irrigation applied in the current day.

- Per-plot NDVI was obtained from Sentinel-2 imagery and processed in Google Earth Engine. Images corrupted by cloud were removed. The NDVI values were interpolated between image acquisition dates.
- Daily meteorological data from the weather station was aggregated. Daily reference evapotranspiration (ET_o) was calculated using the Penman-Monteith equation Walter et al. (2004) and other daily statistics such as rainfall and growth degree days (with base temperature of 12°C) and solar radiation were computed.
- Crop evapotranspiration (ET_c) was calculated from ET_o using $ET_c = K_c \times ET_o = (1.37 \times NDVI - 0.086) \times ET_o$, a standard formula used for cotton (Montgomery et al. (2015)).
- Cumulative data (evapotranspiration, rainfall, temperature etc.) between irrigation dates was calculated. The cumulative sums were reset to zero at each irrigation, so it was correlated with the soil tension, which is reset to near 0kPa after irrigation due to soil saturation.
- A linear regression model was trained to predict the observed soil moisture based on meteorological, NDVI and irrigation history.
- Forecast weather was obtained from an online API (darksky.net), and the evapotranspiration and other meteorological daily statistics were computed.
- The forecast weather data and the most current NDVI was used by the latest model to predict the soil moisture for the next 7 days.
- The historical and forecast data was stored in a CSV file in a Cloud Storage Bucket.

A live visualization was created in Google Data Studio, which accessed the constantly updating database. It showed previous and forecast weather and soil tension among other variables. This website was accessible to collaborators and was used to manage irrigations through the season.

3. RESULTS

3.1 Irrigation characteristics

Table 1 shows the irrigation dates for the M3 plot. It indicates some of the possible factors involved in deciding when to irrigate. Some growers simply schedule irrigation after a fixed number of days. This does not account for high stress periods, such as consecutive high temperature and/or evapotranspiration days. Others use soil moisture probes. For example, in this region, it was found that cotton stomatal conductance began to decrease at soil tension around -60kPa (Ballester et al. (2019)), suggesting this as a suitable threshold for irrigation. Others irrigate off evapotranspiration in millimeters, taking into account the rain. Crop growth stage can also affect irrigation decisions, as cotton for example can tolerate more stress later in the season than during the critical flowering stage. For the case shown, it can be seen that irrigations were generally applied at soil tension around -60kPa, and evapotranspiration around 80-100 mm. More days between

irrigations were allowed when there was rainfall, and when the average daily evapotranspiration was lower.

The soil moisture, rainfall and NDVI data for each of the plots is shown in Fig. 3. The irrigation events are observable when the tension resets to near 0kPa. Rainfall events can also be seen to correlate with jumps in tension. The variability between soil moisture between the plots is evident. Some of this variation is related to irrigation system performance, which will be discussed below. Part of the variation between sensors may also be due to factors local to individual sensors. Such factors include soil cracking and root growth. Multiple sensors are needed to minimise these sources of variability.

The NDVI plot (Fig. 3) follows the expected cotton crop growth profile. Consistent differences are observed between plots, for example W3 and W2 have consistently lower NDVI. M1 NDVI is higher. This variation is likely due to it the order that the plots are irrigated in (with those irrigated later suffering more water stress) and top soil depth variation from laser levelling. The temporal fluctuation in NDVI is due to irrigations. When an image acquisition occurs soon after an irrigation, NDVI is higher as the crop is greener. Conversely, when an acquisition occurs many days after an irrigation, the NDVI is lower due to water stress.

To analyze the performance of the irrigation system, the WiField loggers were set to record soil moisture at 10 minute intervals (instead of the default 1 hour) during an irrigation event. The results are shown in Fig. 4. In this layout, M1 is irrigated first, and once the water reaches a set height, the water is then allowed to flow to M2, then to M3. W1 is irrigated at the same time as M3, and when the water reaches the set level, W2 and finally W3 are irrigated. There was no logger in the W1 plot. The figure shows that water may take almost 2 days to reach W3 from the start of an irrigation event. Also shown in Fig. 4 is the detailed air temperature readings during the irrigation event, and daily evapotranspiration. This irrigation occurred during a very hot period with high evapotranspiration. W3 experienced two days above 40°C and high evapotranspiration while it was already water stressed. A summary of the delay for water to reach each of the plots is given in Table 2. Also shown are NDVI, tension summaries and the yield data from hand harvests around each of the logger locations. The later irrigated plots had lower NDVI and lower yields.

3.2 Soil moisture forecasting

A portion of the dataset for the M3 plot is shown in Table 3. The cumulative crop evapotranspiration (with rainfall subtracted) was reset on each irrigation date. The input variables included daily variables as well as variables accumulated from the start of each irrigation period (reference and crop evapotranspiration, temperatures, rainfall etc.). The input variables were processed with a second order polynomial, producing all squared and cross terms, resulting in 77 variables. The dataset was split by irrigation dates. The algorithm stepped through the irrigation dates, holding out the data starting from the last irrigation in each step for testing, and using the data prior to the irrigation for training. The training data was further

Table 1. Irrigation dates, soil and environmental conditions immediately prior to irrigation for the M3 plot.

Date	Days	Tension (kPA)	ETc total (mm)	ETc daily (mm)	Rain (mm)	GDD (°C)	Max temp (°C)
2018-12-21	16.0	-42.6	63.3	4.0	23.0	216.1	38.1
2019-01-01	10.0	-78.5	85.9	8.6	8.0	167.4	42.8
2019-01-10	8.0	-64.5	84.9	10.6	0.0	126.0	43.4
2019-01-18	7.0	-84.5	97.6	13.9	1.0	143.2	46.1
2019-01-25	6.0	-55.2	77.7	13.0	0.0	114.9	42.6
2019-02-03	8.0	-57.4	79.8	10.0	4.0	126.2	42.9
2019-02-17	13.0	-56.4	121.7	9.4	26.0	153.7	40.0
2019-02-27	9.0	-62.8	95.1	10.6	13.0	102.6	39.6
2019-03-09	9.0	-59.0	53.6	6.0	9.0	70.0	36.2

Table 2. Per plot data for the 2019-01-17 irrigation, and final yield.

Plot	Max NDVI	Water delay (hours)	Min tension (kPA)	Mean tension (kPA)	Yield (bales/ha)
M1	0.89	0	-71.05	-16.58	13.43
M2	0.89	8	-148.82	-45.68	10.63
M3	0.88	21	-84.53	-20.36	10.61
W2	0.88	30	-60.65	-14.03	9.17
W3	0.87	37	-116.56	-40.16	7.68

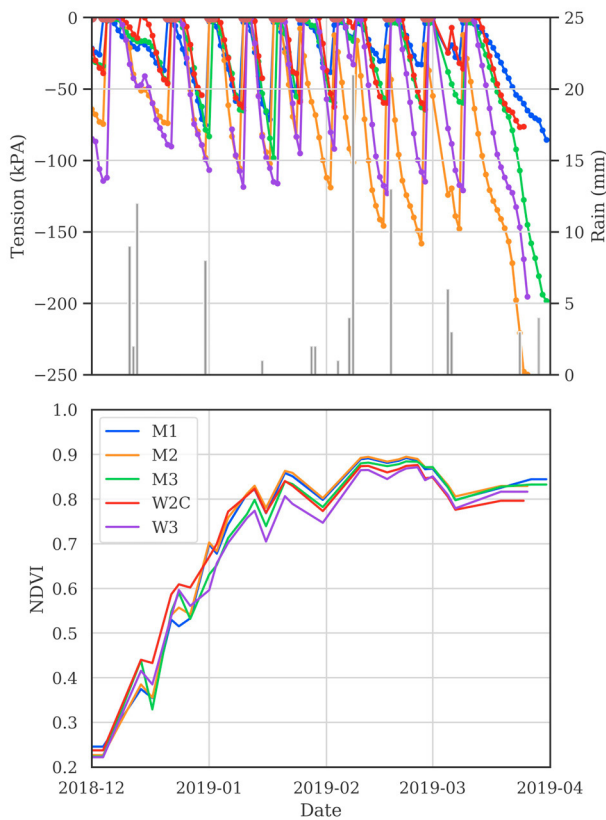


Fig. 3. Soil tension and NDVI for each plot (computed from Sentinel-2 images) and rainfall over the season.

separated for tuning the most important parameters of each regression algorithm using cross validation. Using this methodology, we could assess accuracy of the model in forecasting future soil moisture as the season progressed (and progressively more training data was available to the model).

A number of models were tested, including Lasso (linear regression with variable selection), Decision Trees, Random Forests and Support Vector Machines using the

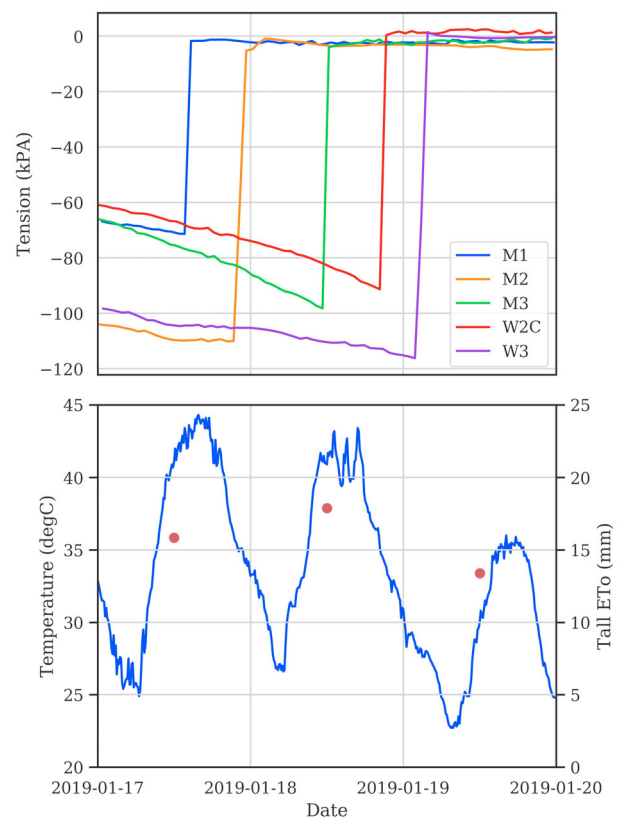


Fig. 4. Soil tension and weather data, focusing on the 2019-01-17 irrigation event, showing additional high temperatures and evapotranspiration experienced by the last plots to be irrigated.

Scikit-learn library in Python (Pedregosa et al. (2011)). The models were assessed by computing the absolute error between measured and predicted soil moisture 7 days from an irrigation (with training data consisting only of data before the irrigation). The absolute error rather than percentage error is used, as when the readings are near 0kPA, the percentage error can be very large but of little consequence. The average of these errors through

Table 3. Portion of the dataset for the M3 field. The soil tension is the predicted variable. ETc is crop evapotranspiration, and GDD is growth degree days. "Reset" indicates variables that were reset on irrigation dates. (f) indicates values computed from weather forecasts, used to predict future soil tension (x).

Date	ETc reset (mm)	GDD reset (°C)	Tension (kPA)
2019-01-17	97.6	143.2	-84.5
2019-01-18	0.0	0.0	-2.3
2019-01-19	13.5	17.4	-0.8
2019-01-20	23.0	32.6	-1.9
2019-01-21	34.4	49.0	-11.3
2019-01-22	49.3(f)	72.4(f)	x
2019-01-23	63.3(f)	93.6(f)	x

Table 4. Mean error of predicted soil tension (kPA) 7 days from each irrigation event for each plot, using multiple modeling methods.

Field	Lasso	Decision Tree	Random Forest	Support Vector Mach.	Linear Reg.
M1	7.9	9.3	9.6	8.0	8.8
M2	27.5	29.7	30.9	25.9	37.8
M3	12.8	11.4	13.3	14.3	8.2
W2	18.4	14.6	15.4	19.9	18.7
W3	31.2	23.4	16.9	33.4	29.9

the season was taken. The results are shown in Table 4. The accuracy in the M2 plot is relatively poor. This is most likely due to sensor issues. The sensors in this plot gave unexpected results, as the tension after irrigations did not reset to 0kPA later in the season (see Fig. 3). In this context, techniques to detect sensor failure or deterioration would be useful, such as the continuous data prediction and comparison method that was used for sewer temperature sensors in Thiyagarajan et al. (2018). There is some variability in the best performing model per plot, but the Random Forest models produce close to the best results across all plots.

The models gave information on the most important predictor variables. The two most consistently observed important predictors were the square of cumulative crop evapotranspiration minus rainfall, and the square of growth degree days (both these are reset on irrigation dates as shown in Table 3). It is intuitively reasonable that the cumulative input variables should be squared, as the tension typically stays more than -10kPA for several days after an irrigation, and begins dropping more quickly relative to evapotranspiration as the soil dries. The results for a simple linear regression using these two most important input variables are comparable with the more complex methods as shown in Table 4. The advantage of using a linear regression model is that it is easily interpretable, and we can compare the model coefficients between plots. The models can also be easily transferred to another software platform or implemented in spreadsheets, which may be useful for agronomists. The regression equation was:

$$\text{Soil tension} = a\left(\sum(ETc - \text{rain})\right)^2 + b\left(\sum GDD\right)^2 \quad (1)$$

The regression intercept was not statistically significant so was omitted. The prediction performance for two example irrigation intervals in the middle of the season are shown

in Fig. 5 for the M3 plot. Both the prediction during the training interval, and during the forecast interval is shown. The model attempts to capture the effect of the 25mm of rain on 2019-02-07 and 2019-02-08 (Fig. 5b). More training data with more rainfall events would be needed to improve the prediction of the effect of rain. The coefficients of equation (1) are also shown, and are relatively stable once the model converges after the first couple of irrigations. The figure also shows the prediction error for the training and forecast data as the season progresses.

To investigate the effect of soil tension prediction errors on determining the time to irrigate, equation (1) can be used. For the M3 plot, $a \approx -0.008$ and $b \approx -0.0006$. The average daily ETc from December to March was around 9mm, and average GDD was 14°C. If irrigation is triggered at -60kPA, (1) says this will occur on average after 8.9 days. A 10kPA error will result in irrigation time error of -19 hours to +17 hours. The errors reduce as the time to irrigate gets nearer. The far off estimate can be used to order water, and the exact time to irrigate refined as the time gets closer.

4. CONCLUSION

This study demonstrated a platform for ingesting soil moisture, weather and crop vigour (NDVI) data in real time. Models were trained on the data continuously through the season, allowing predictions of soil moisture to be made using weather forecast data. After the season, different modeling methods were trialed on the dataset. The system demonstrated variability between the plots due to irrigation system delays. The soil moisture forecasting methodology could be used in irrigation decision support tools, and potentially as part of a closed-loop irrigation automation system.

ACKNOWLEDGEMENTS

This work was funded by the Cotton Research and Development Corporation, project 1819FRP084. The authors are grateful for the support of the Irrigation Research and Extension Committee, and for assistance in field work from Rob Houghton and Wendy Quayle.

REFERENCES

- Adeyemi, O., Grove, I., Peets, S., Domun, Y., and Norton, T. (2018). Dynamic Neural Network Modelling of Soil Moisture Content for Predictive Irrigation Scheduling. *Sensors*, 18(10), 3408.
- Ballester, C., Brinkhoff, J., Quayle, W.C., and Hornbuckle, J. (2019). Monitoring the Effects of Water Stress in Cotton Using the Green Red Vegetation Index and Red Edge Ratio. *Remote Sensing*, 11(7), 873.
- Brinkhoff, J., Hornbuckle, J., Quayle, W., Lurbe, C.B., and Dowling, T. (2017). WiField, an IEEE 802.11-based agricultural sensor data gathering and logging platform. In *2017 Eleventh International Conference on Sensing Technology (ICST)*, 1–6.
- Delgoda, D., Saleem, S.K., Malano, H., and Halgamuge, M.N. (2016). Root zone soil moisture prediction models based on system identification: Formulation of the theory and validation using field and AQUACROP data. *Agricultural Water Management*, 163, 344–353.

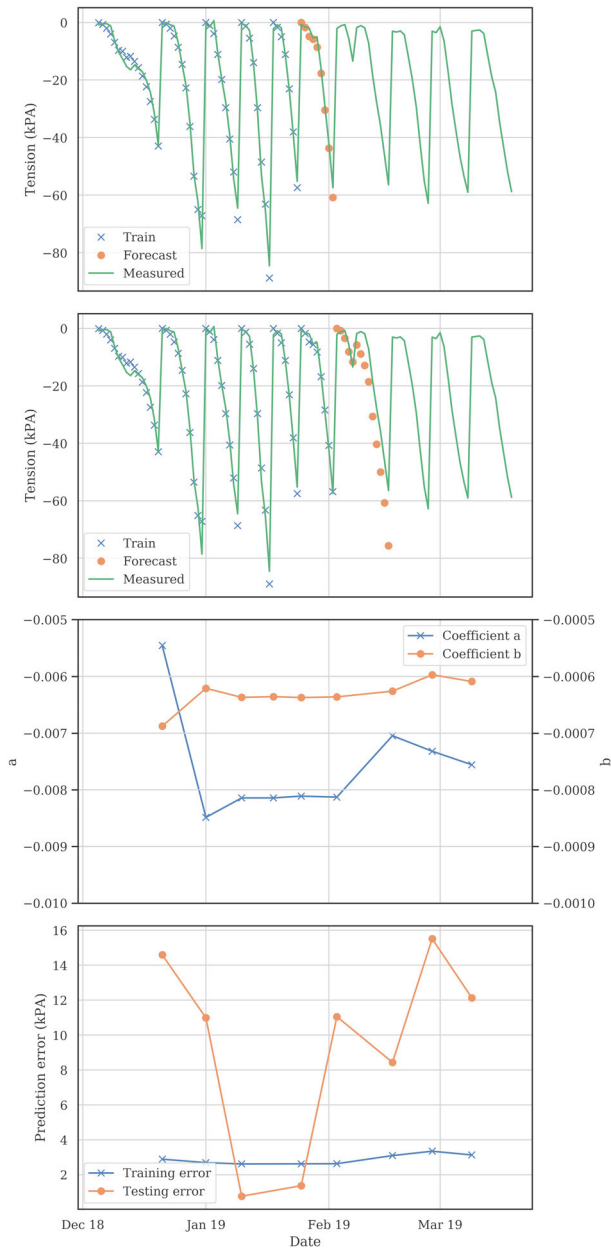


Fig. 5. Linear regression model for the M3 plot, evaluated through the season. (a) Predicting soil tension from the 2019-01-25 irrigation. (b) Predicting soil tension from the 2019-02-03 irrigation. (c) Evolution of the model coefficients as the season progresses. (d) Model error on the training and testing data as the season progresses.

Fereres, E., Goldhamer, D.A., and Parsons, L.R. (2003). Irrigation Water Management of Horticultural Crops. *HortScience*, 38(5), 1036–1042.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., and Snyder, P.K. (2005). Global Consequences of Land Use. *Science*, 309(5734), 570–574.

Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O., and Ravid, G. (2018). Applying machine learning on sensor data for irrigation recommendations: revealing the agronomists tacit knowledge. *Precision Agriculture*, 19(3), 421–444.

Harrison, P. (ed.) (2002). *World agriculture: towards 2015/2030: summary report*. Food and Agriculture Organization of the United Nations, Rome. Food and Agriculture Organization of the United Nations.

Mareels, I., Weyer, E., Ooi, S.K., Cantoni, M., Li, Y., and Nair, G. (2005). Systems engineering for irrigation systems: successes and challenges. *IFAC Proceedings Volumes*, 38(1), 1–16.

McCarthy, A.C., Hancock, N.H., and Raine, S.R. (2014). Simulation of irrigation control strategies for cotton using Model Predictive Control within the VARIwise simulation framework. *Computers and Electronics in Agriculture*, 101, 135–147.

Montgomery, J., Hornbuckle, J., Hume, I., and Vleeshouwer, J. (2015). IrriSAT weather based scheduling and benchmarking technology. 4.

Navarro-Helln, H., Martinez-del Rincon, J., Domingo-Miguel, R., Soto-Valles, F., and Torres-Sánchez, R. (2016). A decision support system for managing irrigation in agriculture. *Computers and Electronics in Agriculture*, 124, 121–131.

North, S., Griffin, D., Grabham, M., and Gillies, M. (2010). Improving the performance of basin irrigation layouts in the southern Murray-Darling Basin. *CRC for Irrigation Futures technical report*, 9/10.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, . (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.

Sauer, T., Havlk, P., Schneider, U.A., Schmid, E., Kindermann, G., and Obersteiner, M. (2010). Agriculture and resource availability in a changing world: The role of irrigation: agriculture and natural resources. *Water Resources Research*, 46(6).

Shock, C.C., Barnum, J.M., and Seddigh, M. (1998). Calibration of Watermark Soil Moisture Sensors for Irrigation Management. *Proceedings of the International Irrigation Show, Irrigation Association*, 139–146.

Thiyagarajan, K., Kodagoda, S., Nguyen, L.V., and Ranasinghe, R. (2018). Sensor Failure Detection and Faulty Data Accommodation Approach for Instrumented Wastewater Infrastructures. *IEEE Access*, 6, 56562–56574.

Thompson, R.B., Gallardo, M., Valdez, L.C., and Fernandez, M.D. (2007). Using plant water status to define threshold values for irrigation management of vegetable crops using soil moisture sensors. *Agricultural Water Management*, 88(1), 147–158.

Walter, I.A., Allen, R.G., Elliott, R., Jensen, M.E., Itenfisu, D., Mecham, B., Howell, T.A., Snyder, R., Brown, P., Echings, S., Spofford, T., Hattendorf, M., Cuenca, R.H., Wright, J.L., and Martin, D. (2004). ASCE's standardized reference evapotranspiration equation. In *Watershed Management and Operations Management 2000*.