Root zone soil moisture prediction models based on system identification: Formulation of the theory and validation using field and AQUACROP data

Dilini Delgoda a, *, Syed K. Saleem b, Hector Malano d, Malka N. Halgamuge b

a Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia
b Department of Electrical and Electronics Engineering, The University of Melbourne, Parkville, Victoria, Australia

ABSTRACT

In model-based irrigation control, the root zone soil moisture deficit (RZSMD) is maintained based on the water balance. To predict RZSMD in real-time, effective rainfall, irrigation and crop evapotranspiration need to be calculated online. Estimating the first two variables is more important yet tedious due to practical limitations of knowing the amount of water actually infiltrated into the soil. In order to solve this problem, we propose to apply system identification on water balance data to obtain a linear time series model. We further investigate how to carry out the modelling (i) under saturated conditions, (ii) when there is a rule-based irrigation control, and (iii) under measurement noise in the soil moisture readings. Using synthetic data we obtained a model fit above 80% in all cases. Additionally, we show the model optimality and applicability with an independent dataset, using residual tests. For two sets of field data, we observed model fits of 84% and 63%, and satisfaction in all residual tests. Simplicity in the model reduces calibration efforts whereas its linearity and adequacy recommend it for real-time irrigation control applications. In summary, the results indicate that a first order linear time series model based on system identification can successfully predict RZSMD in a real-time irrigation control system.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Obtaining a healthy plant growth is important in agriculture, for which it is desirable to maintain soil moisture at a favourable level. This level is decided based on the crop type, soil type and crop growth stage. The soil moisture deficit is the difference between the desirable soil moisture level and the current soil moisture level. When the soil moisture deficit is greater than zero, the plant requires application of water.

Irrigation is the application of water to meet the water demand by using a variety of methods such as sprinklers, furrows and drippers. Real-time irrigation control and scheduling involves irrigation operations on a daily or hourly basis or over a short period usually less than a week (Gowing and Egieji, 2001; Protopapas and Georgakakos, 1990; Rao et al., 1992).

Some real-time irrigation control methods are rule-based. This includes farmers using their empirical knowledge of soil moisture along with meteorological forecasts and soil moisture readings to create heuristic irrigation rules or automatic irrigation software that activate actuators after a predefined threshold in soil moisture level is reached (Hibbs et al., 1992; Hornbuckle et al., 2009; Goldhamer and Fereres, 2003; Allam, 2002). These methods do not assure that the irrigation supply meets the actual demand. Due to their ad-hoc nature, inefficiencies may arise causing water losses or shortages. In model-based systems, the irrigation control aims to achieve a desired level of soil moisture based on the soil response to water. The soil water balance method has been widely used in the literature to explain the soil response to irrigation and other operational variables (Przybyla, 1996; Hess, 1996; de Jager and Kennedy, 1996; Specty, 1996; Bailey, 1996). The concept involves calculating the irrigation amount by balancing inflows to and outflows from a soil profile. The daily water balance can be used to restore the water loss (Hess (1996)) or it can be used to calculate the optimal irrigation amount using methods such as model predictive control, a few days in advance (Saleem et al., 2013).

However, when using the water balance for irrigation control, it is always expected that the effective values of inflows and outflows are known. For example, if the total rainfall amount for a given

* Corresponding author. E-mail addresses: d.delgoda@pgrad.unimelb.edu.au (D. Delgoda).
location is known, the effective value that is going to infiltrate into the soil needs to be calculated. Unless the irrigation system is highly efficient and it is assured that the irrigation is applied directly at the point where the soil moisture measurements are made, knowing the fraction of irrigation amount that contributes to refill the soil moisture is essential. Due to fast drainage and surface sealing, a certain portion of applied water on a given day may run off before infiltrating into the soil profile. Similarly, some of the water may be lost as deep percolation. The same observation applies to evapotranspiration since the estimated crop evapotranspiration may be slightly different from the true crop evapotranspiration. Calculating these effective values is a tedious and time-consuming task that requires experimental procedures. Some studies based on the water balance give attention to uncertainty in evapotranspiration and soil moisture deficit (Aboitiz and Labadie, 1986; Or and Hanks, 1992; Huang, 2004; Wu et al., 2001) which in essence estimates true crop evapotranspiration from estimated crop evapotranspiration. However, calculating the effective values of rainfall and irrigation without experimental procedures has only been given scant attention to-date.

In addition, there may be incompatible measurement units related to different components in the water balance. For example, soil moisture deficit is measured as the depth of water in millimetres per 100mm of soil whereas rainfall may be expressed in litres. Conversion between these components will need more information about soil water characteristics. Further, if the scale of one variable is considerably greater than other variables, large rounding errors may appear. In irrigation control, it is enough to know the soil moisture response to given irrigation amounts and weather conditions without knowing the soil internal water dynamics. We exploit this fact to build a grey-box model of the soil water balance.

There are few instances where the relationship between the variables in water balance is given some attention. The study by Timm et al. (2011) observes the correlation among pairs of variables such as between soil moisture and evapotranspiration. However, the relationship among all variables is not quantified. In their studies, Manabe and Denworth (1990) and Parlangie et al. (1992) represent the water balance as an Auto Regressive AR (1) model that considers water events as white noise. On the other hand, Ooi et al. (2008) use system identification to identify the effect of two variables i.e. temperature and irrigation on the soil moisture level, yet they overlook the impact of rainfall. Also, the hourly time scale used in Ooi et al. (2008) is not readily applicable to irrigation control and limits the modelling to the uppermost soil layer.

In this paper, we propose a grey-box method based on system identification to be used in irrigation control. The advantage of this method is that it allows: (i) the variables in water balance model to be measured at neighbouring locations, (ii) the physical properties of the soil to be unknown and (iii) measurements to be exempted from being converted to a common set of units. We developed and evaluated two linear time series models to represent the water balance. One is the extended form of the other. The two models are compared under few different conditions. As an added benefit, such a system identification method can be useful when predicting the soil moisture deficit when a minimum amount of historical data on evapotranspiration is available. This point will be discussed later in the paper. We further investigate the modelling process under saturated conditions, rule-based irrigation control method that leads to closed-loop identification, and measurement noise in the soil moisture readings.

Section 2 of the paper proposes system identification as a means of modelling RZSMD and then delves into different scenarios with increasing complexity. Section 3 evaluates the model fit and validity of the model using synthetic data generated by the model AQUACROP (Steduto et al., 2012) and field data from Dookie, Victoria, Australia. Application of the proposed method in real-time irrigation control is demonstrated using a sample irrigation scenario. Finally, Section 4 outlines the conclusions and identifies future work.

2. Materials and methods

2.1. Soil moisture physical description

Water usable by plants is held in the soil region called the root zone. The root zone soil moisture level is considered to be optimal if it is maintained at a point called field capacity (FC) or close to it. If the root zone soil moisture level goes below the level called permanent wilting point (PWP) due to outflows such as evapotranspiration and deep percolation, the plants would wilt and the damage to growth is irreversible. The water held between PWP and FC is called total available water (TAW). The saturation point (SP) is situated above FC and is an ephemeral condition in soils with free drainage. The soil moisture level at saturation can occur under high water inflows from irrigation or rainfall. This makes the soil not suitable for plant growth. Thus, it is common practice aiming to maintain soil moisture close to FC.

The water balance approach was first introduced by Jensen et al. (1971) to support irrigation scheduling. Other researchers used it in subsequent years (Aboitiz and Labadie, 1986; Or and Hanks, 1992). It considers water inflows to and outflows from the root zone. Fig. 1 illustrates the root zone water balance, which includes inflows (irrigation and rainfall) and outflows (evapotranspiration, runoff and deep percolation below the plant’s root zone).

As per the water balance model concept, soil moisture level within the soil column M varies as

\[
\frac{dM}{dt} = -\epsilon(t) + P(t) + I(t) - G(t) - R(t).
\]

where \(t\) is the time, \(\epsilon\) is the true crop evapotranspiration which will be explained shortly, \(I\) is the effective irrigation, \(P\) is the effective rainfall, \(G\) is the deep percolation and \(R\) is the runoff (Aboitiz and Labadie, 1986).

Actual evapotranspiration is the quantity of water that is actually removed from a surface due to the processes of evaporation and transpiration (McMahon et al., 2013). In the paper Allen et al. (1998), the term crop evapotranspiration is used to refer to this variable in the context of a vegetation surface, and details how to estimate it using a crop coefficient and a reference evapotranspiration term. The reference evapotranspiration is calculated for a

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t)</td>
<td>time (min)</td>
</tr>
<tr>
<td>(k)</td>
<td>time step index</td>
</tr>
<tr>
<td>(M)</td>
<td>soil moisture level (mm)</td>
</tr>
<tr>
<td>(E)</td>
<td>calculated evapotranspiration (mm)</td>
</tr>
<tr>
<td>(P)</td>
<td>effective rainfall (mm)</td>
</tr>
<tr>
<td>(I)</td>
<td>effective irrigation (l/s, mm)</td>
</tr>
<tr>
<td>(G)</td>
<td>deep percolation (mm)</td>
</tr>
<tr>
<td>(R)</td>
<td>runoff (mm)</td>
</tr>
<tr>
<td>(D), (RZSMD)</td>
<td>root zone soil moisture deficit (mm)</td>
</tr>
<tr>
<td>(P)</td>
<td>measured rainfall (mm)</td>
</tr>
<tr>
<td>(I)</td>
<td>measured irrigation (mm)</td>
</tr>
<tr>
<td>(K_c)</td>
<td>crop coefficient</td>
</tr>
<tr>
<td>(FC)</td>
<td>field capacity (mm)</td>
</tr>
<tr>
<td>(PWP)</td>
<td>permanent wilting point (mm)</td>
</tr>
<tr>
<td>(SP)</td>
<td>saturation point (mm)</td>
</tr>
<tr>
<td>(TAW)</td>
<td>total available water (mm)</td>
</tr>
<tr>
<td>(RAW)</td>
<td>readily available water (mm)</td>
</tr>
</tbody>
</table>
reference crop which is a hypothetical green grass actively growing and well watered. Differences due to the crop canopy and aerodynamic resistance relative to the reference crop are accounted for within the crop coefficient.

The numerical values of above evapotranspiration variables that we calculate using various methods are estimations of their true quantities and may slightly differ from the true quantities. In this paper, we use new terms to denote this subtle difference.

- Estimated reference evapotranspiration $\tilde{t}$: numerical value estimated using Penman-Monteith or any other method by assuming standard conditions and using measurements that could be erroneous.
- True reference evapotranspiration $T$: true value of reference evapotranspiration (which is unobservable) under standard conditions and perfect measurements.
- Estimated crop evapotranspiration $E = K_cT$: numerical value estimated using equations in work such as Allen et al. (1998) using crop coefficient ($K_c$) and estimated reference evapotranspiration, assuming standard conditions.
- True crop evapotranspiration $\varepsilon = K_c K_s T$: true value of crop evapotranspiration (which is unobservable) under real (possibly water stressed) conditions. $K_s$ is the water stress factor.

When considering non-standard conditions where the crop experiences water stress, we make the assumption in this paper that water stress factor is constant throughout the training period. The assumption holds when the same irrigation threshold range is adhered to throughout the season.

When Eq. (1a) is discretised and the soil moisture term is replaced by negative soil moisture deficit ($D$) we obtain,

$$D_{k+1} = D_k + \tilde{\varepsilon}_k - \tilde{r}_k - \tilde{d}_k + \tilde{g}_k + R_k$$

where $k$ is the time index.

It must be noted that, in this paper we consider the soil moisture deficit, rather than actual soil moisture level. This can be explained as follows. When the plants grow, their water requirement expands because the root depth also expands. Therefore, having a greater soil moisture content does not assure a lesser water requirement. On the contrary, having a higher soil moisture deficit always implies that the water requirement is greater. Deficit can be estimated from soil moisture measurements or directly measured using methods such as tensiometer.

As mentioned in Section 1, rainfall and irrigation water can go out of the system boundaries as runoff due to soil saturation, fast water application and vegetation covers. Further, irrigation may not be applied at the same point we measure the soil moisture and the rainfall gauge may be situated some distance away. Therefore their impact would be different at the point of the soil moisture measurement. This leads to the terms effective rainfall and effective irrigation.

We propose two grey-box models based on system identification of the water balance that incorporate measured or estimated daily values of irrigation, rainfall and evapotranspiration with measured RZSM. When selecting the model structure, preference is given to simple models that can give an adequate level of accuracy. Since the model is intended for irrigation control methods such as model predictive control (MPC), a linear structure and relatively low order is preferred. A linear structure can be considered assuming a linear response of plant growth to RZSM in the range of interest. At first, we examine the models of order 1 and 2. If their performance is not adequate, then we consider higher order models and systematic methods to estimate the model order (Ljung, 1999). The process resulted in two models referred to in this paper as Model A and Model B.

2.2. Model A description

As Model A, we propose a simple grey-box model expressed as:

$$D_{k+1} = c_1 D_k + c_2 E_k + c_3 P_k + c_4 R_k + \epsilon_k$$

where $E$ is the estimated crop evapotranspiration, $P$ is the measured rainfall, $I$ is the gross irrigation, $c_1$, $c_2$, $c_3$ and $c_4$ are coefficients. The term $\epsilon$ is assumed to be white noise. In proposing the model, we make the following assumptions: irrigation efficiency is a constant; relationships between estimated evapotranspiration and true crop evapotranspiration and that between effective rainfall and actual rainfall are linear. Deep percolation and runoff due to saturation are neglected assuming that the soil moisture level does not reach saturation. Note that this point will be addressed later.

It is possible that some of the rainfall that occurred the previous days also contributes to the soil moisture change on the current day. In this case we should change model such that,

$$D_{k+1} = c_1 D_k + c_2 E_k + c_3 P_k + c_4 R_k + \epsilon_k$$

where $P$ is the measured rainfall during $n$ number of time steps such that $P_k = [P_k, P_{k-1} \cdots P_{k-n}]$, and $c_3 = [c_1, c_2 \cdots c_n]$. The term $c_3P_k$ is similar to the concept of antecedent precipitation index (API) given in rainfall runoff models (Heggen, 2001). However, in our model the decay term associated with a day’s rainfall is included in the coefficient $c_3$ that is determined using system identification. A similar form of array coefficient can be applied in irrigation although it is unlikely that infiltration from an irrigation event occurs over several days. For this reason, we do not consider this factor in our model. Yet, it is always possible to have a coefficient $c_4$ of a form similar to $c_3$ if necessary.

Note that $E$, $P$ and $I$ do not have to be measured at the same point as $D$. They can be measured some distance away provided that the correlation between the variable at the two locations is positive. For $E$ and $P$ this distance is in the range of up to a few kilometres (Huff, 1979; Harcum and Lofis, 1987) while for $I$ we can expect positive correlation within a few metres.

We assume that the ground water table is sufficiently deep to disable capillary rise until the root zone. For the current study, this assumption is shown to be valid as observed from the results. Incorporating capillary rise to the model would be desirable as future work where data is available for this situation.
Despite Ooi et al. (2008) having identified the possibility of a time delay between inflows and the soil moisture response, we have not observed this behaviour in our data. This could be due to the daily time scale used in our study whereas Ooi et al. (2008) use an hourly time step. Considering inflows and outflows in the water balance in Eq. (3), $c_1$ and $c_2$ must be positive whereas the other two coefficients must be negative. The noise term represents the complex inaccuracies in the model.

Since Eq. (3) is a linear model, we use linear least square method to calculate the coefficients (Ljung, 1999).

A comparison of, Model A and Eq. (1b) reveals equivalent terms:

$$
D_{k+1} + \psi = c_1 D_k \\
\varepsilon_k = c_2 E_k \\
-p_k = c_3 P_k \\
-I_k = c_4 I_k
$$

where $\psi$ denotes variations due to root growth. Runoff caused by fast water application and surface sealing is taken into account with in the rainfall and irrigation terms.

### 2.3. Model B description

Model B is a second order extension of Model A where the main improvement lies in the consideration of the soil moisture deficit on the previous day such that,

$$
D_{k+1} = c_1 D_k + c_{01} D_{k-1} + c_2 E_k + c_3 P_k + c_4 I_k + \varepsilon_k.
$$

where $c_{01}$ is an additional coefficient to those in Model A. We use the linear least square method to estimate the coefficients.

It can be hypothesised that the model order can be increased for soil types such as clayey soils, where the soil water movement is slower. To test the validity of this hypothesis, the performance of the two models is compared using soil types with different water retention properties. If there is a performance improvement with Model B, we can conclude that the hypothesis is valid.

### 2.4. Dealing with saturation

The proposed model is intended to be used in real-time calculation of irrigation amount. Thus, irrigation is defined as a function of other terms in the water balance model. Let the saturation amount in the water balance be denoted by $S_k$ that includes deep percolation and part of the surface runoff. It is obvious that $S_k$ cannot be used as an input variable during the real-time irrigation control process. On the other hand, it is assured that if the model is used for closed-loop irrigation control, there will not be over-irrigation. When there is saturation due to rainfall, the root zone soil moisture content will stabilize within a day or two. This justifies not having $S_k$ in the proposed model during the control phase.

However, this is not the case during the training phase where deep percolation and runoff due to saturation may appear. If there is no saturation, training can be carried out normally without $S_k$. If there is saturation, the saturated amount needs to be estimated from the training dataset. Once $S_k$ is estimated, linear least square method is performed on Eq. (2) with an additional term $c_5 S_k$ also added. Once all the coefficients are known, the term $c_5 S_k$ is simply removed from the model. According to Frisch-Waugh-Lovell theorem (Frisch et al., 1933), this does not alter the values of the coefficients already identified. The remaining model is suitable for the irrigation control phase, where saturation amount is not considered as an input variable.

### 2.5. Closed-loop identification

During the training phase, it is possible that the data was obtained from a field where irrigation was driven by a threshold soil moisture level (i.e. rule-based irrigation control systems mentioned in Section 1). This creates a closed-loop in the system model where soil moisture deficit and irrigation amount are dependent on each other. In such cases, the linear least square method still applies given that there is no measurement noise in the soil moisture readings and there is a delay in the closed-loop (Ljung, 1999). The latter condition is fulfilled in irrigation control, however the problem needs further research, when measurement noise is present.

### 2.6. Existence of measurement noise

Existence of noise is highly possible in soil moisture measurements. If that is the case, it will be required to find the relationship between true soil moisture deficit and other operational variables using noisy soil moisture data. Let soil moisture deficit measurement is denoted by $y$ and true soil moisture deficit be denoted by $x$. If the measurements are perfect $y_k = D_k$. If the measurements are noise corrupted, $x_k = D_k$ and $y_k = f_k + v_k$ where $f$ is a constant and $v$ is white measurement noise. Thus, for error-free data, identifying Model A reduces to applying $y_{k+1} = c_1 y_k + c_2 E_k + c_3 P_k + c_4 I_k + \varepsilon_k$ on the dataset. If noisy measurements are present, identification becomes equivalent to obtaining state space equations: $x_{k+1} = c_1 x_k + c_2 E_k + c_3 P_k + c_4 I_k + \varepsilon_k$; $y_k = f_k + v_k$. Prediction error method (Ljung, 1999; Soderstrom and Stoica, 1989) can be used to identify the coefficients.

### 3. Results and discussion

This section is formulated under few motives as follows:

1. To check the validity of the two proposed models using standard performance criteria under variations in:
   - soil types.
   - saturated and unsaturated conditions.
   - open-loop and closed-loop conditions.
   - noise level in the soil moisture readings.

2. To see if the increase of model order has an advantage over the first order model.

Two approaches were used to test the model performance. One is based on idealised numerical experiments using AQUACROP model data and the other is based on field data.

#### 3.1. Idealised numerical experiments – using AQUACROP model data

Using a physically based model to generate soil moisture data is an appropriate method to assimilate the soil water dynamics under idealised conditions because it allows control over model inputs such as irrigation and crop soil parameters. For this purpose, we used the model ‘AQUACROP’ developed by the Food and Agriculture Organization of the United Nations (FAO, 2011; Steduto et al., 2012). We input actual weather data to the model while irrigation was changed as per the scenario. AQUACROP generated a soil moisture trace in response to irrigation and weather conditions.

The default AQUACROP project ‘Foggia’ available in the AQUACROP project library was modified as described in Section 3.1.1. By default, the crop type used is tomato and weather data are given for the years 2000 and 2001 for Foggia, Italy.
3.1.1. Changing the files in Foggia project

The AQUACROP soil file was modified to represent the soil type and characteristics outlined in the AQUACROP manual (FAO, 2011) (see Supplementary Table S1 for parameter details).

For open-loop case, irrigation was set up using a given user defined schedule. For closed-loop case the irrigation rule was set up in AQUACROP to allow the maximum allowable soil moisture depletion percentage $a$ to be equal to a given value. Saturation was obtained by changing the irrigation amount. A sample set of generated data is shown in Supplementary figure S1.

Open-loop identification: The linear least square method was applied using the MATLAB function $lsqlin$ to obtain the model coefficients. It can be observed that the coefficient $c_3$ in both Eqs. (3) and (5) is equal to $c_5$, that implies that the rest of the elements of $c_3$ are zero. This may be due to the infiltration of applied water having already contributed to satisfy the soil moisture deficit of the current day. Hence, we use the scalar notation $c_5$ for this coefficient to indicate that it only scales the current rainfall value.

The system identified models were validated using data for the year 2001. The goodness of model fit is based on three criteria, namely: coefficient of determination, root mean square error and Nash-Sutcliffe coefficient. Their purpose is to measure the agreement between the actual data and the simulated data.

We further use so called residual tests to validate the model. The independence test is used to confirm that the model is able to perform similarly for an independent dataset yielding a residual not greater than the highest residual generated in the training process. The test checks whether the correlation between the inputs and the residual is sufficiently small (Ljung, 1999). It also is a test to indicate that the assumption on model linearity is valid. In addition, the model must satisfy the optimality condition. Otherwise, the soil moisture deficit could have been better predicted from past data. The optimality is checked by using a whiteness test (Ljung, 1999).

Fig. 2 and Supplementary figure S2 show the training and validation results of Model A for the unsaturated and saturated cases for seven soil types. The results for Model B is given in Fig. 3 and Supplementary figure S3. For both cases, the residual tests were successful for all the scenarios except for sandy soil (when saturation was not considered). Given the results of the goodness of model fit criteria and residual tests, one can conclude that both models perform similarly for all soil types. The difference in results for sandy soil indicates that the low water holding capacity that causes drainage flows in sandy soil, must be addressed through the extra step detailed in Section 2.4.

There is no improvement in model fit of Model B over Model A. Further, and contrary to what is expected, Model B does not match the real system better than Model A for clayey soil types. Thus, hereafter the testing is restricted to Model A only.

Modelling for saturated soil is similar to the unsaturated case. However, estimating saturated flows is practically not an easy task. If the saturation by over-irrigation is prevented, estimation reduces to calculating saturation flows by rainfall – which does not occur frequently.

Closed-loop identification: here too the linear least square method was applied during the training phase. The performance in this case is demonstrated in Fig. 4 and is similar to the open-loop case. Further, there is no difference under the change of saturation conditions.

Identification under measurement noise: In this case, white noise $\mathcal{N}(0, 1)$ is added to the AQUACROP soil moisture deficit data. MATLAB function $ssest$ based on prediction error method was used to do the training. The training and validation results for the open-loop and closed-loop cases are shown in Fig. 5 for unsaturated soil. Results are shown to be similar for the saturated soil as well.

For both open-loop and closed-loop identification, the results are shown to be of a good quality. However, there is a slight degradation in results in terms of goodness of fit for the closed-loop identification (with above 80% model fit as opposed to 90% model fit for the open-loop case).
Results based on AQUACROP data are highly encouraging, since they indicate that the model identification methods proposed in this paper are valid under changes in saturation conditions, open-loop and closed-loop identification scenarios, and appearance of noisy measurements. Results pertaining to closed-loop identification under noisy measurements are slightly inferior to those in open-loop identification however it is less significant.

1. **Fig. 3.** Open-loop identification of Model B for different soil types with AQUACROP data. Raw 1: with no saturation. (a) Model parameters. (b) Model fit during training ($R^2$, Nash-Sutcliffe, RMSE). (c) Model fit during validation ($R^2$, Nash-Sutcliffe, RMSE). Raw 2: with saturation. (d) Model parameters. (e) Model fit during training ($R^2$, Nash-Sutcliffe, RMSE). (f) Model fit during validation ($R^2$, Nash-Sutcliffe, RMSE). $R^2$ = coefficient of determination RMSE = root mean square error, Nash-Sutcliffe = Nash-Sutcliffe model efficiency coefficient. Refer to Supplementary Table S1 for the descriptions of the soil types.

2. **Fig. 4.** Closed-loop identification of Model A for different soil types with AQUACROP data and an irrigation rule such that maximum soil moisture depletion percentage from TAW, $a=40$ and it was replaced to $b = FC = 20$ mm. Raw 1: with no saturation. (a) Model parameters. (b) Model fit during training ($R^2$, Nash-Sutcliffe, RMSE). (c) Model fit during validation ($R^2$, Nash-Sutcliffe, RMSE). Raw 2: with saturation. (d) Model parameters. (e) Model fit during training ($R^2$, Nash-Sutcliffe, RMSE). (f) Model fit during validation ($R^2$, Nash-Sutcliffe, RMSE). $R^2$ = coefficient of determination RMSE = root mean square error, Nash-Sutcliffe = Nash-Sutcliffe model efficiency coefficient. Refer to Supplementary Table S1 for the descriptions of the soil types.
3.2. Validation using field data

Two datasets from a Shiraz vineyard and an Apple orchard located in the Dookie campus of The University of Melbourne were used to evaluate the model formulations developed in this study. The growing season at the experimental site spans between the beginning of October and end of February. The field measurements were obtained with an Enviropri® multi depth capacitance probe. Volumetric soil moisture content was measured at four depths from the surface i.e. 20 cm, 40 cm, 60 cm and 80 cm for Shiraz and at the first three depths for Apple. A reading at each depth represents the moisture content of the soil layer in the vicinity of that depth. The data spread is shown in Supplementary figure S4. Based on the total value of soil moisture measurements at the depths, the irrigators are turned ON and OFF by the software iNTELLiTRoL through latching solenoids. The space between two lines is 0.75 m for Shiraz and 1 m for Apple. The rate of water application is a uniform 3.5 l/h using drippers for Shiraz and 40 l/h using sprinklers for Apple. Rainfall data and weather data necessary to calculate evapotranspiration were recorded by the weather stations located at the site. In general, the soil type in the area is a variation of clay loam with the soil getting slightly heavier with depth (Downes, 1949).

Following the irrigation events, water extraction was observed at all depths. Soil moisture levels at two locations are shown in Fig. 6. Firstly, we identified the field capacity (FC) by identifying the points on the soil moisture curve at which the rate of drainage process slows down after an irrigation or rainfall event. RZSMMD model for each layer was calculated from the difference between FC and the soil moisture reading at a given depth. Total RZSMMD is the addition of these values. Crop coefficients for the simulations were obtained from the crop coefficient curves and data given in Allen (1986). As the original curves are given for the northern hemisphere, the curve had to be shifted to match Australian weather and to cover the seasonal period. Inputs to the models are given in Supplementary figure S5. The season spans from November 1st in 2011 to January 31st in 2012. Data from the first set of days in the season are used for model training and the remaining data for model validation.

Subsequently, model fit and residual tests were applied on the two datasets.

Training and validation results for the two datasets are shown in Fig. 7 and Supplementary figure S6. All residual tests were successful with the field data. Model fit is around 84% for Shiraz and 63% for Apple.

The degradation in the results of modelling Apple data could be due to errors in estimating the saturation flows. This signifies the importance of avoiding over-irrigation even from the modelling point of view, in addition to the reasons such as water waste and leaching. The validation fit is average (50%) in both cases. This could be attributed to the change in the irrigation rule within the season (Refer to Fig. 6(a) and (b)). Residuals pass the independence test and the whiteness test, assuring a similar performance with an independent dataset and model optimality. Apart from that, this difference can be ascribed to error in estimating the FC throughout the year. There could be errors in the estimation of the crop coefficients as these are obtained by shifting the FAO crop coefficient curves to match the Southern hemisphere. Nevertheless in overall, the field data support the results obtained with the synthetic data.

In summary, we can conclude that both the synthetic and field data show that Model A is a good approximation for soil moisture deficit when the primary variables of the water balance equation are present.

3.3. A sample application of system identified first order model in irrigation control

From the model validation described in the previous section, it can be observed that the models perform well in terms of optimality (whiteness test) and assuring the same level of performance with an independent dataset (independence test).

The purpose of modelling RZSMMD in this paper is to use it in real-time irrigation control. This section presents a simple real-time closed-loop irrigation control application of the grey-box models. Model A is selected over Model B because of its simplicity given that both models’ performance is similar.
We use a loam soil, with allowable moisture depletion $a = 50$ to estimate the parameters of Model A using the Foggia dataset for year 2000. Then, using the year 2001 dataset, irrigation is managed in real-time under the rule that soil moisture is refilled to maintain a zero soil moisture deficit below FC. Using the year 2001 dataset, the effective irrigation amount for day $k$ is calculated as follows:

$$I_k = D_{k+1} - c_1 D_k - c_2 E_k - c_3 P_k$$

**Fig. 6.** Data from Shiraz vineyard and Apple orchard of The Melbourne University Dookie campus. Raw 1: Variation of soil moisture level within different soil layers. Horizontal lines show the estimated FC for individual layers (a) Shiraz (b) Apple. Raw 2: Other data common to three locations (c) Shiraz (d) Apple.

**Fig. 7.** Identification of Model A using field data: Shiraz data with no saturation, Apple data with saturation. For both cases the irrigation is driven by soil moisture readings. (a) Model parameters. (b) Model fit during training ($R^2$, Nash-Sutcliffe, RMSE). (c) Model fit during validation ($R^2$, Nash-Sutcliffe, RMSE) $R^2$=Coefficient of determination. RMSE=Root mean square error, Nash-Sutcliffe=Nash-Sutcliffe model efficiency coefficient.
shown by a model fit above 99% using AQUACROP model data. We further investigate the performance (i) under saturated conditions, (ii) when there is a rule-based irrigation control, and (iii) under measurement noise in the soil moisture readings and obtained above 80% model fit in these cases. The model was also applied to two field datasets from a Shiraz vineyard and an Apple orchard in Dookie, Australia with model fit values of 84% and 63% respectively. The models met the residual tests in all instances. Whiteness test demonstrates model optimality whereas passing the independent test assures that the model would perform similarly with an independent data set. The use of proposed method is demonstrated by a simple real-time closed-loop irrigation application example.

Acknowledgements

We would like to thank Dr. Kithsiri Dassanayake of School of Land and Environment of The University of Melbourne and Dr. Minh Pham of NICTA Victoria for their advices and Mr. Marc Elliott for providing data from a Shiraz vineyard located in the Melbourne University Dookie campus.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agwat.2015.08.011.

References


Przybyla, C., 1996. A microcomputer scheduling program for supplementary irrigation. In: Irrigation Scheduling in Large-scale Sprinkler Irrigation in the Wielkopolska Region of Poland.


