

Statistical Assessment of a Numerical Model Simulating Agro Hydro-chemical Processes in Soil under Drip Fertigated Mandarin Tree

Phogat V^{1*}, Skewes MA¹, Cox JW^{1,2} and Simunek J³

¹South Australian Research and Development Institute, Australia

²The University of Adelaide, Australia

³Department of Environmental Sciences, University of California, USA

Abstract

Qualitative assessment of model performance is essential because reliable statistical comparison of observed data with simulated behaviour of a model reflects the performance and consistency of the mathematical tool under defined conditions. In this study we compared the measured temporal and spatial distribution of water content, soil solution salinity (EC_{sw}), and nitrate (NO_3^- -N) concentration in the soil beneath a drip-fertigated mandarin tree during a complete season with corresponding HYDRUS-2D simulated values using a range of standard statistical techniques, comprising mean error (ME), mean absolute error (MAE), root mean square error ($RMSE$), paired t -test (t_{ca}), coefficient of determination (R^2), Nash and Sutcliffe model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), relative index of agreement (IA_{rel}), modified E (E_1) and IA (IA_1).

Temporal and spatial values of ME , MAE , and $RMSE$ for water content (-0.04 to 0.05 $cm^3.cm^{-3}$) and salinity (-0.42 - 0.93 dSm^{-1}) were within an acceptable range. However, a relatively wider range in MAE (1.44 - 27.65 $mg.L^{-1}$) and $RMSE$ (2.00 - 39.57 $mg.L^{-1}$) values were obtained for NO_3^- -N concentrations measured weekly or at the 25-cm depth ($MAE = 21.2$ and $RMSE = 30.7$ $mg.L^{-1}$). Temporal and spatial $RMSE$ were higher than MAE , which suggests a slight bias in $RMSE$ due to squared differences between measured and simulated values. Similarly, the paired t -test (t_{ca}) showed significant differences for NO_3^- -N during the mid-season (85-140 DOY) for temporal (weekly) comparison and at several depths for water content (10, 25, 80 and 110 cm), salinity (100 and 150 cm) and NO_3^- -N concentration (25, 100 and 150 cm).

The R^2 values varied in a narrow range (0.5 to 0.59). Similarly, values for E (0.12-0.43), IA (0.80-0.84), and E_1 (0.26-0.32) and IA_1 (0.61-0.69) suggest that the model precisely predicted water content, salinity and nitrate concentration over the season, however, E_{rel} (-319.25) and IA_{rel} (-71.3) values were highly negative for nitrate concentration, indicating a mismatch. It was concluded that none of the evaluated measures described and tested the performance of the model for water, salinity and nitrate ideally. Each criterion had its specific advantages and disadvantages, which should be taken into account. Hence, sound model performance evaluation requires the use of a combination of different statistical criteria, which consider both absolute and relative errors. Judicious use of statistical criteria should lead to improvements in the modelling assessment of water, salinity and nitrate dynamics in soil under cropped conditions.

Keywords: HYDRUS-2D; Water content; Salinity; Nitrate; Statistical errors; Model efficiency

Introduction

The advent of high speed computers, which enable enhanced modelling capabilities and rapid development in mathematical software, has transformed the mathematical evaluation of natural processes. Models are extensively used in almost every field of science for problem solving and decision making, and the vadose zone of agricultural soils is not spared from this revolutionary change. Due to the advancement of micro-irrigation systems such as sprinkler and/or surface/subsurface drip irrigation, which has transformed irrigation and fertilizer practices, there is an increasing interest in evaluating and optimizing these high frequency systems for water and fertilizer use efficiency [1-3].

Development of mathematical tools has contributed towards improving irrigation system design and installation, and the monitoring of water and solute movement through the soil from point source applications. Additionally, models can save time and money because of their ability to perform long term simulations evaluating the effects of root zone processes and management activities on water quality, water quantity, and soil quality. Recently, a number of numerical codes have been developed or improved, such as SWAP, FEHM, HYDROBIOGEOCHEM, RZWQM, TOUGH2, APSIM, HydroGeoSphere, and HYDRUS, to cater to the needs of soil physicists/irrigation experts for evaluating irrigation systems and vadose zone

processes for water movement, climate variability, water quality, and solute transport in the soil [4-11]. Especially, the HYDRUS model has been used extensively for evaluating the effects of soil hydraulic properties, soil layering, dripper discharge rates, irrigation frequency, water quality, and timing of nutrient applications on wetting patterns and solute distribution [1-3,12-22]. The most significant aspect of these studies, which determines the utility of these models, is the evaluation of modelling outputs against field observed values. The model should generate data which closely mirror field observations, so that reliable conclusions can be drawn about real world processes.

There are a number of reasons why there is a need to evaluate model performance: (1) to provide a quantitative estimate of the model's ability to reproduce historic and future behaviour of agricultural/environmental systems; (2) to provide a means for evaluating

*Corresponding author: Phogat V, South Australian Research and Development Institute, Australia, Tel: +61 8 83039567; E-mail: vinod.phogat@sa.gov.au

Received January 19, 2016; Accepted January 22, 2016; Published January 25, 2016

Citation: Phogat V, Skewes MA, Cox JW, Simunek J (2016) Statistical Assessment of a Numerical Model Simulating Agro Hydro-chemical Processes in Soil under Drip Fertigated Mandarin Tree. Irrigat Drainage Sys Eng 5: 155. doi:10.4172/2168-9768.1000155

Copyright: © 2016 Phogat V, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

improvements to the modelling approach through adjustment of model parameter values, model structural modifications, the inclusion of additional observational information, and representation of important spatial and temporal characteristics of the domain; (3) to compare current modelling efforts with previous studies [23]. Field calibration and validation of the model requires conducting tests based on statistical measures, and is the most important aspect of testing the goodness of fit of values generated by the model. This process of assessing the performance of a model requires evaluation of the closeness of the simulated behaviour of the model to field measurements made within the domain.

Exhaustive evaluations and objective analyses have been carried out for models used in various fields of hydrological and hydraulic modelling [23-30]. Accepting the wide recognition and utility of HYDRUS for modelling water and solute movement under irrigation applications, there is a need to assess the performance of HYDRUS for potential sources of deviation using appropriate and simple indicators. Most field evaluation studies using HYDRUS either present graphical comparisons or subjective assessments [15,31-33], and generally considered only limited error and correlation estimates, i.e., an objective assessment [16,18,20,34] of the performance of the model to evaluate water, salt and nitrate movement in soils. These criteria may place emphasis only on a particular behaviour of the model, and may not be able to assess the overall efficacy of the model on a long term basis. Hence, there is a need to evaluate the performance of HYDRUS more vigorously, utilizing different error analyses, test of significance, regression analyses, and efficiency testing to clearly assess the model's sustained performance. It is important to compare the suitability and relative importance of each of these techniques for evaluating modelling predictions of water and solute transport under high efficiency irrigation systems.

In the present investigation, the performance of HYDRUS-2D in simulating water movement, soil solution salinity, and nitrate movement under a mandarin tree during one season was assessed, using eleven statistical measures: mean error (ME), mean absolute error (MAE), root mean square error ($RMSE$), paired t -test (t_{cal}), coefficient of determination (R^2), model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), relative index of agreement (IA_{rel}), modified E (E_f) and modified IA (IA_f). The aim of this comparison was to identify which subset of the statistical measures is most appropriate for evaluating model performance.

Materials and Methods

The statistical tests were employed on the measured and simulated data generated from the field experiment on mandarin and modelling simulations illustrated in our earlier paper Phogat et al. [3]. However, a brief description of experimental details and modelling technique is presented here.

Experimental detail

Modelling evaluation was performed on field experimental data collected at Dareton Agricultural and Advisory Station (34.10°S and 142.04°E), located in the Coomealla Irrigation Area, New South Wales, Australia for one season during 2006-2007. The field experiment involved surface drip irrigation of mandarin, established in October 2005. The trees were planted at a spacing of 5 m x 2 m between rows and plants, respectively. The trees were managed and fertilized following current commercial practices. The total yearly rainfall during the experimental period was 187 mm, which was significantly less than annual potential evapotranspiration (1400 mm).

Irrigation water was supplied through a surface drip system, with drip lines placed at a distance of 60 cm on both sides of a tree line. The laterals had 1.6 L.h⁻¹ pressure compensating drippers spaced at an interval of 40 cm. Irrigation was performed weekly, and the total seasonal irrigation was 432.8 mm. The salinity of the irrigation water (EC_w) was monitored daily, and ranged between 0.09 and 0.19 dS.m⁻¹, well below the EC_w threshold for irrigation of orange, a close relative of mandarin (1.1 dS.m⁻¹). Daily water content measurements were performed using Sentek® EnviroSCAN® capacitance soil water sensors, and soil water was sampled on a weekly basis using SoluSAMPLERS™ [35]. The extracted soil solution was analysed to determine soil solution salinity (EC_{sw}) and nitrate-nitrogen (NO₃⁻-N) content.

Modelling technique

The HYDRUS-2D software package was used to simulate the transient two-dimensional movement of water and solutes in the soil [11]. Refer to the HYDRUS technical manual for a detailed description of the governing equations describing variably-saturated flow using the Richards' equation, solute transport using the advection-dispersion equation, and root water uptake, as well as various initial and boundary conditions that can be implemented. In this approach, the drip tubing was considered as a line source, because in this twin line drip irrigation system the wetted patterns from adjacent drippers merge to form a continuous wetted strip along both sides of the tree [36,37]. Modelled observation nodes corresponded to the locations where EnviroSCAN probes (at depths of 10, 25, 50, 80, and 110 cm) and SoluSAMPLERS (at depths of 25, 50, 100, and 150 cm) were installed.

Soil hydraulic properties were described using the van Genuchten-Mualem constitutive relationships [38]. The spatial root distribution is defined in HYDRUS-2D according to Vrugt et al. [39]. We considered a simple root distribution model, in which the roots of mandarin trees expanded horizontally into all available space between the tree lines ($x_m = 200$ cm), were concentrated mainly below the drip emitter ($x^* = 60$ cm, $z^* = 20$ cm) where water and nutrients were applied, and extended to a depth of 60 cm ($z_m = 60$ cm).

Reduction of root water uptake due to water stress was described using the piecewise linear relation developed by Feddes et al. [40]. The following parameters in the Feddes et al. model were used: $h_1 = -10$, $h_2 = -25$, $h_3 = -200$ to -1000 , $h_4 = -8000$ cm, which were taken from Taylor and Ashcroft for orange. Reduction of root water uptake due to salinity stress, $\alpha_s(h_\phi)$, was described by adopting the Maas and Hoffmann salinity threshold and slope function [40,41]. The salinity threshold (EC_s) for orange (closely related to mandarin) corresponds to a value of the electrical conductivity of the saturation extract (EC_e) of 1.7 dS.m⁻¹, and a slope (s) of 16%.

The longitudinal dispersivity (ϵ_L) was considered to be 20 cm, and the transverse dispersivity (ϵ_T) was taken as one-tenth of ϵ_L , optimised in similar studies involving solute transport in soils [12,20]. Since NH₄NO₃ and mono-ammonium phosphate were the fertilizers used in our study, nitrification of NH₄⁺-N to NO₃⁻-N was assumed to be the main N process occurring in the soil. HYDRUS-2D incorporates this process by means of a sequential first-order decay chain.

A time-variable flux boundary condition was applied to a 20 cm long boundary directly below the dripper, centred on 60 cm from the top left corner of the soil domain. During irrigation, the drip line boundary was held at a constant water flux, q . The atmospheric boundary condition was assumed for the remainder of the soil surface during periods of irrigation, and for the entire soil surface during periods between irrigation. A no-flow boundary condition was

established at the left and right edges of the soil profile, to account for flow and transport symmetry. A free drainage boundary condition was assumed at the bottom of the soil profile. Initial conditions for water, salinity and nitrate simulations were based on measured data which are described in details in Phogat et al. [3].

HYDRUS-2D requires daily values of potential evaporation (E_s) and transpiration (T_p), which were obtained using the dual crop coefficient approach and local meteorological data [42,43].

Statistical indicators

Error estimates: The model's performance was evaluated by comparing measured (M) and HYDRUS-2D simulated (S) values of water content, electrical conductivity of the soil solution (EC_{sw}), and nitrate concentration (NO_3^- -N) in the soil, and calculating a range of error estimates, tests of significance, regression analyses, and dimensionless efficiency tests. The error estimates included mean error (ME), mean absolute error (MAE), and root mean square error ($RMSE$), given by:

$$ME = \frac{1}{N} \sum_{i=1}^N (M_i - S_i) \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |M_i - S_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - S_i)^2} \quad (3)$$

The test of significance was conducted using the paired t -test (t_{cal}) and given as:

$$t_{cal} = \frac{d}{S_m \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; d = \bar{x}_1 - \bar{x}_2 \text{ and } S_m = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \quad (4)$$

Here, n and s are the number of comparable paired points and standard deviation respectively; subscripts 1 and 2 are indicative respectively of measured and predicted values; S_m is the standard deviation of the mean, and t_{cal} is the calculated paired t -test value.

Efficiency criteria: The coefficient of determination (R^2) was applied for testing the proportion of variance in the measured data explained by the model, and is defined as the square of the coefficient of correlation (r) according to Bravais-Pearson, calculated as:

$$R^2 = \left(\frac{\sum_{i=1}^N (M_i - \bar{M})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}} \right)^2 \quad (5)$$

Values of R^2 can vary between 0 and 1, with higher values indicating less variance, and values greater than 0.5 typically considered acceptable [24].

Efficiency measures for the evaluation of model performance investigated in this study were: model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), and relative index of agreement (IA_{rel}).

Model efficiency (E), as proposed by Nash and Sutcliffe [44], is defined as one minus the sum of absolute squared differences between

simulated and measured values, normalized by the variance of measured values during the period under investigation:

$$E = 1 - \frac{\sum_{i=1}^N (M_i - S_i)^2}{\sum_{i=1}^N (M_i - \bar{M})^2} \quad (6)$$

The range of E lies between $-\infty$ and 1.0 (perfect fit). An efficiency value between 0 and 1 is generally viewed as an acceptable level of performance. Efficiency lower than zero indicates that the mean value of the observed time series would be a better predictor than the model, and denotes unacceptable performance [45].

The index of agreement (IA) was proposed by Willmot (1981), and represents the ratio of the mean square error to the potential error:

$$IA = 1 - \frac{\sum_{i=1}^N (M_i - S_i)^2}{\sum_{i=1}^N (|S_i - \bar{M}| + |M_i - \bar{M}|)^2} \quad (7)$$

The value of IA varies between 0 and 1. A value of 1 indicates a perfect agreement between measured and simulated values, and 0 signifies no agreement at all.

Relative efficiency criteria: Various criteria described above (R^2 , E , and IA) quantify the difference between observations and predictions in absolute values. As a result, an over- or under-prediction of larger values has, in general, a greater influence than that of smaller values. To counteract this, efficiency measures based on relative deviations can be derived from E and IA as:

$$E_{rel} = 1 - \frac{\sum_{i=1}^N \left(\frac{M_i - S_i}{M_i} \right)^2}{\sum_{i=1}^N \left(\frac{M_i - \bar{M}}{\bar{M}} \right)^2} \quad (8)$$

$$IA_{rel} = 1 - \frac{\sum_{i=1}^N \left(\frac{M_i - S_i}{\bar{M}} \right)^2}{\sum_{i=1}^N \left(\frac{|S_i - \bar{M}| + |M_i - \bar{M}|}{\bar{M}} \right)^2} \quad (9)$$

where, E_{rel} and IA_{rel} represent the relative efficiency and a relative index of agreement, respectively. These parameters can also range between the values described for E and IA , respectively.

Modified form of E and IA: The modified form of E and IA are extensively used to overcome the problem of squared differences and oversensitivity to extreme values induced by the mean squared error in E and IA as given below:

$$E_j = 1 - \frac{\sum_{i=1}^N (|M_i - S_i|)^j}{\sum_{i=1}^N (|M_i - \bar{M}|)^j}, j \in N \quad (10)$$

$$IA_j = 1 - \frac{\sum_{i=1}^N (|Mi - Si|)^j}{\sum_{i=1}^N (|Si - \bar{M}|) + (|Mi - \bar{M}|)^j}, j \in N \quad (11)$$

where j represents an arbitrary power i.e. a positive integer (N). Especially when $j=1$, the errors and differences are given their appropriate weighting, not inflated by their squared values. Hence E_1 and IA_1 represent modified form of efficiency and index of agreement. Squaring in statistics (E_2 and IA_2) is useful because squares are easier to manipulate mathematically than are absolute values, but use of squares forces an arbitrarily greater influence on the statistic by way of the larger values [23]. These parameters can also range between the values described for E and IA , respectively.

The error measures (ME , MAE and $RMSE$) and t -test were computed temporally (across all measurement/simulation depths on weekly basis), spatially (across all weekly measurement/simulation for each depth), and across all individual measurements/simulations for the entire dataset. However, regression and efficiency (R^2 , E , IA , E_{rel} , IA_{rel} , E_1 and IA_1) measures were only evaluated on the entire dataset.

Results and Discussion

Data sets for error parameter comparison

The total data set of weekly measured (M) and HYDRUS-2D simulated (S) water content, soil solution salinity (EC_{sw}), and nitrate-nitrogen concentration (NO_3^-N) at different depths and their graphical comparisons are described in Phogat et al. [3]. These data sets provide an ideal basis for comparing the range of error parameters, given that the three data sets (water content, salinity and nitrate concentration) represent a range from good matching between simulated and measured data (water content) to relatively poorly matched (nitrate concentration). This allows comparison of the various error parameters across a range of error values and efficiency testing.

Comparison of error indices

Numerous error estimation methods are in use for comparing simulation results with measured data. These indices are valuable tools because they evaluate the error in the units of the constituent of interest, which helps in the analysis of results and describe the performance and utility of the modelling exercise. Weekly computed temporal error indices (ME , MAE , and $RMSE$) on measured and simulated water content, soil solution salinity (EC_{sw}) and nitrate-nitrogen content (NO_3^-N) are depicted in box plots in Figure 1 and range of these parameters are shown in Table 1. Spatial values of error indices are shown in Figure 2. Seasonal values of all statistical measures are displayed in Table 2.

Mean error (ME) is the signed measure of deviations between measured and simulated values, indicating whether the deviations tend to be positive or negative. The ME in temporal data ranged from -0.04 to 0.04 $cm^3.cm^{-3}$, -0.42 to 0.54 $dS.m^{-1}$, and -11.31 to 12.38 $mg.L^{-1}$ for water contents, EC_{sw} , and NO_3^-N concentrations, respectively in Table 1 and Figure 1. Similarly, spatial ME s for water contents, EC_{sw} , and NO_3^-N concentrations ranged from -0.01 to 0.04 $cm^3.cm^{-3}$, -0.24 to 0.34 $dS.m^{-1}$, and -21.09 to 4.21 $mg.L^{-1}$, respectively in Table 1 and Figure 2). Seasonal ME values for water content, EC_{sw} , and NO_3^-N for the entire data set were 0.003 $cm^3.cm^{-3}$, 0.12 $dS.m^{-1}$, and -6.7 $mg.L^{-1}$ (Table 2). This comparison revealed that ME was smallest for water content and greatest for NO_3^-N , with EC_{sw} in between. Although NO_3^-N errors indicated the widest variability, a low value of ME may still conceal simulation inaccuracy due to the offsetting effect of large positive and negative errors, hence the need to also consider MAE and $RMSE$.

Comparing MAE and $RMSE$ in Figure 2 reveals that the magnitude of the inter-quartile range (IQR) was higher in $RMSE$ as compared to MAE . IQR is the difference between the 25th and 75th percentile, and indicates the magnitude of variation in the mid-range of error values. The magnitude of IQR in $RMSE$ was 0.01 $cm^3.cm^{-3}$, 0.36 $dS.m^{-1}$, and 19.05 $mg.L^{-1}$, as compared to 0.01 $cm^3.cm^{-3}$, 0.33 $dS.m^{-1}$, and 14.98 $mg.L^{-1}$ in MAE , for water content, EC_{sw} , and NO_3^-N , respectively. Similarly, the

Soil Parameter	Scale	Range	ME	MAE	RMSE
Water content ($cm^3.cm^{-3}$)	Temporal	Max	0.04	0.04	0.05
		Min	-0.04	0.005	0.01
	Spatial	Max	0.04	0.04	0.05
		Min	-0.01	0.02	0.02
Salinity (EC_{sw}) ($dS.m^{-1}$)	Temporal	Max	0.54	0.76	0.93
		Min	-0.42	0.08	0.09
	Spatial	Max	0.34	0.47	0.56
		Min	-0.24	0.19	0.25
Nitrate (NO_3^-N) ($mg.L^{-1}$)	Temporal	Max	12.31	27.65	39.57
		Min	-11.38	1.44	2.00
	Spatial	Max	4.21	21.24	30.75
		Min	-21.09	5.06	5.89

Table 1: Maximum and minimum values of ME , MAE , and $RMSE$ obtained from spatial and temporal data sets of water content, EC_{sw} , and NO_3^-N

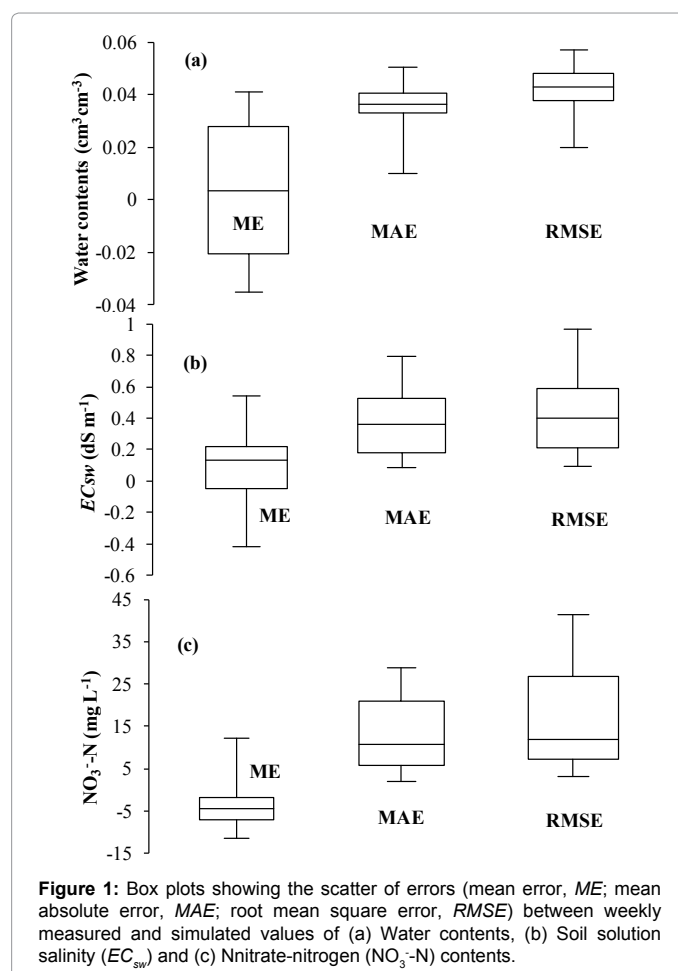


Figure 1: Box plots showing the scatter of errors (mean error, ME ; mean absolute error, MAE ; root mean square error, $RMSE$) between weekly measured and simulated values of (a) Water contents, (b) Soil solution salinity (EC_{sw}) and (c) Nnitrate-nitrogen (NO_3^-N) contents.

Statistical Measures	Moisture content (cm ³ cm ⁻³)	Salinity (EC _{sw}) (dSm ⁻¹)	Nitrate (NO ₃ -N) (mg L ⁻¹)
N [#]	240	188	192
ME	0.003	0.12	-6.70
MAE	0.03	0.34	24.49
RMSE	0.03	0.46	26.76
hSD _m	0.023	0.28	7.35
t _{cal}	0.84	1.79	3.42
R ²	0.50	0.59	0.56
E	0.43	0.30	0.12
E _{rel}	0.36	0.49	-319.25
E _i (j=1)	0.26	0.32	0.28
IA	0.84	0.85	0.80
IA _{rel}	0.82	0.80	-71.3
IA _i (j=1)	0.61	0.69	0.68

[#]Number of measured and simulated data pairs; *Significant at 5% level

Table 2: Values of ME, MAE, and RMSE, half of the standard deviation of measured data (hSD_m), and a paired t-test (t_{cal}), coefficient of determination (R²), model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), relative index of agreement (IA_{rel}), modified E (E_i), and modified IA (IA_i) estimated for the entire (seasonal) dataset

spatial RMSE values were also higher than the MAE values at all depths shown in Figure 2, but their magnitude was much wider at shallow depths (10-25 cm) where RMSE values were 0.01 cm³.cm⁻³, 0.2 dS.m⁻¹, and 9.49 mg.L⁻¹ higher than MAE values for water content, EC_{sw}, and NO₃-N, respectively. However, higher variations in all types of errors at the surface depth (10-25 cm) reflect the assumption in the model of a constant atmospheric boundary flux during daily time steps, which deviates from actual conditions at the surface boundary, particularly the diurnal fluctuation in evaporation, which peaks in day time and decreases during the night [2].

Comparison of MAE and RMSE further indicated that as the magnitude of variations between measured and predicted values increased, RMSE increased disproportionately, as is evident from the NO₃-N values. Similar trends for NO₃-N were also obtained in the MAE and RMSE analysis of the whole data set, where the values of these parameters were 24.49 mg.L⁻¹ and 26.76 mg.L⁻¹, respectively in Table 2. RMSE was always larger than MAE, and varied with the variability of the error magnitude, because the errors are squared in RMSE before they are averaged. RMSE varies with the variability within the distribution of error magnitudes and with the square root of the number of errors (n^{1/2}), as well as with the average-error magnitude (as MAE) [26]. Hence, RMSE gives a relatively high weight to large errors, as obtained in the case of NO₃-N. On the other hand, MAE is a linear measure, which means that all individual differences are weighted equally in the average. Hence, MAE may be preferred over RMSE as a more natural measure of the average error, and for an unambiguous assessment of model predictions. Legates and McCabe expressed similar views as RMSE produces inflated values when large outliers are present [23].

The use of these two measures (RMSE and MAE) suffers from a significant drawback, in that they do not indicate the direction of the error. However, this discrepancy may be ignored where the main focus of the comparison is the magnitude of the error rather than its direction.

There is no universally accepted threshold limit for error magnitude when judging the degree of accuracy of model performance. However, Singh et al. stated that RMSE and MAE values smaller than half of the standard deviation of the measured data (hSD_m) may be considered

low and appropriate for model evaluation [46]. Therefore, RMSE, MAE, and hSD_m values obtained using temporal water content data are compared in Figure 3. It can be seen that these errors (RMSE and MAE) were higher than hSD_m except on a few occasions during mid season (DOY 36 to 64) and during the terminal period (DOY 194 onward). Similarly, MAE and RMSE values were higher than hSD_m in all analyses of the complete data set shown in Table 2. Hence, model performance was relatively poor in view of this criterion.

In the context of such definitive measures of model performance, it is important to consider the natural variability inherent in the measured data against which the simulations are judged [47]. Real world variability of the natural environment, such as soil variations, as well as measurement inaccuracies can cause measured data to vary relative to the best simulation. For example, in our study EnviroSCAN sensors were used to measure the profile water content. The probes were properly calibrated during installation; however, measurements with capacitance probes are highly variable and sensitive to bulk electrical conductivity, temperature, and change in storage estimates [48]. The capacitance sensors used in access tubes may generate consistent errors ≤ 0.05 cm³ cm⁻³, which is similar to the variation observed in our study between EnvironSCAN measured and simulated values.

The error associated with simulated data and outliers in observed data can be further minimised by optimizing the input data by complex weighing techniques and probability distribution based uncertainty analysis techniques like Monte Carlo simulation and Bayesian analysis framework which deals with both random and systemic errors in the simulations [30,49].

Test of significance

The paired t-test was used to evaluate the level of significance between measured and simulated data on water content, EC_{sw}, and NO₃-N content which is shown in Figure 4. It showed non-significant differences (p = 0.05) between mean values of temporal measured and simulated water content. However, positive t_{cal} values during the early period showed that the measured values were higher than the corresponding simulated values, and the opposite was true later in the season shown in Figure 4a. Similarly, insignificant differences were observed for soil salinity, except at DOY 36 and 43, where differences between measured and simulated mean EC_{sw} were significant. However, significant differences were observed in NO₃-N content from DOY 78 to 134, which corresponded to a period from March 2007 to early May 2007.

However, t-test showed significant differences in the spatial data set at depths of 10, 25, 80, and 110 cm for water content, 25, 100, and 150 cm for NO₃-N content and 100 and 150 cm for EC_{sw} (data not shown here). These revelations conform to visual observation of the dataset [3]. Additionally, the t_{cal} values for water content and EC_{sw} for the whole season (Table 1) were non-significant at p = 0.05, whereas the t_{cal} value for NO₃-N showed significant difference, indicating a relatively poor performance of the model for nitrate simulation. However, t-test represents variation between the mean values of measured and simulated data, and therefore a single seasonal figure may not reflect the degree of spatial and temporal heterogeneity across the season. Moreover, t-test assumes that the measured and simulated values are normally distributed, and that both groups have equal variance. These assumptions may not be perfectly satisfied, and this calls into question the reliability of this statistical measure.

It is also important to understand that hypothesis driven tests, such as paired t-test, should not be relied on solely to measure reliability of

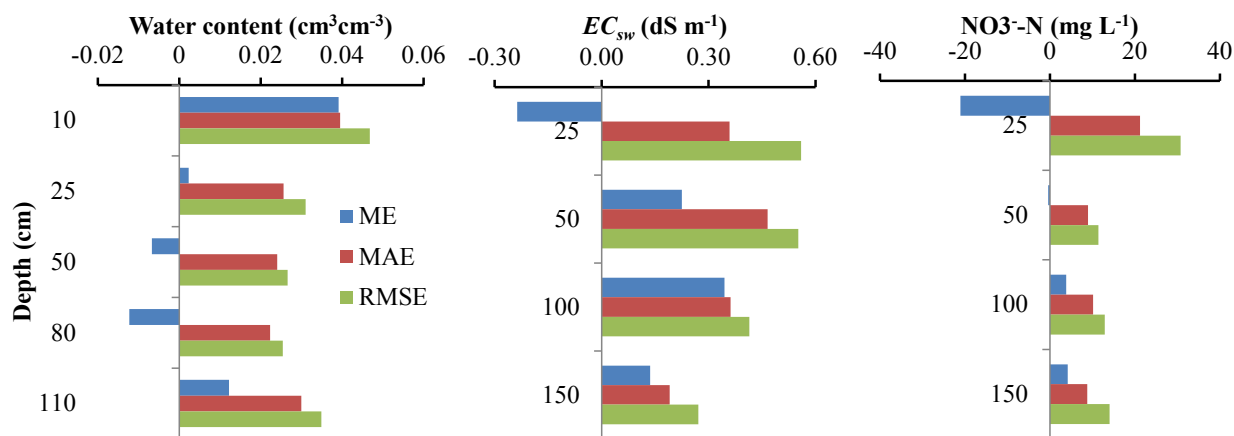


Figure 2: Various errors (ME, MAE, and RMSE) for water contents, soil solution salinities (EC_{sw}) and nitrate-nitrogen (NO_3^- -N) contents at different soil depths

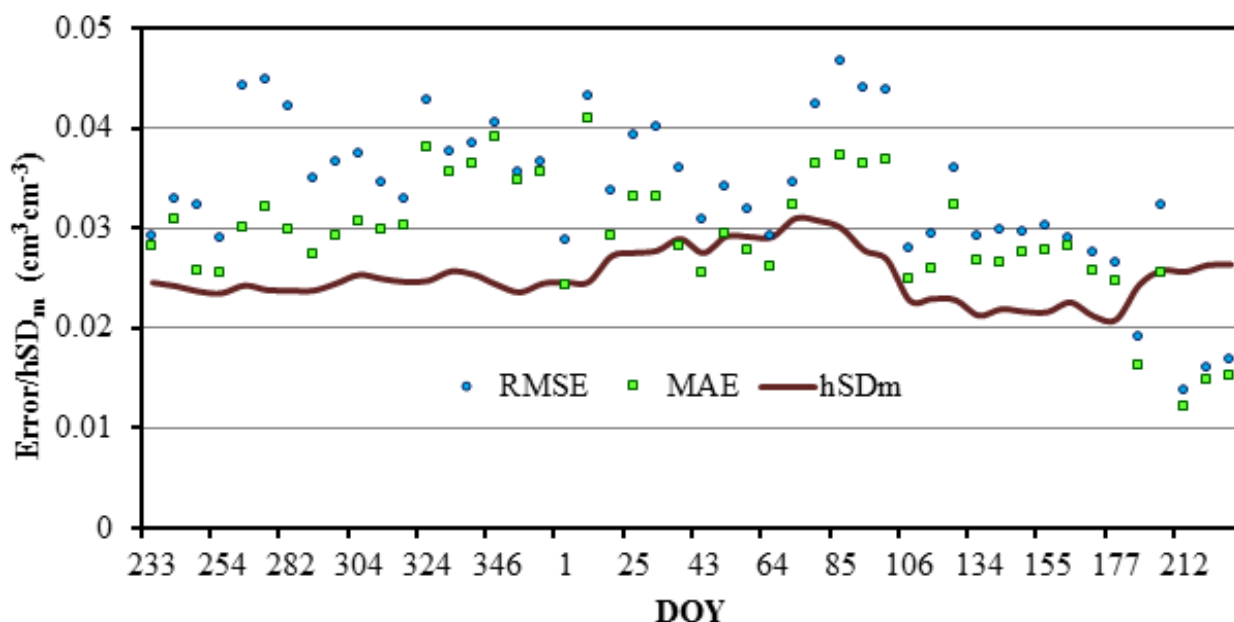


Figure 3: Comparison of error values (RMSE, MAE) and half of standard deviation of measured values (hSD_m) for weakly measured water contents (cm^3cm^{-3})

simulations, as the degree of random variation determines the detection of a significant difference; significant systematic bias will be less likely to be detected if it is accompanied by large random errors [50,51].

Regression analysis and efficiency testing

Model performance was also assessed using regression analysis [coefficient of determination (R^2)] and various efficiency testing indices [model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), relative index of agreement (IA_{rel}), modified form of E (E_I), and modified form of IA (IA_I)]. These measures were computed using the complete data set of measured and corresponding simulated values of water content (240 values), soil solution salinity (EC_{sw} ; 188 values), and nitrate nitrogen content (NO_3^- -N; 192 values) at all spatial and temporal positions (Table 2).

The R^2 value of 0.5 for water content was just at the margin of the satisfactory level [24]. However, its values for EC_{sw} (0.59) and the NO_3^-

-N content (0.56) were within the acceptable limit. Hence, R^2 produced relatively similar results across all simulated processes (water content, EC_{sw} , and NO_3^- -N) (Table 2). This is contrary to our results using error tests, where the variability in water content was comparatively smaller than in NO_3^- -N content (Table 2). This reveals a serious drawback in considering R^2 values alone for model performance evaluation, in that it only quantifies dispersion among values. Krause et al. reported that a model which systematically over- or under-predicts at all times will result in R^2 values close to 1.0, even if all predictions are wrong, which undermines the reliability of R^2 values [25]. Similarly, Legates and McCabe suggested that correlation based measures are inappropriate and should not be used to evaluate the goodness-of-fit of model simulations, as these measures are oversensitive to extreme values and are insensitive to additive and proportional differences between model predictions and observations [23]. Hence, consideration of R^2 alone for model performance assessment sometimes leads to a flawed acceptance of modelling results.

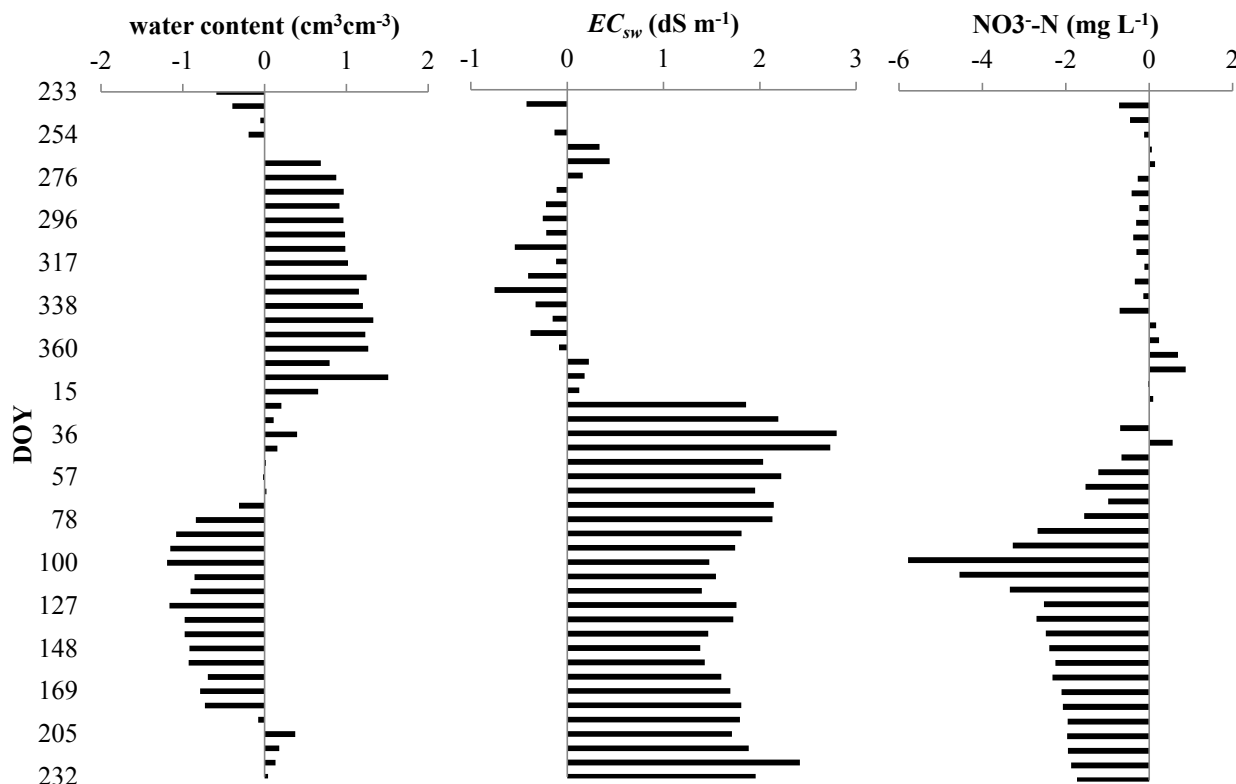


Figure 4: Temporal paired t -test (t_{cal}) values between measured and simulated a) Water content, b) Soil solution salinity (EC_{sw}) and c) Nitrate-nitrogen ($\text{NO}_3\text{-N}$) content

Six efficiency tests were applied to the data set to evaluate their relative performance. The E , E_{rel} , IA_{rel} , E_I and IA_I values for water content were 0.43, 0.84, 0.82, 0.20 and 0.59 respectively (Table 2), suggesting a good match between measured and simulated data, as indicated by previous parameters. On the contrary, IA (0.36) showed relatively poor efficiency of the model for simulating water content distribution. However, values of modified efficiency (E_I) and index of agreement (IA_I) were lower than relative estimates because these statistics utilize absolute values rather than squared differences in their computation which makes them more conservative measures [23]. The E value (0.43) was within the satisfactory limit reported in other studies [24,45]. Similarly, the E , IA , E_{rel} , IA_{rel} , E_I and IA_I values for weekly soil solution salinity (EC_{sw}) data were 0.30, 0.49, 0.85, 0.80, 0.32 and 0.69, respectively, which are well within acceptable limits, and match previous indicators relatively well. Relative efficiency (E_{rel}) for water and relative index of agreement (IA_{rel}) for water and salinity were relatively close to their precursors, E and IA , respectively (Table 2).

Conversely, the nitrate simulation provides a more complicated picture. The Nash and Sutcliffe efficiency ($E = 0.12$) and index of agreement ($IA = 0.80$) values are within acceptable limits, and in fact IA is quite high, in contrast to previous error parameters for nitrate concentration. The modified estimates (E_I and IA_I) fall within acceptable range. However, large negative values of E_{rel} (-319.25) and IA_{rel} (-71.3) reflect the wide divergence between measured and simulated values at certain times during the simulation. However, relative deviations reduce the influence of absolute differences among the measures and simulated values.

It is significant that Nash-Sutcliffe efficiency (E), the most frequently used indicator in hydrologic studies, is much more sensitive

to errors in higher values, as the differences between measured and simulated values are squared. As a result larger values in a data series have a much higher weighting, whereas lower values are neglected [23]. This comparison suggests that E and/or IA may not always be suitable parameters for describing model performance. Additionally, the large negative values of E_{rel} and IA_{rel} for nitrate showed disproportionately high under-prediction, as reported in Krause et al. [25]. Hence, these parameters proved to be sensitive only to large variations in values and not at all to small divergences because, due to the summation of the absolute or squared errors in efficiency testing methods, emphasis is placed on larger errors while smaller errors tend to be neglected. Hence modified efficiency (E_I) and index of agreement (IA_I) could be the more appropriate measures for model's efficiency testing than their precursors (E and IA) and relative statistics (E_{rel} and IA_{rel}). Nevertheless IA_I has advantages due to its bounds between 0.0 and 1.0 [23]. But good modelling efficiency shown by E_I and IA_I statistics is contradictory to the poor modelling simulation for $\text{NO}_3\text{-N}$ revealed in error estimates. Additionally, tests of significance and efficiency measures, similar to t -test, evaluate the mean variability in the domain and are unable to capture the modelling divergence at a particular point.

Overall, it can be stated that none of the efficiency parameters which were evaluated in this study adequately described and tested the reliability of model predictions. Each has specific pros and cons, which have to be taken into account during model calibration and validation. Hence, for a sound model performance evaluation, a combination of different statistical efficiency criteria, complemented by the assessment of the absolute or relative error, may need to be included.

Conclusion

Simulation models have been increasingly used in high efficiency drip irrigation systems to evaluate the water and solute dynamics under cropped conditions, and suggest necessary management options to optimise the system efficiency. In this study, eleven statistical measures were used to compare HYDRUS-2D simulated values of water content, soil solution salinity (EC_{sw}), and nitrate-nitrogen (NO_3^- -N) dynamics with field measured values obtained under drip irrigated mandarin crop over a season. as the statistical parameters compared were mean error (ME), mean absolute error (MAE), root mean square error ($RMSE$), paired t -test (t_{cat}), coefficient of determination (R^2), model efficiency (E), index of agreement (IA), relative model efficiency (E_{rel}), relative index of agreement (IA_{rel}), modified E (E_1), and modified IA (IA_1). The purpose of applying all of these parameters to the same data sets was to evaluate the relative importance of these parameters in model performance testing.

The error parameters (ME , MAE and $RMSE$) remained within acceptable limits when applied to measured and simulated values of water content and EC_{sw} , whilst a wider range of values of MAE (1.44 to 27.65 $mg \cdot l^{-1}$) and $RMSE$ (2.00 to 39.57 $mg \cdot l^{-1}$) obtained for nitrate (NO_3^- -N) indicated poor agreement between simulated and measured values for this data set. Low ME values may conceal simulation inaccuracy due to the offsetting effect of large positive and negative errors. The results revealed that $RMSE$ values were consistently higher than MAE due to squaring of the difference between measured and simulated values. Hence, it was concluded that among the error tests, MAE may be preferred over ME and $RMSE$ for evaluating goodness-of-fit of the simulated values.

Paired t -test values revealed a non-significant difference ($p = 0.05$) between weekly measured and predicted water content and EC_{sw} distributions. However, differences were significant for the NO_3^- -N distribution during the mid-season and at several spatial depths. Similarly, regression analysis (R^2) and efficiency testing methods (E , IA , E_{rel} , IA_{rel} , E_1 and IA_1) also indicated that the model accurately predicted seasonal changes in water and salinity distributions in the soil. However, negative values of E_{rel} (-319.25) and IA_{rel} (-71.3) for NO_3^- -N reflected the relatively poor prediction of NO_3^- -N dynamics in the soil. However, relative deviations reduce the influence of absolute differences among the measures and simulated values.

It was concluded that, for reliable model performance evaluation, a combination of different statistical efficiency criteria, along with the assessment of the absolute or relative volume error, must be included. Taken together, these comparisons were able to provide an objective assessment of the closeness of the simulated behaviour to the observed measurements of water, salinity and nitrate distribution in the soil under mandarin. It is expected that such studies would help in improving the performance evaluation and reliability of modelling data on irrigation and fertigation programme of horticultural crops, and contribute to improving system efficiency, and reducing environmentally harmful agro-hydrological practices.

Acknowledgement

Dareton Agricultural Research and Advisory Station farm staff for managing the site and plantings; and financial support of the National Program for Sustainable Irrigation is gratefully acknowledged.

References

- Hanson BR, Šimunek J, Hopmans JW (2006) Evaluation of urea-ammonium-nitrate fertigation with drip irrigation using numerical modeling. *Agric Water Manage* 86: 102-113.

- Ramos TB, Šimunek J, Goncalves MC, Martins JC, Prazeres A, et al. (2012) Two-dimensional modeling of water and nitrogen fate from sweet sorghum irrigated with fresh and blended saline waters. *Agric Water Manage* 111: 87-104.
- Phogat V, Skewes MA, Cox JW, Sanderson G, Alam J et al. (2014) Seasonal simulation of water, salinity and nitrate dynamics under drip irrigated mandarin (*Citrus reticulata*) and assessing management options for drainage and nitrate leaching. *J Hydrol* 513: 504-516.
- Van Dam JC, Huygen J, Wesseling JG, Feddes RA, Kabat P, et al. (1997) Theory of SWAP version 2.0. Simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment. WAGENINGEN UR.
- Zyvoloski GA, Robinson BA, Dash ZV, Trease LL (1997) User's Manual for the FEHM application- A finite-element heat- and mass-transfer code. Geothermal.
- Yeh GT, Salvage KM, Gwo JP, Zachara JM, Szecody JE (1998) HydroBioGeoChem: A coupled model of hydrologic transport and mixed biogeochemical kinetic/equilibrium reactions in saturated-unsaturated media. UNT digital library.
- Ahuja LR, Hanson JD, Rojas KW, Shaffer MJ (1999) The Root zone water quality model. Water Resources Publications LLC.
- Pruess K, Oldenburg C, Moridis G (1999) TOUGH2 User's Guide, Version 2.0.
- Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, et al. (2003) An overview of APSIM, a model designed for farming systems simulation. *Eur J Agron* 18: 267-288.
- Therrien R, McLaren RG, Sudicky EA, Panday SM (2010) HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport.
- Šimunek J, van Genuchten MTh, Šejna M (2008) Development and applications of the HYDRUS and STANMOD software packages and related codes. *Vadose Zone J* 7: 587-600.
- Cote CM, Bristow KL, Charlesworth PB, Cook FJ, Thorburn PJ (2003) Analysis of soil wetting and solute transport in subsurface trickle irrigation. *Irrig Sci* 22: 143-156.
- Lazarovitch N, Šimunek J, Shani U (2005) System-dependent boundary condition for water flow from subsurface source. *Soil Sci Soc Am J* 69: 46-50.
- Assouline S, Moller M, Cohen S, Ben-Hur M, Grava A, et al. (2006) Soil-plant system response to pulsed drip irrigation and salinity: Bell pepper-Case study. *Soil Sci Soc Am J* 70: 1556-1568.
- Ajdary K, Singh DK, Singh AK, Khanna M (2007) Modelling of nitrogen leaching from experimental onion field under drip fertigation. *Agric Water Manage* 89: 15-28.
- Patel N, Rajput TBS (2008) Dynamics and modeling of soil water under subsurface drip irrigated onion. *Agric Water Manage* 95: 1335-1349.
- Šimunek J, Hopmans JW (2009) Modeling compensated root water and nutrient uptake. *Ecol Model* 220: 505-521.
- Doltra J, Munoz P (2010) Simulation of nitrogen leaching from a fertigated crop rotation in a Mediterranean climate using the EU-Rotate-N and Hydrus-2D models. *Agric Water Manage* 97: 277-285.
- Li J, Liu Y (2011) Water and nitrate distributions as affected by layered-textural soil and buried dripline depth under subsurface drip fertigation. *Irrig Sci* 29: 469-478.
- Phogat V, Mahadevan M, Skewes M, Cox JW (2012) Modeling soil water and salt dynamics under pulsed and continuous surface drip irrigation of almond and implications of system design. *Irrig Sci* 30 : 315-333.
- Phogat V, Skewes MA, Mahadevan M, Cox JW (2013a) Evaluation of soil plant system response to pulsed drip irrigation of an almond tree under sustained stress conditions. *Agric Water Manage* 118: 1-11.
- Phogat V, Skewes M, Cox JW, Alam J, Grigson G, et al. (2013b) Evaluation of water movement and nitrate dynamics in a lysimeter planted with an orange tree. *Agric Water Manage* 127: 74-84.
- Legates DR, McCabe GJ (1999) Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35: 233-241.
- Van Liew MW, Arnold JG, Garbrech JD (2003) Hydrologic simulation on agricultural watersheds: Choosing between two models. *Trans ASAE* 46: 1539-1551.

25. Krause P, Boyle DP, Base F (2005) Comparison of different efficiency criteria for hydrological model assessment. *Adv Geosci* 5: 89-97.
26. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res* 30: 79-82.
27. Schaefli B, Gupta H (2007) Do Nash values have value? *Hydrol Processes* 21: 2075-2080.
28. Jain SK, Sudheer KP (2008) Fitting of hydrologic models: A close look at the Nash-Sutcliffe index. *J Hydrol Eng* 13: 981-986.
29. Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J Hydrol* 377: 80-91.
30. Pechlivanidis IG, Jackson B, McIntyre N, Wheeler HS (2011) Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST J* 13: 193-214.
31. Hassan G, Persaud N, Reneau RB (2005) Utility of HYDRUS-2D in modeling profile soil moisture and salinity dynamics under saline water irrigation of soybean. *Soil Sci* 170: 28-37.
32. Goncalves MC, Šimunek J, Ramos TB, Martins JC, Neves MJ, et al. (2006) Multicomponent solute transport in soil lysimeters irrigated with waters of different quality. *Water Resour Res*.
33. Crevoisier D, Popova Z, Mailhol JC, Ruelle P (2008) Assessment and simulation of water and nitrogen transfer under furrow irrigation. *Agric Water Manage* 95: 354-366.
34. Roberts T, Lazarovitch N, Warrick AW, Thompson TL (2009) Modeling salt accumulation with subsurface drip irrigation using HYDRUS-2D. *Soil Sci Soc Am J* 73: 233-240.
35. Biswas T, Schrale G (2007) Sentek SoluSAMPLER: A tool for managing salt and nutrient movement in the root zone, Instruction Manual Version 2.0. Sentek Pvt Ltd, Australia.
36. Skaggs TH, Trout TJ, Šimunek J, Shouse PJ (2004) Comparison of Hydrus2D simulations of drip irrigation with experimental observations. *J Irrig Drainage Eng* 130: 304-310.
37. Falivene S, Goodwin I, Williams D, Boland A (2005) Introduction to Open Hydroponics. NPSI Fact sheet.
38. Van Genuchten MTh (1980) A closed form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44: 892-898.
39. Vrugt JA, Hopmans JW, Šimunek J (2001) Calibration of a two dimensional root water uptake model. *Soil Sci Soc Am J* 65: 1027-1037.
40. Feddes RAP, Kowalik J, Zaradny H (1978) Simulation of field water use and crop yield.
41. Taylor SA, Ashcroft GM (1972) Physical Edaphology. Freeman and Co, USA.
42. Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper, Italy.
43. Allen RG, Pereira LS (2009) Estimating crop coefficients from fraction of ground cover and height. *Irrig Sci* 28: 17-34.
44. Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I-A discussion of principles. *J Hydrol* 10: 282-290.
45. Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD et al. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50: 885-900.
46. Singh J, Knapp HV, Demissie M (2004) Hydrologic modelling of the Iroquois river watershed using HSPF and SWAT.
47. Harmel RD, Cooper RJ, Slade RM, Haney RL, Arnold JG (2006) Cumulative uncertainty in measured stream flow and water quality data for small watersheds. *Trans ASAE* 49: 689-701.
48. Evett SR, Schwartz RC, Casanova JJ, Heng LK (2012) Soil water sensing for water balance, ET and WUE. *Agric Water Manage* 104: 1-9.
49. Finsterle S, Najita J (1998) Robust estimation of hydrogeologic model parameters. *Water Resour Res* 34: 2939-2947.
50. Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London.
51. Bland JM, Altman DG (1995) Comparing two methods of clinical measurement: a personal history. *Internat J Epidemiol* 24: 7-14.

Citation: Phogat V, Skewes MA, Cox JW, Simunek J (2016) Statistical Assessment of a Numerical Model Simulating Agro Hydro-chemical Processes in Soil under Drip Fertigated Mandarin Tree. *Irrigat Drainage Sys Eng* 5: 155. doi:[10.4172/2168-9768.1000155](https://doi.org/10.4172/2168-9768.1000155)