CSIRO PUBLISHING
www.publish.csiro.au/journals/sr *Soil Research*, 2011, **49**, 77–86

Temporal variability of soil water storage evaluated for a coffee field

L. C. Timm[A,F], *D. Dourado-Neto*[B], *O. O. S. Bacchi*[C], *W. Hu*[D], *R. P. Bortolotto*[B], *A. L. Silva*[E], *I. P. Bruno*[B], *and K. Reichardt*[C]

[A]Department of Rural Engineering, FAEM, UFPel, Pelotas, RS C.P. 354-96001-970, Brazil.
[B]Crop Science Department, ESALQ, USP, Piracicaba, SP C.P. 9, 13418-900, Brazil.
[C]Soil Physics Laboratory, CENA, USP, Piracicaba, SP C.P. 96, 13416-903, Brazil.
[D]Key Laboratory of Water Cycle and Related Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
[E]CTC, Fazenda Santo Antônio, Piracicaba, SP 13400-970, Brazil.
[F]Corresponding author. Email: lctimm@ufpel.edu.br

Abstract. Sampling field soils to estimate soil water content and soil water storage (*S*) is difficult due to the spatial variability of these variables, which demands a large number of sampling points. Also, the methodology employed in most cases is invasive and destructive, so that sampling in the same positions at different times is impossible. However, neutron moderation, time domain reflectrometry, and, more recently, frequency domain reflectrometry methodologies allow measurements at the same points over long time intervals. This study evaluates a set of neutron probe data, collected at 15 positions placed randomly along a coffee crop contour line, over 2 years at 14-day intervals. The temporal stability of *S* was again demonstrated, so that wetter or dryer locations remain so over time, and the definition of such positions in the field reduces the number of sampling points in future *S* evaluations under similar conditions. An analysis was made to determine the minimum number of sampling points to obtain the average *S* of the field within a chosen level of significance. Classical statistical analysis indicated that the 15 measurement positions could be reduced to four or even to one position to obtain a reliable field *S* average. State–time analysis showed *S* estimations depend more on previous measurements of rainfall *P* (52%) than on evapotranspiration *ET* (28%) and *S* (20%). The analysis also showed that *ET* was not realistically estimated from previous measurements of *S*; it was more dependent on previous measurements of *ET* (59%) than on *P* (30%) and *S* (9%). This statistical procedure showed great advantages over classical multiple regressions. Future studies of this type should be carried out at regularly spaced observation points in a grid, in order to allow a 2-D and 3-D state–space–time analysis.

Additional keywords: multiple regression, neutron probe, sampling number, state–space, state–time.

Introduction

Soil water storage (*S*) in agricultural soil profiles is an important parameter for a rational management of any crop, besides giving information on environmental aspects of the water cycle. Spatial variability of *S*, however, imposes serious problems when determining average values over large areas, which are needed for management of water availability to crops. The variability of *S* is a consequence of the erratic rainfall input, differences in crop stand, and natural soil matrix differences that can occur over short distances as well as over large fields due to soil genesis and topography. Knowledge of the characteristics of the variability of *S* help us to understand and predict several hydrologic processes (Western *et al*. 2004) and to improve soil water sampling strategies (Warrick and Nielsen 1980).

Variability of soil physical and chemical properties is not a new research topic. Since the first half of last century, the problem of obtaining representative sampling of agricultural fields has led to the development of new sampling schemes. Initially, scientists based their strategies on classical statistical concepts, which were later complemented with geoestatistics and time–space series analyses, and more recently neural networks (Hills and Reynolds 1969; Mohanty and Mousli 2000; Western *et al*. 2002; Timm *et al*. 2006; Hu *et al*. 2008).

The temporal stability of *S* measurements was first indicated by Vachaud *et al*. (1985), who statistically determined the presence of locations that systematically presented soil water contents above or below the field average. Kachanoski and De Jong (1988) and Moreti *et al*. (2007) also used this concept to show the temporal persistence of spatial patterns of *S*. Reichardt *et al*. (1997) suggested that part of the time stability of soil water content measurements is due to systematic errors introduced by soil water content calibration curves when indirect methods of measurement are employed, such as neutron probes, time domain reflectrometry (TDR), and frequency domain reflectrometry (FDR). Hu *et al*. (2008) verified the time stability of soil water content measurements made using FDR at the soil surface layer of a hill-slope of the Loess Plateau in China, and found significant correlations with several factors

 10.1071/SR10023 1838-675X/11/010077

influencing landscape. More recently, Hu *et al*. (2010) presented a new criterion to identify sites for S determinations based on the mean absolute bias error.

Few studies have analysed the time variability of S as affected by evapotranspiration and rainfall. A comprehensive report was presented by Aboitiz *et al*. (1986), who developed a methodology for estimating and forecasting soil water depletion and evapotranspiration in irrigated fields, using a time-varying state–space model, which we here call 'state–time'. In this paper, we aim to contribute to the improvement of water management practices of natural ecosystems and perennial crops such as the coffee crop, analysing a two-year series of S measurements, giving emphasis to the time stability and spatial variability of this set of data. A new perspective and deeper insight is made through a state–time analysis to better understand the temporal relations between S, rainfall, and evapotranspiration.

Materials and methods

We analysed the temporal variability of S (mm) data collected in a coffee crop grown in Piracicaba, SP, Brazil (22°42′30″S, 47°38′00″W; 580 m asl). Soil water contents θ (i) were measured along a horizontal domain x_i (m) at 15 locations ($i = 1, 2, \ldots, 15$), and at five depths z_k (m), 0.2, 0.4, 0.6, 0.8, and 1.0 m from surface ($k = 1, 2, \ldots, 5$), every 14 days, at times t_j ($j = 1, 2, 3, \ldots, 52$) over a 2-year period beginning 1 September 2003. Soil water content measurements obtained with a neutron probe (model CPN 503 DR) were taken at irregular spacings along a levelled contour line of the horizontal domain corresponding to a coffee row, following the distribution of five fertiliser plots arranged within a 0.2-ha coffee field. Details of the fertiliser trial can be found elsewhere (Fenilli *et al*. 2007). Measurements of θ were made using aluminium neutron probe access tubes installed below crop canopies. Coffee (*Coffea arabica* L.) was of the cultivar Catuaí Vermelho (IAC-144) and is a perennial crop, 3–5 years old during the experimental period, which is the beginning of the yearly coffee production cycles. The spacing between plants was 0.75 m and between rows 1.5 m. Rows were kept bare chemically and manually, as commonly done in coffee plantations.

The soil is a Rhodic Kandiudalf (Soil Survey Staff 1993), locally called 'Nitossolo Vermelho Eutroférrico' (EMBRAPA 2006). The climate is of the Cwa type (Köppen 1931), with dry winter.

Slow neutron counting data were transformed into soil water contents using calibration curves established as suggested by Reichardt *et al*. (1997), taken as valid over all depths. Soil water storages at times j and positions i, $S_j(i)$ (mm) for the 0–1.0 m soil layer were calculated from $\theta_{t,x}(k)$ data by the trapezoidal rule:

$$S_j(i) = [1.5\theta_{i,j}(1) + \theta_{i,j}(2) + \theta_{i,j}(3) + \theta_{i,j}(4) + 0.5\theta_{i,j}(5)]\,\frac{1000}{5} \tag{1}$$

with $\Delta z = 0.2$ m. Soil water contents $\theta_{i,j}(1)$ measured at the depth 0.2 m ($k = 1$) were considered to cover a layer of $1.5\Delta z = 0.3$ m which includes soil surface. The first measurement made at the depth of 0.2 m was evaluated to be deep enough not to lose slow neutrons to the atmosphere. $\theta_{i,j}(5)$ measured at 1.0 m ($k = 5$) covered $0.5\Delta z = 0.1$ m since the lower level of the control volume for water balances was set at 1.0 m, and the total depth L was taken as 1000 mm to obtain data in mm. The coffee root system was assumed not to reach depths below $z = 1.0$ m, which was confirmed by Silva *et al*. (2009).

In order to apply the following statistical procedures, $S_j(i)$ data were tested for normality with respect to space by performing cumulative probability plots.

To reduce the number of observation points so that future evaluations of the soil water status of this perennial coffee field could be made more rapidly and without losing accuracy, two approaches were used: (*a*) performing a time stability analysis to determine which access tube can represent the overall average of the field; and (*b*) establishing the minimum number of observation points that would yield an average value within a pre-established coefficient of variation. To verify the time stability of the measurements, the approach proposed by Vachaud *et al*. (1985) was used. For this, the relative deviation $\delta_j(i)$% of each $S_j(i)$ realisation in relation to the mean soil water storage $\overline{S_j}(i)$ was calculated as follows:

$$\delta_j(i) = \frac{S_j(i) - \overline{S_j}(i)}{\overline{S_j}(i)} \times 100 \tag{2}$$

According to Vachaud *et al*. (1985), very small time variations of $\delta_j(i)$ indicate a time stability of $S_j(i)$, so that consistently wetter or dryer positions (i) can be selected in the field. Therefore, if time averages $\overline{\delta_i}(j)$ of the $\delta_j(i)$ values are plotted in rank, it is possible to find out which sites present systematically $S_j(i)$ values below or above the position time average $\overline{S_j}$ and also those sites that systematically present a negligible $\overline{\delta_i}(j)$ and, therefore, represent $\overline{S_j}$.

To estimate the number of observations N needed in a new sampling event to obtain a mean value $S_t(i)$, within a chosen deviation (%) of the estimated mean value, the suggestion of Warrick and Nielsen (1980) was applied:

$$N = t_\alpha^2 s^2 d^{-2} \tag{3}$$

where t_α is the value of the Student t-distribution considering the level of significance α (for $\alpha = 5\%$ the t value is 1.96) for infinite degrees of freedom; s^2 is the variance of a previous sampling event $S_t(i)$ made with n (15 in our case) replicates; and d any desired deviation from the mean, for example [0.5, 1, 2%, … of $\overline{S_t}(i)$]. Equation 3 assumes that the samples are independent, the central limit applies, and that the true mean deviation σ can be represented by the standard deviation s.

In a second step, the time variability structure of the $\overline{S_j}$ data was studied using the state–time approach (Shumway 1988; Nielsen and Wendroth 2003), which provides opportunities for a suitable identification of temporal relations between soil–atmosphere–plant variables taking into account their temporal association. The state–time analysis characterises the state of a system (set of p unobservable variables) at a time t to its state at a time t–j, $j = 1, 2, 3, \ldots, 52$, in our study. For $j = 1$, the state–space approach is described as follows (called state equation):

$$X_t = \phi X_{t-1} + \omega_{X_t} \tag{4}$$

X_t and X_{t-1} being the state vector (a set of p unobservable variables) at time t and $t-1$; ϕ a $p \times p$ matrix of state coefficients, which indicates the measure of the regression; and ω_{X_t} noises of the system for $t = 1, 2, 3, \ldots, j$. Noise values are assumed to have zero mean, not being autocorrelated and being normally distributed with constant variances. If these X variables were observable, this would be the usual structure of a vector autoregressive model, in which the coefficients of the matrix ϕ could be estimated by multiple regression techniques, taking X_t and X_{t-1} as the dependent and independent variables, respectively. In the case of the state–time model, however, the true state of the variables is considered 'embedded' in an observation equation:

$$Y_t = AY_{t-1} + \nu_{Y_t} \tag{5}$$

the observation vector Y_t being related to the state vector X_t by an observation matrix A (usually known as, for instance, an identity matrix, $p \times p$) and an observation noise vector ν_{Y_t}, also considered of zero mean, not autocorrelated and normally distributed. The noises ω_{X_t} and ν_{Y_t} are assumed to be independent of each other. The state coefficients of the matrix ϕ and noise variances of Eqn 4 are estimated through a recursive procedure given by Shumway and Stoffer (1982). According to Hui *et al*. (1998), if the X_t data are scaled with respect to their mean (m) and standard deviation (s), as follows:

$$x_t = [X_t - (m - 2s)]/4s \tag{6}$$

the transformed values x_t become dimensionless with mean $m = 0.5$ and standard deviation $s = 0.25$. This transformation allows state coefficients of the matrix ϕ have magnitudes directly proportional to their contribution to each state variable used in the analysis. The software Applied Statistical Time Series Analysis (ASTSA) (Shumway 1988) was used for applying the state–space approach.

Concomitantly to $S_j(i)$ measurements, Silva *et al*. (2006) evaluated time series of evapotranspiration $ET_j(i)$, rainfall $P_j(i)$, supplementary sprinkler irrigation $I_j(i)$, surface runoff $RO_j(i)$, and soil water drainage fluxes $Q_j(i)$ below the 1.0 m depth, to establish complete water balances, which were used in the state–time and multiple regression analyses. Irrigation was applied only during the dry winter, in just a few events when the available water capacity reached ~25% of its maximum. For the analysis, I was added to P. For a few 14-day intervals with no rainfall during the rainy season, a negligible value of $P = 0.1$ mm was assumed for this variable, so that the state–time analysis could be performed. It is important to mention that classical multiple regression is based on mean values of each variable throughout the time being investigated and that the magnitudes of each variable at a given time compared to their respective values at a previous or future time are neglected.

Coefficients of variation (CV), cumulative probability plots, and rank plots were also used in the analysis (SAS and R statistical programs).

Results and discussion

Soil water storage $S_t(i)$ data were normally distributed for all 52 measurement dates, as exemplified in Fig. 1 through cumulative probability plots for a wet period (31 January 2005) and for a dry period (1 September 2003). These spatial data presented space coefficients of variation for fixed times j in the range 1.1–5.9%, indicating that the variability in space can be considered low.

Ranges of soil water storage changes $\Delta S_t(i)$ shown in Fig. 2, in which positive values represent maximum soil water recharges occurring in 14-day intervals and negative values represent soil water maximum depletions in 14-day intervals, reflect the great time variability of $S_t(i)$ data observed during the 2 years in this field. Such plots illustrate well the spatial variability of S measurements made in agricultural fields, as in this case for a coffee crop field, justifying the search for good and stable averages of S for water management purposes.

For future measurements of $S_t(i)$ in the same or other fields of similar condition, the minimum number N of observation points was calculated for chosen precision levels according to Eqn 3. Selecting three dates for which the $S_t(i)$ value is of the order of 300 mm: $j = 10$, $S_{10}(i) = 302$ mm, $s_{10}(i) = 18$ mm; $j = 20$,

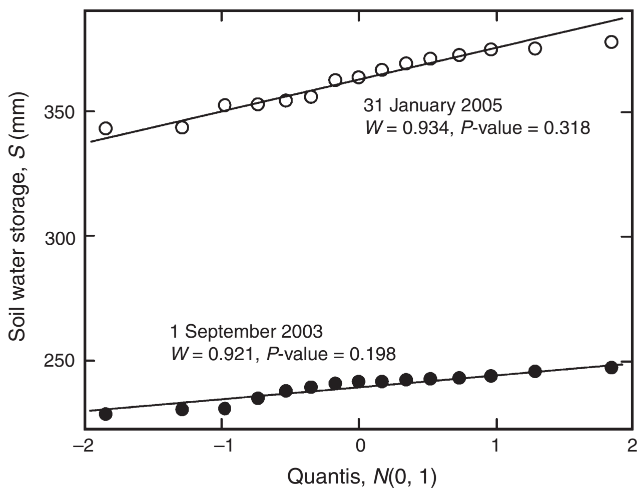

Fig. 1. Cumulative probability plots of soil water storage $S_t(i)$ for two selected dates.

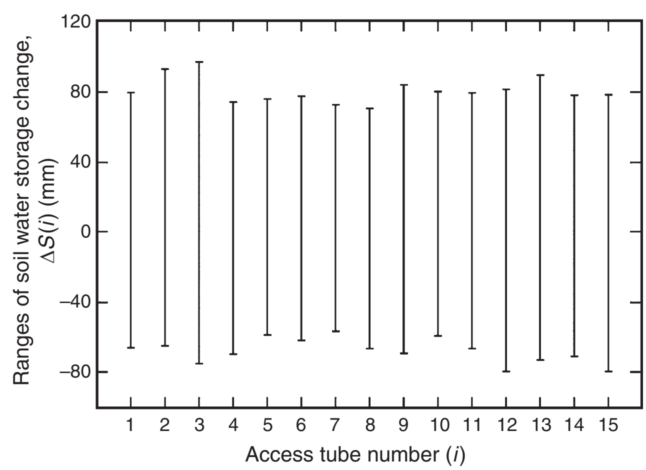

Fig. 2. Ranges of soil water storage changes $\Delta S_t(i)$ observed for the 15 neutron probe access tubes during the 2-year observation period, in a coffee crop field.

$S_{20}(i) = 296$ mm, $s_{20}(i) = 8$ mm; and $j = 30$, $S_{30}(i) = 285$ mm, $s_{30}(i) = 3$ mm; for which $s_t(i)$ were maximum, medium, and minimum, the deviations (d) from the mean are 5.9, 2.7, and 1.1%, respectively. For new samplings according to Eqn 3, if the desired $\overline{S_t}(i)$ of 300 mm should be evaluated within 0.5, 1, or 2% of the correct value, with an average $s_t = 8$ mm, the number of samplings would be 56, 14, and 4, respectively. For this example, the only viable choice to reduce the number of sampling points is to accept a deviation of 2% and make future measurements in four access tubes.

In terms of time stability of the measurements, the rank plot presented in Fig. 3 shows that position 3 best represents the mean $\overline{S_t}(i)$ over the two years of observation, which means that future observations of $\overline{S_t}(i)$ could be performed at this single site or at four sites as discussed above (sites 2, 6, 3, and 10, Fig. 3), with much lower coefficients of variation than 2% used in Eqn 3, since the chosen four points present the least deviation from the mean. Such measurement would represent the mean soil water storage of the whole field, greatly simplifying future experimental field work. This reduction of observation points is very important for long-term experimentation in natural ecosystems or perennial crops such as coffee, when $S_t(i)$ is observed over long periods of time (years), e.g. Silva *et al.* (2006) and Moreti *et al.* (2007). It is important to recall that in the establishment of field water balances, the soil components are the more laborious measurements.

A great shortcoming of the time stability as a criterion to reduce the number of sampling points is the need of representative previous information in space and time, in order to be able to make significant rank plots of mean deviations from the mean. Therefore, the approach presented here is more suitable for long-duration experiments in which costly and time-consuming variables are measured.

As discussed below, the state–time analysis is a step ahead of the previous discussion since it allows a better insight of the relations among the climate variables that determine S. So, in order to better understand the temporal relations between S, P, and ET, a discussion is made comparing the state–time analysis to the classical multiple regression using the same state variables. Figure 4*a* and *b* shows the multiple regression and state–time equations and the value of their coefficients of determination (r^2) from linear regressions between estimated and measured values of scaled (Eqn 6) soil water storage. Classical multiple regression is based on mean values of each variable throughout the time being investigated, in which the magnitudes of each variable at a given time compared with their respective values at a previous or future time are neglected, so that no more 35.8% of the variance of the biweekly measured soil water storage data was explained from the measurements of precipitation and evapotranspiration (Fig. 4*a*). Estimated values by regression are much less variable than those measured, and consistently underestimate the larger, and overestimate the smaller, measured values.

When the temporal associations among soil water storage, precipitation, and evapotranspiration data were considered, 99.8% of the variance of S was explained from the use of the state–time analysis (Fig. 4*b*). We note that nearly 70% of the previous value S_{i-1} contributes to that of S_i while preceding values P_{i-1} and ET_{i-1} contribute only 8 and 20%, respectively.

The major experimental consideration influencing the utility of state–time analyses is the time interval between successive measurements that allows the possibility of state variables to be temporally associated. In other words, measurements taken during very short time intervals will tend to be autocorrelated

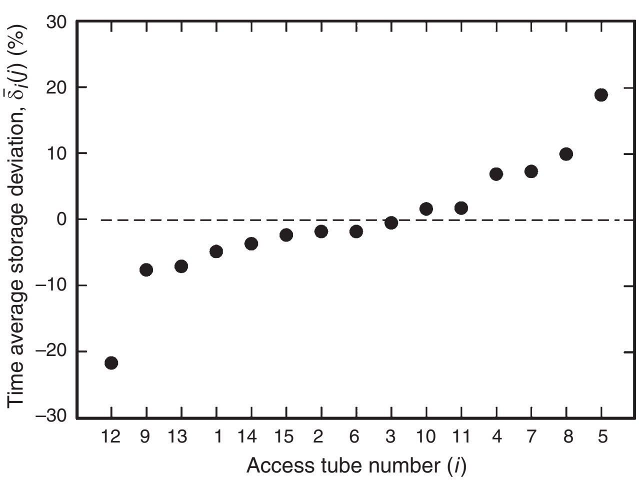

Fig. 3. Rank plots of time average relative spatial storage deviations $\overline{\delta}_i(j)$.

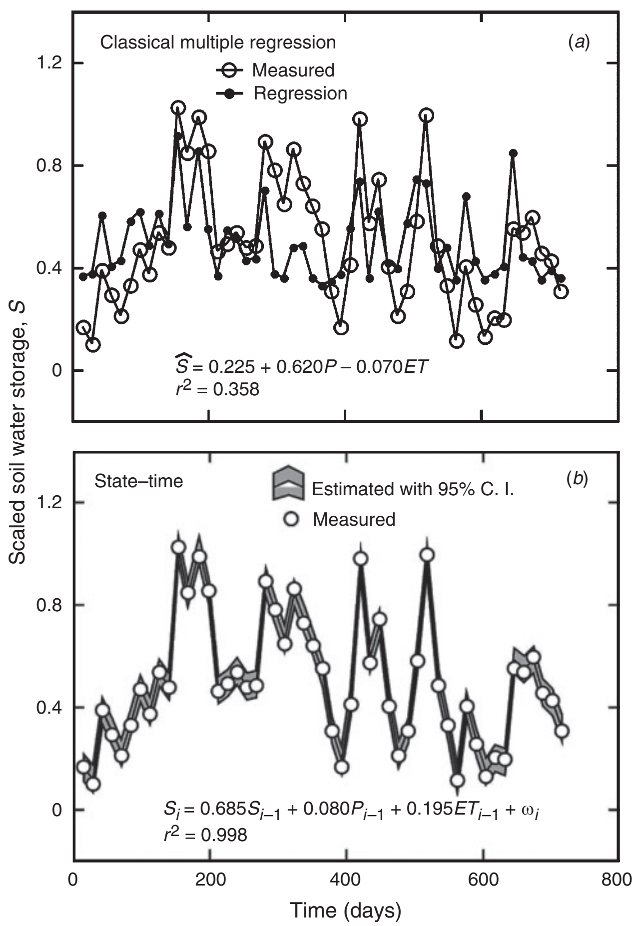

Fig. 4. Estimates of soil water storage measured biweekly for 714 days using (a) classical multiple regression and (b) state–time analysis.

or cross-correlated with each other. However, with increasing time, the state variables change their magnitudes as environmental conditions change. We know that a water balance for a given soil profile is the result of five processes that occur as a function of time: precipitation plus irrigation, surface runoff, evapotranspiration, storage of water in the soil profile, and the drainage of water from the soil profile. Each of these processes quantified by Silva *et al*. (2006), who provide data for this study, indicated that surface runoff was negligible over the 2-year period and that the drainage of water from the soil profile has yielded accurate measurements of water storage S_i in the profile. Hence, neglecting surface runoff, the use of only three state variables (S, P, and ET) in the state–time analysis accounts for the physical processes responsible for a quantitative estimate of S provided that the amounts of water that eventually drain from the 1-m soil profile from occasional large rainfalls can be robustly accounted for in the state variable P. The temporal autocorrelation and cross-correlation functions given in Table 1 indicate that ET, S, and P have autocorrelation lengths of 3, ~2, and <1 lag, respectively. In other words, values of ET are related to each other during more than three consecutive sampling dates (42 days), those of S during no more than two consecutive sampling dates (28 days), and those of P are essentially not related to each other between consecutive sampling dates (14 days). All three values of lag are reasonable, including that for precipitation. Indeed, the general nature of rainfall is more seasonal and does not consistently repeat its relative magnitude with a 2-week periodicity through a 2-year period.

Examining the cross-correlation coefficients in Table 1, we are not surprised to find that ET is related to P for more than three consecutive sampling dates (42 days) and that S is related to P for at least two consecutive sampling dates (28 days). The fact that ET and S showed essentially not to be related to each other between sampling dates is not obvious, since on many occasions the actual value of ET was much below the potential value. However, during the 2-year period, regardless of the daily and biweekly fluctuations of local weather conditions, every effort was made to irrigate the field in a timely manner to provide adequate amounts of water stored in the root-zone.

There are several methods available to examine the reliability of state–time analyses (see for example, Shumway and Stoffer 2000). Here, we choose (on the basis of the information in Table 1) to observe the impact of omitting increasing numbers of observations from the calculations of the state variable being estimated. An example is given in Fig. 5 where the S is estimated with all measurements of P and ET, but with increasing numbers of its biweekly measurements omitted from the state–time analysis.

Figure 5*a* illustrates the results when one-half of the observations of soil water storage were not considered in the calculations. Comparing Figs 4*b* and 5*a*, it can be seen that the coefficient of determination r^2 decreased slightly from 0.998 to 0.957 and that the width of the confidence intervals increased. At each time step when a measured value of S is omitted from the calculation, its forward prediction cannot be compared to its observation, and hence, an update based on its temporal association is precluded and causes a larger confidence interval.

State–time estimates in Fig. 5*b* made while ignoring two of every three observations of soil water storage are not as good as those illustrated in Fig. 5*a*. Nevertheless, a linear regression between state–time estimated and measured values of S yielded a coefficient of determination $r^2 = 0.834$. However, notice that approximately five values omitted in the calculations fall outside of the confidence interval as a result of the state–time analysis judging they did not belong to the distribution of S values used in the calculation.

State–time estimates in Fig. 5*c* made while ignoring three of every four observations of S are definitely not reliable. A linear regression between estimated and measured values of S yielded a coefficient of determination r^2 of only 0.296, and ~16 values omitted in the calculations fall outside of the confidence interval. There are two primary reasons why the state–time estimates illustrated in Fig. 5*c* do not agree with reality. First, during a time period of 56 days (4 lags and nearly equal to 2 months), values of soil water storage are no longer temporally related to each other during the 2-year experiment (Table 1)—a requirement of state–time analyses. Second, the amounts of water that eventually drained from the 1-m soil profile from large rainfalls robustly accounted for in the state variable P occurring within time spans of 56 days could not be ignored. Note in Figs 4*b*, 5*a–c*, as the relative number of ignored observations of S increases, the magnitude of the transition coefficient of S_{i-1} decreases with estimates of S_i depending progressively on the values of P_{i-1}. In other words, with fewer and fewer temporal observations of S_i available, reliable estimates of S_i depend more and more on the temporal association between S and precipitation. This dependence is entirely reasonable since changes in S are generally related directly to amounts of precipitation infiltrating the soil surface during relatively short time periods. Notice, however, that no such consistent trend was

Table 1. Autocorrelation and cross-correlation coefficients for state variables soil water storage *S*, precipitation *P*, and evapotranspiration *ET*

lag h[B]	Autocorrelation coefficient $r(h)$[A]			Cross-correlation coefficient $r_c(h)$[A]		
	S	P	ET	S v. P	ET v. S	ET v. P
0	1	1	1	0.595	0.153	0.359
1	0.551	0.163	0.558	0.370	0.005	0.507
2	0.257	0.119	0.444	0.203	–0.053	0.316
3	0.038	0.024	0.344	0.033	0.036	0.375
4	–0.005	0.082	0.185	–0.050	0.159	0.072
5	0.025	0.081	0.099	–0.027	0.126	–0.028

[A]The 95% significance level of r and r_c is 0.2745. [B]A lag of $h = 1$ is equal to 14 days.

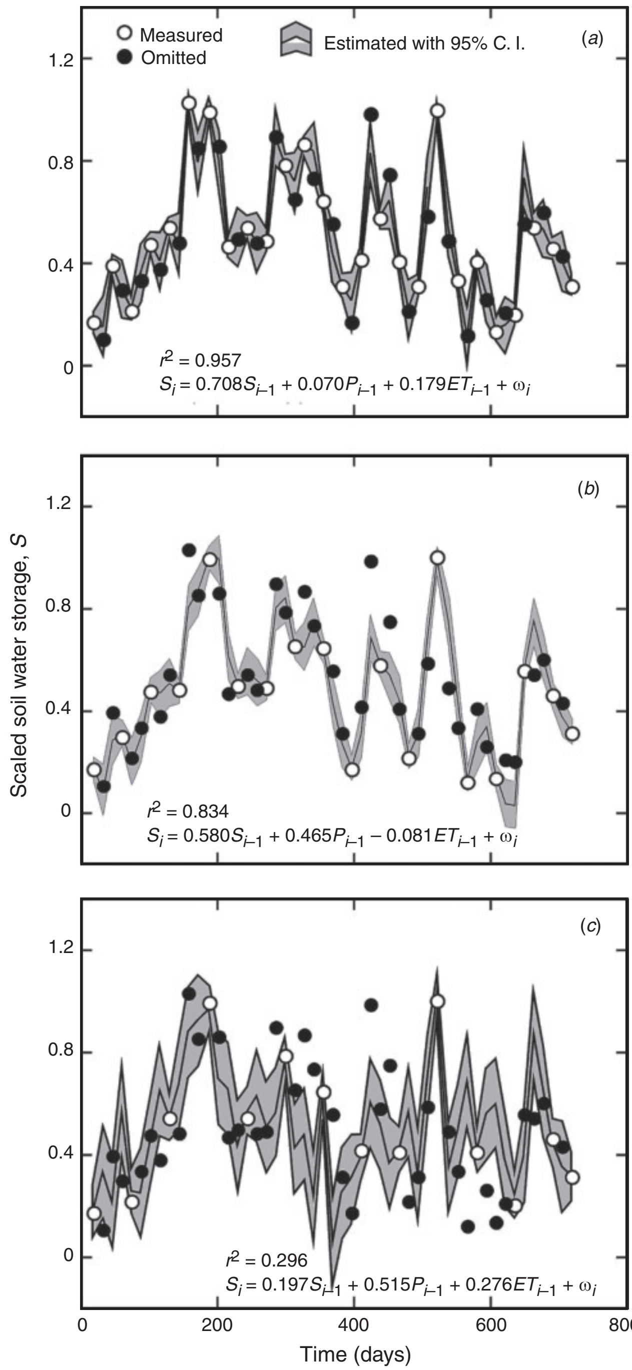

Fig. 5. Soil water storage measured biweekly for 714 days estimated from measurements of precipitation and evapotranspiration with (*a*) one-half, (*b*) two-thirds, and (*c*) three-quarters of the soil water storage observations omitted from the state–time analysis.

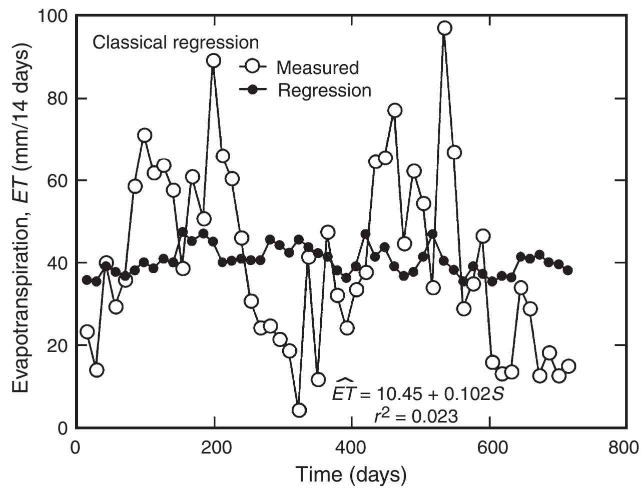

Fig. 6. Evapotranspiration measured biweekly for 714 days estimated using classical linear regression.

manifested during these short time periods by the transition coefficients of ET_{i-1}. This fluctuation is reasonable in that changing local weather conditions can easily cause major shifts in *ET* that do not impose major changes in average soil water storage. We verify the previous statement by examining Fig. 6 where the mean values of *ET* throughout the time investigated are related by simple linear regression to the average amount of water stored in the soil profile. This Figure indicates that measurements of *ET* at any given time are not realistically estimated by the amount of water stored within the root-zone of the soil profile of the coffee crop. Yet, *S* is generally sparingly and inadequately monitored in agricultural fields to assure that there is sufficient water within the root-zone for the crop to sustain an adequate transpiration rate for optimal growth and harvestable yield.

Rather than repetitively measuring the water stored in the soil profile to ascertain *ET* across an agricultural field or even at a location designated as representing the mean (access tube number 3 according to Fig. 3), a common practice has been the measurement of water lost from a Class A evaporation pan (Allen *et al.* 1998). This procedure is convenient and inexpensive, but does not necessarily relate to quantitative measures of *S* at positions related to mean values for the field, or *vice versa*.

With measurements of mean values of *ET*, *S*, and *P* laboriously made biweekly in this study, we are able to examine the estimation of *ET* made by classical multiple regression and state–time analyses. Estimations of *ET* using classical multiple regression based on mean values of each variable throughout the time being investigated can be compared with measured values in Fig. 7*a*. We note that no more than 13.5% of the variance of the biweekly measured evapotranspiration data was explained from the measurements of soil water storage and precipitation. We also note that variations of *ET* with a coefficient of 0.415 were more related to fluctuations of precipitation than those of soil water storage, with a coefficient of only 0.095. A similar relationship was also apparent in the state–time analysis presented in Fig. 7*b*, where the transition coefficient of *S* was only 0.090 while that of *P* was larger, having a value of 0.310. Estimated values of *ET* from the state–time analysis approached those of the measured values, and manifested a coefficient of determination of 0.887. Nevertheless, eight of the 51 estimated values of *ET* fell outside the 95% confidence interval.

In Fig. 8 where ET is estimated with all measurements of *S* and *P*, but with one-half and three-quarters of its biweekly

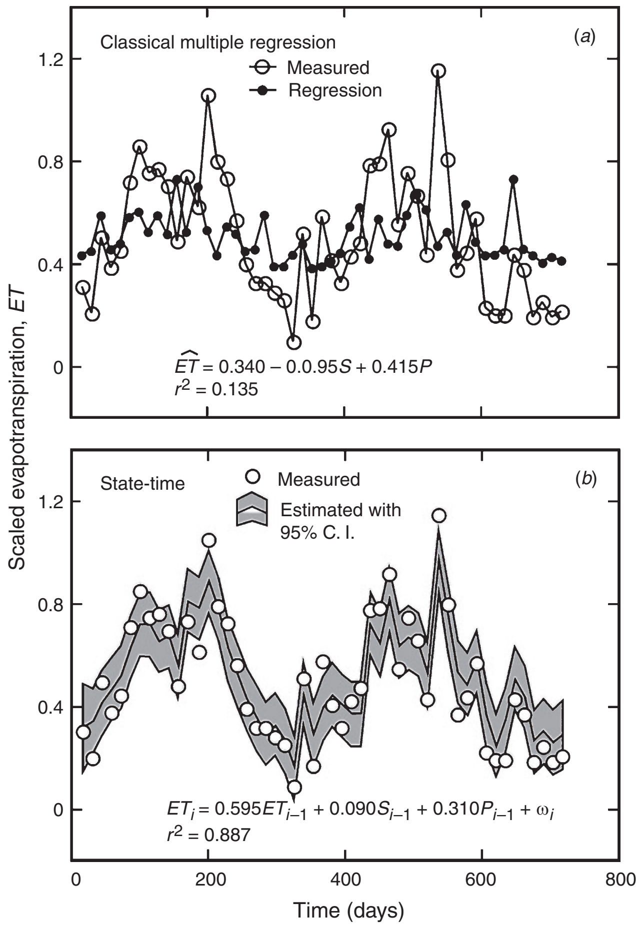

Fig. 7. Estimates of evapotranspiration measured biweekly for 714 days using (*a*) classical multiple regression and (*b*) state–time analysis.

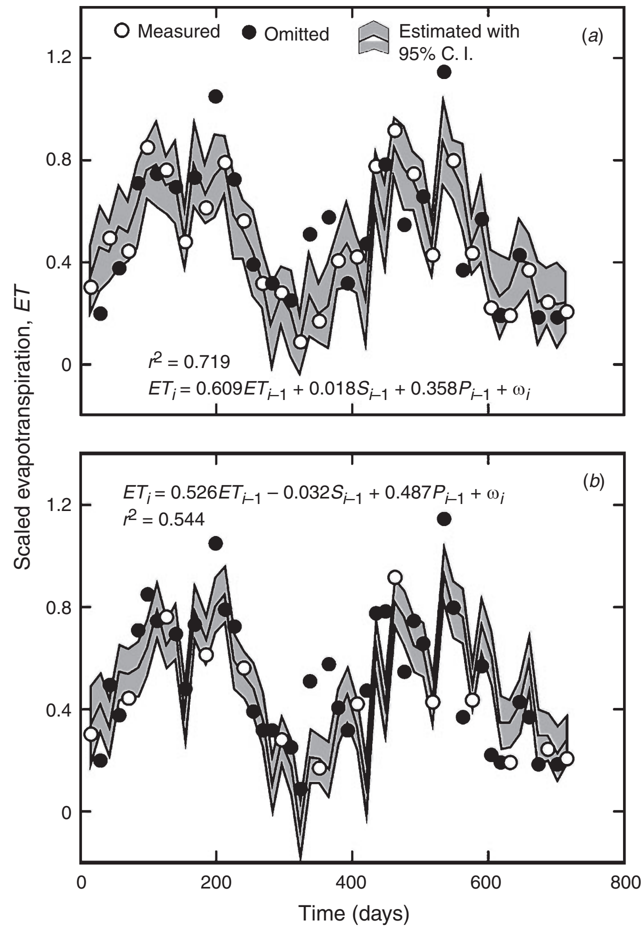

Fig. 8. Evapotranspiration measured biweekly for 714 days estimated from measurements of soil water storage and precipitation with (*a*) one-half and (*b*) three-quarters of the evapotranspiration observations omitted from the state-time analysis.

measurements omitted from the state–time calculations, the coefficient of determination decreases to 0.719 and 0.544. Nine of the 51 estimated values of *ET* in Fig. 8*a* and 21 of the 51 estimated values of *ET* fell outside the 95% confidence interval.

Noting that the contribution from neighbouring values of *S* decreases from 9% in Fig. 7*b* to a mere 2% and 3% in Fig. 8*a* and *b*, respectively, we learn that for the case of this dataset from a coffee crop, the temporal variations in *ET* are not physically caused by variations of *S*. Therefore, we examine the relationships between the two state variables *ET* and *P* in Fig. 9.

Classical regression between *ET* and *P* throughout the time of the investigation yielded a coefficient of determination of only 0.129 (Fig. 9*a*). On the other hand, state–time estimates were much more reliable, with a coefficient of determination of 0.864 (Fig. 9*b*). We expected that the state–time analysis would be superior because *ET* and *P* are significantly cross-correlated with three temporal lags and *ET* has an autocorrelation length of three lags. We note that each preceding value of both state variables more or less equally contributes to the estimated value of *ET*. By omitting one out of two values of measured *ET* (Fig. 10*a*) and three out of four values of measured *ET* (Fig. 10*b*) in the state–time analyses, we learn that the coefficient of determination between estimated and measured values of *ET* reduces from 0.864 (Fig. 9*b*) to 0.694 and 0.554 (Fig. 10*a* and *b*, respectively). Without having neighbouring values of *ET* for the updating procedure in the calculation, the contribution from the neighbouring cross-correlated measured *P* is inadequate to capture estimates of *ET* within an ever-increasing confidence interval. In other words, state variables physically linked to the cause of *ET* fluctuations were not monitored.

During 14-day intervals, what physical processes in addition to precipitation alter the amount of water transpired from the crop and evaporated from the soil surface? From the above information, rainfall (and to a very limited extent, soil water storage) is the only parameter that accounts for some of the 14-day variability of *ET* throughout each year, as illustrated in Fig. 11. Patterns of *ET* for both years are very similar, and indeed have similar spectra yielding significant coherence at several temporal frequencies not presented here. In order to identify the cause of this similarity as well as to improve estimates of evapotranspiration as a function of time, it would be necessary to measure at least one other variable or parameter physically responsible or linked to evapotranspiration, e.g. air temperature, relative humidity, cloudiness, wind velocity, soil temperature, distribution of water within the soil profile, vegetative and

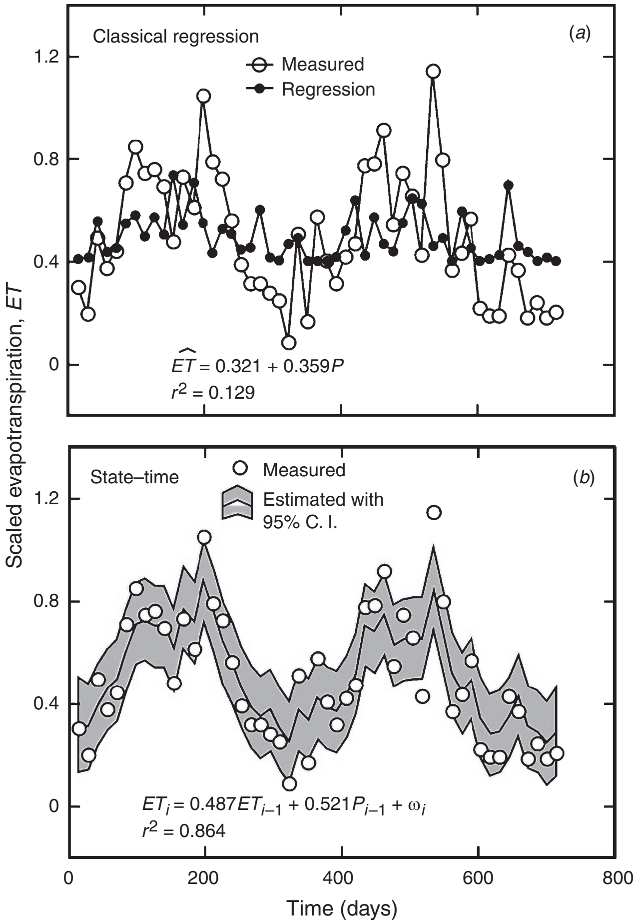

Fig. 9. Estimates of evapotranspiration measured biweekly for 714 days using (*a*) classical multiple regression and (*b*) state–time analysis.

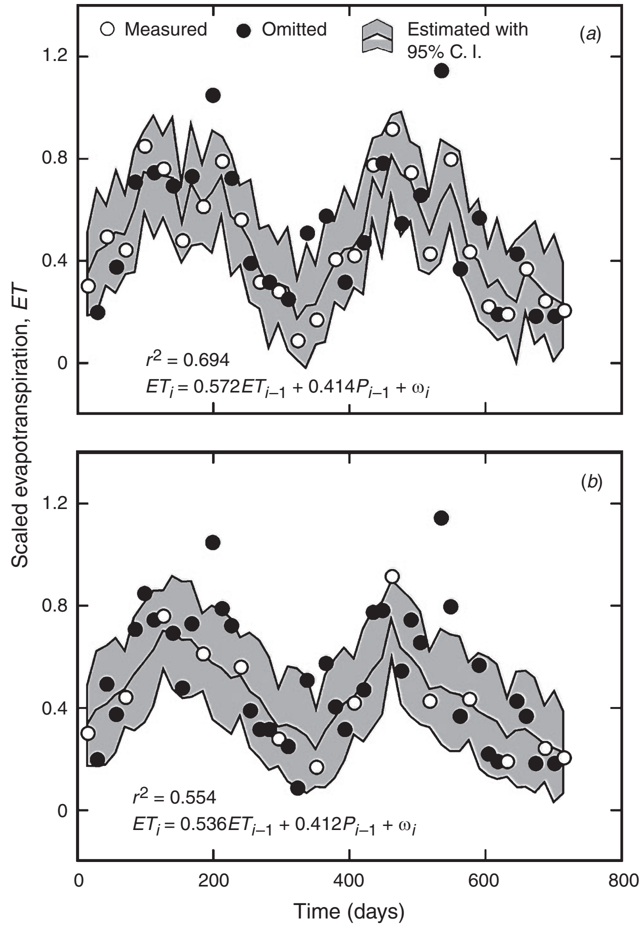

Fig. 10. Evapotranspiration measured biweekly for 714 days estimated from precipitation measurements with (*a*) one-half and (*b*) three-quarters of the evapotranspiration observations omitted from the state–time analysis.

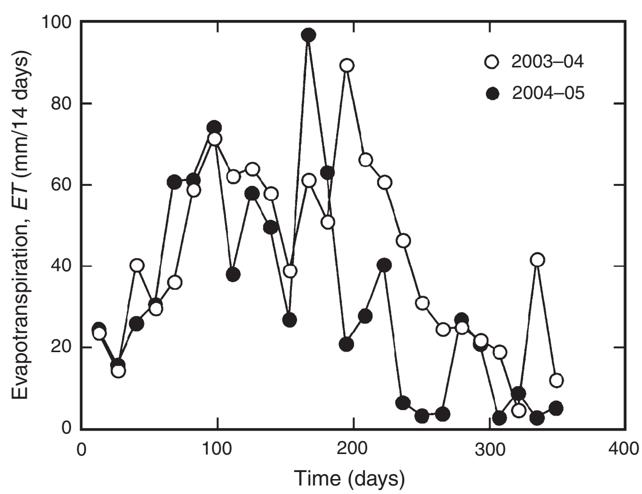

Fig. 11. Biweekly measurements of evapotranspiration during 2003–04 and 2004–05 *v.* time commencing the first week of September.

productive stage of the crop, insect damage, plant diseases, and plant nutrient availability, mainly (Penman 1963; Allen *et al.* 1998).

Previous and present outlook

As mentioned in the Introduction, the estimation of *S* is a difficult task due to the spatial and temporal variability of field soils and their local environment. This presentation focused on characterising the average amount of water stored in the topsoil across a specific field measured at time intervals of 14 days. Because most coffee plant roots were limited to a depth of 1 m within the soil profile, soil water designated as that available to the coffee crop was calculated from soil water content measurements from the soil surface to 1 m deep. The field was irrigated only when it was deemed necessary, i.e. whenever the stored water in the profile reached <20% of its full capacity. This irrigation strategy, embracing the concept that the spatial variation of *S* was invariant in time, allowed the analysis of the distribution of soil water storage measurements within the coffee field to ascertain a unique location consistently manifesting the mean soil water storage regardless of its time of measurement. In addition, the minimum number of locations sampled to achieve an average value within prescribed level of significance was based on the assumption that the sampled values were normally distributed. This strategy has been suggested during the past 25 years. Various other closely related strategies that include the measurement of a threshold minimum soil water storage, a specified integrated matric potential within the root-zone of a plant, and minimum soil water content or matric potential at a specified position within the root-zone were explored and adopted since 1950 (Nielsen and Kutílek 1994). These strategies ignored the spatial distances between sampling locations, and also ignored the temporal correlations between successive soil water storage sampling campaigns. All of them sought and relied on the ability to find a 'good average' to determine when to irrigate a crop. Few strategies were developed to ascertain how seldom measurements could be taken to ascertain when to irrigate for optimal crop production. As a result, published literature will testify that excessive energy and time were spent determining when to irrigate rather than to determine the relative benefit of having irrigated. Hence, the second half of this presentation focused on the temporal association of *S*, *P*, and *ET* at a fixed, hypothetical location assumed to represent the entire field during 14-day time intervals empirically selected for their measurement. The results of this tedious, and time- and energy-consuming, sampling program indicate that the amount of water stored in the soil profile during the empirically designated, 14-day-interval sampling program has little to do with temporal variations of evapotranspiration. However, from locations sampled across the coffee field, it was apparent that the mean values of *ET* during the 14-day sampling intervals are temporally related to *P*, not *S*, and not quantitatively related to infrequent irrigations, including those made in September.

After completing this experiment, we are left with the question: When and where do we take what kind of other measurements to better manage the production of coffee as well as gain information on the environmental impact on its production? During the past two decades, the concept of site-specific farming, precision farming or precision agriculture, has emerged, which emphasises that the quality and quantity of crop production can be improved by simultaneously managing the temporal and spatial variations of crop-dependent processes across an agricultural field during crop growth. In other words, an agricultural field planted to one crop is not considered a unit to be managed or treated uniformly. Instead, based on its local soil and environmental properties, and the nature of physical and biological processes, it is managed as an ensemble of distinct spatial domains each monitored over appropriate scales of space and time. Many methods of statistical analysis (geostatistics, regionalised variable analysis, applied time series, etc.) are available for examining experimental data observed at different points in time and space relative to describing and understanding soil–plant–atmosphere processes within a farmer's field.

Here, we illustrated the utility of state–time analysis to examine the temporal variation of the crop-dependent process of *ET*. We analysed *ET*, considering it to be a random variable, and statistically treated its temporal variation as a function of the time between repetitive observations within a 2-year domain. At any given time, its value was considered to be identical at every location within the experimental area. Although such a consideration is not realistic because *ET* actually varies from one location to the next throughout the entire spatial domain, it is consistent with the common practice of irrigating a field with a given amount of water or also assuming that the rain measured at a specific location falls uniformly across the field.

Having briefly illustrated the utility of state–time analysis in this simple experiment to examine the temporal variation of the crop-dependent process of *ET* within a field, it is clear that many related choices for meaningful field research remain open for immediate application.

One such choice taken by several researchers in the past was to make repeated-measurements of *S*, *ET*, and *P* at the same spatial interval across the experimental area for at least one time. The benefit of state–space analysis to examine the spatial processes of these crop-dependent variables at the time of their measurement should be realised by considering each of them to be a random variable treated statistically, with their spatial association and variation being a function of the distance between their measurements. A spatial soil process is the change of a variable or a vector consisting of several variables across a spatial domain caused by localised conditions; for example, the spatial process of soil water storage considered across a field can be mainly influenced by spatial changes in soil type, topography, vegetation, rainfall, ET, and management.

Obtaining measurements of *S*, *ET*, and *P* repetitively across an experimental area at variable spatial intervals for numerous times as presented here provides another choice. Using 2-dimensional state analysis in both time and space, a complete analysis of the progression of any or all of the three variables occurring at any location in the field at any time during the 2-year experiment would be highly informative. In other words, the analysis would provide 'site-specific' and 'time-specific' management information without the disadvantage of considering average values across the field or during each year.

Further choices could be realised when measurements of coffee plant parameters—locally available soil nutrient and micro-environmental conditions related to potential coffee bean yields—are repeatedly and frequently made across the field during each growing season. With this information, a 2-dimensional state analysis provides quantitative guidelines during the growing season to better manage the crop within specific local field domains to achieve higher yields without a deleterious impact on soil and water resources. As a result, management of the field would be more efficient and sustainable.

Conclusions

Following the most commonly used classical procedure of randomisation to identify sampling locations within a field of small replicated plots, we compared the results of two analyses: classical statistics and one application of applied time series (state–time analysis) to examine the temporal variability of soil water storage in a coffee field. Classical statistical procedures indicated that randomly spaced estimates of *S* averaged across the field can be obtained with a deviation of 2% of the mean using only four of the 15 sampled locations. Time stability analysis of *S* showed that a single specified location would represent the average value of *S* in relation to the average of the 15 locations, and if a standard deviation is required, four specific locations would yield an average with a deviation of only 0.3%.

In contrast to classical multiple regression analysis, the state–time analysis showed that S_i was more dependent on P_{i-1} (52%) than on ET_{i-1} (28%) and S_{i-1} (20%), indicating the low temporal dependence of S in relation to previous measurements. Additionally, the analysis showed that ET_i was not realistically estimated from S_{i-1} measurements inasmuch as it was more dependent on previous estimations ET_{i-1} (59%) than on P_{i-1} (30%) and S_{i-1} (9%). With P and ET easily obtained from automated weather stations, the state–time analysis indicated that S measurements made every 14 days could be reduced to monthly measurements, and that S_i measurements would still be predicted with an r^2 of 0.957, significantly reducing future field work.

We, as well as other researchers with whom we communicate, are conducting field experiments in which measurements of S are being taken at regular intervals in two spatial directions across a cultivated field at specified times that allow a 3-dimensional space–state–time analysis. These experiments should provide improved management without depending on traditional, randomly treated small plots supposedly applicable to an entire field without any sort of experimental verification.

References

Aboitiz M, Labadie JW, Heermann DF (1986) Stochastic soil moisture estimation and forecasting for irrigated fields. *Water Resources Research* **22**, 180–190. doi:10.1029/WR022i002p00180

Allen RG, Pereira LS, Raes D, Smith M (1998) 'Crop evapotranspiration (guidelines for computing crop water requirements).' FAO Irrigation and Drainage Paper No. 56. (FAO: Rome)

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Centro Nacional de Pesquisa de Solo (2006) 'Sistema brasileiro de classificação de solos.' (Embrapa Solos: Rio de Janeiro)

Fenilli TAB, Reichardt K, Dourado-Neto D, Trivelin PCO, Favarin JL, Costa FMP, Bacchi OOS (2007) Growth, development and fertilizer N recovery for a coffee crop growing in Brazil. *Scientia Agricola* **64**, 541–547. doi:10.1590/S0103-90162007000500012

Hills TC, Reynolds SG (1969) Illustrations of soil moisture variability in selected areas and plots of different sizes. *Journal of Hydrology* **8**, 27–47. doi:10.1016/0022-1694(69)90029-8

Hu W, Shao MA, Reichardt K (2010) Using a new criterion to identify sites for mean soil water storage evaluation. *Soil Science Society of America Journal* **74**, 762–773. doi:10.2136/sssaj2009.0235

Hu W, Shao MA, Wang QJ, Reichardt K (2008) Soil water content temporal-spatial variability of the surface layer of a Loess Plateau hillside in China. *Scientia Agricola* **65**, 277–289. doi:10.1590/S0103-90162008000300008

Hui S, Wendroth O, Parlange MB, Nielsen DR (1998) Soil variability: infiltration relationships of agroecosytems. *Journal of Balkan Ecology* **1**, 21–40.

Kachanoski RG, De Jong E (1988) Scale dependence and temporal persistence of spatial patterns of soil water storage. *Water Resources Research* **24**, 85–91. doi:10.1029/WR024i001p00085

Köppen W (1931) 'Grundriss der klimakunde.' (Walter D E Guyter & Co.: Berlin)

Mohanty BP, Mousli Z (2000) Saturated hydraulic conductivity and soil water retention properties across a soil-slope transition. *Water Resources Research* **36**, 3311–3324. doi:10.1029/2000WR900216

Moreti D, Libardi PL, Rocha GC, Lovatti MJ, Aguiar LIG (2007) Avaliação espaço-temporal das armazenagens gravimétrica e volumétrica da água num latossolo com citros. *Revista Brasileira de Ciencia do Solo* **31**, 1281–1290. doi:10.1590/S0100-06832007000600007

Nielsen DR, Kutílek M (1994) 'Soil hydrology.' (Catena Verlag: Reiskirchen, Germany)

Nielsen DR, Wendroth O (2003) 'Spatial and temporal statistics: sampling field soils and their vegetation.' (Catena Verlag: Reiskirchen, Germany)

Penman HL (1963) 'Vegetation and hydrology.' Tech. Comm. 53. (Commonwealth Bureau of Soils: Harpenden, England)

Reichardt K, Portezan-Filho O, Bacchi OOS, Oliveira JCM, Dourado-Neto D, Pilotto JE, Calvache M (1997) Neutron probe calibration correction by temporal stability parameters of soil water content probability distribution. *Scientia Agricola* **54**, 17–21[Special issue]. doi:10.1590/S0103-90161997000300003

Shumway RH (1988) 'Applied statistical time series analysis.' (Prentice Hall: Englewood Cliffs, NJ)

Shumway RH, Stoffer DS (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**, 253–264. doi:10.1111/j.1467-9892.1982.tb00349.x

Shumway RH, Stoffer DS (2000) 'Time series analysis and its applications.' (Springer: New York)

Silva AL, Bruno IP, Reichardt K, Bacchi OOS, Dourado-Neto D, Favarin JL, Costa FMP, Timm LC (2009) Soil water extraction by roots and Kc for the coffee crop. *Revista Brasileira de Engenharia Agrícola e Ambiental* **13**, 257–261.

Silva AL, Roveratti R, Reichardt K, Bacchi OOS, Timm LC, Bruno IP, Oliveira JCM, Dourado-Neto D (2006) Variability of water balance components in a coffee crop grown in Brazil. *Scientia Agricola* **63**, 153–165. doi:10.1590/S0103-90162006000200001

Soil Survey Staff (1993) 'Soil survey manual.' (US Government Printing Office: Washington, DC)

Timm LC, Gomes DT, Barbosa EP, Reichardt K, Souza MD, Dynia JF (2006) Neural network and state-space models for studying relationships among soil properties. *Scientia Agricola* **63**, 386–395. doi:10.1590/S0103-90162006000400010

Vachaud G, Passerat De Silans A, Balabanis P, Vauclin M (1985) Temporal stability of spatially measured soil water probability density function. *Soil Science Society of America Journal* **49**, 822–828. doi:10.2136/sssaj1985.03615995004900040006x

Warrick AW, Nielsen DR (1980) Spatial variability of soil physical properties in the field. In 'Applications of soil physics'. (Ed. D Hillel) pp. 319–344. (Academic Press: New York)

Western AW, Grayson RB, Blöschl G (2002) Scaling of soil moisture: a hydrologic perspective. *Annual Review of Earth and Planetary Sciences* **205**, 20–37.

Western AW, Zhou SL, Grayson RB, McMahon TA, Blöschl G, Wilson DJ (2004) Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *Journal of Hydrology* **286**, 113–134. doi:10.1016/j.jhydrol.2003.09.014

Manuscript received 19 January 2010, accepted 13 July 2010