

Early Estimation of Drought Impacts on Rainfed Wheat Yield in Mediterranean Climate

Dimitris Tigkas · George Tsakiris

Received: 10 October 2014 / Accepted: 28 November 2014 / Published online: 11 December 2014
© Springer International Publishing Switzerland 2014

Abstract Drought events can cause severe damages to agriculture and in particular to rainfed agriculture. The problems can be more evident in semi-arid regions of the Mediterranean. It is of great importance to assess the impacts of drought on the yield at an early stage, in order to be able to take measures on time for mitigating the anticipated losses, supporting the farmers, and enhancing food security. The aim of the paper is to formulate a simple methodology for estimating the impacts of drought on rainfed agriculture, and especially on wheat yield, prior to the harvest. The Reconnaissance Drought Index (RDI) is used as the main independent variable in linear regression models for the assessment of drought effects on wheat yield. To test the proposed methodology, data from two rural areas of Greece are used and the wheat yield is simulated using the AquaCrop model. Multiple regression models with variables that include RDI and the minimum temperature for winter months are tested and evaluated through a cross-validation process. The performance of the models is assessed by various criteria. Indicatively, the correlation coefficients at each study area reach 0.87 and 0.91 for predictions 1 month before harvest, 0.84 and 0.82 for 2 months and 0.67 and 0.77 for 3 months before harvest. The results show that the RDI, for reference periods that represent the critical development stages of the crop, is highly correlated with the wheat yield. A satisfactory prediction of the drought impacts on wheat yield 2 to 3 months before the harvest can be achieved. The proposed methodology can be useful to the authorities, stakeholders and insurance services for facing drought events and mitigating losses in rainfed agriculture.

Keywords Rainfed agriculture · Wheat yield prediction · Drought impacts · RDI · AquaCrop · Mediterranean

D. Tigkas (✉) · G. Tsakiris

Centre for the Assessment of Natural Hazards and Proactive Planning, and Laboratory of Reclamation Works and Water Resources Management, School of Rural and Surveying Engineering, National Technical University of Athens, Iroon Polytechniou 9, Zografou, 157 80 Athens, Greece
e-mail: ditigas@mail.ntua.gr

G. Tsakiris
e-mail: gtsakir@central.ntua.gr

1 Introduction

Drought is a regional recurrent phenomenon, characterised by a temporary severe decrease of water availability, with significant societal, economic and environmental impacts (Tsakiris et al. 2013). The degree of a region's vulnerability depends on many environmental and social factors (Bordi and Sutera 2008). Agriculture is one of the most vulnerable sectors to drought, especially in arid and semi-arid regions, such as the Mediterranean (Kumar 1998; Tsakiris and Tigkas 2007). The impacts of drought on agriculture in a region cannot be easily measured, because there is not a unique way to establish a relationship between a key factor that determines drought and crop yield. The selection of a single crop on which emphasis may be given, may simplify the analysis of agricultural drought for a region (Kumar and Panu 1997).

Wheat is a widely cultivated crop with a major role in world's economy. Winter wheat is a typical rainfed crop in the Mediterranean climate, which is used in this study as the most representative crop for the assessment of drought impacts on crop yield. The early estimation of drought impacts is important to both farmers and responsible organisations in order to take the necessary proactive and supportive measures for mitigating the anticipated consequences.

Many researchers have worked on the field of early prediction of wheat yield, using methods that include meteorological parameters or indices (agrometeorological indices, drought indices, etc.). For example, Salman and Al-Karablieh (2001), and Lee et al. (2013) used empirical models that included precipitation and temperature data, and Kazmi and Razul (2012) used precipitation, temperature and sunshine duration data. Also, Mavromatis (2007) investigated the use of the Standardized Precipitation Index (SPI) and the Palmer Drought Severity Index (PDSI), while Sadat Noori et al. (2012) used SPI along with evapotranspiration and temperature data. Kogan et al. (2013) used an NDVI-based approach to predict winter wheat yield with regression models.

Dalezios et al. (2002) investigated the role of agrometeorological and agrohydrological indices on temporal development of phenological stages of wheat, and found that temperature and precipitation were important parameters. Kristensen et al. (2011) used agroclimatic indices for investigating winter wheat response and indicated the importance of temperature on the yield. Liakatas (1997) underlined the significance of weather parameters (precipitation, temperature and potential evapotranspiration) in various phenological stages of rainfed wheat, especially in arid environments. Porter and Gawith (1999) outlined the effects of climatic variability and temperature extremes on wheat yields.

From the above, it can be clearly concluded that the key determinant for wheat yield assessment is precipitation. However, other climatic variables seem to play a rather important role, as well, in the development stages of the crop. Among them, temperature seems to be the most important, either as a regulator of the crop water needs through the evapotranspiration process, or as a limiting factor that may cause major problems to the plants (e.g., low winter temperatures) (Proutsos et al. 2010).

Therefore, the use of an index that incorporates both precipitation and potential evapotranspiration (PET) should be appropriate for the assessment of drought impacts on crop yield. In this paper, an approach is proposed that uses the Reconnaissance Drought Index (RDI) as the main variable for the early assessment of the impacts on the wheat yield. The predictability efficiency is tested for various time frames (1 to 3 months before harvesting). Critical phenological stages of the crop are considered for using the appropriate reference periods of the RDI in linear regression models. This approach can be suitable for operational purposes, as it requires easily obtainable data (precipitation and temperature) and does not involve complex procedures.

For the case study used in this paper for illustrating the proposed method, precipitation and temperature data from two major agricultural areas of Greece are used. The evaluation of the drought effects is achieved by using the AquaCrop water productivity simulation model for the simulation of wheat yield in these areas.

2 Materials and Methods

2.1 The Reconnaissance Drought Index

The Reconnaissance Drought Index (RDI) is a simple, yet practical index for studies of drought impacts on agriculture, since it takes into account both precipitation and PET, which are key factors for the development stages of the plant (Tsakiris et al. 2010). Also, it can be used as a composite climatic index for identifying climatic variations (Tigkas et al. 2013a). The RDI is the ratio of the cumulative precipitation to the cumulative PET for a specific time period (Tsakiris and Vangelis 2005; Tsakiris et al. 2007). It is expressed either by its initial value (α_k) or by its standardised form (RDI_{st}). The initial value (α_k) of RDI is calculated for the i -th year in a reference period of k (months) as follows:

$$\alpha_k^{(i)} = \frac{\sum_{j=1}^k P_{ij}}{\sum_{j=1}^k PET_{ij}}, \quad i = 1(1)N \text{ and } j = 1(1)k \tag{1}$$

in which P_{ij} and PET_{ij} are the precipitation and PET of the j -th month of the i -th year and N is the total number of years of the available data. The average of the annual values of α ($\bar{\alpha}_{12}$) is equal to the aridity index of the area (UNEP 1992).

The standardised form of RDI (RDI_{st}) is calculated through a standardisation process, assuming that α_k values fit the lognormal or the gamma distribution (Tsakiris et al. 2007, 2008; Tigkas 2008). The RDI_{st} provides values categorised in predefined drought classes, so it can be used as a global index. Drought severity categorisation includes mild, moderate, severe and extreme drought classes, with corresponding boundary values of RDI_{st} (-0.5 to -1.0), (-1.0 to -1.5), (-1.5 to -2.0) and (<-2.0), respectively (Tigkas et al. 2012).

It should be noted, that the method of calculation of PET does not have any significant effect on RDI_{st} , meaning that temperature based methods, such as Hargreaves (Hargreaves and Samani 1982), can be sufficient for producing reliable RDI results (Vangelis et al. 2013).

2.2 Wheat Crop Characteristics and Development Stages

Wheat is one of the most produced crops worldwide, affecting significantly the economy in local and regional levels. Winter wheat is a typical rainfed crop for the Mediterranean region. Winter wheat planting takes place usually in November and the growing period lasts for about 8 months (Allen et al. 1998). The major physical constraints to wheat production in the Mediterranean are the terminal drought and terminal heat stress as well as year to year weather variability. Drought is probably the major cause of yield loss in these environments. For this reason, the wheat grain yield follows the year to year variability of rain (Acevedo et al. 1999).

Total cumulative evapotranspiration (ET) of wheat crop typically ranges from 200 to 500 mm, being around 400 mm for rainfed wheat in Mediterranean conditions (Shimshi 1973; Steduto et al. 2012).

Early in the growing season under Mediterranean conditions, the daily water consumption can be less than 2 mm/day due to relatively low temperatures and high humidity. As the canopy enlarges during tillering and stem elongation, the rate of water consumption increases and typically reaches a peak around anthesis, at rates between 5 and 8 mm/day (Steduto et al. 2012).

There are three periods in which wheat yield is considered to be most responsive to moisture stress: first, the period when tillers are developing and their abortion rates are highest; second, when florets are being formed and grains are set; and third, from early to mid-grain filling when young developing grains can be aborted due to a lack of assimilate (Fischer 1973; Turner 1997). Water deficit during tillering can reduce the number of tillers, affecting the final yield (Shimshi 1973). Important yield variables under drought conditions can be considered the number of spikes/m², the weight of grains/spike, the harvest index and the biological yield (Leilah and Al-Khateeb 2005).

Low winter temperatures (mainly in January and February under Mediterranean conditions) play a significant role in the growth of wheat, affecting the tillering process. In addition, dormancy period may be developed if very low temperatures sustain for many days, while frost may cause leaf death (Kazmi and Razul 2012).

2.3 Yield Simulation with AquaCrop Model

Long-term data availability of wheat yield are usually either limited or of unverified reliability. Further, the actual yield depends on various factors, such as soil characteristics, cultivation techniques, plant diseases etc. Since the purpose of this study is to assess the effects of drought on yield, which is determined by the climatic conditions, the AquaCrop model was employed for the simulation of the crop yield. This way, other influencing factors can be set to have a neutral role in crop growth, without affecting the interpretation of the results.

AquaCrop is a water-driven simulation model that was developed by FAO. The model uses the water production function approach (Doorenbos and Kassam 1979) as a starting point, and evolves from it by calculating the crop biomass based on the amount of water transpired, and the crop yield as the proportion of biomass that goes into the harvestable parts. One of the main characteristics of AquaCrop is the separation of the final yield (Y) into biomass (B) and harvest index (HI), which allows the partitioning of the corresponding functional relations as response to environmental conditions (Steduto et al. 2012):

$$Y = HI \times B \quad (2)$$

The model requires a relatively small number of parameters and input data to simulate the yield response to water for most of the major field and vegetable crops. However, it produces a significant number of output data, including the simulation of canopy cover, biomass and soil water components over the entire growing cycle, and the final harvestable yield (Steduto et al. 2009; Raes et al. 2009). AquaCrop has been used in several applications around the globe (e.g., Andarzian et al. 2011; Abedinpour et al. 2012; Mkhabela and Bullock 2012; Kumar et al. 2014), and its simulation results appear to be satisfactory, despite the simplifications introduced in the model.

In this study, version 4.0 of the model is used. In regard to the recommended values provided for the crop parameters, they are estimated by a calibration – validation process of AquaCrop using experimental data (Raes et al. 2012).

2.4 Basic Notions on the Selection of Linear Models

The assessment-prediction of wheat yield in this study is achieved through linear regression modelling. A multiple linear regression model is described by the following equation, which can be applied to predict a dependent variable y , using a set of independent variables x_j :

$$y = b_0 + \sum_{j=1}^k b_j x_j + \varepsilon \quad j = 1(1)k \tag{3}$$

where b_0 is the intercept, k is the number of independent variables, b_j is the corresponding regression coefficients and ε is the residual error.

It is evident that modelling of a physical process cannot be based solely on statistical criteria, since this would be vulnerable to over-fitting issues, deteriorating the ability to generalise the model. It is important to keep a more inspective approach, taking also into account the physical processes and their underlying interactions. This approach may lead to a more meaningful type of model, rather than a black-box approach.

Stepwise regression can be used to find the most parsimonious sets of predictors that are most effective in predicting the dependent variable (Hocking 1976). A semi-automatic procedure is performed for the selection of independent variables to maximize the model’s prediction efficiency. This procedure may assist in understanding the statistical behaviour of the variables and in identifying possible problems (e.g., collinearity issues). However, the produced results can be biased, as the same data sets are used both to formulate the model and evaluate its goodness of fit (Chatfield 1995).

In order to validate the goodness of fit of the model with an unbiased estimate of its performance and to avoid possible over-fitting effects, cross-validation techniques can be used (Picard and Cook 1984). The main concept of cross-validation is based on the partitioning of the available dataset into subsets, using one subset for calibrating the model, while the other is used for validation.

There are several types of cross-validation, for instance the split-sample, the random sub-sampling, the K-fold and the Leave-One-Out (Arlot and Celisse 2010). In this paper, the K-fold cross-validation is used, in which the dataset is partitioned into K parts. The K-1 parts are used for calibration, while the remaining one part is used for validation purposes. This process is repeated K times. The cross-validation criterion is the average, over each repetition, of the estimates of discrepancy between the dataset and the fitted model (Browne 2000; Zucchini 2000). The main advantage of this approach is that all the data are utilised for both calibration and validation phases of the model.

The criteria that are used for the evaluation of the performance of the models are (Allen 1974; Hocking 1976; Willmott et al. 2012):

- The Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{4}$$

- The Route Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{5}$$

- The Coefficient of Determination (R^2)

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y}) (\hat{y}_i - \bar{\hat{y}}) \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 (\hat{y}_i - \bar{\hat{y}})^2} \quad (6)$$

- The Index of Agreement (d)

$$d = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n \left(\left| \hat{y}_i - \bar{y} \right| + \left| y_i - \bar{y} \right| \right)^2} \quad (7)$$

- The Predicted Residual Sum of Squares (PRESS)

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

where y is the observed value, \hat{y} is the predicted value, \bar{y} and $\bar{\hat{y}}$ are the mean of the observed and the predicted values, respectively, and n is the number of observations.

From the above criteria, *MAE* and *RMSE* are error measures used to represent the average differences between model predicted and observed values. The coefficient of determination (R^2) describes the proportion of the total variance in the observed data that can be explained by the model. The index of agreement (d) is the ratio between the mean square error and the potential error, and measures the degree to which the observed data are approached by the predicted data (Quiring and Papakyriakou 2003). *PRESS* is the sum of squares of differences between the observed and predicted values.

3 Results and Discussion

3.1 Drought Characteristics and Wheat Yield

Meteorological data of monthly precipitation and average monthly mean, maximum and minimum temperature were provided by the Hellenic Meteorological Service, for two agricultural areas of central and northern Greece (Fig. 1): Larissa (47 years) and Alexandroupolis (50 years).

The PET was estimated using the Hargreaves method (Hargreaves and Samani 1982) and the RDI values were calculated using DrinC software (Tigkas et al. 2013b, 2014). The main climatic characteristics of each area are presented in Table 1. As known, the aridity index ($\bar{\alpha}_{12}$) is calculated using Eq. (1).

The annual drought conditions for the two areas, based on RDI_{st} classification, can be briefly characterised as follows:

- In Larissa, 15 drought years (31.9 %) were observed, from which 2 are characterised as severe and 1 as extreme drought. Regarding the drought



Fig. 1 The location of the study areas (Larissa and Alexandroupolis)

- persistence, there were 5 cases with 2 consecutive years of drought and 1 case with 3 consecutive years of drought.
- In Alexandroupolis, there were 15 drought years (30 %), from which 1 is characterised as severe and 3 as extreme droughts. There were 5 cases with 2 consecutive years of drought and 2 cases with 3 consecutive years of drought.

The annual drought characterisation can provide a general image of the conditions of the area. However, the seasonal variations of drought may play a much more important role in the study of agricultural production. In Fig. 2, for instance, it can be seen that for the area of Larissa, in some years the drought conditions are noticeably different if the reference applies to the entire year or the 6-month winter/spring period (in which the main development of wheat takes place).

The data of each area were imported to the AquaCrop model for the simulation of wheat yield. The main outputs of the model appear in Table 2. It should be noted that since RDI takes

Table 1 Main climatic characteristics of the study areas (annual values)

	Precipitation (mm)				PET (mm)				Tmean (°C)				$\bar{\alpha}_{12}$ (aridity index)
	aver.	max.	min.	st.dev.	aver.	max.	min.	st.dev.	aver.	max.	min.	st.dev.	
Larissa	418	764	216	110	1263	1350	1190	40	15.8	17.3	14.8	0.6	0.33
Alexandroupolis	547	923	277	140	1081	1142	1022	28	15.0	17.1	13.6	0.7	0.51

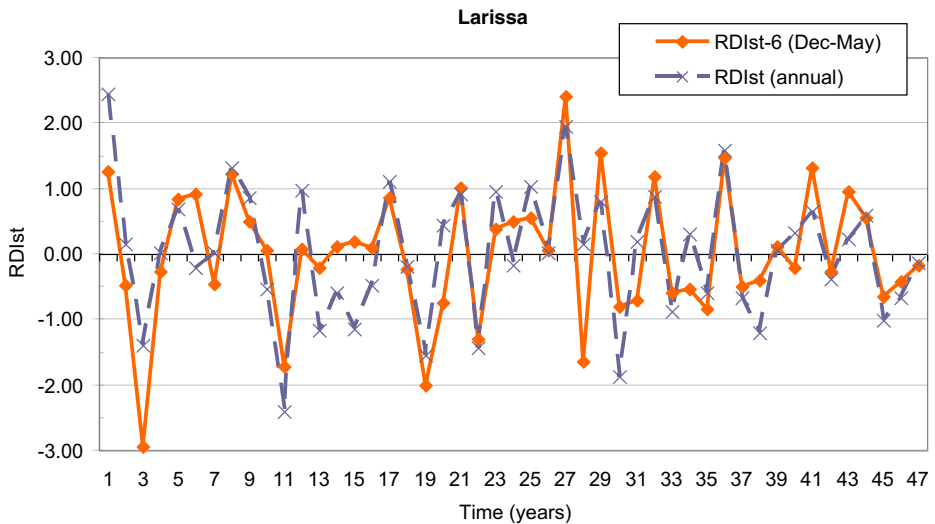


Fig. 2 The seasonal RDI_{st-6} (Dec–May) and the annual RDI_{st} for the area of Larissa

into account only meteorological parameters, other parameters of the model, such as soil characteristics, field management, etc., were set so that they have a neutral effect (not limiting factors) to the crop production:

- no adjustment of biomass was considered in regard to soil fertility stress,
- no surface mulches were considered (0 % cover),
- for the field surface, runoff occurrence was considered and no soil bunds were applied,
- the soil type was considered as sandy clay,
- no shallow groundwater table was considered.

The calculated values of the RDI_{st-6} (Dec–May) and the simulated wheat yield for the 2 study areas are presented in Fig. 3.

3.2 Model Formulation

From the aforementioned characteristics of wheat, we can identify the most important periods regarding the crop water requirements for the Mediterranean conditions. These are from January to March, with gradually increasing water demand, April with a peak on water demand and May with gradually decreased requirements.

Table 2 Simulated wheat yield and biomass with AquaCrop model for each study area

	Average annual biomass (ton/ha)	Average annual yield (ton/ha)	Average value of Harvest Index (%)
Larissa	9.675	2.808	26.2
Alexandroupolis	10.853	3.829	33.0

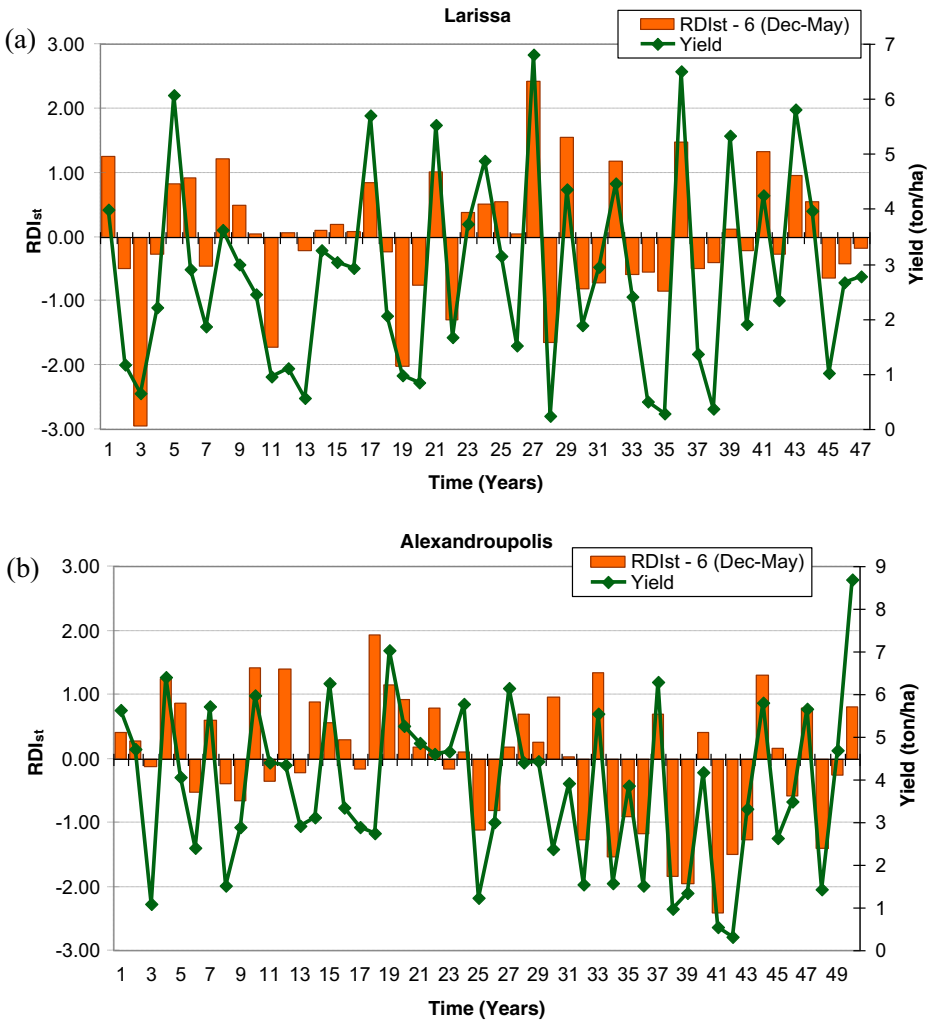


Fig. 3 The RDI_{st-6} (Dec–May) and the simulated wheat yield for: **a** Larissa; and **b** Alexandroupolis

The first step for model formulation is to identify the correlation of yield with each one of the independent variables (RDI_{st} for various reference periods) through simple linear regression. In addition to the RDI, the effect of the average monthly minimum temperatures of the winter months is also tested.

In Table 3, the coefficients of determination between wheat yield and various periods of RDI_{st} and monthly average T_{min} are presented. The annual RDI_{st} (RDI_{st-12}) and the 9-months RDI_{st} (RDI_{st-9} , October to June) seem to have a satisfactory correlation with the yield. However, the RDI_{st-6} for the period December to May, covering the most critical period of crop development, has a significantly better correlation. The RDI_{st-3} for the periods December to February and March to May, have also a good correlation for both areas of study. Regarding the monthly RDI_{st} (RDI_{st-1}), March and April have good correlation for both areas, while February exhibits also high correlation for

Table 3 Coefficients of determination (R^2) between yield and various RDI_{st} and monthly average T_{min}

		R^2	
		Larissa	Alexandroupolis
RDI_{st-12}	Oct–Sep	0.375	0.376
RDI_{st-9}	Oct–Jun	0.342	0.370
RDI_{st-6}	Oct–Mar	0.190	0.026
	Nov–Apr	0.443	0.322
	Dec–May	0.640	0.509
	Jan–Jun	0.504	0.437
RDI_{st-3}	Oct–Dec	0.005	0.038
	Nov–Jan	0.024	0.094
	Dec–Feb	0.385	0.345
	Jan–Mar	0.431	0.255
	Feb–Apr	0.653	0.294
	Mar–May	0.524	0.414
	Apr–Jun	0.128	0.180
RDI_{st-1}	Oct	0.000	0.000
	Nov	0.004	0.004
	Dec	0.012	0.135
	Jan	0.113	0.150
	Feb	0.341	0.073
	Mar	0.277	0.205
	Apr	0.283	0.191
	May	0.085	0.081
	Jun	0.036	0.002
T_{min}	Dec	0.088	0.130
	Jan	0.138	0.312
	Feb	0.102	0.327
	Mar	0.018	0.143

Larissa. It should also be noted that T_{min} for January and February have also good correlation, especially in the case of Alexandroupolis.

The second step is to use the stepwise technique to identify the best fitted set of parameters for a multiple linear regression model. Several model structures are tested, based on different variables sets for several prediction times, from 1 to 3 months before harvest (Table 4). For each model, a version including only RDI variables (model version a) and a version including also T_{min} variables (model version b) were tested. The selected model structures for each candidate set of variables resulted using as criterion the probability (significance) of the F statistic: the variable is incorporated into the model if the probability of F is less than or equal to 0.05, and it is removed from the model if the probability of F is greater than 0.10. The correlation and standard error of estimate for each model appear in Table 5.

For the third step, the evaluation of each model, a 10-fold cross-validation approach was applied for all tested models. In Table 6, the results of the evaluation

Table 4 Tested model structures with various candidate variable sets, including RDI_{st} and average monthly minimum temperature for January and February (Tmin)

#	Prediction time (end of month)	Candidate variables sets	Selected model structures	
			Larissa	Alexandroupolis
1a	May	All RDI	RDI-3 (Feb–Apr), RDI-6 (Dec–May)	RDI-6 (Dec–May), RDI-4 (Dec–Mar)
1b		+ Tmin	RDI-3 (Feb–Apr), RDI-6 (Dec–May), Tmin (Feb)	RDI-6 (Dec–May), Tmin (Jan), RDI-3 (Dec–Feb), Tmin (Feb), RDI-1 (May)
2a	April	All RDI (Nov to Apr)	RDI-3 (Feb–Apr)	RDI-5 (Dec–Apr), RDI-1 (Apr)
2b		+ Tmin	RDI-3 (Feb–Apr), Tmin (Feb)	Tmin (Jan), RDI-2 (Mar–Apr), T (Feb), RDI-1 (Dec)
3a	March	All RDI (Nov to March)	RDI-2 (Feb–Mar)	RDI-4 (Dec–Mar)
3b		+ Tmin	RDI-2 (Feb–Mar)	Tmin (Jan), T (Feb), RDI-1 (Mar)
4a	May	RDI-1 (Nov to May)	RDI-1 (Feb), RDI-1 (Apr), RDI-1 (Mar)	RDI-1 (Mar), RDI-1 (Apr), RDI-1 (May), RDI-1 (Dec), RDI-1 (Jan)
4b		+ Tmin	RDI-1 (Feb), RDI-1 (Apr), RDI-1 (Mar), Tmin (Feb), Tmin (Jan)	Tmin (Feb), RDI-1 (May), Tmin (Jan), RDI-1 (Mar), RDI-1 (Apr)
5a	April	RDI-1 (Nov to April)	RDI-1 (Feb), RDI-1 (Apr), RDI-1 (Mar)	RDI-1 (Mar), RDI-1 (Apr), RDI-1 (Dec), RDI-1 (Feb)
5b		+ Tmin	RDI-1 (Feb), RDI-1 (Apr), RDI-1 (Mar), Tmin (Feb)	Tmin (Feb), Tmin (Jan), RDI-1 (Mar), RDI-1 (Apr), RDI-1 (Dec)
6a	March	RDI-1 (Nov to March)	RDI-1 (Feb), RDI-1 (Mar)	RDI-1 (Mar), RDI-1 (Jan), RDI-1 (Dec)
6b		+ Tmin	RDI-1 (Feb), RDI-1 (Mar)	Tmin (Feb), Tmin (Jan), RDI-1 (Mar)
7a	May	RDI-2 (Nov–Dec), RDI-3 (Jan–Mar), RDI-1 (Apr), RDI-1 (May)	RDI-3 (Jan–Mar), RDI-1 (Apr), RDI-1 (May)	RDI-3 (Jan–Mar), RDI-1 (Apr), RDI-1 (May)
7b		+ Tmin	RDI-3 (Jan–Mar), RDI-1 (Apr), RDI-1 (May), Tmin (Feb), Tmin (Jan)	Tmin (Feb), RDI-1 (May), Tmin (Jan), RDI-1 (Apr), RDI-3 (Jan–Mar)
8a	April	RDI-2 (Nov–Dec), RDI-3 (Jan–Mar), RDI-1 (Apr)	RDI-3 (Jan–Mar), RDI-1 (Apr)	RDI-3 (Jan–Mar), RDI-1 (Apr)
8b		+ Tmin	RDI-3 (Jan–Mar), RDI-1 (Apr), Tmin (Feb)	Tmin (Feb), Tmin (Jan), RDI-1 (Apr), RDI-3 (Jan–Mar)
9a	March	RDI-2 (Nov–Dec), RDI-3 (Jan–Mar)	RDI-3 (Jan–Mar)	RDI-3 (Jan–Mar)

Table 4 (continued)

#	Prediction time (end of month)	Candidate variables sets	Selected model structures	
			Larissa	Alexandroupolis
9b		+ Tmin	RDI-3 (Jan–Mar)	Tmin (Feb), Tmin (Jan), RDI-3 (Jan–Mar)
10a	May	RDI-3 (Dec–Feb), RDI-1 (Mar), RDI-1 (Apr), RDI-1 (May)	RDI-3 (Dec–Feb), RDI-1 (Apr), RDI-1 (Mar), RDI-1 (May)	RDI-3 (Dec–Feb), RDI-1 (Apr), RDI-1 (May), RDI-1 (Mar)
10b		+ Tmin	RDI-3 (Dec–Feb), RDI-1 (Apr), T (Jan), RDI-1 (May), RDI-1 (Mar), Tmin (Feb)	Tmin (Feb), RDI-1 (May), Tmin (Jan), RDI-1 (Mar), RDI-1 (Apr)
11a	April	RDI-3 (Dec–Feb), RDI-1 (Mar), RDI-1 (Apr), RDI-1 (May)	RDI-3 (Dec–Feb), RDI-1 (Apr), RDI-1 (Mar)	RDI-3 (Dec–Feb), RDI-1 (Apr), RDI-1 (Mar)
11b		+ Tmin	RDI-3 (Dec–Feb), RDI-1 (Apr), Tmin (Jan), RDI-1 (Mar)	Tmin (Feb), T (Jan), RDI-1 (Mar), RDI-1 (Apr)
12a	March	RDI-3 (Dec–Feb), RDI-1 (Mar), RDI-1 (Apr), RDI-1 (May)	RDI-3 (Dec–Feb), RDI-1 (Mar)	RDI-3 (Dec–Feb), RDI-1 (Mar)
12b		+ Tmin	RDI-3 (Dec–Feb), RDI-1 (Mar)	Tmin (Feb), T (Jan), RDI-1 (Mar)

Table 5 Correlation coefficients (R), adjusted coefficients of determination (R²) and standard error of the estimate for the tested model structures based on the stepwise regression analysis

Model	R		Adjusted R ²		Standard error	
	Larissa	Alexan.	Larissa	Alexan.	Larissa	Alexan.
1a	0.841	0.766	0.694	0.570	0.998	1.257
1b	0.863	0.909	0.727	0.806	0.942	0.843
2a	0.808	0.697	0.646	0.463	1.075	1.404
2b	0.836	0.818	0.686	0.639	1.012	1.151
3a	0.670	0.587	0.437	0.331	1.354	1.568
3b	0.670	0.777	0.437	0.577	1.354	1.246
4a	0.771	0.772	0.565	0.551	1.190	1.285
4b	0.863	0.913	0.706	0.815	0.979	0.823
5a	0.771	0.701	0.565	0.447	1.190	1.426
5b	0.806	0.822	0.617	0.639	1.118	1.152
6a	0.668	0.624	0.420	0.350	1.374	1.546
6b	0.668	0.777	0.420	0.577	1.374	1.246
7a	0.807	0.715	0.627	0.479	1.102	1.383
7b	0.861	0.887	0.711	0.762	0.971	0.934
8a	0.772	0.626	0.577	0.367	1.174	1.525
8b	0.801	0.772	0.617	0.559	1.118	1.272
9a	0.657	0.505	0.419	0.240	1.376	1.671
9b	0.657	0.729	0.419	0.501	1.376	1.354
10a	0.798	0.763	0.602	0.546	1.139	1.292
10b	0.867	0.913	0.714	0.815	0.966	0.823
11a	0.759	0.700	0.546	0.457	1.216	1.413
11b	0.788	0.802	0.585	0.611	1.163	1.195
12a	0.622	0.603	0.359	0.336	1.446	1.562
12b	0.622	0.777	0.359	0.577	1.446	1.246

criteria for the cross-validation are presented, which are based only on the subset of data that were not utilised in the formulation process of the model for each iteration. This allows an unbiased interpretation of the predictive potential of each selected set of variables.

From the results, it can be seen that there is a very satisfactory prediction of the yield 1 month before harvest (end of May) with almost all model structures, for both study areas. The estimation remains satisfactory even for 2 months before harvest (end of April). For the time frame of 3 months before harvest (end of March) the predictive capacity of the models is decreased considerably, but it is still within acceptable limits. Also, in most of the cases, there is a positive effect on the accuracy of the models when the T_{min} of January and February is added in the candidate sets of variables.

The models #1a and #1b exhibit a very good performance, though, it should be noted that the selected RDI variables sets include overlapping time periods, e.g., RDI_{st-6} (Dec–May) and RDI_{st-3} (Feb–Apr), so they are sharing a substantial amount of information. Therefore, despite the fact that these variables are linearly independent

Table 6 The performance criteria based on the cross-validation process

Model	MAE		RMSE		d		R ²		PRESS	
	Larissa	Alexan.	Larissa	Alexan.	Larissa	Alexan.	Larissa	Alexan.	Larissa	Alexan.
1a	0.828	1.070	0.966	1.273	0.866	0.769	0.729	0.562	49.115	85.270
1b	0.756	0.814	0.936	1.003	0.884	0.880	0.773	0.782	45.474	53.080
2a	0.853	1.108	1.023	1.365	0.837	0.724	0.667	0.496	55.441	104.684
2b	0.772	1.268	0.982	1.498	0.860	0.690	0.735	0.493	50.640	118.645
3a	1.112	1.254	1.308	1.526	0.745	0.599	0.549	0.467	88.445	128.413
3b	1.112	1.045	1.308	1.209	0.745	0.818	0.549	0.647	88.445	82.665
4a	0.990	1.129	1.190	1.333	0.751	0.744	0.566	0.546	72.145	95.086
4b	0.884	0.738	1.069	0.844	0.802	0.935	0.627	0.854	58.823	38.683
5a	0.990	1.191	1.190	1.427	0.751	0.718	0.566	0.474	72.145	110.159
5b	0.911	1.006	1.122	1.189	0.787	0.831	0.656	0.695	64.937	77.035
6a	1.147	1.292	1.328	1.536	0.731	0.625	0.513	0.518	90.960	129.693
6b	1.147	1.045	1.328	1.209	0.731	0.818	0.513	0.647	90.960	82.665
7a	0.964	1.192	1.112	1.381	0.782	0.673	0.652	0.524	68.439	104.670
7b	0.864	0.836	1.026	0.970	0.802	0.904	0.620	0.780	54.582	49.134
8a	0.949	1.205	1.119	1.476	0.791	0.664	0.601	0.429	67.961	119.195
8b	0.859	1.085	1.070	1.264	0.813	0.807	0.639	0.588	61.065	88.370
9a	1.114	1.387	1.305	1.651	0.747	0.486	0.552	0.354	90.219	144.004
9b	1.114	1.175	1.305	1.312	0.747	0.751	0.552	0.531	90.219	98.056
10a	1.031	1.145	1.198	1.315	0.759	0.729	0.596	0.518	73.652	93.116
10b	0.910	0.754	1.063	0.860	0.804	0.932	0.614	0.851	56.618	39.859
11a	1.006	1.129	1.210	1.382	0.767	0.722	0.560	0.489	73.481	106.576
11b	0.936	0.993	1.120	1.191	0.786	0.849	0.598	0.679	67.311	78.119
12a	1.187	1.282	1.439	1.555	0.697	0.583	0.513	0.421	101.499	131.882
12b	1.187	1.045	1.439	1.209	0.697	0.818	0.513	0.647	101.499	82.665

in the specific cases ($p < 0.05$), these sets may not secure the generalisation of the corresponding model structures.

The rest of the models provide similar results, for the same time frame of prediction (March, April or May). The most important month for RDI, based on the significance of the variables in the models, is April, followed by March, May, January, February and December.

The use of RDI-1 for all months gives good results for both study areas. However, the use of RDI-3 for important periods of crop development (e.g., Feb–Mar) seems to be also efficient.

For illustration purposes, the observed and predicted wheat yields and the corresponding diagrams of the coefficients of determination are presented in Figs. 4 and 5, respectively, based on model #4b for the two study areas. The #4b model structure has the following mathematical expression for the two study areas, respectively:

$$\text{Larissa : } \hat{y} = 2.361 + 0.544x_1 + 0.664x_2 + 0.636x_3 + 0.343x_4 + 0.253x_5 + 0.212x_6 \quad (11)$$

$$\text{Alexandroupolis : } \hat{y} = 2.745 + 0.856x_1 + 0.499x_2 + 0.616x_3 + 0.308x_5 + 0.426x_6 \quad (12)$$

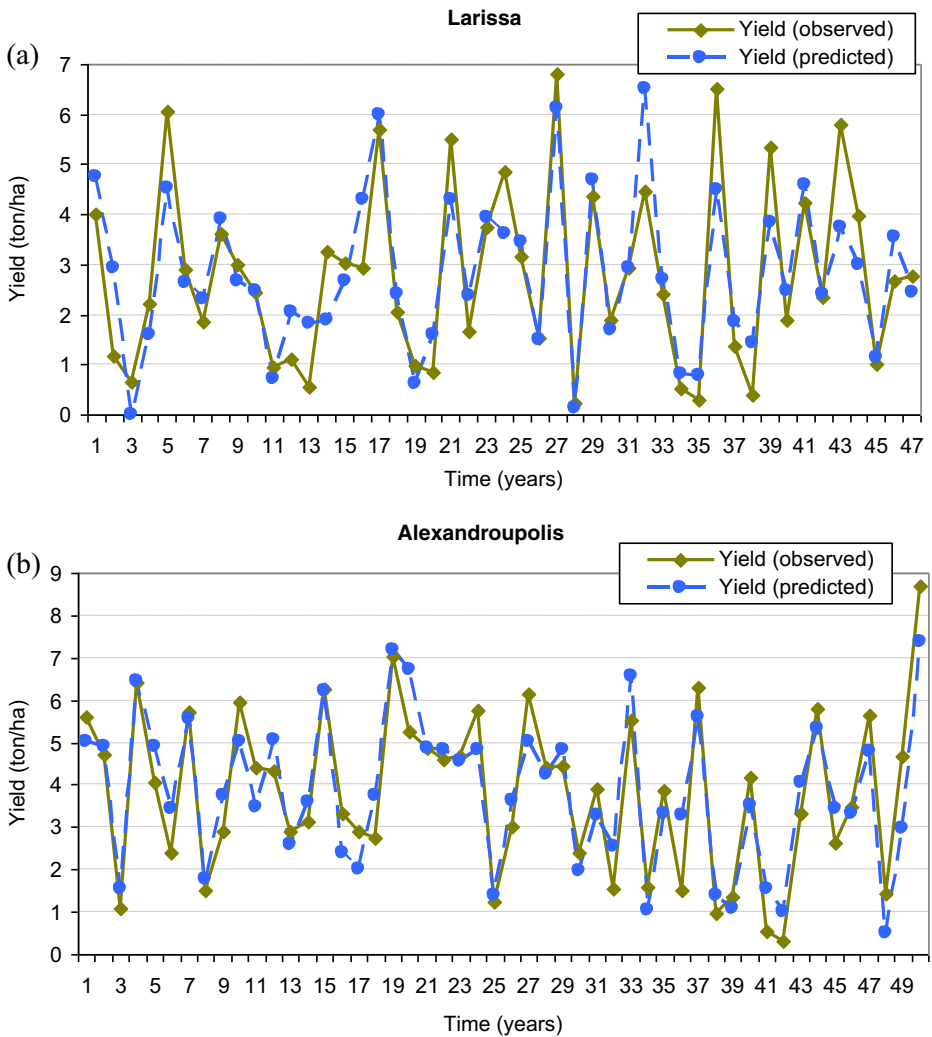


Fig. 4 Observed and predicted (based on model 4b) wheat yield for: **a** Larissa; and **b** Alexandroupolis

where, \hat{y} is the predicted yield, x_1 is RDI_{st-1} (May), x_2 is RDI_{st-1} (Apr), x_3 is RDI_{st-1} (Mar), x_4 is RDI_{st-1} (Feb), x_5 is T_{min} (Feb) and x_6 is T_{min} (Jan).

4 Concluding Remarks

A methodology for the early estimation (prediction) of drought impacts on the expected winter wheat yield is presented. The methodology is tested in two areas with Mediterranean climate. The proposed approach can be useful for operational use, even for areas with limited data availability. The required data are only the monthly precipitation and temperature.

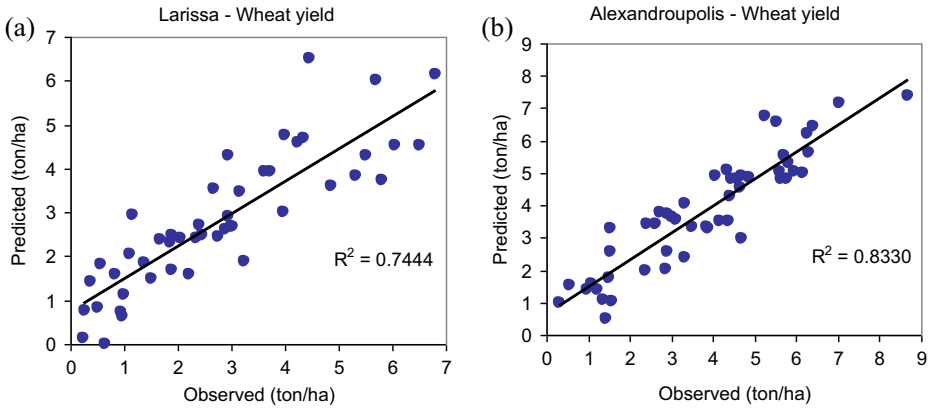


Fig. 5 Comparison of predicted by model 4b wheat yields with observed ones, and coefficient of determination (R^2) obtained from this model, for: **a** Larissa; and **b** Alexandroupolis

The methodology is based on the formulation of linear regression models for the prediction of wheat yield using RDI (of various reference periods) as the key independent variable. The performance of the models was evaluated with a cross-validation process.

From the single linear regression results, it can be concluded that wheat yield has high correlation with RDI_{st-6} (Dec–May), which corresponds to the main development phase of the crop. Further, various multiple regression model structures were tested, for different candidate sets of variables and prediction time frames from 1 to 3 months before harvest. The prediction was very successful for 1 and 2 months before harvest. However, the models could still give reasonable estimates of the final yield for predictions 3 months before harvest. The correlation coefficients in both study areas between the observed and predicted values reached 0.87 and 0.91 for predictions 1 month before harvest, 0.84 and 0.82 for 2 months and 0.67 and 0.77 for 3 months before harvest. Additionally, these results are supported by the fact that the evaluation through the cross-validation process confirmed the stability of the tested models.

Model structures with monthly RDI values (RDI_{st-1}), mainly from January to May, performed very successfully for both areas studied, though the use of longer reference periods for RDI_{st} (e.g., RDI_{st-3}), including important months in crop development, may improve the performance of the model in most cases. Further, the incorporation of the average monthly minimum temperature of January and February seemed to enhance the predictive ability of the models.

All the above lead to the conclusion that RDI can be successfully used in the early estimation of drought impacts on the yield of the winter wheat in Mediterranean climate, which is very important for an early reconnaissance estimation of yield losses due to drought, by organisations and stakeholders. Such information is also very useful in the planning against drought in the agricultural sector and in taking supporting measures (e.g., timely arrangement of imports) ensuring food security. Finally, it is useful for the activities of insurance companies and related organisations for covering the production losses of the affected farmers. Further research on the subject may be useful for extending these outcomes to other climatic zones or different rainfed crops.

References

- Abedinpour M, Sarangi A, Rajput TBS, Singh M, Pathak H, Ahmad T (2012) Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. *Agric Water Manag* 110:55–66
- Acevedo EH, Silva PC, Silva HR, Solar BR (1999) Wheat production in Mediterranean environments. In: Slafer GA, Satorre EH (eds) *Wheat: ecology and physiology of yield determination*. Haworth Press, New York, pp 298–331
- Allen DM (1974) The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 16:125–127
- Allen R, Pereira LS, Raes D, Smith M (1998) *Crop evapotranspiration – Guidelines for computing crop water requirements*. FAO Irrigation and Drainage Paper No 56, Rome
- Andarzian B, Bannayan M, Steduto P, Mazraeh H, Barati ME, Barati MA, Rahnam A (2011) Validation and testing of the AquaCrop model under full and deficit irrigated wheat production in Iran. *Agric Water Manag* 100:1–8
- Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Stat Surv* 4:40–79
- Bordi I, Sutera A (2008) Drought over Europe in recent years. In: López-Francos A (ed) *Drought management: Scientific and technological innovations*. CIHEAM, Options Méditerranéennes: Série A Séminaires Méditerranéens; n 80, Zaragoza, pp 63–68
- Browne MW (2000) Cross-validation methods. *J Math Psychol* 44:108–132
- Chatfield C (1995) Model uncertainty, data mining and statistical inference. *J R Stat Soc A* 158(3):419–466
- Dalezios NR, Loukas A, Bampzelis D (2002) The role of agrometeorological and agrohydrological indices in the phenology of wheat in central Greece. *Phys Chem Earth* 27:1019–1023
- Doorenbos J, Kassam AH (1979) *Yield response to water*. FAO Irrigation and Drainage Paper No 33, Rome
- Fischer RA (1973) The effect of water stress at various stages of development on yield processes in wheat. In: *Plant response to climatic factors*, UNESCO, Paris, pp 233–241
- Hargreaves GH, Samani ZA (1982) Estimating potential evapotranspiration. *J Irrig Drain Div* 108(3):225–230
- Hocking RR (1976) A Biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics* 32:1–49
- Kazmi DH, Rasul G (2012) Agrometeorological wheat yield prediction in rainfed Potohar region of Pakistan. *Agric Sci* 3(2):170–177
- Kogan F, Kussul N, Adamenko T, Skakun S et al (2013) Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *Int J Appl Earth Obs Geoinf* 23:192–203
- Kristensen K, Schelde K, Olesen JE (2011) Winter wheat yield response to climate variability in Denmark. *J Agric Sci* 149:33–47. doi:10.1017/S0021859610000675
- Kumar V (1998) An early warning system for agricultural drought in an arid region using limited data. *J Arid Environ* 40:199–209
- Kumar V, Panu U (1997) Predictive assessment of severity of agricultural droughts based on agro-climatic factors. *J Am Water Resour Assoc* 33:1255–1264
- Kumar P, Sarangi A, Singh DK, Parihar SS (2014) Evaluation of AquaCrop model in predicting wheat yield and water productivity under irrigated saline regimes. *Irrig Drain* 63:474–487. doi:10.1002/ird.1841
- Lee B-H, Kenkel P, Brorsen BW (2013) Pre-harvest forecasting of county wheat yield and wheat quality using weather information. *Agric For Meteorol* 168:26–35
- Leilah AA, Al-Khateeb SA (2005) Statistical analysis of wheat yield under drought conditions. *J Arid Environ* 61:483–496
- Liakatas A (1997) Growth and yield of Rainfed wheat on the seasonally dry Aegean Islands. *Theor Appl Climatol* 58:43–56
- Mavromatis T (2007) Drought index evaluation for assessing future wheat production in Greece. *Int J Climatol* 27:911–924
- Mkhabela MS, Bullock PR (2012) Performance of the FAO AquaCrop model for wheat grain yield and soil moisture simulation in Western Canada. *Agric Water Manag* 110:16–24
- Picard RR, Cook RD (1984) Cross-validation of regression models. *J Am Stat Assoc* 79(387):575–583
- Porter JR, Gawith M (1999) Temperatures and the growth and development of wheat: a review. *Eur J Agron* 10: 23–36
- Proutsos N, Tsagari K, Karetzos G, Liakatas A, Kritikos T (2010) Recent temperature trends over Mountainous Greece. *Eur Water* 32:15–23
- Quiring SM, Papakyriakou TN (2003) An evaluation of agricultural drought indices for the Canadian prairies. *Agric For Meteorol* 118:49–62
- Raes D, Steduto P, Hsiao TC, Fereres E (2009) AquaCrop - The FAO crop model for predicting yield response to water: II. Main algorithms and software description. *Agron J* 101:438–447

- Raes D, Steduto P, Hsiao TC, Fereres E (2012) AquaCrop ver 4.0 – reference manual. FAO Land and Water Division, Rome
- Sadat Noori SM, Liaghat AM, Ebrahimi K (2012) Prediction of crop production using drought indices at different time scales and climatic factors to manage drought risk. *J Am Water Resour Assoc* 48:1–9
- Salman AZ, Al-Karablieh EK (2001) An early warning system for wheat production in low rainfall areas of Jordan. *J Arid Environ* 49:631–642
- Shimshi D (1973) The irrigation of wheat. In: Yaron B et al (eds) *Arid zone irrigation*. Springer Verlag, Berlin, pp 369–373
- Steduto P, Hsiao TC, Raes D, Fereres E (2009) AquaCrop - The FAO crop model for predicting yield response to water: I. Concepts and underlying principles. *Agron J* 101:426–437
- Steduto P, Hsiao TC, Fereres E, Raes D (2012) Crop yield response to water. FAO Irrigation and drainage paper No 66, Rome
- Tigkas D (2008) Drought characterisation and monitoring in regions of Greece. *Eur Water* 23(24):29–39
- Tigkas D, Vangelis H, Tsakiris G (2012) Drought and climatic change impact on streamflow in small watersheds. *Sci Total Environ* 440:33–41
- Tigkas D, Vangelis H, Tsakiris G (2013a) The RDI as a composite climatic index. *Eur Water* 41:17–22
- Tigkas D, Vangelis H, Tsakiris G (2013b) The drought indices calculator (DrinC). In: Maia R et al. (eds) *Proceedings of 8th International Conference of EWRA 'Water Resources Management in an Interdisciplinary and Changing Context'*, 26–29 June 2013, Porto, Portugal, pp 1333–1342
- Tigkas D, Vangelis H, Tsakiris G (2014) DrinC: a software for drought analysis based on drought indices. *Earth Sci Inform*. doi:10.1007/s12145-014-0178-y
- Tsakiris G, Tigkas D (2007) Drought risk in agriculture in Mediterranean Regions - Case Study: Eastern Crete. In: Rossi et al (eds) *Methods and tools for drought analysis and management*. Springer, The Netherlands, pp 399–414
- Tsakiris G, Vangelis H (2005) Establishing a drought index incorporating evapotranspiration. *Eur Water* 9(10):3–11
- Tsakiris G, Pangelou D, Vangelis H (2007) Regional drought assessment based on the Reconnaissance Drought Index (RDI). *Water Resour Manag* 21(5):821–833
- Tsakiris G, Nalbantis I, Pangelou D, Tigkas D, Vangelis H (2008) Drought meteorological monitoring network design for the reconnaissance drought index (RDI). In: Franco-Lopez A (ed) *Proceedings of the 1st International Conference "Drought Management: scientific and technological innovations"*, Zaragoza, Spain: Option Méditerranéennes, Series A, No 80, 12–14 June 2008, pp 57–62
- Tsakiris G, Vangelis H, Tigkas D (2010) Drought impacts on yield potential in rainfed agriculture. In: *Proceedings of 2nd International Conference on Drought Management 'Economics of Drought and Drought Preparedness in a Climate Change Context'*, 4–6 March 2010, Istanbul, Turkey, pp 191–197
- Tsakiris G, Nalbantis I, Vangelis H, Verbeiren B, Huysmans M, Tychon B, Jacquemin I, Canters F, Vanderhaegen S, Engelen G, Poelmans L, De Becker P, Batelaan O (2013) A system-based paradigm of drought analysis for operational management. *Water Resour Manag* 27(15):5281–5297
- Turner NC (1997) Further progress in crop water relations. *Adv Agron* 58:293–338
- UNEP (1992) *World atlas of desertification*. Edward Arnold, London
- Vangelis H, Tigkas D, Tsakiris G (2013) The effect of PET method on Reconnaissance Drought Index (RDI) calculation. *J Arid Environ* 88:130–140
- Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. *Int J Climatol* 32(13): 2088–2094
- Zucchini W (2000) An introduction to model selection. *J Math Psychol* 44:41–61