

Prediction of Evapotranspiration in a Mediterranean Region Using Basic Meteorological Variables

Jato-Espino, D. and Charlesworth, S.

Post-print deposited in Coventry University repository

Original citation:

Jato-Espino, D. and Charlesworth, S. (2016) Prediction of Evapotranspiration in a Mediterranean Region Using Basic Meteorological Variables . Journal of Hydrologic Engineering. DOI: 10.1061/(ASCE)HE.1943-5584.0001485

[http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0001485](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001485)

©2016 American Society of Civil Engineers

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

Journal of Hydrologic Engineering

Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables --Manuscript Draft--

Manuscript Number:	HEENG-3310R1	
Full Title:	Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables	
Manuscript Region of Origin:	SPAIN	
Article Type:	Technical Paper	
Manuscript Classifications:	97.21000: Evapotraspiration	
Funding Information:	Ministerio de Economía y Competitividad (BIA2012-32463)	Not applicable
Abstract:	<p>A critical need for farmers, particularly those in arid and semi-arid areas is to have a reliable, accurate and reasonably accessible means of estimating the evapotranspiration rates of their crops in order to optimise their irrigation requirements. Evapotranspiration is a crucial process due to its influence on the precipitation that is returned to the atmosphere. The calculation of this variable often starts from the estimation of reference evapotranspiration, for which a variety of methods have been developed. However, these methods are very complex either theoretically and/or because of the large amount of parameters on which they are based, which makes the development of a simple and reliable methodology for the prediction of this variable important. This research combined three concepts such as cluster analysis, multiple linear regression and Voronoi diagrams to achieve that end. Cluster analysis divided the study area into groups based on its weather characteristics, whose locations were then delimited by drawing the Voronoi regions associated with them. Regression equations were built to predict daily reference evapotranspiration in each cluster using basic climatic variables produced in forecasts made by meteorological agencies. Finally, the Voronoi diagrams were used again to regionalise the crop coefficients and calculate evapotranspiration from the values of reference evapotranspiration derived from the regression models. These operations were applied to the Valencian Region (Spain), a Mediterranean area which is partly semi-arid and for which evapotranspiration is a critical issue. The results demonstrated the usefulness and accuracy of the methodology to predict the water demands of crops and hence enable farmers to plan their irrigation needs.</p>	
Corresponding Author:	Daniel Jato-Espino, MSc Universidad de Cantabria Santander, SPAIN	
Corresponding Author E-Mail:	jatod@unican.es	
Order of Authors:	Daniel Jato-Espino, MSc Susanne M Charlesworth, PhD Sara Perales-Momparler, PhD Ignacio Andrés-Doménech, PhD	
Suggested Reviewers:	David M Summer United States Geological Survey dmsummer@usgs.gov Expert in evapotranspiration. Sirisha Adamala Indian Institute of Technology Kharagpur sirisha@agfe.iitkgp.ernet.in Expert in evapotranspiration. Jingyu Li China Institute of Land Surveying and Planning	

	lijingyu@mail.clspi.org.cn Expert in evapotranspiration.
	Weiguang Wang Hohai University wangweiguang006@126.com Expert in irrigation.
	John D Valiantzas Geononiko Panepistimion Athinon ival@aua.gr Expert in irrigation.
Opposed Reviewers:	
Additional Information:	
Question	Response
Authors are required to attain permission to re-use content, figures, tables, charts, maps, and photographs for which the authors do not hold copyright. Figures created by the authors but previously published under copyright elsewhere may require permission. For more information see http://ascelibrary.org/doi/abs/10.1061/9780784479018.ch03 . All permissions must be uploaded as a permission file in PDF format. Are there any required permissions that have not yet been secured? If yes, please explain in the comment box.	No
ASCE does not review manuscripts that are being considered elsewhere to include other ASCE Journals and all conference proceedings. Is the article or parts of it being considered for any other publication? If your answer is yes, please explain in the comments box below.	No
Is this article or parts of it already published in print or online in any language? ASCE does not review content already published (see next questions for conference papers and posted theses/dissertations). If your answer is yes, please explain in the comments box below.	No
Has this paper or parts of it been published as a conference proceeding? A conference proceeding may be reviewed for publication only if it has been significantly revised and contains 50% new content. Any content overlap should be reworded and/or properly referenced. If your answer is yes, please explain in the comments box below and be prepared to provide the conference paper.	No
ASCE allows submissions of papers that	No

<p>are based on theses and dissertations so long as the paper has been modified to fit the journal page limits, format, and tailored for the audience. ASCE will consider such papers even if the thesis or dissertation has been posted online provided that the degree-granting institution requires that the thesis or dissertation be posted.</p> <p>Is this paper a derivative of a thesis or dissertation posted or about to be posted on the Internet? If yes, please provide the URL or DOI permalink in the comment box below.</p>	
<p>Each submission to ASCE must stand on its own and represent significant new information, which may include disproving the work of others. While it is acceptable to build upon one's own work or replicate other's work, it is not appropriate to fragment the research to maximize the number of manuscripts or to submit papers that represent very small incremental changes. ASCE may use tools such as CrossCheck, Duplicate Submission Checks, and Google Scholar to verify that submissions are novel. Does the manuscript constitute incremental work (i.e. restating raw data, models, or conclusions from a previously published study)?</p>	No
<p>Authors are expected to present their papers within the page limitations described in Publishing in ASCE Journals: A Guide for Authors. Technical papers and Case Studies must not exceed 30 double-spaced manuscript pages, including all figures and tables. Technical notes must not exceed 7 double-spaced manuscript pages. Papers that exceed the limits must be justified. Grossly over-length papers may be returned without review. Does this paper exceed the ASCE length limitations? If yes, please provide justification in the comments box below.</p>	No
<p>All authors listed on the manuscript must have contributed to the study and must approve the current version of the manuscript. Are there any authors on the paper that do not meet these criteria? If the answer is yes, please explain in the comments.</p>	No
<p>Was this paper previously declined or withdrawn from this or another ASCE</p>	No

<p>journal? If so, please provide the previous manuscript number and explain what you have changed in this current version in the comments box below. You may upload a separate response to reviewers if your comments are extensive.</p>	
<p>Companion manuscripts are discouraged as all papers published must be able to stand on their own. Justification must be provided to the editor if an author feels as though the work must be presented in two parts and published simultaneously. There is no guarantee that companions will be reviewed by the same reviewers, which complicates the review process, increases the risk for rejection and potentially lengthens the review time. If this is a companion paper, please indicate the part number and provide the title, authors and manuscript number (if available) for the companion papers along with your detailed justification for the editor in the comments box below. If there is no justification provided, or if there is insufficient justification, the papers will be returned without review.</p>	
<p>If this manuscript is intended as part of a Special Issue or Collection, please provide the Special Collection title and name of the guest editor in the comments box below.</p>	
<p>Recognizing that science and engineering are best served when data are made available during the review and discussion of manuscripts and journal articles, and to allow others to replicate and build on work published in ASCE journals, all reasonable requests by reviewers for materials, data, and associated protocols must be fulfilled. If you are restricted from sharing your data and materials, please explain below.</p>	
<p>Papers published in ASCE Journals must make a contribution to the core body of knowledge and to the advancement of the field. Authors must consider how their new knowledge and/or innovations add value to the state of the art and/or state of the practice. Please outline the specific contributions of this research in the comments box.</p>	<p>The aim of this paper is to build linear equations for the prediction of evapotranspiration (ET) based on weather forecasts, so that users can estimate the water requirements of their crops and determine when and how much to irrigate. This was achieved through a methodology which combined three tools such as cluster analysis, multiple linear regression models and Voronoi diagrams to enable the estimation and regionalisation of ET using basic meteorological variables.</p> <p>Previous research is limited to the estimation of reference evapotranspiration (ETo) and did not include any regionalisation methodology to group different locations according to their climatic characteristics, which precludes the calculation of ET and therefore the design of irrigation strategies. Furthermore, former approaches do not yield equations that can be used for predicting ETo with new data, which limits their applicability.</p>
<p>The flat fee for including color figures in print is \$800, regardless of the number of</p>	<p>No</p>

<p>color figures. There is no fee for online only color figures. If you decide to not print figures in color, please ensure that the color figures will also make sense when printed in black-and-white, and remove any reference to color in the text. Only one file is accepted for each figure. Do you intend to pay to include color figures in print? If yes, please indicate which figures in the comments box.</p>	
<p>If there is anything else you wish to communicate to the editor of the journal, please do so in this box.</p>	

Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables

Daniel Jato-Espino*¹, Susanne M. Charlesworth², Sara Perales-Momparler³,
Ignacio Andrés-Doménech⁴

¹ GITECO Research Group, Universidad de Cantabria, Av. de los Castros s/n, 39005 Santander, Spain

² Centre for Agroecology, Water and Resilience (CAWR), Coventry University, CV1 5FB, United Kingdom

³ Green Blue Management, Avda. del Puerto, 180-1B, 46023 Valencia, Spain

⁴ Instituto Universitario de Investigación de Ingeniería del Agua y Medio Ambiente (IIAMA), Universitat Politècnica de València, Cno. de Vera s/n, 46022 Valencia, Spain

E-mail addresses: jatod@unican.es (D. Jato-Espino), apx119@coventry.ac.uk (S. M. Charlesworth), sara.perales@greenbluemanagement.com (S. Perales-Momparler), igando@hma.upv.es (I. Andrés-Doménech),

* Corresponding author. Tel.: +34 942 20 39 43; Fax: +34 942 20 17 03.

Abstract

A critical need for farmers, particularly those in arid and semi-arid areas is to have a reliable, accurate and reasonably accessible means of estimating the evapotranspiration rates of their crops in order to optimize their irrigation requirements. Evapotranspiration

25 is a crucial process due to its influence on the precipitation that is returned to the atmos-
26 phere. The calculation of this variable often starts from the estimation of reference evap-
27 otranspiration, for which a variety of methods have been developed. However, these
28 methods are very complex either theoretically and/or because of the large amount of pa-
29 rameters on which they are based, which makes the development of a simple and reliable
30 methodology for the prediction of this variable important. This research combined three
31 concepts such as cluster analysis, Multiple Linear Regression (MLR) and Voronoi dia-
32 grams to achieve that end. Cluster analysis divided the study area into groups based on
33 its weather characteristics, whose locations were then delimited by drawing the Voronoi
34 regions associated with them. Regression equations were built to predict daily reference
35 evapotranspiration in each cluster using basic climate variables produced in forecasts
36 made by meteorological agencies. Finally, the Voronoi diagrams were used again to re-
37 gionalize the crop coefficients and calculate evapotranspiration from the values of refer-
38 ence evapotranspiration derived from the regression models. These operations were ap-
39 plied to the Valencian Region (Spain), a Mediterranean area which is partly semi-arid and
40 for which evapotranspiration is a critical issue. The results demonstrated the usefulness
41 and accuracy of the methodology to predict the water demands of crops and hence enable
42 farmers to plan their irrigation needs.

43

44 **Keywords**

45

46 Cluster analysis; Crop coefficient; Evapotranspiration; Multiple linear regression; Refer-
47 ence evapotranspiration; Voronoi diagrams

48

49 **1. Introduction**

50

51 Evapotranspiration (*ET*) is the sum of two processes whereby water is lost from the soil
52 surface (evaporation) and from the crop (transpiration) (Ayttek 2009). As such, it is an
53 important factor in the formation of clouds and the occurrence of rainfall and plays a
54 relevant role in several different water-related fields, including aquifer recharge (Healy
55 and Scanlon 2010), ecosystem water balances (Sun et al. 2011), global circulation models
56 (Dolman 1993), hydrology (Sorooshian et al. 1993), irrigation systems (Allen 2000; Bos
57 et al. 2008), land surface modelling (Chen and Dudhia 2001) and water resource manage-
58 ment (Biswas 2004). Despite its importance, *ET* is still one of the most misunderstood
59 variables in the hydrological cycle and its characterization remains limited (Brutsaert
60 1982; Naoum and Tsanis 2003).

61

62 As a global average, *ET* is responsible for approximately 60% of the precipitation re-
63 turned to the atmosphere, a figure that increases to up to 90% in arid and semi-arid regions
64 (Brutsaert 2005). Therefore, its measurement is essential in agricultural terms for estimat-
65 ing crop water demand and managing irrigation systems. The calculation of *ET* is fre-
66 quently preceded by the determination of reference evapotranspiration (ET_o) (López-
67 Urrea et al. 2006), which is the rate at which available soil water is lost from a specific
68 crop (Jensen et al. 1990) and which can be estimated using climate data (Xing et al. 2008).

69

70 There are many methods developed to determine ET_o based on climate data, but the FAO
71 Penman-Monteith equation (Monteith 1981) has been recommended by the Food and Ag-
72 riculture Organization (Allen et al. 1998) and the American Society of Civil Engineers

73 (Allen et al. 2005) as the standard method for this calculation. This equation can be used
74 worldwide without requiring any local adjustment thanks to its physical foundations, val-
75 idated by the use of lysimeters (Gocic and Trajkovic 2010). In contrast, the main weak-
76 ness of the FAO Penman-Monteith (PM) method is the large amount of variables it con-
77 tains, some of which might not be available in many locations, especially developing
78 countries (Martinez and Thepadia 2010).

79

80 Several researchers have pointed to the need for simpler methods to estimate ET_o (George
81 et al. 2002; Sabziparvar et al. 2010; Tabari and Talaei 2011). Since the relationships be-
82 tween ET_o and the climate variables on which it depends are nonlinear (Jackson 1985;
83 Kumar et al. 2002; Parasumaran et al. 2007; Wang et al. 2007; Adamala et al. 2014),
84 Artificial Neural Networks (ANNs), Adaptive Neuro Fuzzy Inference Systems (ANFIS)
85 and Genetic Programming (GP) have been the main methods used during the last decades
86 to model it. Kumar et al. (2002) and Adamala et al. (2014) concluded that ANNs outper-
87 formed the PM method for reproducing values of ET_o measured with lysimeters, based
88 on the errors yielded by both approaches. Parasuraman et al. (2007), who went one step
89 further and also included GP in the comparison, demonstrated that both this technique
90 and ANNs performed better than the PM method. Similarly, the results achieved by Wang
91 et al. (2008) and Traore et al. (2010) revealed that ANNs could reach higher accuracy
92 than empirical models such as Hargreaves and Blaney-Criddle in the prediction of ET_o .

93

94 Despite the nonlinear nature of ET_o , the linear combination of climate variables has been
95 found to provide a simpler and still reliable and accurate alternative to predict it. Hence,
96 the results obtained by Tabari et al. (2012) indicated that the differences between Multiple

97 Linear Regression (MLR) models and Multiple Nonlinear Regression (MNL) models
98 were almost negligible, to the extent that MLR outperformed MNL when the number
99 of predictors used was small. In the same line, the studies carried out by [Jain et al. \(2008\)](#),
100 [Mallikarjuna et al. \(2013\)](#) and [Ladlani et al. \(2014\)](#), who compared the capability of MLR
101 to estimate ET_o with that of nonlinear methods such as ANNs and ANFIS, suggested that
102 the performance of both linear and nonlinear approaches was very similar. The predictive
103 power of the models built by [Sanford et al. \(2013\)](#), which explained around 90% of the
104 proportion of the variance in the ratio of ET over precipitation, also provided evidence of
105 the potential of MLR to estimate this variable.

106

107 These previous studies show that although nonlinear methods can be slightly more accu-
108 rate than MLR, the differences between both approaches might not be significant and the
109 linear combination of climate variables can provide accurate predictions of ET_o . Further-
110 more, MLR are simpler and easier to understand and interpret than nonlinear techniques,
111 which are frequently used as “black boxes” without having a clear perception of their
112 internal workings. For instance, ANNs, which represent the most widely used nonlinear
113 method to estimate ET_o , require a series of hidden layers to relate inputs and output that
114 are often added arbitrarily to improve the accuracy of the prediction model. This might
115 lead to overfitting of the model and result in misleadingly high-quality estimates. Besides,
116 ANNs do not directly yield equations to estimate future values of ET_o as MLR do. How-
117 ever, former applications of MLR to predict ET_o did not provide solid evidence of their
118 potential for making new estimates. Moreover, they were limited to the prediction of ET_o
119 and did not include any regionalization methodology to group different locations accord-
120 ing to their meteorological characteristics, which together with the fact that they were not

121 built according to data availability in weather forecasts precludes the calculation of ET
122 and therefore the design of aprioristic irrigation strategies.

123

124 In this context, the aim of this paper was to build linear equations for the prediction of ET
125 based on weather forecasts, so that users can estimate the water requirements of their
126 crops and determine when and how much to irrigate. This was achieved through a meth-
127 odology which combined three tools such as cluster analysis, MLR models and Voronoi
128 diagrams to enable the estimation and regionalization of ET using basic meteorological
129 variables. These tools were applied to the Valencian Region in Spain, a Mediterranean
130 area with semi-arid climate zones wherein evapotranspiration is an essential factor in op-
131 timizing agricultural production.

132

133 **2. Methodology**

134

135 **2.1. Framework**

136

137 Evapotranspiration (ET) and reference evapotranspiration (ET_o) can be related through
138 Eq. (1):

139

$$ET = ET_o \cdot K_c \quad (1)$$

140

141 where K_c is the single crop coefficient (dimensionless), which combines the effect of soil
142 evaporation and crop transpiration into a single coefficient and is recommended for irri-

143 gation planning, design, management and scheduling (Allen et al. 1998). Since K_c aver-
 144 ages evaporation and transpiration, a single crop coefficient is used to determine ET for
 145 weekly or longer periods (Allen et al. 1998). Based on findings from several researchers
 146 on the temporal scale of K_c for different crops under Mediterranean climate (Ferreira and
 147 Carr 2002; Williams et al. 2003; Testi et al. 2004; Amayreh and Al-Abed 2005; Martínez-
 148 Cob A. 2008; Villalobos et al. 2009), a monthly period was chosen for the estimation of
 149 this coefficient. This is a time horizon that suits the purpose of this research, since it
 150 allows the prediction of daily ET for every month.

151

152 The FAO PM method is used in Spain for calculating ET_o (Doorenbos and Pruitt 1976).
 153 The concept of ET_o was defined by the FAO as the rate of ET from an ideal 12 cm high
 154 grass reference crop with a fixed canopy of $70 \text{ s}\cdot\text{m}^{-1}$ and an albedo of 0.23 (Allen et al.
 155 1998). This reference surface resembles an extensive and well-watered green grass cover
 156 of uniform height, actively growing and completely shading the ground (Droogers and
 157 Allen 2002). ET_o (mm) can be estimated through Eq. (2), once the aerodynamic and radi-
 158 ation terms derived from the PM equation are combined:

159

$$ET_o = \frac{0.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot \frac{900}{T + 273} \cdot U_2 \cdot (e_a - e_d)}{\Delta + \gamma \cdot (1 + 0.34 \cdot U_2)} \quad (2)$$

160

161 where R_n is net radiation at the crop surface ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$), G is soil heat flux ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$),
 162 T is mean temperature ($^{\circ}\text{C}$), U_2 is mean wind speed at 2 m above the ground ($\text{m}\cdot\text{s}^{-1}$),
 163 $(e_a - e_d)$ is the difference between the actual (e_a) and saturation (e_d) vapor pressure

164 (kPa), Δ is the slope of the vapour pressure curve ($\text{kPa}\cdot^{\circ}\text{C}^{-1}$) and γ is the psychrometric
165 constant ($\text{kPa}\cdot^{\circ}\text{C}^{-1}$), computed as shown in Eq. (3) (Brunt 2011):

166

$$\gamma = 0.00163 \cdot \frac{P}{\lambda} \quad (3)$$

167

168 where P is atmospheric pressure (kPa) and λ is latent heat ($\text{MJ}\cdot\text{kg}^{-1}$). Eqs. (2) and (3)
169 reveal the complexity of the PM equation and the great amount of parameters required by
170 it, some of which are not provided by meteorological agencies in their weather forecasts.
171 Therefore, there is a justifiable need to develop alternative models to estimate ET using
172 basic meteorological variables.

173

174 **2.2. Overview**

175

176 The Valencian Region is divided into three provinces: Alicante, Castellón and Valencia.
177 Table 1 summarizes their main demographic and climate characteristics and indicates the
178 number of valid agrometeorological stations located in each of them. The Spanish Min-
179 istry of Agriculture, Food and Environment (MAGRAMA) provides historical daily val-
180 ues of ET_o for these stations calculated using the FAO PM equation (see Eq. (2)).

181

182 **Table 1.** Main characteristics of the provinces forming the Valencian Region

183

184 However, conventional weather stations do not record all the information required to
185 complete the equation, which also cannot be used to predict new values of ET , since it is
186 not compatible with the variables that are presented in the daily Spanish Meteorological

187 Agency weather forecasts (AEMET 2016). In accordance with the data included in these
188 forecasting models, predictors that are made available include mean temperature (T_{mean} ,
189 °C), maximum temperature (T_{max} , °C), minimum temperature (T_{min} , °C), mean relative
190 humidity (RH_{mean} , %), maximum relative humidity (RH_{max} , %), minimum relative hu-
191 midity (RH_{min} , %) and mean wind speed (WS_{mean} , $m \cdot s^{-1}$).

192

193 The four main steps carried out to develop a methodology capable of predicting ET for a
194 single day in any month using basic meteorological variables are listed below:

195

- 196 • Acquisition of the daily datasets corresponding to the seven predictors for the 49
197 stations located in the whole region and their subsequent arrangement in months,
198 according to the time horizon of K_c .
- 199 • Categorizing the weather stations based on their recorded values in relation to the
200 predictors. Measures of central tendency and variability were used to characterize
201 these stations for clustering.
- 202 • Development of regression equations to make predictions of daily ET_o for each
203 month and cluster from the combination of the set of predictors.
- 204 • Delimitation of the boundaries associated with both the clusters previously ob-
205 tained and the values of K_c for each station using Voronoi diagrams.

206

207 The fulfilment of these steps enabled daily ET to be determined by multiplying K_c by the
208 regression equation built to estimate ET_o for the month and the cluster corresponding to
209 the coordinates of the study area. The theoretical framework behind the tools on which
210 these last three steps were based is described in the following subsections.

211

212 **2.3. Cluster analysis**

213

214 Cluster analysis, a term first introduced by [Tryon \(1939\)](#), is a multivariate data mining
215 technique that uses different algorithms and methods to group objects based on their sim-
216 ilarity. As a result, objects within a group are related to one another but unrelated to ob-
217 jects in other groups, so that the distinctness of the clusters increases as the similarity
218 within a group and the difference between groups increase ([Tan et al. 2005](#)).

219

220 Even though the notion of “cluster” is clear, the definition of the threshold that differen-
221 tiates two clusters has not been precisely defined. Consequently, many clustering methods
222 have been developed over the years, each of them based on different working principles
223 ([Estivill-Castro and Yang 2004](#)). Among them, k -means is one of the most popular algo-
224 rithms to cluster large datasets in an efficient and simple way ([Forgy 1965](#); [MacQueen](#)
225 [1967](#); [Wu et al. 2008](#)).

226

227 The k -means algorithm seeks to partition a set of observations n into $k(\leq n)$ clusters by
228 minimising the within-cluster sum of squares ($WCSS$), i.e. the sum of distances of each
229 point in the cluster to its centroid. This algorithm proceeds according to the three follow-
230 ing steps ([Tan et al. 2005](#)): (1) choose k initial centroids, where k is the number of clusters
231 desired; (2) assign each observation to the closest cluster according to the Euclidean dis-
232 tance between them, i.e. the square root of the sum of their squared differences; and (3)
233 update the centroid of each cluster based on the points assigned to it. The last two steps

234 are repeated until the results converge and there are no point changes in the clusters. In
235 other words, the algorithm stops when the centroids remain the same (Tan et al. 2005).

236

237 Two pairs of measures of central tendency and variability were proposed to characterize
238 these variables for each weather station depending on whether they were normally dis-
239 tributed or not: mean (\bar{x}) and standard deviation (σ) or median (\tilde{x}) and interquartile range
240 (*IQR*), respectively. The Shapiro-Wilk test (Shapiro and Wilk 1965), which has been
241 found to be more reliable when checking normality than other commonly used tests such
242 as Kolmogorov-Smirnov or Lilliefors (Shapiro et al. 1968), was selected for checking
243 normality.

244

245 **2.4. Multiple linear regression**

246

247 Multiple Linear Regression (MLR) aims to model the relationship between two or more
248 predictors (basic meteorological variables) and a predictand (ET_o) by fitting a linear equa-
249 tion to observed data (see Eq. (4)):

250

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_k \cdot x_k + \varepsilon \quad (4)$$

251

252 where y is the predictand expressed as a linear combination of a set of K predictors x_k ,
253 each of which is multiplied by a coefficient β_k that indicates its relative weight in the
254 equation. The equation also includes a constant β_0 and a random component ε (the resid-
255 uals) which explain everything that cannot be interpreted from the predictors.

256

257 The goodness-of-fit of a MLR model is often measured through the coefficient of deter-
258 mination (R^2) (Hirsch et al. 1993). The standard R^2 is useful to determine how well the
259 model fits the original data, but has several limitations that compromise its validity to
260 make predictions. It does not capture the influence of the number of predictors in fitting
261 the model, so that the addition of a predictor always results in an increase in R^2 . The
262 adjusted R^2 arose as a modified version of the standard R^2 that compares the explanatory
263 power of regression models built with different numbers of predictors. However, although
264 this coefficient improves the reliability of R^2 , it still cannot provide accurate predictions
265 of new data, which is the main goal of this research. Another variant of the coefficient of
266 determination, known as predictive R^2 , was used to overcome this drawback by making
267 estimates on new observations according to three steps: (1) remove each observation from
268 the dataset, (2) estimate the regression equation without the removed observation and (3)
269 determine how well the model predicts the removed observation. The goodness-of-fit of
270 the models was also tested through the standard error of the regression (S), which repre-
271 sents the average distance from the observed values to the regression line.

272

273 Cook's distance was used to show the influence of each observation on the response val-
274 ues and identify erroneous measurements in the predictors (outliers). According to Eq.
275 (5), an observation with a Cook's distance (D_i) larger than three times the mean Cook's
276 distance is considered as an outlier (Stevens 2009):

277

$$D_i = \frac{\sum_{j=1}^n (z_j - z_{j(i)})^2}{p \cdot MSE} \quad (5)$$

278

279 where z_j is the j th fitted response values, $z_{j(i)}$ is the j th fitted response value where the fit
280 does not include observation i , p is the number of coefficients of the regression model
281 and MSE is the mean squared error.

282

283 MLR is based on four assumptions that must be verified to ensure its validity: linearity,
284 independence, homoscedasticity and normality. Violation to these assumptions was diag-
285 nosed through the residual plots and the Durbin-Watson statistic (Osbourne and Waters
286 2002).

287

288 **2.5. Voronoi diagrams**

289

290 The concept of Voronoi diagrams (Voronoi 1908), also known as Dirichlet tessellation
291 (Dirichlet 1850) or Thiessen polygons (Thiessen and Alter 1911), consists of dividing a
292 plane containing a series of points following the nearest-neighbor rule, so that each point
293 belongs to the region of the plane closest to it (Aurenhammer 1991), called a Voronoi
294 cell.

295

296 Analytically, if $X = \{x_1, x_2, \dots, x_n\}$ is a set of point sites in the plane, then the Voronoi cell
297 for a point site x_i ($VC(x_i)$) is defined as the set of points y in the plane that are closer to
298 x_i than any other point site (see Eq. (6)):

299

$$VC(x_i) = \{y \mid d(x_i, y) < d(x_j, y), \forall j \neq i\} \quad (6)$$

300

301 where $d(x, y)$ denotes the Euclidean distance between the points x and y . From a graph-
302 ical point of view, $VC(x_i)$ can also be defined in terms of the intersection of half-planes.
303 The bisector of x and y is equal to the perpendicular line through the centre of the line
304 segment \overline{xy} and separates the plane into two half-planes. Therefore, the Voronoi diagram
305 of X is the tuple of cells $VC(x_i \in X)$. More details about the properties of Voronoi dia-
306 grams can be found in [Aurenhammer and Klein \(2000\)](#).

307

308 **3. Results and discussion**

309

310 The study period for this research was between 2008 and 2014, since the former was the
311 first year in which all the agrometeorological stations in the Valencian Region (see [Table](#)
312 [1](#)) started to work altogether. [Figure 1](#) shows the location of this region in relation to the
313 geography of Spain and the Mediterranean Sea and its division into the provinces of Ali-
314 cante, Castellón and Valencia.

315

316 **Figure 1.** Location and provincial division of the Valencian Region

317

318 The first step in the methodology was the regionalization of the Valencian Region ac-
319 cording to its weather characteristics, which were provided by the values taken by the
320 basic meteorological variables to be used as predictors for building the regression models
321 in the stations. Normality of this set of possible predictors was checked through the
322 Shapiro-Wilk test, which revealed that the null hypothesis was rejected for all of them (p-
323 values < 0.05). Hence, these variables were characterized for clustering through the me-
324 dian (\tilde{x}) and interquartile range (IQR) corresponding to each station. As an exploratory

325 inspection of the variations in ET_o across the Valencian Region, [Table 2](#) lists the monthly
326 values of \tilde{x} and IQR obtained after averaging the stations located in each of the three
327 provinces forming it. The general trend of these data suggested that the highest values of
328 ET_o were recorded in Alicante, which is characterised by having a drier climate than either
329 Castellón or Valencia and therefore, higher temperatures coupled with lower humidity.
330 The Köppen Climate Classification for the Iberian Península ([Chazarra et al. 2011](#)) con-
331 firmed this inference, since Alicante completely belongs to type B (dry), whereas Castel-
332 lón and Valencia also have some type C areas (temperate).

333

334 **Table 2.** Average median (\tilde{x}) and interquartile range (IQR) of ET_o (mm/month) for each province

335

336 Many different methods have been developed to optimize the determination of the num-
337 ber of clusters in a dataset, such as the gap statistic, Hartigan's approach or silhouette
338 ([Tibshirani et al. 2001](#)). However, since cluster analysis preceded the development of the
339 prediction models, the number of clusters chosen was calculated to maximize the predic-
340 tive R^2 of subsequent regression equations. The results demonstrated that the optimal
341 number of clusters was 1 in all cases except in May, June, July and August, when it was
342 2. In other words, the predictive R^2 was maximized for these clusters and then began to
343 decrease its value gradually as the number of clusters increased.

344

345 [Figure 2](#) illustrates the Voronoi regions obtained for each of these months from the pair
346 of values (\tilde{x}, IQR) calculated from each station. These were the warmer months of the
347 year and those in which the combination of weather effects resulted in the highest and
348 most varying values of ET (see [Table 2](#)), justifying the need to partition the whole work-
349 space into two zones. The clustering patterns were consistent with that premise, since

350 they separated the coastal and interior areas of the region, which were the zones wherein
351 such variability became more accentuated.

352

353 **Figure 2.** Clusters obtained for a) May b) June c) July d) August

354

355 From there, multiple linear regression models were built to estimate daily ET_o for each
356 month and cluster by adapting Eq. (4) to the specifics of this research: $y = ET_o$ (mm/day);
357 $x_1 = T_{mean}$ (°C); $x_2 = T_{max}$ (°C); $x_3 = T_{min}$ (°C); $x_4 = RH_{mean}$ (%); $x_5 = RH_{max}$ (%); $x_6 =$
358 RH_{min} (%); $x_7 = WS_{mean}$ (m·s⁻¹). A 95% confidence interval (p-value < 0.05) was set to
359 choose predictors stepwise, whilst Cook's distances were calculated using Eq. (5) to de-
360 tect and remove influential points. Table 3 shows the regression coefficients and good-
361 ness-of-fit measures obtained for the number of days (N) corresponding to each month
362 and cluster (CL) between 2008 and 2014.

363

364 **Table 3.** Summary of the regression models to predict ET_o (mm/day) for each month and cluster

365

366 The results were 16 regression equations consisting of 5 predictors in each case. Varia-
367 tions in the coefficients associated with the predictors (see Table 3) demonstrated the need
368 to build monthly regression models for the prediction of ET_o , because weather attributes
369 vary over the year (e.g. increased temperature in summer). Although the predictors in-
370 cluded in each model varied in some cases depending on the month and cluster, all re-
371 gression models consisted of two temperature-related variables (' T_{mean} AND T_{min} ' OR
372 ' T_{mean} AND T_{max} ' OR ' T_{min} AND T_{max} '), two humidity-related variables (' RH_{mean}
373 AND RH_{min} ' OR ' RH_{mean} AND RH_{max} ' OR ' RH_{min} AND RH_{max} ') and WS_{mean} . The
374 most influential predictors were found to be those related to temperature with an average

375 contribution around 50% to estimate ET_o , except for the colder months, in which the com-
376 bination of relative humidity and wind speed explained up to 80% of the variations in the
377 predictand. The physical relationships between the mean predictors (x_1 , x_4 and x_7), which
378 are the most representative ones for each type of variable (temperature, humidity and
379 wind), and the predictand were logical in all cases. The pores of plants in which water is
380 released open if they are surrounded by warmer air, i.e. there is an increase in transpiration
381 (Crawford et al. 2012). In contrast, relative humidity is inversely proportional to evapo-
382 transpiration, since the evaporation of water into the air is hindered as this becomes more
383 saturated (Thut 1938). As for wind speed, moving air facilitates the process of evapotran-
384 spiration, since it is less saturated than stagnant air and can absorb water vapor more
385 easily (Moore et al. 2003).

386

387 The reliability of the regression models for making predictions was guaranteed by the
388 high and low values of predictive R^2 and S reached, respectively. The values of predictive
389 R^2 indicated that these regression models can make estimates for new values of daily ET_o
390 with an accuracy of at least 83% through a linear combination of basic variables related
391 to temperature, humidity and wind. The ratio between S and the average monthly values
392 of \bar{x} and IQR (see Table 2) was at most 7% and 25%, respectively, which demonstrates
393 that the errors in the regression models were very small in relation to the typical values
394 and spread of ET_o . The relationships between the climate variables used as predictors and
395 ET_o were nonlinear in general. Figure 3 illustrates this circumstance for April, in which
396 the predictand varied nonlinearly in relation to all predictors except RH_{min} , whose rela-
397 tionship to ET_o could be assumed to be linear. Therefore, these results confirmed that the

398 linear combination of climate variables can provide accurate predictions of ET_o , even
399 though their individual correlations are mostly nonlinear.

400

401 **Figure 3.** Relationships between the predictors and the predictand (ET_o) in the regression model for April

402

403 **Figure 4** shows the histograms and scatterplots of standardized residuals against fitted
404 values for two months representing different weather conditions (April (1 cluster) and
405 June (2 clusters)), which provide graphical diagnose verifying whether the assumptions
406 of normality, linearity and homoscedasticity were violated or not. The symmetrical bell-
407 shape of the histograms, which fitted their corresponding theoretical normal curves with
408 high accuracy, suggested that the normality assumption was true. Moreover, the absence
409 of curvilinear distributions and marked trends (e.g. increasing dispersion as the fitted val-
410 ues increase) in the scatterplots confirmed both the linearity and homoscedasticity of the
411 residuals. Finally, the Durbin-Watson statistics were between 1.5 and 2.5 (Durbin and
412 Watson 1950; Durbin and Watson 1951) in all three cases (1.740 for April and 1.596
413 (CL1) and 1.591 (CL2) for June), which involved that there was no time trends nor serial
414 correlations in the residuals and their independence could be assumed too. Furthermore,
415 the values of Variance Inflation Factor (VIF) obtained for the predictors, which were al-
416 ways below 10 (Belsley et al. 1980), ensured that they were not highly correlated to each
417 other and multicollinearity was not an issue.

418

419 **Figure 4.** Histograms and scatterplots of standardized residuals against fitted values for a) April b) June -

420

Cluster 1 c) June - Cluster 2

421

422 The final step was the regionalization of the Valencian Region according to the crop co-
423 efficients (K_c) in each station, in order to obtain a value for ET from ET_o using Eq. (1).
424 Due to space constraints, this last process was limited to only one crop type: midseason
425 potato. This specific crop was selected because it proved to be variable in terms of both
426 location and time. The daily values of K_c provided by the MAGRAMA through its Agro-
427 climatic Information System for Irrigation (SiAR 2016), which were constant for each
428 month during the years of study, reaffirmed the convenience of choosing a monthly period
429 for the estimation of this coefficient. Therefore, the Voronoi regions were drawn as shown
430 in Figure 5 according to the values of K_c for each station and month of the year. The
431 procedure would be the same for any other crop, with the only difference that the Voronoi
432 regions should be particularized to the monthly values of K_c associated with the specifics
433 of the crop under study.

434

435 **Figure 5.** Monthly crop coefficients (K_c) in the Valencian Region for midseason potato

436

437 Knowing the coordinates for where irrigation was planned, the multiplication of crop co-
438 efficients in this area (see Figure 5) by the regression equations summarized in Table 3
439 enabled an estimation to be made of the water demands of this crop for a single day in
440 any month using basic meteorological variables available from official weather forecasts.
441 For instance, Figure 6 particularizes the procedure for the case of a farmer who planted
442 midseason potatoes in April in the geographic coordinates (39°55'57'' N, 1°04'10'' W)
443 and would like to estimate ET in a day in May. To illustrate the example, the historical
444 average values for May recorded in the closest station to the specified coordinates were
445 taken as the climate variables to be acquired from daily weather forecasts. According to
446 the clusters identified in Figure 6a) and Figure 6b), these coordinates corresponded to

447 CL1 and a Voronoi region with a value of K_c equal to 0.8. The application of the regres-
448 sion equation in Table 3 for these predictors, month and cluster yielded a value of ET_o of
449 4.66 mm/day. The multiplication of ET_o by K_c as formulated in Eq. (1) resulted in a final
450 value of ET equal to 3.77 mm/day.

451

452 **Figure 6.** Estimation of ET in May for midseason potato in the coordinates (39°55'57'' N, 1°04'10'' W)
453 a) Cluster b) Monthly crop coefficient (K_c) c) Historical average values for the predictors in the closest
454 station to the coordinates d) Calculation of ET_o (mm/day) e) Determination of ET (mm/day)

455

456 **4. Conclusions**

457

458 This paper presents a methodology for the prediction of daily evapotranspiration based
459 on the combination of cluster analysis, multiple linear regression models and Voronoi
460 diagrams. The first was used to partition the study area according to its weather charac-
461 teristics, so that regression equations to estimate daily reference evapotranspiration could
462 be built for the resultant clusters using basic meteorological variables. Voronoi diagrams
463 enabled regionalization of the workspace in terms of both clusters and crop coefficients
464 associated with it, whose multiplication by reference evapotranspiration yielded the value
465 for real evapotranspiration which was being sought.

466

467 Despite the relationships between climate variables and reference evapotranspiration are
468 generally nonlinear, the results proved that the linear combination of the former can pro-
469 vide accurate estimates of the latter. The models obtained using multiple linear regression
470 analysis met the four hypotheses related to their residuals and reached high predictive
471 coefficients of determination, which ensured their reliability and capability to make new

472 estimates from daily weather forecasts. As for cluster analysis and Voronoi diagrams,
473 their combination was found to be a simple and effective method for local application of
474 the predictive regression equations and regionalization of crop coefficients, which ena-
475 bled determining real evapotranspiration without any need to take into account complex
476 physical considerations.

477

478 This methodology is proposed as a tool to be used by farmers for irrigation planning and
479 scheduling based on the estimation of water demands of their crops. The daily value of
480 evapotranspiration corresponding to a given date, coordinates and crop can be determined
481 through the cluster, regression equation and crop coefficient associated with the day and
482 region under study, since they are based on primary weather variables that are available
483 from the daily forecasts made by meteorological agencies. Although the validity of these
484 results is not compromised by the size of the study area, further research should consider
485 the application of this methodology to larger locations, in order to delimit different cli-
486 mate zones and develop regional prediction equations at larger scales.

487

488 **Acknowledgments**

489

490 This paper was possible thanks to the research project RHIVU (Ref. BIA2012-32463),
491 financed by the Spanish Ministry of Economy and Competitiveness with funds from the
492 State General Budget (PGE) and the European Regional Development Fund (ERDF). The
493 authors also wish to express their gratitude to the Spanish Ministry of Agriculture, Food
494 and Environment (MAGRAMA) for providing the data necessary to develop this study.

495

496 **References**

497

498 Adamala, S., Raghuwanshi, N. S., Mishra, A., and Tiwari, M. K. (2014). "Evapotranspi-
499 ration modeling using second-order neural networks." *J.Hydrol.Eng.*, 19(6), 1131-1140.

500 AEMET. (2016). "Agencia estatal de meteorología." www.aemet.es (August, 2016).

501 Allen, R. G. (2000). "Using the FAO-56 dual crop coefficient method over an irrigated
502 region as part of an evapotranspiration intercomparison study." *J.Hydrol.*, 229(1-2), 27-
503 41.

504 Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). "Crop evapotranspiration -
505 Guidelines for computing crop water requirements." *FAO Irrigation and Drainage Pa-*
506 *per*, 56(9), 1-174.

507 Allen, R. G., Walter, I. A., Elliott, R., Howell, T., Itenfisu, D., and Jensen, M. (2005).
508 "The ASCE standardized reference evapotranspiration equation." Rep. No. Final Re-
509 port, Environmental and Water Resources Institute (EWRI) of the American Society of
510 Civil Engineers (ASCE), Reston, Virginia (U.S.).

511 Amayreh, J., and Al-Abed, N. (2005). "Developing crop coefficients for field-grown to-
512 mato (*Lycopersicon esculentum* Mill.) under drip irrigation with black plastic mulch."
513 *Agric.Water Manage.*, 73(3), 247-254.

514 Aurenhammer, F. (1991). "Voronoi Diagrams - A Survey of a Fundamental Geometric
515 Data Structure." *ACM Computing Surveys*, 23(3), 345-405.

516 Aurenhammer, F., and Klein, R. (2000). "Voronoi Diagrams." *Handbook of Computa-*
517 *tional Geometry*, V. J. Sack, and G. Urrutia, eds., Elsevier, Amsterdam (Netherlands),
518 201-290.

519 AYTEK, A. (2009). "Co-active neurofuzzy inference system for evapotranspiration mod-
520 eling." *Soft Comput.*, 13(7), 691-700.

521 Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying*
522 *Influential Data and Sources of Collinearity*. John Wiley & Sons, New Jersey (U.S.).

523 Biswas, A. K. (2004). "Integrated water resources management: A reassessment." *Water*
524 *Int.*, 29(2), 248-256.

- 525 Bos, M. G., Kselik, R. A. L., Allen, R. G., and Molden, D. (2008). Water requirements
526 for irrigation and the environment. Springer, Dordrecht (The Netherlands).
- 527 Brunt, D. (2011). Physical and dynamical meteorology. Cambridge University Press,
528 Cambridge (U.K.).
- 529 Brutsaert, W. (2005). Hydrology: An Introduction. Cambridge University Press, Cam-
530 bridge (U.K.).
- 531 Brutsaert, W. H. (1982). Evaporation into the Atmosphere. D. Reidel, Dordrecht (The
532 Netherlands).
- 533 Chazarra, A., Mestre Barceló, A., Pires, V., Cunha, S., Mendes, M., and Neto, J. (2011).
534 "Atlas Climático Ibérico. Agencia Estatal de Meteorología and Instituto de Meteorolo-
535 gía de Portugal." Rep. No. Catálogo General de publicaciones oficiales, Agencia Estatal
536 de Meteorología (AEMET) and Instituto de Meteorologia de Portugal, Madrid (Spain)
537 and Lisboa (Portugal).
- 538 Chen, F., and Dudhia, J. (2001). "Coupling and advanced land surface-hydrology model
539 with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and
540 sensitivity." *Mon. Weather Rev.*, 129(4), 569-585.
- 541 Crawford, A. J., McLachlan, D. H., Hetherington, A. M., and Franklin, K. A. (2012).
542 "High temperature exposure increases plant cooling capacity." *Curr. Biol.*, 22(10), R396-
543 R397.
- 544 Dirichlet, G. L. (1850). "Über die Reduktion der positiven quadratischen Formen mit
545 drei unbestimmten ganzen Zahlen." *J. Reine Angew. Math.*, 40 209-227.
- 546 Dolman, A. J. (1993). "A multiple-source land surface energy balance model for use in
547 general circulation models." *Agric. for. Meteorol.*, 65(1-2), 21-45.
- 548 Doorenbos, J., and Pruitt, W. O. (1976). "Guidelines for predicting crop water require-
549 ments." Rep. No. FAO Irrigation and Drainage Paper 24, Food and Agriculture Organi-
550 zation of the United Nations, Rome (Italy).
- 551 Droogers, P., and Allen, R. G. (2002). "Estimating reference evapotranspiration under
552 inaccurate data conditions." *Irrig. Drain. Syst.*, 16(1), 33-45.
- 553 Durbin, J., and Watson, G. S. (1951). "Testing for serial correlation in least squares re-
554 gression. II." *Biometrika*, 38(1-2), 159-177.

- 555 Durbin, J., and Watson, G. S. (1950). "Testing for serial correlation in least squares re-
556 gression. I." *Biometrika*, 37(3-4), 409-428.
- 557 Estivill-Castro, V., and Yang, J. (2004). "Fast and robust general purpose clustering al-
558 gorithms." *Data Min.Knowl.Discov.*, 8(2), 127-150.
- 559 Ferreira, T. C., and Carr, M. K. V. (2002). "Responses of potatoes (*Solanum tuberosum*
560 L.) to irrigation and nitrogen in a hot, dry climate I. Water use." *Field Crops Res.*, 78(1),
561 51-64.
- 562 Forgy, E. W. (1965). "Cluster analysis of multivariate data: efficiency versus interpreta-
563 bility of classifications." *Biometrics*, 21 768-769.
- 564 George, B. A., Reddy, B. R. S., Raghuwanshi, N. S., and Wallender, W. W. (2002).
565 "Decision support system for estimating reference evapotranspiration." *J.Ir-
566 rig.Drain.Eng.*, 128(1), 1-10.
- 567 Gocic, M., and Trajkovic, S. (2010). "Software for estimating reference evapotranspira-
568 tion using limited weather data." *Comput.Electron.Agric.*, 71(2), 158-162.
- 569 Healy, R. W., and Scanlon, B. R. (2010). *Estimating groundwater recharge*. Cambridge
570 University Press, Cambridge (U.K.).
- 571 Hirsch, R. M., Helsel, D. R., Cohn, T. A., and Gilroy, E. J. (1993). "Statistical Analysis
572 of Hydrologic Data." *Handbook of Hydrology*, D. R. Maidment, ed., McGraw-Hill,
573 New York (U.S.), 1-55.
- 574 Jackson, R. D. (1985). "Evaluating Evapotranspiration at Local and Regional Scales."
575 *Proc IEEE*, 73(6), 1086-1096.
- 576 Jain, S. K., Nayak, P. C., and Sudheer, K. P. (2008). "Models for estimating evapotran-
577 spiration using Artificial Neural Networks, and their physical interpretation." *Hy-
578 drol.Processes*, 22(13), 2225-2234.
- 579 Jensen, M. E., Burman, R. D., and Allen, R. G. (1990). *Evapotranspiration and irriga-
580 tion water requirements*. ASCE, New York (U.S.).
- 581 Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., and Pruitt, W. O.
582 (2002). "Estimating evapotranspiration using Artificial Neural Network." *J.Ir-
583 rig.Drain.Eng.*, 128(4), 224-233.

- 584 Ladlani, I., Houichi, L., Djemili, L., Heddami, S., and Belouiz, K. (2014). "Estimation of
585 Daily Reference Evapotranspiration (ET₀) in the North of Algeria Using Adaptive
586 Neuro-Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR) Mod-
587 els: A Comparative Study." *Arab.J.Sci.Eng.*, 39(8), 5959-5969.
- 588 López-Urrea, R., Martín de Santa Olalla, F., Fabeiro, C., and Moratalla, A. (2006).
589 "Testing evapotranspiration equations using lysimeter observations in a semiarid cli-
590 mate." *Agric.Water Manage.*, 85(1-2), 15-26.
- 591 MacQueen, J. (1967). "Some methods for classification and analysis of multivariate ob-
592 servations." *Proceedings of the 5th Berkeley symposium on mathematical statistics and*
593 *probability*, University of California Press, Oakland, California (U.S.), 281-297.
- 594 Mallikarjuna, P., Jyothy, S. A., and Sekhar Reddy, K. C. (2013). "Daily Reference
595 Evapotranspiration Estimation using Linear Regression and ANN Models." *J. Inst. Eng.*
596 *India Ser. A*, 93(4) 215-221.
- 597 Martinez, C. J., and Thepadia, M. (2010). "Estimating reference evapotranspiration with
598 minimum data in florida." *J.Irrig.Drain.Eng.*, 136(7), 494-501.
- 599 Martínez-Cob A. (2008). "Use of thermal units to estimate corn crop coefficients under
600 semiarid climatic conditions." *Irrig.Sci.*, 26(4), 335-345.
- 601 Monteith, J. L. (1981). "Evaporation and surface temperature." *Quarterly Journal, Royal*
602 *Meteorological Society*, 107(451), 1-27.
- 603 Moore, R., Clark, W. D., and Vodopich, D. S. (2003). *Botany*. McGraw-Hill, New York
604 (U.S.).
- 605 Naoum, S., and Tsanis, I. K. (2003). "Hydroinformatics in evapotranspiration estima-
606 tion." *Environ.Model.Softw.*, 18(3), 261-271.
- 607 Osbourne, J. W., and Waters, E. (2002). "Four Assumptions of Multiple Regression
608 That Researchers Should Always Test." *Practical Assessment, Research & Evaluation*,
609 8(2) 1-7.
- 610 Parasuraman, K., Elshorbagy, A., and Carey, S. K. (2007). "Modelling the dynamics of
611 the evapotranspiration process using genetic programming." *Hydrol.Sci.J.*, 52(3), 563-
612 578.

- 613 Sabziparvar, A. -, Tabari, H., Aeini, A., and Ghafouri, M. (2010). "Evaluation of class a
614 pan coefficient models for estimation of reference crop evapotranspiration in cold semi-
615 arid and warm arid climates." *Water Resour.Manage.*, 24(5), 909-920.
- 616 Sanford, W. E., and Selnick, D. L. (2013). "Estimation of evapotranspiration across the
617 conterminous United States using a regression with climate and land-cover data." *J. Am.*
618 *Water Resour. Assoc.*, 49(1) 217-230.
- 619 Shapiro, S. S., and Wilk, M. B. (1965). "An analysis of variance test for normality." *Bi-*
620 *ometrika*, 52(3-4) 591-611.
- 621 Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). "A Comparative Study of Various
622 Tests for Normality." *J. Am. Stat. Assoc.*, 63(324), 1343-1372.
- 623 SiAR. (2016). "Necesidades netas." [portal.magrama.gob.es/websiar/NecesidadesHidri-](http://portal.magrama.gob.es/websiar/NecesidadesHidricas.aspx)
624 [cas.aspx](http://portal.magrama.gob.es/websiar/NecesidadesHidricas.aspx) (August, 2016).
- 625 Sorooshian, S., Duan, Q., and Gupta, V. K. (1993). "Calibration of rainfall-runoff mod-
626 els: Application of global optimization to the Sacramento Soil Moisture Accounting
627 Model." *Water Resour.Res.*, 29(4), 1185-1194.
- 628 Stevens, J. (2009). *Applied Multivariate Statistics for the Social Sciences*. Taylor &
629 Francis, Mahwah, New Jersey (U.S.).
- 630 Sun, G., Alstad, K., Chen, J., Chen, S., Ford, C. R., Lin, G., Liu, C., Lu, N., McNulty, S.
631 G., Miao, H., Noormets, A., Vose, J. M., Wilske, B., Zeppel, M., Zhang, Y., and Zhang,
632 Z. (2011). "A general predictive model for estimating monthly ecosystem evapotranspi-
633 ration." *Ecohydrology*, 4(2), 245-255.
- 634 Tabari, H., Kisi, O., Ezani, A., and Hosseinzadeh Talaei, P. (2012). "SVM, ANFIS, re-
635 gression and climate based models for reference evapotranspiration modeling using lim-
636 ited climatic data in a semi-arid highland environment." *J.Hydrol.*, 444-445 78-89.
- 637 Tabari, H., and Talaei, P. H. (2011). "Local Calibration of the Hargreaves and Priest-
638 ley-Taylor Equations for Estimating Reference Evapotranspiration in Arid and Cold
639 Climates of Iran Based on the Penman-Monteith Model." *J.Hydrol.Eng.*, 16(10), 837-
640 845.
- 641 Tan, P. -, Steinbach, M., and Kumar, V. (2005). "Cluster Analysis: Basic Concepts and
642 Algorithms." *Introduction to Data Mining*, Addison-Wesley, Boston, Massachusetts
643 (U.S.), 487-568.

- 644 Testi, L., Villalobos, F. J., and Orgaz, F. (2004). "Evapotranspiration of a young irri-
645 gated olive orchard in southern Spain." *Agric.for.Meterol.*, 121(1-2), 1-18.
- 646 Thiessen, A. J., and Alter, J. C. (1911). "Precipitation Averages for Larger Areas." *Mon.*
647 *Weather Rev.*, 39(7), 1082-1084.
- 648 Thut, H. F. (1938). "Relative Humidity Variations Affecting Transpiration." *Am.J.Bot.*,
649 25(8), 589-595.
- 650 Tibshirani, R., Walther, G., and Hastie, T. (2001). "Estimating the number of clusters in
651 a data set via the gap statistic." *J.R.Stat.Soc.Ser.B Stat.Methodol.*, 63(2), 411-423.
- 652 Traore, S., Wang, Y. -, and Kerh, T. (2010). "Artificial neural network for modeling
653 reference evapotranspiration complex process in Sudano-Sahelian zone." *Agric.Water*
654 *Manage.*, 97(5), 707-714.
- 655 Tryon, R. (1939). *Cluster Analysis: Correlation Profile and Orthometric (factor) Analy-*
656 *sis for the Isolation of Unities in Mind and Personality.* Edwards Brothers Malloy, Ann
657 Arbor, Michigan (U.S.).
- 658 Villalobos, F. J., Testi, L., and Moreno-Perez, M. F. (2009). "Evaporation and canopy
659 conductance of citrus orchards." *Agric.Water Manage.*, 96(4), 565-573.
- 660 Voronoi, G. M. (1908). "Nouvelles applications des paramètres continus à la théorie des
661 formes quadratiques. deuxième Mémoire: Recherches sur les paralléloèdres primitifs."
662 *J. Reine Angew. Math.*, 134 198-287.
- 663 Wang, Y. -, Traore, S., and Kerh, T. (2008). "Neural network approach for estimating
664 reference evapotranspiration from limited climatic data in Burkina Faso." *WSEAS*
665 *Trans.Comput.*, 7(6), 704-713.
- 666 Williams, L. E., Phene, C. J., Grimes, D. W., and Trout, T. J. (2003). "Water use of ma-
667 ture Thompson Seedless grapevines in California." *Irrig.Sci.*, 22(1), 11-18.
- 668 Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng,
669 A., Liu, B., Yu, P. S., Zhou, Z. -, Steinbach, M., Hand, D. J., and Steinberg, D. (2008).
670 "Top 10 algorithms in data mining." *Knowl.Inf.Systems.Syst.*, 14(1), 1-37.
- 671 Xing, Z., Chow, L., Meng, F. -, Rees, H. W., Monteith, J., and Lionel, S. (2008). "Test-
672 ing reference evapotranspiration estimation methods using evaporation pan and model-
673 ing in maritime region of Canada." *J.Irrig.Drain.Eng.*, 134(4), 417-424.

Table 1. Main characteristics of the provinces forming the Valencian Region

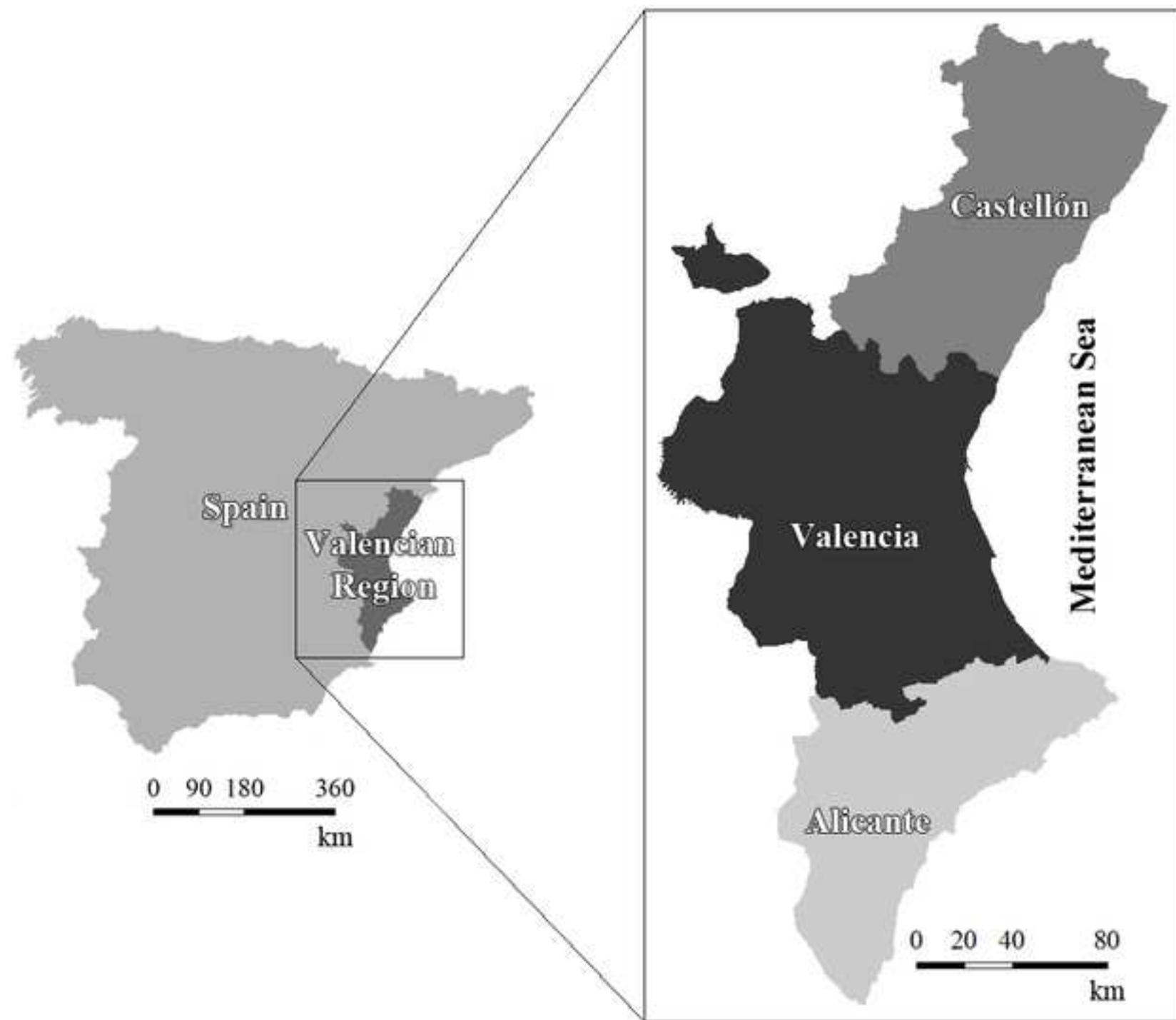
Province	Population	Surface area (km²)	Valid stations	Average Annual Precipitation (mm)	Average Annual Max Temperature (°C)	Average Annual Min Temperature (°C)
Alicante	1,934,127	5,816	16	311.1	23.3	13.2
Castellón	604,344	6,632	10	467.0	22.3	12.7
Valencia	2,578,719	10,763	23	474.9	23.0	13.8

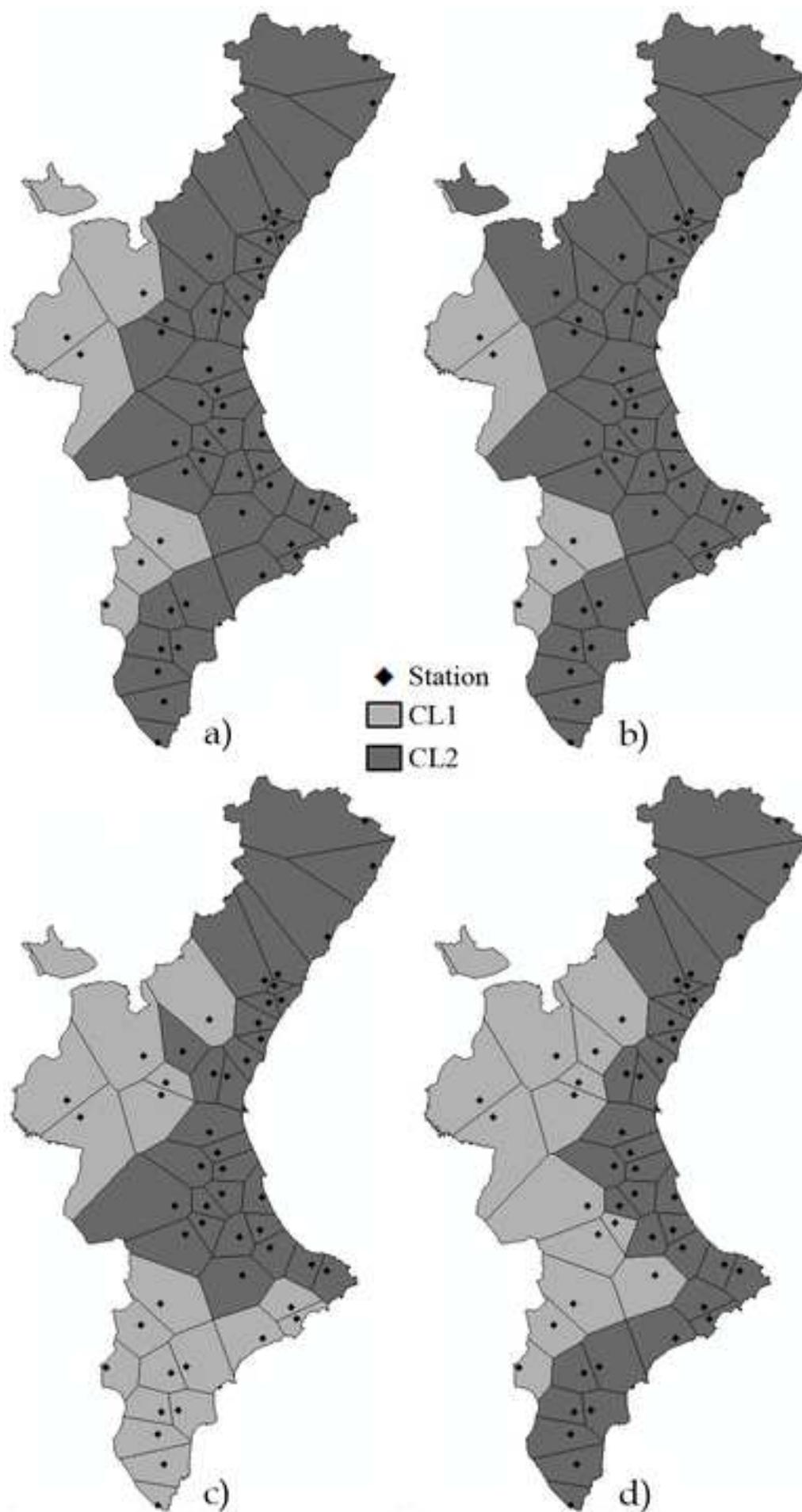
Table 2. Average median (\tilde{x}) and interquartile range (*IQR*) of ET_o (mm/month) for each province

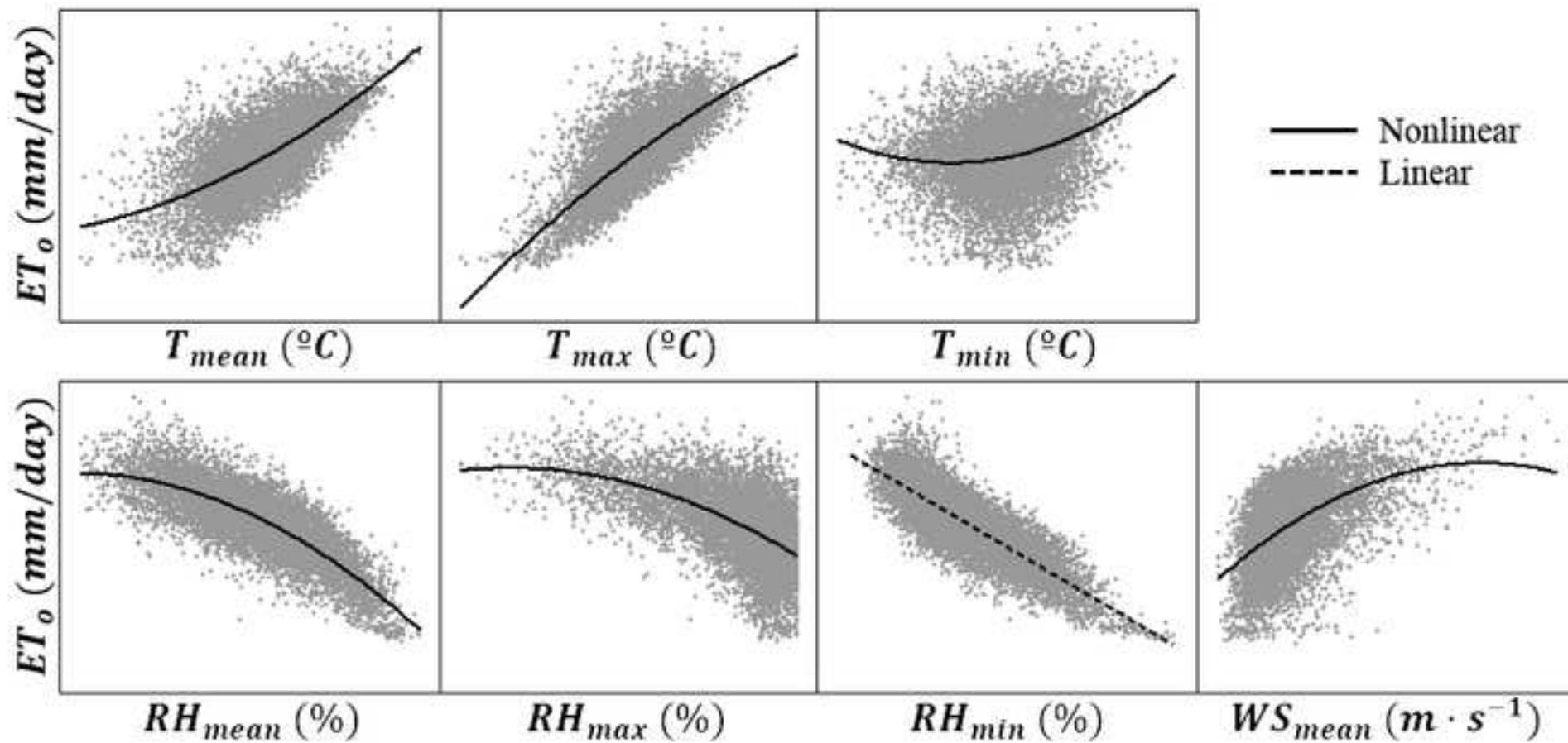
Province	Measure	Month											
		1	2	3	4	5	6	7	8	9	10	11	12
Alicante	\tilde{x}	1.18	1.79	2.74	3.65	4.62	5.45	5.70	5.01	3.67	2.36	1.41	1.05
	<i>IQR</i>	0.70	0.85	1.18	1.41	1.17	0.99	0.75	0.94	1.19	0.86	0.71	0.49
Castellón	\tilde{x}	1.07	1.69	2.51	3.42	4.32	5.12	5.31	4.59	3.44	2.17	1.30	0.99
	<i>IQR</i>	0.64	0.85	1.06	1.28	1.32	1.02	0.89	1.10	1.19	0.88	0.63	0.48
Valencia	\tilde{x}	1.10	1.75	2.68	3.55	4.49	5.30	5.55	4.86	3.55	2.19	1.30	0.98
	<i>IQR</i>	0.81	1.05	1.32	1.47	1.40	1.13	0.77	0.97	1.30	0.94	0.79	0.60

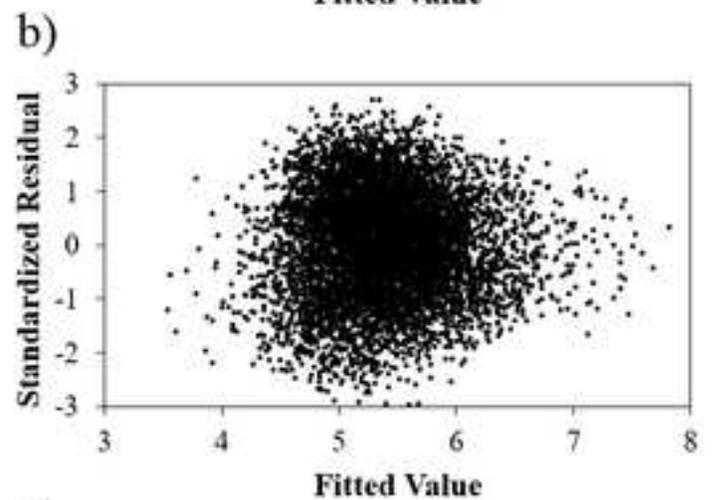
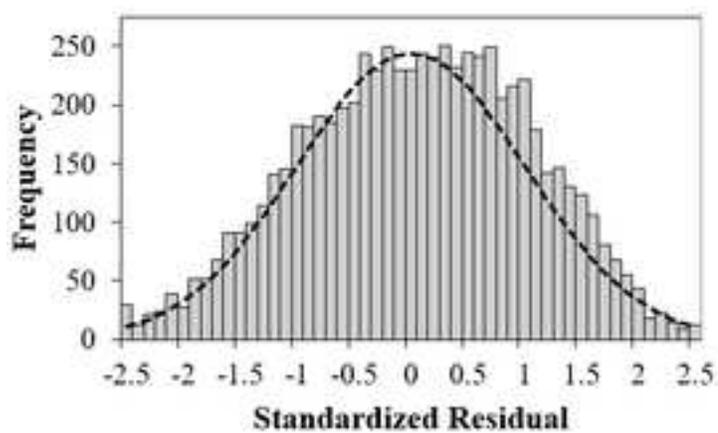
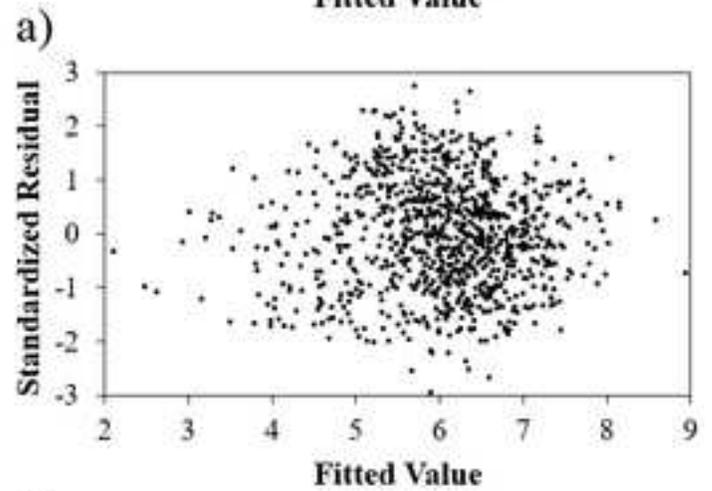
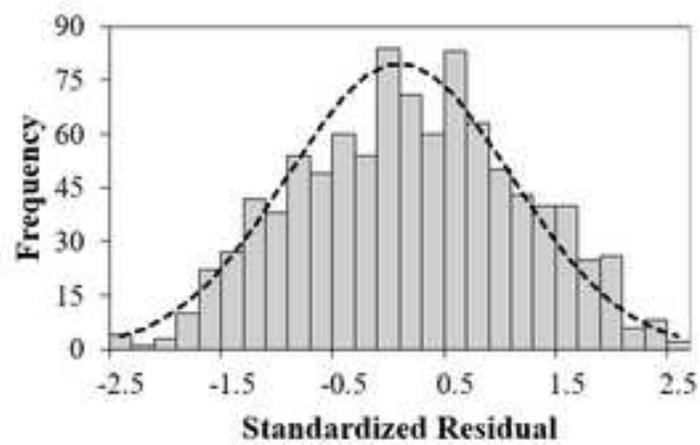
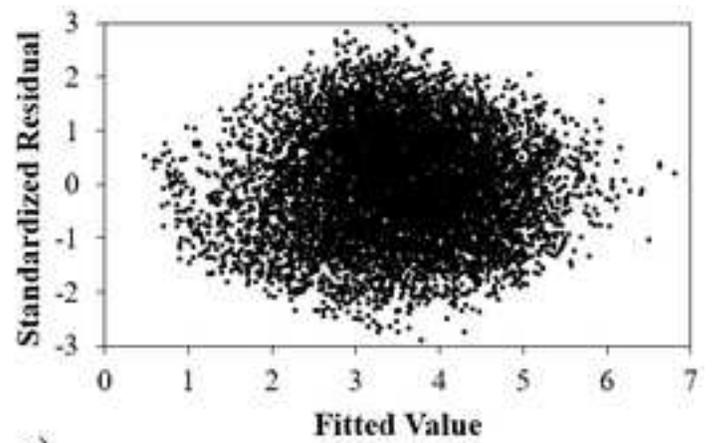
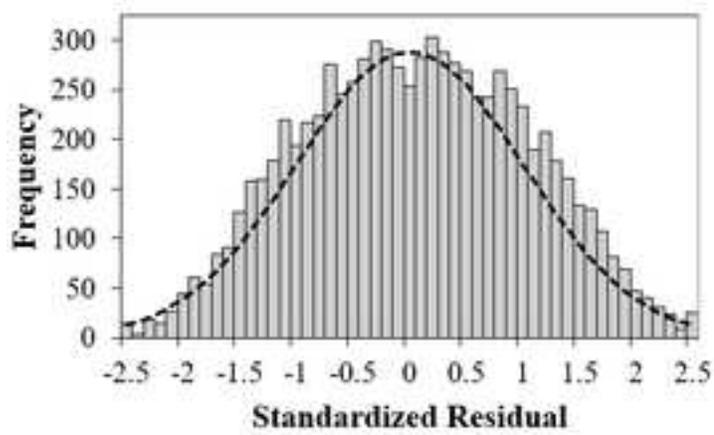
Table 3. Summary of the regression models to predict ET_o (mm/day) for each month and cluster

Month	CL	N	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	S	Pred. R^2
1	1	8309	1.131	-	0.029	0.008	-	-0.007	-0.007	0.464	0.070	96.72
2	1	7473	1.126	-	0.079	-0.006	-	-0.009	-0.009	0.463	0.102	96.96
3	1	8440	1.021	-	0.124	-0.010	-	-0.009	-0.014	0.514	0.174	95.68
4	1	8151	0.881	0.307	-	-0.132	-	-0.007	-0.021	0.553	0.208	95.31
5	1	1027	1.638	-	0.174	-0.035	-0.016	-	-0.020	0.545	0.195	96.19
	2	7070	1.261	0.266	-	-0.114	-0.008	-	-0.016	0.722	0.211	91.45
6	1	967	1.410	0.246	-	-0.069	-	0.002	-0.035	0.616	0.202	95.24
	2	6817	1.969	0.243	-	-0.102	-	-0.007	-0.014	0.673	0.177	89.61
7	1	3092	2.320	0.041	0.077	-	-0.009	-	-0.013	0.830	0.156	93.61
	2	4864	2.751	0.191	-	-0.079	-	-0.007	-0.019	0.692	0.161	86.11
8	1	2467	-0.014	0.038	0.126	-	-0.012	-	-0.009	0.976	0.240	92.64
	2	6175	-0.735	0.353	-	-0.144	-	-0.009	-0.015	0.823	0.237	83.04
9	1	8372	-1.189	0.346	-	-0.149	-	-0.005	-0.018	0.853	0.191	94.53
10	1	8471	-0.452	0.222	-	-0.092	-	0.002	-0.020	0.628	0.168	92.12
11	1	8518	0.568	0.016	0.052	-	-	-0.006	-0.009	0.538	0.082	96.07
12	1	8260	0.755	-	0.028	0.009	-	-0.005	-0.006	0.497	0.043	97.98



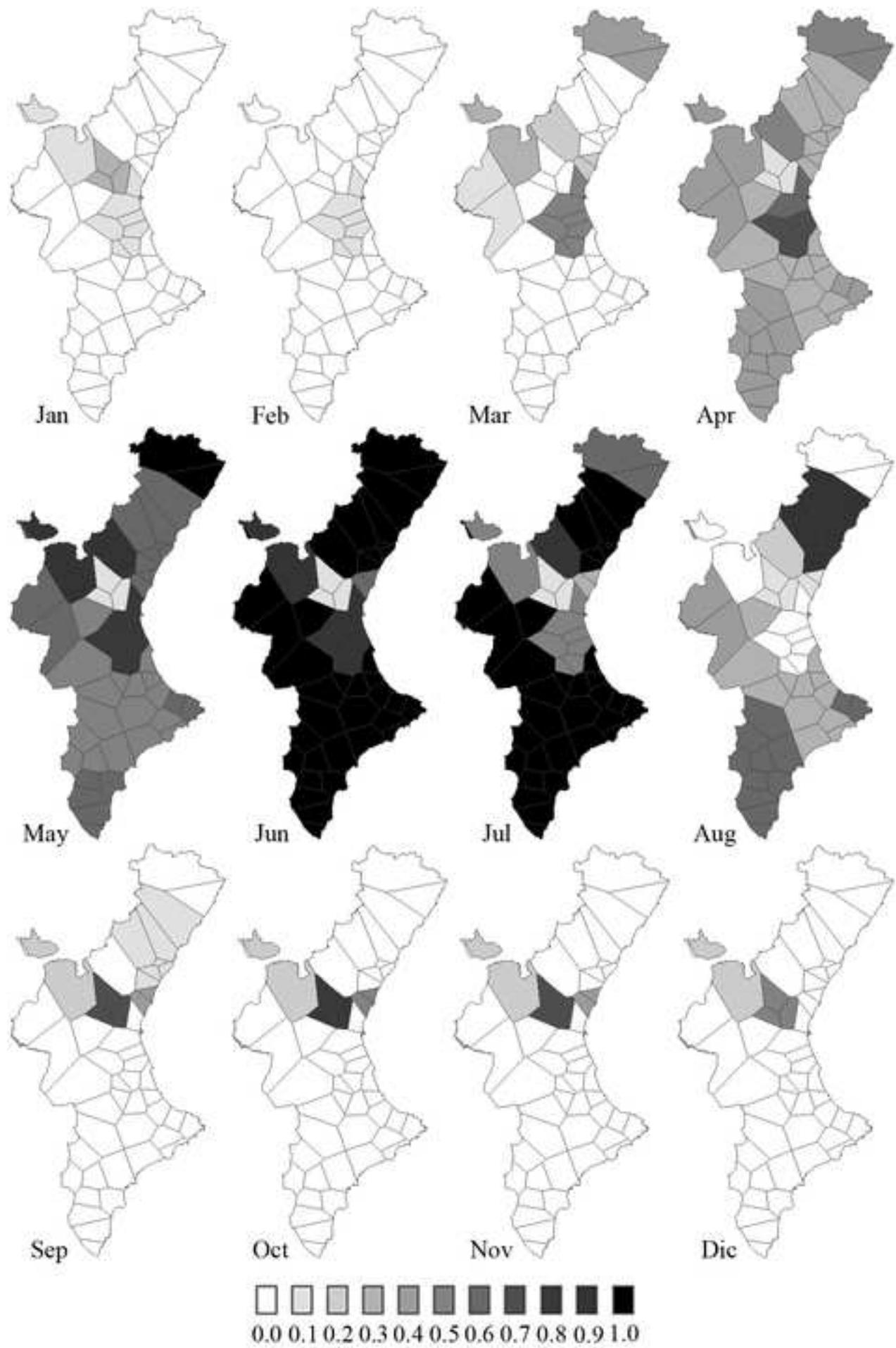


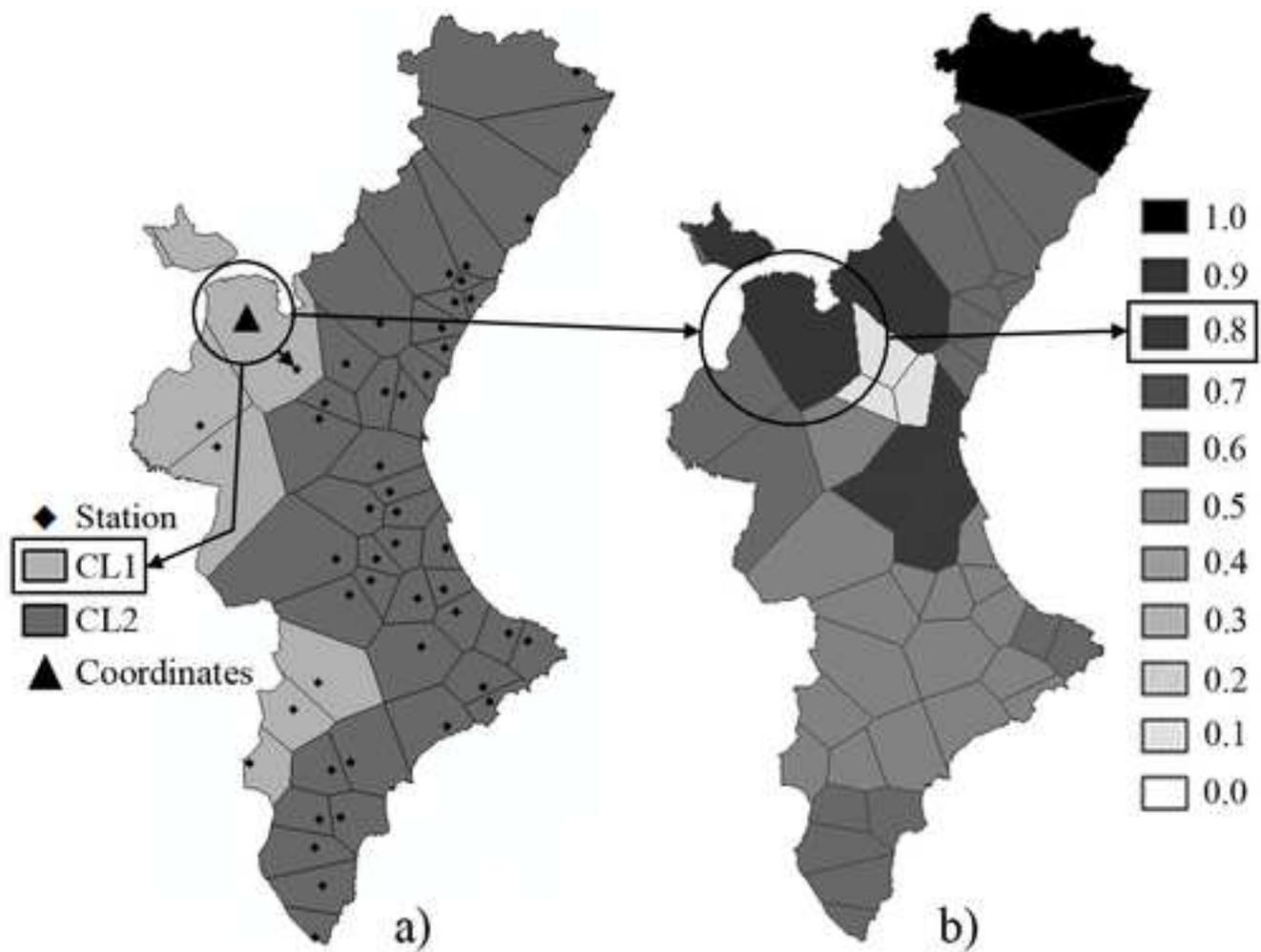




c)

----- Theoretical Normal Distribution





Month	T_{max} ($^{\circ}C$)	T_{min} ($^{\circ}C$)	RH_{mean} (%)	RH_{min} (%)	WS_{mean} ($m \cdot s^{-1}$)
May	23.84	11.23	61.62	32.45	1.66

c)



$$ET_o = 1.638 + 0.174 \cdot T_{max} - 0.035 \cdot T_{min} - 0.016 \cdot RH_{mean} - 0.020 \cdot RH_{min} + 0.545 \cdot WS_{mean} = 4.66 \text{ mm/day}$$

d)



$$ET = ET_o \cdot K_c = 3.73 \text{ mm/day}$$

e)

Figure 1. Location and provincial division of the Valencian Region

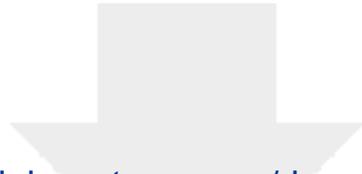
Figure 2. Clusters obtained for a) May b) June c) July d) August

Figure 3. Relationships between the predictors and the predictand (ET_o) in the regression model for April

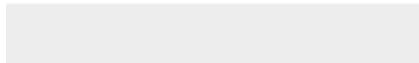
Figure 4. Histograms and scatterplots of standardized residuals against fitted values for a) April b) June - Cluster 1 c) June - Cluster 2

Figure 5. Monthly crop coefficients (K_c) in the Valencian Region for midseason potato

Figure 6. Estimation of ET in May for midseason potato in the coordinates (39°55'57'' N, 1°04'10'' W)
a) Cluster b) Monthly crop coefficient (K_c) c) Historical average values for the predictors in the closest station to the coordinates d) Calculation of ET_o (mm/day) e) Determination of ET (mm/day)



Click here to access/download
Supplemental Data File
M sizing Web Excel.xls



ASCE Authorship, Originality, and Copyright Transfer Agreement

Publication Title: Journal of Hydrologic Engineering

Manuscript Title: Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables

Author(s) – Names, postal addresses, and e-mail addresses of all authors

Daniel Jato-Espino, Av. de los Castros s/n, 39005 Santander, Spain, jatod@unican.es

Susanne M. Charlesworth, CV1 5FB, United Kingdom, apx119@coventry.ac.uk

Sara Perales-Momparler, Avda. del Puerto, 180-1B, 46023 Valencia, Spain, sara.perales@greenbluemanagement.com

Ignacio Andres-Domenech, Cno. de Vera s/n, 46022 Valencia, Spain, igando@hma.upv.es

I. Authorship Responsibility

To protect the integrity of authorship, only people who have significantly contributed to the research or project and manuscript preparation shall be listed as coauthors. The corresponding author attests to the fact that anyone named as a coauthor has seen the final version of the manuscript and has agreed to its submission for publication. Deceased persons who meet the criteria for coauthorship shall be included, with a footnote reporting date of death. No fictitious name shall be given as an author or coauthor. An author who submits a manuscript for publication accepts responsibility for having properly included all, and only, qualified coauthors.

I, the corresponding author, confirm that the authors listed on the manuscript are aware of their authorship status and qualify to be authors on the manuscript according to the guidelines above.

Daniel Jato-Espino



02 AUGUST 2016

Print Name

Signature

Date

II. Originality of Content

ASCE respects the copyright ownership of other publishers. ASCE requires authors to obtain permission from the copyright holder to reproduce any material that (1) they did not create themselves and/or (2) has been previously published, to include the authors' own work for which copyright was transferred to an entity other than ASCE. Each author has a responsibility to identify materials that require permission by including a citation in the figure or table caption or in extracted text. Materials re-used from an open access repository or in the public domain must still include a citation and URL, if applicable. At the time of submission, authors must provide verification that the copyright owner will permit re-use by a commercial publisher in print and electronic forms with worldwide distribution. For Conference Proceeding manuscripts submitted through the ASCE online submission system, authors are asked to verify that they have permission to re-use content where applicable. Written permissions are not required at submission but must be provided to ASCE if requested. Regardless of acceptance, no manuscript or part of a manuscript will be published by ASCE without proper verification of all necessary permissions to re-use. ASCE accepts no responsibility for verifying permissions provided by the author. Any breach of copyright will result in retraction of the published manuscript.

I, the corresponding author, confirm that all of the content, figures (drawings, charts, photographs, etc.), and tables in the submitted work are either original work created by the authors listed on the manuscript or work for which permission to re-use has been obtained from the creator. For any figures, tables, or text blocks exceeding 100 words from a journal article or 500 words from a book, written permission from the copyright holder has been obtained and supplied with the submission.

Daniel Jato-Espino



02 AUGUST 2016

Print name

Signature

Date

III. Copyright Transfer

ASCE requires that authors or their agents assign copyright to ASCE for all original content published by ASCE. The author(s) warrant(s) that the above-cited manuscript is the original work of the author(s) and has never been published in its present form.

The undersigned, with the consent of all authors, hereby transfers, to the extent that there is copyright to be transferred, the exclusive copyright interest in the above-cited manuscript (subsequently called the "work") in this and all subsequent editions of the work (to include closures and errata), and in derivatives, translations, or ancillaries, in English and in foreign translations, in all formats and media of expression now known or later developed, including electronic, to the American Society of Civil Engineers subject to the following:

- The undersigned author and all coauthors retain the right to revise, adapt, prepare derivative works, present orally, or distribute the work, provided that all such use is for the personal noncommercial benefit of the author(s) and is consistent with any prior contractual agreement between the undersigned and/or coauthors and their employer(s).
- No proprietary right other than copyright is claimed by ASCE.
- If the manuscript is not accepted for publication by ASCE or is withdrawn by the author prior to publication (online or in print), this transfer will be null and void.
- Authors may post a PDF of the ASCE-published version of their work on their employers' **Intranet** with password protection. The following statement must appear with the work: "This material may be downloaded for personal use only. Any other use requires prior permission of the American Society of Civil Engineers."
- Authors may post the **final draft** of their work on open, unrestricted Internet sites or deposit it in an institutional repository when the draft contains a link to the published version at www.ascelibrary.org. "Final draft" means the version submitted to ASCE after peer review and prior to copyediting or other ASCE production activities; it does not include the copyedited version, the page proof, a PDF, or full-text HTML of the published version.

Exceptions to the Copyright Transfer policy exist in the following circumstances. Check the appropriate box below to indicate whether you are claiming an exception:

U.S. GOVERNMENT EMPLOYEES: Work prepared by U.S. Government employees in their official capacities is not subject to copyright in the United States. Such authors must place their work in the public domain, meaning that it can be freely copied, republished, or redistributed. In order for the work to be placed in the public domain, ALL AUTHORS must be official U.S. Government employees. If at least one author is not a U.S. Government employee, copyright must be transferred to ASCE by that author.

CROWN GOVERNMENT COPYRIGHT: Whereby a work is prepared by officers of the Crown Government in their official capacities, the Crown Government reserves its own copyright under national law. If ALL AUTHORS on the manuscript are Crown Government employees, copyright cannot be transferred to ASCE; however, ASCE is given the following nonexclusive rights: (1) to use, print, and/or publish in any language and any format, print and electronic, the above-mentioned work or any part thereof, provided that the name of the author and the Crown Government affiliation is clearly indicated; (2) to grant the same rights to others to print or publish the work; and (3) to collect royalty fees. ALL AUTHORS must be official Crown Government employees in order to claim this exemption in its entirety. If at least one author is not a Crown Government employee, copyright must be transferred to ASCE by that author.

WORK-FOR-HIRE: Privately employed authors who have prepared works in their official capacity as employees must also transfer copyright to ASCE; however, their employer retains the rights to revise, adapt, prepare derivative works, publish, reprint, reproduce, and distribute the work provided that such use is for the promotion of its business enterprise and does not imply the endorsement of ASCE. In this instance, an authorized agent from the authors' employer must sign the form below.

U.S. GOVERNMENT CONTRACTORS: Work prepared by authors under a contract for the U.S. Government (e.g., U.S. Government labs) may or may not be subject to copyright transfer. Authors must refer to their contractor agreement. For works that qualify as U.S. Government works by a contractor, ASCE acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce this work for U.S. Government purposes only. This policy DOES NOT apply to work created with U.S. Government grants.

I, the corresponding author, acting with consent of all authors listed on the manuscript, hereby transfer copyright or claim exemption to transfer copyright of the work as indicated above to the American Society of Civil Engineers.

Print Name of Author or Agent

Signature of Author of Agent

Date

More information regarding the policies of ASCE can be found at <http://www.asce.org/authorsandeditors>

Cover Letter

Dear Editor,

We are sending the revised version of the research paper entitled “**Prediction of evapotranspiration in a Mediterranean region using basic meteorological variables**”, for your consideration to be published in **Journal of Hydrologic Engineering**.

Point-to-point responses to every comment made by the four reviewers and the associate editor are attached in five separate WORD files. We have made a great effort to address all the concerns posed by them, since we are very interested in publishing in HEENG. Thanks to their helpful suggestions, we believe that the quality of the manuscript has substantially improved, so we hope you continue considering our work for publication.

We look forward to receiving your opinion.

Yours sincerely,

Daniel Jato-Espino
Susanne M. Charlesworth
Sara Perales-Momparler
Ignacio Andrés-Doménech

The authors.

Response to the Associate Editor

Dear Sir or Madam,

Thank you for your comments on our manuscript. The detailed response to them are given below.

GENERAL COMMENTS

The manuscript has been evaluated by four referees. Reviewers agree that the contribution fits well to the scope of our journal and second version should be significantly improved by considering review comments. The authors should expand the literature section and clearly explain the contribution of their study. Why the authors use linear equations instead of GP or GEP for modeling non-linear ET process? The results section should also be expanded by explaining findings of the study. Based on the reviewer comments, my recommendation is “Revise for Re-Review (Technical Paper). The review comments should be carefully taken into account while preparing the revised version.

The literature review has been extended as suggested by Reviewer #3, in order to give more details about former similar studies (see lines 80-105). A new paragraph has also been included in the Introduction to highlight the contributions of our study in relation to these former similar works (see lines 107-122).

Although the relationships between climate variables and ET_o are nonlinear, as proved in current Figure 3, the linear combination of the former can provide accurate predictions on the latter (see lines 94-95). We agree that nonlinear techniques can be slightly more accurate than MLR, but these differences might not be significant, as demonstrated in the studies described in lines 94-105. In particular, we are presenting average results of pred. R^2 (equivalent to the R^2 obtained in the validation phase in ANNs) of more than 90% (see current Table 3), which speak for themselves. In addition, MLR are simpler and easier to understand and interpret than nonlinear methods such as ANNs, in which hidden layers are often added without really knowing why to improve the quality of the model and obtain higher values of R^2 . Actually, including too many hidden layers might lead to over-fitting of the model and result in misleadingly high values of R^2 .

We have extended and strengthened the discussion about the results obtained through the application of the proposed methodology (see lines 369-377, lines 387-399, lines 414-417 or lines 441-450, to cite some examples).

Reviewer #1

Dear Sir or Madam,

Thank you for your useful suggestions on our manuscript, they have led us to improve the quality of our paper substantially. The detailed responses to your comments are listed below point by point.

GENERAL COMMENTS

The paper is well organized and well written. It offers excellent and sufficient, if not exhaustive, literature review that provides a clear justification for the reported research. The authors discuss the complexity of the FAO- and ASCE-recommended Penman-Monteith equation for estimating the reference evapotranspiration, ETo , that is crucial (along with crop coefficients) for estimating irrigation needs. They correctly note that often there is not enough data to use this equation for estimating the local irrigation needs.

The authors resolve this problem by developing a simpler model with smaller number of input requirements. The price for the relative simplicity is the local nature of the model: as opposed to the Penman-Monteith equation that is valid for arbitrary location, the developed model is valid only locally for the Valencian Region--the overview of this region is well done. Although the proposed model is local, the methodology developed to specify its parameters can be applied to other regions.

In the development of their model, the authors combine skilfully cluster analysis, multiple linear regression, and Voronoi diagrams to basically derive separate submodels for the subregions of the Valencian Region. The subregions are defined by means of Voronoi tessellations.

Advancing applied knowledge, the paper is novel and should be of interest to hydrologists, soil physicists, agronomists, and irrigation engineers, as well as farmers planning their irrigation schedules.

PARTICULAR COMMENTS

Comment #1

Table 1: provide units to the numbers (or columns) for precipitation and temperature.

Table 1 has been modified to include units for precipitation (mm) and temperature (°C) (see line 182).

Comment #2

Table 3: ET_o (mm) ... ET_o is a rate, so it has to be measured in $L^3/TIME$ or $L^3/TIME/L^2 = L/TIME$. Specify the time unit you imply here (your Tables 4 suggests that your time unit is a day).

Current Table 2 (former Table 3) has been modified to specify the time unit of ET_o , which in this case was mm/month (see line 334), since the exploratory analysis associated with this table was done according to the same time horizon used for the calculation of the crop coefficient K_c , as explained in subsection 2.1 (see lines 145-150).

Comment #3

Expand their skimpy discussion of Figure 3, which contains the meat of their results. Expand the discussion of what these six plots tell you and the reader: spell it out; don't imply it.

The discussion of current Figure 4 (former Figure 3) has been expanded as you suggested, including more details about the quality of the regression models and the fulfilment of the assumptions on which they are based (see lines 403-417). Also a new paragraph (see lines 387-399) and a new figure (current Figure 3) have been added to provide more details about the quality of the prediction models determined.

Comment #4

Compare their new model's performance to that of the Penman-Monteith equation with the characteristic literature values assumed for the unmeasured parameters. Which model is better? Which one is more reliable?

The methodology proposed in the paper does not aim to outperform the results provided by the Penman-Monteith method, but rather replicate them using more accessible variables, in order to facilitate the calculation of ET_o from daily weather forecasts which do not include several variables considered in the Penman-Monteith equation. Actually, we state that the Penman-Monteith method is recommended by several organizations, such as FAO and ASCE, as a reliable and worldwide applicable approach for the estimation of

ET_o (see lines 70-73). In fact, we specify in lines 178-180 that the historical daily values of *ET_o* used in the paper to build the regression models were originally calculated through the Penman-Monteith equation.

Reviewer #2

Dear Sir or Madam,

Thank you for your useful suggestions on our manuscript, they have led us to improve the quality of our paper substantially. The detailed responses to your comments are listed below point by point.

Authors have developed a methodology to estimate evapotranspiration (ET) using limited meteorological data in Spain. Initially, they have grouped the weather stations based on their characteristics (central tendency and variability). Then for each group, a monthly regression model has been developed for each month to predict the reference evapotranspiration (ET_0). Finally, the ET was estimated by multiplying the crop coefficient (K_c) of that region, which is obtained using Voronoi diagrams. In general, the manuscript is well written. However, the methodology section deserves some more explanation for clarity and better understanding. The detailed comments are as follows:

Comment #1

Line 87: Recently, many researchers have derived equations (extracted the knowledge (weights and bias) gained during training) from ANN models, and used for future prediction. Other techniques like genetic programming (GP) and gene expression programming (GEP) will directly yield equations. Many studies have been reported using these techniques to predict ET.

We have reworded those lines to clarify that we mean that ANNs do not provide direct equations as MLR do (see lines 115-116).

Comment #2

Line 97: Why linear equations? ET is a non-linear process. A non-linear equation may result in much better prediction. Why MLR is chosen for prediction? Many recent techniques have been proved to be better than MLR.

We have added some explanations about this issue (see lines 94-95 and lines 107-109 and lines 397-399). Although the relationships between climate variables and ET_0 are nonlinear, as proved in current Figure 3, the linear combination of the former can provide accurate predictions on the latter. We agree that nonlinear techniques can be slightly more

accurate than MLR, but these differences might not be significant, as demonstrated in the studies described in lines 94-105. In particular, we are presenting average results of pred. R^2 (equivalent to the R^2 obtained in the validation phase in ANNs) of more than 90% (see current Table 3), which speak for themselves. In addition, MLR are simpler and easier to understand and interpret than nonlinear methods such as ANNs, in which hidden layers are often added without really knowing why to improve the quality of the model and obtain higher values of R^2 . Actually, including too many hidden layers might lead to overfitting of the model and result in misleadingly high values of R^2 . We have added a new paragraph to justify the choice of MLR (see lines 107-122).

Comment #3

Line 121: The value of K_c varies for different stages of crop growth. Does this monthly time period is in agreement with crop growth stages?

Yes, the values of K_c vary according to the month, which in turn depends on the growth stages of the crop.

Comment #4

Line 171-173: How the daily data of each variable were arranged monthly? For example, the January month data of each year was separated and arranged (January month alone) chronologically? Explain in detail.

As we explain in lines 145-150, since a monthly period was chosen for the estimation of K_c , the models for the prediction of daily ET were also built according to such a time horizon. Therefore, we extracted the daily values for the seven predictors for each month of all the years available in each station located in the Valencian Region. The chronological order is not relevant for the application of the methodology, since the aim is to predict ET for a single (and random) future day in a month (January, for example). This “arrangement” is just a division of the whole dataset in each station, which consists of several years of daily data (the exact number of years depends on the station), in months.

Comment #5

It is also not clearly explained why a separate regression models/equations is required for each month. Instead you can have a single equation for all the months with different K_c values for each month.

This can be explained through current Table 3 (former Table 4). The coefficients associated with the predictors and their values vary according to the month, which is consistent with the fact that the weather characteristics change throughout a year (e.g. increased temperature in summer months, etc.). We have added some lines to highlight this (see lines 366-369).

Comment #6

Line 215: I think clustering is done only for ET (based on properties of ET). However, it is not clearly mentioned in the text. What is i , j and p ? 'ip' and 'jp' are confusing. Are you finding the euclidean distance between two points or between the centroid of cluster and a point? In line 318-318, it mentioned that regionalisation was done based according to the weather characteristics. However, it is not clearly mentioned whether it is based on only ET or all the parameters used in this study.

As a result of the suggestions made by other reviewer, we have shortened some explanations in section 2, including the equation related to the Euclidean distance (former Eq. (4)). As specified in lines 231-232, the Euclidean distance is calculated between each point and the centroids of the clusters identified, in order to assign each station to the closest cluster.

Regionalisation (e.g. cluster analysis) was carried out according to the weather characteristics of the stations in terms of the values they recorded for the set of predictors used. The aim of the paper is to provide a methodology for the prediction of *ET* using only basic meteorological variables, so the clustering of the study area must be done according to these parameters. We have reworded that sentence you mention to clarify it (see lines 318-321).

Comment #7

Line 213: Is 'k' subjective/ arbitrary? Have done sensitivity analyses on 'k'. How to fix 'k'? It is mentioned in line 336-338, that 'the number of clusters chosen was calculated to maximise the predictive 2.' How? Do you have any separate method/algorithm for this? How this is done within cluster analysis?

Although k was not set arbitrarily, we did not develop any algorithm to automate the optimization of the number of clusters. As explained in lines 338-342, we built the regression models with different numbers of the clusters and calculated their corresponding pred. R^2 , in order to select the number of clusters that maximised it. We found out that

pred. R^2 was maximised for 1 cluster in all cases and then started to gradually decrease as the number of clusters increased (2, 3, 4...), except for May, June, July and August, where pred. R^2 was maximised for 2 clusters and then started to gradually decrease as the number of clusters increased (3, 4, 5...). We have included an additional sentence to clarify this (see lines 342-343).

Comment #8

Line 236: x_i and x are the smallest value and mean of clusters or whole sample? What is 'n' in Eq 5. Is it number of points in cluster or sample size?

x_i and x are the smallest value and mean value of the whole sample used to test normality. Consequently, n was the number of points in such a sample. However, as a result of the suggestions made by other reviewer, we have shortened some explanations in section 2, including these equations related to the Shapiro-Wilk test.

Comment #9

Multiple linear regression model: Change the variable notations. The manuscript is not consistent with notations. Same notations are used at different places.

Since we have removed the equations corresponding to the Shapiro-Wilk test, x_i and x are now only used for MLR. We have changed the notation in the equation for the Cook's distance (current Eq. (5)), which now includes z_j and $z_j(i)$ instead of y_j and $y_j(i)$ (y was also the response in the regression equation).

Comment #10

There are lot of uncertainties associated with the meteorological variable, which is not properly addressed in this study. Table 4: The random component (ε) is missing, without which how will you estimate the future predictions?

ε refers to the residuals, i.e. the distances from the fitted values to the hyperplane defined by the multiple linear regression models. In other words, they indicate the error of prediction in the regression models, so that $\varepsilon = 0$ if $R^2 = 100\%$. Although $\varepsilon \neq 0$ in our models, the errors are very small because the values of pred. R^2 are around 90% on average, which demonstrates their predictive potential. Furthermore, we have analysed in great detail the characteristics of the residuals (see lines 403-417 and current Figure 4) and guaranteed they met all the hypotheses required to validate multiple linear regression

analysis, so we believe we have thoroughly demonstrated the reliability of our models and their capability to make future predictions.

Lines 441-450 and current Figure 6 have been added to clarify and demonstrate the applicability of the regression models.

Comment #11

The predictive R^2 needs much clear explanation. How it overcomes the drawbacks of standard R^2 . Give equation for estimating predictive R^2 .

We have added some lines to highlight the benefits provided by the predictive R^2 in relation to the standard R^2 and the adjusted R^2 (see lines 259-265). The explanation of how it overcomes the inability of the two other coefficients was already explained in the previous version of the manuscript according to three steps: (1) remove each observation from the dataset, (2) estimate the regression equation without the removed observation and (3) determine how well the model predicts the removed observation. Thus, this process includes the estimation of new data in the calculation of the regression models and their corresponding pred. R^2 , which ensures their capability to predict future values. It uses the same equation than the standard R^2 , so that what changes is related to the inclusion of the abovementioned three-step process. That equation is widely known among engineers, so we believe it is not necessary to specify it, unless the Editor considers it is really necessary. Instead, we have added a reference which can be consulted in case anyone wants more details about the R^2 coefficient (see line 258).

Comment #12

Is your K_c value varies based on clusters or weather stations?

K_c varies depending on the station, as explained in line 423 and line 430.

Comment #13

Some stations have two clusters for summer months, especially the stations in coastal regions. Therefore, there will be two equations for these months and only for these stations. However, the Table 4 doesn't show like this. Are the equations same for all the stations? Also, among two equations, which one has to used and for which station?

Some months (not stations) have two clusters for summer. The purpose for clustering is precisely to group the set of stations according to the similarity in the values they recorded for the predictors. Figure 2 illustrates this pretty well: the polygons are the Voronoi regions associated with the set of stations, whereas the clusters are identified according to the shades of grey (light grey: Cluster 1; dark grey: Cluster 2). Therefore, the equation summarised in the row for the 5th month and the 2nd cluster in current Table 3 (former Table 4) is to be applied in any location enclosed by the dark grey areas in Figure 2a).

The remaining months had only one cluster, which means that all the stations under consideration belonged to the same cluster (the values they recorded for the basic meteorological variables used as predictors were similar enough as to assume that).

Comment #14

Table 4: Why the number of days varies for each cluster in May, June, July and August? How the number of days is obtained for each month?

The number of days in each month depends on data availability in the stations, e.g. one station might have started to work in July and, therefore, it wouldn't include data about May and June of that year. Besides, outliers or influential points were removed from the datasets associated with each month using the Cook's distance. The number of points discarded can also slightly vary depending on the month. In any case, the number of days is about 8,000 in all cases.

Comment #15

Line 363: Why only 5 predictors for each month? What about other two predictors? This also varies for different months? Authors have to do impact analysis of each input variable in their model.

The stepwise process mentioned in lines 358-359 demonstrated that 5 was the optimal number of predictors for each month to maximize the accuracy of the regression models (i.e. the values of pred. R^2) without having problems of multicollinearity (see lines 414-417). This means that including one more predictor resulted in problems of multicollinearity (VIF values above 10), whilst excluding one more predictor resulted in a loss of precision (decrease in R^2). Hence, the process for selecting the number of predictors was accomplished very carefully based on statistical considerations. Although the selected

predictors varied for some months, they always consisted of two temperature-related variables (mean and min, mean and max or min and max), two humidity-related variables (mean and min, mean and max or min and max) and mean wind speed.

We have also added some lines about the contribution of the predictors to the estimation of the predictand (see lines 369-377). The most influential predictors were those related to temperature in general, with the exception of the colder months, wherein relative humidity and wind speed proved to be the greatest contributors for the estimation of *ET_o*.

Comment #16

Line 366-371: These statements are general. Give reference to these sentences?

We have added references to those statements according to your suggestion (see lines 381-385).

Comment #17

References are not according to the style of ASCE. For all web pages in the references, give the date of access.

We have modified both the references in text and the list of references (see lines 496-673) according to the style of ASCE. We have also included the date of access for webpages (see line 500 and lines 623-624).

Reviewer #3

Dear Sir or Madam,

Thank you for your useful suggestions on our manuscript, they have led us to improve the quality of our paper substantially. The detailed responses to your comments are listed below point by point.

The paper presents a methodology for the prediction of evapotranspiration based on weather forecasts. In addition, the authors applied this method to the Valencian region in Spain. The methodology was explained clearly but the following comments need to be addressed:

Comment #1

In Introduction section, literature review should be expanded.

The literature review has been extended as you suggested, in order to give more details about former similar studies (see lines 80-105).

Comment #2

Study area section should be added and explained in detail by using a map which shows the selected area in Spain.

A new figure (current Figure 1) has been added to show the study area in relation to the Map of Spain (see line 316).

Comment #3

The authors should give a brief explanation for the terms of e_a and e_d which are given in Eq.2 on page 6.

The definition of the term $(e_a - e_d)$ has been rewritten, so now it reads: “ $(e_a - e_d)$ is the difference between the actual (e_a) and saturation (e_d) vapor pressure (kPa)” (see line 163).

Comment #4

On page 6, Equation 3 is incorrect. The correct form is given on web page of FAO as [...].

Psychrometric Constant

Reference: Brunt (1952)

$$\gamma = \frac{c_p P}{\epsilon \lambda} \times 10^{-3} = 0.00163 \frac{P}{\lambda}$$

Yes, there was a typo error in Eq. (3) (we “forgot” one zero). It has been modified according to your comment (see lines 166-167).

Comment #5

On page 8, Figure 1 should be more understandable.

Former Figure 1 has been removed as a result of one of the comments made by other reviewer, since the detailed description provided in lines 193-210 is enough to understand the main steps carried out to develop the proposed methodology.

Comment #6

On page 10, Equation 4 is incorrect, as well. It should be corrected.

The Euclidean distance between two points is the square root of the sum of their squared differences, which is what former Eq. (4) represented. In any case, as a result of the comments made by other reviewer, we have shortened some descriptions in section 2, including the equation for the Euclidean distance (former Eq. (4)).

Comment #7

There is no need to write the full name of MAGRAMA on page 17 as it is already given on page 7.

The full meaning of MAGRAMA has been removed from that page (see line 426).

Reviewer #4

Dear Sir or Madam,

Thank you for your useful suggestions on our manuscript, they have led us to improve the quality of our paper substantially. The detailed responses to your comments are listed below point by point.

The paper addresses the prediction of evapotranspiration rates from crops, which is an interesting and very useful study. I therefore find the authors approach interesting and relevant, but I also have some difficulties with the current manuscript. Unfortunately, there are little details on the findings and discussion of the results from the study and very little evaluation if the approach taken is working. I think the paper needs to be tightened up and more focus should be on the results and their application. So my main suggestions are as follows.

Comment #1

You do not need both figure 1 and the detailed description from line 171->. My suggestion would be to remove the figure.

Former Figure 1 has been removed according to your suggestion.

Comment #2

Much of 2.3 – 2.5 contain basic text book information that could be left in the referenced literature. These sections could be reduced to cover only info critical for your use of the methods. E.g. table 2 could be removed since this covers just basic requirements for the regression application. Similar goes for the Cooks distance, the formula is shown but I can't find any results in the paper computed from this formula (except it is used to remove points if necessary – was it necessary?). The Voroni chapter is too long, I think most hydrologists would understand if you just stated you used Thiessen polygons for the cluster boundaries.

We have reduced subsections 2.3, 2.4 and 2.5 as you suggested. The detailed changes are listed below:

- **2.3. Cluster analysis:** removal of bullet points (see lines 230-233) and shortening of the explanation about the Shapiro-Wilk test (see lines 240-243).
- **2.4. Multiple linear regression:** removal of former Table 2 and shortening of the explanation about the assumptions of MLR (see lines 283-286).
- **2.5. Voronoi diagrams:** removal of former Eqs. (7) and (8) and their associated descriptions in text (see lines 304-305).

The equation of the Cook's distance was used to detect and remove influential points (outliers) as specified in lines 359-360.

Comment #3

I am not sure why the Penman-Monteith equation is shown, is it to illustrate the data needs? But on the other hand it is interesting to see the equation since it is an adaptation to the standard version on the P-M equation found in text books.

The Penman-Monteith is shown for two reasons. First, to highlight the great amount of parameters it requires and the need to develop alternative and simpler methods to estimate ET_o . And second, because the historical daily values of ET_o used in the paper to build the regression models were originally calculated through the Penman-Monteith equation. So we are actually trying to replicate them using more accessible variables, in order to facilitate the calculation of ET_o from daily weather forecasts.

Comment #4

I miss a discussion of the accuracy of the method, e.g. by leaving some stations out of the analysis and then testing the prediction of the simplified method on these data. This would strengthen the understanding of the goodness of the method which is important. Now the output is figure 4 with little discussion on it's content and if the values are reasonable.

We precisely used the predictive R^2 as a goodness-of-fit statistic because it is based on that principle you mentioned, as we explain in lines 267-269: "(1) remove each observation from the dataset, (2) estimate the regression equation without the removed observation and (3) determine how well the model predicts the removed observation". So the results we are presenting in Table 3 included these considerations already. To further clarify it, we have extended and strengthened the discussion about the statistical accuracy of the models summarized in current Table 3 (see lines 387-399).

Comment #5

You state (l.366) that the relationships were generally logical. Were there cases where they were not, and if so when and why?

They were logical in all cases. According to the values of the Beta coefficients in current Table 3, the mean values of temperature, relative humidity and wind follow the physical relationships explained in lines 379-385 in all cases (for every month and cluster). We have reworded that sentence to make it clear (see line 379).

Comment #6

Did you consider methods to evaluate the significance of the regression variables with the purpose of reducing the number of variables? Is the difference in results between a model of all predictors and a model of e.g. mean predictors large?

The variables included in current Table 3 as predictors were all statistically significant (p-values < 0.05) (see line 358). The differences between choosing more or less predictors were not high. The number of predictors was selected with the aim of both reaching the highest possible value of R^2 and avoiding multicollinearity (see lines 414-417).

Comment #7

You tested for normality (line 320). You could state that more clearly on line 379 which now states that the graphs suggest that the assumption of normality is ok.

The discussion about Figure 2 has been enhanced, including more details about the fulfilment of the assumption of normality (see lines 403-417).

Comment #8

Space constraints (l.390) limited the study to one crop type. If you can reduce the introductory material could you get space for more crop types, or do I misunderstand this statement?

Yes, you understood it well. Unfortunately, although we have reduced Section 2 and 4, Section 1 and 3 were actually extended as a result of the comments made by the remaining reviewers, so the situation has even worsened in this sense. Anyway, we understand that

limiting the paper to one crop type (as an example) does not limit the scope of our research as both methodology and results are directly replicable for every crop type.

Comment #9

L.402 – 408 is not very clear to me. You talk about water demands and combined regression results with crop factors, but no specific results are shown and no conclusions are drawn from this except that it is possible to do it. Does it produce useful results and based on your results is this a method ready for practical use? This is potentially a central component of the paper that needs more detail.

Lines 441-450 and current Figure 6 have been added to clarify and demonstrate the applicability of this part of the results.

Comment #10

You discuss data availability in more general terms in the intro. Have you considered the application of measured evapotranspiration over reference crops a potential future source? Much work is going into e.g. the fluxnet cooperation. Similarly, many forecasts today provide humidity and basics for radiation estimations, could this combined with reanalysis data be a potential for the future?

Yes, measured evapotranspiration from FLUXNET might be a source from which to build prediction models to estimate *ET*, as we did in this paper with the values of measured *ET* provided by the Spanish Ministry of Agriculture, Food and Environment. Humidity (relative humidity) is already included in the models we are presenting in this paper. As for radiation, it might be estimated as you say and included in the prediction models, but that would involve adding more error in the eventual prediction of *ET*. Besides, very accurate models can be obtained without requiring radiation-related variables, as we prove in the results of this paper.

Comment #11

The conclusion is long and in parts more of a discussion. It should be shorter and more concise, and the elements of discussion or summaries of the work belongs in the results – discussion section.

The conclusions section is 325 words long, which seems quite reasonable from our perspective. In our opinion, a potential reader can get an overview of the whole article with

these conclusions, since paragraph 1 summarises it, paragraph 2 provides evidence of the technical performance of the methods used and paragraph 3 describes the potential uses of the research behind the paper. We have reworded the second paragraph according to your comment (see lines 467-476), in order to avoid giving specific details which might be more characteristic of the results & discussion section.

Comment #12

Line 62: The value of 60% is the global average and the text should state this. E.g. in northern latitudes the percentage is significantly lower than this.

We have modified that sentence according to your comment (see line 62).

Comment #13

I miss a small map of Spain inserted into fig 2 to show where your region is, and as an aid to see inland and coastal areas.

We have added a new figure (current Figure 1) to show the study area in relation to the Map of Spain, including the location of the Mediterranean Sea (see line 316).