Agricultural Water Management 300 (2024) 108903

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Seasonal assessment of the grass reference evapotranspiration estimation from limited inputs using different calibrating time windows and lysimeter benchmarks

Pau Martí [a,*], Ramón López-Urrea [b], Luis A. Mancha [c], Pablo González-Altozano [d], Armand Román [a]

[a] Departament d'Enginyeria Industrial i Construcció, Àrea d'Enginyeria Agroforestal, Universitat de les Illes Balears, Carretera de Valldemossa km 7.5, Palma 07122, Spain

[b] Centro de Investigaciones sobre Desertificación (CIDE), CSIC-UV-GVA, Carretera CV-315, km 10.7, Moncada, Valencia 46113, Spain

[c] Servicio de Regadíos, Consejería de Gestión Forestal y Mundo Rural, Junta de Extremadura, Avda. Luis Ramallo s/n, Mérida 06800, Spain

[d] Departament d'Enginyeria Rural i Agroalimentària, Universitat Politècnica de València, c/Camí de Vera s/n, València 46022, Spain

ARTICLE INFO

Keywords:
FAO 56 methodology
FAO 56 Penman-Monteith
Hargreaves-Samani
Meteorological variables
Large weighing lysimeter
Grass reference surface

ABSTRACT

Models relying on limited inputs are very valuable for estimating reference evapotranspiration, and subsequently irrigation doses, but their accuracy can be very dependent from calibration. This study assessed three versions of the Hargreaves-Samani (HS) and the FAO Penman-Monteith (PM) equations to estimate reference evapotranspiration (ET_o), relying respectively on three input combinations. Further the six models were adjusted each using different time windows for calculating the calibrating constants, namely global, annual, monthly, fortnightly, and weekly constants, while all the models were calibrated and tested using calculated and lysimeter benchmarks. The models relying on mean air temperature and solar radiation tended to be more accurate than those relying on mean air temperature and relative humidity, while these tended to be more accurate than those relying on air temperature difference, but there might be intra annual exceptions according to the monthly indicators. The errors of the PM estimations were just slightly higher than those of the corresponding HS estimations. The accuracy improvement in the calibrated versions was higher the shorter the time window used for averaging the calibrating parameters. Thus, the application of monthly or, at least, seasonal calibrating constant might be recommended for a suitable correction of the bias. During the year, the estimations presented markedly lower errors and lower differences within models during the summer. The error decrease in the calibrated versions was more marked during the winter. The assessment relying on lysimeter benchmarks provided similar qualitative patterns than the assessment relying on calculated benchmarks, but the corresponding error ranges were higher. Finally, 6 examples were presented for visualizing the effect of the method used to estimate ET_o on the corresponding resulting average annual crop water requirements. If irrigation scheduling is based on a soil water balance using crop evapotranspiration estimates, at least, a monthly bias assessment of the ET_o estimates in combination with the crop cycle lengths and dates might contribute to infer if crop water requirement infra-estimation trends are identified during crop sensitive stages to water deficit.

1. Introduction

Evapotranspiration (ET) is a crucial component of the hydrological cycle at all scales, from plot to basin, and its estimation is needed in countless applications from different branches, among others, for improving water resource planning and management in response to climate change. The expected world population growth requires increases in food production derived, among others, from yield increases and better water use efficiency in irrigated agriculture. On the other hand, water scarcity has increased due to a combination of more frequent droughts and increasing competition for water resources among agricultural, industrial and urban users (Schultz et al., 2005; Bachour et al., 2013). Therefore, sophisticated irrigation management is necessary to optimize water use efficiency and maintain levels of crop

* Corresponding author.
E-mail address: pau.marti@uib.es (P. Martí).

https://doi.org/10.1016/j.agwat.2024.108903
Received 6 March 2024; Received in revised form 24 May 2024; Accepted 27 May 2024
Available online 4 June 2024
0378-3774/

productivity and quality (Ortega Farias et al., 2009). A way to contribute to this aim in agricultural water management is using accurate enough estimates of crop water requirements, which calculation relies on accurate knowledge of crop evapotranspiration (ET_c). Due to its simplicity, a very common method for estimating ET_c at field scale, introduced by Doorenbos and Pruitt (1977), consists in calculating ET_c as the product of a crop coefficient (K_c) and the grass reference ET (ET_o) (Allen et al., 1998; Pereira et al., 2021). ET_o is a climatic parameter expressing the evaporation power of the atmosphere, while K_c represents the crop-specific effects on ET. This approach intends to guide and 'protect' against large over- and under-estimation of ET (Pereira et al., 2015).

ET can be measured using weighing lysimeters, which determines ET on the basis of the measurement of some of the components of the water balance in a controlled crop area (Gavilán et al., 2006). Nonetheless, although weighing lysimeters and other measurement systems such as eddy covariance flux towers, if well managed, provide accurate ET data for short time periods, both have a number of limitations and requirements that may hinder their use for monitoring ET (e.g. fetch requirements, high cost of lysimeters, complexity of data processing of eddy covariance systems, etc.) (Allen et al., 2011a). This is not only due to their cost and complexity, but also because the limited area of a typical weather station enclosure does not provide sufficient fetch from a representative surface for these measurements to be meaningful (Sentelhas et al., 2010). As a result, measured ET records are not available in most cases. This lack in combination with an increasing availability of improving networks of meteorological stations has led to the development of a wide variety of calculation methods in the last decades.

Jensen et al. (1990) concluded that a single, physically-based method might be adopted to estimate ET_o. In contrast to FAO 24 (Doorenbos and Pruitt, 1977), the FAO 56 guideline (Allen et al., 1998) recommended as the standard method to estimate ET_o, the FAO 56 Penman-Monteith equation (FAO 56 P-M). This equation is largely accepted to serve as a basis for ET_c calculation globally, as it has been tested worldwide against local ET measurements, while different sensitivity analyses and regional studies have confirmed its applicability to a large variety of environments (Pereira et al., 2015). Allen et al. (1998) define ET_o as the rate of evapotranspiration from a hypothetical reference crop with an assumed crop height of 0.12 m, a fixed daily surface resistance of 70 s m^{-1}, and an albedo of 0.23, closely resembling the evapotranspiration from and extensive surface of green grass of uniform height, actively growing, completely shading the ground and not short of water. The American Society of Civil Engineers standardized reference evapotranspiration equation (ASCE-PM ET_{ref}) (ASCE-EWRI, 2005) resulted from the standardization of the PM equation for clipped grass and alfalfa surfaces with similar parameterizations as FAO 56 for the equation components for daily calculations (Pereira et al., 2015). The parameters of the FAO 56 P-M equation follow standardized procedures. (Nandagiri and Kovoor, 2006) showed the need for strict adherence to the recommended procedures proposed by Allen et al. (1998), especially for estimating vapour pressure deficit and net radiation. Further, weather data should be of good quality a represent weather conditions over a green grass reference area, as previously defined. This equation can be relatively sensitive to error in weather data (Pereira et al., 2015).

The FAO 56 P-M equation requires data on maximum and minimum air temperatures (T_{max} and T_{min}), solar radiation (R_s), air humidity, and wind speed at 2 m height. However, in many locations these weather variables are not observed, are not freely available from the relevant meteorological services, or are of poor quality due to insufficient quality control (Paredes et al., 2020). Typical sensors required for measuring a full set of data in ET_o automated weather stations have high costs (Valiantzas, 2012, 2013, 2018; Exner-Kittridge and Rains, 2010; Exner-Kittridge, 2011), which is particularly dramatic in developing countries. Wind data are often lacking or are of low or questionable quality (Jensen et al., 1990; Allen, 1996), and there are no methods to predict wind speed with total confidence. Solar radiation is not routinely measured at many weather stations, or its measurement are not always reliable, and therefore it may need to be estimated. However, relative humidity (RH) is easily measured, requiring low additional cost (Valiantzas, 2018).

So, it is very usual that only reduced data sets are available, often consisting exclusively of T_{max} and T_{min}. Further, the study and development of temperature-based methods for ET_o estimation is justified for several reasons. First, temperature and solar radiation explain at least 80 % of ET_o variability (Priestley and Taylor, 1972; Samani, 2000). Second, several studies indicate that daily temperature range can be related to relative humidity and cloudiness (Samani and Pessarakli, 1986; Shuttleworth, 1993; Di Stefano and Ferro, 1997). Third, advection depends on the interaction between temperature, relative humidity, vapor pressure, and wind speed, and these variables can be related to the temperature range (Vanderlinden et al., 2004). Finally, temperature is the most wide-spread monitored variable among those needed for ET_o estimation (Mendicino and Senatore, 2013). Further, air temperature can be measured with less errors and by less trained individuals than the other required climate variables used in combination equations (Raziei and Pereira, 2013). Thus, a wide spectrum of alternative methods relying on reduced data sets and considering very different computational approaches have been proposed and tested in different climatic scenarios to overcome the unavailability of data. The validation of alternative equations against FAO 56 P-M computed with full weather data sets has been amply addressed. However, most of them do not consider the basic physics underlying the FAO 56 P-M equation, and aimed to provide simplified tools that produce results similar to FAO 56 P-M. Therefore, rather than modifying the basic method itself or fitting methods to data, FAO 56 recommended estimating missing data and retaining the use of the P-M method, because this retains the physical basis for calculation and interactions among weather parameters (Pereira et al., 2015). But many studies have tended to overlook this recommendation.

When full data sets are not available, ET_o can be estimated according to Allen et al. (1998) either using the empirical Hargreaves-Samani (HS) equation or the FAO 56 P-M with estimations of the missing inputs, including using data from neighbour weather stations (Allen, 1997). The FAO 56 P-M equation, when applied using only measured temperature data (PMT), requires a somewhat heavier computation and data preparation than the HS method. Both the HS and PMT methods have received a continuous attention from research contrarily to the use of neighbour weather data (Raziei and Pereira, 2013).

Actual vapour pressure (e_a) can be computed relying on T_{min}, i.e. assuming that the dew point temperature (T_{dew}) could be replaced by T_{min}, when relative humidity or psychrometric data are not available. This approach is not valid when observations correspond to non-reference sites, and when sites are affected by dryness and local advection, which cause that $T_{min}>T_{dew}$ (e.g. Paredes and Pereira, 2019). In these cases, T_{min} requires correction. In this regard, many applications refer to sites where information on grass cover conditions is limited (Paredes et al., 2020). In agreement with the recommendation of using HS equation for ET_o estimation, Allen et al. (1998) proposed to estimate R_s with such equation too, as it is based on a specific equation for estimating incoming solar radiation from the temperature difference (ΔT). In absence of wind speed data, two options are suggested, namely i) the use of the world average wind speed value (u_2) 2 m s^{-1} as a default estimator (Allen et al., 1998), or ii) the use of average local or regional wind speed data (e.g. Popova et al., 2006; Paredes et al., 2018a; Paredes et al., 2018b).

PMT has been demonstrated to produce low errors if, like HS, a calibrated constant (k_{Rs}) is used to estimate solar radiation, and if temperature is adjusted to overcome the effects of site aridity (Pereira et al., 2015). Many studies assessing the PMT approach reported good accuracies when compared with full data FAO 56 P-M, while other studies reported quite good accuracy of ET_o estimates using estimated values of T_{dew}, R_s and u_2 (Paredes et al., 2020).

Literature seems controversial when comparing HS and PMT results (Raziei and Pereira, 2013, Paredes et al., 2020). A higher accuracy of the PMT equation over HS or other temperature-based equations has been reported by several authors, namely for climates marked by humidity (e.g. Pandey and Pandey, 2016; Ren et al., 2016). But other studies reported the superiority of the HS equation (Singh et al., 2018). In other cases, a better performance of the HS equation over PMT was also reported (e.g. Martinez and Thepadia, 2010).

Due to its simplicity and easy application, and according to the recommendation of Allen et al. (1998), the HS equation (Hargreaves and Samani, 1982, 1985; Hargreaves et al., 1985; Hargreaves, 1994) has become the most popular approach to estimate ET_o when only reduced data sets are available. The HS equation, in its original form, only requires measured mean air temperature and temperature difference, in addition to calculated extraterrestrial radiation. Although accurate daily estimates have been reported with this equation, Hargreaves and Allen (2003) stated that the best HS estimates might be expected for 5-day or longer periods, because daily estimations are subject to higher variability caused by the movement of weather fronts and by large variations in wind speed and cloud cover. Shuttleworth (1993) even recommended not to use shorter periods than one month. Nevertheless, numerous agricultural and hydrological applications require daily ET_o data. This equation has been validated with calculated FAO 56 P-M and lysimeter targets providing a reasonably accurate performance in most climatic regions, with the exception of humid climates, under advective conditions and in mountain or high elevation environments (e.g. Jensen et al., 1990; Itenfisu et al., 2003; Berengena and Gavilán, 2005; Temesgen et al., 2005; Trajkovic, 2007). Other studies found a tendency to overestimate it at low evapotranspiration rates and vice versa (e.g. Droogers and Allen, 2002; Xu and Singh, 2002). In those conditions HS may not fit well, because they might be quite different from those considered for its calibration, i.e. relying on data from arid to sub-humid environments, as well as due to not considering the input parameter wind speed (Raziei and Pereira, 2013). The performance of the original HS equation is strongly influenced by the climatic conditions where it was developed, and should not be extrapolated to different climatic conditions unless it is first calibrated (Samani, 2000; Maestre-Valero et al., 2013). Accordingly, many users calibrate the HS parameters adapting them to the local conditions in order to improve its performance accuracy or even modify the equation itself (Diodato and Bellocchi, 2007). Hargreaves and Allen (2003) revised the history and application of the HS equation and concluded that recalibrating the exponents and coefficients of the HS equation just increased the complexity of the equation. The HS equation that provides estimates of the incoming solar radiation from ΔT usually considers a default parameter (k_{Rs}) of 0.17 $°C^{-0.5}$ (Almorox et al., 2015). Nevertheless, a common calibrating strategy is not only based on exclusively adjusting the bulk parameter (or adjusted Hargreaves coefficient, AHC or k_{Rs}) in the ET_o HS equation, but also on adjusting the original additive constant 17.8 and the exponent 0.5. The adjustment of k_{Rs} and indirectly AHC might be justified because k_{Rs} adjusts the availability of solar radiation at the surface, and subsequently the available energy for evaporation. However, the adjustment of the exponent and the additive constant make the parameterization of the HS equation difficult, because they interact and influence each other, including AHC (Paredes et al., 2020). Further, the relative accuracy improvement is generally low. Indeed, some studies suggested that it might be preferable to adjust only k_{Rs} or AHC, and not the other constants (e.g. Ravazzani et al., 2012; Berti et al., 2014). The calibration can be carried out using lysimeter benchmarks (e.g. Jensen et al., 1997; López-Urrea et al., 2006) or, more commonly, FAO 56 P-M estimates (e.g. Gavilán et al., 2006; Fooladmand and Haghighat, 2007; Tabari, 2010; Ravazzani et al., 2012; Mendicino and Senatore, 2013; Berti et al., 2014), calibrating in most cases a single AHC. Studies considering FAO 56 P-M ET_o targets for calibrating the HS parameters often forget to assess the implications derived from this simplification. Although the soundness of these studies might be only partially affected by this simplification, conclusions should always be drawn bearing this in mind, which is omitted or forgotten in most cases (Martí et al., 2015a).

The main disadvantage of the calibrated equations is that they are still site-specific and cannot be extrapolated to other sites where preliminary calibration is not possible. Indeed, in stations where a local calibration is possible, the FAO 56 P-M equation would actually be used in practice. Accordingly, several authors have proposed the parametric calibration of the HS parameters relying on additional parameters, such as temperature difference, ratio mean temperature to temperature difference, wind speed, relative humidity, rainfall, and/or altitude (e.g. Jensen et al., 1997; Samani, 2000; Droogers and Allen, 2002; Hargreaves and Allen, 2003; Martínez-Cob and Tejero-Juste, 2004; Vanderlinden et al., 2004; Lee, 2010; Martínez and Thepadia, 2010; Thepadia and Martínez, 2012; Ravazzani et al., 2012; Mendicino and Senatore, 2013; Maestre-Valero et al., 2013; Berti et al., 2014; Martí et al., 2015c; Senatore et al., 2015, 2020; Paredes and Pereira, 2019; Paredes et al., 2020).

In most cases, a single parameterization of the HS factors if carried out per station, i.e. the parameters are fitted using the complete patterns of one station, or even group of stations. In contrast, few studies have addressed a monthly calibration of the HS parameters (e.g. Tabari, 2010; Maestre-Valero et al., 2013). However, these studies focussed on providing monthly HS parameters, rather than assessing the monthly performance of the non-calibrated and calibrated equations to find out if monthly or, at least seasonal patterns are identified, and if these parameterizations might be justified. A single calibrating constant per station might be not enough to properly correct the bias in each period. Further, studies which consider lysimeter benchmarks very rarely assess the effect of using calculated FAO 56 P-M calibrating benchmarks while testing with actual lysimeter benchmarks (Martí et al., 2015a). Thus, this study aims at assessing the seasonal performance of different versions of non-calibrated HS and PM equations. Further, a second objective is to evaluate the effect of considering different timescales for providing the calibrating parameters of the HS equation, namely global, annual, monthly, fortnightly, and weekly calibrating parameters in three different versions of the HS equation and the PM versions relying on the same inputs. In both scenarios, FAO 56 P-M calculated values and lysimeter ET_o benchmarks are considered for both calibrating and validating, and compared. Finally, 6 examples are presented for visualizing the possible effect of the ET_o estimation method and the seasonal trends of the estimates on the corresponding crop water requirements.

2. Methods

2.1. Data set

The data from two lysimeter facilities in Spain were considered, namely 'Las Tiesas' (Albacete) and 'La Orden' (Badajoz), Fig. 1. The period 2007–2015 was considered in Albacete, omitting the year 2010. The period January 2007 to December 2016 was considered in Badajoz, omitting the year 2013. During those periods, daily values of maximum (T_{max}), mean (T_{mean}) and minimum (T_{min}) temperature, average wind speed at 2 m height (u_2), maximum (RH_{max}) and minimum (RH_{min}) relative air humidity, solar radiation (R_s) and the grass reference evapotranspiration (ET_o) were measured. In Albacete station, the climate is characterized by pronounced seasonal variation corresponding to its continental nature, with average temperature in the coldest month (January) of 4.6 °C, and of 24.1 °C in the warmest month (July). The annual average precipitation is 314 mm, corresponding to a semi-arid climate, with lower ranges in the summer comparing with Badajoz. In Badajoz station, the local climate is Mediterranean with mild Atlantic influence, with pronounced seasonal variation of temperatures ranging on average between 9 °C (January) and 26 °C (July), and semiarid with an average annual precipitation of 525 mm/year. It can be considered

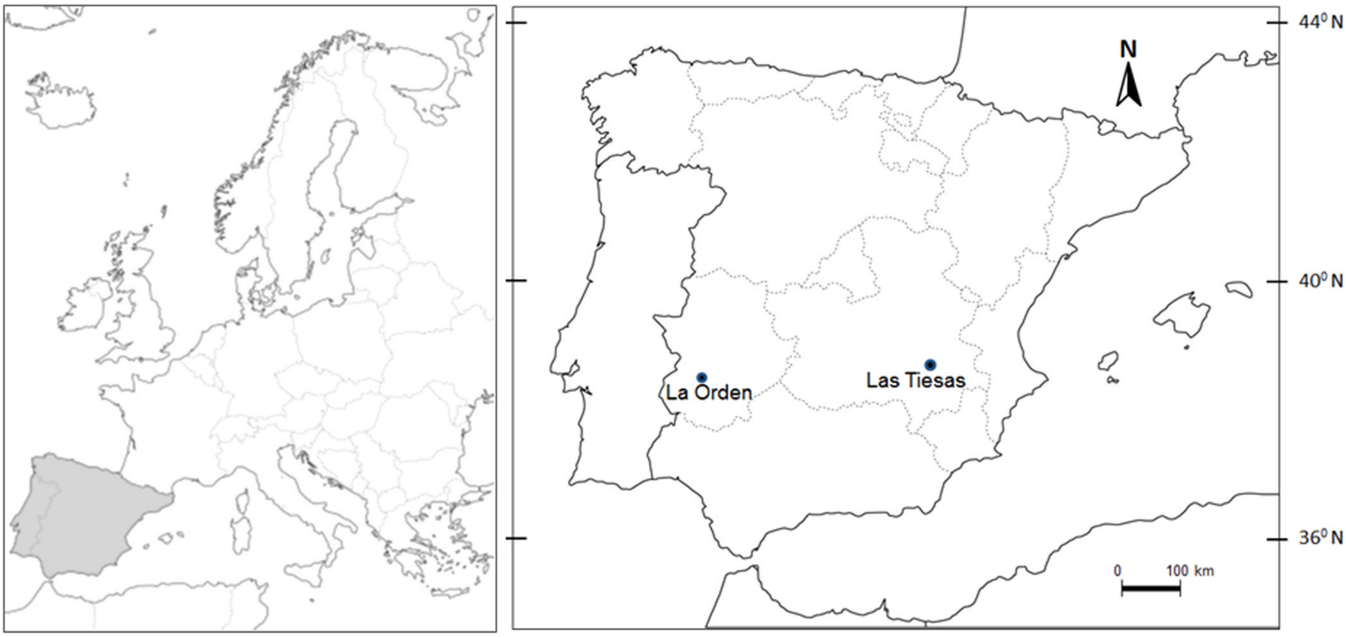

Fig. 1. Geographical location of the considered stations.

that both weather stations are under reference site conditions. Data quality was previously checked in order to detect and exclude outliers. A more detailed climatic characterization of the studied stations can be found in Martí et al. (2015a).

2.1.1. 'Las Tiesas' experimental farm

Grass and alfalfa have been extensively investigated in terms of aerodynamic and surface characteristics, and both crops are widely accepted as a reference surface. To avoid issues with local calibration, which would necessitate time-consuming and costly research, a hypothetical grass reference was selected with well-defined characteristics (Allen et al., 1998). Generally, during the summer months, the ET for a tall crop such as alfalfa is approximately 1.1–1.4 times that of a short crop like grass, due to alfalfa's greater roughness, larger leaf area, lower soil heat flux, and lower surface resistance (Jensen and Allen, 2016). The farm is located near Albacete (SE Spain), in the Castilla-La Mancha region, altitude 695 m above sea level, latitude 39° 3′ North, longitude 2° 5′ West. The surroundings are fully representative of the 110,000 ha irrigated area in the Eastern Mancha region. A large weighing lysimeter was installed in a 1.6 ha plot of grass (*Festuca arundinacea* Schreb.) with uniform height, actively growing, completely shading the ground and well-watered. For this purpose, the grass reference surface was regularly irrigated and mowed to maintain it as near as possible to the reference standard conditions. The soil was maintained close to field capacity and the grass was kept between 0.10 and 0.15 m height, i.e. about 0.12 m.

The lysimeter recipient dimensions are 2.3 m x 2.7 m x 1.7 m (6.21 m^2 surface area) with approximately 14.5 t of total mass. The grass crop was kept in the same condition of growth as the rest of the protection plot in order to be as representative as possible. The lysimeter soil-containing tank sits on a system of beams and counterbalances that offsets the dead weight of the tank with the soil and reduces the load on the weigh beam by 1000:1. This load is communicated to a steel load cell connected to a data logger. All the system was regularly calibrated against known weights. The combined resolution of both load cell and datalogger allowed for the detection of mass changes of about 0.250 kg (0.04 mm water depth). Equipment was programmed to take weight readings every second, and recordings were made every 15 min. Additional information about the technical features of the lysimeter can be found e.g. in López-Urrea et al. (2006) or Martí et al. (2015a). The lysimeter mass change was used to determine ET_o. Irrigation was always carried out at night, and ET_o was determined omitting the hourly values during the irrigation period. Lysimeter readings were checked daily to identify possible errors. Data losses occurred during rainfall events, weight and calibration verifications, and when different works were carried out in the soil of the lysimeter tank. All recordings affected by any kind of incidence were deleted and not used for the analyses. In addition, quality assessment and quality control (QA/QC) procedures for reference ET measurements in the lysimeter were carried out as recommended by Allen et al. (2011a),b).

Meteorological data were recorded by an agro-meteorological station located over the reference surface and close to the lysimeter. The recorded climatic parameters were hourly averages of air temperature at 2 m height, relative air humidity at 2 m, wind velocity at 2 m and net radiation. Net radiation was obtained as net short wave radiation minus net long wave radiation. Net short wave radiation was determined as the difference between incident and reflected shortwave radiation, and was obtained with two pyranometers. Net long wave radiation was determined with a pyrgeometer. All the sensors were connected to a datalogger and read at least every 10 seconds. Hourly and daily values were obtained by averaging these data. Additional information about the technical features of the sensors used can be found in Martí et al. (2015a).

2.1.2. 'La Orden' experimental farm

This farm is located near Badajoz, in the Extremadura Region (South-Western Spain), altitude 198 m above sea level, latitude 38° 51' North, longitude 6° 40' West. It is located in the middle of a 35,000 ha (15 km wide) irrigated area of the low Guadiana river basin called Vegas Bajas del Guadiana. Data were collected from a 1.3 ha plot, uniformly covered with grass (*Festuca arundinacea* Moench) and surrounded by other irrigated crops. The plot was regularly irrigated and clipped during all the measuring period to maintain it as near as possible to the reference standard conditions. Thus, the soil was maintained close to field capacity and with a canopy height between 0.10 and 0.15 m (near to the standard reference crop status).

The large weighing lysimeter has a 6 m^2 squared area (tank dimensions: 2.67 m x 2.25 m x 1.5 m), and it is described in Yrisarry and Naveso (2000). The tank is placed on a balance system with a counterweight system to offset the dead weight. The weighing system was connected to a load cell with a nominal load of 10 kg, and a nominal sensitivity of 2 mV V^{-1}. All the system was regularly calibrated against known weights. The combined resolution of both load cell and datalogger allowed for the detection of mass changes of about 0.20 kg (0.033 mm water depth). Sample time was 0.05 s, and an average weight value was registered in a data logger every 5 min. Thus, hourly ET_o rates were derived from the weight differences recorded in the lysimeter between two consecutive hours. Daily measured ET_o values were obtained by summing up the hourly ones. Irrigation was always carried out at night and ET_o was determined omitting the hourly values during the irrigation period. Only days without incidences (irrigation, precipitation, mowing, any kind of failure, etc.) were used for the analyses. In the same way as in the Albacete lysimeter, a QA/QC of the data was carried out.

Meteorological data were measured in an automatic weather station located over the grass surface and 10 m away from the lysimeter. The data logger recorded hourly averages of air temperature and relative humidity located 1.40 m height, wind speed at 2 m height and net radiation at 1.70 m height. Additional information about the technical features of the sensors used can be found in Martí et al. (2015a).

2.2. Models assessed

Two well-known and broadly used methods for estimating ET_o were assessed. Essentially, one of the major strengths of these methods is that they require few meteorological inputs for their application. First, three versions of the Hargreaves-Samani (Hargreaves and Samani, 1982, 1985; Hargreaves et al., 1985; Hargreaves, 1994; Valiantzas, 2018) equation corresponding to 3 different input combinations were evaluated. Second, corresponding to the previous input combinations of the HS equations, three versions of the Penman-Monteith equation for reduced data sets, i.e. FAO 56 P-M using estimates of the missing data, were evaluated.

2.2.1. Hargreaves-Samani equations

The HS equation for estimating daily reference evapotranspiration is according to Hargreaves and Samani (1985):

$$ET_o^{HS1} = AHC \bullet R_a \bullet (T_{mean} + 17.8) \bullet \sqrt{\Delta T} \tag{1a}$$

where ET_o^{HS1} is the reference evapotranspiration estimation (mm day^{-1}) according to Eq. 1a, R_a is the extraterrestrial radiation (MJ m^{-2} day^{-1}); ΔT is the daily temperature difference (ºC); T_{mean} is the mean daily air temperature (ºC), *AHC* is the adjusted Hargreaves coefficient, equal to 0.0135 k_{Rs}/λ, where 0.0135 is a factor for conversion of units from the American to the International System; k_{Rs} is an empirical radiation adjustment coefficient (º$C^{-0.5}$); λ is the latent heat of vaporization (2.45 MJ kg^{-1}). This equation was developed from:

$$ET_o^{HS2} = 0.0135 \bullet \frac{R_S}{\lambda} \bullet (T_{mean} + 17.8) \tag{1b}$$

where ET_o^{HS2} is the reference evapotranspiration estimation (mm day^{-1}) according to the Eq. 1b, and R_s is the daily shortwave solar radiation (MJ m^{-2} day^{-1}). Eq. 1a is obtained, if instead of being measured, R_s values are computed as follows:

$$R_S = k_{Rs} \bullet R_a \bullet \sqrt{\Delta T} \tag{2}$$

where k_{Rs} is an empirical radiation adjustment coefficient (º$C^{-0.5}$). The historical development of the HS equation can be found in Hargreaves and Allen (2003). Initially, Hargreaves et al. (1985) obtained a value of 0.0022 for *AHC*, after calibrating k_{Rs} using data from four stations in the Senegal river basin in Senegal and Mali, where a value of 0.16 was found. Afterwards, Hargreaves (1994) obtained *AHC*=0.0022 for inland regions, and of 0.0026 for coastal regions. Samani and Pessarakli (1986) obtained k_{Rs} values ranging from 0.119 to 0.212 in the US. A *AHC* value of 0.0023 was accepted for general use (Hargreaves, 1994; Allen et al., 1998). According to Vanderlinden et al. (2004), *AHC* appears to increase in coastal areas, where ΔT decreases due to the sea influence, and decreases in mountainous areas, where air mass movement raises ΔT.

According to Valiantzas (2012), (2018); Exner-Kittridge and Rains (2010), and Exner-Kittridge (2011), if the addition of relative humidity to air temperature data improved the estimation accuracy of R_s, the cost effectiveness of equations relying on temperature and relative humidity could increase dramatically compared to other alternative limited data set methods (i.e. requiring additional wind speed and/or solar radiation methods), because of the low cost of relative humidity sensors. According to Hargreaves and Allen (2003), Hargreaves had suggested in 1977 a formula for estimating R_s from mean relative humidity (RH_{mean}) data alone, namely relying on the term $(1\text{-}RH_{mean}/100)^x$, where x was an empirical coefficient. Thus, combining this formula with the original one (Eq. 2), Valiantzas (2018) proposed an alternative equation for estimating R_s, relying on temperature range and mean relative humidity, namely:

$$R_S = 0.338 \bullet R_a \bullet \Delta T^{0.3} \bullet \left(1.001 - \frac{RH_{mean}}{100}\right)^{0.2} \tag{3}$$

where RH_{mean} is the air mean relative humidity in percent. The coefficients of this equation were calibrated using a global climatic data set, the FAO-CLIMWAT (Smith, 1993). From the full data set, 3588 monthly estimates from 299 stations from 13 countries corresponding to well-watered conditions were used. Valiantzas, 2018 compared Eq. 3 with two other methods, namely with HS (Hargreaves and Samani, 1982), i.e. Eq. 2, and Thornton Running (1999), using daily data from 32 stations from US and Greece, covering a wide range of weather parameters. Eq. 3 performed better than the other two methods for almost all the cases examined. If Eq. 3 is introduced in Eq. 1b, the resulting equation for estimating ET_o would be:

$$ET_o^{HS3} = 0.004563 \bullet R_a \bullet \Delta T^{0.3} \bullet \left(1.001 - \frac{RH_{mean}}{100}\right)^{0.2} \bullet (T_{mean} + 17.8) \tag{4}$$

where ET_o^{HS3} is the reference evapotranspiration estimation according to the Eq. 4 (mm day^{-1}). Thus, HS1 requires measured T_{mean} and ΔT data. HS2 requires measured T_{mean} and R_s data, while HS3 requires measured T_{mean}, ΔT and RH_{mean} data. T_{mean} might be eventually calculated as the mean value of T_{max} and T_{min}.

2.2.2. FAO 56 P-M equation using estimations of missing variables

Allen et al. (1998) proposed a methodology to apply the Penman-Monteith equation when any of the required inputs is/are lacking. It consists of a combination of approaches for estimating: i) T_{dew} from T_{min} or T_{mean}, ii) R_s from ΔT, iii) u_2 using default or regional average values (Paredes et al., 2020). Thus, the values used for these variables were:

i. Relative humidity data or psychrometric observations are missing.

Vapour pressure deficit (VPD) is computed as the difference between the saturation vapour pressure (e_s) and the actual vapour pressure (e_a). e_s is computed as the average of the saturation vapour pressure at T_{max} and T_{min}, while e_a can be calculated in different ways depending on the available data. When RH_{mean} data are available, it can be calculated as follows:

$$e_a = e_s \quad \frac{RH_{mean}}{100} \tag{5}$$

where e_s is calculated from T_{max} and T_{min}. If RH_{mean} data are not available, Allen et al. (1998) recommended to compute e_a assuming that T_{dew} could be replaced by T_{min} as follows:

$$e_a = e^o(T_{\min}) = 0.611 \quad e^{\frac{17.27 \quad T_{\min}}{T_{\min}+237.3}} \tag{6}$$

T_{min} must not be corrected in reference sites, i.e. covered by extensive and actively growing grass crop completely shading the ground and not short of water. In non-reference sites and/or when sites are affected by dryness and local advection, T_{min} should be corrected if it is used as estimation of T_{dew}. A review of the effect of such correction in the estimation of FAO 56 P-M ET_o with limited data sets can be found e.g. in Paredes et al. (2020). In this study, reference sites are considered, and thus T_{min} was not corrected.

ii. Solar radiation data are missing.

According to Allen (1997) and Allen et al. (1998), solar radiation might be estimated from ΔT using Eq. 2, proposed by Hargreaves and Samani (1982), (1985). In particular, in this study a value of 0.17 (°C$^{-0.5}$) was used for the constant k_{Rs}, omitting the distinction between inland (0.16 °C$^{-0.5}$) and coastal (0.19 °C$^{-0.5}$) sites proposed in FAO 56. Other alternatives might be the equation proposed by Bristow and Campbell (1984) or the equation proposed by Thornton and Running (1999).

iii. Wind speed data are missing.

In this case, the world average wind speed value (u_2=2 m s^{-1}) was adopted, in agreement with Allen et al. (1998). Alternatively, local/regional average wind speed values might be used too. According to Allen et al. (1998), the effect of wind speed over ET_o estimates was less important in comparison to other input variables, except for windy and arid areas.

Thus, the FAO 56 P-M equation for reduced datasets was applied relying, respectively, on the same inputs of the previous HS1, HS2, and HS3 equations. Accordingly, PM1 estimates correspond to the FAO 56 P-M equation based on HS1 inputs (i.e. RH_{mean}, R_s, and u_2 measured values were supplanted, e_a was calculated using Eq. 6), PM2 estimates correspond to the FAO 56 P-M equation based on HS2 inputs (RH_{mean} and u_2 measured values were supplanted, e_a was calculated using Eq. 6), and PM3 estimates correspond to the FAO 56 P-M equation based on HS3 inputs (R_s and u_2 measured values were supplanted, e_a was calculated using Eq. 5).

2.3. Calibration approaches

HS estimates should not be overextended to different climatic conditions unless it has been previously locally calibrated (Samani, 2004). In this regard, the performance of the HS equation and its calibrated versions has been widely assessed in different climatic scenarios. Nevertheless, a complete review of such applications is beyond the scope of the present paper. The calibration of the PM equation, when it relies on estimated missing inputs, is less common. However, PM estimates were also calibrated for allowing a fairer comparison with the HS calibrated estimates.

2.3.1. Calibrating and testing benchmarks

Due to the absence of experimental ET_o records, data-driven and conventional empirical models consider in most cases calculated FAO 56 P-M ET_o benchmarks to calibrate and test the models. As the FAO 56 P-M equation is recommended for ET_o estimation and validation of other equations in absence of experimental measurements, studies considering FAO 56 P-M ET_o benchmarks often forget to assess the implications derived from this simplification.

Three scenarios were considered for calibrating and/or validating the performance accuracy of the model estimations. First, calculated FAO 56 P-M values were used as benchmarks for validating the models calibrated with FAO 56 P-M targets. This is the most common procedure in the literature. Second, the process was repeated considering lysimeter ET_o benchmarks, i.e. the models were calibrated and validated using lysimeter ET_o observations. Finally, in a third scenario, the estimations of the models calibrated using calculated FAO 56 P-M benchmarks were validated using the corresponding experimental lysimeter ET_o benchmarks.

The acquisition of lysimeter observations is explained in Section 2.1. The FAO 56 P-M equation (Allen et al., 1998) is directly derived from the original Penman-Monteith equation for a reference crop (clipped grass with 0.12 m height) and assuming standard values of surface resistance, aerodynamic resistance, and albedo, and constant values for air density and for the latent heat of water vaporization (Mendicino and Senatore, 2013). The daily FAO 56 P-M ET_o (mm day^{-1}) is calculated as follows:

$$ET_o^{PM} = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T_{mean}+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{7}$$

where R_n is the net radiation at the crop surface (MJ m^{-2} day^{-1}); G is the soil heat flux density (MJ m^{-2} day^{-1}); T_{mean} is the mean daily air temperature at 2 m height (°C); γ is the psychrometric constant (kPa °C^{-1}); Δ is the slope of vapor pressure curve (kPa °C^{-1}); e_s is the saturation vapor pressure (kPa); e_a is the actual vapor pressure (kPa); and u_2 is the wind speed at 2 m height (m s^{-1}). All the parameters were calculated in the present work by applying the equations provided by Allen et al. (1998).

Only ET_o equations were calibrated, omitting the calibration of some inputs of those equations, e.g. R_s or RH_{mean}, for several reasons. i) If any of the inputs of the ET_o equations is missing, for instance R_s, and it must be estimated using another empirical equation, for instance Eq. 2, any R_s values might be available in practice for calibrating the k_{Rs}; ii) in agreement with i), in this work under 'non-calibrated' HS or PM equations it is assumed that neither the ET_o equations nor the R_s/RH equations (for the inputs) were calibrated; iii) if the inputs of the ET_o equation are previously calibrated, this might affect the resulting values of the calibrated parameters of the ET_o equations. If none of the inputs of the different empirical ET_o equations are calibrated, the comparison between non-calibrated and calibrated estimations of the different ET_o equations is 'fairer', as the calibrations are always exclusively based on ET_o benchmarks.

2.3.2. Calibrating timescales and data management

The 6 approaches mentioned in Section 2.2 were linearly calibrated fitting only the slope term, i.e. for each daily observation the target ET_o value was divided by the non-calibrated ET_o equation, providing a daily calibrating constant per model and station. This procedure was repeated for the 2 types of benchmarks considered in subsection 2.3.1. Subsequently, the complete matrix of data, including the calculated daily calibrating constants, was split per year, month, fortnight, and week. Then, average values of the calibrating constants were calculated for each period, each model and station. The corresponding average calibrating constants were used to provide the calibrated ET_o estimations for each period. The calibrated estimations of the considered equations were provided multiplying the original estimations of the equations by the different mean calibrating constants. This procedure aimed at evaluating if the performance of the calibration might improve by reducing the time window used for calculating the calibrating constant. It is very common in the literature to apply a single average constant per station, which might eventually not allow to properly correct the seasonal bias.

All data were used both for calibrating and validating. The calculation of calibrating constants with generalization ability is beyond the scope of this work. So, it is not required to split data in calibrating and testing data sets. Further, all the considered time windows adopt the same calibration and validation procedure. So, the comparison between time windows can be considered valid. The application of independent test sets for validating purposes might be especially justified for assessing parametric calibrations, where the resulting equations obtained for the calibrating constants should be tested using data series not considered for obtaining those equations (e.g. Paredes and Pereira, 2019; Paredes et al., 2020). Further, previous research (e.g. Shiri et al., 2015; Martí et al., 2015b) stated that, despite a local k-fold validation might be sounder and provide a more reliable assessment of the calibrated estimates, only very small differences might be expected, if enough years are considered in the study. This might be even more marked if the testing period is reduced (e.g. one month, one fortnight etc).

2.4. Performance assessment

Different error parameters were calculated to assess the performance accuracy of the proposed methods (Willmott, 1982). The relative root mean squared error (RRMSE), the mean absolute error (MAE), and the mean bias error (MBE) were obtained according to Eqs. 8–11, respectively, being x_i the actual value of ET_o and the prediction. n was the total number of data in the ET_o matrix. The RRMSE is unitless. The units of MAE and MBE are mm day^{-1}.

$$RRMSE = \frac{1}{\overline{x}} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2} \tag{8}$$

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n}|x_i - \widehat{x}_i| \tag{9}$$

$$MBE = \frac{1}{n} \cdot \sum_{i=1}^{n}(\widehat{x}_i - x_i) \tag{10}$$

Finally, the determination coefficient R^2 was calculated as follows, where and are the standard deviations of observed and predicted ET_o values, respectively.

$$R^2 = \left(\frac{\mathrm{cov}(x_i, \widehat{x}_i)}{\sigma_{x_i} \cdot \sigma_{\widehat{x}_i}}\right)^2 \tag{11}$$

The previous parameters were calculated for the complete data matrix and for the split matrices corresponding to the different time-scales considered. Accordingly, a part from global indicators referred to the complete data set, split values were calculated for each year, month, fortnight, and week in order to assess the seasonal performance of the calibrated and non-calibrated equations. A scheme summarizing the calculation dimensions that were adopted is represented in Fig. 2.

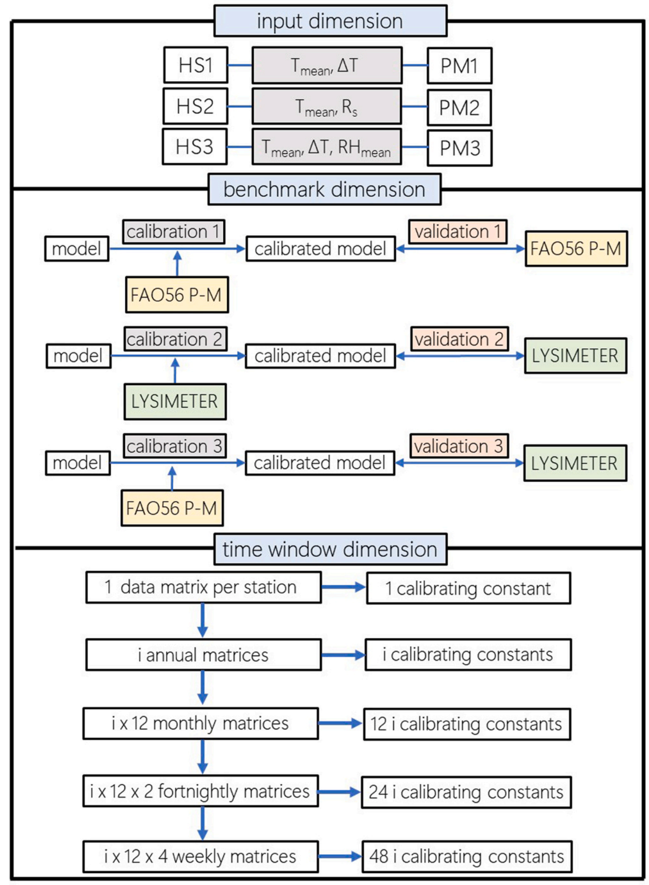

Fig. 2. Calculation dimensions of the applied procedures. (i=8 in Albacete; i=9 in Badajoz).

3. Results and discussion

3.1. Global performance indicators of HS models

Tables 1 and 2 present the global statistical indicators of the HS and PM models, respectively. Each table presents the error parameters for each considered calibrating timescale of the models and each type of benchmark used for calibrating/testing. Accordingly, regarding Table 1 and HS estimates, three main patterns can be stated. First, HS2 (based on T_{mean}, R_s) tends to provide more accurate estimates than HS3 (based on T_{mean}, RH_{mean}), and HS3 tends to provide more accurate estimates than HS1 (based on ΔT) for both calculated FAO 56 P-M and lysimeter benchmarks, respectively. In Albacete, according to the results based on calculated benchmarks, HS2 estimates provided RRMSE values in the range 0.1463–0.1169 (from maximum non-calibrated to minimum weekly average calibrating constants), while HS3 estimates provided RRMSE values in the range 0.1723–0.1289, and HS1 estimates provided RRMSE values in the range 0.1790–0.1481. Thus, attending to non-calibrated equations, HS2 presented a RRMSE 0.026 lower than HS3 and 0.0327 lower than HS1. The RRMSE differences between HS versions decreased comparing the weekly calibrated versions to 0.012 between HS2 and HS3, while they presented a similar range to the differences between the non-calibrated models (0.0312) between HS2 and HS1. Similarly, the MAE range (mm day^{-1}) between the non-calibrated equation and the weekly calibrated version ranged between 0.4513 and 0.3504 (HS2), 0.5468–0.3944 (HS3) and 0.5718–0.4554 (HS1). In Badajoz, similar patterns can be stated based on calculated benchmarks. However, the RRMSE values are in average between 0.0236 (i.e. 2.36 % for non-calibrated estimations) and 0.032 (i.e. 3.20 % for weekly estimations) lower than in Albacete. Thus, HS2 estimates provided RRMSE values in the range 0.1239–0.0771 (from maximum non-calibrated to minimum weekly average calibrating constants), while HS3 estimates provided RRMSE values in the range 0.1471–0.1030, and HS1 estimates provided RRMSE values in the range 0.1558–0.1191. So, similar accuracy differences than in Albacete took place within the non-calibrated HS versions, i.e. HS2 presented a RRMSE 0.023 lower than HS3 and 0.032 lower than HS1. These RRMSE differences were similar in weekly calibrated versions (HS2 was 0.0259 lower than HS3, HS2 was 0.042 lower than HS1). Similar conclusions can be drawn from the analysis of the R^2 values.

Second, the calibrations considering global and annual mean AHCs provided in general very slight accuracy improvements. On the other hand, the calibrations considering monthly to weekly mean constants provided more relevant accuracy improvements. As could be expected, the improvement was more marked when the time window considered for averaging was shorter, i.e. HS estimates based on monthly constants provided less improvement than HS estimates based on fortnightly constants, while HS estimates based on fortnightly constants provided less improvement than HS estimates based on weekly constants. Thus, in Albacete the global and annual calibrations provided RRMSE values quite similar to the non-calibrated HS equations, especially for HS2 and HS1 (with RRMSE of 0.1463 and 0.1441 for the global and annual calibrations of HS2, respectively, vs. 0.1463 for the non-calibrated version

Table 1
Global performance indicators of the HS models for the period 2007–2015 in Albacete and 2007–2016 in Badajoz. (LYS: lysimeter, PM: FAO 56 P-M Penman-Monteith).

station	benchmarks		Calibrating timescale	HS1				HS2				HS3			
	cal	test		MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2	MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2	MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2
A	PM	PM	without calibration	0.5718	0.1790	-0.0071	0.8867	0.4513	0.1463	0.0334	0.9245	0.5468	0.1723	0.2701	0.9131
			global	0.5712	0.1793	-0.0328	0.8867	0.4512	0.1463	0.0332	0.9245	0.4969	0.1571	-0.0324	0.9131
			annual	0.5677	0.1790	-0.0294	0.8868	0.4419	0.1441	0.0268	0.9267	0.4936	0.1562	-0.0300	0.9140
			month	0.5249	0.1673	-0.0100	0.9010	0.4028	0.1305	0.0353	0.9403	0.4592	0.1465	-0.0288	0.9243
			fortnight	0.4923	0.1580	-0.0020	0.9118	0.3800	0.1246	0.0362	0.9458	0.4267	0.1375	-0.0199	0.9331
			week	0.4554	0.1481	-0.0001	0.9225	0.3504	0.1169	0.0313	0.9523	0.3944	0.1289	-0.0147	0.9412
	LYS	LYS	without calibration	0.7114	0.2215	-0.1254	0.8410	0.5775	0.1871	-0.0848	0.8855	0.6640	0.2017	0.1519	0.8691
			global	0.7100	0.2191	-0.0487	0.8410	0.5815	0.1855	0.0168	0.8855	0.6462	0.1990	-0.0443	0.8691
			annual	0.6924	0.2153	-0.0448	0.8463	0.5615	0.1789	0.0163	0.8935	0.6287	0.1949	-0.0412	0.8742
			month	0.6380	0.1986	0.0050	0.8690	0.5130	0.1614	0.0497	0.9141	0.5768	0.1793	-0.0138	0.8929
			fortnight	0.5982	0.1878	0.0121	0.8831	0.4856	0.1536	0.0487	0.9224	0.5390	0.1690	-0.0061	0.9048
			week	0.5565	0.1759	0.0082	0.8975	0.4497	0.1447	0.0387	0.9310	0.4980	0.1581	-0.0067	0.9167
	PM	LYS	without calibration	0.7114	0.2215	-0.1254	0.8410	0.5775	0.1871	-0.0848	0.8855	0.6640	0.2017	0.1519	0.8691
			global	0.7133	0.2227	-0.1510	0.8410	0.5775	0.1871	-0.0851	0.8855	0.6530	0.2031	-0.1506	0.8691
			annual	0.7180	0.2234	-0.1476	0.8393	0.5866	0.1876	-0.0914	0.8850	0.6588	0.2040	-0.1482	0.8674
			month	0.6957	0.2150	-0.1282	0.8493	0.5719	0.1803	-0.0829	0.8931	0.6471	0.1990	-0.1471	0.8733
			fortnight	0.6749	0.2087	-0.1202	0.8577	0.5645	0.1788	-0.0820	0.8948	0.6277	0.1937	-0.1381	0.8795
			week	0.6582	0.2032	-0.1183	0.8651	0.5613	0.1772	-0.0869	0.8968	0.6174	0.1899	-0.1330	0.8838
B	PM	PM	Without calibration	0.6064	0.1558	-0.1863	0.8531	0.4790	0.1239	-0.0248	0.9038	0.5659	0.1471	0.2980	0.8892
			global	0.5884	0.1516	-0.0120	0.8531	0.4897	0.1248	0.0362	0.9038	0.5082	0.1313	-0.0195	0.8892
			annual	0.5482	0.1404	-0.0158	0.8738	0.3607	0.0952	-0.0036	0.9417	0.4759	0.1225	-0.0210	0.9037
			month	0.5178	0.1344	0.0348	0.8885	0.3318	0.0868	0.0286	0.9523	0.4459	0.1168	0.0133	0.9136
			fortnight	0.4948	0.1290	0.0330	0.8968	0.3202	0.0839	0.0267	0.9554	0.4241	0.1120	0.0124	0.9203
			week	0.4463	0.1191	0.0265	0.9115	0.2896	0.0778	0.0232	0.9616	0.3856	0.1030	0.0095	0.9323
	LYS	LYS	without calibration	0.8572	0.2562	0.2333	0.7111	1.0137	0.2785	0.3948	0.6878	1.0215	0.2943	0.7175	0.7255
			global	0.8327	0.2526	-0.0757	0.7111	0.8800	0.2611	0.0157	0.6878	0.8028	0.2467	-0.0737	0.7255
			annual	0.8214	0.2487	-0.0734	0.7193	0.7122	0.2226	-0.0661	0.7779	0.7916	0.2430	-0.0732	0.7328
			month	0.7212	0.2246	0.0268	0.7721	0.5977	0.1985	0.0174	0.8196	0.6902	0.2180	0.0109	0.7832
			fortnight	0.6861	0.2141	0.0276	0.7926	0.5660	0.1879	0.0184	0.8384	0.6539	0.2073	0.0122	0.8038
			week	0.6454	0.2025	0.0211	0.8135	0.5333	0.1790	0.0143	0.8532	0.6154	0.1965	0.0087	0.8234
	PM	LYS	without calibration	0.8572	0.2562	0.2333	0.7111	1.0137	0.2785	0.3948	0.6878	1.0215	0.2943	0.7175	0.7255
			global	0.9073	0.2672	0.4075	0.7111	1.0425	0.2839	0.4557	0.6878	0.8728	0.2605	0.4000	0.7255
			annual	0.9585	0.2757	0.4038	0.6906	0.8702	0.2497	0.4160	0.7536	0.9176	0.2679	0.3986	0.7080
			month	0.9516	0.2751	0.4544	0.7074	0.8712	0.2520	0.4482	0.7567	0.9126	0.2670	0.4329	0.7195
			fortnight	0.9437	0.2740	0.4526	0.7092	0.8688	0.2523	0.4463	0.7554	0.9070	0.2664	0.4319	0.7206
			week	0.9287	0.2698	0.4461	0.7172	0.8634	0.2516	0.4427	0.7566	0.8990	0.2633	0.4291	0.7271

of HS2, and 0.1793 and 0.1790 vs. 0.1790 for HS1, respectively). In HS3, there was a more marked improvement (RRMSE of 0.1571 and 0.1562 vs. 0.1723). In contrast to this, the monthly, fortnightly and weekly calibrated estimates provided RRMSE decreases in comparison to the non-calibrated estimates around 0.01, 0.02 and 0.03 (i.e. 1 %, 2 %, and 3 %), in HS2 and HS1, while the RRMSE decrease was slightly more marked in HS3 (0.026, 0.035 and 0.043, respectively). In Badajoz, similar patterns can be stated. The global and annual calibrations provided RRMSE values quite similar to the non-calibrated HS equations (with RRMSE of 0.1248 and 0.0952 vs. 0.1239 in HS2, 0.1516 and 0.1404 vs. 0.1558 in HS1, and 0.1313 and 0.1225 vs. 0.1471 in HS3). So, in HS2 and HS1, there was a more marked improvement in the annual calibration in comparison to Albacete, while in HS3 the global and annual calibration provided again more marked improvements, similarly to Albacete. On the other hand, the monthly, fortnightly and weekly calibrated estimates provided, respectively, RRMSE decreases in comparison to the non-calibrated estimates of 0.021, 0.027 and 0.037, in HS1, 0.037, 0.040 and 0.047, in HS2, and 0.0303, 0.0351, and 0.0441, in HS3. The analysis of the R^2 values is consistent with such trends. So, the accuracy improvements were slightly better in Badajoz than in Albacete. Consequently, the calibration accuracy of the HS equation considering a single constant per station or year will depend on the presence or not of a clear under- or over estimation trend of the non-calibrated estimates. If this is the case, a single average constant might correct all points in the right direction and thus allow to improve the accuracy of the calibrated estimates. If not, some points might be properly corrected, while others not, resulting globally in a very scarce accuracy improvement. The same pattern applies to shorter calibration time windows. However, reducing the calibrating time window increases the probability that a same bias in the non-calibrated estimates takes place, and consequently allows to correct them in the correct direction. Thus, the application of monthly or, at least, seasonal calibrating constants would be desirable to properly adjust the bias of the original estimates.

Third, lysimeter targets provide similar qualitative conclusions than calculated targets regarding HS rankings and accuracy improvement derived from calibrating constants with decreasing time windows. However, the error range considerably increased, especially in Badajoz. This increasing seems reasonable and is related to the consideration of experimental benchmarks. In the former case, HS tried to approximate an already existing function, namely the FAO 56 P-M equation. A detailed analysis of the obtained errors is provided as follows. In Albacete, based on lysimeter benchmarks, HS2 estimates provided RRMSE values in the range 0.1871–0.1447 (from maximum non-calibrated to

Table 2
Global performance indicators of the PM models for the period 2007–2015 in Albacete and 2007–2016 in Badajoz. (LYS: lysimeter, PM: FAO 56 P-M Penman-Monteith).

station	benchmarks		Calibrating timescale	PM1				PM2				PM3			
	cal	test		MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2	MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2	MAE (mm day^{-1})	RRMSE (-)	MBE (mm day^{-1})	R^2
A	PM	PM	without calibration	0.6024	0.1852	0.0615	0.8802	0.4828	0.1532	-0.1029	0.9245	0.5700	0.1759	-0.3291	0.9133
			global	0.6015	0.1881	-0.0874	0.8802	0.4803	0.1519	-0.0838	0.9245	0.5361	0.1742	0.1418	0.9133
			annual	0.5969	0.1874	-0.0838	0.8804	0.4793	0.1521	-0.0812	0.9237	0.5300	0.1724	0.1406	0.9150
			month	0.5387	0.1717	-0.0017	0.8960	0.4194	0.1343	-0.0123	0.9362	0.4604	0.1472	-0.0071	0.9237
			fortnight	0.5068	0.1623	0.0061	0.9073	0.3913	0.1266	-0.0041	0.9433	0.4303	0.1392	-0.0051	0.9317
			week	0.4694	0.1524	0.0067	0.9184	0.3615	0.1186	-0.0027	0.9503	0.4003	0.1319	-0.0035	0.9388
	LYS	LYS	without calibration	0.7373	0.2264	-0.0567	0.8330	0.6404	0.2042	-0.2211	0.8804	0.7207	0.2216	-0.4473	0.8744
			global	0.7382	0.2280	-0.1012	0.8330	0.6236	0.1953	-0.0993	0.8804	0.6682	0.2066	0.1365	0.8744
			annual	0.7214	0.2240	-0.0973	0.8381	0.6105	0.1924	-0.0939	0.8830	0.6530	0.2033	0.1356	0.8785
			month	0.6527	0.2034	0.0143	0.8632	0.5399	0.1687	0.0020	0.9053	0.5726	0.1774	0.0080	0.8959
			fortnight	0.6120	0.1922	0.0208	0.8781	0.5014	0.1588	0.0088	0.9162	0.5348	0.1677	0.0079	0.9069
			week	0.5693	0.1801	0.0155	0.8929	0.4674	0.1489	0.0048	0.9263	0.4932	0.1576	0.0034	0.9177
	PM	LYS	without calibration	0.7373	0.2264	-0.0567	0.8330	0.6404	0.2042	-0.2211	0.8804	0.7207	0.2216	-0.4473	0.8744
			global	0.7480	0.2340	-0.2056	0.8330	0.6368	0.2025	-0.2020	0.8804	0.6490	0.2000	0.0236	0.8744
			annual	0.7496	0.2339	-0.2020	0.8320	0.6464	0.2040	-0.1994	0.8772	0.6537	0.2028	0.0224	0.8710
			month	0.7064	0.2187	-0.1199	0.8434	0.6064	0.1889	-0.1305	0.8850	0.6334	0.1942	-0.1254	0.8775
			fortnight	0.6848	0.2120	-0.1121	0.8527	0.5898	0.1843	-0.1224	0.8901	0.6198	0.1908	-0.1233	0.8817
			week	0.6677	0.2062	-0.1116	0.8606	0.5820	0.1812	-0.1209	0.8937	0.6121	0.1880	-0.1217	0.8851
B	PM	PM	Without calibration	0.6636	0.1681	-0.1378	0.8239	0.4696	0.1178	0.1779	0.9203	0.5766	0.1488	-0.2796	0.8782
			global	0.6547	0.1657	-0.0454	0.8239	0.4369	0.1114	-0.0226	0.9203	0.5269	0.1406	0.0321	0.8782
			annual	0.6129	0.1551	-0.0489	0.8460	0.4354	0.1113	-0.0214	0.9204	0.5009	0.1310	0.0290	0.8932
			month	0.5559	0.1435	0.0482	0.8753	0.3898	0.0995	0.0218	0.9374	0.4646	0.1239	0.0259	0.9052
			fortnight	0.5287	0.1375	0.0465	0.8851	0.3706	0.0952	0.0204	0.9426	0.4385	0.1178	0.0225	0.9137
			week	0.4756	0.1266	0.0373	0.9014	0.3290	0.0866	0.0154	0.9523	0.3941	0.1074	0.0157	0.9275
	LYS	LYS	without calibration	0.9138	0.2664	0.2818	0.6928	0.9965	0.2757	0.5975	0.7352	0.8274	0.2541	0.1399	0.7092
			global	0.8837	0.2627	-0.1029	0.6928	0.7929	0.2425	-0.0743	0.7352	0.8164	0.2518	-0.0094	0.7092
			annual	0.8730	0.2588	-0.0998	0.7008	0.7360	0.2271	-0.0919	0.7723	0.8107	0.2531	-0.0130	0.7069
			month	0.7422	0.2279	0.0374	0.7667	0.6278	0.2012	0.0099	0.8148	0.7099	0.2260	0.0251	0.7696
			fortnight	0.7047	0.2176	0.0386	0.7873	0.5915	0.1910	0.0117	0.8331	0.6680	0.2143	0.0243	0.7923
			week	0.6619	0.2059	0.0299	0.8083	0.5563	0.1813	0.0072	0.8494	0.6233	0.2018	0.0168	0.8147
	PM	LYS	without calibration	0.9138	0.2664	0.2818	0.6928	0.9965	0.2757	0.5975	0.7352	0.8274	0.2541	0.1399	0.7092
			global	0.9395	0.2718	0.3742	0.6928	0.8977	0.2561	0.3970	0.7352	0.9104	0.2741	0.4517	0.7092
			annual	0.9874	0.2804	0.3707	0.6705	0.8973	0.2563	0.3981	0.7350	0.9320	0.2746	0.4486	0.7065
			month	0.9737	0.2799	0.4678	0.7008	0.8803	0.2548	0.4414	0.7492	0.9324	0.2743	0.4455	0.7082
			fortnight	0.9625	0.2784	0.4661	0.7037	0.8777	0.2550	0.4400	0.7484	0.9216	0.2723	0.4420	0.7116
			week	0.9439	0.2736	0.4568	0.7121	0.8701	0.2530	0.4349	0.7520	0.9073	0.2677	0.4353	0.7200

minimum weekly average calibrating constants), while HS3 estimates provided RRMSE values in the range 0.2017–0.1581, and HS1 estimates provided RRMSE values in the range 0.2215–0.1759. Thus, attending to the non-calibrated equations, HS2 presented a RRMSE 0.0146 lower than HS3 and 0.0344 lower than HS1. The RRMSE differences between HS versions decreased comparing the weekly calibrated versions to 0.0134 between HS2 and HS3, while they presented a similar range to the differences between the non-calibrated models (0.0311) between HS2 and HS1. Similarly, the MAE range (mm day^{-1}) between the non-calibrated equation and the weekly calibrated version ranged between 0.5775 and 0.4497 (HS2), 0.6640–0.4980 (HS3) and 0.7114–0.5565 (HS1). In Badajoz, in contrast to the case of calculated benchmarks, the RRMSE ranges are considerable higher than in Albacete. HS2 estimates provided RRMSE values in the range 0.2785–0.1790 (from maximum non-calibrated to minimum weekly average calibrating constants), while HS3 estimates provided RRMSE values in the range 0.2943–0.1965, and HS1 estimates provided RRMSE values in the range 0.2562–0.2025. Thus, in non-calibrated estimations, HS1 performed more accurately than HS2 and HS3. However, the calibrated estimations of HS2 and HS3 were again more accurate than those of HS1. Comparing RRMSE differences between HS versions, non-calibrated the HS1 equation presented a RRMSE 0.0223 lower than HS2, and a RRMSE 0.0381 lower than HS3. However, comparing RRMSE differences between weekly calibrations, HS2 equation presented a RRMSE 0.0175 lower than HS3 and a RRMSE 0.0235 lower than HS1. Regarding the comparison between calibrating time windows, in Albacete, the global and annual calibrations provided, respectively, RRMSE values of 0.1855 and 0.1789 vs. a RRMSE of 0.1871 for the non-calibrated version in HS2, 0.1990 and 0.1949 vs. 0.2017 in HS3, and 0.2191 and 0.2153 vs. 0.2215 in HS1. In contrast to this, the monthly, fortnightly and weekly calibrated estimates provided RRMSE decreases in comparison to the non-calibrated estimates around 0.02, 0.03 and 0.04 (i.e. 2 %, 3 %, and 4 %), in all HS versions. In Badajoz, the global and annual calibrations provided in comparison to the non-calibrated estimation, respectively, RRMSE values of 0.2611 and 0.2226 vs. 0.2785 for HS2, 0.2467 and 0.2430 vs. 0.2943 for HS3, and 0.2526 and 0.2487 vs. 0.2562 for HS1. So, in HS2 and HS3, there was a more marked improvement in the global and, especially, in the annual calibration in comparison to Albacete. On the other hand, the monthly, fortnightly and weekly calibrated estimates provided, respectively, RRMSE decreases in comparison to the non-calibrated estimates of 0.0316, 0.0421 and 0.0537, in HS1, 0.0800, 0.0906 and 0.0995, in HS2, and 0.0763, 0.0870, and 0.0978, in HS3. So, the accuracy improvements were considerably better in Badajoz than in Albacete. The analysis of the R^2 values is consistent with such trends.

Thus, again, reducing the calibrating time window increases the probability to be able to correct the bias of the non-calibrated estimations in the correct direction.

Regarding the bias of the non-calibrated estimations there is not a clear general pattern of over- or underestimation. If the calibration and validation was based on calculated benchmarks, in Albacete, HS1 presented a slight negative MBE (-0.0071 mm day^{-1}, i.e. underestimation), while HS2 and HS3 presented a positive MBE (0.0334 and 0.2701 mm day^{-1}, respectively, i.e. overestimation). In Badajoz, HS1 and HS2 presented negative MBE (-0.1863 and −0.0248 mm day^{-1}, respectively), while HS3 presented a positive MBE (0.2980 mm day^{-1}). If lysimeter benchmarks were used for calibrating and testing, in Albacete HS1 and HS2 presented a negative MBE (-0.1254 and – 0.0848 mm day^{-1}, respectively), while HS3 presented a positive MBE (0.1519 mm day^{-1}). In Badajoz, the three versions of HS presented positive MBE values (0.2333, 0.3948 and 0.7175 mm day^{-1} for HS1, HS2 and HS3, respectively). Comparing the MAE values between Albacete and Badajoz for both type of benchmarks, it can be stated that their ranges are similar when calculated benchmarks are used, while there is an important increase in Badajoz for non-calibrated estimates when lysimeter benchmarks are used. MAE ranges (in mm day^{-1}) from maximum non-calibrated to minimum weekly calibrations in Albacete and Badajoz relying on calculated benchmarks are, respectively, 0.5718–0.4554 vs. 0.6064–0.4463 (HS1), 0.4513–0.3504 vs. 0.4790–0.2896 (HS2), and 0.5468–0.3944 vs. 0.5659–0.3856 (HS3). Similarly, the MAE ranges (in mm day^{-1}) relying on lysimeter benchmarks are 0.7114–0.5565 vs. 0.8572–0.64545 (HS1), 0.5775–0.4497 vs. 1.0137–0.5333 (HS2), and 0.6640–0.4980 vs. 1.0215–0.6154 (HS3). Finally, attending to models that were calibrated using calculated benchmarks, but tested using lysimeter ones, the performance patterns are similar to the scenario where lysimeter benchmarks are used for calibrating and testing. In this case, the estimations provide higher errors, as could be expected. The error increase in comparison to models that were calibrated using lysimeter benchmarks is lower in Albacete than in Badajoz, while increases for lower time windows. Thus, in Albacete, the RRMSE values for global and weekly calibrated estimations using lysimeter vs. calculated calibrating benchmarks is 0.2191–0.1759 vs. 0.2227–0.2032 (HS1), 0.1855–0.1447 vs. 0.1871–0.1772 (HS2), and 0.1990–0.1581 vs. 0.2031–0.1899 (HS3). Similarly, in Badajoz the corresponding RRMSE ranges are 0.2526–0.2025 vs. 0.2672–0.2698 (HS1), 0.2611–0.1790 vs. 0.2839–0.2516 (HS2), and 0.2467–0.1965 vs. 0.2605–0.2633 (HS3).

3.2. Global performance indicators of PM models

Attending to Table 2, a comparison between PM1, PM2 and PM3 models provides similar qualitative patterns than those found between HS1, HS2, and HS3 models for all types of benchmarks and calibration timescales. The error parameters of the PM models are just very slightly higher than the error indexes corresponding to the HS estimations. Accordingly, PM2 (based on T_m, R_s) tends to provide more accurate estimates than PM3 (based on T_m, RH_{mean}), and PM3 tends to provide more accurate estimates than PM1 (based on ΔT) for both calculated FAO 56 P-M and lysimeter benchmarks, respectively. In Albacete, according to the results based on calculated benchmarks, PM2 estimates provided RRMSE values in the range 0.1532–0.1186 (from maximum non-calibrated to minimum weekly calibrated), while PM3 estimates provided RRMSE values in the range 0.1759–0.1319, and PM1 estimates provided RRMSE values in the range 0.1852–0.1524. Similarly, in Badajoz, PM2 provided RRMSE in the range 0.1178–0.0866, PM3 provided RRMSE in the range 0.1488–0.1074, and PM1 provided RRMSE in the range 0.1681–0.1266. So, again, Badajoz presented lower error values than Albacete when using calculated benchmarks, similarly to the HS models. Similarly, the MAE range (mm day^{-1}) between the non-calibrated equation and the weekly calibrated version ranged between 0.4828 and 0.3615 (PM2), 0.5700–0.4003 (PM3) and 0.6024–0.4694 (PM1) in Albacete, and between 0.4696 and 0.3290 (PM2), 0.5766–0.3941 (PM3) and 0.6636–0.4756 (PM1) in Badajoz. So, similarly to the HS models, despite the MAE values were higher in Badajoz, their RRMSE values were smaller. This might be due to a higher order of magnitude of ET values in Badajoz.

On the other hand, the calibrations considering global and annual mean calibrating constants provided in general very slight accuracy improvements. In Albacete, the RRMSE values for global and annual calibrations in comparison to non-calibrated ones were 0.1519 and 0.1521 vs. 0.1532 (PM2), 0.1742 and 0.1724 vs. 0.1759 (PM3), and 0.1881 and 0.1874 vs. 0.1852 (PM1), while in Badajoz the corresponding RRMSE ranges were 0.1114 and 0.1113 vs. 0.1178 (PM2), 0.1406 and 0.1380 vs. 0.1488 (PM3), and 0.1657 and 0.1551 vs. 0.1681 (PM1). On the other hand, the calibrations considering monthly to weekly mean calibrating constants provided more relevant accuracy improvements. Again, the improvement was more marked when the time window considered for averaging was shorter. The RRMSE reduction of monthly to weekly calibrations in comparison to non-calibrated estimations ranged between 0.0189 and 0.0346 (PM2), 0.0287–0.044 (PM3), and 0.0135–0.0328 (PM1) in Albacete, while it ranged between 0.0183 and 0.0312 (PM2), 0.0249–0.0414 (PM3), and 0.0246–0.0415 (PM1) in Badajoz. Thus, despite PM models relying on lacking variables are usually not calibrated, their calibration based on reducing the calibrating time window increases the probability that a same bias in the non-calibrated estimates takes place, and consequently allows to correct them in the correct direction. So, the application of monthly or, at least, seasonal calibrating constants would be desirable to properly adjust the bias of the original estimates.

Regarding models that were calibrated and tested using lysimeter benchmarks, similar conclusions can be drawn on behalf of rankings and accuracy improvement derived from calibrating the models with decreasing time windows in comparison to HS models and PM models calibrated and tested using calculated targets. In comparison to models calibrated and tested with calculated benchmarks, the RRMSE values of the non-calibrated PM estimations were around 4–5 % higher in Albacete, and around 10 % higher in Badajoz (even 15 % higher for PM2). The weekly calibrations presented RRMSE values around 3 % higher in Albacete, and around 8–10 % higher in Badajoz. In comparison to HS models calibrated and tested with lysimeter benchmarks, the PM models calibrated and tested with lysimeter benchmarks present RRMSE values between 0.5 % and 2 % higher in Albacete when they were not calibrated, and between 1 % and 4 % in Badajoz (with the exception of PM2, which present similar RRMSE than HS2). The weekly calibrated equations presented very similar RRMSE in both cases. As mentioned above, the increase might be related to the consideration of experimental benchmarks. There is also not a clear general pattern of over- or underestimation attending to the bias of the non-calibrated estimations. Using calculated benchmarks, in Albacete, PM1 presented a slight positive MBE (0.0615 mm day^{-1}), while PM2 and PM3 presented a negative MBE (-0.1029 and −0.3291 mm day^{-1}, respectively). In Badajoz, PM1 and PM3 presented negative MBE (-0.1378 and −0.2796 mm day^{-1}, respectively), while PM3 presented a positive MBE (0.1779 mm day^{-1}). Using lysimeter benchmarks for calibrating and testing, in Albacete the three versions of PM presented negative MBE values (-0.0567, −0.2211 and −0.4473 mm day^{-1} for PM1, PM2 and PM3, respectively). In Badajoz, the three versions of PM presented positive MBE values (0.2818, 0.5975 and 0.1399 mm day^{-1} for PM1, PM2 and PM3, respectively). MAE ranges (in mm day^{-1}) from maximum non-calibrated to minimum weekly calibrations in Albacete and Badajoz relying on calculated benchmarks are, respectively, 0.6024–0.4694 vs. 0.6636–0.4756 (PM1), 0.4828–0.3615 vs. 0.4696–0.3290 (PM2), and 0.5700–0.4003 vs. 0.5766–0.3941 (PM3). Similarly, the MAE ranges (in mm day^{-1}) relying on lysimeter benchmarks are 0.7373–0.5693 vs. 0.9138–0.6619 (PM1), 0.6404–0.4674 vs. 0.9965–0.5363 (PM2), and 0.7207–0.4932 vs. 0.8774–0.6233 (PM3). Thus, it can be stated that the MAE ranges are similar in Albacete and Badajoz when calculated benchmarks are used, while there is a slight increase in Badajoz when lysimeter benchmarks are used. In models that were calibrated using calculated benchmarks, but tested using lysimeter ones, the performance patterns are close to the scenario where lysimeter benchmarks are used for calibrating and testing, providing higher errors. The error increase in comparison to models that were calibrated using lysimeter benchmarks is lower in Albacete than in Badajoz, while increases for lower time windows, similarly to the HS models.

3.3. Monthly performance of the non-calibrated HS and PM models

Fig. 3 and Fig. 4 present the monthly average RRMSE values of the non-calibrated models for both type of benchmarks in Albacete and Badajoz, respectively. In Albacete, Fig. 3, attending to the models assessed with calculated benchmarks, the RRMSE values of the models per month fluctuated between minimum ranges of 0.0969 (HS2) −0.1313 (PM2) in July (i.e. a difference of 0.0344), or 0.0992 (PM2) −0.1576 (HS3) in June (i.e. a difference of 0.0584), and maximum ranges of 0.2612 (HS2) - 0.4843 (PM1) in December (i.e. a difference of 0.2231), or 0.2789 (HS2) - 0.4026 (PM3) in January (i.e. a difference of 0.1237). Attending to models assessed with lysimeter benchmarks, RRMSE values of the models per month fluctuated between minimum ranges of 0.1085 (HS1) - 0.1575 (PM2) in June (i.e. a difference of 0.0490), or 0.0985 (HS1) - 0.1737 (PM2) in July (i.e. a difference of 0.0752), and maximum ranges of 0.2187 (PM1) - 0.6410 (PM2) in December (i.e. a difference of 0.4223), or 0.2946 (PM1) - 0.5415 (PM2) in January (i.e. a difference of 0.2469). In Badajoz, Fig. 4, there were no available climatic records in January, except for the year 2016. Therefore, this month was omitted from the comparison, because the rest of month relied on estimations covering 9 years. Thus, attending to the models assessed with calculated benchmarks, the RRMSE values of the models per month fluctuated between minimum ranges of 0.0834 (PM2) - 0.1456 (PM1) in July (i.e. a difference of 0.0622), or 0.0930 (PM2) - 0.1343 (HS3) in June (i.e. a difference of 0.0413), and maximum ranges of 0.1951 (PM3) - 0.3825 (PM1) in November (i.e. a difference of 0.1874), or 0.1755 (HS3) - 0.2428 (PM1) in January (i.e. a difference of 0.673). The values of December are abnormally high, and were disregarded. As mentioned above, in this scenario, the errors in Badajoz were lower than in Albacete. Attending to models assessed with lysimeter benchmarks, RRMSE values of the models per month fluctuated between minimum ranges of 0.1434 (PM1) - 0.1958 (HS3) in July (i.e. a difference of 0.0524), or 0.1515 (PM1) - 0.2296 (HS3) in June (i.e. a difference of 0.0781), and maximum ranges of 0.2740 (PM1) - 0.6681 (PM2) in November (i.e. a difference of 0.3941), or 0.3187 (PM1) - 0.4631 (HS3) in January (i.e. a difference of 0.1444). The values of December are again slightly high, and were disregarded. In this scenario, the models provided, in general, lower error ranges in Albacete.

Thus, attending to these figures it can be stated that there is a significant fluctuation of the model performance accuracy during the year, showing considerably lower errors and lower differences within models during the summer, while presenting higher errors and higher differences within models during the winter. Thus, the models present a higher mapping ability during the summer, where the considered inputs might have a higher effect on the ET patterns. Further, the higher order of magnitude of ET_o during the summer might contribute to this pattern. Or, conversely, the low ET_o values during the winter (especially those under 1 mm day^{-1}) might contribute to provide higher RRMSE values. On the other hand, the model ranking fluctuates within months, but models HS2 and PM2 tend to provide the most accurate estimations in Albacete when calculated benchmarks are considered, while PM3 and PM1 provide the highest errors, respectively, during January to June, and during July to December. However, the RRMSE differences between the models might be slight. In Badajoz, again PM2 and HS2 tend to provide the optimum estimations, while PM1 tends to provide the highest monthly errors. When the models are tested using lysimeter benchmarks, HS1 and PM1 might provide the most accurate estimations, respectively, from April to November, and from December to March. The highest errors are provided by PM2 and PM3. However, as mentioned the monthly RRMSE differences between some models might be very slight, and it seems difficult to provide absolute rankings.

Fig. 5 and Fig. 6 present the average monthly absolute MBE values of non-calibrated models for both types of benchmarks in Albacete and Badajoz, respectively. In Albacete, Fig. 5, the models PM3 and PM2 tend to present negative MBE values during year, and thus to underestimate FAO 56 P-M ET_o values (upper plot). The models HS2, HS3 and PM1,

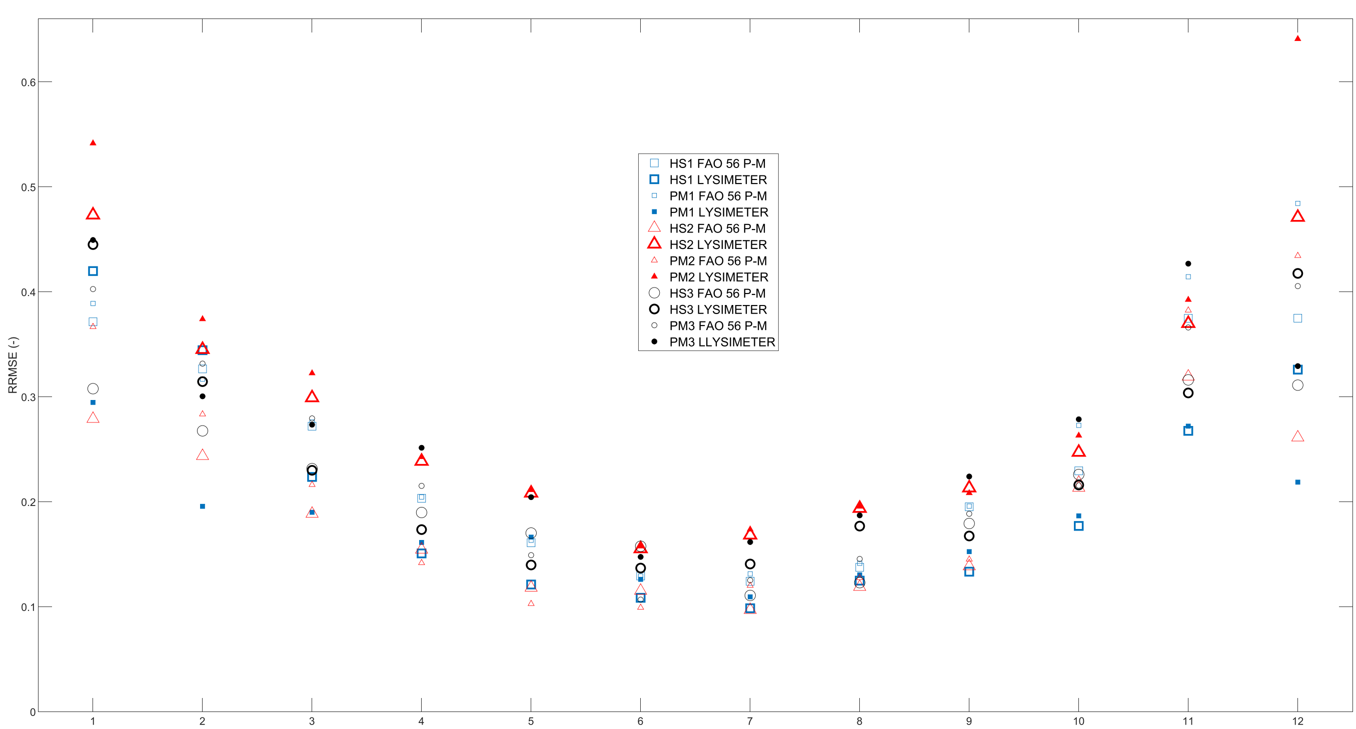

Fig. 3. Average RRMSE of non-calibrated HS and PM estimations against calculated and measured benchmarks per month in ALBACETE.

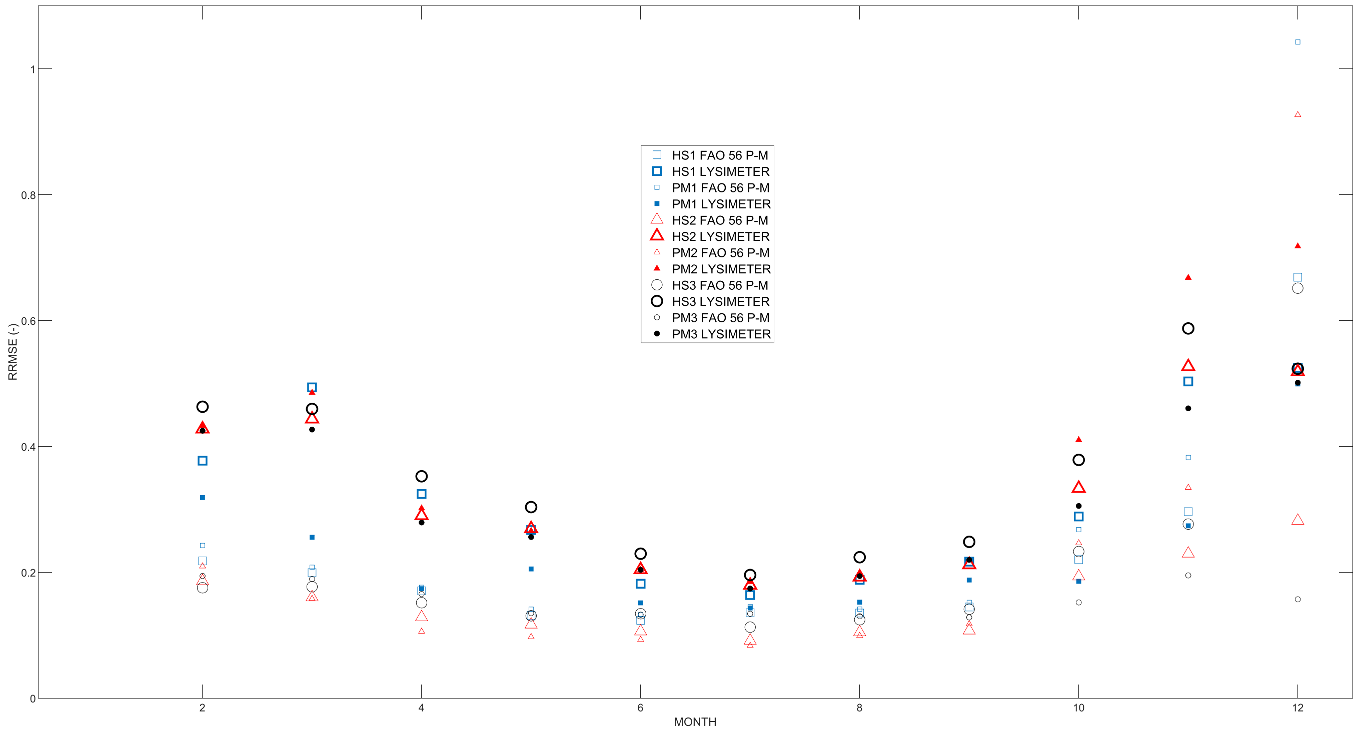

Fig. 4. Average RRMSE of non-calibrated HS and PM estimations against calculated and measured benchmarks per month in BADAJOZ.

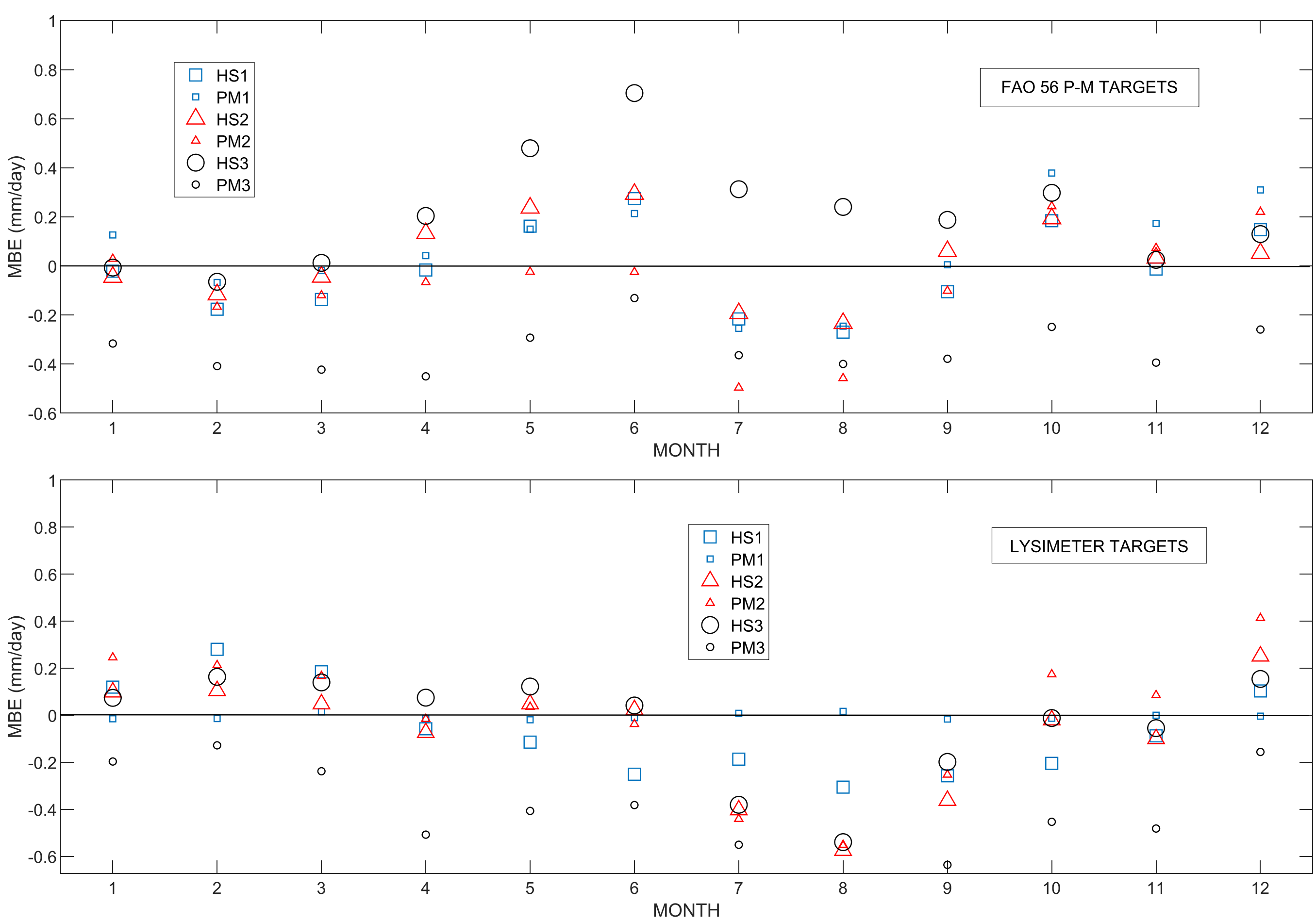

Fig. 5. Average absolute MBE of non-calibrated HS and PM estimations against calculated and measured benchmarks per month in ALBACETE.

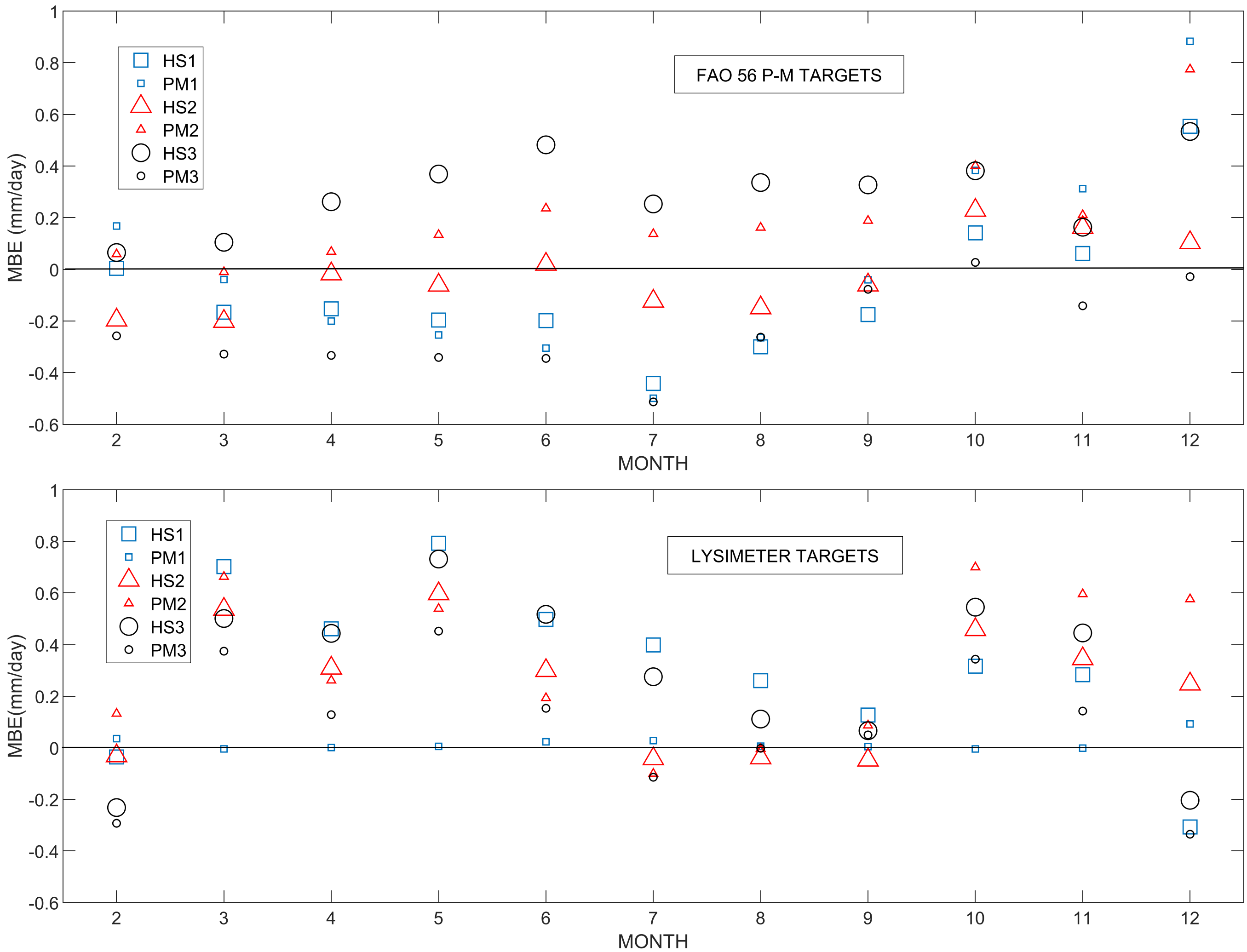

Fig. 6. Average absolute MBE of non-calibrated HS and PM estimations against calculated and measured benchmarks per month in BADAJOZ.

with the exception of some months, tend to present positive MBE values, and thus to overestimate FAO 56 P-M ET_o, while HS1 estimations do not show a clear pattern. When lysimeter benchmarks are considered, lower plot of Fig. 5, PM3 estimations tend to underestimate ET_o. The rest of models do not show a single pattern during the year. Most of them tend to underestimate during the summer, while they tend to overestimate during the winter. PM1 presents MBE values near 0 during the complete year. In Badajoz, Fig. 6, the models HS3 and PM2 tend to overestimate FAO 56 P-M ET_o values (upper plot) while PM3 and HS1 tend to underestimate ET_o. HS2 and PM1 tend to underestimate ET_o during the summer, while they tend to overestimate ET_o during the winter. According to the lower plot of Fig. 6, all the models tend to overestimate the lysimeter ET_o values, although HS3, PM3 and HS1 present some exceptions mainly during the winter months.

These trends are translated into the corresponding calibrating constants, providing, in general, values over 1 if MBE was negative, and values under 1 if MBE was positive. Tables 3 and 4 present for each month, station and calibrating benchmark the annual maximum, minimum and mean calibrating constant for HS and PM models, respectively. Given that the PM equations are normally not calibrated, the calibrating constants presented in Tables 3 and 4 refer to the factor that should multiply the non-calibrated ET_o estimation in order to be calibrated. The parameters of the HS equations might have been adjusted to the original scale (i.e. 0.0023, 0.0135 and 0.004563 in the corresponding original equations), i.e. multiplying the values of Table 3 by the original values. However, this was not possible in the PM equation, because it is not an empirical model, and therefore there is not an adjusted original constant. So, the presented values in Tables 3 and 4 correspond directly to the average ratio target ET_o value/non-calibrated ET_o value in order to allow a comparison of the parameters of the HS and PM equations. As mentioned above, Badajoz only presented climatic records in January for 2016. So, in this scenario maximum, minimum and mean values coincide. It can be stated that, in general, the annual variation range of the calibrating constants is higher in the winter months, while it decreases in the summer months. As showed in Figs. 3 and 4, the estimation accuracy of the models increased from winter to summer. So, this might contribute to provide a more stable monthly bias within the years during the summer, in contrast to the winter performance and bias. Further, the use of monthly average calibrating constants might allow a more suitable adjustment of the estimations during the summer, or, in general, in those months where the annual variation range is small. The same applies if annual or global calibrating constants are used. When a single calibrating constant is applied per station, the seasonal/monthly bias cannot be corrected. Thus, it seems reasonable to use monthly or, at least, seasonal calibrating constants to ensure a proper correction of the bias. On the other hand, the shorter the time window used for averaging the calibrating constants, the higher the potential variability of the bias within the years during that time window.

Table 3
Monthly ranges of the HS calibrating constants per station and type of benchmark for the period 2007–2015 in Albacete and 2007–2016 in Badajoz.

MODEL	MONTH	calibrating constant											
		FAO 56 P-M benchmarks						LYSIMETER benchmarks					
		ALBACETE			BADAJOZ			ALBACETE			BADAJOZ		
		min	max	mean	min	max	mean	min	max	mean	min	max	mean
HS1	1	0.8408	1.6581	1.0910	0.7092	0.7092	0.7092	0.5889	1.0737	0.9282	1.0509	1.0509	1.0509
	2	0.8116	1.4249	1.0950	0.7724	1.1007	0.9122	0.5740	1.1102	0.8784	0.6130	1.1622	0.8641
	3	0.8485	1.1705	1.0366	0.9621	1.1974	1.0568	0.8686	1.1714	0.9796	0.6209	1.1486	0.8249
	4	0.7734	1.1452	1.0042	0.8907	1.1536	1.0513	0.7971	1.2327	1.0343	0.7785	1.3795	0.9182
	5	0.8883	1.0515	0.9581	0.9646	1.1168	1.0421	0.8660	1.1404	0.9848	0.7712	1.0284	0.8924
	6	0.9352	0.9834	0.9541	0.9711	1.1436	1.0424	0.9064	1.0785	0.9960	0.8596	1.0527	0.9642
	7	1.0007	1.1046	1.0379	0.9441	1.1552	1.0782	0.8773	1.1779	1.0682	0.9452	1.2236	1.0178
	8	0.9652	1.1196	1.0504	0.9589	1.1133	1.0599	0.9640	1.1997	1.1030	0.8828	1.2465	1.0149
	9	0.8984	1.0641	1.0141	0.9383	1.1580	1.0496	0.9499	1.1649	1.0794	0.9397	1.0901	1.0132
	10	0.8448	0.9742	0.9246	0.7578	1.0869	0.9557	0.8788	1.1057	1.0040	0.6439	0.9553	0.8378
	11	0.7817	1.3362	1.0085	0.7710	1.2204	0.9523	0.9010	1.3422	1.1156	0.5868	1.1299	0.8120
	12	0.7206	1.3988	0.9877	0.4103	0.6183	0.5143	0.6575	2.9487	1.1545	0.0524	0.8922	0.4723
HS2	1	0.8041	1.2601	1.0717	0.9411	0.9411	0.9411	0.7003	1.1700	0.9383	1.4394	1.4394	1.4394
	2	0.8566	1.2535	1.0356	0.7493	1.2356	1.0091	0.4927	1.2145	0.8428	0.5947	1.3345	0.9700
	3	0.8377	1.1563	1.0271	0.9348	1.1761	1.0909	0.8868	1.1564	0.9738	0.6717	1.3216	0.8583
	4	0.8429	1.1433	0.9794	0.9020	1.1771	1.0168	0.8715	1.0967	1.0008	0.6885	1.4183	0.8977
	5	0.8984	0.9936	0.9553	0.9162	1.1738	1.0239	0.9350	1.0381	0.9775	0.7205	1.1686	0.8784
	6	0.8696	1.0322	0.9567	0.9143	1.1928	1.0132	0.8956	1.0656	0.9974	0.7420	1.2275	0.9414
	7	0.9846	1.0799	1.0319	0.9526	1.1661	1.0273	0.9202	1.1385	1.0608	0.8111	1.1590	0.9736
	8	1.0151	1.1235	1.0485	0.9575	1.1745	1.0388	0.9473	1.2321	1.1006	0.8265	1.2564	1.0000
	9	0.9141	1.1481	1.0011	0.9484	1.1519	1.0255	0.8794	1.2954	1.0689	0.8567	1.3076	1.0034
	10	0.8679	1.0504	0.9511	0.8712	1.0201	0.9365	0.9504	1.1019	1.0289	0.6166	1.1974	0.8359
	11	0.7876	1.3943	0.9865	0.8618	1.0793	0.9093	0.8703	1.3110	1.0880	0.5542	0.9814	0.7986
	12	0.7950	1.3387	0.9830	0.7476	0.9663	0.8569	0.7387	1.7893	1.0818	0.0954	1.4462	0.7708
HS3	1	0.9088	1.4464	1.0453	0.6980	0.6980	0.6980	0.5863	1.0584	0.8968	1.0051	1.0051	1.0051
	2	0.7766	1.2626	1.0065	0.7749	1.0193	0.8972	0.5364	0.9810	0.8104	0.6150	1.0767	0.8447
	3	0.7848	1.1383	0.9759	0.8995	1.0522	0.9645	0.8064	1.1881	0.9298	0.5588	1.0742	0.7570
	4	0.7606	1.0465	0.9365	0.8389	1.0201	0.9479	0.7832	1.1255	0.9634	0.6894	1.2442	0.8305
	5	0.8215	0.9737	0.8960	0.8810	0.9960	0.9383	0.8003	1.0553	0.9208	0.7102	0.9376	0.8045
	6	0.8725	0.9100	0.8900	0.8675	1.0007	0.9367	0.8517	1.0053	0.9287	0.7859	0.9646	0.8669
	7	0.9262	1.0144	0.9599	0.8752	1.0271	0.9725	0.8188	1.0805	0.9874	0.8576	1.0879	0.9186
	8	0.8798	1.0223	0.9646	0.8833	1.0075	0.9530	0.8948	1.1066	1.0120	0.7875	1.1094	0.9132
	9	0.8389	0.9851	0.9428	0.8634	0.9995	0.9384	0.8857	1.0855	1.0039	0.8477	1.0112	0.9090
	10	0.8163	0.9259	0.8811	0.7242	0.9712	0.8781	0.8407	1.0540	0.9571	0.6046	0.8543	0.7715
	11	0.7594	1.2146	0.9604	0.7334	1.0794	0.8862	0.9002	1.2767	1.0688	0.5744	1.0410	0.7598
	12	0.7483	1.2504	0.9462	0.4340	0.6341	0.5340	0.6983	2.6421	1.1022	0.0554	0.9210	0.4882

3.4. Fortnightly performance of calibrated HS and PM models with different calibrating time windows

Fig. 7 and Fig. 8 present the fortnightly RRMSE values of the non-calibrated and calibrated versions of the models using calculated FAO 56 P-M benchmarks in Albacete and Badajoz, respectively. Thus, these plots allow to visualize the error decrease caused by the different calibrating time windows throughout the year for the different models. Similarly, Fig. 9 and Fig. 10 present the same comparison relying on lysimeter benchmarks for calibrating and testing. Fortnights 1 and 2 were missing in Badajoz. In general, the following patterns can be stated. First, as mentioned above, the performance accuracy of the models increases from winter to summer, and is markedly higher when calculated benchmarks are considered, similarly to Fig. 3 and Fig. 4. High errors of estimation might be found in the winter months. There might be two factors causing these high errors: i) during winter the daily ET_o estimates are lower than 1 mm day^{-1}, and the RRMSE is a relative measure; ii) during a fortnightly period the number of observations is very reduced; iii) the probability of data gaps is higher during winter, and therefore the number of observations might be lower than 15. Second, in agreement with Tables 1 and 2, a smaller calibrating time window provides higher error decreases in comparison to the non-calibrated models, while annual calibrating constants might not decrease, or even increase, the estimating error. Third, the error decreases of the calibrations are more marked when the non-calibrated models were less accurate, i.e. usually during winter. This is particularly clear in the fortnights 21–24 (November and December) in Albacete (Fig. 7) and in the fortnights 3 (February) and 23–24 (December) in Badajoz (Fig. 8). When lysimeter benchmarks are considered, the period where the error decreases are more marked is longer and might comprise the fortnights 1–7 and 20–24, because the performance of the non-calibrated models is also worser during more months than if calculated benchmarks are considered. Fourth, the effect of the calibrating time windows in the error decreases is similar within models, despite there are accuracy differences between models, according to the rankings already mentioned based on Tables 1 and 2. The error decrease due to calibration and shown in Figs. 7–10 can be quantitatively summed up in Table 5, where the average fortnightly RRMSE decrease for each calibrating time window, station and benchmark type is presented. For calculated targets, the RRMSE decrease ranges in Albacete between 0.0018 (PM2) - 0.0275 (PM3) for the annual calibration, 0.0272 (HS1) - 0.0572 (PM3) for the monthly calibration, 0.0402 (HS2) - 0.0709 (PM3) for the fortnightly calibration, and 0.0520 (HS2) - 0.0858 (PM3) for the weekly calibration. In Badajoz, the RRMSE decrease ranges between 0.0040 (HS2) - 0.0413 (HS3) for the annual calibration, 0.0263 (PM3) - 0.1162 (PM1) for the monthly calibration, 0.0308 (HS2) - 0.1262 (PM1) for the fortnightly calibration, and 0.0389 (HS2) - 0.1374 (PM1) for the weekly calibration. When lysimeter calibrating and testing benchmarks are used, the RRMSE decrease ranges in Albacete between 0.0006 (HS2) −0.0183 (PM3) for the annual calibration, 0.0308 (HS2) - 0.0691 (PM1) for the monthly calibration, 0.0383 (HS2) - 0.0843 (PM1) for the fortnightly calibration, and 0.0564 (HS2) - 0.1046 (PM1) for the weekly calibration. In Badajoz, the RRMSE decrease ranges between 0.0167 (PM3) - 0.0807 (HS3) for the annual calibration, 0.0493 (PM3) −0.1334 (PM2) for the

Table 4
Monthly ranges of the PM calibrating constants per station and type of benchmark for the period 2007–2015 in Albacete and for the period 2007–2016 in Badajoz.

MODEL	MONTH	calibration constant											
		FAO 56 P-M benchmarks						LYSIMETER benchmarks					
		ALBACETE			BADAJOZ			ALBACETE			BADAJOZ		
		min	max	mean	min	max	mean	min	max	mean	min	max	mean
PM1	1	0.7296	1.4903	0.9693	0.6268	0.6268	0.6268	0.5115	0.9673	0.8241	0.9358	0.9358	0.9358
	2	0.7606	1.3295	1.0272	0.7356	1.0212	0.8503	0.5390	1.0353	0.8244	0.5838	1.0788	0.8049
	3	0.8242	1.1157	0.9936	0.9046	1.1626	1.0189	0.8195	1.1421	0.9396	0.5941	1.1150	0.7954
	4	0.7688	1.1331	0.9899	0.8886	1.1843	1.0720	0.7927	1.2189	1.0203	0.7761	1.5798	0.9409
	5	0.8964	1.0567	0.9623	0.9648	1.1360	1.0570	0.8741	1.1463	0.9894	0.7770	1.0364	0.9046
	6	0.9421	0.9962	0.9647	0.9880	1.1706	1.0630	0.9157	1.0920	1.0072	0.8721	1.0773	0.9832
	7	1.0045	1.1182	1.0456	0.9487	1.1720	1.0902	0.8854	1.1927	1.0762	0.9493	1.2378	1.0289
	8	0.9652	1.1176	1.0480	0.9426	1.1114	1.0562	0.9704	1.1890	1.1005	0.8791	1.2432	1.0108
	9	0.8789	1.0425	0.9891	0.8951	1.1343	1.0222	0.9324	1.1454	1.0530	0.9067	1.0563	0.9857
	10	0.7934	0.9027	0.8648	0.6963	1.0008	0.8861	0.8319	1.0337	0.9392	0.5921	0.8786	0.7767
	11	0.7057	1.1829	0.9045	0.6743	1.1006	0.8408	0.7910	1.2013	1.0021	0.4934	1.0039	0.7169
	12	0.6201	1.2251	0.8605	0.3588	0.5046	0.4317	0.5679	2.5885	1.0078	0.0458	0.7336	0.3897
PM2	1	0.7931	1.5994	1.0499	0.6830	0.6830	0.6830	0.5459	1.0553	0.8946	1.0290	1.0290	1.0290
	2	0.8728	1.3556	1.0663	0.7328	1.0764	0.8831	0.5415	1.1418	0.8609	0.5816	1.1438	0.8404
	3	0.8719	1.1577	1.0286	0.9579	1.1033	1.0030	0.8804	1.1582	0.9737	0.6029	1.1492	0.7861
	4	0.8383	1.1037	1.0136	0.9344	1.0628	0.9875	0.8657	1.1387	1.0413	0.6930	1.4061	0.8691
	5	0.9472	1.0451	0.9942	0.9280	1.0382	0.9801	0.9572	1.1289	1.0194	0.7427	1.0534	0.8397
	6	0.9440	1.0431	1.0028	0.9253	1.0444	0.9721	0.9469	1.0908	1.0463	0.7529	1.1011	0.9013
	7	1.0372	1.0999	1.0784	0.9487	1.0398	0.9853	0.9432	1.1708	1.1090	0.8193	1.0500	0.9321
	8	1.0280	1.1170	1.0814	0.9353	1.0353	0.9801	0.9974	1.2446	1.1353	0.7967	1.1024	0.9409
	9	0.9481	1.0881	1.0131	0.9297	1.0163	0.9679	0.9360	1.2264	1.0807	0.8363	1.0884	0.9392
	10	0.8656	0.9354	0.9083	0.7392	0.9738	0.8774	0.8939	1.0706	0.9856	0.6234	0.9416	0.7729
	11	0.7748	1.2490	0.9623	0.7185	1.1308	0.8884	0.8371	1.2809	1.0665	0.5225	1.0678	0.7583
	12	0.6630	1.4100	0.9590	0.3904	0.5448	0.4676	0.6201	2.9791	1.1294	0.0498	0.8001	0.4250
PM3	1	1.2696	1.5705	1.3690	1.0695	1.0695	1.0695	0.8572	1.4537	1.2020	1.4884	1.4884	1.4884
	2	1.0179	1.4870	1.2563	0.9541	1.2107	1.0763	0.6258	1.2572	1.0180	0.7572	1.2817	1.0074
	3	0.9728	1.2880	1.1599	1.0323	1.2159	1.1174	0.9706	1.3231	1.0997	0.6215	1.3440	0.8851
	4	0.9798	1.2797	1.1422	1.0273	1.1559	1.0950	1.0082	1.3780	1.1753	0.7809	1.4435	0.9623
	5	0.9882	1.1399	1.0561	0.9668	1.1398	1.0711	0.9631	1.2353	1.0854	0.8147	1.0639	0.9183
	6	0.9818	1.0447	1.0220	0.9980	1.1364	1.0688	0.9783	1.1622	1.0660	0.8951	1.1203	0.9895
	7	0.9811	1.1274	1.0622	0.9774	1.1527	1.0884	0.9043	1.1992	1.0916	0.9443	1.2158	1.0275
	8	0.9804	1.1309	1.0801	0.9512	1.1352	1.0520	1.0276	1.2184	1.1313	0.8627	1.2241	1.0079
	9	1.0028	1.1545	1.0964	0.9220	1.0660	1.0244	1.0441	1.2992	1.1691	0.9189	1.0894	0.9945
	10	1.0453	1.1473	1.1161	0.9063	1.1024	0.9917	1.0754	1.3599	1.2138	0.7128	1.0570	0.8775
	11	1.1401	1.4929	1.3233	0.9957	1.1673	1.0749	1.1987	1.7707	1.4926	0.6825	1.2926	0.9494
	12	1.1064	1.6333	1.3497	0.6037	1.0188	0.8113	0.9143	3.4496	1.5512	0.0771	1.4376	0.7573

monthly calibration, 0.0662 (PM3) - 0.1469 (PM2) for the fortnightly calibration, and 0.0855 (PM3) - 0.1635 (PM2) for the weekly calibration.

Finally, a comparison with previous applications of the HS and PM equations that can be found in the bibliography seems difficult, due the particular scope adopted in this work. The main differences in scope presented here can be summarized as follows: i) the intra annual performance of the non-calibrated models is assessed; ii) monthly, fortnightly and weekly calibrations were provided in addition to annual calibrations; iii) the intra annual improvement rate of the different calibration time steps is assessed; and iv) the effect of using FAO 56 P-M calculated benchmarks for calibrating and testing is compared with lysimeter ones.

3.5. Comparison with related previous research results

The consideration of monthly AHCs for calibrating the HS1 equation was already presented in previous studies (e.g. Maestre-Valero et al., 2013; Tabari and Hosseinzadeh Talaee, 2011) in contrast to the great part of studies, which just considered a single calibrating constant per station. Maestre-Valero et al. (2013) reported an average relative error of 19.8 % for the original parameterization of the HS1 estimates in the Murcia region (Spain). A single global calibrating constant reduced the average relative error to 9.6 %, while the application of monthly calibrating constants reduced the average relative error to 7.71 %, quite lower to the presented RRMSE values in the current study. However, that study was based on the application of monthly averaged values in the HS1 equation. Similarly, Tabari and Hosseinzadeh (2011) reported in Iran a reduction of the percent error from 24.7 % to 1.42 % in arid stations, and from 17.4 % to 1.59 % between non-calibrated and locally calibrated HS1 equations in cold stations. In recent years, (Martí et al., 2015c) assessed the application of the HS1 equation relying on averaged inputs considering different time scales (day, week, fortnight, and month) in the Mediterranean coast of Spain. Those intervals provided RRMSE (unitless) reductions between non-calibrated and calibrated estimations of, respectively, 0.223 vs. 0.197 (day), 0.168 vs. 0.133 (week), 0.145 vs. 0.104 (fortnight), and 0.141 vs. 0.095 (month). As could be expected, the error measures decreased when the timescale increased, due to the variability reduction associated to the use of mean values. The reduction in RRMSE values for the calibrated HS1 estimates with respect to the non-calibrated ones was higher for the weekly, fortnightly and monthly timescales than for the daily timescales. Similarly, considering daily, weekly and monthly estimates, Bachour et al. (2013) found RMSE decreases between non-calibrated and calibrated HS1 estimates of 0.600 vs. 0.481 mm day^{-1} (day), 0.387 vs. 0.333 mm day^{-1} (week), and 0.295 vs. 0.253 mm day^{-1} (month), respectively. These results are in agreement with Hargreaves and Allen (2003), who found optimal accuracies for five-day or longer timescales. However, the application of monthly or shorter calibrating constants for adjusting daily HS and PM estimates could not be found. López-Urrea et al. (2006) incorporated the seasonal assessment of different non-calibrated ET_o equations using lysimeter benchmarks, analysing their performance in two periods, namely: period of high evaporative demand (defined as April to September), and period of low evaporative demand (defined as

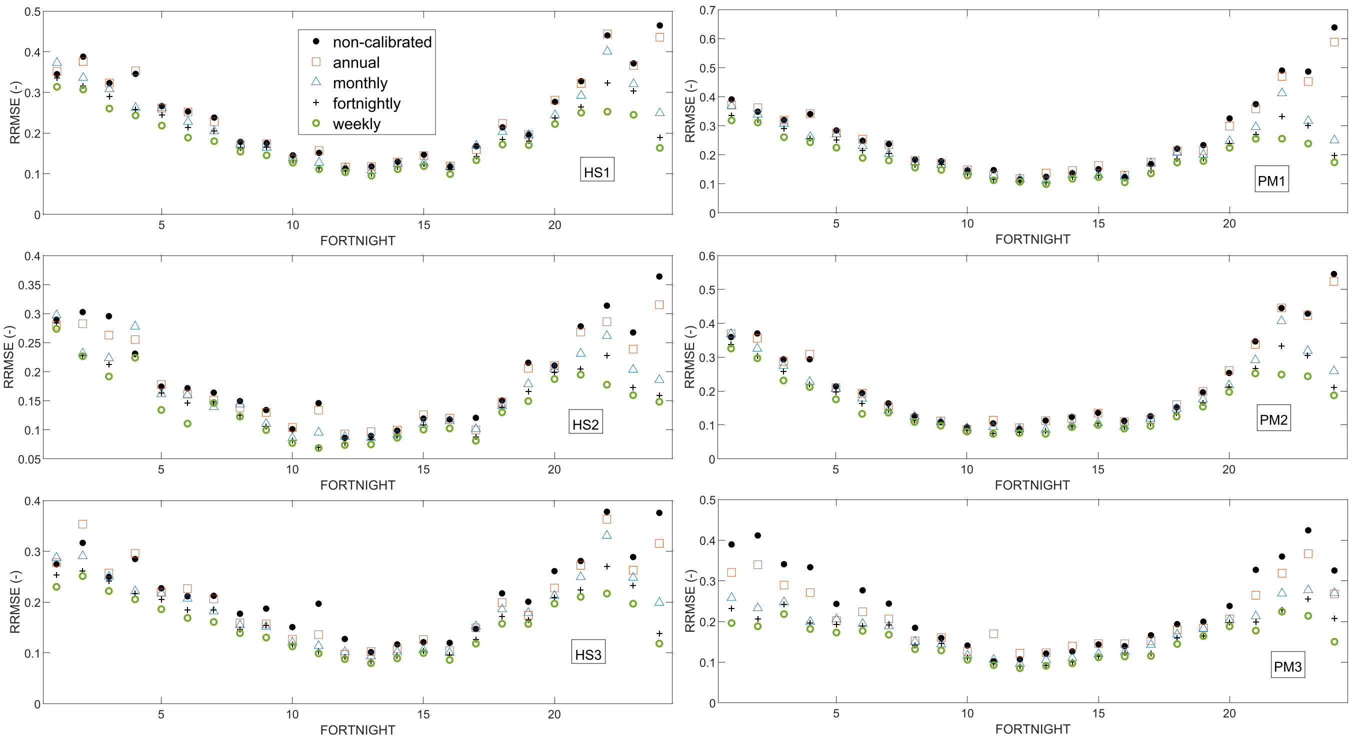

Fig. 7. Average RRMSE of calibrated HS and PM estimations against FAO 56 P-M benchmarks per fortnight in ALBACETE.

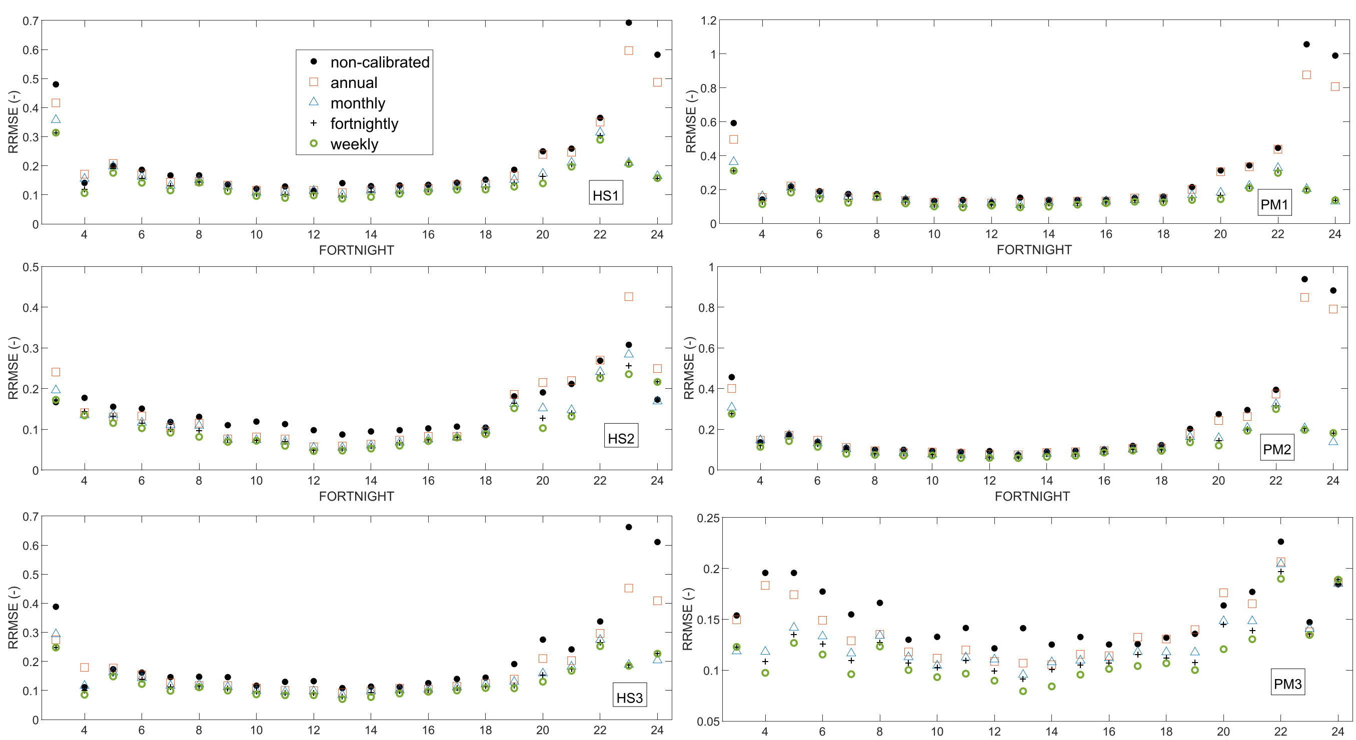

Fig. 8. Average RRMSE of calibrated HS and PM estimations against FAO 56 P-M benchmarks per fortnight in BADAJOZ.

October to March). The error indicators decreased in the period of high evaporative demand (e.g. HS1 estimates presented average RRMSE values of 20.52 % in April-September vs. 34.32 % in October-March) in agreement with the currently presented results. Senatore et al. (2020) assessed, among others, the performance of non-calibrated and locally fitted HS1 and PM1 estimates in 101 stations in northeastern Spain. The global mean absolute percentage error for all stations decreased from 29.8 % to 26.4 % between non-calibrated and calibrated HS1 estimates, and from 34.4 % to 33.9 % between non-calibrated and calibrated PM1 estimates, considering a single constant per station and a daily time-scale. On the other hand, Martí et al. (2015a) compared, among others, the HS1 equation with a model based on Gene Expression Programming

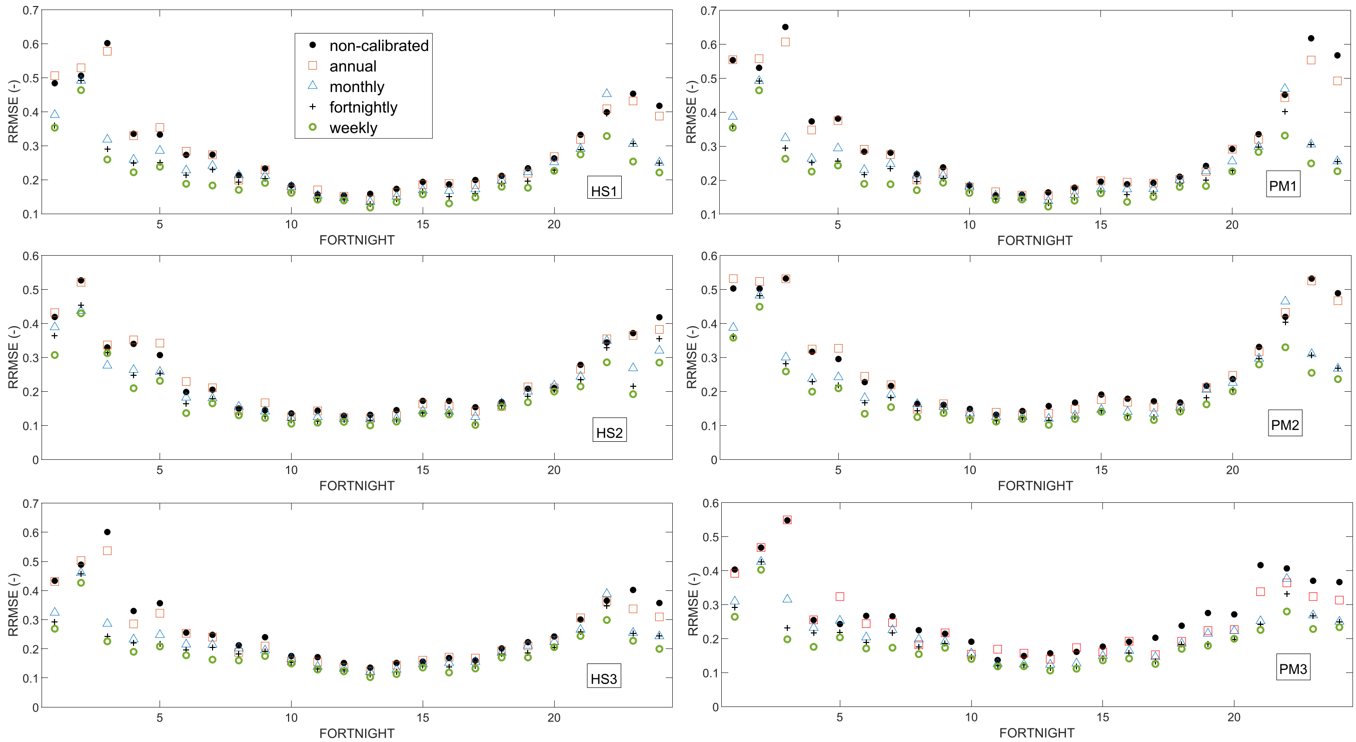

Fig. 9. Average RRMSE of calibrated HS and PM estimations against lysimeter benchmarks per fortnight in ALBACETE.

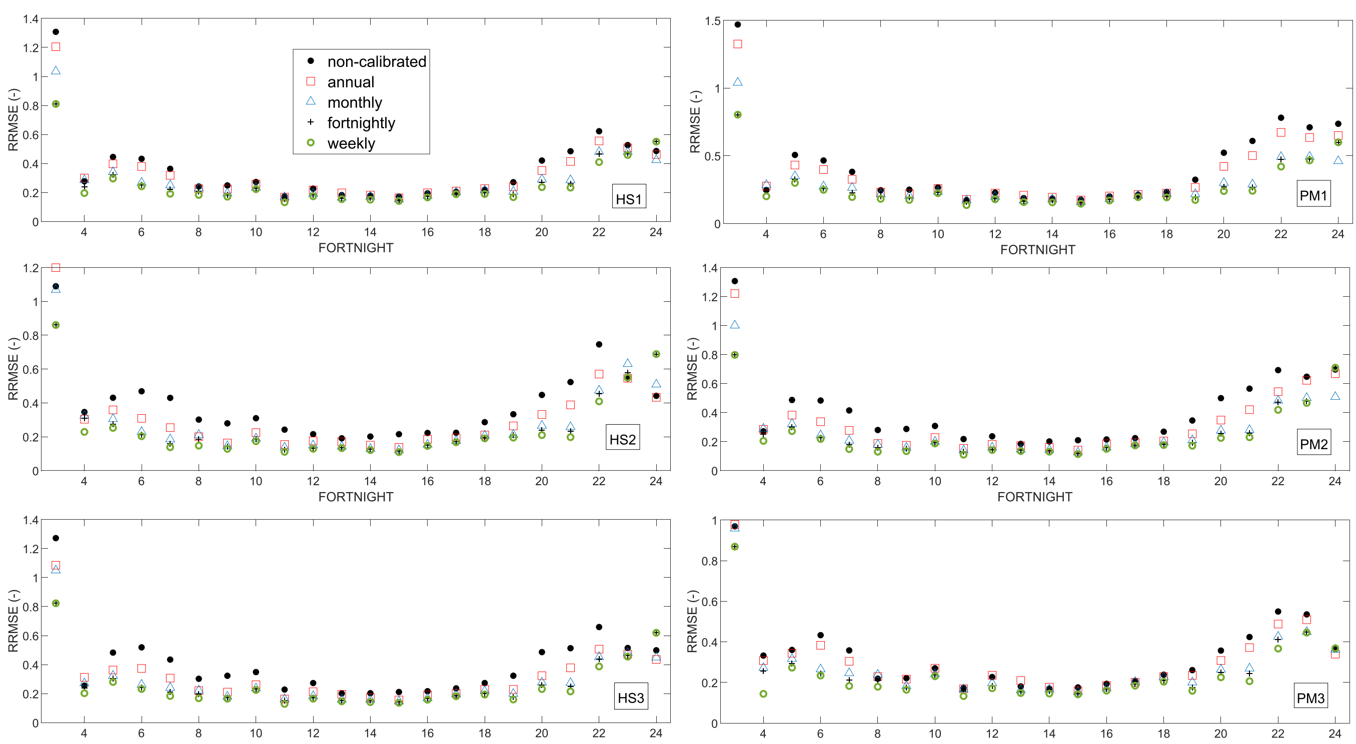

Fig. 10. Average RRMSE of calibrated HS and PM estimations against lysimeter benchmarks per fortnight in BADAJOZ.

relying on the same inputs (GEP4). These models were evaluated locally (i.e. training and testing in the same station) and externally (i.e. training in one station and testing in the other one). The RRMSE values (unitless) of the local performance were 0.1720 (FAO 56 P-M benchmarks) and 0.1515 (lysimeter benchmarks) in Albacete, and 0.1235 (FAO 56 P-M benchmarks) and 0.1236 (lysimeter benchmarks) in Badajoz. The RRMSE values (unitless) of the external performance increased to 0.2847 (FAO 56 P-M benchmarks) and 0.2762 (lysimeter benchmarks) in Albacete, and to 0.1514 (FAO 56 P-M benchmarks) and 0.2506 (lysimeter benchmarks) in Badajoz. Finally, regarding the application of independent test sets for assessing the calibration performance of HS1 estimates, Shiri et al. (2015) found only very slight performance

Table 5
Average fortnightly RRMSE decrease per calibrating time window, station and benchmark type.

STATION	MODEL	FAO 56 P-M BENCHMARKS				LYSIMETER BENCHMARKS			
		ANNUAL	MONTH	FORTNIGHT	WEEK	ANNUAL	MONTH	FORTNIGHT	WEEK
Albacete	HS1	0.0023	0.0272	0.0447	0.0627	0.0038	0.0451	0.0597	0.0796
	HS2	0.0086	0.0278	0.0402	0.0520	0.0006	0.0308	0.0383	0.0564
	HS3	0.0137	0.0320	0.0480	0.0626	0.0163	0.0515	0.0657	0.0838
	PM1	0.0092	0.0460	0.0630	0.0812	0.0112	0.0691	0.0843	0.1046
	PM2	0.0018	0.0349	0.0499	0.0658	0.0020	0.0532	0.0660	0.0843
	PM3	0.0275	0.0572	0.0709	0.0858	0.0183	0.0565	0.0721	0.0903
Badajoz	HS1	0.0204	0.0659	0.0748	0.0850	0.0250	0.0684	0.0869	0.1037
	HS2	0.0040	0.0271	0.0308	0.0389	0.0711	0.1031	0.1187	0.1363
	HS3	0.0413	0.0727	0.0790	0.0880	0.0807	0.1121	0.1284	0.1444
	PM1	0.0287	0.1162	0.1262	0.1374	0.0397	0.1140	0.1336	0.1504
	PM2	0.0187	0.0954	0.1006	0.1100	0.0773	0.1334	0.1469	0.1635
	PM3	0.0132	0.0263	0.0315	0.0404	0.0167	0.0493	0.0662	0.0855

differences in Iran between HS1 calibrated estimates, when they were assessed reserving one year for testing through a k-fold validation, and when all timeseries were used for both calibrating and testing. In particular, the k-fold assessment of the calibrated estimates provided mean absolute relative errors (unitless) of 0.187 and 0.148 for coastal and inland stations, respectively, while without reserving independent data for testing the errors were, respectively 0.185 and 0.154. Martí et al. (2015b) presented similar results for the Mediterranean coast of Spain. The parametric calibration of monthly or, at least, seasonal constants might be tackled in further research.

4. Effect of the ET_o estimation method and its seasonal trends on crop water requirements. Examples

In order to visualize the possible effect of the ET_o estimation method on the annual crop water requirements (CWR), six scenarios were assessed in Las Tiesas station, namely: almond, maize (two cycles), wheat (two cycles) and onion. These are 4 common crops in Albacete. Therefore, the theoretical annual CWR of these crops were calculated according to the specific crop coefficients, the lengths of crop development stages, and the plant dates proposed in Allen et al. (1998). Further, only lysimeter, FAO 56 P-M, and HS1 ET_o values were considered for simplification purposes. A thorough analysis of all crops and possible cycle lengths is beyond the scope of this section. Table 6 presents the crop cycle information required for the calculations, extracted from FAO 56 (Allen et al., 1998), and the resulting average annual CWR per crop and model. Thus, in Almond (1. March, and 240 cycle days), the consideration of FAO 56 P-M and HS1 ET_o values lead to average annual under-endowments of 277 m^3 (FAO 56 P-M), 297 m^3(HS1), 20 m^3 (HS1 if calculated benchmarks are considered). The average annual CWR would be 7740 m^3 (lysimeter), 7463 m^3 (FAO 56 P-M), and 7442 m^3 (HS1). Average annual CWR based on lysimeter ET_o values and FAO 56 crop coefficients ($ET_o \cdot K_c$) resulted in 7493 m^3 for Maize 1 (plant date 1. April, and 180 cycle days). Similarly, 7229 m^3 and 7143 m^3 would be required if FAO 56 P-M and HS1 ET_o values are considered, respectively. So, the consideration of FAO 56 P-M and HS1 instead of lysimeter values would lead to average annual under-endowments of 264 and 350 m^3, respectively. Further, if FAO 56 P-M values are considered as benchmarks, the consideration of HS1 estimations would be evaluated as an under-endowment of 86 m^3 (instead of 350 m^3). In Maize 2 (plant date 1. June, and 125 cycle days), the average annual under-endowments present a similar order of magnitude, namely: 230 m^3 (FAO 56 P-M), 376 m^3 (HS1), 146 m^3 (HS1 if calculated benchmarks are considered). The average annual CWR would be 5533 m^3 (lysimeter), 5302 m^3 (FAO 56 P-M), and 5156 m^3 (HS1). These values are lower than in Maize 1, because the cycle is 55 days shorter. In spring wheat 2 (plant date 1. July, and 150 cycle days) the average annual CWR would be 4862 (lysimeter), 4606 m^3 (FAO 56 P-M), and 4537 m^3 (HS1). This involves average annual under-endowments of 255 m^3 (FAO 56 P-M), 325 m^3 (HS1), and 69 m^3 (HS1 if calculated benchmarks are considered). In spring wheat 1 (plant date 1. March, and 135 cycle days) the average annual CWR present similar ranges, i.e. 4190 m^3 (lysimeter), 4097 m^3 (FAO 56 P-M), and 4202 m^3 (HS1). However, this is translated into and under-endowment of 93 m^3 if FAO 56 P-M estimations are used, while the consideration of HS1 is translated into over-endowments of 12 m^3 (vs. lysimeter) and 93 m^3 (vs. FAO 56 P-M). Finally, onion (plant date 1. April, and 150 cycle days) presents average annual CWR of 6726 m^3 (lysimeter), 6524 m^3 (FAO 56 P-M), and 6526 m^3 (HS1). This is translated into under-endowments of 202 m^3 (FAO 56 P-M) and 199 m^3 (HS1), while the consideration of HS1 is translated into a slight

Table 6
Average annual crop water requirement estimation for lysimeter, FAO56 P-M, and HS1 ET_o values (CWR: crop water requirements, LYS: lysimeter, PM: FAO 56 Penman-Monteith, HS1: Hargreaves based on temperature range).

	Almond	Maize 1 (Low grain moisture)	Maize 2 (High grain moisture)	Spring wheat 1	Spring wheat 2	Onion (dry)
plant date/start of season	01 March	01 April	02 June	01 March	01 July	01 April
Initial stage length (days)	30	30	20	20	15	15
Development stage length (days)	50	50	35	25	30	25
Middle stage length (days)	130	60	40	60	65	70
Late stage length (days)	30	40	30	30	40	40
Total length (days)	240	180	125	135	150	150
Kc initial stage	0.4	0.3	0.3	0.3	0.3	0.7
Kc middle stage	0.9	1.2	1.2	1.15	1.15	1.05
Kc end	0.65	0.35	0.6	0.25	0.25	0.75
CWR LYS (m^3)	7740.542	7493.676	5533.739	4190.322	4862.500	6726.447
CWR PM (m^3)	7463.381	7229.347	5302.947	4097.082	4606.805	6524.412
CWR HS1 (m^3)	7442.640	7142.893	5156.830	4202.695	4537.257	6526.566
CWR LYS - CWR PM (m^3 ha^{-1})	277.161	264.328	230.792	93.241	255.694	202.035
CWR LYS - CWR HS1 (m^3 ha^{-1})	297.902	350.783	376.910	-12.373	325.242	199.881
CWR PM - CWR HS1 (m^3 ha^{-1})	20.742	86.454	146.118	-105.614	69.548	-2.154

over-endowment of 2 m^3 for HS1 if FAO 56 P-M benchmarks are considered.

The under-/over-endowment trends are consistent with the MBE values presented in Fig. 5. Within the 6 considered crop cycles, the method used to estimate ET_o does not seem to provide large enough differences in the corresponding annual endowments, even if the calculated ET_o estimates presented a lower estimation accuracy of lysimeter values. This should be confirmed in future research covering all the crop cycles proposed in Allen et al. (1998). However, the annual endowment might be hiding eventual relevant differences between the daily requirements calculated based on different ET_o approaches, due to changes in the daily over-/underestimation pattern of the ET_o estimates. The annual endowment trend will depend on the ET_o MBE trend during the specific months (or even days) corresponding to the crop developing stages. So, at least, a monthly MBE assessment of the ET_o estimates in combination with the crop cycle dates might allow to infer if the estimated annual endowment would be accurate, excessive or loss-making.

If irrigation scheduling is based on ET_c estimates, special attention should be paid if negative MBE values are identified during crop sensitive stages to water deficit. It seems difficult to define accurately the exact dates of the theoretical FAO56 crop cycle where the crop might be sensitive to water deficit, because this will depend on many other on-site factors in real-time conditions. However, it might be possible to define potential approximate sensitive periods and to assess the MBE trends of the ET_o estimates during those periods. Accordingly, possible sensitive dates to water stress in annual crops might correspond to the midseason stage. Thus, attending to the cycles presented in Table 6, the sensitive periods might eventually take place approximately during July-August (Maize 1), August (Maize 2), May (Spring wheat 1), September (Spring-wheat 2), June-July (Onion). In Almond, the possible sensitive dates might take place during spring and autumn in Albacete (thus, eventually during April-May, and October-November). The analysis of the monthly MBE values in Albacete (Fig. 5) shows that the HS1 estimates tend to present negative values between April and November if lysimeter benchmarks are considered. Thus, all possible dates where the previous crops might be sensitive to water deficit present negative MBE ET_o values. So, if the irrigation CWR are calculated exclusively relying on a soil water balance, and ET_c is calculated using HS1 ET_o estimates, the calculated irrigation doses might cause water stress during crop sensitive stages. On the other hand, if FAO 56 P-M estimates are used as benchmarks, the HS1 estimates present positive MBE values, among others, during May-June, and October, and negative values during July-September. Thus, even if monthly MBE ET_o values are incorporated to the assessment of the irrigation doses, the user might have a false sense of being overdosing irrigation water in a potentially sensitive period to water stress, when in fact the crop would be receiving less water than required.

In any case, the monthly/fortnightly/weekly assessment of the MBE ET_o estimates in combination with the crop cycle dates might contribute to detect if the theoretical doses based on calculated ET_c values might be lower than those required during stages where the crop might be sensitive to water deficit. On the other hand, positive MBE ET_o trends might indicate that overdoses are being scheduled. Thus, in this regard, the seasonal assessment of the ET_o MBE values might be more relevant than the assessment of the corresponding RRMSE or MAE values.

5. Conclusions

Regarding the non-calibrated HS models, HS2 (based on T_m and R_s) tended to provide more accurate estimates than HS3 (based on T_m and RH_{mean}), while HS3 tended to provide more accurate estimates than HS1 (based on ΔT) for both FAO 56 P-M and lysimeter benchmarks, respectively. Non-calibrated PM estimations provided similar qualitative patterns than those found between HS models for all types of benchmarks and calibrating timescales. The error parameters of the PM estimations were just slightly higher than the error indexes corresponding to the HS estimations.

For both HS and PM models, and both types of benchmarks, the calibrations considering global and annual mean calibrating constants provided in general very slight accuracy improvements. On the other hand, the calibrations considering monthly to weekly mean calibrating constants provided more relevant accuracy improvements. The improvement was more marked when the time window considered for averaging was shorter. Thus, the application of monthly or, at least, seasonal calibrating constants would be desirable to properly adjust the bias of the original estimates. Lysimeter benchmarks provided similar qualitative conclusions than calculated benchmarks regarding rankings and accuracy improvements derived from calibrating constants with decreasing time windows. However, the error range considerably increased. Attending to models that were calibrated using FAO 56 P-M benchmarks, but tested using lysimeter ones, the performance patterns were similar to the scenario where lysimeter benchmarks are used for calibrating and testing. In this case, the estimations provided higher errors.

There was a significant fluctuation of the model performance accuracy during the year, with considerably lower errors and lower differences within models during the summer, while presenting higher errors and higher differences within models during the winter. Thus, the models presented a higher mapping ability during the summer, where the considered inputs might have a higher effect of ET patterns. Regarding the effect of the calibrating time windows during the year, the error decrease of the calibrations was more marked when the non-calibrated models were less accurate, i.e. usually during winter. When lysimeter benchmarks were considered, the period where the error decreases is more marked was longer and might comprise the fortnights 1–7 and 20–24, because the performance of the non-calibrated models was also worser during more months than if FAO 56 P-M targets were considered. The effect of the calibrating time windows in the error decreases was similar within models.

If irrigation scheduling is based on a soil water balance using crop ET estimates, a monthly bias assessment of the ET_o estimates in combination with the crop cycle lengths and dates might contribute to infer if crop water requirement infra-estimation trends are identified during potential crop sensitive stages to water deficit.

CRediT authorship contribution statement

Armand Román: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Data curation, Conceptualization. **Pablo González-Altozano:** Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Luis A. Mancha:** Writing – original draft, Data curation, Conceptualization. **Ramón López-Urrea:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Pau Martí:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

This research has been funded by the Agencia Estatal de Investigación with FEDER (grant numbers PID2021-123305OB-C31 and

PID2020-113498RB-C21), and NextGenerationEU (TED2021-130405B-I00) co-financing.

References

Allen, R.G., 1996. Assessing integrity of weather data for reference evapotranspiration estimation. J. Irrig. Drain. Eng. 122 (2), 97–106.

Allen, R.G., 1997. Self-calibrating method for estimating solar radiation from air temperature. J. Hydrol. Eng. 2 (2), 56–67. https://doi.org/10.1061/(ASCE)1084-0699(1997)2:2(56).

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011a. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manag. 98 (6), 899–920. https://doi.org/10.1016/j.agwat.2010.12.015.

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011b. Evapotranspiration information reporting: II. Recommended documentation. Agric. Water Manag. 98 (6), 921–929. https://doi.org/10.1016/j.agwat.2010.12.016.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage, Paper 56. FAO, Rome.

Almorox, J., Quej, V.H., Marti, P., 2015. Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Koppen climate classes. J. Hydrol. (Amst. 528, 514–522. https://doi.org/10.1016/j.jhydrol.2015.06.057.

ASCE-EWRI, 2005. The ASCE standardized reference evapotranspiration equation. In: Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E.,Snyder, R.L. (Eds.), Report 0-7844-0805-X. Am. Soc. Civ. Eng.—Environ. WaterResources Instit., 69 pp (+App. A-F and Index).

Bachour, R., Walker, W.R., Torres-Rua, A.F., McKee, M., 2013. Assessment of reference evapotranspiration by the Hargreaves method in the Bekaa Valley, Lebanon. J. Irrig. Drain. Eng. ASCE 139 (11), 933–938. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000646.

Berengena, J., Gavilán, P., 2005. Reference evapotranspiration estimation in a highly advective semiarid environment. J. Irrig. Drain. Eng. 131 (2), 147–163. https://doi.org/10.1061/(ASCE)0733-9437(2005)131:2(147).

Berti, A., Tardivo, G., Chiaudani, A., Rech, F., Borin, M., 2014. Assessing reference evapotranspiration by the Hargreaves method in north-eastern Italy. Agric. Water Manag. 140 (7)), 20–25. https://doi.org/10.1016/j.agwat.2014.03.015.

Bristow, K.L., Campbell, G.S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric. For. Meteorol. 31 (2), 159–166. https://doi.org/10.1016/0168-1923(84)90017-0.

Di Stefano, C., Ferro, V., 1997. Estimation of evapotranspiration by Hargreaves formula and remotely sensed data in semi-arid Mediterranean areas. J. Agric. Eng. Res. 68 (3), 189–199. https://doi.org/10.1006/jaer.1997.0166.

Diodato, N., Bellocchi, G., 2007. Modeling reference evapotranspiration over complex terrains from minimum climatological data. Water Resour. Res. 43, W05444. https://doi.org/10.1029/2006WR005405.

Doorenbos, J., Pruitt, W.O., 1977. Guidelines for Predicting Crop Water Requirements. Rev. edn. Food and Agriculture Organization of the United Nations, Rome.

Droogers, P., Allen, R.G., 2002. Estimating reference evapotranspiration under inaccurate data conditions. Irrig. Drain. Syst. 16 (1), 33–45. https://doi.org/10.1023/A:1015508322413.

Exner-Kittridge, M.G., 2011. Closure of "case study on the accuracy and cost/effectiveness in simulating reference evapotranspiration in West-Central Florida" by Exner- Kittridge, M. C., Rains. M. C. J. Hydrol. Eng. 17 (1), 225–226. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000430.

Exner-Kittridge, M.C., Rains, M.C., 2010. Case study on the accuracy and cost/effectiveness in simulating reference evapotranspiration in West-Central Florida. J. Hydrol. Eng. 15 (9), 696–703. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000239.

Fooladmand, H.R., Haghighat, M., 2007. Spatial and temporal calibration of Hargreaves equation for calculating monthly ET_0 based on Penman-Monteith method. Irrig. Drain. 56 (4), 439–449. https://doi.org/10.1002/ird.305.

Gavilán, P., Lorite, I.J., Tornero, S., Berengena, J., 2006. Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. Agr. Water Manag. 81 (3), 257–281. https://doi.org/10.1016/j.agwat.2005.05.001.

Hargreaves, G.H., 1994. Defining and using reference evapotranspiration. J. Irrig. Drain. Eng. 120 (6), 1132–1139. https://doi.org/10.1061/(ASCE)0733-9437(1994)120:6(1132).

Hargreaves, G.H., Allen, R.G., 2003. History and evaluation of Hargreaves evapotranspiration equation. J. Irrig. Drain. Eng. 129 (1), 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53).

Hargreaves, G.L., Hargreaves, G.H., Riley, J.P., 1985. Irrigation water requirements for Senegal river basin. J. Irrig. Drain. Eng. 111 (3), 191–298. https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265).

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from ambient air temperature. Appl. Eng. Agric. 1 (2), 96–99. https://doi.org/10.13031/2013.26773.

Itenfisu, D., Elliott, R.L., Allen, R.G., Walter, I.A., 2003. Comparison of reference evapotranspiration calculation as part of the ASCE standardization effort. J. Irrig. Drain. Eng. 129 (6), 440–448. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:6(440).

Jensen, M.E., Burman, R.D., Allen, R.G., 1990. Evapotranspiration and irrigation water requirements. In: ASCE Manual and Report on Engineering Practice No, 70. ASCE, New York.

Jensen, D.T., Hargreaves, G.H., Temesgen, B., Allen, R.G., 1997. Computation of ET_o under nonideal conditions. J. Irrig. Drain. Eng. 123 (5), 394–400. https://doi.org/10.1061/(ASCE)0733-9437(1997)123:5(394).

Lee, K.H., 2010. Relative comparison of the local recalibration of the temperature-based evapotranspiration equation for the Korea Peninsula. J. Irrig. Drain. Eng. 136 (9), 585–594. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000221.

López-Urrea, R., Martín de Santa Olalla, F., Fabeiro, C., Moratalla, A., 2006. Testing evapotranspiration equations using lysimeter observations in a semiarid climate. Agr. Water Manag. 85, 15–26. https://doi.org/10.1016/j.agwat.2006.03.014.

Maestre-Valero, J.F., Martínez-Álvarez, V., González-Real, M.M., 2013. Regionalization of the Hargreaves coefficient to estimate long-term reference evapotranspiration series in SE Spain. Span. J. Agric. Res. 11 (4), 1137–1152. https://doi.org/10.5424/sjar/2013114-3869.

Martí, P., González-Altozano, P., López-Urrea, R., Mancha, L.A., Shiri, J., 2015a. Modeling reference evapotranspiration with calculated targets. Assessment and Implications. Agric. Water Manag. 149 (2)), 81–90. https://doi.org/10.1016/j.agwat.2014.10.028.

Martí, P., Royuela, A., González-Altozano, P., 2015b. Discussion of "Assessment of Reference Evapotranspiration by the Hargreaves Method in the Bekaa Valley, Lebanon". J. Irrig. Drain. Eng. 141 (6), 07014053 https://doi.org/10.1061/(ASCE)IR.1943-4774.0000739.

Martí, P., Zarzo, M., Vanderlinden, K., Girona, J., 2015c. Parametric expressions for the adjusted Hargreaves coefficient in Eastern Spain. J. Hydrol. 529, 1713–1724. https://doi.org/10.1016/j.jhydrol.2015.07.054.

Martínez, C.J., Thepadia, M., 2010. Estimating reference evapotranspiration with minimum data in Florida. J. Irrig. Drain. Eng. 136 (7), 494–501. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000214.

Martínez-Cob, A., Tejero-Juste, M., 2004. A wind-based qualitative calibration of the Hargreaves ET_0 estimation equation in semiarid regions. Agric. Water Manag. 64 (3), 251–264. https://doi.org/10.1016/S0378-3774(03)00199-9.

Mendicino, G., Senatore, A., 2013. Regionalization of the Hargreaves coefficient for the assessment of distributed reference evapotranspiration in Southern Italy. J. Irrig. Drain. Eng. 139 (5), 349–362. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000547.

Nandagiri, L., Kovoor, G.M., 2006. Performance evaluation of reference evapotranspiration equations across a range of Indian climates. J. Irrig. Drain. Eng. 132 (3), 238–249. https://doi.org/10.1061/(ASCE)0733-9437(2006)132:3(238).

Ortega Farias, S., Irmank, E., Cuenca, R.H., 2009. Special issue on evapotranspiration measurement and modeling. Irrig. Sci. 28 (1), 1–3. https://doi.org/10.1007/s00271-009-0184-x.

Pandey, P.K., Pandey, V., 2016. Evaluation of temperature-based Penman–Monteith (TPM) model under the humid environment. Model. Earth Syst. Environ. 2, 152. https://doi.org/10.1007/s40808-016-0204-9.

Paredes, P., Fontes, J.C., Azevedo, E.B., Pereira, L.S., 2018b. Daily reference crop evapotranspiration in the humid environments of Azores islands using reduced data sets. Accuracy of FAO-PM temperature and Hargreaves-Samani methods. Theor. Appl. Clim. 134, 595–611. https://doi.org/10.1007/s00704-017-2295-2.

Paredes, P., Fontes, J.C., Azevedo, E.B., Pereira, L.S., 2018a. Daily reference crop evapotranspiration with reduced data sets in the humid environments of Azores islands using estimates of actual vapour pressure, solar radiation and wind speed. Theor. Appl. Climatol. 134, 1115–1133. https://doi.org/10.1007/s00704-017-2329-9.

Paredes, P., Pereira, L.S., 2019. Computing FAO56 reference grass evapotranspiration PM-ETo from temperature with focus on solar radiation. Agric. Water Manag. 215, 86–102. https://doi.org/10.1016/j.agwat.2018.12.014.

Paredes, P., Pereira, L.S., Almorox, J., Darouich, H., 2020. Reference grass evapotranspiration with reduced data sets: Parameterization of the FAO Penman Monteith temperature approach and the Hargreaves-Samani equation using local climatic variables. Agric. Water Manag. 240, 106210 https://doi.org/10.1016/j.agwat.2020.106210.

Pereira, L.S., Allen, R.G., Smith, M., Raes, D., 2015. Crop evapotranspiration estimation with FAO56: past and future. Agric. Water Manag. 147, 4–20. https://doi.org/10.1016/j.agwat.2014.07.031.

Pereira, L.S., Paredes, P., Hunsaker, D.J., López-Urrea, R., Jovanovic, N., 2021. Updates and advances to the FAO56 crop water requirements method. Agric. Water Manag. 248, 106697 https://doi.org/10.1016/j.agwat.2020.106697.

Popova, Z., Kercheva, M., Pereira, L.S., 2006. Validation of the FAO methodology for computing ETo with limited data. Application to South Bulgaria. Irrig. Drain. 55, 201–215. https://doi.org/10.1002/ird.228.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large scale parameters monitoring. Weather Rev. 100 (2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.

Ravazzani, G., Corbari, C., Morella, S., Gianoli, P., Mancini, M., 2012. Modified Hargreaves-Samani equation for the assessment of reference evapotranspiration in Alpine River Basins. J. Irrig. Drain. Eng. 138 (7), 592–599. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000453.

Raziei, T., Pereira, L.S., 2013. Estimation of ETo with Hargreaves-Samani and FAO-PM temperature methods for a wide range of climates in Iran. Agric. Water Manag. 121 (4)), 1–18. https://doi.org/10.1016/j.agwat.2012.12.019.

Ren, X., Qu, Z., Martins, D.S., Paredes, P., Pereira, L.S., 2016. Daily reference evapotranspiration for hyper-arid to moist sub-humid climates in Inner Mongolia, china: I. Assessing temperature methods and spatial variability. Water Resour. Manag. 30, 3769–3791. https://doi.org/10.1007/s11269-016-1384-9.

Samani, Z., 2000. Estimating solar radiation and evapotranspiration using minimum climatological data. J. Irrig. Drain. Eng. 126 (4), 265–267. https://doi.org/10.1061/(ASCE)0733-9437(2000)126:4(265).

Samani, Z., 2004. Discussion of 'History and evaluation of Hargreaves evapotranspiration equation.' by G.H. Hargreaves, and R.G. Allen. J. Irrig. Drain. Eng. 130 (5), 447–448 https://doi.org/10.1061/(ASCE)0733-9437(2004)130:5(447.2).

Samani, Z.A., Pessarakli, M., 1986. Estimating potential crop evapotranspiration with minimum data in Arizona. Trans. ASAE 29, 522–524. https://doi.org/10.13031/2013.30184@1986.

Schultz, B., Thatte, C.D., Labhsetwar, V.K., 2005. Irrigation and drainage. Main contributors to global food production. Irrig. Drain. 54 (3), 263–278. https://doi.org/10.1002/ird.170.

Senatore, A., Mendicino, G., Cammarelli, C., Ciraolo, G., 2015. Regional-Scale modeling of reference evaporanspiration: intercomparison of two simplified temperature- and radiation-based approaches. J. Irrig. Drain. Eng. 141 (12), 04015022 https://doi.org/10.1061/(ASCE)IR.1943-4774.0000917.

Senatore, A., Parrello, C., Almorox, J., C., Mendicino, G., 2020. Exploring the potential of temperature-based methods for regionalization of daily reference evaporanspiration in two Spanish regions. J. Irrig. Drain. Eng. 146 (3), 05020001 https://doi.org/10.1061/(ASCE)IR.1943-4774.0001447.

Sentelhas, P.C., Gillespie, J., Santos, E.A., 2010. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. Agr. Water Manag. 97 (5), 635–644. https://doi.org/10.1016/j.agwat.2009.12.001.

Shiri, J., Sadraddini, A.A., Nazemi, A.H., Martí, P., Fakheri Fard, A., Kisi, O., Landeras, G., 2015. Independent testing for assessing the calibration of the Hargreaves–Samani equation: new heuristic alternatives for Iran. Comput. Electron. Agric. 117 (9)), 70–80. https://doi.org/10.1016/j.compag.2015.07.010.

Shuttleworth, W.J., 1993. In: Maidment, D.R. (Ed.), Evaporation. Handbook of hydrology.. Mc Graw-Hill, New York.

Singh, L.K., Jha, M.K., Pandey, M., 2018. Framework for standardizing less data-intensive methods of reference evapotranspiration estimation. Water Resour. Manag. 32, 4159–4175. https://doi.org/10.1007/s11269-018-2022-5.

Smith, M., 1993. CLIMWAT for CROPWAT: a Climatic Database for Irrigation Planning and Management. FAO Irrigation and Drainage Paper, 49, Rome.

Tabari, H., 2010. Evaluation of reference crop evapotranspiration equations in various climates. Water Resour. Manag. 24 (10), 2311–2337. https://doi.org/10.1007/s11269-009-9553-8.

Tabari, H., Hosseinzadeh Talaee, P., 2011. Local Calibration of the Hargreaves and Priestley-Taylor Equations for Estimating Reference Evapotranspiration in Arid and Cold Climates of Iran Based on the Penman-Monteith Model. J. Hydrol. Eng. 16 (10), 837–845. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000366.

Temesgen, B., Eching, S., Davidoff, B., Frame, K., 2005. Comparison of some reference evapotranspiration equations for California. J. Irrig. Drain. Eng. 131 (1), 73–84. https://doi.org/10.1061/(ASCE)0733-9437(2005)131:1(73).

Thepadia, M., Martínez, C.J., 2012. Regional calibration of solar radiation and reference evapotranspiration estimates with minimal data in Florida. J. Irrig. Drain. Eng. 138 (2), 111–119. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000394.

Thornton, P.E., Running, S.W., 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agric. For. Meteorol. 93 (4), 211–228. https://doi.org/10.1016/S0168-1923(98)00126-9.

Trajkovic, S., 2007. Hargreaves versus Penman-Monteith under humid conditions. J. Irrig. Drain. Eng. 133 (1), 38–42. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:1(38).

Jensen, M.E., Allen, R.G. (Eds.), 2016. Evaporation, evapotranspiration, and irrigation water requirements, 2nd ed. ASCE Manual 70. ASCE, Reston, VI, p. 744.

Valiantzas, J.D., 2012. Discussion of "Case study on the accuracy and cost/efectiveness in simulating reference evapotranspiration in West-Central Florida" by Exner-Kittridge, M. G., Rains. M. C. J. Hydrol. Eng. 17 (1), 224–225. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000394.

Valiantzas, J.D., 2013. Simple ET0 forms of Penman's equation without wind and/or humidity data II: comparisons with reduced set-FAO and other methodologies. J. Irrig. Drain. Eng. 139 (1), 9–19. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000502.

Valiantzas, J.D., 2018. Modification of the Hargreaves–Samani model for estimating solar radiation from temperature and humidity data. J. Irrig. Drain. Eng. 144 (1), 06017014 https://doi.org/10.1061/(ASCE)IR.1943-4774.0001275.

Vanderlinden, K., Giraldez, J.V., Van Meirvenne, M., 2004. Assessing reference evapotranspiration by the Hargreaves method in southern Spain. J. Irrig. Drain. Eng. ASCE 130 (3), 184–191. https://doi.org/10.1061/(ASCE)0733-9437(2004)130:3(184).

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteor. Soc. 63 (11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.

Xu, C.Y., Singh, V.P., 2002. Cross comparison of empirical equations for calculating potential evapotranspiration with data from Switzerland. Water Resour. Manag. 16 (3), 197–219. https://doi.org/10.1023/A:1020282515975.

Yrisarry, J.J.B., Naveso, F.S., 2000. Use of weighing lysimeter and Bowen-Ratio Energy-Balance for reference and actual crop evapotranspiration measurements. Acta Hortic. (537), 143–150. https://doi.org/10.17660/ActaHortic.2000.537.14.