Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system

A. Pelosi[1,2], H. Medina[3], P.Villani[1,2], G. D'Urso[4] and G.B. Chirico[4]

[1]Department of Civil Engineering (DICIV) - University of Salerno - Italy

[2]C.U.G.RI. - University of Salerno - Italy

[3]Department of Basic Sciences, Agrarian University of Havana, Cuba

[4]Department of Agricultural Sciences, University of Naples Federico II - Italy

Corresponding authors: A. Pelosi – apelosi@unisa.it, G.B. Chirico - gchirico@unina.it

Highlights

- Reference evapotranspiration (ET_0) forecasted with a limited area ensemble prediction system
- Forecast performances were evaluated with both deterministic and probabilistic indices
- Forecasts were robust and reliable for lead times up to 5 days.

Abstract

The increasing availability of operational limited-area ensemble prediction systems (LEPS) opens up new opportunities for the application of weather forecasts in agriculture and water resource management. This study aims to evaluate the performances of probabilistic daily reference crop evapotranspiration (ET_0) forecasts with lead times up to 5 days and a spatial resolution of 7 km, computed by using COSMO-LEPS outputs (provided by the European Consortium for small–scale modelling, COSMO), in a region of southern Italy known for its complex topography in proximity to the Mediterranean coastline. ET_0 was estimated by means of three different estimation methods, i.e. the Hargreaves-Samani (HS), Priestley-Taylor (PT) and FAO Penman-Monteith (PM) equations, in order to assess the size of the weather forecast errors with models of different accuracies. Forecasts were verified with ground-based data from 18 automatic weather stations, and for two irrigation seasons.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

Performances were assessed with both deterministic indices, including BIAS, RMSE, correlation coefficients and coefficients of variation of the 16-member ensemble forecasts, and probabilistic metrics, such as the Brier skill score, reliability diagrams and relative operating characteristic. ET_0 forecasts with PM equation were robust and reliable, with slight sensitivity to the forecast lead time. High performances were also achieved with HS and PT equations, except for locations close to the coastline, where large systematic errors affect the numerical weather forecasts.

Keywords

reference evapotranspiration, numerical weather prediction, limited area models, ensemble, probabilistic verification

1. Introduction

Predicting evapotranspiration is fundamental in hydrological applications addressing water resources and irrigation management issues. Evapotranspiration is often retrieved as a function of the daily reference crop evapotranspiration (ET_0), which is evapotranspiration from a well-watered hypothetical reference crop. An internationally recognized standard method for computing ET_0 is the FAO-56 Penman-Monteith ($ET_{0\text{-}PM}$) equation (Allen et al., 1998). $ET_{0\text{-}PM}$ is considered the best method for estimating daily ET_0 in all climates, because the FAO-56 Penman-Monteith (PM) equation follows a physically based approach incorporating both physiological and aerodynamic parameters and thus does not require any local calibration (e.g., Garcia et al., 2004). $ET_{0\text{-}PM}$ entails the availability of a complete set of meteorological data, including air temperature, wind speed, solar radiation and relative humidity. These data are often unavailable in many regions of the world or are available with large uncertainty, since they are estimated by spatial interpolation of sparse meteorological ground stations. Other equations have been proposed for estimating ET_0 with a reduced number of meteorological data, but with additional empirical parameters that, where possible, are calibrated at local scale. Allen et al. (1998) proposed the Hargreaves-Samani (HS) equation for estimating ET_0 (hereinafter referred to as $ET_{0\text{-}HS}$) solely from temperature data (Hargreaves and Samani, 1985). The Priestley-Taylor (PT) equation (Priestley and Taylor, 1972) has also been suggested as a valid alternative for estimating ET_0 (hereinafter referred to as $ET_{0\text{-}PT}$) for locations where only temperature and radiation data are available (e.g. Pereira, 2004).

One practical aspect is that ET_0, whatever equation is used for computing it, is only a function of weather variables and thus ET_0 can be regarded as a diagnostic meteorological variable. Forecast performance of numerical weather prediction (NWP) models have considerably improved in the 21st century, making their output a valuable source for estimating ET_0 maps, alternative to the spatial interpolation of spatially coarse ground-based weather datasets (WMO, 2012).

Recent studies have focussed on assessing the performance of ET_0 estimates obtained with output data of regional weather models, also known as limited area models (LAM), which exploit the prediction of global circulation models (GCM) for identifying the initial and boundary conditions of a small region where the meteorological phenomena are explicitly resolved with finer spatial resolution. Nesting NWP models with finer scale into coarser models is equivalent to dynamically downscaling the output of the coarser model, consistently with the physical and empirical laws numerically resolved for describing the main meteorological phenomena.

Cai et al. (2007; 2009) employed weather forecast messages produced by the China Meteorological Administration for estimating daily $ET_{0\text{-}PM}$. Ishak et al. (2010) applied the regional model MM5, nested with ERA-40 reanalysis data provided by the European Centre for Medium-Range Weather Forecast (ECMWF) global model, and found that $ET_{0\text{-}PM}$ was overestimated by 27%-46%. Silva et al. (2010), also applying MM5 outputs, estimated daily $ET_{0\text{-}PM}$ in Central Chile with a root mean square error (RMSE) between 0.99 mmday^{-1} and 1.54 mmday^{-1}. They managed to reduce the RMSE by 10-20% after bias correcting raw NWP model outputs. Er-Raki et al. (2010), to overcome the scarcity of ground data in a semi-arid region of Central Morocco, employed the temperature fields produced by the ALADIN regional NWP model (nested with the ARPEGE global model) and, by applying an uncalibrated HS equation, estimated monthly ET_0 maps with an average RMSE of 16 mm. Srivastava et al. (2013) compared $ET_{0\text{-}PM}$ estimates in southeast England with weather data obtained by nesting the WRF regional NWP model with reanalysis data, respectively provided by ECMWF ERA-interim and the National Centers for Environmental Prediction (NCEP). The study suggested that $ET_{0\text{-}PM}$ estimates obtained by dynamically downscaling ECMWF reanalysis data outperform those obtained with NCEP reanalysis data.

Other recent studies evaluated the possibility to exploit operational numerical weather model outputs for real-time forecasting ET_0 in the short-medium range, i.e. with a lead time up to 1-2 weeks. Perera et al. (2014) applied output data provided by the ACCESS-G global model output operated by the Australian Bureau of Meteorology with a spatial resolution of 80 km, to estimate $ET_{0\text{-}PM}$ with lead times

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

up to nine days. The study showed good forecast performances with average RMSE less than 1 mm day^{-1} for lead time up to four days, after removing systematic bias of the numerical weather output data with respect to the ground weather stations.

In the last two decades, ensemble prediction systems (EPS) have become increasingly popular in operational decision-making processes. Unlike traditional deterministic forecasts where the numerical weather prediction model is run only once, in EPS the NWP model is run several times from very slightly different initial conditions and perturbed model parameters, to produce an ensemble of forecasts that are used to account for uncertainty in initial atmospheric conditions and NWP model errors (Buizza et al., 1999).

Tian and Martinez (2012a,b) employed Global Forecast System (GFS; Hamill et al., 2006) ensemble reanalysis data provided by NCEP to generate 1-15 day probabilistic ET_0 forecasts and then statistically downscale the forecasts by means of the analog approach (Hamill and Whitaker, 2006) in the southwestern United States. The GFS data set consisted of 15 members with a spatial resolution of about 200 km. Since the GFS dataset did not include all meteorological data required for estimating $ET_{0\text{-}PM}$, ET_0 forecasts were produced by using both the PM equation with alternative approximations of some of its main variables as well as the Thornthwaite equation (Thornthwaite, 1948). The statistical downscaling method was calibrated and verified with a set of $ET_{0\text{-}PM}$ produced with a 32 km grid reanalysis dataset provided by the North American Regional Reanalysis dataset (NARR; Mesinger et al., 2006). The results showed that most of the forecasts were skilful in the first five lead days.

Tian and Martinez (2014) replicated the experiment with a second GEFS reanalysis dataset, which was operationally available from 2012 (Hamill et al., 2013), with 11 ensemble members and a spatial resolution of 100 km. Tian and Martinez (2014), compared with the previous experiment (Tian and Martinez 2012a,b), managed to improve the skill of the probabilistic $ET_{0\text{-}PM}$ forecasts as well as the accuracy in estimating the soil water deficit for irrigation scheduling in the first five lead days, thanks to the availability of a complete meteorological dataset produced by a more advanced NWP model at higher spatial resolution.

Compared with the dynamic downscaling, statistical downscaling as the analog method has an advantage in requiring much less computational resources. However, simultaneous ground observations and forecast reanalysis data are required for a long period of time (e.g., about 25 years) in order to achieve a good calibration and verification of the statistical techniques. Such datasets are available with difficulty: indeed, Tian and Martinez (2012 a,b; 2014) resorted to model data generated at higher resolution as a surrogate for ground observations. No studies evidenced that statistical downscaling of forecasts performs better than dynamic downscaling of forecasts. Statistical downscaling is also exposed to limitations in tracking the effects of changing climatic conditions as well as weather conditions that are not represented by the sample data set employed for its calibration.

In recent years, limited area ensemble prediction systems (LEPS) have been developed as dynamic regional downscaling of global ensemble prediction systems. The development of operational LEPS was mainly motivated by the need to support decision makers with forecasts of high-impact weather events and particularly precipitation fields, at higher resolution and greater reliability than what could be achieved with single deterministic regional forecasts. The operational availability of LEPS opens up new opportunities for the application of weather forecasts in agriculture and water resource management, since high resolution probabilistic forecasting allows water irrigation managers to set-up agrometeorological advisory services based on a more reliable risk analysis.

One of the first examples is the limited-area ensemble prediction system, developed by the Consortium for small-scale modelling (COSMO-LEPS), which is now operationally used by several countries in Europe (Montani et al., 2011; Marsigli et al., 2014). COSMO-LEPS is nested on selected members of ECMWF EPS and is designed to combine the advantages of the probabilistic EPS approach with the high–resolution details gained in the mesoscale integrations (Montani et al., 2011).

This study aimed to evaluate the performance of probabilistic reference evapotranspiration forecasts based on numerical weather predictions produced by COSMO-LEPS. To our knowledge this is the first study explicitly examining the probabilistic performance of numerical weather predictions produced by dynamic downscaling of global ensemble forecasts for evapotranspiration studies.

The performance analysis focused on two irrigation seasons in southern Italy where a simultaneous set of meteorological data from 18 ground automatic weather stations and COSMO-LEPS forecasts was collected within a research programme to develop an advanced irrigation advisory service (Vuolo et al., 2015). ET_0 forecasts with lead times up to five days were computed with the PM equation and with uncalibrated HS and PT equations. The forecast performances are presented and discussed herein, using both deterministic and probabilistic indices.

2. Data

2.1. Study area

The study area was the region of Campania, about 14000 km^2 of land in southern Italy, between the Tyrrhenian Sea and the Southern Apennines (Figure 1). Weather forecasting is a challenging task in this region, as in other coastal regions of the central Mediterranean basin, where weather patterns are strongly influenced by the complex topography close to the coastline (e.g. Buzzi et al., 1994).

Under the Köppen-Geiger climate classification, most of the region is characterised by dry-summer subtropical climates, which are often described as being a “Mediterranean climate”. The coastal zone presents warm summers, while the adjacent inland zones are subject to hot summers. The eastern border zone of the region, close to the Apennines range, has a continental climate, as frequently occurs at higher elevations adjacent to areas with a Mediterranean climate (Peel et al., 2007).

The mean monthly temperature ranges from 25°C to 30°C in summer and between 11°C and 17°C in winter. The mean annual precipitation ranges from 800 to 1100 mm: the coastal and central mountainous areas have higher precipitation than the north-eastern side of the region. The maximum monthly precipitation values are recorded during November and December, the minimum during July and August.

Field irrigation starts no earlier than April and lasts till the end of September, although the actual time span of the irrigation season is influenced by the weather fluctuations and specific agricultural practices.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

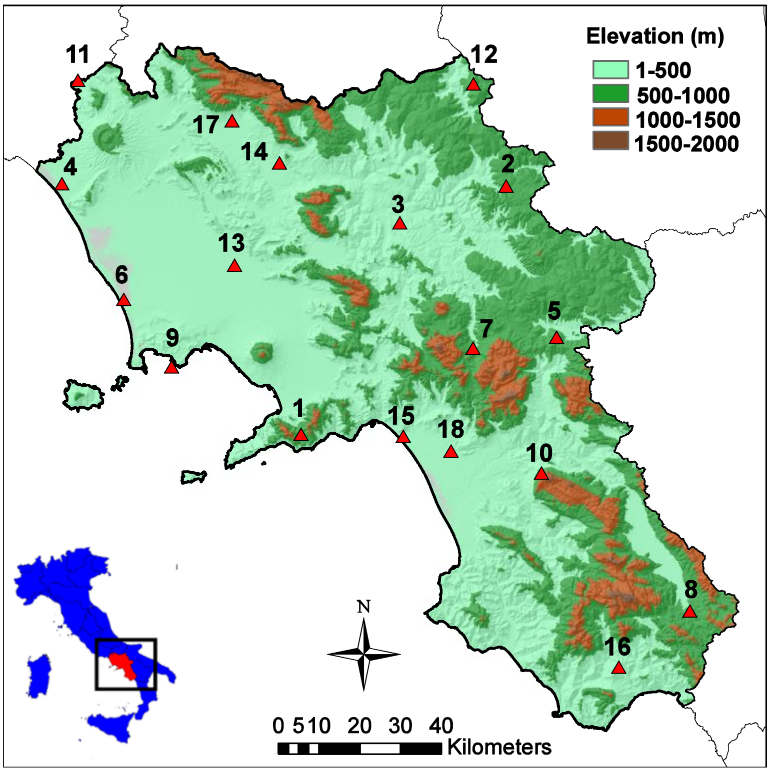

Figure 1. Relief map of the study area along with AWS stations.

2.2. Meteorological data

2.2.1. Observed ground-based data

Meteorological data from 18 ground-based automatic weather stations (AWS) distributed across the region were collected (Figure 1). These stations are part of the reference weather monitoring network of the Regional Meteorological Service. The AWS network complies with the EUMETNET technical specifications (De Leonibus and Vecchi, 1999): each station is equipped with redundant sensors to provide measurements with high accuracy and precision standards.

Table 1 reports a complete list of the AWSs along with their coordinates and elevations, ranging from 1 m a.s.l. to 848 m a.s.l. The AWS sites were chosen to achieve a good representation of the climatic variability within the region, including the coastal areas, the central hilly areas on the west side of the Apennines, as well as the inland side of the region.

No.	Name	Elevation [m]	Latitude [°]	Longitude [°]
1	Agerola METEO	848	40° 38’ 49’’	14° 32’ 28’’
2	Ariano Irpino METEO	631	41° 11’ 49’’	15° 8’ 10’’
3	Benevento METEO	236	41° 6’ 54’’	14° 49’ 30’’
4	Cellole METEO	9	41° 11’ 46’’	13° 50’ 17’’
5	Conza della Campania METEO	770	40° 51’ 43’’	15° 16’ 55’’
6	Lago Patria METEO	1	40 56’ 31’’	14° 1’ 19’’
7	Montella METEO	515	40° 50’ 17’’	15° 2’ 20’’
8	Montesano Marcellana METEO	552	40° 15’ 22’’	15° 39’ 50’’
9	Nisida METEO	88	40° 47’ 38’’	14° 9’ 50’’
10	Postiglione METEO	660	40° 33’ 43’’	15° 14’ 13’’
11	Rocca d'Evandro METEO	62	41° 25’ 30’’	13° 52’ 48’’
12	S.Bartolomeo METEO	750	41° 25’ 19’’	15° 2’ 28’’
13	San Marco Evangelista METEO	31	41° 1’ 12’’	14° 20’ 38’’
14	S.Salvatore Telesino METEO	167	41° 14’ 49’’	14° 28’ 23’’
15	Salerno METEO	13	40° 38’ 38’’	14° 50’ 13’’
16	Torre Orsaia METEO	413	40° 7’ 55’’	15° 27’ 32’’
17	Alife	117	41° 20’ 20’’	14° 20’ 2’’
18	Battipaglia	64	40° 36’ 40’’	14° 58’ 34’’

Table 1. List of Automatic Weather Stations

These AWSs have been operating since 2007. The following data recorded from April to September were considered for the forecast verifications: air temperature and humidity at 2 m; global incoming solar radiation; wind speed at 10 m; barometric pressure. Performance analyses focused on irrigation seasons for the years 2013 and 2014 which did not experience extreme weather conditions. Table 2 summarises some average statistics of the observed data from April to September for 2013 and 2014.

	Min	Max	Mean	Standard deviation
T [°C]	-0.6	40.5	19.9	4.3
RS [W m^{-2}]	20.2	403.5	243.3	71.4
WS [m s^{-1}]	0.3	13.2	2.4	0.9
RH [%]	28.3	100	75.5	10.7

Table 2. Statistics of the weather variable datasets over the region based on data collected during two irrigation seasons (2013 and 2014)

2.2.2. NWP forecast data

The numerical weather prediction outputs used for forecasting daily ET_0 are those given by COSMO-LEPS, which is a limited area ensemble prediction system, implemented by the HydroMeteoClimate Regional Service of Emilia-Romagna, located in Bologna, Italy (ARPA–SIMC). COSMO-LEPS was developed within the Consortium for small–scale modelling (COSMO), whose associates are Germany, Greece, Italy, Poland, Romania and Switzerland. It was the first mesoscale ensemble application to be run on a daily basis in Europe. COSMO-LEPS is based on 16 integrations of the non–hydrostatic mesoscale model COSMO, and combines the advantages of the probabilistic approach by global ensemble systems with the high–resolution details gained in the mesoscale integrations (Montani et al., 2011). The current model configuration has been in operation since 2009. Since December 2011, COSMO-LEPS has run twice a day, at 00:00 UTC and 12:00 UTC. The model has a forecast range of 132 hours, with data available at three-hour intervals, and a spatial resolution of 7.5 km. The locations of the COSMO-LEPS grid points overlaid with the reference AWS sites are shown in Figure 2.

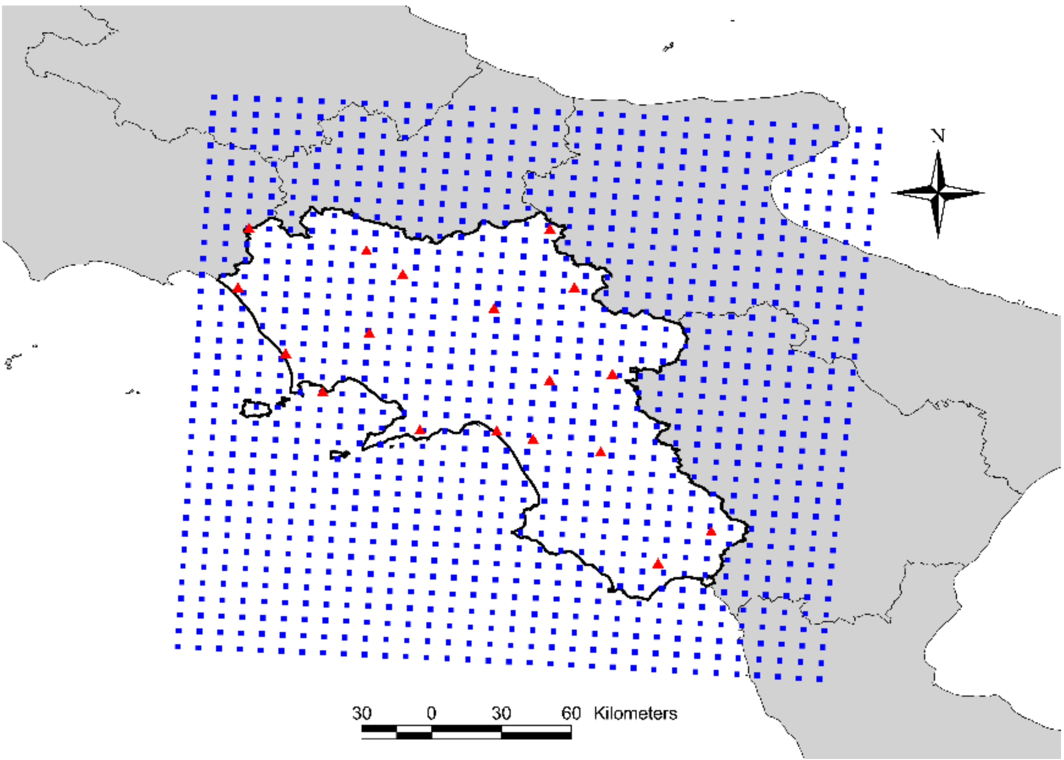

Figure 2. Location of the AWS stations (red triangles) and COSMO-LEPS grid points (blue squares)

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

In this study, the relevant weather variables to calculate ET_0 were extracted from grib files released as output of the 00:00 UTC run: atmospheric pressure reduced to mean sea level, net short wave radiation, albedo, wind speed at 10 m, temperature and relative humidity at 2 m.

The forecast dataset consists of variable ensemble output produced by the operational chain of the COSMO-LEPS from April 1st to September 30th in 2013 and 2014, for a total of 366 days, with lead times from one day to five days.

3. Methods

3.1. Computation of the daily reference evapotranspiration ET_0

The PM equation is that recommended by the Food and Agriculture Organization (FAO), in Paper No. 56, as the standard method for computing reference evapotranspiration ET_0. It applies the energy balance and mass transfer principles to estimate evapotranspiration from a uniform grass reference surface. Specific parameters are employed to model the surface and aerodynamic resistance from the vegetation (Allen et al., 1998). The PM equation is expressed as follows:

$$ET_{0-PM} = \frac{0.408\Delta\left(R_n - G\right) + \gamma \frac{900}{T+273} u_2 \left(e_s - e_a\right)}{\Delta + \gamma\left(1 + 0.34 u_2\right)} \quad \text{(Eq. 1)}$$

where ET_{0-PM} is the daily reference evapotranspiration in (mm day^{-1}), 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 is the daily mean air temperature at 2 m height (°C), u_2 is the wind speed at 2 m height above ground (m s^{-1}), e_s is the saturation vapour pressure (kPa), e_a is the actual vapour pressure (kPa), Δ is the slope of the vapour pressure curve (kPa °C^{-1}) and γ is the psychometric constant (kPa °C^{-1}).

The net radiation (R_n) was calculated as the difference between the incoming net shortwave radiation and the outgoing net longwave radiation. As suggested by Allen (1998) for the reference crop, the incoming net shortwave radiation was calculated by coupling the measured or predicted incoming shortwave solar radiation with an albedo of 0.23. The outgoing net longwave radiation was estimated

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

from the daily maximum and minimum air temperature and relative shortwave radiation, which is computed as the ratio of the incoming shortwave solar radiation and the clear-sky radiation. The soil heat flux density (G) is computed as a fraction of R_n as suggested by Allen et al. (1988) for the reference crop.

Daily mean air temperature was computed as the average of daily maximum and minimum air temperature, instead of computing it by averaging the data at the lowest available time-resolution, which would lead to underestimating the daily $ET_{0\text{-}PM}$, as a result of the non-linear relationship between the saturation vapour pressure and temperature (Allen et al., 1998).

Daily wind speed was computed as the average of the predicted or observed wind data at the highest available temporal resolution. Wind speed (u_2) values, both forecasted and measured at 10 m height above ground, were adjusted at 2 m above ground by employing the logarithmic equation of the wind speed profile suggested by Allen (1998).The actual vapour pressure was computed as a function of the mean air relative humidity.

The PM equation implies the availability of a complete weather dataset, which is normally feasible in a limited numbers of locations. This was one of the main motivations of previous studies exploring the applicability of numerical weather prediction outputs as a proxy of ground weather data, as mentioned in the introduction (e.g. Cai et al., 2007; 2009; Ishak et al., 2010; Silva et al., 2010; Er-Raki et al., 2010; Srivastava et al., 2013). Related to this, another aspect that is worth taking into consideration is that all forecasted weather variables involved in the $ET_{0\text{-}PM}$ estimation are affected by forecast errors, which all contribute to downgrade the $ET_{0\text{-}PM}$ forecasts (Perera et al., 2014). Thus, in this study we also evaluated simpler and uncalibrated methods for estimating reference evapotranspiration, which employs a number of uncertain weather forecast variables smaller than those required for computing $ET_{0\text{-}PM}$. The motivation for assessing the forecast performances with different evapotranspiration methods arises from the purpose of investigating how the uncertainty associated with the input weather variables propagates into the estimated ET_0. Since each forecasted weather variable brings its own uncertainty into the ET_0 equation, we sought to assess to what extent the application of equations based on a reduced number of weather variables for computing ET_0 could compensate the effect of the reduced accuracy deriving from simpler uncalibrated ET_0 estimation methods (Droogers and Allen, 2002; Cai et al., 2007; Bormann, 2011).

In this study we considered a temperature-based model as the HS equation (Hargreaves and Samani, 1982) and a radiation-based model as the PT equation (Priestly and Taylor, 1972). The HS equation is that suggested by Allen et al. (1998) in the FAO guidelines for estimating ET_0, when only temperature data are available, and is given as:

$$ET_{0-HS} = K_{HS}\left(T + 17.8\right)\sqrt{T_{max} - T_{min}}\left(0.408 R_a\right) \qquad \text{(Eq. 2)}$$

where $ET_{0\text{-}HS}$ is the daily reference evapotranspiration in (mm day^{-1}), R_a is the extraterrestrial radiation (MJ m^{-2} day^{-1}), T_{max} and T_{min} are respectively the daily maximum and minimum temperature (°C), and K_{HS} is an empirical coefficient, assumed to be equal to 0.0023 as suggested by Allen et al. (1998). The formula only needs temperature data, since the extraterrestrial radiation is a function of latitude and time of year. The HS equation has been widely used thanks to its simplicity and acceptable results. The term ($T_{max} - T_{min}$) indirectly estimates the effect of the daily radiation, as it is related to humidity and cloudiness (e.g. Shahidian et al., 2012).

Finally, the PT equation was considered:

$$ET_{0-PT} = \alpha\,\frac{0.408\Delta\left(R_n - G\right)}{\Delta + \gamma} + \beta \qquad \text{(Eq. 3)}$$

where $ET_{0\text{-}PT}$ is the daily reference evapotranspiration in (mm day^{-1}), and α and β are empirical coefficients. Wind speed and relative humidity data are not needed, since potential evapotranspiration is estimated in terms of energy fluxes without an aerodynamic component. Parameters α and β are assumed to be equal to 1.26 and 0, respectively, as found by the authors for “advection-free” saturated surfaces and theoretically explained by Lhomme (1997). $ET_{0\text{-}PT}$ mainly depends on solar radiation, but temperature data are also needed for computing R_n, G and Δ.

The empirical parameters of (Eq. 2) and (Eq. 3) can also be specifically calibrated, accounting for the local weather and terrain characteristics, as done in previous studies (e.g. Xu and Singh, 2000; Er

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

Raki et al., 2010; Shahidian et al., 2012). In this study, we used the values recommended for the most general case since, as explained above, our interest was to evaluate the relative impact of the weather forecast uncertainty on the estimated ET_0 values with methods of different levels of accuracy, without any preliminary bias correction.

3.2. Assessment of forecast performances

The COSMO-LEPS forecasted meteorological outputs and ET_0 estimated using the outputs in question were verified with the corresponding ground-based observations. The PM equation was used to compute the reference evapotranspiration with ground-based data (hereinafter referred to as $ET_{0g\text{-}PM}$). $ET_{0g\text{-}PM}$ are hereinafter also denoted as "observed ET_0" and are taken as benchmark values to evaluate the performances of the daily ET_0 forecasts. $ET_{0g\text{-}PM}$ values were compared with those forecasted with lead times from one to five days, respectively computed with the PM, HS and PT equations (Eqs. 1, 2 and 3).

From an operational perspective, two alternative interpolation strategies for estimating ET_0 forecasts at the AWS nodes can be followed: i) interpolating forecasted weather data prior to computing ET_0 values at each node; ii) interpolating ET_0 values computed at the COSMO-LEPS grid nodes. These strategies can lead to different results, since the ET_0 equations employed are non-linear. We preferred the first strategy as we suppose that this better preserves the spatial structure of the weather input variables and their cross-correlation (Van Schaeybroeck and Vannitsem, 2015). A triangle-based bi-linear interpolation method was employed, which consists in interpolating the three grid points closest to the examined site.

The forecast performances were assessed with both deterministic and probabilistic metrics. Deterministic metrics are well-suited for single-valued forecast verifications. Probabilistic metrics are used to verify the forecast probabilities (given by the forecast ensembles) with the observed frequencies. In the case of probabilistic forecasts, deterministic metrics of forecast performance cannot provide a comprehensive assessment of the forecast quality, which instead can be evaluated only through estimation of the joint distribution of forecasts and observations (Wilks, 2011).

3.2.1. Deterministic metrics

Statistical performance indices are computed for all lead times by comparing the median value, $\tilde{P}_i$, of the ensemble of the predicted variables retrieved from the COSMO-LEPS forecasts on the generic *i*-th day, with the predicted variable retrieved from the ground-based weather stations, O_i. We chose the median value as representative of the ensemble forecasts, instead of the mean value, since outliers occasionally make the ensemble distribution strongly asymmetric and thus the mean results become biased.

The first index is the BIAS, which was used as an indicator of accuracy of the ET_0 forecasts:

$$\mathrm{BIAS} = \frac{\sum_{i=1}^{n}\left(\tilde{P}_i - O_i\right)}{n} \qquad \text{(Eq. 4)}$$

where *n* denotes the number of examined days, in this study equal to 366.

The second deterministic performance indicator is the root mean square error, RMSE, which gives insight into both accuracy and precision of the ET_0 forecasts:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\tilde{P}_i - O_i\right)^2}{n-1}} \qquad \text{(Eq. 5)}$$

The correlation coefficient, *R*, was used to measure the linear relationship between the forecasted and observed ET_0:

$$R = \frac{\mathrm{Cov}\left(O,\tilde{P}\right)}{\sqrt{\mathrm{Var}\left(O\right)\mathrm{Var}\left(\tilde{P}\right)}} \qquad \text{(Eq. 6)}$$

where $\mathrm{Cov}\left(O,\tilde{P}\right)$ is the sample covariance between the ensemble forecast medians and their corresponding observed values, and $\mathrm{Var}\left(O\right)$ and $\mathrm{Var}\left(\tilde{P}\right)$ are respectively the sample variances of the observed and forecast medians.

As indicator of the prediction uncertainty due to the ensemble spread, we computed the coefficient of variation, CV, of the forecasted ET_0:

$$\mathrm{CV} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{\tilde{P}_i}\left[\sqrt{\frac{\sum_{j=1}^{m}\left(P_{i,j}-\bar{P}_i\right)^2}{m-1}}\right]\right\} \qquad \text{(Eq. 7)}$$

where m=16 is the number of members in each ensemble, $\bar{P}_i$ is the mean of the ensembles on the i-th day, and $P_{i,j}$ is the j-th member value on the i-th day.

Another deterministic index was employed to compare the ET_0 prediction BIAS due to the weather forecast errors with the prediction BIAS due to the simplification of the reference evapotranspiration estimation method, i.e. $ET_{0\text{-}HS}$ or $ET_{0\text{-}PT}$ as compared with $ET_{0\text{-}PM}$. Let $ET_{0g\text{-}HS,i}$ and $ET_{0g\text{-}PT,i}$ be the reference evapotranspiration estimated with HS and PT equations, respectively, using the data observed with the AWS on the i-th day as input weather variables. Let $ET_{0-HS,i}$ and $ET_{0-PT,i}$ be the corresponding medians of the forecasted values on the i-th day for a generic lead time. The following absolute relative bias indices are then computed:

$$\mathrm{rBIAS_{HS}} = \left|\frac{\sum_{i=1}^{n}\left(ET_{0-HS,i}-ET_{0g-HS,i}\right)}{\sum_{i=1}^{n}\left(ET_{0g-HS,i}-ET_{0g-PM,i}\right)}\right| \qquad \text{(Eq. 8a)}$$

$$\mathrm{rBIAS_{PT}} = \left|\frac{\sum_{i=1}^{n}\left(ET_{0-PT,i}-ET_{0g-PT,i}\right)}{\sum_{i=1}^{n}\left(ET_{0g-PT,i}-ET_{0g-PM,i}\right)}\right| \qquad \text{(Eq. 8b)}$$

The terms at the denominators of the above indices quantify the BIAS of the simplified uncalibrated ET_0 prediction method. The terms at the numerators quantify the BIAS due to the numerical weather forecast errors. The above indices are greater than one if the weather forecast errors dominate the ET_0 model error.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

3.2.2. Probabilistic metrics

For a generic AWS location, let $F_P(p)$ be the cumulative distribution function of the forecasts, given by the ensembles, and let t denote a selected threshold value (in the following, the median value of the observations). Similarly to RMSE in the deterministic case, the Brier score, BS, measures the mean squared probability error (Murphy, 1973) as follows:

$$\mathrm{BS}(t) = \frac{\sum_{i=1}^{n}\left(F_{P_i}(t) - 1\{O_i \leq t\}\right)^2}{n} \qquad \text{(Eq. 9)}$$

where $1\{\cdot\}$ is a step function that is equal to 1 if the condition $\{\}$ is met and zero otherwise. The Brier score ranges from 0 to 1. Values of BS equal to 0 indicate a perfect score.

The Brier skill score, BSS, measures the improvement of the probabilistic forecast relative to a reference forecast:

$$\mathrm{BSS}(t) = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{reference}}} \qquad \text{(Eq. 10)}$$

where $\mathrm{BS}_{\mathrm{reference}}$ is the Brier score of the reference method. In this study, we take as reference probabilistic forecast the one defined by the unconditional distribution of the observations, which is computed by the relative frequencies of the n observations O_i in the verification data set. This distribution is usually called the sample climatological distribution, or simply the *sample climatology* (Wilks, 2011). The Brier skill score ranges from -∞ to 1 and values of BSS equal to 1 indicate perfect skill.

In addition to BS and BSS, reliability and the relative operating characteristic (ROC) diagrams were computed to investigate the forecast quality. The reliability diagram plots the observed frequency of an event (defined by the threshold t) against its forecasted probability. The range of forecast probabilities is divided into k bins. Then, on the x-axis, we plot the average probability of the forecasts that falls in the k-th bin while, on the y-axis, the fraction of the corresponding observations that are below the threshold. Reliability is a measure of systematic and conditional bias. Perfect reliability is achieved along the 45° diagonal line on the reliability diagram when the observed frequency of the given event within each bin equals the average of the corresponding forecast probabilities. The deviation from the diagonal gives the conditional bias. A curve that lies above the diagonal line indicates under-forecasting: the forecasted probabilities related to a given event are too low if compared with the observed frequency of the event; vice versa, points below the diagonal line indicate over-forecasting. The flatter the curve in the reliability diagram, the less resolution it has. Resolution is the ability to distinguish one type of outcome from another. By definition, forecasts from the sample climatology have no resolution and this condition is shown, for comparison, on the reliability diagram by means of a horizontal line. On the same diagram, it is also possible to show the sharpness of the forecast, a measure of the forecast confidence, by means of a histogram representing the frequency of forecasts in each probability bin. A deterministic forecast is infinitely sharp while forecasts from sample climatology have no sharpness.

The ROC diagram is a discrimination-based forecast verification metric (Wilks, 2011), which measures the ability to discriminate between two possible outcomes, not sensitive to bias (i.e. reliability). ROC plots the probability of detection (hit rate), POD, of an event (defined by the threshold t) against the probability of false detection (false alarm), POFD, of the same event. The 45° diagonal line on the ROC diagram represents the line of no skill while high skill is achieved with a curve located in the upper left corner of the plot. The ROC is conditioned on the observations (i.e., given that an event occurred, it shows the corresponding forecast). It is therefore a good companion to the reliability diagram, which is conditioned on the forecasts (Wilks, 2011).

POD (y-axis) and POFD (x-axis) are calculated as follows, having chosen some probability thresholds p_t among the interval [0, 1]:

$$\mathrm{POD}(t, p_t) = \frac{\sum_{i=1}^{n} 1\left\{1 - F_{P_i}(t) > p_t \mid O_i > t\right\}}{\sum_{i=1}^{n} 1\left\{O_i > t\right\}} \qquad \text{(Eq. 11a)}$$

$$\mathrm{POFD}(t, p_t) = \frac{\sum_{i=1}^{n} 1\left\{1 - F_{P_i}(t) > p_t \mid O_i \le t\right\}}{\sum_{i=1}^{n} 1\left\{O_i \le t\right\}} \qquad \text{(Eq. 11b)}$$

4. Results and discussion

The forecast performance of daily ET_0 is obviously influenced by the forecast performance of the weather variables employed as input to the ET_0. Thus we first report the forecast performance of these variables and then those of the ET_0 predictions. The forecast performances of weather variables were evaluated with the deterministic metrics introduced in Section 3.2.1. We do not present any probabilistic metrics of the raw weather variables for the sake of conciseness. The performance of the ET_0 forecasts is instead illustrated with both deterministic and probabilistic metrics.

4.1. Forecast performances of weather variables with deterministic metrics

We verified the daily forecast, with lead time up to five days, in the irrigation seasons from April to September in two years (2013 and 2014), of the following weather variables: T (defined as the mean between the daily T_{max} and T_{min}), solar radiation (RS), wind speed (WS) and relative humidity (RH). We did not show any results concerning the atmospheric pressure since we found almost perfect agreement between observations and forecasts. Unlike previous studies (e.g. Perera et al., 2014), we did not apply any preliminary bias corrections to the weather forecast outputs, beside their bi-linear interpolation at the AWS sites, as outlined above.

The values of BIAS and RMSE for the 18 AWS sites and varying lead times are shown below in Figure 3 and Figure 4, respectively.

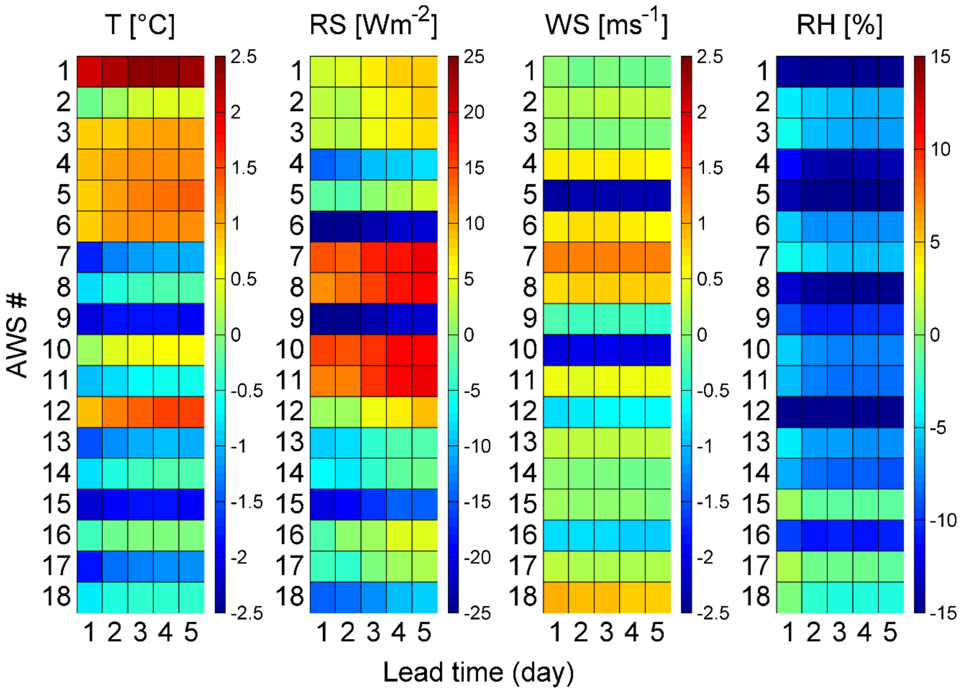

Figure 3. BIAS of forecasted vs. observed daily weather variables for all 18 AWS sites and lead times

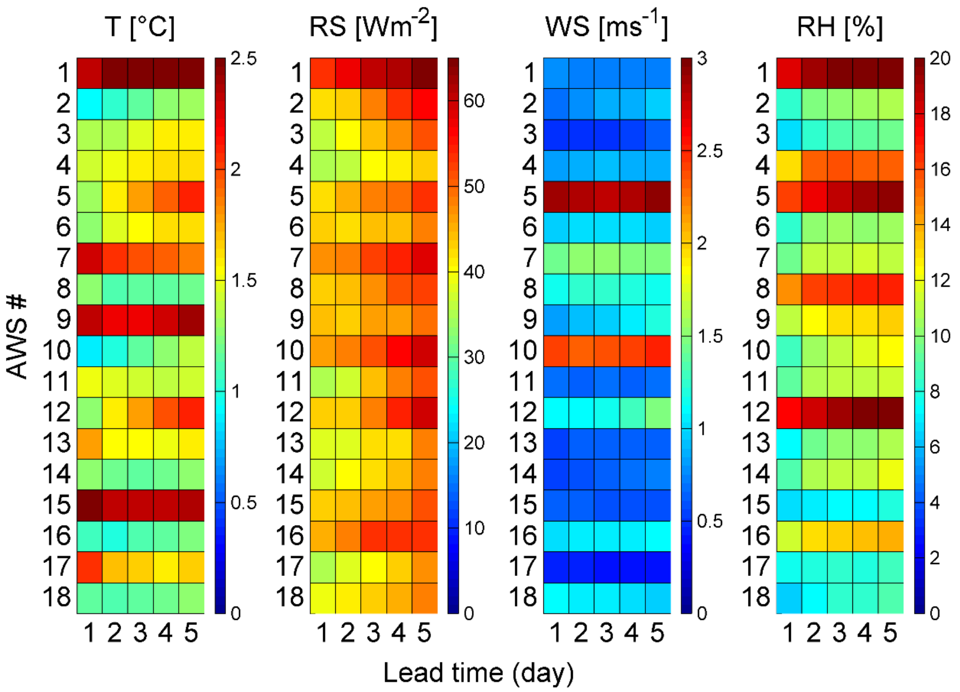

Figure 4. RMSE of forecasted vs. observed daily weather variables for all 18 AWS sites and lead times

As highlighted by Perera et al. (2014), the ground measurements errors may also contribute to reduce the forecast performances. However, we verified that for all weather variables the measurement errors due to the ground sensor accuracies were significantly less than the corresponding forecast uncertainties. This suggested that the effects of measurement errors on the forecast evaluations were negligible.

Air temperature

Air temperature is the only weather variable needed in the computation of $ET_{0\text{-}HS}$ but it is also needed for $ET_{0\text{-}PT}$ since T_{max} and T_{min} are required to compute the net long wave radiation and saturation vapour pressure. T_{max} and T_{min} are also required to calculate the vapour pressure deficit in Eq. (1). Here, for the sake of conciseness, we provide forecast performances only with reference to T, defined as the mean between daily maximum air temperature and daily minimum air temperature.

Figures 3-4 show that the forecast performances for T do not significantly decline with increasing lead time for all locations. Moreover, Figure 3 highlights that there is a broad variation of forecast performances among the AWS sites. The NWP model has no systematic tendency to overforecast or underforecast T. Rather, we found T is overforecasted in half the AWS sites and underforecasted in the other half. BIAS values range between -2.1 °C and 2.3 °C.

At AWSs 9 and 15, T is dramatically underpredicted. These two sites are close to the coastline (Figure 1), where the COSMO-LEPS, with the bilinear interpolation method adopted for estimating the weather forecasts at the AWS sites, is unable to resolve the local weather effects associated with the proximity to sea and thus the forecasts are subject to systematic biases. The higher overprediction of T is found in correspondence to AWS 1, which is located close to the coastline, like AWSs 9 and 15, but on a cliff at an elevation of 848 m. Here, the COSMO-LEPS model with the bilinear interpolation of the values forecasted at the numerical grid is unable to resolve the small scale variability due to steep elevation gradients close to the coastline.

The RMSE ranges between 0.9 °C and 2.7 °C, with an average value over the region of 1.6 °C. These RMSE values indicate very good performances compared with the results of other studies, where these RMSE values for *T* were achieved only after bias-correcting the NWP outputs (e.g. Silva et al., 2010).

Solar radiation

The forecast performances for daily incoming solar radiation, *RS*, show a clear decline with increasing lead time. The BIAS values range from -26.8 W m^{-2} to 19.3 W m^{-2}, the RMSE values range from 34.7 W m^{-2} to 64.1 W m^{-2}. The RMSE values at a 5-day lead time are 20% higher than the RMSE values at a lead time of 1 day. Strong negative BIAS values were observed at AWS sites close to the coastline (i.e., AWSs 6, 9 and 15), while strong positive BIAS values were found in inland areas (i.e. AWSs 7, 8, 10 and 11). In addition, high RMSE values were found at AWSs 1, 2, 12 and 16.

These forecast errors were probably produced by factors (such as the topography), which influence the local global incident radiation (direct and diffuse) and are not properly resolved by NWP model.

Wind speed

The forecast performances related to daily mean wind speed experience great spatial variability in the region of interest. For *WS*, the BIAS values range between -2.4 m s^{-1} and 1.2 m s^{-1}. The RMSE values goes from 0.4 m s^{-1} to 2.9 m s^{-1}. At AWS sites 5 and 10 we found the highest RMSE and BIAS in the region. In these sites, the highest *WS* values were observed during the year, enhanced by local terrain features not resolved by the COSMO-LEPS model, which tends to dramatically underforecast the wind speed. *WS* was instead overforecasted at AWSs 7 and 8, where local terrain features mitigate the wind speed with respect to the dominant wind patterns predicted for the surrounding area.

Air humidity

The air humidity is underforecasted at most of the AWS sites, except for three that are subject to slight overforecasting (i.e. AWSs 15, 17 and 18). The BIAS values range between -17.1 % and 1.1 %, the RMSE values range between 6.4 % and 20.5 %. The worst performances were collected at AWSs 1, 5 and 12.

4.2. Forecast performances of reference evapotranspiration with deterministic metrics

In the examined two irrigation seasons (2013 and 2014), the difference between the daily ET_0 calculated using weather forecasts and the daily $ET_{0\text{-}gPM}$ appears to be stochastically independent of the time of year. Thus we computed the performance indices aggregated for the entire irrigation seasons, avoiding performance assessment within smaller time-spans (e.g. monthly), as the differences were not significant.

The observed long–term monthly mean of $ET_{0g\text{-}PM}$ computed at the 18 AWS sites in Campania region from April to September is shown in Figure 5. The monthly mean $ET_{0g\text{-}PM}$ ranges from 2.3 mm day^{-1} to 5.5 mm day^{-1}. The highest values are reached in July, when the maximum daily air temperatures are registered. The interquartile spread among the examined AWS peaks in July and August, and reaches its lowest point in May. The maximum spread (occurring in July) of the monthly mean $ET_{0g\text{-}PM}$ among the AWS stations is about 1.2 mm day^{-1}.

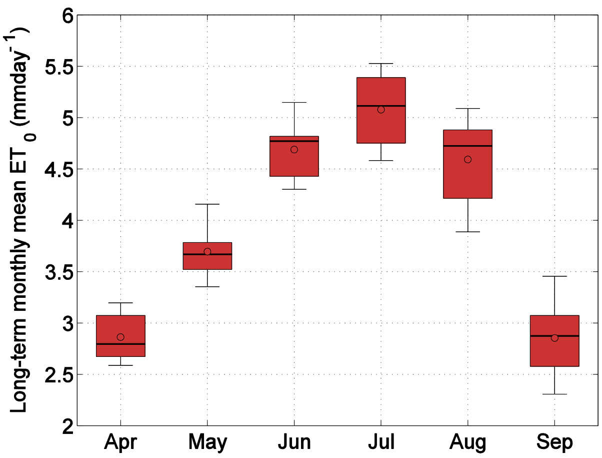

Figure 5. Observed long-term monthly mean ET_0 computed for the 18 AWS sites in Campania

Figures 6a-f show the scatter between the observed and the median of the forecasted daily ET_0 at all AWS sites for the two extreme lead times (i.e. 1 day on the left and 5 days on the right) and for the three different evapotranspiration equations. The linear trends between observed and forecasted change marginally with the increasing lead time, while the scatter increases markedly. Thus, the accuracy of the forecast appears to be more sensitive to the lead time than its precision.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

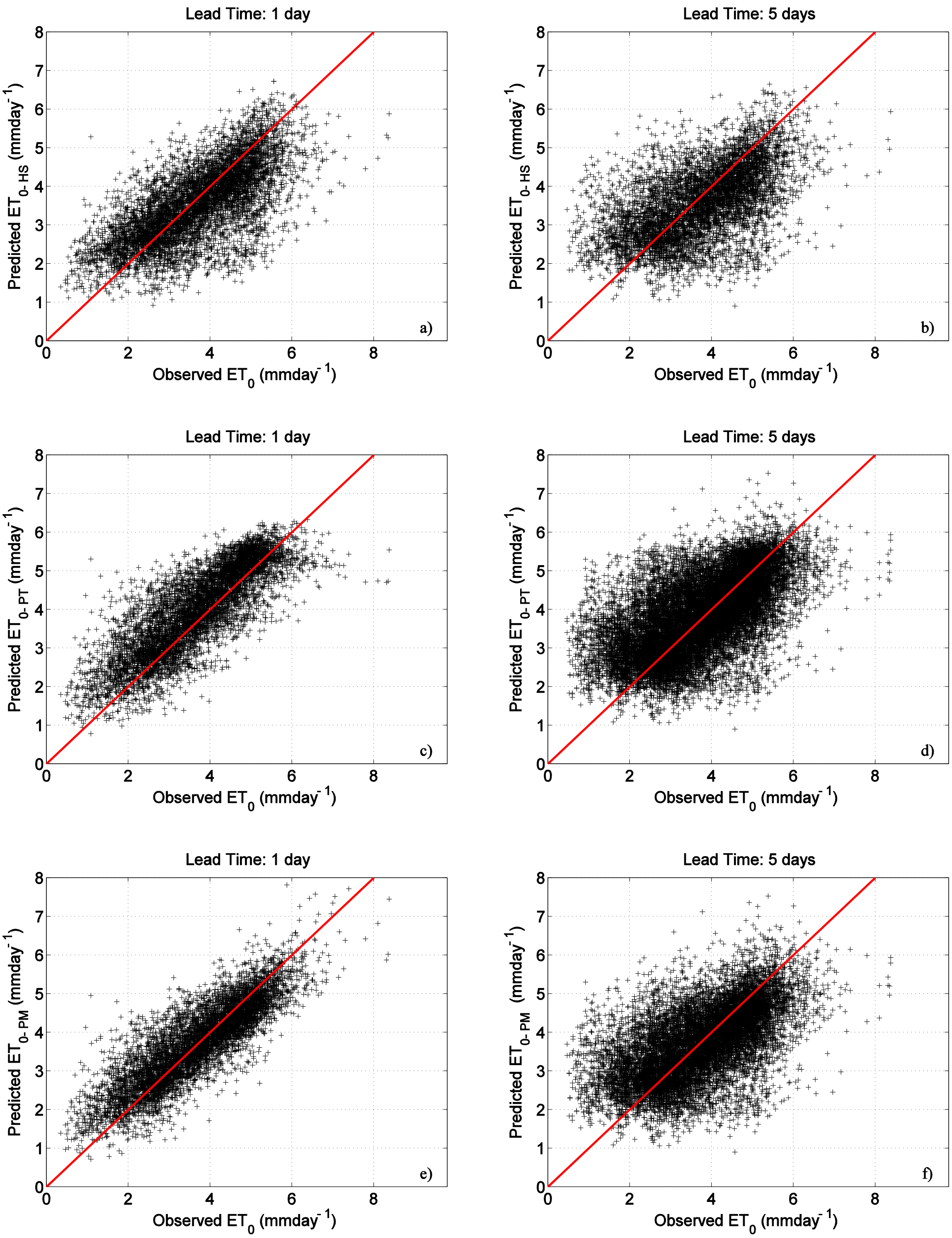

Figure 6. Daily observed $ET_{0\text{-}PM}$ vs daily predicted ET_0 at all AWS stations for 1-day and 5-day lead times and for different evapotranspiration models. The red line is the 45° line.

For low values of observed ET_0 (< 2 mm day^{-1}), the forecasts tend to overpredict ET_0 for all lead times and equations considered. The PT and PM methods overpredict ET_0, while HS exhibits some points with largely underpredicted ET_0. These points correspond to AWSs 6, 9 and 15, as clarified below.

Figures 7 and 8 show the BIAS and RMSE, respectively, for each AWS site (row) and lead time (column). The values of the BIAS range:

i) from -1.96 mm day^{-1} to 0.46 mm day^{-1} for the daily predicted $ET_{0\text{-}HS}$, with an average value over space and lead time of -0.16 mm day^{-1};

ii) from -0.35 mm day^{-1} to 0.76 mm day^{-1} for the daily predicted $ET_{0\text{-}PT}$, with an average value over space and lead time of 0.33 mm day^{-1};

iii) from -0.43 mm day^{-1} to 0.72 mm day^{-1} for the daily predicted $ET_{0\text{-}PM}$ with an average value over space and lead time of 0.123 mm day^{-1}.

The values of the RMSE range:

i) from 0.57 mm day^{-1} to 2.17 mm day^{-1} for the daily predicted $ET_{0\text{-}HS}$, with an average value over space and lead time equal to 0.90 mm day^{-1};

ii) from 0.55 mm day^{-1} to 1.24 mm day^{-1} for the daily predicted $ET_{0\text{-}PT}$, with an average value over space and lead time of 0.81 mm day^{-1};

iii) from 0.48 mm day^{-1} to 1.17 mm day^{-1} for the daily predicted $ET_{0\text{-}PM}$ with an average value over space and lead time of 0.71 mm day^{-1};

Figure 9 shows, at each AWS site and for each lead time, the value of $rBIAS_{HS}$ and $rBIAS_{PT}$ as in Eqs. (8a-b), to highlight the main source of error when simplified evapotranspiration methods (i.e. HS and PT) are used instead of the PM equation. Values of rBIAS greater than one suggest that the forecast error is greater and relatively large improvements can be achieved with the same ET_0 prediction method if the raw forecasts are post-processed for removing systematic prediction errors, which are mainly due to the limited capacity of the NWP model to resolve the effects of small scale variability.

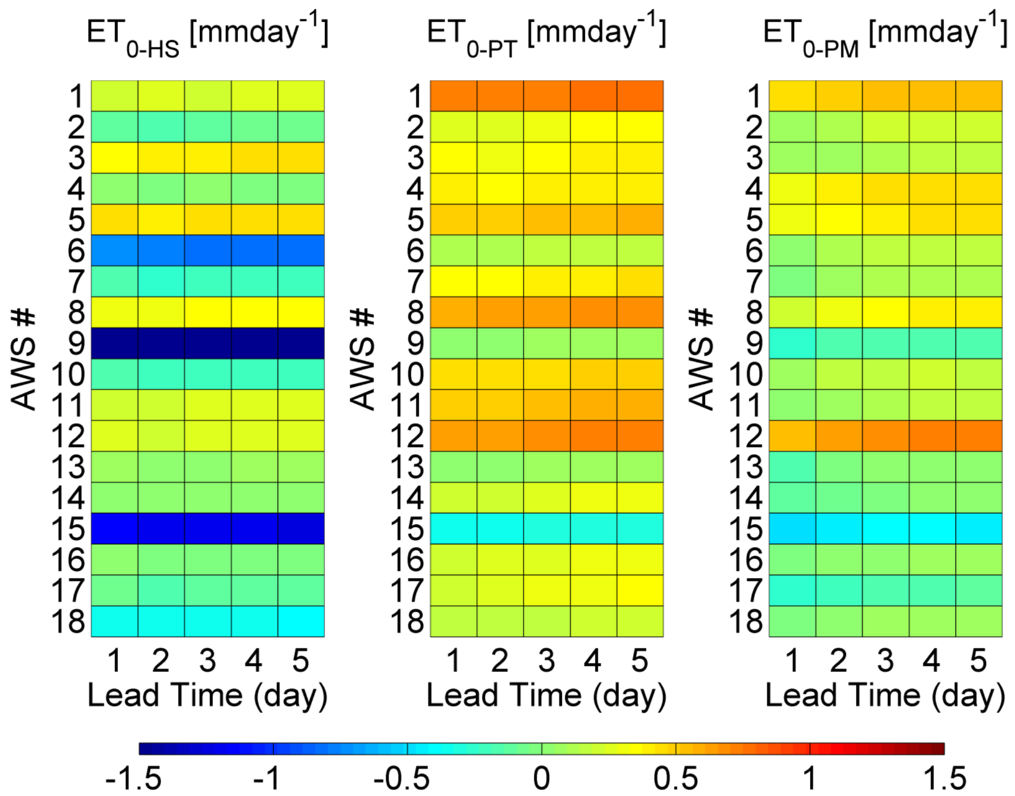

Figure 7. BIAS of forecasted vs. observed daily ET_0 for all 18 AWS sites and lead times

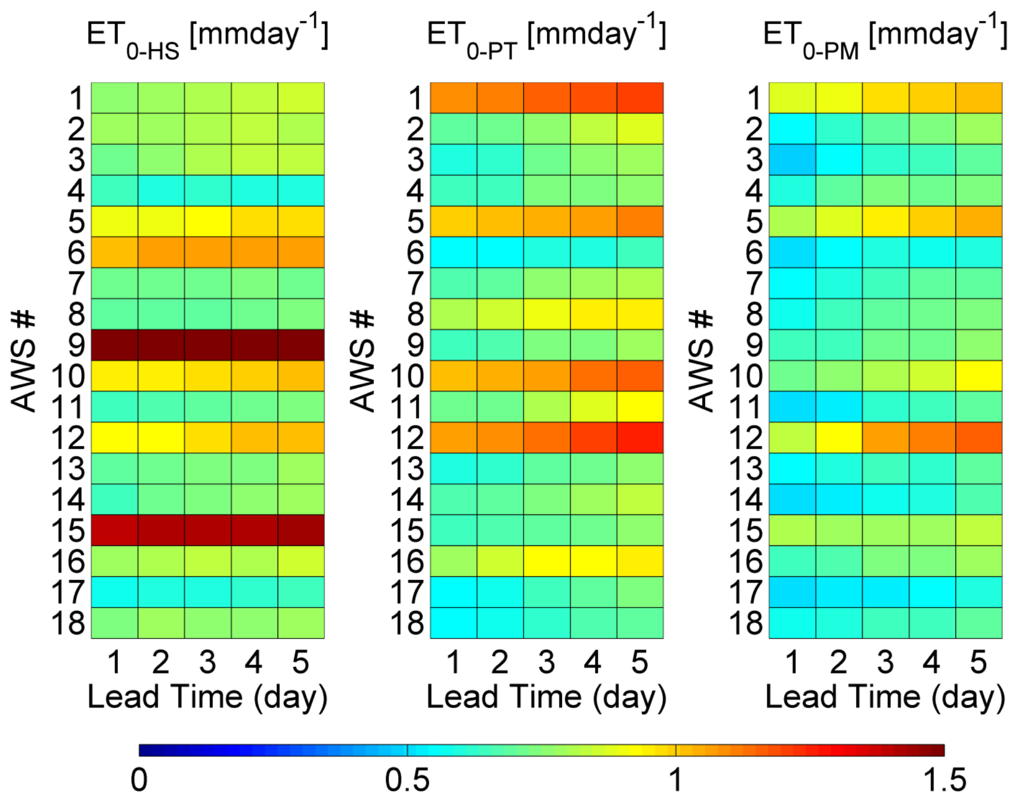

Figure 8. RMSE of forecasted vs. observed daily ET_0 for all 18 AWS sites and lead times

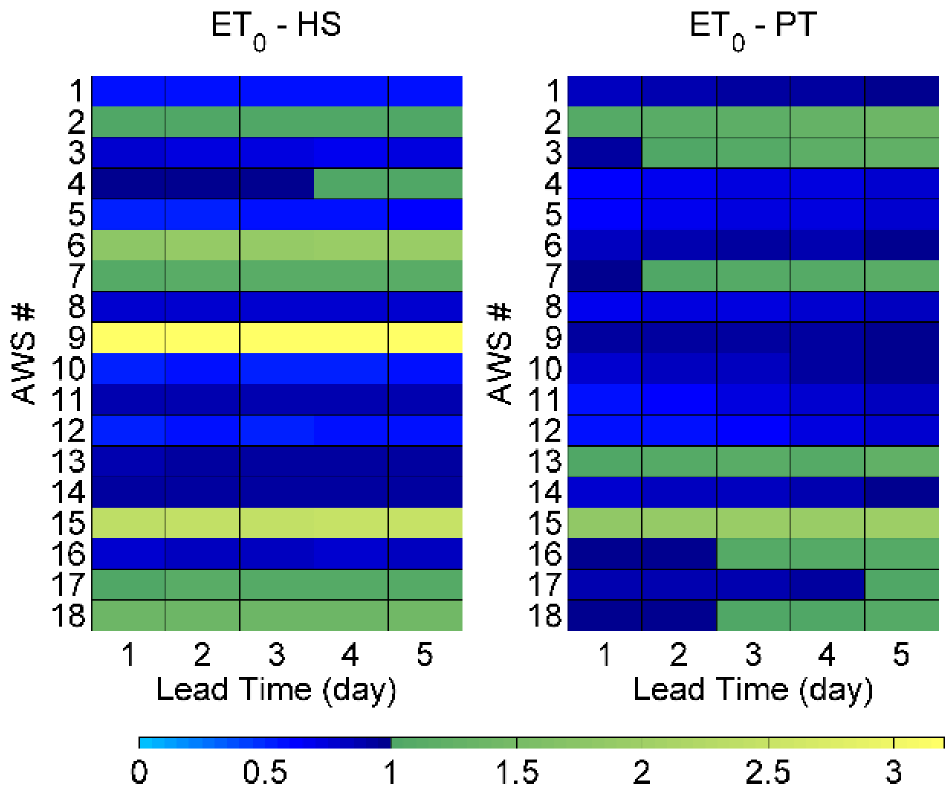

Figure 9. rBIAS of forecasted vs. observed daily ET_0 for all 18 AWS sites and lead times

Overall, based on Figures 7-9, the following considerations can be drawn.Except for AWS 6, 9 and 15, the prediction performances obtained with the simple HS equation are comparable with those obtained by previous studies, which employed the PM equation with bias-corrected NWP outputs for similar lead times (e.g., Silva et al., 2010; Perera et al., 2014). The forecast BIAS at AWS 6, 9 and 15 is mainly due to the temperature forecast errors, which were outlined in the previous section and in Figure 3. At AWS 1, 5, 10 and 12 the forecast BIAS instead appears to be mainly due to model simplification. Since the HS method does not explicitly account for relative humidity, it can overestimate ET_0 in humid regions, and underestimate it in areas of high winds and high vapour pressure deficits. For these sites, a specific calibration of K_{HS} is particularly recommended (Allen et al., 1998).

$ET_{0\text{-}PT}$ *BIAS* is always positive, except for AWS 15, where both temperature and radiation are underestimated. The highest RMSE values are observed at AWS 1, 5, 10 and 12 due to the errors in forecasted temperature and solar radiation. As indicated by Figure 9, at these AWS sites, model error and forecast errors play a similar role. For all other stations, $ET_{0\text{-}PT}$ exhibits absolute *BIAS* smaller than 0.5 mm day^{-1} and RMSE smaller than 0.75 mm day^{-1}, which are excellent forecast performances compared with previous analogous studies.

$ET_{0\text{-}PM}$ forecasts present a pattern of BIAS and RMSE similar to that of $ET_{0\text{-}PT}$ but with smaller absolute values: the only negative BIAS are observed at AWS 9, 15 and 17; the highest RMSE are observed at AWS 1, 5, 10 and 12. AWS 12 exhibits absolute BIAS exceeding 0.5 mm day^{-1} for lead times greater than one day and RMSE greater than 1 mm day^{-1} for lead times exceeding three days.

The impact of the BIAS on the forecasted air temperature at AWS 6, 9 and 15 is mitigated with equations PM and PT, where other weather variables, different from the air temperature, play a more important role and are less affected by proximity to the sea.

Figures 10a-b depict the coefficients of variation (CV) and the correlation coefficient, respectively, across all 18 AWS sites for varying lead times. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data values not considered outliers, and outliers are plotted individually. The points are drawn as outliers if they are larger than $q_3 + 1.5(q_3 - q_1)$ or smaller than $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles, respectively. The circle mark represents the mean value among the AWS sites.

The values of the CV increase with lead times as a result of the increasing ensemble spread. The CV also increases as the number of uncertain variables involved in the ET_0 computation increases, moving from HS to equations PT and PM.

The correlation (Figure 10b) exhibits a marked decrease with increasing lead time. The rate of the decreasing trend is larger for $ET_{0\text{-}PT}$, due to the higher sensitivity of the forecasted radiation to the lead time. In any case, the correlation generally increases from equation HS to PT and PM, except for AWS 1, 5, 10 and 12 where the $ET_{0\text{-}PT}$ correlation is smaller than $ET_{0\text{-}HS}$ for lead times exceeding three days.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

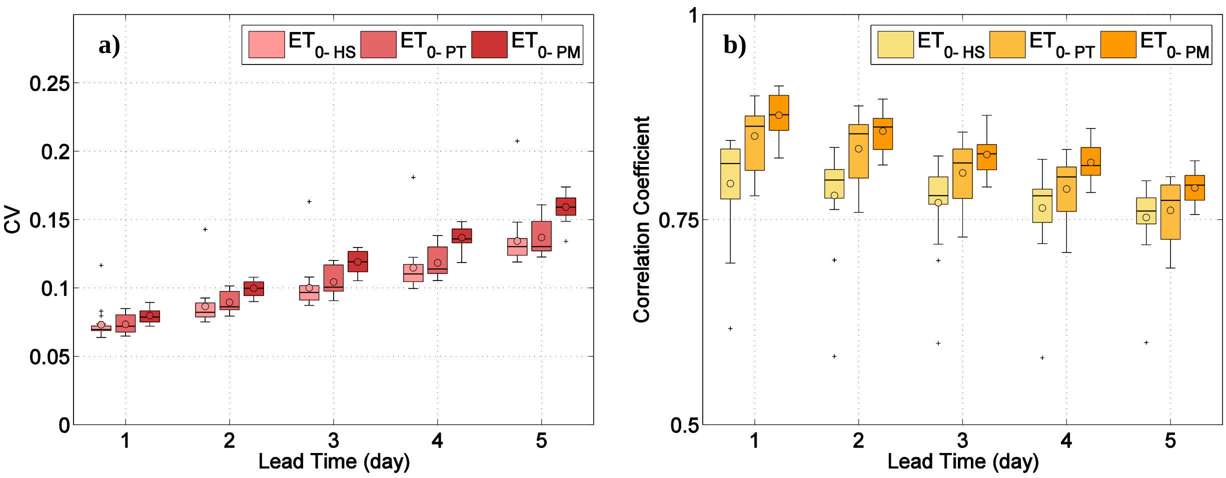

Figure 10. a) Coefficient of variation (CV) and **b)** correlation coefficient across all 18 AWS sites of forecasted vs. observed daily ET_0 (the circles represent the mean values).

An insight into the impact of the forecast errors of different weather variables, i.e. temperature, solar radiation, relative humidity and wind speed, on the daily ET_0 estimation is afforded by Figure 11. It shows the variation of forecast performances in terms of RMSE when we substitute one weather variable forecast with its own observed value. The substitution of the weather variable forecast with the weather variable observation can be useful for highlighting the sensitivity of the ET_0 forecast to errors in the weather forecast for that variable, as suggested by Perera et al. (2014).

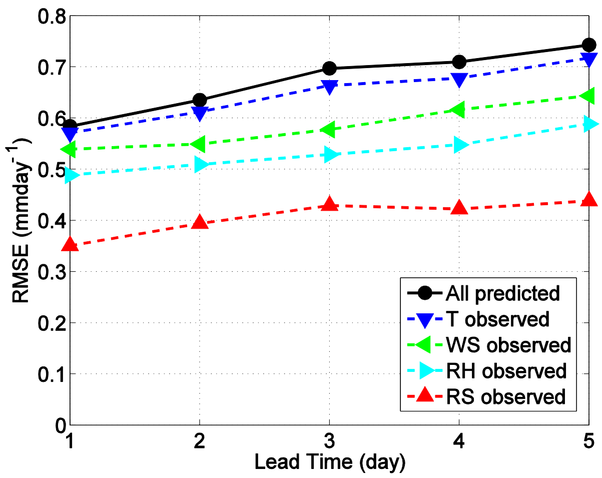

Figure 11. Sensitivity of daily FAO Penman-Monteith ET_0 forecasts to errors in weather predicted variables

Figure 11 suggests that errors in solar radiation forecast have the greatest influence on the $ET_{0\text{-}PM}$ forecast performance (the improvement by using its observed value is the largest), followed by relative humidity and wind speed. For temperature, the improvement by using observed values is

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

negligible, which is also the reason why we have still good performances of $ET_{0\text{-}PM}$ forecasts in those sites where the errors in temperature forecasts lead to severe underestimation of the predicted $ET_{0\text{-}HS}$. The results in Figure 11 are somewhat consistent with the findings of Perera et al. (2014), who showed that forecast errors related to solar radiation are the main source of errors in ET_0 forecasts. On the other hand, in latter's findings, the sensitivity of ET_0 forecasts to errors in the temperature forecasts seems to play a more important role.

4.3. Forecast performances of reference evapotranspiration with probabilistic metrics

The probabilistic metrics of the forecast performances to assess the quality of the ensemble forecasts are reported in Figures 12-14. The metrics are all computed for a threshold t equal to the average (among all the AWS sites) median value of $ET_{0g\text{-}PM}$ (see Eqs. 9-11). Boxplots of the *BSS* among all 18 AWS sites are shown in Figure 12 for increasing lead times. BSS declines with increasing lead time, but the reduction from lead day 1 to lead day 5 is smaller than 30% for all ET_0 forecasting methods herein examined.

All ET_0 forecasts are better than sample climatology, except for $ET_{0\text{-}HS}$ forecasts at AWS 6, 9 and 15, where anomalous BSS values below zero were observed due to significant systematic bias in temperature forecasts, as illustrated above. These BSS outliers also caused a significant bias in the mean BSS (circle marks) of $ET_{0\text{-}HS}$ with respect to the median values (horizontal central line of the whisker). The median BSS of $ET_{0\text{-}HS}$ is greater than 0.45, while its p_{25} exceeds 0.37, for all lead times. $ET_{0\text{-}PT}$ presents the largest spreads in BSS, symptomatic of a lower capacity to forecast solar radiation at a large number of AWS. Its median value is always above 0.47 and its 25th percentile is greater than 0.30. The median BSS of $ET_{0\text{-}PM}$ is greater than 0.50 for all lead times, while its 25th percentile exceeds 0.45. Overall, these BSS values are quite high compared with the findings of Tian and Martinez (2012a; 2014), who presented the first studies with a probabilistic verification of ET_0 forecasts. Tian and Martinez (2014) obtained the best BSS values by statistical bias-correcting and downscaling GFS reanalysis forecasts to a spatial resolution of 12 km^2. In this case, the maximum BSS scores achieved

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

in the warm seasons at 1-day lead time were 0.20 for the middle terciles thresholds and around 0.40 for the upper and lower terciles. Moreover, in their study, BSS radically decreased towards zero for increasing lead times up to five days.

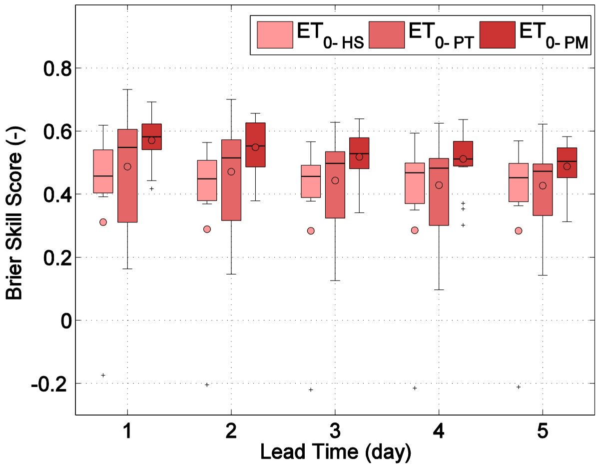

Figure 12. BSS across all 18 AWS sites of forecasted vs. observed daily ET_0 (the circles denote the mean values).

Figures 13a-c show the reliability diagrams for the three examined methods: $ET_{0\text{-}HS}$, $ET_{0\text{-}PT}$ and $ET_{0\text{-}PM}$. In all cases, the forecasts exhibit good sharpness as described by the histograms in the insets (upper left corners) of Figures 13a-c.As indicated by Figure 13c, $ET_{0\text{-}PM}$ ensures good reliability and resolution for all lead times. Slight overforecasting occurs for lead times exceeding three days.

The correspondence between forecasted and observed frequencies worsens when simpler methods such as PT and HS are employed. The curves related to $ET_{0\text{-}PT}$ forecasts (Figure 13b) indicate overforecasting, except for a lead time of five days at high probabilities. The case of $ET_{0\text{-}HS}$ forecasts (Figure 13a) is the worst case with poor reliability and resolution, probably caused by those AWS sites with negative BSS (i.e. AWS 6, 9 and 15).

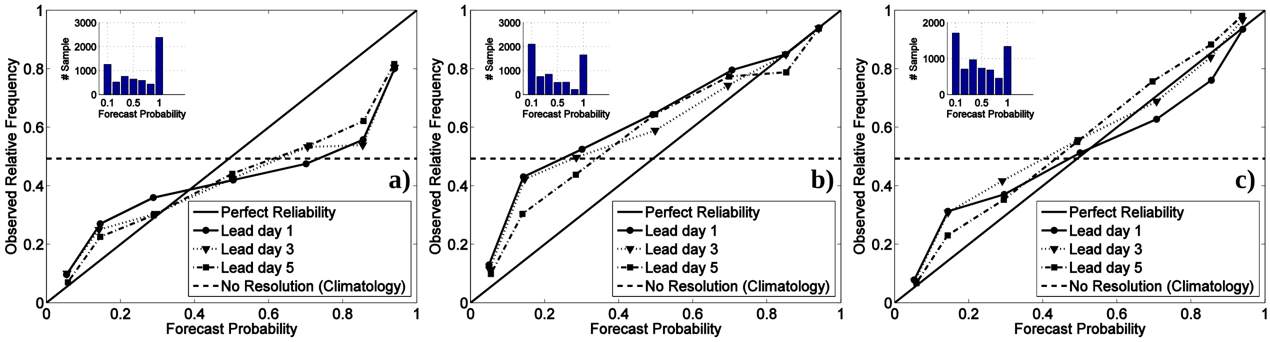

Figure 13. Reliability diagrams for **a)** *ET_{0-HS}*, **b)** *ET_{0-PT}* and **c)** *ET_{0-PM}*

Finally, Figures 14a-c show the ROC diagrams, respectively, for the cases of ET_{0-HS}, ET_{0-PT} and ET_{0-PM}. ROC diagrams clarify how well the probabilistic forecasts discriminate between events and non-events. The dependence on lead time is very clear: the performances on the ROC diagram decline with increasing lead time for all the ET_0 methods. Very slight differences are appreciable between Figure 14b and Figure 14c, which show, respectively, the ROC diagram for ET_{0-PT} forecasts and ET_{0-PM} forecasts. The case of ET_{0-HS} forecasts (Figure 14a) is that which performs worst. Yet it is still very satisfying compared with the results shown by Tian and Martinez (2012a; 2014).

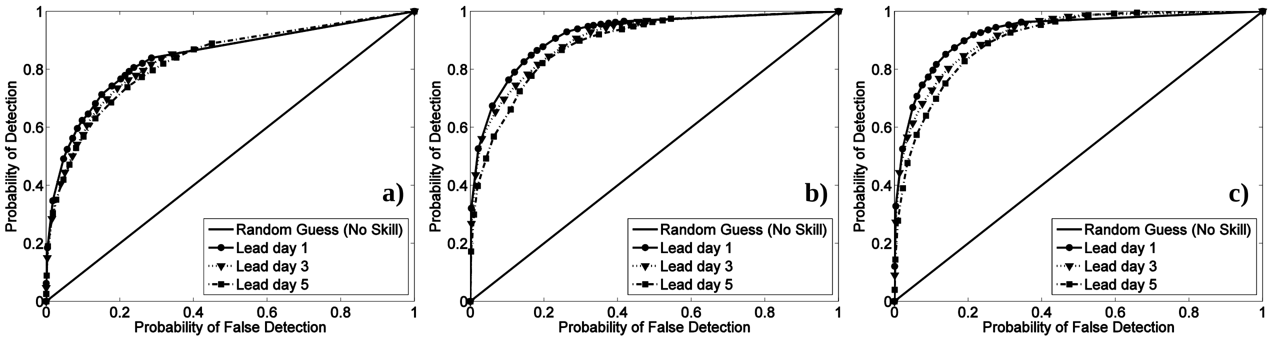

Figure 15. ROC diagrams for **a)** *ET_{0-HS}*, **b)** *ET_{0-PT}* and **c)** *ET_{0-PM}*

5. Conclusions

A more rational and efficient use of water in agriculture can be achieved by supplying accurate forecasts of reference evapotranspiration (ET_0), which is one of the key factors for the assessment of crop water requirements and irrigation needs. A probabilistic approach is recognized as the most appropriate to cope with the uncertainty of weather variability in the short-medium term. Although

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015

statistical downscaling techniques of global ensemble forecasts have been proved to provide reliable forecasts (e.g. Tian and Martinez, 2014), their applicability is hindered by the need of large data sets of ground-based observations for their calibration. The operational availability of weather forecasts by limited area ensemble prediction systems (LEPS) offers new opportunities for developing reliable advisory services for agricultural management, particularly for rural areas where complete ground-based weather data are rare.

To our knowledge, this is the first study to verify the ability of LEPS outputs to forecast reference evapotranspiration in the short-medium range. COSMO-LEPS forecasts with a spatial resolution of 7 km and lead times up to five days were employed for forecasting daily ET_0 in southern Italy, in a region where weather forecasting is quite challenging given its complex topography in proximity to the Mediterranean coastline. The numerical weather outputs were applied without any preliminary post-processing aimed at removing local systematic errors. Forecast performances were assessed with three different empirical methods for estimating ET_0, in order to evaluate the size of the weather forecast errors with models of different accuracies.

ET_0 forecasts with the FAO Penman-Monteith (PM) equation were skillful and reliable, with limited sensitivity to the forecast lead time. Both deterministic and probabilistic scores were better than those presented by analogous studies (e.g. Perera et al., 2014; Tian and Martinez, 2014). Solar radiation forecast errors appear to be the largest source of error for PM forecasts.

High skill scores were achieved also with the simpler and uncalibrated Priestley-Taylor (PT) and Hargreaves-Samani (HS) equations, except for a few locations close to the coastline. Forecasts with the uncalibrated Hargreaves-Samani (HS) and Priestley-Taylor (PT) equations were more vulnerable to local systematic errors of the forecasted temperature and solar radiation, respectively. In almost half of the 18 locations examined, systematic weather forecast errors appear to affect ET_0 forecasts errors more than the application of an uncalibrated equation as an alternative solution to the more complex PM equation. Systematic errors are mainly due to limitations of the numerical weather model to resolve topographic effects on local weather conditions in areas with a complex terrain, as occurs along coastlines surrounded by high mountains.

The performances herein presented are based on data from only two irrigation seasons which did not experience extreme weather conditions. Such conditions could enhance the effects of systematic errors of the forecasting system and thus reduce the accuracy of ET_0 forecasts, particularly if estimated with simpler HS and PT estimation methods.

Since the installation of comprehensive new weather stations is becoming common in modern precision farming, further studies will be devoted to develop adaptive methods for removing systematic biases with ground data in real time. Such methods could offer opportunities to fully exploit the advances in ensemble numerical weather forecasting by developing innovative advisory services based on the optimal combination of LEPS forecasts and ground-based data from newly installed automatic weather stations.

Acknowledgements

The study is part of the PIRAM projects, funded by the European Union and Campania Region, within the rural development program 2007-2013. COSMO-LEPS forecasts were provided by ARPA Emilia Romagna – Servizio Idro-Meteo-Clima. The Regional Meteorological Service provided ground-based data. Special thanks to Carlo Cacciamani, Andrea Montani, Davide Cesari, Matteo Gentilella and Mauro Biafore for their support.

References

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements, Irrigation and Drain, Paper No. 56, FAO, Rome, Italy, p. 300.

Bormann H., 2010. Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change at German climate stations. Climatic Change, 104, 729-753.

Buizza, R., Miller, M., and Palmer, T. N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Q. J. R. Meteorol. Soc., 125, 2887-2908.

Buzzi, A., M. Fantini, P. Malguzzi and F. Nerozzi, 1994. Validation of a limited area model in cases of Mediterranean cyclogenesis: surface fields and precipitation scores. Meteorol. Atmos. Phys., 53, 137-153.

Cai, J., Liu, Y., Lei, T., and Pereira, L.S. 2007. Estimating reference evapotranspiration with the FAO Penman–Monteith equation using daily weather forecast messages. Agric For Meteorol., 145:22–35. doi:10.1016/j.agrformet.2007.04.012

Cai, J. B., Liu, Y., Xu, D., Paredes, P., and Pereira, L. S., 2009. Simulation of the soil water balance of wheat using daily weather forecast messages to estimate the reference evapotranspiration. Hydrol. Earth Syst. Sci., 13, 1045-1059. doi:10.5194/hess-13-1045-2009.

De Leonibus,L., Vecchi, U, 1999. EUMETNET Automatic Weather Station Contract - Technical Specifications.

Droogers, P., and Allen, R.G., 2002. Estimating reference evapotranspiration under inaccurate data condition. J. Irrig. Drain. Syst., 16(1), 33-45.

Er-Raki, S., Chehbouni, A., Khabba, S., Simonneaux, V., Jarlan, L., Ouldbba, A., Rodriguez, J.C., Allen, R., 2010. Assessment of reference evapotranspiration methods in semi-arid regions: Can weather forecast data be used as alternate of ground meteorological parameters? J. Arid Environ. 74 (12), 1587–1596. doi:10.1016/j.jaridenv.2010.07.002

Garcia, M., Raes, D., Allen, R., and Herbas, C., 2004. Dynamics of reference evapotranspiration in the Bolivian highlands (Altiplano). Agric. For. Meteorol. , 125 (1-2), 67–82.

Hamill, T. M., and Whitaker, J. S., 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. Mon. Wea. Rev., 134, 3209–3229.

Hamill, T. M., Whitaker, J. S. and Mullen, S. L., 2006. Reforecasts: An Important Dataset for Improving Weather Predictions. Bull. Amer. Meteor. Soc., 87, 33–46. doi: http://dx.doi.org/10.1175/BAMS-87-1-33

Hamill, T. M., Bates, G.T., Whitaker, J. S., Murray, D. R., Fiorino, M., Galarneau, T. J., Zhu, Y. and Lapenta, W., 2013. NOAA’s Second Generation Global Medium Range Ensemble Forecast Dataset . Bull. Amer. Meteor. Soc ., 94 , 15531565.

Hargreaves, G.H. and Z.A. Samani, 1985. Reference crop evapotranspiration from temperature. Transaction of ASAE 1(2):96-99.

Ishak, A. M., Bray, M., Remesan, R. and Han, D., 2010. Estimating reference evapotranspiration using numerical weather modelling. Hydrol. Process., 24: 3490–3509. doi: 10.1002/hyp.7770

Lhomme, J.-P., 1996. A theoretical basis for the Priestley-Taylor coefficient. Boundary-Layer Meteorol. 82:179-191.

Marsigli, C., Montani, A. and Paccagnella, T., 2014. Perturbation of initial and boundary conditions for a limited-area ensemble: multi-model versus single-model approach. Q.J.R. Meteorol. Soc., 140: 197–208. doi: 10.1002/qj.2128

Mesinger, F. et al., 2006. North American Regional Reanalysis. Bull. Amer. Meteor. Soc., 87, 343-360.

Montani, A., Cesari, D., Marsigli, C., and Paccagnella, T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus A, 63, 605–624.

Murphy, A.H., 1973. A new vector partition of the probability score. J. Appl. Meteor., 12, 595-600.

Peel, M. C., Finlayson, B. L. and McMahon, T. A., 2007. "Updated world map of the Köppen–Geiger climate classification". Hydrol. Earth Syst. Sci. 11: 1633–1644. doi:10.5194/hess-11-1633-2007

Pereira, A.R., 2004. The Priestley–Taylor parameter and the decoupling factor for estimating reference evapotranspiration. Agric. For. Meteorol., 194,50–63

Perera, K.C., Western, A.W., Nawarathna, B., George B., 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. Agric. For. Meteorol., 125, 305–313. doi:10.1016/j.agrformet.2004.04.002

Priestley, C. H. B., and Taylor, R. J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Wea. Rev., 100, 81–92.

Shahidian, S., Serralheiro, R., Serrano, J., Teixeira, J., Haie N., and Santos, F., 2012. Hargreaves and Other Reduced-Set Methods for Calculating Evapotranspiration. Evapotranspiration - Remote Sensing and Modeling, Dr. Ayse Irmak (Ed.), ISBN: 978-953-307-808-3, InTech, Available from: http://www.intechopen.com/books/evapotranspiration-remote-sensing-and-modeling/hargreaves-and-otherreduced-set-methods-for-calculating-evapotranspiration

Silva, D., Meza, F.J., Varas, E., 2010. Estimating reference evapotranspiration (ET_0) using numerical weather forecast data in central Chile. J. Hydrol. 382 (1–4), 64–71.

Srivastava, P. K., Han, D., Rico Ramirez, M. A. and Islam, T. 2013. Comparative assessment of evapotranspiration derived from NCEP and ECMWF global datasets through Weather Research and Forecasting model. Atmosph. Sci. Lett., 14: 118–125. doi: 10.1002/asl2.427

Thornthwaite, C. W., 1948: An approach towards a rational classification of climate. Geogr. Rev., 38, 55-94.

Tian, D. and Martinez, C.J., 2012a. Comparison of two analog-based downscaling methods for regional reference evapotranspiration forecasts. J. Hydrol. 475 (0), 350–364. doi:10.1016/j.jhydrol.2012.10.009

Tian, D, and Martinez, C. J., 2012b. Forecasting reference evapotranspiration using retrospective forecast analogs in the Southeastern United States. J. Hydrometeorol., 13 (6), 1874–1892. doi: http://dx.doi.org/10.1175/JHM-D-12-037.1

Tian, D. and Martinez, C. J., 2014. The GEFS-Based Daily Reference Evapotranspiration (ETo) Forecast and Its Implication for Water Management in the Southeastern United States. J. Hydrometeorol., 15, 1152–1165. doi: http://dx.doi.org/10.1175/JHM-D-13-0119.1

Van Schaeybroeck, B. and Vannitsem, S., 2015. Ensemble post-processing using member-by-member approaches: theoretical aspects. Q.J.R. Meteorol. Soc., 141: 807–818. doi: 10.1002/qj.2397

Vuolo, F., D'Urso, G., De Michele, C., Bianchi, B. and Cutting M., 2015. Satellite-based irrigation advisory services: a common tool for different experiences from Europe to Australia. Agric. Water Manage., 147, 82–95

Wilks, D. S., 2011. Statistical Methods in the Atmospheric Sciences (third ed.). Academic Press., p. 704.

WMO, 2012. Guide to Agricultural Meteorological Practices. WMO No. 134, p. 799.

Xu, C.Y. and Singh, V.P., 2000. Evaluation and Generalization of Radiation-based Methods for Calculating Evaporation, Hydrolog. Processes 14: 339–34.

This is a post-peer-review, pre-copyedit version of an article published in "Agricultural Water Management".
The final authenticated version is available online at: https://doi.org/10.1016/j.agwat.2016.09.015