Gaussian process models for reference ET estimation from alternative meteorological data sources

Daniel Holman [a,*], Mohan Sridharan [b], Prasanna Gowda [c], Dana Porter [d,b], Thomas Marek [e], Terry Howell [c], Jerry Moorhead [c]

[a] Texas A&M AgriLife Research, Lubbock, TX 79403, USA
[b] The University of Auckland, Auckland 1142, NZ
[c] United States Department of Agriculture – Agricultural Research Service, Bushland, TX 79012, USA
[d] Texas A&M AgriLife Extension Service, Lubbock, TX 79403, USA
[e] Texas A&M AgriLife Research, Amarillo, TX 79106, USA

ARTICLE INFO

SUMMARY

Accurate estimates of daily crop evapotranspiration (ET) are needed for efficient irrigation management, especially in arid and semi-arid regions where crop water demand exceeds rainfall. Daily grass or alfalfa reference ET values and crop coefficients are widely used to estimate crop water demand. Inaccurate reference ET estimates can hence have a tremendous impact on irrigation costs and the demands on U.S. freshwater resources, particularly within the Ogallala aquifer region. ET networks calculate reference ET using local meteorological data. With gaps in spatial coverage of existing networks and the agriculture-based Texas High Plains ET (TXHPET) network in jeopardy due to lack of funding, there is an immediate need for alternative sources capable of filling data gaps without high maintenance and field-based support costs. Non-agricultural weather stations located throughout the Texas High Plains are providing publicly accessible meteorological data. However, there are concerns that the data may not be suitable for estimating reference ET due to factors such as weather station siting, fetch requirements, data formats, parameters recorded, and quality control issues. The goal of the research reported in this paper is to assess the use of alternative data sources for reference ET computation. Towards this objective, we trained Gaussian process models, an instance of kernel-based machine learning algorithms, on data collected from weather stations to estimate reference ET values and augment the TXHPET database. Results show that Gaussian process models provide much greater accuracy than baseline least square regression models.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Accurate estimates of daily crop evapotranspiration (ET) are valuable for irrigation management within arid, semi-arid, and semi-humid regions where crop water demand exceeds rainfall. Daily grass/alfalfa reference ET (ETo) values are widely used in conjunction with crop coefficients (Kc) to estimate crop water demand. Hence, the impact of inaccurate ET estimates can hardly be overstated given the increased demands on U.S. freshwater resources, especially within the central Great Plains underlain by the vast but declining Ogallala groundwater aquifer. Reference ET can be estimated using the FAO-56 (Allen et al., 1998) or the recent ASCE Standardized Reference ET Equation (Allen et al., 2005). Areal coverage of ET networks is not universal and there are significant gaps in the spatial coverage. It is further complicated by high spatial variation in air temperature, wind speed, wind direction and other weather parameters due to regional effects such as atmospheric circulation patterns (Buishand and Brandsma, 1997; Knapp, 1992) and local effects such as topography (Goovaerts, 2000), land use (Li et al., 2010), elevation (Dodson and Marks, 1997) and soil properties (López-Granados et al., 2005). It is therefore difficult to use one predetermined weather station for determining daily reference ET for irrigation management over large regions. A sufficiently dense network of weather stations can effectively capture the spatial variability of parameters for use in computing reference ET. However, funding and staffing issues are ongoing concerns of many ET networks, making it difficult to expand coverage through additional weather stations. Although remote sensing based ET estimates are showing promise of expanding areal coverage and integration capabilities, accuracy of these methods is dependent on accuracy of input (i.e., weather

data and/or reference ET). There is hence an immediate need for exploring alternative data sources capable of filling data gaps without high maintenance and field-based support costs. Fortunately, there are other (non-agricultural, non-ET) weather stations and data networks, and some of these networks make data publicly accessible. However, there are concerns that the data may not be appropriate for estimating reference ET due to factors such as weather station siting, data formats, parameters recorded, and data quality control issues.

Fig. 1 from an article by Porter et al. (2012) illustrates the sensitivity of average daily grass reference ET to changes in air temperature, dew temperature, wind speed and solar radiation by 1 °C, 1 °C, 1 m s^{-1} and 25 W m^{-2} respectively, using the representative Bushland ET station in the Texas High Plains. The authors reported that grass reference ET calculations were most sensitive to errors in wind speed and air temperature, and that sensitivity was greater during the mid-summer growing season when greater accuracy levels are required for irrigation scheduling. For example, a 1 °C error in air temperature throughout the growing season in the Texas High Plains would yield $\approx$25 mm error in the total grass reference ET for the typical corn growing season. The cumulative effect of this error over the irrigated corn acreage within the Texas High Plains is approximately equivalent to the total drinking water consumption by the City of Houston for about 22 days (Marek et al., 2010). Erroneous selection of reference weather station and/or wrong values of reference ET rates can thus affect the water use efficiency and producers' net profits, in addition to the loss of mined water of the Ogallala aquifer. This illustrative example demonstrates the need for accurate reference ET data for irrigation scheduling and irrigation management.

Most statistical models reported in the literature on irrigation management are based on ordinary least square regression. Popular models used in regression include: (1) linear; $\mathbf{y} = c + \mathbf{b}\mathbf{x}$; (2) exponential; $\mathbf{y} = \mathbf{a}e^{\mathbf{b}\mathbf{x}}$; (3) power or logarithmic; $\mathbf{y} = \mathbf{a}\mathbf{x}^{\mathbf{b}}$; and (4) a quadratic polynomial; $\mathbf{y} = \mathbf{a}\mathbf{x}^2 + \mathbf{b}\mathbf{x} + c$. In these models, $\mathbf{y}$ represents the desired output vector (e.g., reference ET values from ET stations) and $\mathbf{x}$ is a vector representing input values such as rainfall, irrigation amount, weather parameters or reference ET estimated at non-agricultural weather stations. Values of coefficients $\mathbf{a}$, $\mathbf{b}$ (vectors with the same dimension as $\mathbf{x}$) and c are tuned on training data such that the computed values of output are as close as possible to the given (i.e., known ground truth) output values.
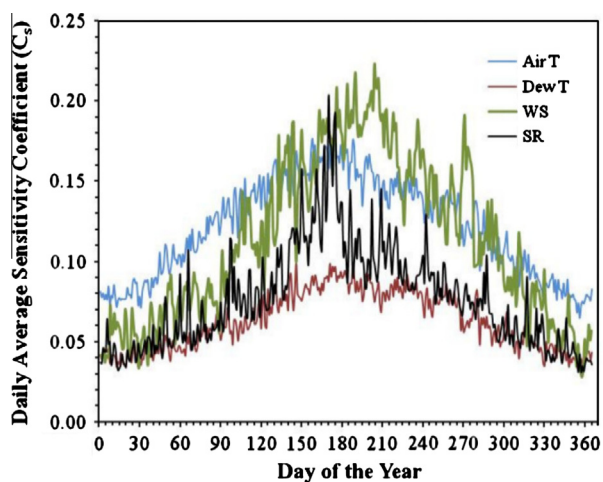

Fig. 1. Sensitivity of daily average grass reference ET to changes in air temperature, dew point temperature, wind speed, and solar radiation by 1 °C, 1 °C, 1 m/s, and 25 W/m^2 respectively, using data from Bushland ET station in the Texas High Plains (Porter et al., 2012).

The ordinary least squares regression formulations tend to fix the basis functions before observing the training data, and the number of basis functions grows exponentially with the dimensions of input space. Furthermore, the basis functions are not adaptable to data and the associated curse of dimensionality makes a strong case for the use of more sophisticated models (Bishop, 2008). In recent years, numerous statistical learning algorithms are being developed and used for inference and prediction. Examples of such methods include artificial neural network (ANN) (ASCE Task Committee on Applications of Artificial Neural Networks in Hydrology, 2000a,b; Buscema and Sacco, 2000; DeRoach, 1989; Hornik et al., 1989), support vector machine (Vapnik, 1995) and Gaussian Process (GP) models (Williams and Rasmussen, 1996). These methods provide substantial benefits over linear (or other) regression models. For instance, ANNs readily adapt to data and can be used to model complex functions between input and output parameters. Although different algorithms can be used to learn ANNs based on a compromise between computational cost and performance, the most popular choice is the back propagation algorithm (Rumelhart et al., 1986). However, ANN formulations can result in local minima, lead to over-fitting, and become computationally expensive in high-dimensional spaces. Further, it is also not easy to extract an understandable interpretation of the functions learned.

Support vector machines (SVMs) have a simple geometric interpretation, avoid overfitting and find global solutions. They project input features to high dimensions, resulting in sparse representations and robust decision boundaries. However, SVMs (and other similar algorithms) require that the parametric function that models the relationship between inputs and outputs be defined in advance—choosing an appropriate function from the infinite space of functions may be difficult in complex problem domains. Gaussian process is a non-parametric kernel-based machine learning algorithm. GP models are well-suited for the estimation step of our study and they capture the temporal evolution of normally-distributed random variables that represent the patterns being tracked. They have been successfully used for large-scale estimation problems with high-dimensional features, e.g., nuclear disasters, climate modeling and sensor placement for surveillance (Higdon et al., 2003; Kennedy and O'Hagan, 2001; Krause et al., 2008). To the best of our knowledge, only a few applications of GP models have been found in the research field of water resources management, including ground water transport (Marrel et al., 2009) and hydrological modeling (Song et al., 2011).

The objective of the study reported in this paper is to identify, evaluate and use alternative meteorological data sources for accurately computing reference ET values, resulting in efficient irrigation management and water resources planning. We trained GP models to estimate reference ET values using data from non-ET National Weather Service (NWS) stations. Although our study focused on the Texas High Plains, the experimental methodology can be used for estimating reference ET from non-agricultural and non-ET weather stations elsewhere in the U.S. and the world. An initial version of this study was reported in (Holman et al., 2013).

The remainder of the paper is organized as follows. Section 2 describes the geographic locations included in our study, while Section 3 presents an overview of the various steps involved in the study. Next, Gaussian processes and the underlying mathematical concepts are described in Section 4. The experimental setup and results are discussed in Sections 5 and 6 respectively, followed by the conclusions in Section 7.

2. Study area

This study covers a 39-county area of the Texas High Plains, as shown in Fig. 2. Most of this region is semi-arid with highly
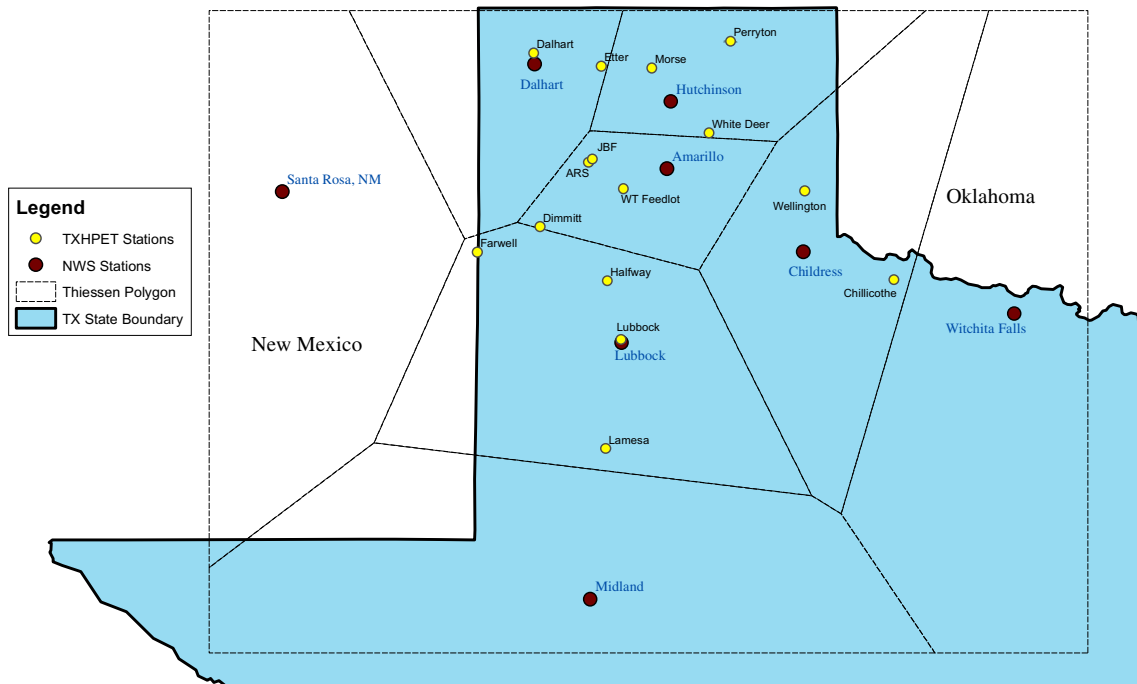
Fig. 2. Locations of 15 weather stations managed by the Texas High Plains ET network and paired NWS stations in the Texas High Plains.

variable precipitation (both temporally and spatially), averaging 400–560 mm from west to east. Most rainfall occurs in scattered thunderstorms—some areas may receive 50–100 mm of daily precipitation while areas a short distance away may not receive any rainfall. In 2010, the Lubbock weather station recorded a total of 672 mm of annual rainfall, while in 2011 the station only recorded 149 mm. This region is also known to have high evaporative demand (approximately 2500 mm/year Class A pan evaporation) due to high solar radiation, high vapor pressure deficit (VPD), and strong regional advection. The grass reference ET data from 15 weather stations managed by the TXHPET network (Porter et al., 2005) (Fig. 1) were used to train and evaluate GP models that estimate reference ET values based on meteorological data from non-ET NWS stations.

3. Materials and methods

This study was implemented in four steps: (1) identify publicly accessible data sources and evaluate whether they provide sufficient data parameters, detail and quality for use in computing daily reference ET; (2) determine whether missing data can be estimated using other available data, trends and associative models and calculate the daily reference ET; (3) assess and validate the quality of reference ET values determined from these data sources compared with valid ET network data sources; and (4) develop and evaluate models based on statistical learning methods to calculate reference ET given data from non-agricultural, non-ET weather stations or networks. In the remainder of the paper, we refer to non-agricultural, non-ET weather stations as "non-ET stations".

The first step involved a thorough assessment of all the weather networks in the Texas High Plains for estimating daily reference ET. Selection criteria include: (a) timely and real-time availability of data to the public at no cost; (b) length and continuity of historic records; (c) availability of measurements of parameters needed for estimating reference ET; and (d) ability to accurately estimate parameters that are missing or not measured. Once suitable non-ET stations were identified, a map showing the location of these

stations was used to create a Thiessen polygon map, which (in turn) was used to pair these stations with selected agricultural weather stations of the TXHPET network (i.e., the "ET stations") for model development and evaluation. The weather parameters considered include: daily maximum and minimum air temperatures, average dew point temperature, average wind speed, average relative humidity, barometric measure and solar radiation.

The second step involved computing missing data using equations provided in the 2005 Standardized reference ET methodology (Allen et al., 2005). For instance, if the solar radiation values were not measured or missing for a non-ET station, they were estimated using the *complex method equation* provided below:

$$R_s = \lambda_{R_s}\sqrt{(T_{max} - T_{min})}R_a \tag{1}$$

where T_{max} and T_{min} are daily maximum and minimum air temperatures, R_a is extraterrestrial radiation, λ_{R_s} is an empirical coefficient for semi-arid climates, and R_s is solar radiation. Barometric pressure is calculated using the elevation of the corresponding weather station. If the missing data cannot be estimated (e.g. wind speed) for a given day, then the data records for that day were removed. After verifying the weather dataset and filling the missing data, unit conversions of the weather parameters were performed (as necessary) to estimate daily reference ET using the ASCE Standardized ET equation-based Bushland Reference ET Calculator (Gowda et al., 2012).

In the third step, statistical relationships were computed between daily reference ET estimates obtained from the ET stations and the corresponding non-ET stations to assess feasibility of using alternative data sources to expand (or increase data intensity within) the coverage area of the Texas High Plains ET network. Coefficient of determination (R^2), slope and intercept of the regression line, Nash–Sutcliffe efficiency (NSE) and root mean square error (RMSE) were used to compare the reference ET values obtained from the ET stations and those obtained from the corresponding non-ET stations. The value of R^2 describes the proportion of variability in the observed data that is explained by the model—R^2 ranges from 0 to 1, with a higher value indicating a better goodness of fit (model explanation). For instance, $R^2 = 1$ with

an intercept of 0 and slope of 1 indicates a perfect fit between the observed and modeled data. The NSE is a common efficiency measure that compares variance in estimations with the measured data variance. NSE ranges between—inf and 1—values closer to 1 are more accurate and an NSE of 1 represents an optimal model. Negative NSE values indicate that the mean of the observations is more accurate than the model estimation. An RMSE of 0 also indicates a perfect fit—it is usually computed as a % of the observed mean, e.g., RMSE < 50% is usually considered low. The RMSE measure is appropriate for our study because it provides values in the same units as the values that are to be estimated (i.e., reference ET). More information on performance statistics can be found in Moriasi et al. (2007).

In the fourth step, the Gaussian process models were trained and evaluated. These models capture the relationship between the meteorological data from non-ET stations and the corresponding reference ET values from the ET stations. The trained models can then be used to estimate reference ET values given new data from the non-ET stations. For evaluating the estimation ability of GP models, the daily reference ET database was divided into two parts. Data corresponding to odd-numbered days of the year were used for model development and data from even-numbered days of the year were used for validation. Performance statistics (R^2, NSE and RMSE) were used to evaluate and compare the estimation capabilities of the GP models with (baseline) linear regression (LR) models.

4. Gaussian processes

Gaussian processes are sophisticated supervised learning models used for regression (Williams and Rasmussen, 1996) and classification (Williams and Barber, 1998). Supervised learning approaches infer a hypothesis function $h(x)$ based on training data. Training data consists of a set of N vectors consisting of inputs: $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ and corresponding target outputs: $\mathbf{T} = \{\mathbf{t_1}, \ldots, \mathbf{t_N}\}$, to generate input–output pairs: $\{(\mathbf{x_i}, \mathbf{t_i}), \mathbf{i} = \mathbf{1} \ldots \mathbf{N}\}$. The supervised learner uses the training data to learn a model that approximates h. To evaluate the learned model's estimation ability, a testing dataset consisting of previously unseen inputs and target outputs is defined: $\{(\hat{\mathbf{x}}_i, \hat{t}_i), i = 1 \ldots N\}$. The learned model processes input vectors of the testing dataset to estimate outputs: $\hat{y}(\mathbf{x})$, which are compared with ground truth (i.e., actual) target outputs. Various statistical error measures can hence be used to compute the accuracy of values estimated by the learned model.

For non-linear regression problems, the unknown function $y(\mathbf{x})$ exists in the infinite-dimensional space of possible functions for $\mathbf{x}$, making it difficult to decide the range of possible non-linear functions. Standard parametric models such as ANNs, linear regression and polynomial regression require that $y(\mathbf{x})$ have an explicitly defined functional form whose parameters are defined in advance—the values of parameters are assigned by learning weights $\mathbf{W}$. Choosing this function from the infinite space of function types and weights can be a challenge. Gaussian processes address this issue by placing a prior $P(y(\mathbf{x}))$ over the space of functions. GP models thus do not need an explicit parametric definition of the function $y(\mathbf{x})$, i.e., they are non-parametric. Instead, (stochastic) random variables define priors for each input vector. Random functions defined over the space of inputs constitute the GP prior, as shown in Fig. 3. During the training phase, the discrete set of inputs are used to modify these functions to pass as close as possible to the target outputs, thus approximating the (unknown) underlying function. Gaussian processes can be viewed as a natural generalization of a Gaussian distribution over a finite vector space to an infinite space of functions. Just as a Gaussian distribution is defined by its mean vector and covariance matrix, a GP is defined by its mean and covariance functions $\mu(x)$ and $C(x, x')$:
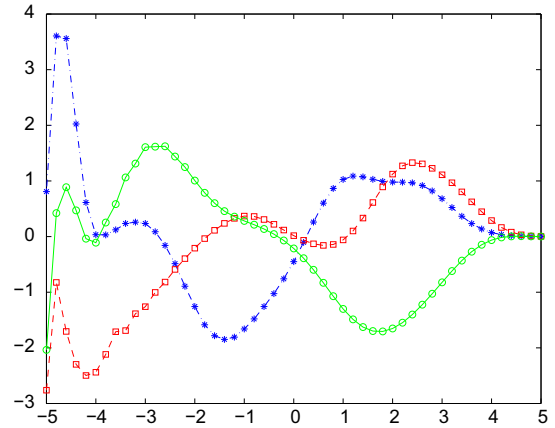

Fig. 3. Illustrative (general) example of a Gaussian process prior—three random functions that define outputs as a function of the inputs. GP models use the input values in the training data to incrementally revise these functions with the objective of minimizing the error in estimating the corresponding output values.

$$f \sim GP(\mu(\mathbf{x}), C(\mathbf{x}, \mathbf{x}')) \qquad (2)$$

where the function f is distributed as a Gaussian process with mean function $\mu(x)$ and covariance function $C(\mathbf{x}, \mathbf{x}')$. In our research, we define the mean function as the zero function. The covariance function expresses the expected covariance of the values at each pair of points $\mathbf{x}$ and $\mathbf{x}'$. Given N input vectors in the training data, the covariance function is a $N \times N$ covariance matrix $K : K_{ij} = C(\mathbf{x_i}, \mathbf{x_j})$. This matrix can be used to estimate output values for new inputs. In general, the estimated distribution is Gaussian with mean and covariance:

$$\hat{y} = \mathbf{k}^T(\mathbf{x})K^{-1}\mathbf{t} \qquad (3)$$

$$\sigma_{\hat{y}}^2(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T(\mathbf{x})K^{-1}\mathbf{k}(\mathbf{x})$$

where $\mathbf{x}$ is a new input vector, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ are the training data input vectors, $\mathbf{k}(\mathbf{x}) = (C(\mathbf{x}, \mathbf{x}^{(1)}), \ldots, C(\mathbf{x}, \mathbf{x}^{(N)}))^T$ denotes the matrix of covariances between the input and training data, K is the covariance matrix for training data, and $\mathbf{t} = (t^{(1)}, \ldots, t^{(N)})^T$. This algorithm has $O(N^3)$ time complexity due to the matrix inversion in Eq. (3). GP formulations can hence become infeasible for data with a large number of samples. Algorithms are being developed to enable GP formulations of domains with large datasets (Bishop, 2008). However, our experiments consist of a few thousand training samples per weather station and the time complexity is (currently) not an issue.

Many different options exist for selecting covariance functions for a Gaussian process. The main requirement is that the function should generate a non-negative definite covariance matrix for any set of inputs $(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$. Graphically, the goal is to define covariances such that points that are nearby in the input space produce similar output estimates. In the research reported in this paper, we chose the popular radial basis function (RBF) kernels (Musavi et al., 1992):

$$C(\mathbf{x}, \mathbf{x}') = e^{-\gamma * (x - x')^2} \qquad (4)$$

The key advantage of using a non-parametric model such as GP is that it does not require any manual parameter tuning. Instead, the covariance function contains *hyperparameters* that are tuned automatically to maximize the likelihood of training data. Eq. (4) contains a single hyperparameter γ. Assigning different values to the hyperparameter results in different GP models. We randomly initialize a finite set of hyperparameters over the space of possible hyperparameter values and compute the estimation error of the corresponding GP models on the training data. This *training error*

is computed by building a GP model using the training data, and comparing the output values estimated by the GP model for the training data inputs with the actual outputs included in the training data. The hyperparameter value that results in the lowest error, or equivalently the highest accuracy, is chosen for subsequent experimental studies. Section 5 illustrates this approach to compute a suitable value for the hyperparameter.

5. Experimental setup

The experiments performed for this project were implemented using the WEKA open source machine-learning library (Hall et al., 2009). WEKA includes Java implementations of popular machine-learning algorithms such as GP, SVM, linear regression and multi-layer perceptron (ANN). The library also has evaluation schemes that can be used to compare performance of different algorithms over different datasets. We adapted the existing implementation of GP to fit our needs. Our application first reads in the training data and trains different GP models corresponding to different values of the hyperparameter (γ). As described in Section 4, γ is selected based on the GP model that results in the lowest RMSE between the estimated reference ET and actual TXHPET reference ET over the training data. This GP model is chosen for further use as it is the most accurate model. The RMSE statistic was used because it represents the actual difference in ET in mm. Fig. 4 illustrates this approach to compute the value of γ for the data obtained from the paired non-ET station and TXHPET station in Lubbock. A linear regression (LR) model was also learned from the same training data to serve as a baseline for comparison. The estimation accuracy of learned GP models was then compared with the estimation accuracy of LR models on separate test data for each non-ET station. For instance, Fig. 4 also shows that the GP models result in significantly lower RMSE in comparison with the LR models. It is also possible to automatically select the best value of γ by computing error measures over a separate *validation* set (Bishop, 2008).

To ensure accurate estimates from the learned models, the test data must be drawn from the same space as the training data, i.e., the probability distributions underlying the datasets must be equivalent. Consider the Lubbock datasets over the years 2001–2010, and consider the data division scheme that uses data from 2001 to 2005 as the training set and data correspond to years 2006–2010 as the test set. Such a data division scheme will not
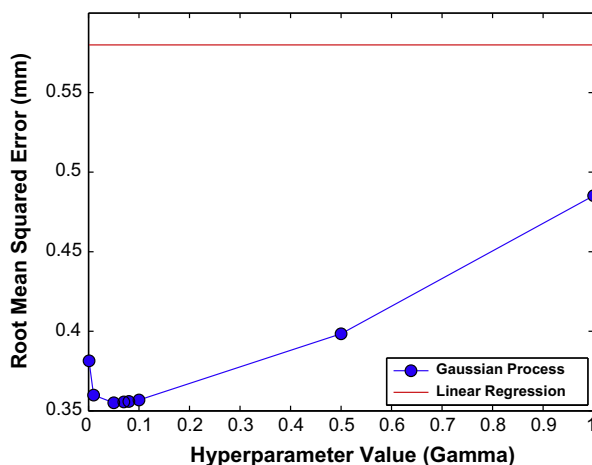
work because the wet and dry years are typically inconsistent—the models trained with data corresponding to dry years will result in high errors on data corresponding to wet years. We therefore split the data evenly across all years by using odd days for training and even days for testing. Although such a division of data into training set and testing set makes it difficult to run a standard cross-validation analysis, we repeated the experiments after swapping the training and test datasets. Overall, we conducted experimental trials using data from 15 non-ET (i.e., NWS) weather stations matched with the TXHPET stations determined by the Thiessen polygon (Fig. 2). We used data consisting of daily measurements over a period of 10 years.

Experimental trials were divided into two groups based on the inputs used to train GP models. In the first set of experiments, reference ET values were computed from non-ET station weather parameters (see Section 3) and used as inputs to GP models—each input is thus a single value. The corresponding GP models capture the relationship between these reference ET values from non-ET stations and the corresponding reference ET values from the paired TXHPET station. In the second set of experiments, weather parameters from non-ET stations were used as inputs to the GP models—each input is thus a vector of weather parameter values. The target outputs were the TXHPET reference ET values. We hypothesized that GP models trained in the second set of experiments would provide more accurate estimates because they can model and account for the uncertainty in computing reference ET from the unreliable observations of weather parameters at non-ET stations.

For each station, the GP model was trained using training data (half the values from the total number of years for each station) and various values for the hyperparameter γ. Once a suitable GP model is selected for further use (as described above), error statistics are computed using the reference ET estimates provided by this trained model over the test set and the actual reference ET values from the paired TXHPET station(s). Measures used for comparison include R^2, NSE, and RMSE.

6. Experimental results

Fig. 5 summarizes the estimation capabilities of the linear regression and Gaussian process models using the R^2 measure. Models that provide highly accurate estimates will result in points that lie on (or very close to) the $Y = X$ line. Fig. 5(a) and (b) shows results (for Lubbock non-ET and TXHPET stations) with LR models and GP models (respectively) that used the reference ET computed from non-ET station as inputs. Similarly, Fig. 5(c) and (d) shows results with LR models and GP models (respectively) that used the weather parameters recorded at the non-ET station as inputs. We observed that the estimates were more accurate when the weather parameters were used as inputs instead of the reference ET computed at the non-ET stations. This observation was true for both LR and GP models, and similar plots were obtained for other stations included in the experimental trials. As hypothesized in Section 5, using the weather parameters as inputs enables the learned models to account for the uncertainty in observations of weather parameters at the non-ET stations. In other words, the models were able to capture the correlations in the data more accurately. The results reported below therefore correspond to experiments in which the weather parameter measurements at non-ET stations were used as inputs to train and test the models.

Another key observation based on the results in Fig. 5 is that the GP models in Fig. 5(b) and (d) result in much greater accuracy than the corresponding LR models. Furthermore, the best performance (i.e., most accurate estimates) were obtained using GP models trained using the weather parameters as inputs, as shown in Fig. 5(d) where most points are along the diagonal line. Similar

Fig. 4. RMSE of GP models trained using data from the paired non-ET station and TXHPET station in Lubbock for different values of the hyperparameter (γ). GP model with lowest RMSE can be selected automatically from training data and used to estimate outputs for the test data. All GP models perform much better than LR models.

(a) LR estimates using calculated NWS ET

(b) GP estimates using calculated NWS ET

(c) LR estimates obtained using individual weather parameters

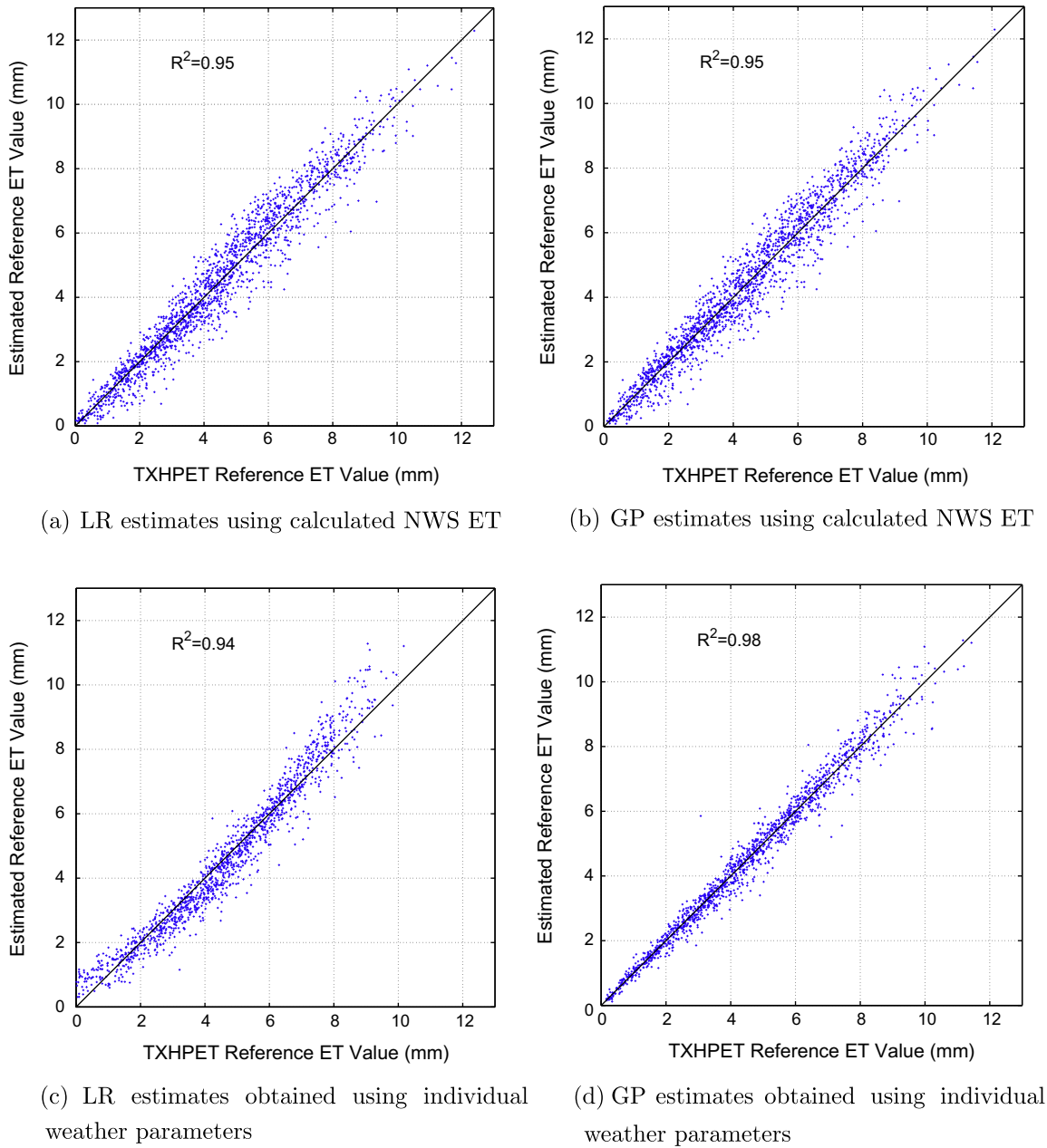(d) GP estimates obtained using individual weather parameters

Fig. 5. Comparison of estimation capability of GP models and LR models for the two sets of experiments using Lubbock-NWS and Lubbock TXHPET data.

results were obtained using the data from other stations included in the study.

Fig. 6 summarizes the results obtained on test data from 15 non-ET (NWS) weather stations, using RMSE as the performance measure. We observe that GP models provide much lower RMSE in comparison with the LR models. In other words, the reference ET values estimated by GP models are much closer to the reference ET values obtained from the corresponding TXHPET stations. This performance improvement is statistically significant and GP models show significant promise in enabling the use of alternative data sources for accurately computing reference ET values.

Although the improvement in the accuracy of GP models (compared with the LR models) is different at different stations, the improvement is significant in all stations considered in our study. Stations such as Lubbock and Dalhart produced highly accurate estimates: $R^2 = 0.98$ and 0.98 respectively and NSE = 0.98 and 0.98, whereas matching the non-ET (NWS) station at Lubbock with

the Farwell (TXHPET) station obtained $R^2 = 0.89$, NSE = 0.89 and RMSE of 0.76 mm for daily reference ET values. This represents $\approx 29\%$ error which is still significantly better than the LR models that result in an RMSE of 0.84 with a relative error of 33%. Future research could consider additional features for input (e.g., elevation) and other GIS selection methods for matching non-ET stations with TXHPET stations. For instance, although our analysis identifies a good correlation between the Farwell TXHPET station and the Lubbock non-ET station based on the Thiessen polygon map, including additional features may help identify stations that are strongly correlated.

The GP models provide more accurate estimates of reference ET than LR models, and this improvement in accuracy has significant practical value. The average difference in RMSE for daily reference ET estimates provided by GP and LR models is $\geqslant 0.2$ mm. For a typical cropping season of about 200 days, this amounts to approximately 40 mm or 1.5 in. over the season. Although this difference
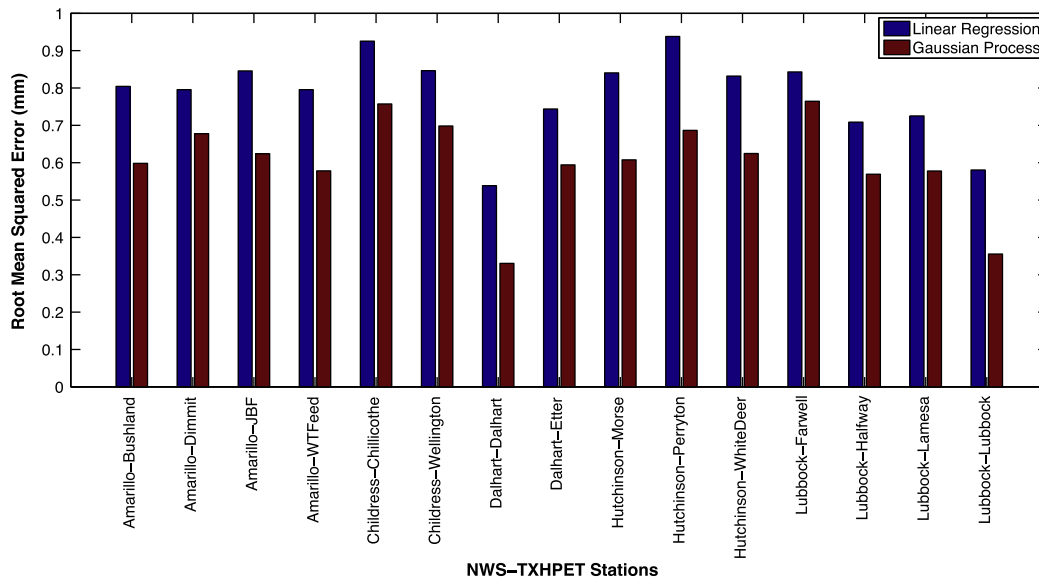
Fig. 6. Comparison of RMSE obtained with the GP models and LR models, with the results averaged over data from each of the non-ET stations used in the study. The GP models result in much lower RMSE compared with LR models.

Table 1
Performance measures for estimates obtained from the LR models and GP models at each non-ET station used in this study. The GP models provide higher accuracy than LR models in estimating reference ET.

NWS – TXHPET station	Linear regression			Gaussian process		
	R^2	NSE	RMSE (mm)	R^2	NSE	RMSE (mm)
Amarillo – Bushland-ARS	0.90	0.89	0.80	0.95	0.95	0.60
Amarillo – Dimmit	0.90	0.88	0.80	0.92	0.92	0.68
Amarillo – Bushland-JBF	0.90	0.89	0.85	0.95	0.95	0.62
Amarillo – West Texas A&M Feedlot	0.90	0.89	0.80	0.95	0.95	0.58
Childress – Chillicothe	0.87	0.85	0.93	0.91	0.91	0.76
Childress – Wellington	0.88	0.87	0.84	0.92	0.92	0.70
Dalhart – Dalhart	0.95	0.95	0.54	0.98	0.98	0.33
Dalhart – Etter	0.92	0.90	0.74	0.95	0.94	0.59
Hutchinson – Morse	0.90	0.89	0.84	0.95	0.95	0.61
Hutchinson – Perryton	0.88	0.87	0.94	0.94	0.94	0.69
Hutchinson – White Deer	0.90	0.89	0.83	0.94	0.94	0.62
Lubbock – Farwell	0.87	0.85	0.84	0.89	0.89	0.76
Lubbock – Halfway	0.91	0.90	0.71	0.94	0.94	0.57
Lubbock – Lamesa	0.91	0.90	0.73	0.94	0.94	0.58
Lubbock – Lubbock	0.94	0.94	0.58	0.98	0.98	0.36

may seem rather small, one acre-inch of water for all the fields of the Texas High Plains results in approximately 24.8 billion gallons of wasted water (Marek et al., 2010), which can be compared to the amount of water supplied to the entire city of Houston for about two and a half months!

Table 1 summarizes the performance of the LR models and GP models at each non-ET station included in our study, using the performance measures (R^2, NSE and RMSE) described in Section 3. With the GP models, each station's R^2 and NSE values are closer to 1 with a lower RMSE. The results show that GP models provide higher accuracy in estimating reference ET than LR models.

7. Conclusion

Efficient water resource management represents a pressing need in agriculture. Accurate estimates of crop evapotranspiration (ET) are essential for irrigation management, especially in arid and semi-arid regions where crop water demands exceed rainfall. Existing ET stations do not provide the required areal coverage and also face funding challenges. This paper presented the results

of a study conducted towards our long-term goal of using data from non-ET stations for filling data gaps in the ET networks.

In the context of data collected in the Texas High Plains, we described the use of Gaussian process models, an instance of sophisticated kernel-based machine learning, to estimate the daily reference ET values based on the corresponding data obtained from National Weather Service stations. Our experiments show that GP models result in significantly more accurate estimates of daily reference ET values than the (popular) linear regression models. We also observe that using the daily weather parameter measurements from the non-ET stations as inputs (instead of the reference ET computed from the these measurements) results in more accurate estimates. The improvement (provided by the GP models) in accurately estimating the daily reference ET values addresses a critical need and has significant practical value. Errors in reference ET estimates can translate to huge costs associated with wasteful use of precious water resources (due to over-watering), in addition to crop stress and even crop loss (due to under watering).

Although our study focused on the Texas High Plains, the models and experimental methodology can be adapted to regions elsewhere in the world. Furthermore, Gaussian process models and

other similar stochastic machine learning algorithms are generic tools for classification and regression in high-dimensional input spaces, with significant potential for addressing key open challenges in water resources management and other sub-fields of agriculture.

USDA EEO disclaimer

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720–2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, DC 20250-9410, or call (800) 795–3272 (voice) or (202) 720–6382 (TDD). USDA is an equal opportunity provider and employer.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements. Technical Report 56. Food and Agriculture Organization of the United Nations. Rome. <http://www.fao.org/docrep/X0490E/X0490E00.htm>.

Allen, R.G., Walter, I.A., Elliot, R., Howell, T.A., 2005. ASCE-EWRI Standardization of Reference Evapotranspiration. American Society of Civil Engineers – Environmental Water Institute.

ASCE Task Committee on Applications of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology I: preliminary concepts. J. Hydrol. Eng. 5, 115–123.

ASCE Task Committee on Applications of Artificial Neural Networks in Hydrology, 2000b. Artificial neural networks in hydrology II: hydrologic applications. J. Hydrol. Eng. 5, 124–137.

Bishop, C.M., 2008. Pattern Recognition and Machine Learning. Springer-Verlag, New York.

Buishand, T.A., Brandsma, T., 1997. Comparison of circulation classification schemes for predicting temperature and precipitation in the Netherlands. Int. J. Climatol. 17, 875–889.

Buscema, M., Sacco, P., 2000. Feedforward networks in financial predictions: the future that modifies the present. Expert Syst. 17, 149–169.

DeRoach, J.N., 1989. Neural networks: an artificial intelligence approach to the analysis of clinical data. Aust. Phys. Eng. Sci. Med. 12, 100–106.

Dodson, R., Marks, D., 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. Clim. Res. 8, 1–20.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. 228, 113–129.

Gowda, P.H., Ennis, J.R., Howell, T.A., Marek, T.H., Porter, D.O., 2012. The ASCE standardized equation based reference ET calculator. In: Proceedings of the World Environmental and Water Resources Congress. ASCE-EWRI, Albuquerque, NM, pp. 2198–2205.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. 11.

Higdon, D., Lee, H., Holloman, C., 2003. Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems. In: Bayesian Statistics. Oxford University Press, p. 181.

Holman, Daniel, et al., 2013. Estimating reference evapotranspiration for irrigation management in the texas high plains. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. AAAI Press.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feed forward network are universal approximators. Neural Networks 2, 359–366.

Kennedy, M., O'Hagan, A., 2001. Bayesian calibration of computer models. J. Roy. Stat. Soc. B 63 (3), 425–464.

Knapp, P.A., 1992. Correlation of 700-mb height data with seasonal temperature trends in the great basin (Western USA). Clim. Res. 2, 65–71.

Krause, A., Singh, A., Guestrin, C., 2008. Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies. J. Mach. Learn. Res. 9, 235–284.

Li, J., Richter, D.D., Mendoza, A., Heine, P., 2010. Effects of land use history on soil spatial heterogeneity of macro- and trace elements in the Southern Piedmont USA. Geoderma 156, 60–73.

López-Granados, F., Jurado-Expósito, M., Pena-Barragan, J.M., Garcia-Torres, L., 2005. Using geostatistical and remote sensing approaches for mapping soil properties. Eur. J. Agron. 23, 279–289.

Marek, T.H., Porter, D.O., Gowda, P.H., Howell, T.A., Moorhead, J.E., 2010. Assessment of Texas Evapotranspiration (ET) Networks. Publication AREC 201011-12, Texas AgriLife Research, Amarillo, TX, 367.

Marrel, A., Iooss, B., Laurent, B., Roustant, O., 2009. Calculations of sobol indices for the Gaussian process metamodel. Reliab. Eng. Syst. Saf. 94, 742–751.

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Musavi, M.T., Ahmed, W., Chan, K.H., Faris, K.B., Hummels, D.M., 1992. On the training of radial basis function classifiers. Neural Networks 5 (4), 595–603.

Porter, D.O., Marek, T.H., Howell, T.A., 2005. The Texas High Plains Evapotranspiration Network (TXHPET) User Manual. (version 1.01), AREC 05-37. Texas A&M System.

Porter, D.O., Gowda, P.H., Marek, T.H., Howell, T.A., Moorhead, J.E., Irmak, S., 2012. Sensitivity of grass and alfalfa reference evapotranspiration to weather station sensor accuracy. Appl. Eng. Agric. 28 (4), 543–549.

Rumelhart, D.E., Hinton, D.E., Williams, R.J., 1986. Learning Internal Representations by Error Propagation. MIT Press, Cambridge, MA, USA.

Song, X., Zhan, C., Kong, F., Xia, J., 2011. Advances in the study of uncertainty quantification of large-scale hydrological modeling system. J. Geograph. Sci. 21, 801–819.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer.

Williams, C.K.I., Barber, D., 1998. Bayesian classification with gaussian processes. IEEE Trans. Pattern Anal. Mach. Intell. 20 (12), 1342–1351.

Williams, C.K.I., Rasmussen, C.E., 1996. Gaussian processes for regression. In: Advances in Neural Information Processing Systems. MIT Press, pp. 514–520.