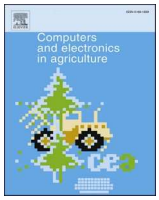
Computers and Electronics in Agriculture 205 (2023) 107608



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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





IoT based dynamic Bayesian prediction of crop evapotranspiration in soilless cultivations

Alexander Kocian [a,*], Giulia Carmassi [a], Fatjon Cela [a], Stefano Chessa [b], Paolo Milazzo [b], Luca Incrocci [a]

[a] Department of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, 56124, Pisa, Italy
[b] Department of Computer Science University of Pisa, Pisa, Italy

ARTICLE INFO

Keywords:
Internet-of-Things
Quality-of-Service
Dynamic modelling
Crop coefficient
Sweet basil

ABSTRACT

As water resources become scarce around the globe, a sustainable irrigation management is essential for increasing the crop use efficiency. Crop EvapoTranspiration (ET) is considered as one of the fundamental variables for correct irrigation management. Current ET prediction models adjust their system parameters: i) during time consuming training cycles; ii) by (destructive) leaf-area measurements; iii) only once during training for a particular meteorological scenario. These parameters are static. They cannot learn over time. On top of that, these models fail when some of the input variables are missing. To overcome these shortcomings, we follow an Internet of Things (IoT) approach to crop water use modeling and prediction for soilless cultivations. First, a stochastic crop coefficient confined to global radiation, Kr, is defined. Then, the evolution of Kr is modeled as state-space system with the daily temperature as input data, and global radiation and crop weight as measurement data. The latter data may also be missing at random. We derive a predictor for the posteriori probability of Kr and show that predicting crop ET via Kr it is immune to propagation of uncertainty over time due to erroneous weather forecasting. Our predictor is implemented in the cloud server of the IoT network. The sensor nodes, connected to a lightweight IoT client, supply our model with information in (soft) real-time. A web interface allows for visualization of the time series in real-time. To discuss Quality of Service (QoS) issues, we model the IoT network as synchronous data flow graph revealing its workflow and timings of the individual nodes.

Sweet basil (*Ocimum basilicum* L.) is used to test the performance in four climatic environments and two different sampling methods: i) sample daily from the same environment in parallel; ii) sample daily from one environment and move on to the next in round robin fashion, emulating reuse of the same set of measurement equipment at a quarter of the costs. It turns out that in both cases, our Kr-based crop ET predictor outperforms the competing Baille technique. The performance of our IoT network is assessed by means of the QoS metrics response time, jitter, and availability.

1. Introduction

Population and economic growth, along with increased climate variability puts water resources under pressure. Focusing on Europe, annual renewable freshwater resources are generous, amounting to a long-term average of 4,560 m^3 per capita. For example, Italy is the third largest agricultural producer contributing to the highest value-added agriculture in the European Union (CREA, 2021) but ranked 84th among 180 countries based on internal freshwater resources per capita according to AQUASTAT. Hence, irrigation management is to be optimized in soil and soilless cultivation. To improve control over the growing environment, we subsequently focus on soilless culture cultivation.

Sweet basil (*Ocimum basilicum* L.) is an annual aromatic herbal plant belonging to the family of Lamiaceae. Basil is an aromatic plant commonly used in the Mediterranean cuisine (Makri & Kintzios, 2008). Due to its high content of essential oils, phenolic compounds, flavonoids (Jayasinghe et al., 2003) and antioxidant activities, basil has also many applications in medicine (Puccinelli et al., 2020). However, little is known about its water requirements in greenhouse soilless culture (Khater et al., 2021).

Crop evapotranspiration (ET) is considered as one of the

* Corresponding author.
E-mail address: kocian@di.unipi.it (A. Kocian).

https://doi.org/10.1016/j.compag.2022.107608
Received 9 August 2022; Received in revised form 14 December 2022; Accepted 27 December 2022
Available online 5 January 2023
0168-1699/© 2022 Elsevier B.V. All rights reserved.

fundamental variables for irrigation management. Many models for ET estimation have been developed over the last thirty years. These models are commonly classified as either "one-step" or "two-step" approach (Shuttleworth, 2006). The first allows direct estimation of ET based on the Penman-Monteith (P-M) equation (Monteith, 1965). The latter is based on defining a reference crop ET and then using an experimental crop coefficient, to quantify the difference between the two crops (Allen et al., 1998). Virtues and vices of both approaches are evident. The P-M equation is universally accepted by the scientific community as the theoretically superior one-step approach. The P-M equation depends on aerodynamic and canopy resistances relative to meteorological controls which are, however, cumbersome to obtain. Hence, the P-M equation is scarcely used, in practice. For applications in greenhouses, (Baille et al., 1994). proposed a simplified version of the P-M equation that only depends on the meteorological parameters Global Radiation (GR) and Vapor Pressure Deficit (VPD) plus the Leaf Area Index (LAI) of the crop. Further simplifications of the P-M equation were proposed subsequently (Bacci et al., 2011; Carmassi et al., 2007). Nevertheless, all these P-M based equations need to be calibrated before their use. Specifically, knowledge of weather-related data, (destructive) LAI and crop ET measurements are required during a minimum of two growing cycles, consequently consuming time and resources. During validation phase, LAI is then assumed to be known i.e., perfectly estimated, which is the hitch of this approach. The FAO-56 version of the P-M ET equation is a practical two-step approach (Allen et al., 1998) comprising the reference crop ETo of a well-watered hypothetical grass times an empirical single crop-coefficient Kc = Kc(t) curve, incorporating crop characteristics and its development stage. During calibration, the former factor can be easily calculated from data on temperature, relative humidity, and GR. The latter, behaving similar to the LAI over time, can be simply computed as the ratio of actual crop ET to the previously obtained reference ETo. Before the previously obtained Kc curve is ready for crop ET estimation, however, we must not only identify the time point in the respective crop stage but also adjust the Kc curve for climatic conditions, limiting the merit of this two-step approach. Some authors (Fernández et al., 2011; Gallardo et al., 2013) reviewed application of the FAO-56 method to Mediterranean greenhouse conditions. Their study on ETo estimation for perennial grass using drainage lysimeter shows that the FAO-56 P-M ET equation underestimates the reference ET by 21 %. Simpler methods for estimating ETo, restricting their input variables to the daily min/max temperature (Hargreaves & Samani, 1985) or, the daily sum of external solar radiation values (Almeria Radiation method), yield similar accuracy. Ergo, there is still lack of a simple yet accurate method for fast crop ET estimation, especially when it comes to sweet basil.

The key to effective model prediction is to approximate the complex dynamics of the underlying dynamics as *stochastic* process. The unsupervised machine learns the model autonomously without external training data, which is time efficient but hard to achieve. The supervised machine builds a statistical model from training data. Once trained, the machine can make repeatable predictions. The model parameters are static. They do not change over time. Following the latter approach, (Goap et al., 2018) propose the use of support vector regression machine for crop ET prediction. Recently, Artificial Neural Networks (ANNs) comprising many simple connected nodes dubbed neurons can be used to predict parameters related to precision farming (Balducci et al., 2018). Among the many types of available ANNs, it is worth pointing out the recurrent neural network (RNN). Its feedback structure allows for space–time prediction of system parameters such as the LAI (Chen et al., 2015). Though the prediction accuracy is promising, these ANNs require large amount of data along with the ground truth for training, which requires manual work of experts to be obtained. To overcome this drawback, recent research has focused on semi supervised machine learning methods. Starting from a few training points, the underlying system model exploits forthcoming observations in the same data frame to adjust itself as time passes thanks to modern information and communication technology. For example, the dynamic Bayesian network in (Kocian et al., 2020a; Kocian et al., 2020b) models the time-series of crop ET values as Markov process with the weather-related parameters as (known) disturbance. This approach, however, yields good results only when future weather events are anticipated correctly. This is not always the case, especially in the Mediterranean winter and spring time, characterized by strong variability in the climate.

With the development of Internet-of-Things (IoT), machine learning applications for precision farming have been grown rapidly over the past years (Mekonnen et al., 2020). IoT senses huge amount of heterogeneous data from smart sensors and devices ("things") related to crop, soil, and weather. The resulting data stream is transmitted over the wireless channel to some cloud server for advanced analytics, real-time visualizations, and machine learning. The latter feeds the mathematical crop water model. An agent, making smarter predictions over time, might return control information to the actuators in the greenhouse (Burchi et al., 2018). In such a closed-loop system, sensor data might feed in the forward loop a machine learning algorithm. The predicted nominal soil moisture from weather forecast acts as control information in the feedback loop. The sprinkler starts at a predefined minimum soil moisture and stops at the returned nominal value (Goap et al., 2018). The controller might also choose one out of many tasks in order to maximize the cumulative reward (Bu & Wang, 2019). Clearly, each task (such as the sprinkler) in the IoT network must respond within a predefined deadline, on average at least. Quality of Service (QoS) in the field of IoT manages system capabilities and its resources but has gained only low attention until recently (Brogi & Forti, 2017). One way to handle the performance of IoT networks is to model them as synchronous data flow (SDF) graph that is transformed into a directed acyclic workflow graph, revealing the timing of individual tasks (Kocian & Chessa, 2022) QoS aware IoT has not entered as an application in smart farming, yet.

In this contribution, we follow a dynamic Bayesian approach to modeling a crop coefficient confined to radiation and derive a predictor for its a posteriori probability over time. Input data from only three Commercial Off-The-Shelf (COTS) sensors, measuring temperature, global radiation and crop weight, are required. Treating GR as parameter rather than a variable, a guess of crop ET can be retrieved directly from the predicted time-series by scaling. The result is immune to errors from weather forecasting in contrast to common predictors that directly operate on crop ET. Neither time consuming training cycles nor destructive LAI measurements are necessary for this model to work correctly. For the hardware realization, a few simple sensors, measuring temperature, global radiation, and weight balance, are connected to a lightweight IoT client, operating in soft real-time. Measurement data is forwarded to the IoT cloud where this model is implemented. To get basic insight into the end-to-end QoS behavior, we represent the soft real-time IoT network as synchronous data flow graph. The performance of our model is tested for sweet basil and compared with that of the common (static) Baille technique across climates and different data sampling techniques. The performance of our IoT network is assessed by means of the QoS metrics response time, jitter, and availability. The results are compared with the respective requirements.

2. Evapotranspiration models

2.1. Static standard approach

We preliminary recap a standard approach to static ET estimation in greenhouse horticulture. It is shown in (Carmassi et al., 2007) that the daily average $\overline{ET}$ [mm d^{-1}] can be approximated by a linear combination of the daily global radiation $\overline{GR}$[MJ m^{-2} d^{-1}], daily *LAI* [1] and latent heat of water vaporization λ [MJ kg^{-1}],

$$\overline{ET} = a(1 - exp\{-k \times LAI\}) \frac{\overline{GR}}{\lambda} + b. \quad (1)$$

Finally, k is the light extinction coefficient of the canopy, which is important for an appropriate partitioning of radiant energy between crop canopy and soil surfaces. The (static) coefficients a [1] and b [mm d^{-1}] are the regression coefficients that are to be computed by means of linear regression analysis during some training cycle. On top of that, the expression for $\overline{ET}$ in (1) requires knowledge on daily LAI of the crop. The evolution of the latter can be obtained by (destructive or non-destructive) field measurements or regression models which we follow subsequently. A standard approach to LAI modeling requires a nonlinear, typically exponential, logistic, or Gaussian, regression with the cumulative Growing Degree Days (GDD) as explanatory variables (Thornley & Johnson Ian, 1990),

$$LAI = c\exp\left\{d\sum_t GDD_t\right\} \tag{2}$$

from planting to harvesting. The (static) regression coefficients c and d are obtained by measurement during the same training cycle. Note that the expression for $\overline{ET}$ in (1), dubbed "simplified Baille model", is valid for the calibrated crop type within the training range of temperature and global radiation, only. The regression coefficients are static; they do not adapt over time.

2.2. Dynamic Bayesian approach

Most of current prediction models for evapotranspiration consider the crop ET as figure of merit. Following this approach, ET can be predicted with high quality if future weather events can be anticipated correctly. This is often cumbersome when weather changes quickly, especially in the (Mediterranean) winter season. To overcome this shortcoming, we consider

$$Kr_t \triangleq \frac{\lambda \times \overline{ET}_t}{\overline{GR}_t}. \tag{3}$$

a (dimensionless) crop coefficient Kr, confined to global radiation at time instant t, as figure of merit. From the simplified Baille model in (1), it can be readily seen that the Kr resembles the LAI which is related to the growing season rather than weather phenomena. The way in which LAI is modeled in (2) suggests that the Kr evolves over time according to

$$Kr_t = \nu + \kappa e^{\mu(t-t_0)} \tag{4}$$

with the system parameters ν, κ, t_0, and μ. Note that for leafy plants, such as basil, the evolution of the Kr is near-linear (Guerra et al., 2016) i.e., $\mu \approx 0$. Sampling (4) with granularity Δ, it follows

$$\frac{Kr_{t+1} - Kr_t}{\Delta} = \mu \times Kr_t - \nu\mu \tag{5}$$

Slightly rearranging (5), we have

$$Kr_{t+1} = (1+\mu\Delta)Kr_t + \nu\mu\Delta. \tag{6}$$

Note that above expression is solution to a first-order linear difference equation. Let $\boldsymbol{Kr}$ be a time-discrete Markov process containing the random variables $Kr_1 \ldots Kr_T$ that cannot be observed directly, but only through noisy measurement $\boldsymbol{y} = \overline{ET}/\overline{GR}$. Then, the probability of moving from state Kr_t to state Kr_{t+1} is given by

$$p(Kr_{t+1}|Kr_t) = \mathcal{N}(Kr_{t+1}; A \times Kr_t + Bu_t, \boldsymbol{\Sigma}_n) \tag{7}$$

$t = 1 \ldots T$. Here, $\mathcal{N}(-; m, \Sigma)$ denotes a Gaussian distribution of the first argument having mean m and covariance. Moreover, $A, B \in \mathbb{R}$ are the system parameters. The deterministic (known) control data u_t comprises the ambient temperature at time instant t. The conditional probability of measuring y_t given state Kr_t has the form

$$p(y_t|Kr_t) = \mathcal{N}(y_t; C \times Kr_t, \boldsymbol{\Sigma}_w) \tag{8}$$

$t = 1 \ldots T$. Here, $C \in \mathbb{R}$ denotes a system parameter. For the distribution of the initial state, we have

$$p(Kr_1) = \mathcal{N}(Kr_1; \mu_1, \boldsymbol{\Sigma}_1) \tag{9}$$

The state-space system (7) - (9) corresponds to a Dynamic Bayesian Network. The iterative Expectation-Maximization (EM) algorithm is used to track the posterior state sequence as well as the set of system parameters $\boldsymbol{\theta} = \{A, B, C, \Sigma_n, \Sigma_w, \mu_1, \Sigma_1\}$. Mean and covariance of the free running q-step predictor without response are, after a few straightforward algebraic operations,

$$\widehat{Kr}_{T+q} = \widehat{A}\widehat{Kr}_{T+q-1} + \widehat{B}u_T\widehat{V}_{T+q} = \widehat{A}\widehat{V}_{T+q-1}\widehat{A}^T + \widehat{\Sigma}_n, \tag{10}$$

respectively. In the derivation of (10) we treated missing data at random (Rubin, 1976). The hat operator on the respective symbols indicates their estimate at iterative convergence.

We modeled Kr as stochastic process that account for variations in climate. As these variations are small compared to weather variations, we expect that our Kr approach predicts crop evapotranspiration with much higher accuracy than the Baille based ET approach does.

3. The IoT network

Usually crop water models require input data. We could measure the parameters manually in the field. This approach, however, has several hitches ranging from data storage to security issues. Scheduling of manual measurement 24/7 itself is a challenge. We therefore follow an automatic approach using IoT condition monitoring. Also, our crop water models shall be updated automatically at the end of the day.

3.1. Requirements

The IoT platform shall meet the following requirements:

- **Autonomy**: The IoT network can carry out its tasks alone without human interaction.
- **Consistency**: The sensed data must be reproducible.
- **Economic**: COTS components shall be used. A standard Lithium-ion Polymer battery (3.7 V DC, 1800mAh) shall be sufficient to power the IoT client.
- **Memory**: The executable must fit in the flash memory of the IoT client, which is about 100kB.
- **Real-Time**: Real-time support opens up the possibility of using threads, timers and pre-emptive multitasking.
- **Backup and restore**: To recover from failure, the IoT application must be able to back up its state during normal operation and restore its last state after reboot from error.
- **Robustness**: The IoT application must be able to recover autonomously from error.

3.2. System design

Following this approach, Fig. 1 shows a high-level concept of the system architecture. A few sensors are connected to the Particle™ microcontroller with cellular capability, forming an embedded IoT client. Once a while, the micro-controller wakes up, reads the sensor signals, processes their data and pushes the result to the cloud using event driven webhooks, and then goes to sleep again. To ensure data privacy over air transmission, the Constrained Application Protocol (CoAP) with Datagram Transport Layer Security (DTLS) lets our sensors communicate on the IoT. In the cloud, the ThingSpeak™ IoT platform feeds our machine learning algorithm using MATLAB™ functions and toolboxes.

The firmware in the IoT client is subdivided into tasks that are scheduled synchronously in a periodic fashion by the real-time operating system FreeRTOS, maintaining certain quality-of-service (QoS) functionalities. Though FreeRTOS is deterministic, the communication with all the sensors occurs asynchronously in separate threads, introducing uncertainty in the tasks' response time. On top of that, delay

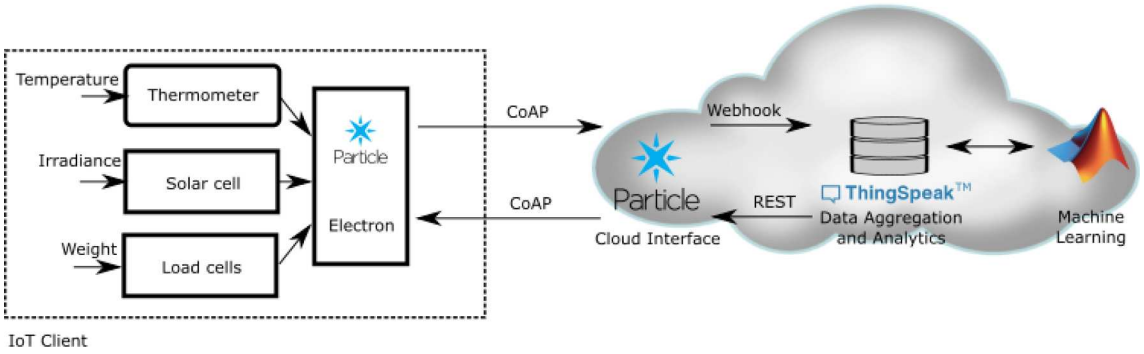


Fig. 1. IoT-based system architecture comprising sensors connected to a client with cellular connectivity and a cloud platform performing data aggregation, analytics, and machine learning.

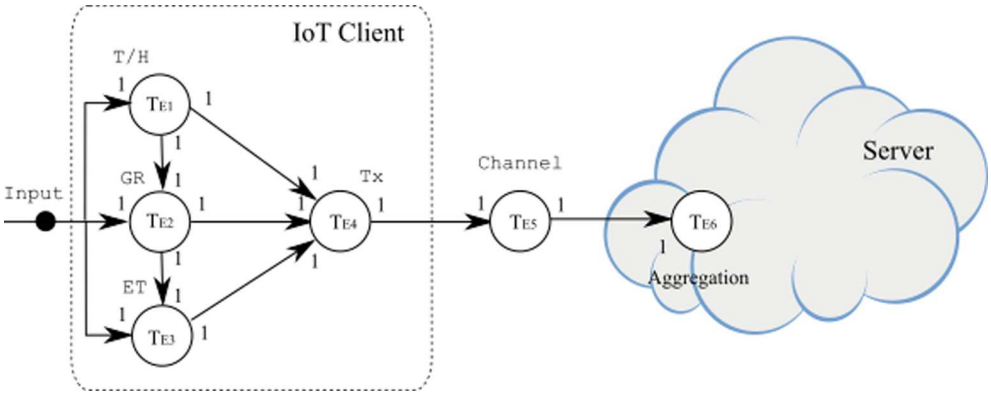


Fig. 2. SDF graph composed of one IoT client connected to one cloud server during one period of execution.

caused by the wireless channel is inherently stochastic, as well. As each task consumes and produces data samples, the concept of Synchronous Data Flow (SDF) with random execution times is ideally suited for workflow modelling of our application (Kocian & Chessa, 2022). The SDF graph for one period of execution is depicted in Fig. 2. Each node in the graph represents one independent task a_n having (stochastic) execution time TE*n*. Edges in the graph indicate the dependence among the nodes. The number on the stem and the arrow on each edge indicate the respective incoming and outgoing rates.

Inspecting Fig. 2, the tasks are to be executed in sequential order $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ within an infinite loop, starting with temperature/humidity (`T/H`) sensing node. The second node senses solar radiation (`GR`) sensor. The third node senses the crop's weight to compute its daily evapotranspiration (`ET`). The transmitter (`Tx`) prepares for transmission, and metadata are sent over the wireless channel (`Channel`) to the cloud server where the data is aggregated (`Aggregation`). In the same order the individual nodes' execution times are to be added, to form the execution profile of the entire IoT network. In Section 4, we will use above SDF graph to evaluate several QoS metrics related to our IoT network. Note that our mathematical models are executed asynchronously at the end of the day and hence, are not part of above SDF graph.

3.3. Experimental design

Several experiments were conducted in two subsequent years (2019 and 2020) at the Department of Agriculture, Food and Environment at the University of Pisa, Pisa, Italy (latitude 43°42′ N, longitude 10°24′ E). Ventilation in our greenhouses occurred by openings along the roof edges. Sweet Basil (*Ocimum basilicum* L.) plants were cultivated in plastic pots (18 cm diameter; 3 L of volume) filled with perlite-peat mixture (1:1 v/v) at a density of 12 plants m^{-2}. Sweet basil seeds (Franchi Seeds, Bergamo, Italy) were sown in 240-cell plug-trays filled with rockwool (Grodan®) and vermiculite; the trays were placed in a growth chamber at 23–25 °C and 60–70 % of relative humidity for 4–5 days. Basil plants were put it the greenhouse at four leaves stage (approximately 20 days). The plants were irrigated on daily basis (1–2 min) using drip irrigation with a nutrient solution with the following nutrient ion composition expressed as mol m^{-3}: $N\text{-}NO_3^-$ 10.0, $N\text{-}NH_4^+$ 0.5, $H_2PO_4^-$ 1.0, K^+ 6.0, Mg^{2+} 2.0, Ca^{2+} 4.0, Na^+ 0.5, SO_4^{2-} 3.5, Cl^- 0.5, and as micro-mol m^{-3} Fe^{2+} 40, Cu^{2+} 1.0, Zn^{2+} 5.0, Mn^{2+} 10.0, B 25.0, Mo^{3+} 1.0. The Electrical Conductivity (EC) of the nutrient solution was 2.30 dS m^{-1}, and the pH was adjusted to 5.50 with diluted sulfuric acid. Above experiments were

Table 1
Descriptive statistics of the environmental control parameters for modelling the evapotranspiration (ET) of sweet basil grown in substrate culture under different climatic conditions.

Cycle (start date)	Days of observation	Mean air temperature [°C]	Daily solar radiation [MJ m^{-2} d^{-1}]	Cumulated solar radiation [MJ m^{-2}]
Cycle 1 (15 November 2019)	26	25.77	5.37	139.61
Cycle 2 (10 January 2020)	29	25.85	6.67	193.41
Cycle 3 (10 February 2020)	22	26.81	8.58	188.86
Cycle 4 (05 March 2020)	19	27.57	11.64	221.17

conducted in a warm greenhouse (heating threshold 15 °C) between 15th November 2019 and 23rd March 2020 with four different basil cultivation cycles in four different microclimatic environments, labeled as "Cycle 1" to "Cycle 4"., Table 1 lists the descriptive statistics of the control parameters during the experiments.

For the sensors, we use the Si7021 by Silicon labs to sense temperature and air humidity; the solar radiation sensor by Netsens, Sesto Fiorentino, Tuscany, Italy; and the load cells in form of a Wheatstone bridge, supporting 20 kg. For the IoT client, we deploy the STM32F205 ARM Cortex M3 micro-controller by the Particle™ Electron series. The mathematical models are implemented in Matlab™. The number of iterations in the Expectation Maximization algorithm is limited to 100.

3.4. Calibration issues

3.4.1. Simplified Baille model

The simplified Baille model described in Section 2.1 must be calibrated, to determine the regression parameters *a, b, c* and *d*. To obtain robust parameter estimates that are valid for a large range of environmental parameters, we generated two independent training datasets apart from the actual test dataset. Trials for model calibration were conducted between 1st April 2019 and 14th May 2019 in two different experimental environments. Environment 1 (dubbed "ENV 1") and Environment 2 (dubbed "ENV 2") were a warm greenhouse with a high level of solar radiation and a shaded greenhouse with emergency heating, respectively. Specifically, the daily average of cumulative global radiation and the air temperature were 11.78 MJ m^{-2} day^{-1} and 22.32 °C; 4.27 MJ m^{-2} day^{-1} and 17.21 °C, respectively, in ENV 1 and ENV 2. Table 2 shows the descriptive statistics of the control parameters during training.

A planimeter (DT Area Meter MK2, Delta *T*-Devices, Cambridge, U. K.) has been used to determine the leaf area in both environments. Five plants were randomly chosen on each of the 0th, 21st and 43rd Growing Day after Transplant (GDT). The resulting data samples are shown in Fig. 3, and so is the minimum mean square error LAI curve fit with respect to the GDD for a base temperature of 11 °C (Walters & Currey, 2019).

Following this approach, the optimum regression coefficients in the linear regression of (2) are $c = 0.01281$ and $d = 0.006627$ ($R^2 = 0.98$; $n = 6$). The calibration dataset (n = 86) has then been used to compute the regression coefficients a and b of the ET model in (1) using the calibrated LAI regression along with a light extinction coefficient $k = 0.8$ (Solbach et al., 2021) Following this approach, the optimum regression coefficients in the linear regression of (1) result in $a = 0.4484$ and $b = 0.1259$ mm d^{-1}.

3.4.2. Dynamic Bayesian model

The EM algorithm, in contrast, requires only a few training data points embedded in the measurement data of the current cultivation cycle. However, inappropriate initialization parameters drive the EM to a sub-optimal local maximum of the log-likelihood function (Hu et al., 2004), instead. To obtain a robust initial estimate, though, let us revisit the (non-linear) system model in (4) containing three unknown parameters. Suppose three training points y_1, y_2, and y_3 embedded

Table 2
Descriptive statistics of the environmental control parameters for training the simplified Baille model.

Environment (start date)	Days of observation	Mean air temperature (°C)	Daily solar radiation (MJ m^{-2})	Cumulated solar radiation (MJ m^{-2})
ENV 1 (02 April 2019)	43	22.32	11.68	506.67
ENV 2 (02 April 2019)	43	17.21	4.27	183.51

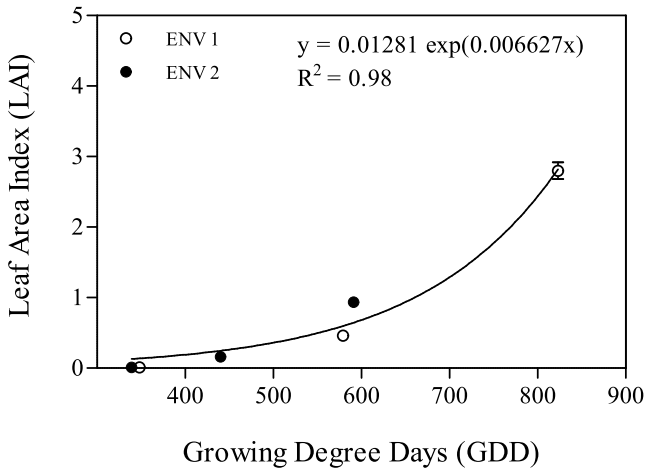


Fig. 3. Relationships between leaf area index (LAI) and growing degree days (GDD) for greenhouse basil grown in soilless substrate. The plants were cultivated in two different climatic environments. Environments ENV 1 (white circle), ENV 2 (black circle).

anywhere in the data frame. Inserting these points in (4) at EM iteration 0, we have

$$\mu^{[0]} = \frac{\log(y_1 - \nu) - \log(y_2 - \nu)}{t_2 - t_1}$$

$$\kappa^{[0]} = \frac{y_3}{1 + e^{\mu^{[0]} t_3}}$$

$$\nu^{[0]} = \kappa^{[0]} \tag{11}$$

as a possible solution to the regression problem. Comparing coefficients between the discrete system model in (6) and the corresponding state-space system in (7)-(9), it follows

$$\boldsymbol{A}^{[0]} = \left(1 + \Delta\mu^{[0]}\right)\boldsymbol{I}$$

$$\boldsymbol{B}^{[0]} = \nu^{[0]}\mu^{[0]}\Delta diag\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t\right)^{-1} \tag{12}$$

The remaining parameters are less crucial for the performance of the EM algorithm and hence, are chosen simply as $\mathbf{C}_0 = \mathbf{I}$, $\sum_n^{[0]} = \sum_w^{[0]} = \varepsilon\boldsymbol{I}$, $\varepsilon \ll 1$.

3.5. Figures of merit

In this section, we define figures of merit that measure quality of the mathematical models as well as QoS of the IoT network.

For the mathematical models, the gap between estimated (E) and measured (M) data is evaluated by the following standard figures of merit (League & Green, 1991):

- Root Mean Square Error (*RMSE*)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(E_i - M_i)^2}{n}}. \tag{13}$$

- Nash-Sutcliffe Model Efficiency coefficient (*EF*)

$$EF = 1 - \frac{\sum_{i=1}^{n}(E_i - M_i)^2}{\sum_{i=1}^{n}(M_i - \overline{M})^2}. \tag{14}$$

- Coefficient of Residuals Mass (*CRM*)

$$CRM = 1 - \frac{\sum_{i=1}^{n} E_i}{\sum_{i=1}^{n} M_i} \quad . \tag{15}$$

The estimated data corresponds to the measured when RMSE, EF and CRM are zero, one, and zero, respectively.

For the IoT network, we focus on the QoS metrics average execution time μ_E [s], jitter σ_E^2 [s^2] and availability in [%]. When the nodes in the graph are executed in successive order, it can be readily shown that the execution time T_E of an IoT application is the sum of the individual execution times. From the linearity property of expectation, the application's expected execution time is the sum

$$\mu_E \equiv \mathbb{E}\{T_E\} = \sum_n \mathbb{E}\{T_{En}\} \tag{16}$$

where $\mathbb{E}\{\bullet\}$ denotes the expectation of the argument. Analogously, the application's variance in execution time, the jitter, is given by

$$\sigma_E^2 \equiv Var\{T_E\} = \sum_n Var\{T_{En}\}\,. \tag{17}$$

For the system availability, we follow the common definition of uptime to runtime ratio in per cent.

4. Experimental results and discussion

Given the environmental parameters in Table 1, we observed the following agronomic crop parameters during the four experiments. The total fresh weight was 225.36 ± 8.96 g $plant^{-1}$, 325.74 ± 15.96 g $plant^{-1}$, 340.46 ± 14.52 g $plant^{-1}$ and 347.98 ± 21.64 g $plant^{-1}$ in Cycle 1, 2, 3 and 4, respectively. The electric conductivity in the drainage water oscillated between 3.56 and 4.11 dS m^{-1} with a mean value of 3.93 ± 0.15 dS m^{-1} during all experiments. The measured cumulative water consumption was 10.05, 16.58, 13.45, and 12.66 mm day^{-1} in Cycle 1, 2, 3 and 4, respectively.

To get started, we investigate the performance the simplified Baille technique in (1). Having trained the regression coefficients during two separate training cycles, we are now ready to deploy the model.

Fig. 4 shows the relationship between estimated and measured crop ET for all four cultivation cycles. A linear regression represented as solid line has also been added to the plot.

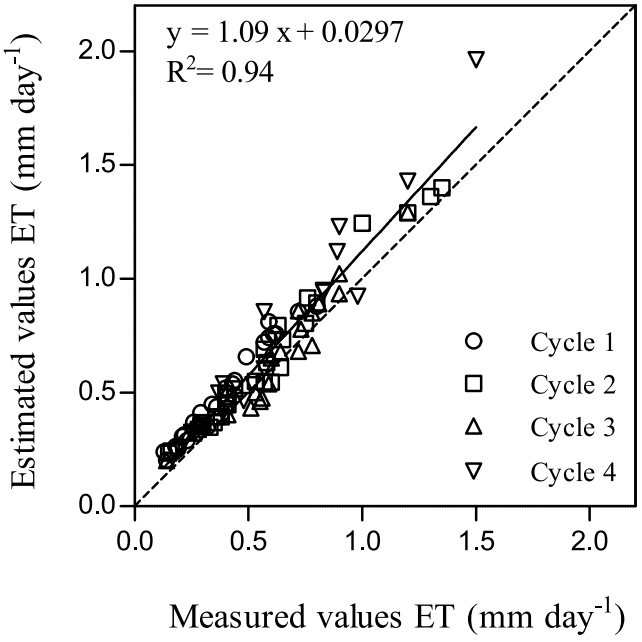


Fig. 4. Relationship between estimated and measured values of crop transpiration (ET) for sweet basil grown in substrate culture in different climatic environments using the simplified Baille technique. The solid line represents the linear regression between estimated and measured ET while the dashed line is the 1:1 relationship; n = 96.

RMSE, EF, and CRM of the linear regression model are given by 0.20, 0.63 and −0.15, respectively. The coefficient of determination (R^2) is 0.94 (n = 96 values) with a slope of 1.09 and a point of intercept of 0.0297, indicating a negligible bias but slight overestimation of ET. Nevertheless, the simplified Baille model overcomes the challenges arising from determination of plant related variables such as leaf resistances. Only (destructive) LAI and weather-related measurement data are required. Finally, it is worth pointing out that the correlation coefficient r between GR and VPD is given by r = 0.69 justifying the missing VPD term in the expression for the crop ET in (1). Models that skip some of the parameters in the P-M ET equation implicitly assume a correlation among those missing (Katsoulas & Stanghellini, 2019).

The performance of our state-space system is evaluated next. To get started, let us plot the time-series of data samples in Fig. 5 for a sampling granularity $\Delta = 1$ day and a prediction length $q = 1$ day. Starting from the planting time, 0 days after planting (DAP = 0), the EM algorithm trains the parameters of our state-space system and tracks its states until DAP = 9. The re-estimated Kr time-series, dubbed "tracking", is depicted as dotted curve in the graph. Beyond DAP = 9, the Kr predictor in (10) is in action. Its mean value is depicted as solid line that is surrounded by shaded confidence intervals of 68 %. The ET time-series provided by the simplified Baille technique model in (1), has been added to the graph as dashed line. The measured crop ET values are represented by circle symbols. To be able to compare our Kr predictor with the ET estimator by Baille, we scaled Kr by the measured $\overline{GR}$ of the respective day and added the result to the plot as dash-dotted line dubbed "Predicted ET".

For all four cultivation cycles, we observe a steadily increasing Kr curve of small curvature but with ripples vs time, mainly due to the variations in ambient temperature. The difference in Kr value reflects the difference in growth state in the respective climatic environment. The predicted ET curve matches the measured to a great extent even when the weather conditions change drastically from one day to another, especially in "Cycle 3". The main reason for this behavior is that any error in the reconstruction of crop ET has no impact on the underlying Kr curve, on which the EM algorithm operates. For the simplified Baille curve, we observe that it correctly catches the trend of crop ET with slight overestimation in Cycle 1 and Cycle 4.

Fig. 6 visualizes the predicted vs measured crop ET data with the cultivation cycle as parameter. The trendline has a slope of 0.99 and a point of intercept of 0.0023 (R^2 = 0.84) indicating high average prediction accuracy over the entire time frame for all growing cycles representing different climatic environments. The statistical figures of merit RMSE, EF, and CRM are equal to 0.108, 0.845, and −0.0019, respectively, for all cultivation cycles, indicating a good model efficiency and a near optimum residual mass. The proposed state-space model clearly outperforms the simplified Baille model.

So far, we investigated the performance during four successive growing cycles requiring 96 consecutive days of measurement. It would be interesting to know the performance during four (pseudo) parallel sessions. This approach would only require one and the same set of measurement equipment for a duration of 26 consecutive days of measurement (see Table 1). Note that the longest cycle determines the overall duration in a parallel system. Following this approach, we start with the same dataset but take only one data sample from each cultivation cycle in round robin fashion. This scenario corresponds to a (virtual) sampling period of $\Delta = 4$ days with missing samples during 3 consecutive days in between. The first three data points of each cycle are used for training so that the EM algorithm may start off between DAP = 9 (Cycle 4) and DAP = 12 (Cycle 1). Fig. 7 shows the Kr and ET times-series using the same line-style and symbols as Fig. 5 does. For comparison purposes, the simplified Baille model from the previous example (ingeniously knowing the environmental parameters occurring every day) has also been added to the plot. We observe that the reconstructed

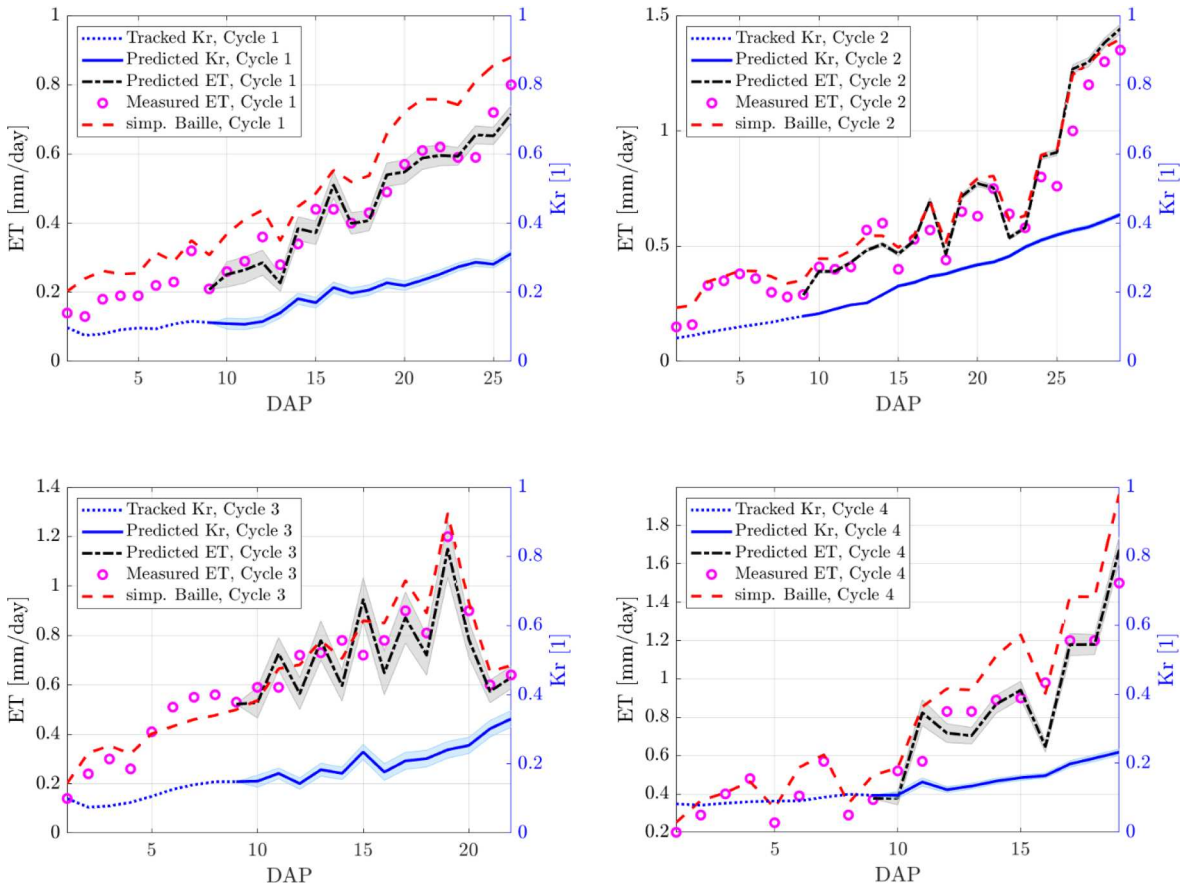


Fig. 5. Comparison between measured (symbols), trained (dotted line), and predicted (solid line) values of daily crop evapotranspiration (ET) for the proposed state-space system, and the simplified Baille technique (dashed line). Sampling granularity $\Delta = 1$ day, prediction length q = 1 day.

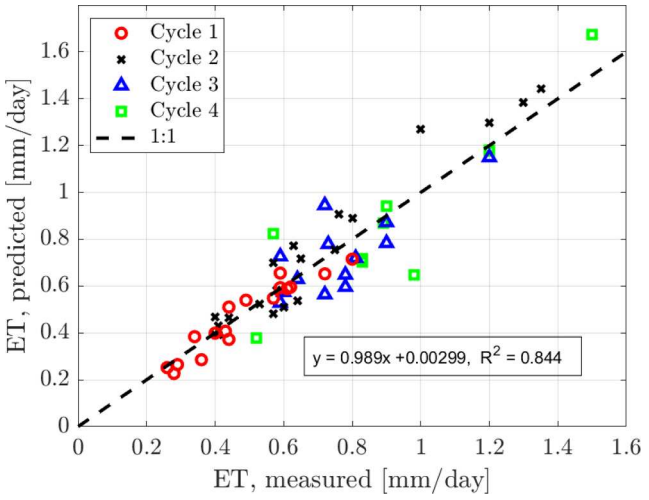


Fig. 6. Predicted vs measured crop ET data for the proposed state-space system. Parameters: sampling granularity $\Delta = 1$ day, prediction length q = 1 day.

crop ET values by our state-space model (via Kr) still follow the measured ET values, though sensing information is missing every time for three days. The main reason for this behavior is that weather hardly impacts the shape of the Kr time-series. The statistical figures of merit RMSE, EF and CRM are 0.099, 0.59, and 0.032, respectively which are only slightly worse than in the previous scenario where data was available each day.

Finally, we test the IoT performance. For the IoT client, the firmware resembles the functionality outlined in Fig. 2. On top of that, each node in the graph adds a timestamp to its data. The result occupies 22,036 Bytes of Flash memory and 3816 Bytes of RAM, satisfying the requirements. From the timestamps of the individual nodes in the graph, the IoT client's execution time profile can be readily obtained. The timings' first and second order statistics are summarized in Table 3. The bottleneck in this implementation is clearly Node a_3 (ET) as well Node a_4 (Tx). Sensing from load cells and processing its data in the former node takes most of the time while transmitting data to the cloud in the latter node has most uncertainty.

The execution time profile of the entire IoT application is shown in Fig. 8. It can be seen that the distribution is centered around $\mu_E = 23$ s. The standard deviation $\sigma_E = 1.6$ s (jitter $\sigma_E^2 = 2.65$ s^2), implying that the wireless channel adds a few more seconds of delay and some second square of jitter to the overall execution time. In any case, almost all data is aggregated in the cloud within a window of 25 s. Hence, we set the maximum response time to 25 s, to limit power consumption of our (battery powered) IoT client. For the system availability, we observed that sometimes the internet networks failed, at other times the Thingspeak™ cloud was out of order. In total, the system was up 233.2 h out of a window of 240 h (10 days) running time, corresponding to an availability of 97.2 %. Considering that our state-space model is immune to random missing data, this figure is excellent.

5. Conclusions

Over the last thirty years there have been proposed many crop water use models, to predict irrigation requirements. Unfortunately, the majority of state-of-the-art approaches "freeze" the model parameters during time-consuming training cycles or require destructive leaf-area-

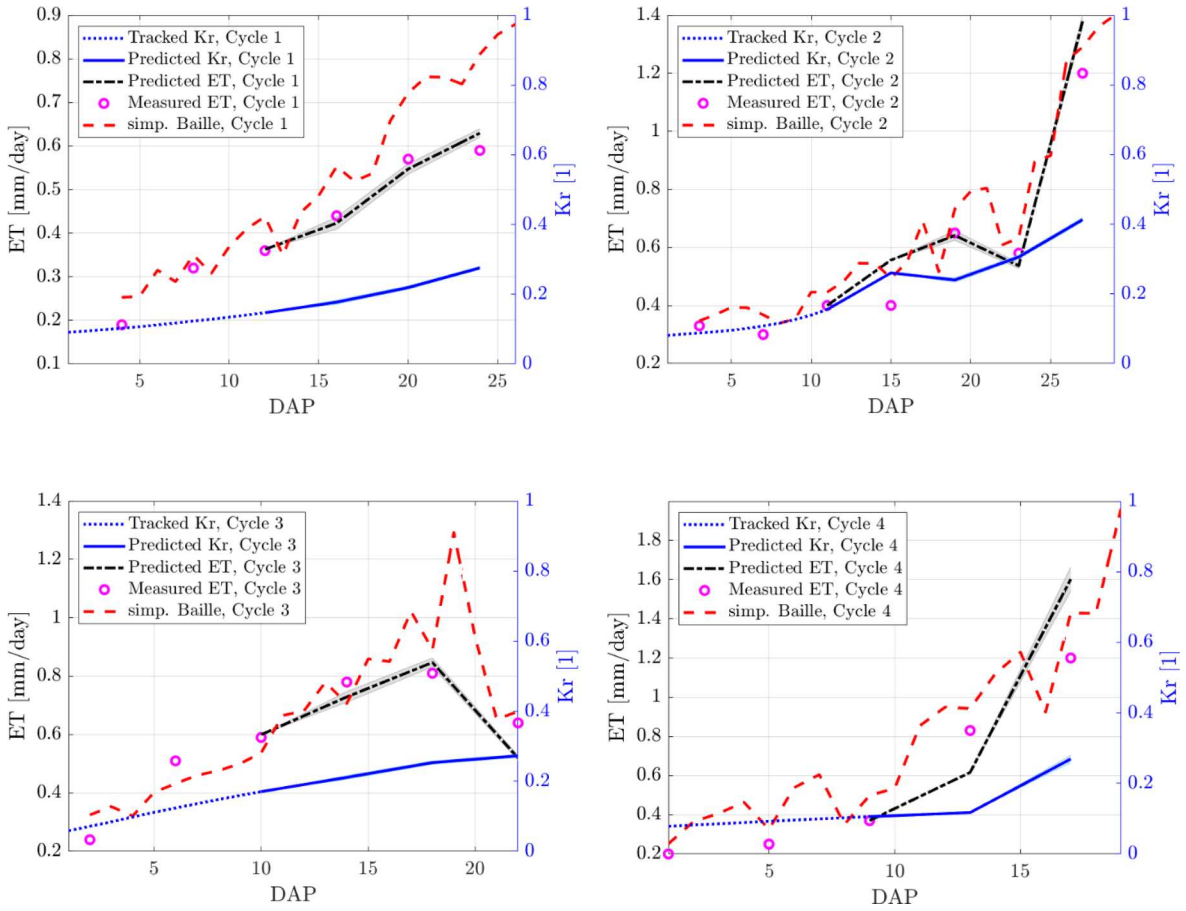


Fig. 7. Comparison between measured (symbols), trained (dotted line), and predicted (solid line) values of daily crop evapotranspiration (ET) for the proposed state-space system, and the simplified Baille technique (dashed line). Parameters: sampling granularity $\Delta = 4$ days, prediction length $q = 4$ days.

Table 3
Execution time profile of all tasks in the IoT client. Node 1: Temperature/Humidity, Node 2: Global radiation, Node 3: Evapotranspiration, Node 4: Data transmission.

Node	μ[s]	σ[s]	Node	μ[s]	σ[s]
a_1 (T/H)	0.051	0.007	a_3 (ET)	19.27	0.038
a_2 (GR)	1.027	0.015	a_4 (Tx)	0.95	0.60

index measurements. To overcome these drawbacks, we proposed to model a crop coefficient confined, Kr, to global radiation as Markov process with the daily air temperature as (known) disturbance. A state-space system describes the time-series of Kr for soilless cultivation. The Expectation-Maximization algorithm was used to compute the parameters of the system as well as to track its states. Finally, we derived a predictor for the posteriori probability of Kr. It turned out that the predicted Kr values are immune to erroneous weather forecasting. Treating GR as parameter rather than a variable, a guess of crop ET can be retrieved directly from the predicted time-series by scaling (without influencing the underlying Kr curve).

We executed our model within the Internet of Things (IoT) paradigm. A lightweight IoT-client from the Particle™ family, operating in soft real-time, collected data from a temperature/humidity sensor as well as a weighing scale, processed and transmitted the result over the wireless channel which is inherently stochastic. The ThingSpeak™ cloud is used to aggregate and visualize the collected data via web-interface. Our water crop use model, implemented in Matlab™ interacted with the ThingSpeak™ channel at the end of each day. To manage its resources, our IoT client was made quality-of-service (QoS) aware. The IoT

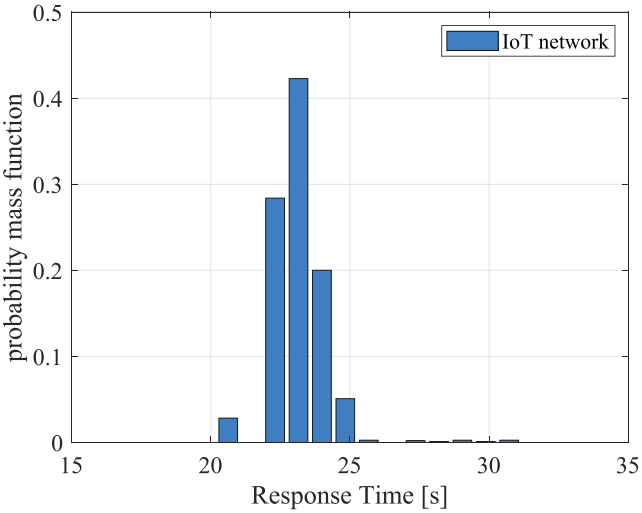


Fig. 8. Execution time profile for a scenario of one IoT client connected to ThingSpeak™ cloud service via Particle™ clout interface.

network was modeled as synchronous data flow graph, to easily evaluate the QoS metrics response time, jitter and availability.

Sweet basil (*Ocimum basilicum* L.) was used to test the performance of approach in four different climatic environments and two different data sampling strategies: i) sample daily from the same environment in parallel; ii) sample daily from one environment and move on to the next in round robin fashion, emulating reuse of the same set of measurement

equipment. We found that the latter strategy is the most economic approach. On top of that, our Kr-based crop ET predictor outperforms the competing Baille model, in both cases.

The proposed predictor is a first step to robust and fully autonomous crop water use anticipation without the need of an expert nor manual training cycles in contrast to the renown Baille technique.

CRediT authorship contribution statement

Alexander Kocian: Writing – original draft, Methodology, Software, Validation, Formal analysis. **Giulia Carmassi:** Writing – original draft, Validation, Formal analysis, Data curation. **Fatjon Cela:** Writing – original draft, Data curation, Validation. **Stefano Chessa:** Writing – original draft, Funding acquisition, Project administration, Resources, Methodology. **Paolo Milazzo:** Writing – original draft, Funding acquisition, Methodology. **Luca Incrocci:** Writing – original draft, Funding acquisition, Project administration, Resources, Methodology, Supervision.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

References

Allen, R. G., Pereira, L. S., & Raes, D. (1998). M. Smith. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56, Rome. Food and Agricultural Organization. Roma, 300(9).

Bacci, L., Battista, P., Cardarelli, M., Carmassi, G., Rouphael, Y., Incrocci, L., Malorgio, F., Pardossi, A., Rapi, B., Coll, G., 2011. Modelling Evapotranspiration of Container Crops for Irrigation Scheduling. In Evapotranspiration - From Measurements to Agricultural and Environmental Applications. https://doi.org/10.5772/18697.

Baille, M., Baille, A., Laury, J.C., 1994. A simplified model for predicting evapotranspiration rate of nine ornamental species vs. climate factors and leaf area. Scientia Horticulturae 59 (3–4). https://doi.org/10.1016/0304-4238(94)90015-9.

Balducci, F., Impedovo, D., Pirlo, G., 2018. Machine learning applications on agricultural datasets for smart farm enhancement. Machines 6 (3). https://doi.org/10.3390/machines6030038.

Brogi, A., Forti, S., 2017. QoS-aware deployment of IoT applications through the fog. IEEE Internet of Things Journal 4 (5). https://doi.org/10.1109/JIOT.2017.2701408.

Bu, F., Wang, X., 2019. A smart agriculture IoT system based on deep reinforcement learning. Future Generation Computer Systems 99. https://doi.org/10.1016/j.future.2019.04.041.

Burchi, G., Chessa, S., Gambineri, F., Kocian, A., Massa, D., Milazzo, P., Rimediotti, L., & Ruggeri, A. (2018). Information technology controlled greenhouse: A system architecture. 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany, IOT Tuscany 2018. https://doi.org/10.1109/IOT-TUSCANY.2018.8373044.

Carmassi, G., Incrocci, L., Maggini, R., Malorgio, F., Tognoni, F., Pardossi, A., 2007. An aggregated model for water requirements of greenhouse tomato grown in closed rockwool culture with saline water. Agricultural Water Management 88 (1–3). https://doi.org/10.1016/j.agwat.2006.10.002.

Chen, B., Wu, Z., Wang, J., Dong, J., Guan, L., Chen, J., Yang, K., Xie, G., 2015. Spatio-temporal prediction of leaf area index of rubber plantation using HJ-1A/1B CCD images and recurrent neural network. ISPRS Journal of Photogrammetry and Remote Sensing 102. https://doi.org/10.1016/j.isprsjprs.2014.12.011.

CREA. (2021). Italian Agriculture in Figures. https://www.crea.gov.it/en/web/politiche-e-bioeconomia/-/agricoltura-italiana-conta.

Fernández, M.D., Bonachela, S., Orgaz, F., Thompson, R.B., López, J.C., Granados, M.R., Gallardo, M., Fereres, E., 2011. Erratum to: Measurement and estimation of plastic greenhouse reference evapotranspiration in a Mediterranean climate. Irrigation Science 29 (1). https://doi.org/10.1007/s00271-010-0233-5.

Gallardo, M., Thompson, B., Fernández, M.D., 2013. Water requirements and irrigation management in Mediterranean greenhouses: the case of the southeast coast of Spain. In Good Agricultural Practices for Greenhouse Vegetable Crops. PrInciple for Mediterranean Climate Areas (Issue November 2015).

Goap, A., Sharma, D., Shukla, A.K., Rama Krishna, C., 2018. An IoT based smart irrigation management system using Machine learning and open source technologies. Computers and Electronics in Agriculture 155, 41–49. https://doi.org/10.1016/J.COMPAG.2018.09.040.

Guerra, E., Ventura, F., Snyder, R.L., 2016. Crop Coefficients: A Literature Review. Journal of Irrigation and Drainage Engineering 142 (3). https://doi.org/10.1061/(asce)ir.1943-4774.0000983.

Hargreaves, G.H., Samani, Z.A., 1985. Reference Crop Evapotranspiration from Temperature. Applied Engineering in Agriculture 1 (2). https://doi.org/10.13031/2013.26773.

Hu, B., Kocian, A., Piton, R., Hviid, A., Fleury, B.H., Rasmussen, L.K., 2004. Iterative joint channel estimation and successive interference cancellation using a SISO-SAGE algorithm for coded CDMA. Conference Record - Asilomar Conference on Signals, Systems and Computers 1, 622–626. https://doi.org/10.1109/acssc.2004.1399209.

Jayasinghe, C., Goto, N., Aoki, T., Wada, S., 2003. Phenolics composition and antioxidant activity of sweet basil (Ocimum basilicum L.). Journal of Agricultural and Food Chemistry 51 (15). https://doi.org/10.1021/jf034269o.

Katsoulas, N., Stanghellini, C., 2019. Modelling crop transpiration in greenhouses: Different models for different applications. In Agronomy Vol. 9, Issue 7. https://doi.org/10.3390/agronomy9070392.

Khater, E.S., Bahnasawy, A., Abass, W., Morsy, O., El-Ghobashy, H., Shaban, Y., Egela, M., 2021. Production of basil (Ocimum basilicum L.) under different soilless cultures. Scientific Reports 11 (1). https://doi.org/10.1038/s41598-021-91986-7.

Kocian, A., Carmassi, G., Cela, F., Incrocci, L., Milazzo, P., Chessa, S., 2020a. Bayesian sigmoid-type time series forecasting with missing data for greenhouse crops. Sensors (Switzerland) 20 (11). https://doi.org/10.3390/s20113246.

Kocian, A., Chessa, S., 2022. Iterative Probabilistic Performance Prediction for Multiple IoT Applications in Contention. IEEE Internet of Things Journal 9 (15), 13416–13424. https://doi.org/10.1109/JIOT.2022.3142324.

Kocian, A., Massa, D., Cannazzaro, S., Incrocci, L., di Lonardo, S., Milazzo, P., Chessa, S., 2020b. Dynamic Bayesian network for crop growth prediction in greenhouses. Computers and Electronics in Agriculture 169. https://doi.org/10.1016/j.compag.2019.105167.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. Journal of Contaminant Hydrology 7 (1–2). https://doi.org/10.1016/0169-7722(91)90038-3.

Makri, O., Kintzios, S., 2008. Ocimum sp. (Basil): Botany, cultivation, pharmaceutical properties, and biotechnology. In. Journal of Herbs, Spices and Medicinal Plants Vol. 13, Issue 3. https://doi.org/10.1300/J044v13n03_10.

Mekonnen, Y., Namuduri, S., Burton, L., Sarwat, A., Bhansali, S., 2020. Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture. Journal of The Electrochemical Society 167 (3). https://doi.org/10.1149/2.0222003jes.

Monteith, J., 1965. 205–234. In Symp Soc Exp Biol. Evaporation and environment. 19th Symposia of the Society for Experimental Biology Vol. 19.

Puccinelli, M., Pezzarossa, B., Rosellini, I., Malorgio, F., 2020. Selenium enrichment enhances the quality and shelf life of basil leaves. Plants 9 (6). https://doi.org/10.3390/plants9060801.

Rubin, D.R., 1976. Inference and missing data. Biometrika 63 (3), 581–592.

Shuttleworth, W.J., 2006. TOWARDS ONE-STEP ESTIMATION OF CROP WATER REQUIREMENTS. Transactions of the ASABE 49 (4). https://doi.org/10.13031/2013.21738.

Solbach, J.A., Fricke, A., Stützel, H., 2021. Seasonal Efficiency of Supplemental LED Lighting on Growth and Photomorphogenesis of Sweet Basil. Frontiers in Plant Science 12. https://doi.org/10.3389/fpls.2021.609975.

Thornley, J.H.M., Johnson Ian, R., 1990. Plant and Crop Modelling. The Blackburn Press.

Walters, K.J., Currey, C.J., 2019. Growth and development of basil species in response to temperature. HortScience 54 (11). https://doi.org/10.21273/HORTSCI12976-18.