

Article

A Stochastic Bayesian Artificial Intelligence Framework to Assess Climatological Water Balance under Missing Variables for Evapotranspiration Estimates

Vitor P. Ribeiro ^{1,*}, Luiz Desuó Neto ², Patricia A. A. Marques ³, Jorge A. Achcar ⁴, Adriano M. Junqueira ¹, Adilson W. Chinatto, Jr. ⁵, Cynthia C. M. Junqueira ⁵, Carlos D. Maciel ² and José Antônio P. Balestieri ¹

¹ School of Engineering and Sciences, São Paulo State University (UNESP), Guaratinguetá 12516-410, SP, Brazil; adriano.junqueira@unesp.br (A.M.J.); jose.perrella@unesp.br (J.A.P.B.)

² Department of Electrical and Computer Engineering, University of São Paulo (USP), São Carlos 13566-590, SP, Brazil; luiz.desuo.neto@usp.br (L.D.N.); carlos.maciell@usp.br (C.D.M.)

³ Department of Biosystems Engineering, University of São Paulo (USP), Piracicaba 13418-900, SP, Brazil; paamarques@usp.br

⁴ Medical School, University of São Paulo (USP), Ribeirão Preto 14049-900, SP, Brazil; achcar@fmrp.usp.br

⁵ Espectro Ltd., Campinas 13084-012, SP, Brazil; chinatto@espectro-eng.com.br (A.W.C.J.); cyjunqueira@uol.com.br (C.C.M.J.)

* Correspondence: vitor.p.ribeiro@unesp.br

Abstract: The sustainable use of water resources is of utmost importance given climatological changes and water scarcity, alongside the many socioeconomic factors that rely on clean water availability, such as food security. In this context, developing tools to minimize water waste in irrigation is paramount for sustainable food production. The evapotranspiration estimate is a tool to evaluate the water volume required to achieve optimal crop yield with the least amount of water waste. The Penman-Monteith equation is the gold standard for this task, despite it becoming inapplicable if any of its required climatological variables are missing. In this paper, we present a stochastic Bayesian framework to model the non-linear and non-stationary time series for the evapotranspiration estimate via Bayesian regression. We also leverage Bayesian networks and Bayesian inference to provide estimates for missing climatological data. Our obtained Bayesian regression equation achieves $0.087 \text{ mm} \cdot \text{day}^{-1}$ for the RMSE metric, compared to the expected time series, with wind speed and net incident solar radiation as the main components. Lastly, we show that the evapotranspiration time series, with missing climatological data inferred by the Bayesian network, achieves an RMSE metric ranging from 0.074 to $0.286 \text{ mm} \cdot \text{day}^{-1}$.

Keywords: irrigation planning; decision support system; artificial intelligence; stochastic modeling; Bayesian inference



Citation: Ribeiro, V.P.; Desuó Neto, L.; Marques, P.A.A.; Achcar, J.A.; Junqueira, A.M.; Chinatto, A.W., Jr.; Junqueira, C.C.M.; Maciel, C.D.; Balestieri, J.A.P. A Stochastic Bayesian Artificial Intelligence Framework to Assess Climatological Water Balance under Missing Variables for Evapotranspiration Estimates. *Agronomy* **2023**, *13*, 2970. <https://doi.org/10.3390/agronomy13122970>

Academic Editors: Mo Li and Fan Zhang

Received: 28 September 2023

Revised: 30 October 2023

Accepted: 2 November 2023

Published: 30 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water maintains life. From a broad perspective, water resources are at the core of several vital elements such as ecology [1], agriculture [2], transportation [3], electricity generation [4], and thirst quenching [5]. Water resources also impact many sociopolitical and economic areas and are essential for human survival, alongside other natural and human resources. Given the importance of these factors for the development of humanity, the United Nations (UN) proposed sustainable development goals (SDGs). These goals consist of seventeen components with their underlying objectives proposed as guidelines for the rational and sustainable development of all UN member states [6]. In Brazil, many socioeconomic factors are directly tied to the use of water resources. For example, hydroelectric plants generated the majority of the country's electricity generation capacity of 181.6 GW for 2022, with a share of 60.2% [7]. Brazil is also a key contributor to global food security, ranking in 2021 as the world's largest producer of coffee beans, oranges,

and soybeans; the second-largest in cattle meat; and the third in maize and chicken meat, among several other agricultural and livestock products [8]. The intricate interactions between soil, plants, and atmospheric systems play a pivotal role in food production [9]. Therefore, techniques that aid in producing nutritious food based on limited components of soil and water resources are paramount due to climate change and the scarcity of water resources [10].

The productivity of crops is one of the predominant factors in the maintenance of food security. The total yield of a crop is highly correlated with factors such as nutrient satisfaction, the application of pesticides, sunshine, and water availability [11]. Wheat is one of the three main cereals that underpin global security—alongside maize and rice—and serves as an example to highlight the amount of water required to produce a significant quantity of such an essential commodity [12]. In 2021, mainland China was the world's largest wheat producer, with an estimated harvested area of 23.5 million hectares, according to the UN's Food and Agriculture Organization (FAO-UN) [8]. Based on a rough approximation where the estimated area is composed solely of winter wheat, which requires approximately 120 mm ($1200 \text{ m}^3 \cdot \text{ha}^{-1}$) for optimal production under regular climate conditions [13], the water required solely for this crop during its production cycle amounts to about 2.828×10^{10} mm as a volume measurement over a standardized area of 1 m^2 . This water volume for the consumptive use of one crop in one country underscores the importance of policies that promote the sustainable and rational use of water resources.

One method to minimize the waste of water in irrigation is to perform the climatological water balance for the cultivated area. This task consists of monitoring climatological events—mainly evapotranspiration and precipitation—to estimate the availability of water resources to the crops in the area [14]. The Penman–Monteith equation is the gold standard for estimating a reference evapotranspiration metric under standardized conditions [15]. The method demands the acquisition of climatological variables—net solar radiation, wind speed, temperature, and relative humidity—whose measurements may not be available for a given area [16]. Several other empirical equations are present in the literature, most as simplifications from the gold standard for specific regions or to deal with the lack of a given climatological variable [17]. Many authors have presented different approaches to tackle the issue of evapotranspiration estimates from data leveraging computational power and algorithmic resources [18–22]. Moreover, the method proposed in [23] models and discusses evapotranspiration on a global scale unaided by soil and vegetation data, and the review in [24] covers many evaporation and evapotranspiration estimate methods from the standpoint of several applications, their requirements, and characteristics.

Reliably assessing the parameters of the climatological water balance is paramount for achieving sustainable water usage and providing food security for the population [25]. Despite the multitude of techniques and approaches to estimate these parameters—mainly evapotranspiration—most methods presented in the literature lack flexibility when a required climatological factor is unavailable, becoming inapplicable [18]. Another characteristic of many methods that provide forecasts based on climatological or agrometeorological data is their inability to cope with the nonlinearity and non-stationarity inherent to these elements [26,27]. As such, the specialized literature reveals a knowledge gap for stochastic methods with a degree of flexibility that allows them to perform even with incomplete information.

A method class that resonates with the purported characteristics is the stochastic Bayesian regression analysis, which estimates the probability distribution functions that best fits the data [28]. This process has the innate ability to assimilate the non-deterministic and non-stationary characteristics in the data and provide a robust tool for modeling real-world data in several fields, such as medicines [29] and power planning [30]. Another method based on Bayesian statistics is the Bayesian network (BN), which is an approach based on probabilistic graph models encompassing association relationships between stochastic variables, which allows for the query of the most probable outcome of a given variable based on a complete or incomplete observation of the other variables

in the model [31]. Examples of fields of study that benefit from BN modeling include environmental science [32], supply chain analysis [33], and engineering reliability [34]. The main drawback of this method relies on finding the directed acyclic graph that presents the best adherence to the data, which is a super-exponential problem based on the number of observed variables and belongs to the NP-hard class of problems [35].

The contributions of this paper are three-fold: first, we present and discuss a fully stochastic computational framework to determine the climatological variables that are most relevant to estimate daily evapotranspiration exclusively from data based on semi-parametric modeling, Bayesian statistics, and MCMC methods. Second, we establish a Bayesian logistic regression equation to incorporate climatological data characteristics of nonlinearity and non-stationarity. Thirdly, we present a BN structure to support the inference task of missing variables, which provides flexibility to the proposed model and allows for its usage under different scenarios of data availability. We also show the performance of the proposed BN to infer any of the missing relevant climatological factors based on the evidence as samples of the other variables and achieve reasonable error metrics for evaporation estimate from the logistic regression with inferred data against the expected value. Lastly, we compare the behaviors of measured precipitation, estimated evapotranspiration—both from the Penman–Monteith equation and the proposed method—climatological water balance, and in situ tensiometer data to verify the adherence of the theoretical models to real-world applications.

The remainder of this paper is structured as follows: Section 2 provides the theoretical background relevant to the conducted experiments and discussions for both the climatological models and fundamentals of Bayesian modeling. Section 3 presents the employed data, evaluated models, and descriptions of the main algorithms. Section 4 presents the results and discussions pertinent to the proposed experiments. Lastly, Section 5 presents the conclusions of this paper, alongside remarks about the advantages and drawbacks of the proposed methodology and a few identified gaps for future works.

2. Theory

This section provides a theoretical background of the topics discussed throughout this work. Sections 2.1 and 2.2 present a theory for the climatological and agrometeorological components of the discussion, while Sections 2.3 and 2.4 focus on the Bayesian inference and BNs.

2.1. Reference Evapotranspiration—Penman–Monteith (FAO-56) Method

Estimating the amount of water evaporated and transpired by the culture and evaporated from the soil is vital to determine the irrigation volume required for optimal production. However, it is difficult—or in some cases, practically impossible—to measure directly due to various climatological factors.

In light of the significance of evapotranspiration, the literature offers diverse analytical methods [36–39] to estimate it from climatological variables based on various environmental conditions [17]. These traditional methods utilize regression analysis to model and extrapolate the evapotranspiration metric conditioned on the availability of specific climatological variables, such as the Hargreaves–Samani equation, which relies solely on mean temperature and extraterrestrial solar radiation [36]. Another instance is the Linacre method, calibrated according to the Australian climate [37].

The Food and Agriculture Organization [15] has designated the Penman–Monteith equation as the gold standard for estimating standardized evapotranspiration for a grass crop (vegetation height = 0.12 m, albedo = 23%) on soil with a fixed surface resistance of $70 \text{ S} \cdot \text{m}^{-1}$ without water restrictions. The original Penman–Monteith equation is presented here as Equation (1), which estimates the ET_0 reference evapotranspiration in millimeters per day.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \left(\frac{900}{T_{mean} + 273} \right) U_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

The Penman–Monteith equation necessitates several climatological measurements, including net solar radiation (R_n), mean daily temperature (T_{mean}), on-site wind speed at 2 m (U_2), and actual (e_a) and saturation (e_s) vapor pressure. The slope of the vapor pressure curve (Δ), the soil heat flux density (G), and the psychrometric constant (γ) are additional parameters.

Despite its high correlation with the observed value for the evapotranspiration metric, the primary limitation of this equation is its dependence on several climatological factors. Weather stations that can automatically provide the required data for the Penman–Monteith equation are unattainable to most rural producers, resulting in this method being unfeasible to estimate [40].

One analytical approach employed throughout the literature to estimate daily evapotranspiration based on fewer climatological variables is the Hargreaves–Samani method [36], presented here as Equation (2). This method requires the mean, maximum, and minimum daily temperatures—respectively T_{mean} , T_{max} , and T_{min} —as well as an estimate for the extraterrestrial incident solar radiation (R_a). Another common analytical approach is the Benavides and Lopez method [39], presented as Equation (3), whose equation only requires data on the mean daily temperature (T_{mean}) and mean daily relative humidity (RH_{mean}).

$$ET_{0Hg} = R_a \times 0.408 \times 0.0023(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} \quad (2)$$

$$ET_{0BL} = 1.21 \times 10 \left(\frac{7.42T_{mean}}{234.7 + T_{mean}} \right) (1 - 0.01RH_{mean}) + 0.21T_{mean} - 2.30 \quad (3)$$

The variety of phytomorphology properties for crops used for plantations implies a high variance for the water amount evaporated and transpired by the plants. For this reason, and to provide a generic framework to estimate evapotranspiration under scenarios of different cultures and development stages, the analytical equations—including the Penman–Monteith approach—provide a standardized reference evapotranspiration ($ET_0 \triangleq ET$). To estimate the evapotranspiration value for a given culture and development stage, one can use the culture coefficient (k_c) to obtain the crop evapotranspiration (ET_c) according to Equation (4).

$$ET_c = k_c \cdot ET_0 \quad (4)$$

Throughout this text, the word “evapotranspiration” refers to the reference evapotranspiration $ET_0 \triangleq ET$ unless stated otherwise.

2.2. Climatological Water Balance and Water Requirements

The water balance for a given region obeys the mass conservation principle summarized by Equation (5), which describes the equilibrium of water inflows and outflows from a control volume [41].

$$P = ET + Q + \Delta S \quad (5)$$

In Equation (5), the water inflow from precipitation (P) balances the outflows from evapotranspiration (ET) and surface and subsurface runoff (Q), promoting a variation in soil moisture (ΔS).

When considering a plantation scenario of groundwater flow equilibrium, a linear or exponential function describes the surface runoff as a function of precipitation [42]. As such, Equation (6) describes the soil moisture dynamics as a function of evapotranspiration and precipitation.

$$P = ET + f(P) + \Delta S \quad (6)$$

Assessing soil water content plays a fundamental role in establishing the water available to crops, which impacts plantation production. The water amount required for optimal production depends on the culture, its development stage, and efforts to minimize water waste.

A simple approach to estimating the water content in a given region relies on climatological data for precipitation and evapotranspiration, the so-called climatological water

balance, to estimate the available water deficit (awd), the percentage of water stored in soil (st), and the net water balance [43]. Algorithm 1 presents the heuristics to assess the awd and st metrics solely from the climatological factors.

Algorithm 1 Algorithm to estimate the accumulated water deficit $curr_awd$ and water stored $curr_st$ based on evapotranspiration ET , precipitation P , and the previous values of these estimates.

Require: $ET, P, prev_awd, prev_st$

- 1: **if** $ET > P$ **then**
- 2: $curr_awd \leftarrow prev_awd + (P - ET)$
- 3: $curr_st \leftarrow 100 \cdot e^{curr_awd/100}$
- 4: **else**
- 5: $curr_st \leftarrow prev_st + (P - ET)$
- 6: **if** $curr_st > 100$ **then**
- 7: $curr_st \leftarrow 100$
- 8: **end if**
- 9: $curr_awd \leftarrow 100 \cdot e^{curr_st/100}$
- 10: **end if**

return $curr_awd, curr_st$

Given all these nuances, several methods estimate soil water content based on readings of low-cost sampling equipment, such as tensiometers and capacitors, to provide decision-support for the irrigation responsible [44]. One of these methods is the van Genuchten equation, which estimates soil water content for a specific soil (θ , in percentage) based on soil moisture content estimated by suction pressure (ψ , in kPa) and experimental coefficients [45]. Equation (7) provides the generic expression for the van Genuchten equation.

$$vG(\psi) = \theta_r + \frac{\theta_s - \theta_r}{(1 + (\alpha|\psi|)^n)^m} \quad (7)$$

where θ_s is the saturated soil water content or field capacity, and θ_r is the residual soil water content or wilting point under the physical limitations of the tensiometers. The parameters α [cm^{-1}], n , and $m = 1 - 1/n$ are constants obtained experimentally based on the soil's characteristic water retention curve.

2.3. Bayesian Regression and Modeling

Regression analysis is a tool used to estimate the influence of a set of input variables over an output (response) variable [46]. Let y be the target variable and $X = (X_1, X_2, \dots, X_n)$ be the set of components (or covariates) that may contribute to the values of the outcome. Equation (8) presents a multiple linear regression of y as a linear combination of elements X_i plus a constant value as the series' DC component. The main task of this regression is to estimate $A = \{a_0, a_1, \dots, a_n\}$ as the set of regression parameters from the dataset.

$$y \approx a_0 + \sum_{i=1}^n a_i X_i + \varepsilon \quad (8)$$

A random variable usually assumed with a Gaussian $\mathcal{N}(0, \sigma^2)$ distribution models the regression error term ε (non-observed) in Equation (8). Under the frequentist approach, the estimators of the regression parameters a_0, a_1, \dots, a_n are usually obtained using the minimum squares method, i.e., the set of coefficients that minimizes the square sum of the errors [47]. The variance parameter σ^2 should also be estimated from the data.

Under a Bayesian approach (Bayesian regression model), we assume the same regression model from Equation (8), while considering Gaussian errors, with the regression parameters β_i representing random quantities with specified probability density functions—also denoted as prior distributions—with hyperparameters φ_i . These hyperparameters may be univariate or multivariate, whose values are usually known, as represented

in Equation (9). If the hyperparameters are unknown, which comes from a hierarchical Bayesian approach, each hyperparameter also requires—over itself—a prior distribution.

$$y \approx \beta_0(\varphi_0) + \sum_{i=1}^n (\beta_i(\varphi_i) \cdot X_i) + \varepsilon \quad (9)$$

Bayes' theorem provides the means to obtain the posterior distributions for the regression parameters β_i [48]. Let D represent the dataset, $\pi(B)$ represent the joint prior distribution for the vector of regression parameters, $B = (\beta_0, \beta_1, \dots, \beta_n)$, and $L(B | D)$ represent the likelihood function for B based on the data D . Under the assumption of Gaussian distribution for the error ε in Equation (9), the joint posterior distribution for B , denoted as $\pi(B | D)$, is given by the combination of the prior distribution $\pi(B)$ with the likelihood function $L(B | D)$, as described in Equation (10). The integral in the denominator represents multiple integrals with dimension $\dim(B)$.

$$\pi(B | D) = \frac{\pi(B)L(B | D)}{\int_B \pi(B)L(B | D)dB} \quad (10)$$

The integral in Equation (10)'s denominator represents multiple integrals with dimension $\dim(B)$, which presents the main drawback of the Bayesian regression analysis. An analytical approach to estimate the term $\int_B \pi(B)L(B | D)dB$ usually becomes unfeasible.

Popular approaches extensively used in applied Bayesian analysis are Markov Chain Monte Carlo (MCMC) method, such as Gibbs sampling or Metropolis–Hastings algorithms, to simulate samples for the joint posterior distribution of interest [48]. The MCMC method generates values from each conditional posterior distribution $\pi(\beta_i | \beta_{(i)}, D)$, where $\beta_{(i)}$ denotes the vector of all parameters, except β_i , and is available in existing Bayesian software.

From the simulated samples of the joint posterior distribution of interest, we obtain the posterior summaries of interest, such as Monte Carlo point Bayesian estimates for the parameters of the model (usually the posterior means) or 95% credible intervals for the parameters of interest.

The authors acknowledge the intricacies and hardships of the MCMC methods, such as chain convergence, the sampler chosen, and initial values for the parameters in the iterative method. These characteristics, albeit relevant, lie outside the scope of the present work and were omitted. Please refer to reference [49] for more details on MCMC methods.

2.4. Bayesian Networks

Let D represent the dataset sampled from a system, which comprises n random variables, i.e., variables whose values follow a probability distribution function. A BN is a probabilistic graph model representing the stochastic association relationships between all variables in a given system [31].

Let Θ be a characterization of the joint probability function, according to a product measure for a system's n random variables $X_i, i = \{1, 2, \dots, n\}$ based on arbitrary ordering. Equation (11) provides a generalization for such a joint distribution.

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i | X_j = x_j, \forall j < i) \quad j \in \mathbb{N}^* \quad (11)$$

It is possible to factor out the independent terms for each X_i variable in Equation (11), which relies on conditional probability distributions. Let $pa(X_j)$ represent the set of random variables whose pairwise joint probability function satisfies the relation in Equation (12), given an ordering for the system's variables. The set $pa(X_i)$ defines the set of *parent* variables for X_i .

$$X_i \in pa(X_j) \iff P(X_i = x_i, X_j = x_j) < P(X_i = x_i)P(X_j = x_j), \quad i > j; \quad i, j \in \{1, 2, \dots, n\} \quad (12)$$

Equation (13) presents a simplified version of Equation (11), based on the variables X_i conditioned to their parents' set $pa(X_i)$.

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i | pa(X_i)) \quad (13)$$

A BN Θ is a probabilistic graphic model that represents the joint probability function in Equation (13), based on a directed acyclic graph (DAG) G [35]. The BN's graphical structure contains one vertex for each random variable X_i , and its edges correspond to the perceived mathematical dependency of a variable on its parents. Based on this definition, each node X_i has one directed path, starting from each $X_j \in pa(X_i)$ and ending on itself.

A relevant comment regarding the definition of the parent sets from Equation (12) that reflects on the direction of the edges in G is that it establishes an *ordering* for the random variables. Based on the symmetrical nature of the right-hand side of Equation (12)— $P(X_i)P(X_j) = P(X_j)P(X_i)$ —whether X_i belongs to $pa(X_j)$ or X_j belongs to $pa(X_i)$ is entirely dependent on the ordering of the variables. Indeed, the edges in the graph of a BN do not imply a causal relationship between random variables but rather association relationships based on the variables' ordering [50].

Each vertex on the BN has an underlying representation of the probability distribution σ_i for its values, conditioned to each realization for the elements on its parent's set. This mechanism—usually a set of conditional probability tables (Σ) when all observed variables are discrete—allows for inference queries for the most probable outcome of any given variable, given a complete or partial observation of all the elements on this component's parent set. As such, both the DAG G and the collection of CPTs over the vertices Σ are factors that fully describe a BN $\Theta = (G, \Sigma)$ [31].

Equation (14) presents a generalization for Equation (10) based on Bayes' theorem, and the most right-handed side component is valid if the dataset contains only independent and identically distributed (i.i.d.) samples.

$$P(\Theta | D) = \frac{P(D | \Theta)P(\Theta)}{P(D)} \propto P(D | \Theta)P(\Theta) \quad (14)$$

The model that has the best adherence to the underlying system is the one that provides the best posterior probability based on the provided dataset under the (usually naive [51]) assumption that the provided data fully represent the system's behavior. As such, Equation (15) provides a mathematical expression based on the right-hand side Equation (14) to find the best model for the underlying system based on data.

$$\max_{\Theta} P(\Theta | D) \propto \operatorname{argmax}_{\Theta} P(D | \Theta)P(\Theta) \quad (15)$$

In other words, Equation (15) stipulates that the model parametrized by Θ that provides the best likelihood of data given its parameters $P(D | \Theta)$ is the one that provides the best predictive posterior distribution $P(\Theta | D)$ for the system's components.

Finding the best BN parametrized by $\Theta = (G, \Sigma)$ that maximizes the likelihood of the data distribution according to the model, satisfying Equation (15), entails a graph optimization problem [52]. This characteristic is due to the structural component of a BN being a DAG, and the representation of the joint probability for each node relies on its parent set.

There are three main approaches to obtaining a BN to maximize the data likelihood of the model: score-and-search, constraint-based, and hybrid methods. Score-and-search methods employ a measure to score the model's adherence to the data and search on the graph space for a structure that improves this metric. Constraint-based methods rely on conditional independence tests over the data variables to establish the edges on G , similar to those presented in Equation (12). Lastly, hybrid methods start from a constraint-based

approach to scale down the graph search space and, afterward, a score-and-search method to scour the reduced space.

The significant amount of time and computational power required to derive a Bayesian Network (BN) from data are among the main disadvantages of this method. Despite these substantial requirements, the inference queries on the BN structure rely on basic matrix operations considering the CPT nodes.

3. Material and Method

3.1. Time Series for Climatological Data

The data employed throughout this paper originate from the Brazilian National Institute of Meteorology (INMET—Instituto Brasileiro de Meteorologia). The dataset consists of time series of seven climatological variables, i.e., temperature, atmospheric pressure, relative humidity, dew point temperature, wind speed, incident solar radiation, and precipitation.

The particular monitored location is the city of Patrocínio in the state of Minas Gerais (18°59'48" S, 46°59'09" W), at 978.11 m above sea level, and a warm temperate with a hot summer climate profile (Cwa), according to the Köppen classification. Figure 1 presents the geographical location of the state of Minas Gerais (MG) within the Brazilian map and the Patrocínio municipality within the former, alongside four other municipalities whose climatological time series that the authors employed as validation sets. Patrocínio is Brazil's largest coffee producer and the fourth-largest producer of bovine milk, among other goods. The initial dataset consists of 5473 daily samples for each climatological variable from Patrocínio/MG. The data span from 22 August 2008 to 31 January 2022, with a 97% rate of valid elements.

Figure 2a–g present graphical representations of the time series, with temperature, atmospheric pressure, relative humidity, and dew point temperatures sub-divided into maximum, mean, and minimum values for each day. Figure 2h presents the estimated evapotranspiration according to the Penman–Monteith equation over the available climatological data.

An additional dataset with the same climatological variables from Patrocínio/MG acts as a test set to evaluate the proposed method. This extra set spans from 1 February to 31 December 2022, with a total of 334 samples. Furthermore, four supplementary datasets, each containing 334 samples from the same period previously described, from four different cities across Brazil, serve to evaluate the model when considering distinct regions and climate profiles. Table 1 presents the names of cities, their geographic coordinates, the climate profiles according to the Köppen classification, and their distances from Patrocínio/MG. The only criterion for choosing these cities is to provide good coverage across Brazil's territory.

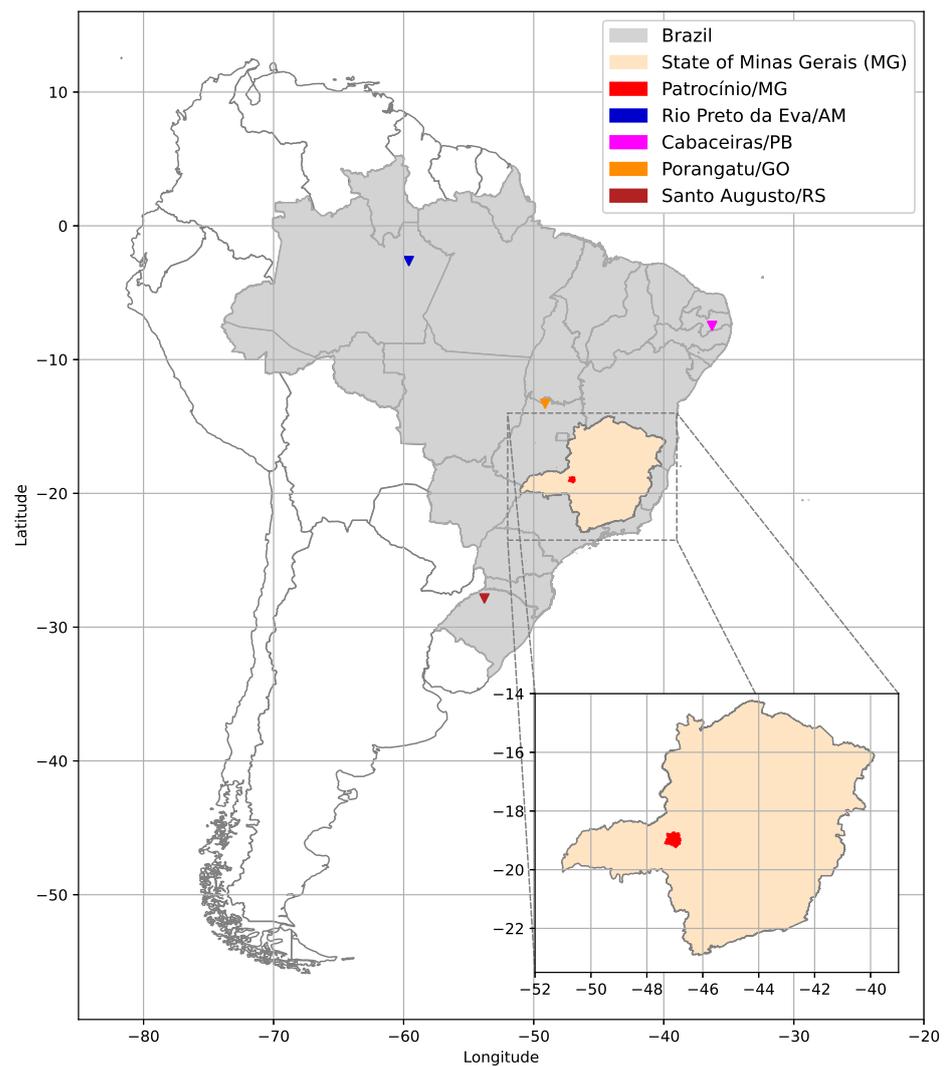


Figure 1. The geographical location of the Minas Gerais (MG) state is highlighted in yellow within Brazil's map. The red detail within the blue area highlights the Patrocínio/MG municipality's geographical location. The other four markers represent cities with different climate profiles from Patrocínio/MG's, whose climatological time series were employed by the authors to validate and deepen the discussion around the proposed model.

Table 1. Localization and profile climate of four different cities within Brazil, alongside their distances to Patrocínio/MG.

City's Name	Coordinates	Climate Profile	Distance to Patrocínio/MG (km)
Rio Preto da Eva/AM	2°41'56" S 59°42'00" W	Tropical rain forest (Af)	2278
Cabaceiras/PB	7°29'20" S 36°17'13" W	Hot semi-arid (BSh)	1723
Porangatu/GO	13°26'27" S 49°08'56" W	Tropical wet (Aw)	659
Santo Augusto/RS	27°51'03" S 53°46'37" W	Humid subtropical (Cfa)	1202

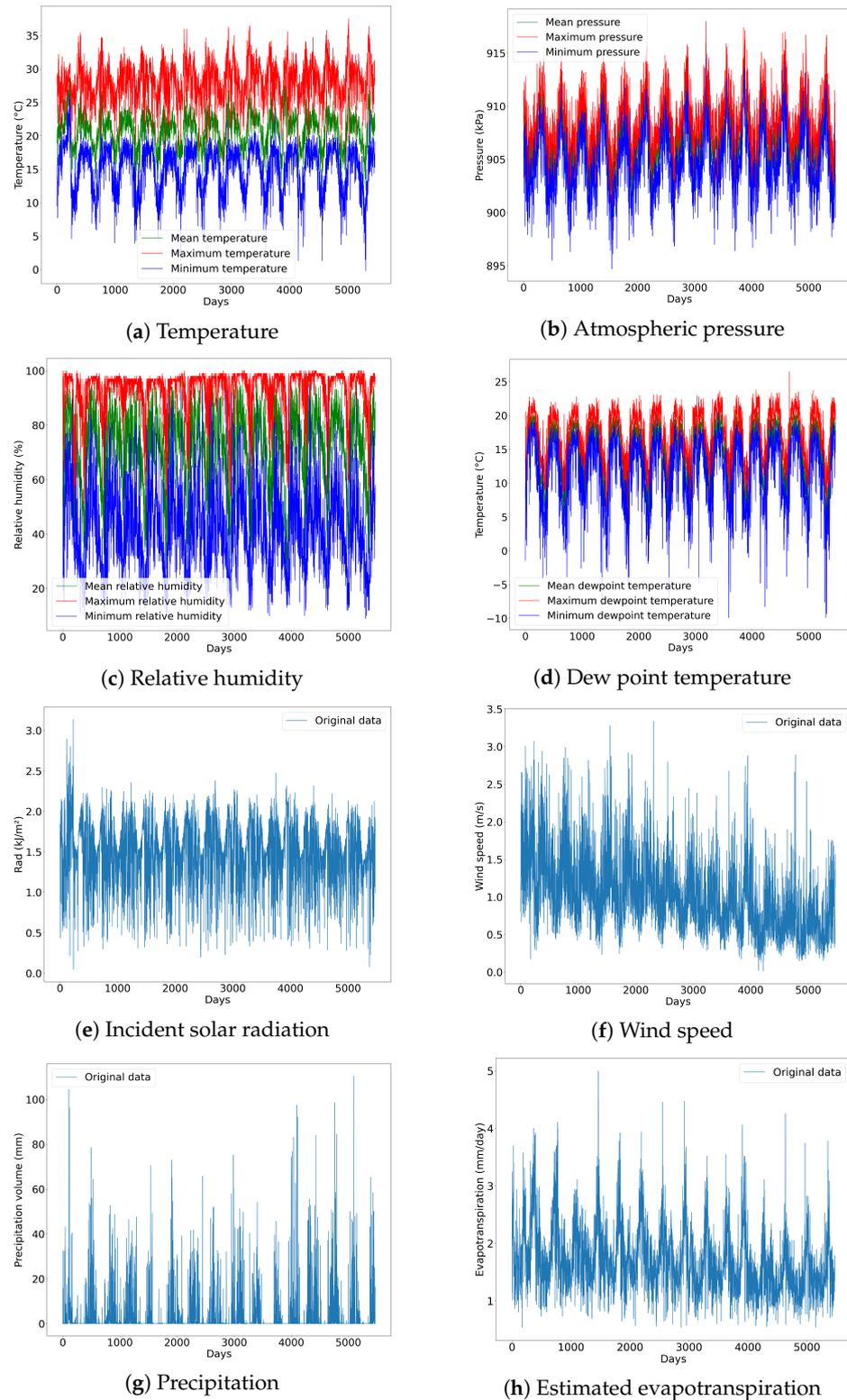


Figure 2. Visualization of the complete time series for each of the fifteen climatological variables observed for the experiments in this paper, totaling 5473 samples; (a–d) follow the same colour pattern, where the red line represents the maximum value for that variable, green for the mean observed value, and blue for the minimum value; (e–h) represent the daily observations for each variable.

3.2. Data Collectors

INMET employs the commercial Vaisala HydroMat™ model MAWS301 as a standard climatological data collector for automatic stations, responsible for data acquisition and communication with the governmental systems [53]. The climatological factors monitored hourly by INMET are as follows:

- Temperature [°C], available at maximum, minimum, and instant values;
- Relative humidity [%], available at maximum, minimum, and instant values;
- Dew point temperature [°C], available at maximum, minimum, and instant values;
- Atmospheric pressure [hPa], available at maximum, minimum, and instant values;
- Wind speed and gust [$\text{m} \cdot \text{s}^{-1}$], measured at 10 m height;
- Wind direction [degrees];
- Incident solar radiation [$\text{kJ} \cdot \text{m}^{-2}$];
- Precipitation [mm].

INMET provides climatological data hourly within an open-access policy from their automatic collectors after acquisition, communication, and pre-processing. The final dataset is available at <https://tempo.inmet.gov.br/TabelaEstacoes/xxxx> (accessed on 1 July 2023) (site in Brazilian Portuguese), where xxxx is the required station's code. The station code for the automatic collector in Patrocínio/MG is A523.

3.3. Soil Characterisation

The soil characteristics around the Patrocínio/MG municipality correspond to those of the Ferralsol reference soil group, according to the International Union of Soil Sciences classification, with a high concentration of hematite [54,55].

The physical water parameters for this soil (field capacity (FC) and wilting point (WP)) alongside the water retention curve were obtained via the pressure plate method applied to three different soil samples, with one from each layer, i.e., from 0 cm to 20 cm, from 20 cm to 40 cm, and 40 cm to 60 cm.

The experimental results suggest uniform physical water parameters throughout the three layers. As such, Figure 3 presents the resulting water retention curve for the three strata.

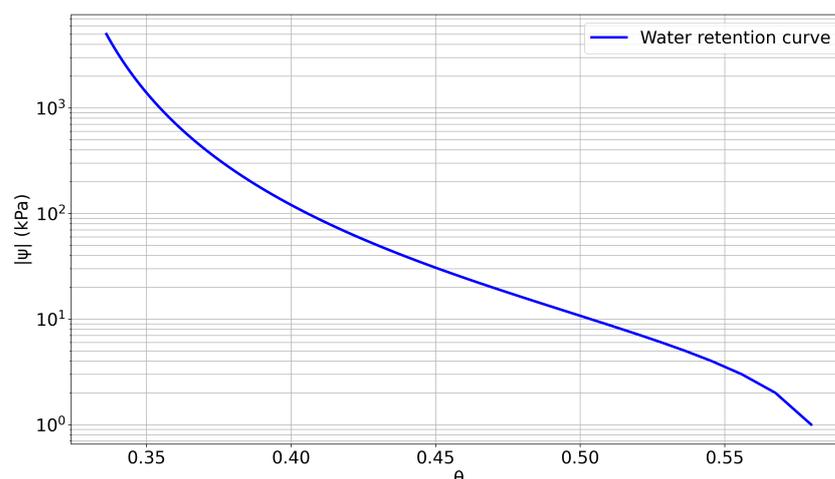


Figure 3. Semi-log representation for the water retention curve obtained from three soil samples from different depths near the Patrocínio/MG municipality. The uniform distribution of soil characteristics throughout the layers allows for the representation of a single water retention curve.

Lastly, Table 2 presents the van Genuchten equation parameters based on the observed soil retention curve in Figure 3. The authors include the parameters for each of the three layers as a matter of completeness despite their uniformity.

Table 2. Parameters for the van Genuchten equation obtained via the water retention curve and pressure plate method.

Layer	θ_s ($\text{cm}^3 \cdot \text{cm}^{-3}$)	θ_r ($\text{cm}^3 \cdot \text{cm}^{-3}$)	α (cm^{-3})	n	m
0–20 cm	0.59	0.31	0.25	1.33	0.25
20–40 cm	0.59	0.31	0.25	1.33	0.25
40–60 cm	0.59	0.31	0.25	1.33	0.25

3.4. Stochastic Modeling

The statistical models employed in this paper address three factors to assess the quality of the Bayesian regression for the evapotranspiration metric: (1) Which mathematical transformation applied to the target variable yields regression results with the lowest error metric? (2) The influence of autoregressive components on the quality of the regression. (3) An evaluation concerning the impact of the sample size on the regression learning process and its results.

The experiments evaluate five distinct operations over the evapotranspiration metric: linear— $f(x) = x$, Box-Cox transformation, square root, natural logarithm, and exponential. Equation (16) presents the definition of the one-parameter Box-Cox transformation [56] with an estimation of the value of the parameter λ from the dataset, which leads to better normality of the transformed series or is assumed to be known in each application. The number of autoregressive elements is, at most, three. Lastly, experiments employed the last 1000 or 2500 samples or the entire set with 5473 entries. A complete permutation of these components results in a total of sixty experiments.

$$f(x, \lambda) = \begin{cases} \frac{(x^\lambda - 1)}{\lambda}, & \lambda \neq 0 \\ \log(x), & \lambda = 0 \end{cases} \quad (16)$$

Equations (17)–(23) establish the experiments' framework based on the posterior distribution in Equation (10). Let $N \in \{1000, 2500, 5473\}$ represent the number of samples for each experiment and X_i represent each of the fifteen climatological variables employed, according to Section 3.1. Equation (17) establishes the target series y composed of the evapotranspiration series after applying one of the five equations presented, while Equation (18) provides a prior probability distribution assumed for this series based on a Gaussian density function.

Equations (19a) and (19b) describe the parametrization of the hyperparameter μ_i as a linear combination of the climatological elements. Equation (19a) applies when no autoregressive window is present, while Equation (19b) denotes the experiments with $w \in \{1, 2, 3\}$ of such components.

Lastly, Equations (20)–(23) provide initial, approximately non-informative prior distributions for the parameter τ and regression parameters β_j . Equation (23) only applies to the experiments with at least one autoregressive component.

The authors employed MultiBUGS 2.0 software [57] (available at <https://www.multibugs.org/>, accessed on 15 August 2023). Each experiment contained five chains, 1000 samples as the burn-in, and one in every ten values for the next 5000 obtained for the estimates.

$$y_i = f(ET_{0i}) \quad i = \{1, 2, \dots, N\} \quad (17)$$

$$y_i \sim \mathcal{N}(\mu_i, \tau) \quad i = \{1, 2, \dots, N\} \quad (18)$$

$$\mu_i = \beta_0 + \sum_{j=1}^{15} \beta_j X_{ji} + \varepsilon \quad i = \{1, 2, \dots, N\} \quad (19a)$$

$$\mu_i = \beta_0 + \sum_{j=1}^{15} \beta_j X_{ji} + \sum_{k=1}^w \beta_k y_{(i-k)} + \varepsilon \quad i = \{1, 2, \dots, N\},$$

$$w \in \{1, 2, 3\} \quad (19b)$$

$$\tau \sim \Gamma(1, 1) \quad (20)$$

$$\beta_0 \sim \mathcal{N}(4, 0.01) \quad (21)$$

$$\beta_j \sim \mathcal{N}(0, 1) \quad j = \{1, 2, \dots, 15\} \quad (22)$$

$$\beta_k \sim \mathcal{N}(0, 1) \quad k = \{1, \dots, w\} \quad (23)$$

3.5. Structural Learning

To build a BN that presents the best likelihood to the climatological and evapotranspiration data, the authors employed both the standard K2 greedy algorithm [58] and a bio-inspired score-and-search algorithm [59].

The relevant nodes and the amount of required data were consonant with the results of the set of sixty experiments described in Section 3.3, resulting in eight climatological variables, alongside evapotranspiration as the target component and a dataset with 1000 samples.

The parameters for both methods were the default ones, with the Bayesian Dirichlet equivalence with the uniform prior metric (BDeu) [60] as a scoring function and k-fold (k = 5) cross-validation. Both methods resulted in the same graphical structure.

4. Results and Discussion

This section presents the results and relevant discussions about the proposed experiments. Appendix A presents the complete tables for the numeric results of all the experiments, and presents the graphical comparisons of the evapotranspiration metric, considering the inference task of each climatological variable.

4.1. Selection of Climatological Variables

The root mean square error (RMSE) metric for each of the sixty experiments against the expected time series for the Penman–Monteith equation over the original dataset ranges from $0.086 \text{ mm} \cdot \text{day}^{-1}$ to $3.502 \text{ mm} \cdot \text{day}^{-1}$. These results range from the experiment with logarithmic transformation, one autoregressive element, and 1000 samples, to the test with the exponential function, no autoregressive terms, and 5473 samples, respectively. Table A1 presents the RMSE results for the sixty experiments proposed in Section 3.4, alongside their mean absolute error (MAE) metrics; the closer the value of each error metric to zero, the better adherence between the two series.

The total time required to reach convergence for each experiment on MultiBUGS depends mainly on the amount of data provided to the software. Experiments with 1000 samples took 12.2 min, 2500 samples required 36.3 min, and the complete dataset needed 67.3 min, on average. These times were achieved on a desktop with an Intel® Core™ i5-6600 CPU at 3.30 GHz, 8 GB of RAM, and operating a Windows® 10 Pro 64-bit.

The first finding is that the RMSE metric exhibits minor fluctuations within the same transformation but higher fluctuations across functions. For instance, the experiments within the logarithm function present RMSE from $0.086 \text{ mm} \cdot \text{day}^{-1}$ to $0.100 \text{ mm} \cdot \text{day}^{-1}$, while experiments within the Box–Cox transform—with each λ parameter automatically obtained by the transformation algorithm for each series—achieve RMSEs from $0.097 \text{ mm} \cdot \text{day}^{-1}$ to $0.107 \text{ mm} \cdot \text{day}^{-1}$. These results suggest that the mathematical function choice is a relevant first step to minimize the estimation error based on this approach.

Within the same mathematical transformation, when comparing the error metric according to the number of samples of each experiment, the error metric consistently presents the lowest value for experiments with 2500 samples. Although this improvement is marginal compared to the errors from those with 1000 samples, this result suggests the

existence of an optimal data volume for the inference task on these variables, and a large volume of data may impair the inference task by introducing more noise into the model.

Finally, when observing the error performance based on the amount of past information provided to the model, there is a marginal improvement across all sixty experiments for each past value added. This result is consistent with the observed meteorological dynamics, as the amount of water evaporated and transpired from the previous day plays a role in the maintenance of relative humidity, dew point temperature, and surface temperature for the current day, which are factors that directly impact the evapotranspiration phenomenon.

The exponential transformation is the one that consistently presents inference results with the highest error metric, according to Table A1. The results and discussions from this point onward do not consider the experiments with the exponential function, given their errors within one or two orders of magnitude greater than those obtained by the other four transformations.

The output of the MultiBUGS software includes the interval at 95% credibility for each beta regression parameter and the most probable value within this interval. As such, a climatological element is relevant if the credible interval for its beta coefficient does not encompass zero, as any value for the climatological variable would represent no impact in the regression.

Table A2 contains a complete overview of the occurrence of each climatological variable deemed relevant by the regression model, grouped by transformation. Across all forty-eight experiments—excluding the twelve from the exponential function—precipitation and all three measurements for atmospheric pressure do not contribute to the regression model. The variables for maximum dew point temperature and minimum temperature and dew point temperature contributed to a few experiments.

Eight variables consistently supported the regression analysis in more than half of the twelve experiments for each function. Wind speed and net solar radiation are relevant across all forty-eight experiments; mean temperature and dew point temperature, maximum temperature and relative humidity, and minimum relative humidity are significant for several experiments. Mean relative humidity is the last of the relevant climatological variables, although this component did not contribute to any regression with the square root function.

The high relevance observed from the wind speed and net solar radiation throughout all experiments is consistent with the observed meteorological phenomena. Both components directly contribute to the regional climate, represented by temperature, relative humidity, and dew point temperature. The irrelevance of the precipitation is also within expectation, given the assumption that elements within a much wider region contribute to this component with their dynamics not considered here. The atmospheric pressure components directly impact the dew point temperature, and the absence of this variable from the set of relevant variables is likely due to a discrepancy in the magnitude of the pressure measurements compared to the other elements.

While Table A1 shows that the natural logarithm is the function that provides a regression with the lowest errors, Table A2 presents this function as the most assertive to which variables are most relevant to the stochastic regression for evapotranspiration. Further, the error improvements within this transformation are negligible after providing more data and autoregressive components, given the trade-off of obtaining these. Based on these considerations, the authors establish that the stochastic model with a natural logarithm transformation and no autoregressive elements, trained with 1000 samples, shows the best adherence to the evapotranspiration metric.

Table 3 presents the numerical values achieved for the beta coefficients from the experiment utilizing natural logarithm transformation on the evapotranspiration data, with 1000 samples for the parameter learning and no autoregressive component. The coefficients and their respective climatological variables—whose 95% credible intervals do not contain zero and, as such, characterize a relevant variable in this context—are highlighted in blue. The independent coefficient β_0 for this experiment is $\beta_0 \approx 0.6373$.

Based on the maximum a posteriori (MAP) values of each beta parameter and the relationships detailed in Equations (17) and (19a), Equation (24) presents the approximation for the hierarchical Bayesian regression for the evapotranspiration metric; this is based on the experiment identified as the best among the sixty conducted. Although this representation for Equation (24) is similar to linear regression, each beta coefficient corresponds to a density distribution function, and its best point estimator value lies within the non-symmetrical 95% credible interval.

$$\begin{aligned} \ln(ET_0) \approx & 0.6373 - 0.024 \cdot T_{mean} + 0.022 \cdot T_{max} \\ & - 0.006 \cdot RH_{mean} - 0.005 \cdot RH_{max} - 0.010 \cdot RH_{min} \\ & + 0.028 \cdot DP_{mean} + 0.498 \cdot W_{speed} + 0.113 \cdot Rad \end{aligned} \quad (24)$$

Table 3. Credible interval at 95% for the beta regression coefficients obtained from the experiment with natural logarithm transformation, using a data partition of 1000 samples, and excluding the autoregressive window feature. The relevant beta coefficients (i.e., those that do not include zero on the interval or whose value is close to zero) are highlighted in blue.

Variable	Coefficient	2.5%	97.5%	MAP Value
Tmean	β_1	−0.035	−0.010	−0.024
Tmax	β_2	0.014	0.028	0.022
Tmin	β_3	−0.008	0.003	0
RHmean	β_4	−0.009	−0.004	−0.006
RHmax	β_5	−0.005	−0.004	−0.005
RHmin	β_6	−0.011	−0.008	−0.010
DPmean	β_7	0.014	0.039	0.028
DPmax	β_8	−0.005	0.006	0
DPmin	β_9	−0.008	0.004	0
Pmean	β_{10}	−0.001	0.002	0
Pmax	β_{11}	−0.003	0.001	0
Pmin	β_{12}	−0.003	0.001	0
Wspeed	β_{13}	0.484	0.513	0.498
Rad	β_{14}	0.093	0.133	0.113
Prec.	β_{15}	−0.001	0.000	0

Figure 4a contains a graphical comparison of the evapotranspiration time series based on the Penman–Monteith equation (blue continuous line) against the estimated time series from Equation (24) (red line) throughout 1000 samples. This result suggests that the proposed method and subsequent equation properly conciliated the series' non-stationarity.

It is relevant to evaluate if the error component follows a Gaussian distribution for the estimated time series against the expected one, as the proposed Bayesian regression in Equation (24) is similar to the model in Equation (19a). To this end, Figure 4b presents a quantile–quantile plot for the distribution of the residual errors. As the blue dots lie close to the red 45-degree line, the distribution for the error is close to a Gaussian distribution, as required by the theoretical model.

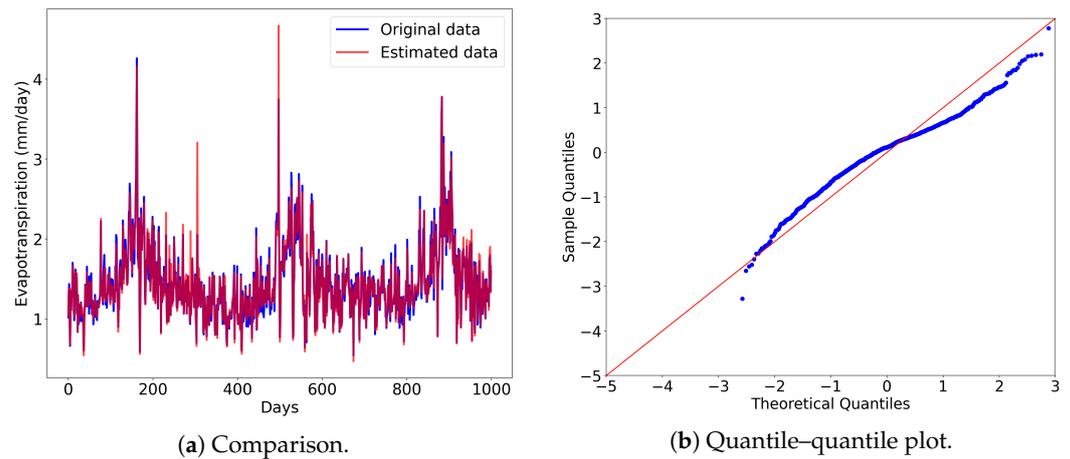


Figure 4. Graphical representation of the results obtained from the experiment with natural logarithm transformation, using a data partition of 1000 samples, and excluding the autoregressive window feature; (a) presents the original time series as the blue line and the estimated time series as the red line; (b) presents the quantile–quantile plot of the residuals for the estimated time series.

One relevant comparison is to ascertain the proposed method’s adherence to the expected evapotranspiration according to the gold-standard FAO56 method against the evapotranspiration estimated by other classical and deterministic approaches. As such, Figure 5 presents a graphical comparison between the natural logarithm of the evapotranspiration from the FAO56 method (blue line), the Benavides and Lopez method (orange line), and the Hargreaves–Samani method (green line); all three are based on the same dataset, alongside the estimate from the proposed method (red line). This result confirms the ability of the proposed approach to closely model the complex behavior of the FAO56 estimate, with an adherence better than the ones presented by other classical equations.

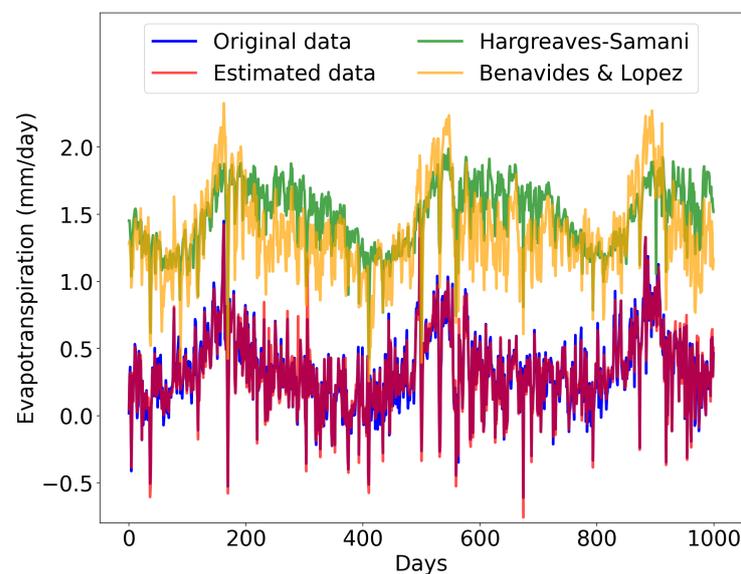


Figure 5. Graphical comparison of the results obtained from the experiment with natural logarithm transformation, using a data partition of 1000 samples, and excluding the autoregressive window feature (as the red line). The blue line represents the natural logarithm of the reference evapotranspiration according to the FAO-56 equation; the orange and green lines represent the natural logarithm for the reference evapotranspiration according to the Benavides and Lopez and the Hargreaves–Samani equations, respectively. The latter two methods represent some of the analytical approximations commonly employed under the unavailability of the climatological components required by the FAO-56 equation.

4.2. Stochastic Climatological Water Balance Comparison with In Situ Measurements

A proxy measurement for the water balance dynamics is necessary, mainly due to the characteristics of the evapotranspiration metric. Given the measurement hardship, this paper employs an estimate for the available soil water content on a coffee plantation in the proximity of Patrocínio/MG based on tensiometer readings and the van Genuchten equation.

Although the authors cannot divulge the precise location of this plantation due to confidentiality concerns, Equation (25) presents the estimated soil water content θ as a function of the matric potential ψ (kPa), according to the original Equation (7). For this particular soil, $\theta_s \approx 59\%$ and $\theta_r \approx 31\%$, $\alpha \approx 0.25 \text{ cm}^{-1}$, $n \approx 1.33$ and $m \approx 0.25$, based on the experiments described in Section 3.3. Tensiometer readings were provided by a PalmaFlex digital tensiometer, model SL-PF-TNSE.

$$\theta = vG(\psi) \approx 0.31 + \frac{0.28}{\left(1 + (0.25|\psi|)^{1.33}\right)^{0.25}} \quad (25)$$

Figure 6 presents the relationships between rainfall, estimated soil water content, and estimated evapotranspiration on the top panel, based on in situ measurements from a tensiometer monitoring the substrate from the surface to a depth of 20 cm, and from a pluviometer, with both sampling data from the 9th to the 25th January 2021. In the highlighted period (from the 9th to the 15th), the observed rainfall is enough to keep the soil water content close to its maximum capacity ($\approx 59\%$). From the 16th onward, when there is no observed rainfall, the curves for soil water content and evapotranspiration have a symmetrical relationship: the higher the estimated evapotranspiration, the faster the reduction in soil water content. This characteristic against the in situ measurements reinforces the Penman–Monteith method as adequate to model the consumptive use of water due to evaporation and transpiration.

The middle panel of Figure 6 presents the influence on the estimated evapotranspiration by the two most relevant climatological variables, according to the MAP values in Equation (24). This panel exhibits the correlations for evapotranspiration trends with wind speed ($\beta_{13} \approx 0.113$) and net incident solar radiation ($\beta_{14} \approx 0.498$). These relationships of the target variable with the two most relevant climatological factors, according to the Bayesian regression model, combined with the adherence observed in the first panel against the in situ measurements, suggest the adequacy of the proposed method to represent the hydrological phenomenon, even under scenarios where the traditional method cannot be employed.

Lastly, the third panel compares the analytical time series for evapotranspiration according to the Penman–Monteith metric and the results from Equation (24) over the same dataset. These three panels suggest that this method models the expected analytical time series and the underlying real-world dynamics.

These results support the adequacy of the proposed method in tracking the analytical FAO-56 equation and functioning as a framework to model dynamic and complex phenomena based on non-linear, non-stationary time series and their underlying factors.

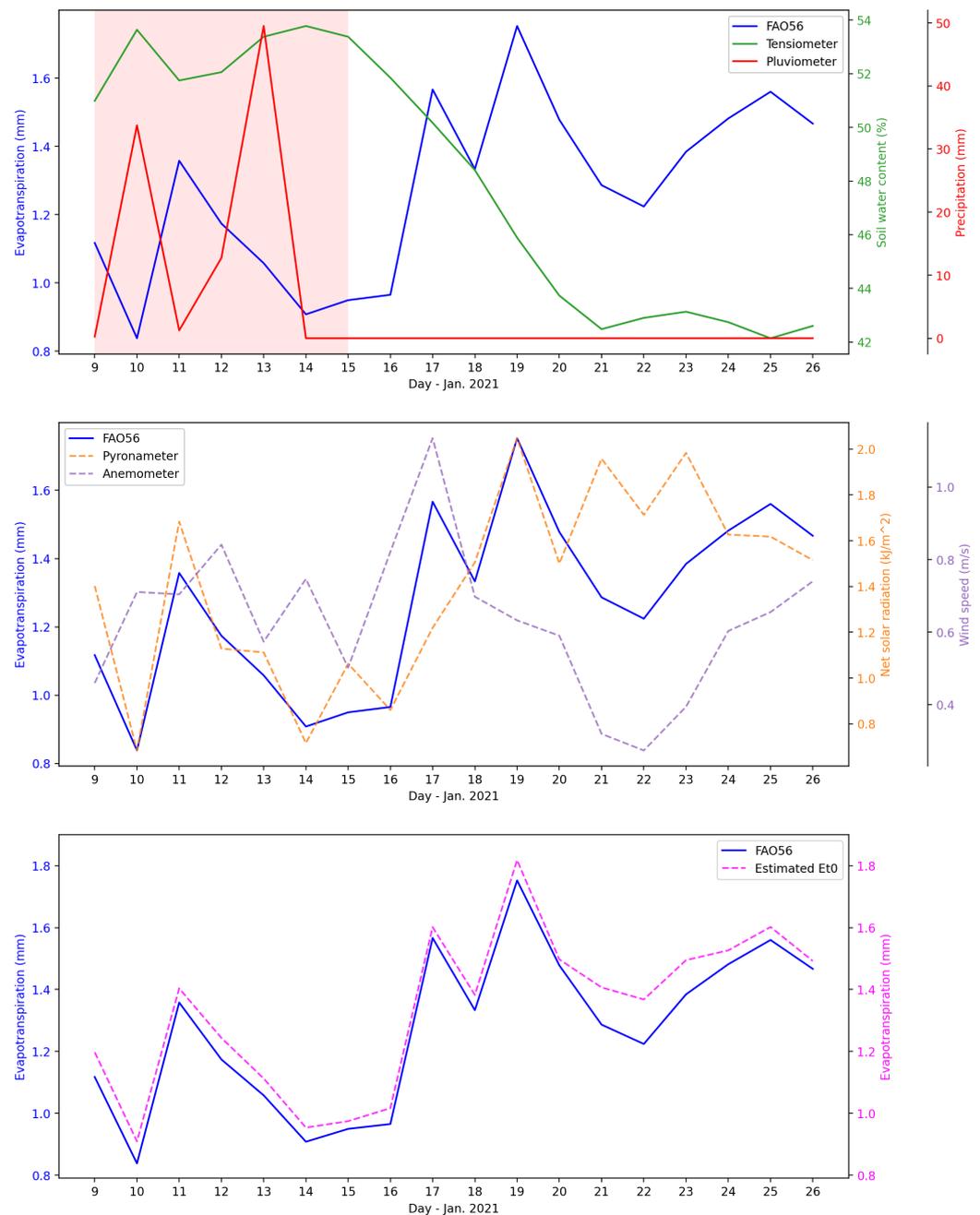


Figure 6. Graphical representation of the climatological water balance compared to in situ samplings from the 9th to the 26th of January 2021. The top panel contains a time series for the two principal factors of the climatological water balance—rainfall and evapotranspiration—compared to the dynamics of the soil water percentage. The area highlighted in red represents the days when the water intake to the plantation due to rainfall was sufficient to counteract the amount of water evaporated and transpired. The middle panel presents the interaction between the estimated evapotranspiration, wind speed, and incident net solar radiation, the latter two being the most relevant factors for evapotranspiration estimation based on the results presented in Section 4.1. Lastly, the bottom panel presents the estimated evapotranspiration based on the FAO-56 method compared to the values obtained by Equation (24), applied to the same dataset during the established timeframe.

4.3. Stochastic Evapotranspiration Inference with Complete or Incomplete Data

To evaluate the model's adequacy in estimating evapotranspiration over a dataset with new information, Figure 7 presents a graphical comparison of the natural logarithm of the evapotranspiration calculated with the Penman–Monteith equation against the time series obtained from Equation (24). This new dataset, comprising 334 samples not used in the learning process, contains complete information on all the climatological variables presented in Section 3.1.

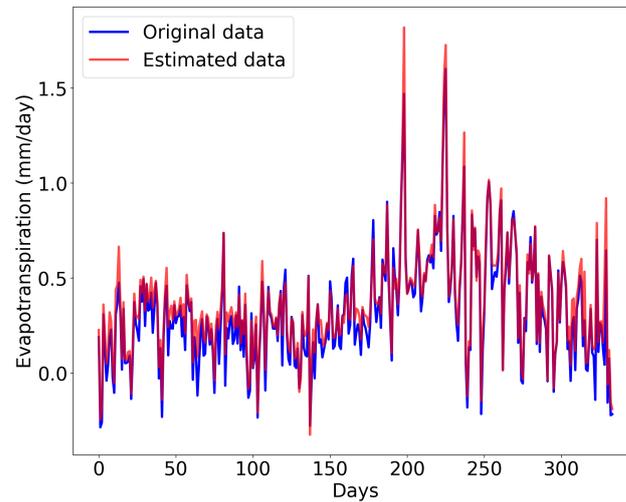


Figure 7. Graphical comparison of the reference evapotranspiration from the Penman–Monteith equation over the complimentary dataset against that obtained via the regression presented in Equation (24). The RMSE metric between the natural logarithm of both series is $0.087 \text{ mm} \cdot \text{day}^{-1}$, consistent with the results in Table A1.

Another way to check the model's performance is to compare the expected time series for the Penman–Monteith equation against the result from Equation (24) when dealing with climatological data from regions with different climate profiles. Figure 8a–d provide a graphic comparison similar to that in Figure 7, but this time it is applied to the cities described in Table 1, with RMSE metrics and distances from Patrocínio presented in Table 4. As expected, the RMSE error for each region increases as the distance from the original city increases, reinforcing the notion that Penman–Monteith equation relies on local climatological factors for an accurate estimate.

Table 4. Error metrics for the natural logarithm of the reference evapotranspiration derived from the Penman–Monteith equation versus the values from Equation (24), based on the climatological data from different cities in Brazil with different climate profiles compared to Patrocínio/MG.

City	RMSE ($\text{mm} \cdot \text{day}^{-1}$)	Distance to Patrocínio/MG (km)
Rio Preto da Eva/AM	0.202	2278
Cabaceiras/PB	0.140	1723
Porangatu/GO	0.102	659
Santo Augusto/RS	0.185	1202

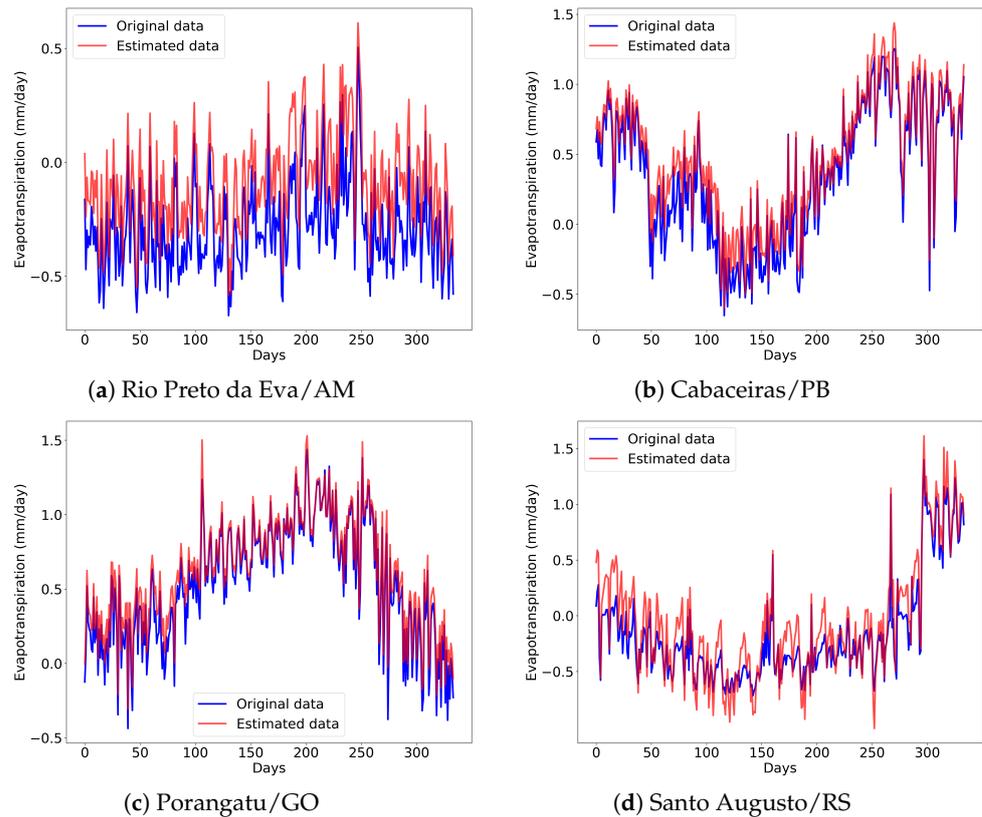


Figure 8. Graphical comparison of the reference evapotranspiration from the Penman–Monteith equation over the complimentary dataset for each different city presented in Table 1, against the value obtained via the regression presented in Equation (24) over that city’s climatological time series; (a) presents the comparison for Rio Preto da Eva/AM, which achieved $RMSE = 0.202 \text{ mm} \cdot \text{day}^{-1}$; (b) presents the comparison for Cabaceiras/PB, which achieved $RMSE = 0.140 \text{ mm} \cdot \text{day}^{-1}$; (c) presents the comparison for Porangatu/GO, which achieved $RMSE = 0.102 \text{ mm} \cdot \text{day}^{-1}$; lastly, (d) presents the comparison for Santo Augusto/RS, which achieved $RMSE = 0.185 \text{ mm} \cdot \text{day}^{-1}$.

One shortcoming of every analytical approach to estimating evapotranspiration is their prerequisite of having all the required climatological variables available. Under the Bayesian framework, the BN provides a probabilistic approach to supply the most probable value for missing data based on complete or incomplete evidence from the system.

Figure 9 presents a graphical representation of the BN’s DAG, which provides the best posterior likelihood estimates based on the original dataset containing 1000 samples. The nodes represent only the random variables relevant to the Bayesian regression from Table 3 instead of the complete set of fifteen variables. This approach reduces the graph search space for the learning algorithm and employs only the relevant climatological variables, which decreases the number of required components and sensors. The evapotranspiration node is highlighted in blue; the omission of the conditional probability tables from the BN is intentional due to the high dimensionality of each one.

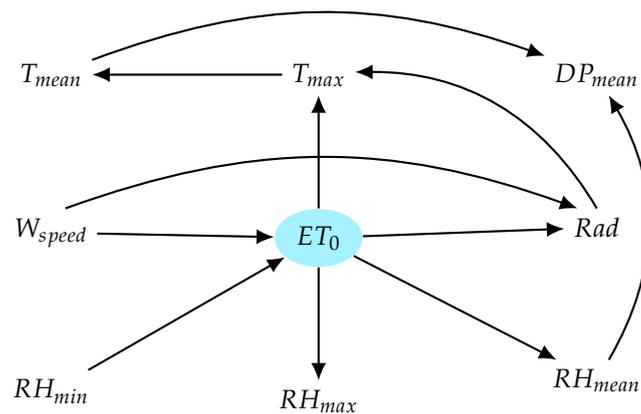


Figure 9. Bayesian network for estimating the natural logarithm of reference evapotranspiration from the climatological variables most frequently present as relevant, according to the set of experiments with natural logarithm transformation.

After learning the conditional probability tables based on the maximum likelihood of the model over the training dataset, the BN framework represented in Figure 9 allows for queries about the most probable state of a climatological variable, providing the Bayesian regression analysis with a new level of flexibility to deal with missing data. For instance, if a meteorological data collector experiences a failure in its anemometer, an estimate of the most probable value for wind speed may allow for the use of the underlying methods that require this variable, albeit within an error margin. Figure A1 presents the natural logarithm of the expected evapotranspiration time series against the values obtained from Equation (24) after replacing each climatological variable from the original dataset with values inferred from the BN.

As expected, wind speed estimates based on the BN provide a time series with the lowest adherence to the baseline time series. This result occurs because the wind speed is one of the independent variables in the BN, resulting in the inference task for this variable becoming a query for the most probable value that would induce the observed values for the other variables in the system. This type of inverted query increases the error margin for the estimated value and, alongside the relatively high MAP value for wind speed's beta coefficient, culminates in the amplification of the inference error on the final estimate for evapotranspiration when wind speed data are lacking.

Table 5 presents the RMSE metric for the evapotranspiration estimates according to Equation (24), considering the removal and subsequent inference of each one of the climatological variables from the BN in Figure 9 and the expected evapotranspiration time series. The variables presenting the highest error values are wind speed and minimum relative humidity, which is reasonable within the BN structure as both components are independent elements of the BN. The lowest error values are from the maximum relative humidity component, likely due to its position on the graphical structure, where only one random variable influences it, and it does not influence any others, and due to the lowest absolute MAP value of the beta parameter for maximum relative humidity in Equation (24).

The results in Table 5 suggest that the BN approach provides efficient estimates for the required variable under scenarios lacking a single component and allows for the maintenance and usage of methods that may require these missing factors.

Table 5. Error metric for the natural logarithm of the reference evapotranspiration metric from the Penman–Monteith equation against the value from Equation (24), after the removal of each relevant climatological variable and the inference of its value from the BN in Figure 9.

Removed and Inferred Variable	Evapotranspiration RMSE (mm · day ⁻¹)
T_{mean}	0.087
T_{max}	0.092
RH_{mean}	0.087
RH_{max}	0.074
RH_{min}	0.156
DP_{mean}	0.085
W_{speed}	0.286
Rad	0.096

5. Conclusions

In this paper, the authors provided and discussed a complete stochastic and Bayesian framework to model a real-world application based on non-deterministic and non-stationary data.

The first result shows how the Bayesian regression analysis is affected by the employed mathematical pre-processing operation, the amount of provided data, and the number of available autoregressive components for this application. Based on the RMSE metric, the natural logarithm function is the best candidate to model the evapotranspiration time series.

The established baseline Bayesian regression equation provides a close approximation for the evapotranspiration metric modeled by the Penman–Monteith equation, and for the dynamics of soil water moisture measured by tensiometers. As such, the results from the entirely data-driven approach are coherent with the traditional analytical method and the real-world phenomenon it proposes to model.

The Bayesian regression analysis also highlights the subset of random variables relevant to the stochastic model. We leverage this newfound insight to build another Bayesian structure in the form of a BN, as the knowledge of relevant variables allows for a reduction in the graph’s search space.

Lastly, we show the adequacy of probabilistic queries over the BN structure to estimate climatological variables required for the analytical methods. Our results show that the BN toolset provides synthetic data with a high degree of adherence to real-world data and, in turn, allows for the application of subsequent methods. As expected from the theoretical basis, this inference task presents the worst performance when estimating a value for independent nodes on the graphical structure. The flexibility the BN provides for analytical methods may even offset the time and computational power required to learn an adequate configuration for the query process.

Author Contributions: Conceptualisation, C.C.M.J. and C.D.M.; data curation, V.P.R. and L.D.N.; formal analysis, V.P.R., P.A.A.M., J.A.A. and C.D.M.; investigation, V.P.R., A.W.C.J. and C.C.M.J.; methodology, V.P.R., L.D.N., P.A.A.M., J.A.A., C.D.M. and J.A.P.B.; project administration, C.D.M. and J.A.P.B.; software, V.P.R. and L.D.N.; supervision, C.D.M. and J.A.P.B.; validation, P.A.A.M., J.A.A. and C.D.M.; writing—original draft, V.P.R., A.M.J., C.D.M. and J.A.P.B.; writing—review and editing, L.D.N., P.A.A.M., J.A.A., A.W.C.J., C.D.M. and J.A.P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the National Council for Scientific and Technological Development (CNPq), grant nos. 142390/2019-4 (V.R.), 406774/2022-6 (P.M.), 301923/2019-1 (J.A.), 465755/2014-3 (C.M.), and 301853/2018-5 (J.B.), and by the São Paulo Research Foundation (FAPESP), grant nos. 2014/50851-0 (C.M.), USP-FAPESP-IBM 2019/07665-4 (L.N. and C.M.), and BPE-FAPESP 2018/19150-6 (C.M.). The APC was funded by Espectro Ltda., under grant no. 001/2019.

Data Availability Statement: Publicly available datasets were analyzed in this study concerning climatological variables. These datasets can be found at <https://tempo.inmet.gov.br/TabelaEstacoes/xxxx> (accessed on 1 July 2023) (site in Brazilian Portuguese), where xxxx is the required station code. Please note that there are restrictions on the availability of data from tensiometer readings. Data were obtained from a proprietary coffee plantation site under the condition of anonymity and cannot be made publicly available.

Acknowledgments: The authors thank the Foundation for UNESP's Development (FUNDUNESP), UNESP's Innovation Agency (AUIN), and the Centre for Artificial Intelligence (C4AI-USP).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Appendix A.1. Complete Table for the Error Metrics

Table A1. The RMSEs and MAEs of all sixty experiments, combining each proposed transformation, original dataset partition, and autoregressive window. From a scoring standpoint, every experiment achieved error metrics within the same order, except for those with the exponential transformation.

Transformation	N° of Samples	Window	RMSE (mm · day ⁻¹)	MAE (mm · day ⁻¹)
Linear	1000	0	0.115	0.074
		1	0.112	0.073
		2	0.112	0.073
		3	0.112	0.073
	2500	0	0.118	0.076
		1	0.116	0.075
		2	0.116	0.075
		3	0.116	0.075
	5473	0	0.139	0.090
		1	0.134	0.089
		2	0.134	0.088
		3	0.133	0.088
Box-Cox	1000	0	0.107	0.054
		1	0.105	0.054
		2	0.105	0.054
		3	0.105	0.054
	2500	0	0.097	0.059
		1	0.095	0.058
		2	0.095	0.058
		3	0.095	0.058
	5473	0	0.101	0.070
		1	0.098	0.069
		2	0.097	0.068
		3	0.097	0.068

Table A1. Cont.

Transformation	N° of Samples	Window	RMSE (mm · day ⁻¹)	MAE (mm · day ⁻¹)
Square root	1000	0	0.094	0.061
		1	0.092	0.061
		2	0.092	0.061
		3	0.092	0.061
	2500	0	0.093	0.062
		1	0.092	0.062
		2	0.092	0.061
		3	0.092	0.061
	5473	0	0.108	0.074
		1	0.105	0.072
		2	0.104	0.072
		3	0.104	0.072
Natural logarithm	1000	0	0.088	0.054
		1	0.086	0.053
		2	0.086	0.054
		3	0.086	0.053
	2500	0	0.088	0.058
		1	0.087	0.057
		2	0.086	0.057
		3	0.086	0.057
	5473	0	0.100	0.070
		1	0.098	0.069
		2	0.097	0.068
		3	0.097	0.068
Exponential	1000	0	2.408	1.010
		1	2.296	0.979
		2	2.283	0.988
		3	2.282	0.993
	2500	0	2.190	1.029
		1	2.129	1.007
		2	2.130	1.010
		3	2.128	1.009
	5473	0	3.502	1.609
		1	3.458	1.560
		2	3.430	1.561
		3	3.429	1.560

Appendix A.2. Complete Table for the Relevant Variables

Table A2. The frequency of each climatological variable found relevant for each set of twelve experiments, according to the transformation applied to the evapotranspiration metric. The variables most frequently present are highlighted in blue.

Variable	Linear	Box-Cox	Square Root	Natural Logarithm
T_{mean}	6	12	8	12
T_{max}	9	12	12	12
T_{min}	3	4	1	4
RH_{mean}	8	10	0	11
RH_{max}	11	8	4	12
RH_{min}	11	12	12	12
DP_{mean}	6	12	8	12
DP_{max}	2	1	0	0
DP_{min}	8	1	1	2
P_{mean}	0	0	0	0
P_{max}	0	0	0	0
P_{min}	0	0	0	0
W_{speed}	12	12	12	12
Rad	12	12	12	12
$Prec.$	0	0	0	0
$ET_0[-1]$	9/9	9/9	9/9	9/9
$ET_0[-2]$	4/6	4/6	4/6	4/6
$ET_0[-3]$	1/3	1/3	1/3	2/3

Appendix A.3. Graphical Comparison between the Expected Time Series and Inferred Data

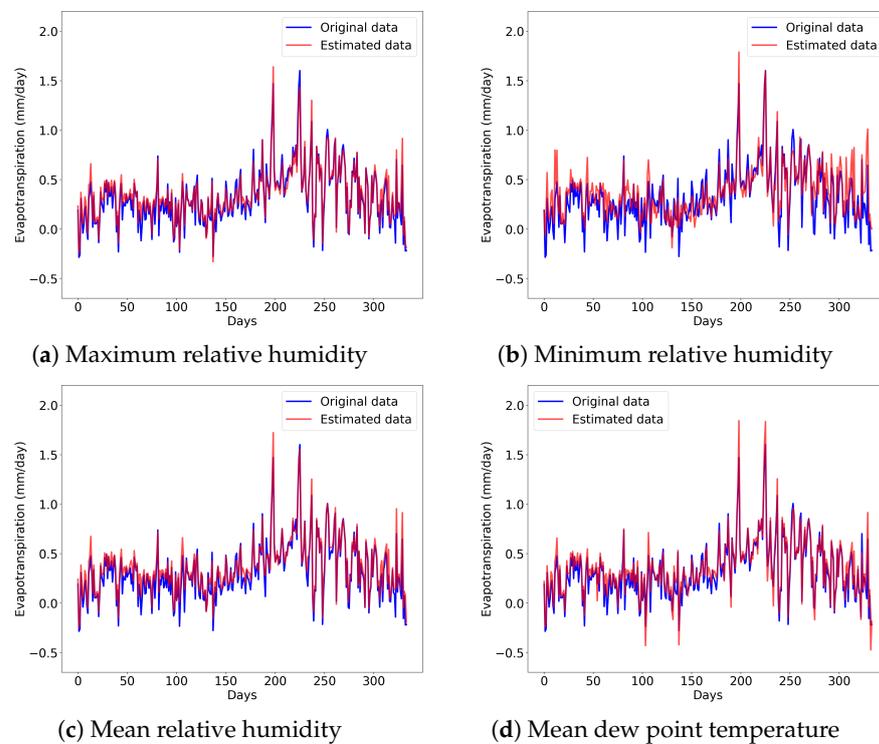


Figure A1. Cont.

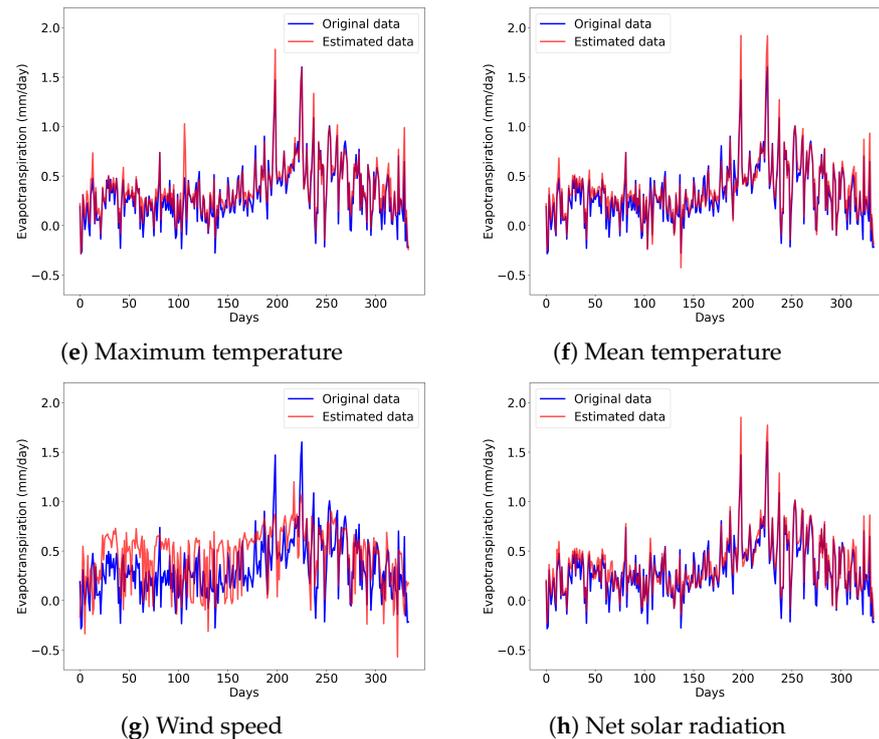


Figure A1. A graphical comparison between the expected time series and the results obtained after each climatological variable was removed from the dataset; subsequent inference of the components based on the Bayesian Network structure.

References

- Lu, Y.; Yuan, J.; Lu, X.; Su, C.; Zhang, Y.; Wang, C.; Cao, X.; Li, Q.; Su, J.; Ittekkot, V.; et al. Major threats of pollution and climate change to global coastal ecosystems and enhanced management for sustainability. *Environ. Pollut.* **2018**, *239*, 670–680. [CrossRef]
- Chen, B.; Han, M.; Peng, K.; Zhou, S.; Shao, L.; Wu, X.; Wei, W.; Liu, S.; Li, Z.; Li, J.; et al. Global land-water nexus: Agricultural land and freshwater use embodied in worldwide supply chains. *Sci. Total Environ.* **2018**, *613–614*, 931–943. [CrossRef]
- Jägerbrand, A.K.; Brutemark, A.; Barthel Svedén, J.; Gren, I.M. A review on the environmental impacts of shipping on aquatic and nearshore ecosystems. *Sci. Total Environ.* **2019**, *695*, 133637. [CrossRef]
- Hecht, J.S.; Lacombe, G.; Arias, M.E.; Dang, T.D.; Piman, T. Hydropower dams of the Mekong River basin: A review of their hydrological impacts. *J. Hydrol.* **2019**, *568*, 285–300. [CrossRef]
- Koumparou, D. The Right of Thirst: Water as a Human Right and as a Commons. *Glob. Nest J.* **2018**, *20*, 637–645. [CrossRef]
- United Nations Department for Economic and Social Affairs. *Sustainable Development Goals Report 2020*; United Nations: New York, NY, USA, 2020.
- Empresa de Pesquisa Energética. Anuário Estatístico de Energia Elétrica 2022. 2022. Available online: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-160/topico-168/Fact%20Sheet%20-%20Anu%20C3%A1rio%20Estat%20-%20ADstico%20de%20Energia%20El%20C3%A9trica%202022.pdf> (accessed on 14 July 2023).
- FAO. FAOSTAT: Crops and Livestock Products. 2021. Available online: <https://www.fao.org/faostat/en/#data/QCL> (accessed on 2 February 2023).
- Rasera, J.B.; Silva, R.F.D.; Piedade, S.; Mourão Filho, F.D.A.A.; Delbem, A.C.B.; Saraiva, A.M.; Sentelhas, P.C.; Marques, P.A.A. Do Gridded Weather Datasets Provide High-Quality Data for Agroclimatic Research in Citrus Production in Brazil? *AgriEngineering* **2023**, *5*, 924–940. [CrossRef]
- Bwambale, E.; Abagale, F.K.; Anornu, G.K. Data-driven model predictive control for precision irrigation management. *Smart Agric. Technol.* **2023**, *3*, 100074. [CrossRef]
- Porter, J.; Xie, L.; Challinor, A.; Cochrane, K.; Howden, S.; Iqbal, M.; Lobell, D.; Travasso, M. Chapter 7: Food Security and Food Production Systems. In *Food Security and Food Production Systems. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2014; pp. 485–533.
- Neupane, D.; Adhikari, P.; Bhattarai, D.; Rana, B.; Ahmed, Z.; Sharma, U.; Adhikari, D. Does Climate Change Affect the Yield of the Top Three Cereals and Food Security in the World? *Earth* **2022**, *3*, 45–71. [CrossRef]
- Zhang, C.; Xie, Z.; Wang, Q.; Tang, M.; Feng, S.; Cai, H. AquaCrop modeling to explore optimal irrigation of winter wheat for improving grain yield and water productivity. *Agric. Water Manag.* **2022**, *266*, 107580. [CrossRef]

14. Silva Fuzzo, D.F.; Carlson, T.N.; Kourgialas, N.N.; Petropoulos, G.P. Coupling remote sensing with a water balance model for soybean yield predictions over large areas. *Earth Sci. Inform.* **2020**, *13*, 345–359. [[CrossRef](#)]
15. Allen, R.; Pereira, L.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper 56; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
16. Salam, R.; Islam, A.R.M.T. Potential of RT, bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *J. Hydrol.* **2020**, *590*, 125241. [[CrossRef](#)]
17. Ghiat, I.; Mackey, H.R.; Al-Ansari, T. A Review of Evapotranspiration Measurement Models, Techniques and Methods for Open and Closed Agricultural Field Applications. *Water* **2021**, *13*, 2523. [[CrossRef](#)]
18. Goyal, P.; Kumar, S.; Sharda, R. A review of the Artificial Intelligence (AI) based techniques for estimating reference evapotranspiration: Current trends and future perspectives. *Comput. Electron. Agric.* **2023**, *209*, 107836. [[CrossRef](#)]
19. Brekel, J.; Thorp, K.R.; DeJonge, K.C.; Trout, T.J. Version 1.1.0—pyfao56: FAO-56 evapotranspiration in Python. *SoftwareX* **2023**, *22*, 101336. [[CrossRef](#)]
20. Aghelpour, P.; Norooz-Valashedi, R. Predicting daily reference evapotranspiration rates in a humid region, comparison of seven various data-based predictor models. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 4133–4155. [[CrossRef](#)]
21. Chia, M.Y.; Huang, Y.F.; Koo, C.H.; Fung, K.F. Recent Advances in Evapotranspiration Estimation Using Artificial Intelligence Approaches with a Focus on Hybridization Techniques—A Review. *Agronomy* **2020**, *10*, 101. [[CrossRef](#)]
22. Jing, W.; Yaseen, Z.M.; Shahid, S.; Saggi, M.K.; Tao, H.; Kisi, O.; Salih, S.Q.; Al-Ansari, N.; Chau, K.W. Implementation of evolutionary computing models for reference evapotranspiration modeling: Short review, assessment and possible future research directions. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 811–823. [[CrossRef](#)]
23. Ma, N.; Szilagyi, J.; Zhang, Y. Calibration-Free Complementary Relationship Estimates Terrestrial Evapotranspiration Globally. *Water Resour. Res.* **2021**, *57*, e2021WR029691. [[CrossRef](#)]
24. McMahon, T.A.; Peel, M.C.; Lowe, L.; Srikanthan, R.; McVicar, T.R. Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1331–1363. [[CrossRef](#)]
25. Liu, J.; Hertel, T.W.; Lammers, R.B.; Prusevich, A.; Baldos, U.L.C.; Grogan, D.S.; Frolking, S. Achieving sustainable irrigation water withdrawals: Global impacts on food security and land use. *Environ. Res. Lett.* **2017**, *12*, 104009. [[CrossRef](#)]
26. Paek, J.; Pollanen, M.; Abdella, K. A Stochastic Weather Model for Drought Derivatives in Arid Regions: A Case Study in Qatar. *Mathematics* **2023**, *11*, 1628. [[CrossRef](#)]
27. Chidzalo, P.; Ngare, P.O.; Mung'atu, J.K. Pricing weather derivatives under a tri-variate stochastic model. *Sci. Afr.* **2023**, *21*, e01768. [[CrossRef](#)]
28. Oliveira, R.P.D.; Achcar, J.A.; Chen, C.; Rodrigues, E.R. Non-homogeneous Poisson and linear regression models as approaches to study time series with change-points. *Commun. Stat. Case Stud. Data Anal. Appl.* **2022**, *8*, 331–353. [[CrossRef](#)]
29. Elmer, J.; Coppler, P.J.; Jones, B.L.; Nagin, D.S.; Callaway, C.W.; on behalf of University of Pittsburgh Post-Cardiac Arrest Service. Bayesian Outcome Prediction After Resuscitation From Cardiac Arrest. *Neurology* **2022**, *99*, e1113–e1121. [[CrossRef](#)] [[PubMed](#)]
30. Shedbalkar, K.H.; More, D.S. Bayesian Regression for Solar Power Forecasting. In Proceedings of the 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP), Vijayawada, India, 12–14 February 2022; pp. 1–4. [[CrossRef](#)]
31. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.
32. Hui, E.; Stafford, R.; Matthews, I.M.; Smith, V.A. Bayesian networks as a novel tool to enhance interpretability and predictive power of ecological models. *Ecol. Inform.* **2022**, *68*, 101539. [[CrossRef](#)]
33. Ojha, R.; Ghadge, A.; Tiwari, M.K.; Bititci, U.S. Bayesian network modelling for supply chain risk propagation. *Int. J. Prod. Res.* **2018**, *56*, 5795–5819. [[CrossRef](#)]
34. Li, H.; Guedes Soares, C.; Huang, H.Z. Reliability analysis of a floating offshore wind turbine using Bayesian Networks. *Ocean. Eng.* **2020**, *217*, 107827. [[CrossRef](#)]
35. Neapolitan, R. *Learning Bayesian Networks*; Pearson: Upper Saddle River, NJ, USA, 2003. [[CrossRef](#)]
36. Hargreaves, G.H.; Samani, Z.A. Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* **1985**, *1*, 96–99. [[CrossRef](#)]
37. Linacre, E.T. A simple formula for estimating evaporation rates in various climates, using temperature data alone. *Agric. Meteorol.* **1977**, *18*, 409–424. [[CrossRef](#)]
38. Turc, L. Estimation of irrigation water requirements, potential evapotranspiration: A simple climatic formula evolved up to date. *Ann. Agron.* **1961**, *12*, 13–49.
39. Benavides, J.G.; Lopez, D. Formula para el calculo de la evapotranspiracion potencial adaptada al tropico (15° N–15° S). *Agron. Trop.* **1970**, *20*, 335–345.
40. Cruz, T.A.C.D.; Marques, P.A.A. Low-cost irrigation management system: Improving data confidence through artificial intelligence. *Eng. Agrícola* **2023**, *43*, e20210164. [[CrossRef](#)]
41. Silva-Júnior, R.O.; Souza-Filho, P.W.M.; Salomão, G.N.; Tavares, A.L.; Santos, J.F.; Santos, D.C.; Dias, L.C.; Silva, M.S.; Melo, A.M.Q.; Costa, C.E.A.d.S.; et al. Response of Water Balance Components to Changes in Soil Use and Vegetation Cover Over Three Decades in the Eastern Amazon. *Front. Water* **2021**, *3*, 749507. [[CrossRef](#)]
42. Raghunath, H.M. *Hydrology: Principles, Analysis and Design*, 2nd ed.; New Age International (P) Ltd.: New Delhi, India, 2006.
43. Mutti, P.R.; Dubreuil, V.; Bezerra, B.G.; Arvor, D.; Funatsu, B.M.; Santos E Silva, C.M. Long-term meteorological drought characterization in the São Francisco watershed, Brazil: A climatic water balance approach. *Int. J. Climatol.* **2022**, *42*, 8162–8183. [[CrossRef](#)]

44. Saha, A.; Sekharan, S.; Manna, U. Evaluation of Capacitance Sensor for Suction Measurement in Silty Clay Loam. *Geotech. Geol. Eng.* **2020**, *38*, 4319–4331. [[CrossRef](#)]
45. Van Genuchten, M. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils¹. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [[CrossRef](#)]
46. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2020.
47. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; Wiley Series in Probability and Mathematical Statistics; Wiley: New York, NY, USA, 1981.
48. Gelman, A.; Hill, J.; Vehtari, A. *Regression and Other Stories*; Cambridge University Press: Cambridge, UK, 2020.
49. Doucet, A.; Wang, X. Monte Carlo methods for signal processing: A review in the statistical signal processing context. *IEEE Signal Process. Mag.* **2005**, *22*, 152–170. [[CrossRef](#)]
50. Cox, L.A., Jr. Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. *Crit. Rev. Toxicol.* **2018**, *48*, 682–712. [[CrossRef](#)] [[PubMed](#)]
51. Pearl, J.; Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*, 1st ed.; Basic Books: New York, NY, USA, 2020.
52. Gross, T.J.; Bessani, M.; Darwin Junior, W.; Araújo, R.B.; Vale, F.A.C.; Maciel, C.D. An analytical threshold for combining Bayesian Networks. *Knowl.-Based Syst.* **2019**, *175*, 36–49. [[CrossRef](#)]
53. Moura, A.D.; Lucas, E.W.M.; Rodrigues, J.E.; de Rezende, J.M. Nota Técnica No. 001/2011/SEGER/LAIME/CSC/INMET. 2019. Available online: http://www.cemtec.ms.gov.br/wp-content/uploads/2019/02/Nota_Tecnica-Rede_estacoes_INMET.pdf (accessed on 31 July 2023).
54. Santos, H.G.d.; Jacomine, P.K.T.; Anjos, L.H.C.d.; Oliveira, V.A.d.; Lumbrreras, J.F.; Coelho, M.R.; Almeida, J.A.d.; Araujo Filho, J.C.d.; Oliveira, J.B.d.; Cunha, T.J.F. *Brazilian Soil Classification System*; Embrapa: Brasília, Brazil, 2018.
55. IUSS Working Group WRB *World Reference Base for Soil Resources. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*, 4th ed.; International Union of Soil Sciences (IUSS): Vienna, Austria, 2022.
56. Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1964**, *26*, 211–243. [[CrossRef](#)]
57. Goudie, R.J.B.; Turner, R.M.; De Angelis, D.; Thomas, A. MultiBUGS: A Parallel Implementation of the BUGS Modeling Framework for Faster Bayesian Inference. *J. Stat. Softw.* **2020**, *95*, 1–20. [[CrossRef](#)] [[PubMed](#)]
58. Cooper, G.F.; Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **1992**, *9*, 309–347. [[CrossRef](#)]
59. Ribeiro, V.P.; Pereira, B.R., Jr.; Maciel, C.D.; Balestieri, J.A.P. An Improved Bayesian Network Super-Structure Evaluation using Physarum polycephalum Bio-inspiration. In Proceedings of the XXIV Brazilian Congress on Automation, Fortaleza, Brazil, 16–19 October 2022.
60. Buntine, W. Theory Refinement on Bayesian Networks. In Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence, Los Angeles, CA, USA, 13–15 July 1991; UAI'91, pp. 52–60.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.