

Modeling of reference evapotranspiration by multiple linear regression

Helder J. F. da Silva^{*}, Marconio S. dos Santos^{***}, Jório B. Cabral Junior^{*}, Maria H. C. Spyrides^{**}

^{*} Federal University of Rio Grande do Norte - UFRN, Postgraduate Program in Climate Sciences - PPGCC, Campus Universitário, Lagoa Nova, Natal-RN, Brazil. Phone: 55+84 32153883. E-mail: helderlagoia@gmail.com (Corresponding author)

^{**} Federal University of Rio Grande do Norte - UFRN, Department of Atmospheric and Climate Sciences - DCAC, Campus Universitário, Lagoa Nova, Natal-RN, Brazil. Phone: 55+84 32153787.

^{***} Federal University of Rio Grande do Norte - UFRN, Department of Mathematics - DMAT, Campus Universitário, Lagoa Nova, Natal-RN, Brazil. Phone: 55+84 32153883.

Received 22 November 2015; accepted 20 April 2016

Abstract

Evapotranspiration is an important parameter for many projects related to climate characterization, hydrological modeling and water resources. This work was established as the first study in Rio Branco, eastern Acre, in order to derive empirical relations to estimate the reference evapotranspiration in the annual range from meteorological data readily available using the multiple linear regression analysis. Meteorological data of mean temperature (maximum and minimum), wind speed and insolation were obtained from the National Institute of Meteorology for the period 1980-2014, which can be considered representative of the local climate. To estimate the reference evapotranspiration was used the Penman-Monteith-FAO, and multiple regression analysis was used as a selection process of significant variables for the model fit. Generated values by the proposed evapotranspiration models were compared to observed values for validation. Results indicated that the model with three variables (mean temperature, wind speed and insolation) satisfactorily estimated reference evapotranspiration for Rio Branco, AC, with great performance for annual data. Models with one variable (insolation) and two variables (mean temperature and insolation) showed less accuracy. However, they have advantage due to simplicity, since they can estimate the reference evapotranspiration from a few climatic parameters. From a practical point of view, these models can be regarded as a method to estimate the reference evapotranspiration when the input weather variables are insufficient to other methods.

Keywords: Variability, Rio Branco, numerical modeling.

1. Introduction

Evapotranspiration (ET) is one of the most fundamental processes, which influence climate and weather, both global and local scales, consisting of the link between energy, climate and hydrology (Braun et al., 2001). It is estimated, through evapotranspiration, that 60% to 80% of total precipitation returns to the atmosphere, a fact which gives it the status of primary regulator of water availability, superficial and subterranean, as well as from various human

activities, for example, agriculture (Victoria, 2004).

Knowledge about variability of ET is essential for climatic characterization of a region. Along with other factors, it conditions the understanding about energetic and hydrologic partitioning between surface and atmosphere for different terrestrial biomes, it is used by managers to predictions and weather forecasts as well as the trends analysis studies (Alkaeed et al., 2006; Some'e et al., 2012).

The reference evapotranspiration (ET_o) is one that occurs in a large area with low

vegetation in active growth, fully covering the ground with height between 0.08 and 0.15 m, without water restriction. For the definition, it is noteworthy that the undergrowth "grass" is specifically defined as reference culture and this culture is considered free of water stress and diseases (Mendonça et al., 2003).

The ETo depends on the power supply and heat that, in turn, depend on meteorological factors such as temperature, relative humidity, wind speed and solar radiation (Xu and Singh, 1998; Brutsoert, 2005). In hydrological practice, ETo can be obtained by direct (measured) and indirect (estimated) methods. Direct methods, using lysimeters (water balance), they are generally used in extensive research projects due to the high cost of equipment and maintenance. Indirect methods are less expensive and are based on the application of mathematical methods using climatic data measured at weather stations (Pereira et al., 1997).

Among the various indirect methods to estimate ETo, there is the Penman-Monteith method, parameterized and recommended by the Food and Agriculture Organization of the United Nations (FAO) as the standard method for ETo estimation and calibration of empirical methods, due to its better performance when applied in various types of climates, and good accuracy when compared with lysimeters measures (Allen et al., 1998). Camargo and Camargo (2000), Bernardo (2002) and Carvalho et al. (2011) also emphasized that Penman-Monteith-FAO method estimates efficiently ETo under different conditions of atmospheric moisture. The main limitation of this method is the difficulty in obtaining all necessary meteorological data (about 10 variables), which are not always available in some weather stations, as well as a complex mathematical calculation involving approximately 16 other necessary mathematical subroutines to obtain the final value of ETo.

Thus, it is important to apply statistical techniques in order to develop simple empirical equations to estimate ETo. As a result, many methods have been developed over the past decades to an estimated ETo based on empirical equations and/or physical reasons. Many of these methods, however, have variations, for reasons of local adjustments and calibrations, raising further uncertainties, requiring thus regional checks and calibrations. For example, Ferraz (2008) and

Souza (2009) compared ETo values in southwestern Amazon, east of Acre, obtained by nine and four estimation methods, respectively. The authors used the Penman-Monteith-FAO model as default and, through linear regression, showed that the most simplified methods such as Hargreaves-Samani (Samani, 2000) and Ivanovi (Dorfman, 1977) demanded regional arrangements.

This work focused on investigating the empirical relations between the estimates of reference evapotranspiration in Rio Branco city, eastern of Acre, in annual scale as function of meteorological data easy to obtain and measured at weather stations of the National Institute of Meteorology (Instituto Nacional de Meteorologia – INMET), in the period 1980-2014, using multiple regression analysis. For the development of statistical analysis, it was used the R software, created in 1996 by Ross Ihaka and Robert Gentleman (R, 2005), it is a language and environment that allows data manipulation for statistical computing and graphics. Thus, this paper also intends to present in parallel using the R software, explaining step by step each function and commands employee. Moreover, the performance was examined for each of the selected models, comparing them to select the best fit with more accurate estimates.

2. Materials and methods

2.1 Study Area

The study area corresponds to Rio Branco city, capital of Acre (AC), which has an area of 9,222.58 km², located in the eastern portion of the state, between the coordinates 9°58'29" South latitude and 67°48'36" West longitude and altitude 159 m, as showed in Figure 1. This municipality comprises 336,038 habitants and its population density is 38.03 hb/km² (IBGE, 2010).

2.1 Climatic Characterization

Local climatology, considering the period of the last 35 years (1980-2014) in Rio Branco - AC, is characterized by mean maximum temperature 31.5 °C and mean minimum temperature 20.5 °C. The mean compensated temperature varies between 23.5 °C in July and 26.0 °C in October, with annual mean 25.1 °C (Silva, 2015). The mean of monthly cumulative rainfall ranges from 35.6 mm (August) to 293.2

mm (March), with mean of annual accumulated 1963.4 mm during the study period. Two distinct seasons occurs: the rainy and the less rainy seasons. The less rainy season begins in June and ends in October. The rainy season begins in December and runs until April, characterized by

constant and abundant rainfall. May and November can be characterized as months of transition from the rainy to the less rainy season, and from the less rainy to the rainy, respectively (Duarte, 2006; Silva, 2015).

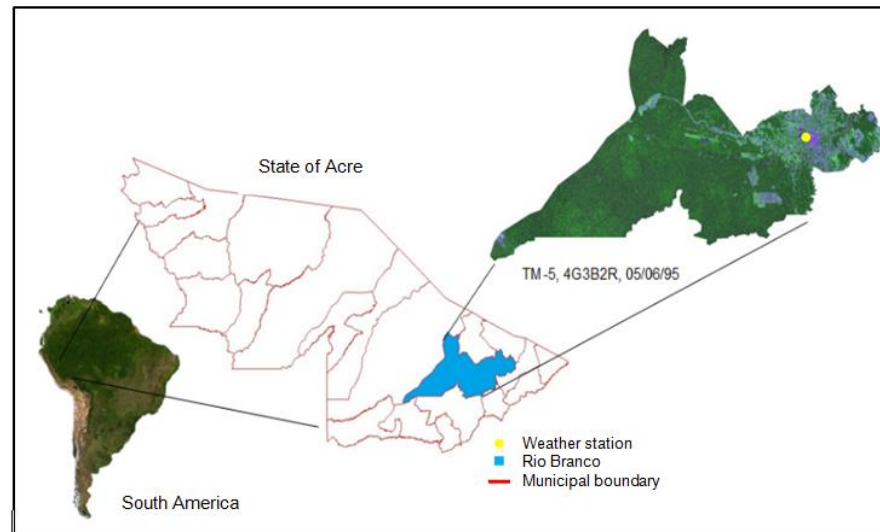


Figure 1 - Location of the study area - Rio Branco (AC).

The main weather systems operating in the region and contributing to the formation of clouds and rain in the rainy season are convective mechanisms and moisture transport from the Atlantic Ocean evaporation and evapotranspiration in the Amazon rainforest, this time cumulus of great vertical development are intensified. For less rainy season the greatest contributions to any rain are local convection and cold fronts, coming from southern Brazil, causing the phenomenon known as cold waves (known locally as “friagens”). These are the result of the advancement of a Polar Front driven by a Polar Atlantic air mass advancing the Chaco Plain to the Western Amazon, bringing down temperature, reaching up to 10 °C minimum in the region (Ferraz, 2008; Duarte, 2006). The mean of monthly relative air humidity varies from 77.6% (August) to 87.6% (February), with annual mean 84.3%.

2.3 Data and procedures

For this research were used meteorological data from the period 1980-2014 of a conventional weather station of the National Institute of Meteorology (INMET), specifically the unit located in Rio Branco, AC, with geographic coordinates: 9° 58' 12''S latitude and 67° 52' 12''W longitude and altitude 160 m. Figure 2 show the time series of collected monthly meteorological data, and converted to annual basis, which were: mean air temperature (Tmed, °C); mean maximum temperature (Tmax, °C); mean minimum temperature (Tmin, °C); wind speed measured at 2 m height (U2, m s⁻¹) and insolation (n, hours).

Regarding the missing data on weather variables, it was used the MTSDI (Multivariate Time Series Data Imputation) which is an algorithm to fill missing data in multivariate normal time series based on the Expectation Maximisation (EM) algorithm proposed by Junger and Leon (2012) and described by Silva (2015).

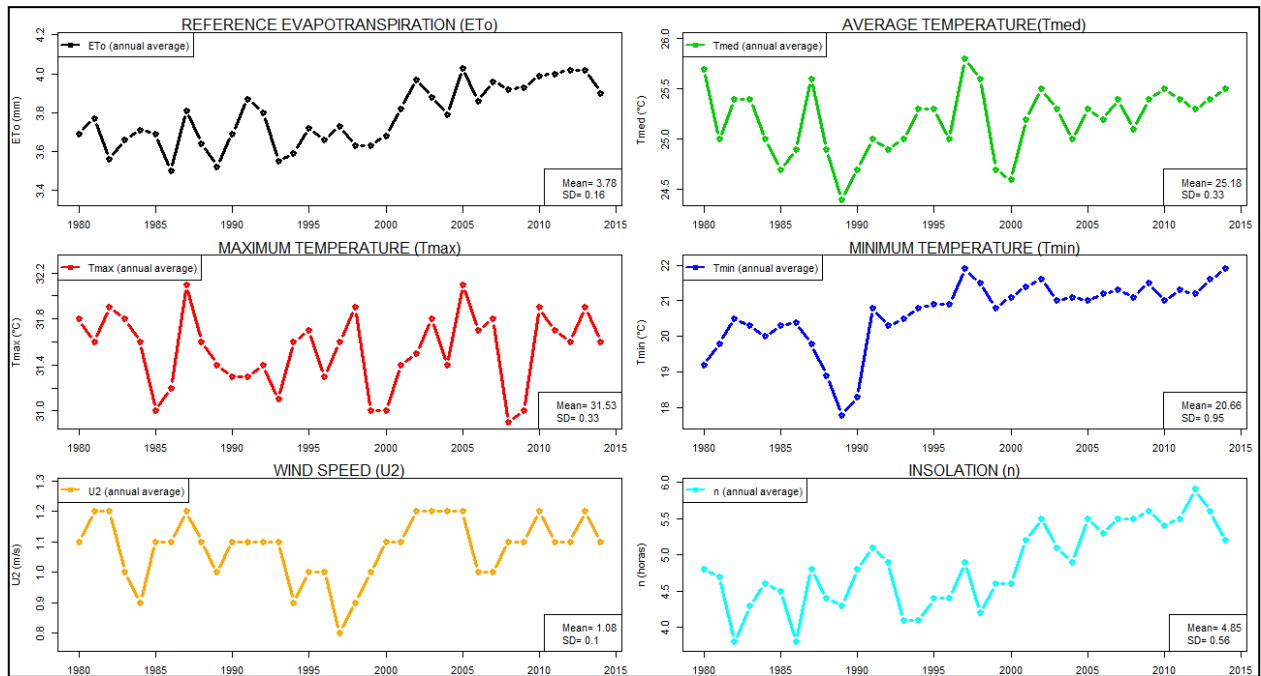


Figure 2 - Time series of annual mean of reference evapotranspiration (ETo), mean temperature (Tmed), mean maximum temperature (Tmax), mean minimum temperature (Tmin), wind speed (U₂) and insolation (n) in Rio Branco, AC - 1980-2014.

After completing the data, it was calculated the ETo estimate according to the Penman-Monteith-FAO model. This considers the grass height fixed in 0.12 m, stomatal

resistance of 70 m s⁻¹ and albedo of 0.23. The value of the heat flow in soil is considered to be zero, its estimate is calculated by Equation 1 (Allen et al., 1998):

$$ETo = \frac{0,408\Delta(R_n - G) + \gamma \left(\frac{900U_2}{T_{med} + 273} \right) (e_s - e_a)}{\Delta + \gamma(1 + 0,34U_2)} \quad (1)$$

where “ γ ” is the slope of the vapor pressure curve with temperature (kPa°C⁻¹); “ R_n ” it is the daily radiation balance (MJm⁻²dia⁻¹); “ G ” is the daily total heat flow in the soil (MJm⁻²dia⁻¹); “ Δ ” is the psychrometric coefficient (kPa°C⁻¹); “ U_2 ” is the wind speed at 2 m height (m s⁻¹); “ e_s ” is the vapor saturation pressure (kPa); “ e_a ” is the current pressure steam (kPa); “ T_{med} ” is the mean temperature (°C).

The values of “ U_2 ” and “ T_{med} ” are measured at the weather station, it is need to calculate the values of “ R_n ”, “ G ”, “ γ ”, “ Δ ”, “ e_s ” and “ e_a ”, which in turn require other meteorological parameters, such as: relative humidity, maximum and minimum temperature, atmospheric pressure, solar radiation and others. Details of calculations for these parameters are

found in the technical circulaire No 65 issued by EMBRAPA (Conceição, 2006) and also Silva (2015).

Then, it was used the Pearson correlation analysis (r) to find out the degree of relationship between independent variables and ETo (dependent variable), as well as the existence of multicollinearity and influential points. The correlation coefficient (r) assumes only values between -1 and 1. If the absolute value of this correlation closes to 1, one say that the relationship between the dependent and independent variables is more intense. The mathematical expression used for obtaining the correlation coefficient (r) is suggested as Spiegel (1998), given by:

$$r = \frac{S_{xy}}{S_x S_y} \quad (2)$$

with,

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3)$$

where S_{xy} = sample covariance; S_x = standard deviation of the independent variable data series; S_y = standard deviation of dependent data; x_i = weather variable for the period under study; $\bar{x}$ = mean weather variable for the period under study; y_i = daily number of ETo; $\bar{y}$ = average ETo.

The regression analysis is mainly used with the purpose to estimate (estimating the value of Y from the value of X) and estimate how much X affects the variability of Y. In this study, it was used the multiple linear regression model (Equation 4) with k independent variables:

$$Y_j = \beta_0 + \sum_{i=1}^k \beta_i X_{ij} + \varepsilon_j, \quad (4)$$

where, Y_j is the observed value of ETo; β_0 it is the linear coefficient; β_i is the regression coefficient of the independent variable X_i ; X_{ij} are the j-th observations of the independent variables X_i ; and ε_j is the error associated with the variable Y on the observation j. In this model, the variable Y is a linear function of the parameters. The error values are approximately normally distributed with zero mean, are homoscedastic (constant variance) and independent, these are the conditions to use of linear regression (Draper and Smith, 1996; Souza, 1998).

The parameters of multiple linear regression equations were estimated assuming that ETo was the dependent variable and the others were independent variables. The most used method for estimating the parameters (β_0 , β_i) is the method of least squares. In simple line regression, this method ensures that the straight line obtained is the one for which the vertical distances (squared) between the observed values

Y and the fitted values are smallest. (Souza, 2007). Linear regression has the advantage of including many variables simultaneously, and checking how each of these variables affects independent variable Y, when all other variables are assumed constant in the model.

It is noteworthy that a variable selection method *forward* was used, so that the variables were included in the model according to the significance order shown in their individual analysis with ETo. Thus, they are included in the model, one by one (Model_1).

For model adjustment purposes, it was used the set of annual data from 1980 to 2009, with 30 observed values; and the annual data from 200 to 2014, with 5 observed values, was used to validate the final model chosen.

After model adjustment, the assessment of its quality may be effected by the coefficient of determination (R^2) of multiple regression, given by:

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_{\text{obs}} - Y_{\text{est}})^2}{\sum_{i=1}^N (Y_{\text{obs}} - \bar{Y})^2}, \quad (5)$$

A high value of R^2 , close to 1.0, indicates a good model fit to the observed data. However, R^2 also increases as the number of explanatory variables used in the model increases, even if these are not

significant variables to explain the variability of the dependent variable. Therefore, it is preferable to use the adjusted R^2 , defined by:

$$R_a^2 = 1 - \frac{n-1}{n-p} \times (1 - R^2), \quad (6)$$

where R_a^2 is the adjusted coefficient of determination; n is the sample size and p is the number of explanatory variables.

In addition the adjusted coefficient of determination, the performance of the models was evaluated by statistical tests, such as: Mean

Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), agreement coefficient ("d") and index of performance ("c"). These

rates can be calculated according to equations proposed by Wilmott et al. (1985):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{E_i - O_i}{E_i} \right| \times 100 \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - O_i)^2} \quad (8)$$

$$d = 1 - \frac{\sum_{i=1}^n (E_i - O_i)^2}{\sum_{i=1}^n (|E_i - O_i| + |O_i - O|)^2} \quad (9)$$

$$c = |r \times d| \quad (10)$$

where, O_i e E_i are the observed and the simulated values (estimated), respectively. For better model fit, MAPE and RMSE statistics should be near zero. For the values "c" of Equation (10), the

performance of the estimates were categorized according to Camargo and Sentelhas (1997), Table 1.

Table 1- Values of index of performance ("c") proposed by Camargo and Sentelhas (1997).

Value of "c"	Performance
> 0.85	Great
0.76 to 0.85	Very good
0.66 to 0.75	Good
0.61 to 0.65	Median
0.51 to 0.60	Tolerable
0.41 to 0.50	Bad
≤ 0.40	Very Bad

To test the hypothesis that there are no $\beta_1 = \beta_2 = \dots = \beta_k = 0$ is equivalent to testing the hypothesis that there is no linear association

between the values of the independent variables and the dependent variable. This hypothesis is tested by the F-test, given by:

$$F = R^2 (n-p-1) / (1-R^2) p \quad (11)$$

where $(n - p - 1)$ is the degree of freedom, n is the sample size and p is the number of model parameters. When the calculated value of F is greater than F tabulated for a given significance

level (α) (it was used 5%), it is rejected the hypothesis that all β_i are equal to zero. The F-test also indicates whether the regression equation is significant, that is, if the established functional

relationship between the dependent variable and the combined effects of the independent variables is relevant.

A second test was applied, the t-Test, to verify the significance (correlation) of regression equation coefficients ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$) individually. The purpose of this test is to verify if the estimated value for each coefficient is significantly different from 0 (zero). These definitions can be seen in Draper and Smith (1996) and Spiegel (1998).

After performing the regression, it is relevant to verify if the fitted model is in agreement with the assumptions: the residuals are normally distributed with mean zero, homoscedastic (constant variance) and independent. This is done with the graphical analysis of the residuals (difference between observed value and predicted values by the regression model) and through some statistical tests such as Kolmogorov-Smirnov (Thode Junior, 2002), it which tests normality and uses the null hypothesis that the data are normally distributed, and also Breusch-Pagan tests (Kramer and Sonnberger, 1986) and Durbin-Watson (Racine e Hyndman, 2002) for constant variance (homoscedasticity) and independence of the data, respectively.

3. Results and discussion

Figure 3 contains the correlation matrix between the ETo and the independent variables (Tmed, Tmax, Tmin, U2 and n), with their respective Pearson correlation coefficient values (r) and statistical significance (p-value) of this correlation obtained through commands in Appendix 1. It is observed that all the correlations between the explanatory variables and the ETo

were significant by t-test at 5% significance level, except the maximum temperature (Tmax) which had the lowest correlation ($r = 0.28$), showing no statistical significance. The strongest correlation occurred with the insolation ($r = 0.95$) followed by the minimum temperature ($r = 0.52$), mean temperature ($r = 0.43$) and wind speed ($r = 0.40$). It was not surprising that insolation had the greater correlation with ETo, since this variable is indicative of power availability. If the number of hours of sunshine is greater, also is greater the energy available for evaporation and transpiration processes in a region (Cunha and Escobedo, 2003; Lemos Filho et al., 2010).

With respect to the independent variables, the correlations between the mean temperature, wind speed and insolation were not significant at 1%, while among the other variables significant correlation was found, indicating the multicollinearity that occurs when the observed sample of the explanatory variables are correlated significantly. Thus, in order to solve this problem the variable minimum temperature and maximum temperature were eliminated from the linear regression analysis.

Table 2 shows the precision measurements in the construction of possible models. For the model with one variable, it is noted that the best model, based on the evaluation criteria, is the model which contains the insolation (in hours), with excellent performance, now called Model_1. For all models with two variables, the best combination of two variables occurred between the mean temperature (Tmed) and insolation (in hours), called Model_2. The Model_3 that considered all the variables obtained excellent performance and accuracy according to the adopted indicators.

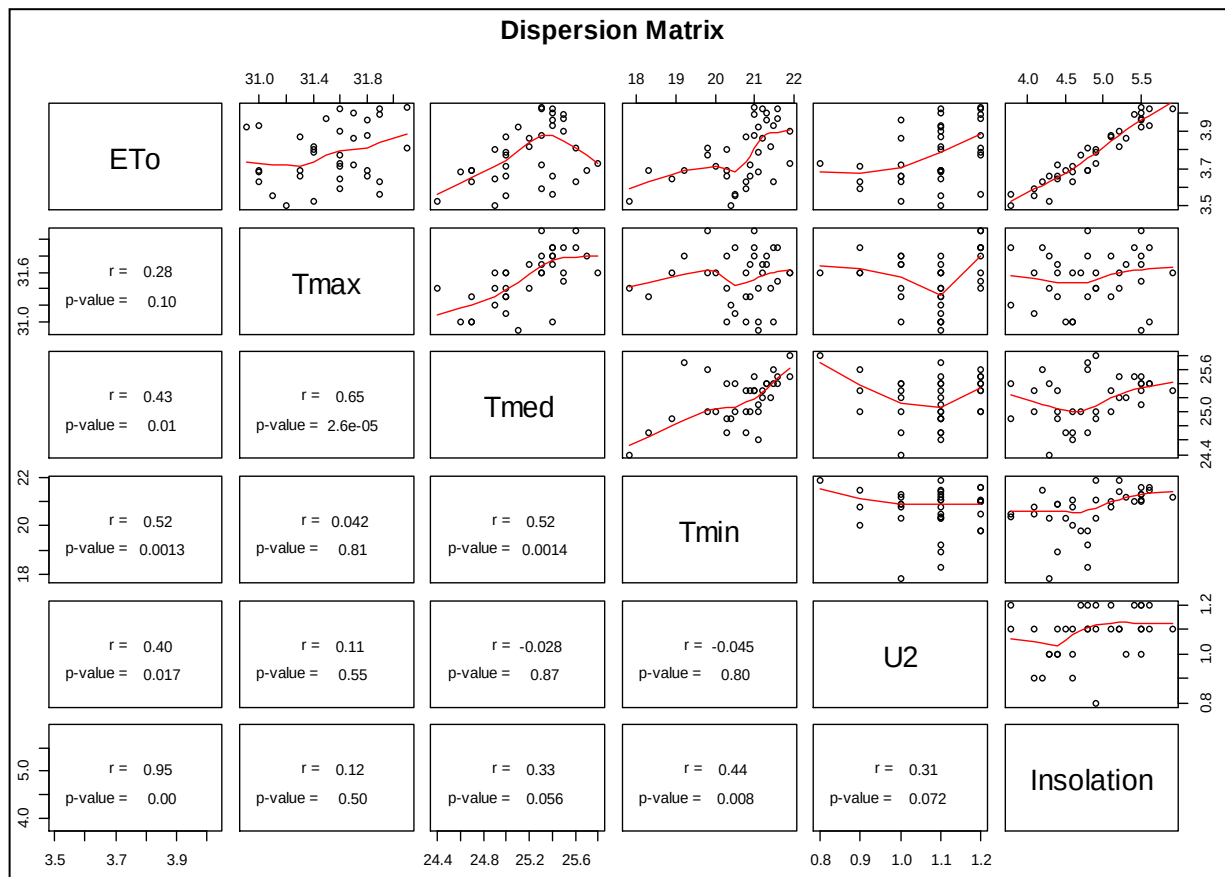


Figure 3 - Dispersion matrix with the weather variables and their Pearson correlation coefficients (r) and p-values of t-test for Rio Branco, Acre - 1980-2009 (calibration period).

Table 2 - Accuracy of regression models using average annual weather data for Rio Branco (AC), 1980-2009 (calibration period).

Variable	R^2_a	MAPE (%)	RMSE (mm)	"c"	Performance
Models of one variable					
Tmed	0.09	11.0	0.13	0.44	Tolerable
U2	0.08	11.0	0.13	0.42	Tolerable
Insolation (n)	0.89	4.0	0.05	0.97	Great
Models of two variables					
Tmed, U2	0.20	9.0	0.12	0.63	Median
Tmed, Insolation (n)	0.90	3.0	0.04	0.98	Great
U2, Insolation (n)	0.89	4.0	0.04	0.97	Great
Models of three variables					
Tmed, U2, Insolation (n)	0.92	3.0	0.04	0.98	Great

Among the three models selected, the Model_3 was considered the best, based on all assessment indicators shown in Table 3. The Model_3 showed the best coefficient of determination (R^2_a), and 92% of the annual

variation of ETo can be explained by the combination of main temperature, wind speed and insolation, which gives an explanatory significant degree for this set of variables.

The Model_1, considering only the insolation, had very good accuracy, reinforcing the high correlation between this and the ETo, and presenting advantage for simplicity, since it can estimate the ETo from just one climatological parameter, insolation. From a practical point of view, this model can be considered as a method to estimate ETo when the input weather variables are insufficient for other methods. The Model_2 excludes wind speed, taking the advantage of improved accuracy among the models with two variables.

The F-test and its significance and the regression results are presented for the three models in Table 3, which showed strong evidence of a successfully significant linear relation between ETo and other variables, being 1% for Model_2 and Model_3, and 6% for Model_1. However, the result of F-test should not be individually considered, that is, only its result must not discard and/or fully accept the regression equation, since the regression coefficients may present significant correlation.

Table 3 - Regression results for the suggested models fitted with average yearly weather data for Rio Branco (AC), 1980-2009 (calibration period).

Model	Parameters	Coefficients	Standard error	Value t	Value-p	F	Significance F
Model_1	Intercept	2.51017	0.08153	30.79	0.000	3.76	0.063
	Insolation	0.25988	0.01710	15.19	0.000		
Model_2	Intercept	1.22707	0.61343	2.00	0.056	131.8	0.000
	Tmed	0.05260	0.02495	2.11	0.044		
	Insolation	0.25158	0.01661	15.14	0.000		
Model_3	Intercept	0.80112	0.58119	1.38	0.179	108.7	0.000
	Tmed	0.06338	0.02305	2.75	0.011		
	U2	0.19318	0.07463	2.59	0.016		
	Insolation	0.24063	0.01568	15.35	0.000		

Also according to Table 3, it is observed that the p-value of t-Test for each parameter of the regression models was highly significant at 5% (p-value < 0.05) for all models, that is, at 95% confidence interval, the regression equation has all parameters (β_1 , β_2 , β_3) positively correlated with the dependent variable (ETo). Since the Modelo_3 can be considered as the best model to estimate the annual ETo based on the indicators adopted in Tables 2 and 3, and agreed the assumptions, one can interpret the estimated parameters.

It appears that all estimated parameters in Model_3 are positive. This suggests that for each increment of one unit in β_1 (mean temperature, Tmed), ETo rate increases 0.063 mm; for each increment of one unit in β_2 (wind speed, U2), ETo rate increases 0.193 mm; and for each increment of one unit in β_3 (insolation, n), ETo rate increases 0.241 mm. The intercept value 0.801 (β_0) (intercept with the ordinate axis) of the regression line represents the ETo expected rate when the other variables are equal to 0.

Adjusted multiple linear regression model is shown in Equation 12:

$$Y = 0,801 + 0,063 * \text{Tmed} + 0,193 * \text{U2} + 0,241 * \text{n} \quad (12)$$

where Y is the annual ETo estimated by Model_3; 0.801 is the value of the intercept; Tmed is the annual mean temperature; U2 is the wind speed measured at 2 m; n is the isolation.

To verify the suitability (quality) of the fit, it is need to verify whether the assumptions of normality, constant variance (homoscedasticity) and independence of residuals has been satisfied. The R software can be done it with the commands

given in Appendix 2, which generated the graphs showed in Figure 4.

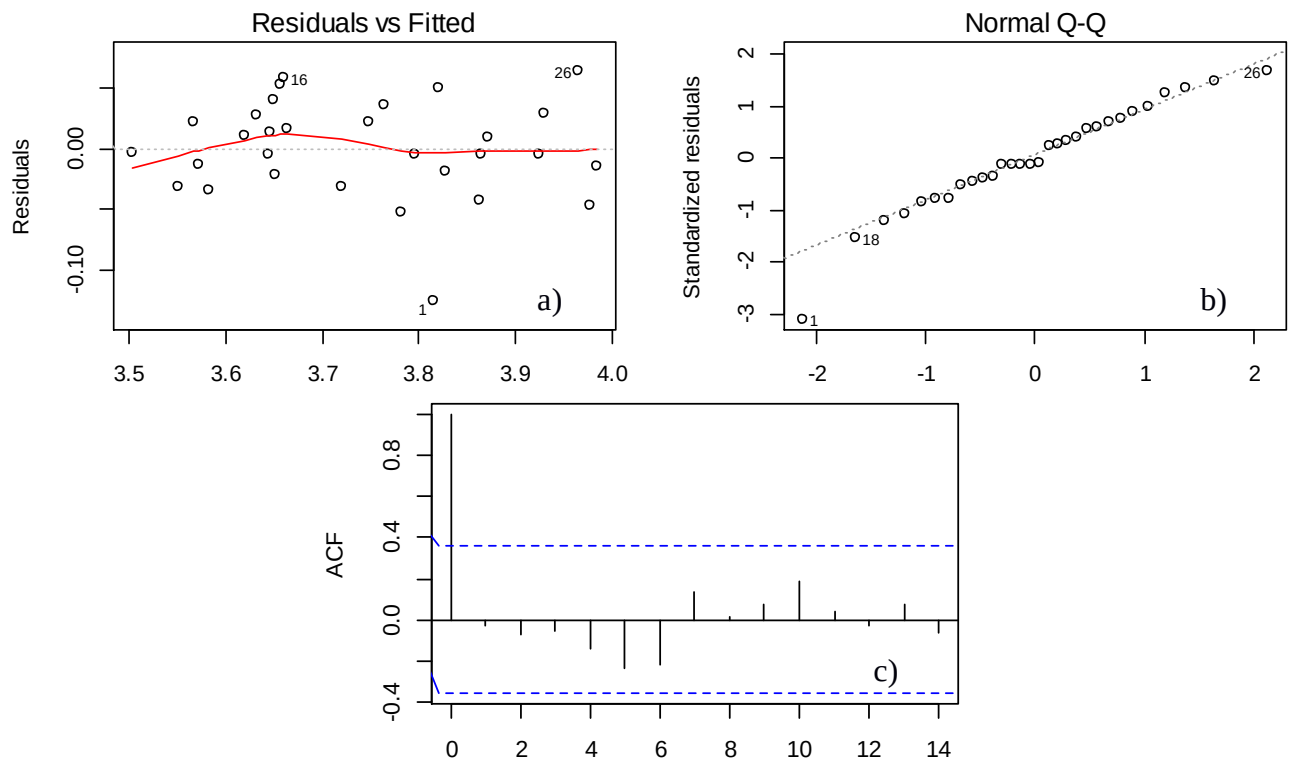


Figure 4 - Graphics of residuals vs. predicted values (a), probabilistic standardized residuals (b) and autocorrelation function (c) generated by R.

Analyzing the graphs of Figure 4, it can be inferred that the residuals of the fitted model are normally distributed because, in Figure 4b, it is observed that the points are almost over the line. The residuals are also in agreement with the constant variance (homoscedasticity) and independence assumptions, as shown in Figures 4a and 4c, respectively. Also applied the Kolmogorov-Smirnov, Breusch-Pagan and Durbin-Watson tests to verify if in fact the residuals are normally distributed, homoscedastic and independent, respectively. According to the results in Appendix 3, the p-values were greater than the significance level 0.05. So, there are no evidences to reject the null hypothesis, that is, one can infer with 95% confidence that the residuals are normally distributed homoscedastic and independent.

Initially, the accuracy of the model was verified using the range of ETo data from 2000-2014, with measures of five years, which has not been used for model calibration. Regarding this period, the model produced, with accuracy, values of R^2 , MAPE, RMSE and index of performance ("c") equal to 0.89, 3%, 0.03 mm and 0.88 (very good), respectively.

Figure 5 shows a good fit between observed ETo values (black solid line) for the entire period 1980-2014, both with respect to predicted values (1980-2009) and the verification and calibration period of the model (2010 to 2014) representing the interannual variability over time satisfactorily.

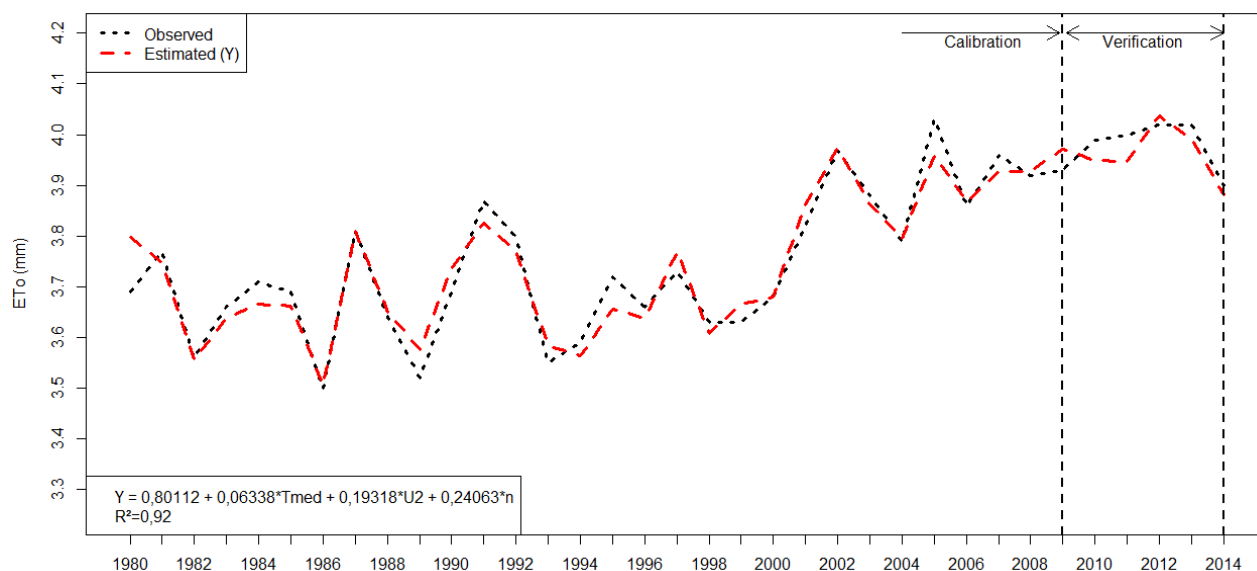


Figure 5 - Graph of observed annual ETo (black solid line) and ETo estimated by Model_3 (red dashed line). Calibration was performed during the period 1980-2009 (30 years) and verification during 2010-2014 (5 years) - Rio Branco, AC, 1980-2014.

Almedeij (2012) used the multiple regression analysis in Western Asia to estimate the daily and monthly evaporation using as independent variables the air temperature, relative humidity and wind speed for a period of 17 years (1993 -2009). The author concluded that only the model that contained the temperature and relative humidity showed the best results, with $R^2 = 0.84$. In Brazil, Pimentel (2007) and Cargnelutti Filho et al. (2008) used multiple regression to estimate monthly averages of minimum and maximum temperatures, decadal mean maximum temperature and decadal mean temperature, respectively, using as independent variables latitude, longitude and altitude for a period exceeding 60 years, with weather stations of Rio Grande do Sul. According to the authors, the model obtained good estimates with small absolute errors.

4. Conclusions

This study derived from empirical relationships for the reference evapotranspiration modeling obtained by the Penman-Monteith-FAO for application in Rio Branco (AC). The mean temperature, wind speed and insolation were significant for the estimation of the reference evapotranspiration, because there was high positive correlation primarily verified with insolation.

In general, three models recommended in this study provided acceptable results with

considerable accuracy. The Model_3 involving mean temperature, wind speed and insolation was regarded as the best, based on the evaluation criteria used, being able to simulate the local interannual variability. However, in absence of complete information, Model_1 and Model_2 take advantage of simplicity, since they require fewer climatic parameters for estimation of the reference evapotranspiration, although they produced lower values of accuracy.

The multiple regression models using weather variables demonstrated its usefulness as method to obtain accurate estimates of ETo, when one can not obtain measurements. Although these relationships are applied locally to the meteorological data of Rio Branco, it is expected that this work can inspire other researchers to examine whether these correlations exist in other spatial scales (regional, global) for data collected from other sources.

Acknowledgements

To the Federal University of Rio Grande do Norte (UFRN) for the financial support and encouragement of research through the Postgraduate Program in Climate Science (Programa de Pós-Graduação em Ciências Climáticas – PPGCC), and to the Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior - CAPES).

References

- Almedeij, J., 2012. Modeling Pan Evaporation for Kuwait by Multiple Linear Regression. *The Scientific World Journal* 2012, 1-9.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop water requirements. FAO, Rome. (Irrigation and Drainage Paper, 56).
- Alkaeed, O., Flores, C., Jinno, K., Tsutsumi, A., 2006. Comparison of several reference evapotranspiration methods for Itoshima Peninsula area, Fukuoka, Japan. *Memoirs of the Faculty of Engineering, Kyushu University* 66, 1-14.
- Bernardo, S., 2002. Manual de Irrigação, 6th ed. UFV, Viçosa.
- Braun, P., Maurer, B., Muller, G., Gross, P., Heinemann, G., Simmer, C., 2001. An integrated approach for the determination of regional evapotranspiration using mesoscale modelling, remote sensing and boundary layer measurements. *Meteorology and Atmospheric Physics* 76, 83-105.
- Brutsaert, W., 2005. *Hydrology: an introduction*. Cambridge University Press, New York.
- Camargo, A.P., Camargo, M.B.P., 2000. Uma revisão analítica da evapotranspiração potencial. *Bragantia* 59, 125-137.
- Camargo, A.P., Sentelhas, P.C., 1997. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. *Revista Brasileira de Agrometeorologia* 5, 89- 97.
- Cargnelutti Filho, A., Maluf, J.R.T., Matzenauer, R., 2008. Coordenadas geográficas na estimativa das temperaturas máxima e média decendiais do ar no Estado do Rio Grande do Sul. *Ciência Rural* 38, 2448-2456.
- Carvalho, L.G., Rios, G.F.A., Miranda, W.L., Neto, P.C., 2011. Evapotranspiração de referência: uma abordagem atual de diferentes métodos de estimativas. *Pesquisa Agropecuária Tropical* 41, 456-465.
- Conceição, M.A.F., 2006. Roteiro de cálculo da evapotranspiração de referência pelo método de Penman- Monteith-FAO. Embrapa Uva e Vinho, Bento Gonçalves. (Circular Técnica, 65).
- Costa, R.L., Silva, F.D.S., Sarmanho, G.F., Lucio, P.S., 2012. Imputação Multivariada de dados diários de precipitação e Análise de Índices de extremos climáticos. *Revista Brasileira de Geografia Física* 3, 661-675.
- Cunha, A.R., Escobedo, J.F., 2003. Alterações micrometeorológicas causadas pela estufa plástica e seus efeitos no crescimento e produção da cultura de pimentão. *Revista Brasileira de Agrometeorologia* 11, 15-26.
- Dorfman, R., 1997. Critérios de avaliação de alguns métodos de cálculo da evapotranspiração potencial. Rio Grande do Sul, Brasil. Thesis (Master). Porto Alegre, UFRGS.
- Draper, N.R., Smith, R., 1996. *Applied Regression Analysis*. 3rd ed. Wiley, New York.
- Duarte, A.F., 2006. Aspectos da climatologia do acre, Brasil, com base no intervalo 1971 – 2000. *Revista Brasileira de Meteorologia* 21, 308-317.
- Ferraz, P.A., 2008. Estimativa da evapotranspiração de referência (ET_o) para a região de Rio Branco – ACRE. Thesis (Master). Rio Branco, UFAC.
- IBGE. Instituto Brasileiro de Geografia e Estatística, 2010. Cidades. Rio Branco. Available: <http://ibge.gov.br/cidadesat>. Access: feb., 14, 2014.
- Junger, W.L., 2008. Análise, imputação de dados e interfaces computacionais em estudos de séries temporais epidemiológicas. Thesis (Doctoral). Rio de Janeiro, UFRJ.
- Junger, W., Leon, A.P., 2012. *mtsdi*: Multivariate time series data imputation. R package 0.3.3.
- Krämer, W., Sonnberger, H., 1986. *The Linear Regression Model under Test*. Physica-Verlag, Heidelberg.
- Lemos Filho, L.C.A., Carvalho, L.G, Evangelista, A.W.P, Alves Júnior, J., 2010. Análise espacial da influência dos elementos meteorológicos sobre a evapotranspiração de referência em Minas Gerais. *Revista Brasileira de Engenharia Agrícola e Ambiental* 14, 1294–1303.
- Mendonça, J.C., Sousa, E.F., Bernardo, S., Dias, G.P., Grippa, S., 2003. Comparação entre métodos de estimativa da evapotranspiração de referencia (ET_o) na região Norte Fluminense, RJ. *Revista Brasileira de Engenharia Agrícola e Ambiental* 7, 275-279.
- Nunes, L.N., Kluck, M.M., Fachel, J.M., 2010. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de

- risco para mortalidade cirúrgica. *Revista Brasileira de Epidemiologia* 13, 596-606.
- Pereira, A.R., Villa Nova, N.A., Sedyama, G.C., 1997. Evapotranspiração. FEALQ, Piracicaba.
- Pimentel, M.G.P., 2007. Método de estimativa de temperaturas mínimas e máximas médias mensais climatológicas do ar no Rio Grande do Sul. Thesis (Master). Pelotas, UFPEL.
- R, C.T., 2005. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org/>. Access: feb., 14, 2014.
- Racine, J., Hyndman, R., 2002. Using R to teach econometrics. *Journal of Applied Econometrics* 17, 175–189.
- Samani, Z., 2000. Estimating solar radiation and evapotranspiration using minimum climatological data. *Journal of Irrigation and Drainage Engineering* 126, 265-267.
- Silva, H.J.F., 2015. Análise de Tendência e caracterização sazonal e interanual da Evapotranspiração de Referência para o sudoeste da Amazônia Brasileira: Acre, Brasil. Thesis (Master). Natal, UFRN.
- Some'e, B.S., Ezani, A., Tabari, H., 2012. Spatiotemporal trends and change point of precipitation in Iran. *Atmospheric Research* 133, 1-12.
- Sousa, N.M.N., 2007. Influência de variáveis meteorológicas sobre a incidência do Meningite, Meningite e Pneumônia em João Pessoa-PB. *Revista Brasileira de Meteorologia* 22, 183-192.
- Souza, G.S., 1998. Introdução aos Modelos de Regressão Linear e Não-linear. Embrapa-SPI/Embrapa-SEA, Brasília.
- Souza, M.L.A., 2009. Comparação de métodos de estimativa da evapotranspiração de referência (ET_o) em Rio Branco, Acre. Thesis (Master). Rio Branco, UFAC.
- Spiegel, M.R., 1998. Estatística. Mc Graw – Hill, São Paulo.
- Thode Junior, H.C., 2002. Testing for Normality. Marcel Dekker, New York.
- Victoria, D.C., 2004. Estimativa da evapotranspiração da bacia do Ji-Paraná (RO) através de técnicas de sensoriamento remoto e geoprocessamento. Thesis (Master). Piracicaba, ESALQ/USP.
- Willmott, C.J., Rowe, C.M., Mintz, Y., 1985. Climatology of the terrestrial seasonal water cycle. *Journal of Climatology* 5, 589-606.
- Xu, C.Y., Singh, V.P., 1998. Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrological Processes* 12, 429–442.

APPENDIX

APPENDIX 1: SCATTERING MATRIX

```
##### function to generate correlation matrix
```

```
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...) {  
  usr <- par("usr"); on.exit(par(usr))  
  par(usr = c(0, 1, 0, 1))  
  r <- cor(x, y)  
  txt <- format(c(r, 0.12), digits = digits)[1]  
  txt <- paste0(prefix, txt)  
  text(0.63, 0.6, txt)  
  text(0.4, 0.6, "r =")  
  r2 <- cor.test(x, y)  
  txt2 <- format(c(r2$p.value, 0.12), digits = digits)[1]  
  txt2 <- paste0(prefix, txt2)  
  text(0.65, 0.33, txt2)  
  text(0.27, 0.33, "valor-p =") }  

```

```
##### GRAPH 1: SCATTERING MATRIX
```

```
cor.eto=data.frame(ETo, Tmax, Tmed, Tmin, u2, n)  
pairs(cor.eto, lower.panel = panel.cor, upper.panel = panel.smooth,main=" Scattering Matrix ")
```

APPENDIX 2: LINEAR REGRESSION AND VALIDATION GRAPHS

```
#### ESTIMATING THE PARAMETERS OF MODEL_3
```

```
Model.3=lm(ETo ~ Tmed + U2 + n)  
summary(Model.3)
```

```
#### CREATING AN ENVIRONMENT TO RECEIVE FOUR FIGURES  
#### AND ONLY THE FIRST TWO ARE USED
```

```
par(mfrow=c(2,2))
```

```
plot(Model.3)
```

```
##### creating a graph of the autocorrelation function
```

```
acf(Model.3$residuals)
```

APPENDIX 3: RESULTS OF THE STATISTICAL TEST

```
#### DATA NORMALCY TEST (Kolmogorov-Smirnov)  
### Load package (North)
```

```
lillie.test (Model.3 $ residuals)
```

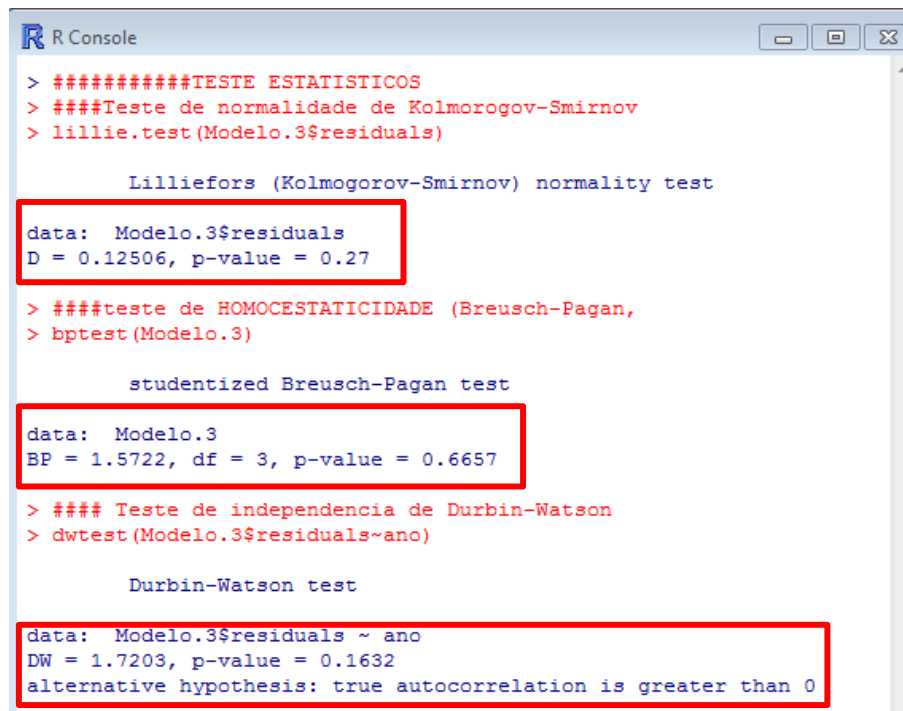
```
#### HOMOCESTATICIDADE TEST DATA (Breusch- Pagan)  
### Load package (lmtest)
```

```
bptest(Modelo.3)
```


INDEPENDENCE TEST DATA (Durbin-Watson)

Load package (lmtest)

```
dwtest(Model.3$residuals~ano)
```



```
R Console

> #####TESTE ESTATISTICOS
> ####Teste de normalidade de Kolmogorov-Smirnov
> lillie.test(Modelo.3$residuals)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Modelo.3$residuals
D = 0.12506, p-value = 0.27

> ####teste de HOMOCESTATICIDADE (Breusch-Pagan,
> bptest(Modelo.3)

      studentized Breusch-Pagan test

data:  Modelo.3
BP = 1.5722, df = 3, p-value = 0.6657

> #### Teste de independencia de Durbin-Watson
> dwtest(Modelo.3$residuals~ano)

      Durbin-Watson test

data:  Modelo.3$residuals ~ ano
DW = 1.7203, p-value = 0.1632
alternative hypothesis: true autocorrelation is greater than 0
```