I Workshop Internacional de Inovações Tecnológicas na Irrigação
&
I Conferência sobre Recursos Hídricos do Semi-Árido Brasileiro
26 a 28 de Setembro de 2007
Sobral - CE

PRINCIPAL COMPONENT ANALYSIS OF THE EVAPOTRANSPIRATION PROCESS FOR THE CURU PARAIPABA VALLEY, CEARA, BRAZIL.

OLIVEIRA, C.W.[1]; COSTA, G. M. DA[2]; CAVALCANTE, L. C.[2] & GOMES, P. M. C.[2]

[1] Ph. D. Biosystems Engineering, Faculdade de Tecnologia – CENTEC/Limoeiro do Norte-CE, Rua. Estevão Remígio 1145, CENTRO CEP 62.930-000 Fone:(88) 3423 6911.
[2]Graduando em Tecnologia da Irrigação – CENTEC/Limoeiro do Norte.

RESUMO: Neste trabalho a importância relativa das diferentes variáveis climáticas envolvidas no processo da evapotranspiração foram avaliadas numa base diária. Os dados foram coletados próximo a cidade de Paraipaba na bacia do Curu Paraipaba, no estado do Ceará, Brasil. As variáveis testadas foram dia do ano (dia), radiação solar (Mj/m^{2}day), tempertura do ar de 24-h (°C), temperatura máxima (°C), temperatura mínima (°C), temperatura ponto de orvalho (°C), umidade relativa de 24-h (%), umidade relativa máxima (%), umidade relativa mínima (%), velocidade vento (m/s), razão do vento dia/noite, média vento diurna (m/s), deficiti da pressão de vapor (kPa), vapor de pressão real (kPa) e precipitação (mm). Desses dados temperatura ponto de orvalho (°C), deficiti da pressão de vapor (kPa), e vapor de pressão real (kPa) foram estimados dos dados originais registrados. O modelo com todas variáveis estimadas e medidas foi capaz de explicar 89% da variação dos dados, com 63% de correlação com ET medida pelo lisimetro de pesagem. Um modelo reduzido com menos variáveis foi capaz de explicar 99,71% da variação dos dados com 68,57% de correlação com ET.

Palavras Chave: componentes principais, correlação, evapotraspiração.

ABSTRACT: On these work the relative importance of different climatologic variables involved on the evapotranspiration (ET) process on a daily basis were evaluated. The data was collect near the city of Paraípaba on the Curu Paraipaba watershed, in the state of Ceará, Brazil. The variables tested were day of the year (days), solar irradiance (Mj/m^{2}day), 24-h average air temperature (°C), maximum air temperature (°C), minimum air temperature(°C), dew point temperature (°C), 24-h average relative humidity (%), maximum relative humidity (%), minimum relative humidity (%), wind speed (m/s), day/night wind ratio, average wind speed day time (m/s), vapor pressure deficit (kPa), actual vapor pressure (kPa) and precipitation (mm/d). Out of this data set, dew point temperature (°C), vapor pressure deficits (kPa) and actual vapor pressure (kPa) were estimated from the original recorded data. The model with all measured and estimated variable was able to explain 89% of the total variance, with 63% correlation with ET measured by a weight lysimeter. A reduced model with fewer variables was able to explain 99.71% of the total variance with 68.57% correlation with measured ET.

Key words: principal component, correlation, evapotranspiration.

INTRODUCTION:

Evapotranspiration (ET) is one of the major processes within the hydrologic cycle. Both the long and short term water budgets of the soil profile and basin depend to a significant degree on the evapotranspiration rate. The ideal approach is to measure ET and weighing lysimeters are considered the most practical and reliable way to measure direct evapotranspiration in short time periods, e.g., daily or hourly (Howell et al., 1985). Another possible step is ET estimation using equations that utilize climatologic data. Those equations uses several parameters which, to be obtained, also require a number of ongoing equations to be plugged in. These peculiarities allow such equations be found in the literature in several different forms depending on methodology used, available data, and assumptions taken (Allen et al., 1998). Most of ET estimations methodologies diverge from one to another in complexity, accuracy, and number of inputs required. The selection of the method to use is the one that gives the minimum error with the available variables (Oliveira et al., 1998). Other point that come across is how to select the variables that best fits a model to estimate ET, and to assure the quality (integrity) of the assess data. Some work has been done to assure quality of data (Allen, 1996). The objective of this work was to evaluate a technique for assessing the inputs most relevant in the ET process. A relation between measured ET by a weighing lysimeters and meteorological data such as, air temperature, humidity, solar irradiance, wind, etc is to be found for the Curu Paraipaba conditions.

MATERIAL E METHODS:

The data for this work is from January to December of 1998 collected at the weather station installed at the Curu Valley Experimental Station near Paraípaba, in the state of Ceará, Brazil (Latitude 3°45’ S, longitude 39°21’ W Greenwich. and 20 m above sea level). This region has a rainy season from March though June and a dry season for the remainder of the year. We had 365 observations of which some had to be discarded because of maintenance on the weighting lysimeters, to result with 310 observations. The width and length of the lysimeter is 1.5 m, giving a surface area of 2.25 m^2, and the depth is 1.0 m, with 4500kg (10,000 lb) capacity, and accuracy of 0.45 kg (1 lb or 0.2041 mm of water). There were fifteen variables analyzed which were day of the year (days), solar irradiance (Mj/m^2day), 24-h average air temperature (°C), maximum air temperature (°C), minimum air temperature(°C), dew point temperature (°C), 24-h average relative humidity (%), maximum relative humidity (%), minimum relative humidity (%), wind speed (m/s), day/night wind ratio, average wind speed day time (m/s), vapor pressure deficit (kPa), actual vapor pressure (kPa) and precipitation (mm/d). Out of this data set, dew point temperature (°C), vapor pressure deficit (kPa), actual vapor pressure (kPa), were estimated from the original recorded data, using the following relations (Howel and Dusek, 1995)

Tdew = 237.3/((1/((LN(RH/100)/17.27)+(T/(237.3+T))))-1) (1)

where: Tdew is the current dew point temperature(°C),

LN is the Neperian logarithmic

RH is the current relative humidity (%)

T is the current air temperature (°C)

The saturated vapor pressure e(T) was computed using hourly air temperature according with the base equation for vapor pressure used in the FAO-56 (Allen et al., 1998). The actual vapor pressure ea and VPD was computed using its relation with relative humidity as:

ea = 0.611*EXP ((17.27*T)/(T+237.3))*(Hourly Relative Humidity/100) (2)

VPD = 0.611*EXP ((17.27*T)/(T+237.3))*(1 – Hourly Relative Humidity/100) (3)

The daily ea value was obtained by the 24-h average of the computed hourly ea values. To find the relative importance on the ET process, a coefficient of correlation was computed between each variable and the ET measured. The correlation coefficient (r) assesses the linear dependence between two random variables raging from $-1<r<+1$ (Ott, 1992).

r = Sxy/ (SxxSyy)^0.5 (4)

where: Sxy is the standard deviation between x e y given by $E[(x-\mu x)(y - \mu y)]$

Sx is the standard deviation of x given by $\Sigma x^2 - (\Sigma x)^2/n$

Sy is the standard deviation of y given by $\Sigma y^2 - (\Sigma y)^2/n$

If a linear relation is found between the variables and the ET measured, a linear regression curve can be adjusted which will then explain these relationship. When the variables are correlated with each other, principal component analysis (PCA) can be used. PCA decomposes the data matrix X (m samples, n variables) as the sum of the outer product plus a residual matrix E. The number of principal components k, to retain in the model will be less than the number of inputs if the data is correlated.

RESULTS AND DISCUSSION:

Compared with measured ET the highest correlation was found with solar irradiance (0.68). The other variables are fairly related, maximum temperature (0.55) and vapor pressure deficit (0.57). Since solar irradiance and maximum temperature are not independent variables (0.77). This may be an indicative that we are dealing with an ill-posed problem. Ill-conditioned

Table 1. Correlation coefficients of the climatic variables.

	Irad	Tavg	Tmax	UR24	URmn	VPDf	ET
JDay	0.4581	-0.1064	0.2286	-0.6842	-0.6055	0.6575	0.3366
Irad	1	0.4552	0.7717	-0.7365	-0.7568	0.7706	0.6764
Tavg	0.4552	1	0.6942	-0.2669	-0.2093	0.3273	0.435
Tmax	0.7717	0.6942	1	-0.579	-0.708	0.6523	0.5527
Tmin	-0.2688	0.5457	-0.0705	0.2764	0.4093	-0.2751	-0.1122
Tdew	-0.53	0.2218	-0.2685	0.8771	0.8368	-0.8401	-0.3269
UR24	-0.7365	-0.2669	-0.579	1	0.9114	-0.9859	-0.5247
URmx	-0.2876	-0.1974	-0.1286	0.5713	0.3598	-0.5249	-0.2094
URmn	-0.7568	-0.2093	-0.708	0.9114	1	-0.9236	-0.5019
Vent	0.6023	0.2056	0.4051	-0.8724	-0.7373	0.8379	0.3954
vd/n	-0.0234	0.0982	0.0533	0.3515	0.2484	-0.2995	0.0148
Vdia	0.3952	0.3452	0.3532	-0.4192	-0.3355	0.4371	0.3455
VPDf	0.7706	0.3273	0.6523	-0.9859	-0.9236	1	0.5686
ea24	-0.5308	0.2232	-0.262	0.8708	0.8269	-0.8346	-0.3302
PPT_	-0.6076	-0.3821	-0.5759	0.5619	0.5694	-0.5743	-0.4138
ET	0.6764	0.435	0.5527	-0.5247	-0.5019	0.5686	1

means that the samples are only marginally independent and the condition number is high. The condition number is defined as the ratio of the largest eigenvalue to the smallest eigenvalue. This number should be below 100 for stability, and a number less than 10 is best, for these data set the condition number found was 8.3749e+007.

Table 2. Results variance and accumulated variance, and components

Components	% Explained	% Exp. Ac.	Variables	PC1	PC2	PC3	PC4
PC1	55.437	55.437	JDay	0.2361	-0.2822	0.0865	-0.2653
PC2	17.145	72.582	Irad	0.3116	0.1793	-0.1481	0.0737
PC3	10.519	83.101	Tavg	0.1904	0.4648	0.1873	0.2292
PC4	6.380	89.480	Tmax	0.284	0.263	-0.136	0.2352
PC5	3.486	92.966	Tmin	-0.1011	0.3234	0.6091	0.0574
PC6	2.647	95.614	Tdew	-0.2937	0.3046	0.0701	0.07
PC7	1.867	97.481	UR24	-0.3408	-0.0135	-0.0615	-0.0462
PC8	0.897	98.378	URmx	-0.1923	-0.0004	-0.4893	0.3866
PC9	0.703	99.081	URmn	-0.3313	0.0199	0.113	-0.1428
PC10	0.490	99.571	Vent	0.3158	-0.0611	0.21	-0.0325
PC11	0.287	99.858	vd/n	-0.0601	0.3761	-0.4396	-0.3687
PC12	0.083	99.940	Vdia	0.1186	0.3403	-0.0829	-0.6725
PC13	0.034	99.974	VPDf	0.3399	0.0418	0.0407	0.056
PC14	0.022	99.996	ea24	-0.2878	0.3137	0.0691	0.0968
PC15	0.004	100.000	PPT_	-0.2349	-0.2096	0.1858	-0.1854

We use PCA to reduce the number of variables and deals with the problem of collinearity. The total variance found on the data was 20, and the first principal component (PC) explains 55.44% of it. The first four components explain together 89.48% of the total variance of the data. From the eighth PC forward they explain less than 1% of total variability, which it is worst than the original variables values.

The first PC correlates 63% with measured ET. Selecting a data set with the best correlated variables in respect with measured ET, which were solar irradiance, 24-h air temperature, maximum air temperature, 24-h relative humidity, minimum relative humidity, and vapor pressure deficit, a new principal component analysis, was found.

Table 3. Results variance and accumulated variance, and components with few variables

Components	% Explained	% Exp. Ac.	Variables	PC1	PC2	PC3	PC4
PC1	84.1549	84.155	Irad	0.4225	-0.0034	0.3827	-0.8153
PC2	10.3794	94.534	Tavg	0.3308	0.8241	-0.3311	0.0514
PC3	3.4823	98.017	Tmax	0.4173	0.2341	0.5611	0.4121
PC4	1.699	99.716	UR24	-0.4228	0.274	0.4799	0.046
PC5	0.2125	99.928	URmn	-0.4178	0.3749	-0.2066	-0.4005
PC6	0.0719	100.000	VPDf	0.4296	-0.2245	-0.3952	0.0178

These six variables, originated six components where the first four components were sufficient to explain 99.71% of the total variance of the data which was down to 8. For these components the first PC correlates 68.57% with measured ET. These results was better than the one found by Silveira (2000) which for the Metropolitan area, Acaraú and Curu watershed had a model that explain respectively, 89%, 91% and 93% of the total variance of each site on the estimated ET.

CONCLUSION:

For the Curu Paraipaba Valley the model with all measured and estimated variable was able to explain 89% of the total variance, with 63% correlation with measured ET by weight lysimeters. A reduced model with fewer variables was able to explain 99.71% of the total variance with 68.57% correlation with measured ET. The results revealed that solar irradiance and vapor pressure deficit were the climatologic variables with highest weight on the evapotranspiration process.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLEN, R.G. Assessing integrity of weather data for reference evpotranspiration estimation. Journal of irrigation and drainage engineering. v. 122.(2) p 97-106. 1996.

ALLEN, R.G.; PEREIRA, L.S.; RAES, D.; SMITH, M. Crop evapotranspiration, Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. Rome. 1998.

HOWELL, T.A.; Mccormick, R.L.; Phene, C.J. Design and instalation of large weighing lysimeters. Transactions of the ASAE. v. 28, p. 106-112, 117. 1985.

HOWELL, T.A.; DUSEK, D.A. Comparison of vapor-pressure-deficit calculation methods – Southern High Plains. Journal of Irrigation and Drainage Engineering, v. 121. (2), p 191-198. 1995.

OLIVEIRA, C.W.; SOUZA, F.; YODER, R.E.; RIBEIRO, R.S.F.; MIRANDA, F.R. Estimating reference evapotranspiration in northeast Brazil. Orlando: ASAE, 14p. (ASAE Paper, 982184). 1998.

OTT, R.L. An introduction to statistical methods and data analysis. 4th Edition. Belmont, CA: Duxbury press, 1051p. 1992.

SILVEIRA, S. S. Análise de componente principal como ferramenta para estimar os parâmetros meteorológicos de mais influência para o processo de evapotraspiração no Estado do Ceará. Fortaleza, 60p. Dissertação (Mestrado) – Departamento de Engenharia Agrícola, Universidade Federal do Ceará. 2000.