Journal of the Nigerian Association of Mathematical Physics
Volume35,(*May, 2016*), pp175 – 184
© J. of NAMP

Dimension Reduction and Clustering of Micrometeorological Variables Using P mode Principal Component and Hierarchical Cluster Analyses

Mojisola Oluwayemisi Adeniyi

Department of Physics, University of Ibadan, Ibadan, Nigeria.

Abstract

Variations in meteorological variables greatly influence socioeconomic activities, especially in developing nations such as Nigeria. However, routine meteorological data are limited to a few variables. Though, unavailablevariables can be estimated from the available ones, the transfer functions connecting the available routine parameters and the unavailable ones do not exist. This paper provides the transfer functions by grouping 36 observed micrometeorological parameters from Nigerian Micrometeorological Experiment 1 (NIMEX_1) for subsequent estimation of parameters in each group from a member of the group. The NIMEX_1 was carried out at Ile-Ife, Nigeria in 2004 between Day of years 56 and 68. Principal component analysis is used to reduce the dimensionality in the observed data while cluster analysis is used to group them into variables with similar variances. Thereafter, polynomial regression is used to estimate some variables in a cluster from a member of the cluster.The cross-validated coefficient of determination between observed and estimated parameters ranges between 0.64 and 0.99.Moreover, the cross-validated root mean square errors are mostly less than the standard deviations in the observed data. Therefore, the obtained transfer functionsare applicable in estimatingmeteorologicalvariables in areas with similar weather conditionsto the prevailing weather during NIMEX_1 experiment.

Keywords:P mode PCA, cluster analysis, micrometeorological variables, polynomial regression

1.0 Introduction

Variations in micrometeorological parameters have great influence on the socioeconomic, agriculture, industry, healthand climate [1]. Most importantly agricultural productivity is directly linked with variations in agro-meteorological variables. To this end, routine data of meteorological variables are usually available through the Meteorological Agencies in different countries. However, the routine data are mostly limited in variables. Comprehensive data are only available during specialized field Campaign, which are mostlydone by Atmospheric Scientists. These are not continuous because they are usually set up to achieve a specific aim and after the aim is achieved the experiment will be stopped since the field campaigns are not to duplicate the efforts of the Meteorological Agencies in routine measurements of meteorological parameters. Preservation of the meteorological sensorsis also important to the Atmospheric Scientist for them to last longer and be applicable in future experiments, knowing the difficulties usually encountered in acquiring the equipment.

In a screen house usually employed for all year round production of seasonal crops such as tomatoes, it is of paramount importance to monitor the variations in meteorological parameters for prompt detection and action against any unsuitable condition for the plants. At times, the sensors and experts may be available to monitor the variations in and outside the screen house comprehensively, but this is not always the case. Only simple variables such as temperature can be monitored by the agriculturalist.Nevertheless, dimension reduction and clustering of a pool of micrometeorological variables will permit the estimation of some variables from other variables, if a few parameters can be measured.

A number of estimation studies have been done for global radiation and soil surface temperature using other meteorological variables. Among which is the workon estimation of global radiation over Iraq [2] using maximum temperature, relative humidity and clearness index. The same parameter was estimated over Indiausing total precipitable water content in a vertical column [3]. Over Nigeria also, routine meteorological variables were used to estimate solar radiation [4]. Soil surface temperature which is not available as routine measurement was estimated from air temperature and soil temperature at 5 cm

Corresponding author:Mojisola Oluwayemisi Adeniyi, E-mail:mojisolaadeniyi@yahoo.com, Tel.: +2348033579081

depth [5].This was based on the established physical relationship between the solar radiation, soil surface and soil depth temperatures.Only a few parameters have been estimated from other parameters [2-5], so a comprehensive transfer functions for estimating the unavailable parameters in routine measurements is still needed.

In 2004,a number of Atmospheric Physicists and Meteorologist collaborated with German Scientist to carry out micrometeorological experiment at Obafemi Awolowo University in Ile-Ife (7^{o} 33'N and 4^{o} 33' E). The experiment was tagged 'Nigerian Micrometeorological Experiment (NIMEX_1)'. The NIMEX_1was a comprehensive experiment where over 40 metrological variables were measured [6]. These variables co-vary together, so the covariance can be used to group them into clusters.This experiment provides enough number of micrometeorological variables for this type of analysis. Such experiments are scarce. Details of the NIMEX_1 experiment, the micrometeorological parametersand the equipment used in observing themare well documented[6,7,8]. This paper therefore aims at providing transfer functions for relating the routine observed meteorological data to the unavailable parameters for subsequent estimation of such parameters.

2.0 Data and Methodologies

2.1 Data

The data used in this study are the profile measurements from NIMEX_1 experiment. The variables used and their profile height or depth are listed in Table 1. However, details of the experiment and the equipment used are well documented[6, 7, 8]. Fifteen days data at 30 minutes interval, from Day of Year (DOY) 54 to 68 in 2004 (25 February to 8 March) are used in this study. The quality of the data is tested following Adeniyi and Ogunsola [8]. Visual check is also done to remove any unreliable data [9]. DOYs 54 and 55 are only used for cross-validation.

Table 1: Micrometeorological parameters used in this study and their profile heights

Parameter	Abbreviation	Levels (m)
Wind speed	W1- W8,	0.7, 1.2, 2.2, 3.3, 5.2, 7.2, 10.2, 14.8
Wind direction	Wdr	14.8
Wet bulb temperature	Tw1-Tw3	0.9, 4.9, 10.0
Dry bulb temperature	Td1-Td3	0.9, 4.9, 10.0
Soil temperature	Ts1-Ts3	0.05, 0.10, 0.30
Soil heat flux	Hf1-Hf3	0.02, 0.10, 0.30
Soil moisture	Smt	
Global radiation	Glob	1.5
Net radiation	Rnet	1.7
Longwave downwelling Radiation	Lwd	1.7
Longwave upwelling Radiation	Lwu	1.7
Shortwave downwelling Radiation	Swd	1.7
Shortwave upwelling Radiation	Swu	1.7
Air pressure	Press	1.0

2.2 Methodology

The PCA is used to reduce the dimensionality in the diurnal micrometeorological variables. Thereafter,hierarchical cluster analysisis used to group the observed data into variance related clusters. These are done to be able to represent many variables with only a few without loss of significant information about the original data. Any needed variance of any particular variable in a cluster can then be obtained from a variable in such a group. Without measuring all the variables, information about them can be obtained from only the few that are measured.

2.2.1 Principal Component Analysis

The diurnal meteorological variables are first subjected to Kaiser–Meyer–Olkin Measureof Sampling Adequacy (KMO) and Bartlett's Test ofSphericity (BTS). Details of the significance of the two tests are stated explicitly in Adeniyi [10]. A significance value of 0.05 is necessaryfor a Principal Component Analysis (PCA). To identify modes of variability in thediurnal meteorological variables,a P-mode PCA was conducted.PCA can only be conducted on normally distributed variables so highly skewed variables such as rainfall are exempted from the analysis.

The P-mode PCA is employed in this study to reduce the dimension of the observed micrometeorological variables with little or no loss of the variability in the data. In this mode, the different micrometeorological variables, all from one station are the

columns while the rows are the time in half hours. This PCA mode usually results in component loadings for micrometeorological parameters which identify groups of parameters that co-vary together. The resulting component scores are time and the scores reveal the time of important variations in the parameters.The P mode PCA is conducted on 13 days (DOYs 56-68 in 2004) half hourly data of 36 micrometeorological variables (Table 1). The 36 variables are the columns while rows are time (half hours).
Since the variables are all from one station, time dependent real valued scalar fields $s_i(t)$ are considered, t denotes time. The time dependent fields $s_i(t)$ are represented as finite dimensional time series

$$s_i(t) = (s_1(t), s_2(t).....s_N(t))^T \tag{1}$$

for $1 \le i \le N$, where N is the number of micrometeorological variables (36).The P-mode PCA makes use of a transformation in variable spaceto identify modes of variations such that an approximate expansion of the data uses only a small number $n<N$ of variables without much loss of information. It identifies the dominant modes of variation in the dataset. The expansion of the dataset in the time series is given by:

$$s(t) = \sum_{k=1}^{n} a_k(t) p_k \tag{2}$$

Where p_k are orthogonal variables.

The time expansion coefficients $a_k(t)$ are calculated by projecting the data onto the orthogonal variables.

The principal components are derived from correlation matrix in order to standardize the covariance from the different scales of the different micrometeorological variables. This will give equal weight to each variable such that the original variable with the highest variance will not dominate the principal components. Varimax (orthogonal) rotation is used for easy interpretation of the PCs [10-13]. The uncorrelated components output from orthogonal rotation satisfy the requirement of the cluster analysis. A number of criteria have been used to truncate the PCs such that only the significant ones are retained. They are: The scree plot [14], the PCs with eigenvalue greater than one [15] and number of PCs that explain a threshold percentage variance in the original data [16], to mention a few.

2.2.2 Cluster Analysis

Cluster analysis is carried out on the principal componentloadings of the retained modes of variability to group the diurnal meteorological data. Cluster analysis usesa measure of distance to group observations within adataset thereby classifyingit[17].The retained principal component loadings are subjectedto hierarchical methods of clustering and latermerged based on the intra- and inter-class similarity. The squared Euclidean distance is used as the similarity measure. Ward method of clustering [18] is used based on its reported effectiveness in clustering of meteorological variables [10, 11].The method is based on minimum sum of squares within the resulting clusters has been found to outperform other clustering methods such as average linkage; nearest neighbor and centroid [10]. The number of clusters to retain is based on the rule of Clark and Hoskingin equation (3) [19];

$$k = 1 + (3.3\log_{10} N) \tag{3}$$

where k = maximum number of clusters and N is the number of items (micrometeorological variables) to be clustered.

2.2.3 Polynomial Regression

Linear regression models, which are usually employed in estimating a parameter from one or more other parameters is applied to each cluster of micrometeorological variables [10, 20, 21].Polynomial regression models are developed to provide the transfer function for each meteorological variable based on any of the variables in the same group with the variable. If the relationship between the variables is not linear polynomial regression of higher order is applied. Discrete Fourier transform is also applicable when physical relationship is established amongst variables. In this paper, any order of polynomial that best represent the relationship between two variables will be established and can later be applied to estimate the predictand from the predictor.
Let the predictand (dependent variable) be y and the predictor (independent variable) be x. The relationship between x and y can be found by fitting the polynomial of order 1, 2………n until the best fit is found. Polynomial of order k can be represented as matrix equation (4a).

$$\begin{bmatrix} n & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \cdots & \sum_{i=0}^{n} x_i^k \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \cdots & \sum_{i=0}^{n} x_i^{k+1} \\ \sum_{i=0}^{n} x_i^2 & \sum_{i=0}^{n} x_i^3 & \sum_{i=0}^{n} x_i^4 & \cdots & \sum_{i=0}^{n} x_i^{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{n} x_i^k & \sum_{i=0}^{n} x_i^{k+1} & \sum_{i=0}^{n} x_i^{k+2} & & \sum_{i=0}^{n} x_i^{2k} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_3 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^2 y_i \\ \vdots \\ \sum_{i=0}^{n} x_i^k y_i \end{bmatrix} \tag{4a}$$

. where k is the degree of the polynomial, and *n* is the number of data points.

The matrix equation (4a) represents k+1 simultaneous equations having (k+1)unknowns. Each side of the equation has dimension (k+1x1). The coefficients are obtained by solving the matrix depending on the available x and y data points.

Equation (4a) is equivalent to

$$[x]^T [x]\alpha = [x]^T y \tag{4b}$$

The solution to the equation is obtained by left- multiplying both sides of the equation by the inverse of the first matrix on the left hand side of the equation.

$$([x]^T [x])^{-1} [x]^T [x]\alpha = ([x]^T [x])^{-1} [x]^T y$$

$$[I]\alpha = ([x]^T [x])^{-1} [x]^T y$$

Cross-validation

Thirteen days data from DOY 56-68 are used to train the polynomial regression models while DOYs 54 and 55 data are used to validate the models. This is to ensure the transferability of the models and their performance on data outside the training period. Cross-validated correlation, coefficient of determination (R^2) and Root Mean Square Error (RMSE)

3.0 Results

3.1 Data Quality

The surface temperature data has some missing values on DOYs 55,57 and 61 so the variable is excluded from the analysis. Also,rainfall and Hf2 are highly skewed (skewness = 17 and -4.8respectively), so they are also excluded in the analyses. The skewnesses of the selected variables for the analysis are mostly less than 1 and greater than -1. Only a few are outside the range, so the selected variables are approximately normally distributed.

3.2 PCAand Cluster Analysis Results

The KMO and BTS reveal the suitability of a PCA analysis in the diurnal micrometeorological data from NIMEX-1 station. The KMO value of 0.914 obtained is a superb value for PCA which reveals the existence of distinct groups of variables in the data set which can be identified by PCA. The significance of the BTS test is also high ($P < 0.001$), this means that there are common variance among the micrometeorological variables for the PCA to identify. The PCA conducted on 624 half hourly data points for36 variables results in five PCswith eigenvalue greater than one.Furthermore, the cumulative percentage of explained variance by the PCs is above 90% so the Kaiser rule is applied [15].Hence, five PCs with eigenvalue greater than one which explain 92.4% of variance in the original data matrix are retained for cluster analysis. The retained rotated PCs 1 to 5 explain 41.7%, 18.6%,17.9%,10.5% and 3.7% of variance, respectively in the 36 micrometeorological variables. The PC1 is the most important mode of variation having wind speed as its variable of importance. The PC1 loadings for the important variables are$\geq$0.95. This mode is evident on DOY 57 in the evening (8:30 pm local time). At this time convection ceases because there is no input from the sun. Therefore, wind variations take charge of the atmospheric mixing and turbulence. The variable of importance in PC2 is the soil temperature at 5cm. The other important variables have PC2 loadings >0.86. The mode of variation is evident on DOY67 at 5:30 pm local time. Based on the synoptic weather observation, there was no rain on DOY67 but poor visibility persisted throughout the day, this reveals the presence of high level aerosol on that day which can act as green house gas. By 5:30 pm local time, top soil layers on that day have relatively high temperature because they are been heated by the Longwave downwelling radiation from the aerosols (greenhouse gas effect), they have also been heated slightly by the sun and they have not yet released the whole heat to warm the atmosphere. Though there is no input any longer from the sun, the soil is hot because of the residual heat that it has not yet released based on its higher soil thermal conductivity than air. The Longwave downwelling radiation is also important on that day but in the negative sense (PC2 loading <-0.68). By definition, the longwave downwelling radiation is the thermal irradiance that reaches the earth's surface in the infrared spectrum (4-100 μm), not in the shortwave form that is directly from the sun. The poor visibility on DOY 67 justifies the use of greenhousegas effect to explain the significance of longwave downwelling radiation on that day. The radiation variables are the important variables in PC3 mode of variation; while the net radiation, global radiation, short wave upwelling, short wave downwelling and longwave upwelling radiations are important in the positive sense, the longwave downwelling radiation is important in the negative sense, such that when the value of longwave downwelling radiation is reducing the values of the other radiation variables are increasing and vice versa.This reveals the ability of clouds and other atmospheric constituents to absorb and re-emit solar radiation. They block solar radiation from reaching the earth's surface on days when the atmosphere consists of them in high proportion. The effect of PC3 mode of variation is evident on DOY 67 at 12:30 pm local time which is the peak time of solar irradiance in this area[22].Wet bulb temperature dominates the PC4 mode with PC loading $\geq$ 0.95. The highest score for PC4 occurs at 11:30 am local time on DOY 58. This is usually the period when the high humidityin the early morning hours begins its decent (turning point) towards its minimum value as a result of the increasing solar irradiance towards maximum solar irradiance and the associated evaporation. The PC5 is the least significant PC, the important variables have PC loadings >0.4 and the leading variable is soil temperature at depth 30 cm (Ts3). This mode of variation is evident at 8:00 pm local time on DOY 56. Synoptic condition recorded for the day shows that it rained shortly before 8:00 pm on that day, the rain was short lived. This would have cooled down the top soil, but because it is short lived the deeper soil layer is not likely to be affected. So the high and dominant temperature at that time would be the one measured from the deepest soil layer,Ts3.

3.3 Cluster Analysis Results

Varimax rotated PC loadings are used for the cluster analysis and based on the rule of Clark and Hosking [19] six clusters are specified. Cluster one consist of moisture related variables Tw1, Tw2, Tw3 and Smt. Cluster 2 consists of Td1, Td2, Td3, Ts1, Ts2 and Hf3. Cluster 3 consists of Ts3, Lwd and Press. Cluster 4 consists of the wind speed w1 to w8, standard deviation of wind 1 (sw1) to standard deviation of wind 8 (sw8) and Lwu. The wind direction is left alone in cluster 5. Cluster 6 consists of Hf1, Swd, Swu, Glob and Rnet.

3.4 Polynomial Regression Results

Variables that co-vary together and fall within the same cluster are examined.Physically clear relationships amongst variables are then explored for estimation of a variable from any of the remaining variables in each cluster. The obtained regression coefficients (Table 2) are then applied to another data set not included in the training of the regression model, precisely DOYs 54 and 55. This was done for cross-validation of the models. Wet bulb temperature at 5 m (Tw2) and 10m (Tw3) are estimated from observed wet bulb temperature at 1 m height, Tw1. The cross-validated correlations obtained are very high and significant at all levels of significance (0.1% to 5%). The R2 is greater than 0.93 (Figs. 1a and b, Table 3). The RMSEs are less than the standard deviation in the observed Tw2 and Tw3 data (Table 3). This means that the estimated values agree with the observed at 99.7% confidence level.

Table 2: Polynomial coefficients obtained for the polynomial regression models. Tw2, Tw3 are estimated from Tw1. Td2, Td3, Ts1 and Ts2 are all estimated from Td1. W8 is estimated from W1 while Hf1, Rnet, Swd and Swu are estimated from global radiation.

Tw2	Tw3	Td2	Td3	Ts1	Ts2
-1.81E-13	-2.22E-13	6.13E-17	-1.04E-16	-9.70E-16	-7.94E-16
5.58E-11	6.84E-11	-1.91E-14	3.79E-14	3.37E-13	2.75E-13
-7.97E-09	-9.77E-09	2.64E-12	-6.26E-12	-5.28E-11	-4.31E-11
6.98E-07	8.56E-07	-2.15E-10	6.13E-10	4.92E-09	4.01E-09
-4.19E-05	-5.144E-05	1.12E-08	-3.95E-08	-3.02E-07	-2.46E-07
0.00183	0.0022458	-3.91E-07	1.74E-06	1.27E-05	1.03E-05
-0.059896	-0.0735395	8.93E-06	-5.2E-05	-0.00036	-0.0003
1.496453	1.837712	-0.00012	0.001029	0.006841	0.005556
-28.7384	-35.296565	0.000833	-0.01093	-0.0695	-0.05642
423.5308	520.20166	0	0	0	0
-4737.759	-5818.8306	0.002975	1.308774	7.64111	6.19873
39302.18	48262.812	0	0	0	0
-231455.7	-284153.78	-28.4293	-419.373	-2257.02	-1830.16
886934	1088484.9	611.537	6343.306	32829.45	26616.38
-1759740	-2158647	-4284.23	-34951.5	-174141	-141165
0	0	0	0	0	0
4898719	6001972.6	83992.04	498792.4	2311029	1872983

W8	Hf1	Rnet	Swd	Swu
4.877015	-6.40E-39	1.60E-38	5.13E-39	6.53E-39
-97.4153	4.52E-35	-1.14E-34	-3.25E-35	-4.35E-35
872.4505	-1.45E-31	3.65E-31	9.09E-32	1.31E-31
-4623.05	2.78E-28	-6.99E-28	-1.46E-28	-2.32E-28
16098.73	-3.56E-25	8.91E-25	1.47E-25	2.73E-25
-38655.7	3.22E-22	-7.96E-22	-9.14E-23	-2.22E-22
65173.76	-2.10E-19	5.13E-19	3.00E-20	1.29E-19
-76626.8	1.01E-16	-2.41E-16	1.45E-18	-5.30E-17
60029.25	-3.52E-14	8.30E-14	-6.60E-15	1.54E-14
-26458.2	8.87E-12	-2.08E-11	3.35E-12	-3.07E-12
0	-1.57E-09	3.75E-09	-9.13E-10	3.90E-10
8403.706	1.84E-07	-4.79E-07	1.48E-07	-2.73E-08
-5797.65	-1.3E-05	4.26E-05	-1.3E-05	5.76E-07
2025.356	0.000471	-0.00252	0.000573	2.59E-05
-382.525	-0.00388	0.088021	-0.00357	0.000272
35.47775	0.211697	-0.66645	0.839895	0.112591
-0.73683	-39.3992	-34.3209	2.054129	4.608185

Table 3: Standard deviation (STD) in observed and cross-validated coefficient of determination (R^2)between observed and estimated data and root mean square error (RMSE) in estimated data

Parameter	STD (Observed)	RMSE (Estimated)	R^2
Tw2	1.23	0.25	0.954
Tw3	1.26	0.32	0.939
Td2	3.86	0.83	0.959
Td3	3.82	1.03	0.933
Ts1	4.12	3.02	0.918
Ts2	2.62	3.38	0.643
W8	0.39	0.25	0.706
Hf1	45.25	16.79	0.960
Rnet	172.34	81.67	0.897
Swd	303.25	12.03	0.999
Swu	70.62	76.07	0.998

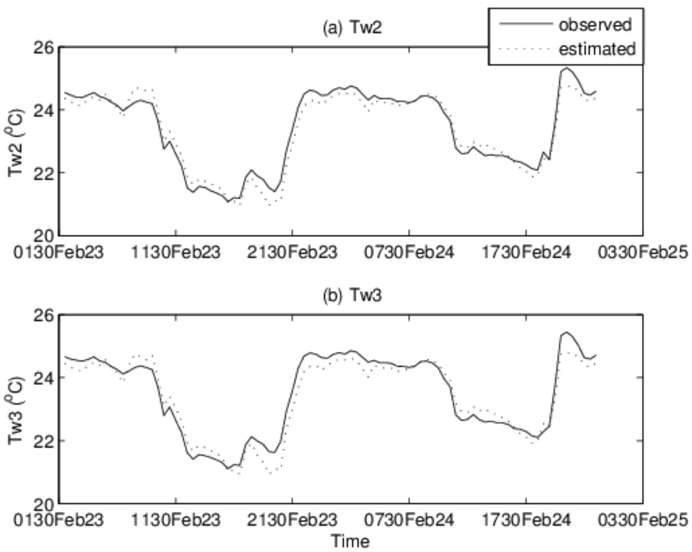

Fig. 1: Half hourly values of observed and estimated wet bulb temperature at (a) 5 m (Tw2) and (b) 10 m (Tw3) heightfor the cross-validated period (DOYs 54 and 55 in 2004). Tw2 and Tw3 are estimated from Tw1.

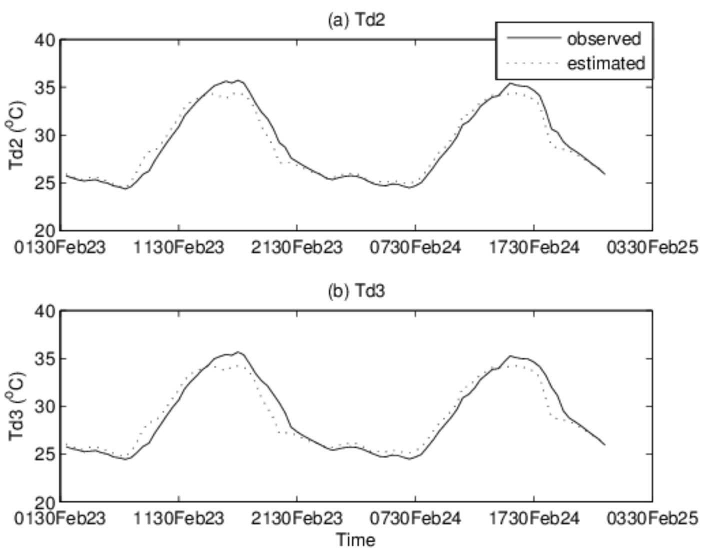

Fig. 2: Half hourly values of observed and estimated dry bulb temperature at (a) 5 m (Td2) and (b) 10 m (Td3) height for the cross-validated period (DOYs 54 and 55 in 2004). Both Td2 and Td3 are estimated from Td1.

Dry bulb temperature at 5 and 10 m, Td2 and Td3 are also estimated from dry bulb temperature at 1 m, Td1. The cross-validated correlation is also highly significant wih R^2 greater than 0.93 (Figs. 2a and b). The root mean square errors in the estimated are far less than the standard deviations in the observed Td2 and Td3 data. So they agree at 99.7% confidence level. Td1 is also used as the predictor in the estimation of soil temperature at depths 5 cm and 10 cm (Figs 3a and b). The cross-validated correlationsare highly significant at all levels of significance (0.1 to 5%) and the R^2 value >0.64. Adeniyi and Nymphas[5] estimated soil surface temperature from air temperature at 1 m and soil depth temperature at 5 cm. Their linear regression models yielded estimated soil surface temperature with very low bias.Wind speed at 14.8 m (W8) is estimated from the observed wind speed at 1 m (Fig 3c).The cross-validated R^2 is > 0.7 and the RMSE is less than the standard deviation in the observed W8 (Table 3) So the estimated W8 agrees with the observed at 99.7% confidence level. Global radiation is used as the predictor for Heat flux at depth 2 cm (Hf1), Rnet (Fig.4), Swd and Swu (Fig. 5).The cross-validated R^2 are all greater than 0.9 and their RMSE are all less than one standard deviation in the observed values except Swu which has RMSE a little higher than the standard deviation in the observed data. The observed and estimated data sets agree at 95 % or higher confidence level depending on the parameter.

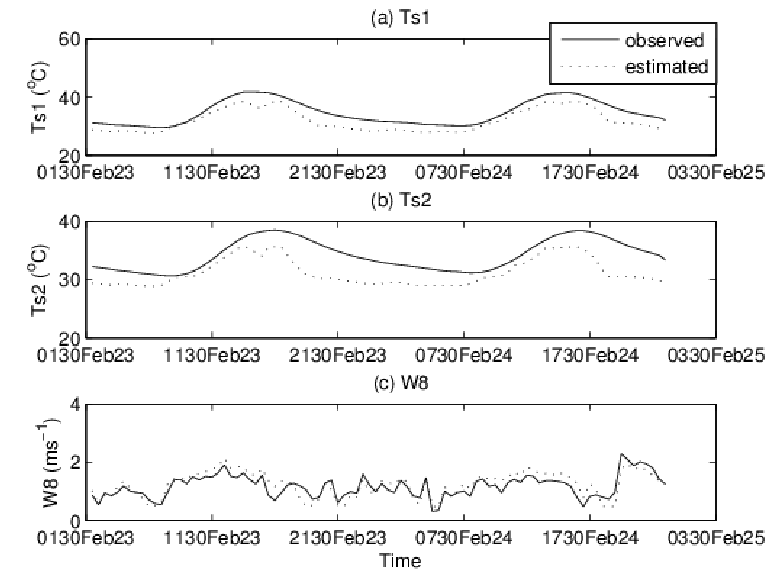

Fig. 3: Observed and estimated (a)Ts1 from Td1, (b) Ts2 from Td1 and (c) W8 from W1 for the cross-validated period (DOYs 54 and 55 in 2004)

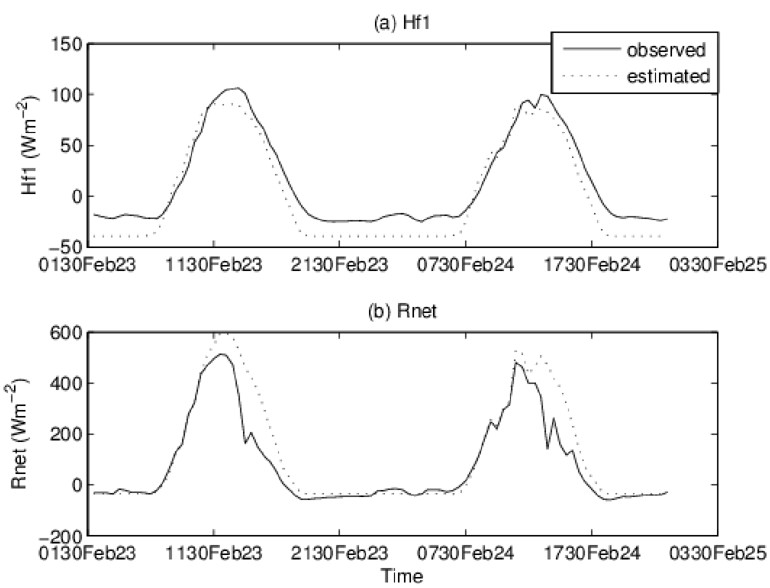

Fig. 4: Observed and estimated (a) Hf1 and (b) Rnet for the cross-validated period (DOYs 54 and 55 in 2004). Both Hf1 and Rnet are estimated from global radiation.

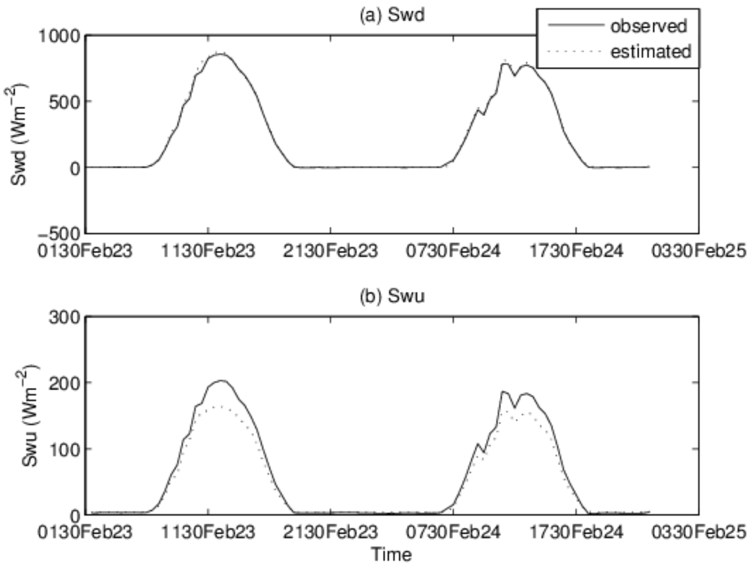

Fig. 5: Observed and estimated (a)Swd and (b) Swu for the cross-validated period (DOYs 54 and 55 in 2004). Both Swd and Swu are estimated from global radiation.

Reliable estimates of most parameters are found possible with the use of the polynomial regression model of order 16. The high order of the regression polynomial is applied in order to increase the precision of the estimated values. This is allowed because a polynomial of order k can be used to fit k+1 data points [23]and the number of data points used in fitting the polynomials is far more than 16. In a day with 30 minutes interval data, there are 48 data points. So there is a total of 624 data points in 13 days used for training the polynomial regression models. High cross-validated correlation, R^2 and low cross-validated RMSE are obtained for most parameters with the use of 92 data points not included in training the polynomial regression models. The obtained R^2 are comparable with those obtained in previous estimation works. For example, the R^2 between estimated and observed global radiation over Iraq [2] is between 0.79 and 0.94.Over Nigeria [4] obtained R^2 ranges from 7052 to 7447, the least $R^2$was found over the Guinea coast. The error in estimated global radiation over India [3] is within 20 % of the observed data, whereas the errors in the estimated data in this study are all within 5% of the observed values. This is because at 95% confidence level, most of the data are within two standard deviations and the RMSE in all the estimated data are less than two standard deviations (Table 3). The obtained regression polynomial coefficients (Table 2)can be applied in the estimation of unavailable meteorological parameters in areas where similar weather variations to those at Nimex-1 site prevail.

4.0 Conclusion

Meteorological data are usually available at Meteorological Agencies. However, there is a limit to the available parameters in the Agencies. More comprehensive parameters are measured for specific research purposes by the Atmospheric Scientists but such measurements are not routine in nature, they are usually for specific periods. The need for data on micrometeorological parameters that are not included in routine measurements can be met by estimation based on the available parameters. In this paper, 36 micrometeorological parameters from NIMEX_1 site are subjected to PCA and cluster analyses in order to group them according to their common variances. Six clusters are found and one parameter from each cluster is used as a predictor to estimate the other members of the cluster. Some parameters could not be estimated, for example the wind direction which is alone in a cluster. Non-coherent clusters are also neglected in the regression analysis for estimation. For example the cluster 3that consists of Ts3, Lwd and Press; there is weak correlation between the parameters and so no estimation is made. Other clusters allow the estimation of each of the members from one parameter in the cluster. The estimated datasets all have high R^2 and RMSE within 5% of the observed data. The polynomial regression coefficients are available for use in the humid tropical areas with the same weather conditions that prevailed during the NIMEX_1 experiment. The limitation of lack of data on some parameters can be eliminated with the application of the regression coefficients obtained in this study.

5.0 Acknowledgements

The author is grateful to the other members of the Nigerian Micrometeorological Experimental Group for making the NIMEX_1 experiment a reality. The NIMEX_1 experiment was sponsored by the International Programs in the Physical Sciences of Uppsala University, Sweden.

6.0 References

[1] Sombroek, W, Gommes R(1993). The climate change-agriculture conundrum. Paper presented at the Expert Consultation on Global Climate Change and Agricultural Production: Direct and Indirect Effects of Changing Hydrological, Soil and Plant Physiological Processes, FAO, 7-10 December 1993, 300 pp.

[2] Ali MA, Maylaa MK, Ali JM (2010). Estimation of Global Solar Radiation on Horizontal Surface Using Routine Meteorological Measurements for Different Cities in Iraq. *Asian Journal of Scientific Research 3: 240-248.*DOI**:** 10.3923/ajsr.2010.240.248.

[3] Kumar R, Umanad L (2005). Estimation of global radiation using clearness index model for sizing photovoltaic system. *Renewable Energy* 30:2221-2233. DOI: 10.1016/j.renene.2005.02.009.

[4] Babatunde A, Olusola O (2015). Regional estimation of solar radiation using routinal meteorological variables. *Journal of Environment and Earth Science* 5(17): 147-153.

[5] Adeniyi MO, Nymphas EF (2011). Estimation of bare soil surface temperature from air temperature and soil depth temperature. *International Journal of Natural and Applied Sciences* 7(4): 429-437.

[6] Jegede OO, Mauder M, Okogbue EC, Foken T, Balogun EE,Adedokun JA, Oladiran EO, Omotosho JA, Balogun AA,Oladosu OR, Sunmonu LA, Ayoola MA, Aregbesola TO, Ogolo, EO, Nymphas EF, Adeniyi MO, Olatona GI, Ladipo KO,Ohamobi SI, Gbobaniyi EO, Akinlade GO (2004). The Nigerianmicrometeorological experiment (NIMEX-1): an overview. *Ife JournalScience* 6(2):191–202.

[7] Mauder M, Jegede OO, Okogbue EC, Wimmer F, Foken T (2007). Surface energy balance measurements at a tropical site in WestAfrica during the transition from dry to wet season. *Theor ApplClimatol* 89:171–183.

[8] Adeniyi MO, Ogunsola, EO (2012). Assessment of the drag and bulk transfermethod in estimating sensible and latent heat fluxes in a tropical station', *Theoretical andApplied Climatology* 107(3-4): 511-518, DOI: 10.1007/s00704-011-0496-7.

[9] Foken T. (2003). Angewandte Meteorologie. MikrometeorologischeMethoden. Springer, Heidelberg, p 289.

[10] Adeniyi MO (2014). Variability of daily precipitation over Nigeria. *Meteorology and Atmospheric Physics*. 126: 161-176, DOI 10.1007/s00703-014-0340-6.

[11] Phillips ID, Denning H (2007) Winter daily precipitation variabilityover the south West Peninsula of England. *Theor Appl Climatol.* 87:103–122. DOI:10.1007/s00704-005-0226-0

[12] Chen GTJ, Jiang ZH, Wu MC (2003) Spring heavy rain events in Taiwan during warm episodes and the associated large-scale conditions. *Mon Weather Rev.* 131:1173–1188.

[13] Brito-Castillo L, Diaz-Castro S, Salinas-Zavala CA, Douglas AV. (2003). Reconstruction of long-term winter streamflow in theGulf of California continental watershed. *J Hydrol*. 278:39–50.

[14] Cattell RB. (1966). The scree test for the number of factors. MultivarBehav Res 1:245–276. DOI:10.1207/s15327906mbr0102_10.

[15] Kaiser HF (1960). The application of electronic computers to factoranalysis. *Educ Psychol Meas* 20:141–151. DOI:10.1177/001316446002000116.

[16] Jolliffe IT (2002). Principal component analysis, 2nd edn. Springer,New York.

[17] Gong X, Richman MB (1995). On the application of cluster analysis togrowing season precipitation data in north America east of theRockies. *J Climatol.* 8:897–931.

[18] Ward JH (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 58(301): 236–244.

[19] Clark WAV, Hosking PL (1986). Statistical methods for Geographers.Wiley 518, pp ISBN 0471818070.

[20] Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate changemodelling to impacts studies: recent advances in downscalingtechniques for hydrological modelling. *Int J Climatol*. 27(12):1547–1578. DOI:10.1002/joc.1556.

[21] Spak S, Holloway T, Lynn B, Goldberg R (2007) A comparison ofstatistical and dynamical downscaling for surface temperature inNorth America. *J Geophys Res Atmos.* 112:D08101. DOI:10.1029/2005JD006712.

[22] Adeniyi MO, Nymphas, EF, Oladiran, EO (2012).Characteristics of total solar radiation in an urban tropical environment. *International Journal of the Physical Sciences* 7(30): 5154-5161. DOI: 10.5897/IJPS09.253.

[23] Schrevens E, Portier K (2005). Experimental design in product development and optimization. In: Handbook of Food Science, Technology and Engineering, Edited by Hui YH, Sherkat F. Volume 3 Published by Taylor and Francis, pp 143_2:143:29.