Research Journal of Mathematics and Statistics 2(1): 23-31, 2010
ISSN: 2040-7505

Submitted Date: October 15, 2009 Accepted Date: November 12, 2009 Published Date: March 20, 2010

# Short and Long Memory Time Series Models of Relative Humidity of Jos Metropolis

[1]M.A. Chiawa, [2]B.K. Asare and [3]B. Audu
[1]Department of Mathematics and Computer Science, Benue State University, Makurdi, Nigeria
[2]Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria
[3]Department of Mathematics, Gombe State University, Gombe, Nigeria

**Abstract:** The percentage monthly relative humidity of Jos metropolis is examined in this study. Two models, a short memory seasonal autoregressive integrated moving average model [SARIMA(1,0,1)(2,1,2)] and long memory autoregressive fractional integrated moving average [ARFIMA(1,0.29,1)] are used to fit the same humidity data. Even though both models fit the data well, forecasts obtained from the ARFIMA(1,0.29,1) capture the swing in the data and resemble the actual values better than the forecasts using SARIMA(1,0,1)(2,1,2) model. This result shows that the Jos metropolitan data is better fitted by a long memory time series which captures the long swing in the weather data better than the short memory time series models whose effect quickly dies down.

**Key words:** Precipitation, autocorrelation function, autocovariance, spectrum, periodogram, fractional integration

## INTRODUCTION

The amount of water vapor contained in the air at a particular time is called its humidity. Even though there are three different ways to measure humidity, the most frequently used is the relative humidity. Relative humidity indicates the likelihood of precipitation, dew, or fog. The relative humidity if it is high makes us feel hotter outside. This is usually experienced in the summer, that is, between June and September. It also reduces sweating which should cool the body thus preventing the evaporation of perspiration from the skin. This is why geographers usually state that ‘warmer air holds more moisture’. In warmer air, there is more energy for more water molecules to hold themselves and overcome hydrogen bonds which seek to pull water molecules together.

In the development of space modeling of time series data with long memory dependence, Haslett and Raftery (1989) assessed Ireland’s wind power resources using autoregressive fractional integrated moving average (ARFIMA). Since then it has become usual to model meteorological time series with an ARFIMA model (Hsu *et al*., 1998; Hunt and Nason, 2001). Haslett and Raftery (1989) and Boutte *et al.* (2004) fit an ARFIMA model to the Irish wind data and found seasonal component in the series with the residuals exhibiting an autoregressive moving average (ARMA) process. Boutte *et al.* (2004) then went ahead to conclude that ARMA, Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA), are especially suited for modeling short memory processes with little or no persistence shock.

To our knowledge, ARFIMA models have not been used substantially to model humidity data in Nigeria. This is therefore one of the contributions of this research. The second contribution of the research is to be able to predict the humidity pattern in Jos metropolis with some degree of accuracy. This enables the meteorologist warn people in advance the next weather pattern to expect within a couple of months.

**Theoretical framework:** Meteorological studies suggest that meteorological variables exhibit characteristics that are more consistent with long memory (Haslett and Raftery, 1989; Montanari *et al*., 1996; Hunt and Nason, 2001; Bhardwaj and Swanson, 2004). In a long memory process, the autocorrelation of a variable decays very slowly, in other words, the autocorrelation function of a long memory process typically decays at a hyperbolic rate (Haslett and Raftery, 1989). That is to say that these processes have autocovariances that are not absolutely summable (Hurst, 1951). A sequence of matrices $\{\Psi_s\}_{-\infty}^{\infty}$ is absolutely summable if each of its element forms an absolutely summable scalar sequence, that is $\sum_{s=0}^{\infty}\left|\psi_{ij}^{(s)}\right| < \infty$ for i,j = 1,2, ..., n, where $\psi_{ij}^{(s)}$ is the row i column j of $\psi_s$.

**Corresponding Author:** M.A. Chiawa, Department of Mathematics and Computer Science, Benue State University, Makurdi, Nigeria

Long memory processes have been popularized by Grange and Joyeux (1980) and Hosking (1981). A long memory autoregressive fractional integrated moving average (ARFIMA) process is a process that must be differenced d-times, where d can take on non-integer values, to achieve a stationary process. This process also has the property that its spectrum is unbounded at the origin. The notion of fractional integration has proven to be quite important in modeling meteorological data since the study of Haslett and Raftery (1989). Whitcher (1998) use the discrete wavelet transform to construct a test for homogeneity of variance in time series exhibiting long memory characteristics and applied the result to the Nile river water levels.

## MATERIALS AND METHODS

This study was conducted at the Jos International Airport by the Meteorological Department of the Federal Ministry of Aviation in Nigeria. Monthly percentage relative humidity time series data was collected from January 1985 to December 2004. Thereafter, the data generating process of the SARIMA and ARFIMA models are tested. To apply the ARFIMA and SARIMA tests the variables are first examined for unit root and stationarity. The Augmented Dickey Fuller (ADF) test and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test are used for this purpose. These preliminary tests are necessary in order to determine the order of nonstationarity of the data. If the series is integrated of order one then using ARFIMA and SARIMA models on the data results in good specification of these models. These tests are discussed below.

**Adjusted Dickey-Fuller (ADF) test:** The ADF regression equation due to Dickey and Fuller (1979) and Said and Dickey (1984) is given by:

$$\Delta y_t = \mu_0 + \mu_1 t + \phi y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \varepsilon_t$$

$$t = p+1, p+2, \ldots, T. \qquad (1)$$

where $\mu_o$ is the intercept, $\mu_1^t$ represents the trend incase it is present, ϕ is the coefficient of the lagged dependent variable . $y_{t-1}$ and p lags of $\Delta y_{t-j}$ with coefficients $a_j$ are added to account for serial correlation in the residuals. The null hypothesis $H_O : \phi = 0$ is that the series has unit root while the alternative hypothesis $H_1: \phi \neq 0$ is that the series is stationary. The ADF test statistic is given by

$$ADF = \frac{\hat{\phi}}{SE(\hat{\phi})}$$

where $SE(\hat{\phi})$ is the standard error for $\hat{\phi}$, and ˆ denotes estimate. The null hypothesis of unit root is accepted if the test statistic is greater than the critical values.

**Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test:** The integration properties of a series $y_t$ may also be investigated by testing the null hypothesis that the series is stationary against a unit root. Kwiatkowski *et al.* (1992) derived a test for this pair of hypotheses. Assuming no linear trend term, the data generating process is given as

$$y_t = x_t + z_t \qquad (2)$$

where $x_t$ is a random walk, $x_t = x_{t-1} + v_t$, $v_t \sim iid(0, \sigma_v^2)$ and $z_t$ is a stationary process. Kwiatkowski *et al.*, (1992) propose the following test statistic

$$KPSS = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\infty^2} \qquad (3)$$

where $S_t = \sum_{j=1}^{t} \hat{w}_j$ with $w_j = y_t - \bar{y}$ and $\hat{\sigma}_\infty^2$ an estimator of the long run variance of

$$z_t, \sigma_\infty^2 = \lim_{T \to \infty} T^{-1} Var\left(\sum_{t=1}^{T} Z_t\right).$$

The null hypothesis of the test is $H_0 : \sigma_v^2 = 0$ against the alternative hypothesis $H_1 : \sigma_v^2 \neq 0$. This test uses the Bartlett window with a lag truncation parameter

$$l_q = q\left(\frac{T}{100}\right)^{\frac{1}{4}} \qquad (4)$$

where

$$\hat{\sigma}_\infty^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{w}_t^2 + 2 \sum_{j=1}^{l_q} w_j \left( \frac{1}{T} \sum_{t=j+1}^{T} \hat{w}_t \hat{w}_{t-j} \right)$$

and

$$w_j = 1 - \frac{j}{l_q + 1}$$

Reject the null hypothesis if the test statistic is greater than the asymptotic critical values.

**The ARFIMA model:** ARFIMA models were first introduced by Granger and Joyeux (1980) and Hosking (1981). An ARFIMA *(p,d',q)* process is a stationary process that satisfies:

$$\phi_p(L)(1-L)^d y_t = c + \theta_q(L)\varepsilon_t, \quad t = 1,2,\ldots,T \tag{5}$$

where *d* is the parameter of fractional differentiation, c is a constant and $\phi_p$ and $\theta_q$ are autoregressive and moving average polynomials of order p and q, respectively. The autocorrelations, $p_k$ for an ARFIMA process for large k and d<1/2 are given by the following approximation (Granger and Joyeux, 1980).

$$\rho_k \approx \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \tag{6}$$

which is a monotonic, hyperbolic function. The autocorrelation function of ARFIMA decreases slowly to zero while its spectral density is infinite (Hosking, 1981).

**Estimating ARFIMA Models:** ARFIMA is a more general modeling of the autoregressive integrated moving average (ARIMA) model that allows non-integer d values. A typical ARFIMA model is given by the equation

$$(1-L)^d \phi(L) y_t = \theta(L)\varepsilon_t, \tag{7}$$

where

$$(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \ldots \tag{8}$$

In this study a simple method of estimating d is used. This method is due to Geweke and Porter-Hudak (1983) and starts the estimation of d with the spectral density function of $y_t$

$$f_x(\lambda) = \left(\sigma^2 / 2\pi\right)\left\{4\sin^2(\lambda/2)\right\}^{-d} f_u(\lambda) \tag{9}$$

Take logarithms and evaluate at harmonic frequencies

$$\lambda_j = \frac{2\pi j}{T}, \quad j = 0, 1, 2, \ldots, T-1$$

add the periodogram $I(\lambda_j)$ at ordinate j, to both sides of (9) yields

$$\begin{aligned}\ln\{I(\lambda_j)\} &= \ln\{\sigma^2 f_u(0)/2\pi\} \\ &- d\ln\{4\sin^2(\lambda/2)\} + \ln\{I(\lambda_j)/f_x(\lambda_j)\} \\ &+ \ln\{f_u(\lambda_j)/f_u(0)\}\end{aligned} \tag{10}$$

The last term on the right-hand side of (10) becomes negligible when low-frequency ordinates $\lambda_j$ are close to zero. Equation (10) then results in the following simple regression equation

$$\ln\left[I(\lambda_j)\right] = \beta_0 + \beta_1 \ln\left[4\sin^2(\lambda_j/2)\right] + \eta_j \tag{11}$$

where the intercept $\beta_0 = \ln\{\sigma^2 f_u(0)/2\pi\}$, with parameter $\beta_1 = -d$ and the error term is

$$\eta_j = In\{I(\lambda_j)/f_x(\lambda_j)\}$$

The number of ordinates j to be used in the estimation of the regression truncates at n = g(T), where g(T) satisfy the following conditions:

(i) $\lim_{T\to\infty} g(T) = \infty$ and

(ii) $\lim_{T\to\infty}\left(\frac{g(T)}{T}\right) = 0$

The function $g(T) = T^{\mu}$, with $0 < \mu < 1$ is the number of ordinates used to estimate *d* and this satisfies both conditions. The estimator of *d* is therefore consistent (Hurvich *et al*., 1998). The test of hypothesis for the parameter *d* can be done based on the asymptotic distribution of $\tilde{p}$ derived by Geweke and Porter-Hudak (1983).

$$\hat{d} \to N\left(d, \pi^2/6\sum_{i=1}^{n}(y_i - \bar{y})^2\right) \tag{12}$$

where $y_i$ is the regressor *ln* [4 Sin$^2$ ($\lambda_j$ /2)] . Since $\lim_{T\to\infty} g(T) = \infty$ and $\lim_{T\to\infty}[g(T)/T] = 0$ it follows that *p* lim $s^2 = \pi^2/6$, where $s^2$ is the sample variance of the residuals from the regression Eq. (11).

The value of the power factor, $\mu$ is the determinant of ordinates included in the regression. Traditionally the number of ordinates is chosen from the interval [$T^{0.45}$,

$T^{0.55}$]. However, Hurvich *et al.* (1998) recently showed that the optimal lag m is $O(T^{0.8})$.

**SARIMA model:** A Seasonal Autoregressive Integrated Moving Average (SARIMA) model denoted by SARIMA $(p,d,q)(P,D,Q)$ is given as

$$(1-L)^d (1-L^s)^D \phi(L)\Phi_s(L) y_t = a + \theta(L)\Theta_s(L) e_t \tag{13}$$

where $p$ denotes the number of autoregressive terms, $q$ the number of moving average terms and $d$ is an integer which denotes the number of times the series must be differenced to attain stationarity. $P$ is the number of seasonal autoregressive components, $Q$ denotes the number of seasonal moving average terms and $D$ the number of seasonal differences required to induce stationarity. $a$ is a constant, $e_t \sim NID(\mu, \sigma^2)$ i.e. $e_t$ is a sequence of uncorrelated normally distributed random variables with the same mean $\mu$ and variance $\sigma^2$, $L$ is the lag operator defined by $L_k X_i = X_{t-k}$ ,

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots \theta_p L^p,$$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p,$$

$$\Phi(L) = 1 - \Phi_{1s} L^{1s} - \Phi_{2s} L^{2s} - \ldots - \Phi_{ps} L^{ps}$$

and

$$\Theta(L) = 1 + \Theta_{1s} L^{1s} + \Theta_{2s} L^{2s} + \ldots + \Theta_{Qs} L^{Qs}.$$

**Residual analysis:** The model specified is statistically tested for adequacy using a barrage of tests. These tests include: the Portmanteau test, Lagrange multiplier (LM) test, Jarque-Bera test, autocorrelations and partial autocorrelations, and the Ljung-Box test. Details of these tests can be found in Brown, Durbin and Evans (1975), Ljung and Box (1979), Jarque and Bera (1987) and Harvey (1990, 1993). Detailed discussions for these tests are skipped in this section to reduce the size of the paper.

**Analysis of data:**

**Initial analysis of data:** A time plot of the original series is conducted in Fig. 1. A visual inspection of the plot shows that the series has constant mean and variance. The plot however does not show any evidence of stationarity.

We proceed to examine the autocorrelation function (Fig. 2). The autocorrelation function (ACF) is strongly periodic with period 12 and very persistent. Clearly the

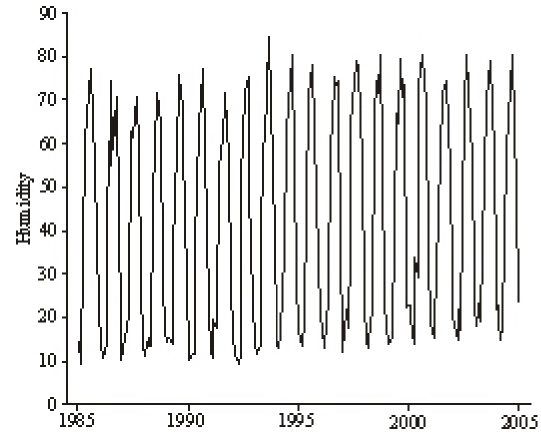


Fig. 1: Time plot of the monthly percentage relative humidity of Jos

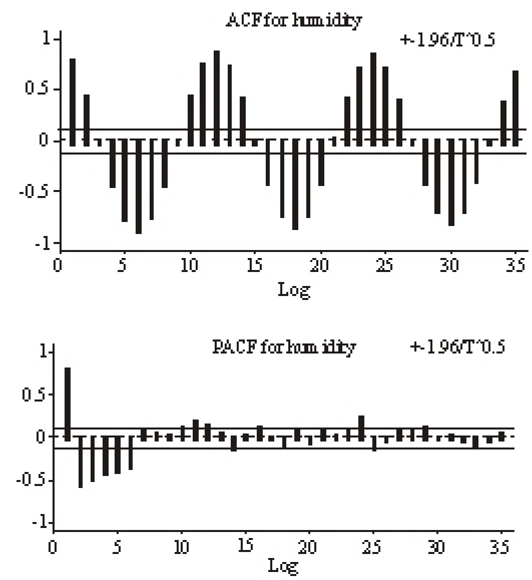


Fig. 2: The ACF and the PACF of the monthly percentage relative humidity of Jos

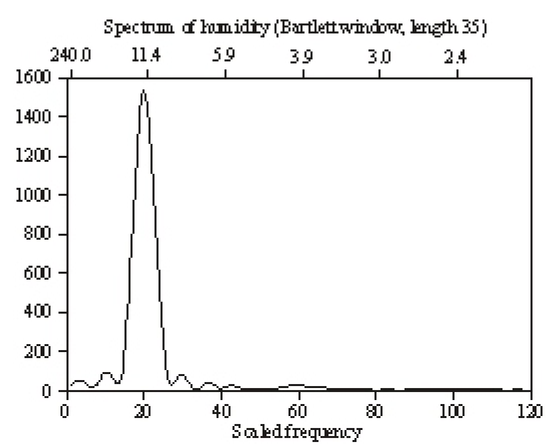


Fig. 3: The Periodogram of the monthly percentage relative humidity of Jos

data shows seasonal behaviour. The spectral analysis is then used to test the existence of periodicity in the time series.

Figure 3 shows the periodogram of the series. It can be seen that there is a frequency of 1 cycle every 12 months. We therefore conclude that the data generating process is seasonal and the length is 12 months, which coincides with the result of the ACF.

**Unit root test for the humidity data:** The ADF test is conducted at levels with seasonal dummies because the time series is strongly periodic. The result of the test given in Table 1 shows that the series at lag 4 is stationary since at 5% the test statistic is -5.14 with the critical value of -2.86.

**Formulation of the SARIMA model for humidity data:** The order of seasonal integration as shown in Fig. 3 of the series is one as there is a single spike in the periodogram. Since the series is stationary (see results of the ADF test) d = 0. The result of the spectral analysis shows that s = 12 and the order of seasonal integration of the series is 1. The optimal lag lengths for both the AR and MA are selected using the information criteria. These are also used to determine the lag lengths of seasonal integration for the AR and MA. Three candidates models are obtained using p = 1, 2; q = 1, 2; P = 1, 2; Q = 1, 2. Out of these models a parsimonious model is obtained which has the lowest AIC and log likelihood function. The value of the log likelihood function and the AIC for three candidates SARIMA models are computed and reported in Table 2.

Before the choice of our model, standard diagnostic test are conducted using the residual ACF and PACF of all the candidate models. As seen in Fig. 4, 6 and 8 all the models pass the standard diagnostic testing criteria as their residual values lie within the 95% confidence interval band. The residuals are also plotted against time to ascertain whether or not the models have passed the standard test criteria of being white noise. These are reported in Fig. 5, 7 and 9. The results show that the residual plots for all the models are white noise. Since all the residuals of the candidates models behave well, we proceed to select SARIMA(1,0,1)(2,1,2) which has the lowest AIC and log likelihood as the best model.

**Parameter estimates of the model using humidity data:** After the best model has been chosen, the parameters of the model are next estimated. The result of the parameter estimates of the optimal model are shown in Table 3. After fitting the model, we check the model for adequacy. This is what we are going to see in the next section.

**Model checking for SARIMA(1,0,1)(2,1,2):** The model is next tested for adequacy using three diagnostic tests, Ljung-Box test, Jarque-Bera test and ARCH-LM test. The results of these tests are given in Table 4.

Table 1: Results of the ADF test down procedure

| Actual no. of lags | 5% | Test statistics | Decision |
|---|---|---|---|
| 4 | -2.86 | -5.14 | Stationary |

Table 2: Results of SARIMA model identification for the humidity of jos

| Model | Loglikelihood | AIC |
|---|---|---|
| SARIMA(2,0,2)(2,1,2) | -723.44 | 1466.88 |
| SARIMA(2,0,2)(1,1,1) | -737.16 | 1490.31 |
| SARIMA(1,0,1)(2,1,2) | -724.65 | 1465.30 |

Table 3: Results of SARIMA model estimation for humidity time series

| Variable | Coefficient | Std. Error | T-Stat | P-Value |
|---|---|---|---|---|
| Constant | 0.2702 | 0.120 | 02.246 | 0.02500 |
| Phi-1 | 0.9300 | 0.070 | 13.033 | 0.00001 |
| Phi-1 | - 1.0140 | 0.080 | - 13.121 | 0.00001 |
| Phi-2 | - 0.0150 | 0.076 | - 00.200 | 0.84400 |
| Theta-1 | - 0.8600 | 0.100 | - 08.960 | 0.00001 |
| Theta-1 | 0.0700 | 0.060 | 01.183 | 0.24000 |
| Theta-2 | - 0.8700 | 0.050 | - 16.800 | 0.00001 |

Table 4: Summary of Diagnostic Tests of SARIMA (1,0,1) (2,1,2) Model

| Test | p-value ($\chi^2$) |
|---|---|
| Ljung-Box | 0.3580 |
| Jarque-Bera | 0.4137 |
| ARCH-LM | 0.0621 |

Table 5: Results of SARIMA (1,0,1) (2,1,2) forecast values for two years

| In sample forecast | | Out of sample forecast | |
|---|---|---|---|
| Humidity | Prediction | Humidity | Prediction |
| 22 | 16.54 | 20 | 17.69 |
| 18 | 17.12 | 27 | 21.69 |
| 23 | 21.45 | 19 | 30.77 |
| 58 | 48.06 | 45 | 50.46 |
| 70 | 65.17 | 59 | 70.97 |
| 72 | 70.16 | 72 | 83.56 |
| 80 | 81.72 | 76 | 86.10 |
| 80 | 80.66 | 80 | 91.31 |
| 66 | 69.61 | 76 | 81.69 |
| 59 | 53.75 | 40 | 69.00 |
| 26 | 28.86 | 22 | 50.44 |
| 16 | 19.98 | 24 | 42.41 |
| 14 | 17.92 | 17 | 36.21 |
| 13 | 17.52 | 15 | 29.80 |
| 20 | 19.60 | 20 | 36.57 |
| 20 | 37.59 | 42 | 27.97 |
| 66 | 62.19 | 61 | 82.68 |
| 76 | 65.35 | 66 | 104.14 |
| 72 | 79.10 | 74 | 113.98 |
| 82 | 77.32 | 82 | 102.29 |
| 62 | 70.90 | 66 | 91.94 |
| 30 | 42.79 | 50 | 65.41 |
| 20 | 23.03 | 24 | 34.35 |
| 18 | 15.08 | 24 | 27.24 |

The results of the tests show that the model is adequate at 5% as all the p-values are greater than 0.05. The model is then used to forecast future values of the series.

**SARIMA(1,0,1)(2,1,2) Forecast Evaluation:** After a good SARIMA model has been fitted, its ability to forecast the time series data is tested. This further testifies

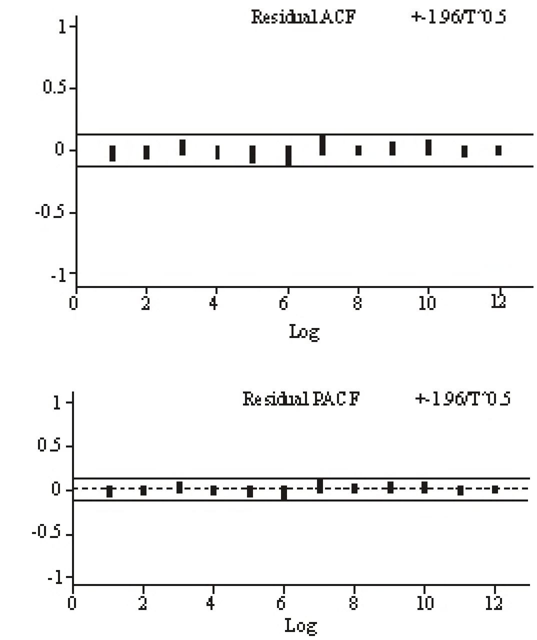


Fig. 4: ACF and PACF of SARIMA(2,0,2)(2,1,2)

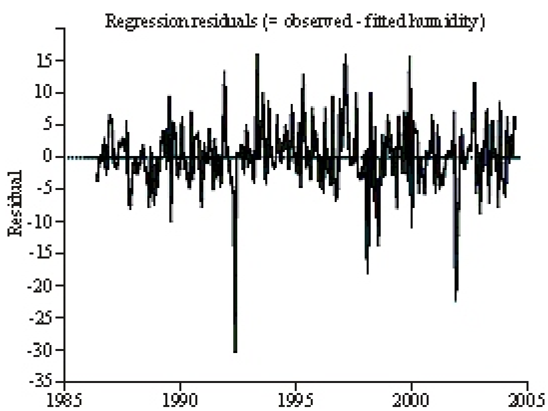


Fig. 5: Residual plot of SARIMA(2,0,2)(2,1,2)

the validity of this model. Table 5 contains the two years in-sample and out of sample humidity forecasts using two benchmarks.

The prediction is consistent with the observed humidity and therefore the model is adequate, valid and good. Forecast errors are then computed in order to assess the performance of the forecasts, in particular, the smaller the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) the better the forecasts. These values are reported in Table 6 for both in-sample and out-of-sample forecast performance of the SARIMA(1,0,1)(2,1,2) model to determine its forecast ability and to make inference on which mode of forecast is better for the model.

Since the RMSE, MAE and the MAPE of the in-sample forecast are smaller than those of the out-of-sample forecast, the result suggests that the in-sample

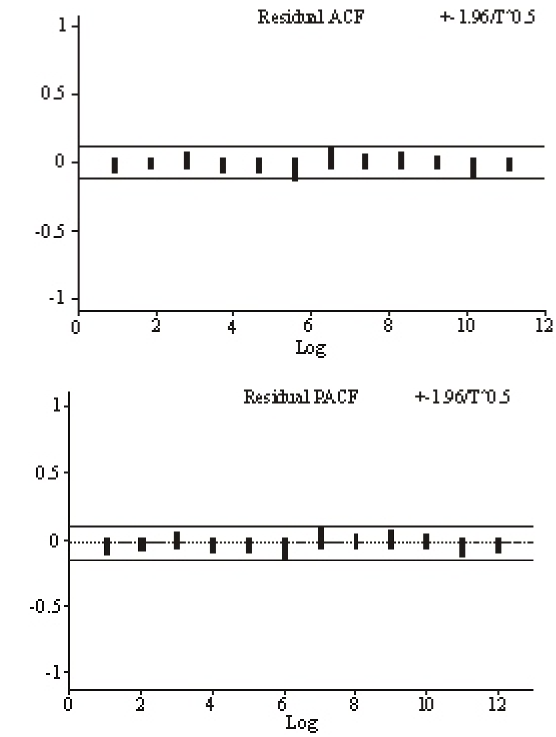


Fig. 6: Residual ACF and PACF of SARIMA (2,0,2) (1,1,1)

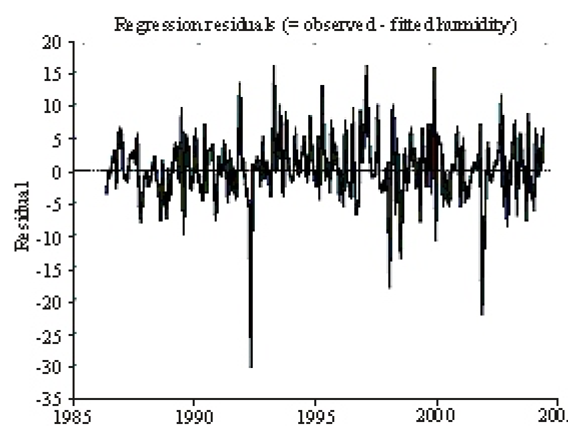


Fig. 7: Residual plot against time of SARIMA (2,0,2) (1,1,1)

forecasts are better than the out-of-sample forecasts. This shows one of the inadequacies of the model even though it has passed all the tests. One of those things that may cause such behavior is choosing d=0 when it is a fraction (Haslett and Raftery, 1989).

An alternative way of modeling the meteorological data is by fitting an ARFIMA model (Haslett and Raftery, 1989). In the subsequent section this task is accomplished and the result compared with the SARIMA model.

**ARFIMA modelling of the relative humidity data:** Here, an ARFIMA model is fitted to the meteorological data. Before estimating the *d* the ACF of the monthly

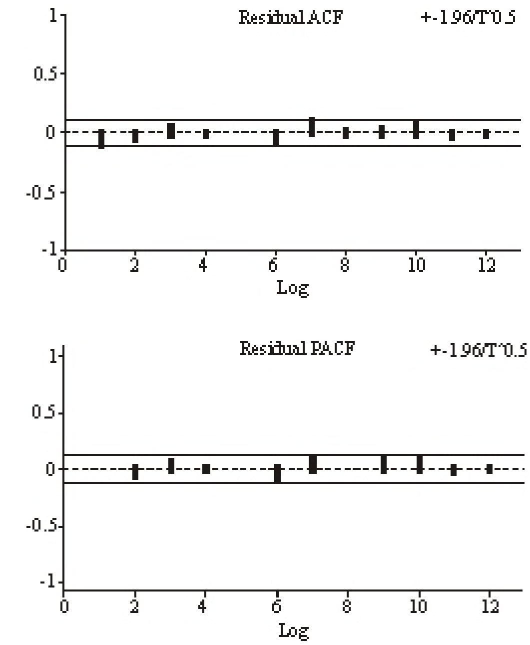


Fig. 8: Residual ACF and PACF of SARIMA (1,0,1) (2,1,2)

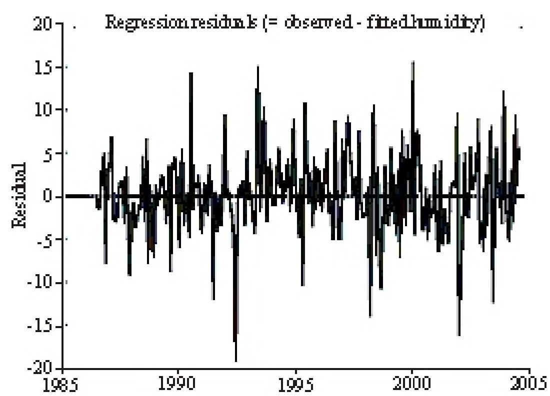


Fig. 9: Residual plot against time of SARIMA (1,0,1) (2,1,2)

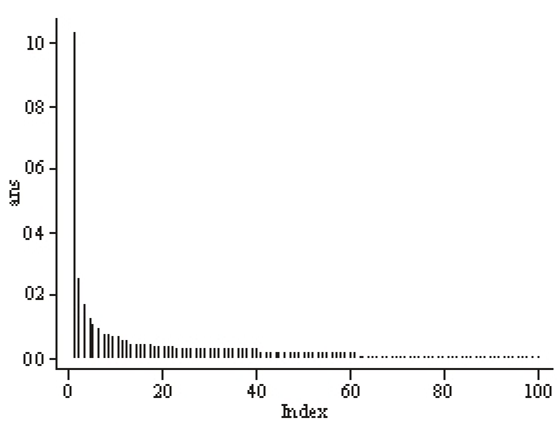


Fig. 10: The autocorrelation function of the humidity of Jos Metropolis

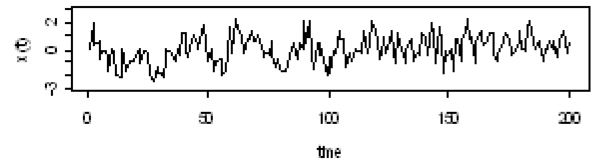


Fig. 11: The smoothed Periodogram

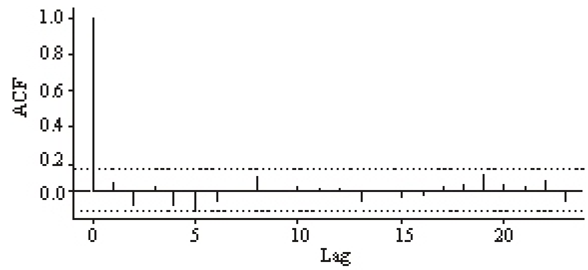


Fig. 12: The ACF of the fractionally differenced humidity Time Series

Table 6: Result of forecast comparison of SARIMA (1,0,1) (2,1,2) Model

| Mode of forecast | RMSE | MAE | MAPE |
|---|---|---|---|
| In-sample | 6.56 | 4.82 | 0.16 |
| Out-of-sample | 19.1 | 16.71 | 0.46 |

Table 7: Results of Estimating the *d* parameter

| Coefficient | Estimate | Asymptotic standard deviation | Standard error deviation |
|---|---|---|---|
| *D* | 0.2911554 | 0.2196708 | 0.1770737 |

relative humidity of the long memory process is conducted as shown in Fig. 10. The autocorrelation function of the humidity decreases slowly at a hyperbolic rate in conformity with a fractional integrated series (Haslett and Raftery, 1989; Gil-Alana, 2008).

A spectral density of the series is then conducted to determine the homogeneity of variance the smoothed periodogram. Figure 11 is the result of the smoothed periodogram.

The smoothed periodogram of the series show that it has constant variance. This is a third property of a long memory series. (Haslett and Raftrey, 1989). The degree of autocorrelation in the fractionally differenced humidity series is examined using the autocorrelation function as shown in Fig. 12.

Even though the series has been fractionally differenced. and is stationary with a zero infinite long memory characteristics, it still exhibits the behavior of an ARMA process (Haslett and Raftery, 1989).

**Estimating d using Geweke Porter-Hudak method:** The long memory parameter d is estimated using the Geweke and Porter-Hudak (1983) method. The estimated value of the parameter, its asymptotic deviation value and regression standard deviation values are reported in Table 7.

After estimating the long memory parameter *d* and using it to fractionally difference the humidity time series, we move ahead to model the series as an ARFIMA process. First we start by identifying the ARFIMA model that best fits the series.

**The ARFIMA model identification:** The particular optimization routine of the log-likelihood function shows that among the ARFIMA models, the optimal model for the fractionally difference series is ARFIMA (1, 0.29, 1) since this model exhibit the smallest values of AIC and log likelihood (Table 8).

The results in the Table 8 show that ARFIMA (1, 0.29,1) model is the best candidate model with the least value of AIC and the log likelihood. Before we draw any conclusion, it is necessary to examine the residuals of all the models to see if they pass the diagnostic tests. Table 8 shows the results of the diagnostic tests.

The diagnostic test results in Table 9, shows that all the models have passed the diagnostic tests of normality, serial correlation and residual autocorrelation, since the p-values are greater than 0.05. We therefore proceed to estimate the parameters of the ARFIMA(1, 0.29,1) model selected.

**Estimating the parameters of ARFIMA (1,0.29,1) model:** The results of the estimated parameters of ARMA (1,0.29,1) model are shown in Table 10. Having fitted a model to the fractionally differenced humidity time series, we check the model for adequacy.

**ARFIMA(1,0.29,1) model checking:** The residual plots of the ARFIMA(1,0.29,1) model is first performed to examine the adequacy of the model. The residual ACF and PACF of the fitted ARFIMA(1,0.29,1) model of the fractionally differenced humidity time series are given in Fig. 13. The plots show that there is no serial correlation observed in the residuals of the series therefore the model is adequate and good.

The results of the diagnostic tests are shown in Table 11. The existence of serial correlations in the residuals is rejected based on the results of the Lagrange multiplier tests. The Ljung-Box test also rejects the presence of residuals correlation. The results of the Jarque-Bera test, the skewness and kurtosis which are tests for normality of the residuals, show that the residuals are normally distributed as the skewness is close to zero and kurtosis close to 3.

**Forecast evaluation of ARFIMA(1,0.29,1) model:** After checking the model for adequacy, we finally study its forecast values. Both the in-sample and out-of-sample forecasts are computed and shown in Table 12. The observed and forecasted values are very close which that ARFIMA(1,0.29,1) model is adequate, valid and good.

We next compare the in-sample forecast with the out-of-sample forecast of the ARFIMA(1,0.29,1) to evaluate their performance. RMSE, MAE and MAPE are used for this purpose (Table 13).

Table 8: The result of AIC and log likelihood tests

| Model | AIC value | SE | Log likelihood |
|---|---|---|---|
| ARFIMA (2, 0.29, 1) | 727 | 1.09 | -359.59 |
| ARFIMA (2, 0.29, 2) | 727 | 1.09 | -358.90 |
| ARFIMA (3, 0.29, 3) | 723 | 1.07 | -354.62 |
| ARFIMA (1, 0.29, 1) | 725 | 1.09 | -359.88 |

Table 9: Results of diagnostic tests

| Model | Normality (p-value) | ARCH-LM (p-value) | Portmanteau (p-value) |
|---|---|---|---|
| ARFIMA(2, 0.29, 1) | 0.7621 | 0.2010 | 0.3523 |
| ARFIMA(2, 0.29, 2) | 0.7446 | 0.2044 | 0.4221 |
| ARFIMA(3, 0.29, 3) | 0.6645 | 0.2068 | 0.3693 |
| ARFIMA(1, 0.29, 1) | 0.7850 | 0.2341 | 0.4252 |

Table 10: Results of the estimated ARFIMA (1, 0.29, 1) parameters

| Estimates | Coefficient | SE | T-ratio | Approx. prob. |
|---|---|---|---|---|
| AR1 | -0.455 | 0.39 | -1.15 | 0.25 |
| MA1 | -0.546 | 0.37 | -1.54 | 0.12 |
| Constant | 0.055 | 0.08 | 0.72 | 0.47 |

Table 11: Misspecification tests

| | Type of test | Test stat | p-value | Decision rule |
|---|---|---|---|---|
| Serial Corr | LM-type | 0.8661 | 0.2341 | Reject |
| Resid Corr | Ljung-Box (LB)-type | 7.7556 | 0.4252` | Reject |
| ARCH | ARCH-LM | 8.0547 | 0.5286` | Reject |
| Normality | Skewness | 0.1555 | 0.7850 | Accept |
| Normality | Kurtosis | 3.0440 | 0.7850 | Accept |
| Normality | JB test | 0.9872 | 0.7850 | Accept |

Table 12: Results of forecast evaluation of ARFIMA (1,0.29,1) model

| In Sample forecast | | Out of sample forecast | |
|---|---|---|---|
| Humidity | Prediction | Humidity | Prediction |
| 22 | 21.7020 | 20 | 19.83 |
| 18 | 17.4852 | 27 | 26.99 |
| 23 | 23.4589 | 19 | 18.88 |
| 58 | 57.5854 | 45 | 45.02 |
| 70 | 70.2214 | 59 | 58.91 |
| 72 | 71.9539 | 72 | 72.04 |
| 80 | 79.8470 | 76 | 75.92 |
| 80 | 80.2920 | 80 | 80.05 |
| 66 | 65.6393 | 76 | 75.93 |
| 59 | 59.3617 | 40 | 40.06 |
| 26 | 25.7297 | 22 | 21.93 |
| 16 | 16.1518 | 24 | 24.06 |
| 14 | 14.0085 | 17 | 16.94 |
| 13 | 12.8771 | 15 | 15.06 |
| 20 | 20.2443 | 20 | 19.94 |
| 20 | 20.2621 | 42 | 42.06 |
| 66 | 66.2726 | 61 | 60.94 |
| 76 | 75.8219 | 66 | 66.06 |
| 72 | 72.1016 | 74 | 73.94 |
| 82 | 82.0332 | 82 | 81.95 |
| 62 | 61.8988 | 66 | 65.95 |
| 30 | 30.1976 | 50 | 50.05 |
| 20 | 19.8100 | 24 | 23.95 |
| 18 | 18.2024 | 24 | 24.05 |

Table 13: Results of forecast for ARFIMA(1, 0.29, 1) model

| Mode of forecast | RMSE | MAE | MAPE |
|---|---|---|---|
| In-sample | 0.280 | 0.230 | 0.008 |
| Out-of-sample | 0.069 | 0.063 | 0.002 |

Table 14: Results of the benchmark evaluation for the two types of models

| Model forecast | RMSE | MAE | MAPE |
|---|---|---|---|
| SARIMA(1,0,1)(2,1,2) | 6.56 | 80.36 | 0.1621 |
| ARFIMA(1,0.29,1) | 0.28 | 00.226 | 0.0080 |

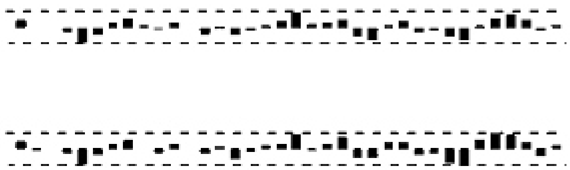

Fig. 13: Residuals ACF and PACF of the ARFIMA (1,0.29,1) Model

We notice that for the ARFIMA(1, 0.29, 1) model the out-of-sample forecast is better than the in-sample counterpart since it has smaller values for the RMSE, MAE and MAPE. This shows that ARFIMA(1, 0.29, 1) model has good prediction power.

We next compare the SARIMA(1,0,1)(2,1,2) short memory model and the ARFIMA(1,0.29,1) long memory model. The three benchmarks used are the RMSE, MAE and MAPE. The results of these tests are given in Table 14. The results show that ARFIMA(1,0.29,1) has smaller values of three benchmarks hence it is a better model for humidity data than SARIMA(1,0,1)(2,1,2).

## CONCLUSION

A short memory model SARIMA(1,0,1)(2,1,2) and a long memory model ARFIMA(1,0.29,1) are used to fit the same humidity data. Even though both models fit the data well, forecasts obtained using the long memory model tend to capture the swing in the data and resemble the actual values better than the forecasts using the short memory model. The conclusion we draw from these results is that the Jos humidity data is better fitted by a long memory time series which captures the long swing in the weather data better than the short memory time series models whose effect quickly dies down.

## REFERENCES


Bhardwaj, G. and N.R. Swanson, 2004. An empirical investigation of the usefulness of arfima models for predicting macroeconomic and financial time series. J. Econometrics, 131(2): 539-578.

Bouette, J.C., J.F. Chassagneux, D. Sibai, R. Terron and A. Charpentier, 2004. Wind in Ireland: Long memory or seasonal effect? Stoch Environ. Res. Risk Assess., 1: 12.

Brown, R.L., J. Durbin and J.M. Evans, 1975. Techniques for testing the constancy of regression relationships over time. J. Royal Stat. Soc., Series B37: 149-192.

Dickey, D.A. and W.A. Fuller, 1979. Distributions of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc., 74: 427-431.

Gil-Alana, L.A., 2008. Fractional integration and structural breaks at unknown periods of time. J. Time Series Anal., 29(1): 163-185.

Granger, C.W.J. and R. Joyeux, 1980. An introduction to long memory time series models and fractional differencing. J. Time Series Anal., 1: 15-29.

Geweke, J. and S. Porter-Hudak, 1983. The estimation and application of long memory time series models. J. Time Series Anal., 4(4): 221-238.

Harvey, A.C., 1990. The Econometric Analysis of Time Series. 2nd Edn., Philip Allan, Hemel Hempstead.

Harvey, A.C., 1993. Time Series Models. 2nd Edn., MIT Press.

Haslett, J. and A.E. Raftery, 1989. Space-time modelling with long memory dependence: Assessing Ireland wind power resources. Appl. Stat., 38(1): 1-50.

Hosking, J., 1981. Fractional differencing. Biometrica, 68(1): 165-176.

Hsu, N.J., B.K. Ray and F.J. Breidt, 1998. Bayesian Estimation of Common Long Range Dependent Models. Prob. Theory Math. Stat. B. Grigelionis *et al*. (Eds.). VSP/TEV.

Hunt, K. and G.P. Nason, 2001. Wind speed modelling and short-term prediction using wavelets. Wind Eng., 25: 55-61.

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Eng., 116: 770-779.

Hurvich, C.M., R.S. Deo and J. Brodsky, 1998. The mean squared error of geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. J. Time Ser. Anal., 19: 19-46.

Jarque, C.M. and A.K. Bera, 1987. A test of normality of observations and regression residuals. Int. Stat. Rev., 55: 163-172.

Kwiatkowsi, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. J. Econometrics, 54: 159-178.

Ljung, G.M. and G.E.P. Box, 1979. On measure of lack of fit in time series models. Biometrika, 65: 297-303.

Montanari, A., R. Rosso and M. Taqqu, 1996. Some long-run properties of rainfall records in Italy. J. Geophys. Res. Atmosphere, 101(23): 431-438.

Said, S.E. and D. Dickey, 1984. Testing for unit roots in autoregressive moving average models with unknown order. Biometrika, 71: 599-607.

Whitcher, B., 1998. Assessing nonstationary time series using wavelets. Unpublished Ph.D. Thesis, University of Washington.