Forecasting solid waste generation in Juba Town, South Sudan using Artificial Neural Networks (ANNs) and Autoregressive Moving Averages (ARMA)

1David Lomeling* and 2Santino Wani Kenyi

1,2Department of Agricultural Sciences, College of Natural Resources and Environmental Studies (CNRES), University of Juba, P.O. Box 82, Juba, South Sudan

Prediction of solid waste generation is critical for any long term sustainable waste management, especially of a fast-growing municipality. Lack of, or inaccurate solid waste generation records poses unparalleled challenges in developing cohesive and workable waste management strategies for any concerned authorities, as this is influenced by several interlinked demographic, economic, and socio-cultural factors. The objective of this study was to compare two models in forecasting of MSW generation and how this would be built into an effective MSW management program. Two models, the Autoregressive Moving Average (ARMA 1,1) and the Artificial Neural Networks (ANNs) were tested for their ability to predict weekly waste generation of 14 households in Juba Town, Central Equatoria State (CES), South Sudan. Results showed that both the artificial intelligence model ANNs and the traditional ARMA model had good prediction performances; where for ANNs the RMSE, MAPE and $r^2$ were 0.080, 10.64%, 0.238 respectively, whereas for ARMA the RMSE, MAPE and $r^2$ were 0.102, 6.98% and 0.274 respectively. Both models showed no significant differences and could be therefore be used for Solid Waste (SW) forecasting. Based on the results, the weekly SW generated 52 weeks later (end of year) had reached 0.365 kg/capita indicating a 18.2% rise from 0.3 kg/capita at the start of the study. Under the current consumption rate, the weekly SW per capita in Juba Town is expected to reach 0.596 kg by 2020.

Keywords: Artificial Neural Networks, Autoregressive Moving Averages, Continuous Wavelet Transform, Waste Generation Forecasting,

INTRODUCTION

South Sudan witnessed rapid economic growth rates with Gross Domestic Product (GDP) estimates of over $16billion (World Bank Report 2008) immediately after the signing of the Comprehensive Peace Agreement, CPA in 2005. Huge oil revenues from crude oil sales attracted investments and rapid population growth especially in urban centers like Juba. Conversely, this economic and population growth put enormous strain on the local environment and on the availability of natural resources (Lomeling et al., 2016) in terms of increased demand for areas for settlement as well as depleting the forest resources as cheap energy source. With the rapid population increase in Juba, waste generation inadvertently also increased. This meant that more land and resources would have to be required for planned waste disposal site in the short to long term, if serious environmental pollution through indiscriminate waste disposal is to be abated.

*Corresponding author: David Lomeling, Department of Agricultural Sciences, College of Natural Resources and Environmental Studies (CNRES), University of Juba, P.O. Box 82, Juba, South Sudan. Email: dr.david_lomeling@gmx.net
Prediction or forecasting of solid waste generation in Juba Town has become indispensable and crucial, if available resources are to be effectively deployed in the sustainable management of SW. Time series offers an important area of stochastic forecasting in which past observations of a specific variable are analyzed to develop a model that can be used to make future projections. Over the past decades, much effort has been undertaken in the development and improvement of time series models that can be applied for forecasting like the Artificial Neural Networks (ANNs) and the Auto-Regressive Moving Averages (ARMA) as compared to classical and traditional methods like linear and multiple regressions or polynomial functions. During the last two decades, several stochastic models have been used to predict SW waste like the support vector machine, (Abbasi et al., 2013); artificial neural networks (Abdoli et al., 2011); (Antanasijevic et al., 2013); hybrid procedure (Xu et al., 2013); time series analysis, (Mwenda et al., 2014); multi-step chaotic model (Song and He, 2014); principal component analysis and gamma testing (Noori et al., 2010); grey fuzzy dynamic modeling (Chen and Chang, 2010); Fourier series (Darko et al., 2016); simulated annealing based hybrid forecast (Song et al., 2014). In general, most ANN prediction models have clearly outlined architecture with specific number of input variables at the input layer and corresponding number of expected outputs at the output layer. Depending on the problem to be modeled, several input variables may be chosen for a given number of anticipated outputs. The choice of any one single, all-purpose model under any prevailing conditions is therefore unrealistic. Structurally, such a multi-purpose model would require more complex algorithms capable of handling several calculations simultaneously for some desired number of input variables and then come up with optimal predictions.

The objective of this study was to compare the effectiveness of machine learning method (Artificial Neural Networks ANNs) and the stochastic linear model (Autoregressive Moving Average – ARMA) for medium to long-term weekly forecasting of solid waste generation in some households of Juba Town, South Sudan. The integration of Continuous Wavelet Transform (CWT) with both ANNs and ARMA models was primarily used for easy visualization and interpretation of input signals as well as frequencies in the time series. Using both models, a good estimate of the weekly SW per capita was to be made during the 2012-2020 forecasting period.

Forecasting Models

Artificial Neural Networks (ANN)

ANNs is basically a computational approach whose architecture is mostly composed of three layers: an input, hidden and output layers mimicking the way biological neurons receive, transfer and output signals. Each neural unit or perceptron is linked with many others and can either be enforcing once the summation function has surpassed some threshold value to be propagated or inhibitory in their effect once below the summation function value. The single neuron is illustrated by the McCulloch-Pitts Model (1943).

Mathematically, the output variable, \( y \) is the sum of the individual weighted variables and bias that influences the activation function \( S \):

\[
(x_1 w_1 + x_2 w_2 + \cdots x_n w_n) + b = y = S \sum_{j=1}^{n}(x_i w_j + b)
\]

Equation (1)

The underlying concept is to arrive at a function that minimizes the error (E) between the input (actual) and output (forecasted) variables thereby enhancing the accuracy in the forecasting or prediction.

\[
E_{\min}(x, b) = f[\sum_{j=1}^{n} x_i w_j b - y]^2
\]

Equation (2)

Usually the sums of each input signal \( (x_1, x_2, \ldots x_n) \) and intensity or weighted values \( (w_1, w_2, \ldots w_n) \) are passed on through a non-linear function known as an activation or transfer function that usually has a sigmoid shape, that is bounded, and differentiable as:

\[
S(x) = \frac{1}{(1+e^{-x})}
\]

Equation (3)

Many theoretical and experimental works have shown that a single hidden layer (with one or more several hidden nodes) is sufficient for ANN to approximate any complex nonlinear function (Dreiseitl and Ohno-Machado, 2002; Chattopadhyay and Bandyopadhyay 2007; Matias et al., 2013; Vishwakarma and Gupta, 2011; Aggarwal and Kumar 2015). A more plausible argument is that, the low number of nodes in the hidden layer directly linked to the input neuron have a low bias (b) and would tend to increase the input of the activation function. This in turn would enhance large changes in their weights and learn very quickly and so incur less errors as manifested by the high correlation coefficients for both training and test sets. In this study, a model based on a feedforward neural network with a single hidden layer was used. Hereby, the learning process in understanding hidden and strongly non-linear dependencies in the time series of the observed
and modeled data in the training and test were faster and the forecasting made easier. However, such forecasts can only be made for shorter prediction times and not for extensively longer future times as the error would tend to increase.

**Auto-Regressive Moving Average, ARMA model**

The first step in developing the ARMA model was determining the stationarity of the time series in which case the mean and variance are time invariant. The autocorrelation function (ACF) may signify stationarity of the time series, if it cuts off or decomposes quite rapidly towards zero. Conversely, if the ACF decomposes very slowly and gradually towards zero, this would indicate non-stationarity and would need to be transformed or differenced to obtain stationarity by stabilizing the variance of a time series. Differencing helps stabilize the mean of a time series by removing changes in the level of a time series, and so eliminating trend and seasonality.

On the other hand, a time series that shows seasonality as in Figure 2a can be exponentially smoothened by an exponentially weighted non-parametric value ($\alpha$) to “smoothen” the value $X_t$ to a new value $X^*$ recursively (Figure 2b):

$$\bar{X} = \alpha X_t + (1 - \alpha) X_{t-1} \quad \text{Equation (4)}$$

where $0 \leq \alpha \leq 1$

Our data showed that the best seasonal exponential smoothing was when $\alpha=0.2$ with $r^2=0.26$; $\alpha=0.5$ with $r^2=0.08$ and $\alpha =0.8$ with $r^2=0.04$. The value $\alpha = 0.2$ best approximated the mean value and regression constant of the time series and therefore gave a better trend analysis.

Owing to its simplicity, the exponential smoothing only “de-seasonalizes” a time series thereby assuming the nature of a linear regression equation between two variables. However, it’s predictive ability is inadequate in more complex stochastic and non-linear processes with more
than 3 or more input variables. Although some authors (Mwenda et al., 2014; Petridis et al., 2016; Rimaitye et al., 2011; Karpušenkaitė et al., 2016) have mentioned the theory behind single exponential smoothing, however, not much in terms of its practical applicability have yet been reported.

The ARMA \((p,q)\) model used herein is made up of an Autoregressive \(AR(p)\) and Moving Average \(MA(q)\) components. This means that, the forecast value of \((X)\) at time \((t)\) in a time series is a function of both linear combination of past \(X\)-innovations and a moving average of series \((\epsilon_{t})\), known as white noise process characterized by zero mean (\(\mu\)) and variance (\(\sigma\)).

\[
X_t = c + \epsilon_t + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}
\]

Equation (5)

With the values \(p\) and \(q\) identified from the ACF and PACF, the model parameters \((\alpha_i)\) and \((\beta_i)\) can then be estimated. Once we had confirmed the stationarity of the time series, the autocorrelation (ACF) and partial autocorrelation functions (PACF) were used to determine the correlation and model structure of the data.

MATERIALS AND METHOD

This study focused on solid waste generated by single persons in 14 households in Kator residential area of Kator Payam, Juba County of Central Equatoria State in South Sudan. Collected waste was placed into a container whose tare weight was initially determined using the hanging scale. The net weight of the solid waste was then determined. The data were collected weekly over a period of 31 weeks as from June 2010 till January 2011. Each household had on average 6 persons with monthly income of about 650 SDG (Sudanese Guine equivalent to 145$/month as of June 2010). About 40-60% of the collected waste was made up of predominantly degradable organic component consisting mainly of food residues and partly cartons and newspapers. The rest was made up of PET plastic water bottles. The waste was collected at the end of each week, weighed and the daily amounts per capita generated (kg/capita/week) was then calculated. This work attempted to predict the weekly waste generated in the remaining 21 weeks till July 2011 based on the previous data. For the training set, data from the first 20 weeks were used representing 90% of the actual data. The rest 10 weeks representing 10% of the actual data were then used as test set.

Data description and analysis

From the weekly solid waste data reports, the Excel-based Alyuda Forecaster XL software was used to make future projections in the time series. Its algorithm allows an easy data preprocessing of the neural networks. Additionally, the Continuous Wavelet Transfer (CWT) using the PAST3 software was used to illustrate through the spectral power the peaks or spikes of weekly solid waste disposal in the time series.

Model identification

The effect of model choice on both correlation functions is shown in Figure 3 (a) and (b). The spikes presented in the ACF and PACF showed a correlation in the data every 2 lag units. The model identification revealed that with the cut-off at lag 1, the autoregressive of order \(p\) and the moving average of order \(q\) was also 1 as the ACF was got below zero after first lag. For illustrative purposes, the ARMA (1,2) as opposed to ARMA (1,1) was also used to compare the parameter values and how these influenced the model choice. The ARMA (1,2) in Figure 4 (a) and (b) as compared to ARMA (1,1) showed dissimilar AC and PAC functions at lag 1 and was outside the 95% confidence limits. The ARMA (1,1) was then chosen and there was therefore no need for any differencing.

Training Set: From this, 32 data entries of the weekly solid waste disposed per household were trained through a process of finding values for the weights \((w)\) and biases \((b)\) whereby the error between the measured and predicted values was minimized. The back-propagation algorithm was used here. The accuracy of the resulting training process was then applied for making projections in neural network model. Figure 5 shows the accuracy of the training set between the observed and forecasted values with a high \(r^2=0.99\).

Testing Set: This data set consisted of entries of the last weekly solid waste disposed per household and was used to test whether, or not the accuracy derived from the training process would provide the most appropriate solution and hence confirm the predictive power of the network model. Similarly, the test set showed high correlation coefficient of \(r^2=0.99\) as shown in the training set (Table 2).

Table 2: Comparison of the learning ability (MSE) between the training and test set of a ANN.

<table>
<thead>
<tr>
<th></th>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of rows</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>Nr. of Good Forecast</td>
<td>30(94%)</td>
<td>12(100%)</td>
</tr>
<tr>
<td>Average MSE</td>
<td>8.91E-05</td>
<td>1.25E-05</td>
</tr>
<tr>
<td>(r^2)</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Validation Set: Generally, this data set is used to minimize overfitting in the model. As in our case, given the high correlation coefficients of both the training and test sets and the high predictive power of the neural network model, there was therefore no need for a validation set. Figure 6 shows simulation runs of both the training and test sets and confirmed the accuracy of the ANN in predicting future solid waste data.
Figure 3. Autocorrelation plot of weekly per capita solid waste disposed in Juba (red dotted lines are upper and lower 95% confidence limits). ACF plot of residuals of ARMA (1,1) with $\alpha_1 = -0.999$ and $\beta_1 = -0.800$ (a) and PACF plot of residuals of ARMA (1,1) with $\alpha_1 = -0.999$ and $\beta_1 = -0.800$ (b).

Figure 4: ACF and PACF plot of residuals of ARMA (1,2) with AR (1) at $\alpha_1 = -0.999$ and AM (2) at $\beta_1 = -0.526$ and $\beta_2 = -0.800$. 
Figure 5. Scatter plot of an ANN for both training and test set of a weekly per capita solid waste in Juba Town.

Figure 6. Simulation of the training and test data sets.

Figure 7. Error distribution of training and test sets.

**RESULTS**

**ANN Model**

Our simulation was based on a simple network architecture that returned the smallest MSE and therefore the best prediction accuracy. The MSE and \( r^2 \) recorded for both the training and test sets have already been...
Table 3. Parameters of ARMA (1,1) and (1,2) models of weekly SW generated in Juba Town, South Sudan

<table>
<thead>
<tr>
<th>ARMA</th>
<th>Parameters</th>
<th>Standard Error, SE</th>
<th>t-statistics</th>
<th>p-Value</th>
<th>AIC</th>
<th>LLF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA (1,1)</td>
<td>AR (1)</td>
<td>-0.999 (α₁)</td>
<td>0.073</td>
<td>7,643</td>
<td>7,66E-10</td>
<td>-96,56</td>
</tr>
<tr>
<td></td>
<td>AM (1)</td>
<td>-0.888 (β₁)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA (1,2)</td>
<td>AR (1)</td>
<td>-0.999 (α₁)</td>
<td>0.000</td>
<td>12,738</td>
<td>1,08E-16</td>
<td>-96,34</td>
</tr>
<tr>
<td></td>
<td>AM (1)</td>
<td>-0.526 (β₁)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AM (2)</td>
<td>-0.278 (β₂)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA (1,3)</td>
<td>AR (1)</td>
<td>-0.999 (α₁)</td>
<td>0.000</td>
<td>13,464</td>
<td>3,304E-17</td>
<td>-99,05</td>
</tr>
<tr>
<td></td>
<td>AM (1)</td>
<td>-0.632 (β₁)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AM (2)</td>
<td>-0.420 (β₂)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AM (3)</td>
<td>0.289 (β₃)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

presented in Table 2, from where we observed 1-1-1 (1 input layer, 1 hidden layer, and 1 output layer) gives an accurate prediction of the weekly solid waste output. Applying the rule-of-thumb method for estimating the number of neurons in the hidden layer reported by Karsoliya (2012), we can assume that the number of neurons in the hidden layer is approximately 1. (or 70-90%) of the input layer. Noticeably, the number of hidden neurons is equal to the number of input nodes whereby the larger number of neurons would ostensibly lead to “overfitting” whereas with a relatively smaller number of neurons would lead to “underfitting”. The good fit \( r^2 \) for both training and test sets shows that the ANN modeled the observed data quite accurately. There is generally no “golden rule” for the number of hidden layers that is applicable for all non-linear time series. Whereas during the training process other problems are best predicted with two hidden layers (Srinivasan et al., 1994; Zhang, 1994; Baron, 1994). Other studies (Wanas, et al., 1998) showed that the best performance of a neural network occurred when the number of hidden neurons was equal to log (N), where N is the number of training samples. Another study conducted by Mishra and Desai (2006) showed that the optimal number of hidden neurons is \((2n+1)\), where n is the number of input neurons.

ARMA Model

Although model identification as per Box-Jenkins methodology clearly showed an ARMA (1,1) autoregressive \( (p) \) and moving average \( (q) \), we experimented with different \( q \) values to see to what extent this influenced not only model estimation but also the diagnostic checking and consequently the forecast. Three different model MA values were varied while the AR was kept constant. i.e. ARMA (1,1); (1,2) and (1,3) respectively. The best model parameters were selected based on the model that gave the least Akaike Information Criterion (AIC) value and highest likelihood estimation here denoted as Logarithmic Likelihood Function (LLF) Table 3.

Judging by the values of AIC and LLF, it is evident that experimented values at ARMA (1,2) and (1,3) are close to the actual ARMA (1,1) values and suggested the adequacy of the ARMA (1,1) model.

The scatter plot in Figure 8 shows a positive trend with several points around the trend line. The relatively low \( r^2 \) values suggested that there was neither an under- or overestimation for both ANN and ARMA models with most points between the 0.3-0.4 kg/week for both the observed and ANN-ARMA models. Whereas the ANN model had a \( r^2 \)-value of 0.238 this was 0.274 for ARMA model, these \( r^2 \)-values for both models were not significantly different from each other. The comparatively lower \( r^2 \)-value of the ANN would suggest the inability of a linear function in describing an entirely non-linear time series data.

Performance evaluation

The performance of either of the models was determined by measuring the difference between the observed and predicted values in the time series. Best estimates were those error values that were closer to zero, indicating less differences between the measured and observed values.

Three accuracy measures were used as follows:

1. Mean Absolute Percentage Error (MAPE):

\[
\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{P_{\text{obs}} - P_{\text{pred}}}{P_{\text{obs}}} \right| \times 100
\]

Equation (6)

2. Root Mean Square Error (RMSE):
Forecasting weekly SW generation using ANNs and ARMA models in Juba Town, South Sudan

**Figure 8.** Comparing the ARMA (p, q) and ANN for the observed and forecasted weekly SW per capita

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum (P_{\text{obs}} - P_{\text{pred}})^2}
\]

**Equation (7)**

Weekly SW generation for the first 32 weeks which included both the training and test sets were later combined with the lead time 20 weeks and the errors between observed and predicted assessed as in Table 3. Based on the MAPE, RMSE performance comparison between both models, there was however no significant difference between both models. The slightly higher MAPE value for the ANN model could be due to inherent drawback of the MAPE in overestimating error value especially when the difference between observed and predicted values is zero. (\(P_{\text{obs}} = 0\)).

**Table 4.** Comparison of the model performances in terms of MAPE and RMSE of both ANNs and ARMA models.

<table>
<thead>
<tr>
<th></th>
<th>ANN</th>
<th>ARMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE (%)</td>
<td>10.60</td>
<td>6.98</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.080</td>
<td>0.102</td>
</tr>
</tbody>
</table>

Ideally, the ARMA model would have shown poor performances in both MAPE and RMSE due to its inability to model non-linear variables as the ANN model. In complex and non-linear data in the time series, the ARMA model would inevitably lose predictive accuracy as it is unable to adequately capture the errors between the observed and forecasted values in the time series and so the prediction errors would increase (see Figure HH). Conversely, the ANN model is capable of identifying nonlinearity in the time series by having an additional computation or hidden layer that allows for better curve-fitting with minimal errors between the forecasted and observed data.

**Diagnostic checking**

Based on the autocorrelation plot, the Ljung-Box (1978) test attempts to establish the overall data randomness within a time series at some chosen or predetermined lags. Basically, the null hypothesis (\(H_0\)) assumes that the data are random or independently distributed whereas this is the contrary for the alternative hypothesis (\(H_a\)) that assumes the non-randomness nature of the data. We then performed the Ljung-Box statistic test as:

\[
Q(LB) = \sum_{k=1}^{N_k} \frac{\hat{p}_k^2}{n-k}
\]

**Equation (8)**

Where \(n\)=sample size, \(\hat{p}_k^2\)=sample correlation at lag \(k\), \(N_k\)=number of lags being tested. Choosing the \(\hat{p}_k^2\) at lag 7 and testing at the \(p=0.05\) confidence level, the Q(LB) value at 0.645 was less than the (chi-square) at 2.013 thereby reaffirming the \(H_0\) and adequacy of the ARMA (1,1) model. As aforementioned the Q(LB) for randomness is reinforced by the autocorrelation functions as in **Figures 3** and **4**. Hereby, all the autocorrelation at the subsequent lags fall within the 95% confidence limits, other than that at lag 0.

**Model verification**

Model verification dealt with ascertaining whether the residuals of the ARMA model as expressed by the ACF and PACF had any discernible and systematic patterns.
with respect to the lags. Our study showed that, none of these correlations was significantly different from zero at the 95% confidence limit indicating the goodness of fit and appropriateness of the ARMA model.

**Forecasting**

The two forecasting models ARMA and ANNs presented in this paper (Figures 9 and 10) allowed us to predict the weekly generated SW with mean value of about 0.35 kg/household and upper and lower limits of about 0.63 and 0.16 kg/household respectively. Towards the 51st week, the generated SW was well below the 0.4 kg level and would under constant economic and political conditions remain below the 0.5 kg level till 2020. This forecasting is useful in mobilizing and optimizing available financial resources and personnel needed for effective SW management.

**2.2 Continuous Wavelet Transform (CWT)**

One of the main goals of a CWT is to enable the easy visualization and interpretation of input signals and frequencies as a function of time. The CWT decomposes a continuous time function of a time series into the components called wavelets each with a different localized frequency. Although CWT are usually applied for non-stationary signals, we tried to apply both the Morlet wavelet transform and the Derivative of Gaussian (DOG) to account for the high instantaneous amplitude or signal outbursts in the time series as well as in the recognition of inherent frequency patterns. The Morlet wavelet transform (Goupillaud, et al., 1984) is given as:

\[
y = 0.0011x + 0.3077
\]

\[
r^2 = 0.04
\]
\[ \varphi_{w_0}(t) = Ke^{i w_0 t} e^{-\frac{t^2}{2}} \]  
Equation (9)

Where \( w_0 \) is the non-dimensional frequency and the vertical scale corresponds to the length of wavelet, \( i.e., \) the number of time steps used for the CWT. The cone of influence (COI) is the region where the wavelet power spectra are limited due to the influence of the end points of finite length signals also known as the \( e\)-folding time. Here, the signal discontinuity drops by a factor of \( (e^2) \) and ensures that edge effects are negligible beyond this point. The DOG with \( m= \) derivative was set at 6 and expressed as:

\[ \text{DOG} = \frac{(-1)^{m+1} a^m}{\sqrt{(m+2)} d_m} (e^{-\frac{t^2}{2}}) \]  
Equation (10)

Values of \( m=2 \) or 4 using DOG basis functions did not describe the spectral decomposition in the time series adequately. The choice of the wavelet used for time-frequency decomposition in a time series is critical. The use of the Morlet wavelet for example, showed that the frequency resolution was "lumped" together and was localized within a 95\% confidence limit. On the contrary, using Derivative of Gaussian (DOG) wavelet with \( m=6 \), (Figure 11 a and b), the result was good time localization with strong frequencies. The Morlet and Derivative of Gaussian (DOG) wavelets are plotted below.

The choice for the type of wavelets in interpreting the frequency and intensity of data entries \((x_n)\) in a time series are shown in Figures 11 and 12. We used the forecasted data from both models as input data to generate the respective wavelets. For both models, (Figure 11a and b), the dominant power spectrum was characterized by smaller spikes between log scale 1.2 and 3.2 from week 15 to 30. This time coincided with the highest weekly waste generation around Christmas season where expectedly, more households had much disposable incomes enabling an increased consumption and so increased solid waste generation. However, the ARMA as opposed to the ANNs model showed certain areas outside the COI, clearly an indication of model overestimation by the former. In both cases, using the Morlet basis function (Figure 11c and d) showed poor spectral decomposition as a function of time than the DOG \( m=6 \).

As aforementioned with the Ljung-Box randomness test in a time series with the ARMA model, the resulting power spectrum using the CWT can as well exhibit coherent and significant structures. A test of significance can be used to distinguish between significant and random structures (Mohr, 2003) in which case values in a power spectrum may be considered as statistically significant at the 95\% level and therefore not random.

For the null hypothesis \((H_0)\), we assumed that the time series had a mean power spectrum. Spikes or peaks in the wavelet power spectrum above this background spectrum were shown as black contoured spectrum "significant at the 5\% level" or equivalent to "the 95\% confidence level". Therefore, if the peak in the power spectrum of the Morlet and DOG wavelets are significantly larger than the background spectrum, it is then assumed to be a true feature. There was better interpretation of the spectral decomposition for the observed data as well as for both models when the DOG \( m=6 \) as opposed to Morlet basis function was applied.

**SUMMARY AND CONCLUSION**

In this study on time series models, we analyzed and compared the Artificial Neural Networks (ANNs) and the Autoregressive Moving Averages (ARMA) in forecasting the weekly amounts of solid waste generated by single persons in fourteen households of Kator residential area of Juba town. For the ANNs model, the input training data used were the average weekly amounts of solid waste collected from during the month of June 2011. The test data of July 2011 were then used to validate the ANNs model. The result showed that ARMA (1,1) slightly outperformed the ANNs model in terms of MAPE but not in terms of the RMSE. However, considering both performance indicators, there was no significant differences between both models and so either model could be used to forecast the amount of the solid waste generation for the next weeks and years. Using both models is, however, for comparative reasons imperative in order qualify and quantify the extent of deviation of the estimated values from the observed mean. The projected values showed that by 2020, the weekly generation according to both models will have reached about 0.596 kg/capita with a 95\% confidence interval lying between 0.2-0.6 kg/capita. Such projections may be used by Juba municipal or town council in the proper planning and management of solid waste.

**ACKNOWLEDGEMENTS**

My special gratitude goes to Mr. Santino Wani Kenyi for data collection as part of his BSc dissertation. We the authors also wish to extend our thanks to the households in Kator Payam for allowing the conduction of the interviews.
Figure 11. (a) Weekly solid waste output signals for the ARMA model when using the DOG $m=6$ (b) for the ANNs models. Whereas when using the Morlet basis function for ARMA (d) for ANNs model. Wavelet power spectrum showing cone of influence (COI) at the 95% confidence interval with areas of intense peaks and signals in red contoured in (black) while those with poor signals in (blue).
Forecasting weekly SW generation using ANNs and ARMA models in Juba Town, South Sudan

Figure 12. (a) Weekly solid waste output signals using the Morlet basis function for the observed data and (b) using the DOG $m=6$ basis function. Wavelet power spectrum showing cone of influence (COI) at the 95% confidence interval with areas of intense peaks and signals in red contoured in (black) while those with poor signals in (blue).

REFERENCES


Forecasting weekly SW generation using ANNs and ARMA models in Juba Town, South Sudan


Tanzania.


Accepted 23 August 2017

Citation: Lomeling D and Kenyi SW (2017) Forecasting solid waste generation in Juba Town, South Sudan using Artificial Neural Networks (ANNs) and Autoregressive Moving Averages (ARMA). Journal of Environment and Waste Management 4(2): 211-223.

Copyright: © 2017 Lomeling and Kenyi. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.