The objective of this paper is to present the results obtained from the methodology applied to the analysis of time series of water table; and to examine the possibility of using linear models of different complexity to generate daily forecasts of this variable.

The water table series used corresponds to the locality of Ezeiza, in the Matanza River watershed, Province of Buenos Aires, Argentina (34º48’S, 58º32’O), in a subwet-wet region with water surpluses in excess of 150 mm per year.

The proposed methodology consists in applying linear parametric models with different exogenous variables to forecast daily water tables. Transfer function models make it possible to assess the dynamic response of the water table to rainfall and useful water. The variable accounts for recharge and discharge is useful water, after a calculation of potential evapotranspiration values using the Priestley-Taylor equation estimated from a daily water balance.

The methodology proposed by Box and Jenkins includes the following stages: identification of the time structure by means of the cross-correlation function; estimation of parameters with the maximum likelihood method; verification by means of residuals analysis and fitting of extra coefficients (overfitting). Finally, daily forecast errors of the selected models are analyzed. The statistics used to evaluate the estimated forecast errors are: mean and maximum error, root of mean square error (RECM), absolute error (EA), and relative error (RE).

Transfer function models are used to forecast the daily water table variable, particularly the rainfall-water table model, where the error statistics are comparatively less than those obtained with the univariate ARMA model (1,1,0).

Other results show that the errors statistics of forecast water table are improved when transfer function model and National Oceanic Atmospheric Administration rainfall forecasts one day in advance are used.
1. Introduction

The objective of this paper is to present the results obtained from the methodology applied to the analysis of time series of water table; and to examine the possibility of using linear models of different complexity to generate daily forecasts of this variable. The method used is the one proposed by Box and Jenkins (1976). Calibration, verification and forecast of transfer function models are presented. They are identified from the cross-correlation function of the rainfall-water table and useful water-water table variables.

2. Background information

The water table series used corresponds to the locality of Ezeiza, in the Matanza River watershed, Province of Buenos Aires, Argentina (34º48'S, 58º32'O), in a subwet-wet region with water surpluses in excess of 150 mm per year. The groundwater source is a complex aquifer made up of: Puelche Section with homonymous sands; Pampeano section with limes with sandy intercalations; and the water table with fine textured sand.

3. Methodology

The proposed methodology consists in applying linear parametric models with different exogenous variables to forecast daily water tables. Transfer function models make it possible to assess the dynamic response of the water table to rainfall and useful water. The methodology proposed by Box and Jenkins includes the following stages: identification of the time structure by means of the cross-correlation function; estimation of parameters with the maximum likelihood method; verification by means of residuals analysis and fitting of extra coefficients (overfitting). Finally, daily forecast errors of the selected models are analyzed.

The time series analysis method is widely used in hydrology for the mathematical representation of different types of processes. Viswanathan (1984) proposes rainfall-water table transformation models through dynamic regressions; and Calderón Loaiza and Bermudes (1994) use the time series analysis to estimate the same variable.

3.1 Input and output variables

Rainfall

For the set of data used for identification of the several models, accumulated rainfall is 855.3 mm/year. This value corresponds to a year with less rainfall than the annual mean (950 mm/year) and an annual index (annual rainfall/mean annual rainfall) less than the unit. Yearly rainfall distribution is shown in Figure 1.

According to this analysis, the rainiest season is autumn and the driest is winter. On a daily scale, rainfall has no temporal structure and behaves like a noise.

Useful water

The hydrologic cycle is a system with input/output components, and the occurrence of meteorological phenomena produces modifications in the runoff and storage. Water table variations occur when there is a difference between recharge and discharge volumes. Recharge is produced by rainfall infiltration. The variable accounts for recharge and discharge is useful water, after a calculation of potential evapotranspiration values using the Priestley-Taylor equation (Chow et al., 1988) obtained from a daily water balance:
where: $A_t$ useful water on day $t$
$P_t$ rainfall on day $t$
$E_t$ evapotranspiration on day $t$.

According to Thornthwaite-Mather (1957), soil moisture–storage capacity is the maximum amount of water the soil can hold in the rooting zone at field capacity. In this study, it is assumed to be 100 mm since this is a case of fine sandy soil with pastureland.

A water balance is developed with daily rainfall values and evapotranspiration estimates from which the useful water variable is obtained (equation 1). Figure 2 shows the behavior of the series estimated for the useful water variable. It is a nonstationary series according to its simple autocorrelation function.

### Water table

The behavior of the water table has fluctuations associated to the meteorological variables involved in the hydrologic cycle.

Evolution of the water table through time may be represented with an autoregressive-integrated model (Seoane and Arensburg, 1996) of the series in the form of:

$$N_{tf} = N_{tf-1} + \phi_{t}N_{tf-2} - \phi_{s}N_{tf-3} + z_{t}$$  

(2)

where: $N_{tf}$ water table on day $t$
$\phi_{t}$ autoregressive parameter
$z_{t}$ residuals of the autoregressive integrated moving average model, ARIMA (1,1,0).

### 3.2. Transfer function model

Transfer function models consist of a linear relationship between two stationary time series. Given stationary $X_t$ and $W_t$, the model is expressed as:

$$W_{f} = \nu(B)X_{f} + \theta_{f}$$  

(3)

where: $\nu(B)$ backward shift operator
$\nu(B) = \alpha_{0}(B)\delta(B)$
\[ t_0 = t_1 = \ldots = t_{p* - 1} = 0 \]
\[ \alpha_t(B) = \alpha_0 - \alpha_1 B - \ldots - \alpha_{p*} B^{p*} \]
\[ \phi_t(B) = 1 - \phi_1 B - \ldots - \phi_{q*} B^{q*} \]
\[ X_t = \phi_t^{-1}(B) \alpha_t, \text{ ARMA (p*, q*) process} \]
\[ \eta_t = \phi_t^{-1}(B) \alpha_t, \text{ ARIMA (p**, d**, q**)} \text{ process independent from } X_t \]
\[ d** \text{ number of differencing operation that is required to obtain the stationary series} \]
\[ \alpha_t \text{ y } \alpha_t \text{ noise (random independent variable identically distributed with expected value zero).} \]

### 3.2.1. Cross-correlation function

Under the hypothesis that two variables are stationary, with mean and constant variance, the cross-correlation function for time-delay k is estimated as:

\[
r(k) = \frac{E[(X_{t+k} - E(X))(W_t - E(W))]}{s_X s_W} \tag{4} \]

where:  
- X: input variable  
- W: output variable  
- E: expected value  
- \( s_X \) and \( s_W \): standard deviations from each variable.

The \( r(k) \) function is significantly different from zero for a certain k value when variable X exerts a linear influence on W.

On the basis of the input/output variables used, four options are analyzed:

- **Cross-correlation function between rainfall and differenced water table (d =1).** The water table variable is not stationary because its simple autocorrelation function shows a non-exponential decrease. This series is differenced and the results are shown in Figure 4.

- **Cross-correlation function between rainfall and residuals of the ARIMA (1,1,0) model estimated for the water table.** This option is used when one or both variables have an autocorrelation structure (Granger, 1977).

- **Cross-correlation function between differenced useful water (d =1) and differenced water table (d =1).**

- **Cross-correlation function between differenced useful water (d =1) and residuals of the ARIMA (1,1,0) model estimated for the water table series.**

![Figure 4. Cross-correlation function between rainfall–differenced water table (d=1).](image)

In all cases, the coefficients are significantly different from zero for k = 0, 1 (Table 1). From the estimated cross-correlation functions, the orders for the corresponding transfer function models (b,r,s) are identified. The transfer function models (0,0,1) are also identified.

### 3.2.2. Transfer function estimate

**Preliminary estimate**

The preliminary estimate of parameters is made by calculating the \( v_t \) shown in equation 3 through:

\[
v_t = r(k)^* s_W / s_X \tag{5} \]
where: \( u_i = 0 \)
\[
\begin{align*}
u_t &= \delta_1 v_{t-1} + \ldots + \delta_{b} v_{t-b} + \eta_t, \quad i = b + a < b + s + 1, \quad a = 1, \ldots, s \\
u_t &= \delta_1 v_{t-1} + \ldots + \delta_{b} v_{t-b} \\
& > b + s
\end{align*}
\]

In the transfer function models \((0,0,1)\) equation 5 is as follows:
\[
u(B) = \omega_0 + \omega_1(B) = \omega_0 = r(0) * s \cdot w / s \cdot x \\
\omega_1 = r(1) * s \cdot w / s \cdot x
\]

In the transfer function models \((0,0,1)\) equation 5 is as follows:
\[
u(B) = \omega_0 + \omega_1(B) = \omega_0 = r(0) * s \cdot w / s \cdot x \\
\omega_1 = r(1) * s \cdot w / s \cdot x
\]

Table 1 shows the input and output variables and the cross-correlation coefficients that are used to obtain the initial estimate with equation 5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables</th>
<th>Standard deviation</th>
<th>Significant coefficients of the cross correlation function</th>
<th>( \omega_0 ) (m / mm)</th>
<th>( \omega_1 ) (m / mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Input</td>
<td>Rainfall (X)</td>
<td>3.910</td>
<td>0.5970</td>
<td>0.4620</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>Differenced water table (W)</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Input</td>
<td>Rainfall (X)</td>
<td>9.310</td>
<td>0.4556</td>
<td>0.4456</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>ARIMA (1,1,0) model residuals of water table (z)</td>
<td>0.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Input</td>
<td>Differenced useful water (Y)</td>
<td>7.090</td>
<td>0.3782</td>
<td>0.5313</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>Differenced water table (W)</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Input</td>
<td>Differenced useful water (Y)</td>
<td>7.090</td>
<td>0.4048</td>
<td>0.4314</td>
</tr>
<tr>
<td></td>
<td>Output</td>
<td>ARIMA (1,1,0) model residuals of water table (z)</td>
<td>0.047</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Preliminary estimate of parameters.

Maximum likelihood estimates

At this stage, the transfer function models are estimated using the maximum likelihood method, and they are checked for consistency with data according to the following hypotheses: \( H_1 \): \( a_t \) is noise; \( H_2 \): \( X_t \) and \( a_t \) are independent. For \( H_1 \) it should be ascertained whether the coefficients of the simple (fast) and partial (flap) autocorrelation function of the residuals series are not significantly different from zero. For \( H_2 \) it should be ascertained whether the coefficients of the cross-correlation function between \( X_t \) and \( a_t \) are not significant. If both hypotheses are verified, the model is correct. Otherwise, the time structure of the residuals should be identified and the parameters should be estimated anew for maximum likelihood.

Model I (Table 1), rainfall-differenced water table is expressed as:
\[
W_t = \omega_0 X_t + \omega_1 X_{t-1} + \eta_t
\]

where: \( W_t = Nf_t - Nf_{t-1} \), is differenced water table \((d = 1)\) on day \( t \).

If \( \eta_t \) does not verify the hypothesis on residuals, a model is identified based on the simple partial autocorrelation function of the series:
\[
\eta_t = W_t - \omega_0 X_t - \omega_1 X_{t-1}
\]

The model parameters are estimated using maximum likelihood, and the hypotheses are tested again. Table 2 shows the results obtained with maximum likelihood using rainfall as the input variable and the differenced water table as the output variable. The difference between the two models is that the second verifies the hypotheses on residuals.
I

\[ W_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \eta_t \]

\( \alpha_0 = 0.0024 \quad s_{\alpha_0} = 0.0002 \)

No time structure

\( (\eta_t = \alpha_t) \)

0.0012

\[ W_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \beta_1 \text{dW}_{t-1} - \beta_2 \text{dW}_{t-2} \]

\( \alpha_0 = 0.0024 \quad s_{\alpha_0} = 0.0002 \)

ARIMA (0,1,1) with parameters:

\( \hat{\beta}_1 = -0.989 \quad s_{\beta_1} = 0.0076 \)

0.0012

Table 2: Comparison of models estimated on the basis of the maximum likelihood with the rainfall and differenced water table variables.

The same procedure is followed with the other options presented in the preliminary estimate (Table 1); the results obtained are shown in Table 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Model for ( \eta_t )</th>
<th>( S_{\text{res}}^2 )</th>
</tr>
</thead>
</table>
| II | \( z_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \beta_1 \text{dW}_{t-1} - \beta_2 \text{dW}_{t-2} + \eta_t \) | ARIMA (0,1,2) with parameters:

\( \hat{\beta}_1 = -0.218 \quad s_{\beta_1} = 0.0504 \)

\( \hat{\beta}_2 = -0.292 \quad s_{\beta_2} = 0.0504 \) | 0.0012 |

| III | \( W_t = \alpha_0 Y_t + \alpha_1 Y_{t-1} + \eta_t \) | No time structure

\( (\eta_t = \alpha_t) \) | 0.0015 |

| IV | \( z_t = \alpha_0 Y_t + \alpha_1 Y_{t-1} + \eta_t \) | No time structure

\( (\eta_t = \alpha_t) \) | 0.0015 |

| ARIMA (1,1,0) | \( N_{t+1} = \phi N_{t} + \theta N_{t-1} + \text{dW}_{t-1} + \eta_t \) | \( \phi = 0.3771 \quad s_{\phi} = 0.0438 \) | 0.0023 |

Table 3: Comparison of the transfer function and ARIMA models estimated based on maximum likelihood.

From Tables 2 and 3 it may be inferred that the simpler rainfall – differenced water table models have the smaller residuals variance estimators (\( S_{\text{res}}^2 \)). A comparison of these models and ARIMA (1,1,0) shows that when an exogenous variable is introduced, there is a better model fit.

3.3. Forecast

A forecast consists in estimating the water table on a given day, \( N_{t+1}^{(1)} \), and in contrasting it with observations, \( N_{t+1} \), for a thirty-day period immediately following the period used for identification and estimation. In order to estimate \( N_{t+1}^{(1)} \) the models with the best fit were selected; i.e., smaller \( S_{\text{res}}^2 \) and smaller number of parameters in Tables 2 and 3.

The statistics used to evaluate the estimated forecast errors are: mean and maximum error, root of mean square error (RECM), absolute error (EA), and relative error (RE).
The rainfall forecast estimator, $X_{t+1}$, is obtained from the analysis performed by the National Oceanic Atmospheric Administration (NOAA) of the corresponding coordinates one day in advance ($X_{t+1,\text{NOAA}}$) or from the season’s daily mean ($X_{t}^\text{NOAA}$). The daily rainfall predictive error determines an EA of 0.991 and 1.164 respectively. Therefore, the two models shown in Table 2 use $X_{t+1,\text{NOAA}}$.

Table 4 shows the results obtained using the historical rainfall series with forecast ($X_{t+1}=X_{t}$) and a “weak” forecast ($X_{t+1}=0$) to compare the statistics as a function of the rainfall model selected.

<table>
<thead>
<tr>
<th>Forecast model of the water table (one day)</th>
<th>ARIMA (1,1,0)</th>
<th>Transfer function model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{t+1} = M_{t} + \alpha X_{t+1} + \alpha M_{t}$</td>
<td>$X_{t+1} = X_{t}^\text{NOAA}$</td>
<td>$X_{t+1} = 0$</td>
</tr>
<tr>
<td>Error $= (N_{t+1} - \bar{N}_{t+1})$</td>
<td>0.2081</td>
<td>0.0992</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.0176</td>
<td>0.0029</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0544</td>
<td>0.0366</td>
</tr>
<tr>
<td>RECM $= \left( \sum \frac{(N_{t+1} - \bar{N}_{t+1})^2}{n} \right)^{1/2}$</td>
<td>0.0052</td>
<td>0.0046</td>
</tr>
<tr>
<td>EA $= \left[ \sum \frac{(N_{t+1} - \bar{N}_{t+1})}{n} \right]^2$</td>
<td>0.0032</td>
<td>-0.0005</td>
</tr>
</tbody>
</table>

Table 4: Statistics on water table forecast errors.

4. Conclusions

The analysis of the results obtained makes it possible to assess the response of the water table to some variables of the hydrologic cycle, especially rainfall.

Transfer function models can be used to forecast the daily water table variable, particularly the rainfall-water table model, where the error statistics are comparatively less than those obtained with the univariate ARMA model (1,1,0); and for simulation purposes when historical rainfall records are longer than the rainfall-water table simultaneous series.

Transfer function models of useful water-water table models may be better fitted with the observations if more complex water balance methods are developed.

The results show that the errors statistics of forecast water table are improved when transfer function model and National Oceanic Atmospheric Administration rainfall forecasts one day in advance are used.

Bibliography


