A STATISTICAL CASCADE MODEL TO GROUNDWATER MANAGEMENT

C. Rampolla
carla.rampolla@gmail.com

Index

ABSTRACT

Literature shows that the water balance and water quality are characterized by changes over time whose origin must be carefully evaluated, in order to determine which values should be used as reference. In this paper a statistical method (called cascade model) is studied to analyze the variations in water table, in withdrawals and in concentrations and to formulate a fairly accurate forecast of their values over time. In fact their values are such as to show a clear evolutionary trend that easily allows to make reasonable predictions about their progress in short periods.
Particularly a time series of monthly municipal water use Q(t) is studied by a "cascade model" of four transformations in order to obtain a final random error series.

RIASSUNTO

Il monitoraggio dei fenomeni che determinano il bilancio idrico e la qualità delle acque evidenza che essi sono caratterizzati, nel tempo, da variazioni la cui consistenza e natura deve essere attentamente valutata ai fini di determinare quali siano i valori da prendere come riferimento. Le variazioni di piezometria, dei prelievi e delle concentrazioni sono spesso tali da mostrare un chiaro trend evolutivo che permette di formulare ragionevoli previsioni sul loro andamento, quanto meno nel breve periodo. In questo lavoro è descritto un metodo statistico (cascade model) che consente una razionale analisi della natura dei fenomeni osservati e un previsione dei loro valori nel tempo. In particolare l'uso mensile comunale di acqua Q(t) è stato studiato attraverso questo metodo attraverso quattro trasformazioni che hanno portato ad ottenere un *random error* finale.

1. FOREWORD

The spread of the urban areas involves some significant environmental problems, that require a careful planning of the containment of the most dangerous consequences.
This paper deals with the applicability in a medium-large city of a method to evaluate the actual water demand, knowledge of which is essential for the management of water resources and for the prediction of groundwater levels.
To calculate the water demand Q, Maidment & Parzen (1984) and Mays & Yeou-Koung Tung (1992) proposed the "cascade model"; the analysis of this approach lead to some consideration on the suitability of this statistical method for the solution of the problems of groundwater management in urban areas. Here a methodology for time series analysis of the amount of water used in a city Q(t) is described. This amount of municipal water Q(t) can be predicted with a statistical model developed from past data.
The variations of water withdrawals over time is due to (1) climatic factors such as precipitation, evaporation and the temperature, to (2) socio-economic factors, such as the urban population, the price of water and household income and to (3) other urban activity such as restriction in water use. Water use Q(t) responds to these factors on different time scales. The climatic factors such as air temperature and evaporation vary sinusoidally over the year and the water use responds having a seasonal cycle. The socio-economic factors such as

population and household income vary slowly over the years and their effects on the use of water are adequately measured by annual average values.

2. THEORY

The cascade model used in this paper may be summarized as in Figure 1.

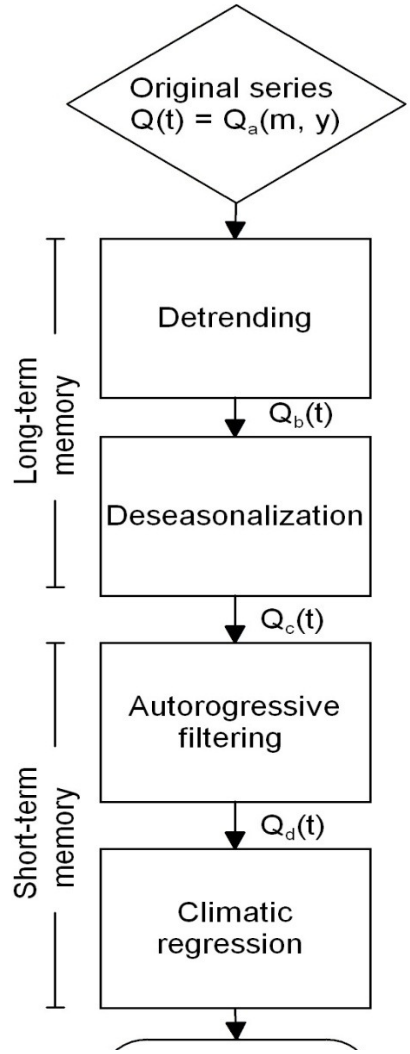

Fig. 1 – Cascade model for the monthly municipal water use Q(t)

Mean monthly municipal water use Q(t) is expressed as the sum of a long-memory component $Q_L(t)$ (or deterministic component) and a short-memory component $Q_S(t)$ (or stochastic).

So the monthly municipal water use Q(t) can be expressed as:

$$Q(t) = Q_L(t) + Q_S(t) \qquad t = 1, 2 \ldots.. T$$

Or

$$Qa(m,y) = Q_L(m,y) + Q_S(m)$$

Where m=1,….,12 is the monthly index within each year, y= 1, …., Y is the index of years and Y is the total number of years.

With:

Q(t) = Qa(m,y)

t = 12(y-1)+m

- The long-memory component QL(t) comprises trend and seasonality.
  The long-memory or deterministic components are described by functions that, once their parameters have been estimated, are considered to operate independently of the values of water use observed in any particular year.
- The short-memory component comprises autocorrelation, climatic correlation, and random error (Fig. 1). The short memory or stochastic components are not completely predictable, they depend upon the current and previous observations of water use and climatic variables.

In this model the different components (trend, seasonality, autocorrelation and climatic correlation) are identified and their effects are eliminated from the water use series; after passing through these four filters, only a random error series remains.

The four steps are now described and the examples refer to a city of large medium-size (about 900,000 inhabitants).

2.1. DETRENDING

The first step is to recognize the *over-year water use trend $Q_a(y)$* in the historical time and then remove it. The annual mean of monthly water use in year y can be estimated as

$$\mathrm{Qa(y)} = \frac{1}{12}\sum_{m=1}^{12} \mathrm{Qa(m,y)} \qquad y = 1, \ldots, Y$$

Instead the socioeconomic factors, like water price, population, number of water connections or household income may be identified with a multiple regression:

$$\mathrm{Qa(y)} = \varphi_0 + \varphi_1 Z_1(y) + \varphi_2 Z_2(y) + \varphi_3 Z_3(y) + \varphi_4 Z_4(y)$$

Where:

$Z_1(y)$ is real average effective buying income per household,

$Z_2(y)$ is real marginal water price

$Z_3(y)$ is number of water connections,

$Z_4(y)$ is population

Income and water price are the two main economic factors influencing personal water consumption, but their inclusion here is simply to account for their influence on the trend in water use. In application, neither of these variables was found significant. Population and number of connections are alternative measures of the number of users. Although population and connections are highly correlated, in application only one or the other of these variables was found significant by the stepwise regression. From the historical monthly water use Q(t) = Q_a(m,y), we remove the just identified trend $Q_a(y)$. Detrended time series Q_b(m,y) are:

$$Q_b(m,y) = Q_a(\mathrm{m,y}) - Q_a(y)$$

Exemple:

Table 1 - Monthly water use time series Q_a(m,y)

Year	Month	Q_a (m,y)	Year	Month	Q_a (m,y)	Year	Month	Q_a (m,y)	Year	Month	Q_a (m,y)	Year	Month	Q_a (m,y)
2000	1	7,3	2001	1	4,3	2002	1	4,5	2003	1	4,3	2004	1	5,1
2000	2	5,8	2001	2	3,8	2002	2	3,9	2003	2	4,3	2004	2	4,9
2000	3	4,7	2001	3	4,9	2002	3	4,4	2003	3	5,2	2004	3	5,5
2000	4	3,9	2001	4	5,6	2002	4	4,6	2003	4	5,5	2004	4	6,3
2000	5	5	2001	5	5,4	2002	5	5	2003	5	5,3	2004	5	6,4
2000	6	6,6	2001	6	7,2	2002	6	6,6	2003	6	9,3	2004	6	5,9
2000	7	7,9	2001	7	10,3	2002	7	6,6	2003	7	12	2004	7	8,3
2000	8	7,8	2001	8	7,9	2002	8	7,1	2003	8	9,4	2004	8	7,6
2000	9	6,4	2001	9	5,2	2002	9	6,4	2003	9	6,8	2004	9	5,9
2000	10	3,7	2001	10	6,3	2002	10	6,9	2003	10	6	2004	10	6,8
2000	11	4,5	2001	11	4,5	2002	11	4,7	2003	11	5,4	2004	11	6
2000	12	4,3	2001	12	4,3	2002	12	4,4	2003	12	4,9	2004	12	6,2
2000	**Tot**	**67,9**	**2001**	**Tot**	**69.71**	**2002**	**Tot**	**65.1**	**2003**	**Tot**	**78.4**	**2004**	**Tot**	**74,9**

$$\mathrm{Qa(y)} = \varphi_0 + \varphi_1 Z_1(y) + \varphi_2 Z_2(y) + \varphi_3 Z_3(y)$$

y = 2000, 2001,2002,2003

where:

$Qa(y)$ = monthly average water use in a year y

$Z_1(y)$ = population in year y

$Z_2(y)$ = water price(in € or \$) in year y

$Z_3(y)$ = average income (in € or \$)in year y

φ_0 = unknow parametres

$$\hat{\varphi} = (X^T X)^{-1} X^T Q_a$$

Where Q_a is the column vector of monthly average water use and X is the matrix containing observations of independent variables, namely :

$$X = \begin{array}{c} \quad\ \ Z_1 \quad\ Z_2 \quad\ Z_3 \\ \begin{bmatrix} 1.0 & 31.42 & 0.87 & 10.44 \\ 1.0 & 32.43 & 0.81 & 10.49 \\ 1.0 & 33.47 & 1.10 & 10.68 \\ 1.0 & 34.54 & 1.05 & 10.83 \\ 1.0 & 35.44 & 0.96 & 11.06 \end{bmatrix} \end{array} \quad Q_a = \begin{bmatrix} 67.90 \\ 69.71 \\ 65.10 \\ 78.40 \\ 74.90 \end{bmatrix}$$

$$\hat{\varphi} = \begin{bmatrix} -1.94 \\ 0.23 \\ 0.03 \\ 0.02 \end{bmatrix}$$

So the monthly water use for 2000-2004 is

$$Q_a(y) = X_{00-04}\,\hat{\varphi} = \begin{bmatrix} 1.0 & 31.42 & 0.87 & 10.44 \\ 1.0 & 32.43 & 0.81 & 10.49 \\ 1.0 & 33.47 & 1.10 & 10.68 \\ 1.0 & 34.54 & 1.05 & 10.83 \\ 1.0 & 35.44 & 0.96 & 11.06 \end{bmatrix} \begin{bmatrix} -1.94 \\ 0.23 \\ 0.03 \\ 0.02 \end{bmatrix} = \begin{bmatrix} 5.56 \\ 5.8 \\ 6.05 \\ 6.30 \\ 6.51 \end{bmatrix}$$

The detrended water use series $Q_b(m,y) = Qa(m,y) - Qa(y)$ is reported in table 2

Table 2 - Detrended monthly water use Qb(m,y)

Year	Month	Qb(m,y)	Year	Month	Qb(m,y)	Year	Month	Qb(m,y)	Year	Month	Qb(m,y)	Year	Month	Qb(m,y)
2000	1	1,74	2001	1	-1,50	2002	1	-1,55	2003	1	-2,00	2004	1	-1,41
2000	2	0,24	2001	2	-2,00	2002	2	-2,15	2003	2	-2,00	2004	2	-1,61
2000	3	-0,86	2001	3	-0,90	2002	3	-1,65	2003	3	-1,10	2004	3	-1,01
2000	4	-1,66	2001	4	-0,20	2002	4	-1,45	2003	4	-0,80	2004	4	-0,21
2000	5	-0,56	2001	5	-0,40	2002	5	-1,05	2003	5	-1,00	2004	5	-0,11
2000	6	1,04	2001	6	1,40	2002	6	0,55	2003	6	3,00	2004	6	-0,61
2000	7	2,34	2001	7	4,50	2002	7	0,55	2003	7	5,70	2004	7	1,79
2000	8	2,24	2001	8	2,10	2002	8	1,05	2003	8	3,10	2004	8	1,09
2000	9	0,84	2001	9	-0,60	2002	9	0,35	2003	9	0,50	2004	9	-0,61
2000	10	-1,86	2001	10	0,50	2002	10	0,85	2003	10	-0,30	2004	10	0,29
2000	11	-1,06	2001	11	-1,50	2002	11	-1,35	2003	11	-0,90	2004	11	-0,51
2000	12	-1,26	2001	12	-2,00	2002	12	-1,65	2003	12	-1,40	2004	12	-0,31

2.2. DESEASONALIZATION

The second transformation in the cascade model removes the *within year seasonal water use* in the detrended water use series $Q_b(m,y)$ (Table 2). The detrended monthly time series can be used to compute the arithmetic average $Q_b(m)$ for each month within a year that contain a seasonality component:

$$\mathrm{Qb(m)} = \frac{1}{\mathrm{Y}}\sum_{\mathrm{y}=1}^{\mathrm{Y}} \mathrm{Q_b(m,y)}$$

m = 1, 2,......, 12

Or represented as a Fourier series:

$$\mathrm{Qb(m)} = \sum_{\mathrm{k}=0}^{6} \mathrm{a_k cos}\left(\frac{2\pi \mathrm{km}}{12}\right) + \mathrm{b_k sen}\left(\frac{2\pi \mathrm{km}}{12}\right)$$

m = 1, 2,......, 12

where

$$a_k = \frac{1}{6}\sum_{m=1}^{12}\left[\frac{1}{Y}\sum_{y=1}^{Y} Q_b(m,y)\right]\cos\frac{2\pi km}{12} \qquad \text{for k= 1,2,...,6}$$

$$b_k = \frac{1}{6}\sum_{m=1}^{12}\left[\frac{1}{Y}\sum_{y=1}^{Y} Q_b(m,y)\right]\sin\frac{2\pi km}{12} \qquad \text{for k= 1,2,...,6}$$

The seasonality of water use rapresented by $\mathrm{Q_b(m)}$ is subtracted from the detrended water use series Qb(m,y)(Table xxxx) resulting in the deseasonalized water use series Qc(m,y), as:

$$\mathrm{Q_c(m,y)= Q_b(m,y) - Q_b(m)}$$

m = 1, 2,......, 12 y= 1, 2,.....Y

Exemple:

We can calculate a_1, a_2, ...,a_6 and b_1, b_2, ..., b_6 with k=1 as:

$$a_1 = Q_b(1)\cos\left(\frac{\pi}{6}\right) + Q_b(2)\cos\left(\frac{2\pi}{6}\right) + \cdots + Q_b(12)\cos\left(\frac{12\pi}{6}\right)$$

....

....

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} = \begin{bmatrix} -0.099 \\ -1.229^* \\ 0.220^* \\ 0.059 \\ -0.186^* \\ 0.019 \\ 0.014 \\ 0 \\ -1.101^* \\ 0.733^* \\ -0.383^* \\ 0.337^* \\ 0.109 \\ 0 \end{bmatrix}$$

Only the * coefficients are significatively different from zero

So $\mathrm{Qb(m)}$ become:

$$\mathrm{Qb(m)} = -1.229\cos\left(\frac{\mathrm{m}\pi}{6}\right) + 0.220\cos\left(\frac{2\mathrm{m}\pi}{6}\right) - 0.186\cos\left(\frac{4\mathrm{m}\pi}{6}\right) - 1.101\sin\left(\frac{\mathrm{m}\pi}{6}\right) + 0.733\sin\left(\frac{2\mathrm{m}\pi}{6}\right) - 0.383\sin\left(\frac{3\mathrm{m}\pi}{6}\right) + 0.337\sin\left(\frac{4\mathrm{m}\pi}{6}\right)$$

The regenerated values of $\mathrm{Qb(m)}$ are calculated as:

$$\begin{bmatrix} Q_b(m=1) \\ Q_b(m=2) \\ Q_b(m=3) \\ Q_b(m=4) \\ Q_b(m=5) \\ Q_b(m=6) \\ Q_b(m=7) \\ Q_b(m=8) \\ Q_b(m=9) \\ Q_b(m=10) \\ Q_b(m=11) \\ Q_b(m=12) \end{bmatrix} = \begin{bmatrix} -0.868 \\ -1.242 \\ -1.124 \\ -0.698 \\ -0.592 \\ 1.263 \\ 3.128 \\ 1.894 \\ 0.311 \\ -0.020 \\ -0.854 \\ -1.195 \end{bmatrix}$$

The deseasonality monthly water use (short-memory process) represented by $\mathrm{Q_c(m)}$ = Qb(m,y) - $\mathrm{Q_b(m)}$ is given in the following table 3.

Table 3 - Deseasonalized monthly water use Qc(t)

Year	Month	Qc(m,y)	Year	Month	Qc(m,y)	Year	Month	Qc(m,y)	Year	Month	Qc(m,y)	Year	Month	Qc(m,y)
2000	1	2.61	2001	1	-0.64	2002	1	-0.67	2003	1	-0.46	2004	1	-0.24
2000	2	1.56	2001	2	-0.69	2002	2	-0.94	2003	2	-1.08	2004	2	-0.48
2000	3	0.34	2001	3	0.22	2002	3	-0.54	2003	3	-0.79	2004	3	-0.39
2000	4	-0.89	2001	4	0.49	2002	4	-0.73	2003	4	-0.02	2004	4	0.08
2000	5	0.04	2001	5	0.26	2002	5	-0.44	2003	5	-0.07	2004	5	0.44
2000	6	-0.19	2001	6	0.13	2002	6	-0.70	2003	6	-0.39	2004	6	0.41
2000	7	-0.76	2001	7	1.45	2002	7	-2.61	2003	7	1.72	2004	7	-1.89
2000	8	0.42	2001	8	0.21	2002	8	-0.2	2003	8	2.62	2004	8	-1.35
2000	9	0.54	2001	9	-0.87	2002	9	-0.82	2003	9	1.22	2004	9	-0.84
2000	10	-1.79	2001	10	0.51	2002	10	0.06	2003	10	0.19	2004	10	-0.93
2000	11	-0.23	2001	11	-0.4	2002	11	0.91	2003	11	-0.22	2004	11	0.27
2000	12	-0.04	2001	12	-0.28	2002	12	-0.47	2003	12	-0.03	2004	12	0.33

2.3. AUTOREGRESSIVE FILTERING

The third transformation in the cascade model (Fig. 1) accounts for the autocorrelation (in a time series is a measure of persistence of observation) of Qc(t) by fitting the autoregressive model.

$$Q_c(t) = Q_c(m, y) \quad with \quad t = 12(y-1)$$

$$Q_c(t) = \sum_{i=1}^{I} \phi_i Q_c(t-i) + Q_d(t) \qquad t = 2, \ldots\ldots, T$$

Where I is the maximum number of lags, ϕis the unknown parameters and $Q_d(t)$ is the residual of autoregression. The parameters ϕ and $Q_d(t)$ are estimating using a linear regression analysis.

$$t = 2 \qquad Q_c(2) = \phi_1 Q_c(1) \qquad + Q_d(2)$$

$$t = 3 \qquad Q_c(3) = \phi_1 Q_c(2) + \phi_2 Q_c(1) \qquad + Q_d(3)$$

$$t = 4 \qquad Q_c(4) = \phi_1 Q_c(3) + \phi_2 Q_c(2) + \phi_3 Q_c(1) \qquad + Q_d(4)$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$t = T \quad Q_c(T) = \phi_1 Q_c(T-1) + \phi_2 Q_c(T-2) + \cdots + \phi_I Q_c(T-I) + Q_d(4)$$

The lag-1 autocorrelation in $Q_c(m)$ is the correlation coefficient of this regression equation:

$$Q_c(t) = \phi_0 + \phi_1 Q_c(t-1) + Q_d(t)$$

For t = 2,3, …, 60

The residual $Q_d(t) = Q_c(t) - [\phi_0 + \phi_1 Q_c(t-1)]$ became (table 4):

$$Q_d(t) = Q_c(t) - [-0.096 + 0.43\, Q_c(t-1)]$$

Table 4 - Monthly water use Qd(t), after autoregression has been subtract

Year	Month	Qd(m,y)	Year	Month	Qd(m,y)	Year	Month	Qd(m,y)	Year	Month	Qd(m,y)	Year	Month	Qd(m,y)
2000	1	0	2001	1	-0.52	2002	1	-0.46	2003	1	-0.78	2004	1	-0.28
2000	2	0.52	2001	2	-0.31	2002	2	-0.56	2003	2	-0.22	2004	2	-0.09
2000	3	-0.23	2001	3	0.61	2002	3	-0.04	2003	3	0.41	2004	3	0.35
2000	4	-0.94	2001	4	0.48	2002	4	-0.40	2003	4	0.03	2004	4	0.50
2000	5	0.52	2001	5	0.14	2002	5	-0.03	2003	5	-0.26	2004	5	0.32
2000	6	-0.11	2001	6	0.11	2002	6	-0.42	2003	6	1.98	2004	6	-1.98
2000	7	-0.58	2001	7	0.50	2002	7	-2.22	2003	7	1.97	2004	7	-0.44
2000	8	0.84	2001	8	0.32	2002	8	0.40	2003	8	0.18	2004	8	-0.16
2000	9	0.46	2001	9	-0.87	2002	9	0.50	2003	9	-0.24	2004	9	-0.47
2000	10	-1.93	2001	10	0.99	2002	10	0.99	2003	10	-0.21	2004	10	0.77
2000	11	0.64	2001	11	-0.54	2002	11	-0.77	2003	11	0.16	2004	11	0.31
2000	12	0.15	2001	12	-0.01	2002	12	-0.16	2003	12	-0.13	2004	12	0.85

2.4. CLIMATIC REGRESSION

The final transformation of the cascade models evaluates the dependence of the monthly water use $Q_d(t)$ with the climatic factors: rainfall, evaporation, and air temperature.
This dependency relationship can be modeled as:

$$Qd(t) = \sum_{l=1}^{L} \beta_l X_l(t) + Q_e(t)$$

t = 1, 2, …, T
X_l = l^th climatic factor
β = unknown parameters estimable by linear regression

L = total number of climatic factor
$Q_e(t)$ = residual from the climatic regression (is a pure random error)

Exmple:

If X_1 is the monthly precipitation and X_2 the maximum temperature, the model for the climatic regression is:

$$Q_d(t) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + Q_e(t)$$

t = 1, 2, …, 60
the results of a regression analysis yield the following model parameters

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

And $Q_e(t)$ become:

$$Q_e(t) = Q_d(t) - \left[\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2\right]$$

CONCLUSIVE REMARKS

For the applicability of the cascade model are necessary:

1. Annual values of socioeconomic variables (for the detrending);
2. Monthly values of water use from all source;
3. Monthly values of climatic variations.

The first and last one are easily recoverable from existing data bases, however the seconds are more difficult to find. In effect, frequently only the annual values of the volumes of water used in cities are available and not the monthly ones. The annual values are not significant for this model, because, for example, the seasonal cycles are lost.

For these reasons, it is believed that the use of this technique is currently possible with the details indicated by Maidment and other authors for groundwater management in the Italian cities where the piezometric surveys and sampling are carried out on a monthly basis. The method can also be advantageously used for the analysis of the piezometric fluctuations and of the concentrations of the contaminants.

REFERENCES

- Maidment, D. R. and Parzen E. (1984b). "A Cascade Model of Municipal Water Use." Water Resources Research 20(1): 15-23
- Mays L. W. and Tung Y.-K. (1992). HydroSystems Engineering and Management. McGraw-Hill Inc, New York (1992).