MASHWIN

MODEL FOR ANALYSIS OF HYDROLOGICAL SERIES

USER’S MANUAL

DEPARTAMENTO DE INGENIERIA HIDRAULICA Y MEDIO AMBIENTE
UNIVERSIDAD POLITECNICA DE VALENCIA

ACKNOWLEDGEMENTS

This manual has been updated and translated during the development of WAM-ME project (Water Resources Management under Drought Conditions: Criteria and tools for the conjunctive use of conventional and marginal waters in the Mediterranean region) EUROPEAN COMMISSION - Research DG – INCO-MED. Contract nº ICA3-CT-1999-00014

MODEL FOR ANALYSIS OF HYDROLOGICAL SERIES: MASHWIN

INDEX

1. INTRODUCTION. 1

2. THEORETICAL BASIS.......... 2

2.1. STATISTICAL ANALYSIS 2
2.2. NORMALIZATION OF THE SERIES.......... 5
2.3. ADJUSTMENT IN FOURIER SERIES (MONTHLY SERIES) 6
2.4. TYPIFICATION OF THE SERIES AND CORRELATIONS 9
2.5. MULTIVARIATE STOCHASTIC MODELS.......... 11

3. BRIEF DESCRIPTION OF THE ORGANIZATION AND FUNCTIONING OF THE MODEL. 25

4. GRAPHIC INTERFACE.......... 26

4.1. Content of the screens at the different levels. 26
4.2. Generalization.......... 36
4.3. Graphs. 37

5. CALCULATION MODULES 38

5.1. LEVEL 1: MONTHLY AND ANNUAL ANALYSIS.......... 38
5.2. LEVEL 2(a): MONTHLY NORMALIZATION.......... 39
5.3. LEVEL 2(b): ANNUAL NORMALIZATION.......... 41
5.4. LEVEL 3(a): MONTHLY TYPIFICATION.......... 42
5.5. LEVEL 3(b): ANNUAL TYPIFICATION.......... 43
5.6. LEVEL 4: MONTHLY AND ANNUAL CALIBRATION 44
5.7. LEVEL 5: GENERATION OF SYNTHETIC SERIES.......... 47

BIBLIOGRAPHY 54

MODEL FOR ANALYSIS OF HYDROLOGICAL SERIES: MASHWIN

1. INTRODUCTION.

MASHWIN is a stochastic analysis model for the study of the temporal (monthly) inflow series in a hydrological system. It is a useful and powerful calculation and analysis tool, which helps to define the stochastic structure that better adapts to a group of temporal inflow series. It helps to achieve a better knowledge of the analyzed system and it is a previous result for the generation of conditioned synthetic series, which can be created either within MASHWIN or further on with other models.

For the analysis of the temporal inflow series in a hydrological system; MASHWIN combines a stochastic periodic modeling by an autoregressive and moving average multivariate method (ARMA), and a spatial segregation by the condensed model of Lane. MASHWIN also performs a wide variety of adjustment tests for all the analysis phases.

The model has been structured in a set of applications for a detailed analysis of a group of hydrological series. It is also coordinated in all the partial steps by a user’s interface which allows automatic access to all the analysis phases.

2. THEORETICAL BASIS

The following sections explain the theoretical basis of the analysis performed in MASHWIN model.

2.1. STATISTICAL ANALYSIS

The model calculates three groups of statistics: a) *basic statistics*, which include the averages, the standard deviations and the skew coefficients of the flow series, b) *draught statistics*, which include the averages, maximums and standard deviations of the duration, the intensity and the magnitude of the droughts for different thresholds; and c) *storage statistics*, which are the storage capacities of the flow series for different thresholds, as well as the adjusted rank, the adjusted re-scaled rank and the Hurst coefficient for each flow series.

2.1.1. Basic statistics

If the series are given by monthly periods, the averages, standard deviations and skew coefficients are calculated for each of the 12 months of the year, according to the equations (1).

$$\overline{Q}_\tau = \frac{\sum_{v=1}^{N} Q_{v,\tau}}{N} \qquad s_\tau = \sqrt{\frac{\sum_{v=1}^{N}\left(Q_{v,\tau}-\overline{Q}_\tau\right)^2}{N-1}} \qquad g_\tau = \frac{N\sum_{v=1}^{N}\left(Q_{v,\tau}-\overline{Q}_\tau\right)^3}{(N-1)(N-2)s_\tau^3} \qquad (1)$$

Where: $\overline{Q}_\tau$: Monthly average of the flows for the month τ.

$Q_{v,\tau}$: Flow for the month τ , year *v* (*v*=1,…,N).

N: Total number of years.

s_τ: Standard deviation for month τ.

g_τ: Skew coefficient for month τ.

If the modeling is made over annual series, these statistics are calculated according to the equations (2), and there will be one only value for each flow series.

$$\overline{Q} = \frac{\sum_{v=1}^{N} Q_v}{N} \qquad s = \sqrt{\frac{\sum_{v=1}^{N} \left(Q_v - \overline{Q}\right)^2}{N-1}} \qquad g = \frac{N \sum_{v=1}^{N} \left(Q_v - \overline{Q}\right)^3}{(N-1)(N-2)s^3} \quad (2)$$

Where $\overline{Q}$: Annual average of the flow series.

Q_v: Flow in year *v* (*v*=1,…,N).

N: Total number of years.

s: Annual standard deviation of the flow series.

g: Annual skew coefficient of the flow series

2.1.2. Drought statistics

To describe the statistics of this category, it is convenient to define previously some concepts used in their determination.

- THRESHOLD: Fraction of the average flow from the whole historical flow series, which is used to determine a situation of *drought* or to determine the volume of a reservoir in which to store a certain amount of water. A threshold can be assimilated to a horizontal line overlying the complete series hydrogram at a height equal to that fraction of the average flow mentioned.

- DEFICIT: Flow below the threshold.

- DROUGHT: Succession of consecutive periods (years or months) in state of deficit. It is basically characterized by three elements: *duration, intensity and magnitude*, which are the statistics calculated in this category.

Therefore, the drought statistics are the following:

- DURATION: Number of consecutive periods in state of deficit.
- INTENSITY: Maximum deficit among all of the drought deficits.
- MAGNITUDE: Total volume of the deficits during the drought.

The program creates the drought series which correspond to the flow series for each station and threshold. Then it calculates the averages, maximums and standard deviations of duration, intensity and magnitude, Eqs. (2a).

$$\bar{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i, \qquad \gamma_{máx} = \text{Máx}(\gamma_1, \gamma_2, \gamma_3, ..., \gamma_n), \qquad s_\gamma = \sqrt{\frac{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}{n-1}} \tag{2a}$$

$$\gamma = \begin{cases} \text{Duration} \\ \text{Intensity} \\ \text{Magnitude} \end{cases}$$

Where n is the number of droughts for a given station and threshold.

2.1.3. Storage statistics

In this group, the *storage capacity* correspondent to one (or several) threshold is calculated. This storage capacity is the volume of water required to completely supply a constant demand equivalent to the value of the threshold. This is made by the algorithm of the *Sequential Peaks*, according to which the storage capacity is the maximum of the values calculated with Eq. (3).

$$S_t = \begin{cases} U_t - Q_t + S_{t-1}, & \text{if } S_t > 0 \\ 0, & \text{if } S_t < 0 \end{cases} \quad \text{for} \quad t = 1, 2, 3, ..., 2N \tag{3}$$

Where:

S_t: Storage volume in the period t (S_0=0).

U_t: Threshold in the period t. It is constant, and it can not surpass the average flow value of the whole series.

Q_t: Flow in the period t. From N+1 to 2N the series is repeated, in order to take account of those cases in which the sequence of critical flows (below the threshold) are at the end of the series.

N: Number of periods (months or years) of the series.

In addition to the storage capacity, the model calculates the adjusted rank, the re-scaled adjusted rank and the Hurst coefficient, according to the equations (4) to (6).

$$R^* = \mathrm{Max}(S_0, S_1, \ldots, S_N) - \mathrm{Min}(S_0, S_1, \ldots, S_N) \qquad (4$$

with

$$S_t = S_{t-1} + (Q_t - Q_m), \qquad t = 1, 2, 3, \ldots, N$$

$$R^{**} = R^*/s_N, \quad \text{being} \quad s_N = \sqrt{\sum_{t=1}^{N} (Q_t - Q_m)^2 / N} \qquad (5)$$

$$K = \frac{\ln(R^{**})}{\ln(N/2)} \qquad (6)$$

Where:

- Q_m: Average flow of the whole series.
- R^*: Adjusted rank.
- R^{**}: Re-scaled adjusted rank.
- s_N: Standard deviation of the whole series.
- K: Hurst coefficient.

2.2. NORMALIZATION OF THE SERIES

This phase consists on transforming the original flow series ($Q_{v,\tau}$ or Q_v, depending on the series being monthly or annual) in normalized series, i.e. series with normal distribution. ($x_{v,\tau}$ or x_v, depending on the series being monthly or annual), by using the functions indicated in (7).

$$\begin{gathered} x_{v,\tau} = \sqrt{Q_{v,\tau}} \quad x_{v,\tau} = \ln(Q_{v,\tau} + 1) \quad x_{v,\tau} = \ln[\ln(Q_{v,\tau} + 1) + 1] \quad x_{v,\tau} = (Q_{v,\tau} - a)^b \\ x_v = \sqrt{Q_v} \quad x_v = \ln(Q_v + 1) \quad x_v = \ln[\ln(Q_v + 1) + 1] \quad x_v = (Q_v - a)^b \end{gathered} \qquad (7)$$

The use of logarithmic functions avoids negative results and mathematical indeterminations in the normalized series.

When the series are monthly it is possible to apply different normalization functions to the different months. Although, the only fact of applying these functions does not guarantee that the transformed series are normalized. To verify that condition, the program calculates the skew coefficient of the normalized series and, by the expression (8), it evaluates if the skew is within the normality interval (Snedecor and Cochran, 1967, quoted by Salas et al., 1980).

$$\begin{aligned} &\left[-3.9601\mathrm{N}^{-0.4598},\ +3.9601\mathrm{N}^{-0.4598}\right] && \text{for} \quad \mathrm{N} < 150 \\ &\left[-1.96\sqrt{\frac{6}{\mathrm{N}}},\ 1.96\sqrt{\frac{6}{\mathrm{N}}}\right] && \text{for} \quad \mathrm{N} \geq 150 \end{aligned} \qquad (8)$$

Where N is the number of years of the series (even if it is monthly).

It is considered that if the value of the skew coefficient is within the interval (8), the series is normally distributed.

To apply the potential normalization (last of the four methods indicated in equation (7)) the parameters *a* and *b* must be previously calibrated, taking care that *a* does not take values above the minimum flows in the series. There is also the option of calibrating in an iterative way, giving out values to *a* and *b* and observing in the results files if the skew coefficient is within the normality interval

2.3. ADJUSTMENT IN FOURIER SERIES (MONTHLY SERIES)

When analyzing monthly flow series, it is usual to find that the variations of monthly averages and standard deviations throughout the year present periodicities; in other words, their graphic representation can be similar in some way to sinusoidal type functions. This property allows representing those statistics by approximate functions by the development of Fourier series, in order to reduce the global number of parameters of the stochastic model. It is important to have in mind that the statistics used to typify the flow series are parameters for the model: for each series (monthly) there are 12 averages and 12 standard deviations, in other words 24 parameters. With the Fourier adjustment, those parameters can be appreciably reduced.

This technique is applied in order to respect the principle of statistic parsimony. It is important to pay attention to this principle especially when dealing with multivariate models having an important number of stations.

The procedure in the program consists on calculating by the Eqs. (9) to (11) the Fourier coefficients Aj and Bj for all the harmonics, that in this case, because they are monthly series, equals to $\omega/2=12/2=6$ harmonics.

$$Aj = \frac{2}{\omega}\sum_{\tau=1}^{\omega} u_\tau \cos\left(\frac{2\pi j\tau}{\omega}\right), \quad j = 1,\ldots,h-1 \quad (9)$$

$$Aj = \frac{1}{\omega}\sum_{\tau=1}^{\omega} u_\tau \cos\left(\frac{2\pi j\tau}{\omega}\right), \quad j = h \quad (10)$$

$$B_j = \frac{2}{\omega}\sum_{\tau=1}^{\omega} u_\tau \operatorname{sen}\left(\frac{2\pi j\tau}{\omega}\right), \quad j = 1,\ldots,h \quad (11)$$

with

$$u_\tau = \begin{cases} \overline{x}_\tau \\ s_{x_\tau} \end{cases}$$

Where: ω: Number of periods of the year (12, because they are monthly series).

τ: Order of the month in the year.

u_τ: Monthly average (or standard deviation) of the normalized series for month τ.

j: Order of the harmonic.

h: Total number of harmonics. It is $\omega/2$, since ω is even.

The representation in Fourier series of the values of the monthly average and standard deviation from the normalized series is calculated. The equation used is (12) in which all of the harmonics are considered.

$$v_\tau = \overline{u} + \sum_{j=1}^{h}\left[A_j \cos(2\pi j\tau/\omega) + B_j \operatorname{sen}(2\pi j\tau/\omega)\right], \quad \tau = 1,\ldots,\omega \quad (12)$$

with

$$\bar{u} = \frac{1}{\omega}\sum_{\tau=1}^{\omega} u_\tau$$

Where: v_τ: Monthly average (or standard deviation) of the normalized series adjusted by Fourier for the month τ.

$\bar{u}$: Average value of the monthly averages (or standard deviations) of the normalized series.

With the Fourier coefficients, the variance explained by each harmonic (Ec. (13)) is calculated, in order to build the *accumulated periodogram* (graphic in which the accumulated percentage of explained variance is represented as a function of the number of harmonics) and decide the number of *significative harmonics* (those which contribute to the accumulated percentage of explained variance) for the monthly average and standard deviation of the normalized series.

$$C_j^2 = A_j^2 + B_j^2 \qquad (13)$$

Where C_j^2 is the variance explained by the harmonic j.

With the explained variance and the total variance we obtain the percentage of the variance explained by each of the harmonics, as it is indicated in Eq. (14).

$$\%V_{exp} = \frac{C_j^2}{2s_u^2} \qquad (14)$$

with

$$s_u^2 = \frac{\sum_{\tau=1}^{\omega}\left(u_\tau - \bar{u}\right)^2}{\omega}$$

Where: $\%V_{exp}$: Percentage of the total explained variance of the monthly averages (or standard deviations) from the normalized series before the adjustment by Fourier.

s_u^2: Variance of the monthly averages (or standard deviations) from the normalized series.

Once defined the accumulated percentage of variance which is accepted to be explained by the statistics adjusted by Fourier, the accumulated periodogram of the average is entered, the number of significative harmonics is selected and then the adjusted average is calculated as a function of those harmonics, as indicated in Eq. (15). The same is applicable for the standard deviation.

$$\hat{v}_\tau = \bar{u} + \sum_{j=1}^{h_s}\left[A_j \cos\left(\frac{2\pi j\tau}{\omega}\right) + B_j \,\mathrm{sen}\left(\frac{2\pi j\tau}{\omega}\right)\right], \qquad j = h_1, \ldots, h_s \tag{15}$$

with

$$\hat{v}_\tau = \begin{cases} \hat{x}_\tau \\ \hat{s}_{x_\tau} \end{cases}$$

Where $\hat{x}_\tau$: Monthly average from month τ of the normalized series.

$\hat{s}_{x_\tau}$: Monthly standard deviation from month τ of the normalized series.

h_1: First significative harmonic.

h_s: Last significative harmonic.

In order to know the quality of the adjustment, it is recommendable to represent in the same graph the averages before and after being adjusted, as well as the standard deviations, and judge from that graph if it is worth it to use the adjusted statistics. It is important to be very careful taking this decision, because the differences among the statistics can create important distortions in the values of the synthetic flows, and therefore, damage the calculations during the application of the model.

2.4. TYPIFICATION OF THE SERIES AND CORRELATIONS

After having the flow series normalized they will be typified by the Eq. (16), in order to eliminate the periodicities –in monthly series- and take them to the same ‘scale’, since that way the differences between them are more evident and explainable.

This way the correlations between the different series can be evaluated, in order to establish the type of modeling from the cross correlation matrixes and the autocorrelation functions.

Monthly series: $z_{v,\tau} = \dfrac{x_{v,\tau} - \hat{x}_\tau}{\hat{s}_{x_\tau}}$ Annual series: $z_v = \dfrac{x_v - \hat{x}}{\hat{s}_x}$ (16)

2.4.1. Autocorrelation functions

The program calculates the autocorrelation functions from the typified series (sample autocorrelation series) and the limits of the temporal independence interval (Anderson, 1941. reference in Salas et al., 1980), by Eqs. (17) and (18) respectively. These functions are useful to know better the structure of the temporal dependence in the series. This gives an idea about the order or size of the model which has to be adjusted to them. This is made by comparing the theoretical autocorrelation functions of the models with the sample ones to know, depending on the similitude between them, which one is more suitable.

$$r_k(z) = \frac{\sum_{t=1}^{N-k} (z_{t+k} - \overline{z})(z_t - \overline{z})}{\sum_{t=1}^{N} (z_t - \overline{z})^2} \qquad (17)$$

$$\left[\frac{-1 - 1.96\sqrt{N-k-1}}{N-k}, \frac{-1 + 1.96\sqrt{N-k-1}}{N-k} \right] \qquad (18)$$

Where: $r_k(z)$: Autocorrelation coefficient for a time interval k.

z_t: Normalized and typified flow in the period (month or year) t. In monthly series it equals $z_{v,\tau}$ with $t=12(v-1)+\tau$. (v: year, v=1,...,N); in annual series it is z_v.

$\overline{z}$: Average of the whole typified series.

N: Total number of flows. It is the same for all of the series.

2.4.2. Cross correlation matrixes

Next, the program calculates (19) the cross correlation matrixes $\mathbf{M}_k$ for the different time intervals k, specified by the modeler.

$$\mathbf{M}_k(z)=\begin{bmatrix} r_k^{1,1}(z) & \cdots & r_k^{1,n}(z) \\ \vdots & & \vdots \\ r_k^{n,1}(z) & \cdots & r_k^{n,n}(z) \end{bmatrix} \qquad r_k^{ij}(z)=\frac{\sum_{t=1}^{N-k}\left(z_{t+k}^{(i)}-\bar{z}^{(i)}\right)\left(z_t^{(j)}-\bar{z}^{(j)}\right)}{\sqrt{\sum_{t=1}^{N}\left(z_t^{(i)}-\bar{z}^{(i)}\right)^2\sum_{t=1}^{N}\left(z_t^{(j)}-\bar{z}^{(j)}\right)^2}} \tag{19}$$

Where: $r_k^{ij}(z)$: Cross correlation coefficient between the series i and j for a time interval k.

$z_t^{(i)}$: Normalized and typified flow from the series i in the period (month or year) t.

n: Number of stations in the multivariate model.

The calculation of the cross correlation matrixes is made in order to: a) know the interdependence existent between the series and decide wither the model must be multivariate or not; b) select the groups of stations in a spatial segregation scheme; and c) calculate the parameter matrixes for the stochastic models. The evaluations correspondent to a) and b) are made over the matrix k=0, and the calculations for c) are made over the matrixes k=0, 1, 2.

2.5. MULTIVARIATE STOCHASTIC MODELS

MASHWIN has three types of stochastic models available for the generation of synthetic series: a) type ARMA (Auto-Regressive Moving Average); b) SPATIAL SEGREGATION; and c) TEMPORAL SEGREGATION.

2.5.1. ARMA type models

The program has three models that belong to the group of ARMA models with (*p*, *q*) order, i.e., with *p* autoregressive parameters and *q* mobile average parameters: ARMA(p, q). These models tend to keep the statistics of first order from the sample series: average, standard deviation and the first *p* autocorrelations. Due to these characteristics they are very useful when generating future series of flows equiprobable to the historic series.

The program uses ARMA models with constant parameters, the general formulation of which is given by the Eq. (20). When the series are monthly, the models with constant parameters, in opposition to those with periodic parameters, are more consequent with the principle of statistic parsimony (less parameters) and in most cases they are adequate enough for the modeling of those series.

$$\{Z\}_t = [\Phi]_1\{Z\}_{t-1} + [\Phi]_2\{Z\}_{t-2} + \cdots + [\Phi]_p\{Z\}_{t-p} + [\Theta]_0\{\varepsilon\}_t - [\Theta]_1\{\varepsilon\}_{t-1} - \cdots - [\Theta]_q\{\varepsilon\}_{t-q} \quad (20)$$

Where: $\{Z\}_t$: Flows vector in period (month or year) t.

$[\Phi]_p$: Autoregressive parameters matrix.

$[\Theta]_q$: Moving average parameters matrix.

$\{\varepsilon\}_t$: Residuals vector in period t (random values, independent and normally distributed with zero average and unitary variance).

Model AR(1)

It is the most simple pure autoregressive model, Ec. (21). It represents the flow from the time interval t as a function of the flow from the precedent period t-1 and an independent random variable ε, normally distributed with average zero and unitary variance.

$$\{Z\}_t = [\Phi]_1\{Z\}_{t-1} + [\Theta]_0\{\varepsilon\}_t \quad (21)$$

The program calculates the matrixes with parameters $[\Phi]_1$ and $[\Theta]_0$ by an estimation procedure based on the momentum method (Matalas, 1967; reference in Salas et al., 1980); those matrixes are given by the Eqs. (22) and (23). It is necessary to solve this last one for $[\Theta]_0$ using the Cholesky decomposition algorithm.

$$[\Phi]_1 = \mathbf{M}_1\mathbf{M}_0^{-1} \quad (22)$$

$$[\Theta]_0[\Theta]_0^T = \mathbf{M}_0 - [\Phi]_1\mathbf{M}_1^T \quad (23)$$

The components of the matrix $[\Theta]_0$ are calculated with the following procedure.

We make

$$[\Theta]_0 [\Theta]_0^T = \mathbf{D}$$

This equation has an infinite number of solutions for $[\Theta]_0$, but if we assume that it is an inferior triangular matrix, and if **D** is a matrix defined positive or semidefined positive, the elements from $[\Theta]_0$ can be solved (Lane, 1979, quoted by Salas et al., 1980) by the algorithm given by the Eqs. (24).

$$\theta_{ki} = 0 \quad \forall\ k < i$$

$$\theta_{ki} = 0 \quad \forall k \geq i, \quad \text{when} \quad d_{ii} - \sum_{j<i} (\theta_{ij})^2 \leq 0 \tag{24}$$

$$\theta_{ki} = \frac{d_{ki} - \sum_{j<i} \theta_{ij}\theta_{kj}}{\left[d_{ii} - \sum_{j<i} (\theta_{ij})^2 \right]^{1/2}} \quad \forall\ k \geq i \quad \text{when} \quad d_{ii} - \sum_{j<i} (\theta_{ij})^2 > 0$$

Model AR(2)

In this model, Eq. (25), the flow in a period t is function of the flows in the two previous periods t-1 and t-2, and of an aleatory variable, ε, normally distributed with average zero, unitary variance and not correlated.

$$\{Z\}_t = [\Phi]_1 \{Z\}_{t-1} + [\Phi]_2 \{Z\}_{t-2} + [\Theta]_0 \{\varepsilon\}_t \tag{25}$$

The model is calibrated with the solutions of the matritial equations (26) to (28) (Salas y Pegram, 1978. reference in Salas et al., 1980). They give out as a result the parameter matrixes $[\Phi]_1$, $[\Phi]_2$ and $[\Theta]_0$.

$$[\Phi]_1 = \left[M_1 - M_2 M_0^{-1} M_1^T\right]\left[M_0 - M_1 M_0^{-1} M_1^T\right]^{-1} \tag{26}$$

$$[\Phi]_2 = \left[M_2 - M_1 M_0^{-1} M_1\right]\left[M_0 - M_1^T M_0^{-1} M_1\right]^{-1} \tag{27}$$

$$[\Theta]_0[\Theta]_0^T = M_0 - \left([\Phi]_1 M_1^T + [\Phi]_2 M_2^T\right) \qquad (28)$$

The program evaluates the components of the matrix $[\Theta]_0$ in the same way as in the model AR(1).

Model ARMA(1,1)

Unlike the two previous models, this one, Eq. (29), includes a moving average component.

$$\{Z\}_t = [\Phi]_1\{Z\}_{t-1} + [\Theta]_0\{\varepsilon\}_t - [\Theta]_1\{\varepsilon\}_{t-1} \qquad (29)$$

For the calculation of the parameter matrixes, the program completes the following process (O'Connell, 1974, reference in Bras y Rodríguez-Iturbe, 1985).

First it calculates the matrix $[\Phi]_1$, as indicated in Eq. (30).

$$[\Phi]_1 = \mathbf{M}_2\mathbf{M}_1^{-1} \qquad (30)$$

Then, by a Cholesky decomposition, it gives an initial value to the matrix $[\Theta]_0$, and then it solves the Eq. (31) in an iterative way starting from that initial value.

$$\left([\Theta]_0[\Theta]_0^T\right)_j = \mathbf{F} - \mathbf{G}\left([\Theta]_0[\Theta]_0^T\right)_{j-1}^{-1}\mathbf{G}^T \qquad (31)$$

with

$$\mathbf{F} = \mathbf{M}_0 - [\Phi]_1\mathbf{M}_1^T + \mathbf{G}[\Phi]_1^T$$

and

$$\mathbf{G} = [\Phi]_1\mathbf{M}_0 - \mathbf{M}_1$$

After finding the matrix $[\Theta]_0$, the program calculates the matrix $[\Theta]_1$ by the Eq. (32).

$$[\Theta]_1 = \mathbf{G}\left([\Theta]_0^T\right)^{-1} \qquad (32)$$

Theoretical autocorrelation functions

The program calculates the multivariate theoretical autocorrelation functions (FACT) correspondent to the multivariate models AR(1) and AR(2), for its comparison with the sample autocorrelation functions (FACM). This provides another criterion to select the most adequate ARMA model. Among two models, the one that shows more similarities between FACT and FACM tends to explain better the behavior of the sample series.

For this calculation, the program evaluates the cross correlation theoretical matrixes (33) from each model for as many time intervals k as the modeler specifies, and then it extracts the elements from their diagonals to conform the multivariated theoretical correlograms (FACT).

$$\mathbf{M}_k(z)=\begin{bmatrix} r_k^{1,1}(z) & r_k^{1,2}(z) & r_k^{1,3}(z) & \cdots & r_k^{1,n}(z) \\ r_k^{2,1}(z) & r_k^{2,2}(z) & r_k^{2,3}(z) & \cdots & r_k^{2,n}(z) \\ r_k^{3,1}(z) & r_k^{3,2}(z) & r_k^{3,3}(z) & \cdots & r_k^{3,n}(z) \\ \vdots & \vdots & \vdots & & \vdots \\ r_k^{n,1}(z) & r_k^{n,2}(z) & r_k^{n,3}(z) & \cdots & r_k^{n,n}(z) \end{bmatrix} \quad (33)$$

- Model AR(1)

$$\mathbf{M}_k = \{\Phi\}_1 \mathbf{M}_{k-1} \quad \text{for} \quad k > 0 \qquad (34)$$

- Model AR(2)

$$\mathbf{M}_k = \{\Phi\}_1 \mathbf{M}_{k-1} + \{\Phi\}_2 \mathbf{M}_{k-2} \quad \text{for} \quad k > 0 \quad (35)$$

In both cases, the calculation begins from the cross correlation sample matrix with order zero, $\mathbf{M}_0$. When the matrix $\mathbf{M}_1$ is evaluated in the model AR(2), we must have in mind that

$$\mathbf{M}_{-k} = \mathbf{M}_k^T$$

The Eqs. (34) and (35) are the application the Yule-Walker equations under a multivariate focus.

Modeling of the residual series

The selection of the best generation model is partially based on a group of adjustment tests over the residual series. MASHWIN obtains these series from the models previously calibrated and calculates the statistics required for those tests, so that the modeler has all this information to take his decision.

The residual series are obtained from the matrix equations correspondent to the models, as the Eqs. (36) to (38) indicate.

- Model AR(1)

$$\{\varepsilon\}_t = [\Theta]_0^{-1}\left(\{Z\}_t - [\Phi]_1\{Z\}_{t-1}\right) \quad (36)$$

- Model AR(2)

$$\{\varepsilon\}_t = [\Theta]_0^{-1}\left(\{Z\}_t - [\Phi]_1\{Z\}_{t-1} - [\Phi]_2\{Z\}_{t-2}\right) \quad (37)$$

- Model ARMA(1,1)

$$\{\varepsilon\}_t = [\Theta]_0^{-1}\left(\{Z\}_t - [\Phi]_1\{Z\}_{t-1} + [\Theta]_1\{\varepsilon\}_{t-1}\right) \quad (38)$$

After calculating the residual series, the program evaluates the average, standard deviation and skew coefficient of each one, (Eqs. 39), and the skew normality interval, (expression 8).

$$\bar{\varepsilon} = \frac{\sum_{t=1}^{N}\varepsilon_t}{N} \qquad s_\varepsilon = \sqrt{\frac{\sum_{t=1}^{N}\left(\varepsilon_t - \bar{\varepsilon}\right)^2}{N-1}} \qquad g_\varepsilon = \frac{N\sum_{t=1}^{N}\left(\varepsilon_t - \bar{\varepsilon}\right)^3}{(N-1)(N-2)s_\varepsilon^3} \quad (39)$$

Where: $\bar{\varepsilon}$: Average of the residual series.

ε_t: Residue from the period t.

N: Total number of residues from the series.

s_ε: Standard deviation of the residual series.

g_ε: Skew coefficient of the residual series.

Afterwards, the program calculates the autocorrelation functions and the limits of the temporal independence interval of the residual series, according to the expressions (40) and (18) respectively.

$$r_k(\varepsilon) = \frac{\sum_{t=1}^{N-k} \left(\varepsilon_{t+k} - \bar{\varepsilon}\right)\left(\varepsilon_t - \bar{\varepsilon}\right)}{\sum_{t=1}^{N} \left(\varepsilon_t - \bar{\varepsilon}\right)^2} \qquad (40)$$

Where $r_k(\varepsilon)$ is the autocorrelation coefficient for a temporal interval k in the residual series.

Finally, the program calculates the cross correlation matrixes from the residual series (41) for the same number of temporal intervals used in the calculation of the autocorrelation functions.

$$\mathbf{M}_k(\varepsilon) = \begin{bmatrix} r_k^{1,1}(\varepsilon) & \cdots & r_k^{1,n}(\varepsilon) \\ \vdots & & \vdots \\ r_k^{n,1}(\varepsilon) & \cdots & r_k^{n,n}(\varepsilon) \end{bmatrix} \quad r_k^{ij}(\varepsilon) = \frac{\sum_{t=1}^{N-k} \left(\varepsilon_{t+k}^{(i)} - \bar{\varepsilon}^{(i)}\right)\left(\varepsilon_t^{(j)} - \bar{\varepsilon}^{(j)}\right)}{\sqrt{\sum_{t=1}^{N} \left(\varepsilon_t^{(i)} - \bar{\varepsilon}^{(i)}\right)^2 \sum_{t=1}^{N} \left(\varepsilon_t^{(j)} - \bar{\varepsilon}^{(j)}\right)^2}} \qquad (41)$$

The statistics in Eqs. (39) to (41) are required to make the adjusment tests of the stochastic model. The program presents most of the information required at the results files, so that the modeler can review those tests, as it is now described.

- THE AVERAGE EQUAL TO ZERO. The averages of the residual series must be statistically equal to zero. To verify this test, the modeler must calculate the limits of the confidence interval for the hypothesis test referred to the equality of the average to a certain value (in this case zero). For a confidence level of 95%, the limits of that interval are the ones given by the expression (42).

$$\left[-1.96\frac{\sigma}{\sqrt{N}},\ 1.96\frac{\sigma}{\sqrt{N}}\right] \qquad (42)$$

Where: σ: Standard deviation of the residual series.

N: Number of residues. It is the same for all of the series.

- EQUALITY OF THE STANDARD DEVIATION TO ONE. The standard deviation of the residual series must be statistically equal to one. To verify this test, the modeler must calculate the limits of the confidence interval for the hypothesis test referred to the equality of the standard deviation to a certain value (in this case one), for example the test Chi-square.

- NORMALITY OF THE RESIDUAL SERIES. The probability distribution of the residual series must be normal. To review this condition, the program calculates the limits of the confidence interval with 95% probability, Eq. (8), in which the obliqueness coefficient must be so that we can statistically consider that the series has a normal distribution.

- TEMPORAL INDEPENDENCE OF THE RESIDUAL SERIES. The residual series must not be correlated in time. To make this test (Anderson Test), we use the results of the expressions (40) and (18). If the residues of the series are inside the Anderson interval, that series are considered to be independent in time.

- SPATIAL INDEPENDENCE OF THE RESIDUAL SERIES. The residual series must not be correlated in space. The modeler must test this verifying that the elements from the cross correlation matrix with order zero are inside the limits of the spatial independence confidence interval, for a given probability level, (1-α), according to (43) (Jenkins y Watts, 1969, quoted by Salas et al., 1980).

$$\left[-u_{1-\alpha/2}/\sqrt{N},\ u_{1-\alpha/2}/\sqrt{N}\right] \qquad (43)$$

Where: $u_{1-\alpha/2}$: Quantile of the typified normal distribution.
N: Number of residues.

Once the tests are made, the modeler has to decide which one of the three calibrated stochastic models will be finally adopted for the later generation of the synthetic flows series.

2.5.2. Spatial segregation model

One of the models available in MASHWIN is the LANE SPATIAL SEGREGATION model (Lane, 1979; reference in Salas et al., 1980), based on the segregation model from Valencia and Schaake (1973, reference in Salas et al., 1980).

This model is especially suitable for multivariate problems with a great number of stations, in other words, cases in which there would be a great number of parameters if they were approached with only one ARMA type multivariate modeling.

Formulation of the model

This is a linear model which gives out series of flows (Y) in a group of *secondary stations*, from a group of flows (X) coming from several *principal stations*. Historic series of flows with the same number of data in all the stations are required for the calibration of the model. They must also correspond to the same period. The Eq. (44) shows the formulation of the Lane model.

$$\{Y\}_t = [\mathbf{A}]\{X\}_t + [\mathbf{B}]\{\varepsilon\}_t + [\mathbf{C}]\{Y\}_{t-1} \qquad (44)$$

Where: Y: Vector of the flows from the secondary stations.
X: Vector of the flows from the principal stations.
ε: Vector of random, normally distributed values with zero average, unitary standard deviation and independent in time.
A, **B** and **C:** Matrixes of parameters.

The model can work at a monthly or annual scale.

The Lane spatial segregation model is designed to preserve the averages and the variances of the historic series in the newly generated series, as well as the zero-order correlations in the new series.

Estimation of parameters

The parameters of the segregation models are usually estimated using the momentums method, which is simpler and faster than other techniques (maximum verisimilitude, for example), since it is not recurrent to iterative processes and it does not give out implicit solutions. This characteristic makes the momentums method strongly attractive when selecting a method for the estimation of parameters, even although it does not provide the best estimators. There are more sophisticated techniques which can produce better results in this sense, but their practical application to the segregation models is not completely proved yet. At any rate, it is a fact that the momentum method, which is the one implemented in the program, is precise enough for segregation purposes, since a great part of the stochastic nature of the series to be generated has already been preserved automatically in the previous generation of the series from the principal stations (Salas et al., 1980). In MASHWIN, this is made with ARMA models.

The parameter matrixes for the model are given as a function of the covariance matrixes of the sample series, as it is indicated in the group of Eqs. (45).

$$\mathbf{A} = \left[\mathbf{S}_{YX} - \mathbf{S}_{YY}(1)\mathbf{S}_{YY}^{-1}\mathbf{S}_{XY}^{T}(1)\right]\left[\mathbf{S}_{XX} - \mathbf{S}_{XY}(1)\mathbf{S}_{YY}^{-1}\mathbf{S}_{XY}^{T}(1)\right]^{-1}$$

$$\mathbf{C} = \left[\mathbf{S}_{YY}(1) - \mathbf{A}\mathbf{S}_{XY}(1)\right]\mathbf{S}_{YY}^{-1} \tag{45}$$

$$\mathbf{B}\mathbf{B}^{T} = \mathbf{S}_{YY} - \mathbf{A}\mathbf{S}_{XY} - \mathbf{C}\mathbf{S}_{YY}^{T}(1)$$

Where $S_{UW}(1)$ (the letters U and W are only for illustration purposes) is the covariance matrix between the U and W series. There is a time interval of retard between U and W series. When W is not retarded (k=0), the covariance matrix is simply denoted as S_{UW}.

The i,j element (row i and column j) from the covariance matrix is calculated as indicated in the Eq. (46).

$$S_{U_t W_{t-k}}(i,j) = \mathrm{Cov}\left(U_t^{(i)}, W_{t-k}^{(j)}\right) = \frac{1}{N-1-k}\sum_{t=1+k}^{N}\left(U_t^{(i)} - \overline{U}_t^{(i)}\right)\left(W_{t-k}^{(j)} - \overline{W}_{t-k}^{(j)}\right) \tag{46}$$

with

$$\overline{U}_t^{(i)} = \frac{1}{N-k}\sum_{t=1+k}^{N} U_t^{(i)} \qquad \text{and} \qquad \overline{W}_{t-k}^{(j)} = \frac{1}{N-k}\sum_{t=1+k}^{N} W_{t-k}^{(j)}$$

Where:

Ut(i): Normalized and typified flow in the station i and for the period t (month or year).

Wt(j): Normalized and typified flow in the station j and for the period t (month or year).

N: Total number of stations from the series.

We must note that, as in the ARMA models, in this model the program works with the normalized and typified flows.

Likewise, the residual series are modeled similarly to ARMA models: they are calculated according to the Eq. (47).

$$\{\varepsilon\}_t = [\mathbf{B}]^{-1}\left(\{Y\}_t - [\mathbf{A}]\{X\}_t - [\mathbf{C}]\{Y\}_{t-1}\right) \quad (47)$$

For these series, the program calculates their averages, standard deviations, skew coefficients, Snedecor and Cochran limits (to evaluate the normality of the residual series), auto-correlation functions and Anderson limits (for the temporal independence test). This is made in order to perform the same tests of adjustment that are made with the ARMA-type models, in case that the modeler considers it necessary. Salas et al. (1980) appoint that for the segregation models this type of tests are 'weak and low valuable'.

2.5.3. Temporal segregation condensed model

MASHWIN also has a temporal segregation model, i.e. a model to segregate series of annual flows in their respective series of monthly flows. In particular, the program works with the LANE TEMPORAL SEGREGATION CONDENSED MODEL (Lane, 1979. reference in Salas et al., 1980), which is also based in the Valencia and Schaake model (reference in Salas et al., 1980), and it is basically a summarized version of it, with a drastic reduction of parameters.

Formulation of the model

This one is also a linear model in which the flow in one month, Y_τ, is calculated as a function of the flow in the previous month, $Y_{\tau-1}$, the flow in the year t, X_t and an random component ε_τ.

$$\{Y\}_\tau = [\mathbf{A}]_\tau \{X\}_t + [\mathbf{B}]_\tau \{\varepsilon\}_\tau + [\mathbf{C}]_\tau \{Y\}_{\tau-1} \quad (48)$$

Equation (48) is applied to calculate the flow for each month τ, in other words, there is a total of 12 equations or, which is the same, there are 12 groups of parameter matrixes, A τ, B τ and C τ (τ =1, 2,...,12). It is evident that this model, unlike the two previous ones, has a greater amount of parameters, but even so, among the segregation models it is the one which requires the least amount of parameters; and that is why it is called condensed.

For its calibration, the model requires series of annual flows with their correspondent monthly values.

This model preserves the covariances between the annual values and their respective monthly values, and the variances and covariances with gap 1 (k=1) between the monthly values.

Estimation of parameters

As in the spatial segregation model, in the temporal segregation model the program calculates the parameter matrixes as a function of the covariance matrixes of the normalized and typified flow series. Equations (49) are the result of the application of the momentum method in the parameter estimation process.

$$\mathbf{A}_\tau = \left[\mathbf{S}_{YX}(\tau,\tau) - \mathbf{S}_{YY}(\tau,\tau-1)\mathbf{S}_{YY}^{-1}(\tau-1,\tau-1)\mathbf{S}_{YX}(\tau-1,\tau)\right]$$
$$\left[\mathbf{S}_{XX} - \mathbf{S}_{XY}(\tau,\tau-1)\mathbf{S}_{YY}^{-1}(\tau-1,\tau-1)\mathbf{S}_{YX}(\tau-1,\tau)\right]^{-1}$$

$$\mathbf{C}_\tau = \left[\mathbf{S}_{YY}(\tau,\tau-1) - \mathbf{A}_\tau \mathbf{S}_{XY}(\tau,\tau-1)\right]\mathbf{S}_{YY}^{-1}(\tau-1,\tau-1) \quad (49)$$

$$\mathbf{B}_\tau \mathbf{B}_\tau^T = \mathbf{S}_{YY}(\tau,\tau) - \mathbf{A}_\tau \mathbf{S}_{XY}(\tau,\tau) - \mathbf{C}_\tau \mathbf{S}_{YY}(\tau-1,\tau)$$

The Eqs. (50) to (54) present the way in which the program calculates the elements from the covariance matrixes. The superscripts i and j represent two different stations (remember that in the program the modeling is multivariate). The letter Y will always correspond to the monthly series and the letter X to the annual series.

$$S_{YX}^{ij}(\tau,\tau) = \mathrm{Cov}\left(Y_\tau^{(i)}, X^{(j)}\right) = \frac{1}{N-1}\sum_{t=1}^{N}\left(Y_{\tau,t}^{(i)} - \overline{Y}_\tau^{(i)}\right)\left(X_t^{(j)} - \overline{X}^{(j)}\right) \quad (50)$$

$$S_{YY}^{ij}(\tau,\tau-1) = \mathrm{Cov}\left(Y_\tau^{(i)}, Y_{\tau-1}^{(j)}\right) = \frac{1}{N-1}\sum_{t=1}^{N}\left(Y_{\tau,t}^{(i)} - \overline{Y}_\tau^{(i)}\right)\left(Y_{\tau-1,t}^{(j)} - \overline{Y}_{\tau-1}^{(j)}\right) \quad (51)$$

$$S_{XY}^{ij}(\tau,\tau-1) = \mathrm{Cov}\left(X^{(i)}, Y_{\tau-1}^{(j)}\right) = \frac{1}{N-1}\sum_{t=1}^{N}\left(X_t^{(i)} - \overline{X}^{(i)}\right)\left(Y_{\tau-1,t}^{(j)} - \overline{Y}_{\tau-1}^{(j)}\right) \quad (52)$$

$$S_{XX}^{ij} = \mathrm{Cov}\left(X^{(i)}, X^{(j)}\right) = \frac{1}{N-1}\sum_{t=1}^{N}\left(X_t^{(i)} - \overline{X}^{(i)}\right)\left(X_t^{(j)} - \overline{X}^{(j)}\right) \quad (53)$$

$$S_{YY}^{ij}(\tau,\tau) = \mathrm{Cov}\left(Y_\tau^{(i)}, Y_\tau^{(j)}\right) = \frac{1}{N-1}\sum_{t=1}^{N}\left(Y_{\tau,t}^{(i)} - \overline{Y}_\tau^{(i)}\right)\left(Y_{\tau,t}^{(j)} - \overline{Y}_\tau^{(j)}\right) \quad (54)$$

The program calculates the rest of the covariance matrixes as a function of the previous ones, as it shown in the Eqs. (55) to (57).

$$\mathbf{S}_{YX}(\tau-1,\tau) = \left[\mathbf{S}_{XY}(\tau,\tau-1)\right]^T \quad (55)$$

$$\mathbf{S}_{XY}(\tau,\tau) = \left[\mathbf{S}_{YX}(\tau,\tau)\right]^T \quad (56)$$

$$\mathbf{S}_{YY}(\tau-1,\tau) = \left[\mathbf{S}_{YY}(\tau,\tau-1)\right]^T \quad (57)$$

In the Eqs. (50) and (54),

$$\overline{Y}_\tau = \frac{1}{N}\sum_{t=1}^{N} Y_{\tau,t}$$

In the Eq. (51),

$$\overline{Y}_\tau = \frac{1}{N-1}\sum_{t=2}^{N} Y_{\tau,t}$$

In the Eqs. (51) and (52),

$$\overline{Y}_{\tau-1} = \frac{1}{N-1}\sum_{t=2}^{N} Y_{\tau-1,t-1}$$

In the Eqs. (50) and (53),

$$\overline{X} = \frac{1}{N}\sum_{t=1}^{N} X_t$$

In the Eq. (52),

$$\overline{X} = \frac{1}{N-1}\sum_{t=2}^{N} X_t$$

In the previous equations, N is the number of years of the series.

3. BRIEF DESCRIPTION OF THE ORGANIZATION AND FUNCTIONING OF THE MODEL.

Each analysis will be defined as a project, to which the user will supply a historical series of hydrological inflows to be analyzed with the model. To keep the standards, the file with the inflow series must have the format used in the SIMGES model for simulation of water resources management.

The program will create a directory with the name of the project. This directory will hold all the data generated under it, including the studied alternatives. The information will be ordered as a tree of files, in such manner that each level (understanding by level an isolated part of the analysis) will hold a phase of the work with all of the partial data from it. This tree will have the level structure described next:

Level 0: This is the parent directory, defined by the name of the project. The inflows file to be analyzed will be placed here.

Level 1: The series selected for analysis are located at this level, together with the information on its statistical analysis. The subfolders at this level will hold the analysis performed on each set of stations selected from the original inflows file.

Level 2: This level represents the normalization method. There will be a subdirectory for each type of normalization studied for the group of series in the previous level.

Level 3: This level holds the results of the Fourier treatment. For the analysis of annual series, this step also calculates the standardized series.

Level 4: The tests of distributed ARMA models and Lane spatial segregation are included at this level. This analysis is performed using the standardized series of the previous level.

Level 5: The last level holds the synthetic series and their analysis. The model will generate synthetic series and it will evaluate their statistics.

4. GRAPHIC INTERFACE.

The user interface controls the access to the analysis of results through a browser similar to the Windows explorer directories tree, located at the left margin of the screen. It allows the user to navigate between the different alternatives of the analysis. When the user selects an element from the tree, the program loads the options at the correspondent level and a text showing the written report on the option selected. The user can also access to a great variety of graphs related to the analysis selected.

4.1. CONTENT OF THE SCREENS AT THE DIFFERENT LEVELS.

4.1.1. Level 0.

When the user begins the work he must define a “new project” or load one already created. A name must be supplied to the new project, as well as the path and name of the inflows data file.

Once these data are available, the program creates a directory for the project, named with the 8 first characters of the project name, and the series of inflows to analyze is copied on it.

A file "bdproyecto.mdb" is created in this directory, containing the names and paths of the files with the analysis alternatives (it is a ‘historical’ record or working ‘index’). This file has the format of a Microsoft Access data base.

In the screen we can visualize a summary of the inflows file and the list of the inflow series included in the file. There is also the option of visualizing the file itself.

The user must select the series to use and press the button “Analyze” to begin the statistic analysis. It is also possible to obtain graphs on the data series, monthly or annually.

The system will automatically generate a name in the control tree which identifies the analyzed series. When this element is created in the tree, the system will pass the control to level 1.

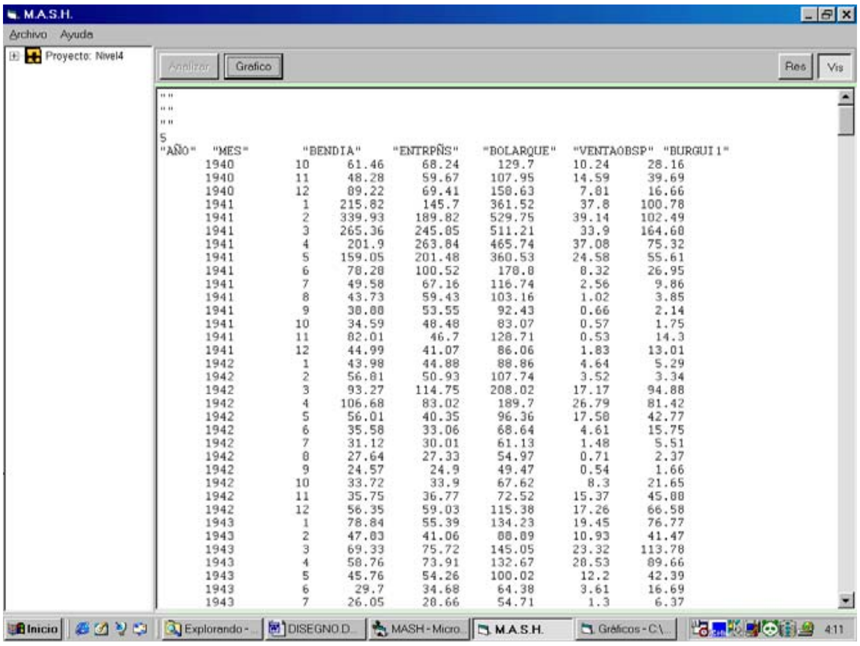

Figure 1: MASHWIN screen using the visualize data option.

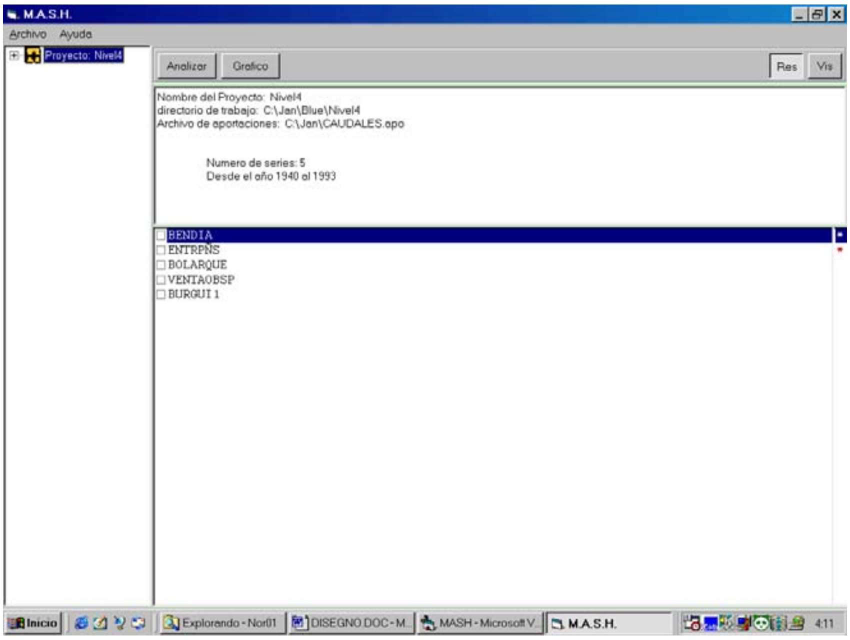

Figure 2: MASHWIN screen at Level 0

4.1.2. Level 1.

Once that a group of series have been defined for the analysis, the program creates a subdirectory (level 1) with a new inflows file containing only the inflows to be analyzed, and starts the module which calculates the average, variance, skew, etc., both for monthly and annual values. When the calculations are finished, the system will issue and display a written report.

The interface allows graphing the results of monthly average, standard deviation and skew, either separately or grouped in any combination. Also, it gives the option to graph simultaneously the results of different hypothesis at the same level.

In this phase it is also possible to continue the analysis with the annual series. The following phases of analysis will differ according to the selection of either monthly or annual series.

The user at this level will decide the type of normalization to apply in the following phase. When a normalization method is selected, a level 2 is automatically generated.

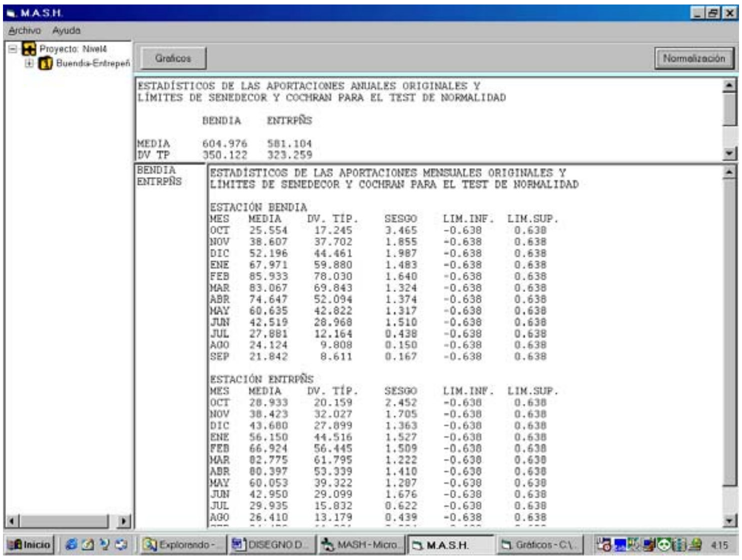

Figure 3: Aspect of the MASHWIN screen in level 1.

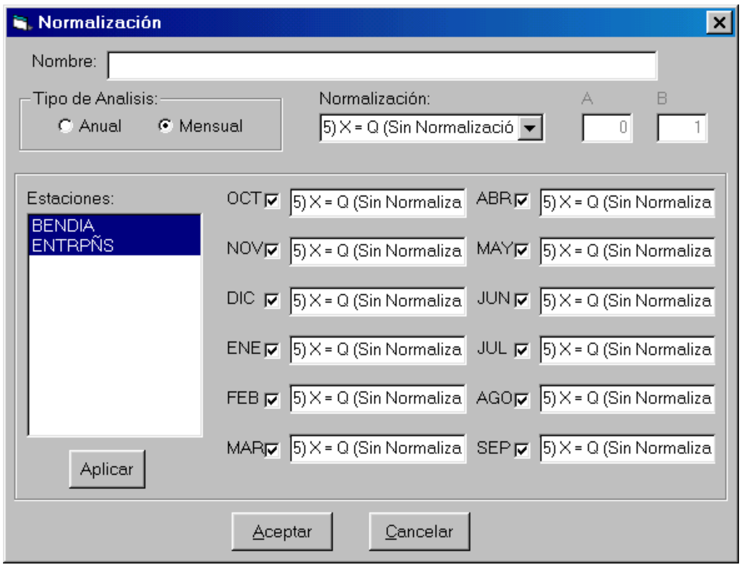

Figure 4: Window with the data to fill to apply the Normalization.

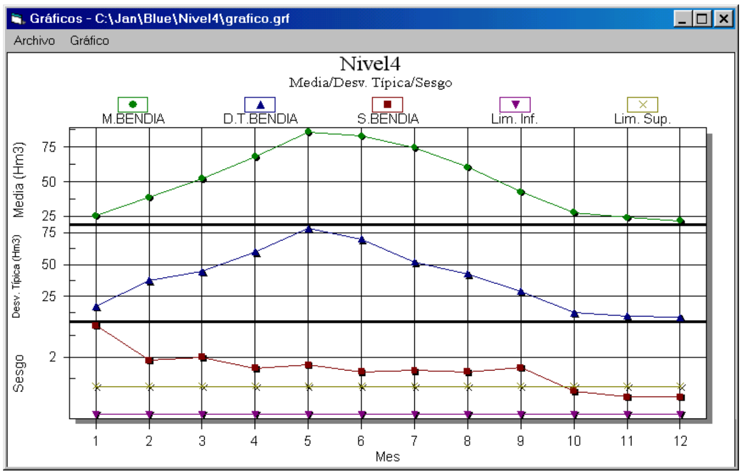

Figure 5: Graphs of the statistic results (Average, variance and skew) from one of the analyzed series

4.1.3. Level 2.

A level 2 is created for the first time when a new type of normalization is selected in level 1. A new series of ‘normalized inflows’, is generated and written at this level. At the same time, the system will start the module which will recalculate all the parameters obtained at level 1, but with the new series (if the user has decided not to make normalization, it will just copy the data from the previous directory). If it is an annual analysis, the annual data series will be generated.

The utility of graphs in this phase will be the same as in the previous phase, but using the normalized data.

The next phase will be the Fourier analysis, which will give access to level 3, generating a subdirectory for each election of Fourier parameters. It will also do it if the decision is not to make the Fourier treatment.

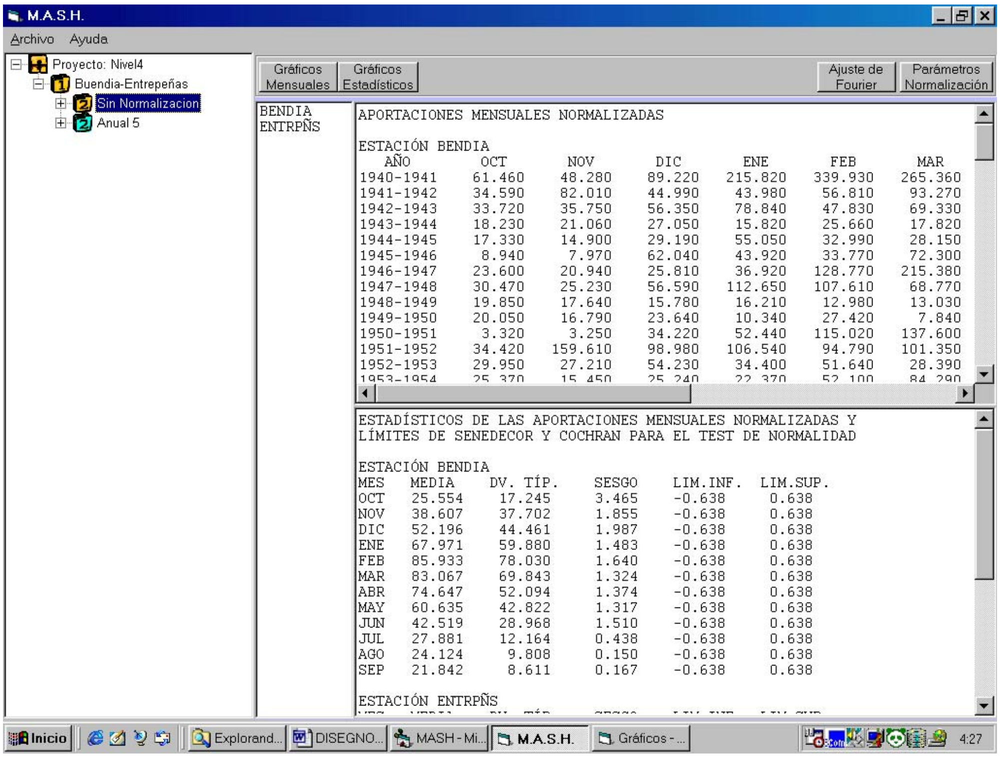

Figure 6: MASHWIN screen in level 2. Monthly analysis.

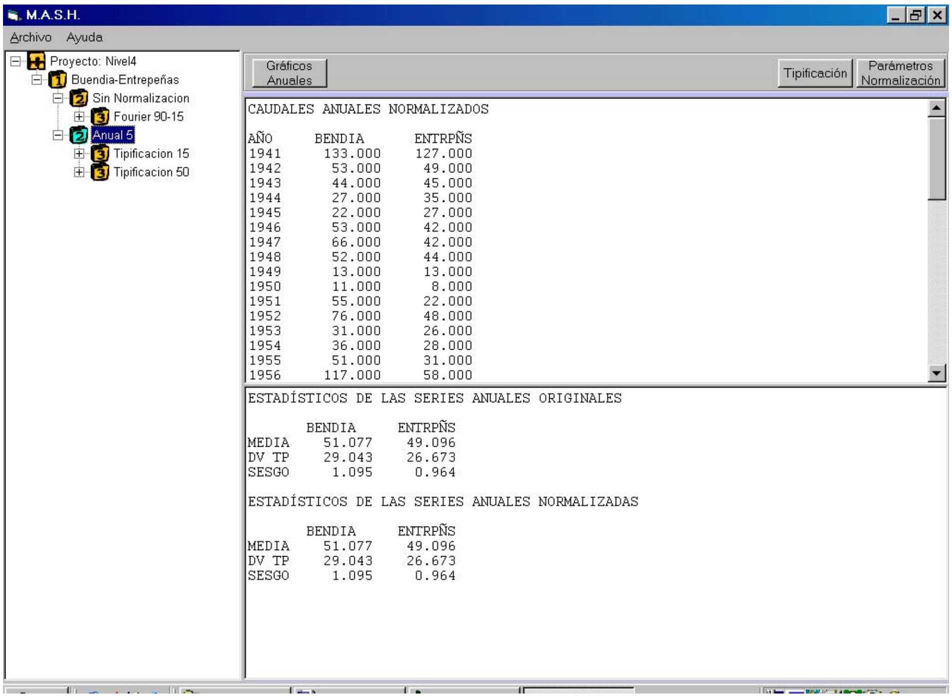

Figure 7: MASHWIN screen in level 2. Annual analysis.

4.1.4. Level 3.

The access to level 3 starts the calculation of the Fourier averages and variances, and with these parameters, the model generates the standardized data series. If the Fourier adjustment is not demanded, the standardization will be made with the sample averages and variances.

From the standard series the model will calculate all the auto-correlation and crossed-correlation parameters.

Inside this level both the normalized series and the statistics can be graphed (both for the monthly and the annual options). It will also add the correlation parameters to the graphs.

With this information, the user can go on to level 4, defining the ARMA and Lane models to study.

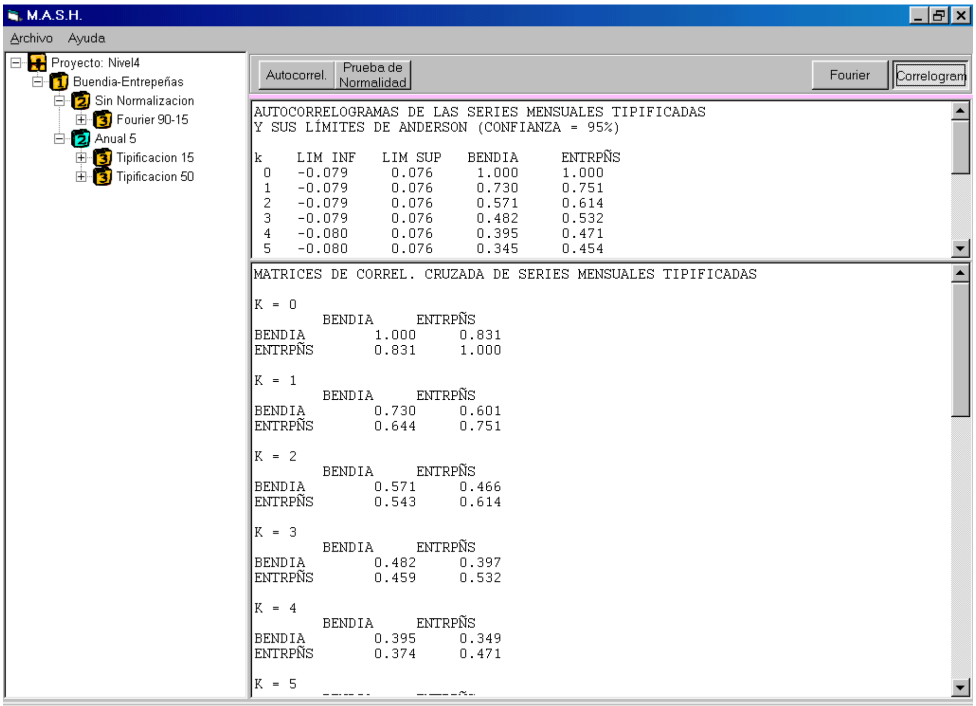

Figure 8: MASHWIN screen in level 3. Correlation results

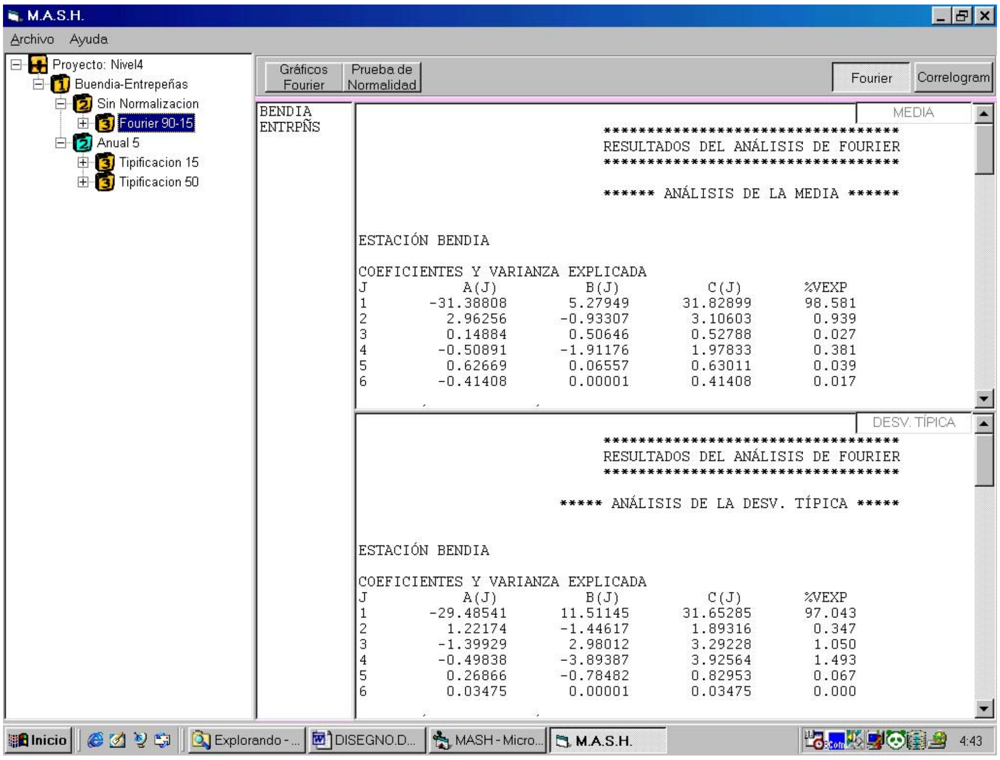

Figure 9: MASHWIN screen in level 3. Fourier results

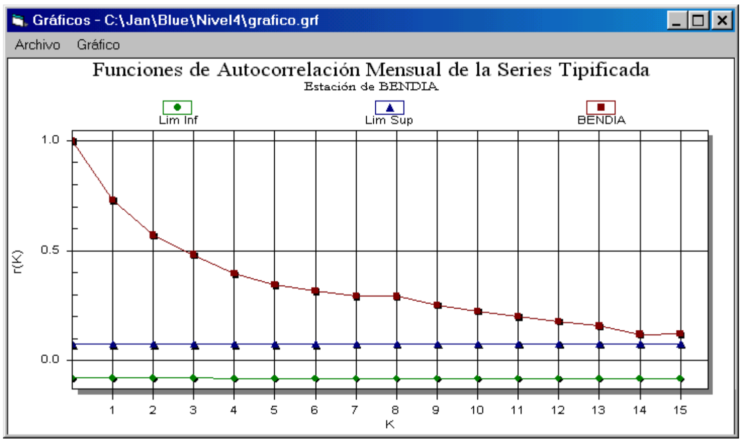

Figure 10: Auto-correlation graph

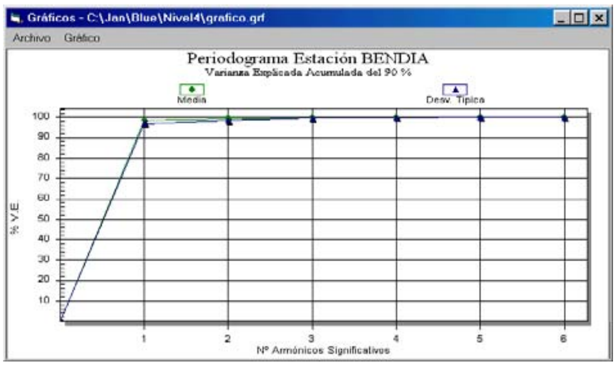

Figure 11: Fourier adjustment graph.

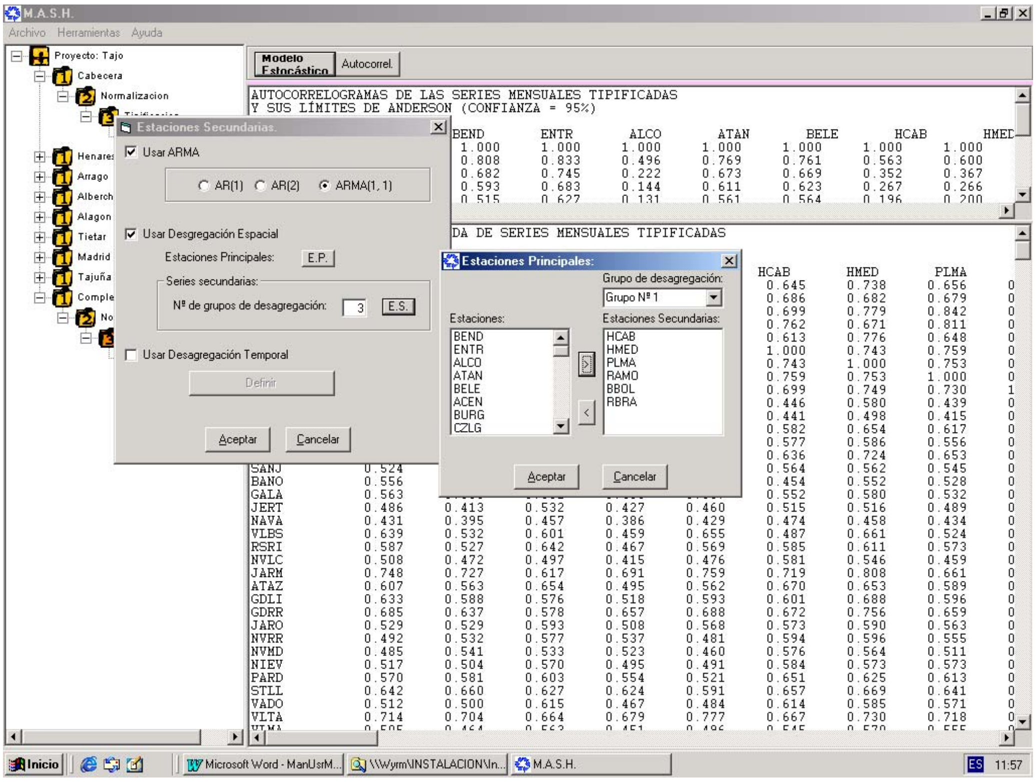

Figure 12: Window for the definition of the stochastic model.

4.1.5. Level 4.

The access to this level requires the introduction of dimensions ‘p’ and ‘q’ for the ARMA (p,q) model, and the identification of the subgroups of principal (ARMA model) and secondary stations (LANE model, in the case that it is used). Once defined, the program will calculate the model parameters, issuing a report with the calculation results.

The residuals of the historical series will be calculated with the model parameters. All the statistics (averages, variances, correlation coefficients...) already calculated in the previous levels for the original series are now calculated for the residuals.

Once the user has selected a model (or several), the last study phase will begin, going on to level 5, in which the contrast series are generated.

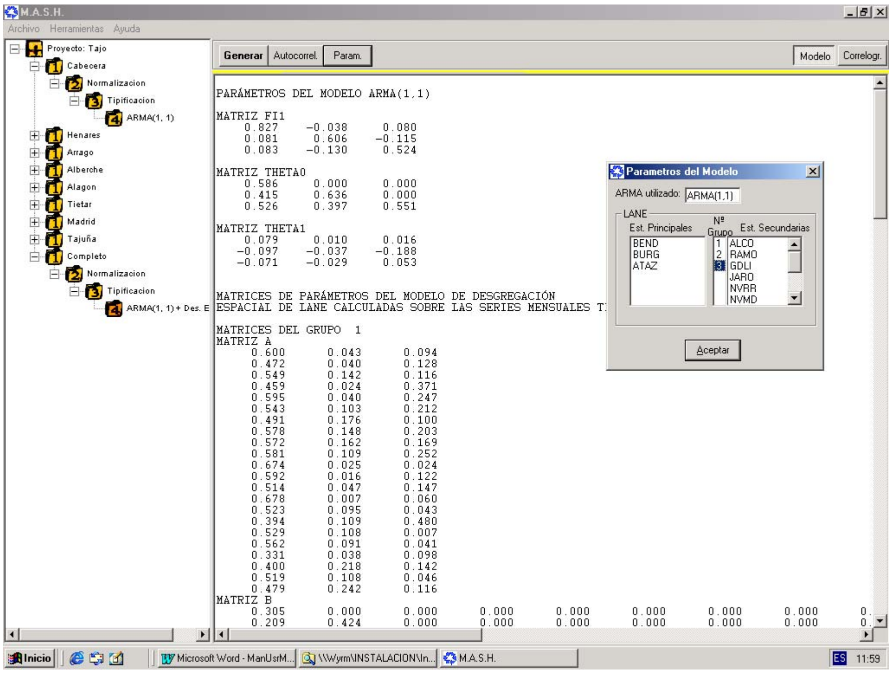

Figure 13: MASHWIN screen in Level 4, with the model parameters.

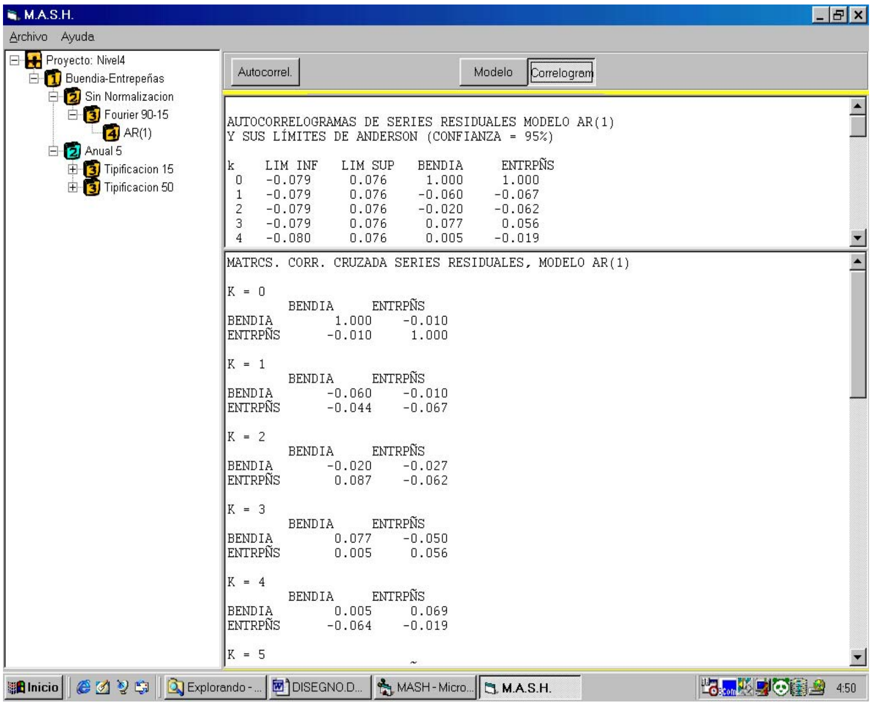

Figure 14: MASHWIN screen in Level 4 showing correlograms of the residuals.

4.1.6. Level 5.

To access this level two data must be introduced: length of the series to generate and number of series to generate (by default the length will be equal to the synthetic series).

With these arguments the interface will start the synthetic series generation module. When the new series are generated, the system will make a new statistic analysis, from which it will issue a complete report containing the evaluation of all the parameters (averages, variances, drought indicators…) and comparing them with the historic series, which have already been calculated at level 1.

The content of this directory will be similar to the previous level, but for the generated series.

4.2. GENERALIZATION.

All the phases described can be generalized using the project browser, which allows the user to return to a previous or parallel phase at any moment, in order to begin the study of a different hypothesis without losing the information of the studies made previously.

Another utility of this model is the option of automatic recalibration of the whole project when the historic flow data are extended. For this there must be an original data file defined in the parent directory.

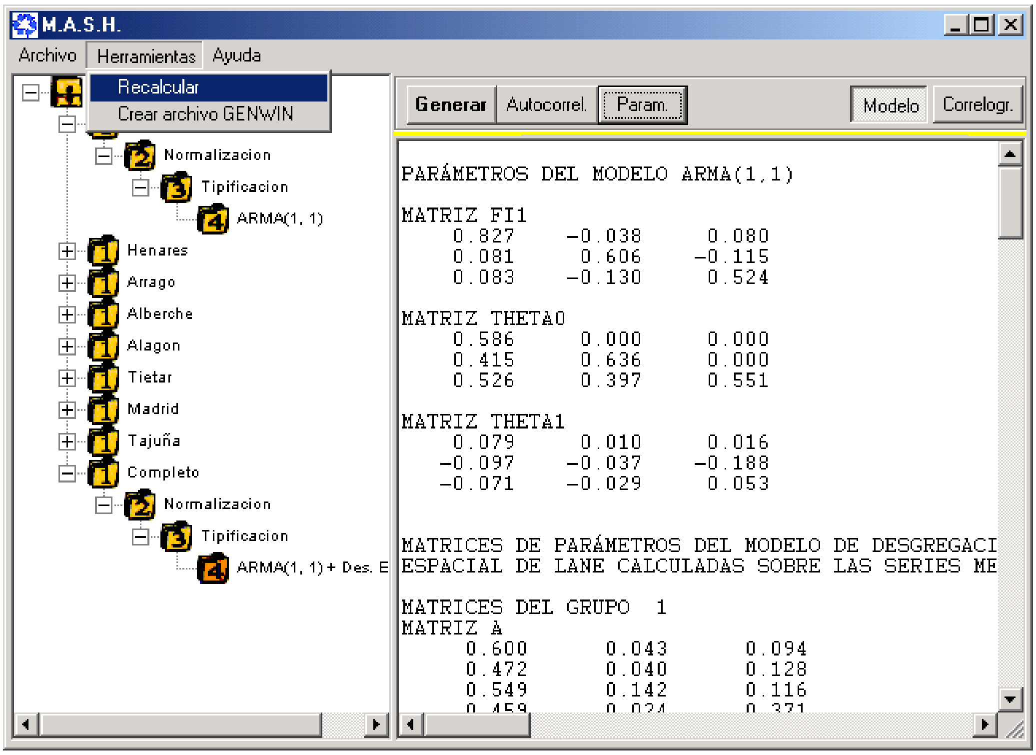

Figure 15: Recalculate option in the MASHWIN model.

4.3. GRAPHS.

For the generation of graphs, an independent interface has been developed that receives the data to graph from a text file. All the graph design can be defined and modified by the user acceding to the graph data file.

5. CALCULATION MODULES

The function of MASHWIN interface is to prompt the user for data and to display and organize the results, while the calculations and analysis described generically at the previous section are made by different applications developed in FORTRAN language. The following section describes the data and result files used by each of the modules in coordination with the graphic interface.

5.1. LEVEL 1: MONTHLY AND ANNUAL ANALYSIS

SUBROUTINES

- MODESAN2: Main program.
- PRNQTAB: Prints a table of monthly flows to a file.
- CALEST2: Calculates and prints monthly statistics.
- ESTADSRA: Calculates statistics of the aggregated annual series starting from the monthly series.
- PRNESQTA: Prints the annual series statistics.
- SEQUIAS: Calculates the drought statistics for as many thresholds as the user requires. These thresholds are expressed as a percentage of the average flow from the flow series. These statistics are: number of droughts, average duration, average intensity, average magnitude, maximum duration, maximum intensity and maximum magnitude. When the series are monthly, the statistics are calculated at a monthly and annual level; and only at annual level when they are annual.
- ALMACNTO: Calculates the rank, the re-scaled rank and the Hurst coefficient for the series at annual level; it also calculates the reservoir volume necessary to satisfy constant demands equal to percentages of the average flow at an annual and monthly level.

DATA FILES

- CAUDALEM.DAT: Monthly flows.
- UMBRALES.DAT: In this file, the number of thresholds and their values are specified, expressed as a fraction of the average flow. These thresholds determine the droughts and the reservoir volumes.

RESULTS FILES

- ANUAL.TXT: Statistics of the annual series.
- MENSUAL.TXT: Statistics of the monthly series.
- SERIES.AP2: Tabulated original monthly flows (includes also the annual flows).
- SEQ-MN.RES: Statistics of the monthly droughts.
- SEQ-ANMN.RES: Statistics of the annual droughts, obtained from the monthly flow series.
- ALM-MN.RES: Water storage statistics of the monthly series.
- ALM-ANMN.RES: Water storage statistics of the annual series, obtained from the monthly flow series.

5.2. LEVEL 2(A): MONTHLY NORMALIZATION

SUBROUTINES

- MODENORM: Main program.
- PRNQNTAB: Prints the normalized flows to a table.
- CALEST2: Calculates and prints the statistics of the normalized flows.

DATA FILES

- ..\CAUDALEM.DAT: Monthly flows. (Located at the immediately superior directory level).

- NORMALIM.DAT: Normalization functions.

The model has four normalization functions and also offers the option of not normalizing the series. The normalization series can differ from a month to another.

The file contains the following information.

- The eight first lines are only informative. The seven first ones never vary. The eighth record will vary if the beginning month of the hydrological year is other than October.
- Following, there are as many lines as stations, each one formed by thirteen fields: the first one is the name of the station and the following twelve contain the indicators of the type of function to use in the normalization, according to the list presented at the beginning of this file (records 2 to 6).
- After that, there are three lines that do not vary and a fourth one which is formed by the months of the year; which will vary only when the starting month is other than October.
- Next there are as many lines as stations, each one having 13 fields: the first one is the name of the station and the others contain the values, for the different months, of the parameter '*a*' for the potential normalization.
- In the following two lines, the first one does not vary, and the second one is again formed by the months of the year.
- After that there are as many lines as stations. Each of them has 13 fields: the first one is the name of the station and the rest contain the values of the parameter '*b*' for the potential normalization for each month.

It is important to note that this file must always have that structure, no matter whether the potential normalization is applied or not. When it is not applied, the values of the parameters *a* and *b* will not be used, since the variables which store those values are only included in the calculations when that normalization is used.

Another important aspect is the order of appearance of the stations both in the section for the indicators of the normalization type, and in the sections in which the values of the parameters *a* and *b* for the potential normalization are specified. This order must be the same as in the flows file.

RESULTS FILES

- QNTAB-MN.RES: Tabulated normalized flows
- ESTQN-MN.RES: Statistics of the normalized flows.
- QNOR-MN.$$$: Tabulated normalized flows. File for internal use of the program, used as data source by the monthly typification module. It records the initial year, initial month, number of stations and number of years.
- MDTNORMN.$$$: Statistics of the normalized flows. File for internal use of the program, used as a data source by the monthly typification module.

5.3. LEVEL 2(B): ANNUAL NORMALIZATION

SUBROUTINES

- MODENORA: Main program.
- ESTADSRA: Calculates the original and normalized annual statistics.
- PRNSTSRA: Prints the statistics of the original and normalized annual flows.

DATA FILES

- ..\CAUDALEA.DAT: Monthly flows (Located at the immediately superior directory level).
- NORMALIA.DAT: Normalization functions.

Their format is similar to the one described for level 2(a)

RESULTS FILES

- QANNOR.RES: Normalized flows.
- ESTADQAN.RES: Statistics of the original and normalized flows.
- MDTNORAN.$$$: Statistics of the normalized flows. File for internal use of the program, used as data source by the annual typification module.
- QNOR-AN.$$$: Tabulated normalized flows. File for internal use of the program, used as data source by the annual typification module. It records the initial year, number of stations and number of flows.

5.4. LEVEL 3(A): MONTHLY TYPIFICATION

SUBROUTINES

- MODETIPM: Main program.
- FOURIER2: Adjustment of the monthly statistics in Fourier series.
- PRNFOU: Prints a part of the Fourier calculations.
- AUTOCOR2: Calculates the auto-correlations.
- CORRCRZ2: Calculates the crossed correlations.
- PRNAUTC: Prints correlograms.
- PRNCOCR: Prints the crossed correlation matrixes.
- QTABULA2: Prints in tabular way the typified flows.

DATA FILES

- ..\QNOR-MN.$$$: Normalized flows. This file is created by the program at the level 2(a) of monthly normalization.
- ..\MDTNORMN.$$$: Statistics of the normalized flows. This file is created by the program at the level 2(a) of monthly normalization.
- VEX.DAT: This file has three records:

    First: Significative percentage of explained variance.

Second: Indicates with Yes (Y) or No (N) if the user wants to typify with statistics adjusted by Fourier.

Third: Number of correlation intervals.

RESULTS FILES

- FOURIER.RES: Accumulated periodograms and statistics before and after being adjusted in Fourier series.
- QTIPTAB.RES: Tabulated typified flows.
- AUTCORMN.RES: Correlograms of the typified series.
- CORCRZMN.RES: Crossed correlation matrixes.
- QTIPIF.$$$: Typified flows. File for internal use of the program, which is used as data source by the models calibration module.
- MATCOCR.$$$: Crossed correlation matrixes. File for internal use of the program, which is used as data source by the models calibration module.

5.5. LEVEL 3(B): ANNUAL TYPIFICATION

SUBROUTINES

- MODETIPA: Main program.
- AUTOCOR2: Calculates the autocorrelations.
- CORRCRZ2: Calculates the crossed correlations.
- PRNAUTC: Prints the correlograms.
- PRNCOCR: Prints the crossed correlation matrixes.

DATA FILES

- ..\QNOR-AN.$$$: Normalized flows. Created by the program at level 2(b) of annual normalization.

- ..\MDTNORAN.$$$: Statistics of normalized flows. Created by the program at level 2(b) of annual normalization.
- LAGSUP.DAT: It has only one record in which the number of correlation intervals is specified.

RESULTS FILES

- QTIP-AN.RES: Typified flows.
- AUTCORAN.RES: Correlograms of the typified series.
- CORRCRAN.RES: Crossed correlation matrixes.
- QTIPIF.$$$: Typified flows. File for internal use of the program, which is used as data source by the models calibration module.
- MATCOCR.$$$: Crossed correlation matrixes. File for internal use of the program, which is used as data source by the models calibration module.

5.6. LEVEL 4: MONTHLY AND ANNUAL CALIBRATION

SUBROUTINES

- MODELOS2: Main program. It calibrates the models AR(1), AR(2), ARMA(1,1), the LANE SPATIAL SEGREGATION model and the LANE TEMPORAL SEGREGATION CONDENSED model.
- ARARMA2: Adjusts the models AR(1), AR(2) and ARMA(1,1).
- SEPARAR2: Separates the series of the main stations from those of the secondary stations.
- DESGRSPC: Calibrates the Lane spatial segregation model.
- CHOLESKY: Solves the decomposition of the grammian matrixes.
- CORRCRZ2: Calculates the crossed correlation matrixes.
- ESTADSRA: Calculates the statistics of the residual series. It is also used to calculate the statistics of the annual series.
- INVERSA: Calculates the inverse of a matrix.

- AUTOCOR2: Calculates the auto-correlation functions.
- PRNAUTC2: Prints the auto-correlation functions to a file.
- COVARZA: Calculates the covariance matrixes.
- PRNCOCR2: Prints the crossed correlation matrixes to a file.
- PRNSTSR2: Prints to a file the statistics of the residual series.
- DSGRTMP2: Calibrates the Lane temporal segregation model.

DATA FILES

- MODELOS.DAT: It has three records.

  First: Indicates with Yes (Y) or No (N) if the ARMA group model is calibrated.

  Second: Indicates with Yes (Y) or No (N) if the spatial segregation model is calibrated.

  Third: Indicates with Yes (Y) that the temporal segregation model is calibrated (in the case that the modeling is performed on a monthly basis) and with No (N) that it is not calibrated (when working with annual modeling).

  The program can calibrate in one only execution any of the following schemes at monthly level:

  - Only ARMA (AR(1), AR(2) or ARMA(1,1)).
  - Only LANE SPATIAL SEGREGATION (DSPC).
  - Only LANE TEMPORAL SEGREGATION (DTMP).
  - ARMA and DSPC.
  - ARMA and DTMP.
  - DSPC and DTMP.
  - ARMA+DSPC+DTMP.

  The program can calibrate in one only execution any of the following schemes at an annual level:

- Only ARMA (AR(1), AR(2) or ARMA(1,1)).
- Only LANE SPATIAL SEGREGATION (DSPC).
- ARMA and DSPC.

- DATOS.DAT: It has three (or two) records.

  First: Index of the ARMA model to calibrate: AR(1) is 1, AR(2) is 2 and ARMA(1,1) is 3. Only required when calibrating ARMA models, in other words, when the first record in MODELOS.DAT is "y".

  Second: Number of correlation intervals.

  Third: Indicates if the modeling is Monthly (M) or annual (A).

- DGLANES.DAT: Data of the spatial segregation scheme. Only needed when modeling the spatial segregation. This file has the following contents.

  - Number of principal stations.
  - Order of the principal stations in the flows file.
  - Amount of spatial segregation groups.
  - And finally, as many lines as groups of secondary stations, indicating in each one the number of secondary stations in the group and then, at the same line, the order of each of those stations inside the flows file.

- ..\QTIPIF.$$$: Typified flows with two records at the second line: Number of stations and number of years. This file is produced by the program at level 3.

- ..\MATCOCR.$$$: Crossed correlation matrixes, order 0, 1 and 2. This file is produced by the program at level 3.

The following data files are only necessary when the temporal segregation model is calibrated.

- ..\..\NORMALIA.DAT: Normalization data of the annual series.

- ..\..\QNOR-MN.$$$: The program uses this file at this level only to read the initial month and the number of years of the series.

- ..\..\..\CAUDALEM.DAT: Series of original monthly flows.

RESULTS FILES (XX=MN, if the modeling is monthly. XX=AN, if the modeling is annual).

- ARMA-XX.RES: Parameter matrixes of the ARMA model.
- ATCMT-XX.RES: Theoretical correlograms of the ARMA model (only available for AR(1) and AR(2)).
- R-AUTCXX.RES: Correlograms and statistics of the residual series from the ARMA model.
- R-COCRXX.RES: Crossed correlation matrixes of the residual series from the ARMA model.
- R-LANES.RES: Correlograms and statistics of the residual series from the Lane spatial segregation model.
- DSGESPXX.RES: Parameter matrixes from the Lane spatial segregation model.
- DSGTMPAM.RES: Parameter matrixes from the Lane temporal segregation condensed model. There are 12, one per month.

5.7. LEVEL 5: GENERATION OF SYNTHETIC SERIES

This module generates synthetic flows by any of the following schemes:

- Only ARMA (AR(1), AR(2) or ARMA(1,1)) at a monthly or annual level.
- ARMA and LANE SPATIAL SEGREGATION (DSPC) at a monthly or annual level.
- ARMA and LANE TEMPORAL SEGREGATION (DTMP) at annual or monthly scale.
- ARMA+DSPC+DTMP annual to monthly scale.

SUBROUTINES

- MODGENE3: Main program.
- GENMENSU: Manages the whole generation of monthly synthetic series.

- LECDATO3: Reads the general data.
- NRTPQPRV: Normalizes and typifies the previous flows.
- SEPARDAT: Separates the principal series from the secondary ones and their respective data.
- GENARMAM: Generates the monthly synthetic series with the ARMA models.
- GENRESI3: Generates the series of random numbers, normally distributed and non-correlated with average zero and unitary standard deviation.
- GQSARMA: Generates the typified flows (z) of the ARMA models.
- DTDNEIM3: Detypifies and denormalizes the monthly synthetic series, calculates their statistics and prints everything to a file.
- ATCQSINT: Calculates the average autocorrelation functions of the synthetic series.
- CCRQSINT: Calculates the average crossed correlation matrixes of the synthetic series.
- GNLANSM2: Generates the monthly synthetic series by the Lane spatial segregation model.
- GQSLANES: Generates the typified flows by the Lane spatial segregation model.
- LECQSTIP: Reads the typified flows used as source data for generating the synthetic series when it is exclusively made by the Lane spatial segregation model.
- GENANUAL: Manages the whole generation of annual synthetic series.
- GENARMAA: Generates the annual synthetic series with the ARMA models.
- GNLANSA2: Generates the annual synthetic series by the Lane spatial segregation model.
- DTDNEIA2: Detypifies and denormalizes the annual synthetic series, calculates its statistics and prints everything to a file.
- GENANAMN: Generates the monthly synthetic series starting from the annual synthetic series previously generated by the program. This is made by the Lane temporal segregation condensed model.

- GQSLANET: Generates the typified monthly flows from the Lane temporal segregation model, starting from typified synthetic annual flows.
- STADQSAN: Calculates the statistics from the synthetic annual series.
- MMMSTQAS: Calculates the average, maximum and minimum values of the statistics from the annual series generated.
- PRNSTQAS: Prints to a file the statistics from the annual series generated, with their respective average values.
- SEQUIAS: Calculates the average values of the drought statistics for as many thresholds as the user desires. These thresholds are expressed as a percentage of the average flow from the series of flows. Those statistics are: number of droughts, average duration, average intensity, average magnitude, maximum duration, maximum intensity and maximum magnitude. When the series are monthly, the statistics are calculated at a monthly level and at an annual level; when they are annual, only at an annual level.
- ALMACNTO: It calculates the average values of the rank, the re-escaled rank and Hurst coefficient for the series at an annual level, and the average values of reservoir water storage necessary to satisfy constant demands equal to percentages of the average flow at an annual and monthly level.

DATA FILES

- DATGEN.DAT: Contains the general data of the generation scheme: type of model(s) to use in the generation, order of the stations in the spatial segregation scheme, number of generation years, initial year of the synthetic series, initial month, parameter matrixes of the models, previous flows, detipification statistics, denormalization functions and maximum number of negative normalized flows admitted in the generation.
- MOA.DAT: It has two records.

  First: Indicates whether the option of generation by temporal segregation is going to be used (Y) or not (N).

  Second: Indicates whether the generation is monthly (M) or annual (A). When the first registration is Y, the second one does not exist.

- UMBRALES.DAT: The number of thresholds and their values are specified, expressed as a fraction of the average flow. Those thresholds determine the droughts and the reservoir volumes.
- DATSMEN.DAT: Data of the monthly series when the generation is made by annual to monthly segregation. It contains: initial month, previous flows, parameter matrixes (there are 12), detypifying statistics, denormalization functions and maximum number of negative normalized flows admitted in the generation. It is only required when the series are generated with the temporal segregation model.
- QSTIPCOL.GEN: Synthetic flows (annual or monthly) typified at the principal stations. It is only required when the generation is made exclusively by the Lane spatial segregation model.

RESULTS FILES

Only for generation with ARMA models:

Monthly generation:

- QMSNTCOL.RES: Synthetic monthly flows written in columns. One for each series generated.
- QMSNTTAB.RES: Synthetic monthly flows written in tables. One for each series generated.
- QMSNTSRK.RES: Synthetic monthly flows written in SIMRISK format.
- ESTQMSNT.RES: Statistics of all the monthly series generated.
- MEDQMSNT.RES: Average values of the statistics from the synthetic monthly series.
- EQASNTSM.RES: Statistics of the annual series (aggregated from the synthetic monthly series).
- SEQ-MN.RES: Average values of the statistics from the monthly droughts.
- SEQ-ANMN.RES: Average values of the statistics from the annual droughts, obtained from the series of monthly flows.
- ALM-MN.RES: Average values of the storage statistics from the monthly series.

- ALM-ANMN.RES: Average values of the storage statistics from the annual series starting from the monthly series.
- ATCQMSNT.RES: Average autocorrelation functions from the monthly series generated.
- CCRQMSNT.RES: Crossed correlation average matrixes from the monthly series generated.

Annual generation:

- QANSINT.RES: Synthetic annual flows.
- QASNTSRK.RES: Synthetic annual flows written in SIMRISK format.
- ESTQASIN.RES: Statistics from the synthetic annual flows and their respective average values.
- SEQ-AN.RES: Average values of the statistics from the annual droughts.
- ALM-AN.RES: Average values of the storage statistics from the annual series.
- ATCQASIN.RES: Average autocorrelation functions from the synthetic annual flows.
- CCRQASIN.RES: Crossed correlation average matrixes from the annual series generated.

For generation with the ARMA and LANE SPATIAL models (X=M, if the generation is monthly. X=A, if the generation is annual):

- QXSEPCOL.RES: Synthetic flows (monthly or annual) from the principal stations written in columns. One for each series.
- QXSESCOL.RES: Synthetic flows (monthly or annual) from the secondary stations written in columns. One for each series.
- QMSEPTAB.RES: Synthetic monthly flows from the principal stations written in tabular way.
- QMSESTAB.RES: Synthetic monthly flows from the secondary stations written in tabular way.
- QXSEPSRK.RES: Synthetic flows (monthly or annual) from the principal stations written in SIMRISK format.

- QXSESSRK.RES: Synthetic flows (monthly or annual) from the secondary stations written in SIMRISK format.
- ESTQXSEP.RES: Statistics of the series generated in the principal stations.
- ESTQXSES.RES: Statistics of the series generated in the secondary stations.
- MEDQMSEP.RES: Average values of the statistics from the monthly series generated in the principal stations.
- MEDQMSES.RES: Average values of the statistics from the monthly series generated in the secondary stations.
- EQASSMEP.RES: Statistics (and their average values) from the aggregated annual series, starting from the synthetic monthly series, in the principal stations. It only appears when the generation is monthly.
- EQASSMES.RES: Statistics (and their average values) from the aggregated annual series, starting from the synthetic monthly series, in the secondary stations. It only appears when the generation is monthly.
- ATCQXSEP.RES: Average autocorrelation functions from the synthetic series of the principal stations.
- ATCQXSES.RES: Average autocorrelation functions from the synthetic series of the secondary stations.
- CCRQXSEP.RES: Crossed correlation average matrixes from the synthetic series of the principal stations.
- CCRQXSES.RES: Crossed correlation average matrixes from the synthetic series of the secondary stations.

For generation with TEMPORAL SEGREGATION MODEL:

- QMSDTCOL.RES: Monthly flows generated by temporal segregation written in columns. One for each series.
- QMSDTTAB.RES: Monthly flows generated by temporal segregation written in tables. One for each series.
- QMSDTSRK.RES: Monthly flows generated by temporal segregation written in SIMRISK format.
- ESTQMSDT.RES: Statistics of the synthetic monthly flows.

- MESQMSDT.RES: Average values of the statistics from the monthly synthetic series.
- ATCQMSDT.RES: Average autocorrelation functions from the monthly series.
- CCRQMSDT.RES: Crossed correlation average matrixes from the monthly series generated.
- Also, the files previously described: QANSINT.RES, ESTQASIN.RES, ATCQASIN.RES, CCRQASIN.RES.

BIBLIOGRAPHY

Bras, Rafael L. and Rodríguez-Iturbe, Ignacio [1985]. *Random Functions in Hydrology*. Addison-Wesley Publishing Company. Massachusetts.

Haan, Charles T. [1977]. *Statistical Methods in Hydrology*. The Iowa University Press Ames. Iowa.

Helsel, D. R. and Hirsch, R. M. [1992]. *Statiscal Methods in Water Resources*. ELSEVIER. Amsterdam

Hipel, K. W. and Mc Leod, A. I. [1994]. *Time Series Modeling of Water Resources and Environmental Systems*. ELSEVIER. Amsterdam.

Loucks, Daniel P., Stedinger, Jery R. and Haith, Douglas A. [1981]. *Water Resources System Plannig and Analysis*. Prentice Hall. New Jersey.

Mood, Alexander M., Graybill, Franklin A. and Boes, Duane C. [1974]. *Introduction to the Teory of Statistics*. McGraw-Hill. New York.

Salas, Jose D., Delleur, J. W., Yevjevich, V. and Lane, W. L. [1980]. *Applied Modeling of Hydrologic Time Series*. Water Resources Publications. Littleton, Colorado.