

**Record Number:** 25110  
**Author, Monographic:** Ouarda, T. B. M. J.//Bobée, B.  
**Author Role:**  
**Title, Monographic:** Regional estimation of floods and droughts in the State of Minas Gerais, Brazil  
**Translated Title:**  
**Reprint Status:**  
**Edition:**  
**Author, Subsidiary:**  
**Author Role:**  
**Place of Publication:** Québec  
**Publisher Name:** INRS-Eau, Terre & Environnement  
**Date of Publication:** 2003  
**Original Publication Date:** Septembre 2003  
**Volume Identification:**  
**Extent of Work:** 27  
**Packaging Method:** pages  
**Series Editor:**  
**Series Editor Role:**  
**Series Title:** INRS-Eau, Terre & Environnement, rapport de recherche  
**Series Volume ID:** 619  
**Location/URL:**  
**ISBN:** 2-89146-482-6  
**Notes:** Rapport annuel 2002-2003  
**Abstract:** Numéro de rapport et ISBN fourni par Geneviève le 7 juin 2002. Confirmé par jdb, 20030917  
**Call Number:** R000619  
**Keywords:** rapport/ ok

**REPORT**

**REGIONAL ESTIMATION OF FLOODS AND  
DROUGHTS IN THE STATE OF MINAS GERAIS,  
BRAZIL**

*Research Report No R-619*

*September 2003*

**Regional estimation of floods and droughts in the State of  
Minas Gerais, Brazil**

**T.B.M.J. Ouarda  
B. Bobée**

submitted to :

**A. Robitaille  
SYNEXUS GLOBAL INC.  
1801 McGill College, Suite 1060,  
Montréal, Québec H3A 2N4 Canada**

**NSERC/Hydro-Québec Chair in Statistical Hydrology  
Institut National de la Recherche Scientifique, INRS-ETE  
2800 Einstein, C.P. 7500, Sainte-Foy (Québec) G1V 4C7**

Research Report No. R-619

ISBN : 2-89146-482-6

**September 2003**

## **RESEARCH TEAM**

### **SYNEXUS GLOBAL INC.**

A. Robitaille

### **Chair in Statistical Hydrology INRS-ETE**

Taha B.M.J. Ouarda  
Bernard Bobée

# **Table of contents**

---

<b>TABLE OF CONTENTS .....</b>	<b>4</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>5</b>
<b>CHAPTER 1 : Introduction .....</b>	<b>6</b>
<b>CHAPTER 2 : Evaluation of the method currently used by IGAM.....</b>	<b>8</b>
<b>CHAPTER 3 : Proposition of alternative regional estimation methodologies.....</b>	<b>10</b>
<b>3.1 Brief literature review .....</b>	<b>10</b>
3.1.1 Homogeneous region delineation.....	10
3.1.2 Regional estimation.....	13
<b>3.2 Selection of methods .....</b>	<b>14</b>
<b>3.3 Adaptation of currently available tools and software.....</b>	<b>15</b>
<b>3.4 An alternative methodology for the regional estimation of drought characteristics.....</b>	<b>20</b>
<b>CHAPTER 4 : Conclusions.....</b>	<b>22</b>
<b>CHAPTER 5 : References.....</b>	<b>24</b>

## **Acknowledgments**

---

This study was funded through a contract with SYNEXUS GLOBAL INC., Montréal, Québec, Canada. The authors are grateful to Mr. A. Robitaille for his capable oversight of this work, and for his encouragement and support. The authors would also like to acknowledge the help of Ms. Melissa Breton in this regard.

# 1. Introduction

---

In regional flood/drought frequency analysis, the objective is to estimate flood/drought characteristics, such as the T-year events  $Q_T$  at any location in a given region, mainly at ungauged sites. In the coming sections we will focus on regional flood frequency analysis, without loss of generality. Indeed, regional flood and drought frequency analysis methodologies are very similar and are based on the same principles.

In regional flood frequency analysis, the flood characteristics of ungauged basins are estimated by using relationships between physiographical and meteorological variables and characteristics of the maximum annual floods or the partial duration series of a set of gauged basins with hydrological regimes similar to those of the ungauged basin.

The usual steps of the estimation are:

- (1) Determination of a set of similar basins («homogeneous region»).
- (2) Regional estimation of the flood distribution of the ungauged basin.

Several methods for the determination of homogeneous regions and for regional estimation have been proposed in the literature. The objective of the present report is to critically evaluate the method currently used by the IGAM (Instituto Mineiro de Gestao das Aguas) in the state of Minas Gerais, Brazil, for the regional estimation of floods and droughts. The present report aims also at presenting a brief review of the various methodologies that are available for regional estimation of floods and droughts, and proposing some alternatives to the methodology currently used by IGAM. The report makes also some recommendations concerning the methodologies that could be adopted in the state of Minas Gerais, and concerning future efforts aiming at the development of

regional estimation tools that are up-to-date and that utilise the most prominent and powerful methods that are currently available in the literature.

We also focus in the present report on the potential to use and adapt currently available hydrologic regional estimation tools and software that utilise sophisticated methodologies, rather than develop new tools that may be costly, untested and inappropriate.

## **2. Evaluation of the method currently used by IGAM**

---

IGAM currently uses a procedure for the estimation of drought flows and design flows, that relies on previous experience and on a series of maps that define a set of geomorphological, meteorological and hydrological characteristics for the State of Minas Gerais. A very elaborate set of maps provides a geographic subdivision of the State of Minas Gerais into a large number of small regions for which the characteristics are the same.

One of the most important maps provides a family of iso-lines for which the minimum specific monthly streamflow is constant (corresponding to a return period of 10 years). The second map defines, in a fine resolution, a subdivision of the State into regions for which three characteristics P, D, and I are homogeneous; where P is the precipitation, D is the slope and I is the permeability. These maps provide a relatively refined description of the rainfall regime, the geomorphologic characteristics and the hydrologic regime in each location in the State.

The empirical approach currently used in the State of Minas Gerais allows to obtain a very fast and relatively reliable estimates of drought flows and design flows. However, these values remain rough estimates of the desired streamflows. It is possible to obtain more accurate estimates of these variables by using more sophisticated but widely accepted regional estimation techniques. These new techniques are nowadays standard approaches for the estimation of streamflow characteristics in ungauged locations. It is also possible to use more recent approaches (such as the canonical correlation analysis method) which were shown to lead to an even better accuracy. This aspect will become apparent as we review,

in chapter 3, the state of the art of hydrological regional frequency analysis. Indeed, the approaches presented in chapter 3 are straight forward, easy to use, accepted by the hydrological community all over the world, and are already available in the form of a user friendly software.

Furthermore, it must be mentioned that the maps currently used by IGAM need to be constantly updated with new information (every few years). Indeed, land use parameters, for example, change constantly and this information needs to be fed to the maps. New precipitation information is available every year, and should be used to update the precipitation index map. This allows to avoid any bias resulting from previous wet periods or drought spells that may have taken place in the past and were used in defining the current map. This systematic updating of these maps can be quite costly and unjustified compared to the costs of using other regionalisation approaches as will be presented in chapter 3.

## **3. Proposition of alternative regional estimation methodologies**

---

### **3.1 Brief literature review**

As indicated in chapter 1, The two main steps of a hydrological regional estimation methodology are the identification of groups of hydrologically homogeneous basins (or “homogeneous regions”) and the application of a regional estimation method within each delineated region. These two steps will be considered separately in the following two sections.

#### ***3.1.1 Homogeneous region delineation***

Homogeneous regions may be defined in the space of geographical coordinates. This definition, however, has the disadvantage that it is not applicable to small areas and that contiguous basins may not be hydrologically similar (Linsley 1982, Cunnane 1986, Wiltshire 1986). To overcome this difficulty, some researchers have defined homogeneous regions in the space of flood-related variables (Mosley 1981, Gottschalk 1985, Wiltshire 1986). This definition has both advantages and disadvantages. The primary advantage is that it is based on variables directly related to the flood phenomenon; its disadvantages are firstly the difficulty of relating the characteristics of hydrologically defined homogeneous regions to the topographical, physiographical and meteorological conditions of the area and secondly the fact that homogeneous regions in this definition are usually determined by cluster analysis, the purpose of which is to discover «natural clusters» (Dillon and Goldstein 1984) based on the assumption that such clusters exist; however, the existence of

such clusters cannot be taken for granted without prior testing (Rogers 1974, Dubes and Zeng 1987). The final set of homogeneous regions depends on the clustering method, the initial partitioning of the space and the metric used. For this reason, some researchers attempted to relate this type of homogeneous region to the geographical coordinates empirically (Mosley 1981, Gottschalk 1985), or to introduce the concept of fractional membership of a basin to a homogeneous region (Wiltshire 1986).

An entirely different concept of homogeneous region is the «neighbourhood» or «region of influence». Here, a homogeneous region is defined in the space of physiographical, meteorological and hydrological variables and centered on the basin under investigation (Acreman and Wiltshire, 1989, Burn, 1990ab, Zrinji and Burn, 1994, Ouarda et al., 1998, 2000, 2001). This type of region avoids the difficulties related to the existence of «real» clusters but, in contrast to a priori regions, it has to be determined specifically for each basin under investigation. Two specific methods can be identified: the region of influence method (Burn, 1990ab) and the canonical correlation analysis method (Ouarda et al., 1998, 2000, 2001).

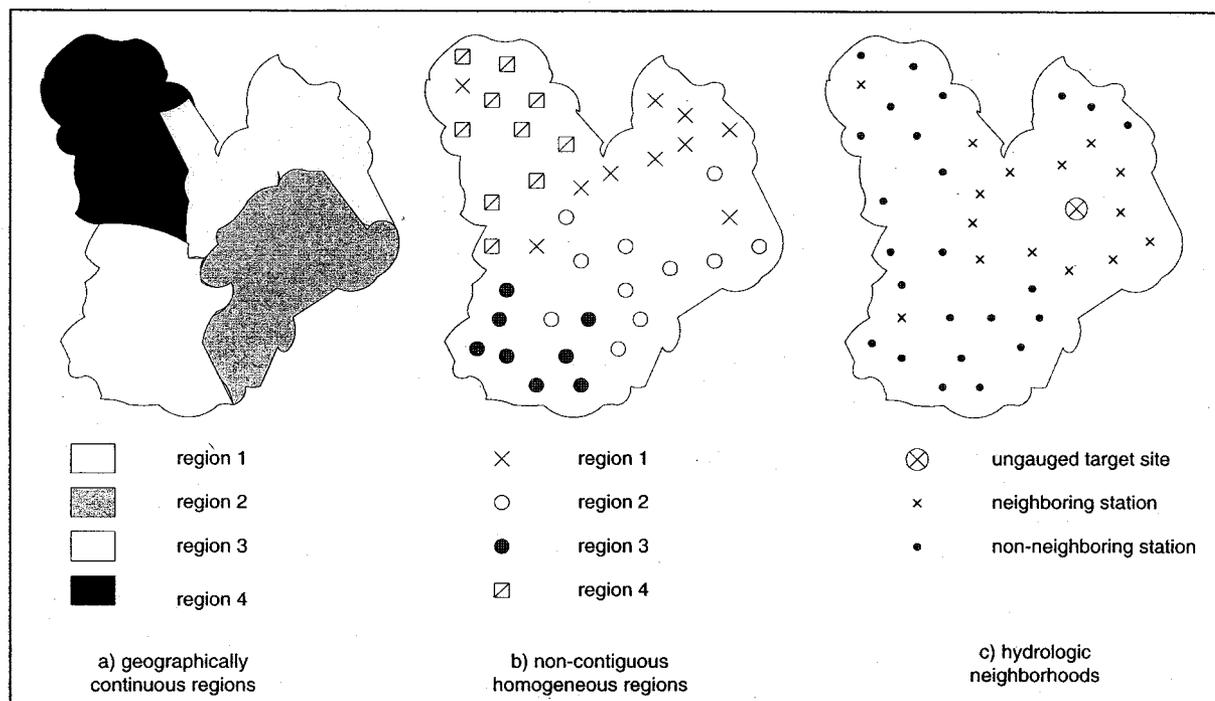
According to the region of influence method, the gauged basins enter the region of influence in the order of their weighted Euclidean distances from the ungauged basin in the space of the physiographical and meteorological variables where the weights of the variables are selected by the user. At every step, a homogeneity test, based on the flood distributions of the gauged basins of the homogeneous region is used to determine whether the boundary of the homogeneous region has been reached. The «region of influence» approach has the following limitations:

- (a) It requires a choice of an arbitrary weight for each basin variable.
- (b) It uses weighted Euclidean distances that do not take into account the correlations between the basin variables.

(c) The region of influence is determined using both the basin and the flood variables without taking into account the relations between these two sets of variables.

The other approach for determining basin-centered homogeneous regions, uses the multivariate method of canonical correlation analysis which takes into account the relationships between the physiographical and meteorological variables and the characteristics of the distribution of the maximum annual floods (Ouarda et al., 1998, 2000, 2001). This approach has several strengths compared to all other methods of delineation of homogeneous regions. Among these strengths is the limited number of assumptions the method makes (only one hypothesis, and absolutely no assumptions concerning the statistical distribution of floods/droughts), the robustness of the method to changes in the configuration of the hydrometric network and to at-site estimation, the solid mathematical background of the method, and the capacity to jointly estimate several variables at the same time (for example flood peak and flood volume quantiles). A detailed description of the mathematical background of the method and its characteristics is available in Ouarda et al. (2001).

In a recent intercomparison study (GREHYS, 1996a,b) the following methods of determining homogeneous regions for flood estimation were compared for the basins of the Canadian provinces of Ontario and Quebec: Correspondence analysis, hierarchical clustering, canonical correlation and L-moments. The study concludes that “the specific use of the canonical correlation technique yielded the best results for the delineation of homogeneous regions. Figure 1 illustrates the various approaches for the determination of homogeneous regions that were discussed in this section.



**Figure 1. Various approaches for the determination of homogeneous regions**

### 3.1.2 Regional estimation

Essentially, a regional estimation method is a technique for transferring information from the sites deemed to possess pertinent hydrological information to the site at which a quantile estimate is needed. Many regional estimation methods have been suggested in the literature.

The index flood method for regional flood estimation was devised by Dalrymple (1960). The main assumption is that floods at different sites within a region are identically distributed except for a scale factor which is a function of the physiographical basin characteristics. Usually, the catchment area is the most important factor for explaining the scale of floods. The index flood method was abandoned for some years, mainly because it was found that the catchment area affected not only the scale, but also other characteristics of flood distributions, such as the coefficient of variation and the coefficient of skewness, which according to the assumptions should be constant. In the last decade, however, the index flood method has regained its early popularity,

and today it is judged to be one of the most efficient regionalisation techniques (Cunnane, 1986; Potter and Lettenmaier, 1990).

*The direct multiple regression* model is based on the relationship between the at-site estimates of flood quantiles and catchment characteristics obtained by multiple regression. An advantage of the method is the flexibility in choosing the type of distribution to represent floods/droughts (Ouarda and Ashkar, 1994). In practice, the power-form function is the most commonly-used model to describe the relationship between the flood quantile  $Q_T$  and the hydrometeorologic and basin characteristics  $A_1, A_2, \dots, A_n$  for a given region. To estimate the parameters, a common procedure consists in linearizing the regression equation by a logarithmic transformation, and then estimating the parameters of the linearized model by an ordinary least squares technique. The usual procedure is therefore straightforward, because one can make use of simple linear regression techniques to identify the parameters of a nonlinear model. However, linear regression cannot be used to estimate the parameters of the power-form model with an additive error term. In such cases, a nonlinear optimisation method must be employed to estimate the parameters.

GREHYS (1996a,b) demonstrated that the approaches of index flood and multiple regression lead to equivalent results. The performances of these two approaches, in association with a wide range of homogeneous region delineation techniques were very similar when applied to a data base composed of all stations in the provinces of Quebec and Ontario in Canada.

### **3.2 Selection of methods**

Based on the discussion presented above, the superiority of the canonical correlation analysis approach for the delineation of “homogeneous” regions is clear. In fact we can generalise that methods that are based on the hydrologic “neighbourhood” approach are superior to the methods that are based on the hydrologic “region” approach. If the user wants also to incorporate a simple method that is based on the geographically contiguous or non-contiguous region approach, the clustering approach can then be recommended. Indeed, this approach is easy to implement, has a sound statistical foundation, and was demonstrated by GREHYS (1996a, b) to lead to good results.

As indicated earlier, regional estimation can be carried out using the index flood/drought approach or the multiple regression method. These two approaches are quite similar and lead to equivalent results.

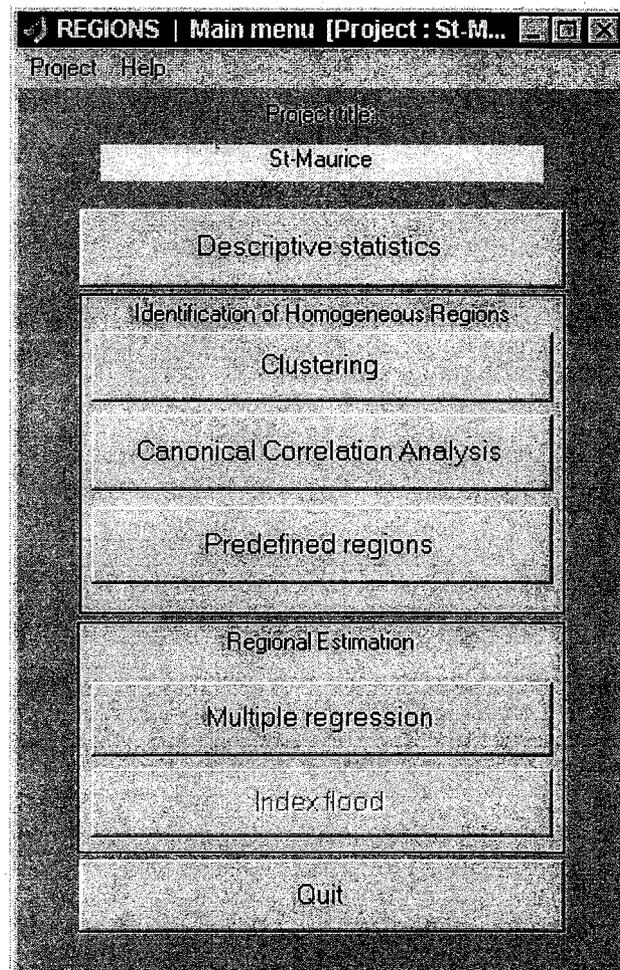
### **3.3 Adaptation of currently available tools and software**

It is also important to focus on the potential to use and adapt currently available hydrologic regional estimation tools and software that utilise sophisticated methodologies. This would avoid the high costs associated with the development of new tools, and would allow to use other tools that were already tested and approved.

The team of the Chair in Statistical Hydrology at INRS-ETE has developed a software **REGIONS** which implements several approaches for the regionalisation of hydrological variables (Ouarda et al., 2002). **REGIONS** is characterized by a user-friendly interface, which was developed in Matlab 6.1. The interface was programmed to assist users who may not be trained in statistical techniques so that they can complete all steps required in a regionalisation procedure. The software **REGIONS** allows to analyse data sets composed of hydrological variables (i.e. max or min discharges and volumes) as well as physiographical and or meteorological variables (e.g. drainage area, main channel slope, percentage of forest, mean annual precipitation, etc.) measured from stations usually located within the same region. It allows to calculate descriptive statistics such as mean, standard deviation, correlation, and offers the possibility to use one of the following methods to define groups of “homogenous stations”:

- User defined regions (geographical regions)
- Hierarchical clustering
- Canonical Correlation Analysis

Figure 2. illustrates the main menu of the **REGIONS** software.

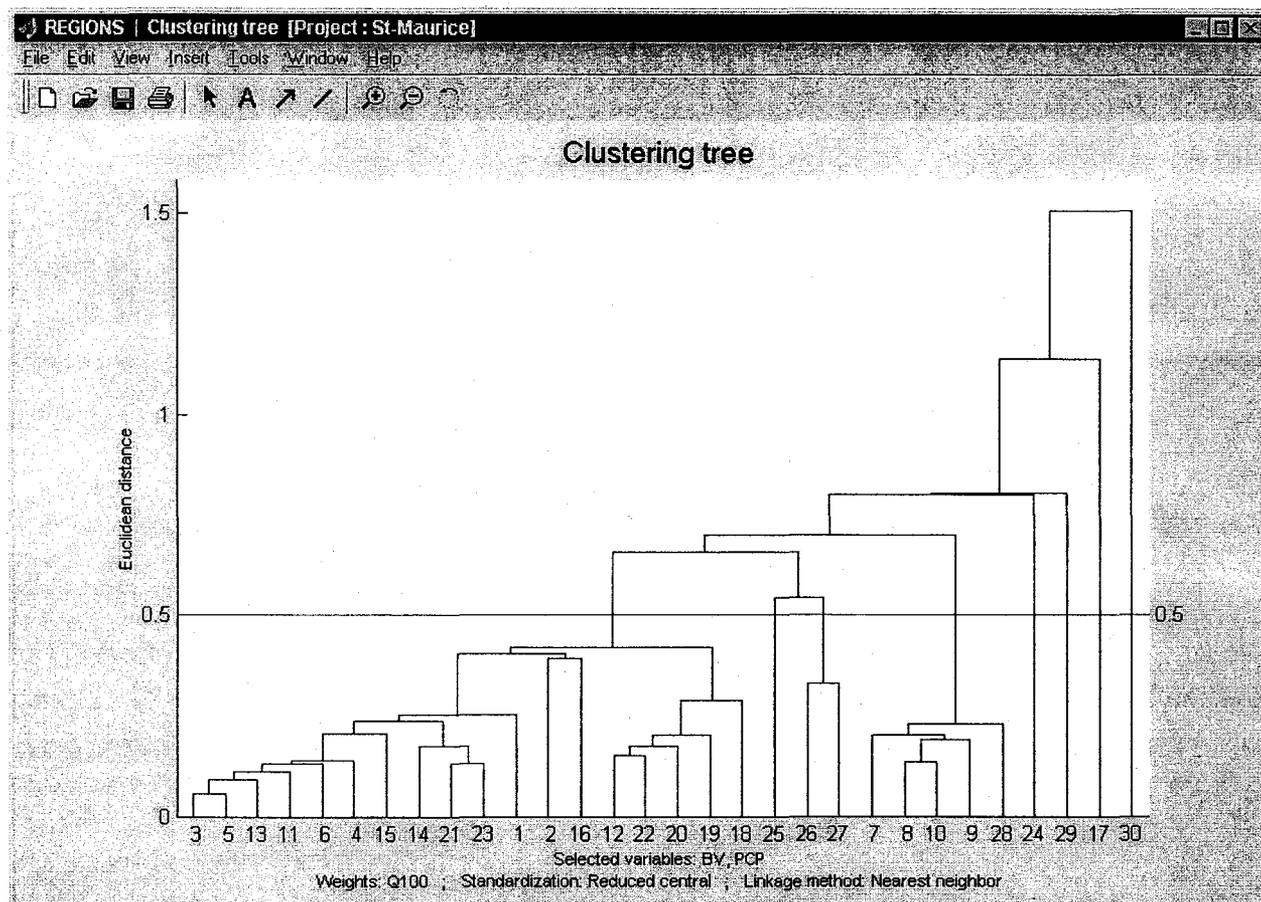


**Figure 2. Main menu of the REGIONS software**

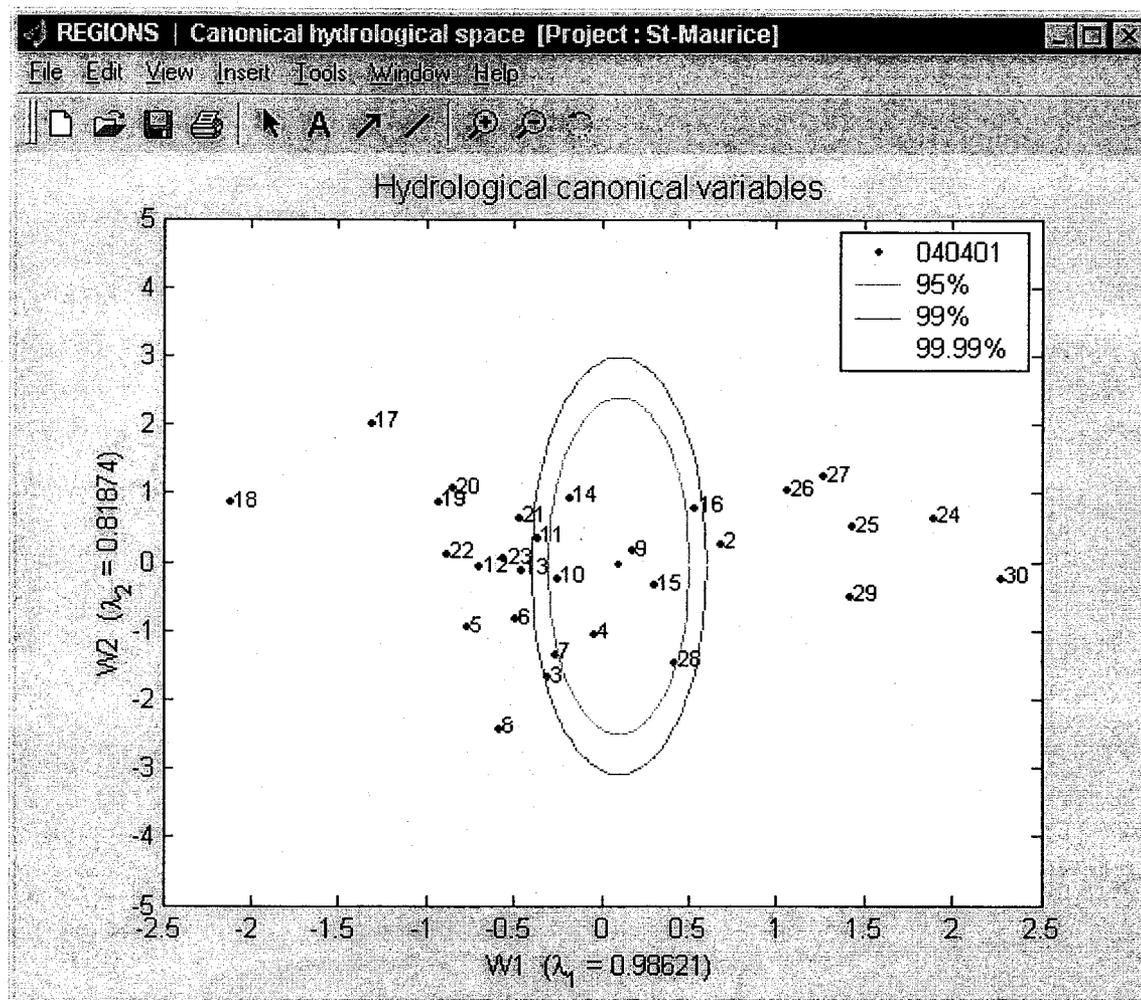
In the final stage, the software allows to carry out the regional estimation based on one of two methods :

- Multiple regression, or
- The Index Flood approach

Figure 3. presents a screen representing the hierarchical clustering approach, and figure 4. illustrates the canonical correlation analysis method for the delineation of homogeneous groups.



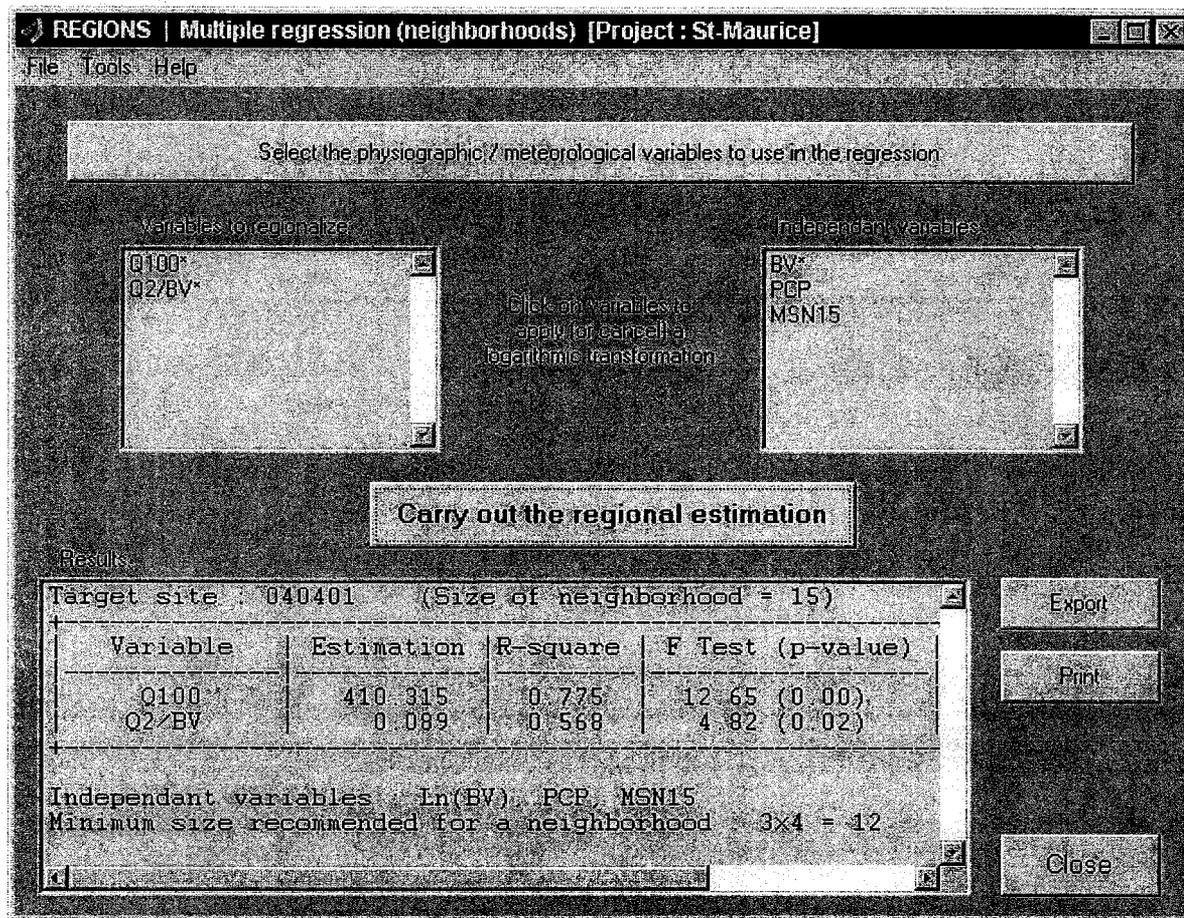
**Figure 3.** A screen representing the hierarchical clustering approach



**Figure 4.** A screen representing the canonical correlation analysis approach

Figure 5. presents a screen from the REGIONS software illustrating the multiple regression approach for regional estimation. The REGIONS software has already been adopted by Hydro-Quebec for all their regionalisation needs. Other improvements and additional features are planned in the software REGIONS by the developing team at INRS-ETE, including the development of a data-import module that will enable the user to import data in various formats. A link with the HYFRAN software will also be developed. This link will enable the users to

perform an initial frequency analysis of data for quantile estimation at gauged sites prior to estimating quantiles at ungauged sites by regionalisation.



**Figure 5.** A screen illustrating the multiple regression approach for regional estimation

Other additions include the development of a specific module for the regional estimation of drought characteristics that is characterised by its generality, and the

development of a module which allows to improve the estimation of flood/drought quantiles in sites that are already gauged, by using available regional information.

### **3.4 An alternative methodology for the regional estimation of drought characteristics**

Another approach that was presented in the literature and which deals with the regional estimation of drought characteristics is the regional drought tail model based on scale invariance (Hamza et al., 2001). This approach lies on the combination of two procedures:

1. The simple scale invariance method for regional series of drought flows, based on the analysis of the relationship moments-surfaces or the relationship quantiles-surfaces, explains the spatial variability of drought processes by their indexation on a series of scale parameters, essentially the size of the drainage basin. This procedure principally aims at the delineation of homogeneous regions, and this characterizes the first condition of a regional model.
2. frequency analysis of minimum flows, with the approach using tail conditional models, consists in fitting a probability distribution to values that are smaller than or equal to some given threshold. As a fact, this procedure gives some more weight to the lower part of the distribution by defining *a priori* a level of censoring, mainly a threshold  $u_i$  called ceiling value. This procedure establishes the second condition of a regional model, that is the determination of a regional estimation approach.

This methodology has been applied by Hamza et al. (2001) to analyse drought flow characteristics for the hydrometric stations of the province of Quebec, Canada. Results of the analysis indicate the robustness of this approach compared to the other approaches that are available in the literature for the regional estimation of drought flows. Regional

drought quantile estimates obtained with this model were the least biased and corresponded to the least root mean square error.

## 4. Conclusions

---

IGAM currently uses an empirical approach for the regionalisation of drought flows and design flows, which relies on previous experience and on a series of maps that define a set of geomorphological, meteorological and hydrological characteristics for the State of Minas Gerais. This approach allows to obtain useful estimates of the hydrological variables of interest for the planning and the management of water resources in the State. However, this approach presents several limitations that are related to the accuracy of estimates and to the cost of estimation. Furthermore, the availability of alternative methods that are tested, approved, that rely on solid theoretical background, and that are available in ready to use software, make it difficult to spend more resources and energy on the updating of the approach currently adopted by IGAM.

As an alternative to this approach, we propose to adopt a regionalisation methodology that is based on:

- The hierarchical clustering approach combined with multiple regression if fast estimates are needed, if a global mapping of the whole state is requested, or if only a limited amount of information is available in the region of study (physiographical and meteorological data).
- The canonical correlation analysis approach combined with either multiple regression or the index flood/drought method if reliable estimates are requested in a specific site, and a sufficient amount of information is available regionally to carry out the whole regional estimation procedure.

These two approaches are already implemented in the software REGIONS developed by the team of the Chair in Statistical Hydrology (Ouarda et al., 2002) and are hence easily available for applications with data from the State of Minas Gerais. Our recommendation is to adopt the use of

these methods, and to take advantage of future developments that are planned in the software REGIONS. Such developments include the development of a data-import module that will enable the user to import data in various formats, a link with the HYFRAN software, the development of a specific module for the regional estimation of drought characteristics that is characterised by its generality, and the development of a module which allows to improve the estimation of flood/drought quantiles in gauged sites through regional estimation. The use of the approach of regional drought tail model based on scale invariance (Hamza et al., 2001) for the regional estimation of drought characteristics represents another planned addition to the software REGIONS.

## 5. References

---

- Acreman, M.C. and Wiltshire, S.E. (1989). The regions are dead: long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. IAHS Publ. no. 187, 175-188.
- Burn, D.H. (1990a). An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrological Sciences Journal*, 35 (2) 149-165.
- Burn, D.H. (1990b). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research* 26 (10) 2257-2265.
- Cunnane, C. (1986). Review of statistical models for flood frequency estimation. Keynote paper in: *International Symposium on Flood Frequency and Risk Analysis* (Baton Rouge, May 1986). Reidel.
- Dalrymple, T. (1960). Flood frequency analysis. *USGS Water Supply Paper 1534 A.*, p. 60.
- Dillon, W.E. and Goldstein, M. (1984). *Multivariate Analysis*, p. 139. John Wiley.
- Dubes, R. and Zeng, G. (1987). A test for spatial homogeneity in cluster analysis. *Classification* 4, 33-56.
- Gottschalk, L. (1985). Hydrological regionalization in Sweden. *Hydrol. Sci. J.* (30) (1).
- GREHYS (1996a). Presentation and review of some methods of regional flood frequency analysis. *J. Hydrol.* 186 (1-4), 63-84.

- GREHYS (1996b). Intercomparison of flood frequency procedures for Canadian rivers. *J. Hydrol.* 186 (1-4), 85-103.
- Hamza, A., Ouarda, T.B.M.J., Durrans, S. R. and Bobée, B. (2001). Développement de modèles de queues et d'invariance d'échelle pour l'estimation régionale des débits d'étiage, *Can. J. Civ. Eng.* 28: 291-304.
- Linsley, R.K. (1982). Flood estimates. How good are they? *Wat. Resour. Res.* 22 (9).
- Mosley, M.P. (1981). Delimitation of New Zealand hydrological regions. *J. Hydrol.* 49, 173-192.
- Ouarda, T. B.M.J. and Ashkar, F., (1994). Regional multiple regression flood frequency estimation by the Peaks-Over-Threshold method. Internal Report, Department of Mathematics, University of Moncton, Moncton, New Brunswick, Canada. Research report for the Strategic Grant N° STR0118482 of NSERC, 20 pp.
- Ouarda, T.B.M.J., Haché, M. and Bobée, B. (1998). Regional estimation of extreme hydrological events, INRS-Eau, Research Report No. R-534 (In French), Quebec, Canada, p. 181.
- Ouarda, T.B.M.J., Haché, M., Bruneau, P. and Bobée, B. (2000). Regional flood peak and volume estimation in Northern Canadian Basin, *Journal of Cold Regions Engineering of the ASCE*, 14(4): 176-191.
- Ouarda, T.B.M.J., Girard, C., Cavadias, G. S. and Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis, *Journal of Hydrology*, 254: 157-173.
- Ouarda, T.B.M.J., Gingras, H., Kouider, A., Ristic-Rudolf, Z. and Bobée, B. (2002). User manual for the software REGIONS, INRS-ETE, Ste-Foy, Canada.
- Potter, K.W. and Lettenmaier, D.P., (1990). A comparison of regional flood frequency estimation methods using a resampling method, *Water Resour. Res.*, 16(3): 415-424.

Rogers, A. (1974). *Statistical Analysis of Spatial Dispersion*. Pion Ltd.

Wiltshire, S.E. (1986). Regional flood analysis II: multivariate classification of drainage basins in Britain. *Hydrol. Sci. J.* 31 (3).

Zrinji, Z. and Burn, D.H. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology* 153, 1-21.