HISTORICAL FLOOD DATA ANALYSIS USING A GIS: THE PALAEOTAGUS DATABASE

M. FERNÁNDEZ DE VILLALTA and G. BENITO
CSIC-Centro de Ciencias Medioambientales,
Serrano 115 bis, 28006 Madrid, Spain.

and

A. DÍEZ HERRERO
Facultad de Ciencias Biológicas, Universidad S.E.K.,
40003 Segovia, Spain.

1. Introduction

Technical measures designed to mitigate natural flood disasters may be broadly classified into three main types: *predictive* measures based on spatial and temporal estimations of the phenomenon and which, in the case of fluvial flood events, are usually approached from a meteorological perspective; *preventive* measures aimed at minimizing human and material losses; and *corrective* measures, which attempt to recover the previous state of the area affected by the flood, including protective structures. Given the uncertainties of many predictive techniques including mid-term prediction of rainfall intensity, and the negative social impact of corrective measures, research efforts are mainly directed towards measures of prevention. Recent investigations have focussed on low frequency floods from a deterministic perspective based on probable maximum precipitation (PMP), and on statistical analysis using precipitation and flood frequency distribution functions.

Extraordinary flood events show periodicities which escape conventional hydrological records and there is thus a need for new methods of estimating events occurring over periods of thousands of years. Several sources of information permit the reconstruction of the magnitude and frequency of flood events which took place before the systematic recording of hydrological data at gauging stations. Basically, there are two types of data: documented information on historical floods (generally covering the last 1500 years) and geological flood indicators i.e., deposits accumulated during different flood events (records covering 100-10,000 years). If we take into account all the possible hydrological and climatic scenarios, these past floods are the best evidence of the probability that a certain event may occur again in the future. Moreover, historical and paleoflood data may successfully be applied in risk analysis and consequently in hydraulic and engineering design, to complete the information supplied by the empirical, statistical methods of conventional hydrology.

In countries of documentary tradition such as those of Europe, there is so much information on extraordinary floods or climatic events of the past that a method of

T. Glade et al. (eds.), The Use of Historical Data in Natural Hazard Assessments, 101–112.

systemization is needed to enable the comparison and analysis of data at different spatial and temporal scales. The search and analysis of this information on historical floods is undertaken using databases which efficiently collate the relevant data. Preexisting past flood databases include that successfully created by Frank M. Law, Andrew R. Black, Robert M. J. Scarrott and John B. Miller under the auspices of the British Hydrological Society: the Chronology of British Hydrological Events (at http://www.dundee.ac.uk/geography/cbhe). This is fully searchable for floods and related events by date (year, month, day), river basin or by text entry. Other countries such as the Peoples Republic of China, have a specific program for the compilation of historical flood data within the so-called national network of historical flood water level marks (Chen Chia-Chi, et al, 1975; Hua Shi-Qian, 1985).

One of the main shortcomings of preexisting databases is a lack of geographic referencing of information which will permit the later handling of data using geographic information systems (GIS). The database created for the Tagus basin "PaleoTagus" is of the relational type and is backed up by a standardized, compatible database management system which not only permits the storage of information, but also its quantitative and qualitative analysis through a GIS. The GIS includes graphical covers and applications plus the necessary tools for numerical analysis. Previous studies have reported the conceptual model for this database (Díez et al., 1998, Fernández de Villalta et al., 1998). In the present report, we describe details corresponding to the implementation of this model within a GIS, its search and analysis tools, and applications in the study and prevention of flood risks.

2. The Tagus River Basin: Present and historical flood data

The Tagus river drains the central part of the Spanish Plateau (Meseta) and has an E-W elongated basin. Originating in the Iberian Range, the Tagus flows into the Atlantic Ocean at Lisbon. It is the longest river of the Iberian Peninsula (1200 km) and the third largest in catchment area (81,947 km^2) passing both through Spanish (54,769 km^2) and Portuguese (27,178 km^2) territory. Hydrologically, the Tagus is characterized by extreme seasonal and annual variability including severe floods with peak discharges over 30 times the average discharge. The river regime is influenced by Atlantic fronts crossing the Iberian Peninsula mostly during winter. Eastern and northeastern tributaries are of a mixed hydrological regime and are affected by snowmelt and rain water from the Iberian and eastern Central Range areas, while southern and northwestern tributaries are dominated by rain water. General discharge characteristics are: (1) maximum discharge from February to March; (2) minimum discharge in August; (3) a peak in December; and (4) reduced discharge in January.

The information available on the past flooding of the Tagus basin dates back to AD 849, although more continuous and reliable data start to emerge in the year AD 1100. As for any historical information, these data may be biased since the majority of written documents correspond to inhabited areas and/or cities. These are scattered along the entire river, the largest being Aranjuez, Toledo, Talavera de la Reina and Alcántara. Other cities located at the main tributaries of the Tagus which boast extensive historical records are Madrid (Manzanares river) and Alcalá de Henares (Henares river). The sources of information for the PaleoTagus database were provided by our own search of historical archives (including the *Archivo General de Simancas, Archivo Diocesano de Toledo, Biblioteca Municipal de Madrid, Biblioteca Nacional*), and several recompilation works of historical floods such as those published by Fontana Tarrats (1977), Font (1988), Canales (1989) and the Comisión Técnica de Inundaciones (1985). Over 375 historical flood entries are included in the database (Benito et al., 1996).

In order to standardize information, a form is completed before introducing the data into the database. The completed form includes information on the year, month and day(s) of the flood, stream, reach/site, discharge, cause(s) and damage. The damage incurred is indicative of the severity of the flood for those events which lack data on flood magnitude. It is nevertheless difficult to qualitatively or qualitatively estimate flood magnitude from historical records. For some locations, the height reached by a given flood is reflected by marks on buildings and can be related to peak discharges using rating curves (discharge/height relationship) of corresponding cross-sections. However, information on most past flood events was documented merely because of the fact that bankfull discharge was exceeded causing damage to agricultural land, bridges and buildings. Further, there were more records of flood events causing damage to property during the second half of the XIX century, since this period saw an increase in human activities conducted on the river floodplain. It is thus clear that this type of record needs to be analyzed in detail to distinguish between human and/or hydroclimatic causes for the greater number of floods recorded.

3. Database structure

The historical database is of the relational type, such that information is not stored in a single table, but rather in several tables related to each other by identification codes. On the one hand, this permits the maintenance of a certain degree of homogeneity in data registration criteria - introduction of data using standard codes, maintenance or editing of data and the expansion of new records and fields of information - and on the other hand, enables the end-user to rapidly and directly consult the data contained in these tables and codes. In the PaleoTagus database, the *primary table* is denoted "historical" and contains information corresponding to a flood event in coded format and the *secondary tables* serve to define the meaning of these codes (Figure 1). The information contained in the primary table is thus simple in terms of size and volume of data, but complete in terms of information on the event in question.

In the "historical" data table, each flood event is assigned a registry code or event identifier which is made up of the combination of date+reach+location of the flood. This registry number, and therefore event number, is unique in the database. Note that the definition of the term "event" within the database differs from the strictly meteorological or hydrological concept of the term. In PaleoTagus, the subsequent spatial and temporal analysis of floods is aided by the fact that floods affecting different localities and reaches occurring on one particular date are considered as different floods. This new definition constitutes one of the main contributions of the PaleoTagus database, since it permits the coherent storage of information with assignment of flood data to particular geographical points. This is the basis of a GIS in which spatial data may be referenced using geographic co-ordinates. Given that the greatest degree of standardization on a national level is sought, the identification codes used are those established by other administrative bodies responsible for the compilation of databases (hydrological, territorial boundary etc.) For streams, the arcs encoded by the decimal river code (DRC) used by the *Centro de Estudios y Experimentación de Obras Hidraúlicas* (CEDEX) were used. Some unregistered streams had to be added to this

DRC according to criteria used by CEDEX if there was evidence of historical flooding. For the municipal districts, the codes of the *Instituto Nacional de Estadística*, INE (Spanish national statistics organization), were used. These are formed by codes which refer to the autonomous community, province and municipal district. For example, a flood occurring on December 20th 1168 (11681220), in the Tagus river (030100000000) as it passes through Toledo (0845168) would be assigned the code: 116812200301000000000845168.

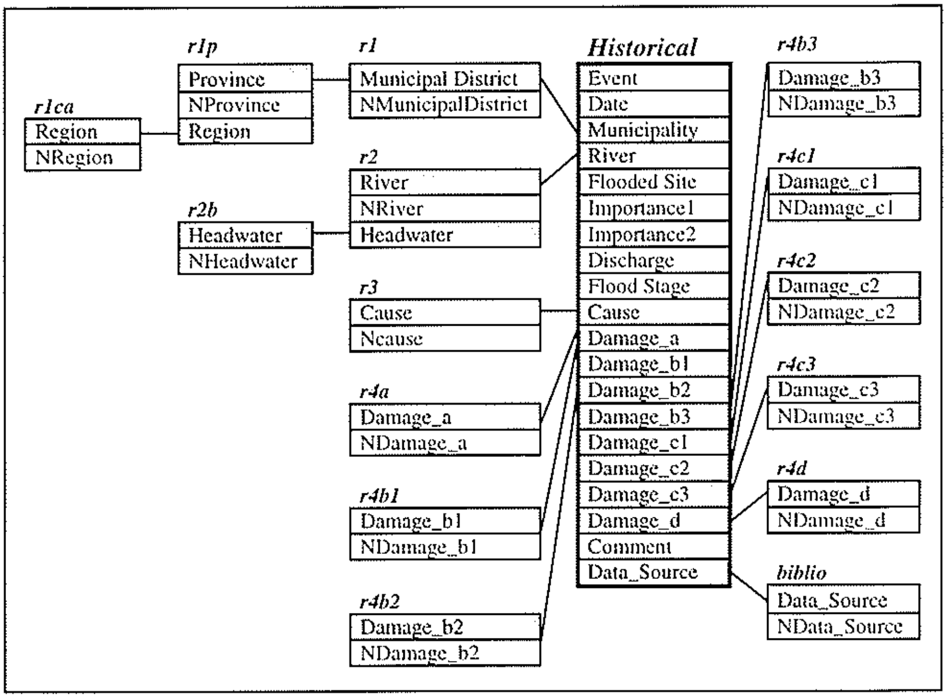


Figure 1. Primary table (Historical) and related secondary tables of the Palaeotagus Database

Once the record of each event is defined, a set of fields containing a further type of information completes the primary table (Figure 1). These are: (1) numerical, (2) encoded alphanumerical (3) and alphanumerical in text format. The numerical information mainly consists of hydraulic flood data such as water level, estimated discharge, flow velocity etc. Alphanumerical data correspond to fields such as causes and damages, encoded according to the pre-established groups. The descriptions obtained from historical documents are included as fields of text and correspond to: (1) the sites or landmarks reached by the flood (e.g., churches, bridges etc.), (2) flooded areas (e.g., orchards, valleys etc.), (3) the relative importance of the event with respect to previous floods, (4) comments and (4) data source of the information.

In this way, the "historical" primary table is drawn up. Associated with this table are the code-relating, or secondary tables (Figure 1), which comprise two fields: code and definition or name. Some of these tables contain a complete list of items: streams

(CEDEX code, name), municipal districts (INE code, name) and, related to the latter, tables of provinces and autonomous communities (INE code, name). Others define the causes leading to the flood (rain, snow, hail, rain+snow, etc.) and the damages produced: human losses, damage to infrastructure (communication, water or urban structures), damage to agricultural land and industries (irrigation and drainage networks, farmland, industrial areas) and geomorphological changes in river channels and floodplains. These damages are assigned codes corresponding to: no damage, minor damage and major damage or destruction.

For the geographic referencing of data, drainage network arc and municipal district covers are used. Flood events are searchable by the stream+municipal district combination which may be obtained from the event code shown in the database. *In situ* rainfall events (flooding by local runoff not related to a stream) are specifically searched by municipal district only.

4. Database implementation

The PaleoTagus database was designed to be implemented by end-users with different profiles and demands (multipurpose exploitation). A further prerequisite is that its use should be simple and flexible. This is achieved through the use of windows which operate the data search and visualization process. Any user unfamiliar with the database or even with GIS should be able to consult the database and obtain information according to his/her needs (scientific, water resource management, civil protection, engineering etc.). Thus, an important feature of Paleotagus is the PaleoTagus Program Manager which aids in the consultation and handling of the database, providing visual results (graphs and maps), text (tables and reports) and simple data analysis (Figure 2). The software was created using the Avenue programming language (directed at objects) based on the GIS ArcView 3.0a program which is widely used worldwide. Using this language, small programs (scripts) were developed which permit the creation of windows, menus, tools, dialogue windows etc. acting as intermediaries between the end-user and the GIS. The general idea was to create a simple interface to permit the future end-user to consult data and obtain all the information required, without the need to directly handle the Paleotagus database.

The program manager consists of a main menu from which windows with the following query options or submenus may be accessed (Figure 3): a) date: specific, interval, season; b) stream: name, headwater area; c) site: town/village, province, autonomous community; d) cause: rain, snow, hail, rain+snow, in situ rain; e) damages: human, material (communication, water and urban networks), to agriculture and industry (irrigation and drainage networks, farming areas, industrial zones), to natural channels. From this menu, the query may be approached via any of these fields or combination of these as required. When an option is chosen, a window appears such that the user may formulate his/her query based on the data available in the database. For example, a complete list of rivers or municipal areas is provided for the end-user to select one or several of the names on the list. Thus, possible errors in introducing names (e.g., spelling, use of capitals, articles, accents etc.), dates (e.g., the order of day/month/year, number of digits etc.) or data that the database lacks are avoided.

Once formulated, the search route may be defined as a new query, be added to the previous query or taken from a previous query thus narrowing the field according to the data of interest for each particular application (Figure 2). Thereafter, another window appears and a name is introduced to identify the new covers generated and the associated database, text files, graphics and maps. In this manner, the information may be called up at a later date without the need to repeat all the steps performed.

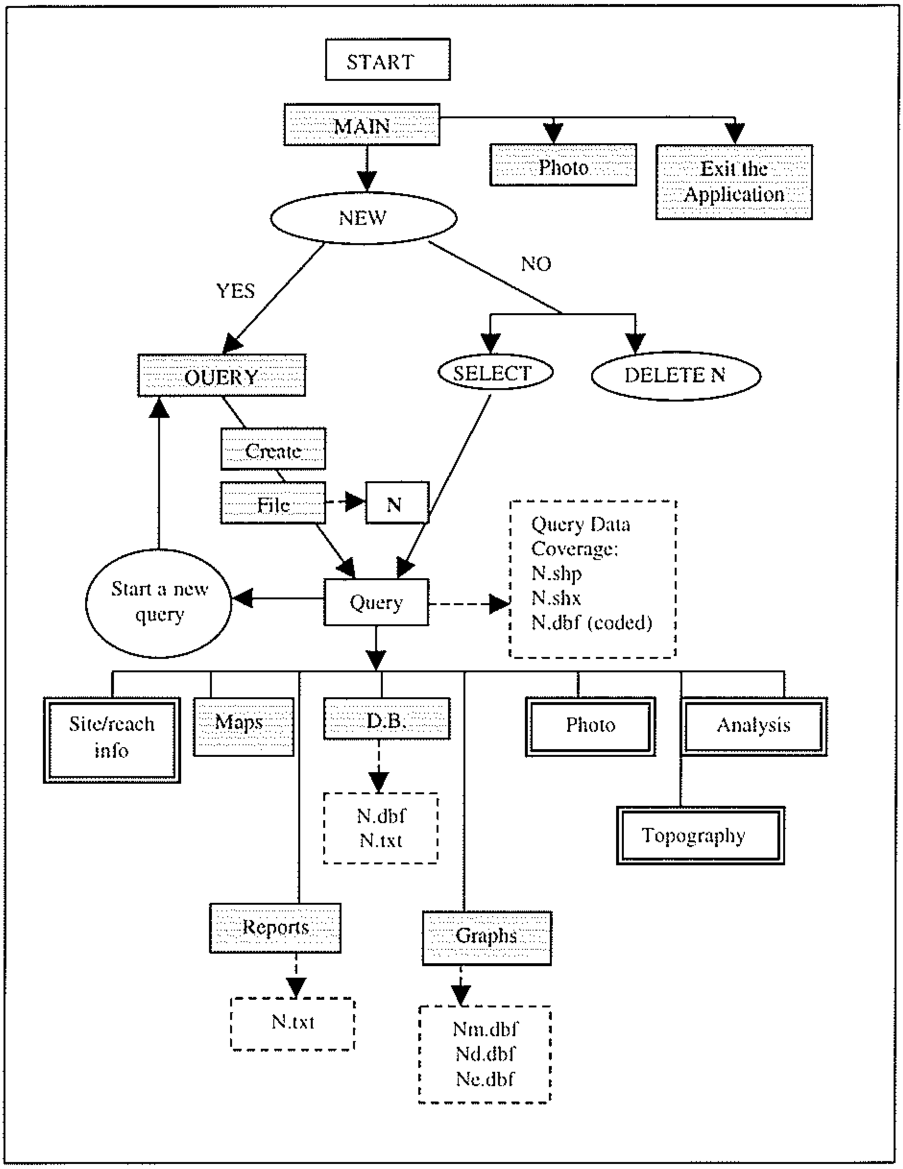


Figure 2. Diagram summarizing in a general way how PalaeoTagus run the ArcView application and the type of information available from a query.

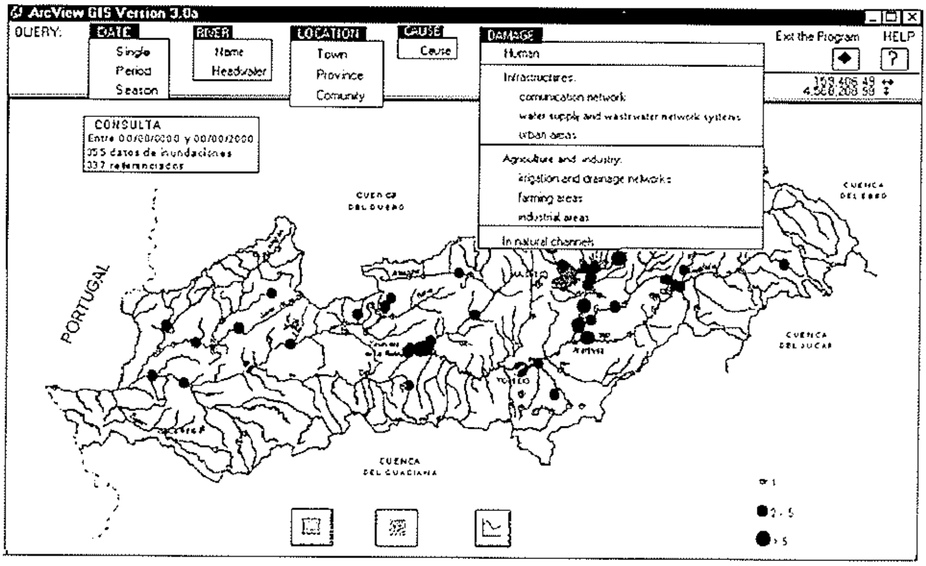


Figure 3. Main menu of the PalaeoTagus program manager and graphical output of a search comprising all the historical floods in the Tagus Basin.

For any type of query, the program extracts data complying with the conditions requested and places them on a geographical cover using a symbol, the diameter and color of which depend upon the number of flood events occurring in the given river reach (1, 2 to 5, or over 5 events) (Figure 3). The basic geographical cover is comprised of the main rivers and streams, the Tagus basin watershed boundary and the main cities.

The final result appearing on screen is a map of flood sites, accompanied by a series of tools and menus which permit the end-user to access the following types of information (Figure 3):

Text

(a) using the "*information*" tool at a particular site, a window appears showing data corresponding to this point.
(b) from the menu option "*Data Table*" a text file containing the results of the query is obtained. This information may be included in any other type of document or be consulted later by definition of the type of information or fields which the user wishes to include.
(c) in the same manner, via the menu option "*Data Table: Export DBF to TXT*", it is possible to obtain a file containing data in the form of a table.

Graphics

Graphs showing the temporal distribution of past flood events by month, season or decade may be called up using the "*Graphics*" option of the menu. At the same time, associated tables are generated such that the information they contain may be used in other applications or programs.

Others

In the lower part of the screen there are three "buttons". When these are selected, a mark appears at the sites where further information is available such as: a) cartographic, enabling the topography of the reach and, when applicable, the longitudinal profile and cross-sections used in hydraulic models to be viewed; b) photographic, providing access to the available photographic file and; c) results of calculations and hydraulic and statistical analyses etc.

5. Database query results

The PaleoTagus database was designed to accomplish multidisciplinary milestones with maximum versatility in mind. In each case, the end-user may seek the information applied to his/her study and formulate the appropriate queries and searches to satisfy his/her needs. Some of the types of searches and applications which may be conducted using the PaleoTagus database are detailed below.

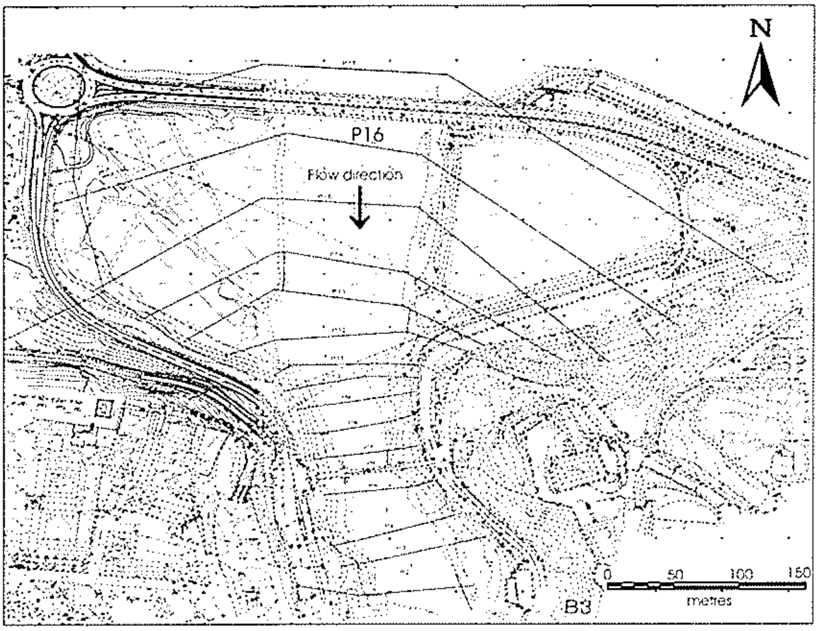
In the most elementary type of search, information is obtained on the distribution and frequency of past floods occurring in the Tagus basin. It should be noted that, since the database is represented in its entirety and includes information on the main towns for which the compilation of data is most complete (Madrid, Alcalá de Henares, Aranjuez, Toledo and Talavera de la Reina), there is always a bias to be considered. A further limitation that needs to be taken into account, although to a lesser extent, is the enhanced number of flood events affecting river reaches with significant construction and/or human activity such as engineering works (bridges, roads, irrigation canals) or agriculture. Although not urbanized, these areas of activity susceptible to interruption or destruction as a consequence of flooding often feature in historical documents. In this search, four periods corresponding to those showing the greatest frequency of extreme flood events are identified: AD 1150-1290, 1550-1600, 1730-1850 and 1850-1910. From AD 1900 onwards, the information is also biased in terms of the number of references made to flood events. This is attributable to an increase in anthropogenic activity in riverside areas and the consequent greater amount of information published in local newspapers.

Apart from general searches, queries which center on specific channel reaches or sites may be made. Here, more detailed information on each documented record is provided along with data on specific hydraulic variables for each past flood event. The type of information supplied consists of photographs of the relevant reach, both at low water times and during recent flood events (when available), detailed topographical maps, hydraulic modeling data and estimated peak discharges of the historical floods. The topographical maps also show cross-sections corresponding to the hydraulic models generated by the HECRAS program (Hydrologic Engineering Center, 1995), and water surface profiles for different discharge values corresponding to flood stages described in the historical record. The relationship between the different historical events and recent events may be visualized through rating curves, where the different flood stages in meters above sea level or other reference datum are related to the peak discharge values estimated by hydraulic modeling of the particular cross-section. Figure 4 shows an example of the type of information and results provided at the local scale for reaches with abundant past flood data. Flood stages are given with one or two reference altitudes and an alphanumerical code. This code refers to one of the following assumptions: (MI) minimum flood stage; (MA) maximum flood stage; (E) exact discharge level (equal to the flood stage); (R) discharge quoted as a range in the case of two recorded levels. Longitudinal and transverse profiles show water surface profiles for the different peak discharges corresponding to historical references.

5.1 CASE STUDY OF THE TAGUS RIVER IN TOLEDO

The river Tagus flows through an alluvial floodplain, upstream from Toledo and narrows as it enters a bedrock canyon which surrounds the city (Figure 4). The Alcántara bridge is located in the area where narrowing commences (P6 in Figure 4), and was probably built during the Roman Empire, although the first documented information on this bridge corresponds to AD 567. Some 150 m downstream from this bridge, there is a mill constructed in medieval times with a dam that diverts water towards the mill at times of low water (P1 in Figure 4). It is of note that a hydraulic jump occurs during flood events at this point permitting critical flow to be selected as the boundary condition to be applied in the hydraulic model. It is known that the geometry of the floodplain at the upper end of this reach has changed in time. However, this type of setting with a hydraulic controlling section in the bedrock canyon and which we assume remains more or less unchanged, determines that changes in the fluvial geometry of the upper reach do not substantially affect the peak discharge values yielded by hydraulic modeling. Peak discharges of past historical floods (Figure 4) reflect at least four flood events (AD 1113, 1168, 1178 and 1181) presenting discharges from 4000 to 5000 m^3sec^{-1}. These floods are considerably greater than those recorded in recent historical times such as those occurring in AD 1876 and 1947, which present peak discharges of 3000 to 3100 m^3sec^{-1}. Other past flood events of relevance such as those of AD 1211 and 1258 have estimated peak discharges *c.* 3000 m^3sec^{-1}, and those occurring in AD 1158, 1200, 1565 and 1942 are estimated at *c.* 2000 m^3sec^{-1}. Peak discharges below those corresponding to these past floods are provided by the Toledo gauging station instrumental data record (from 1972 to the present), which quotes values under 1100 m^3sec^{-1}.

Two important features emerge from these findings. On the one hand, it seems clear that the historical flood record substantially completes the information on extreme floods with recurrence periods beyond gauging station time records. And on the other, the data obtained from the historical record reflect the non-stationary nature of the floods when considering time periods of sufficient length. Annual flood series used in flood frequency analyses are assumed to be stationary in time (all floods are randomly generated from a single probability distribution with stable moments). This assumption is a significant limitation to be kept in mind in the evaluation of hydrological hazards by purely statistical methods.



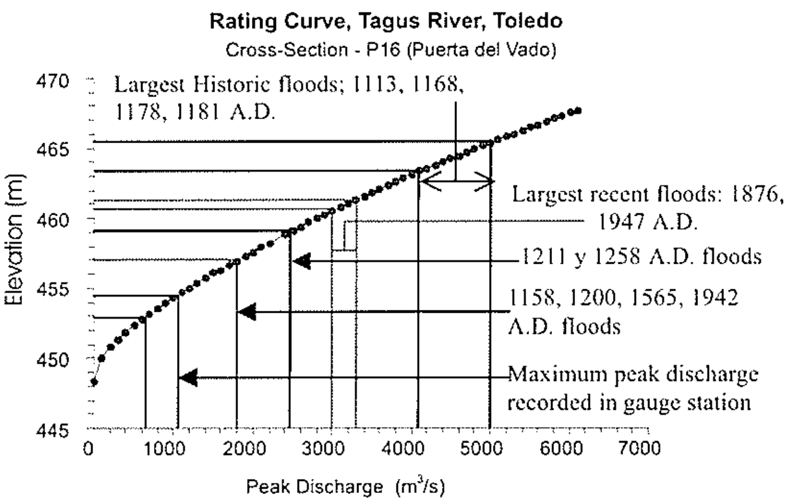


Figure 4. Topographic map of the Tagus river entering a bedrock canyon in Toledo. Cross-section P6 is located in the Alcántara bridge. In a mill dam, located few metres downstream of Cross-section P2, was assumed critical flow for the hydraulic calculations.

6. Conclusions and future perspectives

Palaeoflood and historical flood data are presently being used in several countries at the national scale to draw up a catalogue of past flood events. This use of past flood information is a step-forward since real events occurring at a given basin over several millennia are quantified and may thus be used in flood risk prevention, and particularly in flood frequency analyses. The use of past flood data and statistical tools through database management and geographical information systems considerably increases the potential benefits of spatial and temporal analyses, and the implementation of hydrological models and graphic output, facilitating the assessment of long-term hydroclimatic change and flood risks.

The future applications of this type of database are related to developments in flood planning and flood hazard mapping aimed at dealing with the social concerns of flood impacts. A better understanding of extreme flood events using methods based on past floods that actually occurred in the area will permit the more realistic mapping of flood risks to be applied in civil protection, land use and urban planning strategies. Indeed, the present lack of local legislation concerning land use and water policies is attributable to uncertainty regarding the risks associated with flooding. Tools based on past flood information will no doubt serve to inform decision-makers on the real risks that floods will pose to their citizens.

Acknowledgements

The authors are grateful to Luis Quintas Ripoll (Centro de Estudios Hidrográficos del CEDEX) for supplying GIS covers of the Tagus River Basin. This research was supported by the Spanish Committee for Science and Technology (CICYT), Spanish National Water Resources Programme grant HID99-0850, FEDER Project 1FD97-2110-CO2-02, and by the European Commission (DG XII), through research contract number EVG1-CT-1999-00010 (Systematic, Palaeoflood and Historical data for the improvEment of flood Risk Estimation "SPHERE" Project).

7. References

Benito, G., Machado, Mª.J. and Pérez-González, A. (1996) Climate change and flood sensitivity in Spain, in J. Branson, A.G. Brown and K.J. Gregory (eds.), *Global Continental Changes: the context of Palaeohydrology*. Geological Society of London Special Publication No. 115, pp. 85-98.

Canales, G. (1989) Inundaciones de la Vega Baja del Segura (1875-1925), in A. Gil Olcina, & A. Morales Gil (eds.), *Avenidas fluviales e inundaciones en la cuenca del Mediterráneo*, Alicante, pp. 415-433.

Chen Jia-Qi,, Ye Yong-Yi,, and Tan Wei-Yan, (1974) The important role of historical flood data in the estimation of spillway design floods. *Scientia Sinica* **18** (5), 669-680,

Comisión Técnica de Inundaciones (1985) *Estudio de inundaciones históricas: Mapa de riesgos potenciales*. Comisión Nacional de Protección Civil.

Díez-Herrero, A., Benito, G. and Laín-Huerta, L. (1988) Regional palaeoflood databases applied to flood hazards and palaeoclimate analysis, in G. Benito; V.R. Baker and K.G. Gregroy, *Palaeohydrology and Environmental Change*, John Wiley, Chichester, pp. 335-347.

Dunne, T. and Leopold, L.B. (1978) *Water in Environmental Planning*, W.H. Freeman and Company, New York.

Fernández de Villalta, M., Díez-Herrero, A., Benito, G. and Laín Huerta, L. (1998) Implementation of a regional palaeoflood database: The Palaeotagus Database. *IGU, Study Group on Environmental Change and Extreme Hydrological Events. First Meeting*, August 28 and 29, 1998. Lisbon.

Font, I. (1988) Historia del clima en España. Cambios climáticos y sus causas. Instituto Nacional de Meteorología. Madrid.
Fontana-Tarrats, J.M. (1977) *Entre el cardo y la rosa*, Madrid. (unpublished).
Hua Shi-Qian, (1985) A general survey of flood frequency analysis in China, *U.S.-China Bilateral Symposium on the Analysis of Extraordinary Flood Events*, edited by Anonymous,Nanjing, China:
Hydrologic Engineering Center (1995) *HEC-RAS, River Analysis System, Hydraulics Reference Manual*, (CPD-69).