A web-based open-source geoinformation tool for regional water resources assessment

Susanna Grasso [1],*, Pierluigi Claps [1], Daniele Ganora [1] and Andrea Libertino [1]

[1] Politecnico di Torino, Department of Environment, Land and Infrastructure Engineering, Corso Duca degli Abruzzi, 24 , 10129, Torino, Italy

* Correspondence: susanna.grasso@gmail.com

Abstract: To reduce the impact of droughts and increase the resilience of regional water systems, various demands, such as hydropower, supply and irrigation, need to be reconciled. In this perspective, designers and practitioners must be able to use information tools to define the hydrological constraints for a sustainable management of the resource. In this work, a web-based open-source geoinformation system is presented, that allows to estimate Flow Duration Curves (FDCs) in ungauged basins. The regional statistical model used, developed by Ganora et al., 2013 [1] in North-Western Italy, is based on the characterization of the FDC in a parametric framework where parameters depend on topographic, climatic, land use and vegetation descriptors computed at the basin scale. The software tool, accessible both by web browsers and GIS desktop (e.g. QGIS), pilots the estimations steps by computing the spatial descriptors and applying relations needed to estimate the full FDC via Burr distribution. The developed server-side scripting provides users with always updated data and procedures, being free from software client compatibility issues.

1. Introduction

Increases in the frequency, duration, and severity of extreme drought associated with climate change have raised global concern on the need of better planning of the management of the available water resources. Different studies have shown how global aridity has increased substantially since the 1970s, and climate models predict more droughts in the 21st century over most of Africa, southern Europe and the Middle East, most of the Americas, Australia, and Southeast Asia (a complete review can be found in Dai (2011) [2]). In temperate humid climates, management of resources in regional water systems need to face increasing conflicts of water uses, also due to more strict rules for the preservation of the environment. Tools and methodologies for a systematic assessment of the water availability at the basin scale are thus required, to ensure reliable water allocation. Flood Duration Curves are one of the most common tools for water resource assessment (see e.g. McMahon, 1993 [3]. They provide the frequency distribution of the complete flow regime of a catchment, starting from continuous average daily discharge data. A FDC computed for a given calendar (or hydrologic) year is called annual. If the observations of multiple years are merged together, the FDC refers to a mean annual statistical tool.

FDCs provide the percentage of time that a specified discharge in a river section is equalled or exceeded. This type of information is commonly adopted for hydropower assessment and mimimum instream flow evaluation, but is flexible enough to be the basis for impact assessment and risk management (see e.g. Kim et al, 2016; [4] and work reported therein. Moer importantly, FDC allows to assess mimimum flows as the low end of the curve (Ling Lloyd et al. 2015 [5]).

For a gauged basin, the empirical FDC is easily built by plotting the sorted observations versus their frequency of non-exceedance, computed with a plotting position formula (e.g. Weibull). However, most catchments of interest have no streamflow data. Therefore, regional statistical analyses are devised for FDC prediction in ungauged catchments. Where gauged data are either

limited or not available, regional models for predicting FDCs are commonly adopted (e.g.,Nruthya and Srinivas [6] , Vogel and Fennessey [7]). Ganora et al., 2016 [1] presented a regional model for FDC prediction in ungauged basins with a spatially-smooth approach, developed for the upper Po river region in North Western Italy (Figure 1, left). In Figure 1 (right panel) the basin boundaries and the average annual precipitation map are also depicted to provide a glimpse of the climatic variability of the area under analysis. Details of the methodology are provided in the section below, but climatic, topographic and geomorphologic information are required at the basin scale for the application.

In this paper, we describe how the whole procedure for estimating the mean FDC in a generic watershed has been implemented in a geo-information tool, that largely automates the calculations. In a first step, two Python scripts for the QGIS Processing Toolbox and an Excel worksheet have been developed. Subsequently, with the aim to provide access to GIS data and functionality through standard internet protocols, two WPS (Web Processing Service) procedures have been implemented, with the advantage of allowing users to access calculations independently of the underlying software.

In Section 2, the regional statistical methodology to estimate FDCs at ungauged sites is described. In Section 3, the software implementation is presented, including a brief description of the free and open-source platform architecture realized. The PyWPS procedure is then briey summarized. Finally, in Section 4, the client experience in producing a FDC curve is described, using also visual examples.
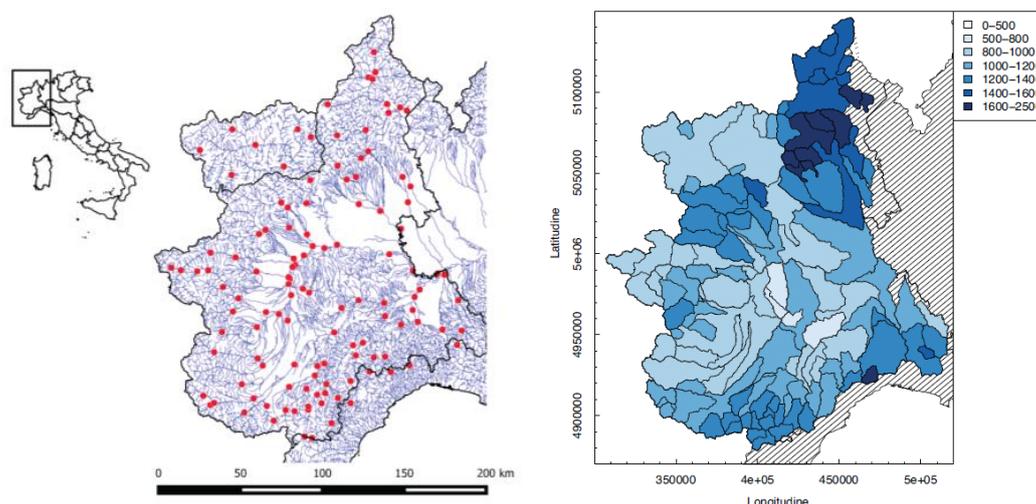

Figure 1. Left: Gauging stations and case-study area of the the FDC 's regional analysis. **Right:** Mean annual precipitation map for Piemont region.

2. The SSEM regional statistical analysis for FDC estimation

In this work, a regional spatially-smooth procedure to evaluate the mean annual FDC in ungauged basins is considered [1]. The statistical framework is based on a spatially-smooth procedure developed by Laio et al., 2011 [8] in the context of the regional flood frequency analysis. The complete FDC curve can be built by fitting a Burr probability distribution function a the set of the first three L-moments estimated for the FDC. They are obtained by multiple linear regression models including, as covariates, some geomorphological and climatic descriptors computed at the basin scale. L-moments are linear combinations of order statistics, which are commonly used in statistical hydrology (Hosking and Wallis, 1993) [9], for parameter estimation of various probability distributions. Here we will refer to μ as the mean value, and to the dimensionless coefficients τ_2 (LCV) and τ_3 (L-skewness or LCA) to completely characterize a three-parameter FDC curve.

To build the regional model, in a first step sample L-moments are estimated from the mean annual FDC computed at each available gauging station. To support the model implementation, about 120 catchment descriptors, including morphologic and climatic characteristics of the basin, soil and vegetation type, land use, etc., have been computed for each basin. Then, in order to select the

most appropriate regional model a larger number of possible models and regressions are considered and performed. To choose a suitable distribution function to represent the FDCs, a comparison between different distributions was done. In the context of the data set analysed, the Burr distribution (in the three-parameter form) resulted as the best choice (for more details see Ganora and Laio, 2015 [10]). The final models for the estimation of L-moments (mean, LCV, LCA), based on some catchment descriptors, are the following:

$$Y = -7.3605 \cdot 10^2 + 1.2527 \cdot \text{MAP} + 3.2569 \cdot 10^{-1} \cdot z_m + 5.2674 \cdot \text{fourierB1} - 6.7185 \cdot \text{clc2}$$

$$\text{LCV} = -2.896 \cdot 10^{-1} - 2.688 \cdot 10^{-3} \cdot \text{clc3} + 9.643 \cdot 10^{-5} \cdot \text{a75} + 1.688 \cdot 10^{-4} \cdot \text{MAP} + 20.941 \cdot c_{int}$$

$$\text{LCA} = 4.755 \cdot z_{max}{}^{-0.2702} \cdot \text{SD(IDFa)}^{0.06869} \cdot \text{CV}_{rp}^{0.2106}$$

where Y represents the general mean annual flow (expressed in mm), MAP is the mean annual precipitation z_m and z_{max} are the mean and maximum basin elevation, fourierB1 and CV_{rp} are rainfall regime parameters, clc2 and clc3 are land use parameters (the percentage of the basin area classified as group2 and group3 in the Corine Land Cover), a75 is the 75th percentile of the hypsographic curve and c_{int} and IDFa are extreme-rainfall statistics. In conclusion, the FDC quantiles can be predicted by the three-parameter Burr distribution (also known as Extended Burr type XII) whose parameters a, b, c, depend on regional predicted L-moments. However there are particular conditions for which the probability distribution becomes either a two-parameter Weibull or a Pareto distribution. This depends on the values of LCA and LCV, as it is possible to see in Figure 2, where the domain of definition of Burr distribution is shown.
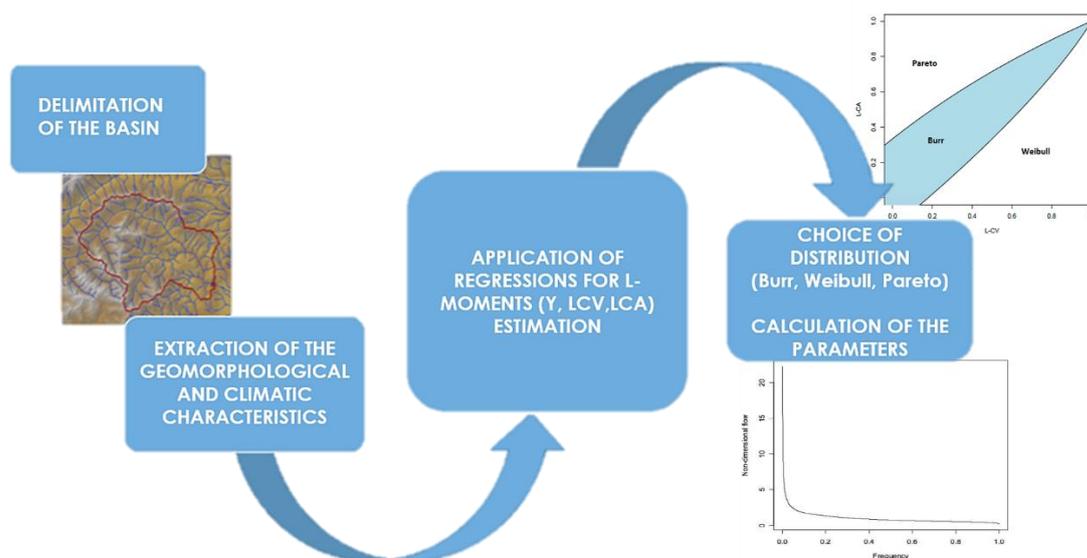

Figure 2. Process scheme to estimate the mean annual FDC in a generic watershed based on its morphometric and climatic descriptors.

3. Implementation of the geoinformation framework

3.1. Platform Architecture

For sharing these procedures a web platform was developed, with free and open-source software, using: PyWPS (Python Web Processing Service), 3.2.5 version; GRASS GIS (6.4.3 version) as backend to access all the geoprocessing functionalities; GisClient3 to build the WebGis (accessible through client browsers); Apache (2.2.14 version) as web server; Ubuntu 14.04 LTS as Server. Python version 2.7 was the one used here. Figure 3a shows a recap of the system platform architecture.
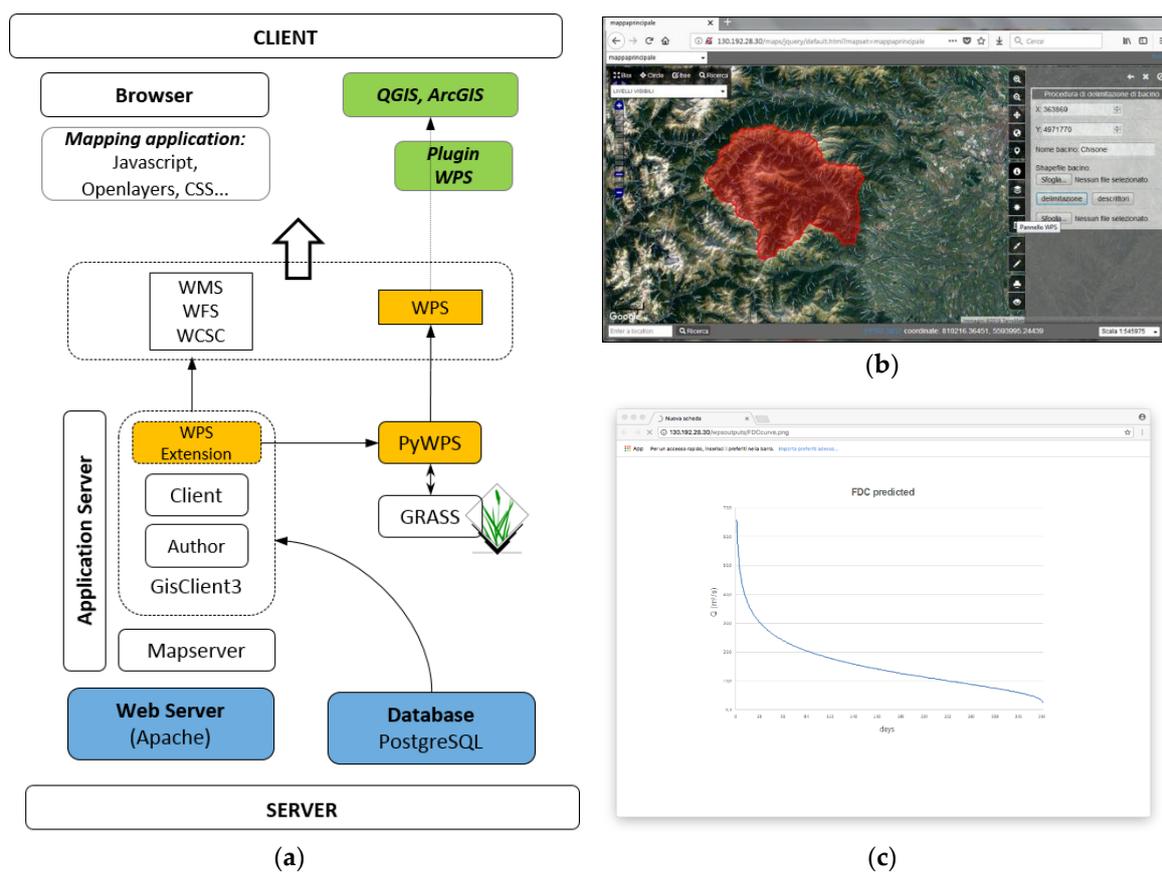
(a)

(b)

(c)

Figure 3. (**a**) The WebGis architecture. (**b**) Figure shows how users can access the procedure by web browser, connecting to the WebGis page. (**c**) Flow duration curve obtained from web browser.

In the following, a brief description of the individual tools making up the system is presented.

Web Processing Service (WPS). WPS is one of the OGC (Open Geospatial Consortium) specifications to provide access to GIS data or functionality over the internet in a standardized way. Other more common specifications provided by the OGC are for example the WMS service, that is used to send back a raster map requested by the client, or the WFS service, used to disseminate vector layers. WPS, instead, is used for serving and executing geospatial processes, algorithms, and calculations. WPS standard defines how a client can request the execution of a process, and how the output from the process is handled. The data that the process uses can be delivered across the network, or made available at the server and can include vector or image formats such as GeoTIFF, GML, KML etc.. Client applications work with a WPS service by appending parameters to the service's URL. The Request may be made as a HTTP GET, or a HTTP POST with an XML request document. The inputs and outputs required depend on the process being executed. Response is delivered as an XML document. There are three key requests that can be submitted to a WPS server. One is GetCapabilities, which first generates a metadata file as a XML document that describes the available processes offered on the server-side. The second is DescribeProcess, which provides more details and a description of a specific process (such as the necessary input data, the targeted output data format, as well as the service title and short abstract). Thirdly, the processing task can be submitted to the server by the Execute request that can answer to the client, through an ExecuteResponse, by returning the created output or storing the results as Web accessible resources.

GRASS (Geographic Resources Analysis Support System) is a free and open-source desktop GIS software. It can handle all geospatial data format such as: raster, vector, image, tabular data, etc. Furthermore, GRASS offers many spatial modeling algorithms focused on hydrological analysis. This software has an intuitive graphical user interface but can also be used through system command line. So, it is possible to use GRASS commands in other services through designated scripts. The GRASS

system data are organized in locations (defined by coordinate system, map projection and geographical boundaries) and mapsets. One reason to use different mapsets is to store maps related to project issues or to support simultaneous access by different users to the layers stored in the same location.

PyWPS is a Python-based WPS implementation. It provides a native support for many geospatial tools like GRASS GIS, R-Project or GDAL. Python is the most geopositive scripting language currently available, therefore most different GIS software take advantage of it.

GISCLIENT3 is an interesting web authoring tool configurator for PostGIS (a spatial extension for the PostgreSQL database) and MapServer and enables both to build up mapfiles and to provide OpenLayers maps. It is an open-source offered by GisWeb s.a.s. (Italian company in Genova, Italy).

3.2. Procedure

Operationally, the proposed procedure, for the estimation of the analytical FDC, requires to:
- evaluate, necessarily through GIS procedures, the geomorphological and climatic characteristics of the basin of interest;
- estimate the regional L-moments starting from the obtained basin descriptors;
- determine the distribution (Burr, Weibull, Pareto) and consequently the distributions parameters based on the Lmoments values;
- construct the annual flow duration curve.

3.2.1. Required input data

Calculation of catchment descriptors requires an input dataset containing pre-determined raster and vector maps, used and analyzed in that study. The most important input is the digital elevation model (DEM), which is necessary for the calculation of the catchment area, the average slope and the hypsographic curve. In the above mentioned work, a *hydrologically conditioned* raster DEM generated from NASA SRTM (Shuttle Radar Topography Mission) in 2000, was used. For *conditioned* DEM we mean a DEM corrected with enforcement technique in order to obtain an accurate flow drainage network from it. Other geospatial input data were set up in GRASS GIS database (e.g. Corine Land Cover vector maps, regime rainfall parameters maps).

3.2.2. Procedure implementation

The PyWPS code implemented can be divided into different segments successively described in this section. In a first step, the PyWPS procedure, keeping the basin shape inserted by the user and using PyWPS-GRASS-bridge, provides for: setting up a GRASS session, importing in GRASS the inserted vector data, extracting the values of basin descriptors from the predetermined raster and vector base maps using *r.univar* and *r.Stats* GRASS algorithms. After that, regional L-moments are estimated according to the regression models and thanks to *openpyxl* library the value of distribution parameters are extracted in relation to the L-moments value. Finally, with *Matplotlib* library, the FDCs graph is implemented. Once the process is completed, the server provides an *URL* pointing to the location of the results.

4. The geoinformation tool interface

The implemented WPS procedure can be called client-side both from the web-mapping application built and from GIS software thanks to the use of WPS plugins. For example, using a special WPS plugin for the open-source software QGIS, the user can execute the process choosing the basin from the list of vector layers displayed into the QGIS project. Process results can be returned into a results console (Figure 4). Figure 3b shows how the user can access the procedure using the WPS panel specifically developed on our WebGis page. In this case, after executing the process, the user can download the result (FDC curve) as an SVG file. In any case, the procedure also returns an URL that allows to see, in a PNG format, a preview of the FDC obtained in a web browser window (Figure 3c).
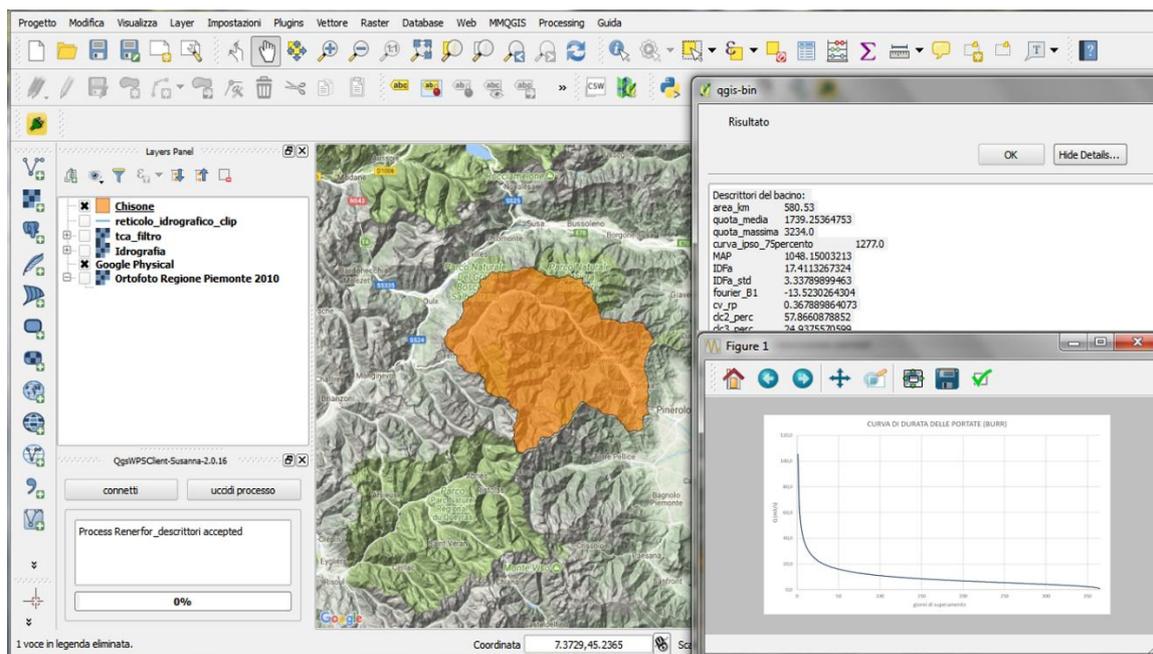
Figure 4. Result of the procedure called up from QGIS using a WPS plugin.

5. Conclusions

A sistematic construction of flow duration curves is a fundamental task for several activities related to water resources management and a useful tool for supporting the management of water systems, particularly in the climate change perspective. Flow duration curves generally need to be estimated and built for ungauged basins, using regional statistical methodologies. In this work a versatile WPS procedure for estimating regional FDCs curves based on PyWPS using GRASS for geoprocessing operations is presented. The tool has been implemented on a case study in North-Western Italy encompassing an area of more than 25000 km^2. The replication possibilities of the geoinformation framework is ensured by the data-based nature of the adopted statistical methodology and the structure of the WPS service, that allows fast remote updates. By the technological side, the use of a server-side scripting offers the great advantage that users can access data and calculations in real time and independently of the underlying software (that can be directly developed by the source organization, responsible for its maintenance), allowing moreover to re-use processes developed by the organization itself. Developments of the proposed web tool can be then envisaged for other regional-scale hydrological applications, as flood frequency analysis.

References

1. Ganora, D.; Laio, F.; Masoero, A. and Claps, P. Spatially-smooth regionalization of flow duration curves in non-pristine basins. Proceedings of the International Association of Hydrological Sciences, 2016, 373, pp. 73–80. DOI: 10.5194/piahs-373-73-2016. Available online: https://www.proc-iahs.net/373/73/2016/piahs-373-73-2016.pdf (accessed 10 Oct 2018).
2. Dai, A. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change* 2011, Volume 2, Issue 1, pp. 45-65.
3. McMahon, T. A. Hydrologicdesign for water use. In *Handbook of Hydrology*; Editor D. R. Maidment, McGraw-Hill, New York, 1993; Chapter 27.
4. Kim, D.; Jung, I. and Chun, J. A. A comparison between parameter regionalization and model calibration with flow duration curves for prediction in ungauged catchments, *Hydrol. Earth Syst. Sci. Discuss.* **2016**, DOI: 10.5194/hess-2016-487.
5. Ling, L.; Yusop, Z.; Askari, M. Assessment of Flow Duration Curve fitting equations. In book: *UTM Press Book*; Chapter Title: Stochastic hydrological modelling in water managementEdition; First Edition, 2015; Publisher: Universiti Teknologi Malaysia, UTM PressEditors, Associate Professor Dr Fadhilah Yus, 2015.

6. Nruthya, K. and Srinivas, V.V. Evaluating methods to predict streamflow at ungauged sites using regional flow duration curves: A case study. *Aquatic Procedia* **2015**, 4, pp. 641–648

7. Vogel, R.M. and Fennessey, N.M. Flow-duration curves. New interpretation and confidence intervals. *Journal of Water Resources Planning and Management* **1994**, 120, pp. 485–504

8. Laio, F.; Ganora D.; Claps, P. and Galeati, G. Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *J. Hydrol* **2011**, 408, pp. 67–77, DOI: 10.1016/j.jhydrol.2011.07.022.

9. Hosking, JRM and Wallis, J.R. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge, 1997.

10. Ganora, D. and Laio, F. Hydrological applications of the burr distribution: Practical method for parameter estimation. *J. Hydrol. Eng*. **2015**, 20, pp. 67–77, DOI: 10.1061/(ASCE)HE.1943-5584.0001203.