# A Web-GIS tool for rainstorm hazard management over large areas

Susanna Grasso[1,2], Andrea Libertino[1] and Pierluigi Claps[1]

[1] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino,
Corso Duca degli Abruzzi, 24, Turin 10129, Italy
[2] susanna.grasso@polito.it

## Abstract

Spatial analysis of rainfall extremes benefits of scientific advances and continuous data increments, that have led to the development of various methodologies in the last decades. In the field of the hydrologic design, due to technical, socio-economic and legislative factors, the availability of new methodologies does not imply the total substitution of the old, consolidated, procedures. This suggests that the hydrologic design is often supported by the opportunity to considering and comparing the results of estimates coming from different estimation methods. In this work, a tool named *MultiRain* is proposed, aimed at providing harmonized regional rainfall estimates for a given point and on areal basis. The software tool is based both on a QGis plugin and on a web-based WPS (Web Processing Service) procedure. Both the interfaces allow to build in a seamless way, from multiple models, the Intensity-Duration-Frequency (*IDF*) curves, either relative to a point or integrated over an area. The computation has been made automatic by manipulating old procedures lacking a map-based access. The procedures adopted for ensuring homogeneous comparison between the older ones and the more recent ones are described, with reference to three different methods and to their implementation over a $25,000\text{km}^2$ area in the North-West of Italy.

Keywords: Web-GIS, Rainfall extremes, Hazard maps, Model averages

## 1 Introduction

Risk assessment related to extreme precipitation has always been a priority of the hydrological sciences, due to the significant consequences of these phenomena on the people safety, the protection of the natural resources and the organization of human activities. Adverse hydro-meteorological events (including floods, mass movements and storms) have led to about 25,000 fatalities per year across the globe [1] and most of them relate directly on intense rainstorms. In scientific literature, many studies have been carried out, aimed at relating rainfall depths with a certain probability of occurrence.

Intensity-Duration-Frequency (*IDF*) or Depth-Duration-Frequency (*DDF*) curves are commonly adopted in water resources engineering for design and operation of flood protection infrastructures. They are mathematical relationships that allow one to estimate the average rainfall intensity *i* (or the average rainfall depth *h*) over a given timescale *d* for a given return period *T* [2] and the consequent design hyetograph (i.e., the rainfall sequence to be used in rainfall-runoff modelling). For the estimation of the *IDF* curves in ungauged sites, or averaged over an area, Regional rainfall Frequency Analysis (*RFA*) are required. They are usually built either by interpolating the data or the distribution parameters estimated at the station points (e.g., [3], [4]), or by pooling the data into statistically homogeneous regions (see, e.g., [5]).

Most of established *RFA* techniques are not framed in a quantitative geographic domain and this causes operational problems in their actual applications, particularly when areal values are required. In this work, a GIS operational tool named *MultiRain* is presented, aimed at producing concurrent regional rainfall estimates at a point or over an area. This tool allows to produce *IDF* curves from multiple existing *RFAs* for a given area, overcoming the operational limitations of the existing studies. The differences of the existing *RFA* studies, that prevent a straightforward comparison, are first analysed. The software tool is later described, with a particular focus on the solutions adopted for harmonizing the considered studies and allowing the applicability to the to the Piemonte region, in North-Western Italy.

## 2   Existing *RFA* methods in Piemonte

In this section, we introduce the operational problems that we had to face in the development of the tool, with reference to the Piemonte region, an area of about 25,000 km$^2$ in the North-West of Italy. In the region, three *RFA* procedures are currently available and all of them are in use, for the reasons described in the introduction. Despite they have been developed considering a more or less common database, the three methodologies adopt different regional statistical approaches and provide extreme precipitation products in a different way. In the following sections a brief description of the methodologies is provided, focusing on the characteristics that make a direct comparison quite cumbersome. This goal has been addressed by the developed *MultiRain* software.

### 2.1   The *PAI-PO* extreme rainfall maps

The "*PAI* Procedure" has been developed by the Po Basin Authority on a 60,000 km$^2$ area covering the Po drainage basin, that in essence includes all the Piemonte region. The analysis has been carried out on 277 rain gauges, distributed over the Po river basin. The time series of annual maxima of rainfall depth for short durations 1,3,6,12 and 24h from 1929 to 1989 collected from the Servizio Idrografico e Mareografico Italiano (*SIMN*) were considered. The study is based on the fitting of rainfall quantiles, that involves that independent *DDF* curves are estimated for the different return periods using the relevant quantiles obtained for the various durations. The basic relation reads:

$$h_T(d) = a_T \cdot d^{n_T} \tag{1}$$

with $a_T$ and $n_T$ coefficients estimated by linear regression among the different quantiles related to each return period *T* at each station point.

The curves at gauged locations have been estimated considering a Gumbel distribution, reporting the parameters for return periods *T*=20, 100, 200 and 500 years. The $a_T$ and $n_T$ parameters are then interpolated on a 2 km grid using ordinary kriging, for estimations at ungauged locations. The *PAI* procedure thus offers a tool for estimating the design rainfall at each gridded point, for the 4 considered return periods. The parameters are provided by means of a table in which the location of each cell is

obtained by sequential ordering of rows and columns indices. For a generic cell, the parameter can also be obtained by querying the online maps at [7].

## 2.2   The *VAPI* Rainfall Maps for NW of Italy

This procedure has been developed for the North-Western Italy by Rosso and De Michele [8] in the framework of the *VAPI* national project. They have considered a database of 366 stations (270 in the Po river basin and 96 in Liguria), with at least 20 years of observation and an average sample length of 34, not very different from the *PAI* database. The rainfall quantiles are estimated by a "index-rainfall method", in which the quantile of the extreme rainfall for a fixed duration $d$ and a defined return period $T$ (i.e., the design rainfall depth $h_T$) comes as the product of an "index value" $\bar{h}$ and a growth curve $K_T$:

$$h_T(d) \ = \bar{h}(d) \cdot K_T(d) \tag{2}$$

The index rainfall $\bar{h}$ can be reconstructed by a power law:

$$\bar{h}(d) = \bar{a} \cdot d^n \tag{3}$$

with $\bar{a}$ and $n$ as coefficients of the average *IDF* curve.

In this procedure the growth factor follows a *GEV* distribution [6], whose parameters are estimated using the *PWM* method for each station [9]. Ordinary kriging interpolation is used for interpolating at ungauged sites the parameters of equation (3) and of the *GEV*. The maps for North-Western Italy (including Lombardia, Piemonte, Valle d'Aosta, Liguria, and Emilia Romagna regions) of $\bar{a}$, $n$ and of the *GEV* parameters are available in [8]. The resolution of the results is not provided, as the map is given in false colours paper format. Local estimates are however available with reference to the sample stations.

## 2.3   The *Patched Kriging* rainfall maps

This procedure is the only one allowing to browse the digital maps [10], included in the "Atlante delle piogge intense in Piemonte". The study includes data from 1913 to 2010, gathered from the *SIMN* and the regional authorities, for the durations 1,3,6,12 and 24h. The methodology involves the sequential application in time of the ordinary kriging equations, producing a homogeneous dataset of synthetic series with uniform lengths [4]. On these series, the index method is adopted and the *IDF* curves are provided according to equations (2) and (3). The authors decided not to choose a "best" probability distribution and provide only the $\bar{a}$ and $n$ parameters of the *DDF* and the first three sample L-moments. This allows the end-users to choose the most appropriate probability distribution for their purposes and estimating parameters on their own. Results are provided on a 250 m gridded domain covering the territory of the Piemonte and Valle D'Aosta regions. Maps are re-sampled on a 1 km grid for the Piemonte region in [10].

# 3   A GIS tool for seamless multi-model rainfall estimates

Application of the above-presented methodologies provide different results, both at-a-point and over an area. In the first case, differences are much simpler to check, even though the older methods require some time to be applied. The tool presented here works in a geographic coherent framework that facilitates the selection of the point in which procedures must be applied. In addition, the *MultiRain* tool returns the areal-averaged *IDF* curves within a basin for all the *RFA* methodologies through the upload of the catchment boundary in vector format.

## 3.1   The *MultiRain* tool: methodological issues

In order to provide an IT tool based on an open-source technology we developed a Python plug-in for the open source software QGis. The main difficulties in the development of the *MultiRain* tool concerns the building of the geospatial database containing all the elements necessary for the application of the considered *RFA*s, due to heterogeneous source of the parameters (e.g. digitized scanned images, tabulated values, ascii raster with different formats and resolutions) and to the methodological differences between the various approaches (e.g., the homogeneous region procedures are based on vector maps while the methods that involve the continuous spatial variation of the parameters refers to raster maps).

More in detail, as the *PAI* methodology provides the $a_T$ and $n_T$ parameters of equation (1) on a 2 km grid, for *T*=20,100,200 and 500 years but not the average values $\bar{a}$ and $n$. To face this issue, cosidering that the procedure adopts a Gumbel distribution, for each cell of the grid we have extracted the values of the $\bar{a}$ coefficient for hourly duration and of the $\theta_1$ and $\theta_2$ parameters of the distribution, using the available quantiles related to *T*=20 and 200 years. In the detail, we exploited the similarities between equation (1) and (3), stating:

$$a_T = \bar{a} \cdot K_T \tag{1}$$

Combining (4) with the equation of the Gumbel distribution:

$$K_T = \left\{ 1 - CV \cdot \left[ 0{,}45 + \frac{\sqrt{6}}{\pi} ln \left( ln \left( \frac{T}{T-1} \right) \right) \right] \right\} \tag{2}$$

where $CV = \dfrac{\theta_2 (\frac{\pi}{\sqrt{6}})}{(\theta_1 + 0{,}5772\theta_2)}$ , comes from the parameter values below:

$$\begin{cases} \theta_1 = \dfrac{y_{500}a_{20} - y_{20}a_{500}}{y_{500} - y_{20}} \\ \quad \theta_2 = \dfrac{a_{20} - \theta_1}{y_{20}} \end{cases} \tag{3}$$

where the subscripts refer to the considered return period *T*, and $y = -ln\left[ ln\left( \frac{T}{T-1} \right) \right]$.

For the evaluation of the equivalent $n$ parameters, we exploited the substantial independence of $n_T$ on *T*. The (constant) $n$ parameters were then estimated for each cell as the average of the $n_T$ values. Raster maps of $\bar{a}$, $n$, $\theta_1$ and $\theta_2$ according to the *PAI* methodology have thus been built, and the Python procedure has been adapted to compute the rainfall quantiles with negligible differences with the original ones (they are only related to the simplification on the parameter $n$).

Regarding the *VAPI* method, the main issue concerned the need of digitizing the maps provided uniquely in a paper format (e.g., see Figure 1 (left)). To this aim we used an Image Processing procedure implemented in Matlab for extracting the colour information from the map, relating it to the provided colour bar and attributing a numeric value to each pixel of the image. The obtained images have then been noise-cleaned (e.g., unnecessary graphical elements shown in Figure 1 have been removed: the grid, the station indicators, etc.), converted in raster format, properly geo-referred and merged (two separate maps were provided for the Po basin and the Ligurian catchments). An example of the resulting map for the $n$ coefficient is shown in Figure 1 (right).

The maps related to the *Patched Kriging* methodology have been requested to the authors of [4] in a raster format. Using the provided sample L-moments, raster maps with a 250 m x 250 m resolution of the $\theta_1$ and $\theta_2$ parameters of the Gumbel distribution and of the $\theta_1, \theta_2$ and $\theta_3$ parameters of the *GEV* have been developed, considering $K_T$ independent on the duration. The maps have been added to the tool, together with the raster maps of $\bar{a}$ and $n$.
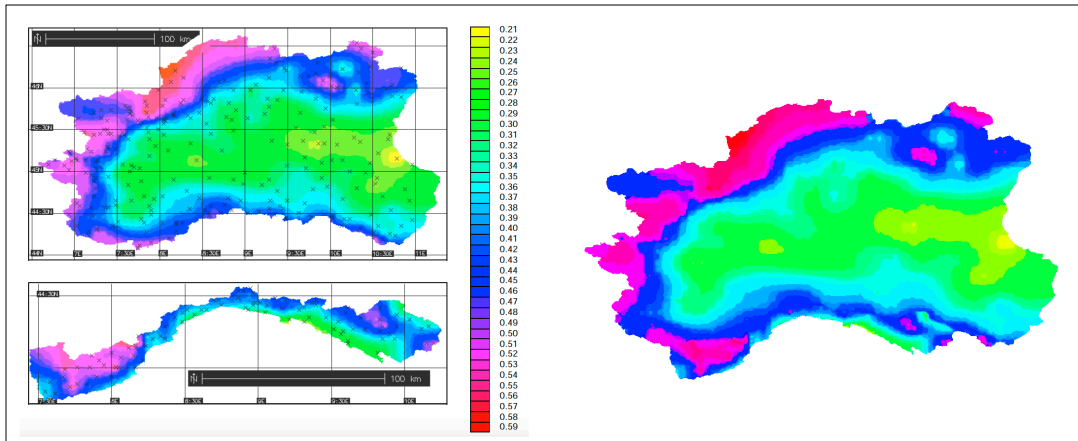
**Figure 1 :** Maps of the *n* parameter of the *VAPI* procedure in the original paper format (left) and in the digitized version (right).

## 3.2   The *MultiRain* tool: technological details

Once digitized, the maps required for the application of the different methodologies have been at first inserted in the plugin configuration package, for being used in the local version of the plugin.

Subsequently, with the aim to provide access to GIS data and functionality over the internet through standard internet protocols, a WPS (Web Processing Service) procedure has been implemented, accessible both by web browser and by some desktop GIS programs. The use of a WPS platform allows the users to access calculations independently of the underlying software and avoid the problem of updating the plugin to keep it compatible with the updates of the various GIS programs. Moreover, data does not need to be stored locally (client side) but are maintained by the hosting entity and can be updated and expanded on the server side, without the need for the users to download an updated version. Furthermore, loading times are usually faster than client-side scripting. For sharing this procedure, a web platform has been developed, with free and open-source software, using PyWPS to set up the WPS processes, GRASS GIS as a backend to access all the geoprocessing functionalities, GisClient3 to build the WebGis (accessible through client browsers) and Apache as web server. GisClient3 is an interesting web authoring tool configurator for PostGIS and MapServer that enables both to build Mapfiles and to provide OpenLayers maps. To be compliant with the WPS procedure all the maps have been uploaded in a GRASS database, to be managed by the procedure on the server.

The *MultiRain* tool, despite the differences in the scripting, has the same capabilities and provide the same user experience for both the developed versions (the QGIS script and the WPS one). The user can provide a point or an area of interest by uploading a proper shapefile or manually inserting the coordinates. In the first case the results are referred to the values of the local pixel, in the second the averaged curves among all the cells falling into the basin are provided. Both the tools execute the following operations for all the available methodologies:

- query of all the available maps, extracting the parameters (point or areal);
- calculation of the index rainfall $\bar{h}$;
- production of a text form the values of the local or areal parameters of the *IDF*s that allow to build curves for any return period *T;*
- production of a graphical representation of the *IDF*s for some significant return periods.

Moreover, if the user also specifies a return period *T* and a duration, the scripts:

- calculate the growth factor $K_T$;
- provide the design rainfall for the duration and return period requested;

- provide a graphical representation of the *IDF*s for the considered return period and duration obtained with the different methodologies.
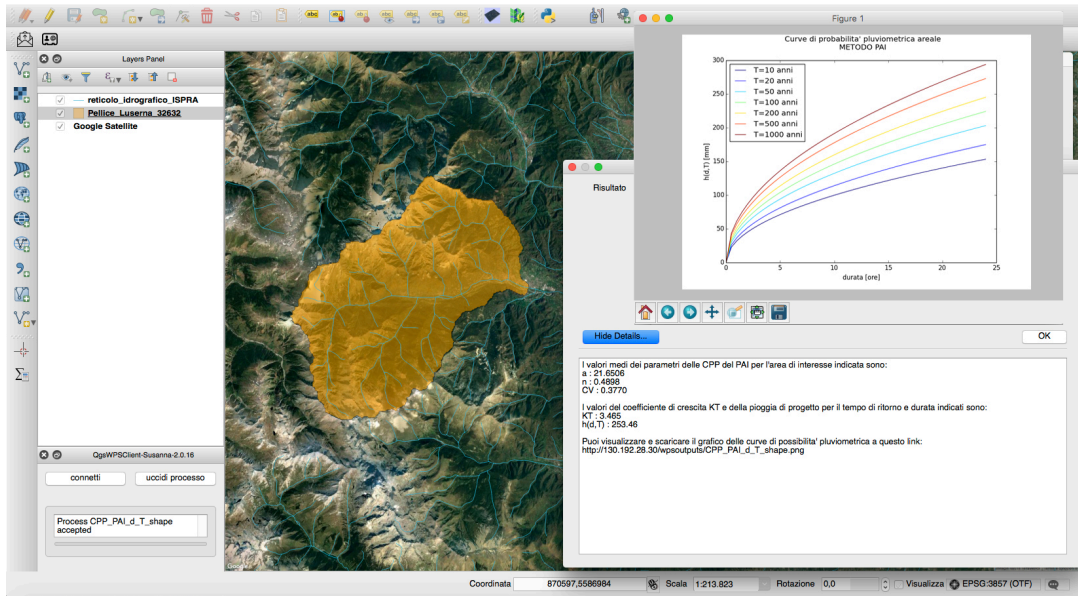


**Figure 2 :** Interface of the *MultiRain* tool in QGIS.

An example of the interface of the tool in QGIS is shown in Figure 2. The yellow basin of interest is inserted as a shapefile. Results for the *PAI* procedure are represented in the graph as areal *IDF*s for some significant return periods, while the areal parameters of the *DDF* and the resulting design rainfall for the selected *T* are reported in the textbox.

# 4  *MultiRain* tool application in Piemonte

A demonstration of the *MultiRain* tool application is presented in this section. Aside of the comparison of curves from different procedures, there are two non-trivial aspects that the application tends to show. The first one refers to the comparison of results as referred to a single station on the one side, and to an area on the other side. The second refers to the so-called *model averaging* (e.g., [11]) as a way to manage parameter and model uncertainty in hydrological statistics.

Let us first address a straightforward case of hydrologic design in a gauged site. The reference site is the station of Luserna San Giovanni, located close to a section of the Pellice river. The application of the MultiRain to all three methods produces four curves (the *Patched Kriging* model can produce estimates with different probabilistic models, and we adopt here both the Gumbel and the *GEV* model based on the L-Moments provided by the procedure).

Figure 3 (left) presents a case in which the curve related to one of the models is clearly separated from the others. This would entail further checking, but we can see that considering the areal curves the difference substantially reduces. To this end, Figure 2 details come into help, as it shows the Pellice basin boundaries above Luserna San Giovanni (the boundary delimitates an area of 216 km$^2$). In that area, individual curves are averaged in space, as can be seen in the exemplification of figure 3 (center). After the application of the spatial average on can notice in Figure 3 (right) that the differences among the curves related to the four methods have got into a substantial reduction.
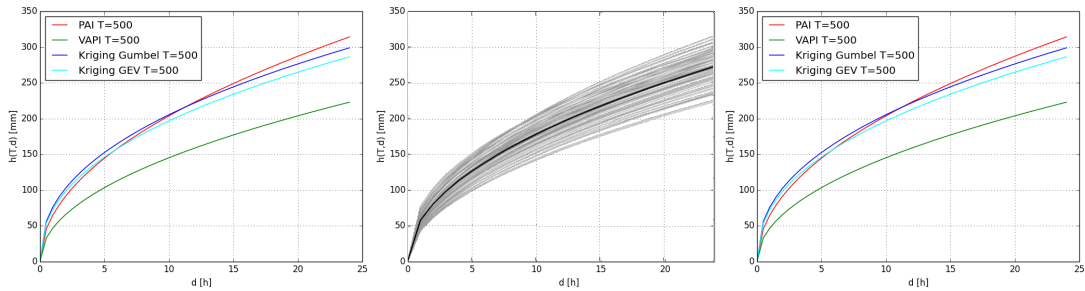
**Figure 3 :** (left) *DDF* curves for T=500 years for the Luserna San Giovanni station and (center) estimation of the areal DDF curve with the PAI model (black) by averaging of all the DDFs of the cells falling within the basin boundary (grey). (right) Areal *DDF* curves for T=500 years for the Pellice basin at Luserna San Giovanni.

For the sake of simplicity and robustness, a method to overcome the need for a detailed comparison of reliability of complex *RFA* frameworks is to adopt a *model averaging* among all methods. This application will be exemplified with respect to the verification of the rarity of an observed event: a significant rainfall occurrence measured in Luserna San Giovanni in November 2016 was then considered [12]. In that occasion, the west of the Piedmont region was subjected to a strong and persistent rainfall system that generated significant flooding in the Pellice river. To assess the rarity of the event, the maxima of the rainfall depths recorded for different durations were gathered, i.e. 17.6 mm, 39.4 mm, 75.6 mm, 140.6 mm, and 278.6 mm for durations of 1,3,6,12 and 24 hours respectively. These rainfall amounts have been related to the *DDF* curves provided by the *MultiRain* tool to assess the maximum return period of the rainfall event. This result comes after the input of the station coordinates (361006, 4963859 in the reference system WGS84 - UTM 32N). The analysis can be made with all the available methods and shows that the return periods for the 24 hours maximum depth ranges between 200 and 500 years, with the exception of the *VAPI* method that suggests a return period of over 5000 years.
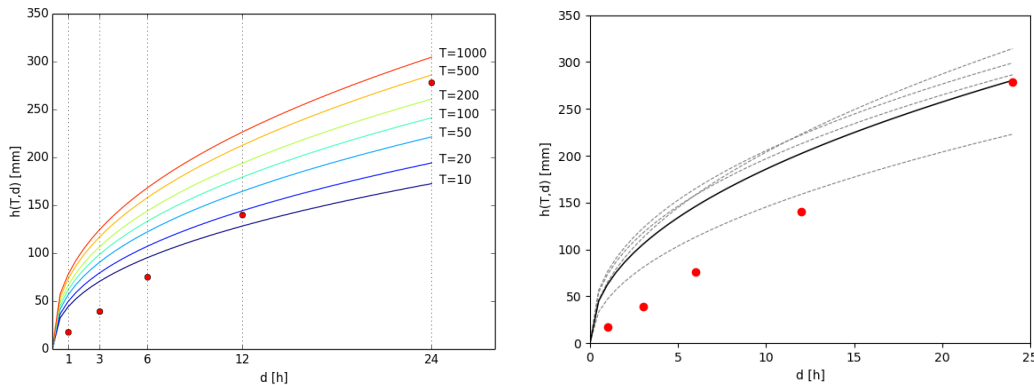


**Figure 4:** Comparison of the event-maxima recorded in Luserna San Giovanni in November 2016 with (left) the *DDF* curves estimated with the Patched Kriging and with (right) the model-averaged *DDF* curve for *T*=500 years (in black) as compared to the individual *T*=500 curves of the three models (in grey).

To exemplify and to give an idea of the return periods of the other rainfall depths of different durations, Figure 4 (left) shows the results of this assessment with the *Patched Kriging* method, with a Gumbel distribution built using the L-Moments provided by the procedure.

If the technician does not have the necessary knowledge to discriminate among the different procedures and then choose the 'best' one, he can provide a robust answer the technical question by considering the model-averaged curve. As a demonstration of the model-average result, the $T$=500 averaged curve is computed and shown in Figure 4 (right). The result of the superposition of the data with the curve shows that a robust estimation of the return period for the 24 hours duration is closer to $T$=500 than to $T$=200 or to $T$=5000 years.

# 5   Conclusions

Merging and harmonizing the information deriving from subsequent set of data and regional statistical estimation procedures is something required in many cases, particularly when hydrological design values face the updates of the estimation procedures.

Even if new approaches are generally based on more data, the institutions in charge of the infrastructure management are interested in analysing the implications of a change in the design values and/or in addressing the changes in the common designing practices. In this framework, the *MultiRain* tool presented here aims at providing a fast operational aid for an exhaustive multi-model approach to the regional rainfall frequency estimation. Accessible both via web-browser and via local GIS software, the *MultiRain* allows the user to carry out a thorough comparative analysis based on the results deriving from different methodologies, and to provide for a model average that can be of great help in many cases. Moreover, a straightforward spatial averaging procedure is accessible to the user by providing the vector file of the area of interest (usually a watershed divide). Additional developments of the tool are under test for extending its functionality to the whole area of Italy.

# References

[1]  Munich RE. NatCatSERVICE. Accessed: 2017-04-01. URL: http://natcatservice. munichre.com/

[2]  Koutsoyiannis D., Kozonis D. and Manetas A. "A mathematical framework for studying rainfall intensity-duration-frequency relationships". In: Journal of Hydrology 206.1 (1998), pp. 118–135.

[3]  Myers DE. "Spatial interpolation: an overview". In: Geoderma 62.1 (1994), pp. 17–28.

[4]  Libertino A., Allamano P., Laio F., Claps P. "Regional-scale analysis of extreme precipitation from short and fragmented records". In: Advances in Water Resources 112 (2018), pp. 147-159.

[5]  Hosking J.R.M. and Wallis J.R. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, 1997. ISBN: 9780511881640.

[6]  Jenkinson A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q.J.R. Meteorological Society, 81(348), 158-171.

[7]  Autorità di bacino distrettuale del fiume Po, Piano stralcio  per l'Assetto Idrogeologico  (PAI) , 2001. Web: http://pai.adbpo.it/, accessed: 01/01/2018

[8]  Rosso R. & De Michele C. (2001). Rapporto sulla valutazione delle piene Italia nord occidentale.

[9]  Hosking J. R., Wallis J. R. and Wood E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27(3), 251-261.

[10] ARPA Piemonte, Atlante delle piogge intense in Piemonte.

[11] Di Baldassarre G., Laio F. and Montanari A. (2009). Design flood estimation using model selection criteria. Physics and Chemistry of the Earth, Parts A/B/C, 34(10-12), 606-611.

[12] ARPA Piemonte, Rapporto d'Evento Novembre 2016.