Background Paper prepared for the Global Assessment Report on
Disaster Risk Reduction 2013

# Improvement of the Global Flood Model for the GAR 2013 and 2015

**Christian Herold**
UNEP-GRID

**Roberto Rudari**
CIMA Research Foundation International

Geneva, Switzerland, 2013

# FINAL REPORT

***Objective 1***: Improvement of the Global Flood Model for the GAR 2013 and 2015
***Expected outcome:*** Improved Statistical Analysis for Discharge quantile regression
***Indicators of achievement:*** Final Report and Regression dataset

## 1.1 Background

Global Flood Modelling is one of the targets pursued within the UNISDR Global Assessment Report. The current approach, though valid, has some limitation that can be overcome by incrementally improving the existing model and changing it in some parts.

One of the key improvements identified for the GFM resides in having a more robust identification of the quantiles of maximum monthly stream flows on one side as well as providing multiple quantile evaluations referred to different return periods on the other. This final improvement is a necessary precondition for future application of the GFM results to the CAPRA approach.

The methodology applied in the last edition of the GAR develops on the following work flow:
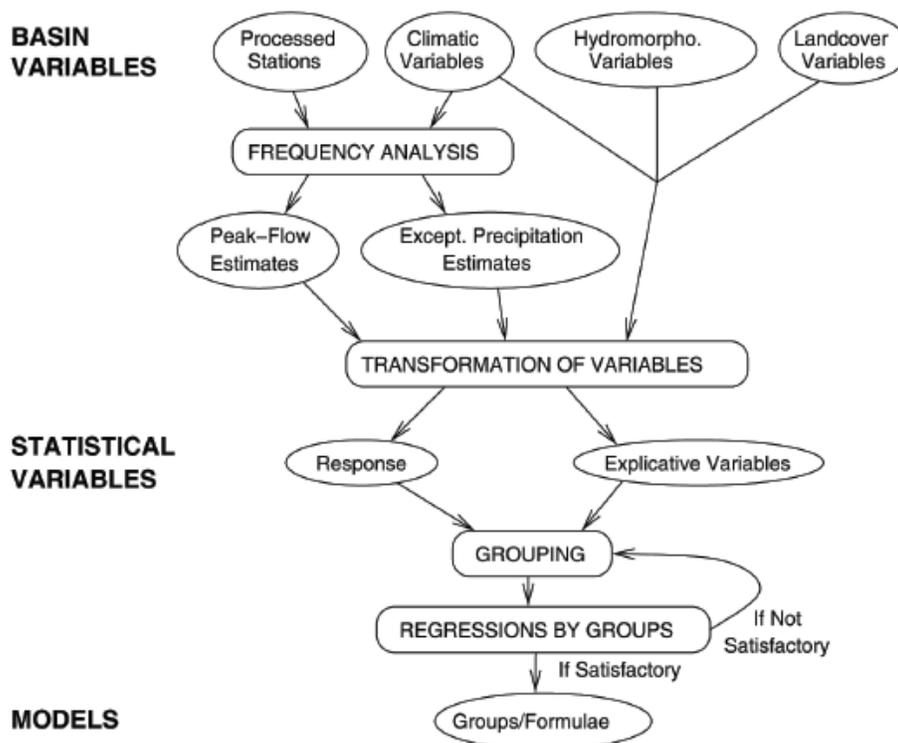


**Figure 1 workflow for quantiles determination GAR 2009 & 2011**

One of the weak points of this workflow is that the statistical evaluation of the quantiles for assigned return period is done at each single site, where observations are available, and the regression is performed on the quantiles through logarithmic regression on the basin variables. In this way for each return period a new regression is needed to transfer quantiles estimations

to sites where observations are not available. Secondly, the robustness of the quantile estimation is dependent on the single site observed time series, which are not often long enough to offer a reliable estimate of rare quantiles.

## 1.2  Methodology

The methodology here proposed is based on a different approach that analyzes all statistical data of an identified homogeneous region within the same statistical framework and tries to overcomes both limitations.

Regionalization techniques for discharge quantiles estimation are a common procedure in hydrology (REF#). Different approaches exist for Regional Flood Frequency Analysis: e.g. Quantiles Regression Method (QRM), Parameter Regression Method (PRM), Index Flood Method (IFM) (see e.g., Tailor et al., 2011). Each method has advantages and drawbacks. However, the specific advantages offered by IFM are crucial assets in the case of the GFM development that has in its workflow the need to estimate at-site flood frequency with the highest confidence possible even for very low frequency quantiles.

At gauging stations, accepted statistical methods can be applied to evaluate flood discharge magnitude for given probability (see, e.g. Bobe´e et al., 1995; Bras, 1990; Chow et al., 1988). They are usually based on the availability of annual discharge maxima (ADM) time series (in this case on a monthly time window), locally recorded for a sufficient number of years. In other cases, and particularly when records are short, regional analyses are used to extend in time available local observations by merging them in one single time series, with the purpose of increasing estimates reliability especially when low frequency quantiles are of interest (Gabriele & Arnell, 1991; Kottegoda & Rosso, 1997). These methods produce regional growth curves for reduced variables. Working with reduced variables makes the different time series statistically comparable. In case they are proven to come from the same statistical distribution through statistical tests they can be analyzed as a single time series having as lengths the sum of the lengths of the single time series. The resulting growth curve counts on a larger sample size.

The so called 'index flood', usually estimated by the ADM expected value based on local observations, is then used to render regional curves dimensional with reference to specific sites. For reasoning on index flood estimation methods see Bocchiola et al., 2003. In un-gauged sites, i.e. sites for which no stage-discharge station and relative records are available, flood frequency can be only estimated by extrapolation of the frequency evaluation made for gauged sites. Since monthly annual maximum probability is of concern, methods used to estimate the magnitude of flood discharges associated with given frequencies can be grouped into two categories. The first collects all methods based on statistical or regression analyses, performed on data pertaining to the same hydrologic region, e.g., recorded ADM at gauging stations, drained area, channel slope, basin shape, location and elevation. Regional frequency analysis belongs to this group. The most diffused regional techniques are the direct regression and the index flood ones (see, e.g., WICP-ACWI, Hydrologic Frequency Analysis Work Group, http://acwi.gov/). The second groups all methods based on the modeling of relevant hydrologic processes, such as rainfall‐runoff, flood formation, and flood propagation (Bougthon & Droop, 2003). A combined rainfall-runoff approach is in many cases more informative, especially in river basins with anthropogenic impact. However, due to the global nature of the work it is difficult to imagine an implementation of this second type at this stage given the time constraints of the overall project, while the first group of methods remains the more straightforward choice.

Within the panorama of regionalization techniques the IFM show advantages that appear crucial in this context. As already stated, sample size used for estimation increases as

homogeneous group time series can be used together forming a longer non-dimensional time series. As a second advantage, the performance of the regression on the basin variables is expected to improve for the expected value compared to higher quantiles (resembling extreme events), as it is expected to be more intimately linked to the local climatology. Climatology indicators can be computed world wide and can be used to set up the regression. A third advantage is that only one regression on the expected values needs to be computed, while the growth curve (non-dimensional CDF) will remain valid for the statistically homogeneous areas (i.e., Station Groups). In this way, once the index value regression is available any quantile can be derived by the growth curve.

Similar disclaimers with respect to the 2011 implementation hold. As this method is based on large river discharge time-series, it is supposed to represents events that affect corresponding floodplains. The model is not expected to properly represent events triggered in different conditions, for instance costal or flash flooding. The final maps have to give satisfactory results in the case of this new undertaken global risk analysis. It will not provide level of precision required for local analysis or land use planning.

## 1.3  Data Used

### 1.3.1  River discharge datasets:

As for the 2011 implementation, the discharge station dataset is based on various data sources providing station time-series of monthly mean discharge values. These sources provide various compilations of national or regional station datasets. In places where spatial coverage is still considered as low, some efforts have been made again to obtain data directly from national authorities, but with an unsatisfactory success rate.

Finally the new dataset is similar to the one used in 2011, except that GRDC recent updates were included in the final compilation.

- Long-term mean monthly discharge dataset. The Global Runoff Data Centre (GRDC), 56002 Koblenz, Germany.

- R-ArcticNET, A Regional, Electronic, Hydrographic Data Network For the Arctic Region. Water Systems Analysis Group. Complex Systems Research Center. Institute for the Study of Earth, Oceans and Space. University of New Hampshire.

- The Global River Discharge Database (RivDIS v1.1). Water Systems Analysis Group. Complex Systems Research Center. Institute for the Study of Earth, Oceans and Space. University of New Hampshire.

- Monthly Discharge Data for World Rivers (except former Soviet Union). DE/FIH/GRDC and UNESCO/IHP, 2001: Monthly Discharge Data for World Rivers (except former Soviet Union). Published by the CISL Data Support Section at the National Center for Atmospheric Research, Boulder, CO (ds552.1).

- Russian River Flow Data by Bodo, Enhanced. Monthly river flow rates for Russia and former Soviet Union countries in ds553.1 are augmented with data from Russia's State Hydrological Institute (SHI) and a few sites from the Global Hydroclimatic Data Network (GHCDN).

- Discharge of selected rivers of the world. World Water Resources and their use, a joint SHI/UNESCO product. International Hydrological Programme. UNESCO's intergovernmental scientific programme in water resources.

**Global distribution of available river discharge stations:**

AVAILABLE STATIONS
● 7552 gauging stations
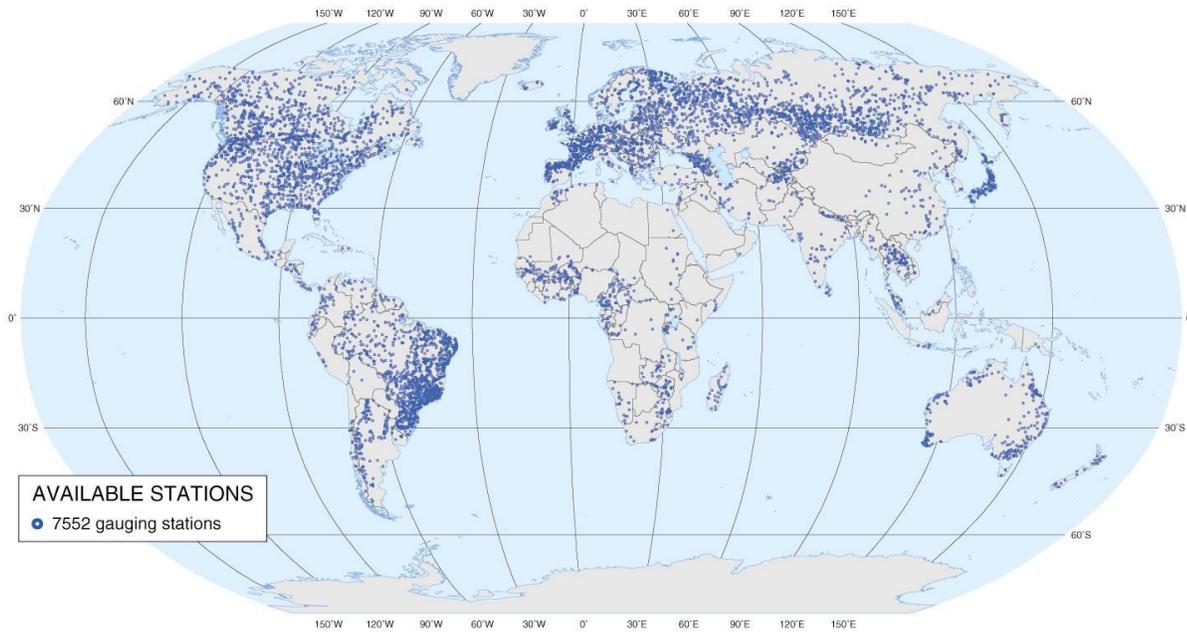
### 1.3.2 Reservoir and Dam Database

The access to a global dam database is essential to estimate the effect of dam networks on downstream river flow.

- The Global Reservoir and Dam Database (GRanD), Global Water System Project (GWSP).

### 1.3.3 Digital Elevation Model and hydrological derived datasets:

The whole GIS processes are now based on SRTM Digital Elevation Model at 90 meter resolution and other important derived products. The processes and spatial treatments previously based on HYDRO1k Digital Elevation Model (1 kilometer of resolution) show substantial improvement in term of output precision.

- NASA Shuttle Radar Topography Mission (SRTM) SRTM version 2. National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA).
- NASA Shuttle Radar Topography Mission (SRTM) Water Body Data. National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA).
- HydroSHEDS, WWF. In partnership with USGS, CIAT, TNC, CESR.

### 1.3.4 Land cover datasets:

Land cover dataset are used to generate specific basin characteristics for statistical analysis, and to estimate roughness coefficient required by the hydraulic model as an input parameter.

- Global land cover GLC_2000 version 1. Institute for Environment and Sustainability, Joint Research Centre.

- ESA's global land cover map 2009. ESA/ESA Glob Cover Project, led by MEDIAS France/Postel.
- Global Lakes and Wetlands Database (GLWD). WWF and the Center for Environmental Systems Research, University of Kassel, Germany.

### 1.3.5 Climatic datasets:

Following climatic datasets are used to generate specific basin characteristics for statistical analysis.

- CRU TS 3.1 monthly precipitation. Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena.

- CRU TS 3.1 monthly mean temperatures. Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena.

- Updated world map of the Köppen-Geiger climate classification. The University of Melbourne, Victoria, Australia.

### 1.3.6 Recorded flood event dataset:

Flood events footprints are used to validate results in term of flood extent for specific return periods.

- World Atlas of Flooded Lands. Dr. G. Robert Brakenridge, Ms. Elaine Anderson. Dartmouth Flood Observatory.
- Specific recent floods footprints merged from different Satellite Imagery (e.g., in the case of the 2011 Thailand flood), courtesy of UNOSAT (Unitar's …)

### 1.3.7 Data preparation

Standard GRASS hydrological functions are applied on HydroSHEDS digital elevation model to generate the stream network and derivates layers. Geographic projection is not modified but a real surface raster, based on an ellipsoid, is used during these treatments to maintained real drainage area in the various outputs. This real surface raster is used when needed in subsequent spatial analyses to avoid re-projecting layer in an equal-area projection.

Stations are individually adjusted on stream network comparing recorded and modelized drainage area values. When a difference threshold between these two values is reach, the station is considered as situated on the right stream section. This treatment is essential for establishing spatial correspondence between river station as recorded in a database, and the digital elevation model and its derivates hydrological layers. Additionally, this process allows excluding some stations, in case of duplicates or inappropriate location on the stream network.

A similar procedure is applied on dam database in order to better fit the dam point to the adequate stream section.

Each station drainage basin is then characterized by a set of variables based on above described global datasets:

**Hydromorphometric:**
- Drainage area
- Mean basin elevation
- Mean basin slope
- Basin shape
- Main channel length
- Main channel slope
- Drainage frequency
- Distance to final outlet

**Land cover:**
- Surface water storage
- Forest cover
- Impervious cover

**Climatic time-series:**
- Mean annual precipitation
- Temporal mean of monthly maximum precipitation
- Minimum mean monthly temperature

**Climatic zones:**
- Percentage area of Köppen-Geiger climatic zones

**Upstream dam network:**
- Dam characteristics
- Temporal mean of monthly maximum precipitation in dam network catchment

## 1.4  Methodology implementation

### 1.4.1  Single site distribution fit and choice of the "parent" distribution

The first step in the methodology implementation is to identify a proper statistical distribution for the description of the single site data and the growth curve after that. Various distributions have been considered parent distributions and several distribution fits were performed on the different gauged stations with reasonably long time series (see e.g. Figure 2). Eventually the choice fell on the GEV distribution (see e.g., Hoskin and Wallis, 1993), that scored better in the overall distribution fit exercise and represent a good compromise between flexibility and robustness.
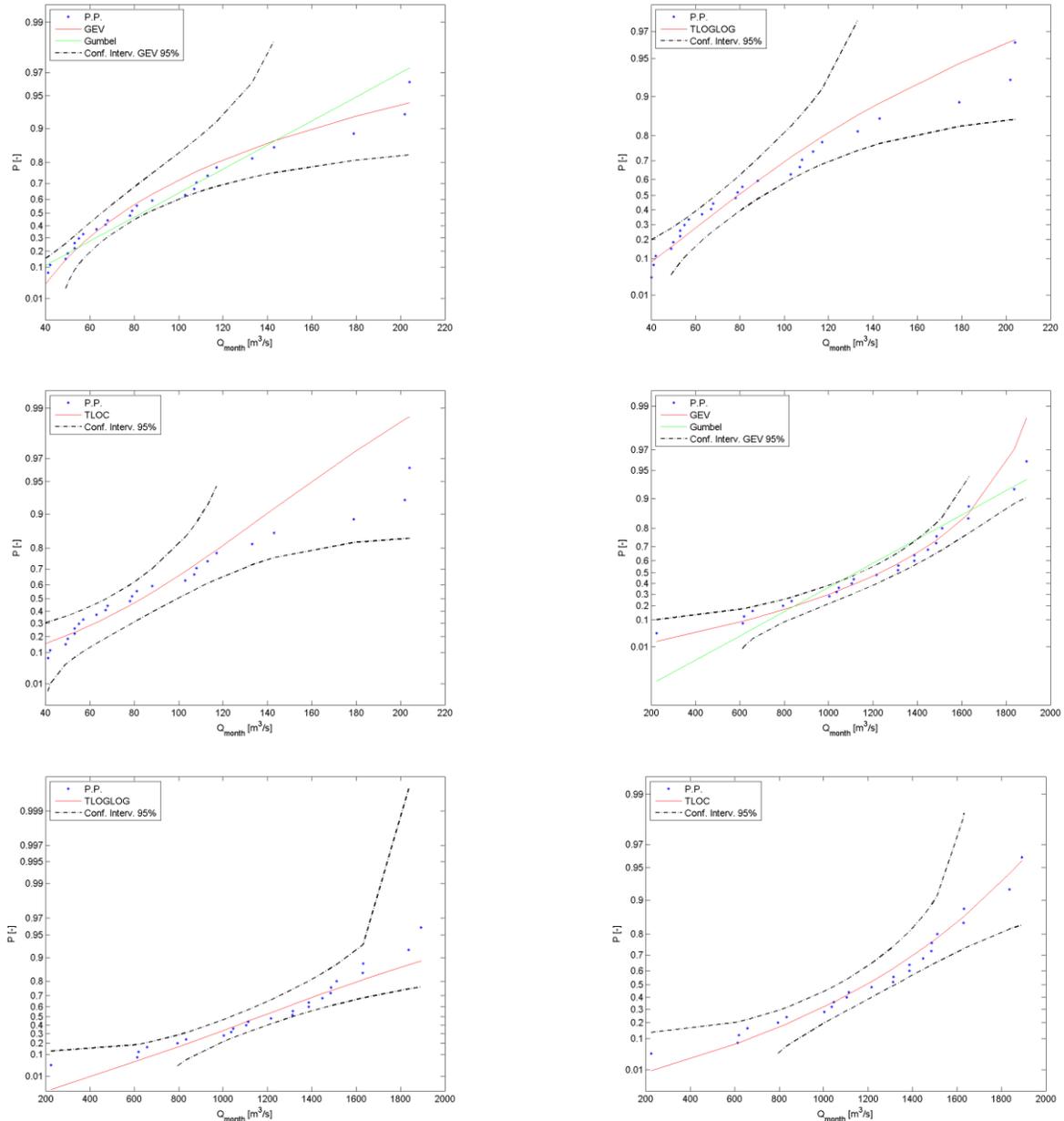
**Figure 2 Different at site distributions fit to data. Here a comparison for two stations and some statistical distributions are shown.**

The main reasons for this choice can be summarized as follows:

- It is widely used for the description of extreme values of physical processes (REFs#)
- It has three parameters and therefore allows for a good flexibility so that can it be adapted to a large casuistry of observed distributions
- It is easy to implement since an explicit formulation of the quantiles exists

The Cumulative Distribution Function (CDF) is given by the following equation:

$$F_X(x) = \Pr[X \le x] = \exp\left\{-\left[1 - \frac{k(x-\varepsilon)}{\alpha}\right]^{1/k}\right\}$$

Where k, α, ε represent respectively the shape, position and scale parameters.

The GEV distribution fit is carried out on each site where a reasonable length of the sample is available. After an analysis of the available samples we considered N=20 as the minimum number of years of data.

Figure 3 and Figure 4 show the curve together with the data for the station considered, the confidence intervals are also reported. The two sections located in South Asia. It is evident that in single site estimations for quantiles above .9-0.95 the confidence intervals tend to explode, providing a very uncertain estimation of the T=100 years quantile and above.

When the regionalization approach is used and the growth curves are considered this problem is considerably smoothed out. This will be clearer in the following paragraphs.
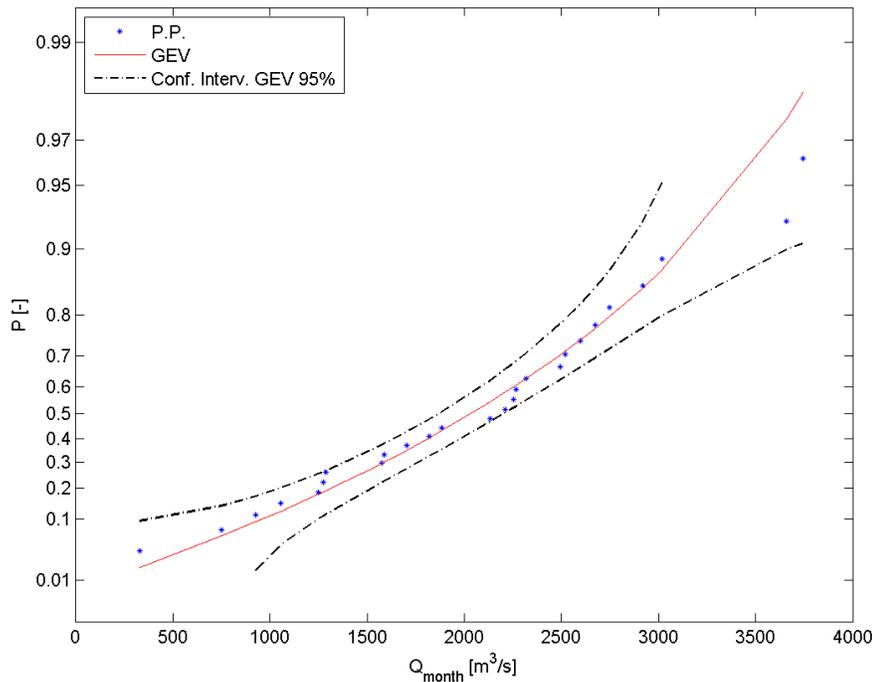


**Figure 3 Distribution fit to data. Station n° 2469120; Gev distribution**
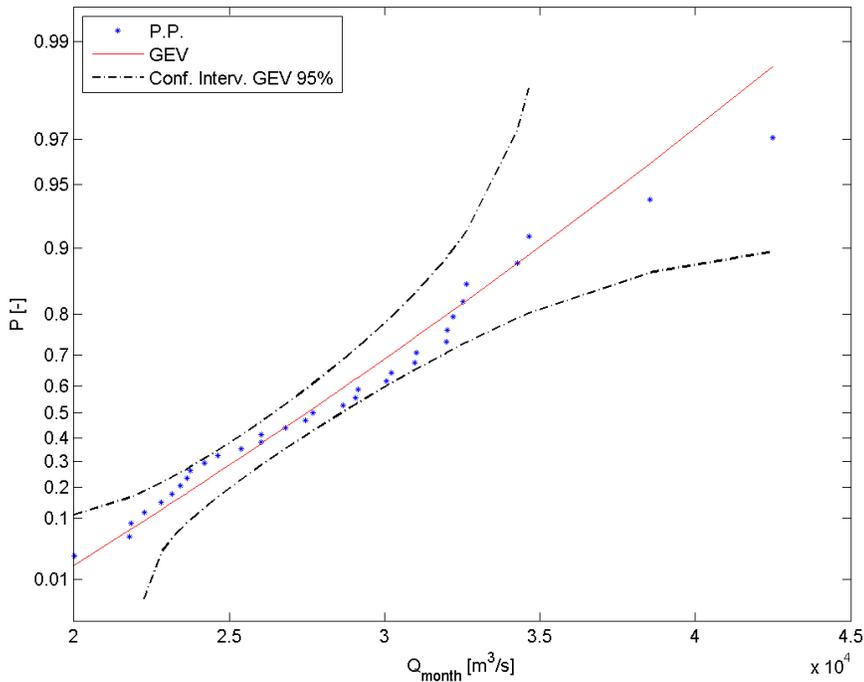
**Figure 4 Distribution fit to data. Station n° 2469260; Gev distribution**

## 1.4.2 Identification of the homogeneous regions

The homogeneous regions identification is the most difficult, and less objective step of the regionalization procedure. In this study they have been determined starting from considerations and analysis about climatology and hydrological regime of the basins pertaining to the region.

As a start the Köppen Geiger (KG) classification (Peel et al., 2007) has been used to identify the main climate zone of the considered area. Five primary classes from the KG classification have been considered: Tropical, Arid, Temperate, Cold and Polar as depicted in Figure 5.
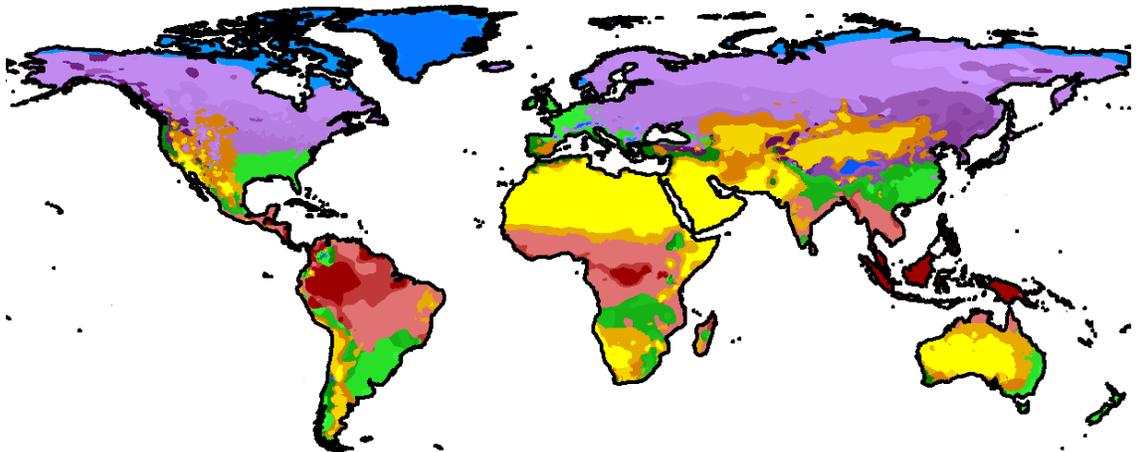


**Figure 5 Map of Köppen Geiger climatic zones classification: tropical in red tones, temperate in green tones, arid in yellow tones, cold in violet tones and polar in blue tones.**

This classification has been used as main driver of the final classification. The classification need to modified as it does not refer to consistent hydrologic units, namely basins and sub basins. The hydrologic response of the basins and as a consequence on its rain regimes has to be considered. In fact, basins that are on different climatic zones might show a similar behaviour in terms of hydrologic response, especially when they are located in transition areas between two main climatic zones. On the other hand, the same climatic zone can have sub-zones with different rain regimes, these differences are due not only on the different precipitation amounts that occur on predefined temporal windows (e.g. 1 year, 1 month…), but also on the variability of these amounts. This can reflect in a noticeable variability of the hydrologic response. Hydrologic response variability is a key factor in homogeneous zone identification.

The variability of the hydrologic regime can be synthetically described by the Coefficient of Variation[1] (CV) of the maximum monthly flow series (QMM). High values of CV indicate that there is a high variability of QMM so that QMM values can be very different from one year to another. This parameter can give indication of the driving rainfall regime of the area as arid climates are expected to show more variability than humid climates.

In Figure 6 the map of CV obtained by interpolation of the stations the CV is reported

By comparing Figure 5 and Figure 6 it can be notice that in many cases that pattern of CV follows the climatic KG classification. As an example it is evident how the arid zones are often characterized by high values of CV. It is nevertheless evident that other cases this correspondence is much less clear as the hydrologic interactions are more complex, as an example the north coast of Black sea as well some areas of Siberia show high CV but they are in cold climatic zones and that is of no trivial interpretation.
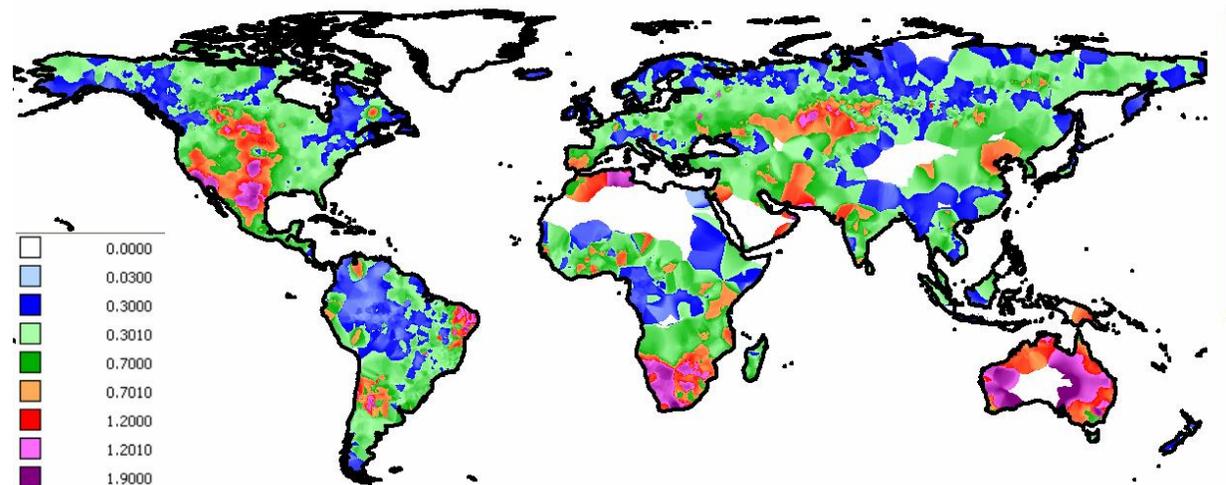


**Figure 6 Map of coefficient of variation of the maximum monthly flow**

This is confirmed by the analysis carried out following the work of Burn (1997) and is summarized in Figure 7. The 12 sectors represent the months of the year while the distance from the centre of the circle is a function of the CV value: the closest the point to the centre the smaller the CV value, the closer to the circle the higher the CV. The graph is built using the data of the North Hemisphere.

The points have different colours according to the KG climatic area. The Climatic area is assigned to the section analyzing its upstream catchment. The climatic area with the maximum percentage of area of the catchment is considered as the dominant area at this stage (if a basin belongs for the 30% to tropical and 70% to temperate is classified as temperate).

[1] CV is espresse as the ratio between standard deviation and mean of the QMM series

The graph shows the distribution of the QMM along the year and their variability; it is quite evident that arid basins have generally high values of CV, the temperate and cold zones have a pronounced seasonality, concentrating the QMM in well defined periods of the years
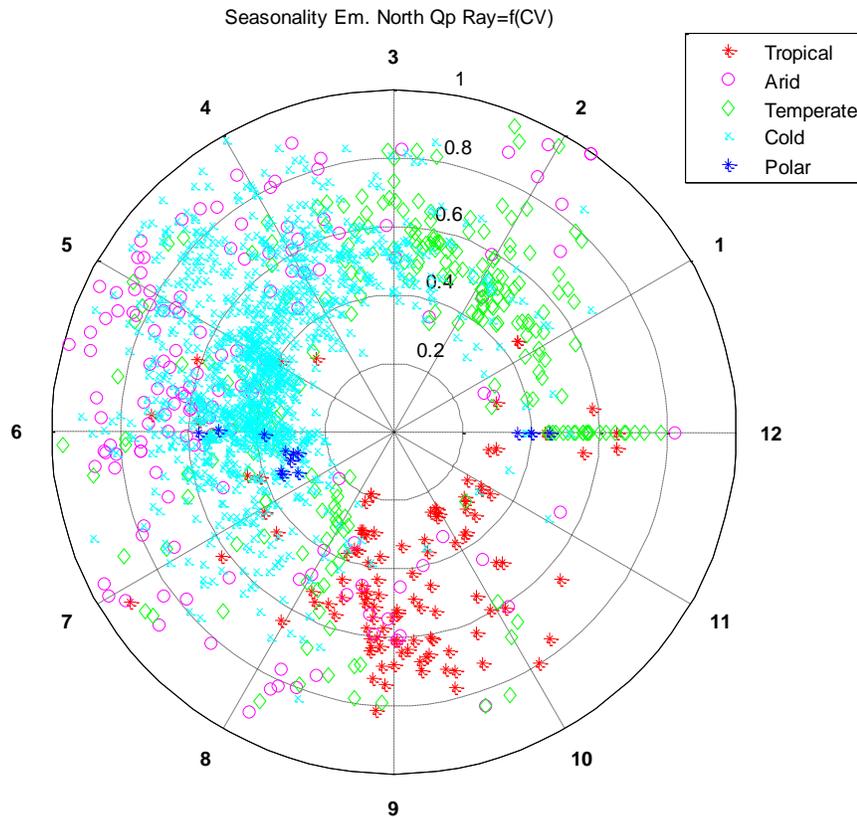


**Figure 7 Seasonal and CV spatial representation Map for the stations of the North Hemisphere.**

Finally, to maintain as often as possible the homogeneity at basin scale, we tried to avoid designing homogeneous regions that cross the basins. This is evidently not always possible, especially when large basins that cross regions that have very different climatology and precipitation regime are considered.

Summarizing, three main criterions have been followed to individuate the homogeneous regions:

1. the Köppen Geiger climatic classification;

2. the rain regime of the region, minimizing the CV variability in the region;

3. minimizing the number of basins that belong to more than one homogeneous region.

The naming of the homogenous regions is given basing on the following rules:

- the first number indicates the main continent;

- the second number the main climatic zone;

- the third number the sub zone (if existing).

Continents:

1:North America; 2: South America; 3: Europe; 4: Africa; 5: Asia; 6: Australia.

Climatic zones:

> 1:Tropical; 2: Arid; 3:Temperate; 4:Cold.

Example:

> Region 531; 5:Asia; 3: Temperate; 1: first Sub-Region

A unique polar zone has been considered for the entire world.

### 1.4.3 Final Section assignment to homogeneous regions.

The problem of assigning a certain section along the stream network to a homogeneous region is of concern when the river crosses more than one homogeneous region. These region are delimited by geographic boundaries and it is not infrequent that the upstream basin of a certain section lays for its larger part in a different region in respect the region of the section.

We needed an automated procedure to assign each section to a homogeneous region. The easiest way would be to assign the region where the largest part of the upstream basin pertains (as done in the first step). However, the hydrological behaviour in terms of discharge is in fact strongly influenced by those parts of the basin where the largest amount of precipitation is collected.

The adopted methodology is therefore the following.

Be A1, A2,..Ai...An the part of drainage area A of a certain basin closed in the section s that crosses the homogeneous regions 1,2,..n.

Be P1, P2,..Pi...Pn the mean annual precipitation occurring in the areas A1, A2,..Ai...An as derived from observations.

The homogeneous region assigned to the section s ($HR_s$) is the one that satisfies the equations:

$HR_s=imx$ with $imx$ such that $P_{imx}*A_{imx}=\max(P_i*A_i)$.

Once the homogeneous regions are defined all stations pertaining to that area are grouped, rendered dimensionless so to build a time series. On the basis of the time series it is possible to estimate the parameters of the GEV and build a growth curve of the area that describes the growth factor. An example is given in Figure 8 while Figure 9 shows a synoptic overview of growth curve fit to data for the different homogeneous region
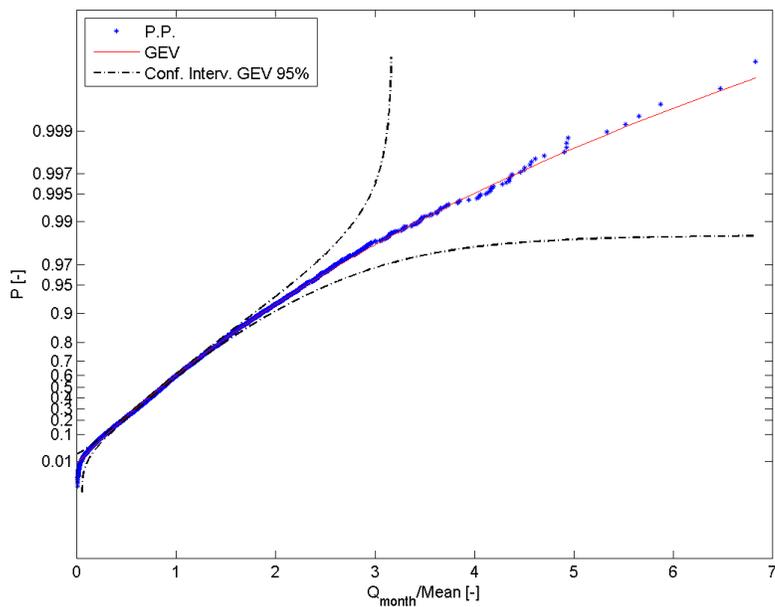
**Figure 8 Example of growth curve fit to data. Group 521 (Asia Arid, Sub Region 1) – GEV distribution**
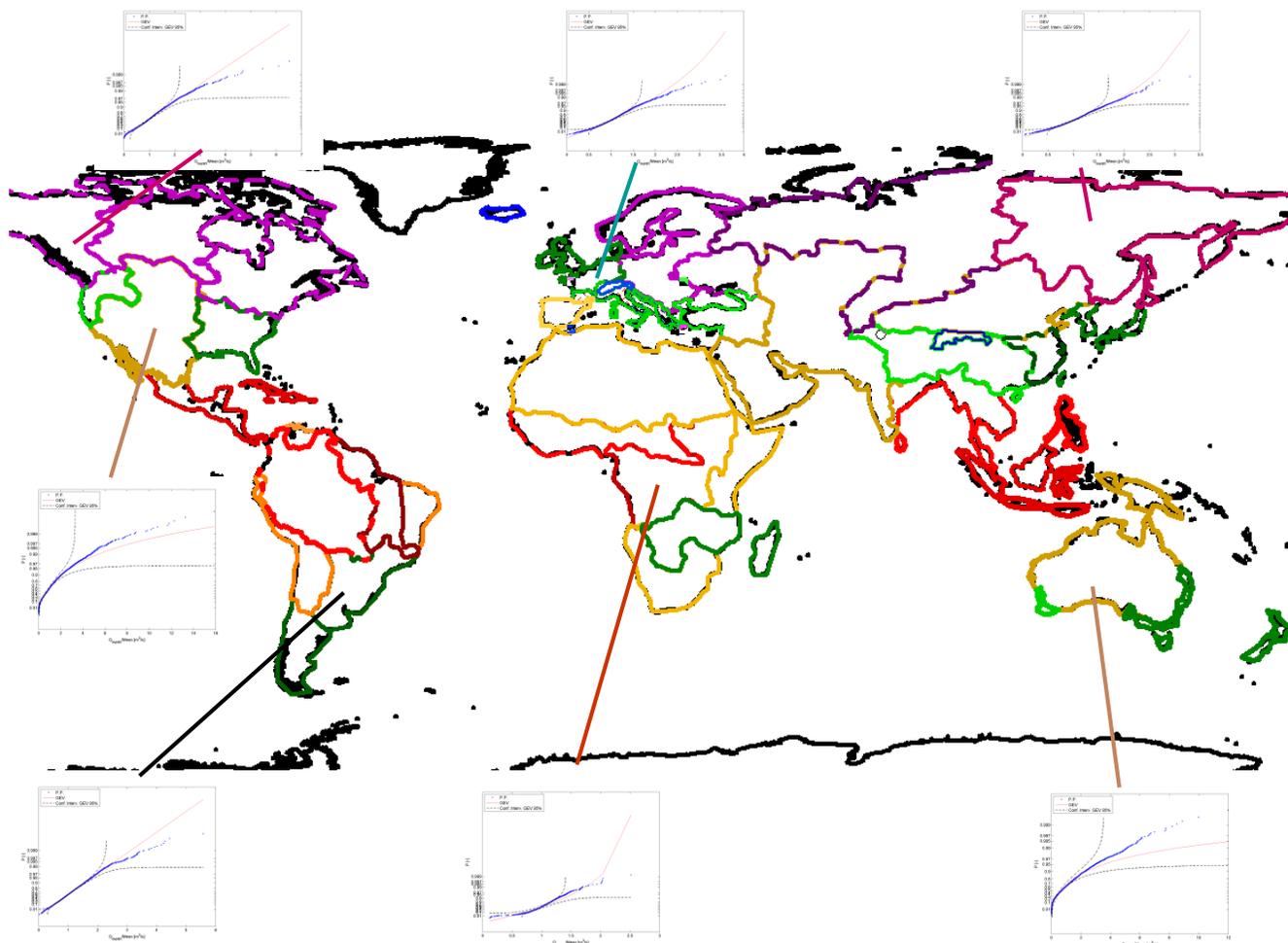


**Figure 9 Synoptic overview of growth curve fit to data for the different homogeneous regions.**

### 1.4.4 Homogeneity testing:

Dalrymple (1960) recommended a test which compares the variability of 10-year flood estimates from each site in the region with that expected if sampling error alone were responsible for between site differences. There have been several applications of this test: Dalrymple (1960) in Pennsylvania and Maryland, USA; Cole(1966) in England and Wales; Biswas & Fleming (1966) in Scotland; and Chong & Moore (1983) in Illinois. In these applications the regions being studied have been found in each case to be homogeneous suggesting either that the test may not be particularly powerful or that a wide variety of flood series and of basin types is consistent with homogeneity.

A more powerful statistical test will involve homogeneity of the analyzed region in terms of higher order moments such as CV and Skewness. Particularly, a homogeneous flood frequency region will contain annual maximum monthly flow populations whose flood frequency relationships have similar slopes on a probability plot. Therefore, variations of Coefficient of Variation (CV) and Skewness (Sk) should be attributed only to sample size limitations. This could be tested via Montecarlo simulations once the parent distribution (growth curve) have been computed. In the specific, series of sample size similar to the observed ones are created from the parent distribution and then treated as observations. A Chi Square test is then used to test if the observed distribution of CVs and Skewnesses can be statistically distinguished from the synthetic one. If they cannot distinguish it is assumed that the observed data can be generated by the same parent distribution and therefore the homogeneity test is passed. The homogeneity test was passed without problems by many initially identified regions (Figure 10 - Figure 11), while others need to be divided in to sub-groups to pass the test, thus refining the homogeneous areas identification.
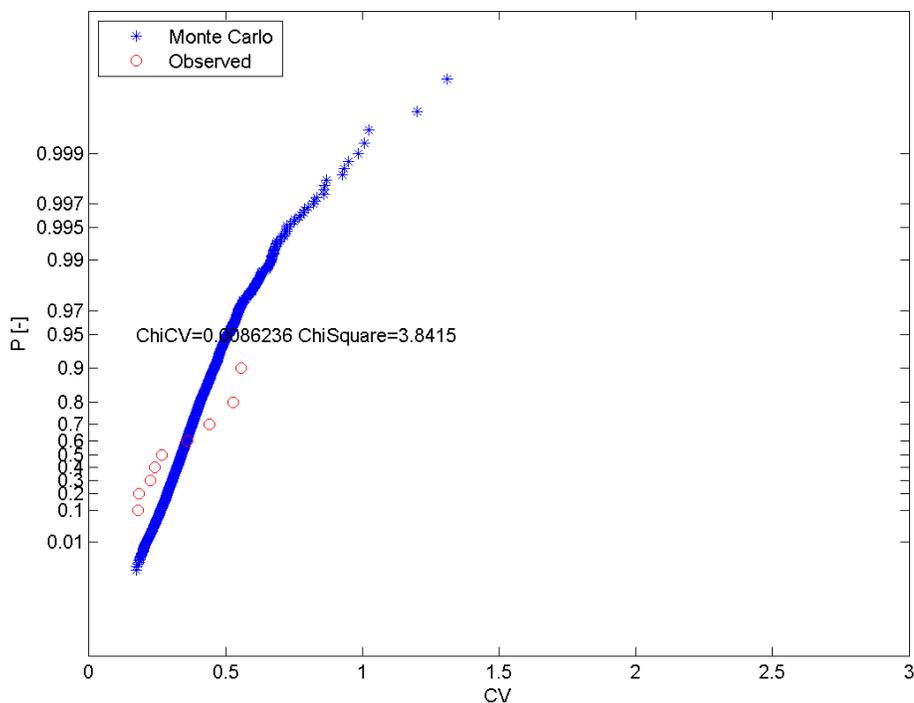


**Figure 10 Example of homogeneity Chi Square test for the Coefficient of Variation by means of Montecarlo technique. Gev distribution**
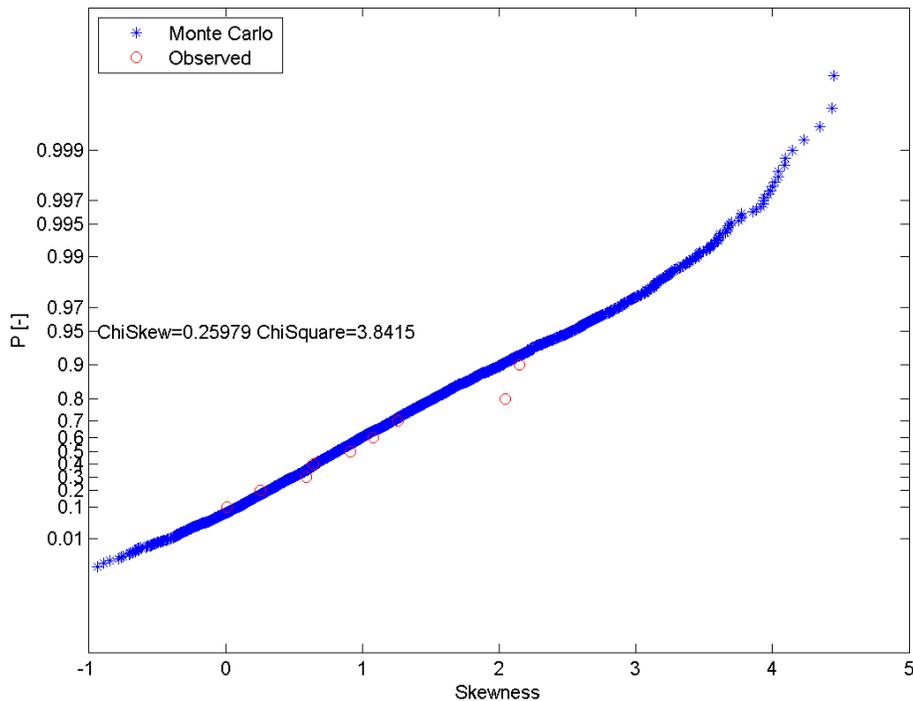
**Figure 11 Example of homogeneity Chi Square test for the Skewness by means of Montecarlo technique. Gev distribution**

### 1.4.5 Regression on the expected values:

Regarding the regression of the expected values we leveraged on previous analysis performed in the GAR 2011, using the same set of basin variables for the new analysis. Similar procedures have been followed (e.g., logarithmic transformation) to remove non linear dependences in the data.

For the variable selection a step-wise method have been used. Stepwise regression is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression. The method begins with an initial model and then compares the explanatory power of incrementally larger and smaller models. At each step, the p value of an F-statistic is computed to test models with and without a potential term. If a term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the term is added to the model. Conversely, if a term is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the term is removed from the model. The method proceeds as follows:

1. Fit the initial model.

2. If any terms not in the model have p-values less than an entrance tolerance (that is, if it is unlikely that they would have zero coefficient if added to the model), add the one with the smallest p value and repeat this step; otherwise, go to step 3.

3. If any terms in the model have p-values greater than an exit tolerance (that is, if it is unlikely that the hypothesis of a zero coefficient can be rejected), remove the one with the largest p value and go to step 2; otherwise, end.

Depending on the terms included in the initial model and the order in which terms are moved in and out, the method may build different models from the same set of potential terms. The method terminates when no single step improves the model. There is no guarantee, however, that a different initial model or a different sequence of steps will not lead to a better fit. In this sense, stepwise models are locally optimal, but may not be globally optimal, because of this reason the regression model have been fit both in an inclusive (from one to all) and exclusive (from all to one) way and the best result in terms of Correlation coefficient are retained. In most of the regions the "all possible regression" method have been applied and results compared with the ones of the stepwise method. Results were similar, due to the predominance of the Drainage area an mean rainfall as parameters dominating the regression in the majority of the hydrological homogeneous areas.
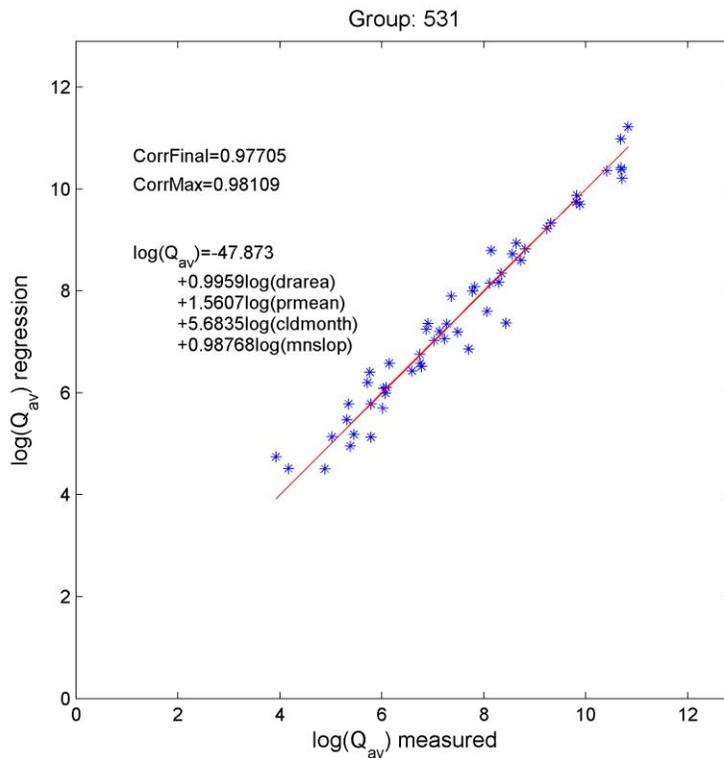


**Figure 12 Regression for Index Flow estimation with basin variables; Group 531**

The fitted regression showed a clear link between the climatological region and the variables selected by the model. Drainage Area as expected is present in all regression models. Annual Precipitation as well is always selected, this is due to the fact that we are considering high frequency quantiles on one side as well as monthly values on the other. In addition to such parameters some other characteristics are selected in line with the expected behaviour. As an example in many cases in regions with prominent orography, temperature of the colder month showed to have an influence in the equation. Where variance in orography, combined with semiarid climate is selected, morphological characteristics such as catchment maximum altitude and stream slope are selected as proxy of the orographic precipitation effect.

## 1.4.6 Quantiles estimation

The final result of the procedure is the estimation of the quantiles associated to different return periods.

In the following the step by step procedure is reported.

The estimation of the regional growing factor is estimated with the following formulation that is valid for the GEV probability distribution.

$$X_T = \varepsilon + \frac{\alpha}{k} \cdot (1 - e^{-k \cdot yT})$$

Where:

$$yT = -\ln\left(\ln\left(\frac{T}{T-1}\right)\right)$$

With T=return period; ε, α and k are parameters valid for the homogeneous region.

The index flood is calculated for the single basin based on the main morpho-climatic characteristics of the catchment, using a formulation that is valid for the whole homogeneous area. It is derived by a regression carried out on the mean flow of the stations that belong to the area in the way described in section 1.4.5.

The estimation of the flood with a defined return period T is calculated with the following expression:

$$Q(T) = X_T \cdot e^{\ln(Q_I)}$$

The following figure shows the comparison of quantiles estimation using single site statistics, the regional distribution using the local mean for dimensionalizing the growth curve and finally the full regional approach using the regression for the index discharge estimation. We notice that the regional estimates provide normally a more consistent estimation with each other while in some cases the single site estimation can produce very different results especially for low frequency quantiles.
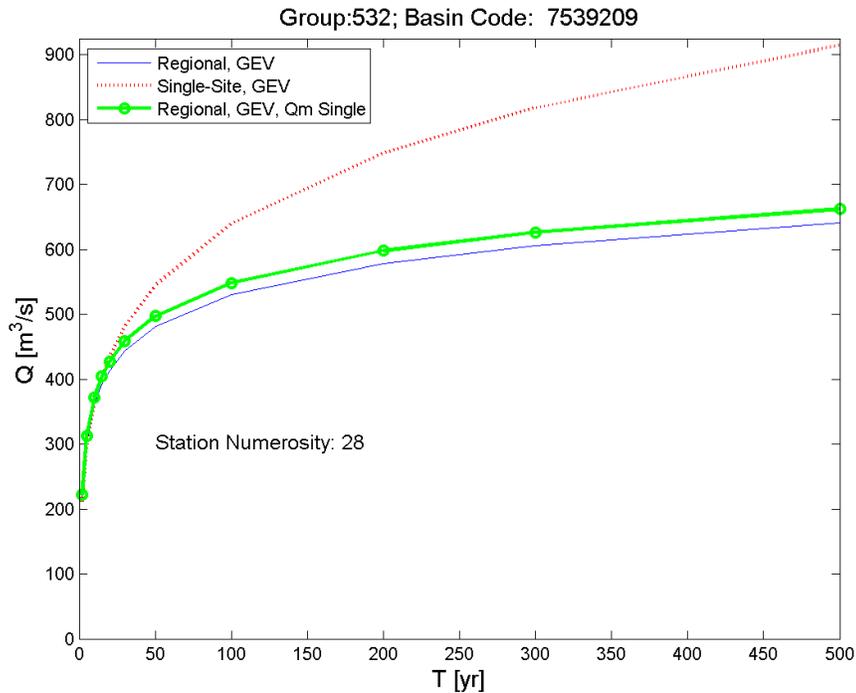
**Figure 13 Quantiles estimation comparison using Regional and single site distributions; Group 532. The green line is the quantile estimation using Regional Approach regarding the growing factor while the $Q_I$ is estimated as the average of the recorded flow data in the section.**

## 1.4.7 Sections affected by dams

Another relevant issue is related to sections placed downstream a dam, or a system of dams. The presence of these hydraulic structures can influence the behaviour of the river in the downstream section even in high flow conditions, modifying the natural flow. During the floods the flood peak is potentially laminated by the artificial reservoirs.

A procedure to account for this effect has been implemented all over the world, starting from information available in the Global Reservoir and Dam (GRanD) database, updated March 2011. the Data base contains all major dams world round.
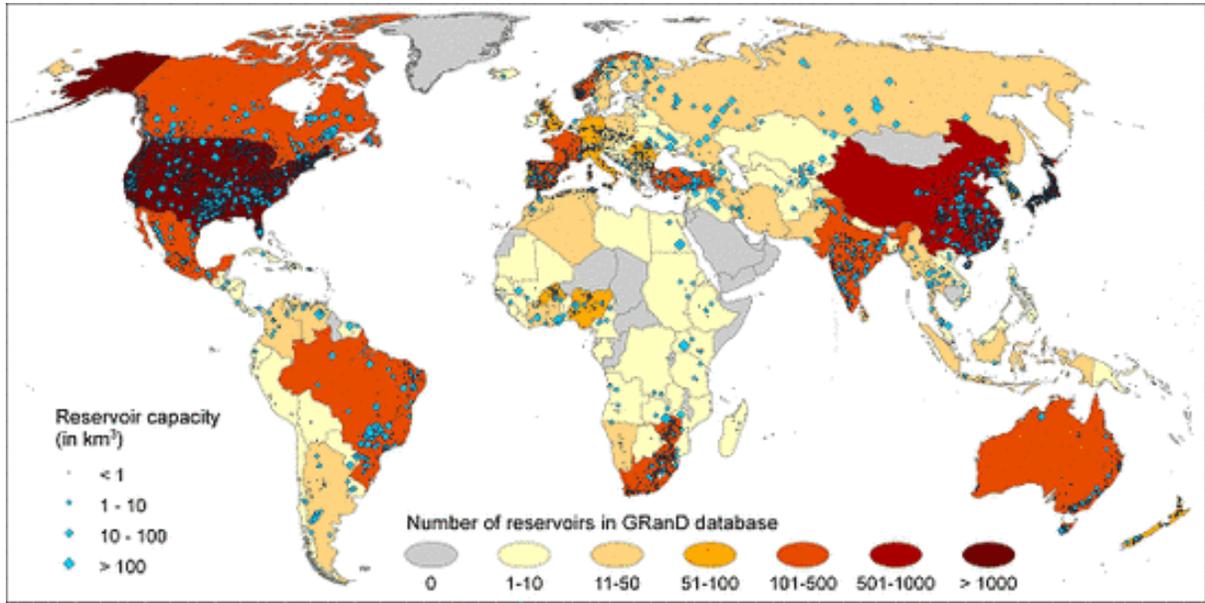
**Figure 14 Distribution of dams in the GRanD Data Base**

First the dams with a low lamination capacity have been filtered out. This has been done referring to the Degree of Regulation (DOR) field in the data base; equivalent to "residence time" of water in the reservoir; calculated as ratio between storage capacity and total annual flow. The total annual flow Long-term (1961-90) average discharge at reservoir location is derived from HydroSHEDS flow routing scheme combined with WaterGAP2 runoff estimates.

As a first approximation, we do not consider the dams that have a DOR<100%.

As a second step we identified those sections that we consider to be affected by the presence of one or more dams in the upstream basin

Let's consider the following quantities:

$P_{ms}$=mean of the monthly maximum precipitation for the section drainage area

$A_s$=drainage area of the basin upstream section

$P_{md}$=mean of the monthly maximum precipitation for the dam drainage area (equivalent dam)

$A_d$=drainage area of the basin upstream the dam (equivalent dam)

Compute the following metric:

$$WP_r = \frac{P_{md} \cdot A_d}{P_{ms} \cdot A_s}$$

WPr should generally be in the range (0-1).

We consider the section affected by the system of the dams upstream the section when WPr> $TWP_r$. $TWP_r$ is fixed to the value 0.1.

In this way all sections potentially affected by the regulation capacity of the dam are marked. The third step is computing the laminated discharge quantile starting from the estimation given by the regional approach.

Basically we considered a synthetic event of rectangular shape of length 1 month and flow equal to the quantile estimation derived by the regional curve, this because we deal with monthly maximum flow. After that the percentage of flood volume that can be laminated by the dams is estimated with the following procedure.

1. Calculate the quantile with regional method in the analyzed section:

   $Q_{S,Reg}(T) = X(T)*Q_I$

   where X(T)= growing factor; $Q_I$=index flow

2. Estimate the percentage quantile flow due to the contribution of the basins upstream the dams:

   $Q_{UD,Reg}(T) = Q_{S,Reg}(T)*WP_r$

3. Estimate the quantile in the section masking the basins upstream the dams

   $Q_{SM,Reg}(T) = Q_{S,Reg}(T)*(1-WP_r)$

4. Estimate the maximum volume that can be laminated by the dam with an empirical function:

   $V_{D,MAX} = K*Cap\_max$

   Where: Cap_max = 'maximum storage capacity' in cubic meters of the dams system derived by GRanD; K=min(0.25,Ks)

$$Ks = 0.1 + 0.15 \cdot \frac{(DOR - 100)}{200}$$

K is thus a function of the DOR. The mean DOR of the system of dams upstream the section is considered.

The flow that can not be laminated for the return period T is estimated as:

$$Q_{DD,Reg}(T) = \max(Q_{UD,Reg}(T) - \frac{V_{D,MAX}}{\Delta t}; 0)$$

Where Δt=30 days (converted in seconds); this because we consider max monthly flow and we refer to a synthetic event of rectangular shape with one month length.

Finally the estimation of the quantile in the section accounting dams' affection is calculated as:

$Q_{SF,Reg}(T) = Q_{SM,Reg}(T) + Q_{DD,Reg}(T) + Q_{Disch}$

Where:

$Q_{Disch}$= average discharge at reservoir location (the max value of the dams system is considered); this terms account for a base flow derived by the outlets of the dams and it is derived by GRanD.
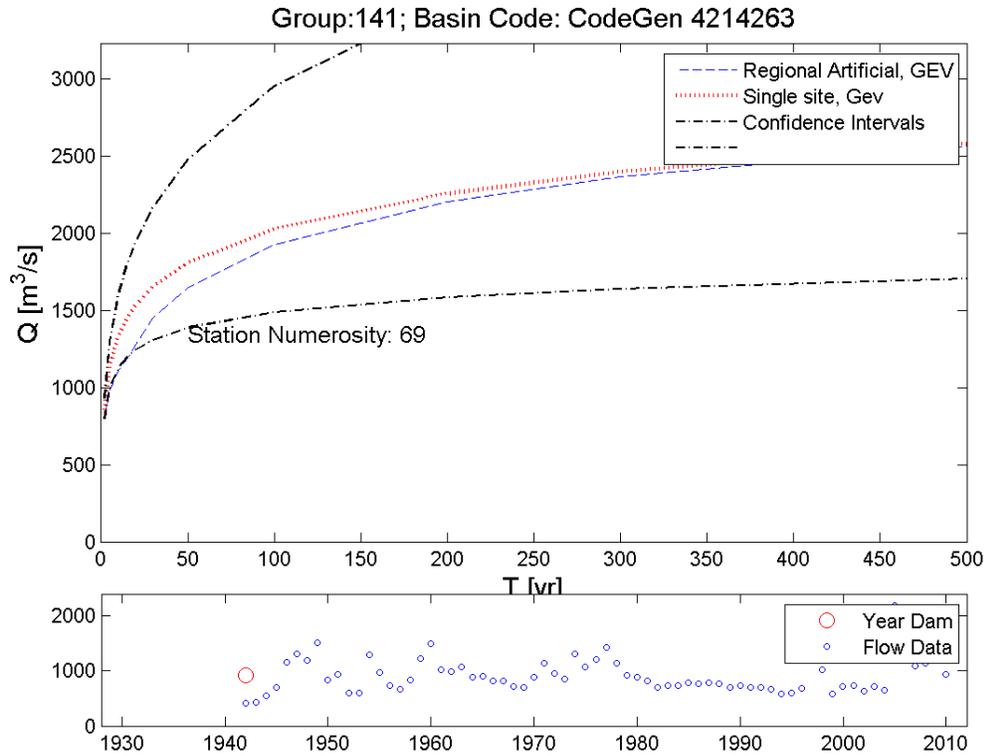
**Figure 15 Example of quantile estimation for a section affected by the upstream dams. The blue dotted line is the regional estimation accounting for the dams, the red line is the single site estimation while the black lines are the confidence intervals (90%) of single site estimation. In the low panel the monthly maximum flow data are reported together with the year of building of the dam (GRanD data).**

There are different levels of validation of the overall procedure. The first on the statistics: the homogeneous region are tested with statistical tests on observations, the growth curve resulting from statistical computations is tested against its confidence limits. A second level of validation, even more constraining, comes from the dam influenced sections. Such sections were taken out of the statistics at first and then after reconstruction of their natural flow are used as at a point validation against observations.
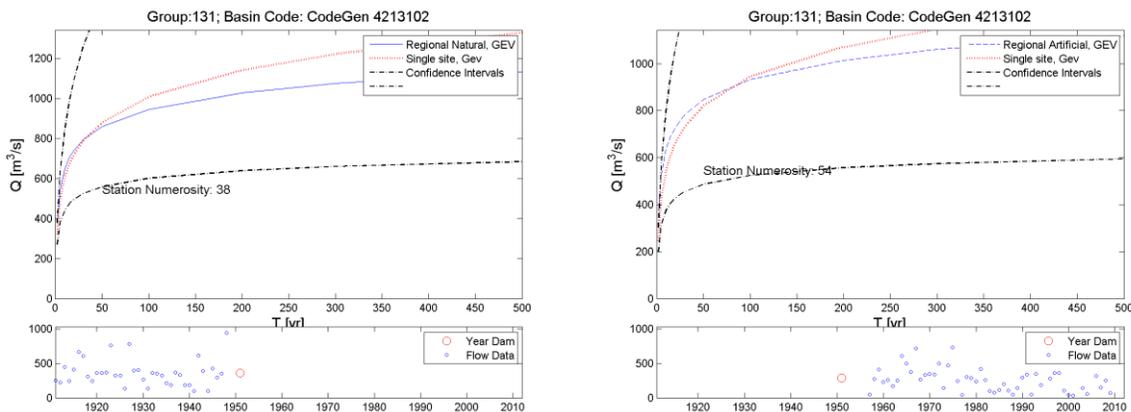


**Figure 16 Example of quantile estimation for a section affected by the upstream dams. On the left: The blue continuous line is the regional estimation, On the right the blue dotted line is the regional estimation accounting for the dams. The red line is the single site estimation while the black lines are the confidence intervals (90%) of single**

**site estimation. In the low panel the monthly maximum flow data are reported together with the year of building of the dam (GRanD data).**

## *1.5  Case study*

### 1.5.1  Introduction

The case study regards the Thailand region and compares the results of GFM to the catastrophic flood that took place in 2011. It is necessary to apply the regional analysis on the entire central and south part of Asia on the identified homogeneous region including Thailand and the ones close to it as some basins cross various homogeneous regions and it is not possible to localize the analysis exclusively on Thailand.

### 1.5.2  Central and South Asia Homogeneous regions

In the following we report the four considered homogeneous regions indentified following the criterions described in paragraph 1.4.2.
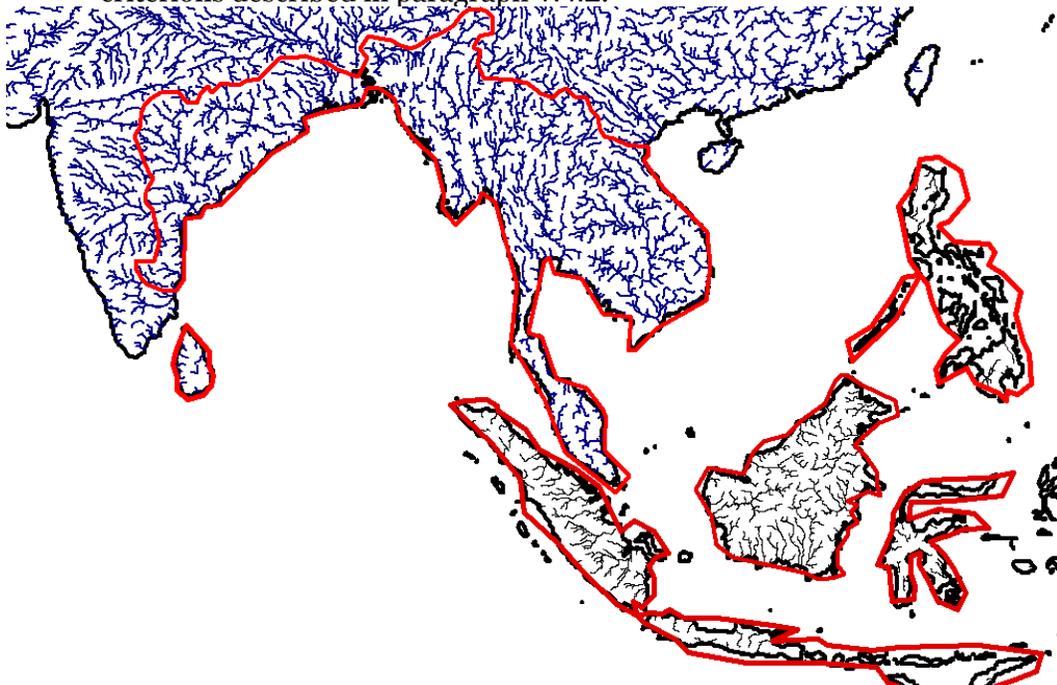


**Figure 17 The red line shows the homogeneous region 511. It includes the tropical basins of Asia.**
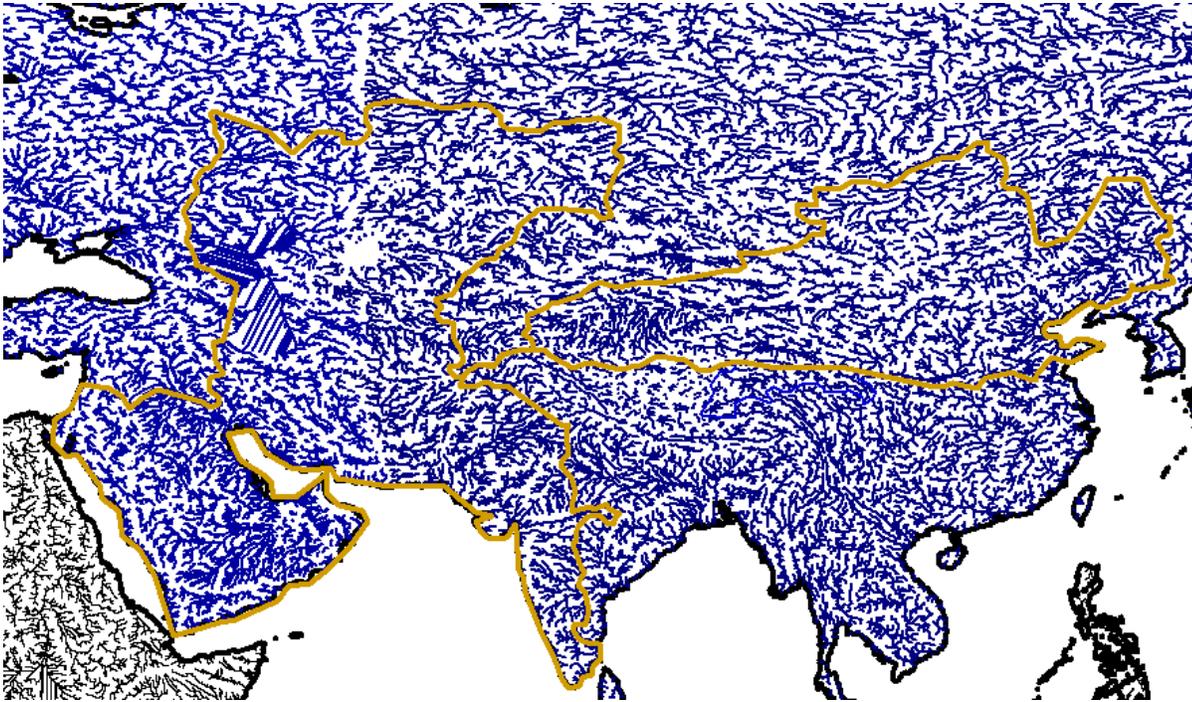
**Figure 18 The yellow line shows the homogeneous region 521. It includes the arid basins of Asia.**
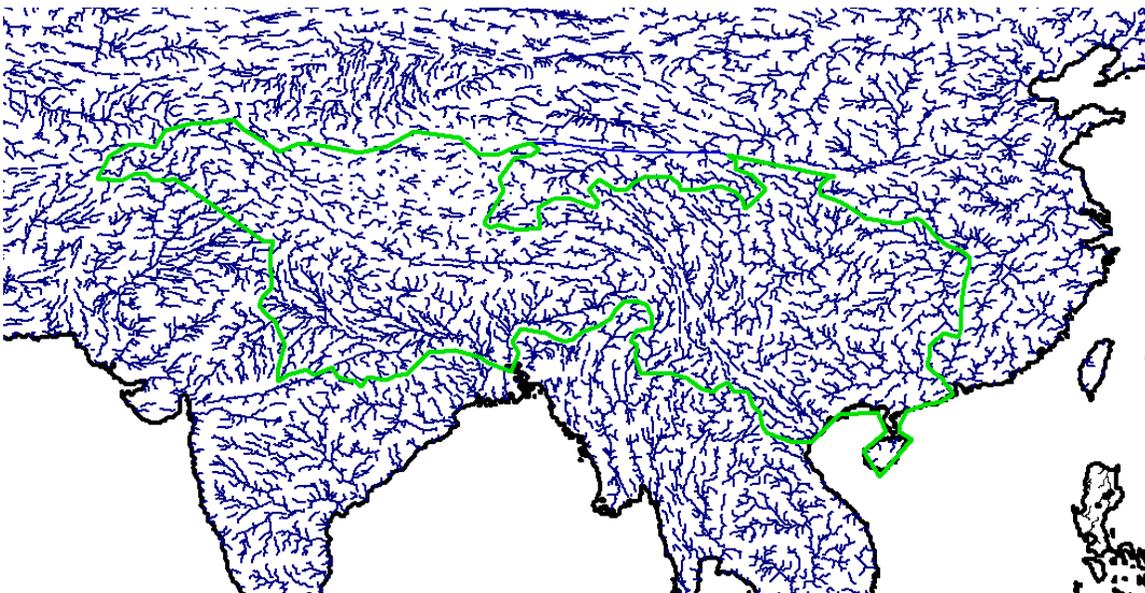


**Figure 19 The light green line shows the homogeneous region 531. It includes the basins that originates by the Central Asian mountain chains and part of the temperate zones.**

**Figure 20 The dark green line shows the homogeneous region 533. It includes the basins of the west temperate zone.**

### 1.5.3  Distribution fitting and homogeneity testing

The GEV distribution has been fitted on the series of flow data available for all the stations that belong to every single region. Every data series has been normalized with its mean before it has been used in the regional data series.

Using the approach described in paragraph 1.4.4 the homogeneity of the region has been verified.

## 1.5.3.1 Region 511

The parameters of the GEV distribution are the following:

$\varepsilon$=0.8452, $\alpha$=0.3123, k=0.0645

In Figure 21 the GEV distribution fitted on data is reported.
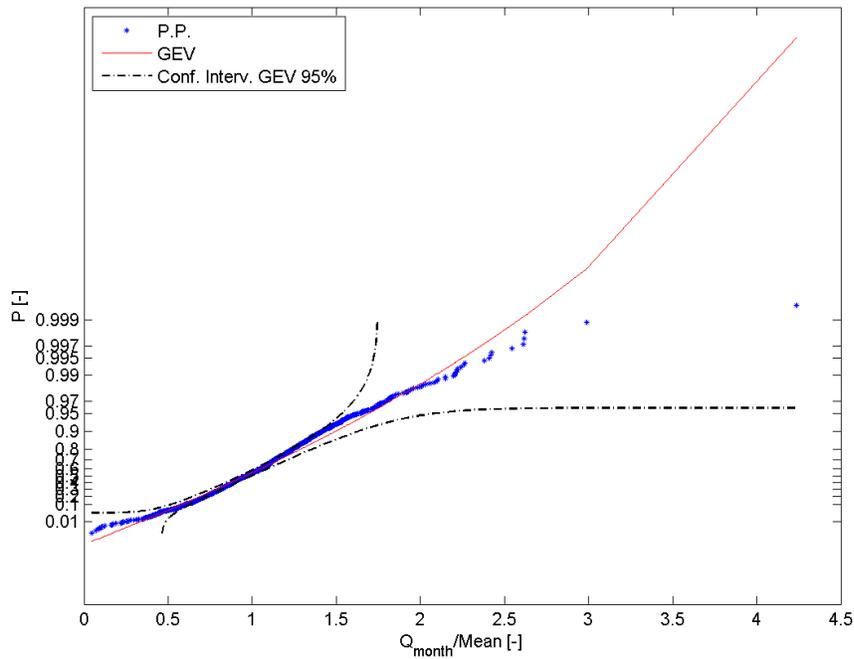
**Figure 21 Gev distribution fitted on flow data for region 511.**

The results of homogeneity on Coefficient of Variation and on Skewness are reported in graphical form.
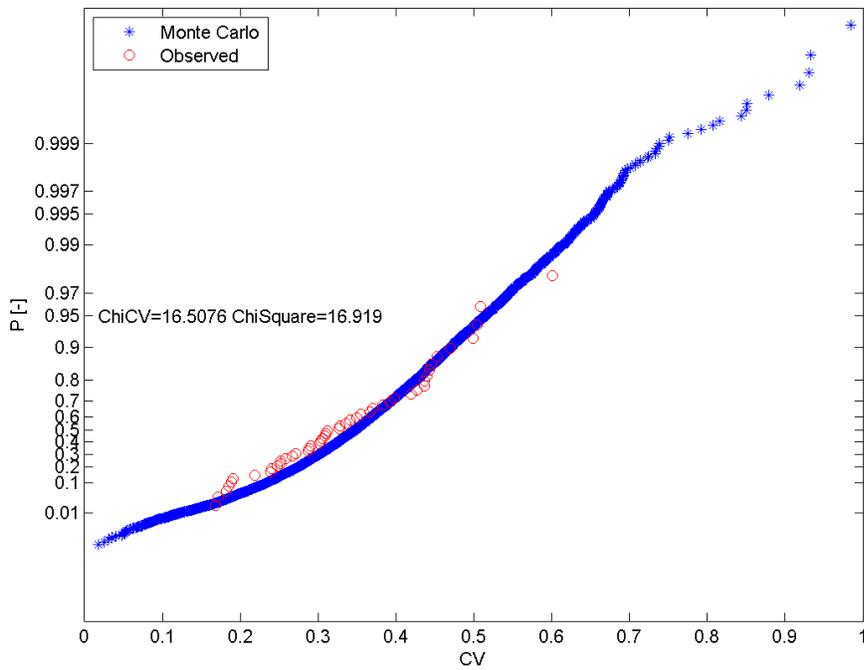


**Figure 22 Homogeneity Chi Square test for the Coefficient of Variation by means of Montecarlo technique. Gev distribution. Regione 511.**
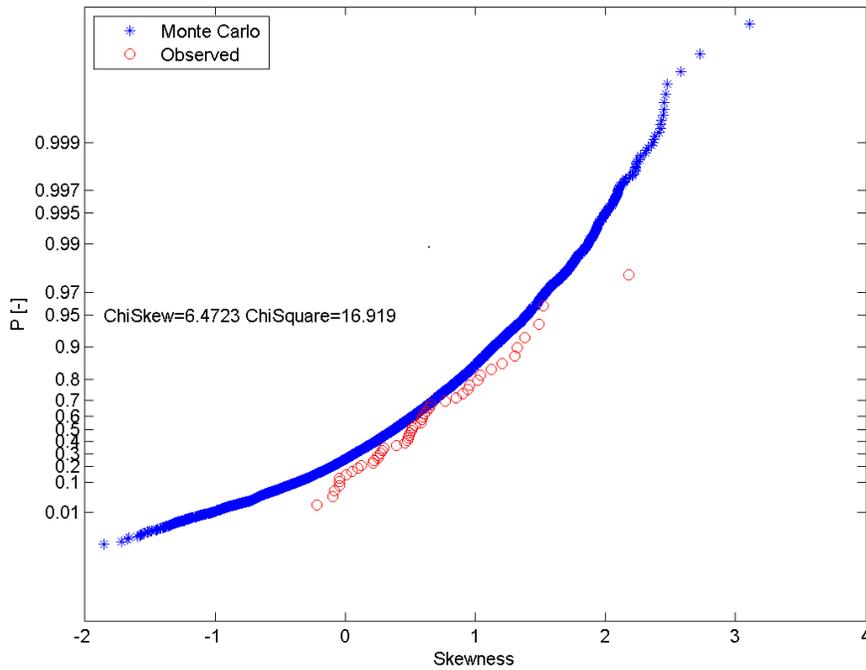
**Figure 23 Homogeneity Chi Square test for the Skewness by means of Montecarlo technique. Gev distribution. Region 511**

## 1.5.3.2 Region 521

The parameters of the GEV distribution are the following:

ε=0.6772, α=0.4668, k=-0.1056

In Figure 24 the GEV distribution fitted on data is reported.
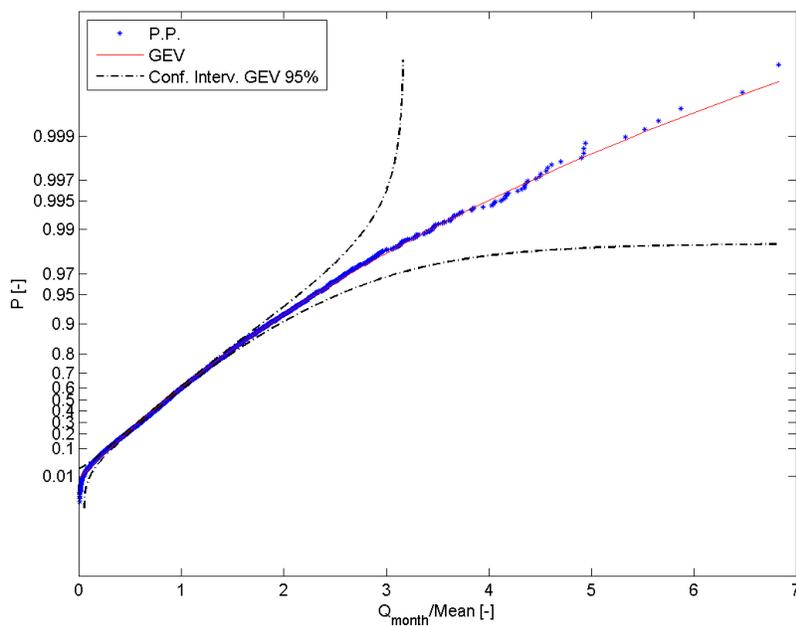


**Figure 24 Gev distribution fitted on flow data for region 521.**

The results of homogeneity on Coefficient of Variation and on Skewness are reported in graphical form.
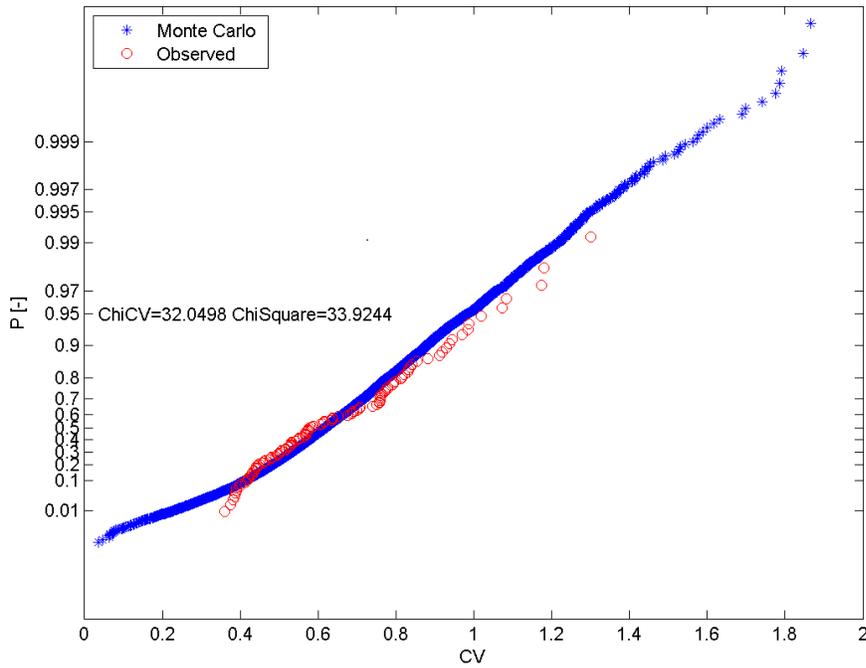


**Figure 25 Homogeneity Chi Square test for the Coefficient of Variation by means of Montecarlo technique. Gev distribution. Regione 521.**
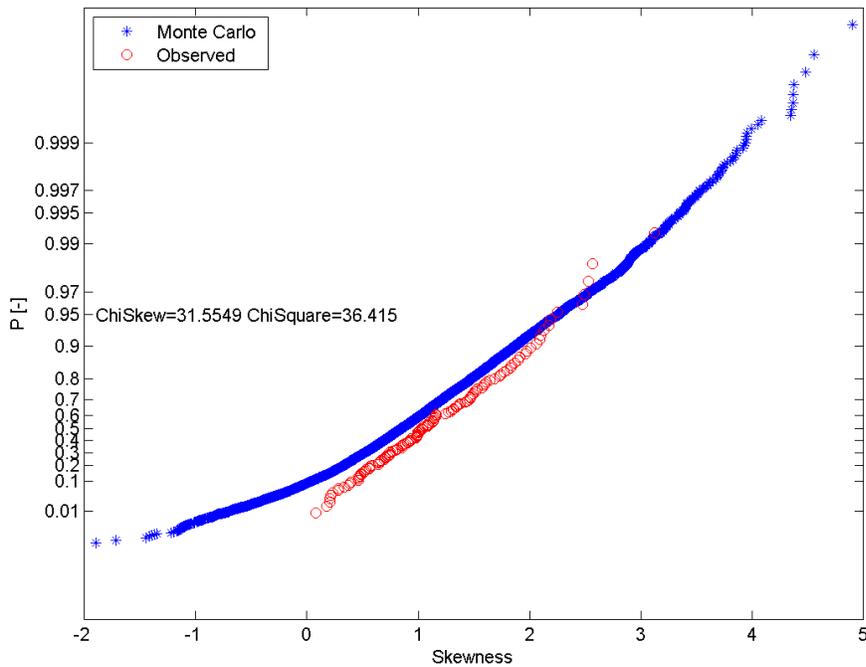


**Figure 26 Homogeneity Chi Square test for the Skewness by means of Montecarlo technique. Gev distribution. Region 521**

### 1.5.3.3 Region 531

The parameters of the GEV distribution are the following:

ε=0.899, α=0.2624, k=0.1999

In Figure 27 the GEV distribution fitted on data is reported.
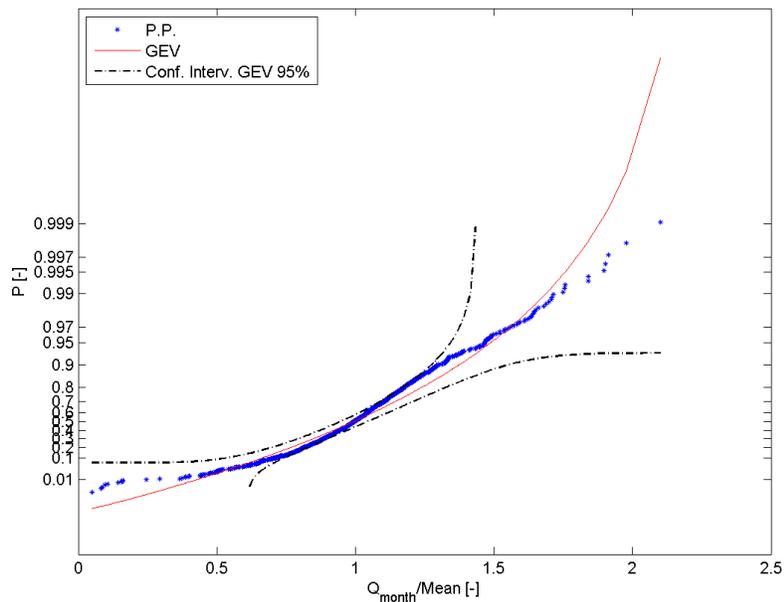


**Figure 27 Gev distribution fitted on flow data for region 531.**

The results of homogeneity on Coefficient of Variation and on Skewness are reported in graphical form.
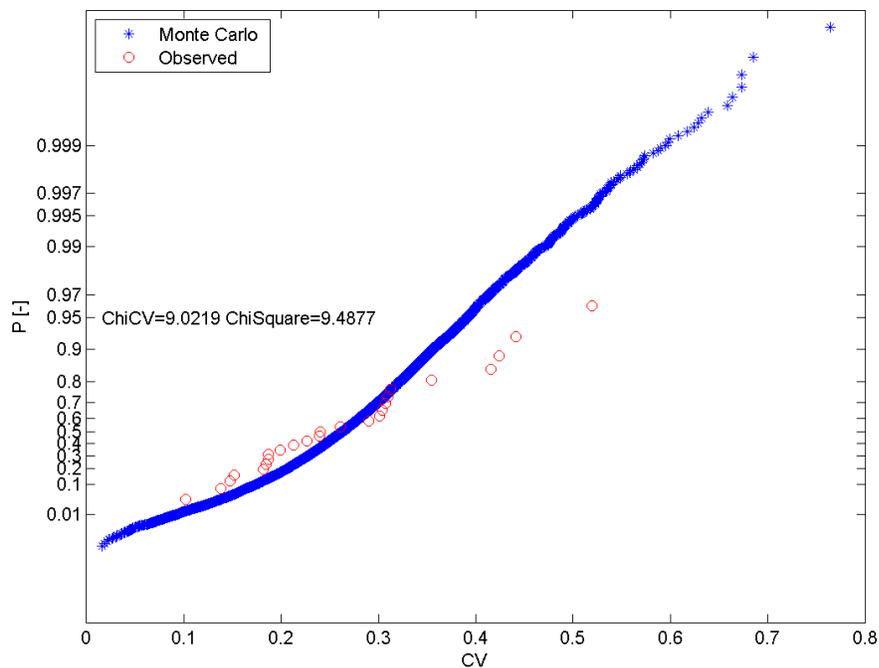
**Figure 28 Homogeneity Chi Square test for the Coefficient of Variation by means of Montecarlo technique. Gev distribution. Regione 531.**
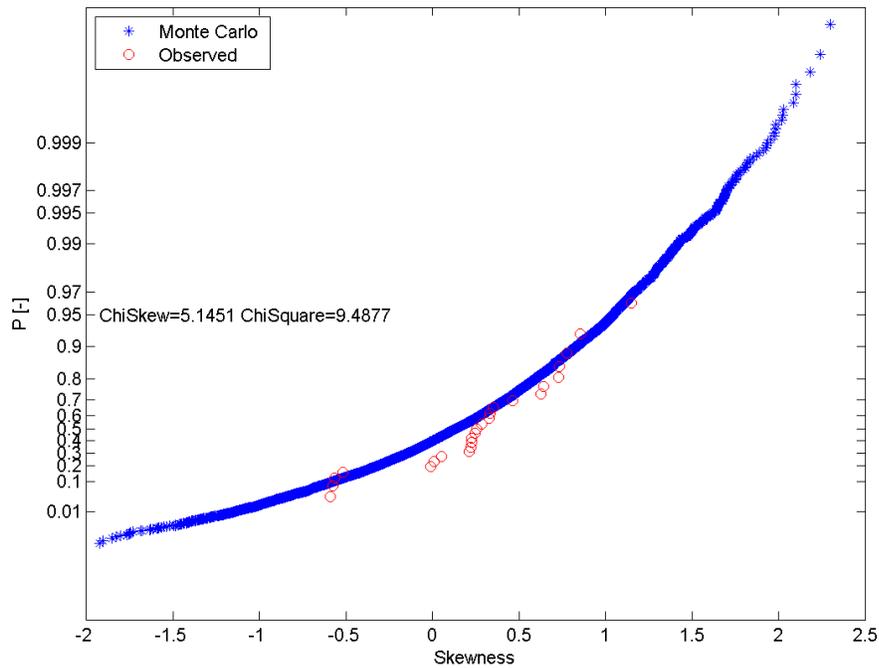


**Figure 29 Homogeneity Chi Square test for the Skewness by means of Montecarlo technique. Gev distribution. Region 531**

## 1.5.3.4 Region 532

The parameters of the GEV distribution are the following:

$\varepsilon=0.808$, $\alpha=0.3512$, k=0.0305

In Figure 30 the GEV distribution fitted on data is reported.
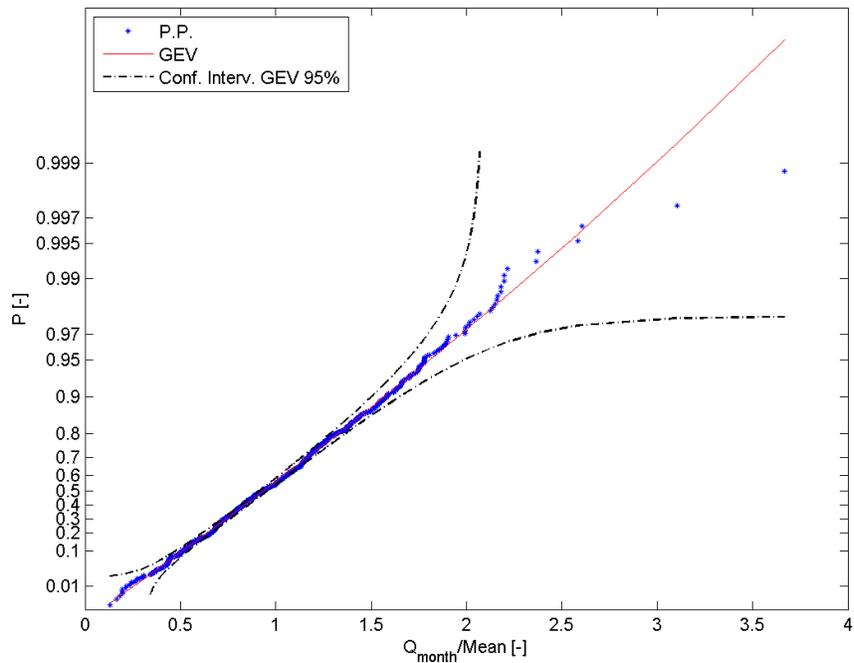
**Figure 30 Gev distribution fitted on flow data for region 532.**

The results of homogeneity on Coefficient of Variation and on Skewness are reported in graphical form.
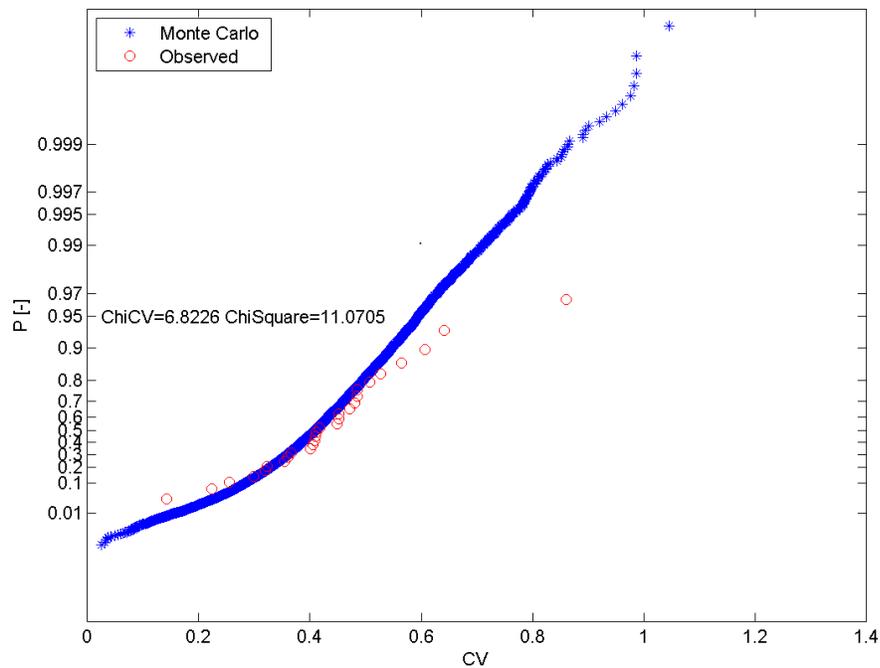


**Figure 31 Homogeneity Chi Square test for the Coefficient of Variation by means of Montecarlo technique. Gev distribution. Regione 532.**
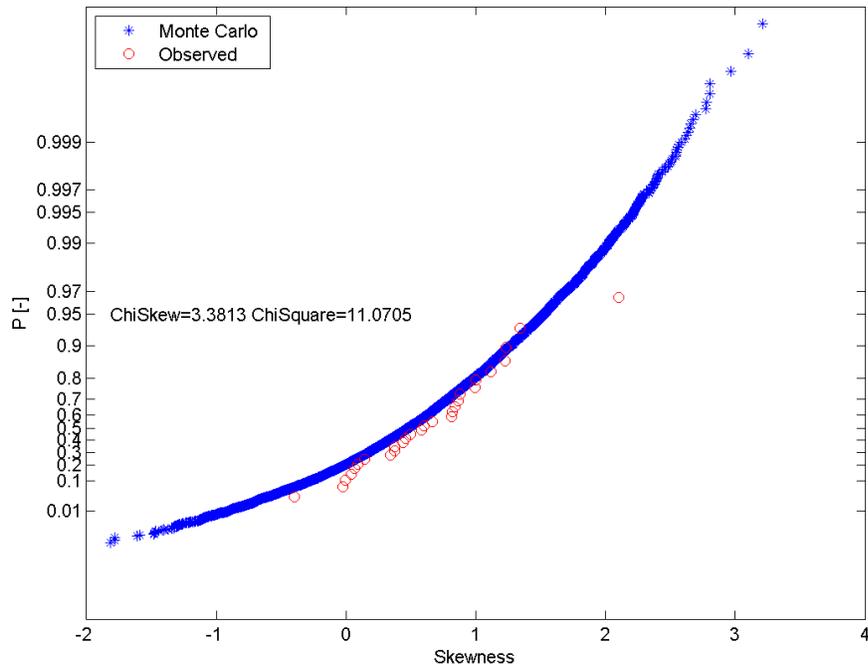
**Figure 32 Homogeneity Chi Square test for the Skewness by means of Montecarlo technique. Gev distribution. Region 532**

## 1.5.4 Regressions

In the following the regression for the estimation of the index flow is reported for each homogeneous region.
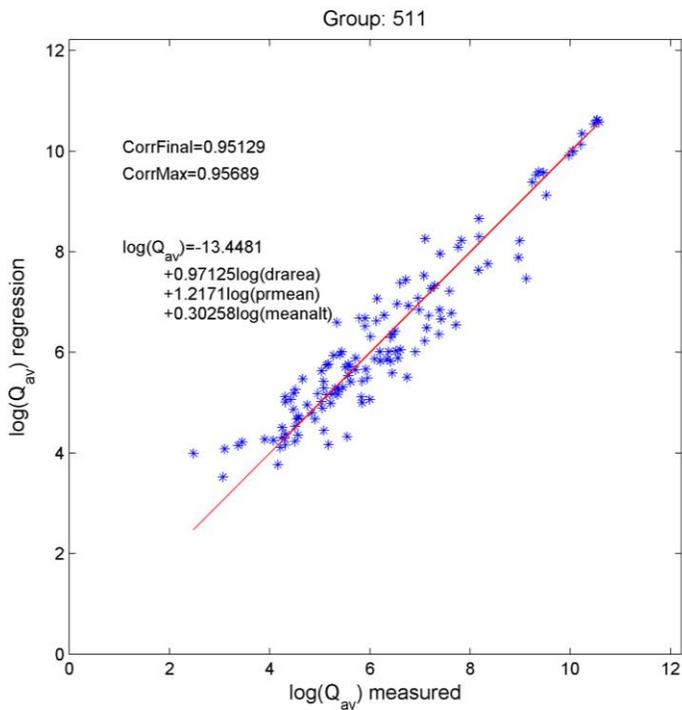


**Figure 33 Regression for Index Flow estimation with basin variables; Group 511**
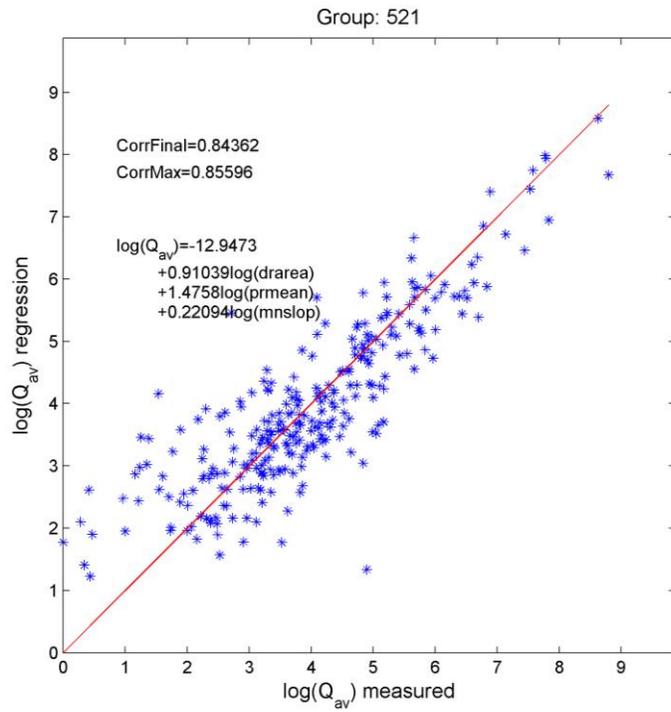
**Figure 34 Regression for Index Flow estimation with basin variables; Group 521**
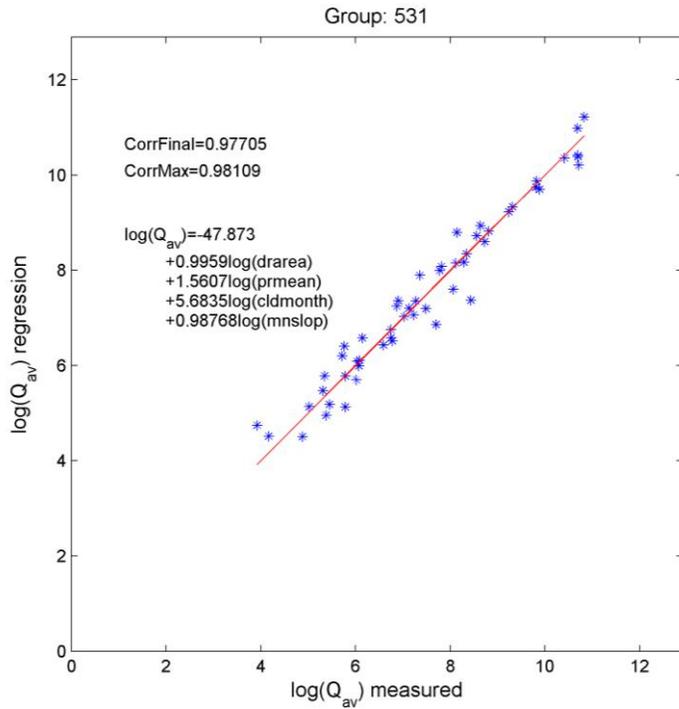


**Figure 35 Regression for Index Flow estimation with basin variables; Group 531**
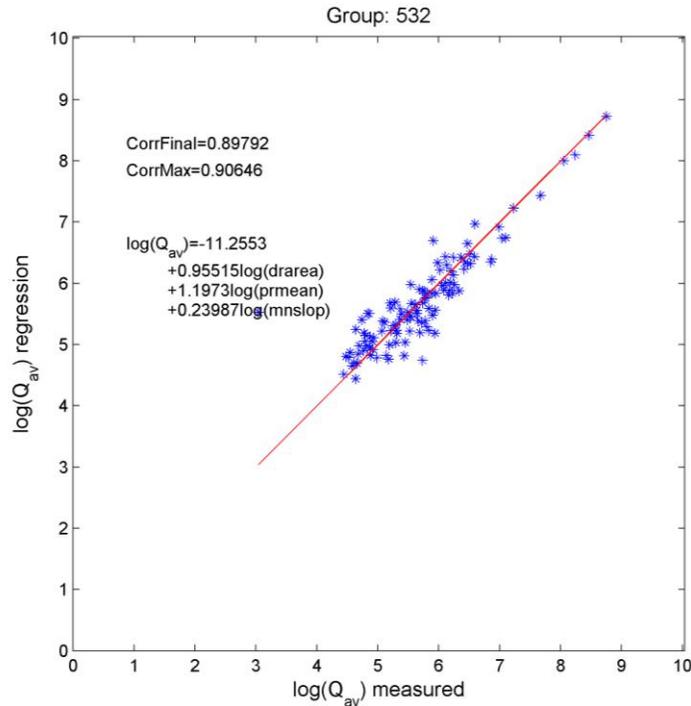
**Figure 36 Regression for Index Flow estimation with basin variables; Group 532**

### 1.5.5 Quantile estimation

Once the parameters of the growth factor and those of index flow formulation are defined the quantiles can be calculated for each section of interest. The section must be assigned to the right homogeneous region with the methodology described in paragraph 1.4.3.

To avoid inconsistencies in the case of rivers that cross more than one homogeneous region, a check on the quantile estimation has been introduced. When there is a passage from a region to another along the river bed from upstream to downstream the flow must always increase, otherwise the quantile estimation of the previous region is considered.

In Figure 37 an example of quantile estimation is shown, the final part of the rivers that flow in the gulf in front of Bangkok are considered.
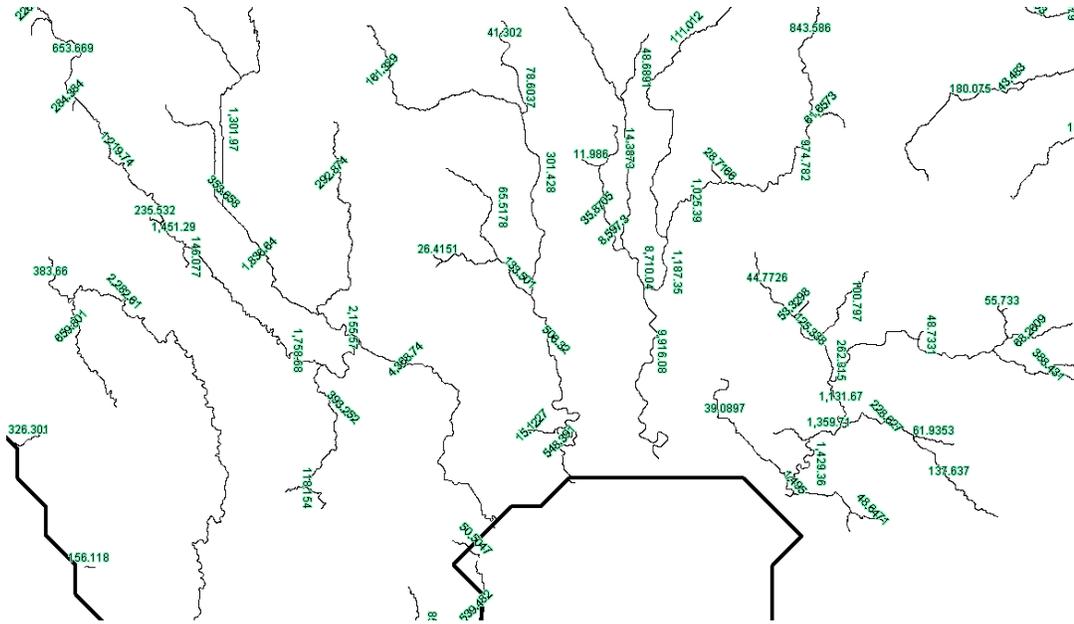
**Figure 37 Quantiles estimation in the final part of the rivers that flow in the gulf in front of Bangkok.**

## 1.5.6 Flooded areas Modeling
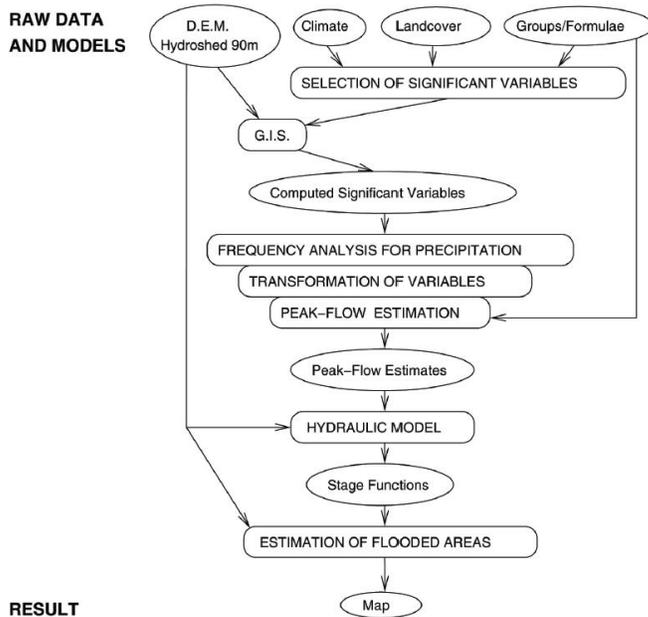


**Figure 38 scheme illustrating the full workflow to obtain flood hazard maps for different return periods as per GAR 2011. the overall scheme can be considered valid for the current release.**

Once the discharge quantiles are determined results are imputed into a simplified hydraulic flood model. The current model (GAR 2013) stems from the one used in 2011 and tries to improve its performance overall and in specific condition where the previous release model is recognized to have problems, mainly in flat and large flood plains as well as when river show a marked braiding.

The hydraulic model starts from an idea provided by the EROS Data Center (EROS/USGS). The original model first generates a relative DEM from HydroSHEDS that set any stream pixel values to 0 as a reference altitude. Then, it generates cross sections of a specified width for each stream section. Each cross section is then used to extract altitude values from relative DEM and generate a specific stage vs. discharge function using Manning's equation. These functions are finally used to calculate river stage from peak flow estimates for a specific recurrence interval, and then generate corresponding flooded areas for each stream section basin, using the generated relative DEM.

## 1.5.6.1 Model improvements

Some of the limitations of the previous model have been removed. In its new version the model does not have the limit of having just one section per stream, that in certain cases could cause serious biases in the computation because of the length of the river steams. The new version of the model uses a sequence of cross sections per stream whose density depends on the morphological characteristics of the catchment segment so that it is possible to capture changes in the floodplain and in the cross section carrying capacity. In figure an example of such setup is showed.

In the original model a single value of the Manning coefficient per stream was assumed. Such an approximation is a critical one as the Manning coefficient is the main parameter to determine the discharge stage transformation. This limitation has been removed and the manning parameter cannot only change between cross sections along the streams, but within the cross section itself where normally the main channel and the floodplain show very different hydraulic roughness.

In the model new version flow in the floodplain is subdivided as the roughness coefficient changes along the cross section. The Manning equation in terms of conveyance is used. For the conveyance computation the cross section is subdivided into segments with limited differences in flow velocity that is determined by the hydraulic roughness. All the incremental conveyances are summed up to obtain a conveyance on the floodplains and in the main channels that ultimately determine the total conveyance of the cross sections and therefore the stage discharge curve.

The Manning coefficient is determined for the different parts of the cross section on the basis of the land use map provided by the Glob Cover map.

A second important modification regards the possibility of determining stage discharge curves in braided rivers that for certain flow conditions present transversal hydraulic disconnections in the cross section. This condition has to be accepted in case of braided rivers and disregarded in other cases arising from specific morphologic conditions. Therefore on the basis of discharge value and slope value a likelihood of the river being braided is determined and a possible extension of the flood plain where hydraulic disconnection is allowed computed on the basis of a simplified steady flow computation based on a regularized triangular representation of the cross section.

In the flood plain especially were the morphological control is low (i.e. flat areas) the concept of relative DEM as a boundary for the flood plain fails. Such limitation is overcome by iteratively recomputed the cross section so that the manning equation is satisfied imposing that as a limit for the flooded area. The cross section points are then connected following the local slope value.

Another evident limitation of the model was the impossibility of representing the backwater effect occurring in certain areas of the globe. It is possible in fact that a flooding plain pertaining to a tributary is not flooded because of the limited carrying capacity of the tributary itself but because of the valley boundary condition imposed by the main river that blocks the

discharge from the tributary diminishing its capacity, and imposes a higher free surface level in the valley par of the tributary. This has a large impact in flood extension computation when the tributary slope is really low. This case has also been solved in the new version. Junctions where a minor tributary showing a limited slope and a major river (with a quantile discharge more than one order of magnitude tributary) intersect are identified. In this case the water height imposed by the main river is propagated backward along the tributary with a rate of decrease that takes slope into account till the height imposed by local calculations on the tributary does not exceed such value. The influence length of the backwater effect can be theoretically computed under certain hypothesis using perturbation theory (Samuels, 1989). Such length is used to validate the procedure results.

## 1.5.6.2 Observed flood events and validation

Actual flood events, as detected by satellite from Dartmouth Flood Observatory (DFO), were used to validate the model. The observed flooding events, based mostly on MODIS satellite sensors at 250 m resolution, provided additional information and were also used for calibration. The data for observed flood events cover only nine years, containing more than 400 events and are not comprehensive. The combination of observed and modelled datasets provides a good picture of the most flood-prone areas.

## 1.6  Case Study: Thailand

Severe flooding occurred during the 2011 monsoon season in Thailand. Beginning at the end of July triggered by the landfall of Tropical Storm Nock-ten, flooding soon spread through the provinces of Northern, Northeastern and Central Thailand along the Mekong and Chao Phraya river basins. In October floodwaters reached the mouth of the Chao Phraya and inundated parts of the capital city of Bangkok. Flooding persisted in some areas until mid-January 2012, and resulted in a total of 815 deaths (with 3 missing) and 13.6 million people affected. Sixty-five of Thailand's 77 provinces were declared flood disaster zones, and over 20,000 square kilometres of farmland was damaged[2]. The disaster has been described as "the worst flooding yet in terms of the amount of water and people affected."

The World Bank has estimated 1,425 billion baht (US$ 45.7 Bn) in economic damages and losses due to flooding, as of 1 December 2011[3]. Most of this was to the manufacturing industry, as seven major industrial estates were inundated by as much 3 meters (10 feet) during the floods. Disruptions to manufacturing supply chains affected regional automobile production and caused a global shortage of hard disk drives, which is expected to last throughout 2012.

The new version of the GFM was therefore run for the Thailand area so to have a fresh comparison of the result with a recent catastrophic event. Figure 40 depicts the flood maps for Thailand for three different return periods (T = 25, 200, 1000 years). There are differences between the three maps both in flooding extension and water depth value although the overall extension is similar. This is due to the climatic specificity of the area. Tropical catchment show in fact small differences between quantiles (small variance in the quantiles) that leads to a very steep growth curve. This in combination with the methodology used to derive the maps results into relatively small differences in flood extensions. However local differences might be striking (see e.g., Figure 39). Figure 40 also shows a comparison between the GFM results and the DFO footprint. The match is very good in mountain areas where the morphologic control is

---

[2] "รายงาน สรุปสถานการณ์ อุทกภัย วาตภัย และดินโคลนถล่ม ฉบับที่ 129 วันที่ 17 มกราคม 2555 (Flood, storm and landslide situation report)" (in Thai). 24/7 Emergency Operation Center for Flood, Storm and Landslide. http://disaster.go.th/dpm/flood/news/news_thai/EOCReport17JAN.pdf

[3] http://www.worldbank.org/en/news/2011/12/13/world-bank-supports-thailands-post-floods-recovery-effort

dominant and the GFM methodology more reliable. Larger discrepancies are found in terms of extension in the large floodplains. Two reasons for that are on the one hand the reliability of DTM and on the other the limitations of the GFM methodology itself. In addition structural defences, that are not considered in the GFM methodology, are more concentrated in the floodplains where stakes are usually denser. However the largest discrepancy between the maps here is concentrated in the Bankog area where the model predicts extended flooding while the DFO footprint does not show any. DFO and satellite derived flood maps in general (both optical and SAR) are quite unreliable in urbanized areas as the algorithms used for flooded area detection suffer limitations in these conditions. In the case of the 2011 flood we know that the Bankog areas had been severely flooded as predicted by the GFM model, so that it is the validation data set that misses the extension in this case, and here, in the most important area of the Country is where the model is giving the larger added value.

A more detailed comparison has been possible in this case thanks to some Hi-RES flooded areas provided by UNOSAT based on a combination of Hi-Res SAR and optical images.

Figure 41 presents the comparison an shows the multi temporal envelope of flooded areas from the beginning to the end of the flood event 2011. The limits are followed quite well by the GFM predictions with differences that can be re-conducted to the comments made previously for the DFO comparison.
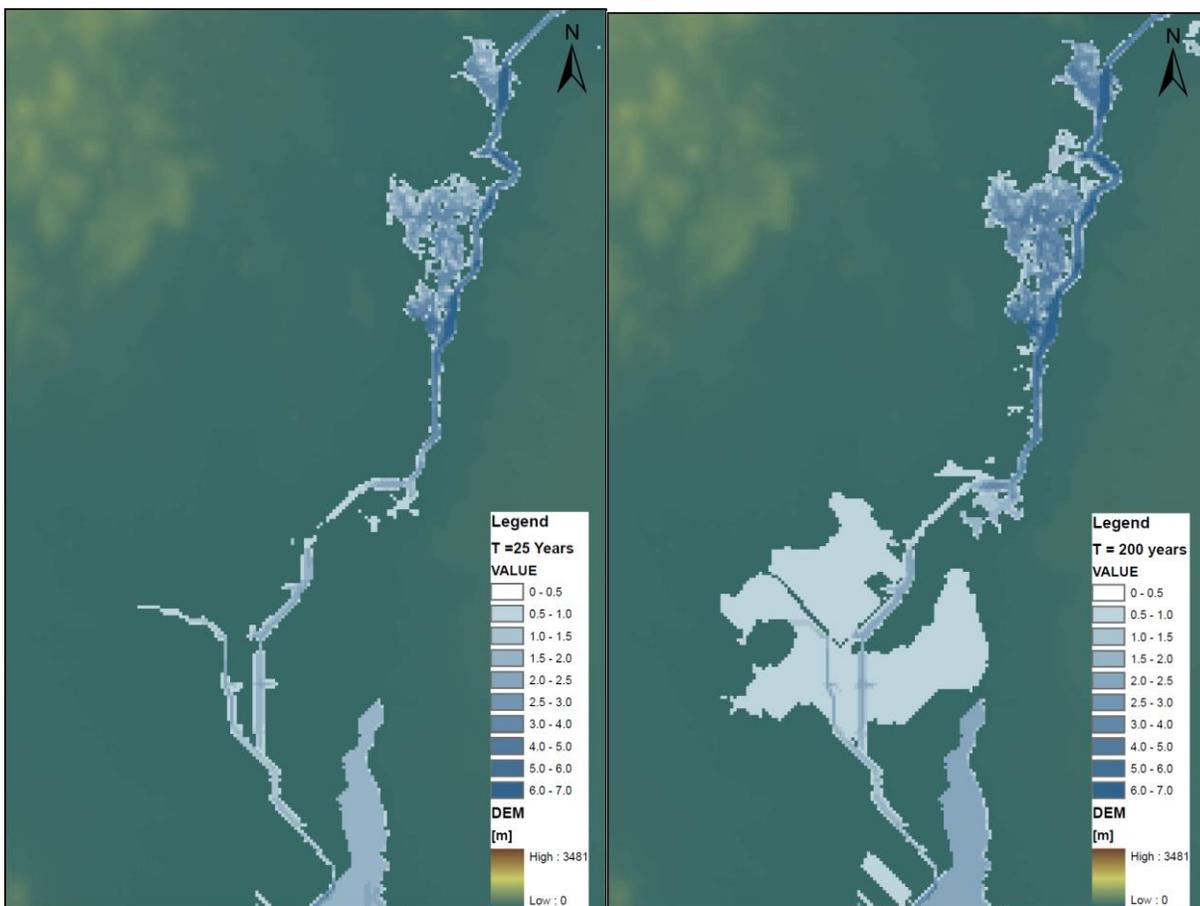


**Figure 39 Flood maps for Thailand for two different return periods ( from left to right: T = 25, 200 years), detail. Reconditioned DEM in the background.**
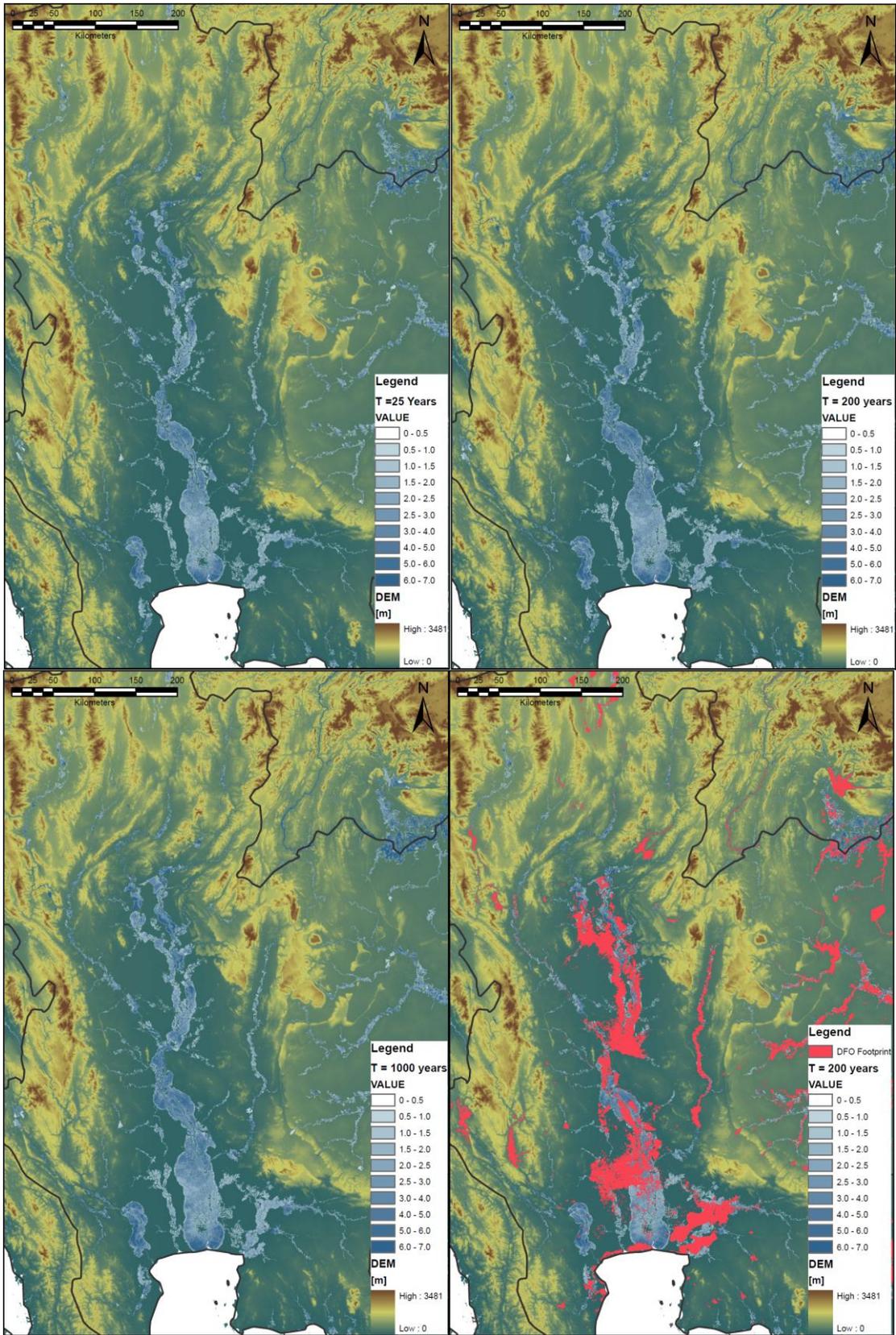
**Figure 40 Flood maps for Thailand for three different return periods ( from left to right from top to bottom: T = 25, 200, 1000 years) and bottom right panel DFO flood footprint envelope layered on the flood map for T = 200 years. Reconditioned DEM in the background.**
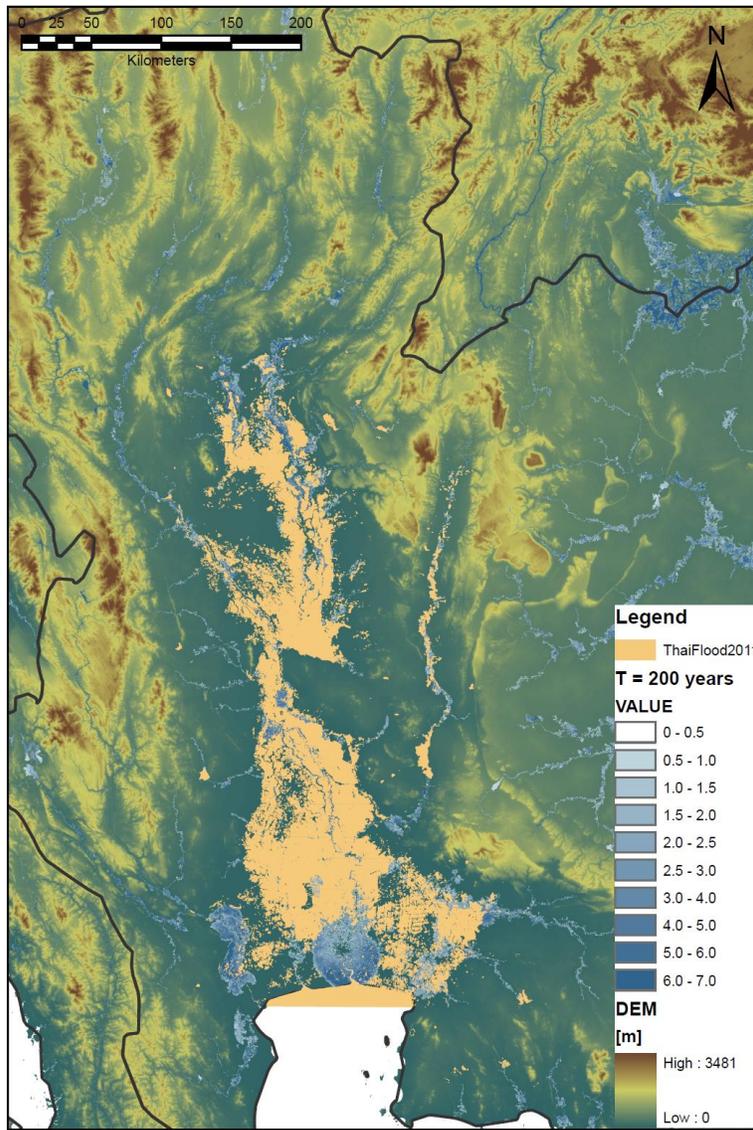
**Figure 41 Flood maps for Thailand UNOSAT flood footprint envelope of 2011 event layered on the flood map for T = 200 years. Reconditioned DEM in the background.**

## 1.7 Bibliography

Arcement, G.J., Schneider, V.R., 1989. Guide for selecting Manning's roughness coefficients for natural channels and flood plains. US Geological Survey, Water Supply Paper No. 2339.

Biswas, A.K. & Fleming, G. (1966) Floods in Scotland: magnitude and frequency. Wat. Wat. Engng, 246-252.

Bobe´e B, Rasmussen PF, Recent advances in flood frequency analysis. U.S. National Report to IUGG, 1991-1994, Rev. Geophys., 1995; 33(Suppl.):1111-6.

Bocchiola, D., De Michele, C., and Rosso, R., (2003), Review of recent advances in index flood estimation, Hydrol. Earth Syst. Sci., 7, 283-296.

Bougthon W, Droop B. Continuous simulation for design flood estimation - a review. Environ Model Software 2003;18:309-18.

Bras RL. Hydrology an introduction to hydrologic science. Addison-Wesley; 1990.

Burn D., H., Catchment similarity for regional flood frequency analysis using seasonality measures, Journal of Hydrology 202, 212–230, 1997.

Brunner, G. W. 2008. "HEC-RAS, River Analysis System hydraulic Reference Manual. US Army Corps of Engineers, Hydraulic Engineering Center USACE HEC". 411Davis.

Chong, S. & Moore, S.M. (1983) Flood frequency analysis for small watersheds in southern Illinois. Wat. Resour. Bull. 19(2), 277-282.

Chow VT, Maidment DR, Mays LW. Applied hydrology. McGraw-Hill; 1988. p. 444-70.

Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena. University of East Anglia Climatic Research Unit (CRU). [Phil Jones, Ian Harris]. CRU Time Series (TS) high resolution gridded datasets, [Internet]. NCAS British Atmospheric Data Centre, 2008, *Date of citation*. Available from http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1256223773328276

Cole, G. (1966) An application of the regional analysis of flood flows. In: Symposium on River Flood Hydrology, Session B, 39-57. Institution of Civil Engineers, London.

Dalrymple, T. (1960) Flood frequency analyses. US Geol. Surv. Wat. Supply Pap. 1543-A.

Gabriele S, Arnell NW. A hierarchical approach to regional flood frequency analysis. Water Resour Res 1991;27(6):1281-9.

Global Land Cover 2000 database. European Commission, Joint Research Centre, 2003. http://gem.jrc.ec.europa.eu/products/glc2000/glc2000.php; BARTHOLOME, E. M. and BELWARD A. S., 2005, GLC2000; a new approach to global land cover mapping from Earth Observation data, International Journal of Remote Sensing, *26*, 1959 - 1977.

Hosking, J.R.M. and J.R. Wallis, 1993: Some statistics useful in regional frequency analysis. Water Resourc. Res., 29, 271-281

Kottegoda NT, Rosso R. Statistics, probability and reliability for civil and environmental engineers. McGraw-Hill; 1997.

Lehner, B., C. Reidy Liermann, C. Revenga, C. Vorosmarty, B. Fekete, P. Crouzet, P. Doll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J.C. Robertson, R. Rodel, N. Sindorf, and D. Wisser. 2011. Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Lehner, B. and Döll, P. (2004): Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296/1-4: 1-22.

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU, 89(10): 93-94.

Manfreda, S., Di Leo, M., and Sole, A.: Detection of Flood Prone Areas using Digital Elevation Models, J. Hydrol. Eng., posted ahead of print 3 January, doi:10.1061/(ASCE)HE.1943-5584.0000367, 2011.

Mitchell, T.D. and Jones, P.D., 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**, 693-712.

Peel, M. C., Finlayson, B. L., McMahon, T. A., Updated world map of the Koppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, 2007

Samuels, P. G., Backwater lengths in rivers, *Proc. Instn Civ. Engrs*, Part 2, 87, Dec., 571-582, 1989

Shuttle Radar Topography Mission (SRTM) Near-global Digital Elevation Models (DEMs). Produced from a collaborative mission by the National Aeronautics and Space Administration (NASA), the National Imagery and Mapping Agency (NIMA), the German Aerospace Center (DLR, Deutsches Zentrum fur Luft-und Raumfart), and the Italian Space Agency (ASI, Agenzia Spaziale Italiana). Available for electronic download and on CD-ROM from the U.S. Department of the Interior, U.S. Geological Survey, Earth Resources Observation Systems (EROS) Data Center (EDC), Distributed Active Archive Center (DAAC), Sioux Falls, South Dakota, USA.

Verdin, K.L., A System for Topologically Coding Global Drainage Basins and Stream Networks, Earth Resources Observation Systems (EROS) Data Center, 1997.