Local and regional flood frequency analyses with BayesianMCMC

Karine Halbert

With this document, we illustrate the use of the functions `BayesianMCMC` and `BayesianMCMCreg`. In the first section, flood frequency analysis is performed using the function `BayesianMCMC` with the data of the gauged station Saint Martin d'Ardèche (south-east of France), for which a lot of information on historical extreme floods is available (see Naulet, 2002). The methodology used for incorporation of historical information in `BayesianMCMC` is described in Payrastre et al. (2011). In the second section, a regional flood frequency analysis is performed using the function `BayesianMCMCreg` with data from several sites in the Ardèche region, which is exposed to frequent and severe flash floods, thus having a lot of records of extreme flood observations at different ungauged sites that were collected under the European Hydrate project (Borga et al., 2011; Gaume et al., 2009). Methodological aspects of this regional flood frequency approach are developed in Gaume et al. (2010); Nguyen et al. (2014).

1 Local flood frequency analysis with historical data

We use here systematic data and historical flood data as described respectively in Nguyen (2012) and Naulet (2002). Load the data available in the **nsRFA** package:

```
> library(nsRFA)
> data(Ardechedata)
> ls()

[1] "Ardeche_areas"          "Ardeche_ungauged_extremes"
[3] "Beauvene_cont"          "Chambonas_cont"
[5] "SaintLaurent_cont"      "SaintMartin_cont"
[7] "SaintMartin_hist"       "Vogue_cont"
```

The systematic data, including the maximum annual flood peaks (m^3/s) for the station Saint Martin d'Ardèche is (Table A.10, page 180 in Nguyen, 2012, catchment size 2240 km^2):

```
> str(SaintMartin_cont)

'data.frame':        43 obs. of  2 variables:
 $ year: int  1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 ...
 $ peak: int  3160 1930 1300 540 1050 1220 1330 3250 1500 1540 ...
```

At this location, a lot of information on extreme historical floods have been collected and analysed in Naulet (2002, Table 5.8, page 200; Fig. 5.30, page 208).

```
> str(SaintMartin_hist)

List of 2
 $ peaks      :'data.frame':        32 obs. of  2 variables:
  ..$ year: int [1:32] 1772 1827 1846 1853 1857 1859 1861 1878 1888 1890 ...
  ..$ peak: int [1:32] 6000 7400 6350 4000 4550 3950 3000 6350 3200 7550 ...
 $ thresholds:'data.frame':         4 obs. of  4 variables:
  ..$ from.yr  : num [1:4] 1645 1772 1827 1892
  ..$ to.yr    : num [1:4] 1771 1826 1891 1962
  ..$ threshold: num [1:4] 7250 6000 5050 2400
  ..$ num.yr   : num [1:4] 127 55 65 71
```

The perception thresholds in `SaintMartin_hist$thresholds` indicate the dicharges that are assumed not to have been exceeded in the indicated time periods except for the historical floods in `SaintMartin_hist$peaks`. The historical period has been stopped here in 1962, i.e. before the beginning of the systematic series.

```
> plot(c(1640,2010), c(0, 8000), type="n", xlab="", ylab="discharge (m3/s)")
>  points(SaintMartin_cont[,"year"], SaintMartin_cont[,"peak"], type="o", pch=21, bg="white")
>  points(SaintMartin_hist$peaks[,"year"], SaintMartin_hist$peaks[,"peak"], pch=16)
>  points(SaintMartin_hist$peaks[,"year"], SaintMartin_hist$peaks[,"peak"], type="h")
>  segments(x0=SaintMartin_hist$thresholds[,"from.yr"],
+           x1=SaintMartin_hist$thresholds[,"to.yr"],
+           y0=SaintMartin_hist$thresholds[,"threshold"], lty=2)
```
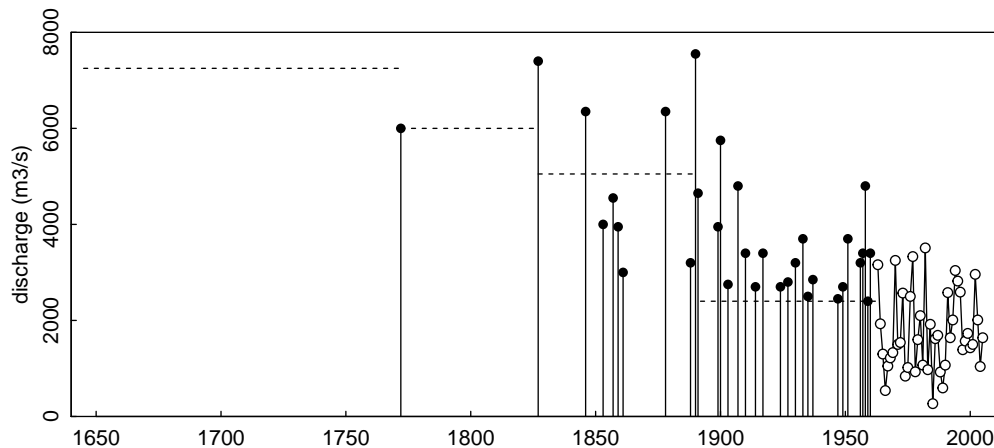

With the function `BayesianMCMC` we fit a flood frequency model to the flood data using a Bayesian Markov Chain Monte Carlo algorithm. The particularity of this function is that it takes into account historical information (see e.g. Payrastre et al., 2011), and that several perception thresholds can be specified. In order to do so, we have to organise the historical data into 3 vectors of equal length:

- `xhist` containing the historical flood peak values;

- `seuil` containing the perception thresholds associated with each historical flood in `xhist` ;

- `nbans` containing the number of years covered by the percception thresholds in `seuil` (number of years during with the thresholds have not been exceeded except for the floods listed in `xhist`).

```
> xhistSM <- c(-1,SaintMartin_hist$peaks[,"peak"])
> seuilSM <- c(7250, 6000, rep(5050, 10), rep(2400, 21))
> nbansSM <- c(127, 55, 65, rep(0,9), 71, rep(0, 20))
> data.frame(xhistSM, seuilSM, nbansSM)
```

```
   xhistSM seuilSM nbansSM
1       -1    7250     127
2     6000    6000      55
3     7400    5050      65
4     6350    5050       0
5     4000    5050       0
6     4550    5050       0
7     3950    5050       0
8     3000    5050       0
9     6350    5050       0
10    3200    5050       0
11    7550    5050       0
12    4650    5050       0
```

13	3950	2400	71
14	5750	2400	0
15	2750	2400	0
16	4800	2400	0
17	3400	2400	0
18	2700	2400	0
19	3400	2400	0
20	2700	2400	0
21	2800	2400	0
22	3200	2400	0
23	3700	2400	0
24	2500	2400	0
25	2850	2400	0
26	2450	2400	0
27	2700	2400	0
28	3700	2400	0
29	3200	2400	0
30	3400	2400	0
31	4800	2400	0
32	2400	2400	0
33	3400	2400	0

This presentation of the dataset gives directly the correpondance between each historical flood and the associated perception threshold. Since the same threshold value is repeated several times (once for each historical flood), the application period of each threshold has to be parted in `nbansSM`: the only requirement is that for each threshold, when you sum up the `nbansSM` values next to it, the sum is the total number of years covered by the threshold. For simplicity reasons, here we just put once in `nbansSM` the total number of years covered by each threshold, the other values being set to 0. Finally, in this step you just have to repeat in `seuilSM` the value of each perception threshold as many times as peak discharge information is available during the application period of the threshold, and then indicate once in `nbansSM` the duration of the application period for each threshold, the other values being set to zero.

You can also notice that the first value in `xhistSM` is equal to -1: this is because the perception threshold defined in `seuilSM` has never been exceeded and no information is available on floods that did not exceed the threshold (see help).

The Bayesian MCMC algorithm is applied as follows:

```
> fit <- BayesianMCMC(xcont=SaintMartin_cont[,"peak"],
+                     xhist=xhistSM,
+                     seuil=seuilSM,
+                     nbans=nbansSM,
+                     nbpas=1000, nbchaines=3, confint=c(0.05, 0.95), dist="GEV",
+                     varparameters0=c(NA, NA, 0.5))
```

Here we choose a generalized extreme value (GEV) distribution as model, with 3 Markov chains, and a confidence interval of 90% (from 5% to 95%). Note that, to have better convergence, higher values of `nbpas` should be used (e.g., 10000).

Usually you need to run it a second time to make sure the Bayesian MCMC algorithm has converged, that is what the `BayesianMCMCcont` function is for:

```
> fit <- BayesianMCMCcont(fit)
```

The results of the fitting and some diagnostic graphs can be plotted with the command:

```
> plot(fit, which=1:4, ask=TRUE)
```

The first plot which can be seen is the plotting position of the data, as well as the fitted distribution (with the maximum likelihood) and the confidence intervals:
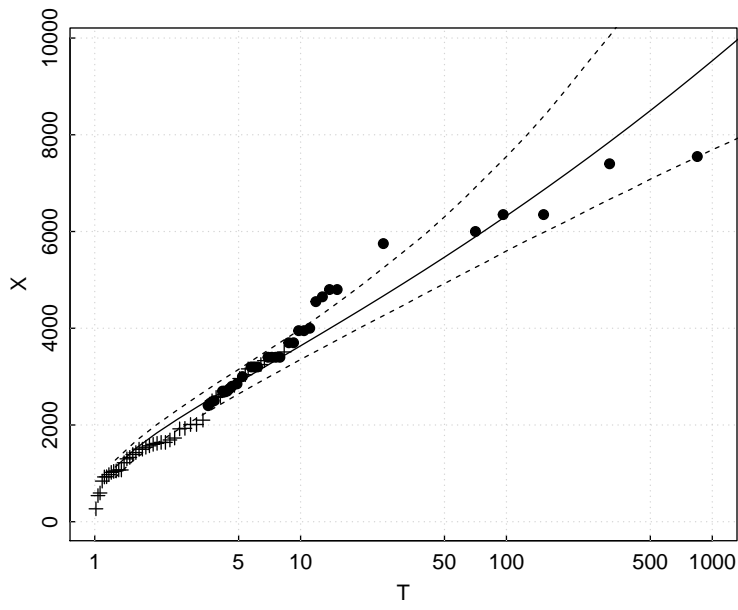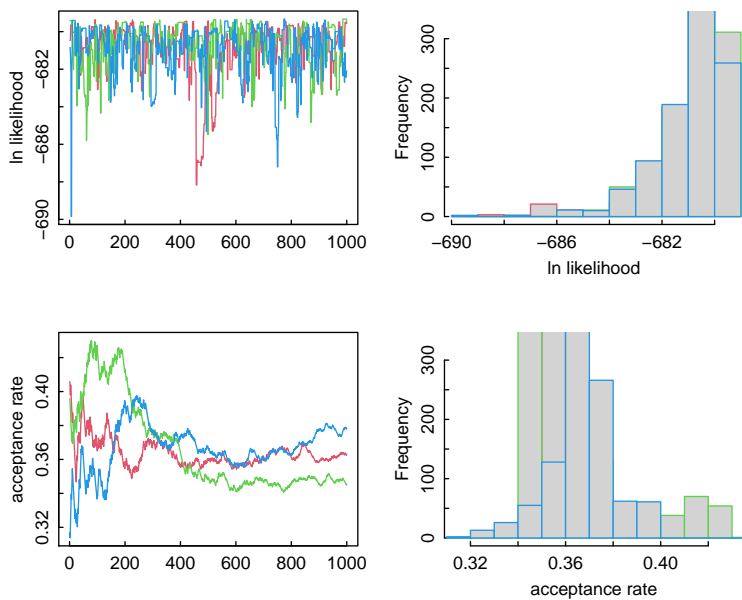
Note that the crosses are standing for the systematic data and the points for the historical data.
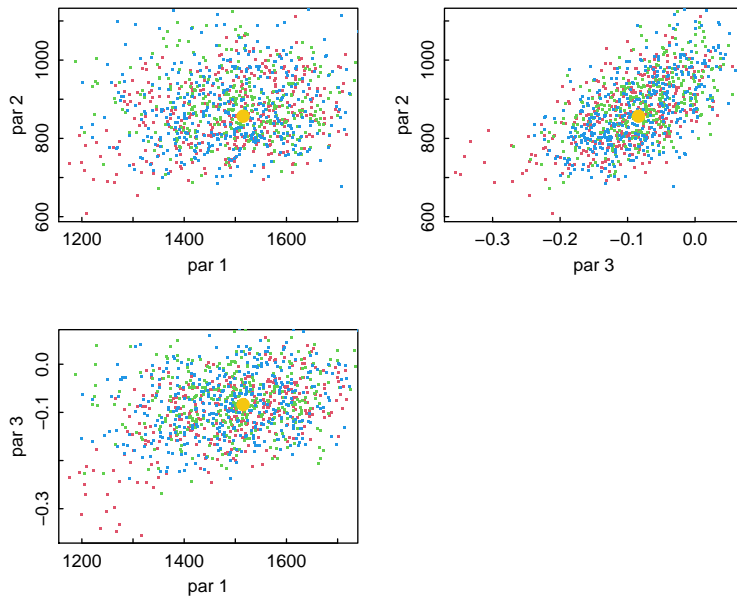
The second plot allows a diagnostic of the MCMC simulation by looking at the parameters:
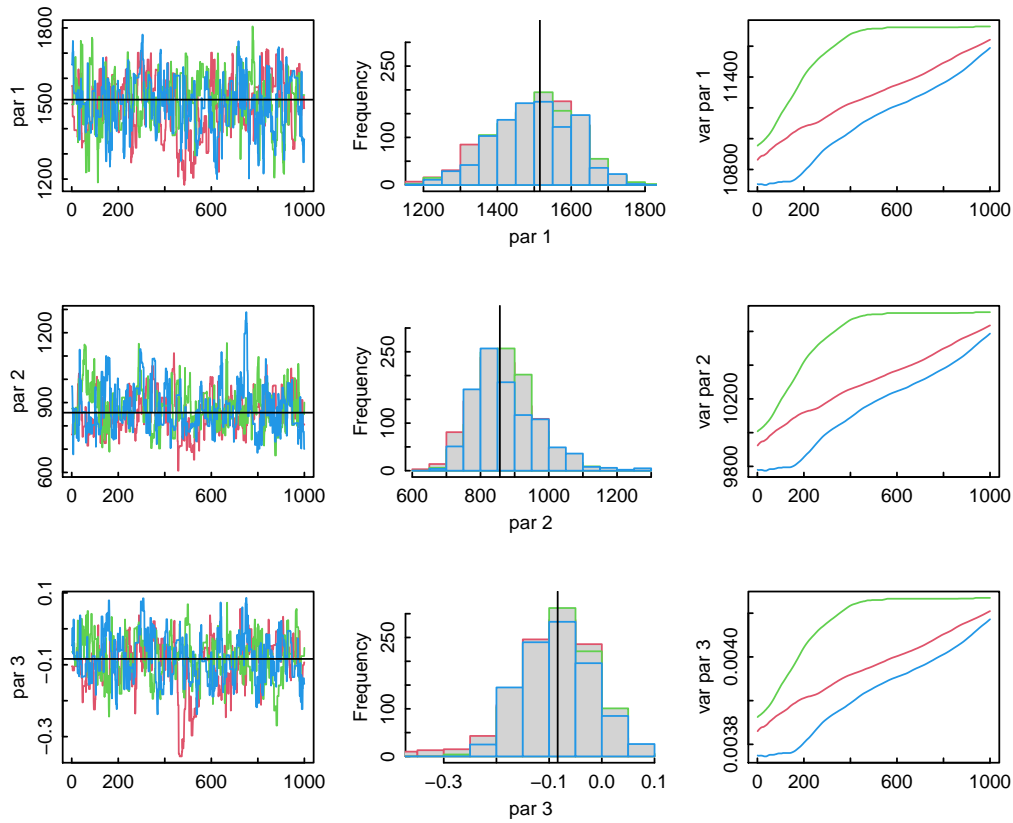
The third plot is also for a diagnostic of the MCMC simulation but looking at the likelihood and MCMC acceptance rate:

Finally, the fourth plot is showing as cloud plots the posterior distribution of parameters obtained with the MCMC simulation:


2 Regional flood frequency analysis with extreme floods in ungauged catchments

We use here data from 5 streamgauges in the Ardèche region as described in Nguyen (2012, pages 180-184).

```
> SaintMartin_cont
> Vogue_cont
> SaintLaurent_cont
> Beauvene_cont
> Chambonas_cont

> plot(SaintMartin_cont[,"year"], SaintMartin_cont[,"peak"], type="o", pch=21, bg="white",
+      xlab="", ylab="discharge (m3/s)", main=" Saint Martin on the Ardeche River (2240 km2)")
> plot(Vogue_cont[,"year"], Vogue_cont[,"peak"], type="o", pch=21, bg="white",
+      xlab="", ylab="discharge (m3/s)", main="Vogue on the Ardeche River (636 km2)")
> plot(SaintLaurent_cont[,"year"], SaintLaurent_cont[,"peak"], type="o", pch=21, bg="white",
+      xlab="", ylab="discharge (m3/s)", main="Saint Laurent on the Borne River (63 km2)")
> plot(Beauvene_cont[,"year"], Beauvene_cont[,"peak"], type="o", pch=21, bg="white",
+      xlab="", ylab="discharge (m3/s)", main="Beauvene on the Eyrieux River (392 km2)")
> plot(Chambonas_cont[,"year"], Chambonas_cont[,"peak"], type="o", pch=21, bg="white",
+      xlab="", ylab="discharge (m3/s)", main="Chambonas on the Chassezac River (507 km2)")
```

Saint Martin on the Ardeche River (2240 km2)

Beauvene on the Eyrieux River (392 km2)

Vogue on the Ardeche River (636 km2)

Chambonas on the Chassezac River (507 km2)

Saint Laurent on the Borne River (63 km2)
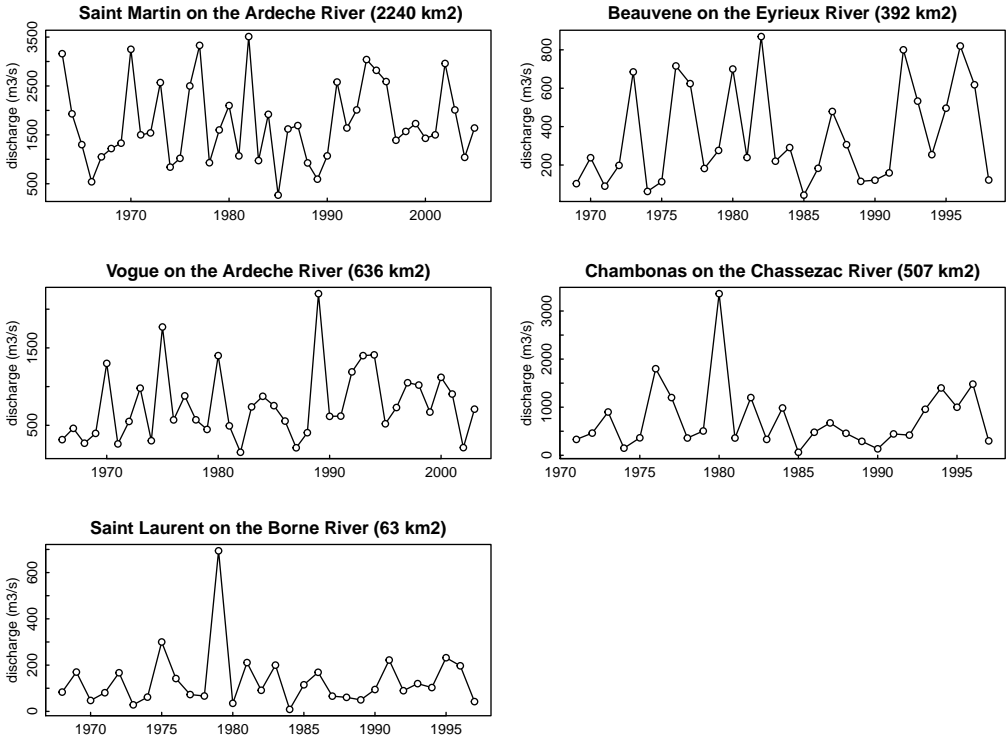
On top of that, extreme flood events have been reconstructed for several ungauged catchments, or for periods preceeding the systematic records in gauged catchments (see Table 4.2, page 127, in Nguyen, 2012).

```
> Ardeche_areas   # km2
```

	station	area
1	Rosieres	210
2	Joyeuse	100
3	Vans	6
4	PontRolandy	150
5	Dorne	78
6	Meyras	130
7	Meysee	56
8	Rieutord	62
9	Burzet	47
10	Lamastre	242
11	Pouzin	140
12	Chambon	11
13	Saliouse	61
14	PontCezar	635
15	Chambonas	507
16	Aubenas	480
17	BarrageCollanges	343
18	SaintMartin	2240
19	Vogue	636
20	SaintLaurent	63
21	Beauvene	392

```
> Ardeche_ungauged_extremes   # m3/s
```

	station	peak	associated.period
1	Rosieres	1820	50
2	Joyeuse	1000	50
3	Vans	130	50

7

```
4        PontRolandy 1150              50
5               Dorne  630              50
6              Meyras  900              50
7              Meysee  500              50
8             Rieutord  444              50
9               Burzet  350              50
10           Lamastre  970              50
11             Pouzin  700              50
12            Chambon  100              50
13           Saliouse  300              50
14           PontCezar 1500              50
15          Chambonas 3360              50
16            Aubenas 2200              50
17 BarrageCollanges 1685              50
18          SaintMartin 4500              50
```

> Ardeche_ungauged_extremes_new <- merge(Ardeche_ungauged_extremes, Ardeche_areas)

In the following, the usage of the function `BayesianMCMCreg` is illustrated to provide a regional flood frequency analysis including both systematic and extreme flood data (see Gaume et al., 2010). This function uses the catchment areas to draw an index flood relationship on the form $\mu = S^{\beta}$. Therefore, the datasets have first to be organised in the following way:

- one vector `xcont` including all the systematic series;

- one vector `scont` including the catchment surfaces for each record in `xcont`;

- one vector `xhist` including the peak discharges of extreme floods;

- one vector `shist` including the catchment surfaces for each extreme flood in `xhist`;

- one vector `seuil` including the perception thresholds associated with each extreme flood in `xhist`;

- one vector `nbans` including the period of application of each perception threshold, i.e. the period during which the threshold has been exceeded only by the extreme flood in `xhist`

```
> xcont <- c(SaintMartin_cont[,2],
+            Vogue_cont[,2],
+            SaintLaurent_cont[,2],
+            Beauvene_cont[,2], Chambonas_cont[,2])
> scont <- c(rep(2240, length(SaintMartin_cont[,2])),
+            rep(636, length(Vogue_cont[,2])),
+            rep(63, length(SaintLaurent_cont[,2])),
+            rep(392, length(Beauvene_cont[,2])),
+            rep(507, length(Chambonas_cont[,2])))
> xhist <- Ardeche_ungauged_extremes_new[,"peak"]
> xhist[Ardeche_ungauged_extremes_new[,"station"]=="Chambonas"] <- -1
> shist <- Ardeche_ungauged_extremes_new[,"area"]
> nbans <- Ardeche_ungauged_extremes_new[,"associated.period"]
> seuil <- Ardeche_ungauged_extremes_new[,"peak"]
```

`BayesianMCMCreg` can then be applied in the following way:

```
> fitreg <- BayesianMCMCreg(xcont=xcont,
+                           scont=scont,
+                           xhist=xhist,
+                           shist=shist,
+                           nbans=nbans,
+                           seuil=seuil,
+                           nbpas=1000, nbchaines=3, confint=c(0.05,0.95), dist="GEV",
```

```
+                                        varparameters0=c(NA, NA, NA, 0.5))
> fitreg <- BayesianMCMCregcont(fitreg)
> fitreg <- BayesianMCMCregcont(fitreg)
```

Note that, to have better convergence, higher values of `nbpas` should be used (e.g., 10000).

The method assumes that, once the data are scaled by the catchment area (to the power of β), the peaks can be pooled into a single sample and the regional growth curve can be estimated using it. Several assumptions are therefore made: that the scaled samples are homogeneous, identically distributed and independent.

One good thing is that the method uses the MCMC algorithm not only for the regional growth curve parameters, but also to estimate β.

Also, the method allows to include historical data (associated with gauged series) as well as extreme discharge events available at ungauged sites, as in this example, which are treated as the historical ones. In the example, it is assumed that extreme floods observed have not been exceeded in the last 50 years (ungauged sites), or during the 50 years before the beginning of the gauged series (historical floods). Therefore, here `seuil=xhist`).

Since one of the extreme floods recorded is already included in the systematic gauged series (1980 flood at Chambonas), its value has been replaced here by -1 in `xhist`, and the corresponding lines in `nbans` and `seuil` are kept unchanged to indicate that this extreme flood has not been exceeded during a 50-year period preceeding the systematic series (see above).

Note that once again you may need to run the Bayesian MCMC algorithm a second time to make sure it converged, this time with the function `BayesianMCMCregcont`.

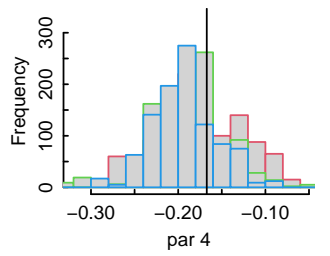The same graphics, as the ones plotted for the local analysis, can be plotted here.
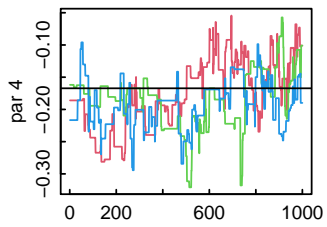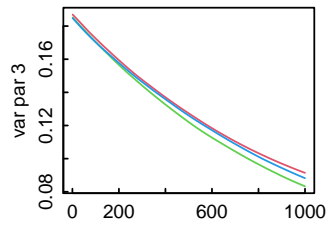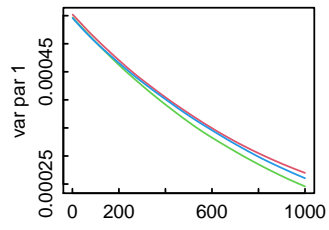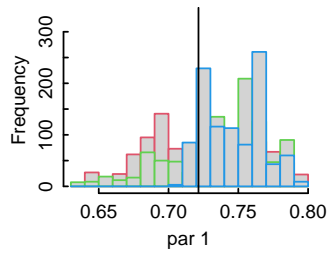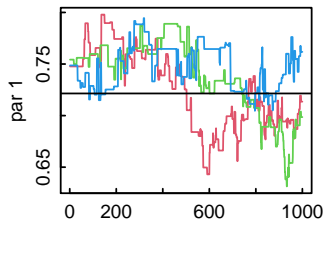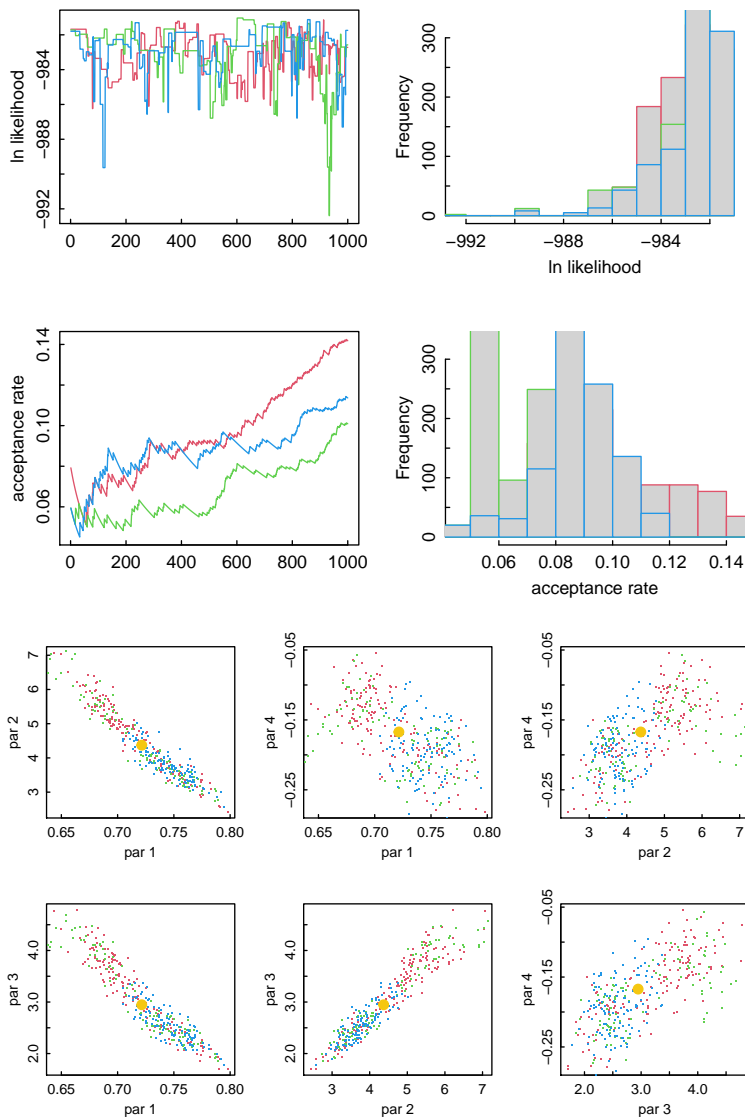
```
> plot(fitreg, which=1:4, ask=TRUE)
```

The first one refers to the regional growth curve, valid for an ungauged site with area 1 km^2:
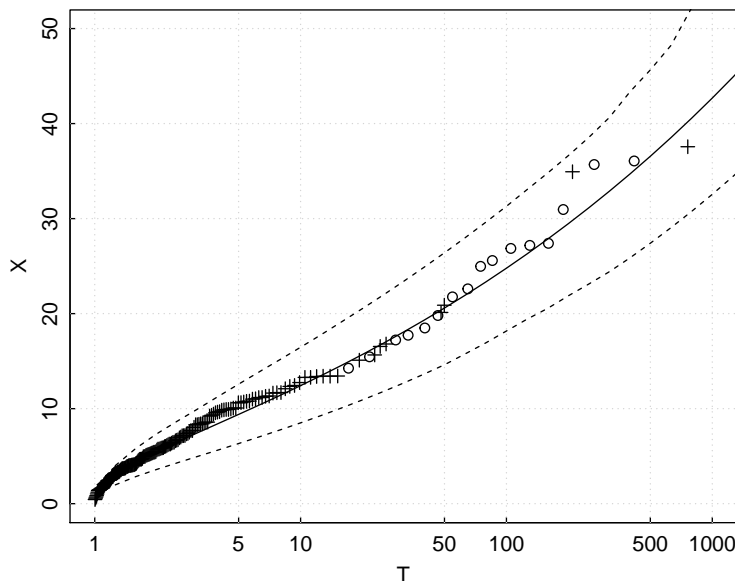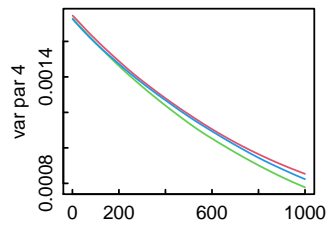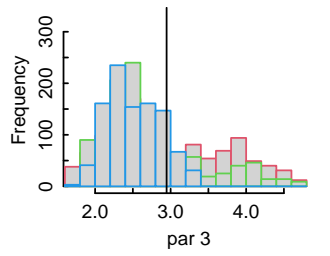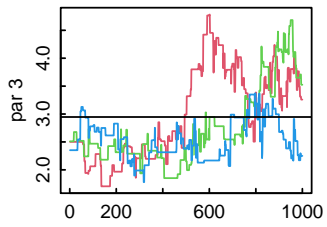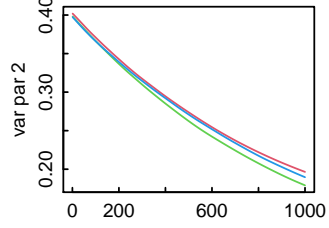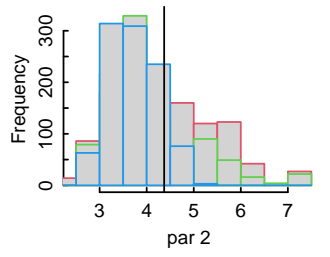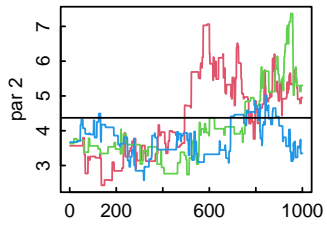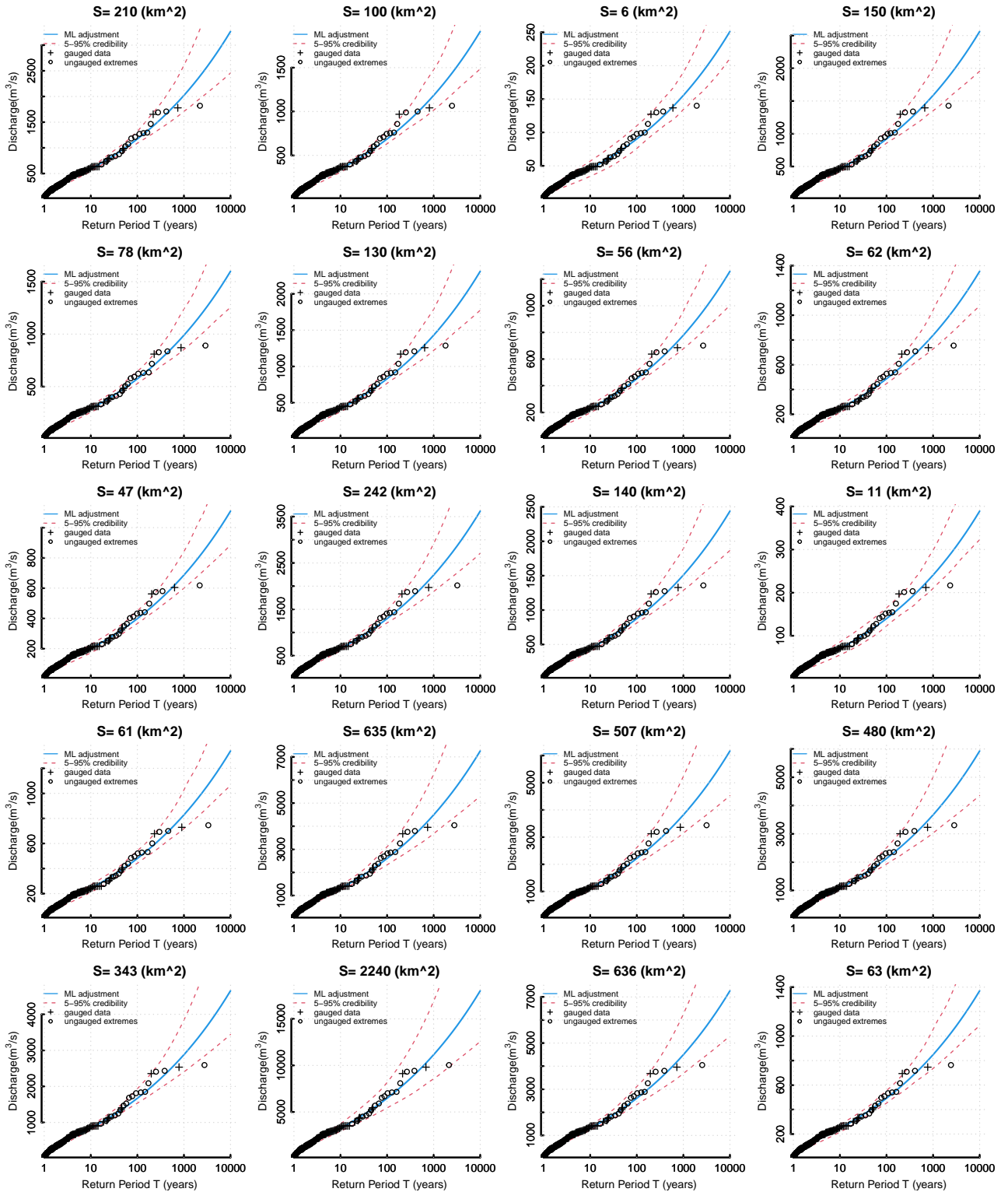
For graphs accounting for catchment areas, the function `plotBayesianMCMCreg_surf` can be used. Here we will do it for every catchment surface of the gauged stations:

```
> plotBayesianMCMCreg_surf(fitreg, surf=unique(Ardeche_areas$area))
```

The last graph produces means of the previous mean discharge and 90% confidence intervals as a function of the surface. It efficiently shows how the surfaces of the chosen catchments have an influence on the algorithm. For more information about how the regional version works, you can see the PhD thesis Nguyen (2012), and Nguyen et al. (2014).

References

Borga, M., E.N. Anagnostou, Blöschl G., and Creutin, J.D. (2011). Flash flood forecasting, warning and risk management: the HYDRATE project, *Environmental Science & Policy*, 14(7):834–844, doi:10.1016/j.envsci.2011.05.017.

Gaume, E., V. Bain, , P. Bernardara, O. Newinger, M. Barbuc, A. Bateman, L. Blaskovicova, G. Bloschl, M. Borga, A. Dumitrescu, I. Daliakopoulos, J. Garcia, A. Irimescu, S. Kohnova, A. Koutroulis, L. Marchi, S. Matreata, V. Medina, E. Preciso, D. Sempere-Torres, G. Stancalie, J. Szolgay, J. Tsanis, D. Velasco, and A. Viglione, (2009). A compilation of data on European flash floods, *Journal of Hydrology*, 367:70–80, doi:10.1016/j.jhydrol.2008.1012.1028.

Gaume, E., L. Gaál, A. Viglione, J. Szolgay, S. Kohnová, and G. Blöschl (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, *Journal of Hydrology*, 394:101–117, doi:10.1016/j.jhydrol.2010.01.008.

Naulet, R. (2002). Utilisation de l'information des crues historiques pour une meilleure prédétermination du risque d'inondation. Application au bassin de l'Ardèche à Vallon Pont-d'Arc et St-Martin d'Ardèche. PhD thesis at CEMAGREF, Lyon, and at the Université du Québec, pp. 322. http://www.lthe.fr/OHM-CV/Documents/theses/these_naulet.pdf

Nguyen, C.C. (2012). *Amélioration des approches Bayésiennes MCMC pour l'analyse régionale des crues (Improvement of BayesianMCMC approaches for regional flood frequency analyses)*. PhD thesis at the University of Nantes, pp. 192. (http://archive.bu.univ-nantes.fr/pollux/show.action?id=178da900-8dd8-491b-8720-011608382b98)

Nguyen, C.C., E. Gaume, and O. Payrastre (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations, *Journal of Hydrology*, 508:385–396, doi:10.1016/j.jhydrol.2013.09.058.

Payrastre, O., E. Gaume, and H. Andrieu (2011). Usefulness of historical information for flood frequency analyses: Developments based on a case study, *Water resources research*, 47, doi:10.1029/2010WR009812.