# EXTREMILES IN ENVIRONMENTAL RESEARCH AND METEOROLOGY

Abdelaati Daouia [1] & Thibault Laurent [2]

[1] *Toulouse School of Economics, France, Abdelaati.Daouia@tse-fr.eu*
[2] *Toulouse School of Economics, CNRS, France, Thibault.Laurent@tse-fr.eu*

**Résumé.** La question du changement climatique s'est imposée comme un enjeu politique mondial. Records de chaleur, multiplication des catastrophes météo, fonte des glaces, déclin de la nature : les preuves de l'impact dévastateur des activités humaines sur la planète s'accumulent. De nombreuses études scientifiques s'appuient sur des données météorologiques, devenues plus nombreuses et facilement accessibles. Dans ce travail, nous appliquons des outils statistiques habituellement utilisés dans l'évaluation des risques dans le secteur de la banque et de l'assurance, à des données environnementales mesurant le changement climatique. Plus précisément, nous calculons les mesures de risque extremiles (Daouia *et al.*, 2019, 2022) sur deux échelles géographiques différentes : la donnée satellite elle-même d'une part et à l'échelle du pays d'autre part. Les mesures de risque sont calculées à différente période entre 1980 jusqu'à 2023. Les résultats montrent une tendance globale à la hausse des risques de fortes précipitations et de vagues de chaleur au cours du temps. Par ailleurs, toutes les zones géographiques ne sont pas exposées de la même façon. Les risques de précipitations élevées sont concentrés autour des tropiques alors que les risques de vagues de chaleurs sont très forts dans les régions proches des pôles, en Afrique centrale, le nord de l'Amérique du Sud, en Europe continentale et en Amérique du Nord.

**Mots-clés.** Environnement et statistique, mesures de risques, quantiles, extemiles, données météorologiques.

**Abstract.** Climate change has emerged as a global political issue. Evidence of the devastating impact of human activities on the planet is accumulating, including record heat, increasing number of weather disasters, melting ice, decline of nature. Many scientific studies are based on meteorological data that has become abundant and easily accessible. In this work, we apply a recent statistical tool from risk handling in finance and insurance to environmental data for assessing some aspects of climate change. More specifically, we use the concept of extremiles, which defines an asymmetric least squares analog to quantiles (Daouia *et al.*, 2019, 2022), as an alternative risk measure to the standard Value at Risk. In order to quantify how extreme precipitation and heatwaves have become over time, we perform extremile estimation and inference at different time periods between 1980 and 2023 on two different geographical scales: the satellite data itself besides the country scale. The results show an overall increasing evolution in heavy precipitation and extreme heatwaves. However, the tail exposure differs following the geographical location. The risks of high precipitation are concentrated around the tropics, while the risks of heatwaves are very strong in regions near the poles, in Central Africa, northern South America, Europe, and North America.

**Keywords.** environment and statistics, risk measures, quantile, extremile, climate change data.

# 1 Introduction

Meteorological data provide nowadays alarming aspects of global warming and climate change with tremendous impact on society, as they are among the most prominent topics of discussion. In statistical decision theory, one is typically interested in the analysis of heavy precipitation (Fischer and Knutti, 2016) and extreme temperatures (Katz and Brown, 1992) due to their adverse effects, such as floods or droughts. The risk of disasters is usually summarized by a risk measure that is estimated from historical data. The most common risk measure used in environmental research and meteorology is Value at Risk, which is a simple tail quantile of the underlying distribution. It is, however, known that quantiles only depend on the frequency of tail observations and not on their severity. Therefore, sample quantiles are either insensitive to the magnitude of infrequent catastrophic events, or they completely breakdown for high tail probability levels.

Extremiles (Daouia *et al.*, 2019, 2022) have recently emerged as an alternative least squares analog of quantiles. They are determined by weighted expectations rather than tail probabilities, and are preferred in this respect over quantiles in terms of alertness and reactivity, as well as resistance to outliers. Their use as an instrument of risk protection in finance and actuarial science has revealed their specific merits and strengths in the axiomatic theory of risk measures. This article is the first work to implement their contribution to mitigating the impact of climate extremes. In recent years, meteorological data has become abundant and easily accessible at extremely fine spatial and temporal scales. In Section 2, we construct, at different geographical scales, two annual indicators of climate change: a first indicator linked to heavy precipitation and another one related to climate warming. Section 3 presents the extremile risk measure, while Section 4 provides our preliminary results.

# 2 Data

We use the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2), provided by the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) from 1980 to 2023. MERRA-2 is fully described in Gelaro *et al.* (2017) and used in many recent scientific articles (see, for instance, `https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/pubs/`). All of the MERRA-2 atmospheric variables are provided at $0.5^\circ \times 0.625^\circ$ spatial resolution: It has 576 points in the longitudinal direction and 361 points in the latitudinal direction, *i.e.* a total of $576 \times 361 = 207,936$ measurement cells covering the whole earth.[1] In addition, several time steps are available (hour, day, month, year) depending on the variable considered.

**Heavy precipitation:** In cities and towns, high precipitation rates overwhelm storm drains and cause flash flooding. They can also be cause for great concern in rural areas by drowning crops, eroding topsoil, and damaging roads. We consider as a measure of heavy precipitation the maximum consecutive three-day precipitation in a year, which is also adopted by

---

[1] Note that the maximum distance between two measurement cells is around $70km$ on the equator line.

the Climate Atlas of Canada (see, for instance, `https://climateatlas.ca/variables`). It is computed using the cumulative sum of the precipitation corrected variable (PRECTOT-CORR, expressed as $kg/m^2/s$ and converted to $mm/day$) over each three consecutive days period. Our first indicator corresponds to the maximum value observed over a period of one year at the cell level.

**Climate warning:** High, persistent temperatures increase the risk of drought, which can severely impact food production and increase the risk of wildfire. High temperatures can also lead to more thunderstorms, which means increased risks of flash flooding, lightning, hail, and perhaps even tornadoes. We consider, as a measure of high temperature, the annual number of days satisfying conventional heatwave criteria. The World Meteorological Organization defines a heatwave as five or more consecutive days during which the daily maximum temperature (variable T2MMAX in Kelvin, converted to Celsius) exceeds the average maximum temperature by $5°C$ or more. The average maximum temperatures were calculated using a running window of $+/-$ 7 days centered on each day of the year for the benchmark climatology decade from 1980 to 1989. Our second indicator corresponds to the total annual count of days satisfying the heatwave criteria at the cell level.

Figure 1 displays the boxplot of the two indicators computed for each year on the $207,936$ cells, along with the yearly quantile plots at levels 0.5, 0.95, 0.99, 0.995, and 0.999. While it is difficult to observe a substantial change in the interquartile range of the precipitations' distribution, the quantile values at hight levels tend to increase significantly across time. As regards heatwaves, although the entire distribution appears to be shifting toward higher values, it is the large-order quantiles that experience the largest increases. This motivates the need for using statistical tools to measure the evolution of extreme values in the data.

# 3 From quantile to extremile risk assessment

Daouia et al. (2019, 2022) have introduced an alternative class to quantiles, called extremiles, which suggest better capability of fitting spread in data points and displaying interesting features of heavy-tailed distributions. Let $Y$ be a random variable with continuous distribution function $F$. Given a $\tau \in (0,1)$, the **quantile** $q_\tau$ of $F$ can uniquely be defined through

$$\begin{aligned} F^{-1}(\tau) &= \inf\{y : F(y) \geq \tau\} \\ &\equiv \text{median}(Z_\tau), \end{aligned}$$

where the random variable $\boldsymbol{Z_\tau}$ has c.d.f. $F_{Z_\tau} = K_\tau(F)$ with $K_\tau(t) = t^{r(\tau)}$ and $r(\tau) = \frac{\log(1/2)}{\log(\tau)}$. The **extremile** of order $\tau$ of $Y$ is then defined as the expectation of $Z_\tau$, $\mathbb{E}(Z_\tau)$, that is,

$$\xi_\tau = \mathbb{E}\left[Y J_\tau(F(Y))\right] = \int_0^1 J_\tau(t)\, q_t\, dt = \int_0^1 q_t\, dK_\tau(t), \tag{1}$$

where $J_\tau(t) = K'_\tau(t) = r(\tau)\, t^{r(\tau)-1}$. When $r(\tau)$ is a positive integer, we then have

$$\xi_\tau = \mathbb{E}\left[\max(Y_1, \ldots, Y_{r(\tau)})\right], \tag{2}$$

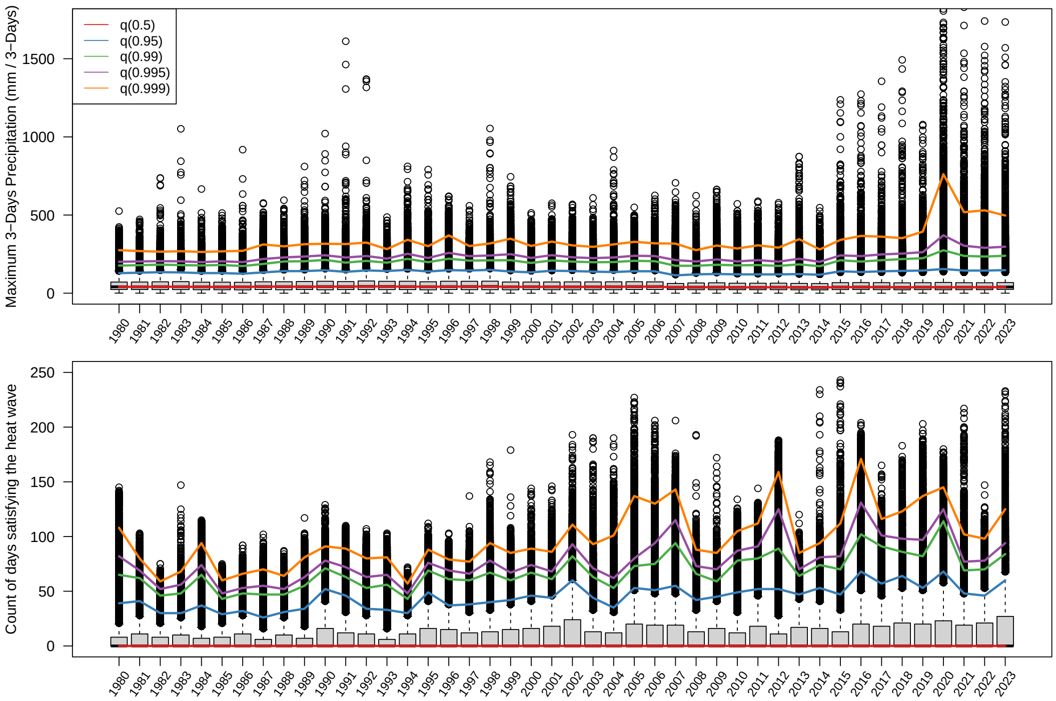

Figure 1: Boxplot of the two indicators (maximum consecutive three-day precipitation on the top and number of days satisfying the heatwave criteria on the bottom) computed for each year on the $207,936$ cells. The quantile plots (of order 0.5, 0.95, 0.99, 0.995 and 0.999) indicate an increase in extreme values.

for independent copies $Y_i$ of $Y$. In our assessments with annual data, the extremile of order $\tau = 0.95$ has the intuitive meaning as the average maximum value that we would obtain every 14 years.

Extremiles also define a least squares analog

$$\xi_\tau = \arg\min_\theta \mathbb{E}\left\{J_\tau(F(Y)) \cdot |Y - \theta|^2\right\}$$

to quantiles

$$q_\tau = \arg\min_\theta \mathbb{E}\left\{J_\tau(F(Y)) \cdot |Y - \theta|\right\},$$

by substituting the squared deviations in place of the absolute deviations. As such, they are motivated via several angles, which reveals their specific merits and strengths, including the following properties:

- The existence of extremiles requires a finite first moment.
- Extremiles have a straightforward interpretation in terms of expected maxima.
- Extremiles are law–invariant and have an explicit integral representation in terms of the quantile function ($L$-statistic).

- Extremiles take into account the whole of the distribution: they depend on the value of each observation and give more weight to extremes (while quantiles only depend on the relative frequency of tail observations).
- Extremiles define a coherent and comonotonically additive risk measure, which is more conservative (pessimistic) than the quantile-Value at Risk (VaR).

# 4 Results

## 4.1 At the cell level

We have 44 observations (one observation per year) at each of the $207,936$ cells. The top panels of Figures 2 and 3 show the cartography of the sample extremiles computed at level $\tau = 0.95$ for our two indicators. We indicate in small triangles the 100 zones with the highest tail exposure. For the top ten risky zones with the highest estimated extremiles, the bottom panels of Figures 2 and 3 exhibit the evolution of both sample $\tau$th extremiles and quantiles for $\tau \geq 0.80$.

One may wonder wether extremiles are high because the indicators are usually abundant (high values every year) or if this is only due to an exceptional value in a given year. It is exactly for this reason why we superimpose the quantile and extremile plots, for the top 10 risky cells, in the bottom panels of Figures 2 and 3. Another interesting question is to track how do risk measures change over time. Since we only have 44 annual observations per cell, we will answer this question below, in subsection 4.2, with data aggregated at the country level so as to have larger datasets.

**Heavy precipitation:** We can see in Figure 2 (top) that the risk areas for heavy precipitation are mainly located between latitude $-30°$ and latitude $30°$, which unsurprisingly corresponds to areas with a tropical climate characterized by heavy precipitation in the summer. Among the Top 10 values, we can observe in Figure 2 (bottom) that some locations are risky because there are heavy precipitations every year (Top 1, 3, 4, or 5 that correspond to cells located in the Independent State of Papua New Guinea, Kenya, Nigeria, and India). By contrast, we observe that some zones (Top 2, 6, or 7 that correspond to cells located in Mexico, Morocco, and Angola) are risky because of only one or two extreme events that influence the extremiles. It is interesting to see that in such heavy-tailed cases, sample quantiles remain much less alert to these catastrophic events than their extremile analogs.

**Climate warning:** We can see in Figure 3 (top) that the most impacted zones by climate warning are localized near the poles, in South America (mostly in Venezuela and Brazil), in Central Africa (the Democratic Republic of the Congo, Angola), continental Europe, and North America. It may seem surprising that certain regions or countries, such as the Sahara or India, do not appear among the risky regions. We recall that heatwaves occur when temperatures exceed the normal thresholds observed between 1980 and 1989 by $5°C$. As temperatures were already high between 1980 and 1989 in these areas, even if temperatures have increased by a few degrees in recent years, they did not exceed $5°C$. Among the Top 10 values, we can observe in Figure 3 (bottom) that all locations are risky not because of just one or two rare events, but due to frequent heatwaves in these zones, which explains why tail quantiles and extremiles are very close (it is established in Daouia et al., 2019, that tail quantiles and extremiles are equivalent for light-tailed distributions).

## 4.2 At the country level

We aim to compute risk measures at the country level and see their evolution over time. In doing so, we consider all cells falling within the borderlines of a given country. A country contains, on average, 240 cells. Russia is the country with the largest number, with more than 10,000 cells. Note that each country has at least 10 cells (for small countries, such as island countries, we add the nearest neighbors until we have 10 cells by country).

Figures 4 and 5 showcase the sample extremiles computed at the country level for four distinct 10-year periods: 1984-1993, 1994-2003, 2004-2013, and 2014-2023. As visualized from Figure 4, precipitations tend to increase across time in the countries close to the tropics. This is particularly visible for Kenya in Central Africa, Bangladesh in South Asia, Nicaragua in Central America, or the Philippines in Southeast Asia, where the highest extremile measurements are observed in 2023. In Bangladesh, the risk of heavy rainfall over 3 days has increased from around $200mm$ thirty years ago to $700mm$ nowadays.

We remark from Figure 5 that extremiles of heatwaves tend to increase across time everywhere, in particular near the poles, in Central Africa (with a peak observed during the 2004-2013 period), northern South America, Europe and North America. For example, in Venezuela, the risk of heatwave has increased from around 5 days per year thirty years ago to 135 days nowadays.

# Bibliographie

Daouia A., Gijbels I. and Stupfler G. (2019), Extremiles: A New Perspective on Asymmetric Least Squares, *Journal of the American Statistical Association*, 114(527), pp. 1366–1381.

Daouia A., Gijbels I. and Stupfler G. (2022), Extremile Regression, *Journal of the American Statistical Association*, 117(539), pp. 1579–1586.

Fischer, E. and Knutti, R. (2016), Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, 6, pp. 986—991.

Gelaro R., McCarty W., Suárez M.J., Todling R., Molod A., Takacs L., Randles C.A., Darmenov A., Bosilovich M. G., Reichle R., Wargan K., Coy L., Cullather R., Draper C., Akella S., Buchard V., Conaty A., da Silva A. M., Gu W., Kim G.-K., Koster R., Lucchesi R., Merkova D., Nielsen J. E., Partyka G., Pawson S., Putman W., Rienecker M., Schubert S.D., Sienkiewicz M., and Zhao B. (2017) The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30(14), pp. 5419–5454.

Katz R.W. and Brown B.G. (1992), Extreme events in a changing climate: variability is more important than averages. *Climatic change*, 21(3), pp. 289–302.

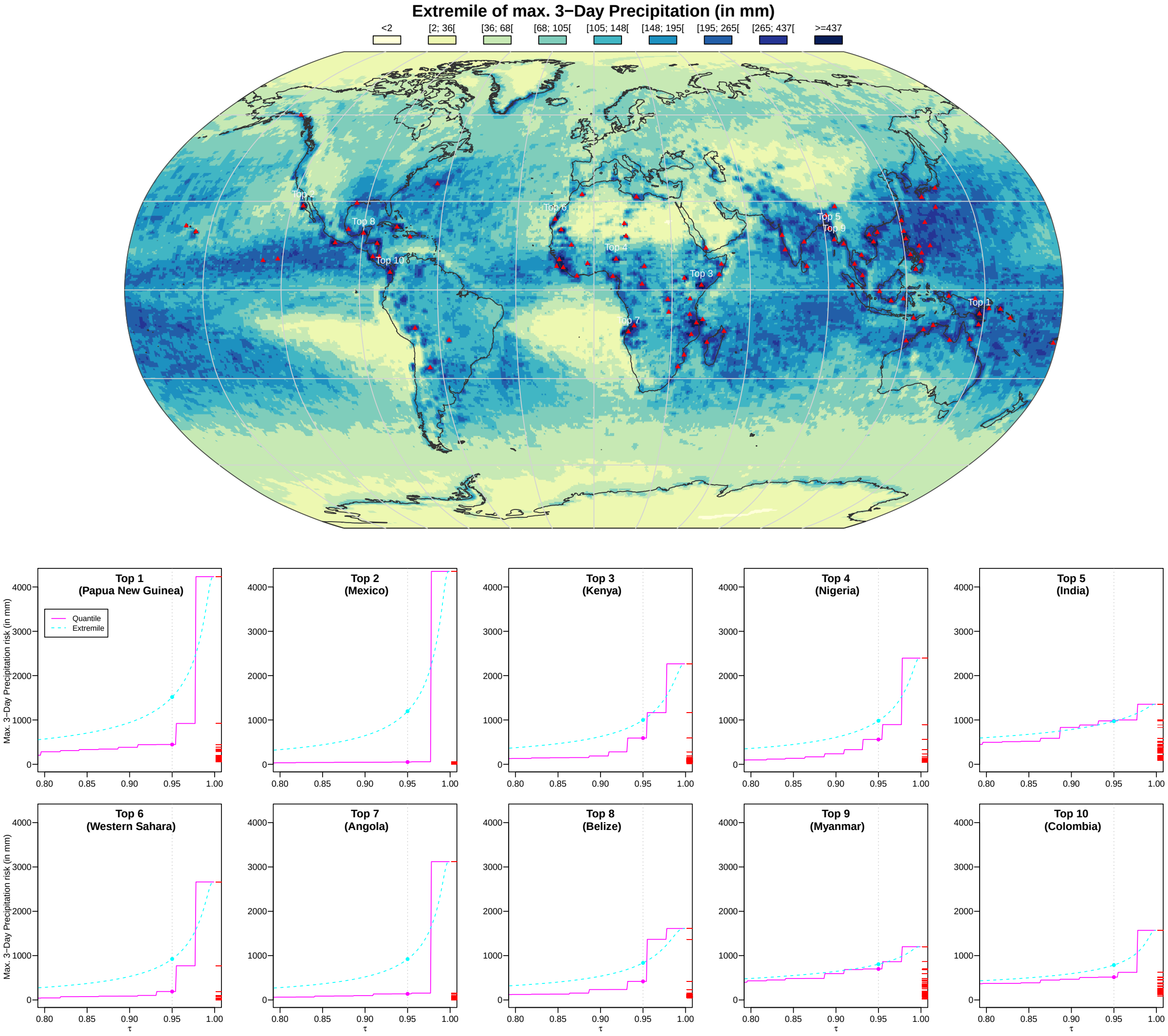

Figure 2: (Top) Map of the sample 0.95th extremile computed at the cell level for the heavy precipitation. The small triangles in red represent the 100 areas with the highest risk values. (Bottom) Plots of the quantile and extremile risks as functions of $\tau \in [0.80, 1]$ for the top ten risky zones with the highest estimated extremiles.

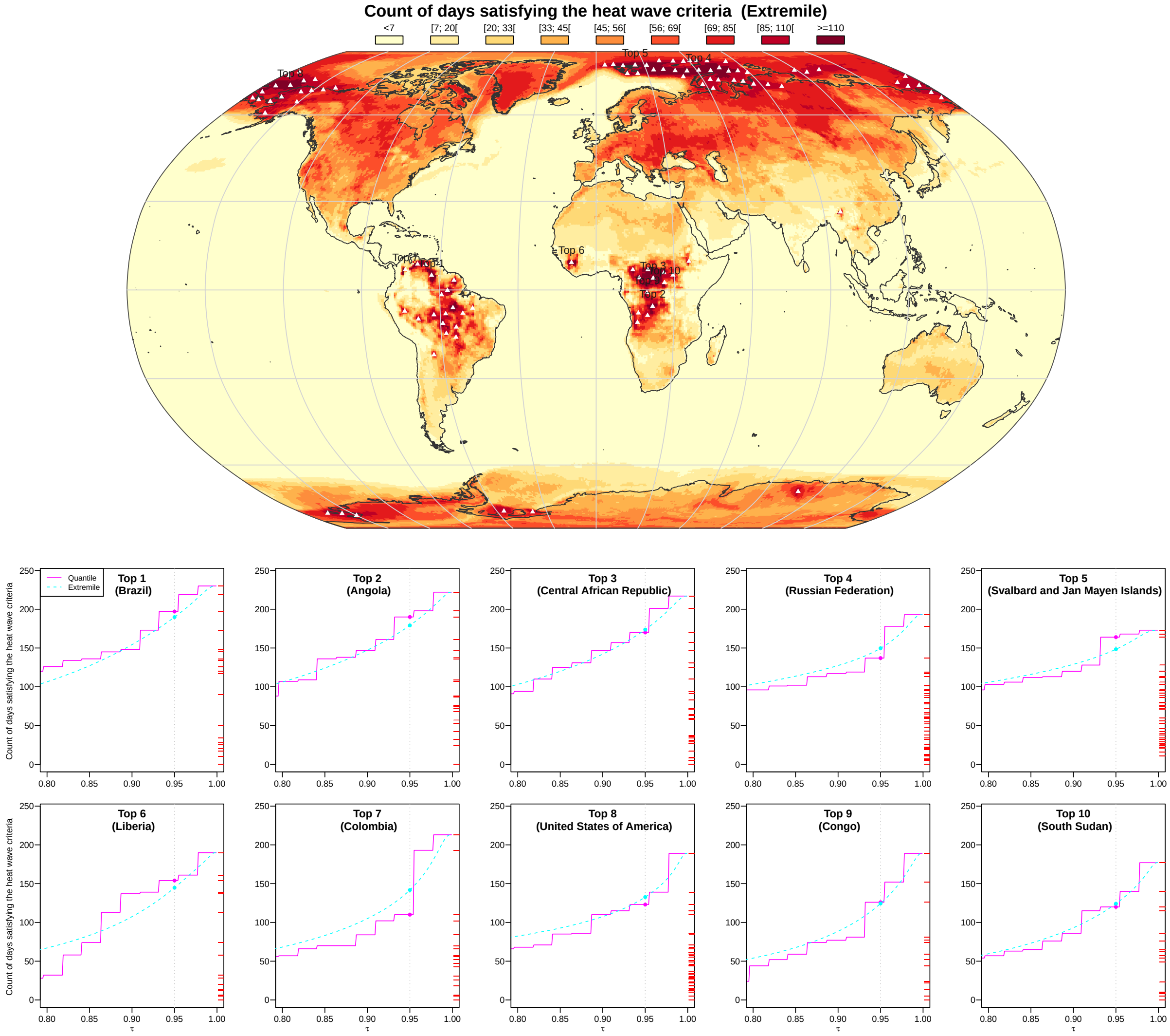

Figure 3: (Top) Map of the sample 0.95th extremiles computed at the cell level for the climate warning indicator. The small triangles in white represent the 100 areas with the highest risk values. (Bottom) Plots of the quantile and extremile risks as functions of $\tau \in [0.80, 1]$ for the top ten risky zones with the highest estimated extremiles .

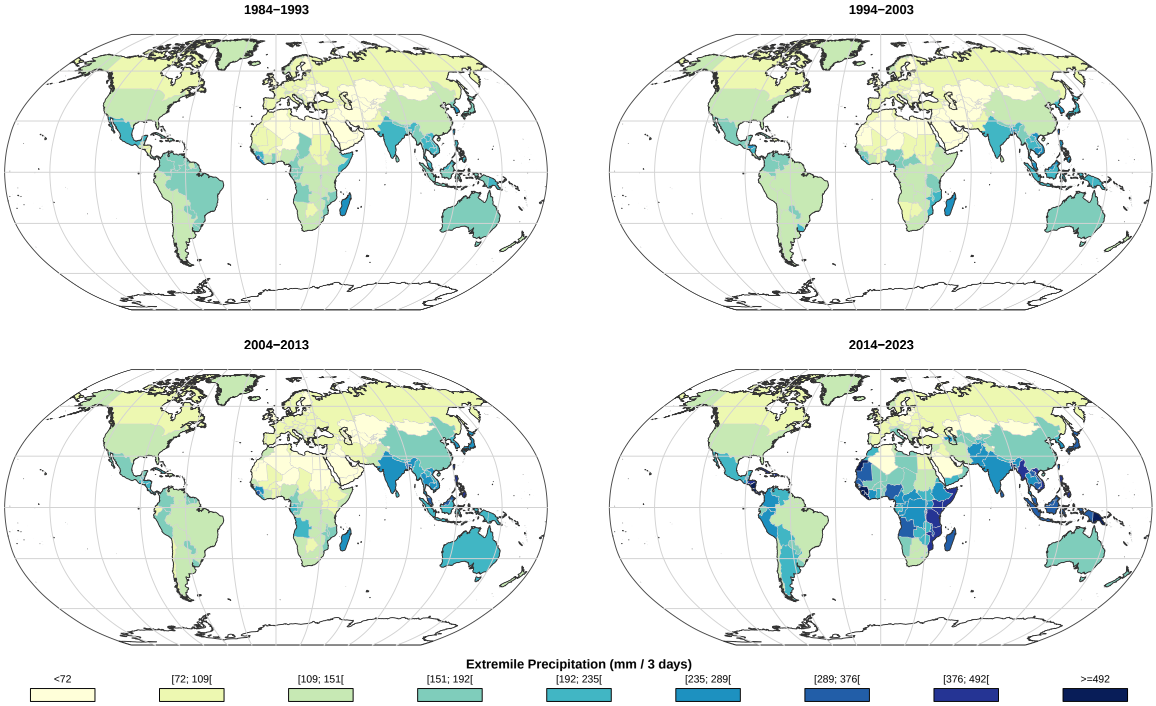

Figure 4: Estimated extremiles for heavy precipitations by country and over periods of 10 years

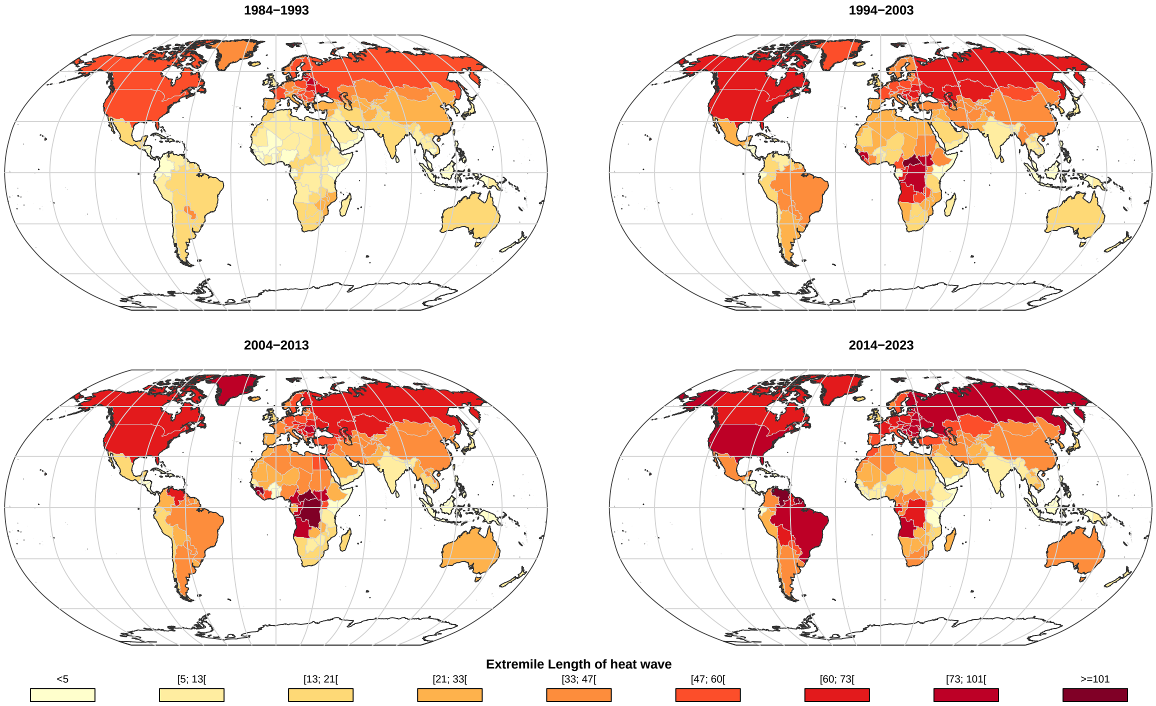

Figure 5: Estimated extremiles for heatwaves by country and over periods of 10 years