RISKADAPT
Modeling of Risks and Adaptation

D3.2 – Extreme value distributions for Europe – present climate (resubmission)

Project name

Asset Level Modelling of RISKs In the Face of Climate Induced Extreme Events and ADAPTtation (RISKADAPT)

Horizon Innovation Actions | Project No. 101093939

HORIZON-MISS-2021-CLIMA-02-03

Funded by the European Union

D3.2 – Deliverable Name	
Dissemination level	Public
Type of deliverable	Report
Work package	WP3 – Climate Data, CC Forcing and Multi-Hazard Modelling
Status - version, date	Final original – V1.0, 30/10/2023 Final revised – V2.0, 30/01/2025
Deliverable leader	Marko Laine (FMI)
Contributing partners	
Contractual date of delivery	31/08/2023
Resubmission date of delivery	31/01/2025
Keywords	extreme value analysis

Quality Control

	Reviewer Name	Organisation	Date
Peer review 1	Dimitrios Bilionis	ERRA	29/01/2025
Peer review 2	Gasper Rak	ULFGG	29/01/2025

Version History

Version	Date	Organisation	Summary of changes
0.1	01/09/2023	FMI	First draft of report structure
1.0	23/10/2023	FMI	Updated version of the report.
1.1	25/10/2023	FMI	Report ready for internal review.
1.2	27/10/2023	FMI	Report (internal reviewers comments addressed)
1.3	30/10/2023	ERRA	Quality reviewed version ready for submission
1.4	15/01/2025	FMI	Revision after EU review.
2.0	30/01/2025	ERRA	Quality check - Ready for submission.

Legal Disclaimer

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

Abbreviation	Meaning
CC	Climate Change
CDS	Copernicus Climate Data Store
CMIP6	Coupled Model Intercomparison Project Phase 6
CORDEX	Coordinated Regional Downscaling Experiment
E-OBS	CDS daily gridded meteorological data
ECMWF	European Centre for Medium Range Weather Forecasts
EO	Earth Observation
ERA5	fifth generation ECMWF reanalysis
GEOSS	Global Earth Observation System of Systems
GEV	Generalised Extreme Value
GPEX	Global Precipitation Extremes data set
HBV	conceptual hydrological model from GloH2O project
IDF	Intensity Duration Frequency
LCC	Life Cycle Cost
LES	Large Eddy Simulation
MCMC	Markov Chain Monte Carlo
MIS	Model Information System
PALM	modeling system for atmospheric and oceanic boundary layer flows
WP	Work Package

Executive Summary

RISKADAPT will provide, in close cooperation with the end-users/other stakeholders, a novel, integrated, modular, interoperable, public and free, customisable user-friendly platform (PRISKADAPT), to support systemic, risk-informed decisions regarding adaptation to Climate Change (CC) induced compound events at the asset level, focusing on the structural system. PRISKADAPT will explicitly model dependencies between infrastructures, which, inter alia, will provide a better understanding of the nexus between climate hazards and social vulnerabilities and resilience. Moreover, this project will identify gaps in data and propose ways to overcome them and advance the state of the art of asset level modelling through advanced climate science to predict CC forcing on the structure of interest, structural analyses, customised to the specific structure of interest, that consider all major CC induced load effects in tandem with material deterioration, novel probabilistic environmental Life Cycle Assessment (LCA) and Life Cycle Cost (LCC) of structural adaptation measures and a new model to assess climate risk that will combine technical risk assessment with assessment of social risks. PRISKADAPT will provide values to a set of indicators for each asset of interest, quantifying primary parameters and impacts, in the form of a Model Information System (MIS) that will provide all required information for adaptation decisions. PRISKADAPT will be implemented in the case studies in the pilots that involve specific assets, however, it will permit customisation with local values of parameters and data, so it can be applicable throughout Europe for CC adaptation decisions involving assets of similar function, exposed to multiple climate hazards.

This report is one of the three deliverables of WP3 "Climate Data, CC Forcing, Multi-Hazard Modelling" and corresponds T3.1 "Climate data for hydrological analyses, wind and rain forcing and material degradation. Extreme Value Analyses" of the RISKADAPT project. To meet the aim of this Task, in this report the terminology, methods, and tools that can be used to perform statistical extreme value analysis on climate data are presented. Input data can mainly come from reanalysis data (like ERA5) or from in-situ time series measurements. In addition, and as the focus of this report is on the analysis of present climate, i.e., approximate for years 1970-2020, pointers to Copernicus Climate Data Store (CDS) data sets and tools are provided that can be used to evaluate risks caused by extreme events for infrastructure, especially those that are related to project's pilots.

This report provides methodological guidance on the application of Bayesian hierarchical modelling techniques to improve the reliability of extreme value analysis by incorporating spatial and temporal dependencies. By quantifying return levels and return periods of extremes of key climate variables, these statistical tools provide essential methods in studying climate change induced risks for infrastructure. We evaluate climate data sources relevant to extreme weather and hydrological events, including precipitation, wind speed, and temperature extremes. The report demonstrates statistical techniques using real-world datasets, with a particular focus on case studies relevant to RISKADAPT pilot sites. This deliverable also includes practical examples of extreme value analysis using Python-based computational tools, enabling application in different infrastructure contexts. The extreme value analysis presented here is relevant for climate scientists, engineers, policymakers, and infrastructure managers involved in climate adaptation planning.

However, while reanalysis data sets presented here, such as ERA5 Land, provide valuable insights, limitations in spatial and temporal resolution necessitate careful methodological considerations, including downscaling approaches, which will be dealt in more detail in other deliverables of the RISKADAPT project.

1. Introduction

In the RISKADAPT project extreme value distributions are used to study climate data for hydrological analyses, wind and rain forcing, and material degradation. Extreme value analyses will be performed to the relevant climatic variables and parameters that have been identified in WP2. For the present climate, focus is especially on the state-of-the-art C3S CDS reanalysis data and E-OBS in-situ data, both available from the C3S CDS. The next deliverable, D3.3, will focus more on the future climate and the effect of climate change on extreme values. For the CC data, the focus will be on the regional CORDEX data as well as on the most recent global climate projection data CMIP6. In WP3 Task 3.1, Bayesian hierarchical approach to extreme value analysis will be applied, allowing us to account for spatial and temporal dependencies on the variables and to study the temporal evolution of the distributional parameters.

1.1 Purpose of the deliverable

The purpose of this Deliverable is to describe the available sources of climate data that can be used as input for extreme value analysis. In doing this, the statistical analysis that can be performed to monthly or yearly maxima of quantities of interest is described; the difficulties in obtaining the data as well the adequacy of available reanalysis data sets in terms of spatial and temporal resolution are explained; and the need for further downscaling of results is also discussed.

The objectives related to this deliverable have been achieved. However, there was a two-month delay in submitting the deliverable, mostly due to not being able to formulate the needs of the partners in terms of extreme value analyses. At the time of writing this deliverable, it is still a work in progress.

1.2 Structure of the deliverable

The Deliverable has been structured as follows:

- Chapter 1 is the introduction
- Chapter 2 describes extreme value analysis and data sources relevant to RISKADAPT project. There are three Python notebooks included in the Annexes that provide examples of analyses that are possible with the current data sources.
- Chapter 3 summarises the main outcomes of this Deliverable and outlines the future work.

1.3 Intended audience

D3.2 is a public document according to the project's Description of Action (DoA). Thus, its intended audience is not limited only to project's partners and officer but it extends outside the consortium. Possible external target groups include public authorities, infrastructure owners and operators, researchers and technology providers, interested in a report that presents the basic terminology related to extreme value analysis.

2. Extreme values and climate

2.1 Introduction

Extreme weather can mean weather phenomena that are rare in climatological sense, e.g., if the average return period of occurrence is more than every 30 years. On the other hand, it can mean weather events that cause extreme damage, like heavy rain, wildfires, or heat waves, even if they are not rare in statistical sense. In this report, we concentrate on climatological features of weather-related variables, mostly for temperature, wind speed and precipitation, as they are relevant in RISKADAPT project's pilots and other tasks. One special quantity of interest is icing, that is target of Task 4.4 in WP4 as well as of Pilot 2. As there are not enough direct observational icing data, we need to use icing models. These models take as input several weather variables, such as temperature and the amount of liquid water in the clouds. The different approaches to study extreme icing will be discussed in more detail below.

The previous deliverable D3.1 for RISKADAPT WP3 described different sources of EO data that can be used in the analyses of the project. The main data source was the Copernicus Climate Data Store (CDS): ERA5 reanalysis, E-OBS observation analysis, and CMIP6 regional climate model runs. Other important source mentioned was the GEOSS Global Earth Observation System of Systems as well as data from national and regional weather institutes that provide open access to their data, both to observations and model output. For extreme value analysis, we will be utilizing the same sources and provide examples on how specific questions can be answered based on the available data. We provide Python code for downloading the data and calculating return periods for quantities of interest based on statistical extreme value theory.

2.2 Statistical extreme value analysis

Statistical extreme value theory is based on the fact that the maximum value in a set of random numbers follows, under quite general assumptions, a distribution called generalised extreme value distribution, GEV. The same applies to the minimum with a change of variables. When we know the statistical distribution of the phenomena of interest, we can draw inference on the occurrence probabilities and risks related to the event. In extreme value terminology, the things of interest are called return periods and return levels. These are defined shortly below.

We describe the classical extreme value theory as it is typically used in climate science. We will, in addition, use Bayesian terminology and methods, which allows us to discuss the relevance of prior information about model parameters and their correlation. Standard reference to statistical extreme value theory is by Coles (2001). Some other relevant literature on extreme value analysis of climate include works by AghaKouchak (2011) and Hamdi (2021).

The cumulative distribution function of GEV can be written with help of three parameters, location μ, scale σ, and shape ξ as

$$G(y;\, \mu, \sigma, \xi) = \exp\left(-\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_{+}^{-\frac{1}{\xi}}\right), \tag{1}$$

where y is the quantity of interest, such as yearly maximum temperature, μ, σ, and ξ are the model parameters and subindex + means taking maximum with zero. This equation directly leads to quantiles of the distribution as

$$y_p = \begin{cases} \mu - \dfrac{\sigma}{\mu}\left[1 - \{-\log(1-p)\}^{-\xi}\right], \text{ for } \xi \neq 0, \\ \mu - \sigma \log\{1 - \log(1-p)\}, \text{ for } \xi = 0. \end{cases} \tag{2}$$

This equation is used to calculate the return level y_p associated with a certain exceedance probability p or equivalently, the return period $T = 1/p$.

There are several approaches to estimate GEV distribution parameters given time series of occurrences of the quantity of interest. In the most basic cases we are assuming a stationary situation, where a set of fixed values of the parameters (μ, σ, and ξ) would explain the behaviour of the phenomena in a given geographical location. If this is the case, the equation (2) would directly give return levels corresponding to given return periods. For the estimation of parameters, we assume having a number of independent realizations of the extreme values which can be used to find a specific version of GEV that can reproduce the observed values. When basing an estimate to a finite number of observations, the estimates will have uncertainty, which is important to quantify for proper use in further analyses. Bayesian analysis is a suitable tool for this. We can base the analysis of uncertainties to Markov chain Monte Carlo (MCMC) simulation, which is a computational technique that provides samples from Bayesian uncertainty distributions of model parameters and model derived quantities.

As an example, consider intensity-duration-frequency (IDF) curves for precipitation. They are standard tools for managing rainfall data in environmental engineering. The curves show rainfall intensities and durations for given return period frequencies. IDF curves are location-specific, meaning that they need to be developed for each specific geographic area based on local rainfall data. They are a valuable tool for assessing and managing the risks associated with precipitation-related events. We provide an example below on calculating IDF curves from the precipitation extremes. Figure 1 below shows IDF curves for Haliacmon river drainage basing calculated using GPEX data. These curves can be used in hydrological analysis related to Pilot 1.

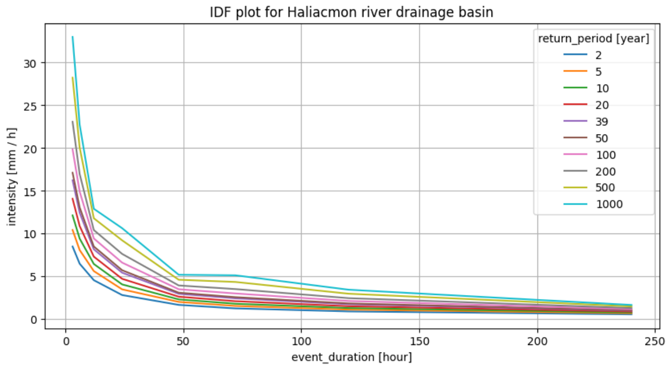


Figure 1: IDF curves that provide information on return periods of precipitation events of different intensity and duration. The data is averaged over the Haliacmon river drainage basin.

2.3 Climate variables

To analyse risks on the infrastructure, we need to consider the effects of weather. For example, when building near shore, we must account for the maximum water levels. As this will depend on random external events as well as on changes in the future climate, we need to think in terms of probabilities and risks.

Extreme weather events, such as extreme temperature or precipitation, are becoming the dominant factor in the causes of disasters that affect the safety of critical assets, like buildings and bridges, with long life span. Climate change will alter the frequency and effect of these phenomena, and it is important to be able to quantify the expected changes in risks and integrate climate change aspects into the adaptation planning process of existing structures and in the planning process of new structures.

In the RISKADAPT project, special attention is given to those climatic variables that affect critical infrastructure. The project pilots target bridges and buildings that are affected by heavy winds and rainfall. In addition, Pilot 2 studies the effect of icing on the energy transmission lines in northern countries. This means that variables of interest will be related to temperature, wind, precipitation, and river discharges, as well as those needed in modelling ice formation, such as cloud water content.

2.4 Available data

We will utilize the data sources described in deliverable D3.1. These include data available from Copernicus services, data provided from previous research projects, and various sources of in-situ time series. Historical reanalyses are useful as they provide complete, both time and space, reconstruction of the climate[1]. The drawback is that the resolution can be quite sparse. For example, ERA5 is delivered in spatial resolution of 0.25° (about 30 km), and ERA5 Land and E-OBS in 0.1° (11 km). In-situ or remote sensing observation archives give more accurate representation at a specific location, but they are spatially and temporally sparse. As an example of the effect of resolution, Figure 2 shows the Pilot 3 location Cattinara Hospital in Trieste, Italy, with data pixels from four different reanalysis data sets. The colouring shows whether the pixel is considered land or sea by the data.

There are some specific high-resolution data sets available from separate previous studies. For example, Pirinen (2022) provide 1 km x 1 km resolution daily gridded evapotranspiration dataset covering Finland over the 40-year period 1981–2020 that is based on in-situ rain gauge observations that are spatially interpolated using Kriging method.

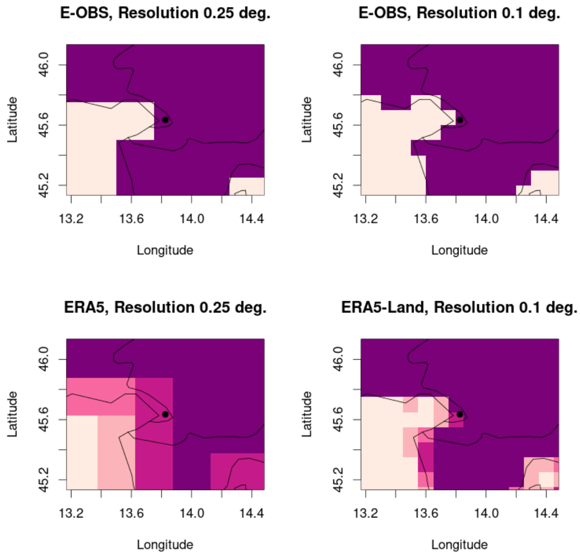


Figure 2: Location of Cattinara Hospital in Trieste, Italy within four different data sets. The darkest colour means land pixel, lightest is sea, while some pixels are considered as being in-between by the data set's land-sea-mask.

For climatic extreme values there are some previous projects that provide data openly for use. One example is GPEX data set (Gründemann, et al. 2023), which contains global extreme precipitation return levels. The data is available for download, and it has precipitation levels for several precipitations durations and return periods ranging from 2 years to 1000 years. From this data set, one can extract the area of interest, such as the Haliacmon river drainage basin (Figure 3) that is used in Pilot 1 of RISKADAPT for studying return times of extreme precipitations events for hydrological

[1] https://www.ecmwf.int/en/about/media-centre/focus/2023/fact-sheet-reanalysis

analysis. Figure 1 shows IDF curves calculated using the precipitation extremes in the GPEX data set and averaging the results over pixels that cover the drainage area.

Figure 3: Haliacmon river drainage basin that is used in RISKADAPT Pilot 1.

The GMEX data is produced using climate data from the GloH2O project (https://www.gloh2o.org), which provides high-resolution, bias-corrected weather data for the past, present, and future. They have several interesting data sets, such as regionalized parameter maps for the conceptual hydrological model HBV covering the entire land surface, https://www.gloh2o.org/hbv/. These will be useful for the hydrological analysis tasks in the RISKADAPT project.

2.4.1 Downscaling

Most climate reanalysis data are provided in spatial resolution of 10 to 50 km, and temporally most typically in hourly intervals. There are very few km-scale reanalysis and the existing ones do not usually allow for extreme value analysis of longer return periods. However, in many studies there is a need for downscaling the results to finer temporal and spatial scale. In RISKADAPT, the need for going from km-scale to even m-scale comes from the study of climate effects on individual constructions, such as buildings, bridges, or power transmission towers. For this reason, special methods are needed. But when the simulation results have been obtained, the statistical analysis of return periods is essentially the same.

If m-scale analysis of wind fields is needed, one approach is based on doing small scale simulation runs with meteorological large-eddy simulation (LES) models, such as PALM (Maronga, 2015). These can be based on larger scale mean wind fields which are rescaled when studying the effect of extreme wind events on local scale (see the work by Hellsten (2021) and the PANOPTIS and PLOTO EU projects[2]). Such high-resolution LES-based downscaling of the HCLIM regional climate model data has been carried out in RISKADAPT for the power-transmission line pilot case. The LES modelling covering the studied 8 km long section of the powerline in Kontiolahti, Finland has 2 m spatial resolution and about 0.1 s temporal resolution. Wind field time series with 3 s intervals covering the powerline section are stored for each 30-degree wind sector. These data are used for downscaling the 3-hourly HCLIM-data which has three orders of magnitude lower resolution. At each HCLIM 3 h time interval the wind direction is identified and matched with the LES wind sector and the corresponding LES wind field is rescaled according to the HCLIM wind speed. This matching is made in the HCLIM grid point horizontally closest to the powerline at 300 m height relative to the local terrain surface. The output

[2] http://www.panoptis.eu/, https://ploto-project.eu/

is the downscaled wind velocity field, stored with 0.33 Hz output frequency for all those 3-hour intervals on which the maximum downscaled gust wind speed exceeds a given threshold. The spatial resolution of the output data is reduced to avoid excessively large output files, resulting in the along-line output interval of 64 m, and the vertical output interval of 4 m.

The wind field downscaled this way provides much more realistic wind statistics for estimating the powerline wind loading and the subsequent risk analysis than the climate model data as such. And still the possible climate change effects are taken into account because the large-scale climate-model data drives the high-resolution downscaled data. The differences of the downscaled and non-downscaled RCM data are large around the powerline. In the downscaled data the top gust speeds are often about 50 % higher than those in the climate model data. The principal reason for this is that the forest canopy surrounding the powerline section has a very strong effect on wind turbulence within the roughness sublayer in which the powerline resides. This effect magnifies the wind gusts remarkably and it is not included in the original climate-model data.

Pilot 3 of the RISKADAPT project estimates the atmospheric load on tall buildings in case of extreme weather events. It will contain a detailed numerical study of the Cattinara Hospital in Trieste, Italy, which was selected as representative of a tall building in Europe exposed to strong wind, i.e. the Bora wind, due to the peculiar local meteorology. The downscaling methodology is based on a chain of multiscale numerical simulations, which are strategically linked together to transport and convey information from the climatic time-spatial scales, the meteorological mesoscale and the very small, local, building scale as the final result. The methodology is described in detail in RISKADAPT project's deliverable D3.5 "Model of high wind loading on a high-rise building" (Cintolesi et al., 2024).

2.4.2 Icing

RISKADAPT Pilot 2 and Task 4.4 consider the effect of atmospheric icing, especially ice formation on electric power lines. For this, a specific icing model is needed that can calculate the accumulation of ice on a surface based on variables that are available from numerical weather prediction models, climate model runs or from historical reanalyses. We are using the icing model defined in ISO 12494 (Makkonen, 2000) that is modified for climate model output. The model produces a time series of accretion of rime ice on a cylinder based on cloud liquid water content, temperature, pressure, and wind speed. For proper analysis, one needs to decide what are the icing events, for example in units of g/h, that are of interest, e.g., the amount of icing or the length of icing periods. What is also important is whether the information on the maximum yearly amount of passive ice formation and its possible trend due to climate change is enough or do we need analysis of finer details. These details need to be decided on before doing further analysis. In addition, the combined effect of icing and wind will be studied.

2.4.3 Obtaining and storing the data

Another challenge in extreme value analysis of climatic variables is that although many services can provide readily computed monthly means, they are not providing monthly maxima. For this reason, one might need to download or otherwise access the whole data in the original temporal resolution and then extract the statistics needed. With hourly resolution and for several decades, this will amount to very large data files. For analysis on high resolution data on the CDS archive, part of the data collection can be done using the CDS Toolbox[3], which saves from downloading large amounts of data. CDS has some useful pre-made applications that can be useful, too[4]. They include:

- Global temperature trend monitor.
- Daily statistics calculated from ERA5 data.

[3] https://cds.climate.copernicus.eu/cdsapp#!/toolbox
[4] https://cds.climate.copernicus.eu/cdsapp#!/search?type=application

- Extreme precipitation statistics for Europe.
- European hydrology seasonal forecast explorer.
- European hydrology and climate data explorer.

However, CDS and the provided CDS Toolbox applications can typically provide only limited amount of data at a time. In many cases, the user must download large data sets, extract extreme events of interest, and perform statistical analysis, which can be quite time consuming. For this purpose, it will be advisable to provide pre-calculated extremes, like monthly maximums of temperature, rain, and wind speed, and store them in RISKADAPT data management system, so that these can be easily accessed by project partners and by the PRISKADAPT platform that will be made later in the project. The use of user-friendly data formats and metadata is important for the usability.

2.5 Extreme value analysis examples

We provide some examples on extreme value analysis related to RISKADAPT Tasks. The Annex has Python notebooks that describe the use of reanalysis data and calculation of extreme value statistics. The examples concentrate on variables and locations that are relevant to the Pilots of the project. The code is available from https://github.com/fmidev/RISKADAPT/. The examples are listed in Table 1. In addition, we provide a set of Python functions that can be used to fit a generalized extreme value distribution to a time series data and calculate return levels for given return periods. The code uses Stan probabilistic language (https://mc-stan.org) to define the statistical model and calculate posterior probabilities for GEV parameters as well as for quantiles of the fitted distribution, which in turn can be used to calculate return levels and their uncertainties. The estimation is based on MCMC sampling. The statistical analysis is similar to the one performed by Räty et al. (2023), who studied maximum water levels for high-risk infrastructure of nuclear power plants located by the shore.

2.5.1 ERA5 extreme values

ERA5 reanalysis data from the Copernicus CDS service is the most comprehensive global reanalysis data. It has a large set of climate variables with 0.25 degree spatial and hourly temporal resolution. The data starts from year 1940 and it is continuously updated for present day values. We use ERA5 here as an example of obtaining extreme value analysis on the past and present climate. For extreme values, one is typically interested in maximum or minimum values over a longer period, like month or year. As the CDS service does not provide these maxima directly, we need to download hourly data and calculate aggregated values locally. Fortunately, the CDS download tools allow for downloading limited geographical areas to make the download size smaller. Here, it is advisable to use already downloaded data when possible. Many institutes store ERA5 and similar reanalysis data sets in their local storage systems. It is also possible to use cloud storages, such as Google Cloud or Amazon AWS for the analysis tasks.

In the examples provided, monthly maxima of 2-meter temperature, dew point temperature, wind speed in 10 m, and total precipitation in 24 hours are calculated for ERA5 pixels over the European domain. This allows us to perform monthly, seasonal, or yearly analysis in selected pixels or over aggregated areas. Generalized extreme value distribution can be fitted for the maxima to obtain the GEV parameters, which can be used to obtain return level estimates for given return periods. The analysis is done using Bayesian framework and MCMC sampling, which provides predictive uncertainty distributions for the quantities of interest. Similar analysis was performed to calculate maximum yearly water levels by Räty (2023) and this reference can be used as introduction to the statistical analysis and software needed. Figure 4 shows August temperature maxima that are expected every 20 years on average. The values are shown for four geographical locations. The density histograms show the uncertainty related to this estimate obtained from the MCMC simulation. This is important for risk management. For example, if one needs 95% certainty on the 20-year maximum August temperature at Trieste, one should be prepared for about 37 degrees Celsius. Using Bayesian analysis, the risk analysis accounts for uncertainty in the estimated extreme value distributions based on limited observational evidence and thus making it more realistic.

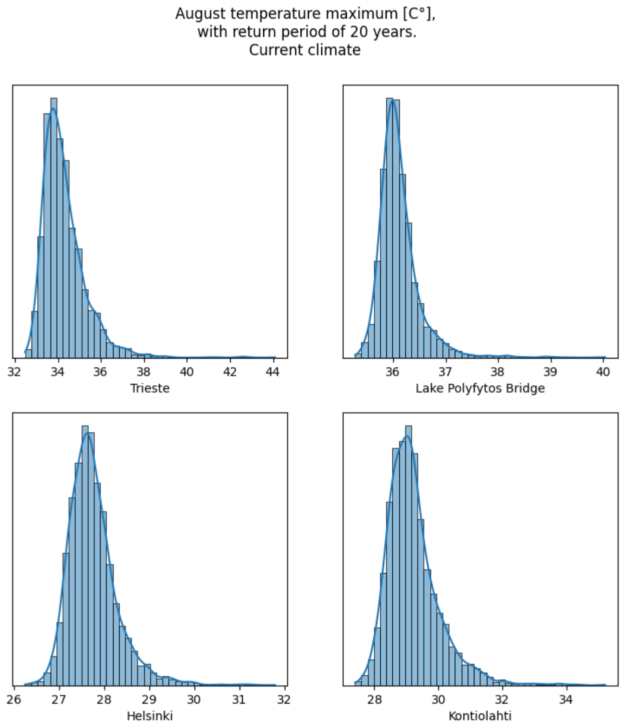


Figure 4: Temperature maxima from ERA5

2.5.2 Accounting for spatial and temporal correlation

The provided ERA5 examples perform the analysis for each spatial location separately. In addition, they assume that the climate does not change significantly during the period of the data used. In statistical terms, we assume stationarity in the distributional properties of the system. This is not always the case. The climate system is changing and analysis on historical observations does not necessarily provide accurate information on future extremes. On the other hand, data from one location has information on the behaviour at neighbouring locations. This can be used as advantage by pooling information from close by sites and to reduce uncertainties. An additional benefit in the case of in-situ observations is that pooling can provide information for locations that are close or between the observation locations. Here, we again refer to the study by Räty (2023) and references therein, where pooled information on water level gauge stations along Finnish coastlines is used. Figure 5 shows the estimated return level curves for yearly to thousand-year return periods. Without pooling the information, the uncertainty for rare events would have been much larger. The analysis in the paper was motivated by the needs of nuclear power plant safety.

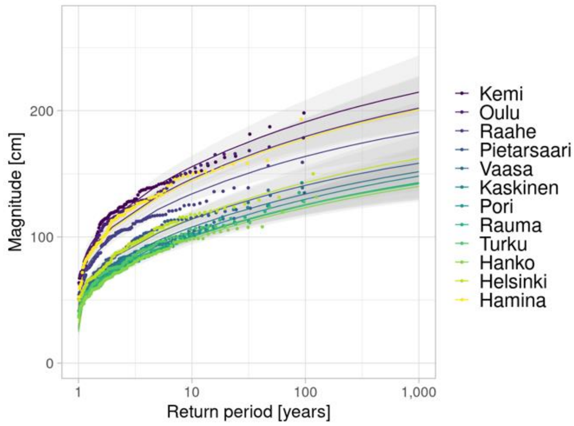


Figure 5: Analysis of extreme water level from Räty et al. (2023). Hierarchical statistical model is fitted to parameters of GEV distribution. The dots show observed maximum yearly water levels. Solid lines are modelled values from GEV model for each 12 stations.

Table 1: Extreme values analysis examples

Example	Variables	Data set used	Years	reference
Extreme value analysis on ERA5 data	temperature, dewpoint temperature, wind speed, total precipitation	ERA5	1980-2019	GitHub link
ERA5 Land examples	temperature and wind speed	ERA5 Land	1980-2019	GitHub link
GPEX data	precipitation	GPEX	1979-2017	GitHub link

3. Conclusions

This deliverable report provided an overview of extreme value analysis on climate data, focusing on methodologies and applications relevant to the RISKADAPT project. Several examples demonstrated the application of openly available climate datasets, showing the potential for analysing extreme events under present climate conditions. The results underscore the importance of advanced statistical techniques in analysing both observational and climate reanalysis data, particularly for supporting infrastructure risk assessments under extreme climatic conditions.

Despite the wide availability of datasets, limitations in spatial resolution remain a challenge for building-level studies within the RISKADAPT project. This necessitates sophisticated downscaling techniques, such as the high-resolution methods applied in the project's pilots. These methods can be used to bridge the gap between regional climate model outputs and asset-specific risk assessments.

For statistical extreme value analysis, the examples highlight the utility of Bayesian hierarchical modelling, which can help us quantify the probability of rare events and address uncertainties inherent in limited observational records, improving the robustness of risk assessments. For instance, by pooling data across spatial locations and leveraging temporal dependencies, these methods enhance predictive power, particularly for rare, high-impact events.

Furthermore, the analysis emphasizes the necessity of targeted data management and computational workflows to streamline extreme value analysis. The development and provision of pre-processed datasets, such as monthly maxima, significantly reduce the time and resources required for subsequent analyses. This highlights the importance of efficient data handling strategies in large-scale projects like RISKADAPT.

The outcomes of this deliverable set the foundation for future work under WP3. The next steps include expanding the scope to the future climate projections and examining the implications of CC on extreme events. This will involve regional climate models, such as Euro-CORDEX, and global CMIP6 datasets, to analyse projected changes in climatic extremes and their impact on critical infrastructure. Additionally, the integration of localized downscaling and risk modelling will further provide tools to support adaptation planning at the asset level.

References

AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (Eds.). (2012). *Extremes in a changing climate: Detection, analysis and uncertainty*. Springer Netherlands.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrol. Earth Syst. Sci., 23, 207–224, https://doi.org/10.5194/hess-23-207-2019, 2019.

Cintolesi et al. (2024), Model of high wind loading on a high-rise building, RISKADAPT project D3.5.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Verlag.

Gründemann, G. J., E. Zorzetto, H. E. Beck, M. Schleiss, N. van de Giesen, M. Marani, and R. J. van der Ent. (2023). Extreme precipitation return levels for multiple durations on a global scale. Journal of Hydrology, 621:129558. https://doi.org/10.1016/j.jhydrol.2023.129558

Hamdi, Y., Haigh, I. D., Parey, S., and Wahl, T.: Preface: Advances in extreme value analysis and application to natural hazards, Nat. Hazards Earth Syst. Sci., 21, 1461–1465, https://doi.org/10.5194/nhess-21-1461-2021, 2021.

Hellsten, A., Ketelsen, K., Sühring, M., Auvinen, M., Maronga, B., Knigge, C., Barmpas, F., Tsegas, G., Moussiopoulos, N., and Raasch, S.: A nested multi-scale system implemented in the large-eddy simulation model PALM model system 6.0, Geosci. Model Dev., 14, 3185–3214, https://doi.org/10.5194/gmd-14-3185-2021, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J-N. (2017): Complete ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S) Data Store (CDS).

Makkonen, L. (2000). Models for the growth of rime, glaze, icicles and wet snow on structures. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 358(1776):2913–2939, nov 2000. https://doi.org/10.1098/rsta.2000.0690.

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., Ketelsen, K., Letzel, M. O., Sühring, M., and Raasch, S. (2015): The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives, Geosci. Model Dev., 8, 1539-1637, DOI:10.5194/gmd-8-2515-2015

Pirinen, P., Lehtonen, I., Heikkinen, R. K., Aapala, K., Aalto, J., (2022). Daily gridded evapotranspiration data for Finland for 1981–2020, Clim. Bull. Res. Lett. 4. https://doi.org/10.35614/ISSN-2341-6408-IK-2022-11-R

Räty, O., M. Laine, U. Leijala, J. Särkkä, and M. M. Johansson (2023). Bayesian hierarchical modelling of sea-level extremes in the Finnish coastal region. Natural Hazards and Earth System Sciences, 23(7):2403–2418, 2023. https://doi.org/10.5194/nhess-23-2403-2023.

Annexes

The following pages contain three examples of obtaining and processing reanalysis data and their preparation for extreme value analysis. It also includes Bayesian extreme value analysis for some variables. The examples are provided as Python language notebooks that contain the computer code used to perform the analysis. The source code is available from GitHub repository https://github.com/fmidev/RISKADAPT/. The examples are listed in Table 1 and described shortly below. The repository contains extreme value analysis code that can be used to further calculate return period of quantities of interest.

Annex 1. Extreme value analysis of ERA5 data

Here we demonstrate extreme values analysis for ERA5 variables. The example Python notebook provides a set of functions to fit a generalized extreme value distribution to a time series data and calculate return levels for given return periods. Parameter estimation and uncertainty quantification is done using Markov chain Monte Carlo sampling. RISKADAPT Pilot locations are used in the examples.

```
import os
import warnings
import calendar
import numpy as np
import xarray as xr

import matplotlib.pyplot as plt
import seaborn as sns

warnings.simplefilter(action='ignore', category=FutureWarning)

DATA = os.path.expanduser('~/DATA/RISKADAPT/')
```

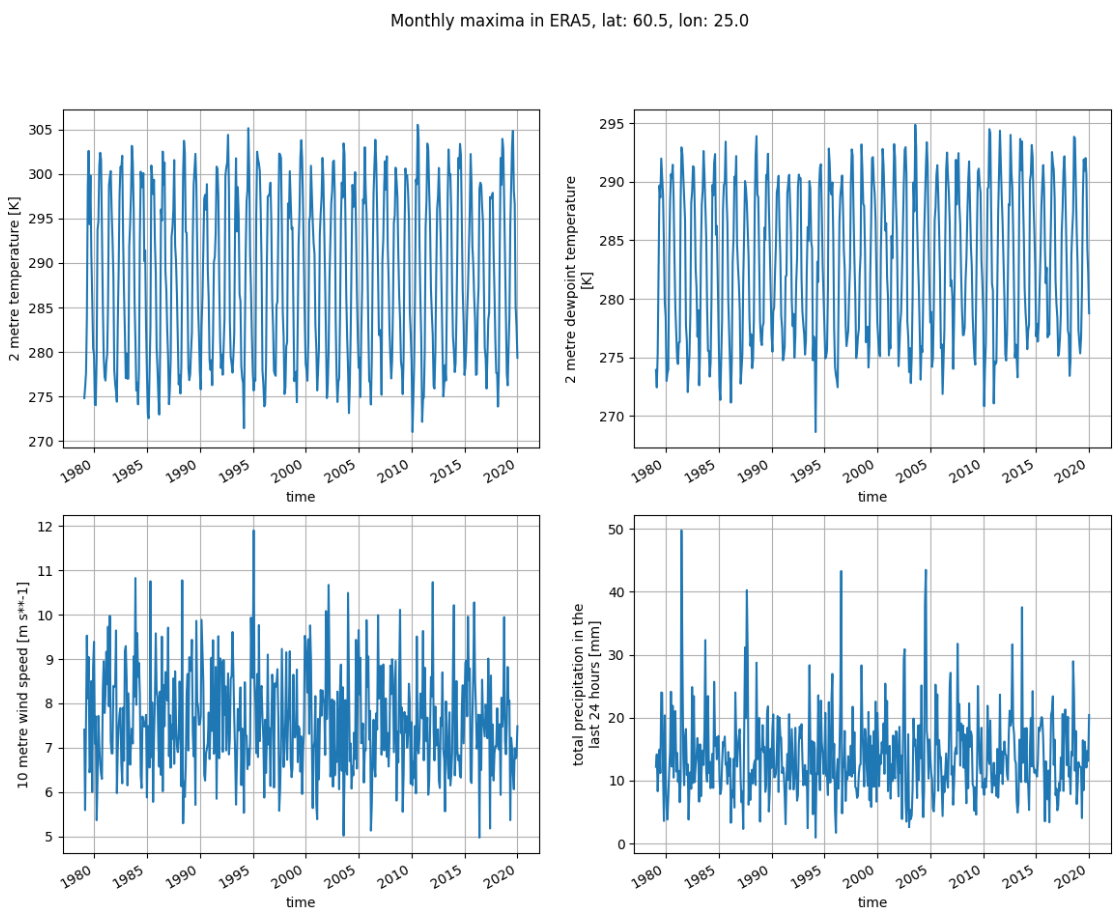
Here, we assume that we have precalculated monthly maxima of the ERA5 variables of interest.

```
era5 = xr.open_dataset(DATA+'ERA5_Europe_monthly_max.nc')
era5

<xarray.Dataset>
Dimensions:    (time: 492, latitude: 187, longitude: 261)
Coordinates:
  * latitude   (latitude) float64 72.0 71.75 71.5 71.25 ... 26.0 25.75 25.
5
  * longitude  (longitude) float64 -25.0 -24.75 -24.5 -24.25 ... 39.5 39.7
5 40.0
  * time       (time) datetime64[ns] 1979-01-31 1979-02-28 ... 2019-12-31
Data variables:
    t2m        (time, latitude, longitude) float32 ...
    d2m        (time, latitude, longitude) float32 ...
    ws10       (time, latitude, longitude) float32 ...
    tp24       (time, latitude, longitude) float32 ...
```

Plot the data at one selected lon lat point.

```
lon, lat = 25.0, 60.5
```

```
fig, axs = plt.subplots(2, 2, figsize=(14, 10))
era5['t2m'].interp(latitude=lat, longitude=lon).plot(ax=axs[0,0])
era5['d2m'].interp(latitude=lat, longitude=lon).plot(ax=axs[0,1])
era5['ws10'].interp(latitude=lat, longitude=lon).plot(ax=axs[1,0])
era5['tp24'].interp(latitude=lat, longitude=lon).plot(ax=axs[1,1])
for ax in axs.ravel():
    ax.grid()
    ax.set_title('')
plt.suptitle(f'Monthly maxima in ERA5, lat: {lat}, lon: {lon}')
plt.show()
```

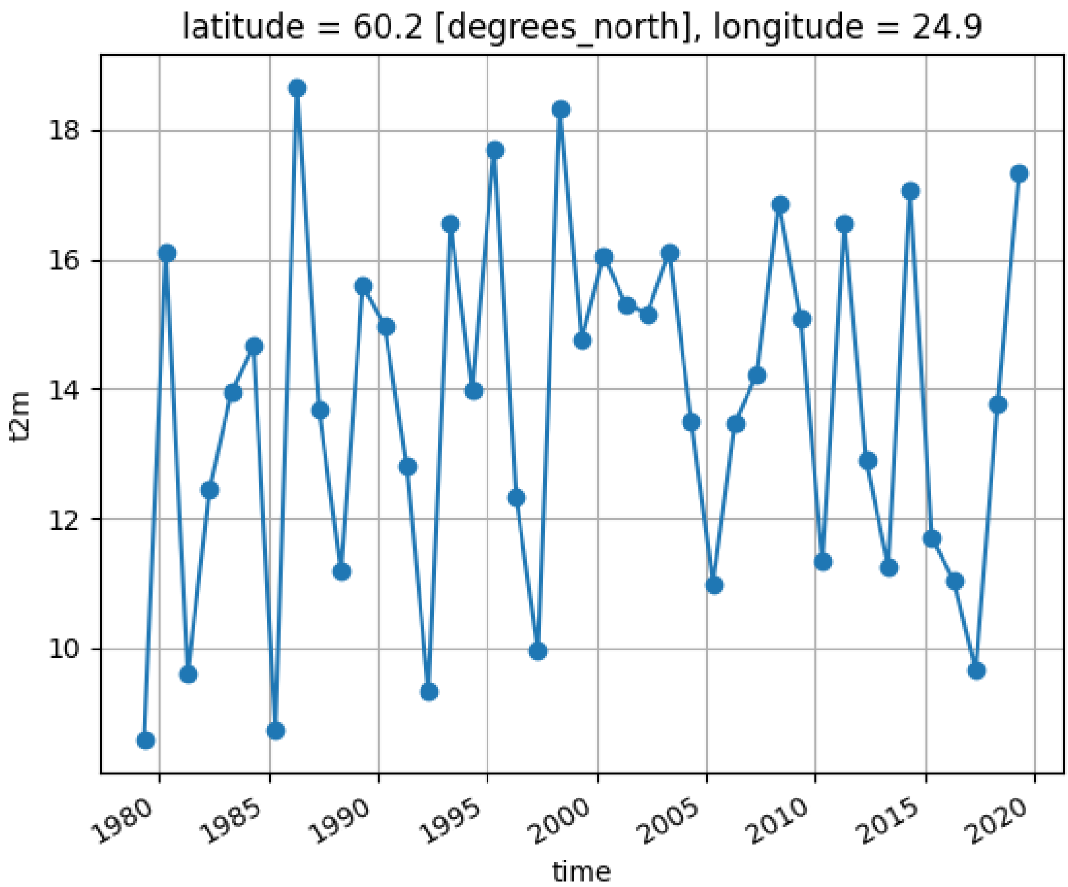


GEV analysis

Here, we provide a set of Python functions that fit a generalized extreme value distribution to a time series data and calculate return levels for given return periods. It uses Stan probabilistic language to define the statistical model and calculate posterior probabilities for GEV parameters as well as for quantiles of the fitted distribution. These can be used to calculate return levels and their uncertainties. The estimation is based on MCMC sampling.

```
import cmdstan_gev as stangev

# select one location and one month
lat = 60.2
lon = 24.9
month = 4
```

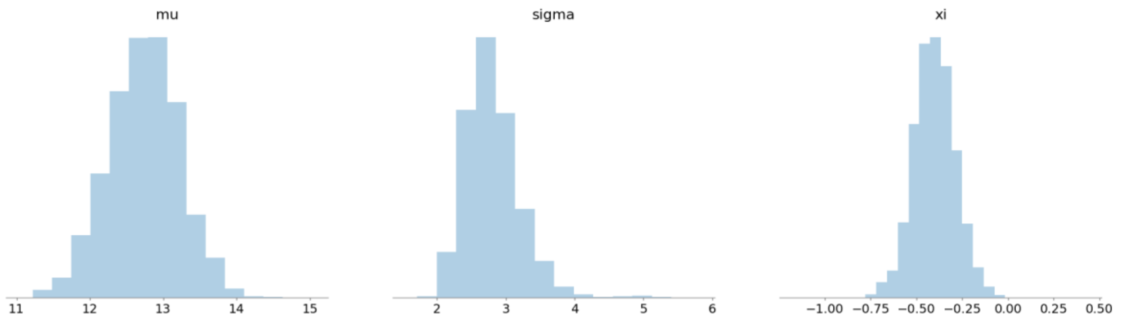
```
months = era5['time.month']
y = era5['t2m'].where(months==month, drop=True).interp(latitude=lat, longi
tude=lon) - 273.15

y.plot(marker='o')
plt.grid()
plt.show()
```




```
fit = stangev.gev_fit(y, hyper={'xi0': -0.3, 'xisig0': 0.2, 'sig0': 3.5},
                      adapt_delta=0.85, show_progress=False)

13:35:33 - cmdstanpy - INFO - CmdStan start processing
13:35:33 - cmdstanpy - INFO - CmdStan start processing
13:35:33 - cmdstanpy - INFO - CmdStan done processing
13:35:33 - cmdstanpy - WARNING - Some chains may have failed to converge.
    Chain 1 had 157 divergent transitions (15.7%)
    Chain 2 had 148 divergent transitions (14.8%)
    Chain 3 had 208 divergent transitions (20.8%)
    Chain 4 had 154 divergent transitions (15.4%)
    Use function "diagnose()" to see further information.
```

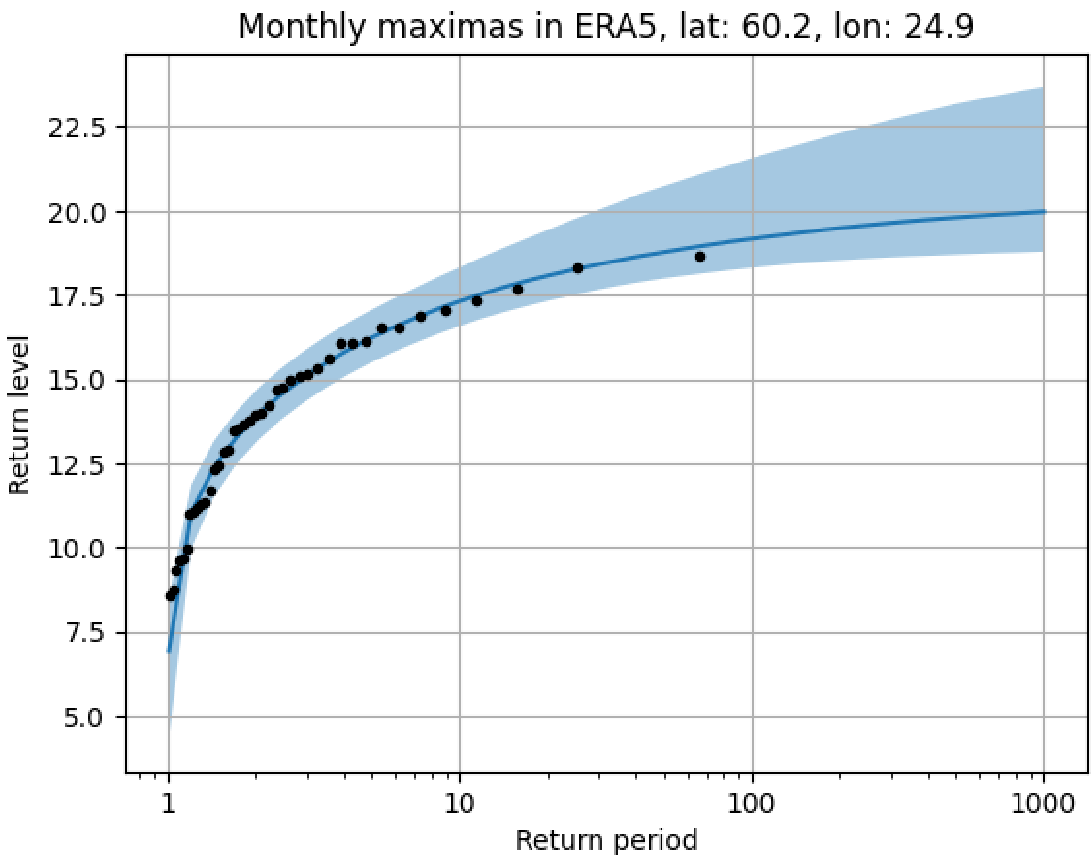
Plot chain histograms of estimated GEV parameters.

```
stangev.gev_posterior(fit)
plt.show()
```



Return level plot with return periods in logarithmic scale in the x axis. The blue shaded area gives 95% predictive uncertainty envelope for the return levels.

```
stangev.gev_qplot(y, fit, maxp=1000)
plt.title(f'Monthly maximas in ERA5, lat: {lat}, lon: {lon}')
plt.show()
```

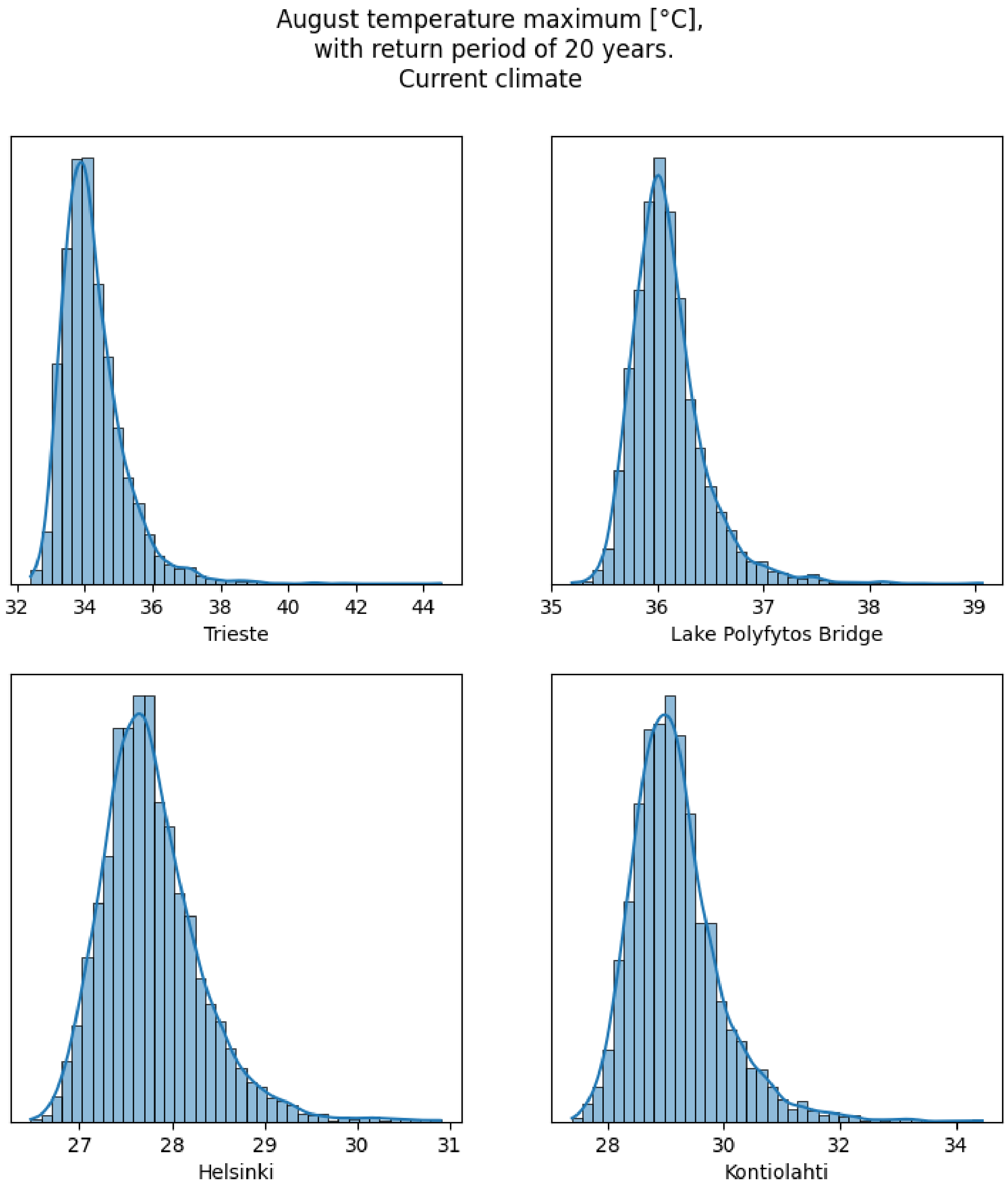


Analysis usingselected coordinates

We can calculate selected return periods for any pixel or collection of pixels in the ERA5 data. Below, we select some locations that are studied in the RISKADAPT project. The same analyses will be extended to the future climate in the next phase.

```
# lon, lat
locations = {
    'Trieste': [13.8, 45.63],
    'Lake Polyfytos Bridge': [21.973889, 40.232778],
    'Helsinki': [24.9375, 60.170833],
```

```
    'Kontiolahti': [30.0, 62.8],
}

return_periods = [20]
month = 8

fig, axs = plt.subplots(2,2, figsize=(9, 9))

for i, l in enumerate(locations):
    ax = axs.ravel()[i]
    lon, lat = locations[l]
    y = era5['t2m'].interp(latitude=lat, longitude=lon).where(era5['time.m
onth']==month, drop=True) - 273.15
    fit = stangev.gev_fit(y, quiet=True, show_progress=False)
    rl_chain = stangev.gev_qpred(fit, yt = return_periods, return_chain=Tr
ue)
    sns.histplot(rl_chain.values[:,0], bins=40, kde=True, stat='probabilit
y', ax=ax)
    ax.set_xlabel(f'{l}')
    ax.set_ylabel('')
    ax.set_yticks([])
plt.suptitle((f'{calendar.month_name[month]} temperature maximum [°C],'
              f'\n with return period of {return_periods[0]} years.'
              '\nCurrent climate'))
#axs[1,1].set_axis_off()
plt.show()
```



The plots above show probability distribution of the estimated return levels of monthly maximum temperature in selected locations based on ERA5 monthly maxima. More specifically, the figures show 20-year return period, i.e., the estimated probability density for the maximum temperature to be observed at most every 20 years assuming that the climate stays the same as it has been during the years 1980-2019.

Annex 2. ERA5 Land examples

These examples demonstrate the use of ERA5 Land data from the Copernicus service. It uses two locations related to RISKADAPT Pilots: Lake Polyfytos Bridge in Greece and Cattinara Hospital in Italy. The demonstration studies one month (August) from year 1980 to 2019 and calculates monthly maxima from the provided hourly data. It shows how total precipitation in a form of hourly data can be accumulated into 24-hour total precipitation, which is more usable information for hydrological studies. For wind speed, the u and v components of wind are interpolated to a given location and transformed to wind speed.

We provide a script *era5land_download.py* to download the data for one location and month over years 1980-2019.

Below, we use two locations related to RISKADAPT Pilots, i.e., Lake Polyfytos Bridge in Greece and Cattinara Hospital in Italy.

```
import os
import numpy as np
import xarray as xr

import matplotlib.pyplot as plt

from glob import glob

DATA = os.path.expanduser('~/DATA/RISKADAPT/')

locations = {
    'Lake Polyfytos Bridge': [21.973889, 40.232778],
    'Cattinara Hospital Trieste': [13.826012, 45.634376],
}
```

Extreme precipitation at Lake Polyfytos

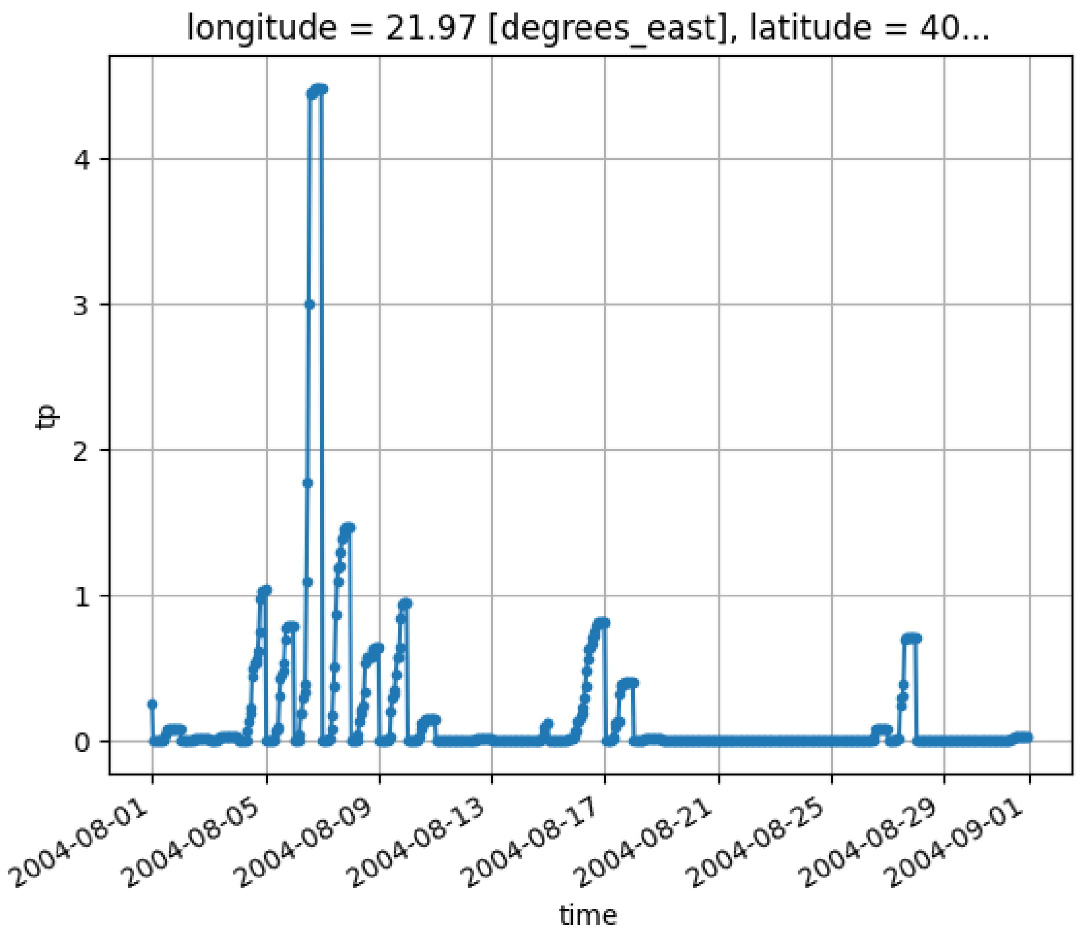
First, we study extreme precipitation near Lake Polyfytos Bridge that is the subject of study in RISKADAPT Pilot 1. Data for August precipitation for years 1980-2019 have been downloaded with the utility *era5land_download.py*.

```
file = DATA+'era5_land_1980-2019_August_tp_Polyfytos.nc'
ds = xr.open_dataset(file)
ds

<xarray.Dataset>
Dimensions:     (time: 1280, step: 24, latitude: 2, longitude: 2)
Coordinates:
    number      int64 ...
  * time        (time) datetime64[ns] 1980-07-31 1980-08-01 ... 2019-08-31
  * step        (step) timedelta64[ns] 01:00:00 02:00:00 ... 1 days 00:00:
00
    surface     float64 ...
  * latitude    (latitude) float64 40.3 40.2
  * longitude   (longitude) float64 21.9 22.0
    valid_time  (time, step) datetime64[ns] ...
Data variables:
    tp          (time, step, latitude, longitude) float32 ...
```

```
Attributes:
    GRIB_edition:            1
    GRIB_centre:             ecmf
    GRIB_centreDescription:  European Centre for Medium-Range Weather Fore
casts
    GRIB_subCentre:          0
    Conventions:             CF-1.7
    institution:             European Centre for Medium-Range Weather Fore
casts
    history:                 2023-10-08T09:01 GRIB to CDM+CF via cfgrib-0.
9.1...
```

The variable *tp* represents total precipitation cumulated hourly, separately for each day. The unit is meters. We interpolate the values to a given location and adjust the coordinates for easier processing.

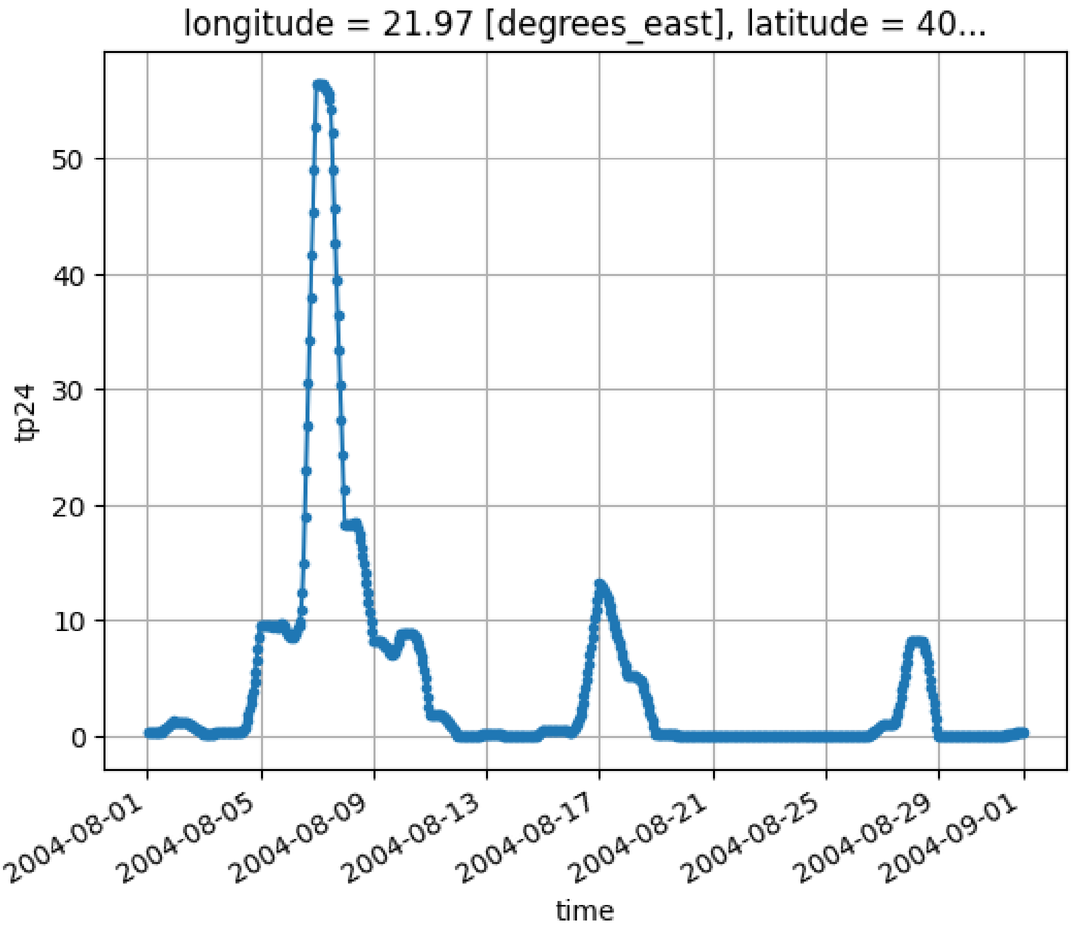
```
lon, lat = locations['Lake Polyfytos Bridge']
tp = ds['tp'].interp(longitude=lon, latitude=lat)

tp = (tp.stack(z=('time', 'step')).
       swap_dims({'z': 'valid_time'}).
       drop(['z', 'number', 'surface', 'time', 'step']).
       rename({'valid_time': 'time'}))
tp

<xarray.DataArray 'tp' (time: 30720)>
array([        nan,         nan,         nan, ..., 1.29357e-05,
       1.29357e-05,         nan])
Coordinates:
  * time       (time) datetime64[ns] 1980-07-31T01:00:00 ... 2019-09-01
    longitude  float64 21.97
    latitude   float64 40.23
Attributes: (12/30)
    GRIB_paramId:                        228
    GRIB_dataType:                       fc
    GRIB_numberOfPoints:                 4
    GRIB_typeOfLevel:                    surface
    GRIB_stepUnits:                      1
    GRIB_stepType:                       accum
    ...                                  ...
    GRIB_shortName:                      tp
    GRIB_totalNumber:                    0
    GRIB_units:                          m
    long_name:                           Total precipitation
    units:                               m
    standard_name:                       unknown

(tp.where(tp['time.year']==2004)*1000).plot(marker='.')
plt.grid()
plt.show()
```



Sum the daily values for the selected month for each year and then take 24-hour differences to get *tp24*, which is the total precipitation for each previous 24 hours. Transform m to mm.

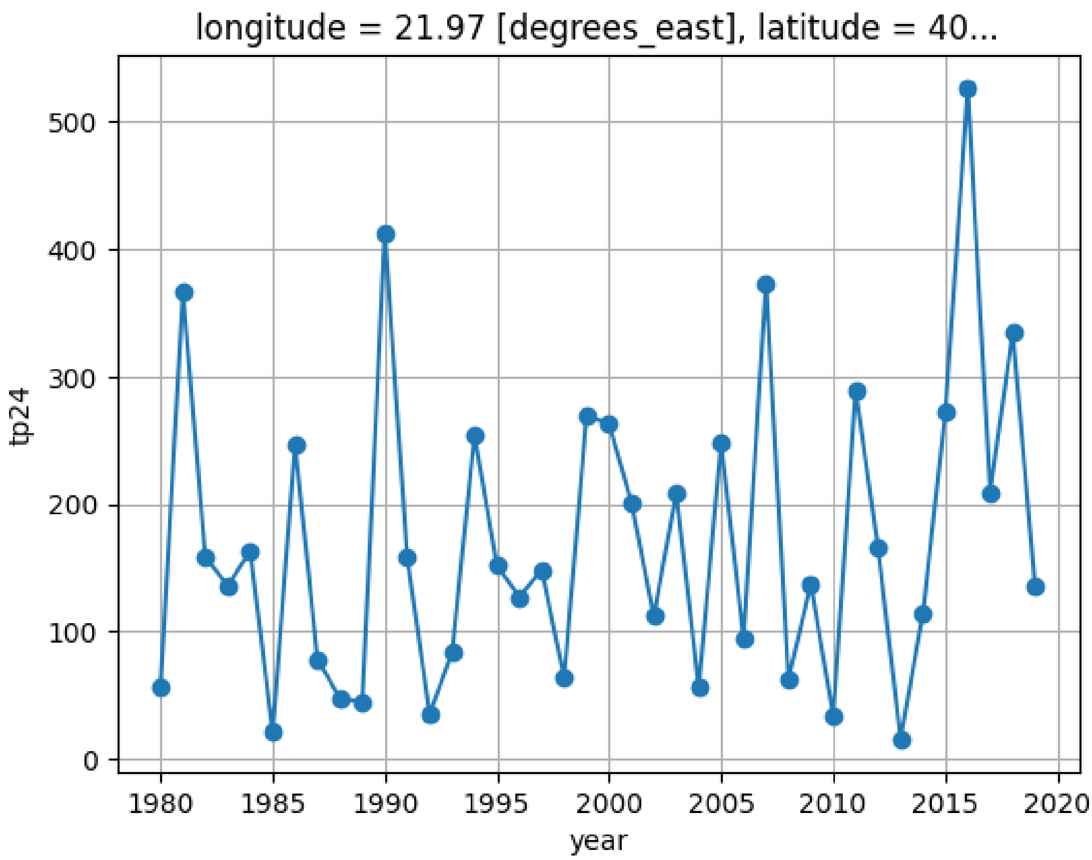
```
tp24 = tp.groupby(tp['time.year']).cumsum() * 1000
tp24.name = 'tp24'

def tp24fun(ds):
       time1 = ds.time[ds.time >= ds.time[0] + np.timedelta64(24, 'h')]
       time2 = time1 - np.timedelta64(24, 'h')
       out = xr.full_like(ds, fill_value=np.nan)
       out.loc[dict(time=time1)] = (ds.sel(time=time1).values -
                                     ds.sel(time=time2).values)
       return out

tp24 = tp24.groupby(tp24['time.year']).map(tp24fun)
tp24

<xarray.DataArray 'tp24' (time: 30720)>
array([     nan,      nan,      nan, ..., 0.17789, 0.18867, 0.18651])
Coordinates:
  * time       (time) datetime64[ns] 1980-07-31T01:00:00 ... 2019-09-01
    longitude  float64 21.97
    latitude   float64 40.23
```

```
tp24.where(tp24['time.year']==2004).plot(marker='.')
plt.grid()
plt.plot()

[]
```



And then we can calculate the maximum 24-hour precipitation for each year and for the selected month.

```
tp24.groupby('time.year').max().plot(marker='o')
plt.grid()
plt.show()
```



Now, we can proceed to extreme value analysis.

Maximum wind speed near Cattinara Hospital

The next example calculates maximum 10-meter wind speed interpolated to the location of Cattinara hospital in Trieste, Italy. We study the August maximum 10-meter wind speed. ERA5 stores the wind as *u* and *v* components. First, we need to calculate 10-meter wind speed from those and do similar coordinate manipulations as for the *tp* above.

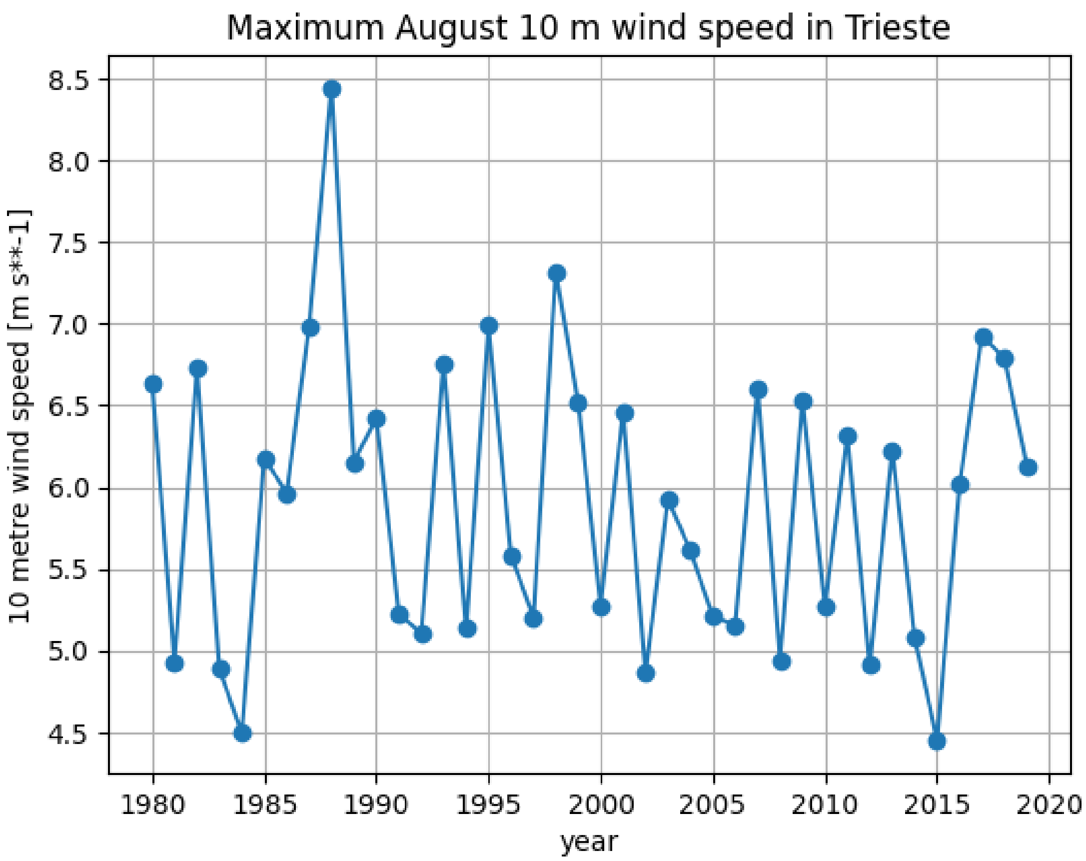
```
file = DATA+'era5_land_1980-2019_August_uv10_Trieste.nc'
ds = xr.open_dataset(file)
ds

<xarray.Dataset>
Dimensions:     (time: 1280, step: 24, latitude: 2, longitude: 2)
Coordinates:
    number      int64 ...
  * time        (time) datetime64[ns] 1980-07-31 1980-08-01 ... 2019-08-31
  * step        (step) timedelta64[ns] 01:00:00 02:00:00 ... 1 days 00:00:
00
    surface     float64 ...
  * latitude    (latitude) float64 45.7 45.6
  * longitude   (longitude) float64 13.8 13.9
    valid_time  (time, step) datetime64[ns] ...
Data variables:
    u10         (time, step, latitude, longitude) float32 ...
    v10         (time, step, latitude, longitude) float32 ...
```

```
Attributes:
    GRIB_edition:             1
    GRIB_centre:              ecmf
    GRIB_centreDescription:   European Centre for Medium-Range Weather Fore
casts
    GRIB_subCentre:           0
    Conventions:              CF-1.7
    institution:              European Centre for Medium-Range Weather Fore
casts
    history:                  2023-10-09T14:26 GRIB to CDM+CF via cfgrib-0.
9.1...
```

```
lon, lat = locations['Cattinara Hospital Trieste']
ds['ws10'] = np.sqrt(np.square(ds['u10']) + np.square(ds['v10']))
ds['ws10'].attrs = ds['u10'].attrs.copy()
ds['ws10'].attrs['long_name'] = '10 metre wind speed'
ws10 = ds['ws10'].interp(longitude=lon, latitude=lat).astype(np.float32)

ws10 = (ws10.stack(z=('time', 'step')).
       swap_dims({'z': 'valid_time'}).
       drop(['z', 'number', 'surface', 'time', 'step']).
       rename({'valid_time': 'time'}))
ws10
```

```
<xarray.DataArray 'ws10' (time: 30720)>
array([     nan,      nan,      nan, ..., 1.88802, 1.86969,      nan],
      dtype=float32)
Coordinates:
  * time       (time) datetime64[ns] 1980-07-31T01:00:00 ... 2019-09-01
    longitude  float64 13.83
    latitude   float64 45.63
Attributes: (12/30)
    GRIB_paramId:                             165
    GRIB_dataType:                            fc
    GRIB_numberOfPoints:                      4
    GRIB_typeOfLevel:                         surface
    GRIB_stepUnits:                           1
    GRIB_stepType:                            instant
    ...                                       ...
    GRIB_shortName:                           10u
    GRIB_totalNumber:                         0
    GRIB_units:                               m s**-1
    long_name:                                10 metre wind speed
    units:                                    m s**-1
    standard_name:                            unknown
```

```
ws10.groupby('time.year').max().plot(marker='o')
plt.title('Maximum August 10 m wind speed in Trieste')
plt.grid()
plt.show()
```


Maximum August 10 m wind speed in Trieste
10 metre wind speed [m s**-1]
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
1980
1985
1990
1995
2000
2005
2010
2015
2020
year

Annex 3. GPEX data

This example uses Global Precipitation Extremes dataset GPEX. The data are described in detail in: Extreme Precipitation Return Levels for Multiple Durations on a Global Scale, Gründemann et al., 2021. 10.1016/j.jhydrol.2023.129558. We use the data over Haliacmon river drainage basin and extract precipitation return levels for several event durations and return periods based on GEV analysis. In addition, Intensity-Duration-Frequency plots are formed over the study area.

The data set GPEX.nc can be downloaded from https://opendap.4tu.nl/thredds/catalog/data2/fig/12764429/4/catalog.html. The data file contains global estimates of extreme precipitation using four extreme value methods (GEV, POT, Gumbel and MEV) for eight durations (3 hours - 10 days).

See page https://data.4tu.nl/articles/_/12764429/4 for more details.

```
import numpy as np
import xarray as xr

import matplotlib.pyplot as plt

import geopandas as gpd
import regionmask
import cartopy.crs as ccrs
from cartopy.io.img_tiles import OSM

imagery = OSM(cache=True)

/usr/local/share/venvs/sci/lib/python3.10/site-packages/cartopy/io/img_til
es.py:113: UserWarning: Cartopy created the following directory to cache G
oogleWTS tiles: /var/folders/8l/qwchy9rd09x6v46zs3_zbmxm0000gn/T/cartopy_c
ache_dir/OSM
  warnings.warn(
```

Open downloaded data set. The spatial resolution is 0.1°.

```
ds = xr.open_dataset('~/DATA/GPEX/GPEX.nc')
ds

<xarray.Dataset>
Dimensions:             (lat: 1480, lon: 3600, dur: 8, tr: 10, year: 38)
Coordinates:
  * lat                 (lat) float32 89.95 89.85 89.75 ... -57.75 -57.85
-57.95
  * lon                 (lon) float64 -179.9 -179.9 -179.8 ... 179.8 179.9
179.9
  * dur                 (dur) int32 3 6 12 24 48 72 120 240
  * tr                  (tr) int32 2 5 10 20 39 50 100 200 500 1000
  * year                (year) int32 0 1 2 3 4 5 6 7 ... 30 31 32 33 34 35
36 37
Data variables: (12/26)
    gev_estimate        (lat, lon, dur, tr) float32 ...
    pot_estimate        (lat, lon, dur, tr) float32 ...
    mev_estimate        (lat, lon, dur, tr) float32 ...
    gumbel_estimate     (lat, lon, dur, tr) float32 ...
    observed_estimate   (lat, lon, dur, tr) float32 ...
```

```
    gev_location         (lat, lon, dur) float32 ...
    ...                   ...
    gumbel_location      (lat, lon, dur) float32 ...
    gumbel_scale         (lat, lon, dur) float32 ...
    annual_maximum       (lat, lon, dur, year) float32 ...
    hydroyear            (lat, lon) float32 ...
    running_parameter    (lat, lon) timedelta64[ns] ...
    mask                 (lat, lon) uint8 ...
Attributes:
    title:             GPEX
    description:       Global Precipitation EXtremes dataset. This data was
use...
    acknowledgment:    Contains modified Multi-Source Weighted-Ensemble Prec
ipi...
    authors:           Gaby Gründemann, Enrico Zorzetto, Hylke Beck, Marc Sc
hle...
    date_created:      2021-08-25
    creator_name:      Gaby Gründemann
    creator_email:     g.j.gruendemann@tudelft.nl
    institution:       Department of Water Management, Faculty of Civil Engi
nee...
    conventions:       CF 1.7
```

For illustration, we use shape file for Haliacmon river drainage basin and select a rectangular area around the basin.

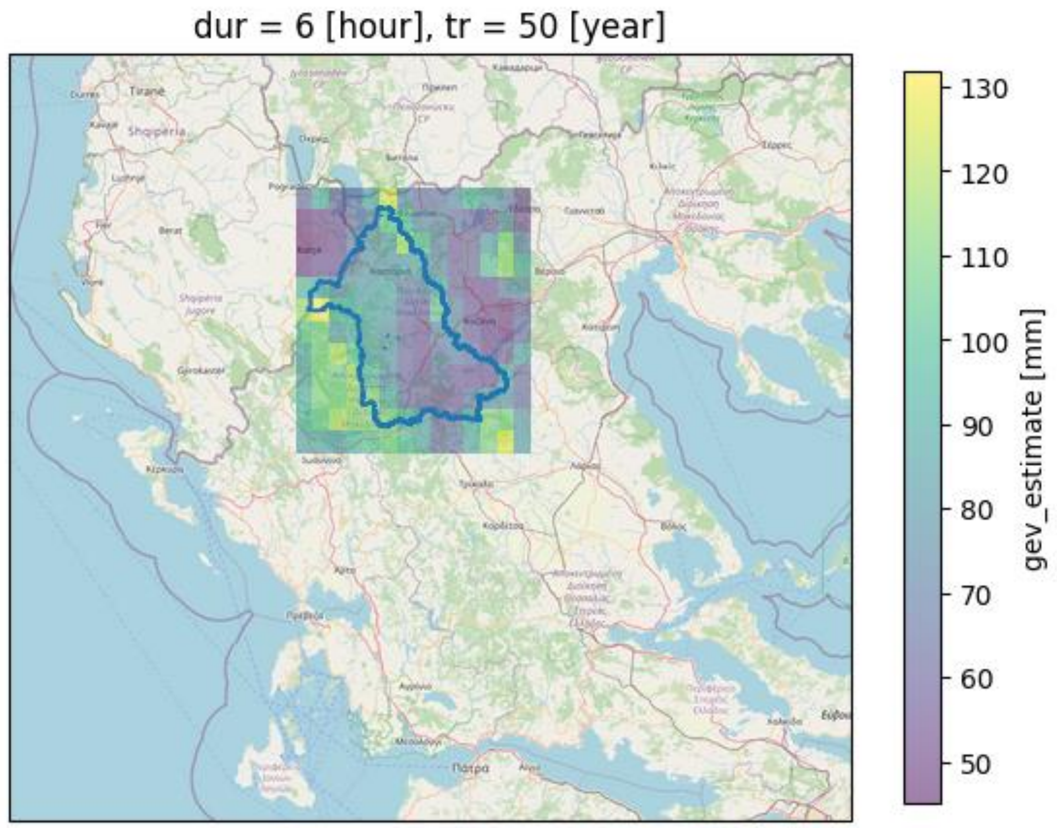
```
shape = gpd.read_file('Drainage_basin/basin.shp')
bb = shape.to_crs('EPSG:4326').bounds.values[0]

bb = bb + [-0.1, -0.1, 0.1, 0.1]
ds2 = ds.sel(lon=slice(bb[0], bb[2]), lat=slice(bb[3], bb[1]))

shape.to_crs('EPSG:4326').bounds
```

	minx	miny	maxx	maxy
0	20.776522	39.818271	21.959666	40.809015

Plot with a background map.

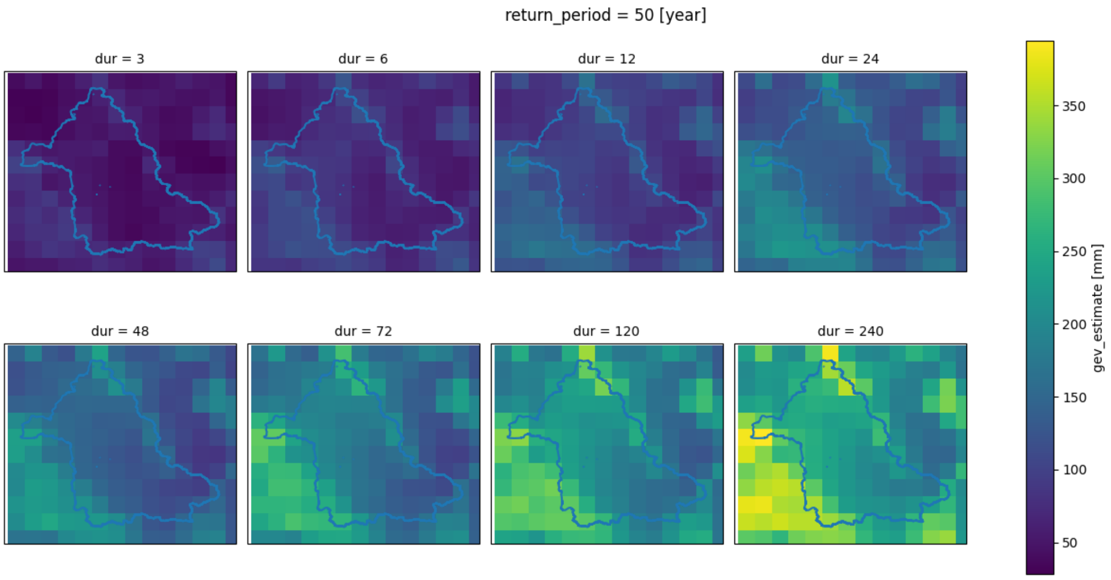
```
tr = 50
dur = 6

p0 = imagery.crs
p1 = ccrs.PlateCarree()

fig, ax = plt.subplots(1, 1, figsize=(7, 7),  subplot_kw={'projection': p0
})
ax.set_extent(np.r_[19.0, 24.0, 38, 41.5], crs=p1)

ax.add_image(imagery, 8, interpolation='spline36')
shape.to_crs(p1).boundary.plot(ax=ax, transform=p1)

gev = ds2['gev_estimate'].sel(tr=tr, dur=dur)
```

```
m = gev.plot.pcolormesh(x='lon', y='lat',
                        transform=p1, ax=ax, alpha=0.5,
                        cbar_kwargs={'shrink': 0.7})
plt.show()
```



Plot one return period for several event durations (dur [h]).

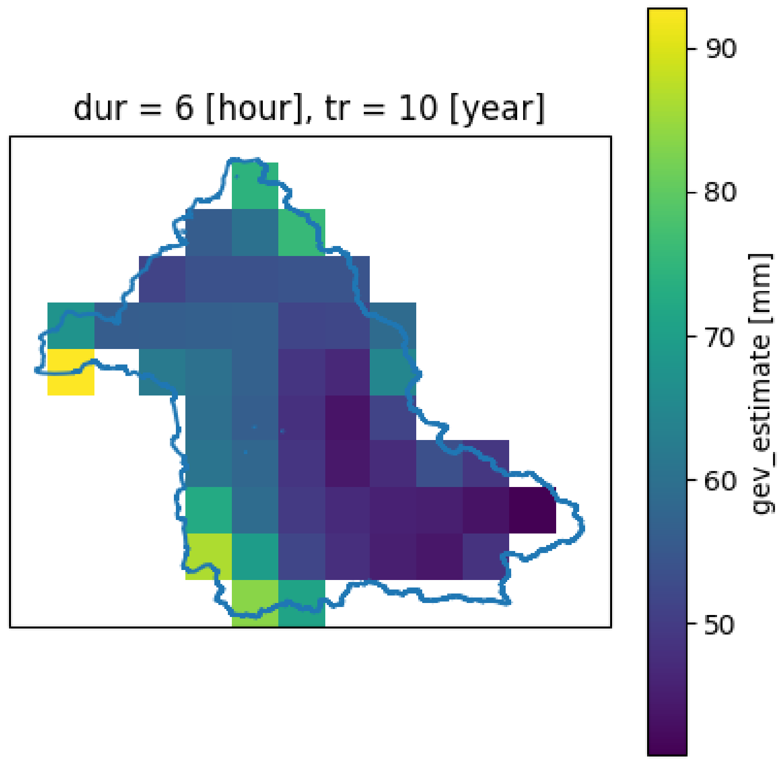
```
tr = 50

gev = ds2['gev_estimate'].sel(tr=tr)
m = gev.plot.pcolormesh(x='lon', y='lat', col='dur', col_wrap=4,
                        transform=p1,
                        subplot_kws={'projection': p1})
plt.suptitle(f'{gev.tr.attrs.get("long_name")} = {gev.tr.values} [{gev.tr.
attrs.get("units")}]', y=1.02)

for ax in m.axs.ravel():
    ax.set_extent(np.r_[bb[0], bb[2], bb[1], bb[3]], crs=p1)
    shape.to_crs(p1).boundary.plot(ax=ax, transform=p1)

plt.show()
```



IDF intensity - duration - frequency curves

IDF curves can be used for various hydrological analyses. In GPEX data, intensity is calculated from precipitation and duration. The data has 10 return periods from 2 to 1000 years. Below, we mask the data to select only pixels for the Haliacmon river basin. Then we calculate average IDF curves over the areas for all provided return periods.

```
mask = regionmask.mask_geopandas(shape.to_crs('EPSG:4326'),
                                 ds['lat'].values,
                                 ds['lat'].values,
                                 lon_name='lon', lat_name='lat')

fig, ax = plt.subplots(1, 1, figsize=(5, 5),  subplot_kw={'projection': p1
})
ds['gev_estimate'].sel(tr=10, dur=6).where(mask==0, drop=True).plot(transf
orm=p1)
shape.to_crs(p1).boundary.plot(ax=ax, transform=p1)
plt.show()
```




```
plt.figure(figsize=(10, 5))

gevmean = ds['gev_estimate'].where(mask==0, drop=True).mean(dim=['lon', 'l
at'])
(gevmean / gevmean['dur']).plot.line(x='dur')
plt.ylabel('intensity [mm / h]')
plt.grid()
plt.title('IDF plot for Haliacmon river drainage basin')
plt.show()
```