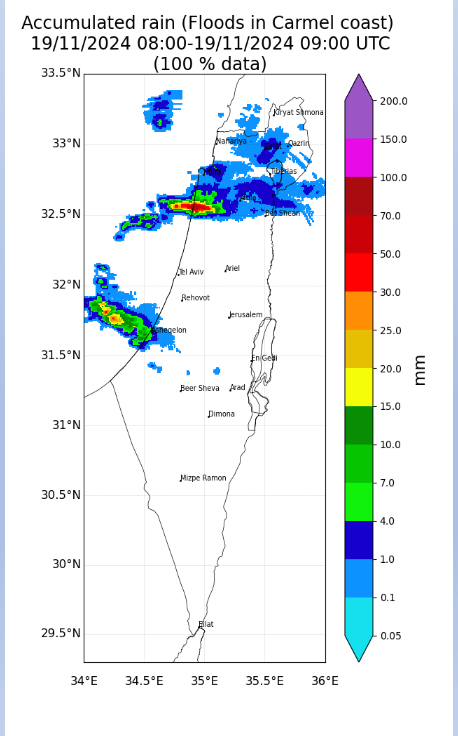
Online Workshop on Global Tools for Flood and Drought Prediction

Extreme precipitation and rain intensities: observed and projected trends - Main key points

Assaf Zipori & Yizhak Yosef



Background

What is rain intensity?

The **amount (height)** of rain (mm, cm, incs…) in a given **duration** (1 minute, 30 minutes, 1 hour, 6 hours…)

Usually it's expressed in mm/h, i.e. if we measure 10 mm of rain in 10 minutes, the rain intensity is 60 mm/h

Why study extreme precipitation?

- While most rainfall events are beneficial to society - supporting ecosystems, water resources, and agriculture, extreme precipitation events carry a significant potential for damage.
- Extreme rainfall events are among the costliest and most disruptive weather hazards globally.
- Impacts: urban & riverine flooding, infrastructure damage, landslides, transport and water management disruptions, agriculture, landslide and flash floods (in rivers and in urban areas)
- Small increases in short-duration intensity can cause non-linear increases in flood risk and failures of drainage systems.
- Relevance: design standards, early-warning systems, hydrological modelling, climate adaptation planning.

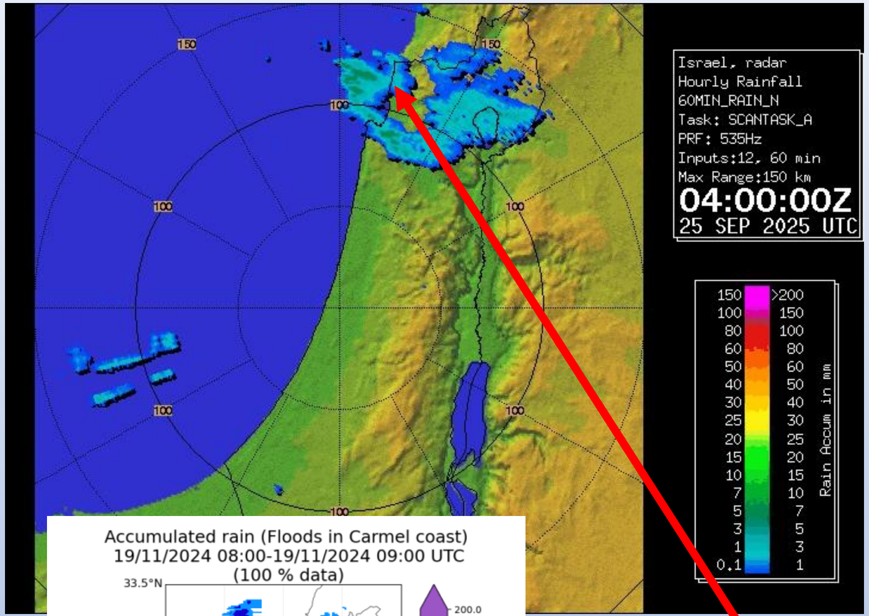
OPB, *Alberto Saiz; Valencia, Oct 2024*

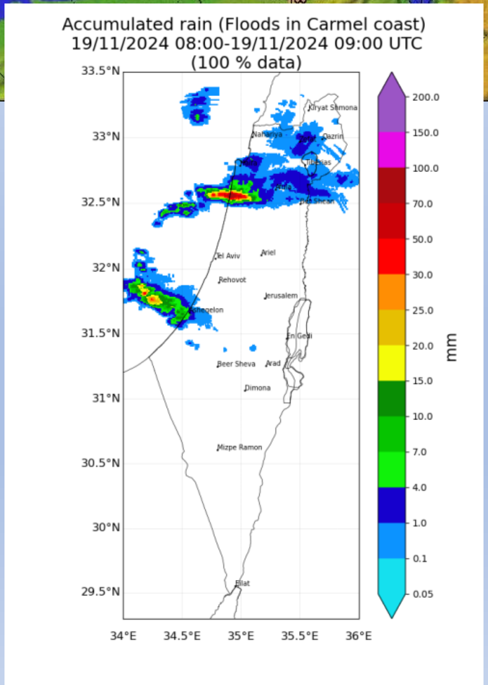
PBS News, Marwan Alfaituri; *Libeya, Sep 2023*

Background

Characteristics of sub-daily extreme rainfall events

- **High spatial variability / random effects**: a single station may measure extreme values while surrounding stations remain near typical levels intensity.
- **Dynamics:** convective organization, storm propagation, mesoscale features (e.g., training, coastal convergence).
- **Local modifiers**: topography, land-use, sea-breeze/coastal effects that are strong in the Eastern Mediterranean.
- **Thermodynamics:** Clausius–Clapeyron scaling affecting potential intensities (~7% per degree).





1h: ~75 mm
3h: ~110 mm
SEP. Record!!!

Before starting – examine your data

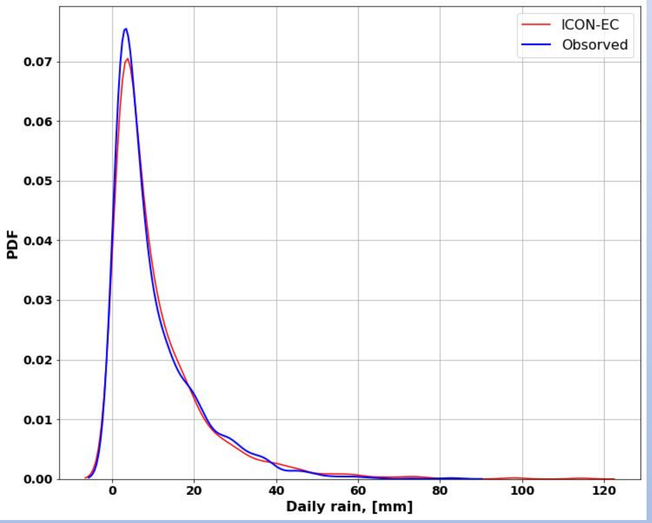
- Statistical results are only as reliable as the data they are based on
  → *Garbage in, garbage out*

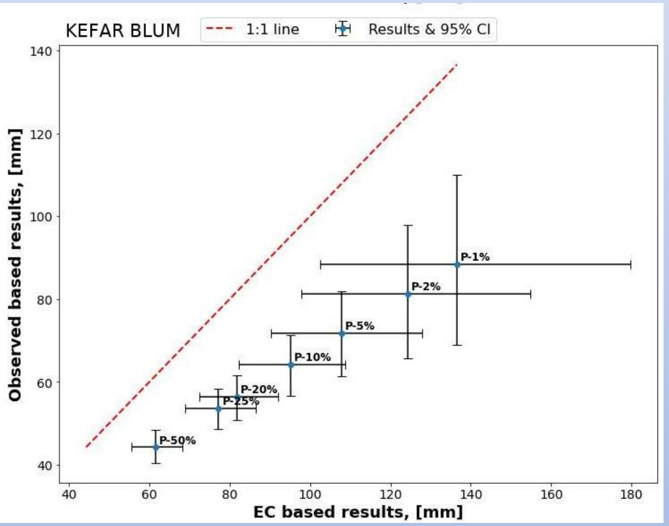
- **Observed (Measured) Data:**
  - **Data integrity:** completeness, false values, outliers, internal inconsistencies
  - **Missing data:** can gaps be reconstructed or reliably imputed?
  - **Instrumentation:** maintenance, calibration, sensor drift, firmware issues
  - **Metadata quality:** station changes, relocation, equipment upgrades
  - **Temporal consistency:** stable sampling intervals, no undocumented breaks

- **For modeled data:**
  - **Verification:** comparison with historical measurements and reference datasets
  - **Data correction:** physical / statistical correction
  - **Plausibility checks:** physical consistency and realistic value ranges
  - **Bias identification:** systematic over- or under-estimation
  - **Uncertainty assessment:** understanding model limitations and spread

Data verification and correction

- Data from model needs to be verified against past measurements:
  - Compare distributions (PDF, CDF, KS test) – Tx, Tn, Daily rain, 10m rain
  - Compare extreme values & probabilities – Txx, Tnn, Rx1day, Max-I
  - Compare trends



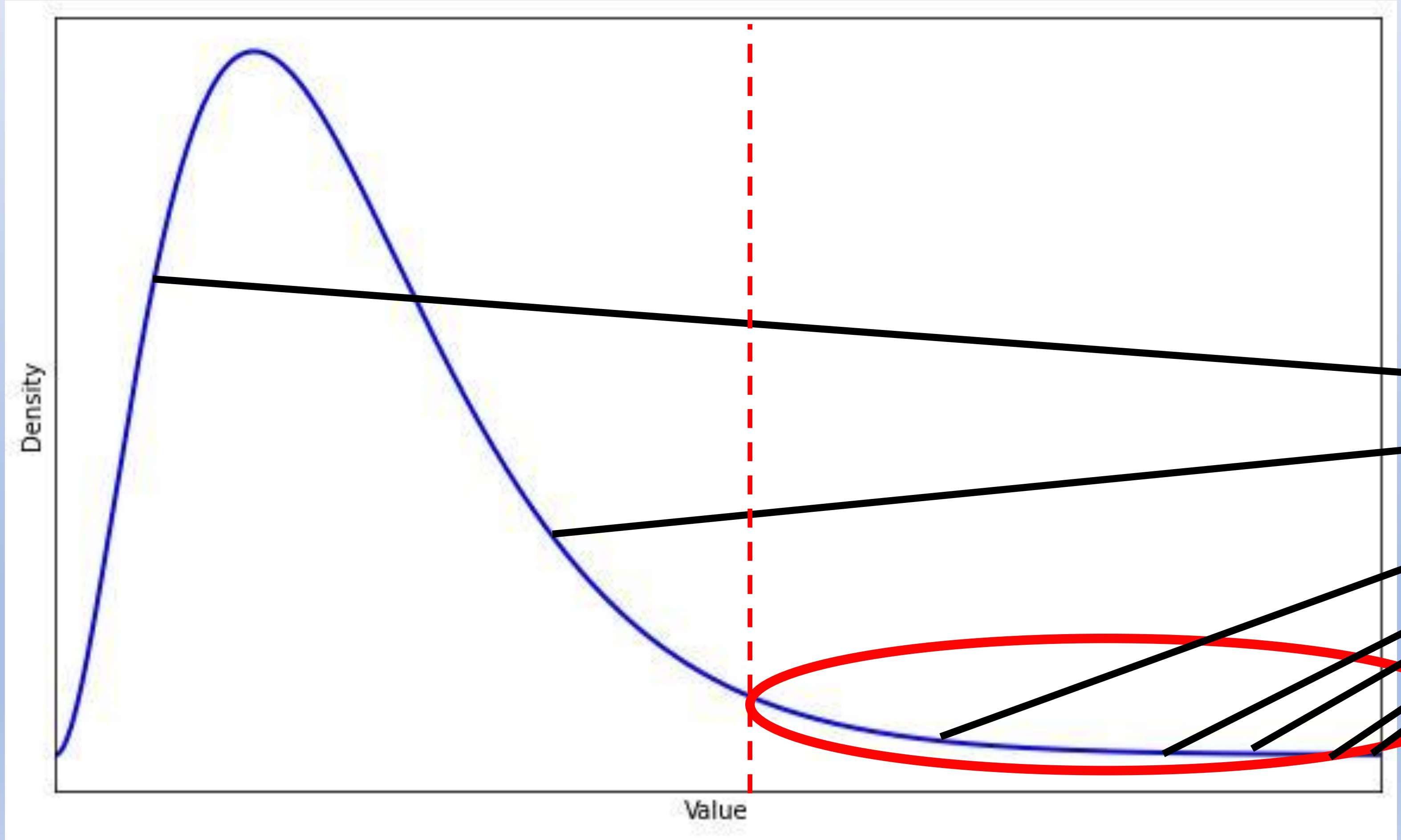


Over estimation of the right tail (left panel) leads to over estimation of events probability (right panel)

- If the verification results are not good enough, a correction for the data needs to be applied:
  - Correct model schemes, parametrizations, etc.
  - Statistical correction for the results (Verification and calibration)

Extreme Value Theory

- **Extreme Value Theory (EVT) is a distinct branch of statistics, built on its own principles and limit theorems. It often behaves very differently from standard statistical methods that focus on the bulk of the data.**



Were going to focus on two main methods:

Block Maxima (BM)

Peak Over Threshold (POT)

Extreme Value Theory (EVT)

	POT	BM
Distribution fit	**Generalized Pareto Distribution (GPD)** for exceedances over a threshold.	**Generalized Extreme Value (GEV)** distribution for block maxima
Fit parameters	**Location (μ)**- sets the central level of extremes; **Shape (ξ)**- tail heaviness; and **Scale (σ or β)**- Spread of values, stretches or compresses the distribution horizontally and must be > 0	
Main principal	Model **all observations that exceed** a suitably high threshold. Uses the *tail of the parent distribution behavior* directly.	Model **only the maximum** from each block (year, season, day…). Treats maxima as a separate process.
Data used	All exceedances above the threshold → usually **many more data points than BM**.	Only the largest value per block!
When it's used	Dense/high-resolution data, use more extreme values information. Common in atmospheric science, hydrology and finance.	When natural block structure. When maxima are naturally defined by physical/organizational constraints.
Advantages	Lower confidence intervals (Uses more data) Allows to use shorter time series Direct modeling of tail behavior	Simple and intuitive (one max per block) Avoids threshold selection Well-established theoretical properties Good when block maxima are physically meaningful (e.g. yearly design floods)
Disadvantages	Requires threshold selection (bias–variance balance) Time consuming Must ensure exceedances are independent Diagnostics more demanding	Significant loss of information Larger uncertainty due to small sample size Block choice can change results May miss significant extremes Requires reliable data to get the real maxima – Data verification

General working process

Stage 1: Building data table - Event-Intensity-Duratio

- Verify data, Reconstruct and estimate past data / events, Combine historical data

	Tm_Beg	Tm_End	RnT	I-10	I-30	I-60	I-120	I-240	T-10	T-30	T-60	T-120	T-240
0	23/04/1971 20:44	24/04/1971 6:04	2.4	7.8	3	1.7	0.9	0.4	23/04/1971 20:54	23/04/1971 21:14	23/04/1971 21:44	23/04/1971 22:44	24/04/1971 0:44
1	25/10/1971 14:27	26/10/1971 7:36	12	36.6	22.6	11.5	5.8	2.9	25/10/1971 14:37	25/10/1971 14:57	25/10/1971 15:27	25/10/1971 16:27	25/10/1971 18:27
2	15/11/1971 21:08	18/11/1971 7:39	34.1	24	8.2	5	4	2.6	17/11/1971 18:18	17/11/1971 18:18	17/11/1971 12:28	17/11/1971 13:04	17/11/1971 15:25
3	06/12/1971 8:50	09/12/1971 7:33	45.3	28.2	11.6	7.1	5.7	3.9	07/12/1971 5:10	07/12/1971 5:26	08/12/1971 8:50	08/12/1971 8:36	08/12/1971 9:14
4	12/12/1971 22:10	13/12/1971 7:54	11.1	12	5	4.1	2.7	1.8	12/12/1971 23:05	12/12/1971 23:05	12/12/1971 23:10	13/12/1971 0:48	13/12/1971 2:10
5	17/12/1971 20:01	19/12/1971 7:37	41.1	10.2	8.2	7	6.8	6.5	17/12/1971 23:10	18/12/1971 3:14	18/12/1971 1:39	18/12/1971 1:43	18/12/1971 3:00
6	20/12/1971 23:35	21/12/1971 6:50	3.3	2.4	2	1.2	0.9	0.6	21/12/1971 1:42	21/12/1971 2:01	21/12/1971 2:32	21/12/1971 2:01	21/12/1971 3:35
7	24/12/1971 17:23	26/12/1971 6:46	18.3	24	16.2	9.4	5.1	2.8	25/12/1971 11:22	25/12/1971 11:25	25/12/1971 11:32	25/12/1971 11:48	25/12/1971 14:17
8	27/12/1971 8:14	28/12/1971 11:42	14.5	16.8	11.6	6.4	3.8	2.1	27/12/1971 9:06	27/12/1971 9:08	27/12/1971 9:21	27/12/1971 10:34	27/12/1971 12:14
9	30/12/1971 16:34	01/01/1972 17:46	9.2	6	3.6	2.3	1.4	0.7	31/12/1971 16:32	31/12/1971 16:35	01/01/1972 17:32	01/01/1972 17:46	01/01/1972 17:46
10	13/01/1972 15:46	17/01/1972 7:27	52.2	19.8	11.6	7.4	3.9	2.7	16/01/1972 17:41	16/01/1972 17:52	16/01/1972 18:13	16/01/1972 18:13	15/01/1972 1:23
11	22/01/1972 8:00	23/01/1972 7:21	9	12.6	5.2	3.8	2.6	1.3	22/01/1972 9:17	22/01/1972 9:12	22/01/1972 9:19	22/01/1972 10:00	22/01/1972 12:00
12	27/01/1972 13:39	28/01/1972 6:03	3.3	14.4	5.6	2.9	1.5	0.7	27/01/1972 13:49	27/01/1972 14:09	27/01/1972 14:39	27/01/1972 15:39	27/01/1972 17:39
13	31/01/1972 10:59	01/02/1972 7:58	39	48	29.2	16.1	8.1	5.2	01/02/1972 7:29	01/02/1972 7:38	01/02/1972 7:55	01/02/1972 7:58	01/02/1972 7:58
14	03/02/1972 8:25	08/02/1972 8:00	71.6	16.8	9.6	7.9	5.9	4.5	06/02/1972 1:41	06/02/1972 2:58	06/02/1972 3:27	06/02/1972 3:31	06/02/1972 4:55
15	16/02/1972 16:53	19/02/1972 1:42	22.1	16.8	11.4	7	3.6	2.4	17/02/1972 11:15	17/02/1972 11:23	17/02/1972 11:23	17/02/1972 11:23	17/02/1972 11:23
16	13/03/1972 18:54	18/03/1972 5:38	82.8	20.4	19.8	13.8	9.6	8.4	16/03/1972 23:44	16/03/1972 23:55	16/03/1972 23:55	17/03/1972 0:43	17/03/1972 2:46
17	20/03/1972 21:38	21/03/1972 7:58	10.2	6.6	4.6	3.1	2.2	1.5	20/03/1972 22:22	21/03/1972 4:40	21/03/1972 4:37	21/03/1972 5:24	21/03/1972 7:29
18	04/04/1972 0:30	04/04/1972 1:55	3.9	13.2	4.8	3.7	2	1	04/04/1972 0:40	04/04/1972 1:00	04/04/1972 1:30	04/04/1972 1:55	04/04/1972 1:55
19	09/04/1972 19:39	11/04/1972 7:15	29.3	31.2	11	6.9	5.1	2.9	10/04/1972 10:12	10/04/1972 10:31	11/04/1972 5:03	11/04/1972 5:03	11/04/1972 6:09
20	29/04/1972 12:32	29/04/1972 13:49	5.7	17.4	6	3.8	2.9	1.4	29/04/1972 12:42	29/04/1972 13:02	29/04/1972 13:32	29/04/1972 13:49	29/04/1972 13:49
21	28/10/1972 8:07	29/10/1972 7:49	8.5	9	3.6	2	1.2	0.9	29/10/1972 1:22	28/10/1972 16:15	29/10/1972 1:22	29/10/1972 1:22	28/10/1972 19:19
22	02/11/1972 23:22	04/11/1972 6:48	11	6.6	5.2	2.9	1.5	1.1	04/11/1972 5:25	04/11/1972 5:44	04/11/1972 5:49	04/11/1972 6:48	03/11/1972 3:22
23	14/11/1972 4:21	14/11/1972 7:16	2.9	5.4	2.8	1.8	1.2	0.7	14/11/1972 6:14	14/11/1972 6:34	14/11/1972 6:39	14/11/1972 6:37	14/11/1972 7:16
24	24/11/1972 8:03	24/11/1972 23:34	2.1	2.4	1.6	0.9	0.5	0.4	24/11/1972 9:03	24/11/1972 12:10	24/11/1972 12:16	24/11/1972 12:16	24/11/1972 12:10
25	27/11/1972 8:01	29/11/1972 7:30	7.4	14.4	6.2	3.5	1.9	1	29/11/1972 6:01	29/11/1972 6:11	29/11/1972 6:01	29/11/1972 6:11	29/11/1972 7:30
26	18/12/1972 16:23	21/12/1972 12:48	56.1	23.4	13	9.7	7.2	4.8	20/12/1972 10:48	20/12/1972 11:10	20/12/1972 11:06	20/12/1972 12:05	20/12/1972 14:06
27	12/01/1973 9:15	13/01/1973 7:36	12.5	13.2	4.6	2.9	2.6	1.4	12/01/1973 20:09	12/01/1973 20:15	12/01/1973 22:03	12/01/1973 21:57	12/01/1973 22:08
28	14/01/1973 8:19	18/01/1973 6:08	63.4	9.6	8.2	8	5.5	3.9	15/01/1973 3:31	15/01/1973 3:08	15/01/1973 3:39	15/01/1973 4:13	15/01/1973 4:13
29	27/01/1973 8:35	28/01/1973 7:14	19.7	22.8	10.4	5.6	3.4	2.2	28/01/1973 5:09	28/01/1973 5:29	28/01/1973 5:35	28/01/1973 6:59	28/01/1973 7:14
30	30/01/1973 12:07	01/02/1973 3:31	28.3	29.4	13.2	8.8	5.7	3.8	31/01/1973 11:43	31/01/1973 11:44	31/01/1973 11:44	31/01/1973 11:44	31/01/1973 13:13
31	22/02/1973 9:58	24/02/1973 6:24	12.7	8.4	4.2	3.6	2.8	1.5	23/02/1973 16:31	23/02/1973 17:35	23/02/1973 17:18	23/02/1973 17:50	23/02/1973 18:09
32	02/03/1973 17:59	03/03/1973 12:52	33.5	24	13.4	7.8	3.9	3	02/03/1973 18:15	02/03/1973 18:35	02/03/1973 18:59	02/03/1973 19:59	03/03/1973 5:12
33	05/03/1973 23:31	08/03/1973 6:10	48.1	28.2	12.8	8.2	4.7	3.1	06/03/1973 19:13	06/03/1973 19:33	06/03/1973 19:50	06/03/1973 7:43	06/03/1973 8:29
34	20/03/1973 9:41	21/03/1973 7:57	3.3	6	2.8	1.6	1	0.6	21/03/1973 5:19	21/03/1973 5:19	21/03/1973 5:49	21/03/1973 6:47	21/03/1973 7:57

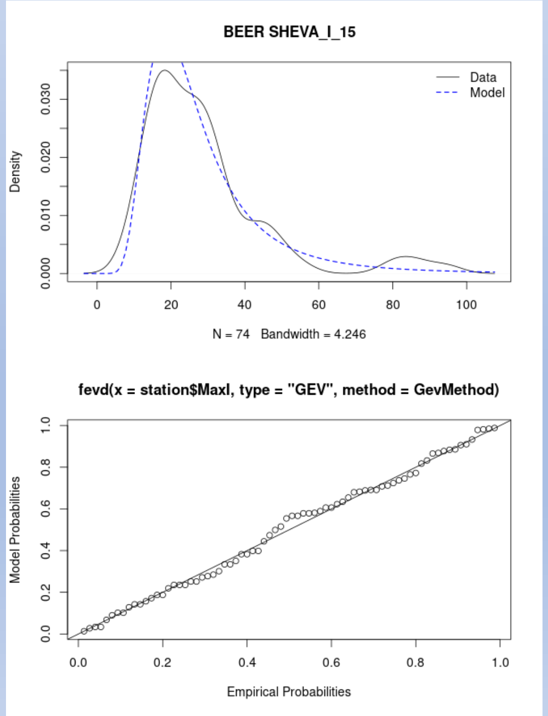
You set the condition to differentiate **INDEPENDENT** rain events!!! (Synoptic, time,…)

Event-Intensity-Duration (EID) table

General working process

$$P = \frac{1}{RP}$$

An event of 1:100 yr can happen year after year!!!



Analyzing rain intensity return periods (probability)

BM

Fit distribution:
GEV (**Fréchet (ξ > 0)**, **Gumbel (ξ = 0)**, **Weibull (ξ < 0))**
Lognormal, Gamma, Weibull (Non EVT)

Verify fit quality

Estimate events probability (RP values) from the fitted distribution

POT

Chose threshold

Fit distribution:
GPD (**Pareto (ξ > 0)**, **exponential (ξ = 0)**, **Uniform/Power (ξ < 0))**
Lognormal, Gamma, Weibull (Non EVT)

Verify fit quality

Verify distribution parameters and threshold

Adjust if needed

Estimate events probability (RP values) from the fitted distribution

POT Thresholds Selection

When using POT, The final estimate is highly sensitive to the choice of the threshold.

The Trade-off:

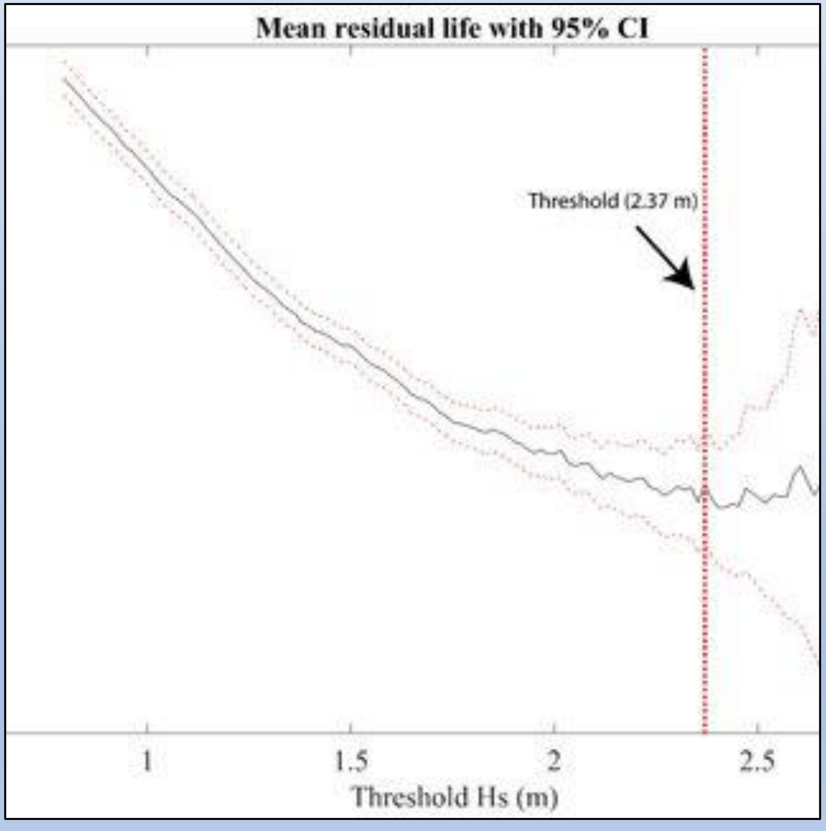
- **High Threshold** → Small sample size: Leads to a high **Variance** (not enough data).
- **Low Threshold** → Large sample size **:** Leads to a high **Bias** (including non-extreme data), Under estimation of events probability

1. **Quantile-Based Threshold:**
   - Chose an upper quantile (80, 90, 95, 99…) as threshold

2. **Mean Residual Life (MRL) plot** :

$$\mathrm{MRL}(u) = \frac{\sum_{i:x_i>u}(x_i - u)}{N_u}$$

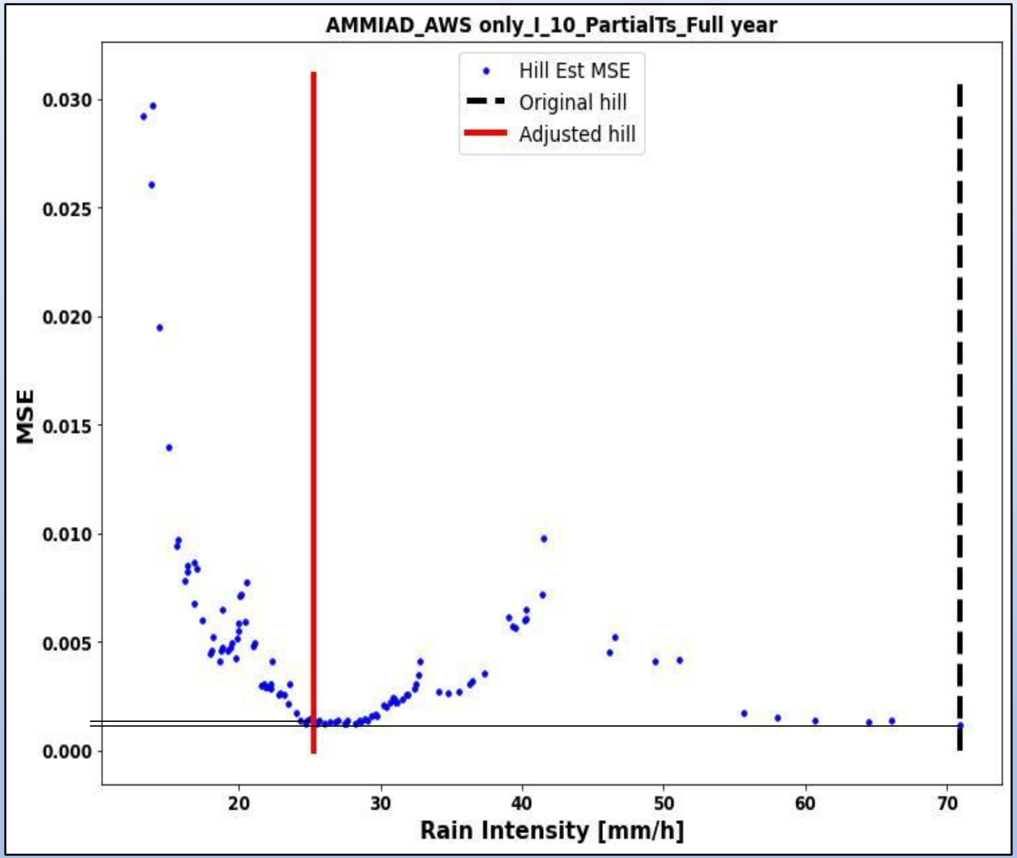
   - Look for a region where the plot is approximately linear, indicating that the GPD is a reasonable fit above that threshold.
   - Threshold is chosen at the beginning of this linear region.
   - X_i – Sample, u-Threshold, N_u- Number of samples above u



3. **Parameter Stability Plot:**
   - Fit a GPD to exceedances for different candidate thresholds.
   - Plot the estimated shape (ξ) and scale (σ) parameters against the threshold.
   - Threshold is chosen where the parameters stabilize (i.e., stop changing significantly with increasing threshold).

POT Thresholds Selection

4. **Automated Methods / Goodness-of-Fit**
   - Fit GPD to exceedances over candidate thresholds and use statistical criteria:
     - Likelihood-based methods: maximize likelihood or use AIC/BIC.
     - Kolmogorov–Smirnov test: test GPD fit for different thresholds.
     - Bootstrap methods: quantify uncertainty of parameter estimates.



5. **Weighted Hill estimation:**
   - Weight the exceedance in a decreasing order
   - Calculate the Mean Square Error (MSE) for each chosen threshold (Kołodziejczyk and Rutkowska 2023)
   - Get the recommended threshold
   - Adjust the threshold

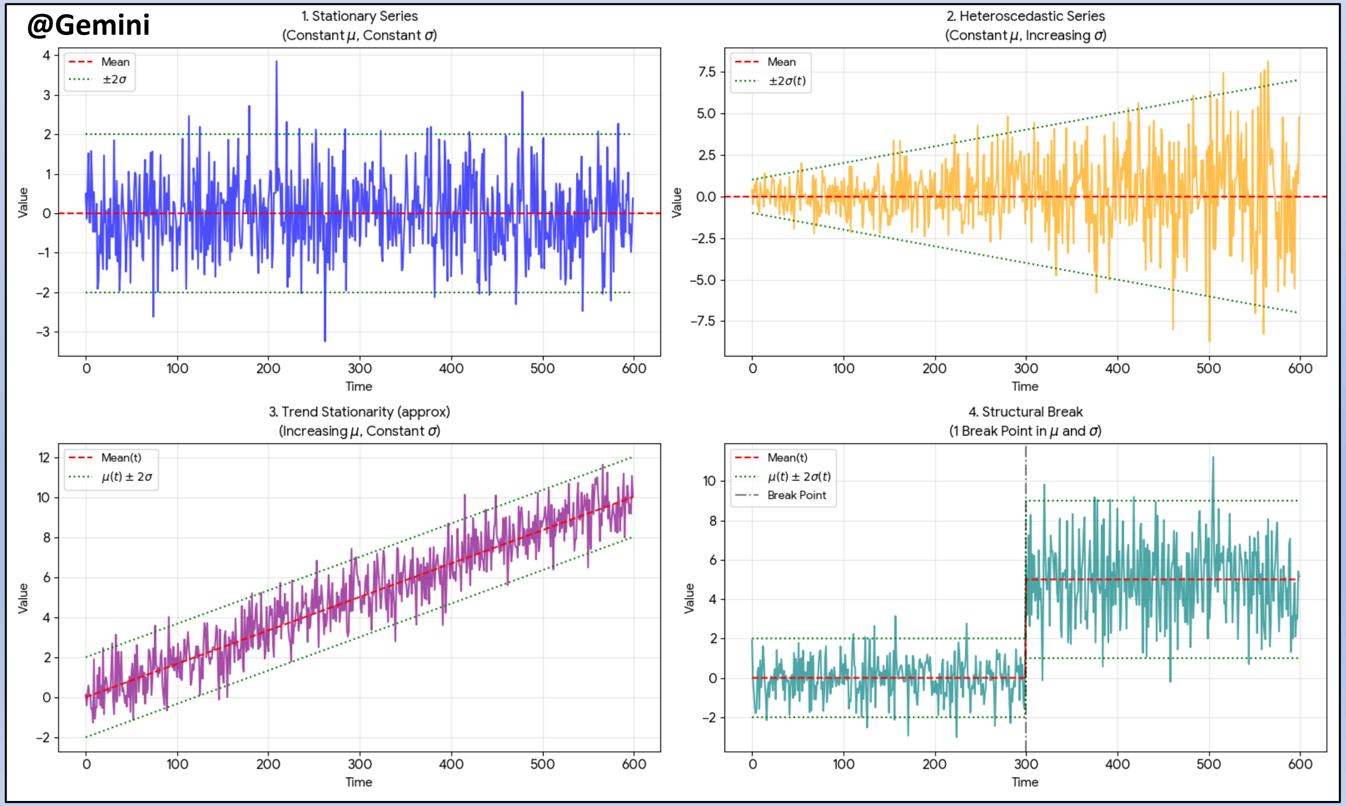
Kołodziejczyk, K.; Rutkowska, A. Estimation of the Peak over Threshold-Based Design Rainfall and Its Spatial Variability in the Upper Vistula River Basin, Poland. Water 2023, 15, 1316. https://doi.org/10.3390/w15071316

Data stationerity

- A key assumption in EVT is data stationarity
- Data stationerity is the uniformity of a data set meaning that it is from a single population under consistent conditions
- For data set with varied parameters (mean, std) over time this assumption is invalid

<u>Sources of non stationerity :</u>

- Inhomogeneity: changes in measurement conditions (e.g., instrument, location, resolution, etc.)
- Climate driven changes → Global warming!!!



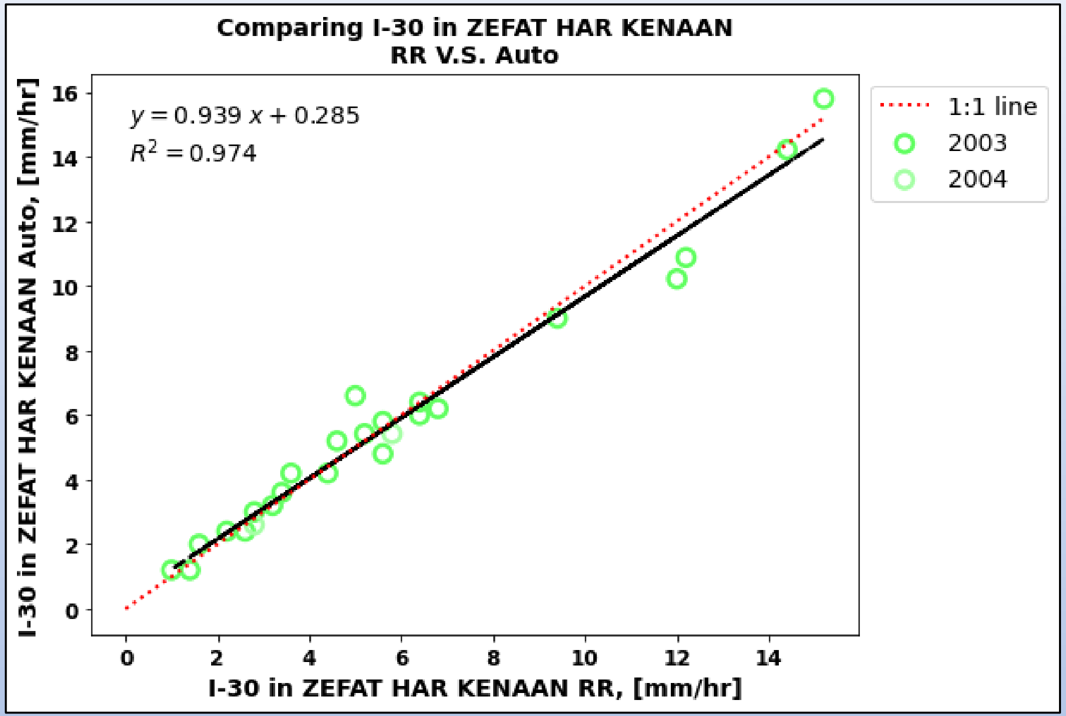
Non stationerity data will produce:

- ➤ **Biased uncertainty ranges**
- ➤ **Biased return levels**

Handling non-stationerity

Homogenization

- Apply homogeneity tests and identify breakpoints in the data (more relevant for past data)
- Examine metadata: station relocation, instrument changes, maintenance activity, etc.
- Apply adjustments/corrections to the data



Non Stationarity tests

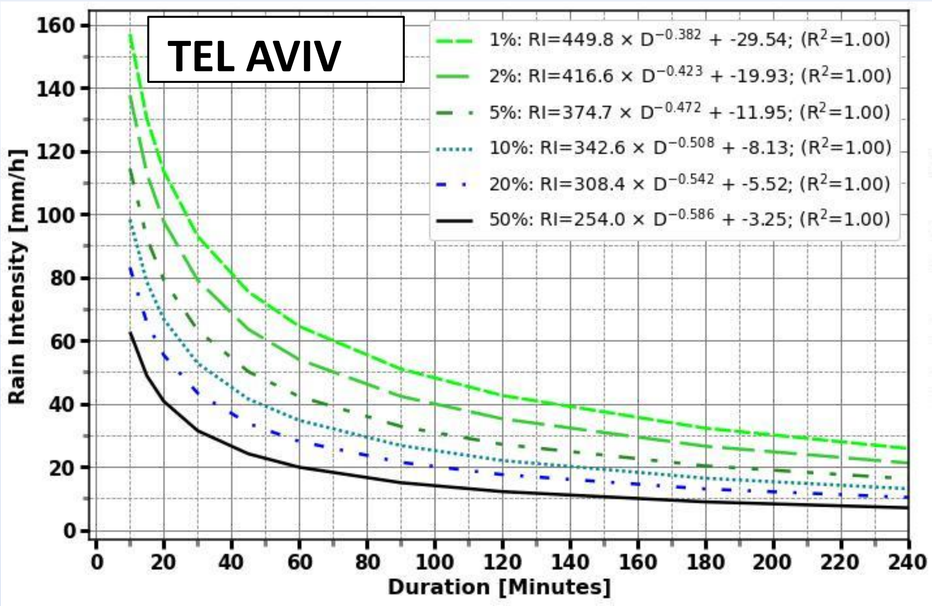
- Fit parameters change over time (for different time blocks)
- Use non stationary EVT models (Let σ, μ, ξ depend on time or covariates (Heffernan and Stephenson, 2018) and examine goodness of fit
- Examine trends, change in mean and std over time

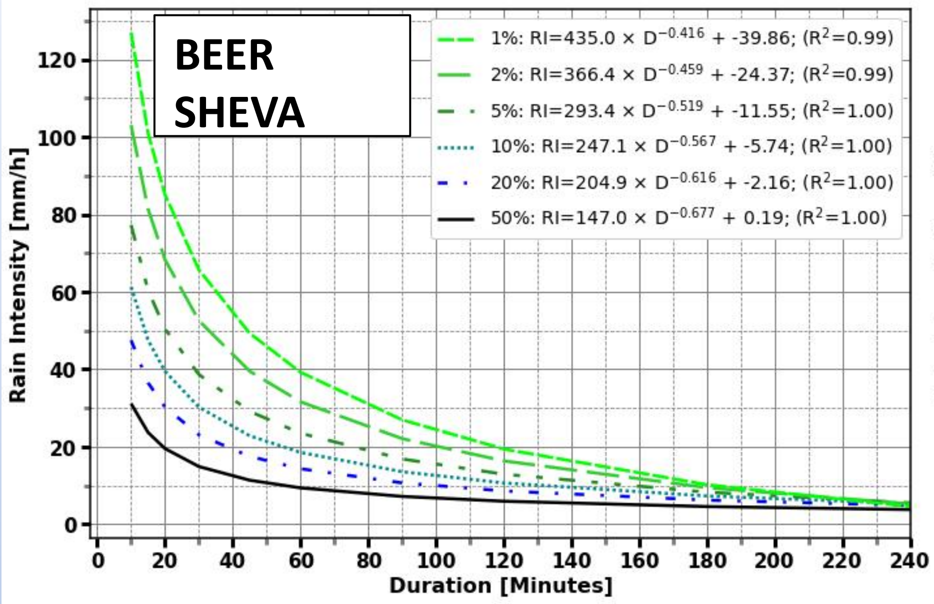
Address non stationarity

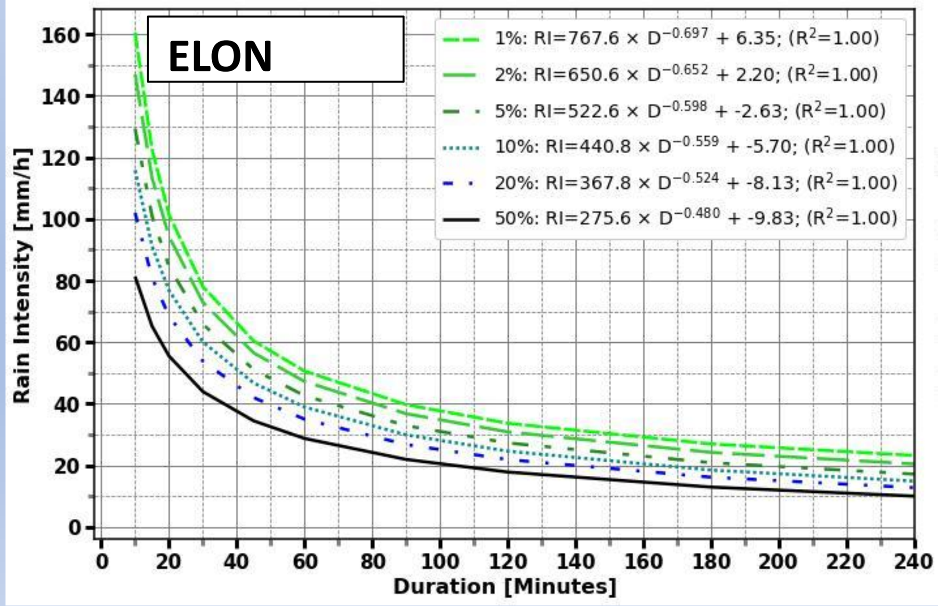
- Split the record into stationary sub-periods (POT / BM)
- Use non-stationary EVT models - Let σ, μ, ξ depend on time or covariates

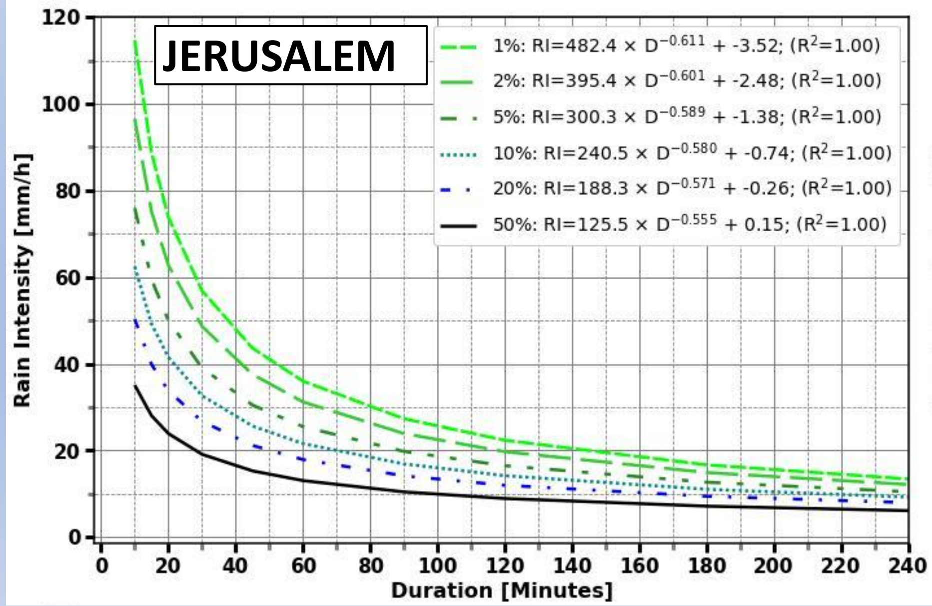
Heffernan, J. E., & Stephenson, A. G. (2018). R package ‘ismev’: An Introduction to Statistical Modeling of Extreme Values.

Finally – IDF's (Intensity-Duration-Frequency)



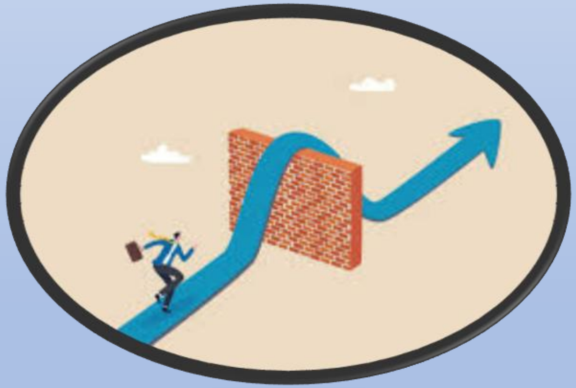


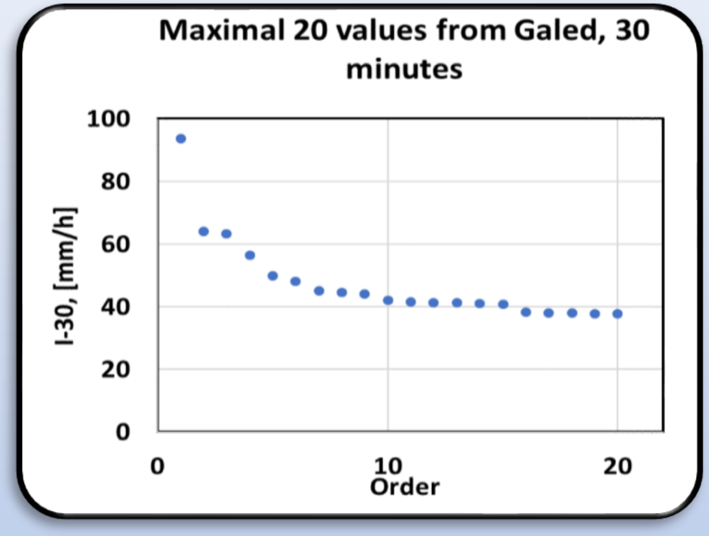


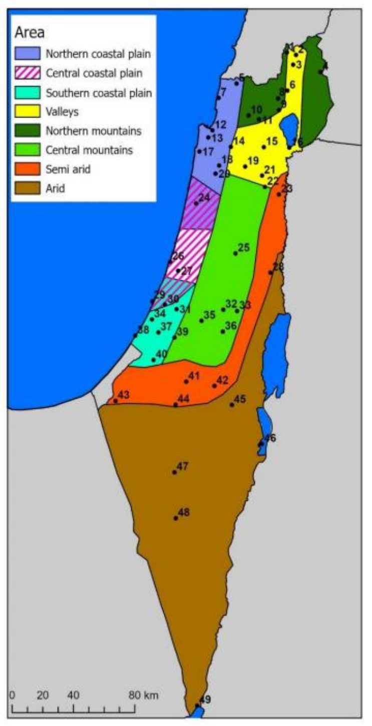


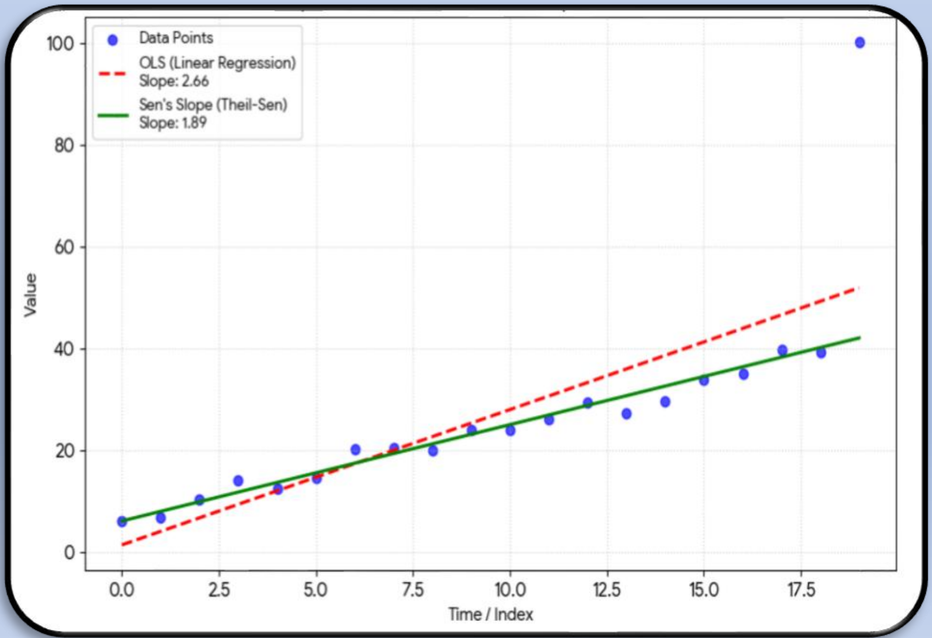
Trend analysis - Main principals

- Rainfall intensity displays strong spatial variability.
- This variability increases as the duration becomes shorter and the probability decreases (return period increases).
- A single extreme event recorded at one station can substantially influence trend estimates, while nearby stations did not experience it.
- Moreover, in many series the maximum value is often far higher than the next highest observation (sometimes by 50% or more), and such large gaps can strongly affect OLS-based trend analyses.









Use areal analysis

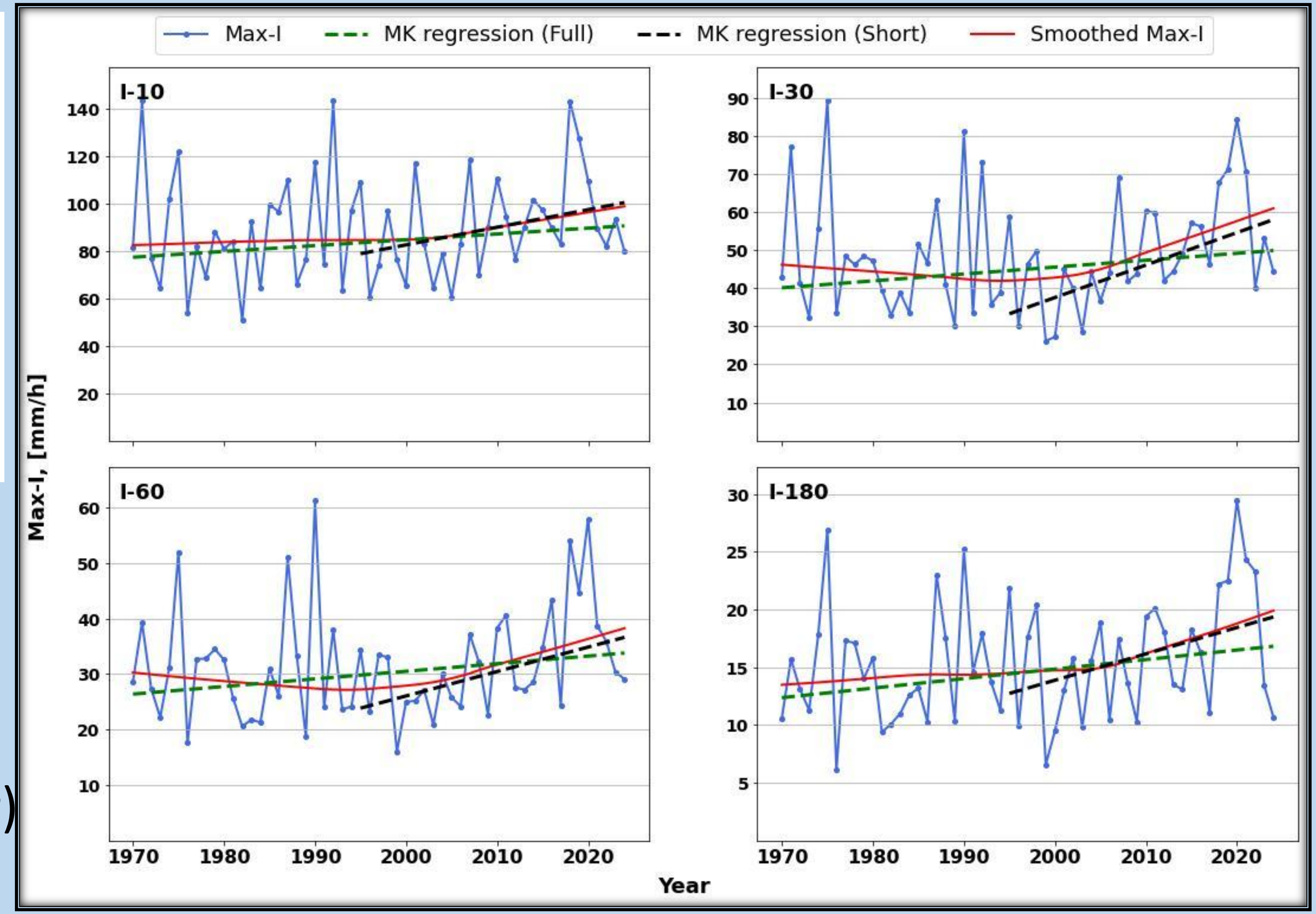
Use non-parametric methods

Trend analysis- Results example

Period	1970-2024		1995-2024	
	Slope [(mm/h)/Decade]	Change rate	Slope [(mm/h)/Decade]	Change rate
I-10	2.5	17%	7.4	27%
I-15	2.6	23%	8.9*	42%
I-20	2.5	26%	8.1*	46%
I-30	1.8	24%	8.5*	74%
I-45	1.4	23%	6.0**	61%
I-60	1.4+	28%	4.4*	53%
I-90	0.9	24%	3.6+	56%
I-120	0.7	25%	3.4+	64%
I-180	0.8+	36%	2.3+	52%
I-240	0.6+	33%	1.8+	53%

p-Value < 0.1 (+); p-Value < 0.05 (*); p-Value < 0.01 (**)

Southern coastal line

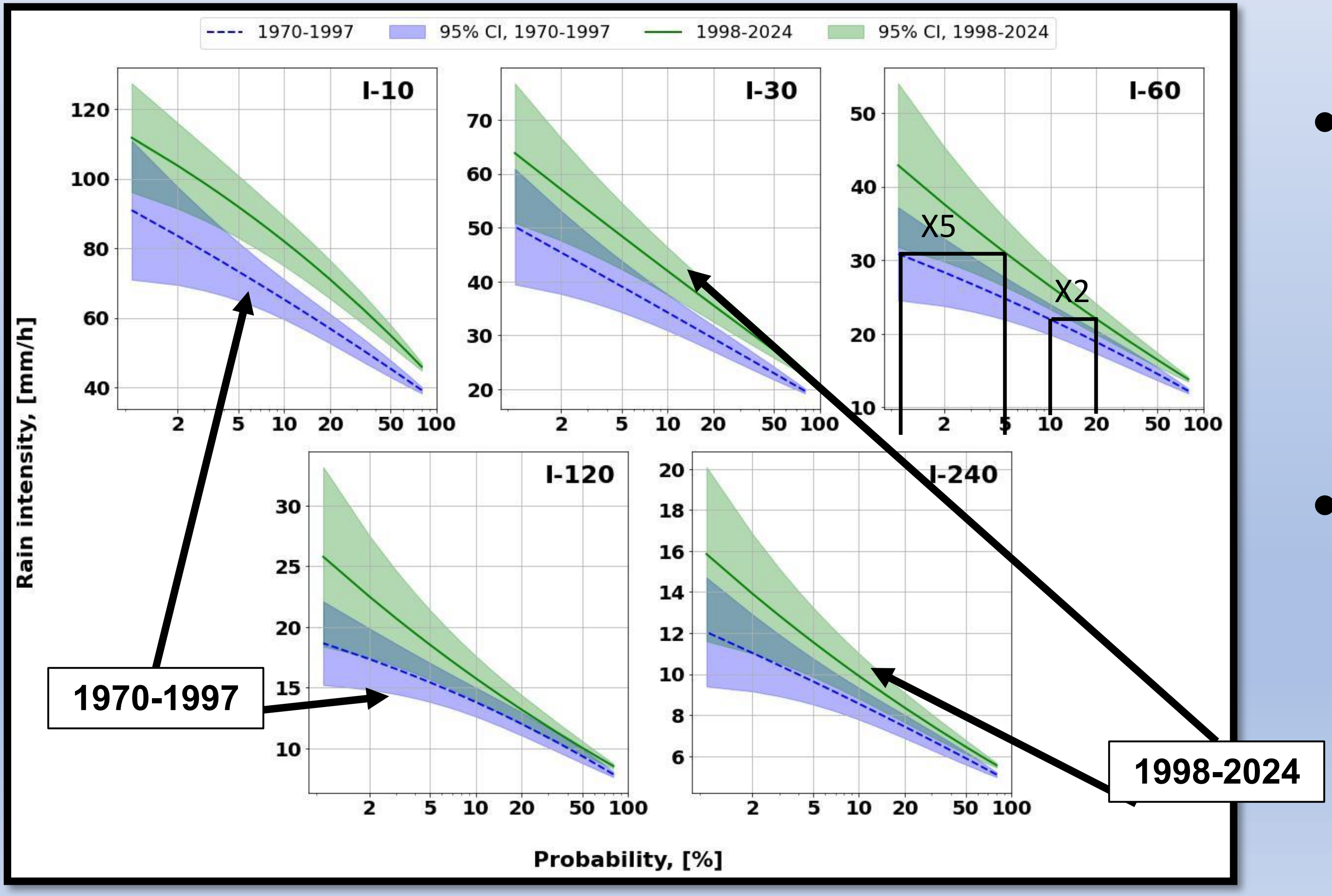


- Stronger trends are observed when shorter periods are analyzed
- Smoothed BM values (LOWESS filter) support these results

Compare periods - Main principals

When splitting a time series into several sub-periods, the resulting segments are often too short for reliable BM analysis, as the large uncertainty usually prevents detecting meaningful differences.
In such cases, use the POT method with a constant quantile-based threshold applied to all periods.

Valleys area

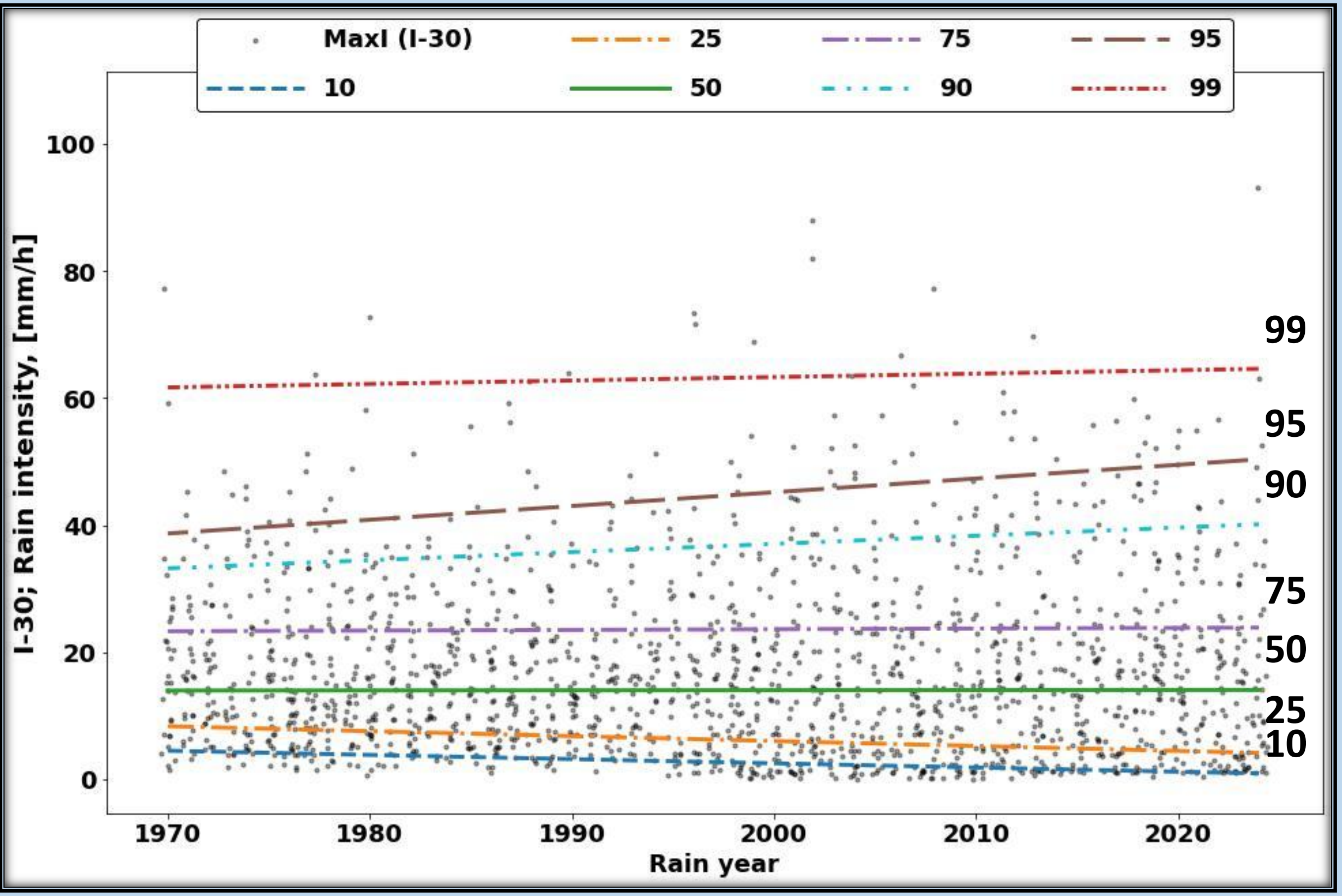


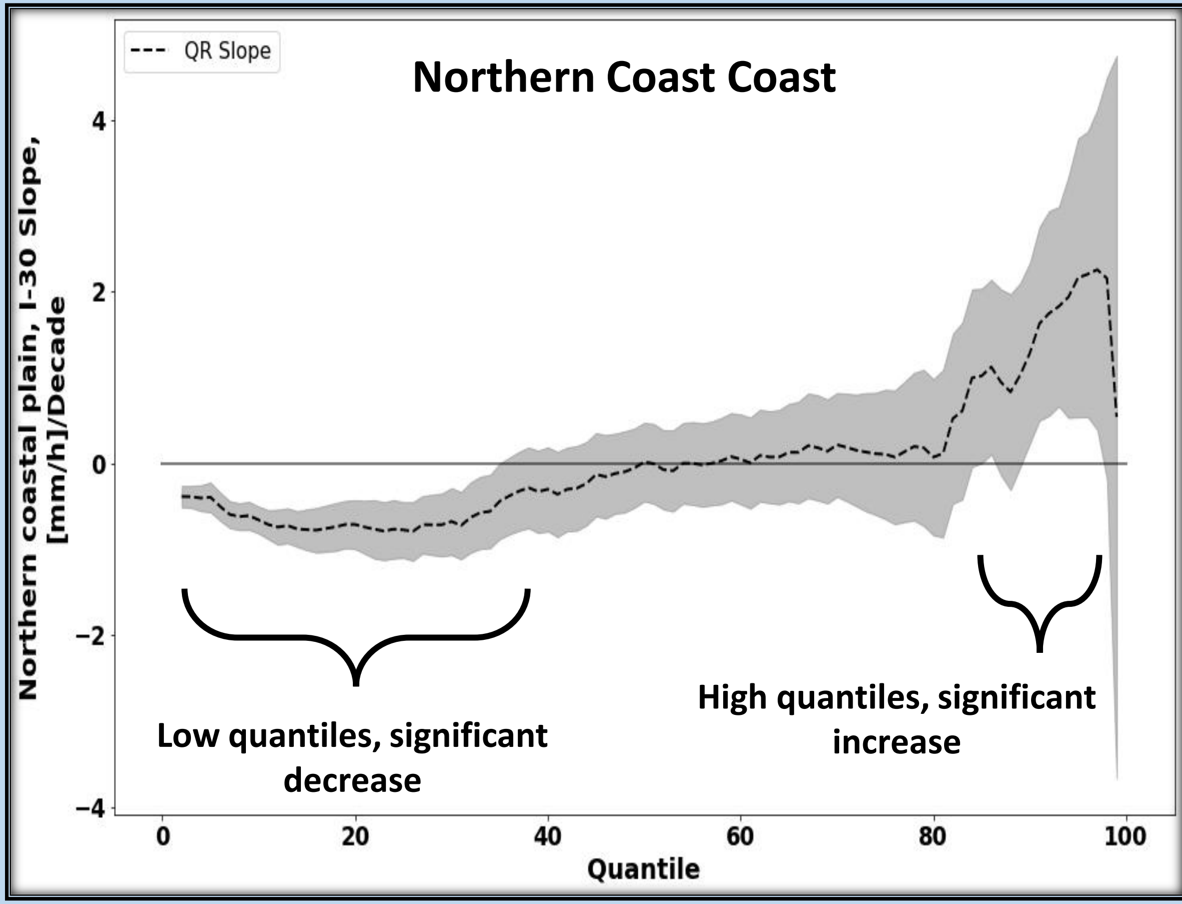
- Uncertainty increases as the duration becomes shorter and the probability decreases.
- The growth factor increases as the probability decreases.

Examine changes in total rain intensities

Quantile regression on EID for duration I-30.

Allows to explore the change in the whole distribution (instead the mean / median / maximal value)





Examine changes in total rain intensities

Summary / Conclusions:

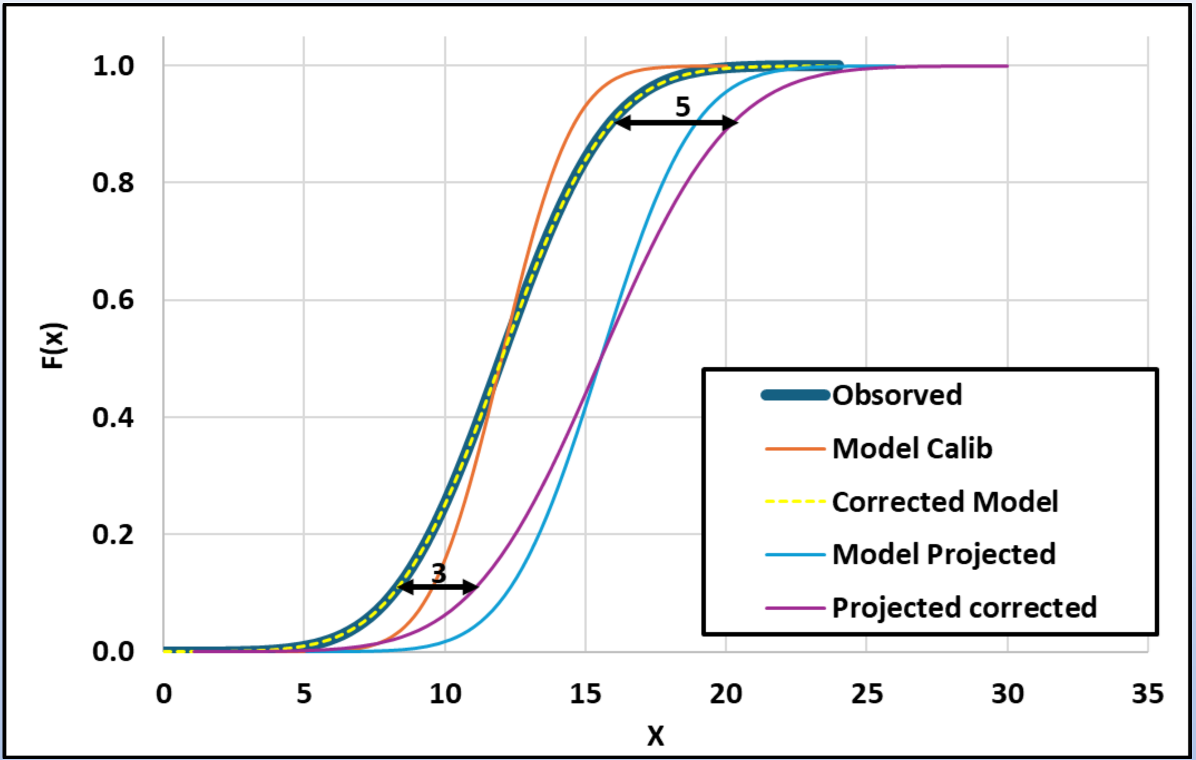
- Influence critical sectors, including infrastructure, energy, transportation, agriculture, and public safety.
- Predicting rain intensities improves preparedness across sectors such as infrastructure, emergency services, and drainage systems.
- High-quality, long, and reliable data are essential for accurate rain-intensity prediction.
- Two extreme-value fitting methods were presented — POT and BM. Both are appropriate for hydrological and atmospheric use, and the decision between them depends on data limitations and local conditions.
- Due to the high spatial variability of rain intensity, an areal analysis is recommended.
- It is also recommended to organize the data in EID tables, as they allow better exploration of the dataset and offer multiple analysis options.
- Due to the changing climate, the non-stationarity of the data must be examined and taken into consideration.

Thank you for listening!!!

Questions??

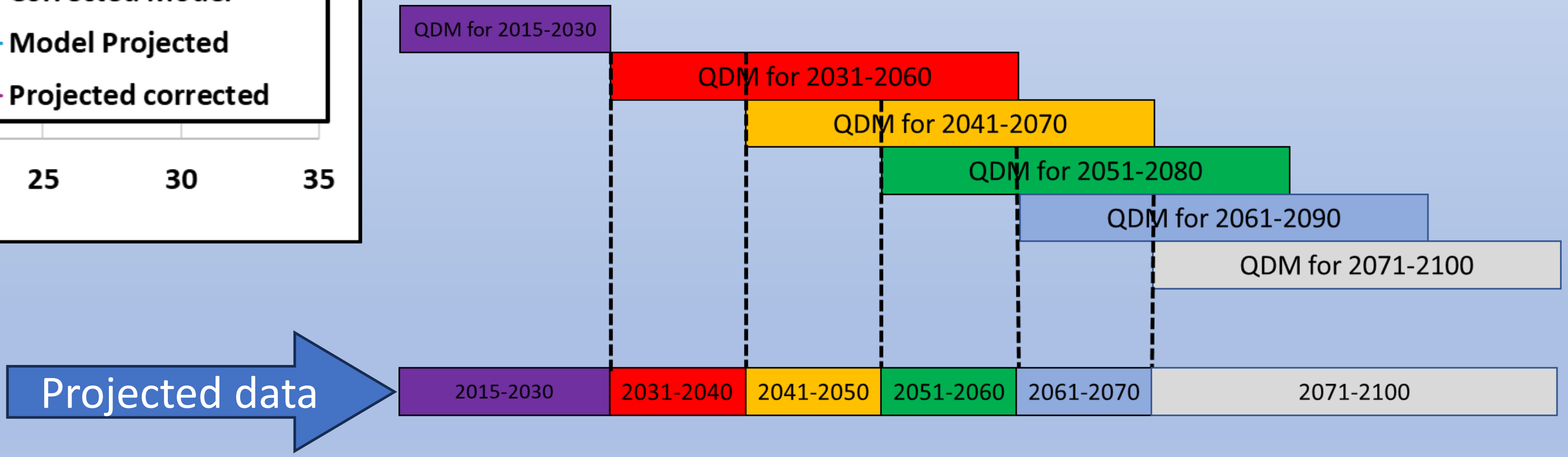
ziporia@ims.gov.il

Quantile Difference Mapping (QDM)



Working process:

i. QDM for calibration period (1980-2004)
ii. QDM for validation period (2005-2014)
iii. Examine performance in validation period
iv. QDM for operative calibration (1980-2014)
v. QDM for projected blocks of time (2015-2030, 2031-2060, 2041-2070, 2051-2080… 2070-2100)
vi. Create projected time series



- Cannon, Alex J., Stephen R. Sobie, and Trevor Q. Murdock. "Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?." *Journal of Climate* 28.17 (2015): 6938-6959.
- Cannon, Alex J. "Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure." *Journal of Climate* 29.19 (2016): 7045-7064.
- Cannon, Alex J. "Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables." *Climate dynamics* 50.1 (2018): 31-49.