Characterization of ERA-5 precipitation: Comparison of ERA5 and EOBS daily precipitation

Pauline Rivoire [1]

Supervision : Olivia Martius [1]
Philippe Naveau [2]

[1]Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CNRS-CEA-UVSQ, Université Paris-Saclay, Gif-sur- Yvette, France

May 6, 2020

UNIVERSITÄT BERN

OESCHGER CENTRE CLIMATE CHANGE RESEARCH

Reanalysis dataset: what & why?

ERA5: ECMWF latest reanalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by European Centre for Medium-Range Weather Forecasts. [C3S, 2017]

Reanalysis dataset: what & why?

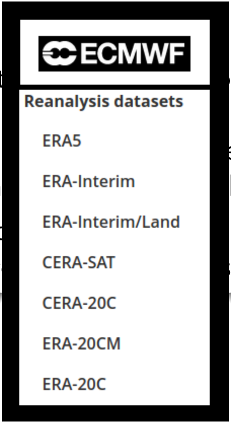
ERA5: **E**CMWF latest **rea**nalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]

Reanalysis dataset: what & why?

ERA5: ECMWF latest [illegible]

- Reanalysis dataset [illegible] observations made in the past with today's numerical [illegible] model, to deliver a complete and consistent picture [illegible] regular temporal and spatial resolution). No m[illegible] bal coverage. [ECMWF, 2019]
- ERA5 = latest glo[illegible] set provided by **E**uropean **C**entre for **M**edium-Rang[illegible] ts. [C3S, 2017]



Reanalysis dataset: what & why?

ERA5: **E**CMWF latest **rea**nalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]

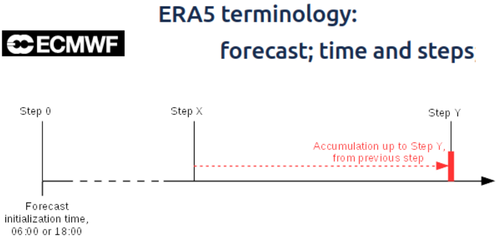
Reanalysis dataset: what & why?

ERA5: **E**CMWF latest **rea**nalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]
- ERA5 precipitation data = forecast data
  "A forecast starts with an analysis at a specific time, and a numerical weather prediction model computes the atmospheric conditions for a number of 'forecast steps', at increasing 'validity times', into the future." [Hennermann, 2020]

Reanalysis dataset: what & why?

ERA5: ECM\



- Reanalys … le in the past with today's r … er a complete and consister … and spatial resolutio … ECMWF, 2019]
- ERA5 = … **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]
- ERA5 precipitation data = forecast data
  "A forecast starts with an analysis at a specific time, and a numerical weather prediction model computes the atmospheric conditions for a number of 'forecast steps', at increasing 'validity times', into the future."

Reanalysis dataset: what & why?

ERA5: **E**CMWF latest **rea**nalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]
- provides precipitation, without relying on observed precipitation, but with forecast derived from the atmospheric conditions

Reanalysis dataset: what & why?

ERA5: ECMWF latest reanalysis dataset

- Reanalysis dataset: combination of observations made in the past with today's numerical weather prediction model, to deliver a complete and consistent picture of past weather (regular temporal and spatial resolution). No missing data and global coverage. [ECMWF, 2019]
- ERA5 = latest global reanalysis dataset provided by **E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts. [C3S, 2017]
- provides precipitation, without relying on observed precipitation, but with forecast derived from the atmospheric conditions
- Hourly precipitation forecast merged to daily precipitation. Spatial resolution 0.25 °.

Comparison with EOBS, an observation based dataset

EOBS: An observation based dataset for Europe

- EOBS: European daily gridded observational dataset for precipitation [Haylock et al., 2008], provided by ECA&D, European Climate Assessment & Dataset.
- Based on station data, spatially interpolated (interpolation of monthly precipitation totals, of daily anomalies)
- Same spatial resolution as ERA5: 0.25 °. Only land precipitation over Europe (including Iceland + part of Nothern Africa, Middle East and Russia)

Comparison with EOBS, an observation based dataset

EOBS: An observation based dataset for Europe

- EOBS: European daily gridded observational dataset for precipitation [Haylock et al., 2008], provided by ECA&D, European Climate Assessment & Dataset.
- Based on station data, spatially interpolated (interpolation of monthly precipitation totals, of daily anomalies)
- Same spatial resolution as ERA5: 0.25 °. Only land precipitation over Europe (including Iceland + part of Nothern Africa, Middle East and Russia)

Comparison with EOBS, an observation based dataset

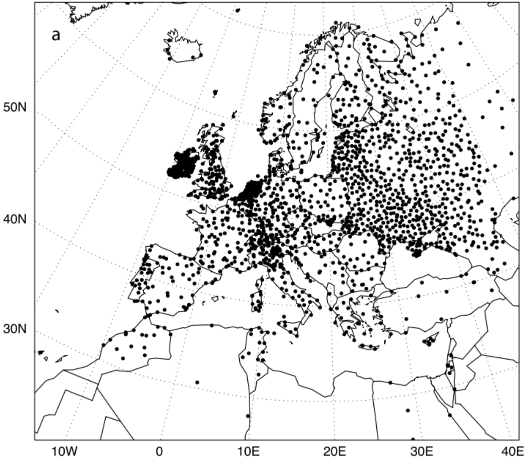
EOBS: An observation based dataset for Europe

- EOBS: European daily gridded observational dataset for precipitation [Haylock et al., 2008], provided by ECA&D, European Climate Assessment & Dataset.
- Based on station data, spatially interpolated (interpolation of monthly precipitation totals, of daily anomalies)
- Same spatial resolution as ERA5: 0.25 °. Only land precipitation over Europe (including Iceland + part of Nothern Africa, Middle East and Russia)

Comparison with EOBS, an observation based dataset

EOBS: An observation

- EOBS: European [Haylock et al., 2(Assessment & Da ... pitation ... e
- Based on station (... monthly precipitation total
- Same spatial resol ... tion over Europe (including ... ast and Russia)



Precipitation station network used in EOBS. [Haylock et al., 2008]
One can notice the uneven spatial distribution.

Comparison with EOBS

Why to compare?

- Both type of data have their uncertainties (from measurement, spatial interpolation (observation), model itself (reanalysis))
  => We don't have access to the "true" value of precipitation.
- But we can approach it, knowing strengths and weaknesses of our precipitation datasets, and comparing them.
- The ERA5 precipitation production process does not include precipitation observation inputs: independance of ERA5 and EOBS.

Comparison with EOBS

Why to compare?

- Both type of data have their uncertainties (from measurement, spatial interpolation (observation), model itself (reanalysis))
  => We don't have access to the "true" value of precipitation.
- But we can approach it, knowing strengths and weaknesses of our precipitation datasets, and comparing them.
- The ERA5 precipitation production process does not include precipitation observation inputs: independance of ERA5 and EOBS.

Comparison with EOBS

Why to compare?

- Both type of data have their uncertainties (from measurement, spatial interpolation (observation), model itself (reanalysis))
  => We don't have access to the "true" value of precipitation.
- But we can approach it, knowing strengths and weaknesses of our precipitation datasets, and comparing them.
- The ERA5 precipitation production process does not include precipitation observation inputs: independance of ERA5 and EOBS.

Comparison with EOBS

In practice

- ERA5 hourly precipitation summed to daily precipitation (starting from 00:00 UTC)
- Comparison from 02.01.1979 to 31.12.2018, land precipitation over Europe.
- Seasonal analysis. Only result maps in autumn, September-October-November (SON) shown here.
- Focus on wet days, defined daily precipitation$> 1mm$.

Comparison with EOBS

In practice

- ERA5 hourly precipitation summed to daily precipitation (starting from 00:00 UTC)
- Comparison from 02.01.1979 to 31.12.2018, land precipitation over Europe.
- Seasonal analysis. Only result maps in autumn, September-October-November (SON) shown here.
- Focus on wet days, defined daily precipitation $> 1mm$.

Comparison with EOBS

In practice

- ERA5 hourly precipitation summed to daily precipitation (starting from 00:00 UTC)
- Comparison from 02.01.1979 to 31.12.2018, land precipitation over Europe.
- Seasonal analysis. Only result maps in autumn, September-October-November (SON) shown here.
- Focus on wet days, defined daily precipitation$> 1mm$.

Table of contents

Table of contents

Hit Rate

- Verification of the co-occurrence of certain events, defined as events above a percentile (30, 50, 75, 90 and 95th percentiles on wet days).
- Only the timing of events, rather than the intensity, is assessed here: the percentile value may vary from one dataset to another.
- Co-occurrence is quantified at every grid point with the hit rate $\hat{\mathcal{H}}$, defined by:

$$\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$$

e_{hit} = number of ERA5 events coinciding with an EOBS events
e_{total} = total number of events (in EOBS or in ERA5).

For one season, between 1979 and 2018.

Hit Rate

- Verification of the co-occurrence of certain events, defined as events above a percentile (30, 50, 75, 90 and 95th percentiles on wet days).
- Only the timing of events, rather than the intensity, is assessed here: the percentile value may vary from one dataset to another.
- Co-occurrence is quantified at every grid point with the hit rate $\hat{\mathcal{H}}$, defined by:

$$\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$$

e_{hit} = number of ERA5 events coinciding with an EOBS events
e_{total} = total number of events (in EOBS or in ERA5).

For one season, between 1979 and 2018.

Hit Rate

- Verification of the co-occurrence of certain events, defined as events above a percentile (30, 50, 75, 90 and 95^{th} percentiles on wet days).
- Only the timing of events, rather than the intensity, is assessed here: the percentile value may vary from one dataset to another.
- Co-occurrence is quantified at every grid point with the hit rate $\hat{\mathcal{H}}$, defined by:

$$\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$$

e_{hit} = number of ERA5 events coinciding with an EOBS events
e_{total} = total number of events (in EOBS or in ERA5).

For one season, between 1979 and 2018.

Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

$$\chi = \lim_{q \to 1} \frac{\mathbb{P}\left[F_1(X_1) > q, F_2(X_2) > q\right]}{1 - q}$$

Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day

  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.

- Remark: link with the tail dependence coefficient, see frame 29

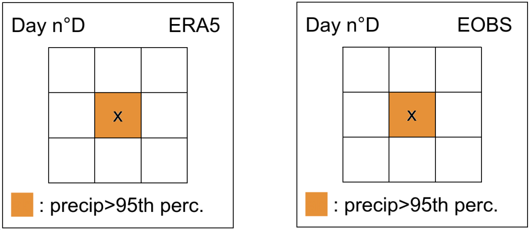
$$\chi = \lim_{q \to 1} \frac{\mathbb{P}\left[F_1(X_1) > q, F_2(X_2) > q\right]}{1 - q}$$

Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

A configuration considered as event coincidence.

The grids represent the 2 gridded datasets, the grid point of interest is the crossed one.

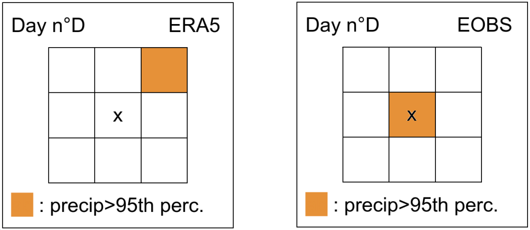


Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

A configuration considered as event coincidence.

The grids represent the 2 gridded datasets, the grid point of interest is the crossed one.

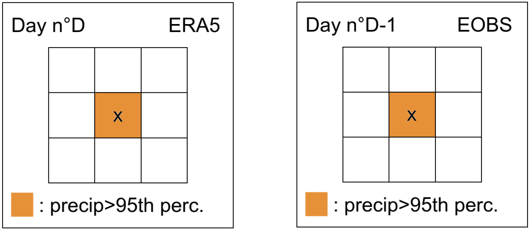


Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

A configuration considered as event coincidence.

The grids represent the 2 gridded datasets, the grid point of interest is the crossed one.

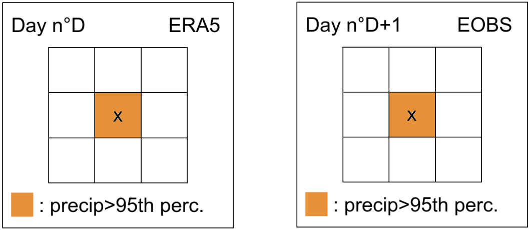


Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

A configuration considered as event coincidence.

The grids represent the 2 gridded datasets, the grid point of interest is the crossed one.



Hit Rate

- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.

- Remark: link with the tail dependence coefficient, see frame 29

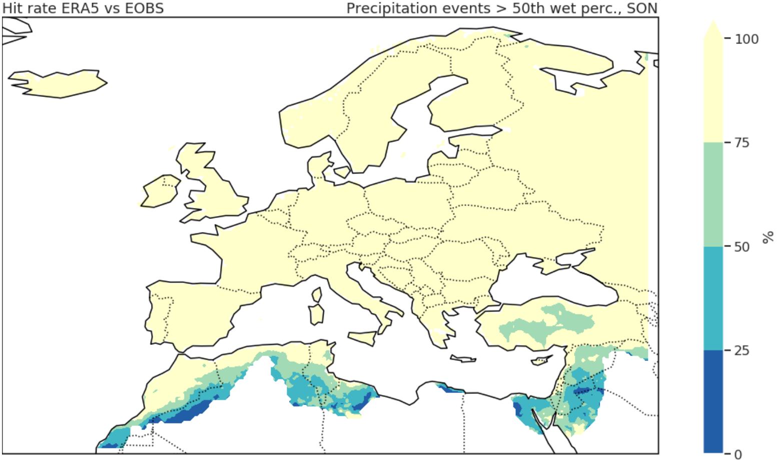
$$\chi = \lim_{q \to 1} \frac{\mathbb{P}\left[F_1(X_1) > q, F_2(X_2) > q\right]}{1-q}$$

Hit Rate

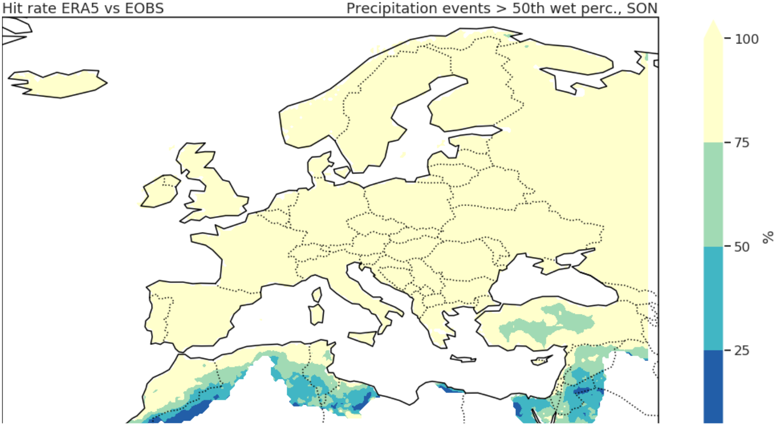
- $\hat{\mathcal{H}} = 100 \times \frac{e_{hit}}{e_{total}}$
- Coincidence = occurrence in the same gridpoint on the same day or +/- 1 day, or one of the 8 surrounding gridpoints on the same day
  The day shift to get rid of accumulation period of daily precipitation discrepancy between ERA5 and EOBS, and in EOBS itself between stations [Haylock et al., 2008]. The spatial shift is to cope discrepancy in coordinates of grids.
- Remark: link with the tail dependence coefficient, see frame 29

$$\chi = \lim_{q \to 1} \frac{\mathbb{P}\left[F_1(X_1) > q, F_2(X_2) > q\right]}{1 - q}$$

Hit Rate: events > 50^{th} perc



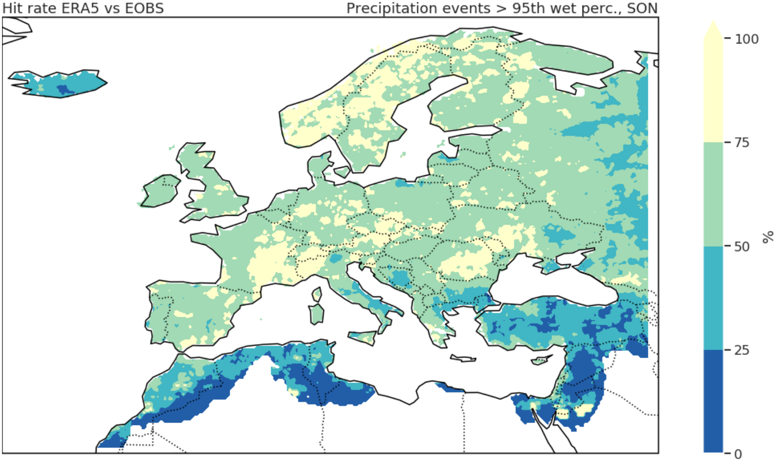
Hit Rate: events > 50^{th} perc



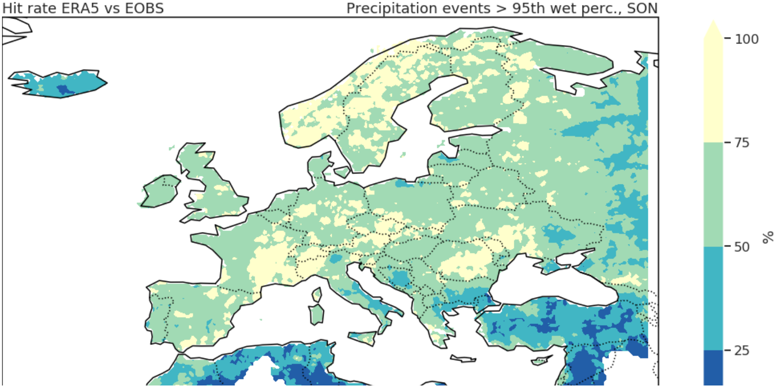
Precipitation events above the 50^{th} perc:

- almost all pixels have more than 75% of events coinciding
- bad performance in arid region, where very few wet days (>1*mm*).

Hit Rate: events > 95^{th} perc



Hit Rate: events > 95^{th} perc



Precipitation events above the 95^{th} perc:

- large majority of pixels have more than 50% of events coinciding
- bad performance in arid region again, Iceland and Eastern Europe, corresponding to regions with very poor station coverage in EOBS.

Intensity comparison

Precipitation intensity comparison

- The assessment of precipitation events co-occurrence was regardless of respective percentile values.
- The two following sections present the comparison of positive precipitation distribution.
- As natural hazards related to extreme precipitation cause consequential casualties and damage, information about rare events are important → Use of extreme value theory [Coles, 2001].
- First comparison of return levels for some non-exceedance probabilities, then comparison of the entire positive distribution.

Intensity comparison

Precipitation intensity comparison

- The assessment of precipitation events co-occurrence was regardless of respective percentile values.
- The two following sections present the comparison of positive precipitation distribution.
- As natural hazards related to extreme precipitation cause consequential casualties and damage, information about rare events are important → Use of extreme value theory [Coles, 2001].
- First comparison of return levels for some non-exceedance probabilities, then comparison of the entire positive distribution.

Intensity comparison

Precipitation intensity comparison

- The assessment of precipitation events co-occurrence was regardless of respective percentile values.
- The two following sections present the comparison of positive precipitation distribution.
- As natural hazards related to extreme precipitation cause consequential casualties and damage, information about rare events are important → Use of extreme value theory [Coles, 2001].
- First comparison of return levels for some non-exceedance probabilities, then comparison of the entire positive distribution.

Intensity comparison

Precipitation intensity comparison

- The assessment of precipitation events co-occurrence was regardless of respective percentile values.
- The two following sections present the comparison of positive precipitation distribution.
- As natural hazards related to extreme precipitation cause consequential casualties and damage, information about rare events are important $\rightarrow$ Use of extreme value theory [Coles, 2001].
- First comparison of return levels for some non-exceedance probabilities, then comparison of the entire positive distribution.

Table of contents

Extended Generalized Pareto Distribution

- Extended Generalized Pareto Distribution (EGPD) =distribution function for the whole distribution of positive precipitation [Tencaliec et al., 2019].
- Extreme value theory for both upper and lower tail of EGPD, but without a threshold selection
- The CDF of a EGPD model member is expressed as $(x > 0)$:

$$F(x) = G\{H_\xi(x/\sigma)\}$$

where $G \; : \; [0;1] \longrightarrow [0;1]$ continuous bijective function, and $H_\xi(\bullet/\sigma)$ = CDF of a GPD:

$$H_\xi(z/\sigma) = \begin{cases} 1 - (1 + \xi\frac{z}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-z\sigma} & \text{otherwise.} \end{cases}$$

σ scale parameter, ξ shape parameter and $\forall a \in \mathbb{R}$, $a_+ = max(a, 0)$.

Extended Generalized Pareto Distribution

- Extended Generalized Pareto Distribution (EGPD) =distribution function for the whole distribution of positive precipitation [Tencaliec et al., 2019].
- Extreme value theory for both upper and lower tail of EGPD, but without a threshold selection
- The CDF of a EGPD model member is expressed as ($x > 0$):

$$F(x) = G\{H_\xi(x/\sigma)\}$$

where $G \; : \; [0;1] \longrightarrow [0;1]$ continuous bijective function,
and $H_\xi(\bullet/\sigma) =$ CDF of a GPD:

$$H_\xi(z/\sigma) = \begin{cases} 1 - (1 + \xi\frac{z}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-z\sigma} & \text{otherwise.} \end{cases}$$

σ scale parameter, ξ shape parameter and $\forall a \in \mathbb{R}$, $a_+ = max(a, 0)$.

Extended Generalized Pareto Distribution

- Extended Generalized Pareto Distribution (EGPD) =distribution function for the whole distribution of positive precipitation [Tencaliec et al., 2019].
- Extreme value theory for both upper and lower tail of EGPD, but without a threshold selection
- The CDF of a EGPD model member is expressed as ($x > 0$):

$$F(x) = G\{H_\xi(x/\sigma)\}$$

where $G : [0;1] \longrightarrow [0;1]$ continuous bijective function,
and $H_\xi(\bullet/\sigma) =$ CDF of a GPD:

$$H_\xi(z/\sigma) = \begin{cases} 1 - (1 + \xi\frac{z}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-z\sigma} & \text{otherwise.} \end{cases}$$

σ scale parameter, ξ shape parameter and $\forall a \in \mathbb{R}$, $a_+ = max(a, 0)$.

Extended Generalized Pareto Distribution

- Extended Generalized Pareto Distribution (EGPD) =distribution function for the whole distribution of positive precipitation [Tencaliec et al., 2019].
- Extreme value theory for both upper and lower tail of EGPD, but without a threshold selection
- The CDF of a EGPD model member is expressed as $(x > 0)$:

$$F(x) = G\{H_\xi(x/\sigma)\}$$

where $G \ : \ [0;1] \longrightarrow [0;1]$ continuous bijective function,
and $H_\xi(\bullet/\sigma)$ = CDF of a GPD:

$$H_\xi(z/\sigma) = \begin{cases} 1 - (1+\xi\frac{z}{\sigma})_+^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-z\sigma} & \text{otherwise.} \end{cases}$$

σ scale parameter, ξ shape parameter and $\forall a \in \mathbb{R}$, $a_+ = max(a, 0)$.

Extended Generalized Pareto Distribution

- The CDF of EGPD : $F(x) = G\{H_\xi(x/\sigma)\}$ $(x > 0)$

  $H_\xi(\bullet/\sigma) =$ CDF of a GPD.

  G : $[0;1] \longrightarrow [0;1]$ continuous bijective function

- For GPD behavior on upper and lower tail, two constraints on G:
  1. $\lim_{u \downarrow 0} \frac{1-G(1-u)}{u}$ is finite and positive
  2. $\lim_{u \downarrow 0} \frac{G(u)}{u^s}$ is finite and positive for some $s > 0$.

- Bernstein polynomial basis approximation to approach G

[Tencaliec et al., 2019]

Extended Generalized Pareto Distribution

- The CDF of EGPD : $F(x) = G\{H_\xi(x/\sigma)\}$ $(x > 0)$

  $H_\xi(\bullet/\sigma)$ = CDF of a GPD.

  G : $[0;1] \longrightarrow [0;1]$ continuous bijective function

- For GPD behavior on upper and lower tail, two constraints on G:
  1. $\lim_{u \downarrow 0} \frac{1-G(1-u)}{u}$ is finite and positive
  2. $\lim_{u \downarrow 0} \frac{G(u)}{u^s}$ is finite and positive for some $s > 0$.

- Bernstein polynomial basis approximation to approach G

[Tencaliec et al., 2019]

Extended Generalized Pareto Distribution

- The CDF of EGPD : $F(x) = G\{H_\xi(x/\sigma)\}$ $(x > 0)$

  $H_\xi(\bullet/\sigma) =$ CDF of a GPD.

  G : $[0;1] \longrightarrow [0;1]$ continuous bijective function

- For GPD behavior on upper and lower tail, two constraints on G:
  1. $\lim\limits_{u \downarrow 0} \frac{1-G(1-u)}{u}$ is finite and positive
  2. $\lim\limits_{u \downarrow 0} \frac{G(u)}{u^s}$ is finite and positive for some $s > 0$.

- Bernstein polynomial basis approximation to approach G

[Tencaliec et al., 2019]

Non parametric bootstrap procedure

- To compare magnitude of return levels in ERA5 and EOBS: computation of confidence intervals of return levels.
- A non parametric bootstrap procedure of size 200 is performed to obtain confidence intervals on return levels associated with probability 0.3, 0.5, 0.75, 0.9 and 0.95, for each season.
- Gridpoints with time series of positive seasonal precipitation shorter than 500 days are discarded.

Non parametric bootstrap procedure

- To compare magnitude of return levels in ERA5 and EOBS: computation of confidence intervals of return levels.
- A non parametric bootstrap procedure of size 200 is performed to obtain confidence intervals on return levels associated with probability 0.3, 0.5, 0.75, 0.9 and 0.95, for each season.
- Gridpoints with time series of positive seasonal precipitation shorter than 500 days are discarded.

Non parametric bootstrap procedure

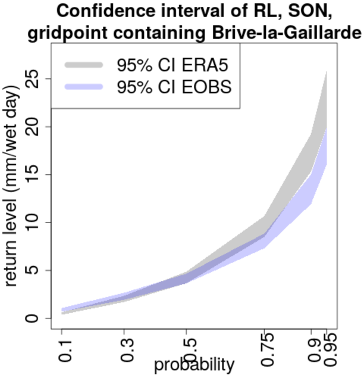
- To compare magnitude of return levels in ERA5 and EOBS: computation of confidence intervals of return levels.
- A non parametric bootstrap procedure of size 200 is performed to obtain confidence intervals on return levels associated with probability 0.3, 0.5, 0.75, 0.9 and 0.95, for each season.
- Gridpoints with time series of positive seasonal precipitation shorter than 500 days are discarded.

Overlap of confidence intervals (c.i.)

We start the result presentation with an example of grid point experiencing the 3 possible configurations of relative position of confidence intervals between ERA5 and EOBS, for different non-exceedance probabilities.

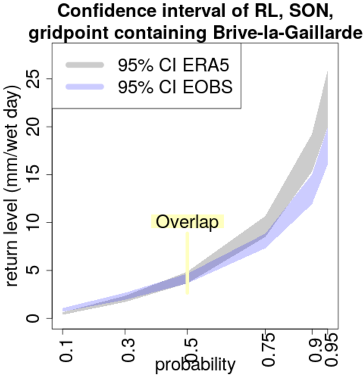
Overlap of confidence intervals (c.i.)

We start the result presentation with an example of grid point experiencing the 3 possible configurations of relative position of confidence intervals between ERA5 and EOBS, for different non-exceedance probabilities.



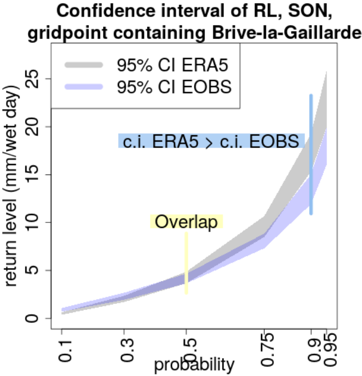
Overlap of confidence intervals (c.i.)

We start the result presentation with an example of grid point experiencing the 3 possible configurations of relative position of confidence intervals between ERA5 and EOBS, for different non-exceedance probabilities.



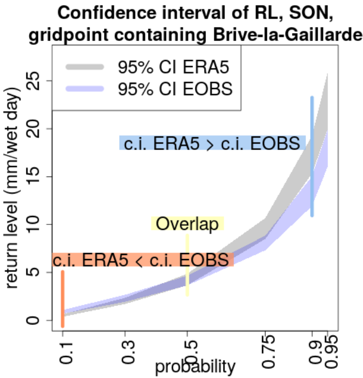
Overlap of confidence intervals (c.i.)

We start the result presentation with an example of grid point experiencing the 3 possible configurations of relative position of confidence intervals between ERA5 and EOBS, for different non-exceedance probabilities.

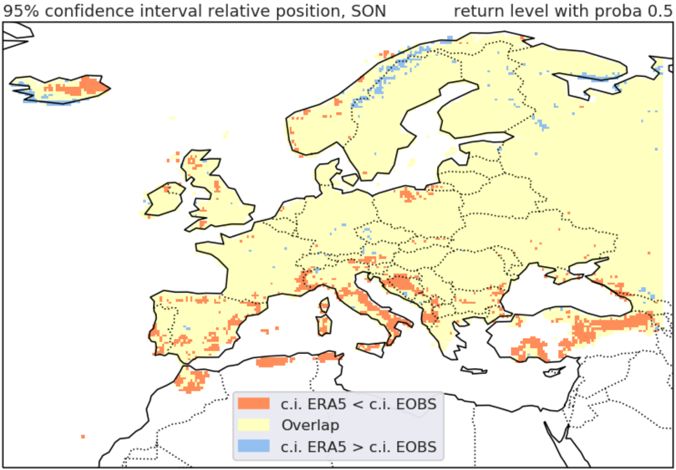


Overlap of confidence intervals (c.i.)

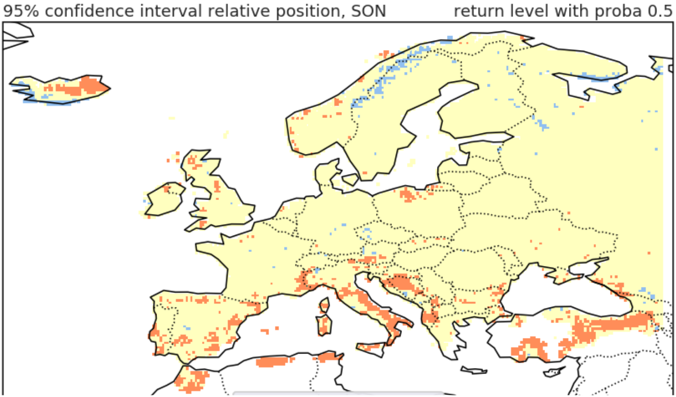
We start the result presentation with an example of grid point experiencing the 3 possible configurations of relative position of confidence intervals between ERA5 and EOBS, for different non-exceedance probabilities.



Confidence intervals (c.i.), return level with proba 0.5



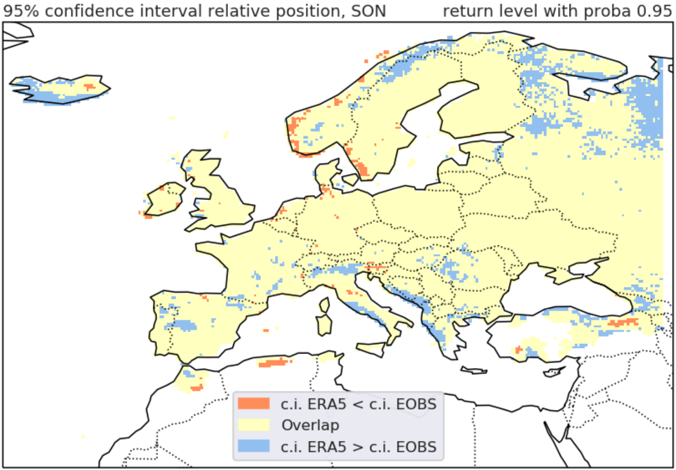
Confidence intervals (c.i.), return level with proba 0.5



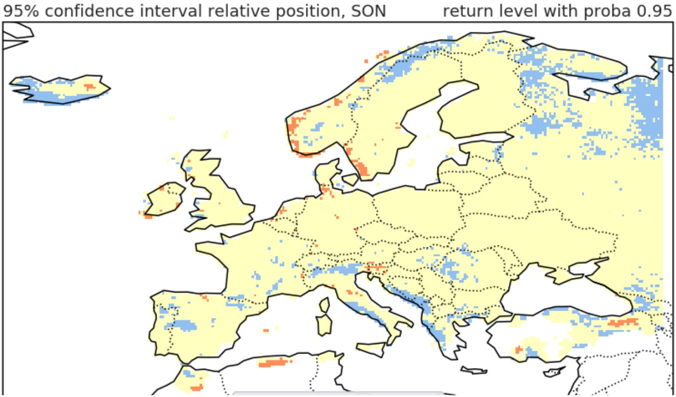
Relative position of c.i., return level with non-exceedance probability 0.5:

- c.i. are overlapping for almost all pixels;
- Around the Mediterranean sea and in Iceland, ERA5 underestimates return levels relatively to EOBS

Confidence intervals (c.i.), return level with proba 0.95



Confidence intervals (c.i.), return level with proba 0.95



Relative position of c.i., return level with non-exceedance probability 0.95:

- c.i. are overlapping for almost all pixels;
- In lot of regions, EOBS underestimates return levels relatively to ERA5. Can be due to underestimation of precipitation extreme because of the spatial aggregation in EOBS [Haylock et al., 2008].

Table of contents

Intensity comparison

Test on the whole distribution

- Only finite number of probabilities have been compared with confidence intervals.
- Strength of EGPD = fitting of the entire distribution.
  To makes use of this property, we conduct here a test to compare directly the whole precipitation distribution in ERA5 and EOBS.
  This test is based on the Kullback-Leibler divergence.

Intensity comparison

Test on the whole distribution

- Only finite number of probabilities have been compared with confidence intervals.
- Strength of EGPD = fitting of the entire distribution.
  To makes use of this property, we conduct here a test to compare directly the whole precipitation distribution in ERA5 and EOBS.
  This test is based on the Kullback-Leibler divergence.

Kullback-Leibler divergence and hypothesis test

- Kullback-Leibler divergence between two distributions F and G (density f and g) defined by:

$$K(f,g) = \mathbb{E}_f \left[\log \left\{ \frac{f(\mathbf{X})}{g(\mathbf{X})} \right\} \right] + \mathbb{E}_g \left[\log \left\{ \frac{g(\mathbf{Y})}{f(\mathbf{Y})} \right\} \right]$$

with $\mathbf{X}$ and $\mathbf{Y}$ random variables following respectively F and G.

- EGPD fitting to get $\hat{f}$ and $\hat{g}$. Then computation of $\hat{K}(X,Y)$:

$$\hat{K}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{\hat{f}(X_i)}{\hat{g}(X_i)} + \frac{1}{m}\sum_{j=1}^{m} \log \frac{\hat{g}(Y_j)}{\hat{f}(Y_j)}$$

- Hypothesis test with null hypothesis:

H_0: X and Y have the same distribution, i.e. $f = g$.

Kullback-Leibler divergence and hypothesis test

- Kullback-Leibler divergence between two distributions F and G (density f and g) defined by:

$$K(f,g) = \mathbb{E}_f \left[\log \left\{ \frac{f(\mathbf{X})}{g(\mathbf{X})} \right\} \right] + \mathbb{E}_g \left[\log \left\{ \frac{g(\mathbf{Y})}{f(\mathbf{Y})} \right\} \right]$$

with $\mathbf{X}$ and $\mathbf{Y}$ random variables following respectively F and G.

- EGPD fitting to get $\hat{f}$ and $\hat{g}$. Then computation of $\hat{K}(X, Y)$:

$$\hat{K}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{\hat{f}(X_i)}{\hat{g}(X_i)} + \frac{1}{m}\sum_{j=1}^{m} \log \frac{\hat{g}(Y_j)}{\hat{f}(Y_j)}$$

- Hypothesis test with null hypothesis:

H_0: X and Y have the same distribution, i.e. $f = g$.

Kullback-Leibler divergence and hypothesis test

- Kullback-Leibler divergence between two distributions F and G (density f and g) defined by:

$$K(f,g) = \mathbb{E}_f \left[\log \left\{ \frac{f(\mathbf{X})}{g(\mathbf{X})} \right\} \right] + \mathbb{E}_g \left[\log \left\{ \frac{g(\mathbf{Y})}{f(\mathbf{Y})} \right\} \right]$$

with $\mathbf{X}$ and $\mathbf{Y}$ random variables following respectively F and G.

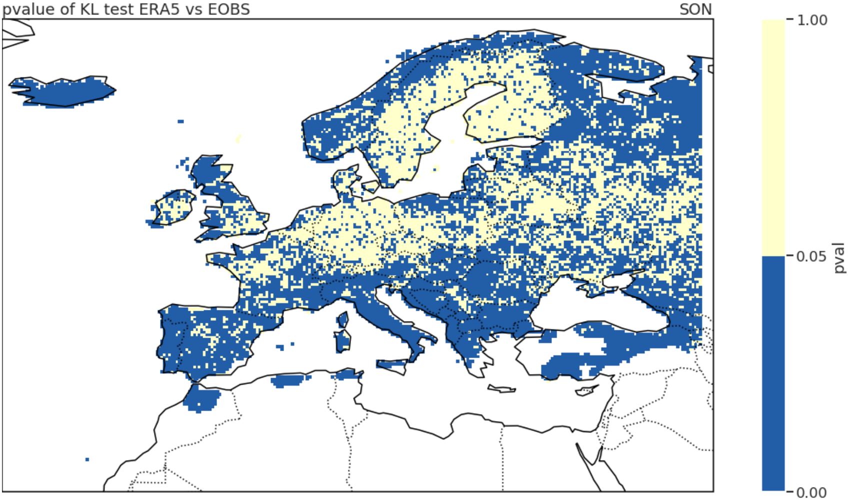
- EGPD fitting to get $\hat{f}$ and $\hat{g}$. Then computation of $\hat{K}(X,Y)$:

$$\hat{K}(X,Y) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{\hat{f}(X_i)}{\hat{g}(X_i)} + \frac{1}{m} \sum_{j=1}^{m} \log \frac{\hat{g}(Y_j)}{\hat{f}(Y_j)}$$

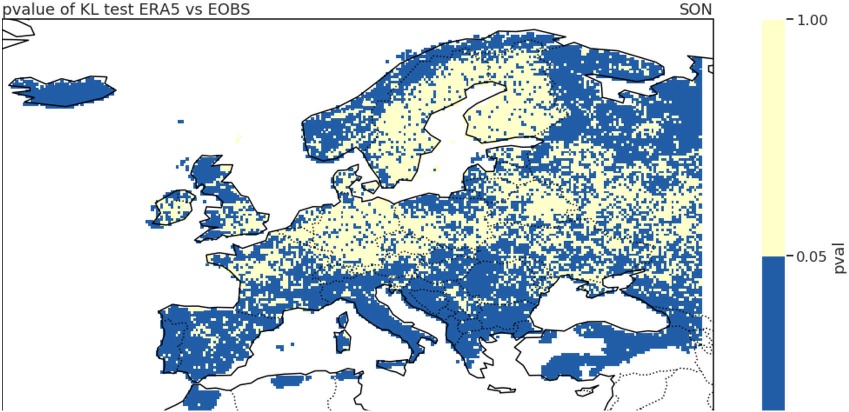
- Hypothesis test with null hypothesis:

H_0: X and Y have the same distribution, i.e. $f = g$.

pvalue of the Kullback-Leibler divergence test

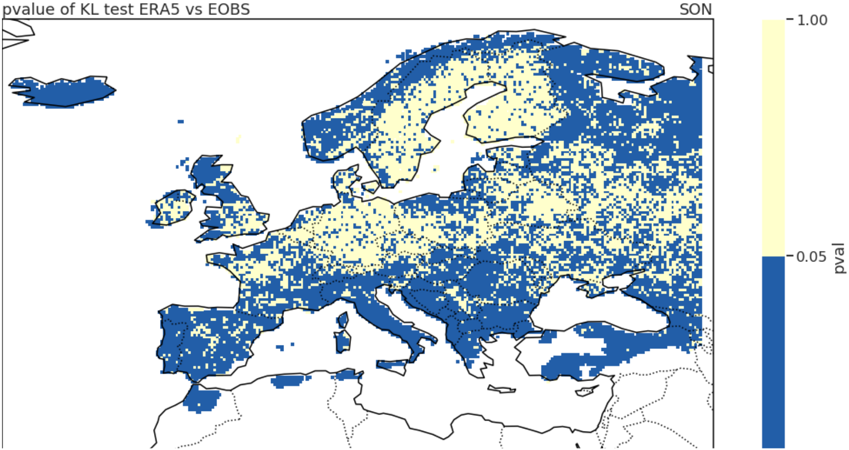


pvalue of the Kullback-Leibler divergence test



- A gridpoint with a pvalue > 0.05 means that the null-hypothesis could not be rejected: the positive precipitation distribution at this location can be considered as identical in ERA5 and EOBS.
- This test is rather strict: the region where null hypothesis could not be rejected is rather large. This region mainly corresponds dense station coverage in EOBS, implying that **the 2 datasets agree where EOBS is most likely to be accurate**

pvalue of the Kullback-Leibler divergence test



- One can notice that some patterns follow borders (Sweden, Finland). Can be an artefact of the fact data assimilated in EOBS are provided by national meteorological centers, potential discrepancy ?

Table of contents

Verification table

Precipitation intensity	SON	
	Intensity	Co-occurrence
Low $p = 0.3$	82%	95%
Median $p = 0.5$	90%	90%
Moderate $p = 0.75$	90%	80%
High $p = 0.9$	87%	68%
Extreme $p = 0.95$	87%	59%
whole distrib. (KL test)	39% $K_{mean} = 0.106$	

Results for september-october-november (SON):

- Good agreement on intensity of precipitation, for all the return levels studied here;
- Decreasing co-occurrence score with increasing rarity of event, but over all good agreement;
- KL test could not be rejected for about 40% of grid points, i.e. almost 40% of grid points identical distribution of positive precipitation, according to the test. This score can be considered as high, because the KL test is quite strict.

Signification of % for:

- *Intensity:* percentage of gridpoints with overlap of confidence intervals;
- *Co-occurrence:* mean value of hit rate;
- *KL test:* percentage of gridpoints with pvalue ≥ 0.05.

K_{mean}=mean value of the Kullback-Leibler divergence.

Verification table

Precipitation intensity	SON	
	Intensity	Co-occurrence
Low $p = 0.3$	82%	95%
Median $p = 0.5$	90%	90%
Moderate $p = 0.75$	90%	80%
High $p = 0.9$	87%	68%
Extreme $p = 0.95$	87%	59%
whole distrib. (KL test)	39% $K_{mean} = 0.106$	

Results for september-october-november (SON):

- Good agreement on intensity of precipitation, for all the return levels studied here;
- Decreasing co-occurrence score with increasing rarity of event, but over all good agreement;
- KL test could not be rejected for about 40% of grid points, i.e. almost 40% of grid points identical distribution of positive precipitation, according to the test. This score can be considered as high, because the KL test is quite strict.

Signification of % for:

- *Intensity:* percentage of gridpoints with overlap of confidence intervals;
- *Co-occurrence:* mean value of hit rate;
- *KL test:* percentage of gridpoints with pvalue ≥ 0.05.

K_{mean}=mean value of the Kullback-Leibler divergence.

Verification table

Precipitation intensity	SON	
	Intensity	Co-occurrence
Low $p = 0.3$	82%	95%
Median $p = 0.5$	90%	90%
Moderate $p = 0.75$	90%	80%
High $p = 0.9$	87%	68%
Extreme $p = 0.95$	87%	59%
whole distrib. (KL test)	39% $K_{mean} = 0.106$	

Results for september-october-november (SON):

- Good agreement on intensity of precipitation, for all the return levels studied here;
- Decreasing co-occurrence score with increasing rarity of event, but over all good agreement;
- KL test could not be rejected for about 40% of grid points, i.e. almost 40% of grid points identical distribution of positive precipitation, according to the test. This score can be considered as high, because the KL test is quite strict.

Signification of % for:

- *Intensity:* percentage of gridpoints with overlap of confidence intervals;
- *Co-occurrence:* mean value of hit rate;
- *KL test:* percentage of gridpoints with pvalue ≥ 0.05.

K_{mean}=mean value of the Kullback-Leibler divergence.

Verification table

Precipitation intensity	SON	
	Intensity	Co-occurrence
Low $p = 0.3$	82%	95%
Median $p = 0.5$	90%	90%
Moderate $p = 0.75$	90%	80%
High $p = 0.9$	87%	68%
Extreme $p = 0.95$	87%	59%
whole distrib. (KL test)	39% $K_{mean} = 0.106$	

Results for september-october-november (SON):

- Good agreement on intensity of precipitation, for all the return levels studied here;
- Decreasing co-occurrence score with increasing rarity of event, but over all good agreement;
- KL test could not be rejected for about 40% of grid points, i.e. almost 40% of grid points identical distribution of positive precipitation, according to the test. This score can be considered as high, because the KL test is quite strict.

Signification of % for:

- *Intensity:* percentage of gridpoints with overlap of confidence intervals;
- *Co-occurrence:* mean value of hit rate;
- *KL test:* percentage of gridpoints with pvalue ≥ 0.05.

K_{mean}=mean value of the Kullback-Leibler divergence.

Verification table

Precipitation intensity	SON		DJF		MAM		JJA	
	Intensity	Co-occurrence	Intensity	Co-occurrence	Intensity	Co-occurrence	Intensity	Co-occurrence
Low $p = 0.3$	82%	95%	82%	91%	81%	99%	39%	100%
Median $p = 0.5$	90%	90%	81%	87%	90%	94%	75%	95%
Moderate $p = 0.75$	90%	80%	72%	77%	89%	84%	93%	84%
High $p = 0.9$	87%	68%	69%	65%	85%	71%	93%	70%
Extreme $p = 0.95$	87%	59%	73%	57%	87%	63%	94%	61%
whole distrib. (KL test)	39% $K_{mean} = 0.106$		29% $K_{mean} = 0.138$		34% $K_{mean} = 0.112$		10% $K_{mean} = 0.167$	

Signification of % for:

- *Intensity:* percentage of gridpoints with overlap of confidence intervals;
- *Co-occurrence:* mean value of hit rate;
- *KL test:* percentage of gridpoints with pvalue ≥ 0.05.

K_{mean}=mean value of the Kullback-Leibler divergence.

Table of contents

Conclusion:

For the four seasons:

- Good event co-occurrence signal, but decreasing performance with increasing relative event intensity.
- Intensity of precipitation $> 1mm$: overall good agreement of return levels on positive daily precipitation (overlap of confidence interval for a large majority of pixels, except low return levels in summer, which affect the result of the KL test)

Conclusion:

For the four seasons:

- Good event co-occurrence signal, but decreasing performance with increasing relative event intensity.
- Intensity of precipitation $> 1mm$: overall good agreement of return levels on positive daily precipitation (overlap of confidence interval for a large majority of pixels, except low return levels in summer, which affect the result of the KL test)

Conclusion:

- The strengths of ERA5's daily precipitation are the expertise from latest atmospheric models, the regular spatial resolution and the long time period availability.
- The users's need should guide the choice of the dataset to use for daily precipitation. For example, the reanalysis dataset should be preferred to observational dataset for regions where observational datasets have a limited coverage.
  E.g. given that ERA5 and EOBS have a good agreement in regions with dense station coverage in EOBS, ERA5 can be preferred to EOBS where the station coverage is poor in EOBS.

Conclusion:

- The strengths of ERA5's daily precipitation are the expertise from latest atmospheric models, the regular spatial resolution and the long time period availability.
- The users's need should guide the choice of the dataset to use for daily precipitation. For example, the reanalysis dataset should be preferred to observational dataset for regions where observational datasets have a limited coverage.

  E.g. given that ERA5 and EOBS have a good agreement in regions with dense station coverage in EOBS, ERA5 can be preferred to EOBS where the station coverage is poor in EOBS.

Conclusion:

- The strengths of ERA5's daily precipitation are the expertise from latest atmospheric models, the regular spatial resolution and the long time period availability.
- The users's need should guide the choice of the dataset to use for daily precipitation. For example, the reanalysis dataset should be preferred to observational dataset for regions where observational datasets have a limited coverage.
  E.g. given that ERA5 and EOBS have a good agreement in regions with dense station coverage in EOBS, ERA5 can be preferred to EOBS where the station coverage is poor in EOBS.

References

Thank you for your attention

Copernicus Climate Change Service C3S. Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate. https://cds.climate.copernicus.eu/cdsapp#!/home, 2017. Accessed: 2020-03-02.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, 2001. ISBN 1852334592. URL https://books.google.fr/books?id=SonbBwAAQBAJ.

Copernicus ECMWF. What is reanalysis? https://www.youtube.com/watch?v=FAGobvUGl24, 2019. Accessed: 2020-03-04.

M. R. Haylock, N. Hofstra, A. M.G. Klein Tank, E. J. Klok, P. D. Jones, and M. New. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research Atmospheres*, 113(20), 2008. ISSN 01480227. doi: 10.1029/2008JD010201.

K Hennermann. Era5: data description. https://confluence.ecmwf.int/pages/viewpage.action?pageId=85402030, 2020. Accessed: 2020-03-06.

P. Tencaliec, A. C. Favre, P. Naveau, C. Prieur, and G. Nicolet. Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, (May):1–20, 2019. ISSN 1099095X. doi: 10.1002/env.2582.

Remark: link with χ parameter

For a given grid point and for $d = ERA5$ or $EOBS$, let X_d be the random variable (r.a.) modeling daily positive precipitation in d, and F_d its cumulative distribution function.

The hit rate $\hat{\mathcal{H}}$ we study here can be seen as the empirical value of
$\mathcal{H} = \mathbb{P}\left[F_{EOBS}(X_{EOBS}) > q \mid F_{ERA5}(X_{ERA5}) > q\right]$, for $q \in \{0.75,\ 0.9,\ 0.95,\ 0.99\}$, i.e.:
$\mathcal{H} = \frac{\mathbb{P}[F_{EOBS}(X_{EOBS}) > q, F_{ERA5}(X_{ERA5}) > q]}{\mathbb{P}[F_{ERA5}(X_{ERA5}) > q]} = \frac{\mathbb{P}[F_{EOBS}(X_{EOBS}) > q, F_{ERA5}(X_{ERA5}) > q]}{1-q}$

Making $q \to 1$, we obtain the defintion of the tail dependance coefficient, that quantify the dependance between two r.a.s:

$$\chi = \lim_{q \to 1} \frac{\mathbb{P}\left[F_1(X_1) > q, F_2(X_2) > q\right]}{1-q}$$

If $\chi = 0$, then X_1 and X_2 are asymptotically independent; If $\chi > 0$, then extreme values of X_1 and X_2 are correlated.

For 40 years of data, between 1 and 20 days above 99th wet percentiles. Out of about 3600/3700 days in total for each season, this can be considered as $q \to 1$.