ERA4CS

An ERA-NET
initiated by JPI Climate

Project: 690462 AQUACLEW

Full project title:

Advancing QUAlity of CLimate services for European Water

Milestone: M2.1

Inventory of indicators and metrics for model evaluation

EVIDENCE FOR MILESTONE COMPLETION

Due date of milestone: 31/07/2018
Date of completion of milestone: 21/12/2018

Title	M2.1: Inventory of indicators and metrics for model evaluation
Creator	Paul Royer Gaspard and Vazken Andréassian (IRSTEA)
Editor	Christiana Photiadou and Berit Arheimer (SMHI)
Brief Description	This milestone contains and inventory of indicators and metrics for model evaluation that will be used in WP2 Advancing Data Quality
Publisher	AQUACLEW Consortium
Contributors	Stefan Achleitner; Andreas Huber (UIBK), Asunción Baquerizo; Manuel Cobos Budia; Agustin Millares; Andrea Lira Loarca (UGR), Herbert Formayer; David Leidinger (BOKU), Peter Berg; Katharina Klehmet; Lorna Little (SMHI), Johannes Lückenkötter; Jörg Peter Schmitt (TUDO), María-José Pérez-Palazón; Rafael Pimentel; María J. Polo (UCO), Jens Christian Refsgaard; Ernesto Pasten Zapata; Torben Sonnenborg: Hans Jørgen Henriksen (GEUS), Guillaume Thirel; Maria-Helena Ramos; Anthony Lemoine (IRSTEA)
Type (Deliverable/Milestone)	Milestone
Format	Report
Creation date	December 2018
Version number	V1.1
Version date	December 2018
Last modified by	Christiana Photiadou (SMHI)
Rights	Copyright "AQUACLEW Consortium". During the drafting process, access is generally limited to the AQUACLEW Partners.
Audience	☐ internal ☒ public ☐ restricted, access granted to: EU Commission
Action requested	☐ to be revised by Partners involved in the preparation of the deliverable ☒ for approval of the WP Manager ☐ for approval of the Internal Reviewer (if required) ☒ for approval of the Project Co-ordinator
Deadline for approval	31/12/2018

Version	**Date**	**Modified by**	**Comments**
v1	17/12/2018	Vasken Andréassian	Creation of document with description of milestone activity
v2	21/12/2018	Christiana Photiadou	Minor revisions to text content

Table of Contents

1. INTRODUCTION

1.1 The AQUACLEW project

The impacts of climate change on the hydrological cycle raises a growing concern among water managers, policymakers and end-users with respect to the evolution of water resources and risks. To face a growing demand in estimating impacts and defining adaptation strategies in various sectors, climate services are blooming across Europe, especially under the umbrella of the Copernicus Climate Change Service (C3S, https://climate.copernicus.eu/) of the European commission. The AQUACLEW European project (2017-2020, www.jpi-climate.eu/nl/25223436-AQUACLEW.html) contributes to improve these services. The overall goal of AQUACLEW is to use innovative research techniques and integrated co-development with users to advance the quality and usability of several climate services for a number of water related sectors. The current work to develop these services indicates that the following crucial improvements are required to increase user adoption and satisfaction: (a) indicators and resolution of the indicators given by the service needs to address a wider range of user needs; (b) large-scale climate service data should be more reliable at the local decision scale and (c) guidance and visualization tools in the services should better reflect the wider range of user needs.

1.2 Objectives of this report

To be useful, climate services must have demonstrated efficiency and robustness: efficiency can be measured by a variety of well-known metrics, but robustness is more difficult to define. To attract and mobilize the attention of all AQUACLEW participants on issues relative to robustness, we conducted a poll under an innovative form (section 2) during our 2018 Copenhagen meeting. Then, we conducted an online inventory (again among AQUACLEW participants) to collect those indicators which were considered relevant by each group (in the limits of the case study they were working on) for assessing the quality of model runs (either climatic or hydrologic).

This report identifies for each case study a set of variables, climate change indicators and associated metrics to assess model performance. An appendix has been added as an in-depth description of the latter metrics, dedicated to users who are not entirely familiar with some parts of the modeling chain. The case studies covered the following domains: pluvial flash floods; hydropower production; fluvial and coastal interactions under Mediterranean climate conditions; agricultural production ; drought and water resource allocation for tourism, agriculture and energy sectors; urban flash floods; biodiversity decline. A more detailed description of the whereabouts of each case study may be found on the project website at http://aquaclew.eu/case-studies/.

2. ON DEFINING ROBUSTNESS

Because robustness is a concept sometimes difficult to grasp or define, we conducted a poll among AQUACLEW participants of the Copenhagen workshop under an innovative form: that of a wall poll using wool strings (hence the name of **Wooly Wall Poll**) to mark the answers of each participant. The advantage of using strings was to provide a direct visual impression concerning the agreement or disagreement of participants. The objective was (i) to attract and mobilize the attention of all AQUACLEW participants on issues relative to robustness, and (ii) to map the perception of robustness related issues within the group.

In designing the poll, we wished to put the robustness issue in a general perspective and show that robustness is not a « narrow » issue only relevant to hydrological modelling, but rather a multifaceted issue. This is why we interspersed questions which were not directly related to hydrological and climatic robustness issues.

Figure 1 presents the poll sheet before submission and Figure 2 the results.

Wooly
Wall
Poll
for AquaClew

1.
In your opinion, which one of these paintings represents best the concept of "robustness"?

2.
In your opinion, which one of these political figures represents best the concept of "robustness"?

3.
You need to make hydrological predictions in a climate change context, if you had to choose one, you would choose...

- ☐ Several CMs, several DSs, several HMs, a single set of HM parameters
- ☐ Several CMs, several DSs, a single HM, several sets of HM parameters
- ☐ Several CMs, a single DS, several HMs, several sets of HM parameters
- ☐ A single CM, several DSs, several HMs, several sets of HM parameters
- ☐ I don't know

4.
You need to travel by bike from Norrköping to Yerevan (ca 3551 km), which bike would you choose...

- ☐ The bike that you use daily to ride to work
- ☐ A brand new bike in carbon alloy that you are going to order because everybody praises it but that you will only get on the day of your departure
- ☐ The nice bike of your neighbour that you have never tested but that he praises all the time
- ☐ Your competition bike that you use only for special occasions
- ☐ I don't know

5.
You need to make hydrological predictions in a climate change context, if you had to choose one, you would choose...

- ☐ A black-box model that fits the present very well
- ☐ A physically-based model that fits the present poorly
- ☐ A model that work well over Europe and poorly in Africa
- ☐ A model that work well in Poland, Zimbabwe and Nigeria, but poorly in Latvia and Congo
- ☐ A model that work mediocrely all over the globe
- ☐ I don't know

6.
You need to calibrate a hydrological model and you are only allowed one parameter set. Which period would you choose for the calibration?

- ☐ The complete time series
- ☐ A selection of wet years
- ☐ A selection of dry years
- ☐ A random patchwork
- ☐ I don't know

7.
You need to calibrate a statistical downscaling method and you are only allowed one parameter set. Which period would you choose?

- ☐ A selection of wet years
- ☐ A selection of dry years
- ☐ A random patchwork
- ☐ I only trust dynamical downscaling methods
- ☐ I don't know

8.
In your opinion, which of these hydrologists best personifies the AQUACLEW project?

- ☐
- ☐
- ☐
- ☐

Figure 1: The *Wooly Wall Poll* as it was presented to the participants

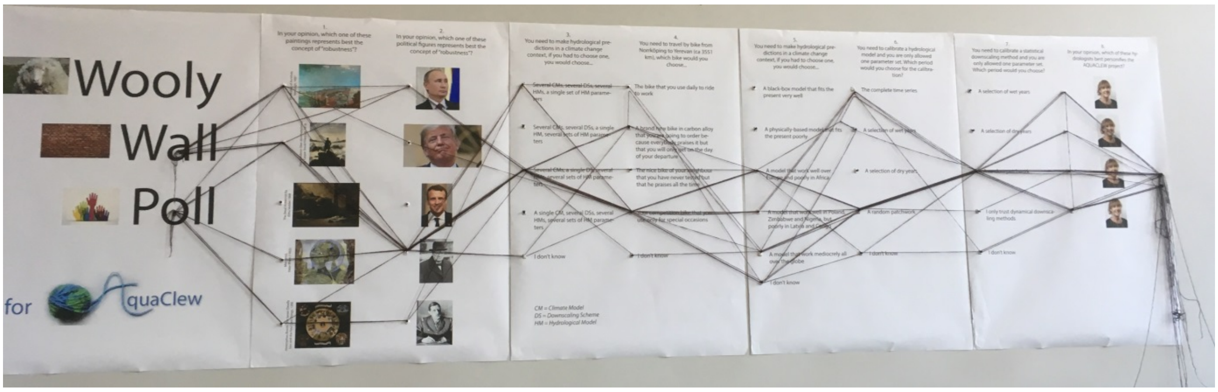

Figure 2: *Wooly Wall Poll* results

Results show that there is little agreement in general on robustness. Perhaps the only agreement was about the political personality best representing the concept (i.e. Winston Churchill). Question 6 and 7 were also relatively consensual:

- to question 6 "*You need to calibrate a hydrological model and you are only allowed one parameter set. Which period would you choose for the calibration?*", most participants answered "*A random patchwork*" showing that they consider that the largest possible variability in calibrating data is an assurance of robustness;
- to question 7 "*You need to calibrate a statistical downscaling method and you are only allowed one parameter set. Which period would you choose?*" most participants answered as for question 6 "*A random patchwork*".

Undoubtedly, the main interest of this poll was in the discussions which it attracted, and in the fact that it helped everyone realize that robustness is probably more complex that he/she thought previously.

3. INDICATORS AND METRICS RECOMMANDED FOR THE CASE STUDIES

The indicators and metrics collected during the inventory phase are summarized in the tables below. The first column is dedicated to variables (precipitation, temperature…) or computed indicators (return periods, drought indices…) that should be tested. The second column gathers either computed or graphical metrics to assess model performance on historical data. Note that:

- The testing framework has been thought to be as exhaustive as possible. Thus some variables may not be available or computable from every model outputs, or only at a higher timescale that the one suggested.
- Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs).
- Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs).

Table 1: Indicators and metrics recommended for an application concerning pluvial flash floods. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	**Associated Metrics**
Precipitation (3-hourly)	IDF curve (graph); Bias
Soil Moisture (daily or less)	RMSE; Bias
3-hr rainfall for the 2, 10, 100-year return periods	RMSE (on space); Bias
Clausius-Clapeyron relationship (slope of the liquid-vapor phase limit in the Clapeyron diagram at a given altitude)	Bias
Runoff (daily time-scale or less)	Annual cycle (graph) ; KGE[Q^2]; RMSE; annual, monthly, daily boxplots (graph)

Table 2: Indicators and metrics recommended for an application concerning drought and water resource allocation for tourism, agriculture and energy sectors. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	**Associated Metrics**
Precipitation (daily)	Bias; RMSE; Annual cycle (graph); Interannual/decadal variations (graph); MKT; Seasonal/decadal boxplots (graph)
Temperature (daily)	Bias; RMSE; Annual cycle (graph); Interannual/decadal variations (graph); MKT; Seasonal/decadal boxplots (graph)
Standardized Precipitation Evaporation index (monthly)	Bias
Runoff (daily)	KGE; NSE; RE; Annual cycle (graph); Annual, monthly, daily boxplots (graph)
Snow Water Equivalent (daily)	Bias; RMSE; Annual cycle (graph); Interannual/decadal variations (graph); MKT on yearly averages; Seasonal boxplots (graph)
Snow Cover Area (daily)	RE

Table 3: Indicators and metrics recommended for an application concerning hydropower production. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue

cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	Associated Metrics
Precipitation (daily, monthly)	Bias; RMSE; CRPS; Annual cycle (graph); Interannual variations (graph); Monthly/annual boxplots (graph)
Temperature (daily, monthly)	Bias; RMSE; CRPS; Annual cycle (graph); Interannual variations (graph); Monthly/annual boxplots (graph)
Daily or weekly rainfall for the 2, 10, 100-year return periods	RMSE (on space); Bias
Standardized Precipitation Evaporation index (monthly)	Bias
North Atlantic Oscillation index (monthly)	RMSE
Runoff (daily, monthly)	KGE; RMSE; Biases (mean, std); CRPS; Annual cycle (graph); Interannual variations (graph)
Snow Water Equivalent (daily)	Biases (annual mean & max); Annual cycle (graph); Monthly boxplots (graph)

Table 4: Indicators and metrics recommended for an application concerning agricultural production. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	**Associated Metrics**
Precipitation (daily, monthly)	Biases (mean, std, max-min, 95th & 99th percentile); Monthly boxplots
Temperature (daily, monthly)	Bias; Annual cycle (graph); Monthly/annual boxplots (graph)
Potential Evapotranspiration (daily, monthly)	Biases (mean, std, max-min);
Standardized Precipitation Evaporation index (monthly)	Bias
Monthly max 1-day precipitation	Bias
Monthly max consecutive 5-day precipitation	Bias
Simple Daily Intensity index	Bias
Annual number of heavy precipitation (> 10,20 mm)	Bias
Maximum number of consecutive dry days (< 1 mm)	Bias
Maximum number of consecutive wet days (> 1 mm)	Bias
Runoff (daily, monthly)	NSE; Biases (monthly & annual means, interquantile, annual std); Flow Statistic Score;
Number of days exceeding low/high daily runoff thresholds	Bias
Groundwater depth (daily)	RMSE; Biases (seasonal means, interquantile, annual std)
Number of days exceeding low/high daily groundwater depth thresholds	Bias
Soil Water Content (daily)	Biases (monthly means, interquantile)
Number of days exceeding low/high daily soil water content thresholds	Bias

Table 5: Indicators and metrics recommended for an application concerning biodiversity decline. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	Associated Metrics
Precipitation (daily)	Bias; RMSE; Annual cycle (graph); Interannual/decadal variations (graph); Seasonal/decadal boxplots (graph)
Temperature (daily)	Bias; RMSE; Annual cycle (graph); Interannual/decadal variations (graph); Seasonal/decadal boxplots (graph)
Temperature min & max (daily)	Bias; RMSE; Interannual/decadal variations (graph);
Snow Cover Area (daily)	RMSE; Bias
Runoff (daily, monthly)	KGE[1/Q]; RMSE; Annual cycle (graph);
River temperature (weekly or less)	Biases (annual mean, annual min & max); Annual cycle (graph); Monthly boxplots (graph)
Snow Cover Area (daily)	RMSE; Bias

Table 6: Indicators and metrics recommended for an application concerning fluvial and coastal interactions. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	Associated Metrics
Wind Speed (6-hourly)	Bias; RMSE; ADT on CDFs
Seal Level Pressure (6-hourly)	Bias; RMSE; ADT on CDFs
Significant Wave Height (6-hourly)	Bias; RMSE; ADT on CDFs
Wave Period (6-hourly)	Bias; RMSE; ADT on CDFs
Seal Level (6-hourly)	Bias; RMSE; ADT on CDFs
100-, 500-year significant wave height return period	Bias; ADT on CDFs of extremes
Annual max of 6-h wave height	Bias; RMSE; Interannual/decadal variations (graph)
Annual max of 6-h sea level	Bias; RMSE; ADT on CDFs
Annual max of 6-h storm surge level	Bias; RMSE; ADT on CDFs

Table 7: Indicators and metrics recommended for an application concerning urban flash floods. Red cells indicate that the variables should be computable from Global Climate Models (GCMs) or Regional Climate Models (RCMs). Blue cells indicate that the variables should come either from Hydrological Models (HMs), Wave Models (WMs) or Hydro-Morphological Models (HMMs)

Climate Change Indicators	Associated Metrics
Precipitation (3-hourly)	IDF curve (graph); Bias
Soil Moisture (daily or less)	RMSE; Bias
2,10,100-year max 3-h rainfall return period	RMSE (on space); Bias
Clausius-Clapeyron relationship	Bias
Runoff (daily or less)	Annual cycle (graph); KGE[Q^2]; RMSE; Annual, monthly, daily boxplots (graph)

4. CONCLUSION

For each case study of the AQUACLEW project, we have presented a set of climate change indicators and associated metrics permitting to assess the performance of climate models and hydrological models.

APPENDIX

Common metrics

- **Bias**

Bias is computed as the ratio of the modeled variable to the observed one. If not specified, it corresponds to ratio of the averaged variable.

$$\mathrm{BIAS} = \frac{\sum_{i=1}^{N} \widehat{X_i}}{\sum_{i=1}^{N} X_i}$$

- **AE (Absolute Error)**

Absolute error is computed as the absolute difference of the simulated and the observed values.

$$\mathrm{AE} = \frac{1}{N}\sum_{i=1}^{N} \left|X_i - \widehat{X_i}\right|$$

- **RE (Relative Error)**

Relative error is computed as the ratio of the absolute error and the observed value.

$$\mathrm{RE} = \left|\frac{\sum_{i=1}^{N} X_i - \widehat{X_i}}{\sum_{i=1}^{N} X_i}\right|$$

- **RMSE (Root-Mean-Square Error)**

Also known as Root-Mean-Square Error, RMSE applies to time series or to spatial grids of an aggregated indicator.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N * M}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(X_{i,j} - \widehat{X_{i,j}}\right)^2}$$

- **CRPS (Continuous Ranked Probability Score)**

It measures the distance between the modeled and the observed probability density function of a variable. It is further described by Hersbach (2000).

$$\mathrm{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left(F_i(x) - \widehat{F_i}(x)\right)^2 dx$$

Metrics more specific to the hydrological community

- **NSE (Nash-Sutcliffe Efficiency)**

NSE is a score to test models’ ability to simulate runoff based on the comparison to the “average” model, which only output is the averaged observed runoff. It is further described by Nash & Sutcliffe (1970).

$$\mathrm{NSE} = \frac{\sum_{i=1}^{N}\left(X_i - \widehat{X}_\iota\right)^2}{\sum_{i=1}^{N}(X_i - \overline{X}_\iota)^2}$$

- **KGE (Kling-Gupta Efficiency)**

KGE is a score to test model ability to simulate runoff based on the computation of three statistical distances to observation (correlation, average bias, standard deviation bias). It is further described by Kling & Gupta (2009). KGE[1/Q] (respectively Q^2) designates the KGE computed on inverted (squared) runoff, excluding zeroes.

$$\mathrm{KGE} = 1 - \sqrt{\left(cor\left(X_i, \widehat{X}_\iota\right) - 1\right)^2 + \left(\frac{std(\widehat{X}_\iota)}{std(X_i)} - 1\right)^2 + \left(\frac{mean(\widehat{X}_\iota)}{mean(X_i)} - 1\right)^2}$$

Graphical metrics

- **Annual cycle**

A curve of the average intra-annual (seasonal) variations of modeled and observed variables (Figure 3 below).

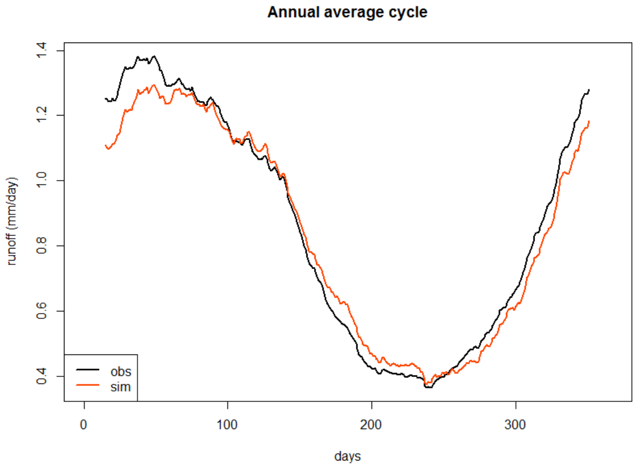

Figure 3. Example of a regime curve for observed and simulated streamflow.

- **Daily (or monthly, or yearly) boxplots**

A boxplot made from the aggregation of a variable on a daily (or monthly, or yearly) timescale depending on the kind of variability that is to be stressed (Figure 4).

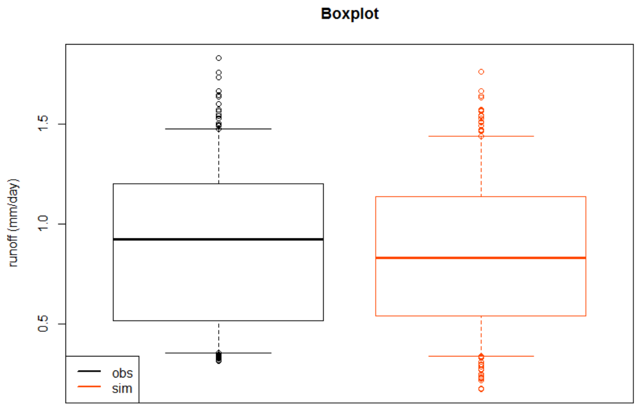

Figure 4. Boxplots comparing observed (in black) and simulated (in red) streamflow. The quantiles synthesized here are 5%, 25% (base of the box), 50%, 75% (top of the box) and 95%.

- **Interannual variations**

A graph with the time series of the yearly (or longer) averages of the modeled and observed variables (Figure 5).

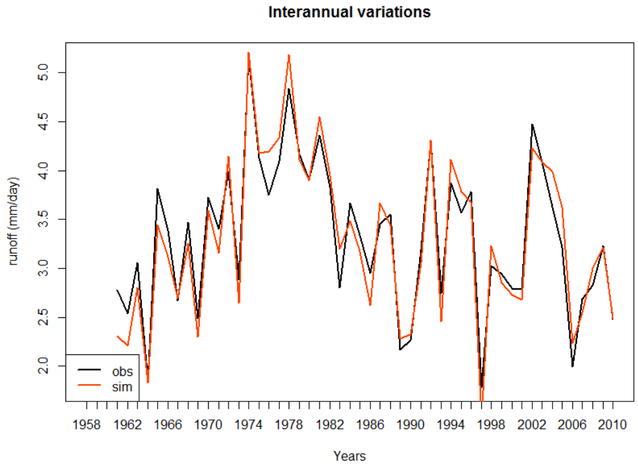

Figure 5. Variations in time of observed (in black) and simulated (in orange) yearly averaged streamflow.

- **IDF curve (Intensity Duration Frequency)**

An IDF curve is a mathematical function that relates the precipitation intensity to its duration and frequency of occurrence. IDF curves may be derived either by empirical or theoretical approaches. A review of existing methods to build IDF curves was written by Koutsoyiannis et al. (1998, Figure 6).

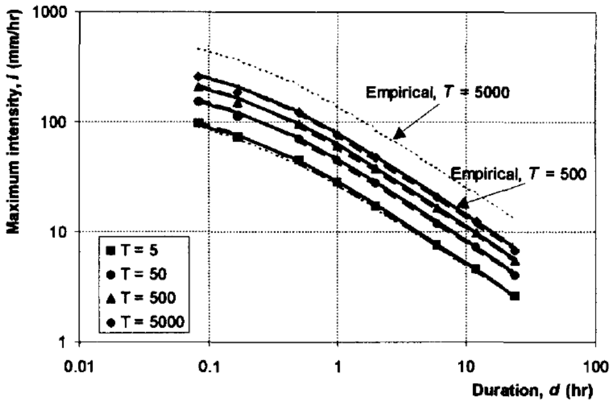

Figure 6. Intensity-Duration-Frequency Curve of rainfall for different return periods (Koutsoyiannis et al., 1998).

Statistical tests

- **MKT (Mann-Kendall test)**

The Mann-Kendall test is a non-parametric test to detect monotonic trends in time series. It does not require any particular assumption on the data (except no serial correlation) or on the errors. An applied example is presented by Vousoughi et al. (2013). A version able to handle autocorrelated series such as daily river flow is presented by Hamed & Rao (1998).

- **ADT (Anderson-Darling test)**

The Anderson-Darling test is designed to test whether a dataset is drawn from a given probability distribution. In its basic form, the AD test is distribution-free, i.e. there are no parameters to be estimated in the distribution being tested (Anderson & Darling, 1954).

REFERENCES

Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, vol. 49, no 268, p. 765-769.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, vol. 377, no 1-2, p. 80-91.

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, vol. 204, no 1-4, p. 182-196.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, vol. 15, no 5, p. 559-570.

Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, vol. 206, no 1-2, p. 118-135.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, vol. 10, no 3, p. 282-290.

Vousoughi, F. D., Dinpashoh, Y., Taghi Aalami, M., & Jhajharia, D. (2013). Trend analysis of groundwater using non-parametric methods (case study: Ardabil plain). *Stochastic environmental research and risk assessment*, vol. 27, no 2, p. 547-559.