Project Number 700099

Call: H2020-DRS-01-2015

Project Title:

ANYWHERE

EnhANcing emergencY management and response to extreme WeatHER and climate Events

Subject:

Deliverable 2.5:
Final report describing: (i) Uncertainty of forecasted weather-related hazards, (ii) robustness under worst-case climate change projections and/or coinciding natural hazards, (iii) operationalizability, and (iv) impact forecasting

Dissemination Level:

PU Public

Delivery date: **31st December 2019**

Month: **Month 43**

Organisation name of lead contractor for this deliverable: **University of Geneva (UNIGE), Switzerland / Wageningen University (WUR), The Netherlands**

This project has received funding from the European Union's H2020 Programme under the topic of potential of current and new measures and technologies to respond to extreme weather and climate events under grant agreement no. 700099.

Document Information

Title	**Final report describing: (i) uncertainty of forecasted weather-related hazards, (ii) robustness under worst-case climate change projections and/or coinciding natural hazards, (iii) operationalizability, and (iv) impact forecasting**
Lead Authors	Henny A.J. van Lanen (WUR, The Netherlands), Simone M. Schauwecker, Markus Stoffel, Juan A. Ballesteros-Cánovas (UNIGE, Switzerland) and Mario B. Rohrer (UNIGE/METEODAT, Switzerland)
Contributors	Marc Berenguer (UPC, Spain) Tuomo Bergman (FMI, Finland) Paolo Ciavola (CFR, Italy) Hannah Cloke (UoR, United Kingdom) Claudia Di Napoli (UoR/ECMWF, United Kingdom) Tomás Fernandez Montblanc (CFR, Italy) Estibaliz Gascon (ECMWF, International) Jarmo Koistinen (FMI, Finland) Tero Niemi (FMI,Finland) Rosa M. Palau (UPC, Spain) Shinju Park (UPC, Spain) Flavio Pignonen (CIMA, Italy) Seppo Pulkkinen (FMI, Finland) Jenna Ritvanen (FMI, Finland) Josias Ritter (UPC, Spain) Daniel Sempere-Torres (UPC, Spain) Heikki Sinisalo (FMI,Finland) Samuel Jonson Sutanto (WUR, The Netherlands) Ryan Teuling (WUR, The Netherlands) Remko Uijlenhoet (WUR, The Netherlands) Claudia Vitolo (ECMWF, International) Annakaisa von Lerber (FMI, Finland) Michalis Vousdouskas (JRC, International) Fredrik Wetterhall (ECMWF, International) Kaisa Ylinen (FMI,Finland)
Distribution	PUBLIC
Document Reference	Van Lanen, H.A.J., Schauwecker, S.M., Stoffel, M., Ballesteros-Cánovas, J.A. and Rohrer, M.B. (2019): Final report describing: (i) uncertainty of forecasted weather-related hazards, (ii) robustness under worst-case climate change projections and/or coinciding natural hazards, (iii) operationalizability, and (iv) impact forecasting. ANYWHERE Report[1].

[1] We revised the original title of Deliverable D2.5: "Final report describing: (i) robustness of algorithms under worst-case climate projections, and (ii) methods implemented to characterize the uncertainty inherent in the hazard and impact forecasting algorithms" to align it better with the objectives: Basically, we swapped the order of the topics of D2.5. Furthermore, we extended the report with operationalizability, and impact forecasting

Document History:

Date	Revision	Prepared by	Organisation	Approved by
23.12.2019	Rev_1_0	Henny van Lanen, Simone Schauwecker, Markus Stoffel, Juan Ballesteros-Cánovas, Mario Rohrer	WUR, UNIGE, METEODAT	Marc Berenguer (UPC)
30.12.2019	Rev_2	Henny van Lanen, Simone Schauwecker, Markus Stoffel, Juan Ballesteros-Cánovas, Mario Rohrer	WUR, UNIGE, METEODAT	Marc Berenguer (UPC)

Related Documents:

This report is closely linked to Deliverable report D2.2 that describes the first assessment of the robustness and uncertainty of models to forecast weather-induced hazards and impacts requirements of the algorithms/tools to be integrated into the ANYWHERE MH-EWS (Ballesteros Cánovas et al., 2017). The first version of the algorithms/tools are described in D2.1 by Van Lanen et al. (2017), and the improved version in D2.3 by Ciavola et al. (2017). The final version of the code of the algorithms/tools to be implemented in the Multi-Hazard Early Warning System (MH-EWS) is reported in D2.4 (Van Lanen et al., 2019).

This report and others are available from the ANYWHERE Project Website at: http://www.anywhere-h2020.eu/.

Summary

This deliverable D2.5 describes details about the uncertainty and robustness (U&R) of the platforms/algorithms/tools that forecast and nowcast weather-related natural hazards. It is a follow-up of the Deliverable report D2.2, which contained a preliminary assessment of U&R. A thorough survey of U&R aspects by developers of the platforms/algorithms/tools showed that a common methodology to assess uncertainty and robustness could not be developed. Reasons for this were among others: (i) the different concepts to forecast natural hazards, (ii) the data requirements (historic period, spatial coverage, type of data), and (iii) state of development of algorithms/tools. However, it came out that there was similarity in the design of the uncertainty assessment for almost all hazards, that is, comparing re-forecasts with observations, or a proxy for observations, using re-analysis data or simulation output. Results from these comparisons on uncertainty are reported for the platforms/algorithms/tools, and summarized in a table. Next, robustness, i.e. if platforms/algorithms/tools are transferable in space and time (towards a future climate), is described based upon existing literature (running the weather forecast platforms was beyond ANYWHERE). The developers concluded that none of the platforms/algorithms/tools uses a fully physically-based approach, which implies that these have to be parameterized. Hence, robustness was approached by exploring if platforms/algorithms/tools can be parameterized in any region in any time. Some algorithms/tools appeared to be more robust (e.g. storm surges) than others. Parameters of some platforms/algorithms/tools are regularly updated and are therefore by default able to cope with new situations in light of a changing future climate. Re-calibration of parameters from weather forecast platforms warrants a re-calibration of some downstream applications, for example, several algorithms/tools that forecast natural hazards. The robustness analysis of forecasting compound weather events brought light that this topic is still rather new on the research agenda. Robustness is an inherent part of weather forecasting platforms, incl. all the feedbacks in the system, and hence these are implicitly included in the weather forecasts that drive possibly coinciding or cascading natural hazards. Examples of compound wet and dry natural hazards and their pathways are described, which could be used later to assess robustness of forecasting these events.

Operationalizability of each platform/algorithm/tool has been presented in terms of development stage, uses, and required data/resources. The development stage of slightly more than 50% of the platforms/algorithms/tools has been proven in an operational environment. About 10% of the algorithms/tools are not validated in relevant environment yet.

The deliverable concludes with some examples on how natural hazards can be translated into impacts for large-scale applications (pan-European, regional) and following different approaches (bottom-up and top-down). Some pan-European maps providing information on vulnerability are described (population density, critical infrastructure). In the bottom-up approach these maps were used to show, for example, how forecasted air quality can be converted into forecasted impacts on people.

Table of Contents

1 Introduction

This deliverable report addresses the following topics:

- an assessment of **uncertainty** of forecasted weather and climate-induced hazards and associated impacts considering various geo-climatic settings across Europe to support existing decision-making processes and complementing and enhancing existing technologies.
- an assessment of **robustness** of the algorithms to forecast natural hazards and their associated impacts under worst-case climate change projections (frequency and severity of future extreme events) and/or coinciding natural hazards.
- **operationalizability** of the platforms/algorithms/tools that forecast and nowcast weather and climate-induced hazards.
- **large-scale impact forecasting**, how to move from hazards to impact over Europe.

Uncertainty reflects a state of imperfect knowledge, which is caused by a lack of data, either not available or not accessible, or from not incorporating all what is known or even foreseeable (e.g. IPCC, 2018). **Robustness** is the ability of a system to generate outcome (e.g. forecasts) that remain valid under different or changing environmental conditions (e.g. future climate under global change, compound or cascading events), or describes the degree to which the system behaviour remains acceptable valid in spite of mutated conditions. The challenge of developing a reliable Multi-Hazard Early Warning System is to include forecast tools/algorithms that have a low uncertainty and high robustness.

The uncertainty of tools/algorithms to forecast natural hazards and their impacts relies on the identification of uncertainty sources and their quantification throughout all the assessment steps. Uncertainty can take multiple forms, for instance, (i) epistemic, and (ii) aleatory uncertainties. Epistemic uncertainties are due to aspects (e.g. mechanism, processes, data) one could in principle know but which are not implemented in practice. This type of uncertainty could be lowered by improving assessment approaches, model structures, and/or including necessary data (e.g. initial conditions, boundary conditions, model parameters). Aleatory uncertainties manifest unknown system outcomes that can differ each time one runs an experiment under similar conditions, e.g. the temporal evolution of a hydrometeorological variable under the same driving forces. Aleatory uncertainties cannot be reduced a priori.

It is assumed that forecasting systems will acceptably perform after a while (allowing improvements) when environmental conditions stay stationary. However, what happens if these environmental conditions start drifting significantly off normal (e.g. global change)? System uncertainty might alter, it may increase or decrease. Development of a forecasting system with an appropriate uncertainty for stationary conditions is already a challenging task, but making it robust as well, adds significant complexity to system design. For example, are model structures still valid, are model parameters sufficiently adapted to the changing environmental conditions? Another

type of robustness, which needs to be addressed in a multi-hazard setting, is whether the forecasting platform is able to acceptably predict coinciding weather-related natural hazards and cascading events or not. This type of robustness requires in particular an appropriate description of feedback processes.

Performance (validation, verification, and evaluation) of ANYWHERE products is addressed at several places, i.e. deliverables (Annex I). The consortium decided to report these analyses as follows:

- D1.4: synthesis of the overall outcome from the tools/algorithms (in terms of impacts) as embedded in the various A4* platforms;
- D2.5: Uncertainty and Robustness (U&R) of the tools/algorithms in terms of forecasted natural hazards, mainly at the pan-European scale, but not exclusively. In cases there are no large scale applications, then U&R knowledge on more detailed scale is presented, if available. Moreover, in some cases, historic events (reforecasts) have been analysed.;
- D3.4: U&R of the tools/algorithms in terms of forecasted natural hazards for the demonstration period in the Pilot Sites (October 2018 – September 2019);
- D6.5: U&R of the tools/algorithms in terms of forecasted impacts for the demonstration period in the Pilot Sites.

In the ANYWHERE project, the Uncertainty and Robustness (U&R) assessments are to be based on using existing probabilistic procedures, including forecast tools/algorithms. Deliverable D2.2 reported about the first findings on uncertainty and robustness within ANYWHERE (Ballesteros Cánovas et al., 2017). One of the main findings is that knowledge on U&R is very diverse among the tools/algorithms. Some algorithms/tools (e.g. weather and hydrological forecasting, flood forecasting) have well-established procedures to investigate and report on uncertainty, whereas others (e.g. pan-European storm surges, heatwaves and health, droughts) were more in an experimental phase. None of the tools/algorithms investigated coinciding natural hazards, and cascading effects of multiple hazards.

The D2.2 report showed that still clear gaps in knowledge on U&R exist. ANYWHERE would benefit from improved knowledge. It was decided to extend WP2 Advanced forecasting models and tools to anticipate Weather and Climate (W&C) event induced impacts, in particular how the tools/algorithms for weather forecasting and nowcasting, hydrological forecasting, as well as forecasting of natural hazards will perform during the demonstration period in the Pilot Sites. U&R of tools/algorithms will be mapped in (qualitative) terms for, as far as possible, dominant pan-European physiographic settings. The U&R of the tools/algorithms will mainly be validated by comparing forecasted natural hazards with observed hazards (e.g. including reforecasts, validation runs, sensitivity analysis). Past events will be identified to investigate impacts of coinciding and cascading effects of multi-hazards. Parallel to investigating historic and ongoing conditions, hazards under a future climate will be explored to test robustness.

When investigating U&R, the various components of the **ANYWHERE** forecasting platform should be considered (Fig. 1). There is uncertainty in the weather forecasting ❶, and in the downscaling ❷ to obtain a higher spatial resolution. Hence, weather forecast products ❸ are presented in a probabilistic way to account for some of the uncertainty. The probabilistic weather forecasts and nowcast products are input to the tools/algorithms ❺ that forecast the various single natural hazards ❻. For some hazards (floods and droughts) a hydrological model driven by the weather forecasts has to be run first to obtain probabilistic hydrological forecasts before the hazard tools/algorithms can be applied. The forecasted single hazards can be combined and jointly be presented to face and cope with compound natural hazards that either are coinciding or cascading. Most of D2.5 focuses on U&R aspects of the forecasting of natural hazards (Steps ❶ to ❻).

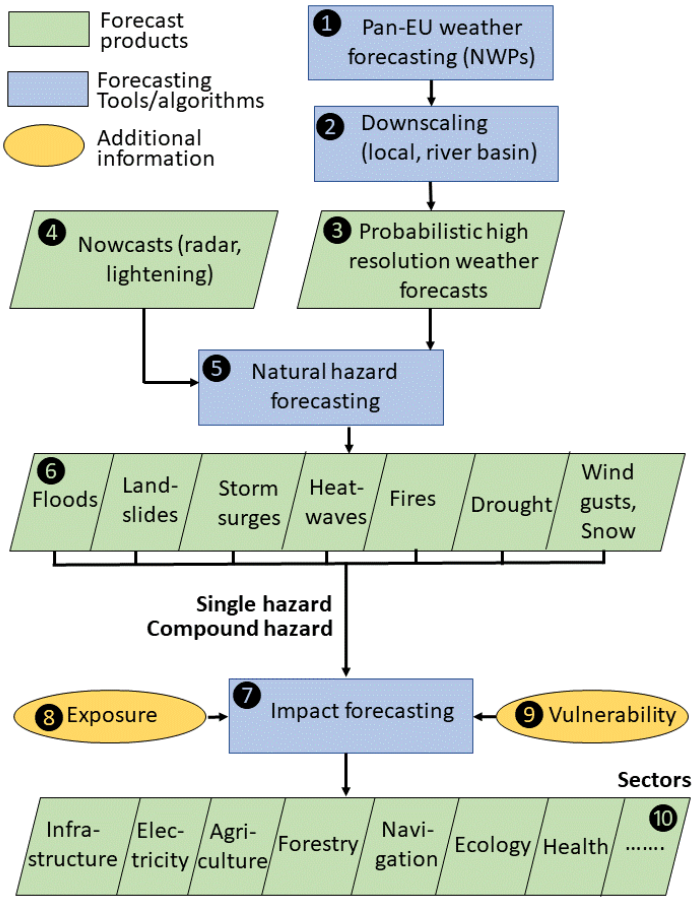


Figure 1: **ANYWHERE** multi-hazard forecasting platform: products and tools/algorithms to forecast and nowcast weather-induced natural hazards and associated impacts.

The original idea for D2.5 was to search for commonalities among algorithms/tools to develop a common methodology to assess U&R aspects. This was assumed to cover all algorithms/tools, or at least those that translate hydrometeorological forecasts into natural hazards ❺. This would provide the basis for improvement of the algorithms/tools, if time would permit, or otherwise it would provide recommendations for improvement. It appeared during the U&R analysis of the natural hazards that a common methodology could not be developed. Reasons for this were among others: (i) the different concepts to forecast natural hazards, (ii) the data requirements (historic period, spatial coverage, type of data), (iii) state of development of algorithms/tools. However, it came out that there was similarity in plans for the uncertainty assessment for all hazards, that is, comparing forecasts with observations, or a proxy for observations, using re-analysis data or simulation. Hence, it was decided to present the current state of knowledge on U&R per algorithm/tool and to make a distinction between algorithms/tools dealing with: (i) weather forecasts and nowcasts, and hydrological forecasts, i.e. the hydrometeorological forecasts, and (ii) natural hazard forecasts. We started the report with the uncertainty in platforms/algorithms/tools forecasting hydrometeorological forecasts (Chapter 2), which is followed by the algorithms/tools that forecast natural hazards (Chapter 3). The chapters on uncertainty are followed by robustness of the algorithms/tools that forecast the natural hazards under a future climate (Chapter 4) and compound hazards (Chapter 5). Chapters 3-5 conclude with a table summarizing per hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether these were observed, proxy or simulated data, (iv) what were the length of time series data, and (v) which comparison method has been applied.

As mentioned above, knowledge on U&R (Ch. 2-5) varies per forecasting platform/ algorithm/tool. Therefore, we present U&R in this report as follows:

- Hydrometeorological forecast platforms: most knowledge was already available before the start of ANYWHERE and likely had to be updated as part of operational routine. Updating is a continuous activity of hydrometeorological organizations. Assessment of uncertainty is beyond ANYWHERE activities. The current state of uncertainty is briefly described with emphasis on references to literature. Robustness under a future climate (Chapter 4) is not dealt with because running the platforms under such new climate was not feasible within ANYWHERE. Forecasting of compound weather events is an inherent part of such platforms and not further described. Though, associated compound natural hazards are explained (Chapter 5);
- Natural hazard algorithms/tools: (i) U&R already reported in D2.2; only conclusions are repeated in D2.5 with reference to D2.2, (ii) U&R obtained during the lifetime of ANYWHERE and published in open access papers; only synthesis is presented in D2.5 with reference to the paper, and (iii) U&R obtained during the lifetime of ANYWHERE and not reported yet; a more comprehensive description is given in D2.5.

As a response to the last review[2], D2.5 continues with a description of the operationalizability of the forecasting algorithms/tools. Operationalizability is linked to experience with using the algorithms/tools, incl. a performance assessment (Chapter 6). The chapter addresses: (i) development stage, (ii) what purpose the algorithms/tools have been used for, (iii) data/resources required to apply at different scales, and (iv) deviation of final algorithm/tool relative to originally promised. The deliverable report concludes with how to move from hazard to impact forecasting at large scale beyond the level of the Pilot Site level (Chapter 7). Pan-European maps with exposure and vulnerability information are presented, which commonly are used to translate hazards into impacts. Some approaches to arrive at large-scale impact forecasting are described, which are supported by examples. The Deliverable D2.5 was originally due in M36 (May 2019). To explain how to translate forecasted hazards into associated impacts at the pan-European scale (Fig. 1,Steps ❼ to ❿), the delivery date was postponed to include information on large-scale exposure and vulnerability, impact forecasting, as well as information on operationalizability. This is a follow-up of the description of impact forecasting by Ciavola et al. (2017).

ANYWHERE nowcast and forecast platforms and the algorithms/tools that forecast and nowcast weather-related natural hazards with their lead time ranges are given in Figure 2 to ease readability of the report.

[2] Result of the Review of H2020 project 700099 — ANYWHERE, Antonio Fernandez-Ranada Shaw, Project Officer, European Commission, Research Executive Agency, Ref. Ares (2019)3316357 - 21/05/2019.

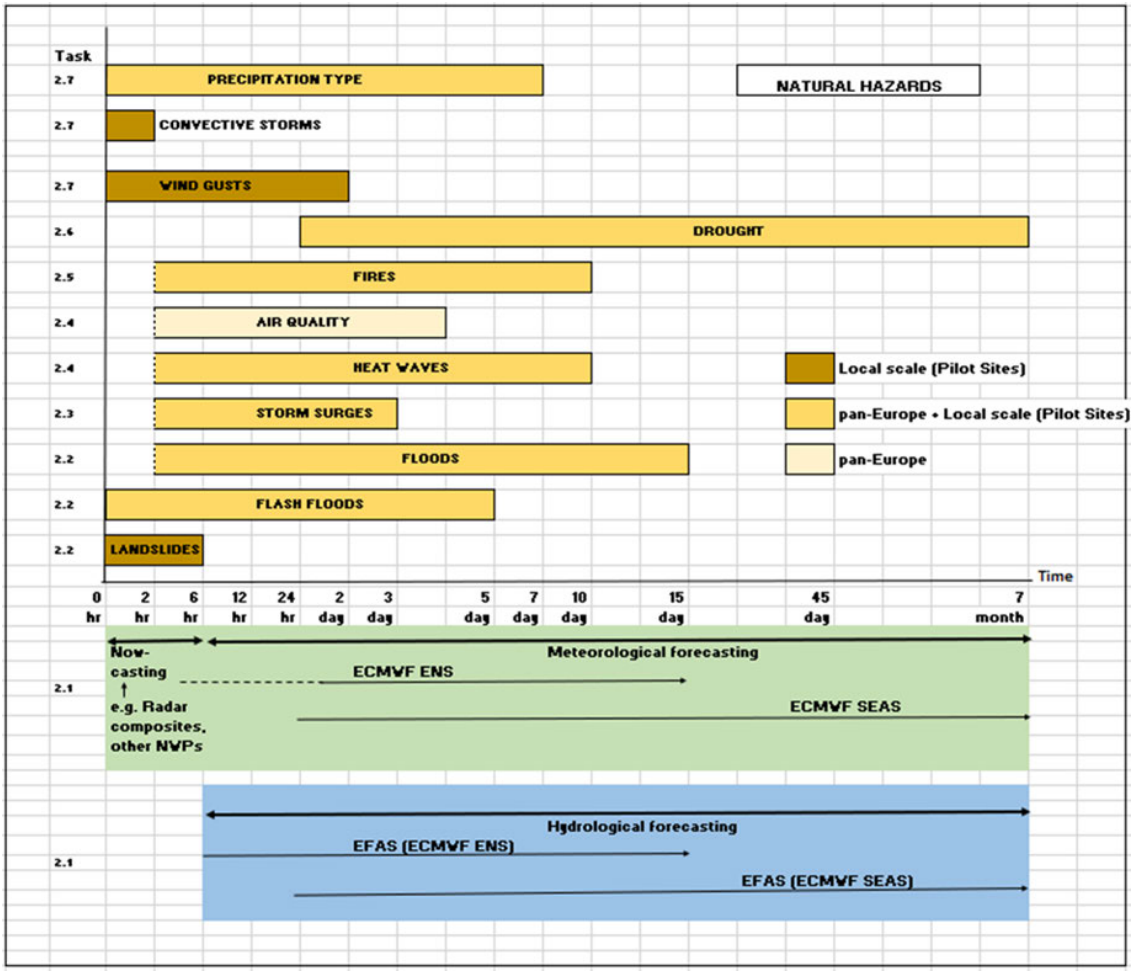


Figure 2: **ANYWHERE** nowcast and forecast platforms (bottom) and the weather-related natural hazards with their lead time ranges (not at linear scale) and spatial coverage.

2 Uncertainty in hydrometeorological and forecast and nowcasts products

This chapter describes uncertainty of using platforms/algorithms/tools that generate hydrometeorological forecast and nowcast products (Fig. 1, ❶ and ❷). The hydrometeorological and forecast and nowcasts are divided in weather forecasts and nowcasts (Section 2.1) and hydrological forecasts (Fig. 1, ❸ and ❹; Section 2.2). Details on the platforms/algorithms/tools and the associated products can be found in D2.3 (Ciavola et al., 2017) and D2.4 (Van Lanen et al., 2019).

Uncertainty of platforms/algorithms/tools is also closely related with lead time. Generally, uncertainty is larger for larger lead times, e.g. several days before a possible event. This does not mean that such forecasts are useless, but, on the contrary, they offer the possibility for issuing a pre-alert. Based on such a pre-alert, the responsible bodies can check the operational readiness and closely monitor the possible event. In this line of thinking these pre-alerts may be an important step in a warning chain, beginning with a pre-alert and eventually ending in a mobilization of public safety organizations.

2.1 Meteorological forecast and nowcasts products[3]

The first part explains uncertainty of platforms/algorithms/tools that produce weather forecasts, which is followed by uncertainty of nowcasts algorithms/tools. The section concludes with an example that describes uncertainty in downscaling.

2.1.1 ECMWF Integrated Forecasting System (ECMWF-IFS)

The ECMWF integrated forecasting (IFS) is the name of the computer software that computes the global forecasts at ECMWF. It is a very complex system consisting of thousands of tasks in different codes. The system is undergoing constant changes and is released in new version in so called cycles. The current cycle of IFS is Cycle 46R1, which was released in 6 June 2019. Each release consists of several new features and upgrades, and they all contribute to the continuous evolution of the system.

The skill and uncertainty of ECMWF's probabilistic forecast is measured continuously as part of the routine. Each cycle is also compared with the previous and a benchmark

3 There is no report on the uncertainty aspects of the radar-based precipitation nowcasting algorithm PhaSt (CIMA; Metta et al, 2009).

The radar-based precipitation nowcasting RAVAKA (FMI) has not been included in this deliverable report. Since the beginning of 2018 there is an internal FMI project to renew its nowcasting methodology to the STEPs method. The research and implementation is ongoing and the estimated launch of the new version is by the end of 2019. Hence there has been no development nor improvements in the RAVAKE method, nor resources to study of its uncertainty. Uncertainty results of the newer version are not available yet.

to test the quality. The testing is done using a plethora of measures, and the cycles are compared with so called “score card”. Some cycle changes may have neutral or even negative effects on a particular score or region depending even though the overall changes are positive.

Uncertainty is inherent in ensemble prediction system, since the whole idea of introducing ensembles is to estimate the forecast uncertainty by perturbing the initial conditions and the physical tendencies in the model. For a well-balanced ensemble, the spread in ensemble should be as large as the mean error for each lead time. The spread-error relationship is something that is being constantly monitored by ECMWF. The overall skill is assessed every year in a technical memo published by ECMWF (Haiden et al., 2018). Examples of measuring and presenting skill scores of ECMWF-IFS are presented in Figure 3 and 4. Examples cover the temporal evolution of the skill of temperature and precipitation forecasts. For more reading on IFS evolution and documentation see: IFS cycles[4].

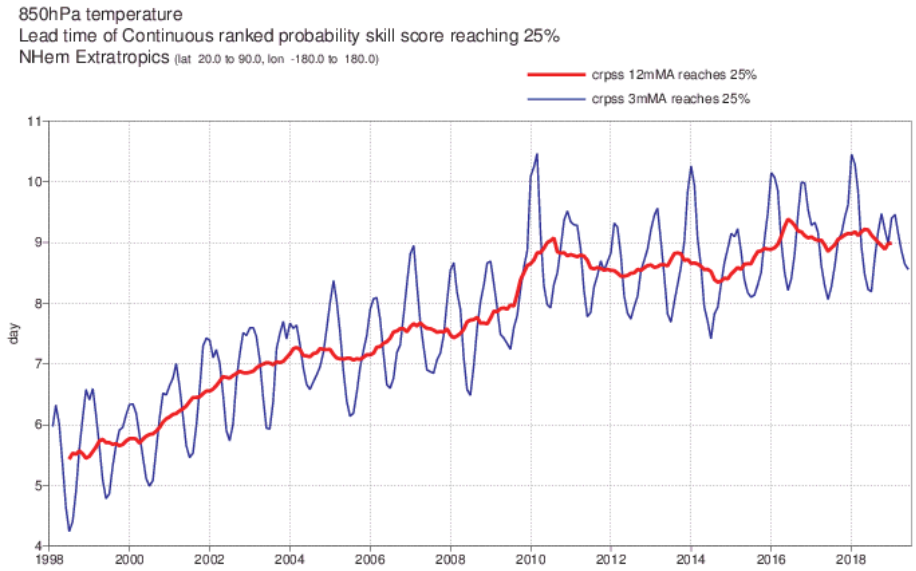


Figure 3: The plot shows for each month the range at which the 3-month mean (blue line) or 12-month mean (red line) centred on that month of the continuous ranked probability skill score of the 850hPa temperature ENS dropped below 25%. This is a primary headline score for the ECMWF ENS.

[4] https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model.

Figure 4: The plot shows for each day the range at which the 365-day mean (centred on the day) of the 24-h precipitation forecast 1-SEEPS score dropped below a fixed threshold. The plot for extratropics shows a supplementary headline score for the ECMWF HRES.

2.1.2 FMI Numerical Weather Prediction models (FMI-NWPs)

The weather forecast products provided by FMI are mainly based on the synoptic scale HIRLAM (High Resolution Limited Area Model) and the mesoscale HARMONIE-AROME model (Harmonie-Arome General description; Termonia et al., 2018). The forecasting product Precipitation Type (Section 3.6.3) utilizes also the synoptic scale GFS (Global Forecast System) and ECMWF (European Centre for Medium-Range Weather Forecasts). The GLAMEPS (Grand Limited Area Model Ensemble Prediction System) weather forecast products are based on HIRLAM and the MetCoOp (Meteorological Co-operation on Operational Numerical Weather Prediction) ensemble products on HARMONIE-AROME. The HIRLAM calculation domain covers Europe and HARMONIE-AROME Nordic countries (Fig. 5).

The main sources of NWP model uncertainty is caused by: i) observations, ii) model and iii) nonlinearity and chaos (Olliaho 2014; Shutts and Pallares 2014). Constantly increasing computing capacity and development of NWP models are decreasing uncertainty, but models have their physical limitations. In the short range (short lead times), the NWP uncertainty is largely an initial value problem. The NWP model are always suffering from insufficient cover of high-quality observations. Nowadays NWP models utilize more and more satellite data, which offers a good cover but only have a tolerable quality. High-quality observations and sophisticated data assimilation are crucial to achieve accurate initial conditions in the data assimilation. Observations are used to correct errors in the short lead time forecasts using an analysis from the preceding period. HARMONIE-AROME utilizes 3-dimensional variational assimilation (3D-Var) and HIRLAM more sophistic 4D-VAR, which covers also the time dimension.

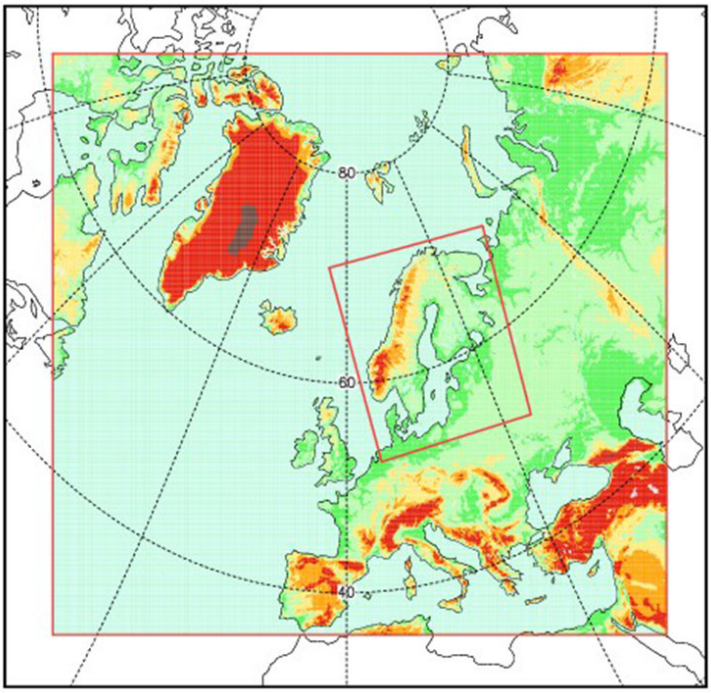


Figure 5: The calculation domains of FMI´s models (outer HIRLAM, inner HARMONIE-AROME).

The other important uncertainty factor is the model structure itself. The uncertainty of the model is based on unresolved scales, theory, numerical calculations and bugs. The unresolved scales lead to need of parametrization schemes (Fig. 6), which always have some weaknesses. A higher resolution decreases the amount of needed parametrization. The closer the parametrizations go towards describing the phenomena in molecular level, the more the knowledge about the processes is lacking.

The need to specify boundary condition increase uncertainties. The details of orography are easily increased with higher computing capacity and resolution, but the changing characters (e.g. soil moisture and water vapor flux) cause uncertainties. In addition mountainous terrain poses a challenge to NWP models, due to physical parametrization that is based on flat terrain. The high resolution HARMONIE-AROME is a model permitting non-hydrostatic convection, which makes it capable to forecast severe weather because of deep convection. Most summer time severe weather in Europe is related to deep convection. It´s also able to depict phenomenon, like downslope wind and mountain waves.

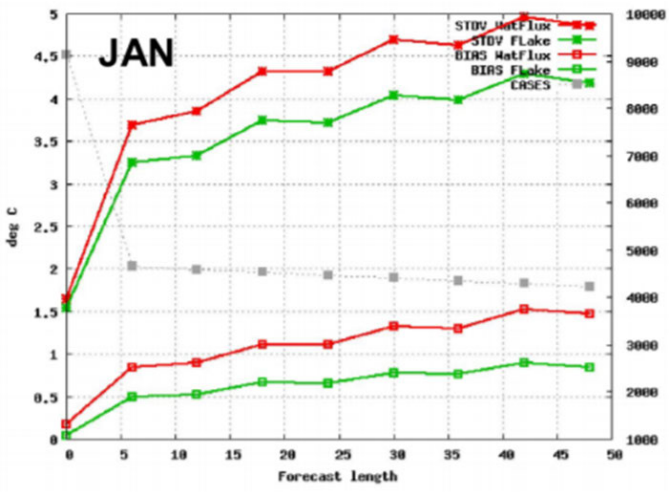


Figure 6: An example on how a lake parametrization schema (lake model Flake) decreases the uncertainty in the HARMONIE-AROME forecasting model. The observations are from 123 SYNOP stations, which are located nearby lake. Source: Ulf Andrae, SMHI.

The chaotic nature of the atmosphere causes a lot of uncertainty. Uncertainty increases with lead time (Fig. 7) due to that small uncertainties grow to large errors in the unstable flow. Present understanding is that chaotic behavior limits accurate weather forecasts to approximately 14 days. In an effort to quantify the large inherent uncertainty remaining in numerical predictions, ensemble forecasts have been used to obtain useful results farther into the future than otherwise possible.

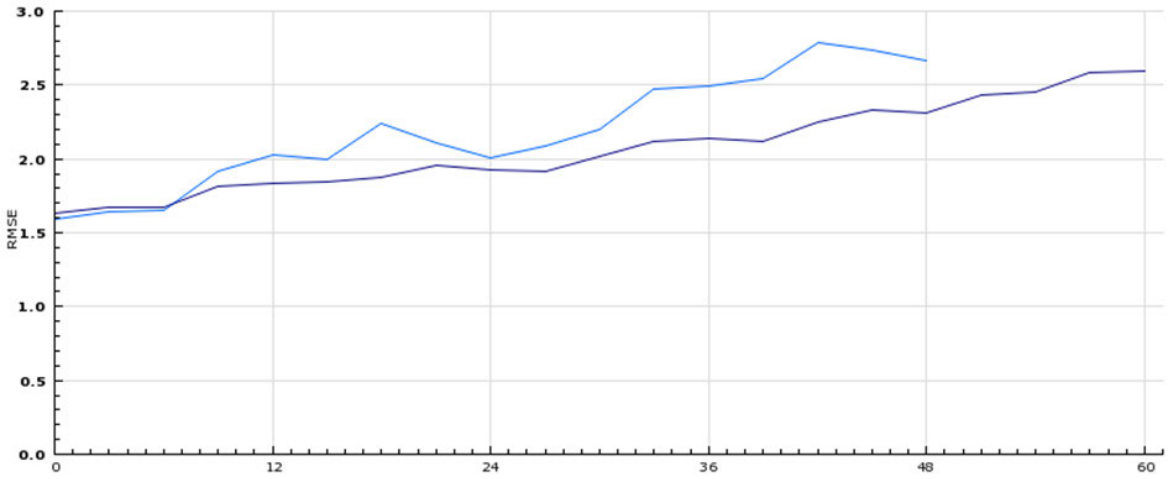


Figure 7: An example how lead time increases gust forecast uncertainty expressed as RMSE (blue: HIRLAM and violet: HARMONIE-AROME). Forecast were compared with Finnish ground-based SYNOP observations. The verification period is 1st March-30th June 2019.

2.1.3 UPC-CRAHI algorithm for precipitation nowcasting

Uncertainty aspects of the algorithm for precipitation nowcasting is reported under Flash Flood forecasting (Section 3.1.1). The nowcasting product drives the flash flood algorithm FF-EWS. Park et al. (2019) provide more details about the precipitation nowcasting algorithm that provides continental precipitation over Europe with high resolution (2 km, 15 min) based upon the operational production of the OPERA composites from the European weather radar networks.

2.1.4 Downscaling precipitation mountainous regions (UNIGE/METEODAT)

High precipitation events are particularly hazardous for mountainous catchments where runoff rapidly increases after strong precipitation events, leaving limited time for warning to first responders. A high-quality forecast of heavy precipitation with long enough lead times (2-5 days) and an adequate spatial resolution is crucial for decision makers and intervention bodies. However, global weather forecast models remain relatively coarse in resolution (Section 2.1.1 and 2.1.2) and higher resolution regional model forecasts have limited lead times. To assess forecast performance for an Alpine region, we estimated daily areal precipitation for three Alpine catchments based on in-situ measurements and the past ECMWF IFS-HRES forecasts (Section 2.1.1) for lead times of 2 and 5 days for January 2010 to September 2018. We then evaluated the potential of a bias correction and downscaling method to remove biases and to increase skill.

Method and Data

We present a post-processing method to bias correct and downscale heavy precipitation forecasts with medium-term lead times. The Model Output Statistics (MOS) consists of a two-step approach. In a first step, we are combining a universal kriging of the daily precipitation data with an indicator kriging of the precipitation occurrence, with the aim to interpolate meteorological station data and forecast model output to a 2 km grid. In a second step, we apply a MOS following a quantile-quantile mapping approach to correct systematic model biases. We used a dataset daily precipitation time series from a dense network of 787 stations from MeteoSwiss and SYNOP reports from some regions. As predictor, we selected daily precipitation from ECMWF historical operational runs. Data are available at a base horizontal resolution of about 9 km up to 10 days in advance for 2010-2018.

Main findings

We found that the uncertainty of heavy precipitation forecasts with medium range lead times (2-5 days) is considerable (Fig. 8) and there are regional patterns of significant positive and negative biases (Schauwecker et al., 2019). This leads to false alarms, but also missed events.

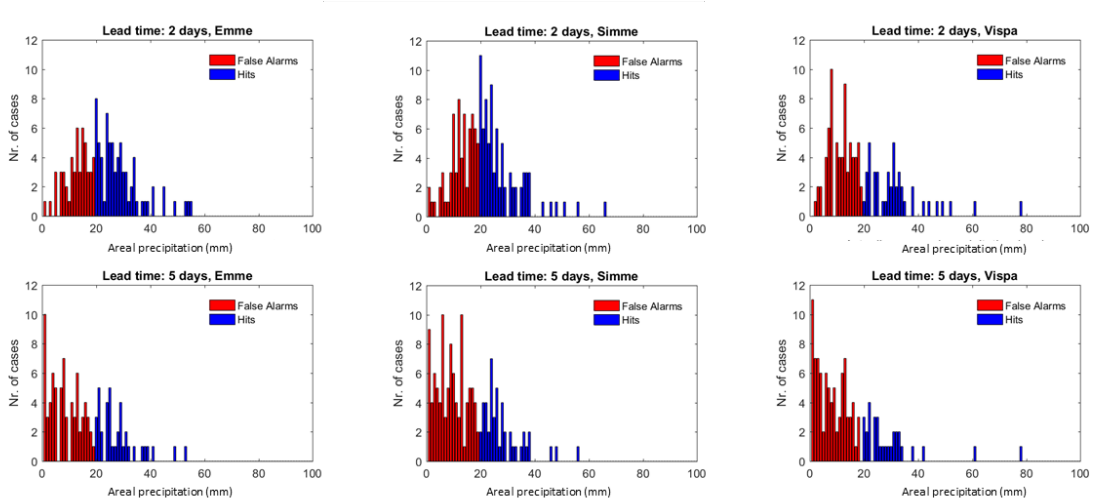


Figure 8: Histograms of interpolated areal precipitation records for three Alpine catchments (Emme, Simme and Vispa) and a lead time of (above) 2 days and (below) 5 days for days with a forecast ≥20 mm.

The maps provided in Figure 9 show the correction factors for 5-days lead time and intensities of 20 mm. A clear spatial pattern of positive correction factors emerges over Northern Switzerland (Jura, Plateau), especially in the winter half year (October – March). The here presented post-processing and downscaling method has the potential to remove biases, improve the mean average errors and lowering the false alarm rate as presented by global weather forecast models. The improvement is however limited for single events of heavy precipitation, in the case the model prediction was low. Our results indicated that despite of the limitations, bias-corrected forecasts with medium range lead times are essential for risk anticipation. Such a forecast might complement other forecast tools and helps improving the reliability of pre-alerts and alerts, which are an important basis for initiating monitoring activities and to timely reach operational readiness.

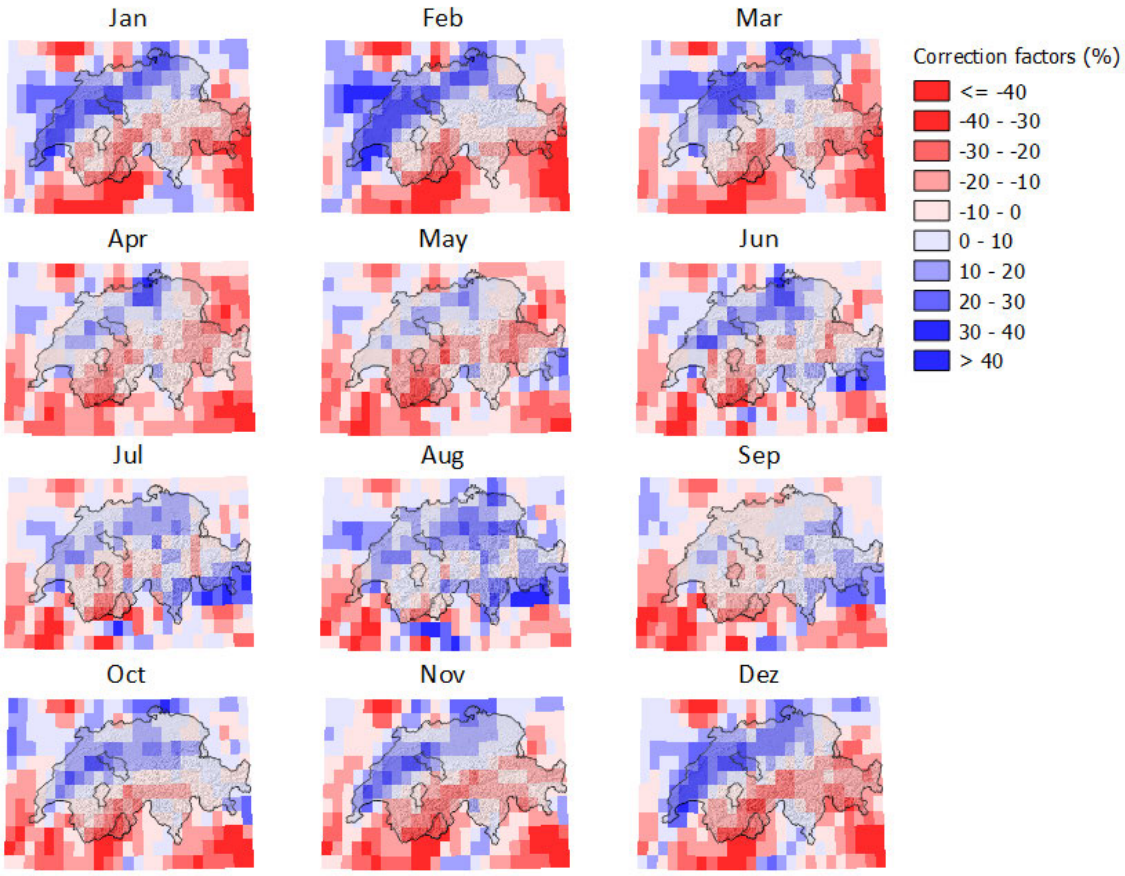


Figure 9: Correction factors (%) for Switzerland and for a lead time of 5 days and a precipitation intensity of 20 mm.

2.2 EFAS Hydrological forecast product (ECMWF)

In ANYWHERE probabilistic hydrological forecasts have been done with EFAS (European Flood Alert System. EFAS has during the last year gone through a major change as the domain was increased, the projection changed, the hydrological model was upgraded, and the parameters were recalibrated. These changes also affected ANYWHERE, since the switch to the new domain was done in January 2019 for all products that were reliant on EFAS, i.e. flood and drought forecast products. This is important to mention in this context, because it means that extra care must be taken when scrutinizing the uncertainty in the model system.

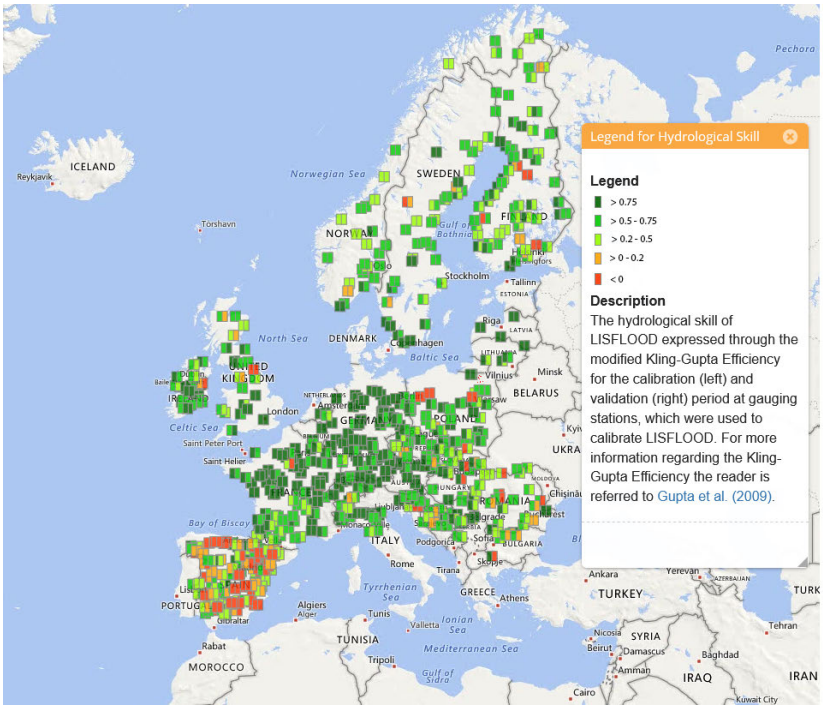


Figure 10: Hydrological skill of EFAS at the calibration locations. Colour coding denotes the quality of the KGE during calibration (left half of square) and validation (right half of the square). Dark green: KGE > 0.75; Green: KGE 0.5 – 0.75, Light green: KGE 0.2 – 0.5; Orange: 0 – 0.2; Red: < 0. Note: the hydrological skill of the calibration will be available as separate layer.

EFAS monitors the performance on a regular basis through two types of scores; qualitative assessment of the issued forecasts and a qualitative skill score assessment of the model performance. The qualitative assessment is relying on assessing whether issued flood warnings were hits or misses, and from there a contingency table can be calculated. However, this has not been carried out recently, and this assessment is currently not available. EFAS gathers feedback from the users after each sent formal notification, but the resulting data is not always reliable.

The update of the system meant a rigorous recalibration of the domain over the period 1991-2016. The main score used was the Kling-Gupta efficiency (KGE). The model performed well in most parts of the domain, except for Spain (Fig. 10). In many parts around the Mediterranean region there were no stations used in the calibration. The reason for the poor performance in Spain is mainly due to uncertainty in terms of reservoir management. It should be noted that a poor performance in the calibration does not necessarily translate into a poor performance of the hydrological forecasts.

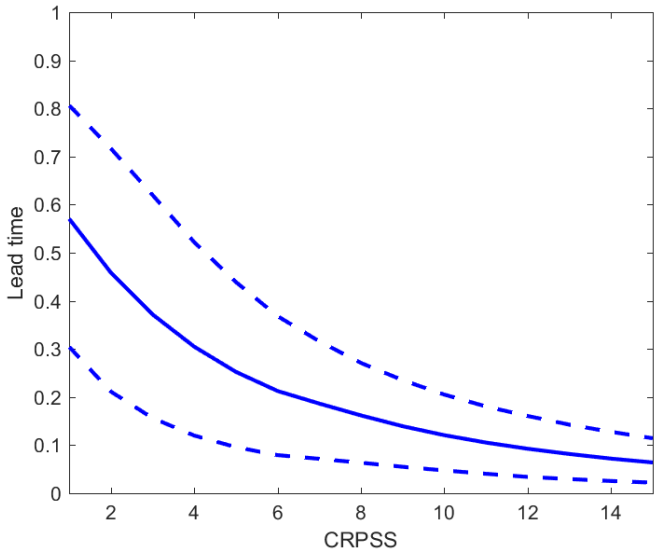


Figure 11: EFAS headline score CRPS Skill score (CRPSS) over the entire period for all grid points which have a upstream area of >=2000 km^2 as a function of lead time. The reference score is the forecast from the previous day, which is a very difficult forecast to beat. The dotted lines represent the 90th and 10 percentiles respectively.

The assessment of skill scores were done over the period 1 June 2018-31 July 2019, well over a year. The skill was computed using the continuous ranked probability skill score (CRPS, Annex III) and the bias in the forecast (Fig. 11). The proxy for observations were the simulations forced with observations. For more information on the scores used and strategy for verification, please see D2,2 (Ballesteros Cánovas et al., 2017).

The performance shows a strong spatial variability, where the model performs generally better in the northern part of Europe than in the south. It can be noted that the Danube river is quite well forecasted (Fig. 12, right), although the areas contributing to the river does not show the same good skill (Fig. 12, left).

The skill also varies greatly over the year, which can be seen in Figure 13. The skill in winter is normally higher than in summer, because most rivers are at their lowest flow during that season. The variability is small, which makes it easier to predict.

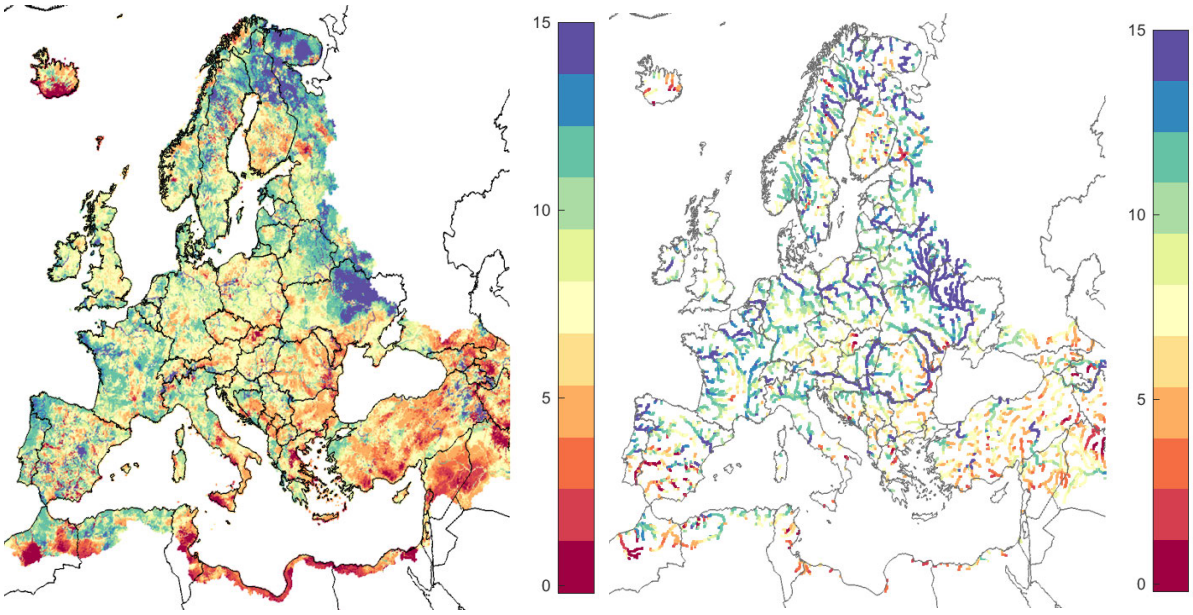


Figure 12: Number of weeks before the CRPSS goes below 0.1 for all the points in Europe (left) and for all major river points (area >=2000km²). Blue colour indicates a good performance, and red colour indicates a less good performance.

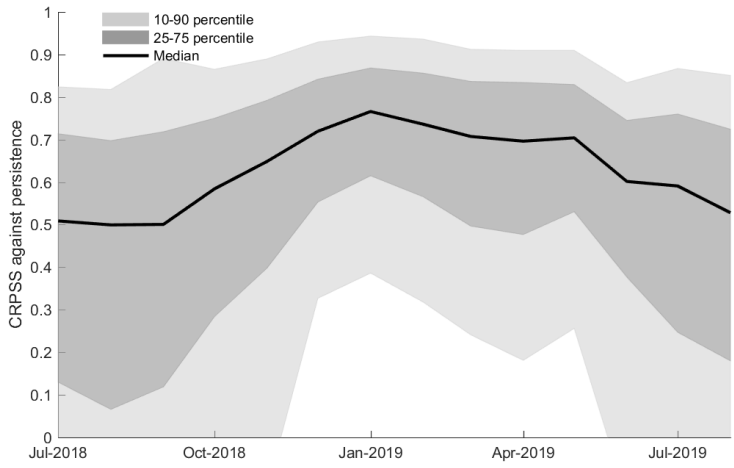


Figure 13: Monthly means of CRPSS for lead time 5 days for all the major river points in Europe over the test period against persistence forecast. The skill is largest during the winter months, when there is less variation in the flow in large parts of Europe, which gives a benefit to persistence as a reference forecast.

Although EFAS originally was developed for flood forecasting (Section 3.1), it was also calibrated taken the water balance into account. It therefore works generally well also under conditions (e.g. low river flow). This is essential for drought forecasting (Section 3.5).

3 Uncertainty in weather-induced natural hazards products

This chapter describes uncertainty of using algorithms/tools (Fig. 1, ❺) that translate hydrometeorological forecast and nowcast products (Fig. 1, ❸ and ❹) into natural hazard forecast products (Fig. 1, ❻). Successively, we describe uncertainty in using algorithms/tools on floods and landslides, storm surges, heatwaves and air quality, fires, droughts, and convective storms, severe winds and heavy snowfall. Details on the algorithms/tools and the associated products can be found in D2.3 (Ciavola et al., 2017) and D2.4 (Van Lanen et al., 2019).

As mentioned in Chapter 1, no common methodology could be applied to assess uncertainty of all algorithms/tools that forecast natural hazards. Hence, a table is given (Table 5, Section 3.7) that summarizes per hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether these were observed, proxy or simulated data, (iv) what were the length of time series data, and (v) which comparison method has been applied.

3.1 Floods, flash floods, landslides and debris flows[5]

The first part explains uncertainty of algorithms/tools that produce flood forecasts, which is followed by uncertainty of algorithms/tools that generate forecasts on landslides and debris flows.

3.1.1 FF-EWS flash flood hazard nowcasting algorithm (UPC)

As described in D2.3, the FF-EWS module (Alfieri et al., 2011; 2017; Corral et al., 2009; 2019; Park et al., 2019) produces flash flood hazard nowcasts using a simple traffic-light index. Within the project, this algorithm has been applied over different domains and different resolutions. The pan-European FF-EWS flash flood hazard algorithm nowcasts flash flood hazard throughout Europe with a resolution of 1 km and 15 minutes, and for lead times up to 6 hours. The algorithm uses the catchment-aggregated rainfall (i.e. the rainfall integrated in the catchment upstream of each point of the drainage network and over its characteristic concentration time) as the main forcing leading to flash floods and compares the observations and nowcasts of this variable with reference values (thresholds) to estimate the flash flood hazard over the drainage network. The sources of uncertainty of the algorithm are related to the quality of the rainfall inputs (both observations and nowcasts), the values of the thresholds related to different flash flood hazard level, and the validity of the hypothesis that the

[5] There is no report on the uncertainty aspects of the Flood-PRObabilistic Operational Forecasting System, Flood-PROOFS (CIMA). However, Flood-PROOFS has been included in a comparison with the FF-EWS flash flood hazard nowcasting algorithm (Section 3.1.1.2).

flash flood hazard can be directly related to the hazard level of the catchment-aggregated rainfall.

The work presented here has focused on analyzing the flexibility of the FF-EWS algorithm (which enables its implementation in different domains, different resolutions and with different inputs) and its robustness performance anywhere in Europe. This is done by using existing regional or national real-time rainfall products (both from radar and gauge networks), increasing the resolution of the drainage network in the pilot site versions (default set is 1km at European scale) or use of existing regional IDF analyses to relate the rainfall amounts and the flash flood hazard (expressed in terms of the return period).

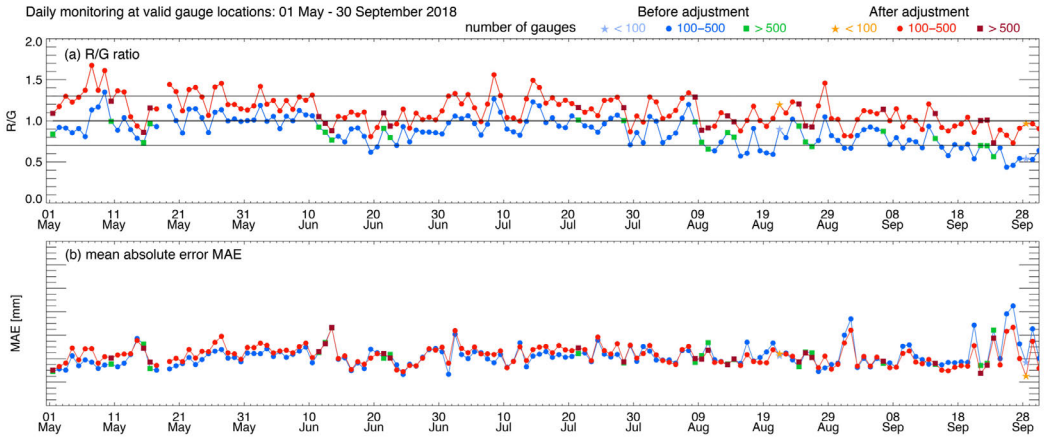
The section has been divided into two subsections, analyzing these implementation efforts of the algorithm and the results from the following perspectives:

- Its ability to assess the flash flood hazard at European scale based on the radar composites produced by the EUMETNET program OPERA (Saltikoff et al., 2019) and the radar-based nowcasts produced with the UPC-CRAHI algorithm for rainfall nowcasting.
- Its ability to assess the flash flood hazard at regional scale in Liguria (Italy), in comparison with the forecasts obtained with the Flood-PROOFS system, based on the rainfall-runoff model Continuum.

3.1.1.1 Flash flood hazard nowcasting at the European scale (UPC)

During the demonstration of ANYWHERE, the European gauge-adjusted radar precipitation estimates based on OPERA radar composites and SYNOP observations have been made available in the different pilot sites of the project through the different configurations of the A4EU platform. Park et al. (2019) performed daily monitoring of these gauge-adjusted precipitation composites in year 2018 and showed that the quality of the input radar has also been improved compared with previous years (e.g. 2015-2017, Fig. 14b). This is possibly due to changes made recently in data collection strategy among the OPERA community to which our feedback has clearly contributed (e.g. Saltikoff et al. 2019). Park et al. (2019) also showed the impact of of adjusting the OPERA rainfall amounts with rain gauges observations on the flash flood hazards estimates (Fig. 15).

(a)



(b)

Skill score quantile [%]			2015 May-Oct	2016 May-Oct	2017 May-Oct	2018 May-Sep
R/G	Non-adjusted	15	0.38	0.39	0.51	0.68
		50	**0.48**	**0.50**	**0.61**	**0.89**
		85	0.60	0.58	0.75	1.06
	Gauge-adjusted	15	0.89	0.91	0.89	0.93
		50	**1.04**	**1.05**	**1.06**	**1.11**
		85	1.27	1.23	1.26	1.31
MAE	Non-adjusted	15	3.59	3.56	3.27	3.41
		50	**5.00**	**5.16**	**4.26**	**4.05**
		85	7.18	7.04	5.61	5.03
	Gauge-adjusted	15	3.25	3.50	3.28	3.42
		50	**4.31**	**4.36**	**3.88**	**4.31**
		85	5.50	5.61	4.95	5.04

Figure 14: (a) Daily monitoring of the ratio R/G (R: radar-OPERA, G: gauge) and Mean Absolute Error (MAE) before and after (blueish and reddish colours, respectively) the gauge adjustment from May to September 2018. (b) Similar as (a) but summarized for different years (figures extracted from Park et al, 2019).

Additionally, in **ANYWHERE**, the effect of the density of the rain gauge network on the adjustment of the OPERA rainfall amounts has been further explored to cope with the uncertainties in the flash flood hazard nowcasting. This is done by adding real-time observations from denser gauge networks to the existing SYNOP network. The test has been made for Spain, where the SYNOP network is quite sparse, particularly in the southern regions. Figure 16 illustrates that the flash flood induced by heavy rainfall could be identified in San Javier (yellow level) after adding the available AEMET gauges in the gauge adjustment of the OPERA radar rainfall estimates used in the FF-EWS.

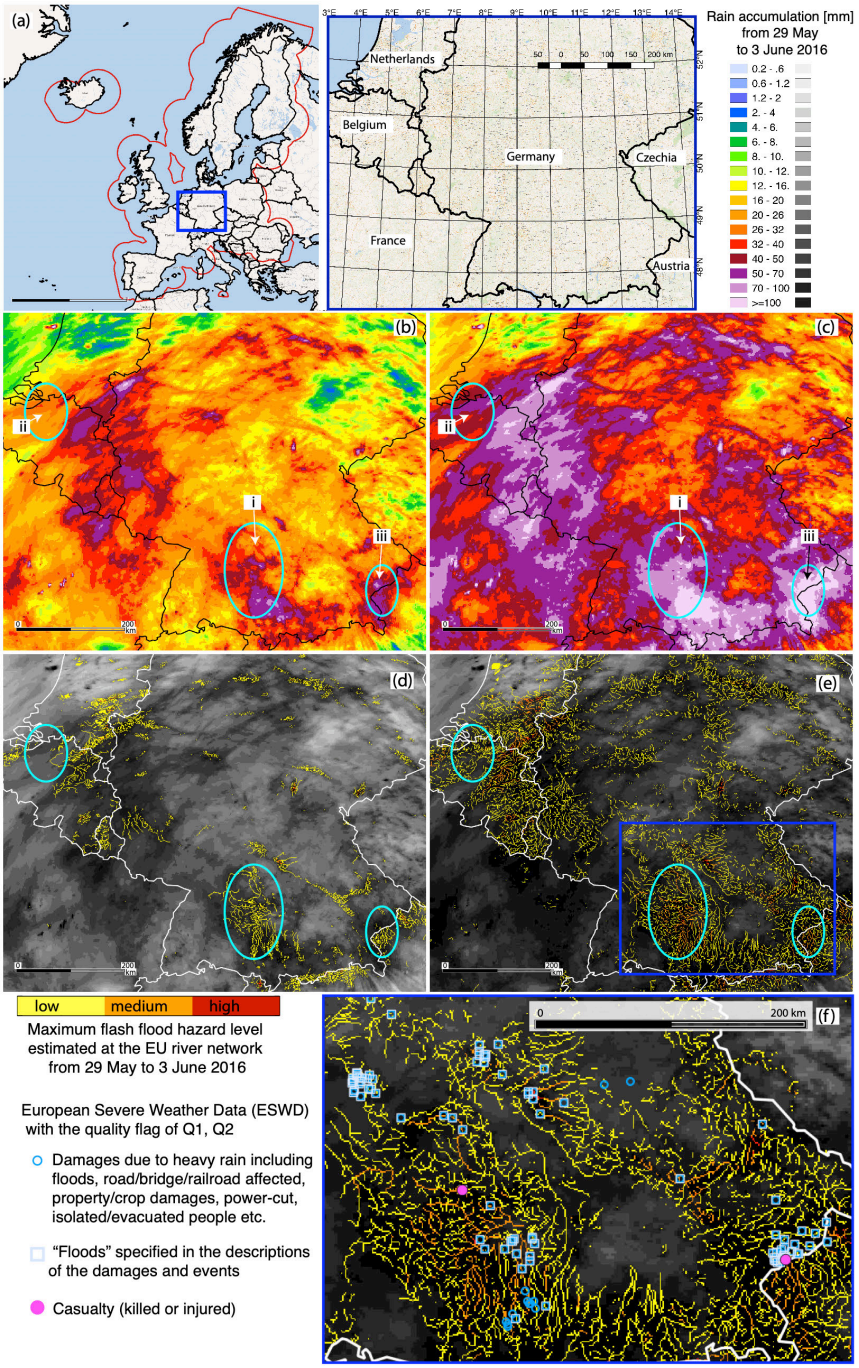


Figure 15: Example of the impact of gauge adjustment on the flash flood hazard assessment. (a) the areas affected in Europe from 29 May to 3 June 2016. The OPERA coverage is outlined in red. (b) OPERA rain accumulations. (c) gauge-adjusted OPERA rain accumulation. A summary of the ERICHA flash flood hazard level (i.e., the maximum level extracted during the periods) obtained with (d) the OPERA accumulation inputs and (e) the gauge-adjusted OPERA rain accumulation. (f) The damage report points from ESWD plotted over the zoomed areas (blue box) of e.

Figure 16: Gauge-adjustment can be by combining SYNOP and existing dense gauge network; e.g. with AEMET over Spain.

To extend the coverage of the European radar products used in **ANYWHERE**, the OPERA composite has been extended with the Italian composites (Fig. 17a), which also enables extended hazard identification over larger areas in Europe (e.g. see Fig. 17b).

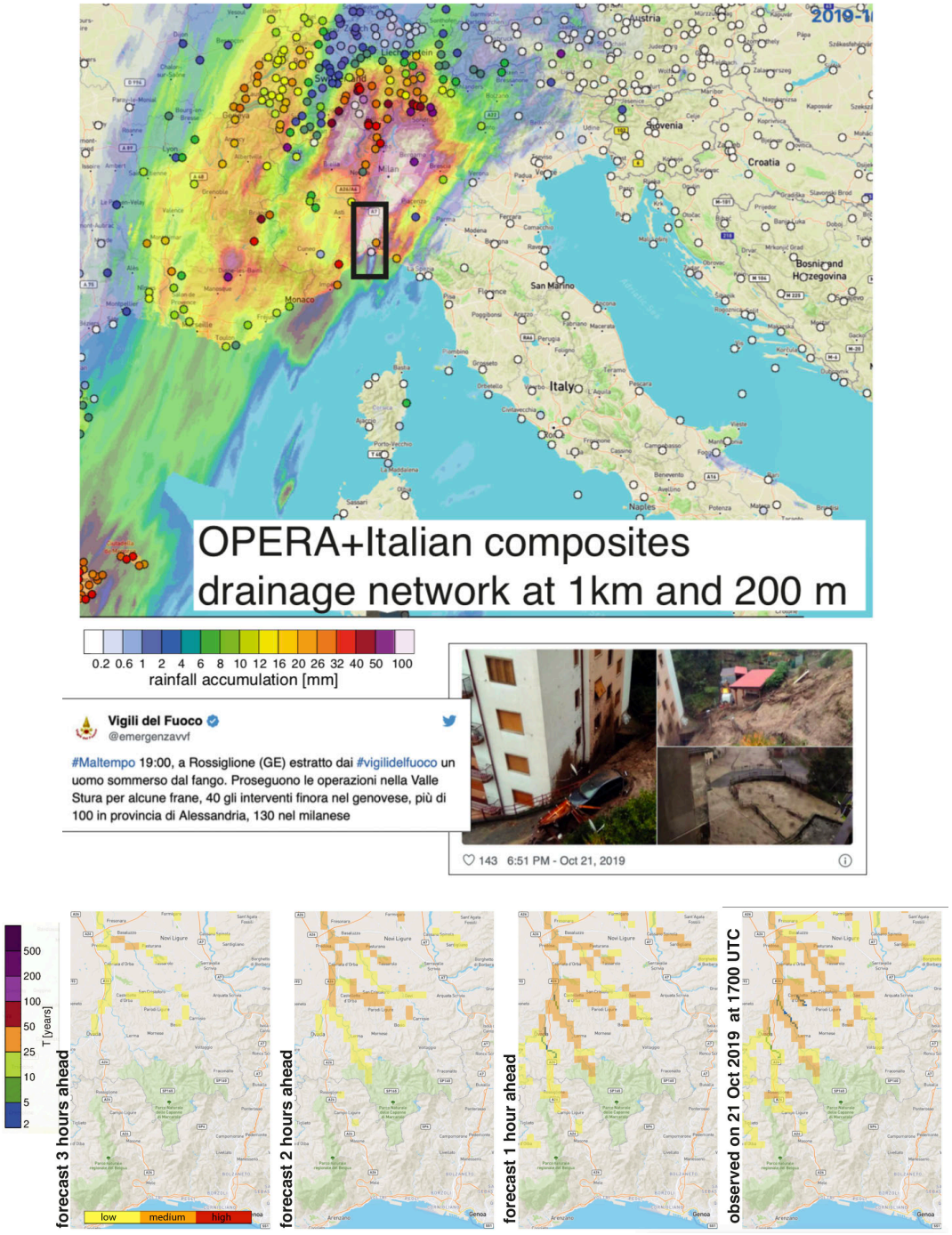


Figure 17: (a) Daily rainfall accumulation for 21 Oct 2019 estimated after combining the OPERA composites with the Italian national composites. Significant flooding occurred near Rossiglione(located inside the black box and shown in twitter at around 1800 LT) and Castelletto d'Orba (e.g. La Republica) (b) The real-time flood hazard level (low, medium, high) nowcasting done in the black box areas with lead time 1, 2 and 3 hours.

During the demonstration phase of the project, the analysis has also focused on assessing the quality of the flash flood nowcasts as a function of lead time. Figures 18 and 19 show the results for two recent events.

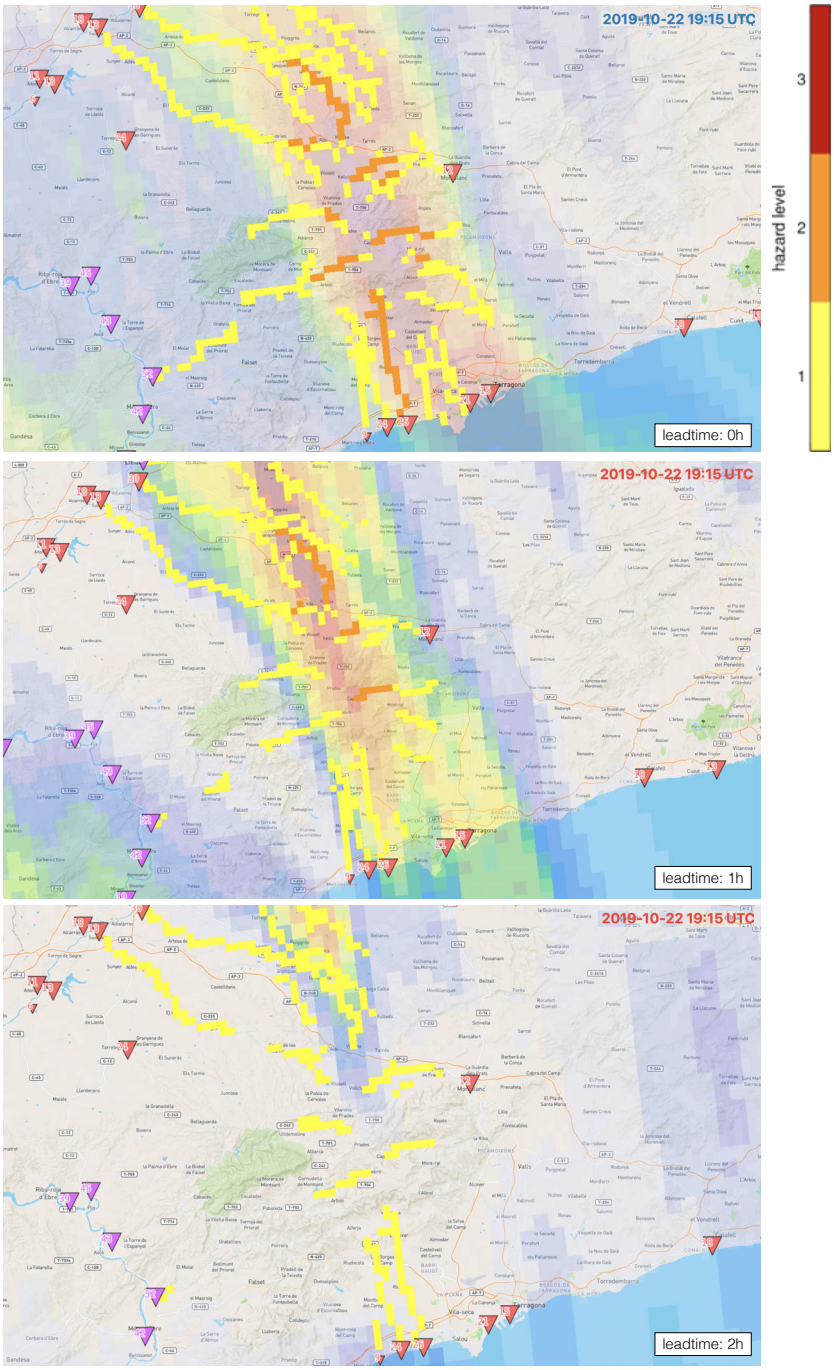


Figure 18: Comparison of the flash flood hazard nowcasts for the event of 22 October 2019 at 19:15 UTC on the coast of Tarragona (NE Spain) based on observations (top), the 1-h rainfall nowcasts (middle) and the 2-h rainfall nowcasts (bottom).

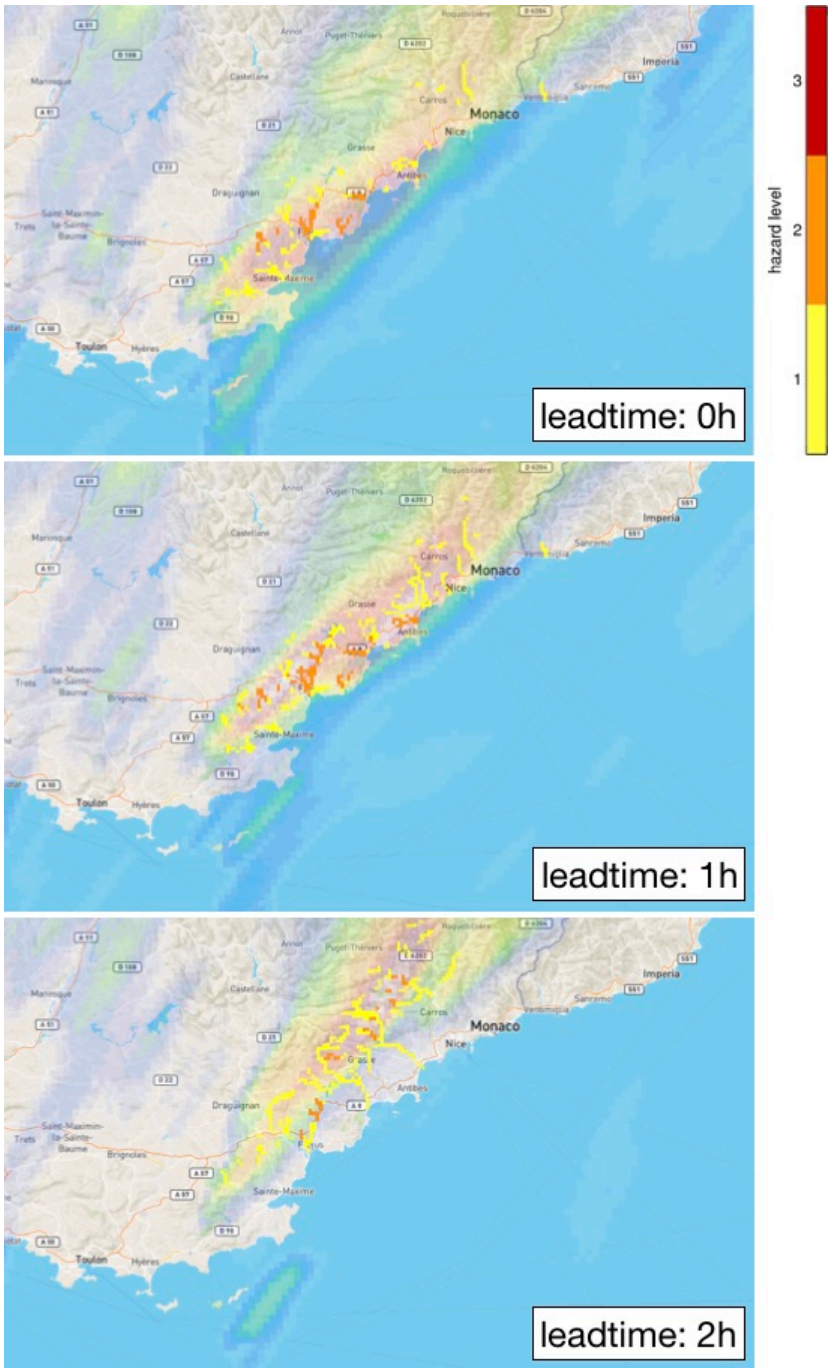


Figure 19: Comparison of the flash flood hazard nowcasts for the event of 01 December 2019 at 19:30 UTC on the French Mediterranean coast based on observations (top), the 1-h rainfall nowcasts (middle) and the 2-h rainfall nowcasts (bottom).

Overall, the conclusions from these analyses are: (i) The 1-h flash flood hazard nowcasts are very good; (ii) The limit of predictability of the flash flood hazard nowcasting algorithm is between 2-3 hours, especially for events with fast-evolving convective rainfalls. Forecasting of impacts caused by flash floods using FF-EWS is presented in Section 7.3.

3.1.1.2 Comparison of the FF-EWS algorithm with the simulations of a runoff-based hazard assessment algorithm

This part of the study is limited to evaluating the performance of the FF-EWS algorithm form the point of view of flash flood hazard assessment (the nowcasting component was turned off) by comparison with the hazard estimated with the Flood-PROOFS system, the flash flood hazard module running operationally for flash flood forecasting at the Regional Environmental Protection Agency of Liguria (ARPAL).

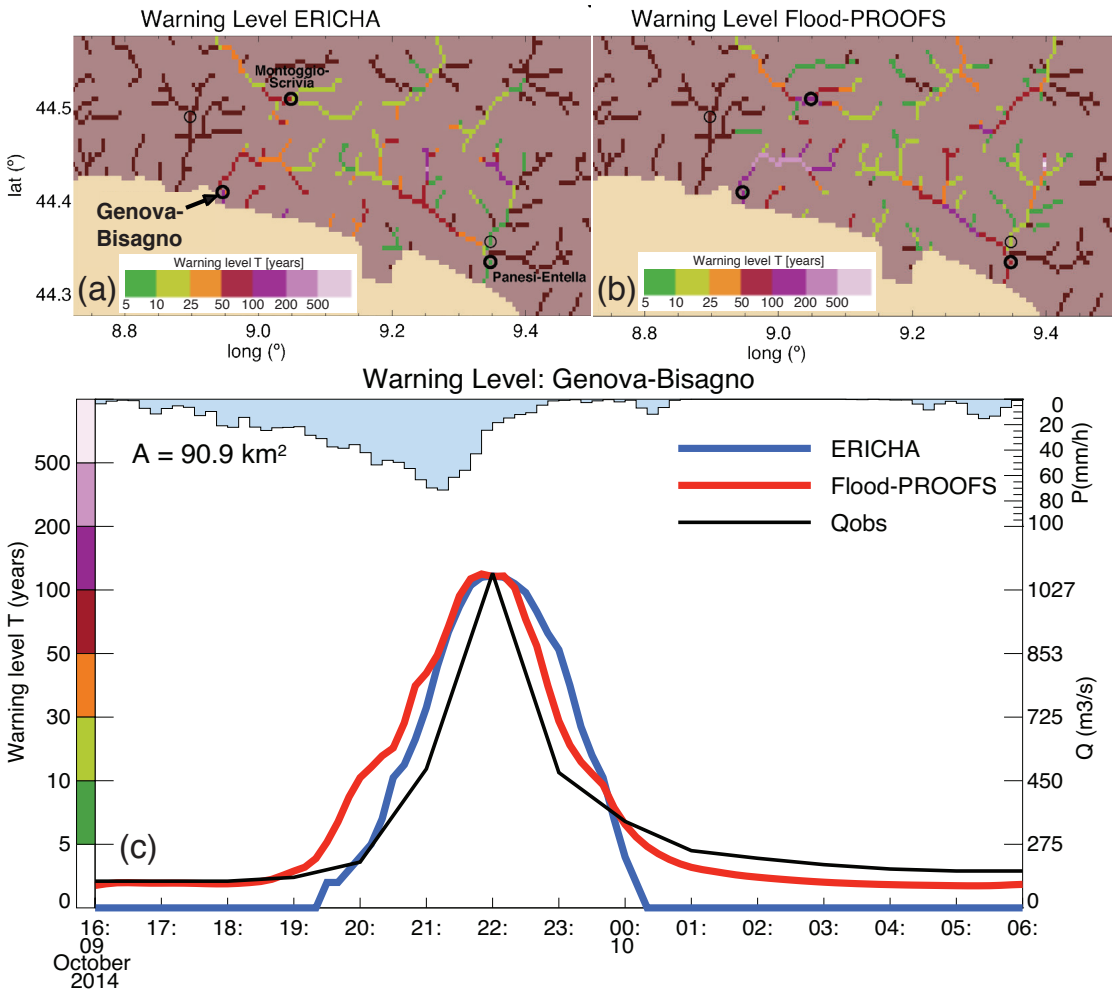


Figure 20: (a, b) Maximum return periods in the drainage network in Liguria (NW Italy), obtained by the FF-EWS system and Flood-PROOFS, respectively, for the event of 09-10 October 2014. (c) Evolution of the return periods estimated form observations at a section of the Bisagno Creek in Genova (in black), and simulated with FF-EWS and Flood-PROOFS (blue and red lines, respectively). From Corral et al., 2019.

The main difference between the two systems is that the FF-EWS algorithm assesses the flash flood hazard based on the rainfall accumulated on the catchment upstream each point of the drainage network (the basin-aggregated rainfall), whereas Flood-

PROOFS uses a full rainfall-runoff model to assess the hazard based on the simulated discharge.

Compared to the pan-European version of FF-EWS, the application of FF-EWS at regional scale enables a refined estimation of the flash flood hazard in to two aspects: (i) increasing the resolution of the drainage network, and (ii) assessing the flash flood hazard in terms of return period, using the regional Intensity-Duration-Frequency study available in Liguria.

FF-EWS and Flood-PROOFS have been run with the same rainfall inputs and over the same drainage network for the most significant rainfall events that affected Liguria in autumn 2014. This enabled a systematic comparison between the two systems, aiming at evaluating the effect of the underlying sources of uncertainty in the results obtained with the two systems. The results of the comparison have been recently published (Corral et al., 2019), and a brief summary is presented here.

The analysis is presented in terms of the estimated hazard (expressed in terms of the estimated return period), showing remarkable similarities between the two systems in the hazard estimates, especially for return periods larger than 10 years, such as the flooding of the Bisagno Creek in Genova on 9 October 2014 (Fig. 20). Flood-PROOFS tended to provide slightly higher return periods than FF-EWS, likely because the catchments were relatively wet before the studied events. Also, this comparison shows better agreement between the two systems in the larger catchments (over 50 km^2).

Also, the evolution and magnitude of the hazard estimates of both systems corresponded very well to the return periods estimated from measured discharges, impeding a conclusion on which of the two systems performed better. Based on these results, the authors recommend to opt for the discharge-based system (Flood-PROOFS) in regions where the rainfall-runoff model can be calibrated well and computational requirements are not a major concern. In regions where these conditions are not met (which is often the case), the FF-EWS system can be a good choice, since it has been shown to be capable of providing hazard estimates of similar quality.

3.1.2 Landslides and debris flows hazard and impact assessment (UPC)

The algorithm of Berenguer et al., 2015 (see also Palau et al., 2020) identifies the areas prone to landslides and debris flows triggered by heavy rainfall and is adapted to real-time performance. The algorithm uses radar-based rainfall estimates and nowcasts (Section 2.1.3) and outputs a map of a qualitative warning level that can be used so support landslide risk management.

The algorithm has been applied at regional scale in the pilot site of Catalonia (Spain), and the results obtained for a period of seven months have been presented by Palau et al. (2020). In this section, the main findings on performance of the system and its uncertainties are summarized.

The landslides warning level is computed combining (i) the susceptibility map of the study area, and (ii) gridded rainfall observations and nowcasts. For Catalonia, the susceptibility has been obtained combining the maps of slope of the terrain and land cover using fuzzy logic; and the global probabilistic rainfall thresholds by Guzzetti et al. (2008) have been adopted to define four qualitative rainfall hazard levels. The combination of these two variables is done using a 2-D warning level matrix classifying the warning level into 4 classes ("very low", "low", "moderate" and "high").

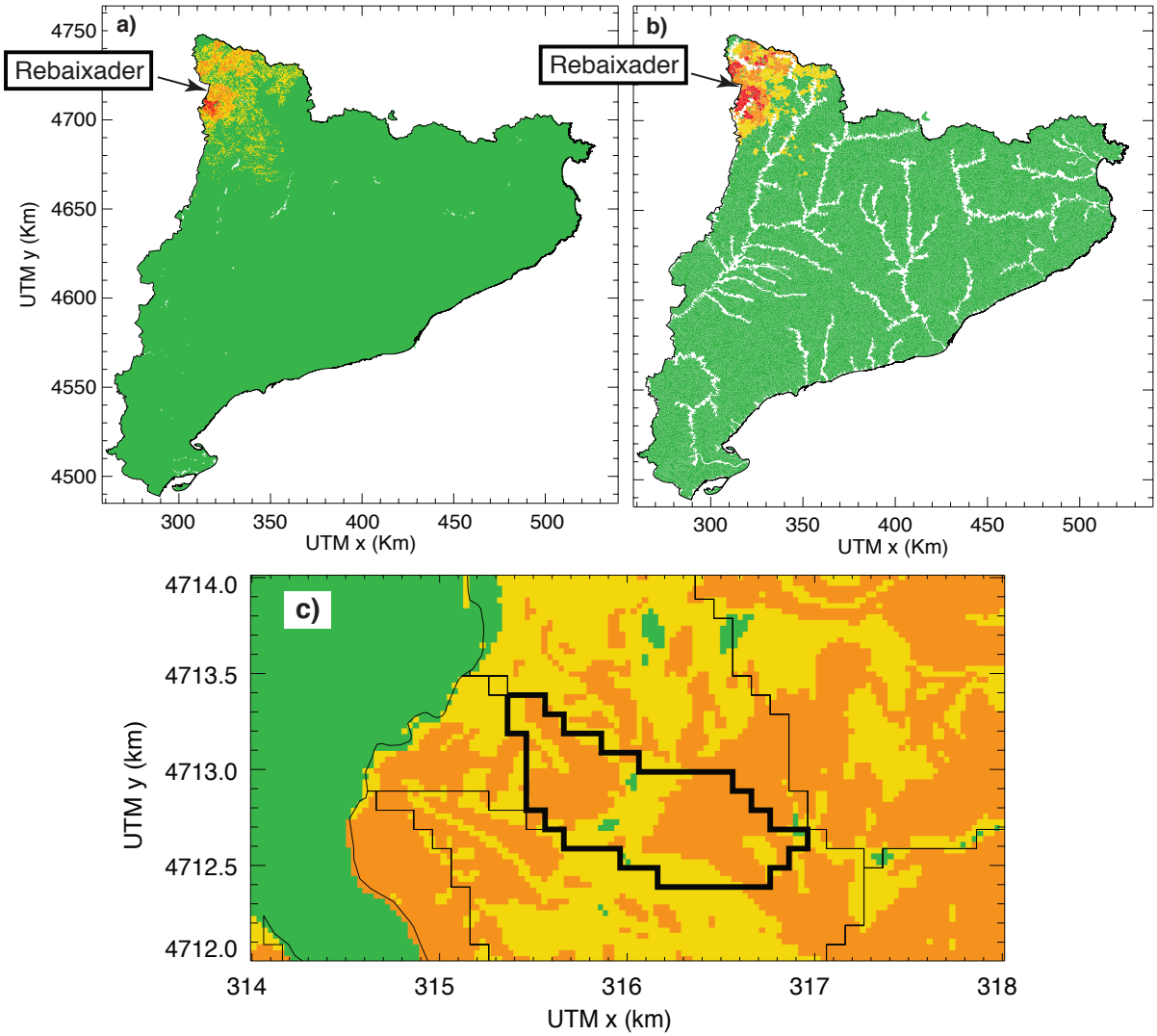


Figure 21: Warning level estimated in Catalonia for the situation of 21 July 2010 at 19:30 UTC. (a) and (b) display the results presented over a 30-m grid, and over a subbasin division. Green, yellow, orange and red represent warning levels "very low", "low", "moderate" and "high", respectively. (c) Warning level around the Rebaixader subbasin with a resolution of 30 m. In this subbasin, a debris flood was reported around this time; the pixels achieving higher warning level represent the possible landslide initiation zones.

The performance of the algorithm shows that landslide warnings were generally located at susceptible areas coinciding with the most significant rainfall events. Over

the studied period of seven months, the area where more than six days with “moderate” or “high” warning level is issued is rather small. However, in many of the locations with “moderate” or “high” warnings, no landslide event was reported (Fig. 21). It needs to be noted, though, that, in Catalonia landslides are typically unreported if no infrastructures, buildings or roads are affected, which makes the validation of the results challenging.

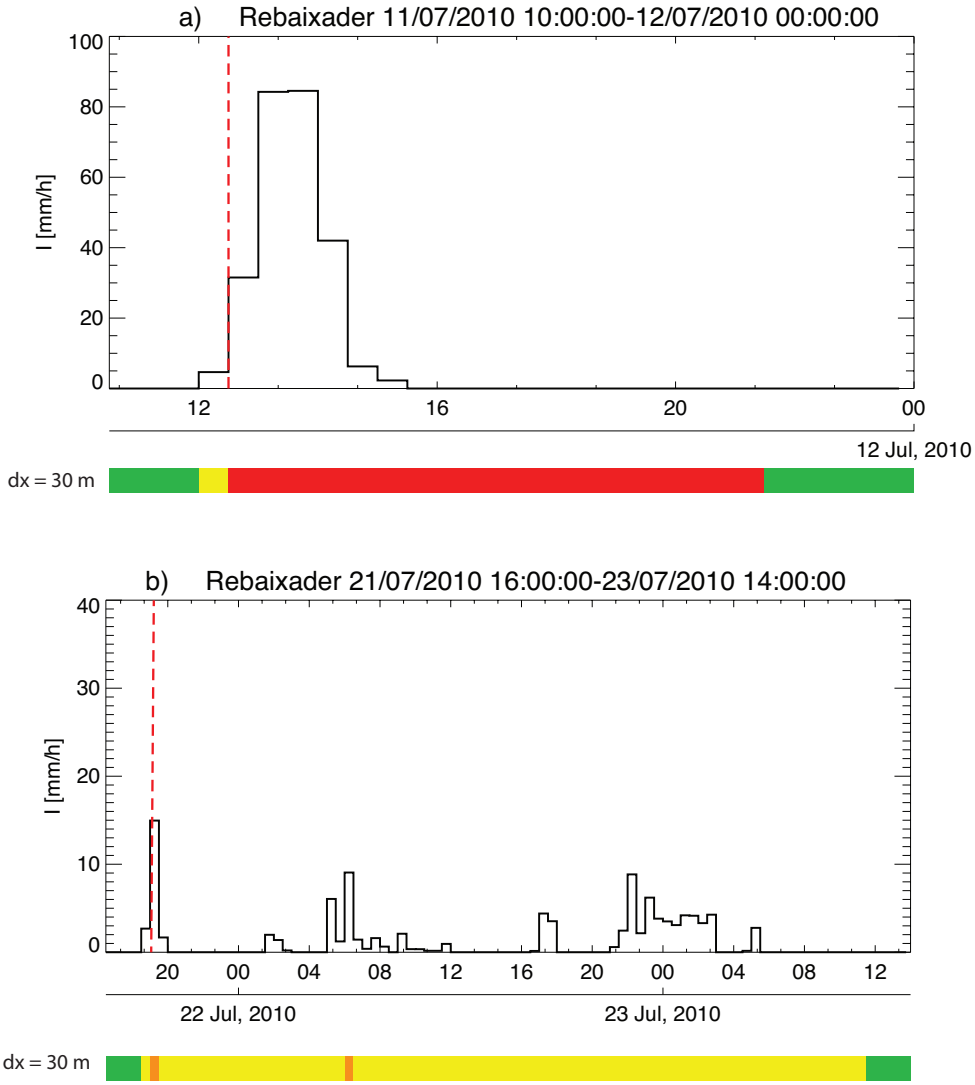


Figure 22: Time series of the warning level issued with the algorithm for two of the landslide events detected at the Rebaixader monitoring site. The black line represents the 30 min rainfall intensities estimated with weather radar in the subbasin: (a) 11 July 2010 and (b) 21 July 2010 landslides. The colour bars represent the maximum warning level observed within the catchment. Green, yellow, orange and red represent warning levels “very low”, “low”, “moderate” and “high” respectively. For each event, the red dashed line indicates the time when de monitoring station detected the landslide event.

Consequently, systematic validation has only been possible in two monitored subbasins (description of the monitoring systems present in these two sites can be

found in Palau et al., 2020; Raïmat et al., 2013; Hürlimann et al. 2014). In these two subbasins, the algorithm has been able to issue a “moderate” or “high” warning level coinciding with the time that four debris flow events were recorded at the two sites (an example is shown in (Fig. 22). However, in these catchments the results also show some false alarms (2 days where the monitoring systems did not detect significant landslides or debris flows).

Additionally, the analysis of the results has also studied the ability of the algorithm to identify three landslide events that affected infrastructures in two different locations.

As part of the study, Palau et al. (2020) have analyzed the sensitivity of the results on the mapping unit used: the algorithm has been run using two types of mapping units: subbasins (the domain was divided into subbasins using the Strahler stream order), and square pixels (configurations with pixels of 30 m and 200 m have been tested). The results obtained are similar for the tested configurations, and in general, all are able to issue warnings for the most significant reported landslide events. However, the high-resolution configurations tend to issue significant warning levels more often than the low-resolution configurations, which results in two false alarms for the former and one missed event for the latter.

All the parts of the algorithm are affected by uncertainties that can have an effect on its performance; namely, the susceptibility classification (both the variables and the rules used in the classification), the input rainfall observations and nowcasts, the rainfall thresholds and the definition of the rainfall hazard levels, and the thresholds used in the classification of the warning levels. Moreover, uncertainties also arise in the evaluation of the warnings at the locations where no landslide reports are available, or in locations where landslide reports are available, but the time of the event is not well known.

Finally, Palau et al. (2020) also discuss possible improvements of the algorithm to enhance its performance (e.g., considering the sediment availability or accounting for the antecedent soil moisture conditions).

3.2 Storm surges

First uncertainty aspects of using the Storm Surge Model at the pan-European scale are described followed by the model at the regional scale. The section concludes with the Inundation and Erosion Model.

3.2.1 European Storm Surge model (ESS) (CFR)

A pan-European Storm Surge Forecasting System (European Storm Surge model, ESS) has been implemented and validated during the ANYWHERE Project. It includes a coupled storm surge, tidal and wave model on an unstructured grid for Europe. The skills to predict tidal, surge and total water levels were evaluated based on measurements from 208 tidal gauge stations. The storm surge forecast system showed satisfactory performance for the two ECMWF atmospheric forcing datasets

tested: a High Resolution Forecast and the ERA-INTERIM reanalysis. For tidal predictions, the total Root Sum of Squares (RSS) is equal to 0.198 m. Thus, the predictive skills of the model outperforms the global tidal model FES2004 (RSS=0.423), but it is overtaken by the performance of the FES2012 (RSS = 0.148 m), which is however a data assimilation-based product. Storm surge validation results show good predictive skill, with RMSE between 0.04 and 0.21 m, and %RMSE within 4%–22% (Fig. 23). Coupling with tides resulted in improved storm surge level predictions, with RMSE reducing to 0.033 m. The North Sea and the English Channel areas benefited the most from coupling storm surge and tidal predictions, resulting in up to 2% reduction of the %RMSE. Increasing the resolution of atmospheric forcing also improved the predictive skills, leading to a reduction of RMSE up to 0.06 m in terms of extremes, especially in shallow areas where the wind is the main driver for surge production. We propose as optimal setup for operational pan-European storm surge forecasting the combination of tidal levels from the FES2012 model and storm surge residuals from the ESS setup which couples meteorological and astronomic tides. (Fernández-Montblanc et al., 2019).

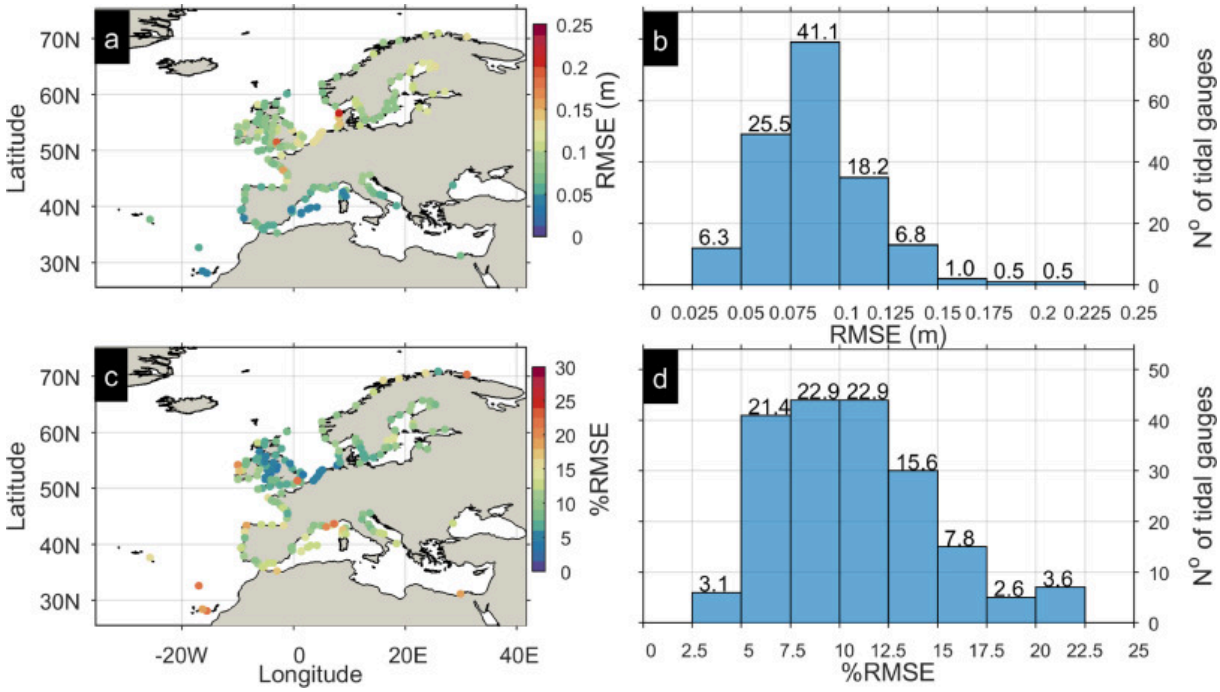


Figure 23: Validation of non-tidal residual (ηRE) results at the tidal gauge stations considered used for the coupled tidal and storm surge simulation (E2). (a,c) Map scatter plots of RMSE, %RMSE respectively; (b,d) RMSE and %RMSE histograms for all tidal gauges with the vertical axis showing the count and the text labels above the bars the percentage of all tidal gauges belonging to the specific bin. (Fernández-Montblanc et al., 2019)

In order to expand the testing of the storm surge model uncertainty a hindcast of Storm Surge Levels (SSL) for the period 1979-2018 was produced (ANYEU-SSL). The dataset covered 40 years (1979-2018) of SSL data along the European coastline with 3-hour temporal resolution. It has been extensively validated for the period spanning from 1992 to 2016, considering the whole time series, and extreme SSL values. Validation against tidal gauge data showed an average RMSE of 0.10 m, and RMSE below 0.12 m in 75% of the tidal gauges. Comparisons with satellite altimetry data (Fig. 24) showed an average RMSE of 0.07 m.

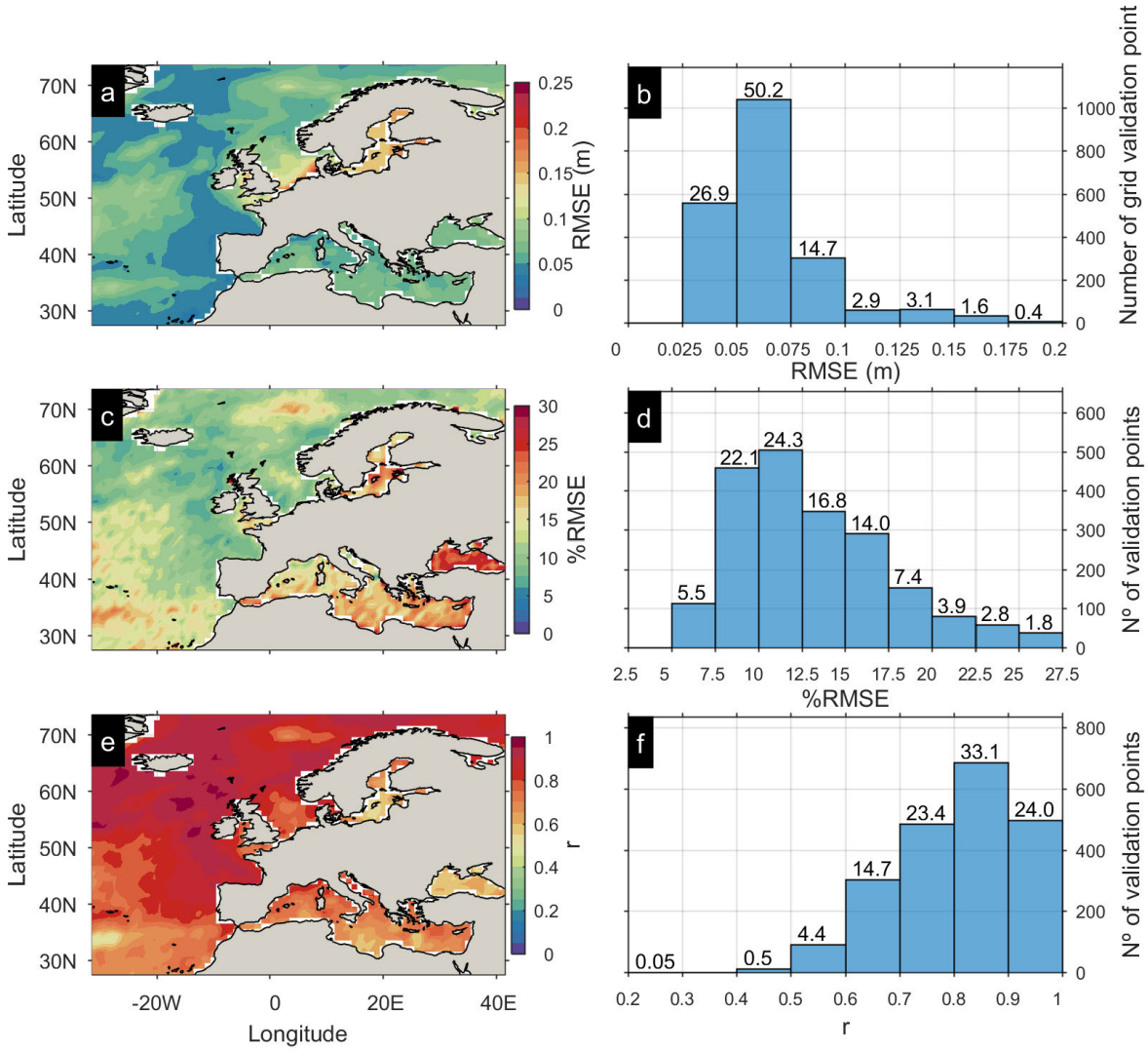


Figure 24: Hindcast validation against satellite altimetry data. (a, c, e). Map scatter plots of RMSE, %RMSE and correlation coefficient (r), respectively; warm colours indicate higher performance; (b, d, f) RMSE, %RMSE and r histograms for all validation cells with the vertical axis showing the count and the text labels above the bars the percentage of all validation cells (1°x1°) belonging to the specific bin. (Fernández-Montblanc et al., in press)

As a potential application of the hindcast dataset, the storm surge trends that might have occurred in last decades along European coastline, were evaluated. The results showed a trend in extreme storm surge magnitude that was controlled by latitude for the period 1979-2016. SSLs appeared to increase in areas at latitudes, exceeding 50 °N, while there was a decrease at low latitudes. Additionally, a seasonal variation of the extreme SSL, particularly strong in the northern areas, has been observed. The dataset is publicly available and aims to provide an important data source for the study of storm surge phenomena and consequential impacts, either at the large or local scales. (Fernández-Montblanc et al, in press).

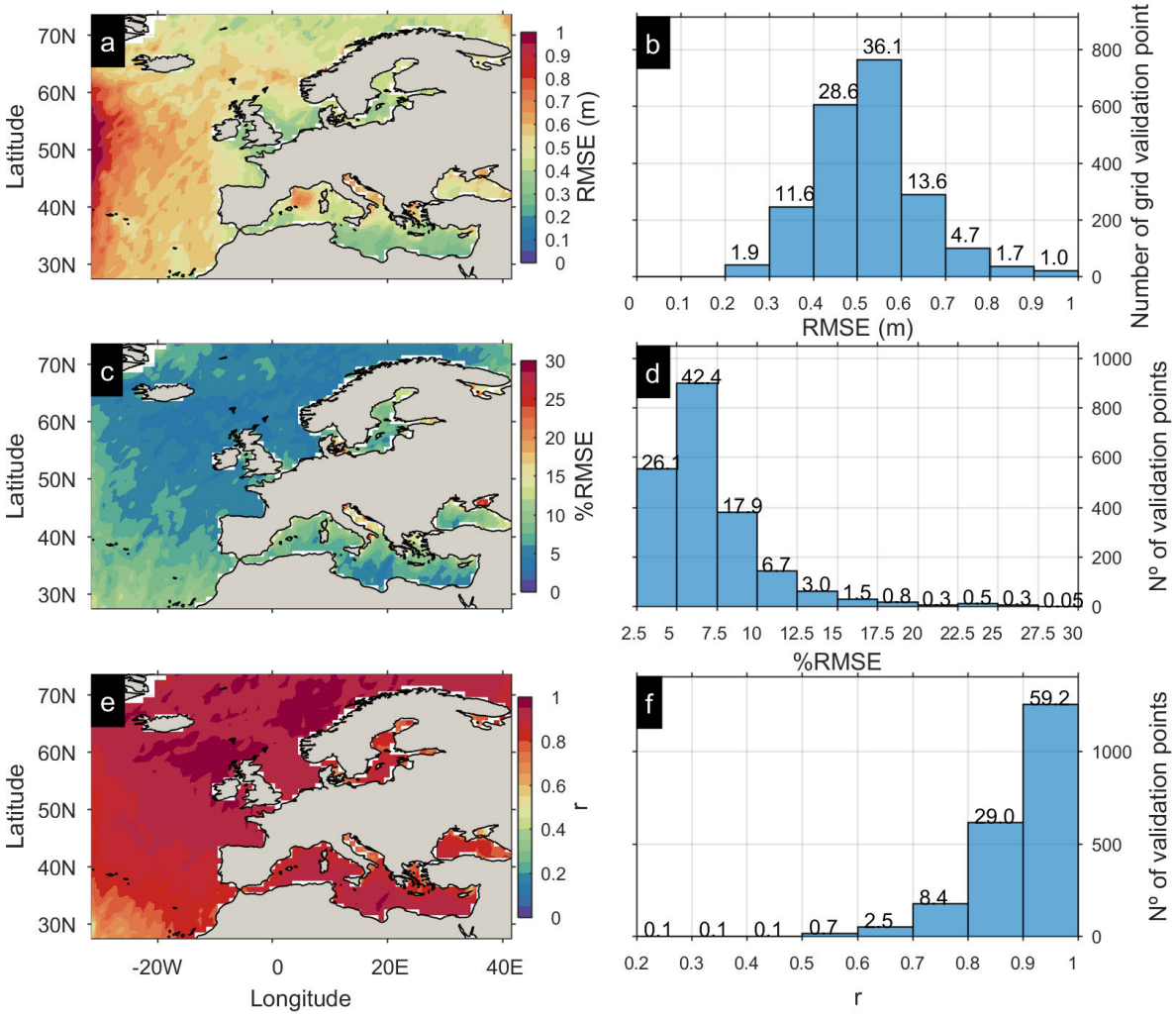


Figure 25: Modelled significant wave height versus satellite altimetry data. (a, c, e) Map scatter plots of RMSE, %RMSE and r, respectively; warm colours indicate higher performance; (b, d, f) RMSE, %RMSE and r histograms for all validation cells with the vertical axis showing the count and the text labels above the bars the percentage of all validation cells (1°x1°) belonging in the specific bin.

The model ability to reproduce significant wave height (SWH) was evaluated by comparison with measurements from 224 wave buoys measurements and satellite measurements for the period 2011 to 2013, using ERA INTERIM, ERA5 reanalysis and high resolution forecast as atmospheric forcing. Skill scores were calculated as a measurement of the model uncertainty to reproduce SWH. The RMSE of SWH was lower than 0.5 for more than 50% of the buoys. The comparison with satellite measurement of SWH showed RMSE around 0.4-0.5 m in most of the validation locations and relative RMSE (%RMSE) below 10% (Fig. 25). The effect of coupling waves, tide and surge was also investigated. Results indicated an improvement of both storm surge level and significant wave height predictions. This improvement was more evident for the storm surge level. When waves were included, the underestimation of extreme SSLs, previously highlighted, was consistently reduced.

In order to reduce the model's uncertainty, it was assessed if the model's shortcomings were related to the atmospheric forcing resolution/accuracy. The results revealed that scatterometer wind data from satellite measurement can be used for statistical corrections of the wind forcing used in the model. Model experiments has demonstrated that the quantile mapping technique for bias correction improved the SSL (up to RMSE reduction of 0.08 m) and SWH (up to RMSE reduction of 0.3 m) predictions, considering the extremes events of surge and waves and in semi-enclosed areas (Baltic and Mediterranean Sea). On the other hand, the contribution by the expansion of the water volume (steric effect) to extreme water levels has been analyzed using altimetry data from satellite measurements. Results have shown that the steric effect could account for around 40% of the non-tidal residual in the Mediterranean and the Black Sea, and 25% in the Baltic Sea. This highlights the importance to incorporate this specific physical process into the algorithms through numerical modeling or data assimilation.

3.2.2 Regional Storm Surge model (CFR)

The total water levels outputs provided by the storm surge regional model has been validated on the basis of historic events. Extreme events, with water level exceeding the 5 years return period were selected from the time series of the water level measured at the Stavanger tidal gauge. The model's performance to reproduce the total water level was satisfactory. The absolute errors of the maximum total water levels during the selected extreme events ranged between 0.05 and 0.1 m. The results highlighted an underestimation of maximum water level. The larger uncertainty was linked to the tidal component of the total water level.

3.2.3 Inundation and erosion model (CFR)

The inundation model for the Stavanger municipality was validated using the historic event occurred on 8 December 1994. As direct inundation measurements were not available, the predicted inundation extent and water level were compared with visual

estimation of the flood depth. Thus, the documentation of the flooding event (pictures and videos) available online was visually analyzed and compared with model's outputs. The inundation level at the base of the first line of buildings in the city center was estimated in the order of ~0.15-0.2 m. The simulated flood depth at that location was on average 0.19 with a standard deviation of 0.03. Along the promenade, flood depths were estimated as ~0.2-0.25 m, while the simulated ones were on average0.23 m with a standard deviation of 0.03 m (Duo et al., in press).

3.3 Heatwaves and air quality (weather-induced health)

Uncertainty of using algorithms/tools to forecast heatwaves and air quality products is explained. These products are associated with weather-related health.

3.3.1 Universal Thermal Climate Index (UTCI) (UOR)

The uncertainty of UTCI forecasts is due to the uncertainty of ECMWF inputs variables– namely 10 metre wind speed, 2 metre relative humidity, 2 metre temperature, and solar radiation (Section 2.1.1) – used for the computation of UTCI forecasts. It has been assessed and described in Deliverable D2.2 (Ballesteros Cánovas et al., 2017) on the basis of the work by Pappenberger et al. (2015) and here is briefly reported. Using reanalysis climate data (1979-2009) at the global scale the UTCI was been shown sensitive to all input variable with: (i) some linear dependencies on air temperature, (ii) a distinct lower boundary for wind (> 17 m/s) justified by the clothing insulation and vapour resistance caused by body movements and the wind itself (Havenith, et al., 2012), and (iii) a lower boundary influenced by the solar elevation angle, solar radiation and thermal radiation (Fig. 26).

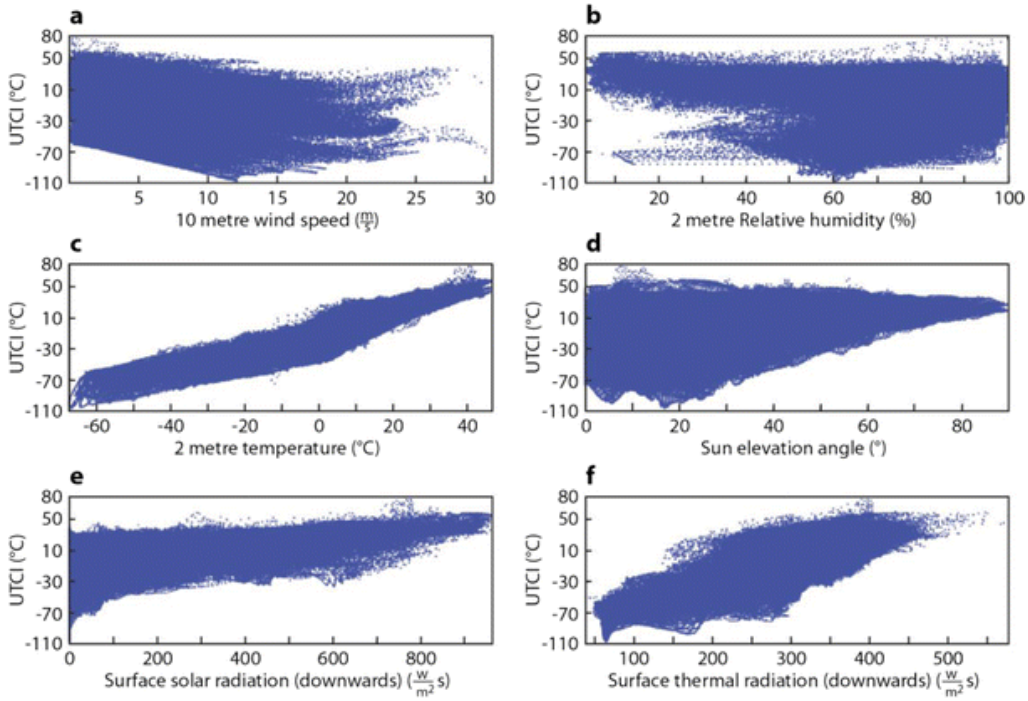


Figure 26: Scatterplots of meteorological inputs against UTCI to illustrate associated dependencies (Pappenberger, et al., 2015).

Similarity between forecast and observed UTCI was also analysed. UTCI forecasts were calculated every day with a lead time of 10 days from 1 January 2009 to 31 December 2012 using both the ECMWF HRES and ENS inputs. The deterministic high-resolution, control and ensemble mean forecasts of the UTCI were then compared with observation using the Anomaly Correlation Coefficient (ACC). Pappenberger et al. (2015) found that the maximum lead time for which ACC stays above 60% is 4-6 days in the Mediterranean Basin (30°N–48°N, 10°W–40°E) and 4-6 days for the ensemble mean forecast in Northern Europe (48°N–75°N, 10°W–40°E). For the latter the deterministic high resolution and control forecast have ACC above 60% for less days, namely 2-4 days (Figure 27).

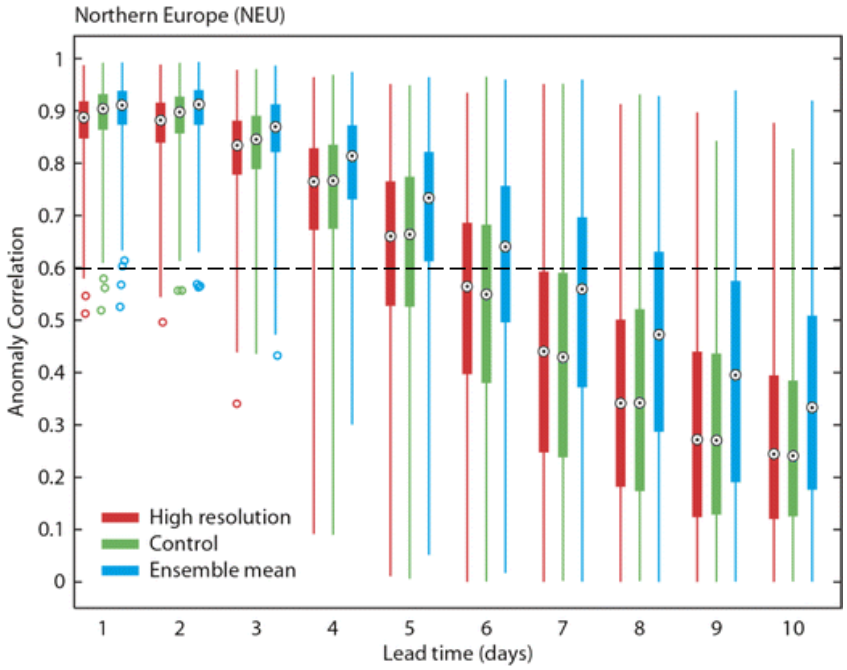


Figure 27: Anomaly correlation for the Northern European area and the three different UTCI forecasts. The *circle* illustrates the mean whilst the *box* indicates the 25th and 75th percentile. The *whiskers* of the box plot extend to the 95th and 5th percentile. *Coloured circles* indicate outliers. (Pappenberger, et al., 2015).

The skill of UTCI forecasts in the prediction of strong heat stress (>32°C) was assessed via the Brier Skill Score (BSS). The lead time at which the BBS drops below zero, i.e. there is no skill compared to climatology, is highest for ensemble predictions rather than the control and the high-resolution forecasts (Pappenberger, et al., 2015).

Studies have also been conducted to assess the UTCI capability to represent thermal perception and associated uncertainty. By definition, the UTCI is universal (Błażejczyk et al., 2013). It has therefore been evaluated across different climate regions (artic, moderate, dry and wet subtropical), as well as on spatial and temporal scales from the micro through to the macro (Coccolo et al., 2016). For instance, Park et al. (2014) compared the thermal perception as expressed by the UTCI with the thermal perception as represented by other thermal indices (as PET, PMV and SET*). Using two different locations – Nanaimo, BC, Canada and Changwon, Republic of Korea – as test areas they demonstrated a large correlation factor exists ($R^2 = 0.95$).

Another correlation that has been tested for the UTCI is the one with health impacts, namely heat-related mortality. Di Napoli et al. (2018), for instance, showed that the strength of the UTCI-mortality correlation depends on bioclimatic conditions. This has been proved both at country (17 different European nations) and city (Paris) level. This correlation is of a particular importance as it is related to the usefulness of the UTCI and UTCI forecasts to predict heat-related mortality. Future, quantitative assessment of the UTCI uncertainty with respect to health impacts will help to shed new lights on this.

3.3.2 Regional Air Quality (RAQ) (UOR)

The uncertainty of RAQ forecasts is inheritably due to the uncertainty of the prediction system they are based upon, namely the ECMWF-IFS (Section 2.1.1) coupled to a seven chemical transport models (see Deliverable 2.3 for further details, Ciavola et al., 2017). The uncertainty of RAQ forecasts has been evaluated against the near-real time air quality (NRT AQ) surface monitoring data as described by Marécal et al. (2015). By considering two high ozone episodes occurred in Europe between 10 and 13 June 2014 as an example, the authors found: (i) a good consistency of the diurnal ozone variations provided by the seven models compared to the NRT AQ station observations from affected areas (Austria, Hungary, Germany and France), and (ii) ozone values from the ensemble mean close to observed ones with the largest deviations due to orography (Fig. 28).

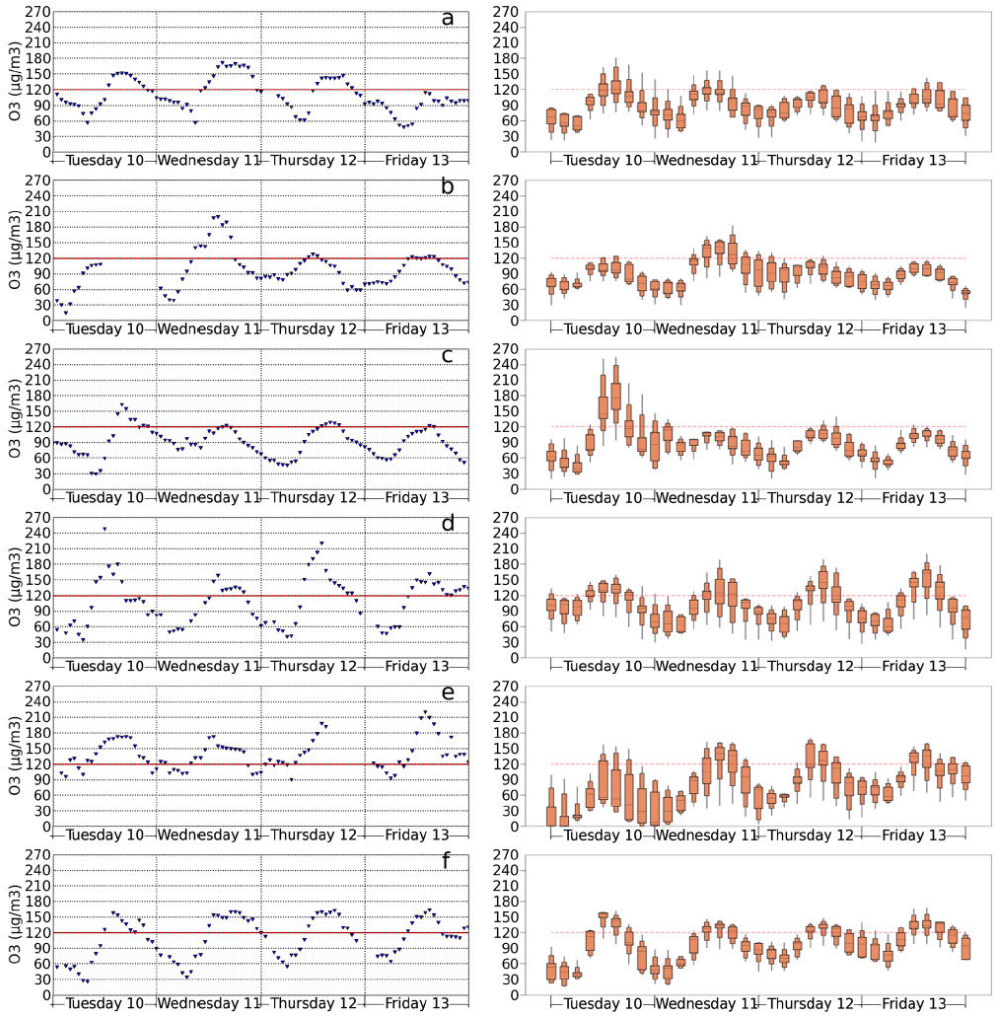


Figure 28: Left panels: ozone measurements from surface stations from 10 June 2014 at 00:00 UTC to 14 June 2014 at 00:00 UTC, located in 6 areas affected by the high ozone episode (Austria, Hungary, Germany, France). Right panels: EPSgrams giving median, 90th percentile, 75th percentile, 25th percentile, 10th

percentile, minimum and maximum from 3 h outputs of the 96 h forecasts of the seven models from 10 June 2014 at 00:00UTC to 14 June 2014 at 00:00 UTC. Model outputs are interpolated at the location of the stations shown in the left panel (Marécal et al., 2015).

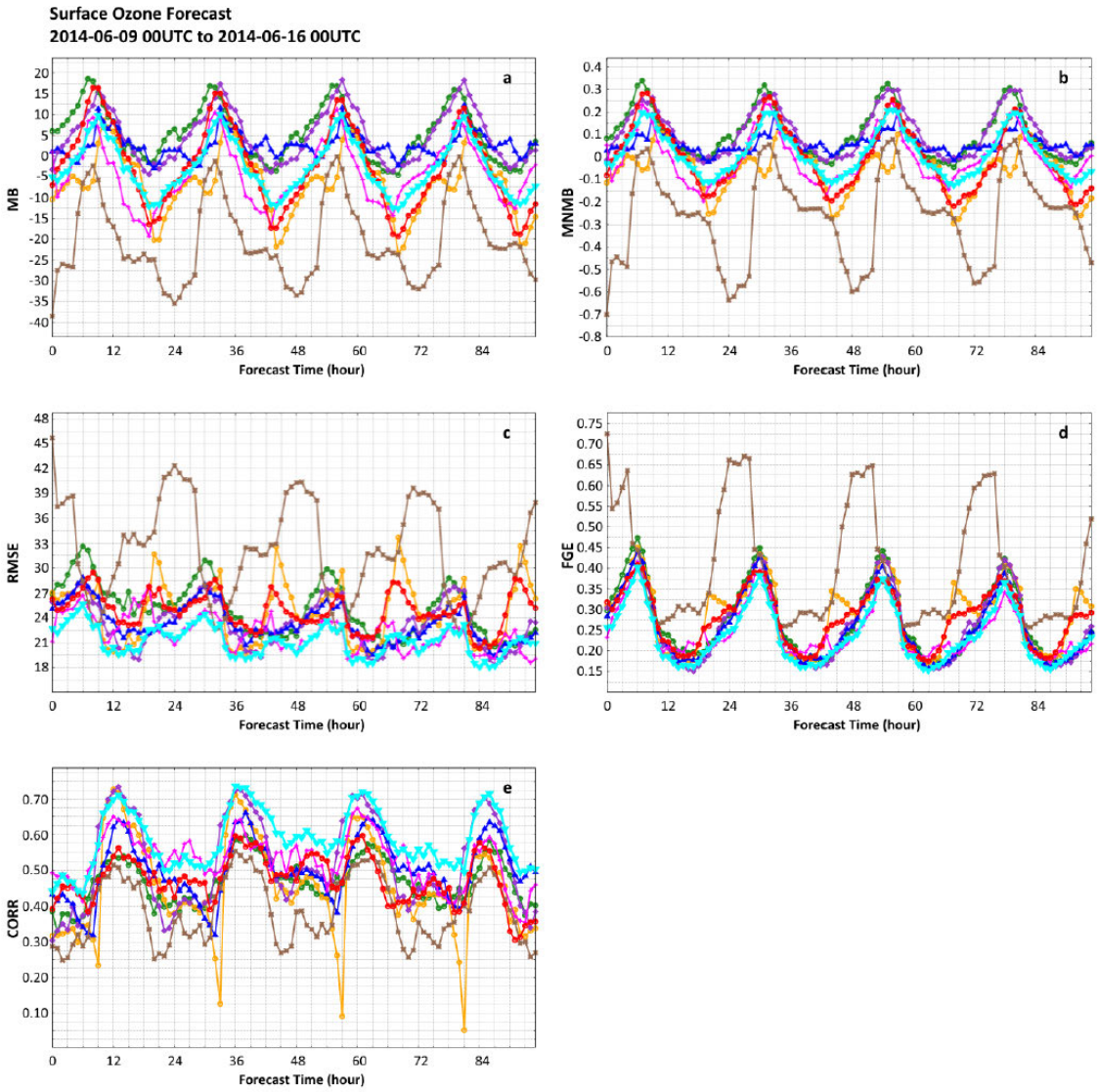


Figure 29: Statistical indicators for ozone as a function of the forecast time (in hours) for the ensemble median (in turquoise) and the seven models (other colours) compared to the hourly surface station measurements available for the period from 9 to 15 June 2014 over the MACC-II European domain. (a) MB (in µg m^{-3}), (b) MNMB, (c) RMSE (in µg m^{-3}), (d) FGE and (e) correlation (Marécal et al., 2015).

For further evaluation, four statistical indicators - mean bias (MB), modified normalised mean bias (MNMB), the root mean square error (RMSE), fractional gross error (FGE) and correlation coefficient (R) - of ozone concentration forecasts from the seven individual models and the ensemble model (ENSEMBLE) were calculated using representative observations available over the whole European domain and

consecutive 96 h forecasts run every day from 9 to 15 June. Figure 29 shows that there is a spread of the seven models and that the ENSEMBLE generally gives the best scores with MNMB between 0.2 and -0.1, FGE between 0.15 and 0.4 and correlations up to 0.75 during daytime.

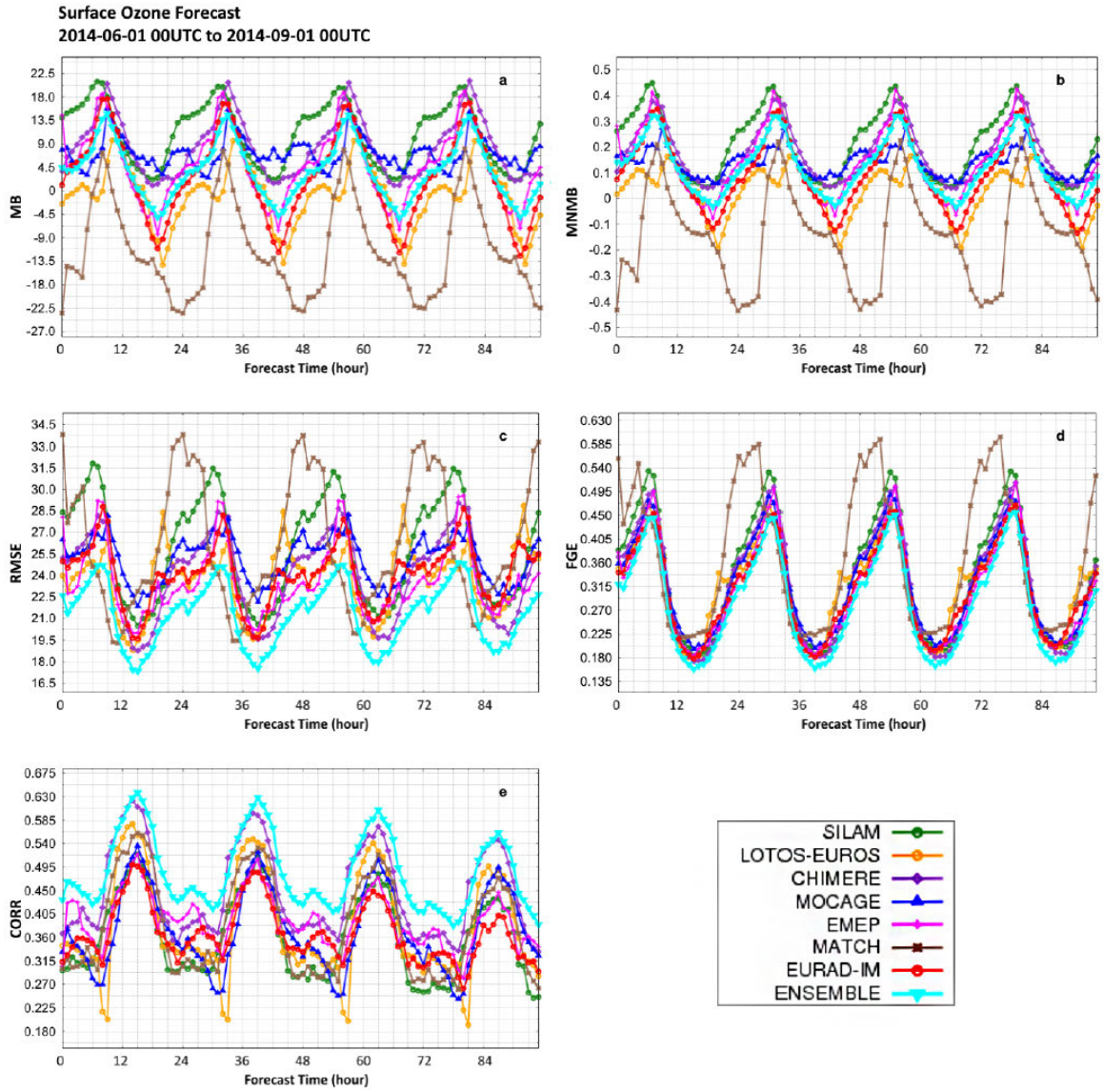


Figure 30: Statistical indicators for ozone as a function of the forecast time (in hours) for the seven models and the ENSEMBLE compared to the hourly surface station measurements available for the period from 1 June at 00:00 UTC to 1 September at 00:00 UTC over the MACC-II European domain for 2014: (a) MB (in µg m^{-3}), (b) MNMB, (c) RMSE (in µg m^{-3}), (d) FGE and (e) correlation (Marécal et al., 2015).

All the models, including the ENSEMBLE, exhibit a diurnal cycle with higher correlations and lower RMSE and FGE during daytime (when ozone is high) than during night-time. Five of the models have a positive MB on average and the other two a negative MB on average. This confirms that the ENSEMBLE, which uses all seven

models, performs generally better on all statistical indicators. For this reason, the output of the ENSEMBLE, i.e. the ensemble forecast median of ozone and other pollutants' concentrations (namely nitrogen dioxide, sulphur dioxide, carbon monoxide, particulate matters), has been chosen for connection to the ANYWHERE platform.

In addition to the production of the daily skill scores just described, statistical indicators were calculated for ozone, NO2 and PM10 at the surface on a seasonal basis for each of the seven models and for the ENSEMBLE. The model's seasonal statistical indicators are also calculated against measurements from the European air quality surface station network available in near-real time.

3.4 Weather-induced fires[6]

The uncertainty aspects of the European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF) is explained.

As mentioned in detail in D2.2 (Ballesteros Cánovas et al., 2017), forecasting wildfires is a complex task because ignition location and time of occurrence cannot be easily predicted, especially when the trigger is due to human behavior (e.g. arson). Therefore, forecasting wildfires means quantifying how dangerous fires could be assuming an ignition occurred and knowing weather conditions from advanced numerical weather models. The European Centre for Medium-range Weather Forecasts (ECMWF) developed and now maintains the modelling engine that works as back-end of the European Forest Fire Information System: the Global ECMWF Fire Forecast (GEFF) model. EFFIS-GEFF is designed to identify fire-favorable weather conditions which can allow sustained fire activity.

The system generates two types of data products: a global reanalysis dataset and daily real-time forecasts up to 10 days ahead. Two types of forecasts are generated: a 9 km deterministic forecast (also called high resolution forecast or HRES) and an 18 km probabilistic forecast (comprising of 51 ensemble members, also called ENS). A detailed description of the reanalysis dataset (based on ERA-Interim), definition of danger levels at country and regional levels and their validation, were already discussed in D2.2 (Ballesteros Cánovas et al., 2017) and have been recently published (Vitolo et al. 2018; 2019a).

Here we focus on the forecast products and their assessment. The forecasting system is described in Di Giuseppe et al. (2016). As the system, developed for the Copernicus Emergency Management Service at ECMWF, has undergone substantial improvements, a new validation at global scale is currently in progress (Di Giuseppe et al., 2019). As results are not published yet, we provide below a description of the

[6] There is no report on the uncertainty aspects of RISICO - fire danger rating system (CIMA) and PROPAGATOR - propagation of a wildfire (CIMA).

methodology currently used to validate wildfire danger forecast products at various scales, from continental to local level.

Each dataset contains numerous fire danger indices, however, the most widely used in Europe is the Fire Weather Index, which is also used herein to assess the performance of the fire danger forecasting system. Below, we predict the FWI values and the probability of detection of fire during 2017. Our assessment focuses on two types of skills: (i) to predict FWI as close as possible to the one calculated using station information, and (ii) to predict fire events, given real occurrences.

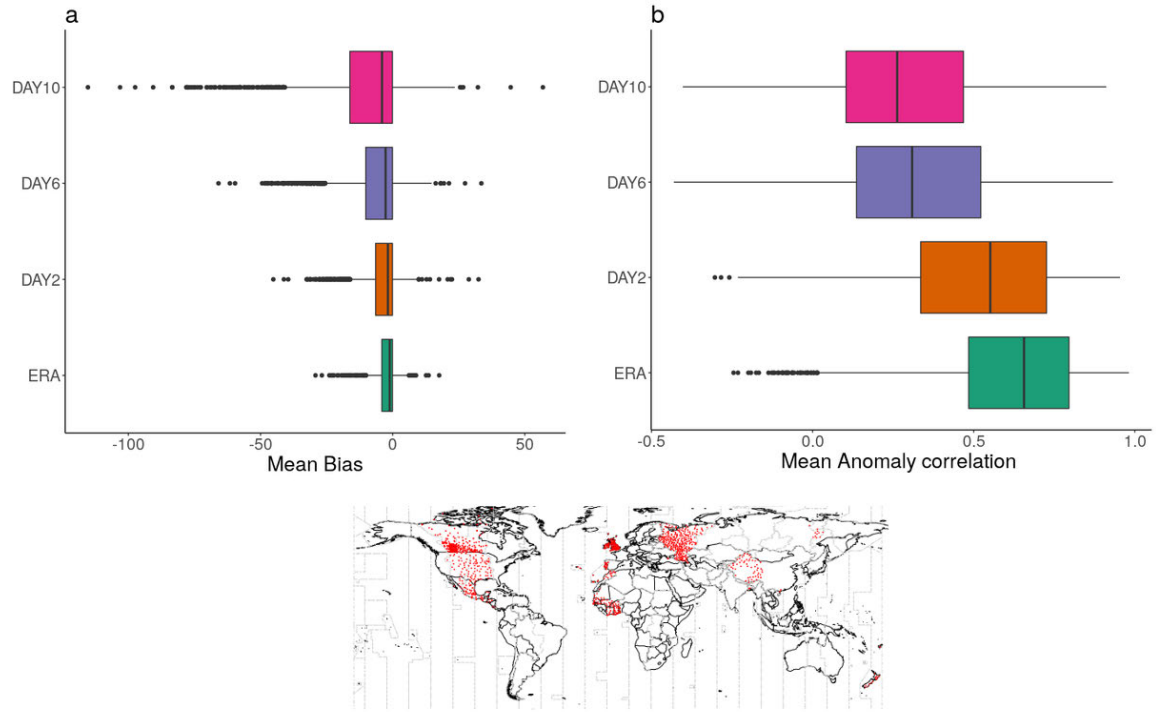


Figure 31: Comparison between modelled FWI and observed FWI value. FWI are calculated using ECMWF reanalysis (ERA-Interim) and HRES forecasts at different lead times. Observed FWI are obtained using recorded weather measurements on the identified SYNOP stations which have at least 30 observation recorded at local noon in 2017. The box plots are used to describe the distribution of mean bias (panel a) and anomaly correlation (panel b) across the observation points in 2017 2017 (red dots in the map at the bottom). Vertical lines show time zones edges.

Skill in predicting FWI

As many forest agencies still rely on observed meteorological data to provide fire danger, the first assessment looks at the capability of the ECMWF fire forecast to reproduce the same FWI values as would be estimated from a network of local stations, 2, 6 and 10 days ahead. Despite several meteorological observations are available through the SYNOP network, only a subgroup has at least 30 days of recordings at local noon during 2017. These selected stations are used to perform an analysis of bias and anomaly correlations at different lead times (Fig. 31). For comparison also

FWI calculation using ERA-Interim is included. This provides an estimation of the limit of predictability when using forcing from model simulations in place of observed values (Di Giuseppe et al., 2016). As expected, there is a performance degradation going towards longer lead times and mean biases (Fig. 31a) are limited to few units even at day 10. However even few units could mean a mismatch in danger level. The anomaly correlations (Fig. 31b) provide information on the capability of the forecasting system to predict high and low anomalies, where the mean value is taken as the climatology. This metric, by comparing anomalies implicitly, removes existing biases that could exist between the forecast and the verifying field. An anomaly correlation coefficient (ACC) below zero could indicate, for instance, a "dangerous" forecast where a low anomaly is forecast in place of a high anomaly event. Values above 0.6 are usually considered skillful. FWI from reanalysis have the largest skills as expected and the mean anomaly correlation rapidly falls below the 0.6 threshold by day 6. However, the distribution of ACC values clearly shows that in selected cases predictive skills can be achievable even at day 10.

Skill in detecting fire events

While national inventories of wildfire activities exist in many countries, they can be heterogeneous and lack the temporal span desirable for the validation of a fire danger system at the global scale. Satellite observations can supply a valid alternative especially as they cover remote areas where in-situ observations are sparse (Flannigan and Haar 1986; Giglio et al., 2003; Schroeder et al., 2008). Daily maps of fire radiative power (FRP) (Kaufman et al., 2003; Wooster et al., 2005) are available from ECMWF since 2003 through the Global Fire Assimilation System (GFAS) (Kaiser et al., 2012; Di Giuseppe et al., 2017; Di Giuseppe et al., 2018). This dataset has been developed in the framework of the Copernicus Atmosphere Monitoring Services (CAMS) and uses observations from the MODIS sensors on board of Terra and Aqua satellite platforms and assumptions on fire evolution to calculate a continuous record of active fires. The GFAS dataset integrates all available FRP observations from the MODIS sensors available in a day over a regular 0.1 deg grid. It therefore provides an indication of the cumulative dry mass available for burning which can be then put into a relationship with fire emissions. In this exercise the FRP products are used to identify fire occurrence. This is achieved by using a minimum detection, FRP> 0.5 Wm^-2 (Kaiser et al, 2012). A "hit" is recorded if the fire forecast predicts high fire danger when a fire really occurred. The model performance is provided as an average over the selected region even if the calculation of the various scores is performed at pixel level.

For an assessment at the continental scale, we use the fire macro-regions defined by the Global Fire Emission Database, GFED4 (Giglio et al, 2013). These macro-regions are characterized by different fire regimes and are very roughly homogeneous in their burning emissions contribution (Giglio et al, 2013). In Figure 32, we use one year of operational service in 2017 to showcase the potential of the use of weather forecasts to support the monitoring of fire danger conditions and planning in case of a potential emergency. By applying a model-based definition of warning levels to the FWI we have shown that it provides a probability of detection (POD, Annex II) for fire activity in

Europe that is above 60%, even at day 10. Mid and high latitude forested areas, where fuel is abundant have the highest predictability while in savanna/shrub-land regions the relationship between FWI and fire occurrence weakens.

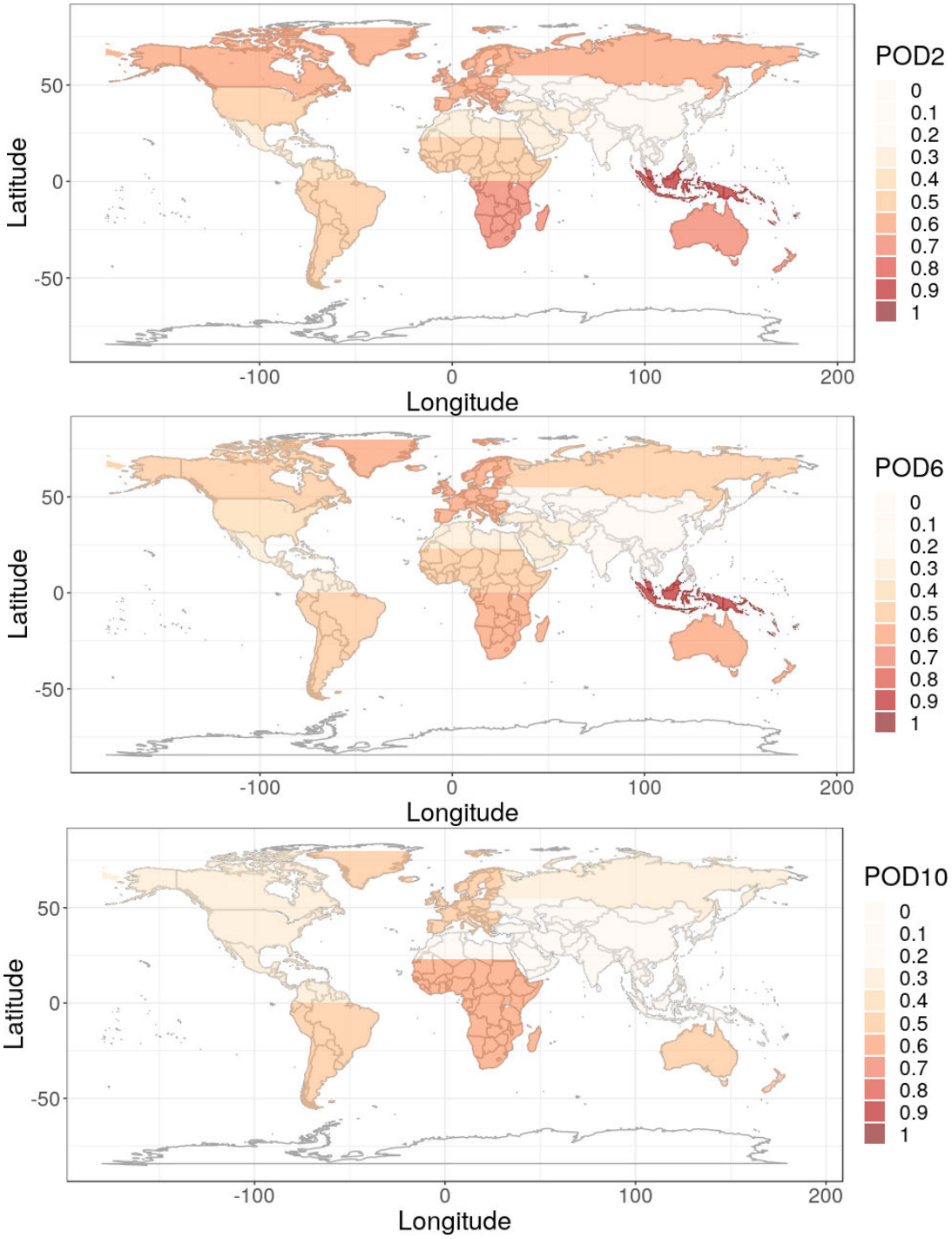


Figure 32: Global area averaged Probability of Detection (POD) for day 2-6-10 forecasts. Pixels where FRP>0.5 Wm^-2 are categorized as 'significant events' and compared to FWI prediction above the high warning level. The global statistic is constructed using all FRP observations detected in 2017 and averaged over the specified regions.

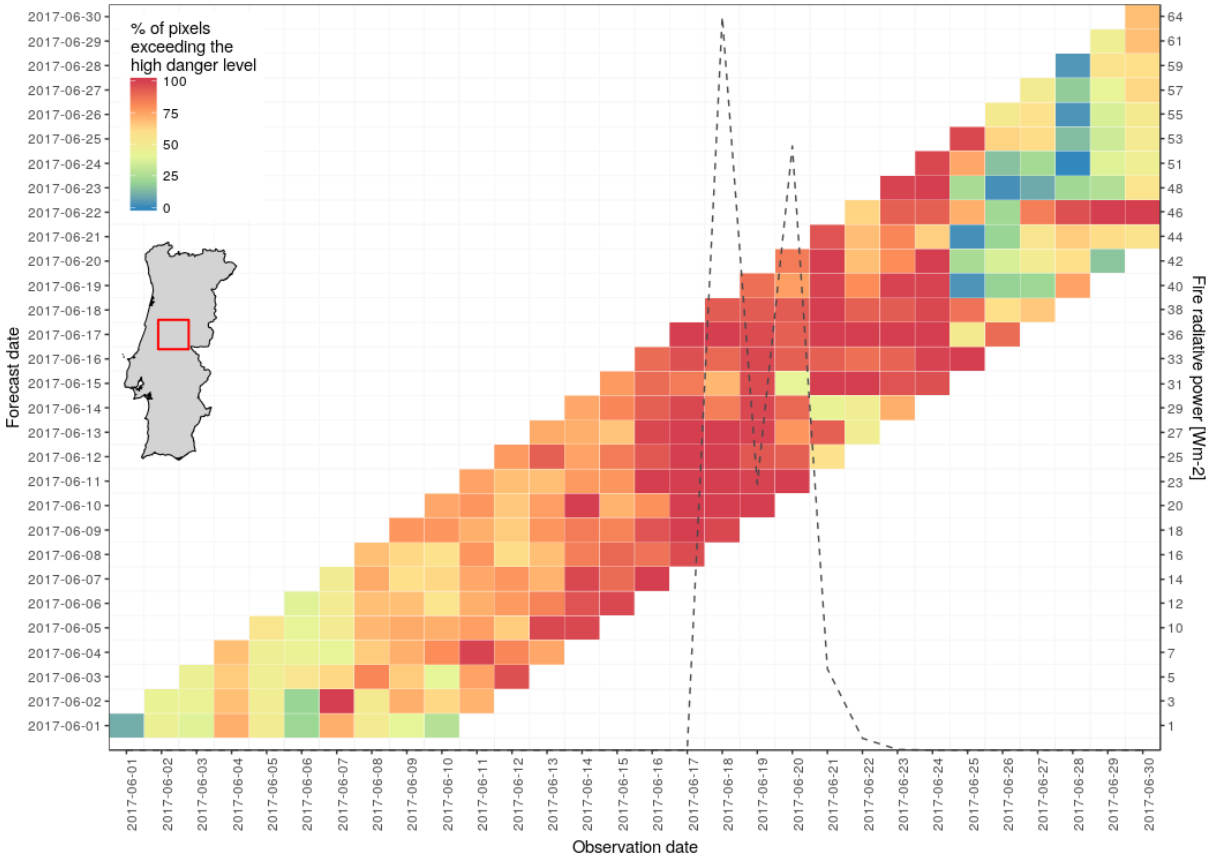


Figure 33: Monthly-summary plot - comparison of Fire Radiative Power (gray dashed line with axis on the right-hand side) with FWI forecasted using the deterministic high resolution model for Portugal during the 2017 June event. FWI is colour coded based on the percentage of pixels exceeding the high danger level calculated at the country/state level. Each of the panel refers to a specific fire event described in the text and the statistics have been calculated over the red box.

For an assessment at the local level, we selected Portugal - which experienced recurrent intense fire episodes and saw major events taking place in 2017. We looked at the Pedrógão Grande fire in June 2017. The affected area is identified as the area including all detected active fires (cells with FRP > 0.5 Wm^-2) during the selected time window. Figure 33 shows the monthly-summary plot of forecasted information that could have been provided for the study areas by the 10-day fire danger high resolution forecasts (HRES), had these been already available. The plot shows on the x-axis the dates in which FRP was observed and, on the y-axis, the dates forecasts were issued. The cell in the bottom left corner shows the percentage of pixels in the study area that were expected to be above the high danger threshold on the first day of the first forecast issued. The forecasts for day 2 to day 10 are on the same row. The forecasts issued on the following day are one row above and so forth. The dashed lines show the observed fire radiative power (see also secondary y-axis).

The reader is reminded that active fires are triggered by highly unpredictable events (ignition) which are not accounted for in the FWI system. The FWI is not supposed to provide the exact localization of the event but an indication of anomalous conditions that could potentially lead to dangerous fire activities. Large areas can be affected by anomalous conditions in the proximity of when the event really occurred. However, it is noticeable the capability of the forecast to detect the increase in fire danger associated to the event. From June 10 over 80% of the area exceeded the high danger threshold (FWI > 20). The FRP spikes (occurred on 17-20 June) highlighted that most of the region was classified at very high danger 10 days ahead. In this case, the persistency of high danger conditions in a given region for successive forecasts tends to increase confidence in the forecasts.

3.5 Droughts[7]

Drought forecasts have been built upon seasonal probabilistic hydrological forecasts (Section 2.2), which are driven by probabilistic weather forecasts issued by ECMWF-IFS (Section 2.1.1), see also D2.3 (Ballesteros Cánovas et al., 2017) and D2.4 (Van Lanen et al., 2019). These hydrometeorological forecasts are obtained from the European Flood Alert System (EFAS). Although EFAS originally was developed for flood forecasting (Section 3.1), it has been revised to improve simulation of hydrological time series under dry conditions. The hydrological model driven by the weather forecasts reproduces the observed low-flow statistics reasonably well, with a general tendency of better performance with increasing catchment size (Feyen and Dankers, 2009). They validated drought characteristics derived from simulated river flow against drought characteristics obtained from observations at hundreds of gauging stations across Europe for which long enough daily river flow time series were available and analyzed: (i) average annual minima, (ii) average river flow deficits, (iii) minimum flows with return period of 20 years, and (iv) maximum deficits with return periods of 20 years. Time series of hydrological variables simulated with the hydrological model in EFAS and the input time series (e.g. precipitation) are the basis for the drought identification. These include the so-called standardized drought indices (Section 3.5.1) and the threshold-based indices (Section 3.5.2) (e.g. Van Loon, 2015). These indices have been calculated using historic hydrometeorological data (from 1990 onwards) and forecasted in a probabilistic way up to 7 month ahead. This section concludes with algorithms/tools addressing areal drought indices (Section 3.5.3). One way to address uncertainty in drought forecasting is by using the ensemble forecasted time series of hydrometeorological variables provided by EFAS. For each drought product (PRD-148 to PRD-160), percentiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th}) are presented, which are obtained from 51 ensemble members (see Table 1, Sutanto et al., 2019a).

[7] There is no report on the U&R aspects of using the European Drought Observatory (EDO) because JRC had no person months to work on this activity.

3.5.1 Standardized Indices (WUR)

The analysis of drought forecast uncertainty using the standardized indices has been carried out at the pan-European scale. The study has been published (Sutanto et al., 2019a).

Season	Class difference	SPI-1 with lead times of (month)					SPI-3 with lead times of (month)					SPI-6 with lead times of (month)					SPI-12 with lead times of (month)				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Winter (DJF)																					
	none	49.8	50.5	48.4	51.8	48.5	67.4	57.3	51.0	51.8	49.4	80.3	70.0	62.5	58.5	54.5	86.2	79.2	74.3	70.6	66.8
	+1	15.3	13.1	8.4	6.4	7.0	10.8	12.0	11.1	7.1	4.0	7.0	9.1	8.9	7.6	5.7	4.9	6.0	5.5	5.1	4.6
	+2	22.8	24.5	29.0	28.4	28.2	17.6	22.1	26.6	29.3	33.0	12.0	17.7	22.4	25.8	29.4	8.7	13.8	17.9	20.8	24.0
	+3	1.3	0.3	0.2	0.1	0.2	1.0	1.1	0.7	0.3	0.2	0.2	0.6	0.5	0.5	0.3	0.1	0.2	0.2	0.2	0.2
	+4	6.2	6.6	7.6	7.1	8.1	2.5	5.2	6.8	7.2	8.2	0.5	2.2	4.5	5.7	6.5	0.1	0.8	1.9	2.9	3.6
	-1	0.2	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
	-2	3.3	3.5	3.9	3.6	4.6	0.3	1.8	2.8	3.1	3.6	0.0	0.2	1.1	1.9	2.8	0.0	0.0	0.1	0.3	0.7
	-3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	-4	1.1	1.6	2.5	2.6	3.3	0.0	0.3	0.8	1.2	1.6	0.0	0.0	0.1	0.3	0.7	0.0	0.0	0.0	0.0	0.0
Spring (MAM)																					
	none	53.7	49.7	47.9	45.2	43.4	67.8	54.0	50.3	47.3	45.9	76.3	64.8	58.9	53.2	52.3	85.5	76.5	70.2	65.4	62.0
	+1	7.3	6.2	12.0	18.7	23.2	5.8	5.3	5.0	10.8	17.3	4.3	3.8	4.8	6.5	9.2	2.9	3.2	3.9	5.8	7.8
	+2	26.4	28.7	26.2	22.9	20.9	23.6	29.8	31.9	28.9	22.2	17.9	26.1	27.9	27.4	26.5	11.4	19.0	22.8	23.8	23.9
	+3	0.4	0.2	0.8	1.1	1.5	0.3	0.3	0.4	1.0	2.2	0.1	0.2	0.2	0.4	0.8	0.0	0.1	0.2	0.3	0.5
	+4	6.5	8.1	7.3	6.6	6.1	3.6	6.9	7.4	7.0	6.0	1.2	4.2	5.7	6.2	6.0	0.2	1.2	2.7	3.8	4.3
	-1	0.1	0.0	0.2	0.4	0.3	0.0	0.1	0.1	0.2	0.5	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.1
	-2	3.5	4.3	3.9	3.5	3.1	0.7	2.8	3.4	3.3	2.8	0.1	0.8	2.0	2.9	3.1	0.0	0.1	0.3	0.8	1.3
	-3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
	-4	1.4	2.7	1.7	1.8	1.4	0.0	0.7	1.4	1.5	1.4	0.0	0.0	0.4	1.1	1.8	0.0	0.0	0.0	0.0	0.1
Summer (JJA)																					
	none	51.4	44.7	44.1	44.8	42.4	64.9	54.7	47.7	43.7	42.7	73.6	62.7	57.6	53.0	50.4	82.8	73.6	67.3	63.4	60.4
	+1	13.1	21.0	22.6	19.5	22.0	11.1	15.4	19.5	23.4	23.1	8.8	13.4	16.6	18.0	18.7	6.5	9.8	13.5	14.7	15.7
	+2	23.6	21.8	21.1	24.0	24.1	19.1	20.0	20.7	19.5	20.7	15.4	17.7	18.0	19.7	21.1	10.1	14.2	15.8	17.0	18.1
	+3	1.3	1.5	1.3	0.8	0.9	1.4	2.2	2.8	2.8	2.4	0.5	1.3	2.2	2.8	2.6	0.3	0.7	1.2	1.5	1.7
	+4	6.2	6.1	6.1	6.6	6.4	2.6	4.5	5.0	5.5	5.8	1.5	3.5	3.4	3.9	4.7	0.3	1.4	1.8	2.6	3.1
	-1	0.3	0.3	0.2	0.2	0.1	0.3	0.5	0.6	0.7	0.6	0.1	0.2	0.5	0.8	0.8	0.0	0.1	0.2	0.2	0.2
	-2	2.9	3.1	3.2	3.0	3.0	0.5	2.1	2.6	3.0	3.2	0.2	1.2	1.4	1.6	1.7	0.0	0.2	0.3	0.5	0.7
	-3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0
	-4	1.1	1.5	1.4	1.0	1.1	0.0	0.4	1.1	1.3	1.2	0.0	0.1	0.2	0.3	0.4	0.0	0.0	0.0	0.0	0.0
Autumn (SON)																					
	none	47.3	43.2	46.6	46.1	49.9	70.0	55.9	45.4	44.7	48.2	77.4	68.6	63.4	57.8	53.5	83.7	76.0	71.5	67.5	65.0
	+1	19.1	19.4	18.4	16.4	14.7	12.0	15.6	18.9	18.3	16.0	10.0	11.9	12.7	13.6	14.7	7.6	9.7	10.6	10.8	11.0
	+2	22.4	25.8	24.5	24.6	23.6	14.5	19.3	22.7	24.3	24.1	11.0	15.7	18.3	20.9	22.0	8.2	12.9	15.2	17.9	19.1
	+3	1.7	0.6	0.5	0.4	0.5	1.3	1.8	1.7	1.2	1.2	0.9	1.3	1.3	1.3	1.5	0.4	0.7	0.8	0.8	0.9
	+4	6.1	6.6	6.0	6.7	6.2	1.6	4.9	6.4	6.4	6.2	0.5	1.9	3.1	4.6	5.5	0.1	0.6	1.7	2.5	3.2
	-1	0.2	0.1	0.1	0.1	0.1	0.3	0.4	0.4	0.3	0.3	0.2	0.3	0.3	0.3	0.4	0.0	0.1	0.1	0.1	0.2
	-2	2.5	3.1	2.9	3.8	3.4	0.2	1.8	3.5	3.4	3.0	0.0	0.3	0.8	1.3	2.0	0.0	0.0	0.2	0.4	0.6
	-3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0
	-4	0.6	1.3	1.1	1.9	1.4	0.0	0.2	1.0	1.2	1.0	0.0	0.0	0.1	0.2	0.3	0.0	0.0	0.0	0.0	0.0

Figure 34: Meteorological drought (SPI) forecasting score for Europe: difference between the drought classes derived from the median of 15 ensemble forecasts and the observed for the pan-European 2002-2008 droughts. Scores, i.e. % of cells that agree (none: no class difference), and disagree (-4 to +4: class differences) are provided for the SPI for the four seasons as the starting forecasted months, with different accumulation periods (1, 3, 6 and 12 months) and lead times (1-5 months). Light green color indicates high forecasting score, light brown color indicates medium forecasting score, light red color indicates low forecasting score, and white color indicates zero % of area (see Sutanto et al. (2019a) for the explanation of the color coding).

For the pan-European scale, the uncertainty analysis was performed using a re-forecast dataset to simulate categorical drought classes for European drought and non-drought years from 2002 to 2008. For this study, we present a simple approach to evaluate drought forecasting score that is well understood by end users compared to common skill metrics (Annex II), such as Brier Skill Score (BSS, Brier,1950), Relative

Operating Characteristic (ROC) curve (Mason, 1982), and the equitable threat score (Rogers et al., 1995; see Sutanto et al., 2019a for detailed information on the method). In other words the drought class (1: no drought, 2: mild drought, 3: moderate drought, 4: severe drought, and 5: extreme drought) derived from re-forecasts was compared with the class taken from observed data. The forecast score was expressed as the difference between the re-forecasted and the observed drought class. Clearly, no difference between the classes is the best score (perfect score) and the worst score is +4 or -4. Figure 34 shows that the best forecasting score for meteorological drought forecasts with perfect forecasts for >60% of the area is up to 3 months ahead in all seasons is achieved for Standardized Precipitation Index (SPI-x) with higher accumulation period (x=12). SPI index with lower accumulation periods (e.g., 1, 3, and 6 months) produces a lower percentage of the area for perfect forecasts. The hydrological drought forecasts, on the other hand, show better skill score than the meteorological ones. The score class of, e.g. the Standardized Runoff Index, SRI-1 is comparable with SPI-3 and SPI-6, and SRI-3 is comparable with SPI-6 and SPI-12, as expressed by similar values of perfect forecasts (Fig. 35). The detailed information of data, methods, and drought forecast skill score is published in Sutanto et al. (2019a).

Season	Class difference	SRI-1 with lead times of (month)					SRI-3 with lead times of (month)					SRI-6 with lead times of (month)					SRI-12 with lead times of (month)				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Winter (DJF)																					
	none	71.4	58.0	51.9	49.8	47.8	84.9	69.6	55.6	52.2	51.3	90.3	80.0	68.8	62.1	57.2	93.6	86.2	79.0	73.4	68.3
	+1	13.2	18.3	16.5	13.7	11.1	7.6	13.4	16.6	14.9	10.2	5.1	9.6	12.7	12.7	10.7	3.3	6.9	9.6	9.6	8.7
	+2	12.0	16.4	20.6	26.0	30.6	6.5	12.2	17.7	22.3	27.7	4.3	8.8	13.6	18.6	23.9	2.9	6.2	9.9	14.7	19.7
	+3	1.5	2.5	3.4	2.8	1.9	0.5	1.9	3.6	3.0	2.2	0.2	0.8	2.3	2.1	2.0	0.0	0.3	0.7	0.9	0.8
	+4	1.4	2.9	4.3	5.2	6.4	0.4	1.5	3.2	4.3	5.6	0.2	0.6	1.5	2.9	4.2	0.1	0.3	0.6	1.2	2.1
	-1	0.3	0.7	1.5	1.1	0.9	0.1	0.5	1.5	1.2	1.0	0.0	0.1	0.7	0.7	0.8	0.0	0.0	0.1	0.2	0.2
	-2	0.2	0.5	1.2	1.4	1.6	0.0	0.2	0.7	1.1	1.6	0.0	0.1	0.2	0.5	0.9	0.0	0.0	0.0	0.1	0.2
	-3	0.0	0.2	0.5	0.4	0.4	0.0	0.1	0.5	0.5	0.4	0.0	0.0	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0
	-4	0.0	0.1	0.2	0.3	0.3	0.0	0.0	0.1	0.2	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Spring (MAM)																					
	none	63.8	52.3	52.7	52.2	54.0	78.3	66.2	60.0	53.4	50.9	84.8	74.6	69.9	65.4	60.9	90.9	83.5	78.9	74.7	71.7
	+1	13.7	11.0	10.6	14.9	19.2	9.7	9.8	8.0	9.7	14.4	7.4	7.9	5.8	6.2	8.0	4.9	6.0	4.7	4.9	6.1
	+2	16.6	27.3	28.6	25.3	20.3	9.3	19.0	26.4	29.4	27.1	6.4	14.4	21.1	24.1	25.0	4.0	9.8	15.4	18.6	19.4
	+3	2.5	2.1	0.7	0.6	0.9	1.7	2.1	0.8	0.5	0.8	1.0	1.4	0.6	0.4	0.6	0.2	0.4	0.3	0.3	0.4
	+4	1.8	5.1	5.6	5.0	3.8	0.4	1.9	4.0	5.7	5.3	0.1	0.8	2.0	3.3	4.2	0.0	0.2	0.6	1.3	2.0
	-1	1.0	1.0	0.2	0.1	0.1	0.5	0.9	0.3	0.1	0.1	0.3	0.6	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.1
	-2	0.3	1.1	1.2	1.1	0.9	0.0	0.3	0.8	1.4	1.3	0.0	0.1	0.2	0.5	0.9	0.0	0.0	0.0	0.1	0.2
	-3	0.2	0.3	0.1	0.1	0.0	0.1	0.3	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1
	-4	0.0	0.2	0.2	0.2	0.2	0.0	0.0	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Summer (JJA)																					
	none	71.5	59.5	55.4	51.5	47.6	87.5	73.2	59.5	51.6	46.5	92.3	84.2	75.3	65.1	55.5	94.8	89.3	83.4	77.2	70.8
	+1	11.5	18.8	23.6	25.9	28.2	5.7	12.3	19.0	24.6	27.4	3.4	7.1	12.1	17.4	21.8	2.6	5.2	8.5	11.8	14.9
	+2	14.6	16.9	16.0	16.7	16.6	6.2	12.0	16.5	16.5	17.0	4.1	8.0	10.5	13.3	15.9	2.5	5.1	7.2	9.2	11.4
	+3	0.7	1.0	1.5	2.4	3.1	0.3	0.9	1.6	2.3	3.3	0.1	0.3	0.9	1.9	2.8	0.1	0.2	0.5	0.9	1.5
	+4	1.3	2.8	2.9	3.3	3.5	0.2	1.1	2.4	3.2	3.5	0.1	0.4	0.8	1.6	2.5	0.0	0.1	0.3	0.6	0.9
	-1	0.1	0.1	0.2	0.4	0.6	0.1	0.1	0.3	0.5	0.8	0.0	0.0	0.2	0.4	0.7	0.0	0.0	0.1	0.1	0.3
	-2	0.2	0.7	0.7	0.8	1.0	0.0	0.1	0.5	0.9	1.2	0.0	0.0	0.1	0.2	0.6	0.0	0.0	0.0	0.1	0.1
	-3	0.1	0.0	0.0	0.1	0.2	0.0	0.1	0.1	0.2	0.3	0.0	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.1	0.1
	-4	0.0	0.1	0.1	0.1	0.2	0.0	0.0	0.1	0.1	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Autumn (SON)																					
	none	68.1	54.9	50.3	51.3	51.5	83.6	68.6	54.8	47.7	46.2	89.2	77.9	69.0	62.0	55.5	93.4	86.1	79.3	73.7	69.0
	+1	16.0	22.2	24.1	22.1	19.8	8.9	15.8	21.4	23.9	23.6	6.4	11.9	15.5	17.5	19.4	4.1	8.4	11.5	13.4	14.8
	+2	12.5	16.7	17.2	17.4	18.4	6.4	12.1	16.6	18.6	19.0	3.8	8.2	11.7	14.6	17.2	2.3	4.9	7.7	10.0	12.3
	+3	1.4	2.1	2.7	2.5	3.0	0.6	1.5	2.6	3.1	3.4	0.4	1.0	1.6	2.2	2.8	0.1	0.4	0.9	1.3	1.7
	+4	1.5	2.8	3.3	3.6	3.6	0.4	1.5	2.9	3.8	4.1	0.1	0.6	1.3	2.1	3.0	0.0	0.1	0.5	1.0	1.5
	-1	0.2	0.4	0.7	0.7	0.9	0.1	0.3	0.7	1.0	1.2	0.1	0.2	0.4	0.6	0.9	0.0	0.0	0.2	0.3	0.5
	-2	0.3	0.6	0.8	0.9	0.9	0.0	0.3	0.9	1.1	1.2	0.0	0.1	0.3	0.6	0.8	0.0	0.0	0.0	0.1	0.2
	-3	0.0	0.1	0.2	0.2	0.3	0.0	0.1	0.2	0.3	0.4	0.0	0.0	0.1	0.2	0.3	0.0	0.0	0.0	0.0	0.1
	-4	0.0	0.1	0.1	0.2	0.2	0.0	0.0	0.1	0.2	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0

Figure 35: SRI-x forecasting scores for pan-European 2002-2008 droughts, with x = 1, 3, 6, and 12 months: difference between the drought classes derived from the median of 15 ensemble forecasts and the observed.

The drought uncertainty analysis has also been carried out at the Pilot Site scale. This is reported by Van Hateren et al. (2019) and a summary is presented in Deliverable D3.4. There, drought forecasts were presented for two river basins namely Guardiola (Llobregat catchment) and Ripoll (Ter catchment) located in Catalonia, Spain. Meteorological drought forecasts do generally not outperform the climatology for short accumulation times (SPI-x, $x \leq 3$). Skill increases as accumulation time increases (SPI-x, $x \geq 6$). The hydrological drought forecasts (e.g. SRI) have good skill up to 4 months and the acceptable skill can be achieved for short accumulation periods. The skill of hydrological forecasts is higher, i.e. lower uncertainty, than the meteorological forecasts.

The studies of drought forecasting performance using standardized drought indices at pan-European scale show that the ANYWHERE meteorological drought products (e.g. SPI) have acceptable forecasting skill up to 3 months, especially for longer accumulation periods. The skill is higher for hydrological drought forecasts. The ANYWHERE hydrological drought products (e.g. SRI, and SGI) have good skill up to 4 months and the acceptable skill can be achieved for short accumulation periods. The capability of ANYWHERE drought forecasts to produce drought hazards up to 3-4 months ahead indicates that the ANYWHERE drought forecasting products have low uncertainty up to 3 months. The uncertainty becomes higher for longer lead times (LTs). In general, drought forecasts show the lowest uncertainty when they are done in winter, whereas forecasts performed in spring show the highest uncertainty.

3.5.2 Threshold-based Indices (WUR)

An uncertainty analysis of algorithms that forecast threshold-based drought indices has not be carried out at the pan-European scale. It has been done at the Pilot Site scale. Details are reported in Van Hateren et al. (2019) and a summary in Deliverable D3.4. The threshold method approach was selected to analyse droughts in discharge at two stations, namely Guardiola and Ripoll located in the Llobregat and Ter river basins (Catalonia, Spain). Hydrological drought reforecasts outperform the climatology up to lead times of 3-4 months LT. Similar to the standardized drought indices, the forecasts of the hydrological threshold-based indices have lower uncertainty (better skill) than the meteorological threshold-based drought indices. Storage of water in soils, groundwater and lakes is the main reason for the differences.

3.5.3 Areal Indices (WUR)

We have not done any specific analysis to assess the uncertainty of algorithms/tools that forecast areal drought forecasts (e.g. area in drought in Europe). However, the analysis presented in Section 3.5.1 (Fig. 34 and 35), uses pan-European maps for each variable and month. Figure 36 shows pan-European maps for meteorological drought (SPI-3 drought class) and hydrological drought (SRI-3 drought class) derived from reforecast data and observed data for a specific month (August 2003). Moreover, the difference between the re-forecasted and the observed drought class is given. We see, that in particular the hydrological drought forecast, is perfect in the majority of Europe (no difference in re-forecasted and the observed drought class, white areas).

This indicates high skill. Figure 36f again shows the higher skill of hydrological drought forecasts relative to meteorological ones (Fig. 36c). We also see that in north Europe the drought class is overestimated (showed by bluish colors), while in central and south Europe, the forecasts under-estimate the drought class (showed by reddish colors).

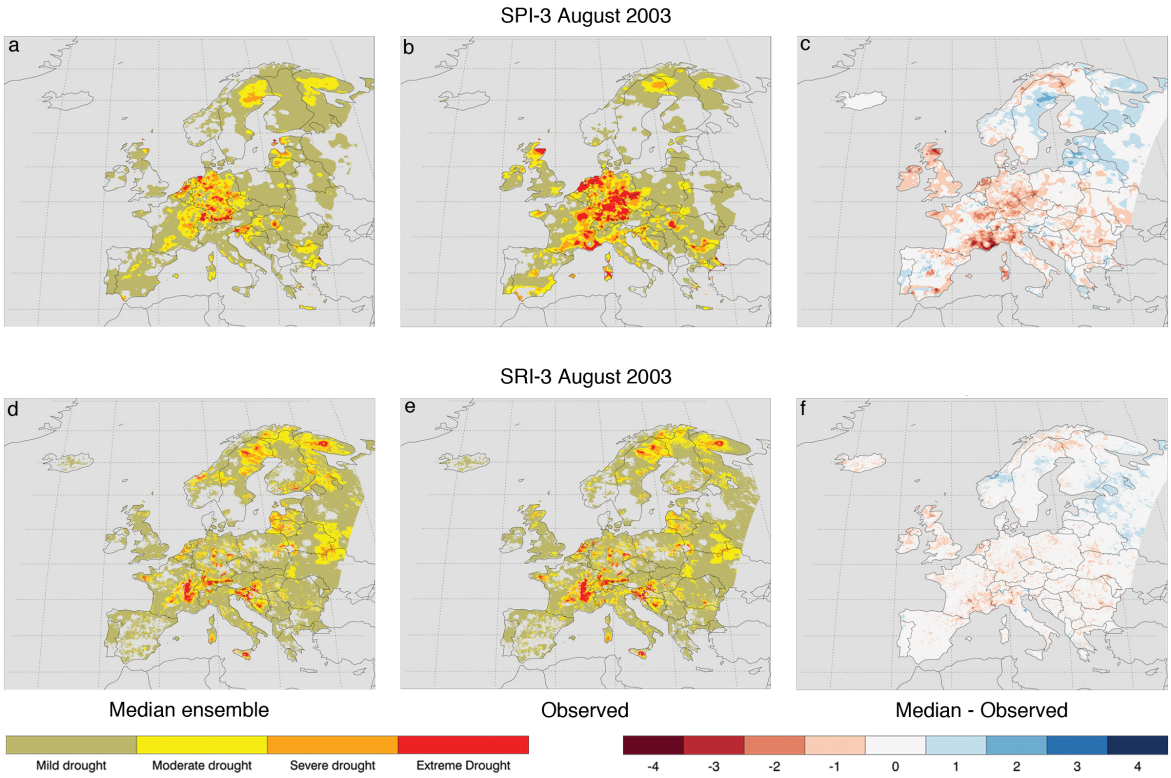


Figure 36: Hydro-meteorological drought in 2003 expressed as drought severity classes using the Standardized Precipitation Index and Standardized Runoff Index accumulated over 3 months (SPI-3, SRI-3): (a) forecasted SPI-3 done early August 2003 (median of 15 ensembles) for a lead time of 1-month, (b) SPI-3 obtained from observations for August 2003 (SFO), (c) drought forecasts score expressed as difference in drought class between the forecast and the observed (a-b). (d,e,f) same as Fig. a,b,c, but for SRI-3. Reddish colors in c and f indicate that the forecast underestimates the drought class and vice versa for bluish colors.

3.6 Convective storms, severe winds and heavy snowfall

The uncertainty of algorithms/tools that forecast typical weather-type natural hazards are described. First, small scale convective storms are explained, which are followed by heavy snow loads. Uncertainty involved in forecasting different types of precipitation is described in the last part. Most of the uncertainty analysis dealing with convective storms, severe winds and heavy snowfall, which has been done during ANYWHERE, has not been published. Hence, the reporting is somewhat more comprehensive than about the previous natural hazards.

3.6.1 Detection and forecasting convective cells (FMI)

The Finnish Meteorological Institute (FMI) pan-European tool used for identification, tracking and nowcasting of convective cells (CC-ITN) is based on the OPERA-composite (Saltikoff et al., 2019) and GLD360 lightning data. The convective cells are identified based on a threshold on the precipitation rate and their tracks are followed from previous positions as well as extrapolated in the form of storm strike probabilities. There are large uncertainties in the nowcasts of convective cells beyond the lead time of approximately 30-60 min because of their huge dynamics (Wilson et al., 1998). Therefore, the identified convective cells are extrapolated in time and space, with a probabilistic forecast that also captures the uncertainty in the future location and severity class. The extrapolation is based on storm movement velocities that are estimated from previously observed cells. For each identified cell, a severity class is assigned based on a statistically-driven climate-adaptive method. A detailed description of CC-ITN algorithms is given in D2.3 (Ciavola et al., 2017).

The uncertainty of the CC-ITN tool has not been tested on real emergency events (continuous monitoring and evaluation) due to the tool being relatively new and therefore not yet much tested and piloted. Instead, the uncertainty of the CC-ITN conducted nowcasts for convective cells was estimated using the FMI radar CAPPI product for eight historical convective precipitation events during July-August in 2016 and 2017 (Table 1). In addition to the inferior spatial and temporal resolution of the FMI composite (1 km and 5 minutes) compared to the OPERA composite (2 km and 15 minutes), the data in the OPERA product is composed of partly dual- and single-polarization radar measurements with varying quality control schemes, whereas the FMI product is composed of only polarimetric data with consistent quality. A radar reflectivity threshold of 35 dBZ corresponding to rain rate of approximately 5 mm/h was used in identifying the storm cells.

Table 1: The precipitation events used in the uncertainty estimation of the CC-ITN conducted nowcasts for convective cells

Event	Date	Start time (UTC)	Duration (hours)
1	27 Jul 2016	04:00	8
2	31 Jul 2016	06:00	15
3	2 Aug 2016	05:00	12
4	3 Aug 2016	07:00	10
5	31 Jul 2017	19:00	8
6	1 Aug 2017	10:00	10
7	12 Aug 2017	14:00	9
8	13 Aug 2017	00:00	11

Two different forecast verification metrics were used in the uncertainty estimation. First metric, the reliability diagram (Bröcker and Smith, 2007) measures the reliability and the sharpness of a probabilistic forecast. For a given intensity threshold for a yes/no prediction, the diagram shows the observed frequencies against the forecast probabilities. The reliability diagram is typically accompanied with a sharpness diagram that shows the distribution of the probabilities. The second metric uses the relative operating characteristics (ROC, Annex II) curve and area under the curve (AUC).

The reliability diagram, ROC curve, and AUC with respect to lead time were computed by pooling the data during the eight events listed in Table 1. The results of the experiments are shown in Figure 37 and 38.

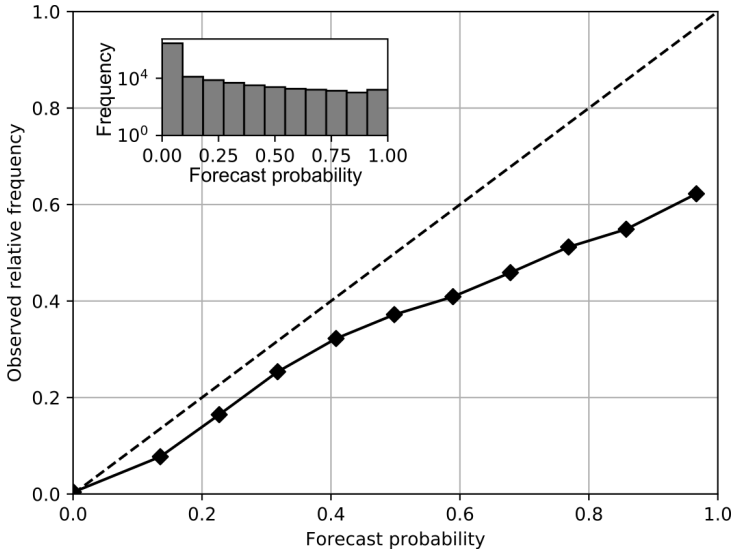


Figure 37: Pooled reliability and sharpness diagram of CC-ITN nowcasts during convective events in Finland in July-August 2016-2017. The intensity threshold is 35 dBZ and the lead time is 15 min.

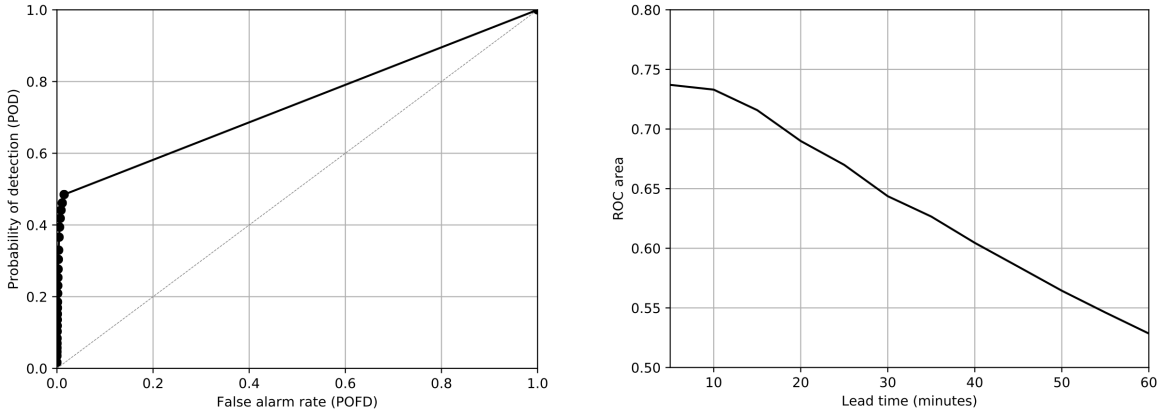


Figure 38: ROC curve with 15-minut lead time (left) and the AUC with respect to lead time (right) for CC-ITN nowcasts during convective events in Finland in July-August 2016-2017. The intensity threshold is 35 dBZ.

The CC-ITN method produces reasonably reliable precipitation nowcasts of convective precipitation up to 15-20 min. At 15 min, the reliability diagram (Fig. 37) shows up to 40% deviation from the diagonal, which becomes more pronounced with probabilities over 0.5. On the other hand, the sharpness diagram shows that the whole probability range between 0 and 1 is represented in the nowcasts with ~10% of the probabilities belonging to the highest range between 0.9 and 1. The above results could be improved by additional tuning of the Kalman filter parameters in the CC-ITN algorithm.

The ROC statistics (Fig. 38) show very small false alarm rates for all probability thresholds. However, the low POD values (less than 0.5) indicate a large number of misses. This is likely because of the inability of the method to predict intensity changes or initiation of new convective cells due to the lack of a life cycle model for individual storm cells, which poses a significant limitation of the method that should be addressed in future work. The AUC remains close to 0.7 up to 20 min, which indicates a potentially useful skill. However, at one hour, this number falls below 0.55, indicating almost no skill.

Based on the analysis, the CC-ITN has troubles producing reliable nowcasts for convective cells beyond 30 min, when all convective cells exceeding the threshold of 35 dBZ are considered independent of their size, as is in the verification above. However, in general the lifetimes of small convective cells are short, and their development is fast (growth or decay, splitting and merging) making the skill rather low beyond lead times of one hour. It should therefore be noted that the largest and most hazardous convective storms, so-called Mesoscale Convective Systems (MCS), have much longer lifetimes and hence, could be forecasted with much better skill for lead times of 1-3 h. These could be also much better observed using the pan-European OPERA radar composite than the FMI CAPPI product utilized here. Uncertainty of the nowcasts is also increased by the lack of a lifecycle model in CC-ITN, which is a limitation of the method that should be addressed in a future work.

The uncertainty of the severity classification was tested based on the extreme rain event that occurred over the Alpine and Central European regions on 11-13 Jun 2018. The CC-ITN algorithm was trained using data from June 2017 and then the event was classified based on the historical data and information of the current situation. An example of the severity classification is presented in Figure 39.

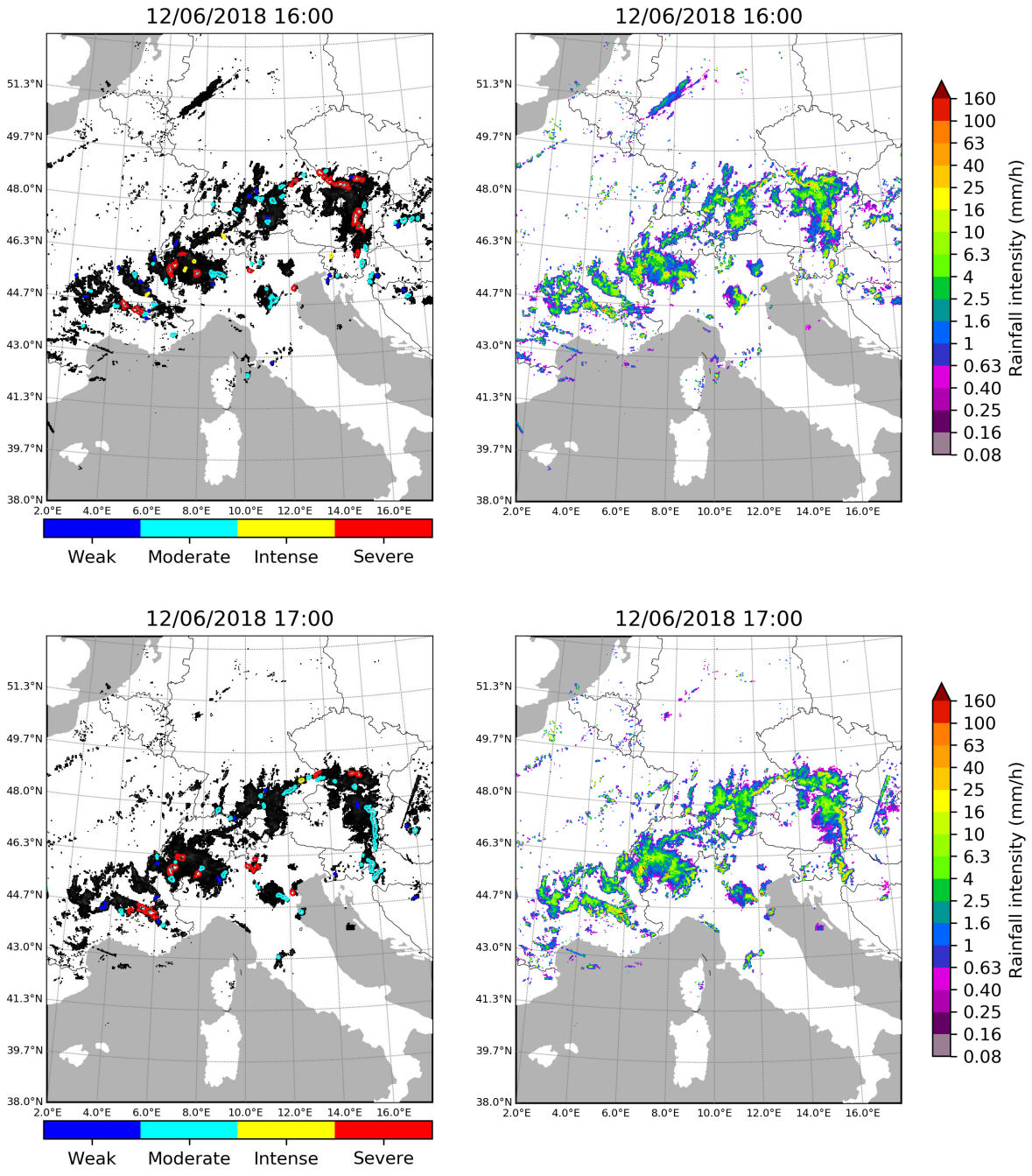


Figure 39: Severity classification of convective cells (left column) based on OPERA radar composite (right column) on 12 Jun 2018 at 16:00 UTC (top row) and at 17:00 UTC (bottom row).

CC-ITN can classify the severity of convective cells based on historical and current information (Fig. 39). As can be seen by comparing the classification results at 16:00 UTC and 17:00 UTC, the severity class of the convective cells changes as a function of time when the event evolves. In this particular case, the event is dissipating, which is seen also in the severity of the cells being on general more severe at 16:00 UTC

than at 17:00 UTC. While the algorithm correctly classifies the cells according to their relative severity, since it has not been tested on real emergency events with continuous monitoring and evaluation, not much can be said about the reliability of the severity classes in terms of actual impacts from the rain event. However, the algorithm correctly classifies cells over Northern Italy, Switzerland and Southern Germany as severe or intense, these being the areas with reported storm induced damage caused by the extreme rainfall.

As reported in D2.2 (Cánovas et al., 2017), so far the CC-ITN algorithm is developed for convective summer storms, and thus the statistical storm severity classes are determined only for summer storms with rainfall making the tool inapplicable for winter storms. Additionally, since the European radar network consists of partly dual-polarization and partly more traditional single-polarization weather radars, to achieve a uniform rain estimate the conversion factors are based on single-polarization measurements, which adds uncertainty to rainfall estimates. The importance of data quality control is emphasized when a statistical analysis of convective storms is conducted using an object-oriented convective storm algorithm such as CC-ITN. Since a relatively small number of storms represent severe, rare events in the statistical analysis, even a few erroneous data points classified as intense can corrupt the extreme-value statistics.

3.6.2 Snow-load and gust algorithms (FMI)

There is a lack of direct measurements of snow load on canopies, which does not allow direct verification or uncertainty of the FMI model on snow load (see Annex III for more background information). Hence, in this study, we have adopted the approach to considering the impact of the snow load on society. The cascading effect caused by snow load accumulation is traced, and the correlation between the forecasted accumulations is compared with the number of emergency tasks of civil protection authorities and electricity fault statistics. Winters 2010-2018 were studied. Three classes, wet, frozen and total snow load of the FMI product are analyzed considering the environmental conditions of relative humidity, temperature, wind speed, precipitation accumulation and height above MSL. The focus of the study is to define the critical conditions, in which the damages to forest and infrastructure are occurring. These conditions are presumably dependent also on other factors, such as the tree age and how the tree during its growth is accustomed to carry snow mass and freezing soil conditions, which, nevertheless, are omitted from this analysis.

FMI Snow load model

In brief, the modeled snow load is classified into four different types: rime, dry snow, wet snow, and frozen snow. The increase of the snow load is introduced by accumulation of rime and snowfall, and in the riming process, the relative humidity acts as input variable. The decrease of snow may occur due to wind removal or melting,

and through a change to another category. The possible transformations are wet snow freezing into frozen snow and dry and frozen snow changing into wet snow. The variables needed in the model calculations are air temperature, precipitation, relative humidity, wind speed, global radiation, and cloudiness. In addition, the elevation of terrain above sea level affects the riming efficiency in the model.

The FMI model uses a time step of one hour. For the parametrization, an exemplar tree is assumed that has a cone-shaped crown with a projected catchment area of 1 m^2 from above and from the side in the direction of the wind. It is further approximated that a 1 mm water layer of melted snow load on a horizontal surface corresponds to a 1 $kg.m^{-2}$ snow load on the tree crown. A more detailed description of the calculation procedure, including equations, can be found in the appendix A of Lehtonen et al. (2014).

Lehtonen et al. (2014) defined risk values for significant snow load based on observed daily average values of temperature, relative humidity, wind speed and precipitation at four different locations in Finland, these are shown in Table 2. The average modeled values of snow load were 5.35 $kg.m^{-2}$ for the total crown snow load and 1.72 $kg.m^{-2}$ for the rime snow load. The table values were used in this study as a proxy of significant snow load events.

Table 2: Threshold values of daily mean 2 m air temperature (T_{mean}), 2 m relative humidity (RH_{mean}), 10 m wind speed (U_{mean}) and total precipitation (P_{day}) that were defined to determine the risk days favorable for heavy snow loading and riming in Lehtonen et al. 2014

Weather variable	**Snow loading**	**Riming**
Temperature	$-3.42°C < T_{mean} < 1.05°C$	$-5.19°C < T_{mean} < -0.16°C$
Relative humidity	$RH_{mean} > 89.44\%$	$RH_{mean} > 95.50\%$
Wind speed	$2.07 ms^{-1} < U_{mean} < 5.63\ ms^{-1}$	$2.00 ms^{-1} < U_{mean} < 4.54\ ms^{-1}$
Daily precipitation	$P_{day} > 6.41$ mm	$P_{day} < 1.11$ mm

In this study, two different snow load types, wet snow load (WSL) and frozen snow load (FSL), were used, as well as the total snow load (TSL). These were examined every hour in a 10x10 km grid consisting of 122 x 76 grid points. The same grid was used for meteorological data, which includes 2 m temperature, wind speed, 3-hour precipitation accumulation and relative humidity. These are stored with a temporal resolution of 3 hours in FMI's climate database. Data was gathered from year 2010 to 2018, with summer months (JJA) excluded. In 2018 the data ends at the end of May.

The grid points were classified based on which municipality they lie in (of total 306 Finnish municipalities). This way, it was possible to define a municipality-wide mean value for each of the meteorological quantities at each time step. Lists containing the mean values in each municipality were written into separate files for each time step,

i.e. 8 times each day. From the hourly snow model files, only the ones corresponding to hours 0, 3, 6, 9, 12, 15, 18 and 21 UTC were used.

To examine the impact of snow load events, two types of impact data were used. Firstly, emergency task description data provided by emergency services, and secondly, power cut data provided by Finnish Energy. Of the emergency task data, the used tasks included categories stating, tasks related to snow and ice or tasks related to wind or storm, explaining the cause of the emergency. The data was available by coordinates, so a daily number of tasks within each municipality could be defined.

The power cut data, however, was only available for much wider areas; Finland was divided into five regions, because of commercial reasons. Thus, the power cut data used in our figures must be considered directional only; each municipality belongs to one of the five regions.

Analysis

The overall image of the emergency tasks resulted from snow during the studied time period of 2010-2018 is shown in Figure 40. There are clear peak periods, which largest can be seen in the new year's time of 2017-2018, described in Annex II. There were almost 3500 emergency tasks in that period. More detailed images were plotted also to each municipality affected (Figure 41). If only events with five or more snow-related tasks in one day were considered, this resulted in 76 images, each showing a timeline of one month in one municipality. Some of these monthly periods contained more than one such event with an increased amount of emergency tasks, and often the minimum of five tasks is exceeded on more than one day in succession.

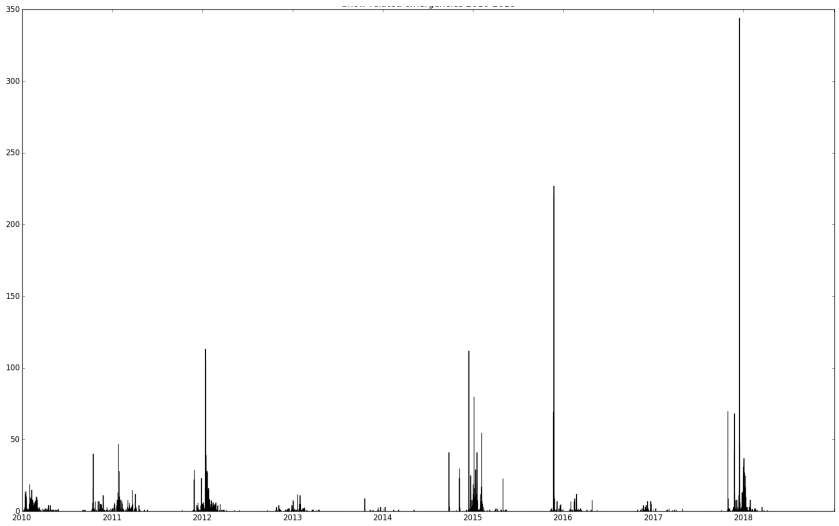


Figure 40: The amount of snow-related emergency tasks during the studied period of 2010-2018.

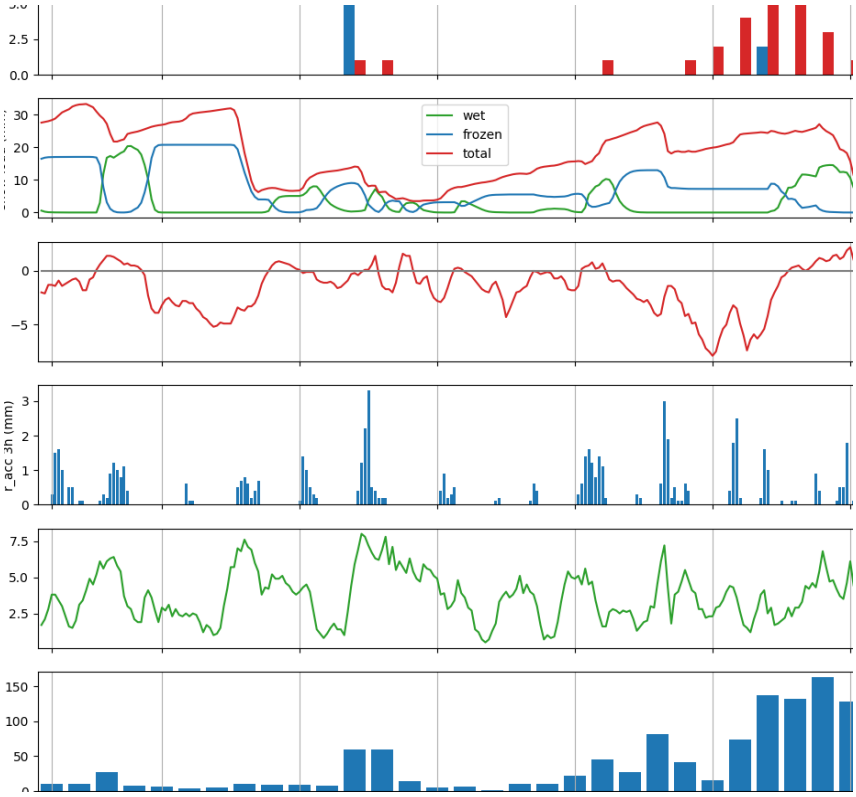


Figure 41: An example report of the Joensuu municipality affected by the snow related emergency tasks in December 2017. Emergency tasks (snow or wind), snow loads (total, wet and frozen), mean temperature (municipality), 3 h precipitation accumulation, wind speed and power cuts on a larger area.

Following the example risk limits presented in Table 2, the different clusters of the increased emergency tasks were searched from the data set by examining different dependencies of the meteorological data, modeled snow load values and the number of tasks. The mean value represents the mean of the eight daily mean values in the municipality during the day in question, and the maximum value is the largest mean value of the eight daily mean values. The examined time period varied between the daily to three-day periods. Figure 42 shows a 2D-histogram of combinations of maximum daily wind and daily mean temperature in respect to snow-related emergencies (Fig. 42, left), as well as three days mean temperature and precipitation (Fig. 42, right). For the three days analysis, it is required that the number of emergencies is above the predefined minimum of 5 on each of those three days, but the snow load data and other meteorological data is collected over a three-day period. This way the idea is to capture the nature of snow load events: often the emergencies caused by changes in snow load are divided over more than one day.

In Figure 42 one can clearly distinguish certain density areas. In the daily plot (Fig. 42, left), the area with the highest density seems to correspond to rime snow load at relatively high winds and cold temperature, whereas the more widely covering area presumably represents emergency tasks related to wet snow load at close to zero degree temperature and varying wind speeds. In the three-day plot (Fig. 42, right),

similarly, one can recognize the area of high precipitation around 20 mm in three days, with the temperature ranging from around -5 degree Celsius to around zero degree Celsius.

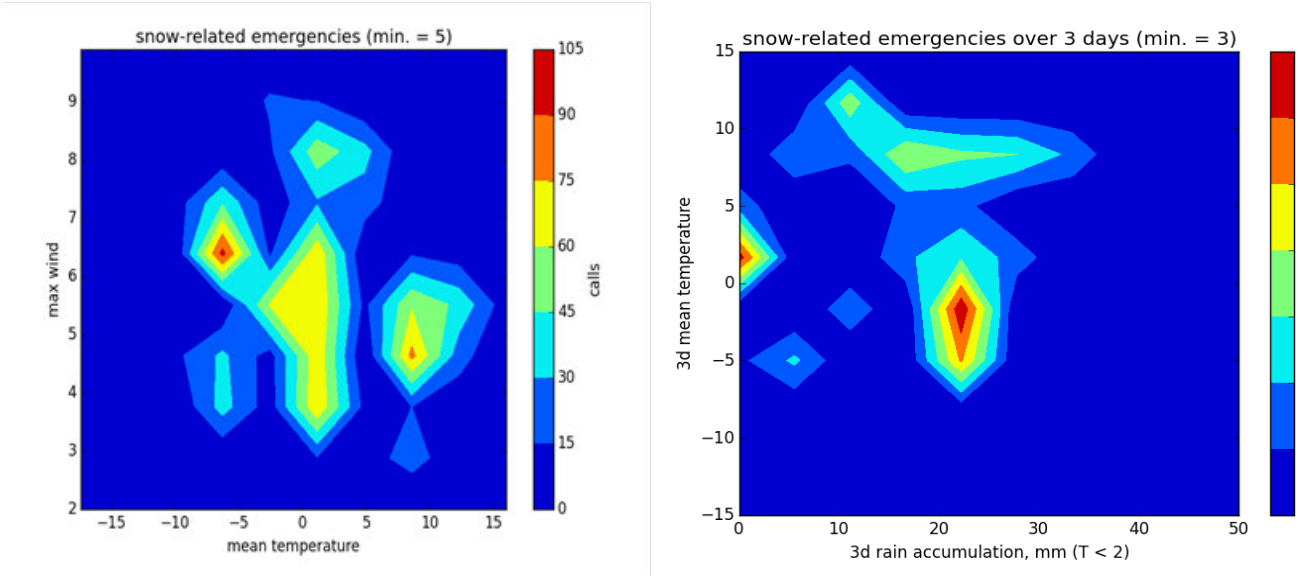


Figure 42: The number of tasks with daily mean temperature and daily maximum wind in municipality at left, and at right with the three day mean temperature and precipitation (temperature below +2 degree Celsius, while precipitating).

To define, if the emergency tasks are caused by a certain type of snow load, we have separated the data with thresholds. If WSL > 3 mm and FSL < 2 mm on the day in question, it is likely that this is a WSL case. If WSL < 3 mm and FSL > 2 mm, it is likely an FSL case. In Figure 43 are shown two example images of two-day analysis. The data is also divided based on daily accumulated precipitation, maximum wind speed and daily mean temperature, as well as the relative humidity thresholds defined in Table 2. In Figure 43, the upper row presents emergency tasks caused by FSL when the mean relative humidity over 2 days is lower than the threshold value, while the lower row shows presumable WSL cases with relative humidity over the given threshold value. In Figure 43 one can see clearly that emergency tasks caused by wet snow load only seem to happen when the precipitation accumulation over two days is more than 5 mm.

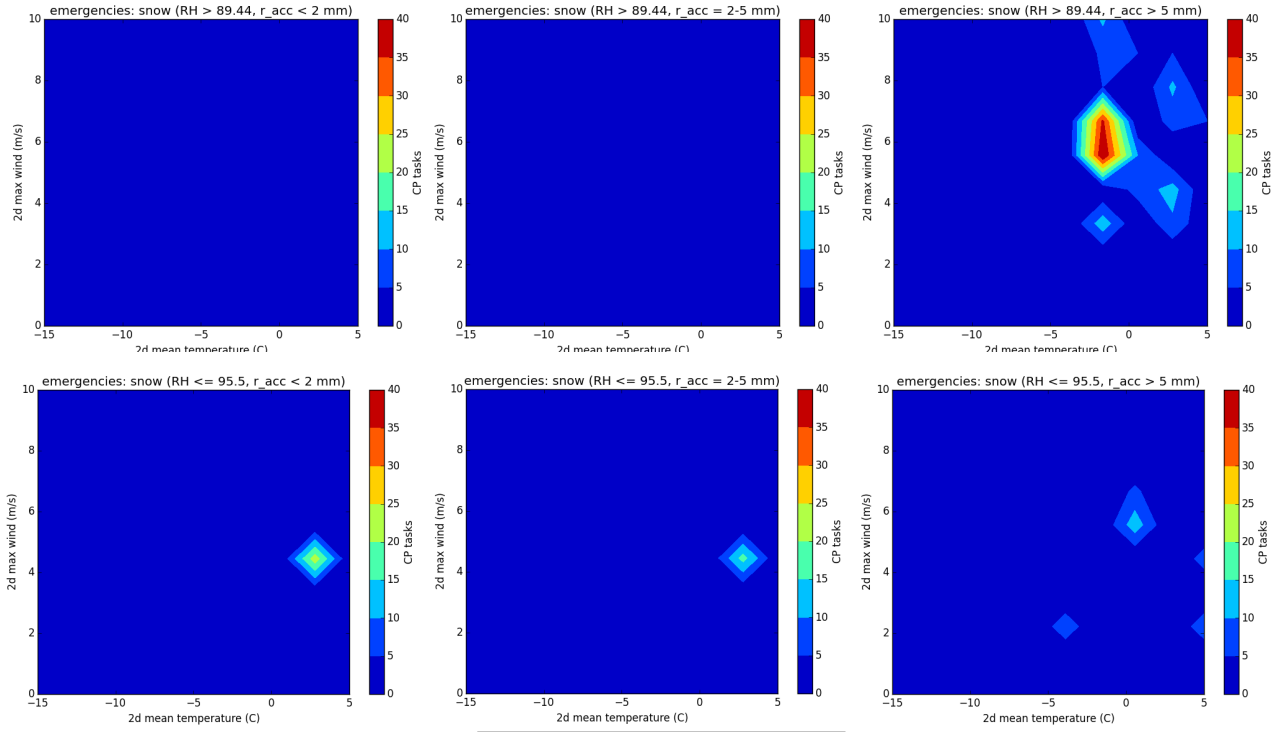


Figure 43: The number of tasks as function of maximum two-day wind and two-day mean temperature divided to three different categories according to precipitation (P_acc < 2 mm, 2 mm < P_acc < 5 mm, and P_acc > 5 mm.) above for rime snow load higher than 2 kgm^{-2} and below for wet snow load higher than 3 kg.m^{-2}

Uncertainty

To clarify, the uncertainty to utilize the snow load product forecast as pre-warning of coming snow-related emergency tasks, the contingency tables of the traditional skill scores were calculated. The data covering years of 2010-2018 was divided to the two types, WSL and FSL, and the daily number of tasks was selected again to be higher than 5.

Below in Table 3 the counts are shown for: (i) snow load has high value, and the number of tasks refers to increased number of events, (ii) snow load has high values, but no increased task number is seen, (iii) there are increased amount of tasks, but snow load model values are not increased, and (iv) no increased amount of tasks, and no increased values of snow load. Here, one count represents one day in one municipality when the given temperature and wind conditions prevail; hence the large overall numbers. The scores false alarm rate (FAR) and probability of detection (POD) are calculated for both snow load types. The POD is 0.2 for WSL and 0.24 for FSL. Numbers are very similar and do not give high scores for detecting an upcoming event. And the FAR is even worse, for both snow load types the value is 0.99. Hence the high snow load value cannot be utilized directly to forecast of an upcoming crisis situation

in civil protection. Additional information to interpret the high values is needed, as suggested at the start of the section.

Table 3: The contingency tables for WSL and FSL related emergency tasks

Snow load types		**Emergency tasks**	
		> 5	**< 5**
Frozen (FSL)	> 5 kg.m^{-2}	12	11746
	< 5 kg.m^{-2}	39	199564
Wet (WSL)	> 5 kg.m^{-2}	11	1828
	< 5 kg.m^{-2}	44	145969

Conclusions

The FMI snow load model has been operational for more than ten years. Finnish meteorologists utilize it to give warnings for society of an upcoming threat of electricity outbreaks because of falling trees. But as such, it cannot be solely used as tool to issue warnings, as the scores for forecasting are low. It may be useful to indicate upcoming risk because of favorable conditions for snow accumulation, as can be seen from the huge amount counts, where no high snow load values also provide no increased number of emergency tasks. However, meteorologists typically use external information of the coming weather situation to scale the potential risk to a correct level.

This is a preliminary analysis of the impact of the snow load model with a surprising result, i.e. the poor correlation to environmental parameters, which were expected to correspond more strongly to the snow load accumulation. The future work is to study the 76 snow load events in more detail and look for the patterns in the snow load events.

To model accumulated snow load is challenging, as it depends in addition to meteorological data on factors, which are not typically available, such as the positioning of the tree, its height and growth patterns. Therefore, the current version of the model cannot be considered to give an accurate representation of the snow load. The model is empirical and still by no means perfect, and to improve the model would require a detailed measurement campaign with actual crown snow loads in different environments and weather situations.

3.6.3 Probability of precipitation type (ECMWF/FMI)

The main sources of uncertainty in the precipitation type forecasting are algorithm choice and model errors. The uncertainty is greatest for freezing precipitation since its

formation needs very specific temperature profile in the lower troposphere. The melting and refreezing processes are the key to the formation of freezing precipitation (Reeves et al., 2014). First, the ECMWF algorithm for the precipitation type is described, which is followed by the description of the FMI algorithm. In the latter the two algorithms are also compared.

3.6.3.1 Probability of precipitation type (ECMWF)

The medium-range ensemble (ENS) from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) (Section 2.1.1) is used to create two new products intended to face the challenges of winter precipitation-type forecasting. The products themselves are a map product that represents which precipitation type is most likely whenever the probability of precipitation is.50% (also including information on lower probability outcomes) and a meteogram product, showing the temporal evolution of the instantaneous precipitation-type probabilities for a specific location, classified into three categories of precipitation rate.

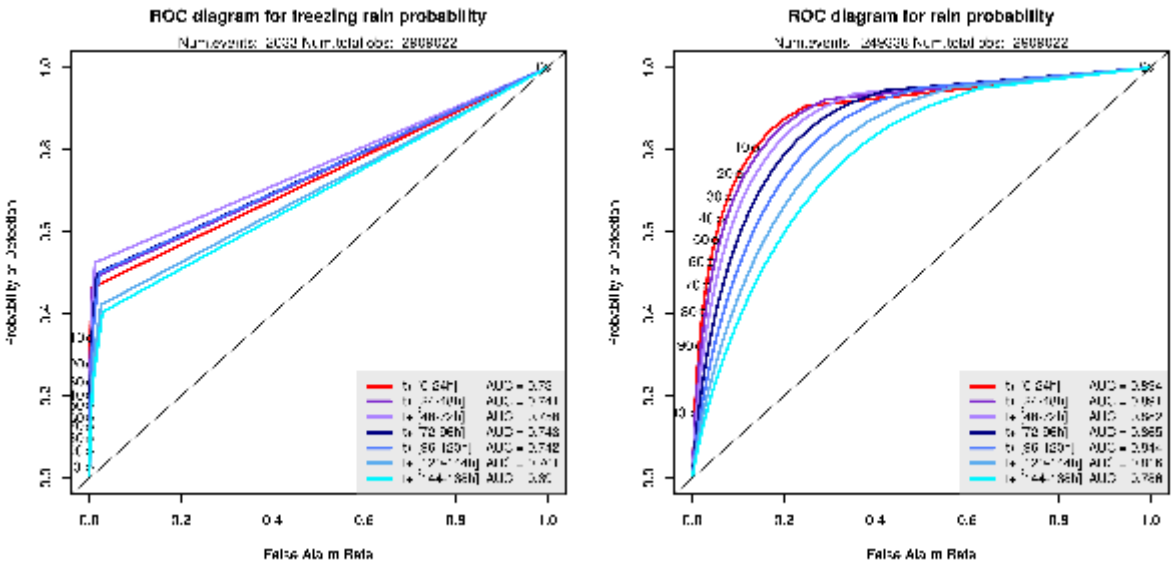


Figure 44: ROC curves at different lead times, up to day 7, freezing rain probability (left) and rain probability (right). The curves are the plots of hit rate vs false alarm rate for each decision threshold (2% interval used). Labels, at 10% intervals, are shown for the day-1 forecasts only (in red). The 458black line represents no skill. The area under the curve (AUC) for each lead time is shown in the grey box.

A minimum precipitation rate is also used to distinguish dry from precipitating conditions setting this value according to type, in order to try to enforce a zero-frequency bias for all precipitation types. The verification of both products was developed using four months' worth of 3-hourly observations of present weather from manual surface synoptic observation (SYNOPs) in Europe during the 2016-19 winter

period (Fig. 44 and 45). This verification shows that the IFS is highly skillful when forecasting rain and snow, but only moderately skillful for freezing rain and rain and snow mixed, while the ability to predict the occurrence of ice pellets is negligible. Typical outputs are also illustrated via a freezing-rain case study, showing interesting changes with lead time [for more information, see Gascón et al. (2018)].

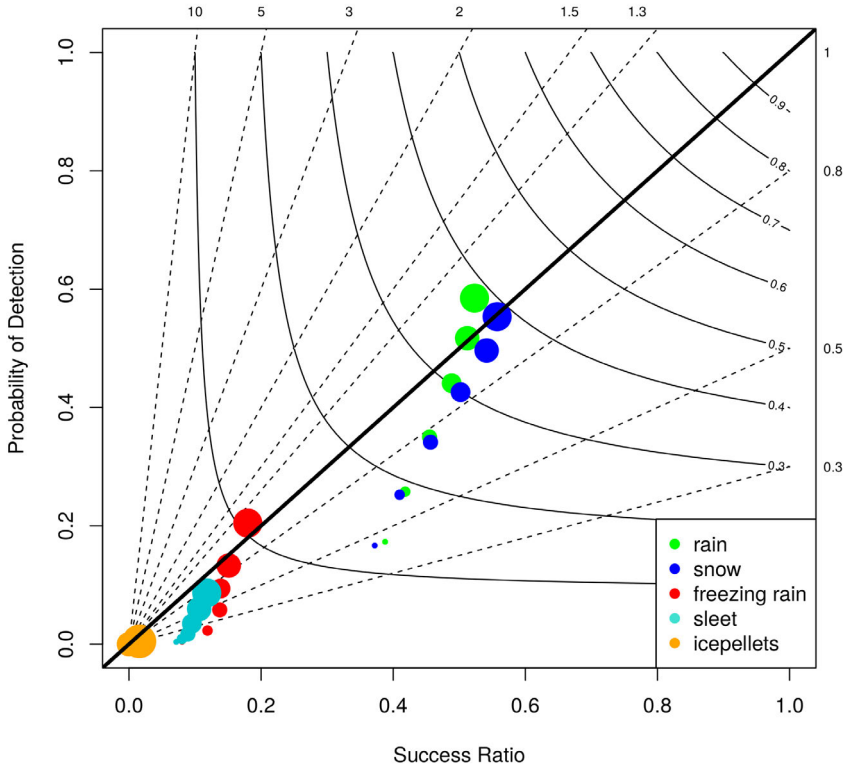


Figure 45: Performance diagram for the PREFptype for each type of precipitation and for multiple lead times. Labelled solid contours represent the CSI and dashed lines are FB with labels along the outward extension of the line. Differing sizes of the points indicate the six different lead times (the bigger the size, the shorter the lead time, from 0–24 to 144–168 h).

3.6.3.2 Probability of precipitation type (FMI)

Prediction models and observation data

In this section, we verify and compare two probabilistic precipitation type models. The first model is based on global European Centre for Medium-Range Weather Forecasts ensemble model (ECMWF-ENS), which uses 51 ensemble members and has approximately 18 km spatial resolution (Section 3.6.3.1). Probability of precipitation type is produced for rain, snow, wet snow, rain/snow mixed (sleet), freezing rain, and ice pellets. In the ECMWF product the current minimum values of precipitation rate that is applied for each precipitation type are 0.12 mm/h for rain, 0.1 mm/h for sleet, and

0.05 mm/h for snow, wet snow, ice pellets and freezing rain. Gascón et al. (2018) give more detailed information of this ECMWF precipitation type model.

The second model has been developed at the Finnish Meteorological Institute (FMI), as part of the TOPLINK SESAR project. This prediction model uses output from different deterministic numerical weather prediction (NWP) models (ECMWF, HIRLAM and GFS) with multiple initial times. The formalization of this process is called a poor man's ensemble prediction system (PEPS) and therefore this model is called FMI-PEPS in this report. The spatial resolution of this multi-model ensemble is 7.5 km which is the resolution of HIRLAM model. The other two models are interpolated to the same 7.5 km resolution because ECMWF model has 18 km and GFS model 13 km spatial resolution. FMI-PEPS model predicts the probability for drizzle, rain, sleet, snow, freezing drizzle, and freezing rain. In FMI product the current threshold for all the precipitation types is 0.1 mm/h except for freezing drizzle/rain 0 mm/h. Since predicted precipitation types between ECMWF and FMI models differ a little, some of the precipitation types are combined to make the comparison possible. In the combined verification we have four different precipitation types: rain, snow, sleet, and freezing rain (Table 4).

Table 4: Precipitation types in both models and their combinations for model verification/comparison. In brackets is the amount of each precipitation type events observed during the verification period 15 Oct 2016 – 15 Feb 2017. Total amount of observations was 38171

Precipitation types for model verification	**FMI-PEPS model precipitation types**	**ECMWF-ENS model precipitation types**
Rain (3003)	Drizzle + Rain	Rain
Sleet (238)	Sleet	Sleet (Rain/Snow mixed)
Snow (2277)	Snow	Snow + Wet snow + Ice pellets
Freezing rain (154)	Freezing drizzle + Freezing rain	Freezing rain

The verification period is winter season from 15th October 2016 to 15th February 2017. Forecasts were verified against METAR ("METeorological Airport Report") observations from 39 airport weather stations in Europe (Fig. 46). Stations are divided into four groups based on their geographical locations to explore if there is regional variability in uncertainties. For the verification, forecasts are interpolated bilinearly to observation points. We used forecasts from the 00 UTC analysis time and lead times from 3 to 120 hours with 3-hour time resolution.

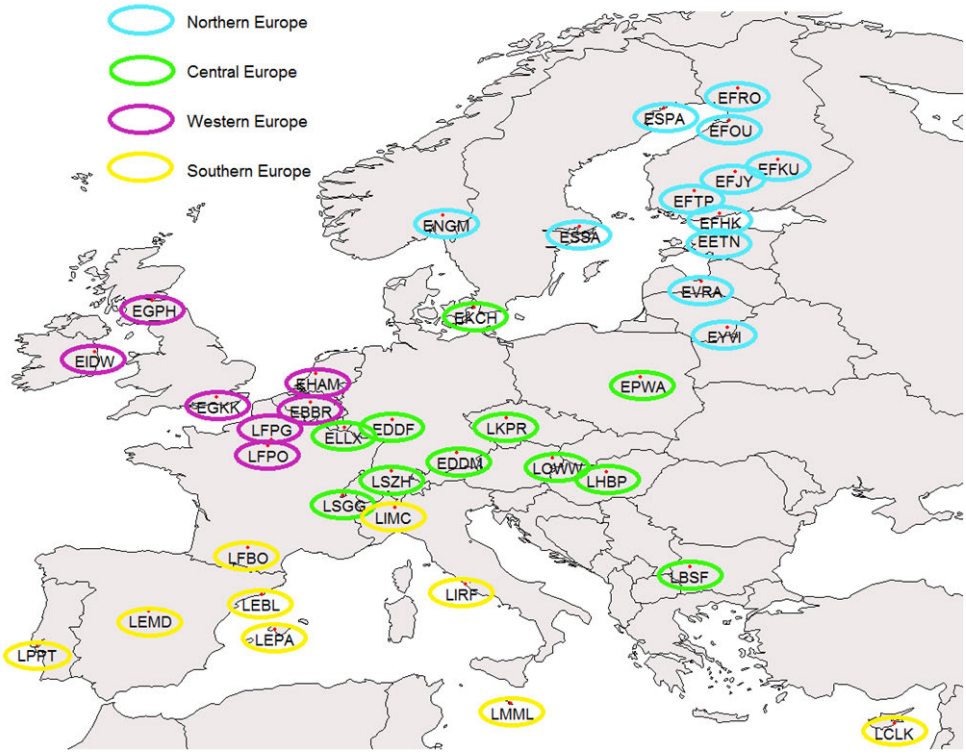


Figure 46: Map of the observation stations (airports) divided into four different groups based on their geographical location.

Verification

The aim of the verification is to examine the uncertainty of the probabilistic precipitation type forecasts and compare the skill of ECMWF and FMI models. As metrices we use reliability diagram, area under ROC-curve (AUC), Brier skill score (BSS), and economic value as verification scores (Annex II). The total amount of each precipitation type events observed during the verification period is shown in Table 4. There are quite few cases of freezing rain (0.4%) and sleet (0.6%) during the period which should be consider when interpreting the verification results.

Figure 47 shows reliability diagrams for each precipitation type forecasts separately with all lead times (3-120 h) combined. Rain forecasts with both models are quite reliable but there is little overforecasting on higher probabilities. Diagram also shows that FMI-PEPS model (blue) never forecasts over 90 % probabilities. This feature might be typical for multi-model ensemble since it usually has more variability between members than ensemble from one NWP model. Reliability curves for sleet lies below the diagonal which indicates that both models overforecast the sleet. Snow is slightly overforecasted with ECMWF-ENS model and underforecasted with FMI-PEPS model. Again FMI-PEPS do not have high probabilities (>80%) at all, and this feature is emphasized when reliability diagrams are looked separately at different lead times (not shown here). Freezing rain forecasts with both models are reliable, since most of the forecasted probabilities are low and lie near the diagonal. ECMWF-ENS forecasted over 60% only three times and FMI-PEPS only once.

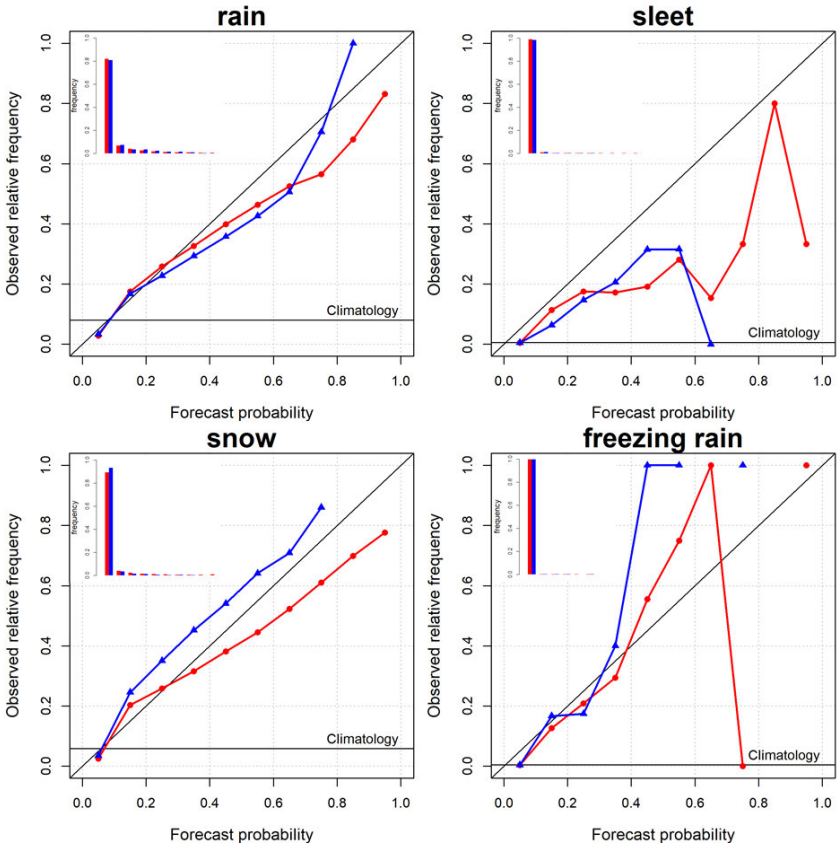


Figure 47: Reliability diagrams for probabilistic precipitation type forecasts (blue: FMI- PEPS, red: ECMWF-ENS) with all lead times combined (3-120h) for rain, sleet, snow, and freezing rain.

Figure 43 shows that for both models AUC is over 0.5 for all precipitation type at all lead times which indicates that forecasts have potential usefulness. AUC is conditioned on the observations, whereas reliability diagram conditions on the forecasts. Therefore it is important to look both verification measures. As reliability diagrams show (Fig. 47), FMI-PEPS model does not forecast the highest probabilities at all which leads to lower resolution compared ECMWF-ENS model for rain and snow. For freezing rain and sleet FMI-PEPS model has better resolution at the shorter lead times and ECMWF-ENS at the longer lead times.

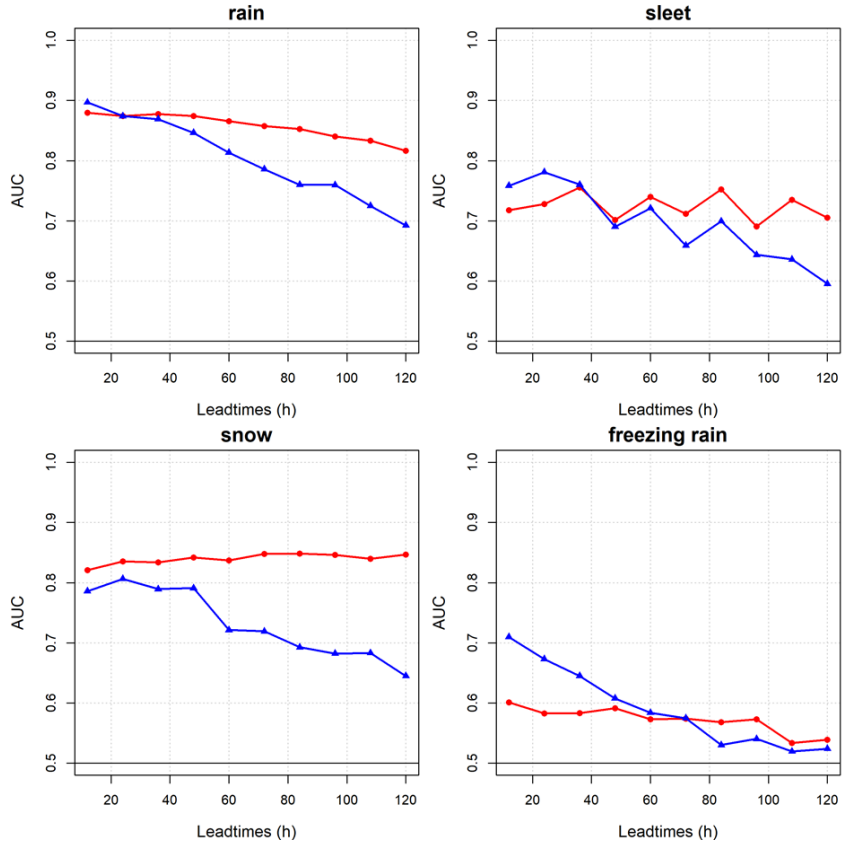


Figure 48: The area under curve (AUC calculated with 2% intervals) at different lead times for rain, sleet, snow, and freezing rain (blue: FMI-PEPS, red: ECMWF-ENS).

The Brier Skill Score for different precipitation type forecasts were plotted as a function of lead time in Figure 49. With clearly better resolution, ECMWF-ENS model has also better BSS for rain and snow. At shorter lead times FMI-PEPS model has better BSS for freezing rain. At longer lead times and for sleet at all lead times there is no clear difference between the accuracy of the models.

When interpreting the BSS, it must consider that sleet and freezing rain occur rarely in this sample (Table 4), which makes the climatological forecast very difficult to beat. For an extremely rare forecast one can be highly accurate simply by forecasting “no” every time. In this situation, a forecast system is severely punished by false alarms.

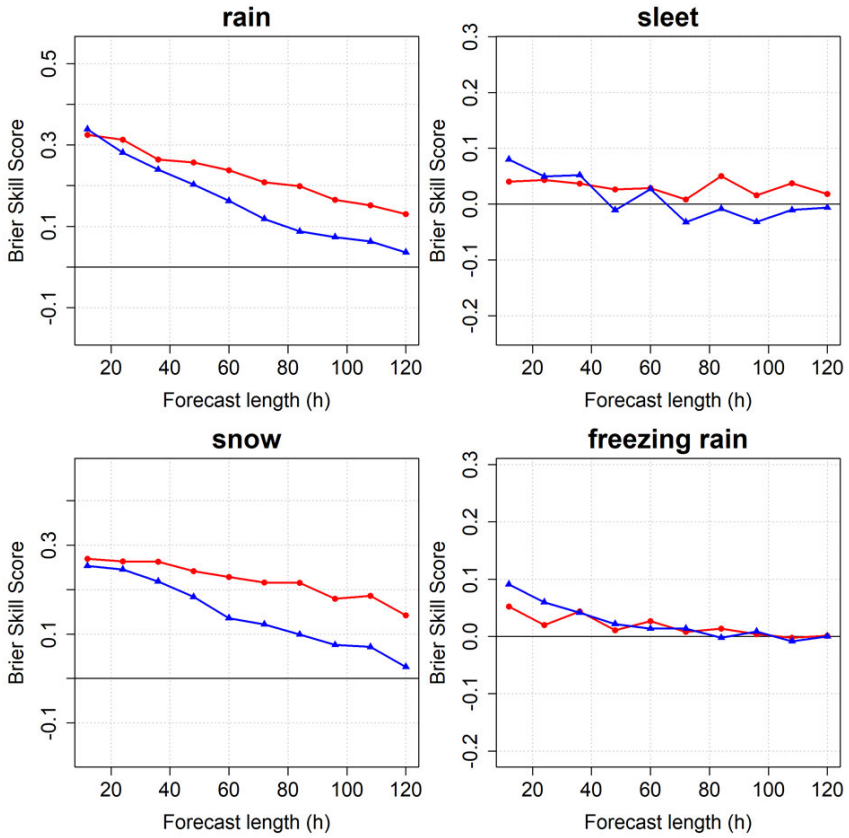


Figure 49: The Brier skill score (BSS) at different lead times for rain, sleet, snow, and freezing rain (blue: FMI-PEPS, red: ECMWF-ENS). Reference model is sample climatology.

Figure 50 shows relative economic value (Annex II) for both models at different lead times for rain and sleet and Figure 51 for snow and freezing rain. All precipitation type forecasts have positive relative economic value at all lead times when loss is much larger than cost. However, the value is not very significant for freezing rain at 96-120 h lead times. When comparing the results between models, ECMWF-ENS has better relative values than FMI-PEPS for rain and snow, especially at longer lead times. For sleet and freezing rain, the economic values are little better for FMI-PEPS model at shorter lead times.

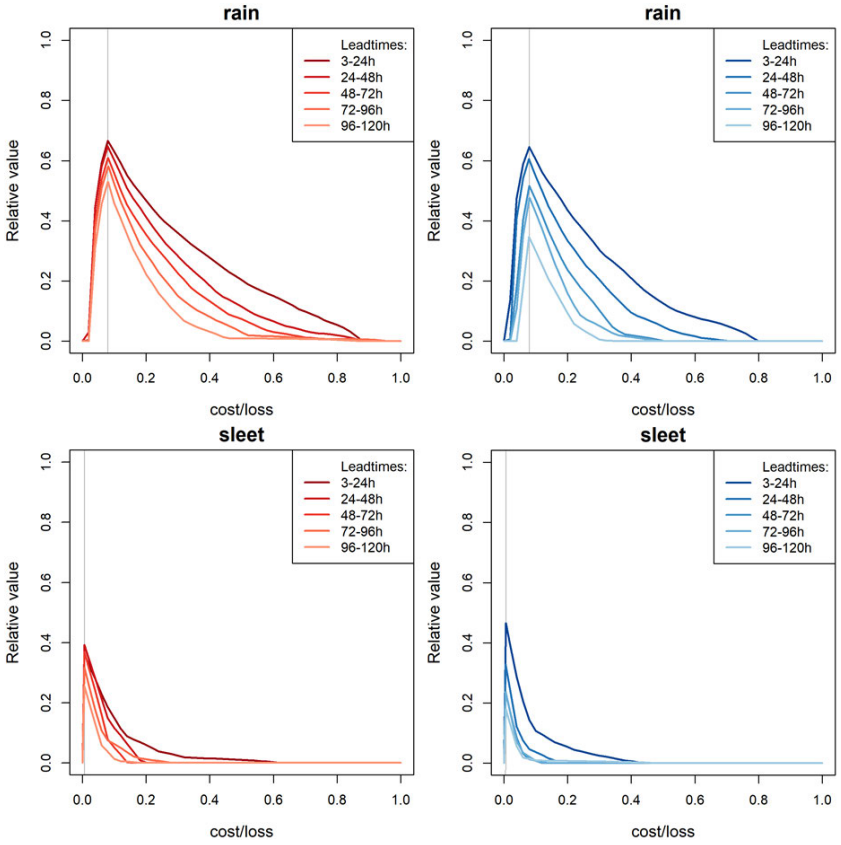


Figure 50: Relative economic value with different cost-loss ratios for ECMWF-ENS (left) and FMI-PEPS (right) rain and sleet probabilistic forecasts. Different lead times are shown with different shades of colors.

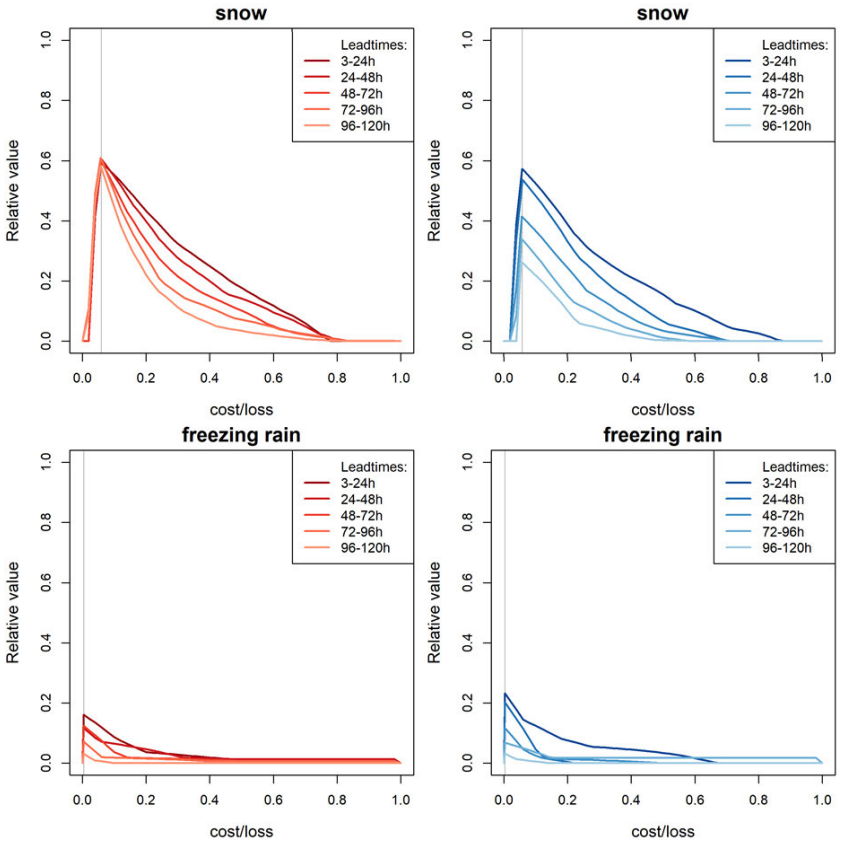


Figure 51: Relative economic value with different cost-loss ratios for ECMWF-ENS (left) and FMI-PEPS (right) snow and freezing rain probabilistic forecasts. Different lead times are shown with different shades of colors.

Lastly, we looked the Brier skill score for different precipitation type forecasts at Northern European stations, only to see if models have different order in skill in that area (Fig. 52, Northern European stations are shown in Fig. 46). Reference model for BSS is again sample climatology but the sample is now different (only observations from Northern European stations), which means that absolute values of BSS cannot be compared between Figure 49 and 52, but the comparison of models relative to each other's is possible. Regarding the BSS, the rank order of the models is quite similar at Northern European than at all European station (Fig. 49 and 52). Only in rain forecasts the FMI-PEPS model has better BSS up to 36 h when at all European stations the model was better up to 12 h. The difference might be caused by the small sample size and having a longer verification period would give more reliable results.

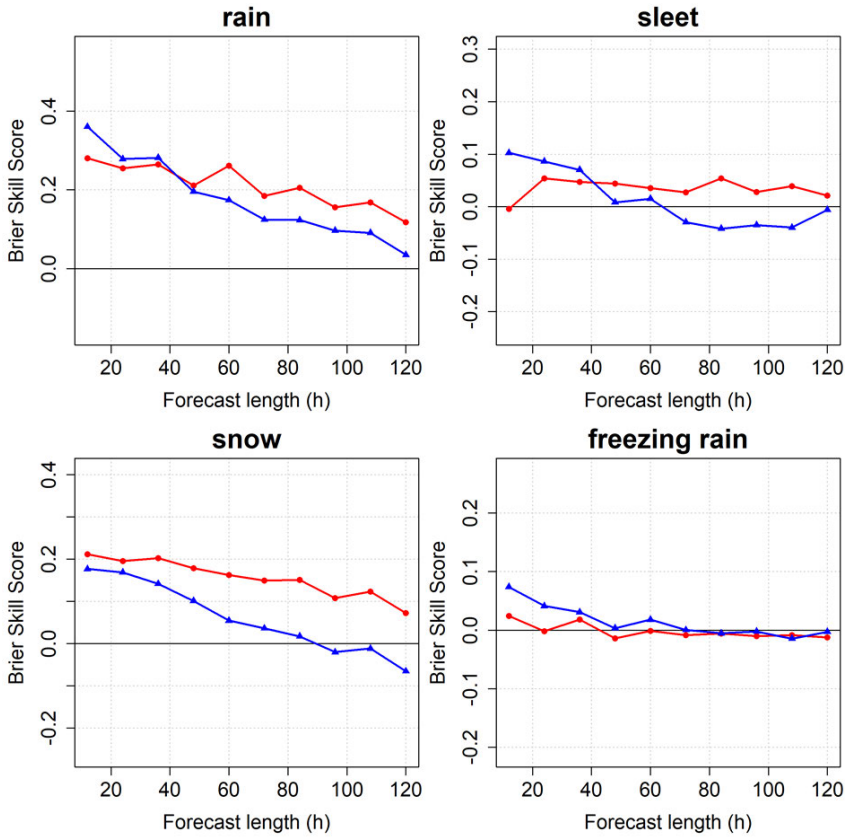


Figure 52: The Brier skill score (BSS) at Northern European stations at different lead times for rain, sleet, snow, and freezing rain (blue: FMI-PEPS, red: ECMWF-ENS). Reference model is sample climatology (only observations from Northern European station included).

Conclusions

Verification results indicate that both models forecast all precipitation types overall quite reliably, although both models tend to slightly over-forecast the sleet, and FMI-PEPS model clearly under-forecasts the snow. FMI-PEPS snow forecasts have also quite low resolution which indicates that the threshold 0.1 mm/h for snow might be too high (e.g. ECMWF model has 0.05 mm/h threshold for snow) and the algorithm itself is capable to determine the snow.

ECMWF-ENS produces more accurate rain and snow forecasts especially at longer lead times. FMI-PEPS model has higher BSS for sleet and freezing rain up to 36 and 48 hours respectively. Still, it must consider that sleet and freezing rain occur rarely in this sample and there might also be interpretation errors in observations between snow, wet snow and sleet.

In this point, the verification results indicate that both models and algorithms give reliable and skilful probabilistic forecasts for rain and snow up to five days, and for sleet and freezing rain up to about 3 days. ECMWF-ENS model is overall better, but at

shorter lead times FMI-PEPS model might be equally or more usable for rain, sleet and freezing rain forecasts.

3.7 Summary of background information for uncertainty assessment of natural hazards

Table 5 summarizes for the uncertainty analysis per hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether observed or proxy data have been used, (iv) length of time series of observed or proxy data, and (v) method that has been used to compare the forecasted with.

Table 5: Overview of background information for uncertainty assessment natural hazards

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy / Simulated)	Length time series	Comparison method	Remarks uncertainty
Floods, flash floods, debris flows, and landslides	FF-EWS	PRD-93 to PRD-94	Simulated flash flood hazard from actual rainfall observations	Simulated	2018-2019 2014	Qualitative comparison The ability of the model to reproduce the flash flood hazard level from observations was evaluated with different skill scores, e.g. CSI	
			Discharge observations	Observed			
	FLOOD-PROOFS	PRD-95 to PRD-97	-	-	-	-	Not reported
	Landslide and debris flow	PRD-98 to PRD-99	Ground truth observations Existing records	Observed	2010	Qualitiative analysis of the model to identify the reported events Also, quantitaitve analysis in terms of hits, misses, and false alarms	

Table 5: Overview of background information for uncertainty assessment natural hazards (cont'd)

Storm surges	ESS	PRD-100 to PRD-104:	Ground-truth dataset from 208 tidal gauges available from the JRC Sea Level Database (http://webcritech.jrc.ec.europa.eu/SeaLevelsDb) and EMODNET web site (http://www.emodnet-physics.eu/Map/))	Observed	1979-2016	The ability of model to reproduce surge and waves was evaluated using different skill scores, e.g. RMSE,%RMSE, correlation coefficient, Normalized BIAS, Normalized RMSE	The model performance was evaluated not only for the whole time series, but also considering extreme surge and wave event (upper tail >99th percentile). Statistics shows a satisfactory performance of the European and storm surge model
			Altimetry data via satellite measurement GLOBAL OCEAN ALONG-TRACK L3 SEA SURFACE HEIGHTS REPROCESSED from CMEMS. Altimeter missions included (Topex-Poseidon; Topex-Poseidon (interleaved orbit); Jason-1; Jason-1 (interleaved orbit); Jason-1 (geodetic orbit); OSTM/Jason-2; OSTM/Jason-2 (interleaved); Jason-3; Sentinel-3A; ERS-1; ERS-2; Envisat; Envisat (extended phase); Geosat Follow On; Cryosat; SARAL/AltiKa; SARAL-DP/ALtiKa; HY-2A; HY-2A (geodetic orbit)). (http://marine.copernicus.eu)		1992-2016		
			Global significant wave height via satellite (Queffeulou and Croizé-Fillon, 2014) Missions included ERS-1&2, TOPEX-Poseidon, GEOSAT Follow-ON (GFO), Jason-1, Jason-2, ENVISAT and CryoSat, SARAL		1991-2016		

			ftp://ftp.ifremer.fr/ifremer/cersat/products/swath/altimeters/waves/data/				
	Regional Storm Surge model	PRD-107 to PRD-110	Tidal-gauge located at Stavanger (Kartverket) http://api.sehavniva.no/tideapi_en.html	Ob-served	1988-2017	Comparison of model performance to reproduce storm surge at Stavanger for selected extreme events	Regional model shows a good performance for the extreme storm surge. Wave measurements were not available for the evaluation of regional model
	Inundation and erosion model	PRD-106; PRD-111 to PRD-116	Documentary sources from the World Wide Web (pictures, videos' snapshots) of extreme events occurred in 1994	Proxy	Historic event	The computed inundations were compared with a visual estimation of the flood depth	Statistics shows a good performance of the local inundation model

Table 5: Overview of background information for uncertainty assessment natural hazards (cont'd)

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy / Simu-lated)	Length time series	Comparison method	Remarks uncertainty
Heatwaves and air quality	Universal Thermal Climate Index (UTCI)	PRD-117	Forecasts (high resolution, ensemble control, ensemble mean)	Reanalysis (ECMWF ERA-Interim)	1 January 2009 to 31 December 2012	Anomaly correlation coefficient (ACC)	Ensemble mean UTCI forecasts have higher skills in predicting heat stress conditions
	Regional Air Quality (RAQ)	PRD-118 to PRD-123	Forecasts (single models, ensemble model	Near-real time air quality surface monitoring data	9 and 15 June 2014	Mean bias (MB), modified normalised mean bias (MNMB), the root mean square error (RMSE), fractional gross error (FGE), correlation coefficient (R)	Ensemble median RAQ forecasts have higher skills than single model forecast both at daily and seasonal level

Table 5: Overview of background information for uncertainty assessment natural hazards (cont'd)

Weather-induced fires	European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF)	PRD-124 to PRD-136	FWI (PRD-124) combines all the other fire indices, therefore is the only one used here for validation	1. SYNOP 2. Fire Radiative Power	1. 1 year (2017) 2. 1 month (June)	1. Probability of detection 2. Visual inspection of FRP time series versus the evolution of fire forecasts	In Europe, the POD is generally above 60%, even at Day 10. At the local scale, the persistency of high danger conditions in a given region for successive forecasts tends to increase confidence in the forecasts.
	RISICO	PRD-137 to PRD-142	-	-	-	-	Not reported
	PROPAGATOR	PRD-143 to PRD-144	-	-	-	-	Not reported

Table 5: Overview of background information for uncertainty assessment natural hazards (cont'd)

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy/ Simu-lated)	Length time series	Comparison method	Remarks uncertainty
Droughts	Drought-Standardised Indices	PRD-148 to PRD-151	Precipitation, Precipitation-Evaporation, runoff, and Groundwater	Observed: meteorological drought Proxy: hydrological drought	Observed/proxy: 1990-2018 Reforecasts: 2002-2010.	Drought class categorical method, and skill score metrics (BSS)	
	Drought-Threshold Indices	PRD-152 to PRD-155	Discharge	Proxy: discharge	Observed/proxy: 1990-2018, Reforecasts: 2002-2010.	Skill score metrics: BSS	
	Drought-Areal Indices	PRD-157 to PRD-160	Precipitation and runoff (presented). Groundwater (not presented)	Observed: meteorological drought Proxy: hydrological drought	Observed/proxy: 1990-2018 Reforecasts: 2002-2010	Drought class categorical method, and skill score metrics (BSS)	

Table 5: Overview of background information for uncertainty assessment natural hazards (cont'd)

Convective storms, severe winds, and precipitation types	Convective cells	PRD-161 to PRD-162	OPERA radar data composite	Observed and created nowcasts	Eight convective events in central Europe in 2016-2017	Nowcasts are compared with observations and skill scores are calculated based on these	
	Snow-load and gust algorithms	PRD-165 to PRD-181	Total snow load, freeze snow load, wet snow load, MET parameters CP tasks log, electricity break statistics	Modeled and observations	2010-2018	Modeled values are compared to observed impacts	
	precipitation type - FMI	PRD-163 to PRD-164	NWP based product probability of precipitation product and METAR observations	3 hourly forecasts and METAR observations from 39 airport weather stations	15th October 2016 to 15th February 2017	Model estimates are compared to METAR observations and comparison is also performed related to similar product of the ECMWF	
	precipitation type - ECMWF	PRD-182 to PRD-200	Observations from manual SYNOP stations in Europe (not automatic). Only stations where the difference between the	3-hourly observations from manual SYNOP stations	3 years, data from 15 October-31 March. Verification done on each year individually, and	Skill of the forecasts were measured with reliability and resolution (ROC curves and Cost/loss ration diagrams). Also	Improvements or changes in the version of the IFS ENS system can have some effect in the precipitation

			height of the closest model grid point and the height of the observation point <200m were used.		all three as a whole	a performance diagram for for all precipitation types and lead times.	type products, so it is recommended to compare only individual years, instead of accumulated , when significant changes in the precipitation field were applied in a new cycle (version) of the model.

4 Robustness of weather-induced natural hazards products under a future climate

The uncertainty of the algorithms/tools to forecast natural hazard products has been described in the previous chapter. However, is an algorithm/tool with an acceptable forecasting skill also robust? An algorithm/tool with an acceptable skill is supposed to perform its intended function when conditions are nominal. As long as design details and environmental conditions remain stable, you can count on such an algorithm/tool to do its task. But what happens if these same design details and environmental conditions start drifting significantly off from nominal? The established acceptable skill might alter. Making an algorithm/tool robust, adds complexity to the design process (e.g. Jensen, 2014).

In this chapter we deal with the question if the ANYWHERE algorithms/tools to forecast natural hazard products with a particular uncertainty under current environmental conditions (Chapter 3) are robust under a future climate. In other words, does the uncertainty change? The description of robustness under a future climate is mainly based on literature. It appeared to be impossible in ANYWHERE to do a robustness modelling experiment to assess natural hazards under a future climate, because this would mean weather and hydrological forecasts (Chapter 2) with future climate as benchmark. This is clearly beyond the scope of ANYWHERE.

The outcome of the investigation whether algorithms/tools are transferable in space and time (towards a future climate) or not, is explained in the following sections. The ANYWHERE developers concluded that none of the algorithms/tools uses a fully physically-based approach, which implies that any algorithm/tool has to be parameterized. Hence, robustness means whether an algorithm/tool can be parameterized in any region in any time. Where does the algorithm/tool works, and where not.

Successively, we describe robustness in using algorithms/tools on floods and landslides, storm surges, heatwaves and air quality, fires, droughts, and convective storms, severe winds and heavy snowfall. Details on the algorithms/tools and the associated products can be found in D2.3 (Ciavola et al., 2017) and D2.4 (Van Lanen et al., 2019). As mentioned in Chapter 1, no common methodology could be developed and applied to assess robustness of all algorithms/tools that forecast natural hazards. Hence, we report separately per hazard, and a table is given (Section 4.6) that summarizes per hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether these were observed, proxy or simulated data, (iv) what were the length of time series data, and (v) which comparison method has been applied.

4.1 Flash floods, landslides and debris flows[8]

In the project, no specific analyses have been done to assess the robustness of the algorithms under a future climate. This section discusses the effect of the changes in future climate scenarios.

The expected increase in the frequency and intensity of heavy rainfall based on climate models (e.g. Kundzewicz et al., 2013) should result in more frequent torrential events.

The algorithm for flash floods nowcasting FF-EWS (UPC) uses on the observed and nowcasted catchment-aggregated rainfall to estimate the flash flood hazard at each point of the drainage network by comparison with reference values (thresholds) linked to a probability of occurrence (or return period). Consequently, the robustness of the algorithm will be mainly affected by the robustness of the thresholds of the catchment-aggregated rainfall used to estimate the return period at each point of the drainage network.

Similarly, the robustness of the algorithm for landslides and debris flows nowcasting (UPC; Palau et al., 2020) in future climate scenarios will be affected by the evolution of the two components of the method: (i) the evolution of the rainfall Intensity-Duration curves used to assess the magnitude of the rainfall event, and (ii) the landslides susceptibility of the terrain (strongly related to the future evolution of the land cover).

Adjusting both algorithms in new climate scenarios would thus require re-calibrating the rainfall thresholds for accurate hazard assessment, and in the case of the landslides nowcasting algorithm, re-calculating the terrain susceptibility with updated land cover scenarios.

Other factors such as the expected lower predictability of the rainfall systems in the future climate scenarios might also have an effect on the performance of both algorithms.

4.2 Storm surges

First robustness aspects of using the Storm Surge Model at the pan-European scale are described followed by the model at the regional scale. The section concludes with the Inundation and Erosion Model.

4.2.1 European Storm Surge model (ESS) (CFR)

In the ANYWHERE project, no specific analyses have been carried out to test the robustness of the European Storm Surge model (ESS) under a future climate. The expected changes in future climate scenarios that could affect to the storm surge and wave predictions are those related with the intensity, frequency and or path of the storm events. Regarding this point, high robustness of the storm surge model is expected,

[8] The robustness of the Flood-PRObabilistic Operational Forecasting System, Flood-PROOFS (CIMA) has not been studied.

since no additional changes should be needed in model parameterization, as proved in the validation described in Section 3.2.1.

Nevertheless, there are several factors that can contribute to reduce robustness of forecasting storm surge level and waves under future sea level scenarios. First, the uncertainty in projections of sea level increase due to global warming (Jevrejeva et al., 2016). Sea level rise will directly affect the storm surge levels (Arns, 2015), as well as wave predictions (Vousdoukas et al, 2012). Thus, semi-enclosed basins with shallow water areas (i.e North Sea, Baltic Sea and/or Northern Adriatic Sea) may be the most affected ones. In these cases, a recalibration of the storm surge model may be necessary. Moreover, rising mean sea levels may modify the tidal phase and amplitude (Idier et al., 2019). Thus, the tidal water level component might need to be modified, finally producing changes in surge and waves. Additionally, in turn it can change the surge-tide interaction, affecting the model performance in specific areas like the North Sea (Fernández-Montblanc et al., 2019). Finally, changes in precipitation and river discharge may increase their effects on coastal hydrodynamics, thus increasing the necessity to model these phenomena in the European Storm Surge forecasting system, which presently does not include these combined interactions.

4.2.2 Regional Storm Surge model (CFR)

In the case of the regional model, there is no requirement for further parametrization due to variation in frequency, or intensity of storm. Nevertheless, the largest changes expected in next century are related to relative sea level rise. Simpson et al. (2012) report changes ranging from 0.32-1.21 m for the high-end scenario. Based on the IPCC AR5 RCP8.5 the projections of relative sea level rise ranges from 0.35 to 0.79 m (Simpson et al., 2015). The most updated projections (IPCC AR5 RCP8.5) show a relative sea level increase from 0.28 to 0.9 m (Simpson et al., 2017). However, the effect of wind on surge and tidal components is inversely proportional to the water depth. Therefore, the effect of sea level changes will be limited in the Stavanger fjord area that is characterized by larger gradients of bathymetry nearshore reaching in the deepest areas up to 700 m. Indeed, the physiography of the area of Stavanger, being located in a fiord confers robustness and diminished the effect of sea level rise effect on water level and wave height forecasts under future scenarios.

Another factor to be considered as potential constrain of the robustness of the regional storm surge model is the concurrence of wet hazards. Indeed, under the high emission (RCP8.5) scenario, for the period 2071-2100 the 200-year river flood will increase ~31-40% in the area of Rogaland (Hanssen-Bauer et al, 2017). It may reduce the robustness of regional storm surge. This could be solved by including the parametrization of new processes, such as the river discharge into the fiord.

4.2.3 Inundation and erosion model (CFR)

Robustness of the local inundation model of Stavanger will be mainly related with the inputs provided to the algorithms at European and regional scale. On the other hand, its robustness will depend on the land-use changes that affect the infiltration and

roughness setup for the urban area. Land-use changes are controlled by socio-economic factors and the implementation of adaptation measures at the local level. The intrinsic characteristics of the local flood model, that is the high resolution of the different parameters, such as topography, roughness or infiltration, provide a low level of robustness to the local algorithm and model parametrization will be needed.

The robustness of flood forecasting under future climate scenarios may be also controlled by the concomitance of heavy rainfall and storm surge phenomena. This can increase the flood hazard, and changes in the parametrization of local flooding models will be required. Indeed, the RCM-based median projections indicate an annual increase in number of days with heavy rainfall of 89% and 49% for RCP8.5 and RCP4.5 by the end of the century. The winter season shows the largest increases (143% for RCP8.5) (Hanssen-Bauer et al,2017), which in turn is the period characterized by the higher frequency of occurrence of extreme storm surge. According to Simpson et al. (2015), the water level height of the 200-year event in Stavanger, based on the reference period 1986-2005, is projected to be exceeded in four out of ten years during the present century.

4.3 Heatwaves and air quality (weather-induced health)

Robustness of using algorithms/tools to forecast heatwaves and air quality products is explained. These products are associated with weather-related health.

4.3.1 Universal Thermal Climate Index (UTCI) (UOR)

Studies on future projections for UTCI have been performed both at the global and local level. At the global level, Kjellstrom et al. (2017) investigated the future UTCI bioclimate 2071-2099 (1981-2010 as baseline, ISI-MIP data, HadGEM and GFDL models) under a RCP6.0 greenhouse gas scenario. The result of the projection is given in Annex IV. The projects show a consistent agreement in predicting conditions of higher heat stress and extend into the future the increase in heat stress observed during recent past decades (Di Napoli et al 2018). It is worth noting, however, that the UTCI cannot not be considered robust on the future climate. This is for two reasons. First, bioclimatic indices rely on the current weather and the ability of its forecast. Second, bioclimatic indices suppose an ability to acclimatisation (i.e. the adaptation to climate) equal to the one nowadays observed in populations. Future acclimatisation is expected to be different from the current one and to decrease vulnerability to heat-related hazards. A few studies have been attempting to include this aspect too (e.g. Ballester et al. 2011) but none of them use bioclimatic indices.

4.3.2 Regional Air Quality (RAQ) (UOR)

Anticipating future air quality is a major concern and it has been the focus of many atmospheric chemistry research projects over the past decades. We here report the results from two recent studies on RAQ forecasts under a future climate (Annex V) for two pollutants considered in ANYWHERE , namely ozone and particulate matter.

One study has been carried out by Colette et al. on ozone (2012). In their analysis an ensemble of air quality models covering both regional and global spatial scales were implemented in a coordinated manner for future projections of anthropogenic emissions at the 2030 horizon. With regards to results' robustness, changes in the concentration of ozone and its precursors are quite consistent across the ensemble (inter-model uncertainty). Furthermore, a statistical bias correction at the location of monitoring stations was applied in order to derive unbiased proxies of future exposure to air pollution.

Future European particulate matter concentrations have also been evaluated under the influence of climate change and anthropogenic emission reductions (Lacressonnière et al., 2017). Large differences are observed between the models used in the study but the decrease of particulate matter over Europe associated with emission reduction was proved robust by the authors conducting the study.

4.4 Weather-induced fires[9]

The robustness of the European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF) is described.

Currently, EFFIS-GEFF does not run under future climate scenarios, as the estimation of fire risk is more valuable at short to medium range and once the ignition has already taken place. However, there are works in the literature that have started investigating fire probability patterns under seasonal and future climate projections. Below we mention the most relevant studies and future research directions.

At the seasonal scale, Turco et al. (2018) tried to estimate the expected fire activity a few months ahead, which would be useful to reduce environmental and socio-economic impacts through short-term adaptation. The authors found out that using currently available operational seasonal climate predictions, the skill of fire seasonal forecasts remains high and significant in a large fraction of the burnable area (~40%). In the future, ECMWF could attempt to reproduce this study, if/when a seasonal fire danger forecasting system will be implemented.

At longer time scale, Guyette et al. (2014) calculated changes in fire frequency and probability under current and future climates using future climate simulations of temperature and precipitations (these are some of the main drivers in EFFIS-GEFF) and found that fire probability increased in cooler northern and high-elevation regions and decreased in some hotter and drier regions. However, they also stated that non-climatic factor that affect fire might be difficult to predict in the distant future. Syphard et al. (2018) used statistically corrected models to map future fire probability patterns under climate change. They found that model predictions are highly sensitive to the

[9] There is no report on the robustness aspects of RISICO - fire danger rating system (CIMA) and PROPAGATOR - propagation of a wildfire (CIMA).

state of vegetation. If ANYWHERE were to provide forecasts of wildfires under future climate we would need to incorporate the vegetation factor (currently omitted in EFFIS-GEFF), as it seems to play an important role in long-term predictions.

4.5 Droughts[10]

There are several global and pan-European studies of drought under a future climate. Van Lanen et al. (2018) provide an overview of these studies, incl. uncertainties (see for a summary Annex VI), from which some observations about robustness can be made. Most of these studies use the threshold-based drought approach (e.g. Van Huijgevoort et al., 2014; Prudhomme et al., 2014; Wanders and Van Lanen, 2015; Wanders et al., 2015; 2019). A smaller number of drought projections are based upon standardized drought indices (e.g. Orlowsky and Seneviratne, 2013; Spinoni et al., 2018).

Drought forecasts use seasonal probabilistic hydrological forecasts (EFAS, Section 2.2) as input, which are driven by probabilistic weather forecasts issued by ECMWF-IFS (Section 2.1.1). This means that robustness of drought forecasts under a future climate firstly depends on the robustness of the seasonal hydrometeorological forecasts. Last step in drought forecasts is the identification of drought events using two approaches, i.e. the standardized drought approach and the threshold-based drought approach, which are driven by the ensemble time series of hydrometeorological variables. Robustness of this last step is reported for the standardized drought indices and threshold-based indices.

4.5.1 Standardized Indices (WUR)

Imperfectness of model structures is one of the reasons for a reduced robustness of forecasts algorithms/tools under future climate. Calibration of model parameters on current environmental conditions, incl. climate, can hide imperfectness. Orlowsky and Seneviratne (2013) have studied past and future meteorological drought using the Standardized Precipitation Index accumulated over 12 months (SPI-12).

[10] There is no report on the U&R aspects of using the European Drought Observatory (EDO) because JRC had no person months to work on this activity.

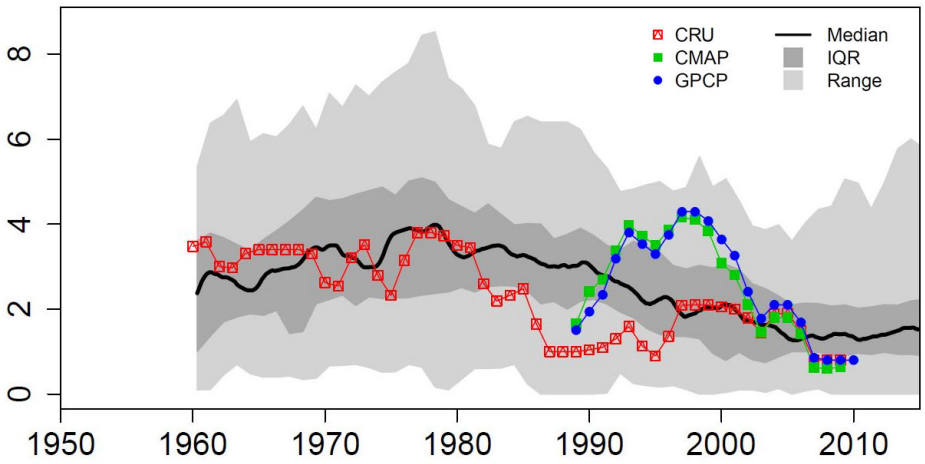


Figure 53: Occurrence frequencies (months per year) in Central Europe of "mild drought" (SPI12<–0.5) in observation-based datasets and CMIP5 simulations (10 yr moving windows). SPI12 values from three observation-based datasets (coloured lines) and median, inter-quartile range (IQR) and total range (Range) across the CMIP5 ensemble (black line, dark grey and grey shading, respectively, total 39 GCMs). SPI12 values are calculated with respect to the 1979–2009 period for all datasets (derived from Orlowsky and Seneviratne, 2013).

They compared the SPI-12 derived from three observed time series of precipitation across the globe ($SPI\text{-}12_{obs}$) for the period 1960-2009 with simulated SPI-12 obtained from 39 different climate models ($SPI\text{-}12_{sim}$) that were not calibrated. Figure 48 gives the outcome for one of the global regions, i.e. Central Europe. The three time series of $SPI\text{-}12_{obs}$ are within the total range of $SPI\text{-}12_{sim}$ time series, but not always within the inter-quartile range (Fig. 53). The spread in the range exposes substantial differences among models for current environmental conditions. An ensemble of a lower number of climate models, which is common for projections or forecasts of hydrological extremes, would result into a different median total and inter-quartile range than using all climate models. Hence, projections of SPI also depend on the selected set of models, i.e. model structures. This also applies to meteorological drought forecasts using SPI implying that robustness is affected by the selected climate model(s).

In ANYWHERE, robustness of drought forecasts has been analyzed at the local scale using SPI-x (x=1, 3, 6, and 12 months) (Van Hateren et al., 2019). The analysis was performed using the BSS as a skill metric (Annex II). A sensitivity study was performed to investigate how robust the forecasts are, if the analysis was performed using other thresholds to identify drought, i.e. SPI = 0 and SPI = -1, instead of SPI = -0.5.

The results of the sensitivity analysis for the drought year 2006-2010 for two catchments in Catalonia are visualized in a color-coded table. Four arbitrary categories were distinguished, based on what was considered to be an acceptable value. The BSS was considered to be good if the values are in between 0.50 and 1.00 (green

color), acceptable if the values are in between 0.00 and 0.50 (yellow color), poor if the values are in between -0.50 and 0.00 (orange color) and bad if the values are in between -∞ and -0.5 (red color).

Table 6 shows the results of the sensitivity analysis to assess an aspect of robustness, i.e. the threshold that determines if the SPI-x in certain month is below the threshold. An acceptable skill is obtained for SPI-12 (BSS > 0.5) for all thresholds and the two catchments (except for the Ripoll, threshold SPI=-1).

Table 6: Summary of the sensitivity analysis of the forecast skills of meteorological drought (SPI) for the Ripoll (RI) and Guardiola (GU) catchments (Spain) to explore robustness. The BSS (entire year) is divided into two lead time periods: per catchment, the upper row shows the average BSS for lead times 1, 2 and 3 months, and the lower row shows the average BSS for lead times 4, 5, 6 and 7 months

Catchment	Lead time	SPI threshold											
		SPI=-0.5 (reference)				SPI=0				SPI=-1			
		SPI-1	SPI-3	SPI-6	SPI-12	SPI-1	SPI-3	SPI-6	SPI-12	SPI-1	SPI-3	SPI-6	SPI-12
RI	1-3	-0.16	-0.09	0.32	0.68	-0.05	-0.17	0.14	0.67	-0.08	0.01	0.41	0.47
	4-7	-0.14	-0.28	-0.28	0.06	-0.06	-0.38	-0.44	0.33	-0.04	-0.14	-0.13	-0.01
GU	1-3	-0.29	-0.26	0.25	0.69	-0.30	-0.33	0.23	0.72	-0.16	-0.03	0.25	0.53
	4-7	-0.25	-0.40	-0.38	0.10	-0.33	-0.69	-0.77	0.22	-0.17	-0.23	-0.17	0.06

We can learn from the sensitivity analysis that the differences in BSS for SPI-12 are in general rather small, irrespective of the selected threshold, which points at relatively high robustness of this model aspect.

Standardized drought indices are based upon probability distributions that eventually convert the empirical distribution into a normal distribution (e.g. McKee et al., 1993). In ANYWHERE we used the gamma distribution to forecast standardized drought indices (SPI, SRI, and SGI), while the three-parameter log-logistic distribution was used to calculate the re-forecasted SPEI (Vicente-Serrano et al., 2010). The gamma distribution has quite a flexible shape parameter, which is applicable to the wide range of accumulated precipitation in Europe (Stagge et al., 2015). A study by Slater et al. (2018) also shows that the gamma distribution can be used for hydrological forecasting of both high and low flows. Drought forecasting under a future climate should investigate if the probability distribution and its parameters are still valid under a future climate.

4.5.2 Threshold-based Indices (WUR)

Several large-scale studies using threshold-based drought indices have investigated the impact of applying different models. Wanders and Van Lanen (2015) present the projected average and standard deviation of the drought duration and deficit volume for different climate regions from three climate models. It is assumed that the smaller the differences between the projected drought characteristics are the higher the robustness is. Prudhomme et al. (2014) and Wanders et al. (2019) use the spread (e.g. inter-quartile range, IQR) in projected drought occurrence from a multi-model experiment (several climate and a number of hydrological models) as a metric, the lower the spread the higher the robustness.

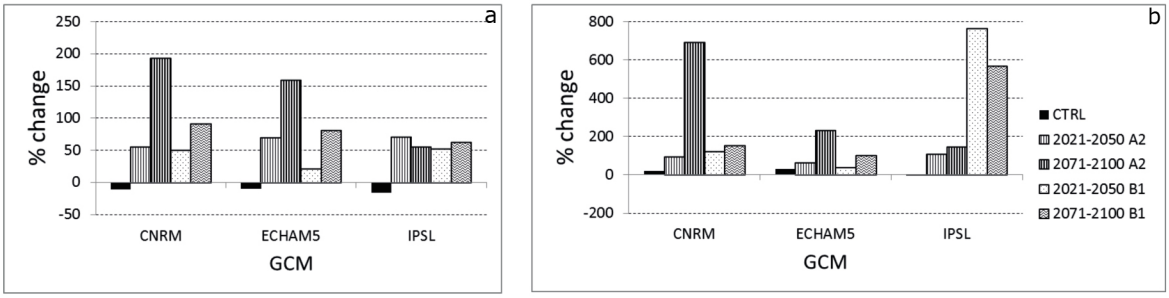


Figure 54: Projected change in drought duration derived from runoff for three GCMs, two emission scenarios and the intermediate (left) and far future (right). CTRL specifies the difference between runoff simulated with the GCM and the runoff simulated with re- analysis data (proxy for observed). Results are shown for the average change of all 244 grid cells. Detailed analysis of projected drought can be seen in Alderlieste et al. (2014).

Another way to approach robustness is by investigating the spread in hydrological drought characteristics from the reference/control period (CTRL) among climate models. Drought characteristics derived from observed hydrological variables (e.g. runoff flow) are compared with simulated runoff from a hydrological model driven by a climate model. Alderlieste et al. (2014) has used this approach to calculate the difference between hydrological drought duration obtained from observed runoff and runoff simulated with climate model input (CTRL) (Fig. 54).

If the projected change in drought duration is larger than the CTRL, it is assumed that the projection is robust, in particular if the hydrological drought duration obtained with all three climate models is pointing in the same direction.

Wanders et al. (2015) have investigated in a global-scale study how robust projected hydrological drought characteristics are for the selected threshold approach. In modelling experiment they used a threshold which is based upon the runoff in the reference period versus a transient threshold that considers a changing hydrological regime in the 21st Century. The projected hydrological drought duration and the deficit volume are expected to increase from 27% to 62% relative to the end of the 20th C, when the benchmark threshold is used. The study illustrates that robustness of drought projections is very much dependent on the selected thresholds.

4.5.3 Areal Indices (WUR)

We did not do a specific robustness analysis for the areal indices. However, we believe that the findings would not deviate from the standardized drought and threshold-based experiences, because basically the areal indices are derived from gridded information, which is described in the two previous sections.

4.6 Convective storms, severe winds and heavy snowfall

The robustness of algorithms/tools that forecast typical weather-type natural hazards are described. First, small scale convective storms are explained, which are followed by heavy snow loads. Uncertainty involved in forecasting different types of precipitation is described in the last part.

4.6.1 Detection and forecasting convective cells (FMI)

The CC-ITN algorithm is based on measured reflectivity and therefore, there is no restrictions of applying it for future storms (Rossi et al., 2014; 2015). However, the statistical determination of storm severity is likely to alter and must be determined based on new statistical data.

4.6.2 Snow-load and gust algorithms (FMI)

The future climate in northern Europe is projected to change during the 21st century (Räisänen and Ylhäisi, 2015). In winter, both temperature and precipitation are projected to increase, hence in the coldest areas, snowfall is generally estimated to increase (Räisänen, 2015), while over milder regions, a larger fraction of total precipitation is expected to fall as rain. How this anticipated change will affect the risk of snow damage for trees and infrastructure is not straightforward given the sensitive nature of snow accumulation to specific weather conditions. In Annex VII the future heavy snow load due to climate change in Finnish forests using the FMI snow load model is presented (Lehtonen et al., 2016). The FMI snow load model is an empirical algorithm, based on a tuned parametrization of accumulation and removal terms of estimated snow load for four different snow load types. The chosen factors are based on the meteorological experience and observations. Therefore, the robustness of the model in future climate is difficult to be addressed.

4.6.3 Probability of precipitation type (ECMWF/FMI)

FMI precipitation type

No analysis of robustness of the FMI precipitation type algorithm under a future climate can be provided. Physical parametrization should be valid also in future climate, but thresholds for the frequency of occurrences are expected to change.

ECMWF precipitation type

The 5th assessment report from IPCC concludes that for Europe, temperature is

increasing for all of Europe, and precipitation is increasing in northern Europe and decreasing in southern Europe (IPCC, 2014). However, heavy precipitation is expected to increase in all of Europe (Fig. 55). The increase in temperature will have an effect on the frequency of freezing rain and sleet (mixed phases) or in the snow, but this will vary heavily for different areas, as it will decrease in some areas, and increase in others (Groisman et al, 2016). However, the overall is an increase, following the general increase in precipitation.

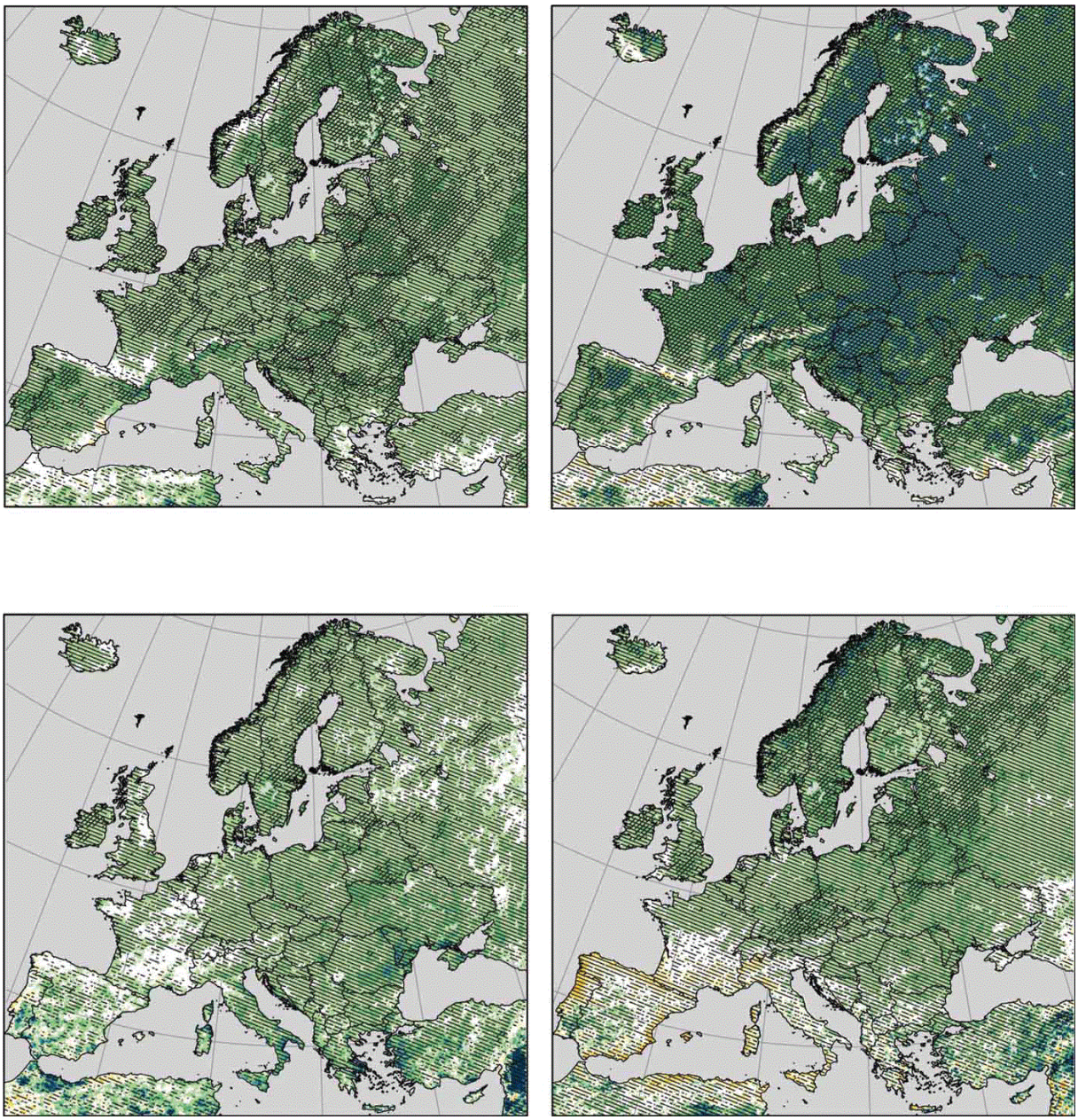


Figure 55: Projected seasonal changes in heavy precipitation defined as the 95th percentile of daily precipitation (only days with precipitation >1 mm day–1 are considered) for the period 2071–2100 compared to 1971–2000 (in %) in the months December to February (DJF, top panel) and June to August (JJA, bottom panel) for the emission pathways 4.5 (left column) and 8.5 (right column).

Whether the precipitation type formulation is robust in terms of future climate is a question as to the ECMWF IFS is a robust model for climate change impacts. The probabilities of precipitation type products are designed in such a way that the precipitation rate variable is involved, indicating different intensities for each type of precipitation as well, so the breakpoints could be re-defined for the different intensities of precipitation if the climate change requires it in the future. It is important to point out that these products are also calibrated for each IFS model cycle and each precipitation type. This calibration consists in defining a minimum precipitation rate threshold for each precipitation type to have a bias=0 and reduce as much as possible the under-prediction and over-prediction of the different precipitation types. This will be a very useful tool in the future climate to optimize the product whatever the climate conditions are. Also, IFS is, as mentioned earlier, constantly changing and is therefore by default able to cope with new situations in light of changing physical properties in a future climate. The precipitation rate thresholds for probability of precipitation type will have to be recalibrated with each new model cycle, but the underlying physics is robust. The IFS is already today very close to a full earth system model and, has been shown to capture observed climate change well. In fact, it is used as both a climate change model within the EC-EARTH consortium and as a reference climate model for the past using the reanalysis ERA5 being the latest version. It is therefore likely that the precipitation probabilities are robust when it comes to climate change.

4.7 Summary of background information for robustness assessment under a future climate

Table 7 summarizes for the robustness analysis per hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether observed or proxy data have been used, (iv) length of time series of observed or proxy data, and (v) method that has been used to compare the forecasted with. It appears that robustness of algorithms/tools that forecast natural hazards under a future climate covers many different aspects, for example, (i) robustness of platforms that forecast weather and hydrology under a future climate, which is input to natural hazard algorithms/tools, (ii) model structure of natural hazard algorithms/tools, (iii) parameters, and (iv) alert/emergency threshold.

Table 7: Overview of background information for robustness assessment of natural hazards under a future climate

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy / Simu-lated)	Length time series	Comparison method	Remarks robustness
Floods, flash floods, debris flows, and landslides	FF-EWS Landslide and debris flow	PRD-93 to PRD-94 PRD-98 to PRD-99	NA	NA	NA	NA	No specific analyses on robustness have been carried on. Section 4.1 discusses what components of the algorithms will be affected by a future climate.
	FLOOD-PROOFS	PRD-95 to PRD-97	-				Not reported
Storm surges	ESS	PRD-100 to PRD-104:	NA	NA	NA	NA	No specific analyses on robustness have been carried on. Section 4.2 anticipates which effect a future climate might have on the model parameters of the storm surge models.
	Regional Storm Surge model	PRD-107 to PRD-110	NA	NA	NA	NA	
	Inundation and erosion model	PRD-106; PRD-111 to PRD-116	NA	NA	NA	NA	

Table 7: Overview of background information for robustness assessment of natural hazards under a future climate (cont'd)

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy / Simu-lated)	Length time series	Comparison method	Remarks robustness
Heatwaves and air quality	Universal Thermal Climate Index (UTCI)	PRD-117	NA	NA	NA	NA	The UTCI, as any other bioclimate index, is not robust
	Regional Air Quality (RAQ)	PRD-118 to PRD-123	Ozone: ensemble of air quality models PM: ensemble of regional chemical transport models	EMEP/EEA air pollutant emission inventory	Ozone: 1998-2007; 2030 PM: 1998-2007, 2050	Ozone: inter-model uncertainty, statistical bias correction PM: inter-model uncertainty	Ozone: decrease in concentration PM: decrease in concentration
Weather-induced fires	European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF)	PRD-124 to PRD-136	NA	NA	NA	NA	Not tested under future climate
	RISICO	PRD-137 to PRD-142	-	-	-	-	Not reported
	PROPAGATOR	PRD-143 to PRD-144	-	-	-	-	Not reported

Table 7: Overview of background information for robustness assessment of natural hazards under a future climate (cont'd)

Hazard	Tool/algorithm	Product numbers (D2.4)	Which data	Data type (Observed / Proxy / Simu-lated)	Length time series	Comparison method	Remarks robustness
	Drought-Standardised Indices	PRD-148 to PRD-151	Orlowsky and Seneviratne: SPI-12 Van Hateren et al.: SPI-x (x=1, 3, 6 and 12)	Orlowsky and Seneviratne: proxy observed, simulations from climate models Van Hateren et al.: proxy observed, simulations from ECMWF and EFAS platforms	Orlowsky and Seneviratne: proxy 1960–2009, climate model simulations 1950–2100 Van Hateren et al.: proxy 1990–2016, ECMWF/ EFAS simulations 2002-2010	Orlowsky and Seneviratne: maps, descriptive Van Hateren et al.: BSS	Orlowsky and Seneviratne: projections: 39 climate models, model structures Van Hateren et al.: test thresholds in standardized indices
	Drought-Threshold Indices	PRD-152 to PRD-155	Prudhomme et al. (2014): drought occurrence runoff Alderlieste et al.: Low flow MAM7, hydrological drought duration and deficit Wanders et al. (2015):	Prudhomme et al. (2014): climate, hydrological models simulations Alderlieste et al.: proxy observed, simulations from climate models, hydrological model Wanders et al. (2015): proxy	Prudhomme et al. (2014): 2070-2099 vs. 1976-2005 Alderlieste et al.: proxy 1971–2000, model simulations 1971–2100 (3 time windows) Wanders et al. (2015): proxy	Prudhomme et al. (2014): maps, descriptive, CDFs Alderlieste et al.: maps, descriptive	Prudhomme et al. (2014): structures of climate, hydrological models Alderlieste et al.: Wanders et al. (2015): test

			hydrological drought duration and deficit	observed, simulations from climate and hydrological models	1971–2000, model simulations 1971–2099	Wanders et al. (2015): maps, descriptive	threshold approach
	Drought-Areal Indices	PRD-157 to PRD-160	NA	NA	NA	NA	Can be derived from standardized and threshold indices
Convective storms, severe winds, and precipitation types	Convective cells	PRD-161 to PRD-162	NA	NA	NA	NA	Not tested under future climate
	Snow-load and gust algorithms	PRD-165 to PRD-181	five global climate models under RCP4.5 and RCP8.5,	Climate models with modelled snow load	Comparison of periods 1980-2009 to 2070-2099	maps, descriptive	Single climate model, snow load algorithm, hard to address robustness
	precipitation type - FMI	PRD-163 to PRD-164	NA	NA	NA	NA	Not tested under future climate
	precipitation type - ECMWF	PRD-182 to PRD-200	NA	NA	NA	NA	Not tested under future climate, but expected to be robust. Because included in ECMWF-IFS update cycles

5 Robustness of compound natural hazards

In this chapter we deal with the question if the ANYWHERE algorithms/tools to forecast natural hazard products are robust when analysing compound natural hazards. In other words does the uncertainty of algorithms/tools (Chapter 3) change when coinciding or cascading hazards occur. This might happen because interactions between natural hazards are not fully covered. Analysis of compound hazards is only just on the agenda of current research on hydrometeorological hazards, which is reflected by the ongoing DAMOCLES COST Action[11]. This multi-disciplinary initiative coordinates activities within Europe to gain a better understanding of compound events. It aims to establish a network of communities including climate scientists, hydrologists, impact modellers, risk modellers, statisticians and stakeholders (Damocles, 2019). Providing the full picture of robustness of algorithms/tools to forecast natural hazard products is challenging. Some experiences with ANYWHERE algorithms/tools on compound hazards are explained in the following sections. These findings will contribute to a more comprehensive view on algorithms/tools to asses compound natural hazards, which eventually provide more insight into robustness, i.e. if interaction between hazards occurs.

Successively, we describe experiences within ANYWHERE with algorithms/tools that assess compound wet hazards (e.g. coastal flooding) followed by compound dry hazards (fires, heat stress, drought). We conclude the chapter with a table (Section 5.3) that summarizes per compound hazard: (i) which forecast products have been assessed, (ii) which data have been used, (iii) whether these were observed, proxy or simulated data, and (iv) what were the length of time series data.

5.1 Wet natural hazards

We start with an explanation of cascading wet events, namely how extreme precipitation triggers a sequence of consequences with significant magnitude. Pathways of cascading in three European regions are described. The section concludes with a description of compound events in the coastal region (combined storm surges and river mouth flooding).

5.1.1 Cascading effects of extreme precipitation with pathway schemes (UNIGE)

Extreme precipitation events with high local precipitation intensities, heavy snowfall or extensive freezing rain can have devastating impacts on society and economy. Not only is the quantitative forecast of such events sometimes difficult and associated with large uncertainties, also are the potential consequences highly complex and challenging to predict. It is thus a demanding task to anticipate or nowcast the impacts

[11] http://damocles.compoundevents.org/stsm.php.

of extreme precipitation, even more so in situations where human lives, or critical infrastructure might be at risk.

In recent years, the term “cascading effects” has been increasingly used to describe events in which an initial trigger leads to a sequence of consequences with significant magnitude. We here analyze three examples for different precipitation types where the initial triggering event generated a cascade of events and impacts, namely a convective precipitation event in the Swiss Prealps, a freezing rain in Slovenia, and a heavy snowfall episode in Catalonia. With the aim to improve process understanding of complex precipitation-triggered events, we assess the prediction of the selected events and analyze the cascading effects that caused diverse impacts. To this end, we use a framework of cascading effects which should ultimately allow the development of a better design risk assessment and management strategies. Readers are referred to Schauwecker et al. (2019) for a comprehensive description of the study.

Our findings confirm that damage of extreme precipitation events is clearly related to the knowledge of potential cascading effects. Major challenges of predicting impacts of cascading effects are, e.g. the high complexity and the interdependencies of the consequences, as well as the increasing uncertainty along the cascade (starting with the uncertainty of the forecast, see Fig. 56). Due to these challenges, we suggest that it is more difficult to assess the robustness in forecasting precipitation-triggered cascading hazards compared to single hazards and different approaches are needed (not evaluated here).

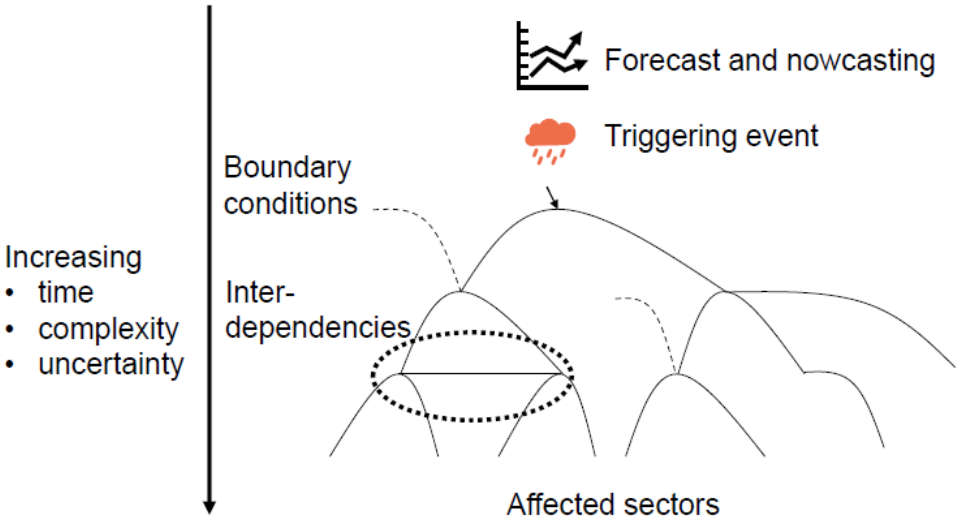


Figure 56: Scheme of challenges in anticipating cascading effects. The forecast of the triggering event is already related to uncertainty (depending on the precipitation type, situation etc.). Along the cascade with increasing time and complexity, the uncertainty increases significantly. Interdependencies appear where several elements are interconnected, and key boundary conditions control the cascade at certain points.

To assess cascading effects of heavy precipitation, we propose a framework including two approaches: (i) one to analyze cascading effects during past extreme precipitation events, which then serves as a basis for a (ii) more generalized approach to increase the preparedness level of operational services before and during future extreme precipitation events and to anticipate potential cascading effects of extreme precipitation. Both approaches are based on pathway schemes that can be used in addition to numerical models or hazard maps to analyze and predict potential cascading effects, but also as training tools.

5.1.2 Coastal flooding (CFR)

To explore coincidence of extreme storm surge events and extreme river discharges, compound effect of extreme storm surge level (SSL) and river discharge (DISC) at river mouth has been analyzed. The analysis has been performed using two reanalysis datasets. The SSL dataset spans the period 1979 to 2016 and covers the pan-European domain with 10 km resolution along the coastline (Fernánez-Montblanc et al., in press). The river discharge dataset covers the period 1990 to 2016 and has a spatial resolution of ~5 km (JRC-LISFLOOD reanalysis, SFO dataset).

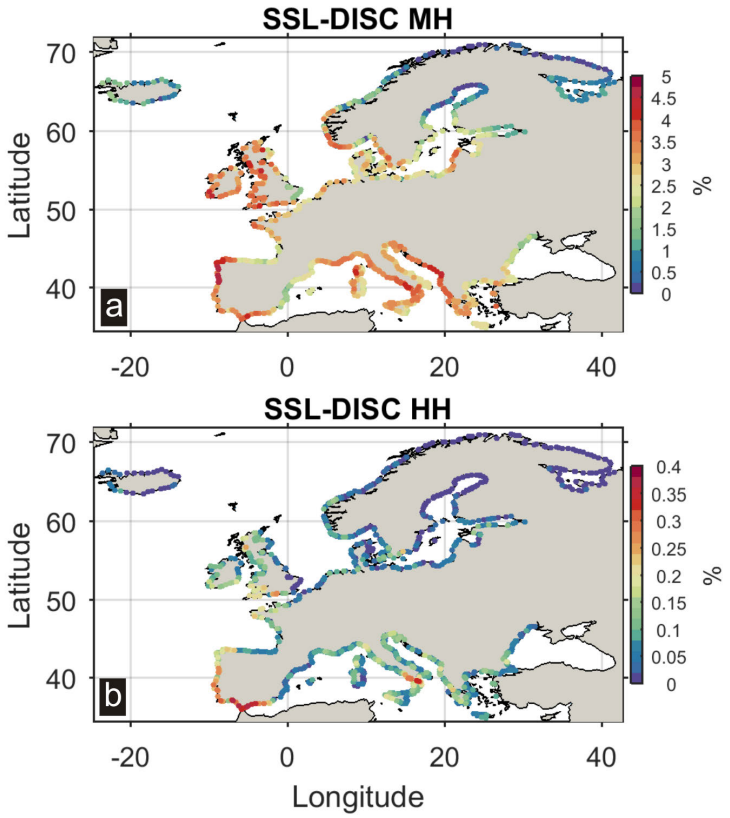


Figure 57: Coincidence of storm surge (SSL) and river discharge (DISC) along the European coastline for the period 1990-2016: (a) % of time medium hazard SSL and DISC level is equalled/exceeded, and (b) % of time high hazard SSL and DISC level is equalled/exceeded.

The hazard level for both datasets was ranked according to a predefined threshold. Thus, low hazard was defined for values of SSL and DISC below 90th percentile, medium hazard was considered when SSL and DISC values were between 90th and 99th percentiles. High hazard occurred when SSL and DISC were above the 99th percentile. Subsequently, 3D (latitude, longitude, time) binary blocks for each hazard SSL and DISC were constructed. Finally, the analysis to identify compound wet hazard hot-spots was done by considering the period 1990-2016. Low-, medium- and high-multi-hazard occurrences were defined when the respective thresholds where exceeded for both datasets.

Compound wet hazard hot-spots were identified for the period 1990-2016. The analysis of storm surge and river discharge equaling/exceeding the medium hazard levels shows a high level of coincidence of the two hazard analyzed with the exception of the northernmost part of Europe (lat. > 60º) (Fig. 57a). The high degree of coincidence along the western European coast facing the Atlantic should also be noted, which can be related with the trajectories of low pressure system, the main driver of both phenomena. The coincidence of storm surge and river discharge equaling/exceeding the high hazard level is significantly lower than for the medium hazard level (Fig. 57b). The maximum % of time of coinciding of these wet events in the coastal region for the high hazard level is about 10 times lower than for the medium hazard level (0.4 versus 5%), which of course is connected to the pre-defined thresholds. Higher level of coincidence of events was observed in the western and southern part of the Iberian Peninsula, and also in the south of Italy and the English Channel. There is a well-document occurrence of compound wet events for the southern regions of Italy, where small mountain catchments in coastal areas of Campania, Calabria and Sicily were flooded regularly, which have caused economic losses, as well loss of lives (Aronica et al., 2009).

The previous analysis has identified areas with high level of coincidences of storm surge and high river discharge at river mouths, especially for the medium hazard level. This preliminary investigation of compound wet hazards in coastal regions along the European coastline is subject to several limitations, which affects robustness. The main constrain is related to the temporal resolution of the river discharge dataset (daily). A more detailed analysis will require higher temporal resolution, accounting for the interaction with tidal dynamics and its time-scale. For areas where surge and river discharge occur simultaneously, their joint probability distribution needs to be determined, adopting more sophisticated numerical approaches like multivariate statistical modeling (Bevacqua et al., 2017).

Especially in large catchments, river discharge could increase the total water level near the river mouth, increasing the SSL hazard in the nearby coastal zones. At the same time, especially in large estuaries, the river outflow can be hampered by the total water level at the river mouth. The near future challenge on forecasting SSL and river discharge is to evaluate their mutual influence considering the interaction between these processes at pan-European scale. Likewise, the assessment of changes in the

spatial pattern and interaction of SSL and river discharge under future climate scenarios needs further investigation (Bevacqua et al., 2017).

5.2 Dry natural hazards

In the first part compound wildfire and heat stress are described, which is followed in the second part by three so-called dry hazards (fires, heatwaves and droughts).

5.2.1 Combined wildfire and heat stress (ECMWF/UOR)

Heat stress and forest fires are often considered highly correlated hazards as extreme temperatures play a key role in both occurrences. This commonality can influence how civil protection and local responders deploy resources on the ground and could lead to an underestimation of potential impacts, as people could be less resilient when exposed to multiple hazards. Earlier a simple methodology to identify areas prone to concurrent hazards was published (Vitolo et al., 2019b). The methodology exemplified with, but was not limited to, heat stress and fire danger. The combined heat and forest fire event that affected Europe in June 2017 was used to demonstrate that the methodology can be used for analyzing past events as well as making predictions, by using reanalysis and medium-range weather forecasts, respectively. The operational use of new spatial layers was proposed, mapping the combined danger and making suggestions on how these could be used in the context of a Multi-Hazard Early Warning System. These products could be particularly valuable in disaster risk reduction and emergency response management, particularly for civil protection, humanitarian agencies and other first responders whose role is to identify priorities during pre-interventions and emergencies.

The new proposed layers include a map of hotspots of combined danger (Fig. 58, focused on Europe) that can be used to analyse the extent of a past event and its spatial correlation with other observed variables (working in retrospect, this makes use of reanalysis data) or to make a prediction for the future (using forecast data). The confidence in the identification of combined danger in a given area is provided by a map of probability of occurrence that is generated by spatial overlap of fire danger and heat stress ensemble forecasts (Fig. 59, with a focus on the Iberian Peninsula). Lastly, an overview of the evolution of the forecasts is provided by the monthly forecast summary plot for combined hazards (similar to Fig. 33).

Assessing the uncertainty and robustness of the newly generated combined danger maps is particularly complex because they are forecasted maps of hazard. The quantification of the risk to the population and environment would require information on exposure and vulnerability. This means any actual occurrence of fire requires the appropriate conditions (expressed by the fire danger index), an ignition source as well as populated areas, presence of fuel, lack of access to transport/means of self-evacuation, no insurance, amongst others. Heat stress is conditional to people being outdoors or not using air conditioning systems indoors as well as the environmental conditions expressed by the index. Nevertheless, uncertainty for a past event can be

assessed in terms of spatial accuracy of hotspot location and temporal span in which a dangerous signal can be predicted.

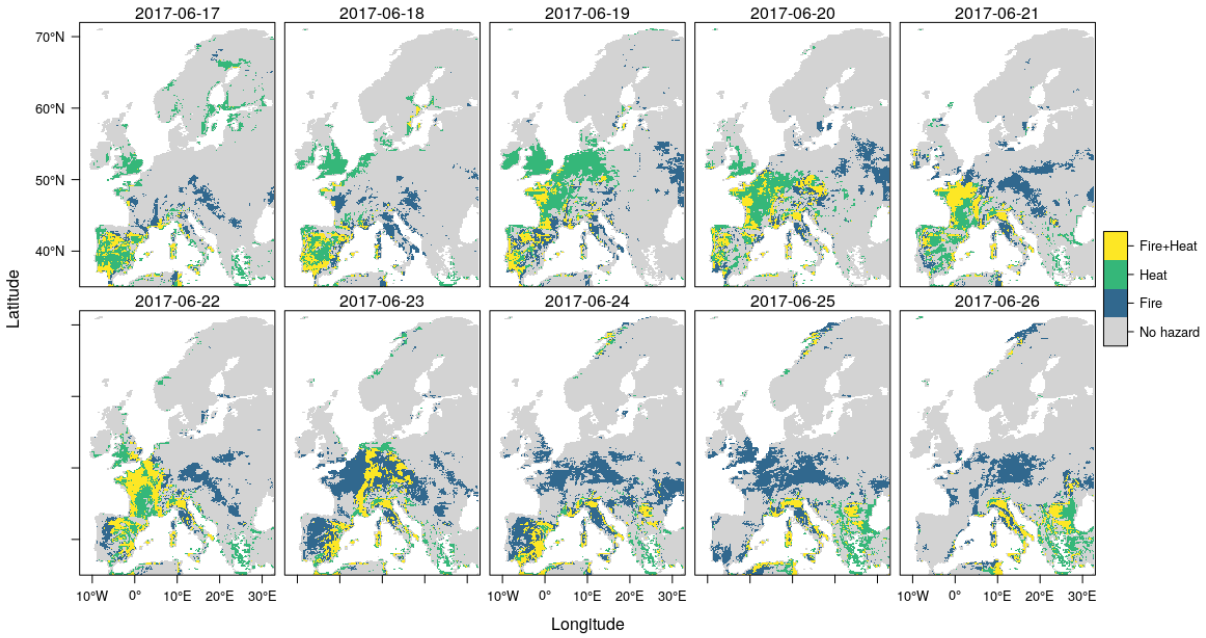


Figure 58: Multi-hazards hotspots forecast maps (forest fire and heat stress) at pan-European scale. The maps are generated on 17th June 2017 and with a 10-day horizon.

A preliminary assessment of the utility of these new layers was made in collaboration with various stakeholders (see Table 2, Vitolo et al., 2019b), amongst them is the Instituto Português do Mar e da Atmosfera (IPMA) who collaborated with decision makers in the 2017 event in Pedrógão Grande. Positive feedback was recorded, with the Portuguese Institute stating: “we liked the new layers, they identify well areas at risk, [...] they would have been useful during the event” [IPMA, Lourdes Bugalho, personal communication]. This statement highlights the importance, for stakeholder, to correctly identify hotspots of danger. It is also important that forecasts provide enough notice to allow preparedness and, in case of emergency, efficient allocation of resources.

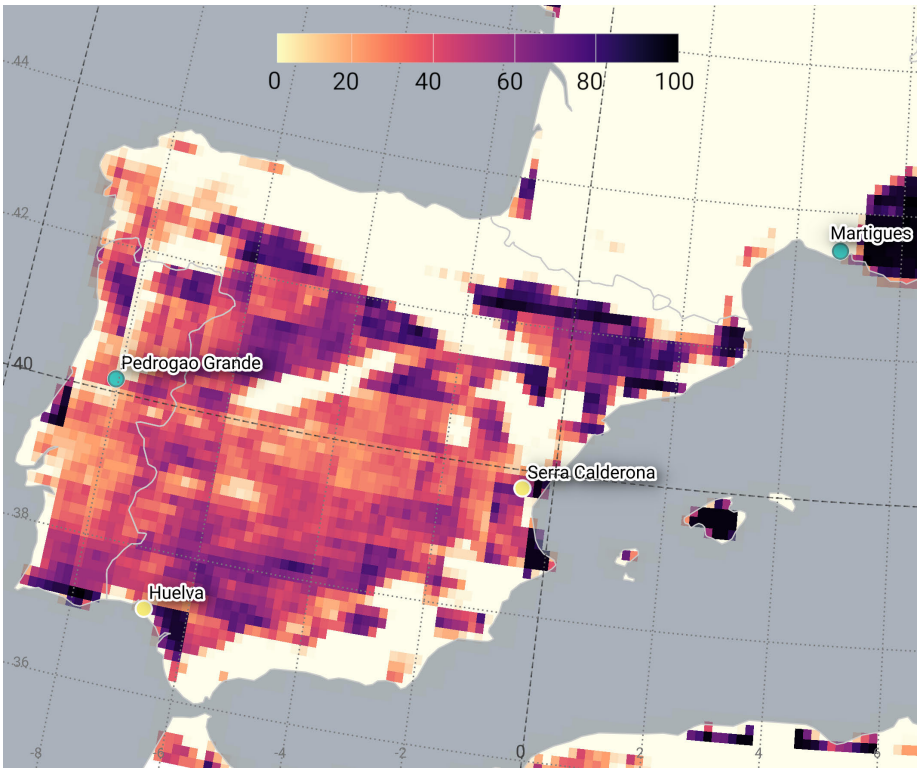


Figure 59: Map of probability of occurrence of very high multi-hazard danger over the Iberian Peninsula, generated on 17th June 2017.

As already mentioned in Sections 4.3.1 and 4.4.1, we currently have no modelling tools to explore the effects of future climate scenarios on either fire or heat stress danger. Consequently, we cannot foresee if/how their interplay will change in the future.

5.2.2 Heatwaves, drought and fires (WUR/ECMWF/UOR)

Coinciding or cascading dry hazards are expected to have more negative impacts than each of the hazards alone (Liu et al., 2016). For example, the drought that occurred in 2003 was not the most severe in Europe (Spinoni et al., 2015). However, in combination with extended heatwaves and fires, it is considered as the most fatal and costly. More than 70,000 people passed away as a result of extreme heat conditions (Robine et al., 2008; Di Napoli et al., 2018) and the economic damage amounted to over 8.7 billion EUR (EC, 2007). In this study on compound dry hazard, we try: (i) to investigate through historical data the coinciding and cascading events of dry hazards, namely drought, heatwaves, and fires across Europe, and (ii) to develop a methodology to explore these events on daily basis. Here we define coinciding hazards as two or more extreme events occurring simultaneously, i.e. on the same day and in the same region. We define cascading events as two or more extreme events (as single and/or as coinciding hazards) occurring successively without being interrupted by a zero-hazard day.

The analysis of coinciding and cascading dry hazards was carried out on daily basis for the summer seasons (June, July, August; referred to as JJA hereafter) from 1990

to 2018 across Europe (for details, see Sutanto et al., 2019b). This is the period when high-impact heatwaves take place and droughts and/or fires may lead to even higher-impact compound events.

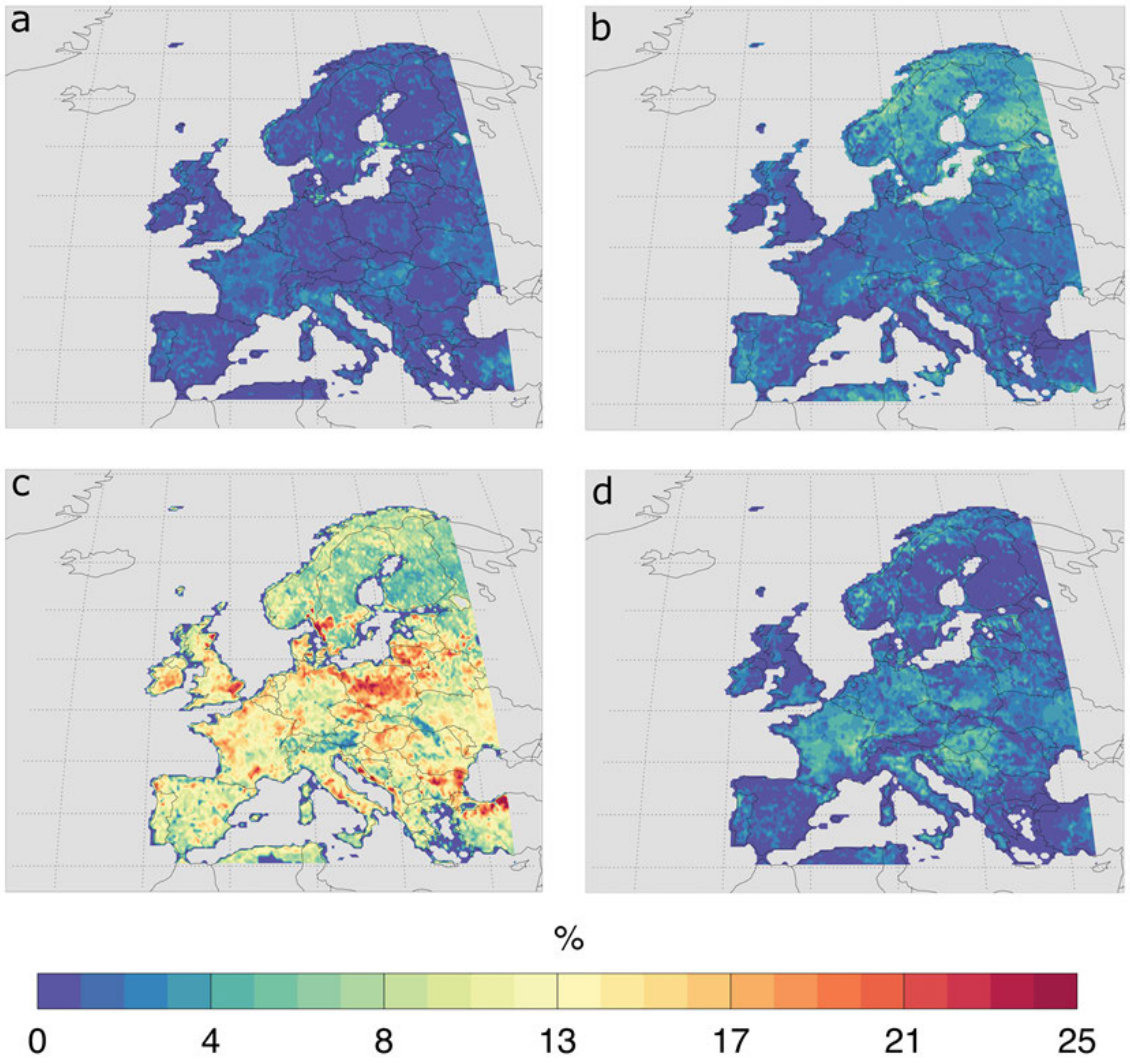


Figure 60: Hotspots of coinciding dry hazards (P90 of yearly % of number of compound days for each year/total number of summer days x 100): a) compound drought-heatwave, b) heatwave-fire c) drought-fire, and d) the three dry hazards all together in Europe obtained from daily proxy observed datasets covering the JJA periods of 1990-2018.

Figure 60 shows hotspot locations that were calculated from the 90th percentile of the yearly number of days with concurrent hazards (P90) divided by the total number of summer days (92 days). Hotspot locations for coinciding drought-heatwave appeared to be spread throughout Europe with a stronger signal in France, Italy, Spain, and east Europe (Fig. 60a). However, the occurrence of drought-heatwave is relatively small, ranging from 0-4%. An interesting result is found for coinciding heatwave-fire

(Fig. 60b). Hotspots are clearly identified in the Scandinavian countries, and to some extent in Portugal and Sicily Italy. The number of heatwave-fire occurrences is twice as high compared to drought-heatwave, especially in northern Europe (up to 8%). The coinciding event with the highest occurrence in Europe is drought-fire (Fig. 60c). Large extents of concurrent drought-fire hotspots are clearly identified in central Ireland, southeastern UK, parts of Germany, southeast France, western Italy, and southeastern Europe. The Iberian Peninsula is not listed as a hotspots area although they also suffer from coinciding drought-fire events (Pausas and Fernández-Muñoz, 2012; Gudmundsson et al., 2014). In this region, the occurrence of drought-fire is less than in central Europe. The hotspots of all three dry hazards concurring mainly occurred in large parts of west, central, and east Europe, from southern UK, France, Germany, Italy, to Romania and Bulgaria, and less frequent in southern Europe, such as Spain, and eastern Scandinavia (Fig. 60d).

Table 8: The most frequent cascading patterns of dry hazards in Europe calculated from daily proxy observed datasets covering the JJA periods of 1990 to 2018. D stands for drought, F stands for fire, and H stands for heatwave

No	Cascading pattern	Number of events (-)	Number of events (%)
1	D-DF	32584	5.9
2	D-DF-F	31247	5.6
3	F-D	24817	4.5
4	D-DF-D-DF	22877	4.1
5	H-HF	21989	4.0
6	F-DF	20501	3.7
7	DF-D	19213	3.5
8	DF-F	14860	2.7
9	F-DF-D	7589	1.4
10	HF-H	7257	1.3

Understandably, the high occurrence of coinciding drought-fire (DF) (Fig. 60c) appears also in the cascading patterns of dry hazards, as obtained by summing up the number of cascading events in the JJA periods of 1990-2018 across Europe, i.e. at all grid cells (Table 8). DF can be found in 7 out of 10 of the most frequently occurring patterns of cascading events. Most cascading events in Europe are dominated by the occurrence of drought in the beginning, i.e. they start with drought, followed by the compound drought-fire (D-DF, Table 8 row 1). This cascading pattern occurred 5.9%, or 32,584 events out of 555,931 events multiplied by 100%, calculated from all land grid cells and from 2668 days. Cascading patterns starting with fire (F) and heatwave (H) are found at rank 3 and 5 with an occurrence of 4.5% and 4.05%, respectively. Interestingly, there is a high number of events that started with fire and ended up with drought (F-D, 4.5%). This cascading event occurred only for short periods as the frequency of fire occurring in a cascading event is relatively low compared to drought.

Our study concludes that dry hazard hotspots were identified largely for an area stretching from west to east Europe, from southern UK, France, Germany, Italy, to Romania and Bulgaria, and with a lower frequency in southern Europe, such as Spain, and eastern Scandinavia. In the study period 1990-2018 (JJA), 0.55% of all cells had an occurrence of all dry hazards in the same day. Droughts dominate in coinciding and cascading dry hazard events and mainly control the number and duration of cascading events, especially in the Mediterranean. In most cascading events, drought appears first as a single hazard, followed by the concurrent drought-fire. This leads to a combination of drought, and drought-fire, as the most frequent cascading pattern of dry hazards in Europe (5.9%). The probability of coinciding and cascading dry hazard occurrence calculated from this study is relatively small. However, it will increase in future due to climate change. The development of MH-EWS that counts for coinciding and cascading hazard therefore is at utmost importance. We note that we had to use proxy data to determine the occurrence of the three selected dry hazards, as long time series of reported dry hazards do not currently exist. The use of real observations, such as from radars, satellites, and gridded in situ observations will provide a more robust analysis. The detailed information of data, methods, and drought forecast uncertainty and robustness is published in Sutanto et al., 2019b.

5.3 Summary of background information on assessment of compound natural hazards

Table 9 summarizes experiences on wet and dry compound hazards (coinciding and/or cascading) obtained during the lifetime of the ANYWHERE project. For each compound event, the following information is provided: (i) type of compound event, (ii) which data have been used, (iii) whether observed, proxy or simulated data have been used, (iv) length of time series, (v) method, and (vi) reference.

Table 9: Overview of background on assessment of compound natural hazards

Compound hazards	Hazards	Type of compound event	Which data	Data type (Observed / Proxy / Simu-lated)	Length time series	Method	Remarks
Wet hazards	Cascading effects of extreme precipitation with pathway schemes	Cascading	We used a large variety of forecast and observation data connected to the three cases	Observations, forecasts	up to approx. 1 week for each of three study cases	Compare past forecasts with in-situ observations for three case studies	Three cases in Switzerland, Slovenia and Spain (Schauwecker et al., 2019)
	Coastal flooding	Coinciding	Storm surge levels, river discharge	Storm surge: simulated along the coastline, river discharge: proxy	Storm surge: 1979-2016, river discharge: 1990-2016	Maps with European coastline, descriptive	Fernánez-Montblanc et al. (in press)
Dry hazards	Combined wildfire and heat stress	Coinciding	Reanalysis, and ECMWF HRES, ENS	Proxy observed, simulated climate data, Local Authority records, Portugal	1 month (June 2017)	Maps, visual inspection	Stakeholders interviewed were satisfied with the information on combined hazards and are willing to test it, operationally, during future events (Vitolo et al., 2019b)
	Heatwaves, droughts and fires	ERA5: daily max and min temperature. Soil moisture SFO data. FWI	Proxy	Daily data from 1990 to 2018 for summer months (June, July, and August).	Combining binary time series.	Maps, descriptive	Sutanto et al. (2019b)

6 Operationalizability of forecast and nowcast tools/algorithms/ products

In the previous chapters we described in a thorough way the uncertainty and robustness of the platforms/algorithms/tools that are used in ANYWHERE to forecast natural hazards. In this chapter we translate for each platform/algorithm/tool the comprehensive information in operational terms, namely: (i) development stage, (ii) uses, (iii) required data/resources to be implemented at the local and pan-European scale, and (iv) possible deviation from what originally was promised at start of the ANYWHERE project (June 2016). The development stage is characterized by the so-called Technology Readiness Level (TRL) scale, which was introduced for the EU funded projects in 2014, as part of the Horizon 2020 framework program (see for a brief description, Annex VIII). We describe the operationalizability aspects at the level of platforms/algorithms/tools, but if there are differences in these aspects for products generated by a particular platform/algorithm/tool then we make a distinction between these forecast/nowcast products.

6.1 Overview of operationalizability

An overview of the development stage for the development stage of ANYWHERE platforms/algorithms/tools is given in Table 10. Slightly **over 50% of the platforms/algorithms/tools that reported development stage is at TRL=9** meaning that these technologies have proven in an operational environment. About 10% is below TRL=5, that is, technology is not validated in relevant environment yet. **From the remaining 40%, almost all algorithms/tools have a technology already demonstrated in a relevant environment (TRL=6).**

Table 10: Overview of development stage of platforms/algorithms/tools to forecast natural hazard

		Platform/ Algorithm/ tool	Product numbers	Develop-ment stage (TRL, Annex VIII)	Remarks
Hydrometeorological forecasts and nowcasts	**Meteorological forecasts and nowcasts products**	IFS-ENS	PRD-1 to PRD-32	**9**	
		IFS-ENS-EXT			
		IFS-ENS-SEAS			
		HIRLAM	PRD-201 to PRD-203	**9**	
		HARMONIE-AROME	PRD-204 to PRD-217		Previously reported as GLAMEPS and LAM- MEPS
		Precipitation nowcasting (UPC)	PRD-33 to PRD-39	**9**	The algorithm is working in real-time in the European Flood Awareness System (EFAS)
		PhaSt	PRD-40 to PRD-42	**9**	Operational in Liguria, Italy
		RAVAKA	PRD-43 to PRD-58	**-**	Has been removed, see footnote section 2.1
		Downscaling (UNIGE)	-	**6**	No product number
	Hydrological forecasts	EFAS		**9**	
Natural hazard forecasts	**Floods, flash floods, debris flows, and landslides**	FF-EWS	PRD-93 to PRD-94	**9**	The algorithm is working in real-time in the European Flood Awareness System (EFAS)
		FLOOD-PROOFS	PRD-95 to PRD-97	**-**	Not reported
		Landslide and debris flow	PRD-98 to PRD-99	**6**	
	Storm surges	ESS	PRD-100 to PRD-104:	**6**	
		Regional Storm Surge model	PRD-107 to PRD-110	**6**	
		Inundation and erosion model	PRD-106; PRD-111 to PRD-116	**6**	

Table 10: Overview of development stage of platforms/algorithms/tools to forecast natural hazard (cont'd)

		Platform/ Algorithm/ tool	Product numbers	Development stage (TRL, Annex VIII)	Remarks
Natural hazard forecasts	**Heatwaves and air quality**	Universal Thermal Climate Index (UTCI)	PRD-117	**6**	
		Regional Air Quality (RAQ)	PRD-118 to PRD-123	**9**	
	Weather-induced fires	European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF)	PRD-124 to PRD-136	**9**	
		RISICO	PRD-137 to PRD-142	**-**	Not reported
		PROPAGATOR	PRD-143 to PRD-144	**-**	Not reported
	Drought	European Drought Observatory (EDO)	PRD-145 to PRD-147	**-**	Not reported in D2.5, but TRL=9 (operational service)
		Drought-Standardised Indices	PRD-148 to PRD-151	**8**	
		Drought-Threshold Indices	PRD-152 to PRD-155	**8**	
		Drought-Areal Indices	PRD-157 to PRD-160	**5**	
	Convective storms, severe winds, and precipitation types	Convective cells	PRD-161 to PRD-162	**3**	
		Snow-load and gust algorithms	PRD-165 to PRD-181	**9**	
		precipitation type - FMI	PRD-163 to PRD-164	**9**	
		precipitation type - ECMWF	PRD-182 to PRD-200	**9**	

6.2 Operationalizability of hydrometeorological and forecast and nowcasts products

This section describes operational aspects of using platforms/algorithms/tools (Fig. 1, ❶ and ❷) that generate hydrometeorological forecast and nowcast products. The hydrometeorological and forecast and nowcasts are divided in weather forecasts and nowcasts (Section 6.2.1) and hydrological forecasts (Fig. 1, ❸ and ❹) (Section 6.2.2).

6.2.1 Meteorological forecast and nowcasts products

The first part explains operational aspects of platforms/algorithms/tools that produce weather forecasts, which is followed by operational aspects of nowcasts algorithms/tools. The section concludes with an example that describes operational aspects of the downscaling example.

ECMWF Integrated Forecasting System (ECMWF-IFS)

a) Development stage (TRL) of the platforms/tools/algorithms/products
The ECMWF Integrated Forecasting System (ECMWF-IFS) has proven in an operational environment (TRL=9).

b) What purposes the platforms/tools/algorithms/products have been used
The ECMWF-IFS is the basis for several algorithms/tools that forecast natural hazards using the ANYWHERE MH-EWS. ECMWF-IFS provides global probabilistic weather forecasts.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The platform is run by the ECMWF and needs lots of inputs. Very specialized knowledge and experience is required run the platform. Downscaling of output is needed for several local applications (e.g. Section 2.6.1). License is requisite to obtain ECMWF-IFS output.

d) Deviation of final platforms/tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
ECMWF implements updates of IFS on a regular basis. Since June 2016 there has been 4 major updates (cycles) implemented of the IFS model. None of these has been a change in resolution of the atmospheric variables, but there has been changes to the physics. It is very difficult to assess the impact on the actual forecasts since there are always natural variability in the skill. However, the ensemble forecasts of precipitation have over the time of the project showed a slow but steady improvement (Fig. 61). For a full update on the model changes of IFS, see weblink [12].

[12] https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model.

Figure 61: The range at which the 12-month mean centred on that month of the continuous ranked probability skill score of the 24-hour precipitation ensemble forecast dropped below 10%. Figure taken from www.ecmwf.int.

FMI Numerical Weather Prediction models (FMI-NWPs)

a) Development stage (TRL) of the platforms/tools/algorithms/products
As stated in Section 2.1.2 the NWP products provided by FMI are based on already operational models of HIRLAM, HARMONIE-AROME and the product of precipitation type utilizes also the synoptic scale from GFS and ECMWF. Therefore these products have reached the majority of TRL 9 (i.e. actual system proven in operational environment). They are produced as part of the FMI operational production chain and are distributed to the end-users in Finland through the FMI ILMANET-service and for ANYWHERE through A4EU-platform. The accuracy of temperature, wind gust and precipitation rate is verified continuously and results are presented in real time in the FMI pages[13]. The forest fire index product is continually performed by the forecasting meteorologists and based on this analysis the products are developed further.

b) What purposes the platforms/tools/algorithms/products have been used
The platforms are used in Finland by the forecasting meteorologists to create a full picture of the weather situation and they are provided to the end-users upon request. Typically, e.g. the civil production authorities are following the wind gust and forest fire index products to estimate the threatening weather-induced hazards and impact. In ANYWHERE these NWP products were applied for the

[13] https://ilmatieteenlaitos.fi/saaennusteen-osuvuus.

A4FINN impact tool to create the weather-related combined impact level for the ISTIKE pilot site.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
These models cover the computational area of the used NWP model. To implement these to wider or different area, this would need the suitable existing NWP model parameters. These are available, but typically open access models are with coarser resolution, therefore similar resolution cannot be guaranteed, if the NWP model assumptions need to be changed. The change for the wind, temperature and precipitation products would require data programming skills, however not scientific research. For the forest fire index feasibility of the algorithm behind should be customized to the governing surface and forest types.

d) Deviation of final platforms/tools/algorithms relative to originally promised (DoA at start of ANYWHERE , June 2016)
In the beginning of project, different NWP model products were offered to the ANYWHERE -project for A4FINN. The change was explained in the D2.3 (Ciavola et al., 2017), the products offered at the start were based on a model, which development is decided to stop, and the newly chosen products that were included in the project are now more likely to have longer lifespan.

UPC-CRAHI algorithm for precipitation nowcasting

a) Development stage (TRL) of the tools/algorithms/products
The UPC-CRAHI algorithm for precipitation nowcasting has proven in an operational environment (TRL=9). The current version of the algorithm has been running to generate some of the products of the flash flood layer of EFAS, and several configurations (one at Continental scale, and four for the pilot sites) have been implemented in the ANYWHERE MH-EWS.

b) What purposes the tools/algorithms/products have been used
The precipitation nowcasts are used as a product that can be accessed on the A4EU platform to monitor the evolution of precipitation situations in the near future. They are the main inputs to the FF-EWS flash-flood hazard nowcasting algorithm.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The algorithm only requires high-resolution gridded quantitative precipitation estimates (QPE), such as those from regional, national or Continental weather radar networks. Typically, the resolutions of these radar products are of the order of 0.5-2 km and 5-15 min. In the context of ANYWHERE , the algorithm has been applied with the regional radar QPE products of the Meteorological Service of Catalonia for the Pilot Site of Catalonia, and the OPERA radar composites (**expanded with the Italian national radar composite for the first time in the framework of ANYWHERE**) over Europe and for the Pilot Sites of

Liguria, Corsica, Canton Berne and Spain. But there is no restrictions to include other national/regional composites where available.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviation.

Radar-based nowcasting precipitation PhaSt (CIMA)[14]

a) Development stage (TRL) of the tools/algorithms/products
The Radar-based nowcasting precipitation algorithm PhaSt has been operational in the Liguria region since 2010 (TRL=9).

b) What purposes the tools/algorithms/products have been used
Phast could be used operationally by a local user or civil protection to better predict precipitation.

Downscaling precipitation mountainous regions (UNIGE/METEODAT)

a) Development stage (TRL) of the tools/algorithms/products
The downscaling and bias correction approach has been calibrated and evaluated for areal precipitation of three study sites in the Swiss mountains. The study has been submitted (Schauwecker et al., 2019) The technical readiness of the forecast downscaling approach in a relevant environment is achieved (TRL=6).

b) What purposes the tools/algorithms/products have been used
The approach has been used to downscale (to a 2 km horizontal resolution) and bias-correct ECMWF IFS-HRES daily precipitation forecasts with a focus on heavy precipitation (>20 mm day^{-1}) and medium lead times (2-5 days) in mountain regions. The main purpose is a medium-range forecast for an operational application of decision makers and intervention bodies.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The approach has the potential to be applied at a pan-European scale. We developed and evaluated the approach for a mountain environment in the Swiss Alps, but this bias-correction and downscaling approach has also the potential to be applied to other mountainous or flat regions. The main strength of the approach is that it is computationally cheap. It is therefore possible to run it for a larger region without exceptional amounts of computational resources. The requirements are: (i) meteorological station data of daily precipitation, and (ii) time series of past daily precipitation forecasts of a certain forecast model.

[14] Point c) What is needed to implement it, and point d) Deviation of final tool/algorithm relative to originally promised at DoA are not reported.

These data are used for the calibration phase to compute the correction factors per month, cell and precipitation intensity.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE, June 2016)
Since it was not foreseen to implement this tool in the MH-EWS, there is no deviation from the DoA (version June 2016).

6.2.2 Hydrological forecast products

a) Development stage (TRL) of the tools/algorithms/products
The European Flood Alert System (EFAS) has proven in an operational environment (TRL=9).

b) What purposes the tools/algorithms/products have been used
Input time series of EFAS (gridded probabilistic weather forecasts derived from ECMWF-IFS) have been used by the drought algorithms (Section 3.5) to forecast meteorological drought across Europe. Output from EFAS (gridded probabilistic forecasts of hydrological time series) drives some algorithms for flood forecasting (Section 3.1.1) and algorithms for hydrological drought forecasting at the pan-European scale (Section 3.5).

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The EFAS platform is run by the ECMWF and needs lots of inputs. Very specialized knowledge and experience is required run the platform. Usually, downscaling is not needed for local applications, if a spatial scale of 5 km acceptable. Users can get access to EFAS output via the Copernicus Emergency Management Service[15].

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE, June 2016)
No deviation, other than an update of the spatial coverage (major update of EFAS) in May 2018. This update was implemented in the operational ANYWHERE platform and the products were recalibrated accordingly. There was no disruption to the service.

6.3 Operationalizability of tools/algorithms for forecasting & nowcasting of natural hazards

This section describes operational aspects of using algorithms/tools (Fig. 1, ❺) that translate hydrometeorological forecast and nowcast products (Fig. 1, ❸ and ❹) into natural hazard forecast products (Fig. 1, ❻). Successively, we describe operationalizability of algorithms/tools on floods and landslides, storm surges,

[15] https://emergency.copernicus.eu/mapping/ems/early-warning-systems-efas-and-effis.

heatwaves and air quality, fires, droughts, and convective storms, severe winds and heavy snowfall.

6.3.1 Floods, flash floods, landslides and debris flows

The first part explains operational aspects of algorithms/tools that produce flood forecasts, which is followed by operationalizability of algorithms/tools that generate forecasts on landslides and debris flows.

Flash flood hazard and impact assessment algorithm (FF-EWS) (UPC)

a) Development stage (TRL) of the tools/algorithms/products
The Flash flood hazard and impact assessment algorithm (FF-EWS) has proven in an operational environment (TRL=9). The current version of the algorithm has been running to generate some of the products of the flash flood layer of EFAS, and several configurations (one at Continental scale, and four for the pilot sites) have been implemented in the ANYWHERE MH-EWS.

b) What purposes the tools/algorithms/products have been used
The algorithm nowcasts the flash flood hazard level based on radar-based rainfall nowcasts. The tool has been used to follow situations of heavy rainfall and their evolution in the near future (few hours).

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The implementation of the algorithm requires processing a Digital Elevation Model to retrieve the flow directions and the drainage network. In the context of ANYWHERE , we have worked with a resolution of 1 km for the European FF-EWS product, and 50-200 m for the setups done in 4 Pilot Sites (Canton Bern, Corsica, Liguria and Catalonia). The algorithm inputs are the rainfall forecasts generated with the UPC-CRAHI algorithm for precipitation nowcasting.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviation.

Landslides and debris flows hazard and impact assessment (UPC)

a) Development stage (TRL) of the tools/algorithms/products
Within the ANYWHERE project, the algorithm for landslides and debris flows hazard and impact assessment has been validated in relevant environment (TRL=5), and the algorithm has just started running in real time as part; i.e. being demonstrated in relevant environment (TRL=6).

b) What purposes the tools/algorithms/products have been used
The algorithm provides a warning of the areas that potentially affected by landslides. The outputs are provided at very high resolution (30 m at regional

scale), and aggregated over polygons to facilitate the identification of the most affected areas.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
Implementation of the algorithm requires the use of an existing landslides susceptibility map, or the production of one [Palau et al. (2020) used a simple methodology for Catalonia using datasets available everywhere in Europe].
The rainfall inputs to the algorithm come from high-resolution weather radar rainfall estimates and nowcasts.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016).
The development of the algorithm has been limited to the Pilot Site of Catalonia. Extensive work has been done to retrieve a landslide susceptibility map for Catalonia. Also, some works have been done to test the transferability of the algorithm to a new domain (in particular, to Canton Berne) using standard datasets available with European coverage.

6.3.2 Storm surges

First operational aspects of using the Storm Surge Model at the pan-European scale are described followed by the model at the regional scale. The section concludes with the Inundation and Erosion Model.

European Storm Surge model (ESS) (CFR)

a) Development stage (TRL) of the tools/algorithms/products
The European Storm Surge model (ESS) has been extensively validated (Fernández-Montblanc et al., 2019; Fernández-Montblanc et al, in press) and demonstrated in the context of the ANYWHERE platforms, so TRL 6 – Technology demonstrated in relevant environment – has been assigned. Figure 62 shows an example of a recent storm surge level forecast during the Hurricane Lorenzo.

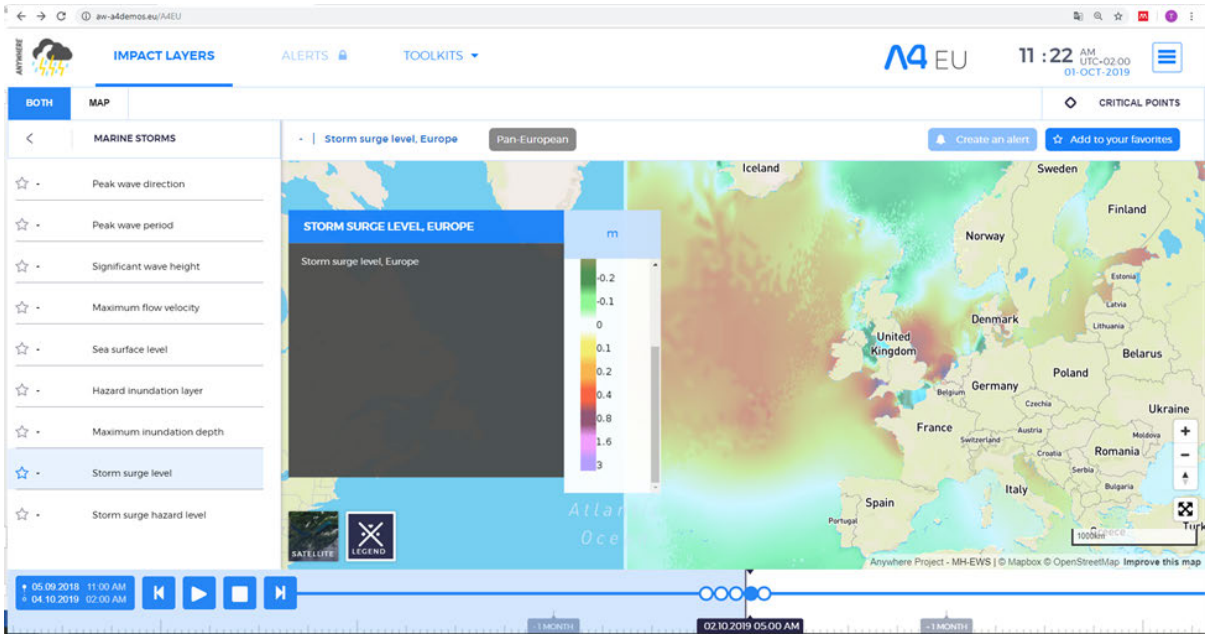


Figure 62: Example of the storm surge level product during the Hurricane Lorenzo showed in A4Demos for 2 October 2019, forecasted on 1 October 2019.

b) What purposes the tools/algorithms/products have been used
The output of the European storm surge model (PRD100-104) has been used to provide 72 hours horizon forecast of the extension, magnitude and duration of the hazardous conditions derived from marine storms at pan-European scale. The algorithm provides not only storm surge or total water level related products, but also the wave characteristics (magnitude, period and direction).They have been included in the ANYWHERE platform (A4Demos and A4Cat).

c) What is needed (data/resources) to implement it (at local and pan-European scale)
As a pan-European-oriented algorithm, the storm surge European model was widely tested and implemented. Further action will be required to maintain the system beyond the ANYWEHRE project and to integrate the outputs in platforms such as the Copernicus Emergency Management Service.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviation for the final algorithms of the European Storm Surge model. Additionally, a second version of the algorithm (~2 km resolution) it is already under validation to enhance the results observed during the demonstration period.

Regional Storm Surge model (CFR)

a) Development stage (TRL) of the tools/algorithms/products

The Regional Storm Surge model has been validated with ground truth data in Rogaland (Norway) and demonstrated in the context of **ANYWHERE** platform, so TRL 6 – Technology demonstrated in relevant environment.

b) What purposes the tools/algorithms/products have been used
The output of the Regional Storm Surge Model (PRD107-110) has been used to provide 72 hours horizon forecast of the extension, magnitude and duration of the hazardous conditions generated by marine storms (storm surges and waves hazards) at Rogaland (Norway). The algorithm provides a high resolution forecast of sea surface level, significant wave height, peak wave period and peak wave direction. Figure 63 provides an example. The product has been included in the **ANYWHERE** platforms A4Demos and provided temporal and spatial information of the marine storm derived hazard to the end user and the Emergency Manager in Rogaland.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The Regional Storm Surge model has been tested and implemented at regional scale in Rogaland Norway. Further action will be required to maintain the system beyond the duration of ANYWEHRE project and to integrate the algorithm outputs in the local natural hazard emergency system.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of **ANYWHERE** , June 2016)
No deviation are observed from the originally promised.

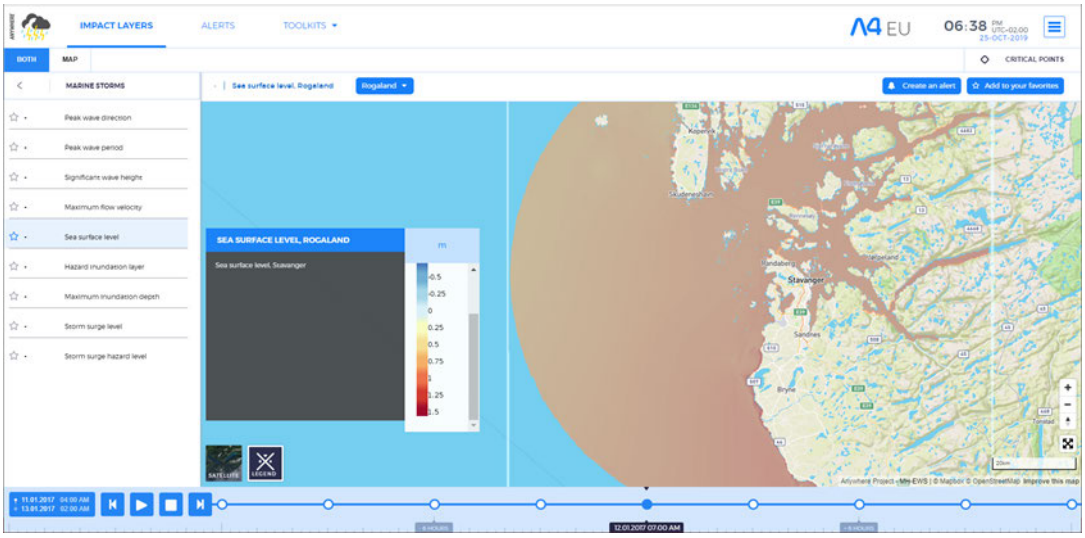


Figure 63: Example of the A4Demos platform showing the sea surface level PRD-107 during an extreme event occurred on 12 January 2017 at Rogaland (Norway).

Inundation and erosion model (CFR)

a) Development stage (TRL) of the tools/algorithms/products
The Inundation and erosion model has been validated and demonstrated in the context of ANYWHERE platform, so TRL 6 – Technology demonstrated in relevant environment.

b) What purposes the tools/algorithms/products have been used
The output of the Local Inundation and Erosion model (PRD112-115) has been used to provide forecast of the extension and magnitude of the coastal flooding in the City of Stavanger (Norway). The algorithm provides high resolution maps of maximum flow velocity, maximum inundation depth and hazard inundation in the City of Stavanger (Fig. 64). The product has been included in the ANYWHERE platforms A4Demos and provided temporal and spatial information to the end user and the Emergency Manager in Stavanger.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
Further needed investment is not requested to implement the system at local level since it has been tested and implemented at local scale in city of Stavanger (Norway). Further action will be required to maintain the system beyond the duration of the ANYWHERE project and to integrate the algorithm outputs in the local natural hazard emergency system.

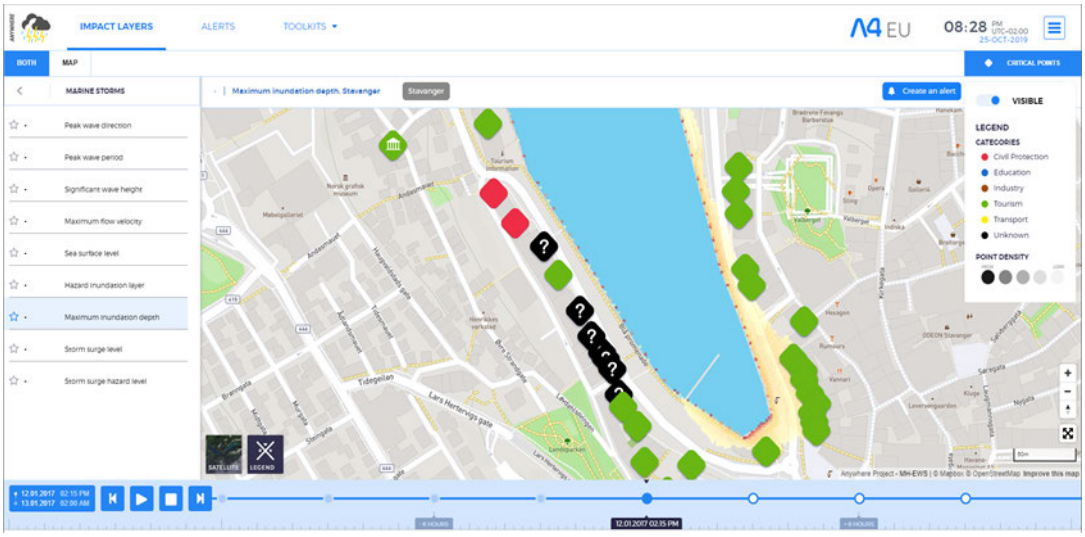


Figure 64: Example of the A4Demos platform showing the maximum inundation depth (PRD-113) in combination with the critical points located in the city center of Stavanger and potentially flooded during the coastal flood occurred on 12th January 2017 Stavanger (Norway).

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviations are observed from the originally promised.

6.3.3 Heatwaves and air quality (weather-induced health)

Operational aspects of using algorithms/tools to forecast heatwaves and air quality products is explained. These products are associated with weather-related health.

Universal Thermal Climate Index (UTCI) (UOR)

a) Development stage (TRL) of the tools/algorithms/products
The development of the UTCI algorithm is at stage 6 of the TRL scale. The algorithm, which delivers UTCI forecasts with a 10-day forecast horizon, is not yet operational within a numerical weather prediction model framework. However, the algorithm has been validated and demonstrated within the ANYWHERE platforms which represent a relevant environment for the purpose.

b) What purposes the tools/algorithms/products have been used
The output of the UTCI algorithm (PDR-117) has been used for the prediction of thermal-related human discomfort during the summer season. Forecasts of the UTCI have been provided to and included in ANYWHERE platforms both at the European and pilot site level.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The UTCI algorithm is already implemented both at the local and the pan-European scale. No further action is needed.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviations.

Regional Air Quality (RAQ) (UOR)

a) Development stage (TRL) of the tools/algorithms/products
The development of the RAQ algorithm is at stage 9 of the TRL scale. The RAQ algorithm, which has been included in the ANYWHERE platforms, corresponds to the regional air quality forecasting system developed by the Copernicus Atmosphere Monitoring Service (CAMS, see D.2.1 Van Lanen et al., 2017, and D2.3 Ciavola et al., 2017). The regional air quality forecasting system is therefore an actual system already proven in operational environment.

b) What purposes the tools/algorithms/products have been used
The RAQ algorithm provides air quality forecasts at the pan-European scale, specifically the predicted concentrations (4-day horizon) of major air pollutants at the surface level (PRD-118 to PRD-123). These are ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide and particulate matters below 10 and 2.5 microns.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
Regional Air Quality at the pan-European scale is obtained from CAMS.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviations.

6.3.4 Weather-induced fires (European Fire Forecasting System and Global ECMWF Fire Forecasting model, (EFFIS-GEFF)

a) Development stage (TRL) of the tools/algorithms/products
The GEFF algorithm runs operationally at ECMWF since 2018, supporting the European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF). The algorithm is open source (https://git.ecmwf.int/projects/CEMSF/repos/geff/browse) and its outputs have been thoroughly documented and validated (Vitolo et al. 2019; Di Giuseppe et al. 2016). As such, the development of the GEFF algorithm is at stage 9 of the TRL scale (has proven in an operational environment).

b) What purposes the tools/algorithms/products have been used
The GEFF algorithm provides fire danger indices (e.g. the Canadian Fire Weather Index) in the form of: (i) 10-day deterministic high resolution forecasts, also called HRES; (ii) 15-day ensemble forecasts, called ENS and (iii) 40 years of renalysis based on ERA-Interim and ERA-5. The full list of data products is available in D3.1 (Smith et al., 2017), from which the HRES version of PRD-124 to PRD-129 were selected by users to be visualised in A4EU and local apps.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The algorithm is already implemented and runs at global scale, incl. Europe. Users can get access to EFFIS-GEFF output via the Copernicus Emergency Management Service[16]. EFFIS provides standard danger classes on its website (https://effis.jrc.ec.europa.eu/about-effis/technical-background/fire-danger-forecast/). For this reason, the interpretation of the FWI values can be considered standardised. First responders using the FWI forecasts are expected to integrate this information with local knowledge of the topography and vegetation.
As EFFIS outputs have a resolution of 9Km (HRES) or coarser (ENS and ERA-based reanalysis), the use of these layers is limited to national and regional levels. For more local-scale studies we suggest making use of RISICO (developed by CIMA foundation).

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE, June 2016)
No deviations.

[16] https://emergency.copernicus.eu/mapping/ems/early-warning-systems-efas-and-effis.

6.3.5 Droughts

Development stage is reported for the standardized drought indices, threshold-based indices and areal drought indices.

Standardized Indices (WUR)

a) Development stage (TRL) of the tools/algorithms/products

The ANYWHERE drought early warning system using the standardized indices is already in the pre-operational mode since 2018 (Fig 65) and has been tested in the pilot site, Catalonia, since September 2017. Hence, the development stage is classified TRL=8 (System complete and qualified). The highest TRL level scale (actual system proven in operational environment), could be achieved if we keep the ANYWHERE MH-EWS operational, e.g. via the ANYWHERE Foundation.

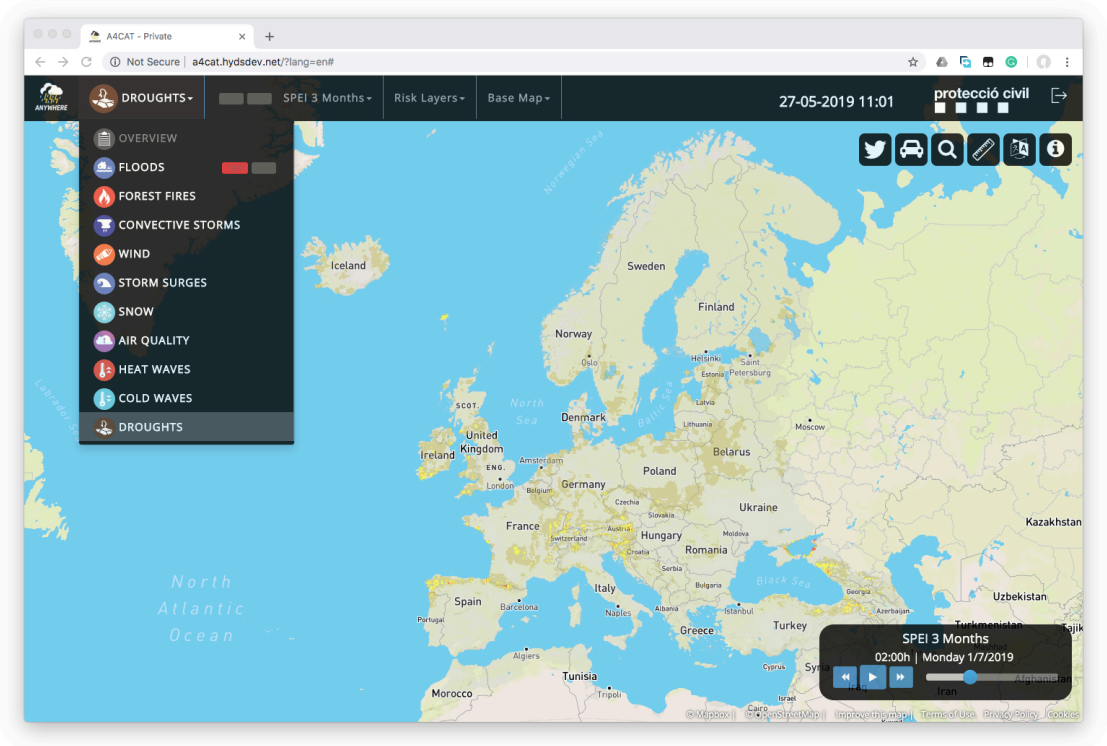


Figure 65: The Standardized Precipitation Evaporation Index (SPEI) for July 2019 forecasted in May 2019 (lead-time: 3-month) for the pan-European region, as shown by the MH-EWS platform. The left panel shows the hazards that are forecasted by the MH-EWS, such as flood, fires, storm surge, heatwaves, droughts.

b) What purposes the tools/algorithms/products have been used

The drought forecasting algorithms was encapsulated in the MH-EWS since September 2017. This system has been tested to predict the 2018 pan-European droughts. All drought forecasting products using different standardized indices produce similar drought pattern in Europe, e.g. the extreme drought in northern and central Europe in July 2018 has been forecasted 3 months ahead (Sutanto et al., 2019a).
We also tested the drought products at the pilot site scale in Catalonia (Spain) together with the Catalonian Water Agency (ACA). The forecasts of meteorological drought (the SPI-12) were compared with the SPI maps based on observed precipitation from the Catalonia Meteorological Agency (MeteoCat). For example, cases for dry conditions (February 2018) and wet conditions (December 2018) showed that the forecasts issued 3 months before are in agreement with the observed SPI-12.
Additionally, we are working closely with ACA and HYDS to forecast reservoir volumes at Llobregat River and Ter River by using the SPI-3 (see also D6.5). This work is still ongoing and an example of using the forecasted SPI-3 to predict the Llobregat reservoir volume is given in Figure 66. It shows that the reservoir volume can be predicted well in advance for short lead times. For longer lead-time the algorithm still requires improvement.

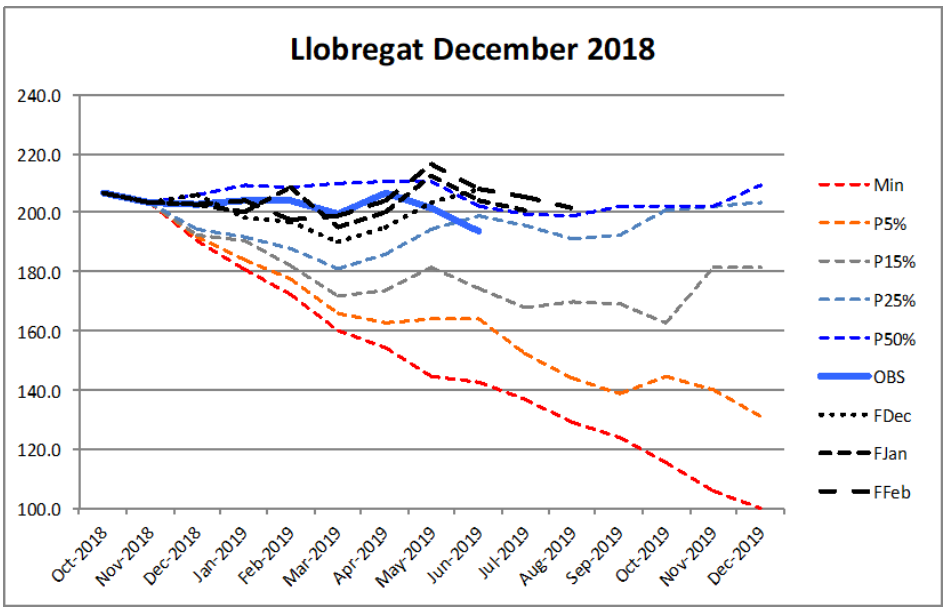


Figure 66: Forecasting the Llobregat reservoir volume using SPI-3 in December 2018 (dotted line), in January 2019 (dashed line), and in February 2019 (long dashed line) with a lead-time of 7 months ahead.

c) <u>What is needed (data/resources) to implement it (at local and pan-European scale)</u>
Drought forecast products (standardized drought indices) are implemented at the pan-European scale. The most important factors now are the continuity of

the data (ECMWF-SEAS5, EFAS forecasts, and proxy), no updates of hydrometeorological platforms that changes, for example new spatial coverage. Forecasted standardized drought indices are also available at the local scale down to 5 km scale. Not many users have sufficient knowledge on the drought forecasts, such ACA. Thus, training on drought forecasts is necessary for a new user to support use of the drought products and to finetune this to the needs.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)

There is no deviation of final algorithms that forecast standardized drought indices relative to what has been promised. On the other hand, the final version has a wider European spatial coverage than was planned at the start of ANYWHERE .

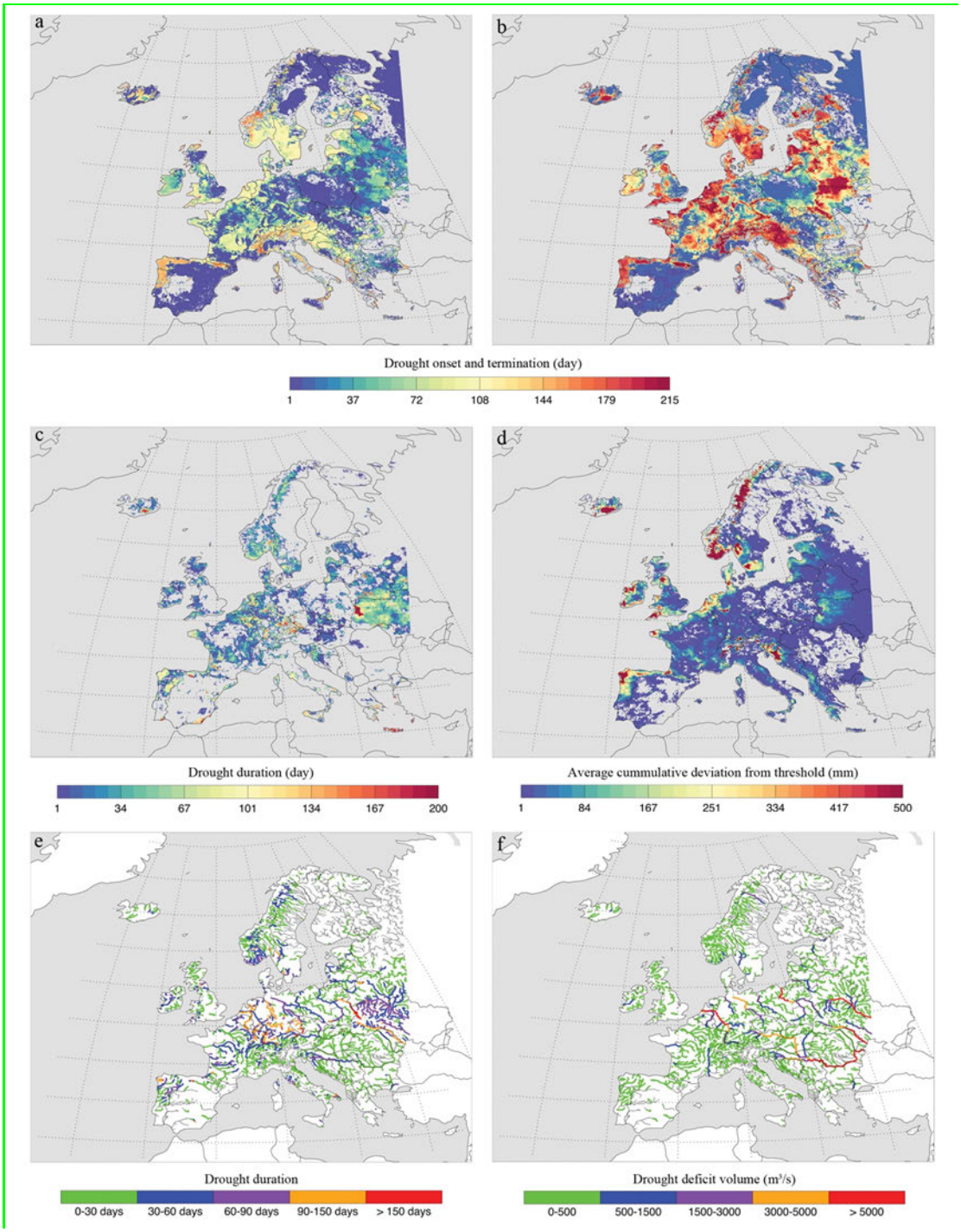


Figure 67: Examples of drought forecast products for different hydrological variables. a) Drought onset for the longest soil moisture drought event, b) drought termination for the longest soil moisture drought event, c) drought duration in the runoff, d) average cumulative deviation from the threshold in groundwater, e) drought duration in discharge, and f) cumulative drought deficit volume in discharge. All data hold for the median of the 51 ensemble members within the forecast period of 7 months and obtained from the forecast from 2nd May 2018.

Threshold-based drought indices (WUR)

In general, the algorithms that forecast the threshold-based drought indices through the ANYWHERE MH-EWS had the same development level as the standardized indices (see above).

a) Development stage (TRL) of the tools/algorithms/products
The drought forecasting algorithms using the threshold-based indices have been tested to predict the 2018 pan-European droughts. The onset and termination of the longest drought (day number) within the forecast period (7 months), the total drought duration (in days rather than in months as for the standardized indices), the deficit volume (for fluxes), and average cumulative deviation from threshold (for state variables) on daily basis for 7 months are forecasted for precipitation, soil moisture, runoff, and groundwater (e.g. Fig. 67). The extreme drought in northern and central Europe in 2018 has been forecasted 3 months ahead (Sutanto et al. 2019a). The drought algorithms using the threshold-based drought indices have been tested at pan-European scale and are in pre-operational state. The test results demonstrate that the system could forecast droughts relatively well months in advance. Similar to the standardized drought indices, the development stage could be classified as TRL=8 (System complete and qualified).

b) What purposes the tools/algorithms/products have been used
The drought forecasting algorithms that forecast threshold-based indices have been tested (see point a). At the pilot site scale we tested operationalizability in Catalonia in close cooperation with ACA and HYDS. Skill in hydrological forecasting is presented by Van Hateren et al. (2019).

c) What is needed (data/resources) to implement it (at local and pan-European scale)
Drought forecast products (threshold-based drought indices) need the same data and resources at the pan-European scale and the local scale, as described under standardized drought indices.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No deviations were reported. Threshold-based drought indices are forecasted for a wider European spatial coverage than was planned at the start of ANYWHERE .

Areal drought indices (WUR)

Area drought indices have not been tested in ANYWHERE by end-users. However, developers have verified at the pan-European scale.

a) Development stage (TRL) of the tools/algorithms/products

The areal drought forecasting products have been tested to predict the 2018 pan-European droughts, which are time series of the average percentage area in drought for each month over the 7-month forecast period (Sutanto et al., 2019a). The development stage of the areal drought indices is not as high as the standardized and threshold-based drought indices, and it is classified TRL=5 (Technology validated in relevant environment).

b) What purposes the tools/algorithms/products have been used
The algorithms that forecast areal drought indices have been used to calculate the percentage area in drought and the statistical summary information of the drought duration and deficit volume in precipitation of the area in drought across Europe.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
Same as point c for the standardized and threshold based indices, see above.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)

e) No deviations were reported. Areal drought indices are forecasted for a larger European domain than was planned at the start of ANYWHERE .

6.3.6 Convective storms, severe winds and heavy snowfall

Continuous development work is performed on all FMI products presented in this section. New versions of the products are regularly produced.

Detection and forecasting convective cells (FMI)[17]

a) Development stage (TRL) of the tools/algorithms/products
The object-oriented convective storm product is still very early stage of development, and it could be characterized with TRL 3 (Experimental proof of concept).

Snow-load and gust algorithms (FMI)

a) Development stage (TRL) of the tools/algorithms/products
The Finnish Meteorological Institute (FMI) provides forecasts of snow load accumulation on canopy and transmission lines based on an experimental model taking as input the NWP data of relative humidity, temperature, wind speed and precipitation estimate. The model is calculated in the FMI operational production line, therefore is interpret to have reached the development stage of TRL 9 (system proven in operational environment). The validation of the snow

[17] Point b) What purposes the tools/algorithms/products have been used, point c) What is needed to implement it, and point d) Deviation of final tool/algorithm relative to originally promised (DoA) are not reported because algorithm development is still in early phase.

load is continuously performed by the forecasting meteorologists and based on this analysis the products are developed further.

b) What purposes the tools/algorithms/products have been used
It is used by the forecasting meteorologists to estimate the possible impacts of snow load for electricity distribution and the building roofs. It is provided to the end-users upon request through FMI-Ilmanet service and for A4EU platform. In ANYWHERE this product is used for the A4FINN impact tool as one of the parameters that create the weather-related combined impact level for the ISTIKE pilot site.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The snowfall product is developed to Finnish conditions, and as it is based upon experimental parametrization, therefore the feasibility to other location is limited and would need scientific resources to calibrate parameters.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE , June 2016)
No changes relative to the DoA.

Probability of precipitation type (FMI)

A first precipitation type product is provided by FMI. Operational aspects are given here.

a) Development stage (TRL) of the tools/algorithms/products
The probability of precipitation type product is in the FMI operational production line, therefore the TRL level is 9 system proven in operational environment). As stated in Section 2.1.2 the NWP products provided by FMI are based on already operational models of HIRLAM, HARMONIE-AROME and the product of precipitation type utilizes also the synoptic scale from GFS and ECMWF.

b) What purposes the tools/algorithms/products have been used
The product is originally developed as part of the SESAR TopLink project targeting for aviation applications. Currently it is used by the forecasters as a tool to estimate the precipitation type, in research it has been applied with radar data to give precipitation type estimate to regions, which do not have dual-polarization radar data coverage. It is provided to the end-users upon request through FMI-Ilmanet service and for A4EU platform. In ANYWHERE this product is used for the A4FINN impact tool as one of the parameters that create the weather-related combined impact level for the ISTIKE pilot site.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
It is a pan-European product.

d) deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE, June 2016)
This product was developed after the DoA was written, and has been added to the ANYWHERE-catalogue, when it was needed in the A4FINN development.

Probability of precipitation type (ECMWF)

A second precipitation type product is provided by ECMWF. Operational aspects are provided below.

a) Development stage (TRL) of the tools/algorithms/products
The two new ECMWF precipitation type products that have been developed in the framework of ANYWHERE project have been operational since end of 2017. These products are closely connected to the ECMWF-IFS and recalibration is part of the update of ECMWF-IFS (Section 4.5.3). The development stage is classed as TRL=9 (Actual system proven in operational environment).

b) What purposes the tools/algorithms/products have been used
Both the probability of precipitation type and the most probable precipitation have been forecasted at the pan-European scale.

c) What is needed (data/resources) to implement it (at local and pan-European scale)
The algorithm to forecast precipitation type is run by the ECMWF and needs input from ECMWF-IFS. Specialized knowledge and experience is required run the algorithm. Downscaling of output is needed for local applications.

d) Deviation of final tool/algorithm relative to originally promised (DoA at start of ANYWHERE, June 2016)
Precipitation type products have not suffered from any change since they started to be operational in 2017. A calibration process is developed to adjust the minimum precipitation rate thresholds considered for each precipitation type, however the values obtained from this calibration had not changed until now.

7 Moving from hazard to impacts

Work Package 2 (WP2) aimed to establish algorithms/tools to nowcast and forecast natural hazards and the resulting impacts. This requires a translation from hazards into impacts. In the conventional approach, knowledge on exposure and vulnerability is needed for the translation. This knowledge is very specific, and commonly people think that appropriate impact assessments only can be done at local scale where detailed small-scale information is available. Local impact assessment is done in close cooperation between end-users, who know and provide relevant information on exposure and vulnerability, and forecasters of natural hazards. This does not mean that there are no large-scale assessments of potential impacts (e.g. Carrão et al., 2017), however, these are very explorative. A comprehensive overview of how hazards are to be translated into impacts, incl. how hazards, exposure, vulnerability, resilience, adaptation risk and impacts are connected, is provided by the MOVE framework (Methods for the Improvement of Vulnerability Assessment in Europe, Birkmann et al, 2013). In an earlier ANYWHERE deliverable, Ciavola et al. (2017) concluded that adequate information on exposure and vulnerability is not available at the pan-European scale, that is, the WP2 scale. They state that within the ANYWHERE framework the forecasts generated with the WP2 algorithms/tools via the MH-EWS (WP3) need to be linked with local knowledge in the Pilot Sites (WP6). Ciavola et al. (2017) provide for each hazard a number of foreseen applications of the algorithms to assess impacts. Nonetheless, we address in this chapter some recent developments within ANYWHERE that elaborated impact assessment/forecasting at scales beyond the Pilot Sites, i.e. the pan-European scale and regional scale.

First, this chapter presents a summary of existing datasets with European coverage that can be used to characterize the vulnerability and exposure. In the context of ANYWHERE , the interest of these datasets is double, in their use: (i) in impact models run at Europe scale; and (ii) as standard datasets that can be used in impact models, especially in areas where no higher-resolution or higher-accuracy datasets are available. Then, the chapter presents a few examples that apply the conventional approach by overlaying hazard, exposure and vulnerability (bottom-up approach). Next we present an alternative approach that starts with the impacts (top-down approach). We conclude with impact assessment at the regional scale.

7.1 Pan-European vulnerability information

To demonstrate the European perspective of the impact forecasting algorithms/tools of the ANYWHERE project, several vulnerability datasets with pan-European coverage have been gathered from various sources. In combination with hazard forecasts, this information facilitates the estimation of socio-economic impacts of weather-induced events across Europe. The datasets can directly be used for the development of pan-European impact forecasts, or for transferring regional impact models to new spatial domains (where regional vulnerability data is not available).

Besides of their use for impact forecasting, it has been found valuable to make these datasets available as vulnerability layers in the A4EU platform, which allows displaying together with weather and hazard forecasts.

The datasets have been integrated in ANYWHERE in collaboration with the EU Joint Research Centre (mainly during a secondment of a researcher from UPC-CRAHI at JRC). Some of them have been generated from existing datasets available at JRC, and information from OpenStreetMaps has also been integrated.

7.1.1 Population density maps

For estimating the number of people potentially affected by a weather event, it can be useful to overlay the forecasts of the hazard with the population density map in the affected area. The project has identified two sources with European coverage:

- The Global Human Settlement Layer (GHSL; Freire et al., 2016) is a static population density map showing the number of people per hectare with a resolution of 100 m. The map was created based on the Eurostat 2011 censuses, Corine Land Cover 2006 (refined), and the European Settlement Map 2016. The coverage includes all EU countries, plus Albania, Andorra, Bosnia and Herzegovina, Iceland, Kosovo, Liechtenstein, North Macedonia, Monaco, Montenegro, Norway, San Marino, Serbia, and Switzerland (Fig. 68).
- The population density maps recently developed in the framework of the H2020 ENACT project[18]. These maps have a lower resolution (1 km) than the GHSL static population density map described above, but they account for day/night dynamics (e.g. due to commuting) and monthly variations (e.g. due to tourism). For each month of the year, two population density maps are provided (one for daytime and one for night-time; i.e. 24 maps in total). The maps are not yet officially published and will meanwhile not be displayed on the A4EU platform. However, we have received permission to use them for the development of impact models within the ANYWHERE project.

[18] https://ghsl.jrc.ec.europa.eu/enact.php.

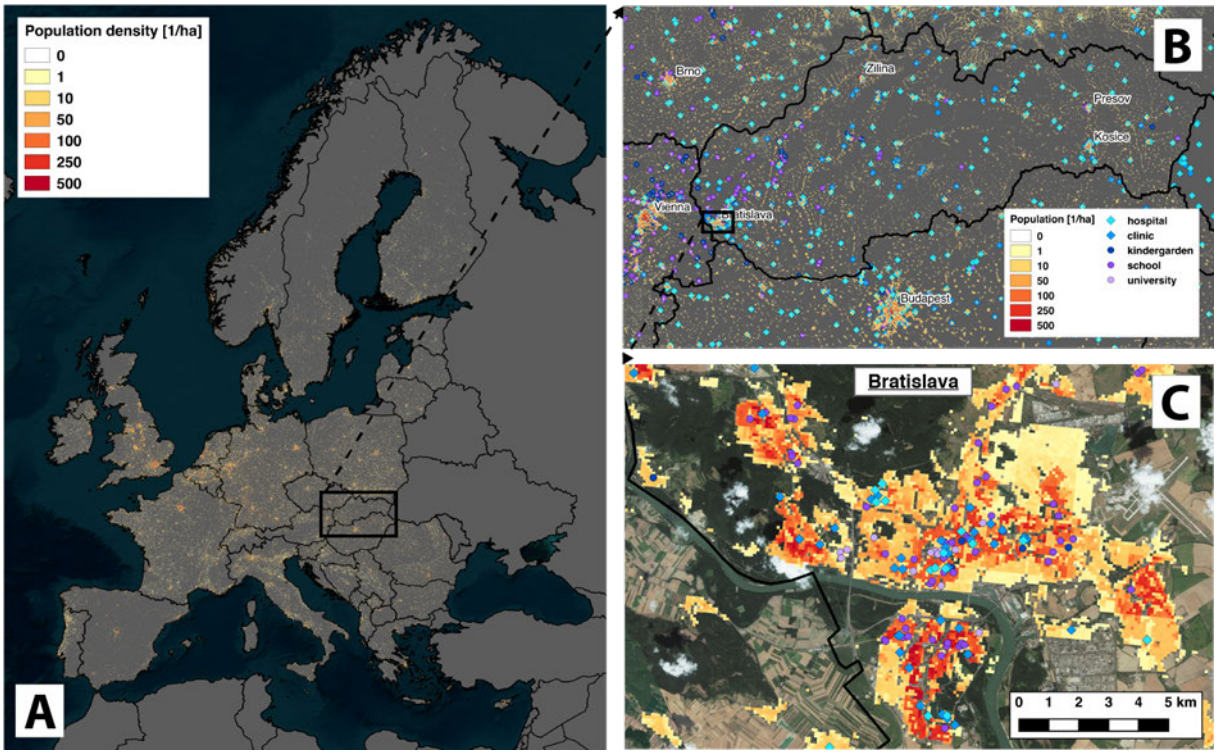


Figure 68: a) GHSL static population density map over Europe (spatial resolution: 100 m). b) Detail of the GHSL population density map over Slovakia; the symbols correspond to the Pan-European critical infrastructure database. c) Same as panel b, but zoomed to Bratislava (Slovakia).

7.1.2 Population exposure to specific hazards

For a refined perspective of the population potentially at risk, the Joint Research Centre (JRC) estimated the number of people exposed to specific weather-related hazards, which is displayed in the Disaster Risk Management Knowledge Centre Risk Data Hub[19]. This information was generated by combining the GHSL static population density map (with a spatial resolution 100 m; see Section 7.1.1) with the areas potentially affected by the individual hazards. The resulting estimates of exposed population are available at the resolution of administrative regions (NUTS), covering all EU countries, plus Norway and Switzerland (see Fig. 69 for an example). The resulting layers are available for river floods, coastal floods, landslides, and forest fires.

River floods: The population numbers potentially affected by river floods were estimated by overlaying the population density map with the JRC EFAS Flood hazards Map[20] of return periods of 10, 50, 100, and 500 years, covering all European rivers

[19] https://drmkc.jrc.ec.europa.eu/risk-data-hub/risks/data_extraction/.

[20] https://data.europa.eu/euodp/en/data/dataset/7dcb3d9dd9b598f2905611d5e22c5ebe69505278.

with an upstream drainage area larger than 500 km^2. One limitation is that defence structures such as dykes were not taken into account.

Coastal floods: Analogously to the method applied for river floods, the population numbers potentially affected by coastal flooding were estimated by overlaying the population density map with coastal flood hazard maps (derived with the LISCoAsT model; Vousdoukas et al., 2017[21]) corresponding to return periods of 10, 50, 100, and 500 years. Also here, defence structures were not considered, resulting in high numbers of potentially exposed population also in locations with high protection standards (e.g. The Netherlands).

Landslides: The population numbers potentially affected by landslides were estimated by combining the ELSUS landslide susceptibility layer (Wilde et al., 2018[22] with daily rainfall maps of different return periods (2, 5, 10, 20, 50, 100, 200, 500 years) from the Global Precipitation Climatology Centre (GPCC), and then overlaying the resulting landslide footprints with the population density map.

Forest fires: The estimates of population potentially affected by forest fires (Fig. 69) are based on three steps: First, the Woodland-Urban-Interfaces (WUI) were derived from Corine Land Cover 2006. Second, the WUI were filtered for areas within a Euclidian distance of less than 10 km from the burnt areas recorded by the European Forest Fire Information System (EFFIS) between 2006-2017, to identify the WUI with high potential for forest fires. Third, the resulting WUI with high fire potential have been overlaid with the population density map.

[21] https://data.jrc.ec.europa.eu/dataset/jrc-liscoast-10009.

[22] https://esdac.jrc.ec.europa.eu/content/european-landslide-susceptibility-map-elsus-v2.

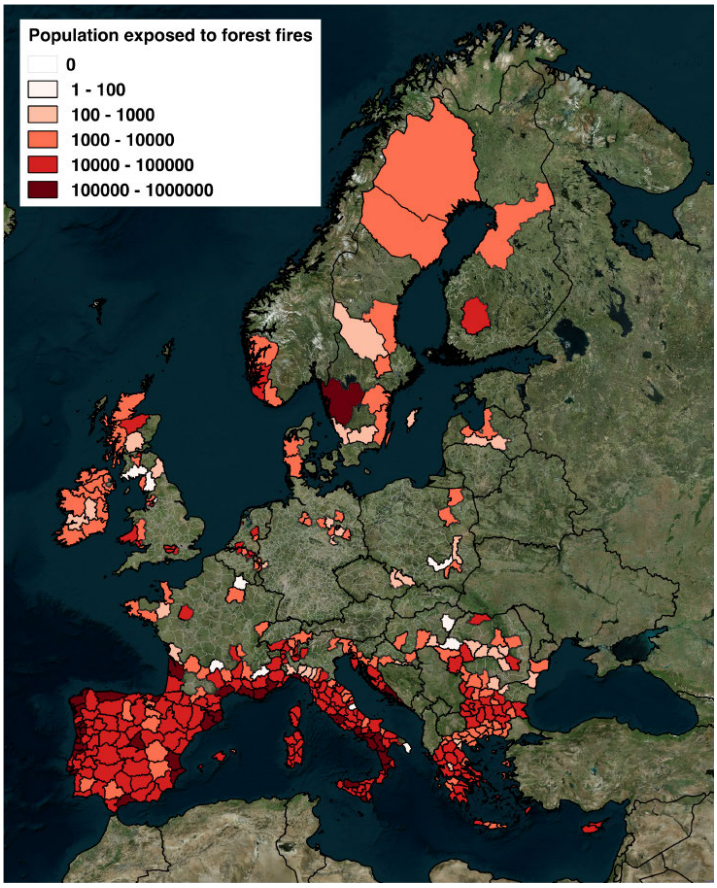


Figure 69: Population potentially exposed to forest fires, aggregated at the level of NUTS regions.

7.1.3 Critical infrastructures

As additional vulnerability information, the locations of critical infrastructures have been collected and harmonised at pan-European scale. The resulting layers include education and health facilities (Fig. 68b and 68c), and the primary road network.

Education facilities: A database containing the locations of education facilities across Europe has been extracted from a OpenStreetMaps dataset[23]. The database includes kindergartens, schools, and universities. While the coverage is very high in some countries (e.g. Germany), the data is less comprehensive in others (e.g. in Slovakia, see Fig. 68b).

Health facilities: A pan-European database of hospitals has been generated based on data from Healthsites.io[24], an open-source mapping project of health facilities at global scale that is based on OpenStreetMaps and public participation. The data have been filtered for the hospitals and clinics to include only the most important facilities.

[23] https://mapcruzin.com/free-europe-arcgis-maps-shapefiles.htm.

[24] https://healthsites.io/.

In a few random test locations, the data appeared to be very comprehensive (we have not found any missing hospitals). For Bosnia and Herzegovina, Czech Republic, Faroe Islands and Northern Cyprus, the Healthsite.io data was not available and has been filled with data extracted from the same OpenStreetMaps dataset used for the education facilities.

Road network: The primary road network across Europe has been extracted from OpenStreetMaps data[25]. The dataset has been filtered for the most important road types of the OpenStreetMaps classification: motorways, trunks, primary, secondary, and tertiary roads. The value of the road network dataset consists in its potential use in impact models (e.g. to automatically identify the roads affected by forecasted trajectories of forest fires), but it does not add much to the base maps from OpenStreetMaps already available in the A4EU platform, and, it will thus not be displayed.

7.2 Impact forecasting at the European scale

This section described examples of continental impact forecasting. Two approaches are explained, i.e. (i) the bottom-up approach that uses the natural hazard forecast maps and the pan-European vulnerability information (Section 7.1), and (ii) the top-down approach that starts with the impacts.

7.2.1 Bottom-up approach – Impact models

To illustrate how the described vulnerability data (Section 7.1) can be employed to model the impacts of weather events at pan-European scale, a first exploratory study has been carried out following a bottom-up approach (i.e. combining hazard forecasts form the ANYWHERE catalogue with the vulnerability and exposure datasets to map the impacts at European scale.

7.2.1.1 Air Quality impact on population

The World Health Organization stated already many years ago the potential of combining air quality estimates with population density information to map the impacts on the health of the population (WHO, 1999). Following this idea, we have explored the possibility to forecast the impact of air quality on the population across Europe to illustrate the use of pan-European vulnerability datasets for simplified impact forecasting.

The proposed approach uses the forecasts of particle concentrations [$\mu g/m^3$] of five individual pollutants that are available in the ANYWHERE MH-EWS (originally provided by Copernicus[26]). To make these forecasts more interpretable for the end-users, the

[25] https://mapcruzin.com/free-europe-arcgis-maps-shapefiles.htm.

[26] http://www.regional.atmosphere.copernicus.eu.

European Air Quality Index (EAQI[27]) of the European Environment Agency has been employed. The EAQI defines concentration thresholds for each of the five pollutants to classify them in five air quality index levels (Table 11). In each location, the poorest index level of any of the five pollutants determines the overall air quality. An example of the resulting pan-European air quality forecast is shown in Figure 70.

Table 11: Particle concentration thresholds for air quality pollutants (Source: European Environment Agency)

Pollutant	Index level (based on pollutant concentrations in $\mu g/m^3$)				
	Good	**Fair**	**Moderate**	**Poor**	**Very poor**
Particles less than 2.5 μm ($PM_{2.5}$)	0-10	10-20	20-25	25-50	> 50
Particles less than 10 μm (PM_{10})	0-20	20-35	35-50	50-100	> 100
Nitrogen dioxide (NO_2)	0-40	40-100	100-200	200-400	> 400
Ozone (O_3)	0-80	80-120	120-180	180-240	> 240
Sulphur dioxide (SO_2)	0-100	100-200	200-350	350-500	> 500

While poor air quality may have only limited consequences in areas with low population densities, the impacts can become very significant when more densely populated areas are affected. To take this consideration into account and estimate the air quality impact on population, the proposed approach combines the EAQI forecast with population density maps. For this task, we have selected the population density maps developed in the ENACT project (Section 7.1.1), which account for some of the dynamics that inherently affect population exposure. At each forecast time step, the population density map of the corresponding month and time of day is used.

[27] https://www.eea.europa.eu/themes/air/air-quality-index/index.

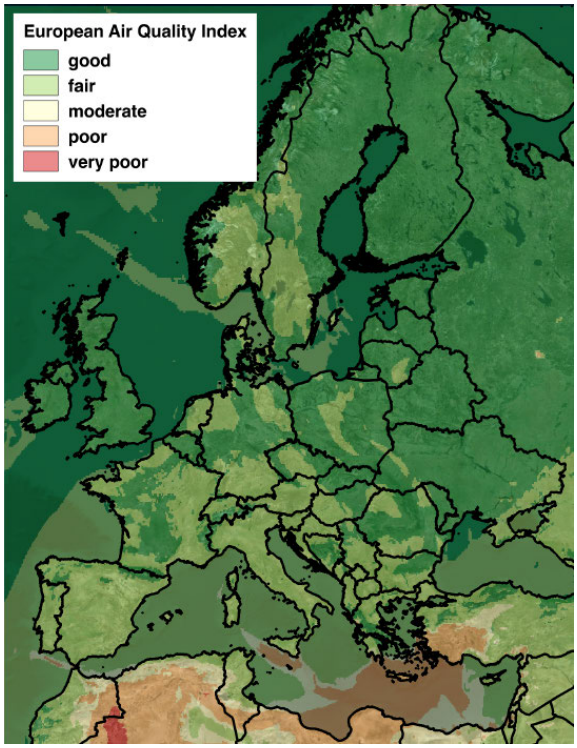


Figure 70: The European Air Quality Index (EAQI) forecast for 11 July 2019 at 10:00 UTC. This product has been generated by selecting the poorest index level of any of the five pollutants of Table 11 in each location.

The presented approach is based on an impact matrix combining the hazard as described by the EAQI forecasts (Fig. 70) and the vulnerability estimated by the population density. With this aim, the population density values have been classified into three population exposure classes: low (10 – 999 people/km^2), medium (1000 – 4999 people/km^2), and high (5000 and more people/km^2). These thresholds are arbitrary chosen and were manually calibrated by testing a variety of threshold values and analysing the resulting population exposure maps in various locations across Europe. The combination of EAQI and population density is done in each cell (with a resolution of 1 km) by means of the impact matrix of Figure 71, resulting in three impact levels: low (yellow), medium (orange) and high (red). It has been assumed that only moderate, poor, and very poor air quality situations cause relevant impacts; in locations with good or fair air quality, no significant impacts are expected. An example of the resulting impact map, zoomed to the Peloponnese (Greece), is shown in Figure 72.

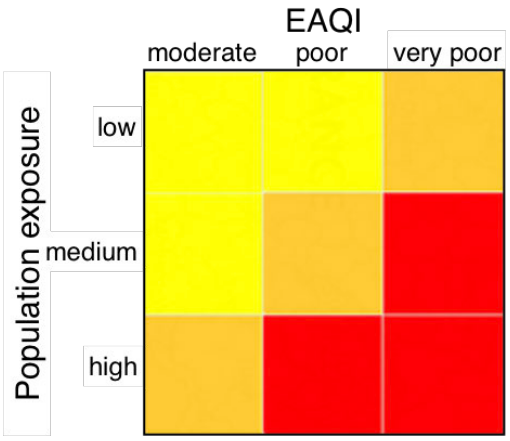


Figure 71: The Impact matrix that estimates the air quality impact on population by combining in each cell the European Air Quality Index (EAQI, Fig. 70) with the population exposure class.

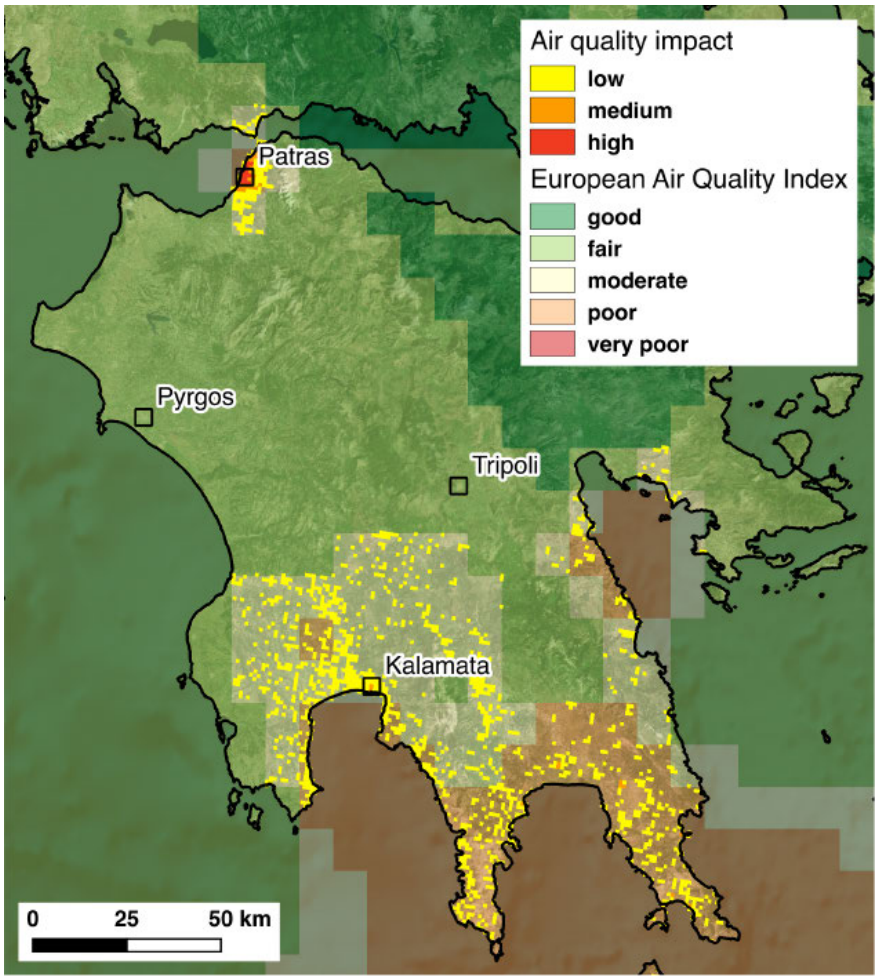


Figure 72: Map of air quality impact on population in the Peloponnese (Greece) on 11 July 2019 10:00 UTC (corresponding to the situation in Fig. 70).

When monitoring the air quality impact forecast with a resolution of 1 km (Fig. 72) at European scale, it can be difficult to identify the locations with forecasted impacts. To point out more clearly the potentially affected areas, the estimated impacts can be aggregated over the administrative regions (NUTS). By overlaying the EAQI forecast (Fig. 70) and the population density map, the total population numbers affected by moderate ($N_{moderate}$), poor (N_{poor}), and very poor ($N_{verypoor}$) air quality are determined for each region. These population numbers are then used to compute the impact index I_{nut} [people/km^2] for the region:

$$I_{nut} = \frac{(1 \cdot N_{moderate}) + (2 \cdot N_{poor}) + (3 \cdot N_{verypoor})}{Area_{nut}}$$

where $Area_{nut}$ is the surface area of the region. The reasoning behind the calculation of I_{nut} is to give higher weight to the population affected by poorer air quality levels. To determine an impact level (low, medium, high) for the region, I_{nut} is compared to thresholds that are based on Mdn_{dens}, the median of population densities of all NUTS regions (Mdn_{dens} = 82.9 people/km^2; see Table 12). An example of the resulting air quality impact map is shown in Figure 73. It can be observed that the Peloponnese (Greece) shows a low impact level, since it has a comparably low population density (33.8 people/km^2) and only a small share of the population is affected by moderate or poor air quality (Fig. 72), which results in a low impact index of I_{nut} = 18.6 people/km^2. In contrast, the more densely populated island of Crete (84.0 people/km^2) is fully affected by poor air quality (Fig. 70), which results in an impact index of I_{nut} = 168.0 people/km^2 and thus in a high impact level (Fig. 73).

Table 12: Threshold definitions to determine the impact levels in the NUTS regions, based on the impact index (I_{nut}) and the median of population densities of all NUTS regions (Mdn_{dens})

Impact level	none	low	medium	high
Thresholds	$I_{nut} = 0$	$0 < I_{nut} \leqslant Mdn_{dens}$	$Mdn_{dens} < I_{nut} \leqslant (2 \cdot Mdn_{dens})$	$(2 \cdot Mdn_{dens}) < I_{nut}$
Threshold values	$I_{nut} = 0$	$0 < I_{nut} \leqslant$ 82.9 people/km^2	82.9 people/km^2 $< I_{nut} \leqslant$ 165.8 people/km^2	165.8 people/km^2 $< I_{nut}$

It should be mentioned that the proposed approach aims at providing a coherent pan-European perspective and making impacts comparable across borders. This pan-European perspective may not be very useful for local decisions. For instance, in densely populated regions (e.g. Paris), the estimated impact level may be high already for moderate air quality. In the same line of thought, the impact model will under no circumstances compute a high impact level in regions with very low population densities (e.g. Iceland).

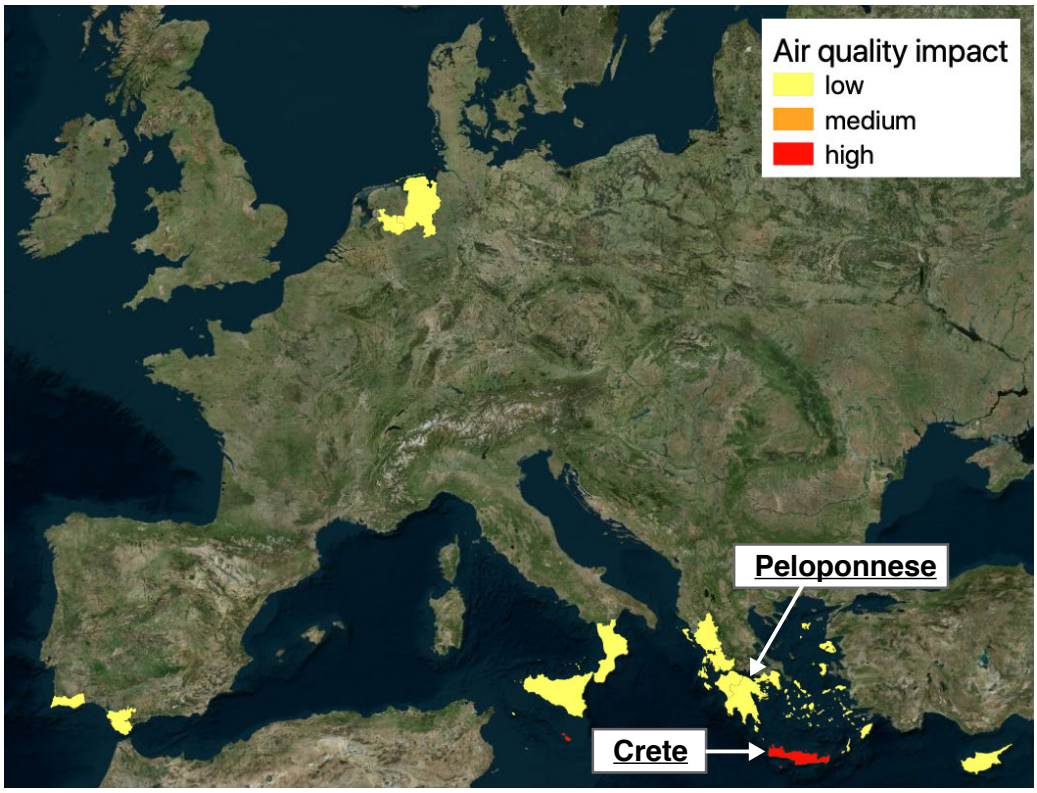


Figure 73: Air quality impact on population at the level of NUTS regions on 11 July 2019 1000UTC (corresponding to the situation in Figures 3 and 5).

7.2.1.2 Heatwave impact on population

A similar approach has been designed to estimate the impact of heatwaves on population at pan-European scale. As a base for defining the heatwave hazard, the Universal Thermal Climate Index (UTCI; see e.g. Di Napoli et al., 2018) has been used. The UTCI employs a multi-node human heat balance model to estimate the heat stress to the human body due to meteorological conditions. One of the UTCI products provided in the A4EU platform is a pan-European map showing the probability ($P_{UTCI>32}$) of exceeding a UTCI value of 32°C, a threshold above which heatwave-related death counts increase significantly. This probabilistic forecasting product has been used as a base for the hazard component in the proposed heatwave impact model (see Figure 74 for an example forecast of a heatwave that affected large parts of Europe in June 2019.

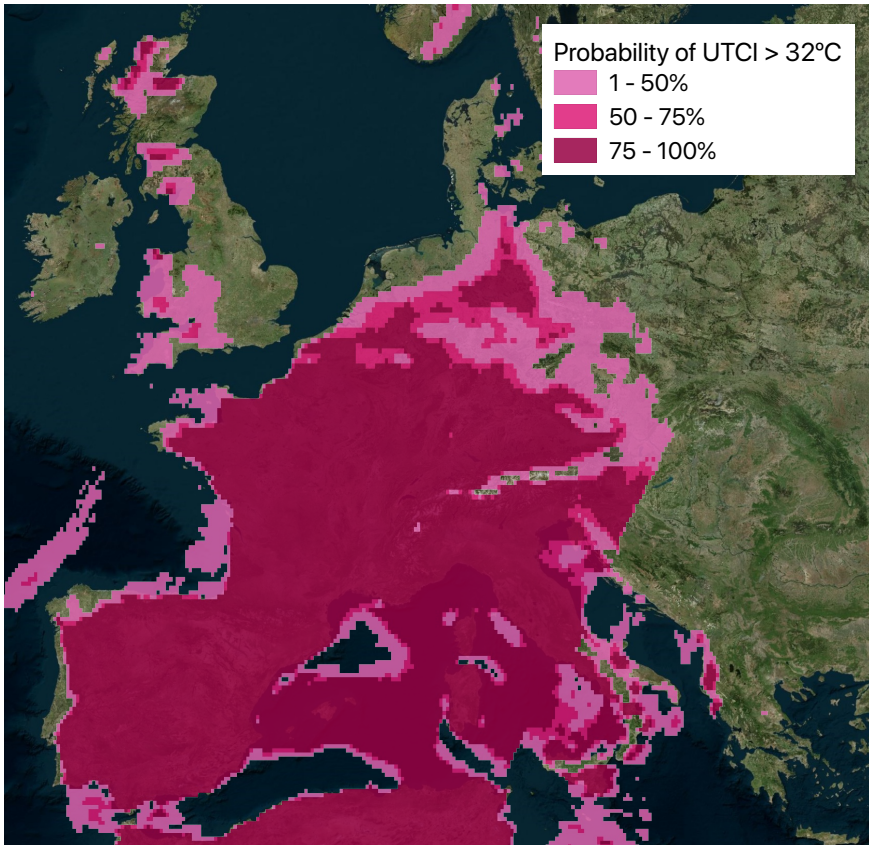


Figure 74: Probabilistic forecast of UTCI exceeding 32°C on 28 June 2019 18:00 UTC (lead time 18 h).

The concept of the heatwave impact approach is very similar to the one described in Section 7.2.1.1 (the main difference is the definition of the hazard classes): the heatwave hazard classes have been defined based on $P_{UTCI>32}$, and the thresholds defining the classes low, medium and high are for values of $P_{UTCI>32}$ of 1%, 50% and 75% (as indicated in the legend of Fig. 74).

For estimating the heatwave impact on population at 1 km resolution, the model applies an impact matrix (such as in Fig. 71, but with the heatwave hazard classes used here). Similarly, an additional output is computed to show the heatwave impact at regional level. Again, this is done by first counting for each NUTS region the total population numbers affected by the three hazard classes, then using them to calculate I_{nut}, and finally determining the impact level according to Table 12. The heatwave impact forecast at regional level, corresponding to the situation of Figure 74, is shown in Figure 75. The Figures 74 and 75 demonstrate the added value of the impact model: In some regions that are affected by a high heatwave hazard, the heatwave impact is only low to medium, since the regions are sparsely populated (see e.g. some rural areas in central Spain). On the other hand, some regions affected by medium heatwave hazard show a high impact due to high population densities (e.g. central Germany).

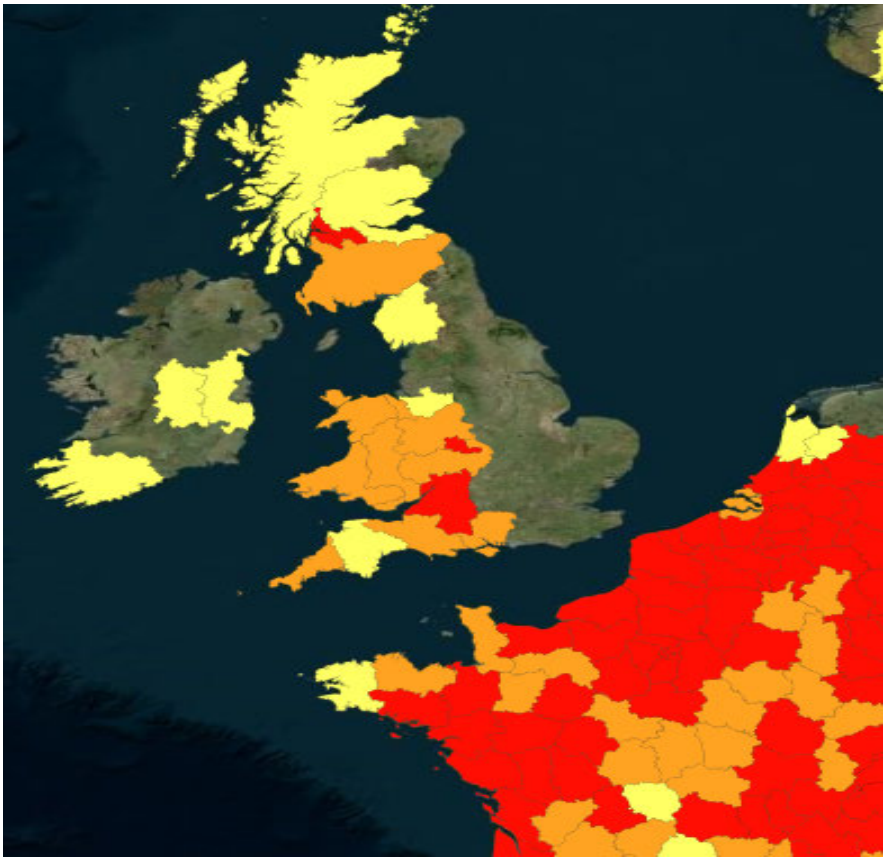


Figure 75: Heatwave impact forecast at the level of NUTS regions, corresponding to the situation in Figure 74.

7.2.2 Top-down approach

A system has been developed that potentially forecasts drought impacts across Europe a few months ahead (Sutanto et al., 2019c). This system provides drought impact forecasts that enable to take measures in time to manage the consequences of droughts, for instance for shipping, agriculture and nature. The study uses a so-called top-down approach implying that this method straight forwards starts with (reported) impacts and tries to link these to environmental factors (i.e. drought indices). The top-down approach deviates from the approach described in Section 7.2.1, which uses the conventional bottom-up approach that tries to find impacts through overlaying risk layers of hazard, exposure and vulnerability.

In the above-mentioned pan-European study, it was found that drought impacts, instead of drought hazards, can be forecasted with substantial skill up to 2-4 months ahead and in some cases even longer. They demonstrate this potentially pan-European approach by forecasting drought impacts, among others, on water-born transport, public water supply, water quality and ecosystems in German NUTS regions.

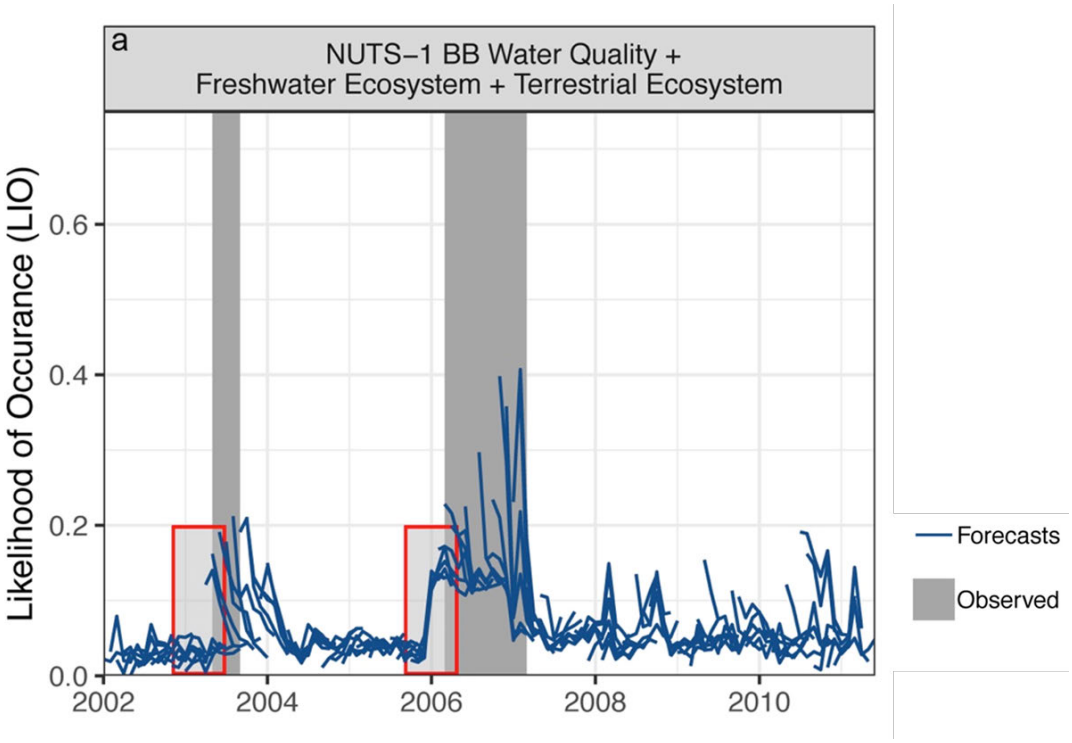


Figure 76: Re-forecasted likelihood of drought impacts and observed drought impact on a specific impact group in the period 2002-2010 for the NUTS region Brandenburg (BB), Germany.

The study used the European Drought Impact Inventory (EDII, Stahl et al., 2016) that includes thousands of historic drought impact reports for several sectors. Machine-learning techniques (Random Forest, Logistic Regression) were applied to connect the impacts in a certain month in the past to drought indices from that time, e.g. the Standardized Precipitation Index (SPI), the Standardized Precipitation-Evaporation Index (SPEI), and the Standardized Runoff Index (SRI). In this way, a relation (drought impact function) between the impact in a certain sector and drought indices was obtained for each European region that has sufficient impact reports in the EDII. In the next step, historic seasonal forecasts of the runoff, evaporation and precipitation (up to 7 months ahead), which are available in the archive of the European Centre for Medium-Range Weather Forecasts (ECMWF), have been used to calculate drought indices (e.g. SPI, SPEI, SRI) for each month in the period 2002-2010. Next, the above-mentioned drought impact functions and the historic seasonal forecasts (i.e. re-forecasts) of monthly drought indices were applied to forecast drought impacts in retrospective for selected regions across Europe (in this study: Germany), for specific sectors and for each month in the period 2002-2010. Eventually, the drought impact forecasts, up to 7 months ahead, were compared with the observed impacts for these regions. Figure 76 provides an example for a specific German NUTS-1 region for the period 2002-2010, that includes the major 2003 and 2006 drought. The grey bar shows observed impacts on a certain impact category (in this case water quality and ecosystems), which occurred in 2003 and 2006. The blue lines shows the re-forecasted Likelihood of drought Impact Occurrence (LIO), that are the median of the

ensembles for lead times of 1 to 7 month and for every months when the new forecast is issued. There are clear peaks in the reforecasts in 2003 and 2006. A more detailed analysis of the months preceding the 2003 and 2006 drought events (red boxes) led to the conclusion that drought impacts forecasts have substantial skill, up to 2-4 months ahead and in some cases even longer (Sutanto et al., 2019c). Ensembles can be used to consider uncertainty in drought impact forecasts.

7.3 Impact forecasting at regional scale

The flash flood hazard estimates obtained by the FF-EWS (Section 3.1.2) can provide valuable decision support at regional scale and high resolution. From such hazard estimates, emergency managers need to estimate the potentially flooded areas and resulting socio-economic effects, to coordinate flood response actions such as warnings or evacuations. In practice, this is commonly done non-automatically, based on the knowledge and experience of the person in charge, which increases the potential for suboptimal decisions and requires time, typically scarce during flash floods. To complement the FF-EWS hazard estimates and nowcasts in this respect, the method ReAFFIRM (Real-time Assessment of Flash Flood Impacts: a Regional high-resolution Method) has been developed and presented in a recently submitted paper (Ritter et al., 2019).

As shown in Figure 77, from the hazard estimates obtained by the FF-EWS along the drainage network, ReAFFIRM estimates the flood extent and depths at very high resolution based on the flood maps produced in the framework of the EU Floods Directive (European Commission, 2007). Finally, the flood depths are combined with several socio-economic exposure and vulnerability layers to quantitatively assess the impacts in three categories (namely, population present in the flooded areas, economic losses, and affected critical infrastructures). Figure 78 presents an example with the flood extent and estimated impacts during the flood that affected some parts of the city of Rubí (Spain) during the event of 15-16 October 2018.

The performance of ReAFFIRM has been evaluated on a number of flash flood events in Catalonia (Spain), and evaluated by comparison against the available information about the effects of the floods reported in newspaper articles, insurance claims, 112 emergency calls, and social media postings. The results show that ReAFFIRM is able to identify the locations of the most significant impacts (e.g. where casualties occurred), whereas some less severely affected locations were missed (in particular for events of return periods smaller than 10 years). The results show that, for the studied cases, the estimated number of people and critical infrastructures in the flooded areas were reasonable, while the economic losses were systematically overestimated.

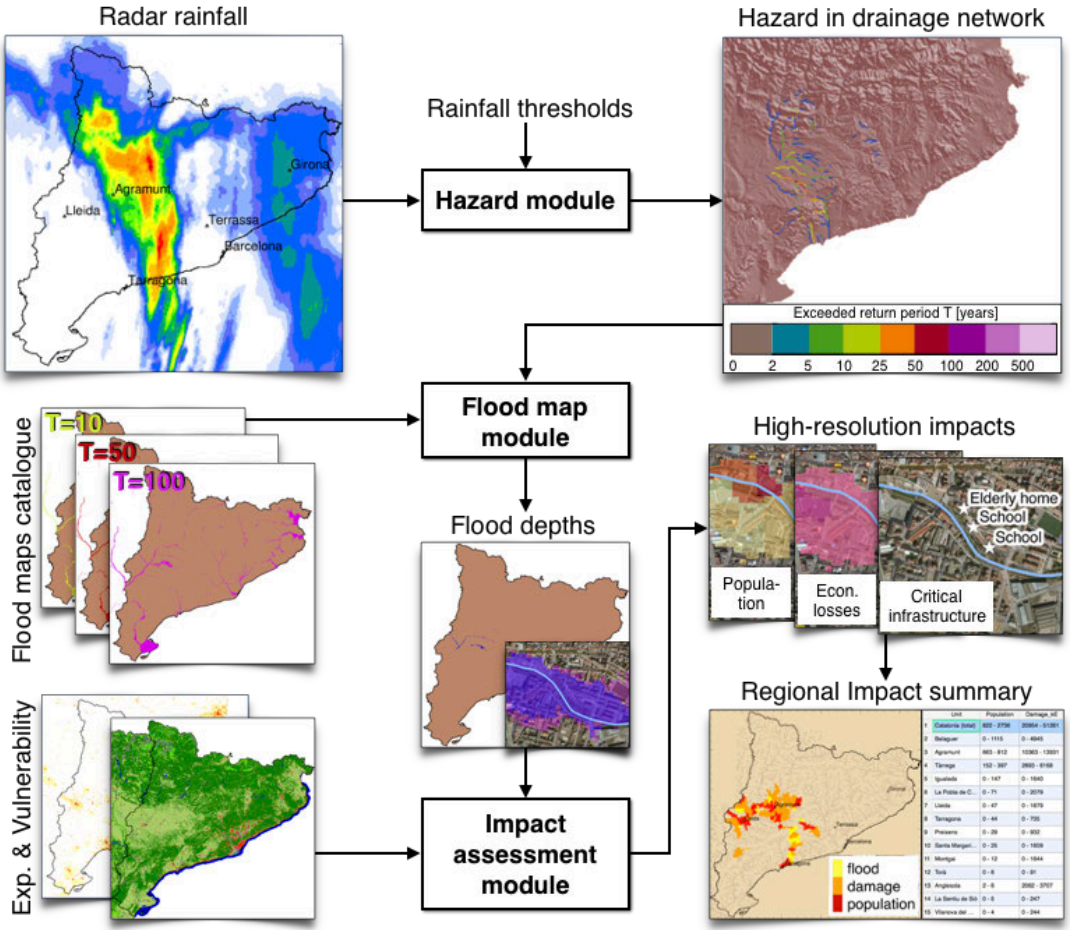
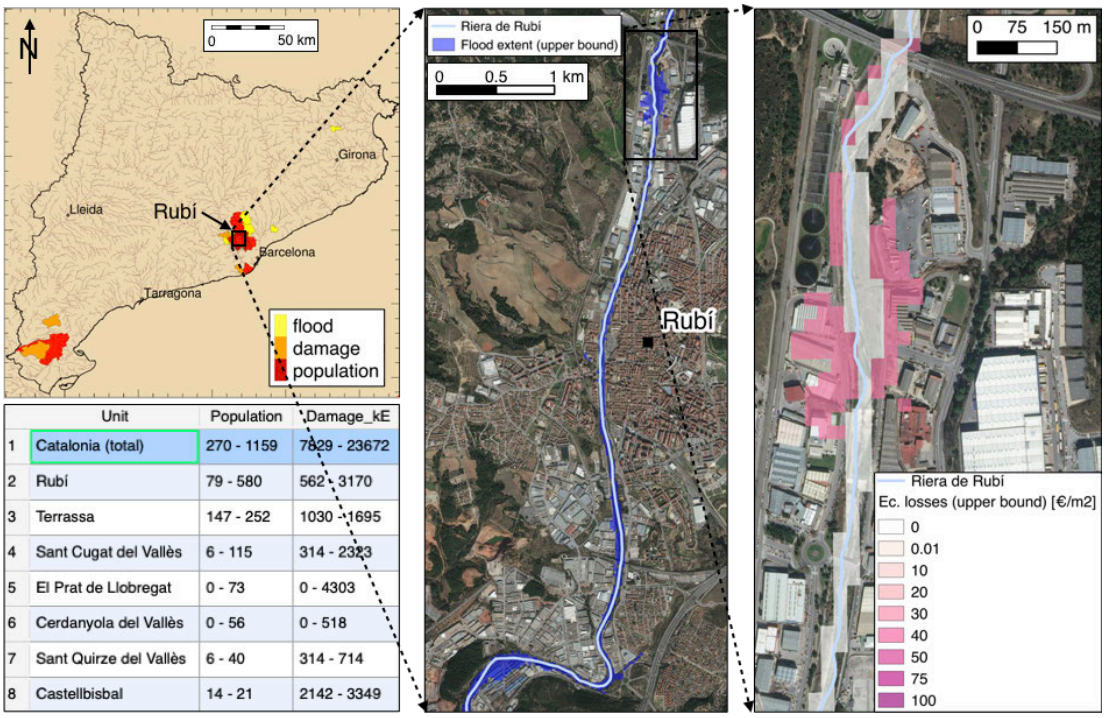


Figure 77: Concept of the method ReAFFIRM transforming the hazard estimates of FF-EWS into FF impacts.

The different components of ReAFFIRM are affected by several sources of uncertainty. In addition to the uncertainties that affect the FF-EWS hazard estimates and nowcasts, the translation from hazard into impacts is affected by different sources of uncertainty, among which the accuracy of the flood maps and the representativeness of the depth-damage curves to estimate the economic effects seem to be the most critical.

ReAFFIRM has been developed aiming at a high practical value: the computational requirements to apply it at regional scale are moderate, and the method can be relatively easily transferred to other regions: Once the FF-EWS system is implemented, the application of ReAFFIRM requires flood maps and exposure and vulnerability information that are nowadays publicly available throughout Europe.



	Unit	Population	Damage_kE
1	Catalonia (total)	270 - 1159	7829 - 23672
2	Rubí	79 - 580	562 - 3170
3	Terrassa	147 - 252	1030 - 1695
4	Sant Cugat del Vallès	6 - 115	314 - 2323
5	El Prat de Llobregat	0 - 73	0 - 4303
6	Cerdanyola del Vallès	0 - 56	0 - 518
7	Sant Quirze del Vallès	6 - 40	314 - 714
8	Castellbisbal	14 - 21	2142 - 3349

Figure 78: Left: Results obtained with ReAFFIRM summarised at municipality level over Catalonia for the event of 15-16 October 2018. Middle: Simulated flood extent in the city of Rubí. At this location, the FF-EWS system estimated return periods between 10 and 25 years. Right: Simulated economic losses in the northern part of Rubí (the location is shown in the middle panel).

The assessment of the impact of flash floods on population by using ReAFFIRM follows a bottom-up approach (Section 7.2.1). A top-down approach, such as in Section 7.2.2 is applied at the national scale (Finland) to connect snow load to calls for emergency tasks of civil protection authorities (Section 3.6.2).

8 Concluding remarks

Within ANYWHERE, no common methodology to assess uncertainty and robustness could be developed. Reasons for this were among others: (i) the different concepts to forecast different natural hazards, (ii) the data requirements (historic period, spatial coverage, type of data), and (iii) state of development of algorithms/tools. However, it came out that there was similarity in the design of the uncertainty assessment for almost all hazards, that is, comparing re-forecasts with observations, or a proxy for observations, using re-analysis data or simulation output. Clearly for some hazards only a few years of observed (or proxy) data were available. This means that the uncertainty and robustness assessment had limitations. In addition, not always re-forecast data were available.

Uncertainty

Most knowledge on uncertainty in products from hydrometeorological forecast platforms was already available at the start of ANYWHERE. In most cases, knowledge has improved or existing knowledge has been confirmed in the update cycles that are part of operational routines. Updating and regular systematic reporting is a continuous activity of hydrometeorological organizations. Specific uncertainty assessment of hydrometeorological forecast platforms was beyond ANYWHERE activities. The same holds for pan-European platforms that already existed prior to ANYWHERE: (i) European Floods Alert System (EFAS), European Fire Forecasting System and Global ECMWF Fire Forecasting model (EFFIS-GEFF), and European Drought Observatory (EDO). The current state of uncertainty of these platforms is briefly described based on the regular reporting. For instance, EFAS performs generally better in the northern part of Europe than in the south. The skill in winter is normally higher than in summer, etc...

During the second part of the ANYWHERE project, the hydrological forecast platform (EFAS) has gone through a major change, as the domain was increased, the projection changed, the hydrological model LISFLOOD was upgraded, and the parameters were recalibrated. These changes also affected all algorithms/tools that were reliant on EFAS, i.e. floods and drought, which required revision. This might affect uncertainty assessments.

Bias correction and downscaling of outcome from global weather forecasting platforms in mountainous regions substantially improves heavy precipitation forecasts in terms of spatial resolution and lead times (2-5 days ahead).

Algorithms that forecast flash flood hazard using a calibrated rainfall-runoff model perform slightly better than those only based on the rainfall accumulated in the upstream catchment (the basin-aggregated rainfall). In regions where insufficient river flow data are available to adequately calibrate the rainfall-runoff model, which frequently happens, the basin-aggregated rainfall is proved to be a good choice for events with frequencies lower than 10-year return period (see Corral et al., 2019).

Testing of landslide forecasting algorithms/tools showed that generally these are able to issue warnings for the most significant reported landslide events. Testing, however, is hampered that only those events that cause damage to infrastructure, buildings or roads are reported.

The pan-European Storm Surge Forecasting System was found to have satisfactory skills to predict tidal, surge and total water levels by testing it against measurements from 208 tidal gauge stations along the European coastline, and using two atmospheric forcing datasets. The regional storm model shows good performance for extreme storm surge. Wave predictions were used for evaluation, so far. The local inundation model showed good performance in the Pilot Site Stavanger.

The Universal Thermal Climate Index (UTCI) that is applied to assess heatwaves apparated to be sensitive to all input data, i.e. forecasted weather variables (wind speed, relative humidity, temperature, solar radiation). Evidence is provided that the UTCI can be used to forecast time periods when heat-related excess mortality is observed.

Air quality is forecasted using seven chemical transport models that are coupled with ECMWF weather forecasts, which calls for reliable weather forecasts. Comparison of observed ozone, as one of the gasses in the seven air quality models, against simulation data learnt that the ensemble of all models performs generally better than individual models. For this reason, it has been decided to use the ensemble forecast median of ozone and other pollutants' concentrations in ANYWHERE.

The Fire Weather Index (FWI) that is included in EFFIS-GEFF, is one of the most common indices to forecast wildfires. FWI quantifies how dangerous fires could be assuming an ignition occurred and knowing weather conditions from weather forecasting platforms. Comparison of FWI derived from stations with observed weather data and forecasted data (ECMWF-IFS) shows that skill exists until day 6, and in selected cases predictive skills can be achievable even at day 10. An analysis of the fatal Pedrógão Grande fire (Portugal) in June 2017 exposes that most of the region was classified at very high danger 10 days ahead.

Uncertainty in seasonal drought forecasting is addressed by using the ensemble forecasted time series of hydrometeorological variables. For each drought product not only the forecasted median is presented, but also the percentiles (10^{th} - 90^{th}). Drought forecasting either using standardized drought indices or threshold-based indices has acceptable skill up to 3-4 months, especially for longer accumulation periods of precipitation or runoff. Drought forecasts show the lowest uncertainty when they are done in winter, whereas forecasts issued in spring show the highest uncertainty.

It is challenging for the convective cell algorithm to produce skilful nowcasts for convective cells beyond 30 min. In general, the lifetime of small convective cells is short, and their development is fast (growth or decay, splitting and merging) making the skill rather low beyond lead times of, for example, one hour.

A preliminary analysis of the uncertainty of the snow load model to forecast number of tasks for emergency services because of expected falling trees and associated power cuts still shows a low correlation to environmental parameters, i.e. high uncertainty. Likely, more factors need to be included in the model, such as more detailed tree characteristics.

Analysis of the uncertainty of ECMWF and FMI algorithms to forecast different precipitation types show that both models forecast all precipitation types overall quite reliably, e.g. models give reliable and skillful probabilistic forecasts for rain and snow up to five days, and for sleet and freezing rain up to about 3 days. However, both models tend to slightly over-forecast the sleet. Uncertainty analysis of sleet and freezing rain could be improved because these rarely occurred in considered time series of observations.

Robustness

Whether weather forecast platforms, such as ECMWF IFS and IMF's HIRLAM and HARMONIE-AROME, are robust models under a future climate, or not, is an intriguing question for the ANYWHERE platforms, because these provide the probabilistic input for almost all algorithms/tools that forecast natural hazards. Positive in this context is that the weather platforms are constantly updated and are therefore by default able to cope with new situations in light of changing physical properties in a future climate. The ECMWF-IFS is already today very close to a full earth system model and, has been shown to capture observed climate change well. In fact, it is used as both a climate change model within the EC-EARTH consortium and as a reference climate model for the past using the reanalysis ERA5 being the latest version. Clearly, running the weather forecast platforms under a series of new climates was not feasible in ANYWHERE.

A high robustness of the storm surge model under future climate is expected, hence no additional changes in model parameterization are anticipated. Nevertheless, there are several "external" factors that can contribute to reduce robustness of forecasting storm surge levels and waves: (i) uncertainty in projections of sea level increase, (ii) quality of wave predictions, particularly relevant for semi-enclosed basins with shallow water areas (e.g. North Sea, Baltic Sea and/or Northern Adriatic Sea), and (iii) changes in precipitation and river discharge that may affect coastal hydrodynamics.

Projections consistently agree on a higher heat stress into the future that continues the increase in heat stress observed during recent past decades. However, the UTCI cannot not be considered robust on the future climate, because: (i) bioclimatic indices rely on the current weather and the ability of its forecast, and (ii) bioclimatic indices suppose an ability to acclimatisation (i.e. the adaptation to climate) equal to the one nowadays observed in populations. Future acclimatisation is expected to be different from the current one and to decrease vulnerability to heat-related hazards.

Fire forecasts in future depend very much on predictions of temperature and precipitation under a future climate (see remarks on robustness of ECMWF-IFS above). However, also non-climatic factors affect fire and are difficult to predict in the future. For example, fire forecasts are highly sensitive to the state of vegetation (i.e. fuel). If ANYWHERE were to provide forecasts of wildfires under future climate, the vegetation factor certainly needed to be included (currently omitted in EFFIS-GEFF).

Robustness of drought forecasts under a future climate firstly depends on the robustness of the seasonal hydrometeorological forecasts, which drive the drought algorithms (see remarks on robustness of ECMWF-IFS above). The model structure of climate models (in ANYWHERE: ECMWF and FMI platforms) also affects robustness of forecasts, as shown through an intercomparison of the Standardized Precipitation Index (SPI) derived from observed precipitation against SPI obtained from tens of climate models. An analysis of the threshold to identify meteorological drought shows that the differences are in general rather small, irrespective of the selected threshold, which points at relatively high robustness of this model aspect. Robustness of drought algorithms has also been investigated by the spread in projected drought from different combinations of climate and hydrological models. The smaller the differences between the projected hydrological drought characteristics are, the higher the robustness is. Robust, i.e. consistent, drought projections were obtained for the Mediterranean region, and robustness is lower in the transition area between lower and higher latitudes in Europe. We learnt from a global scale study that the selection of the threshold, either fixed and obtained from the reference period (last decades of 20th C) or transient and adapting to the gradually changing hydrological regime in the 21st C, has a substantial effect on projected hydrological drought characteristics (about 35% difference in drought duration).

No robustness analysis of FMI weather-type natural hazards (convective storms, snow load, precipitation type) has been performed yet. However, the statistical determination of storm severity is likely to alter and must be determined based on new statistical data. The FMI snow load model is an empirical algorithm with tuned parameters to the current climate. It is based on meteorological experience and observations, and hence robustness is hard to provide. The physical parametrization of the FMI precipitation type algorithm should be valid in future climate, but thresholds for the frequency of occurrences are expected to change.

Whether the ECMWF precipitation type is robust in terms of future climate is a question as to the ECMWF IFS is a robust model for climate change impacts (see remarks on robustness of ECMWF-IFS above). The precipitation type products are designed in such a way that parameters are involved in the recalibration round for each IFS model cycle. This means that breakpoints could be re-defined for the different intensities of precipitation type, if the climate change requires it in the future. This will be a very useful approach in the future climate to optimize forecasting products whatever the climate conditions are.

In summary, robustness assessment of algorithms/tools that forecast natural hazards under a future climate involves many different aspects, for example, (i) robustness of

platforms that forecast weather and hydrology under a future climate, which is input to natural hazard algorithms/tools, (ii) model structure of natural hazard algorithms/tools, (iii) parameters, and (iv) alert/emergency threshold. This is far-reaching and makes it challenging to provide a complete, up-to-date overview for a multi-hazard platform.

Compound natural hazards

Forecasting of compound weather events, incl. all the feedbacks in the system, is an inherent part of weather forecasting platforms, and hence these are implicitly included in the weather forecasts that drive possibly coinciding or cascading natural hazards.

Hotspots of joint occurrence of storm surge and high river discharge have been identified along the European coast. At lower latitudes (Lat. < 60°) these two wet hazards happened about 3-5% of the time in the period 1990-2016 when a medium hazard level was considered. In northern Europe, i.e. central and northern Scandinavia, coincidence is lower. When we increase the hazard level, in particular the southwest and south coast of the Iberian Peninsula show up as hotspots. The same holds for the southwest of Italy and the English Channel. Robustness of findings would increase if river discharge would be available at higher temporal resolution, e.g. hourly or sub-hourly time scale. The same applies to a joint simulation of storm surge levels and river levels at river mouth (in the current version river discharge is not hampered by total water levels at the river mouth, i.e. unconstrained outflow).

A methodology was developed to combine two hazards to identify areas prone to coinciding hazards. It was exemplified with, but it is not limited to, heat stress and fire danger. The methodology can be applied to analyse the extent of a past event and its spatial correlation with other observed variables (working in retrospect) or to make a prediction for the future (using forecast data). The combined heat and forest fire event that affected Europe in June 2017 was used as demonstration.

Another study built upon the previous approach to analyse more generally compound dry hazards (heatwaves, fires and droughts) at the pan-European scale. Probability of occurrence of all three hazards happening at the same day was small (<1% of time, during summers in 1990-2018). Droughts dominate in coinciding and cascading dry hazard events and mainly control the number and duration of cascading events. In most cascading events a drought appears first and last as a single hazard, followed by the coinciding drought-wildfire.

Pathways of wet cascading events have been identified by using three cases with extreme precipitation (amount or type) across Europe, which reveal various challenges for anticipation. It appears that the forecast of the triggering event is already related to uncertainty. Along the cascade with increasing time, complexity and uncertainty increases significantly. Interdependencies appear and key boundary conditions control the cascade at certain points.

Operationalizability

The reported development stage of ANYWHERE platforms/algorithms/tools shows that half of the platforms/algorithms/tools have proven in an operational environment (TRL=9). Most of the others have a technology already demonstrated in a relevant environment (TRL=6). About 10% algorithms/tools is not validated in a relevant environment yet (TRL≤5).

Impact assessment

Large-scale impact forecasting has been explored. Pan-European maps showing population density and critical infrastructure (education and health facilities, road network) are described. These provide spatially-distributed vulnerability information across Europe. Overlaying these vulnerability layers with the spatially-distributed natural hazard across Europe (so-called bottom-up approach) gives maps with impacts. Examples are presented of spatially-distribution of number people exposed to number of hazards (river floods, coastal floods, landslides, forest fires).

Two impact models using the bottom-up approach have been more elaborated: (i) air quality impact on population, and (ii) heatwave impact population. First index classes are defined for the natural hazard; when are hazard levels, for instance, good, poor). The same holds for the population density; when is density low, high. Then an impact matrix is developed that combines the index classes of the hazard and the population density and that defines the impact classes (e.g. low, high). Next the forecasted spatially-distributed natural hazard across Europe is converted into a pan-European map with hazard classes. This hazard map combined with the impact matrix gives the forecasted impacts across Europe. The example of the air quality forecast on a particular day, illustrates that the forecasted impact in Crete is higher than in the mainland (Peloponnese), although the forecasted hazard is more severe in the latter. The same applied to the example of the forecasted impact of heatwaves; although the hazard was more severe in some places (e.g. Central Spain) the impact was lower than in some places in Germany. The bottom-up approach has also been used at the regional level (Catalonia). In this case, forecasted flood depths are combined with several socio-economic exposure and vulnerability layers to quantitatively assess the impacts in three categories (namely, population present in the flooded areas, economic losses, and affected critical infrastructures).

An alternative approach to forecast impacts is also described. This top-down approach starts with reported impacts of natural hazards and combines these with levels of natural hazards to develop hazard impact functions. These functions are unique for a certain impact and region. Knowing these impact functions across Europe allows to forecasts impacts when a forecasted hazard map becomes available. The top-down approach is elaborated to forecast drought impacts in some German regions and specific impacts. The examples show that the drought impacts in some cases can be forecasted 2-4 months ahead.

References

Alderlieste, M.A.A., Van Lanen, H.A.J., Wanders, N., 2014: Future low flows and hydrological drought: how certain are these for Europe? In: Daniell, T.M., Van Lanen, H.A.J., Demuth, S., Laaha, G., Servat, E., Mahe, G., Boyer, J-F, Paturel, JE, Dezetter, A., Ruelland, D. (Eds.), Hydrology in a Changing World: Environmental and Human Dimensions, IAHS Publ. No. 363, pg. 60-65.

Alfieri, L., Velasco, D., Thielen, J., 2011: Flash flood detection through a multi-stage probabilistic warning system for heavy precipitation events. Adv. Geosci. 29, 69–75.

Alfieri, L., Berenguer, M., Knechtl, V., Liechti, K., Sempere-Torres, D., Zappa, M., 2017. Flash Flood Forecasting Based on Rainfall Thresholds. In: Handbook of Hydrometeorological Ensemble Forecasting. Duan, Q. , Pappenberger, F., Thielen, J., Wood, A., Cloke, H. L., Schaake, J. C. (Eds.), Springer Berlin Heidelberg,

Arns, A., Wahl, T., Dangendorf, S., Jensen, J., 2015: The impact of sea level rise on storm surge water levels in the northern part of the German Bight. Coast. Eng. 96, 118–131.

Aronica, G.T., Brigandí, G., Morey, N., 2012: Flash floods and debris flow in the city area of Messina, north-east part of Sicily, Italy in October 2009: the case of the Giampilieri catchment, Natural Hazards and Earth System Sciences, 12(5), 1295-1309.

Ballester, J., Robine, J.-M., Herrmann, F.R., and Rodó, X., 2011: Long-term projections and acclimatization scenarios of temperature-related mortality in Europe, Nature 2, 358.

Ballesteros Cánovas, J.A., Rohrer, M., Smith, P., Vitolo, C., Gascón, E., Di Giuseppe, F., (WUR); Claudia Di Napoli, Sempere-Torres, D., Van Lanen, H.A.J., Sutanto, S.J., Uijlenhoet, R., Berenguer, M., Park, S., Corral, C., Láng, I., Tiainen, P., Bergman, T., Fernández Montblanc, T., Ciavola, P., Fiorucci, P., Pignone, F., Salamon, P. and Stoffel, M. (2017): Assessing the robustness and uncertainty of models to forecast weather-induced hazards and impacts, ANYWHERE Report, Genova, Switzerland.

Berenguer, M., D. Sempere-Torres, and M. Hürlimann, 2015: Debris-flow forecasting at regional scale by combining susceptibility mapping and radar rainfall, Natural Hazards and Earth System Sciences, 15, 587-602.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., Vrac, M., 2017: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701–2723, https://doi.org/10.5194/hess-21-2701-2017.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., Widmann, M., 2019: Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change, Science Advances, 5(9), eaaw5531.

Birkmann, J., Cardona, O.D., Carreno, M.L., Barbat, A.H., Pelling, M., Schneiderbauer, S., Kienberger, S., M.Keiler, Alexander, D., Zeil, P., Welle, T., 2013: Framing vulnerability, risk and societal responses : the MOVE framework, Nat Hazards, 67, 193–211.

Błażejczyk, K., Jendritzky, G., Bröde, P., Fiala, D., Havenith, G., Epstein, Y., Psikuta, A., Kampmann, B., 2013: An introduction to the universal thermal climate index, Geogr Pol 86(1): 5–10.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability, Mon. Weather Rev., 78: 1-3.

Bröcker, J., Smith, L.A., 2007: Increasing the Reliability of Reliability Diagrams. Weather Forecast. 22, 651–661, https://doi.org/10.1175/WAF993.1.

Carrão, H., Naumann, G., Barbosa, P., 2016: Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability, Global Environmental Change 39: 108–124, http://dx.doi.org/10.1016/j.gloenvcha.2016.04.012.

Chi, X., Cubasch, U., Sodoudi, S., 2018: Assessment of human bio-meteorological environment over the Tibetan Plateau region based on CORDEX climate model projections, Theor Appl Climatol, https://doi.org/10.1007/s00704-018-2632-0.

Ciavola, P., Fernandez Montblanc, T., Armaroli, C., Berenguer, M., Bergman, T., Cloke, H., Corral, C., Di Napoli, C., Dotttori, F., Duo, E., Gascon, E., Harri, Kalas, M., A-M, Koistinen, J., Láng, I., von Lerber, A., Llort, X., Park, S., Pylkkö, P., Rebora, N., Rodríguez Ramos, Á., Salamon, P., Sempere, D., Smith, P.J., Van Lanen, H.A.J., Sutanto, S.J., Taufik, M., Teuling, A.J., Uijlenhoet, R., Vitolo, C., Vousdouskas, M., 2017: Improved version of the algorithms to be incorporated into the MH-EWS, ANYWHERE Internal Report, Ferrara, Italy.

Coccolo, S., Kämpf, J., Scartezzini, J.-L., Pearlmutter D., 2016: Outdoor human comfort and thermal stress: a comprehensive review on models and standards, Urban Climate 18: 33–57.

Colette, A., et al., 2012: Future air quality in Europe: a multi-model assessment of projected exposure to ozone, Atmos. Chem. Phys. 12:10613-10630.

Corral, C., Velasco, D., Forcadell, D., Sempere-Torres, D., Velasco, E., 2009. Advances in radar-based flood warning systems. The EHIMI system and the experience in the Besos flash-flood pilot basin. In: Flood Risk Management: Research and Practice. Samuels, P., Huntington, S., Allsop, W., Harrop, J. (Eds.), Taylor & Francis, London, UK, pp. 1295-1303, ISBN 978-0-415-48507-4.

Corral, C., Berenguer, M., Sempere-Torres, D., Poletti, L., Silvestro, F., Rebora, N., 2019: Comparison of two early warning systems for regional flash flood hazard forecasting, Journal of Hydrology, 572, 603–619. https://doi.org/10.1016/j.jhydrol.2019.03.026.

Damocles, 2019: CA17109 Understanding and modeling compound climate and weather events, https://www.cost.eu/actions/CA17109/#tabslName:overview

De Souza Hacon, S., Alves de Oliveira, B.F., and Silveira, I., 2019: A Review of the health sector impacts of 4 °C or more temperature rise. In: Nobre, C.A., Marengo J.A., Soares, W.R. (Ed.), Climate change risks in Brazil, Springer, Switzerland, 72-79.

Di Giuseppe, F., Pappenberger, F., Wetterhall, F., Krzeminski, B., Camia, A., Libertá, G., San Miguel, J., 2016: The potential predictability of fire danger provided by numerical weather

prediction, Journal of Applied Meteorology and Climatology, 55(11): 2469-2491, https://doi.org/10.1175/JAMC-D-15-0297.1.

Di Giuseppe, F., Rémy, S., Pappenberger, F., Wetterhall, F., 2018: Using the fire weather index (fwi) to improve the estimation of fire emissions from fire radiative power (frp) observations, Atmospheric Chemistry and Physics, 18 (8): 5359–5370, doi:10.5194/acp-18-5359-2018, URL https://www.atmos-chem-phys.net/18/5359/2018/.

Di Giuseppe, F., Rémy, S., Pappenberger, F., Wetterhall, F., 2017: Improving forecasts of biomass burning emissions with the fire weather index, Journal of Applied Meteorology and Climatology, 56 (10): 2789–2799.

Di Giuseppe, F., Vitola, C., San Miguel, J., Krzeminski, B., Pappenberger, F., 2019: Fire danger: the prediction skill provided by ECMWF Integrated Forecasting System, Environment International (submitted).

Di Napoli, C., Pappenberger, F., Cloke, H.L., 2018: Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI,. Int. J. Biometeorol. 62: 1155-1165.

Di Napoli, C., Pappenberger, F., Cloke, H.L., 2019: Verification of heat stress thresholds for a health-based heatwave definition, Journal of Applied Meteorology and Climatology 58: 1177-1194.

Duo, E., Fernández-Montblanc, T., Armaroli, C., Ciavola, P.: Coastal Flooding Impact Analysis at Stavanger (Norway), Environment International (in press).

European Commission, 2007: Addressing the challenge of water scarcity and droughts in the European Union, Communication from the Commission to the European Parliament and the Council, European Commission, DG Environment, Brussels.

European Commission, 2007: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks.

European Commission, 2014: TRL Scale in Horizon 2020 and ERC – Explained, Brussels, https://enspire.science/trl-scale-horizon-2020-erc-explained/.

Fernández-Montblanc, T., Vousdoukas, M.I., Mentaschi, L.,Ciavola, P.: A pan-European high resolution storm surge hindcast, Environment International (in press).

Fernández-Montblanc, T., Vousdoukas, M.I., Ciavola, P., Voukouvalas, E., Mentaschi, L., Breyiannis, G., Feyen, L., Salamon, P., 2019: Towards robust pan-European storm surge forecasting, Ocean Modelling 133: 129-144. https://doi.org/10.1016/j.ocemod.2018.12.001.

Feyen, L.,. Dankers, R. 2009: Impact of global warming on streamflow drought in Europe, J. Geophys. Res., 114: D17 116, doi:10.1029/2008JD011438.

Flannigan, M.D., Haar, T.V., 1986: Forest fire monitoring using noaa satellite avhrr, Canadian Journal of Forest Research, 16 (5): 975–982.

Fleig, A.K., Tallaksen, L.M., Hisdal, H., Demuth, S., 2006: A global evaluation of streamflow drought characteristics, Hydrol. Earth Syst. Sci. 10: 535-552.

Freire S., K. MacManus, M. Pesaresi, E. Doxsey-Whitfield, J. Mills, 2016: Development of new open and free multi-temporal global population grids at 250 m resolution. Proceedings of the 19th AGILE Conference on Geographic Information Science. Helsinki, Finland.

Gascón, E., Hewson, T., Haiden, T., 2018: Improving predictions of precipitation type at the surface: description and verification of two new products from the ECMWF ensemble, Weather Forecast. 33, 89–108, https://doi.org/10.1175/WAF-D-17-0114.1.

Giglio, L., Randerson, J.T., Werf, G.R., 2013: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (gfed4),.Journal of Geophysical Research: Biogeosciences, 118 (1): 317–328.

Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y.J., 2003: An enhanced contextual fire detection algorithm for modis, Remote sensing of environment, 87 (2): 273–282.

Gregow H., Puranen U., Venäläinen A., Peltola H., Kellomäki S., Schultz D., 2008: Temporal and spatial occurrence of strong winds and large snow load amounts in Finland during 1961–2000, Silva Fennica 42: 515–534, http://dx.doi.org/10.14214/sf.231.

Groisman, P.Y. et al, 2016: Recent changes in the frequency of freezingprecipitation in North America and NorthernEurasia, Environ. Res. Lett.11 045007.

Guerreiro, S.B., Dawson, R.J., Kilsby, C., Lewis, E., Ford, A., 2018: Future heat-waves, droughts and floods in 571 European cities, Environ. Res. Lett. 13: 034009. https://doi.org/10.1088/1748-9326/aaaad3.

Guyette, R.P., Thompson, F.R., Whittier, J., Stambaugh, M.C., Dey, D.C., 2014: Future fire probability modeling with climate change data and physical chemistry, Forest Science, 60(5): 862-870.

Guzzetti F, Peruccacci S, Rossi M, Stark C.P., 2008: The rainfall intensity-duration control of shallow landslides and debris flows: An update. Landslides, 5, 3–17. doi: 10.1007/s10346-007-0112-1.

Haiden, T., Janousek, M., Bidlot, J., Buizza, R., Ferranti, L., Prates, F., Vitart, F., 2018: Evaluation of ECMWF forecasts, including the 2018 upgrade, Technical Memo 831, ECMWF, Reading.

Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Mayer, S., Nesje, A., Nilsen, J. E. O., Sandven, S., Sandø, A. B., Sorteberg, A. & Ådlandsvik, B., 2017: Climate in Norway 2100 - A Knowledge Base for Climate Adaption. NCCS Report 1/2017, Norwegian Centre for Climate Services, Oslo, Norway.

Harmonie-Arome General description: http://hirlam.org/index.php/hirlam-programme-53/general-model-description/mesoscale-harmonie.

Havenith, G., Fiala, D., Blazejczyk, K., Richards, M., Bröde, P., Holmer, I., Rintamaki, H., Benshabat, Y., Jendritzky, g., 2012: The UTCI-clothing model. Int. J. Biometeorol. 56 (3): 461–470.

Hisdal, H., Tallaksen, L.M., Clausen, B., Peters, E., Gustard, A., 2004: Hydrological Drought Characteristics, In: Tallaksen, L.M., Van Lanen, H.A.J. (Eds.) Hydrological Drought. Processes and Estimation Methods for Streamflow and Groundwater, Development in Water Science, 48, Elsevier Science B.V., pg. 139- 198.

Hürlimann, M., C Abancó, J Moya, I Vilajosana, 2014: Results and experiences gathered at the Rebaixader debris-flow monitoring site, Central Pyrenees, Spain, Landslides, 11, 939-953.

Idier, D., Paris, F., Cozannet, G.L., Boulahya, F., Dumas, F., 2017: Sea-level rise impacts on the tides of the European Shelf, Cont. Shelf Res. 137: 56–71.

IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, (Barros, V.R., Field, C.B., Dokken, D.J., Mastrandrea, M.D., Mach, K.J., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., White, L.L., Eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC (2018): Annex I): Glossary (Matthews, J.B.R., Ed.). In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty (Masson-Delmotte, V. et al. Eds), IPCC, Geneva (in Press).

Jalkanen R., Konôpka B., 1998: Snow-packing as a potential harmful factor on Picea abies, Pinus sylvestris and Betula pubescens at high altitude in northern Finland, European Journal of Forest Pathology 28: 373–382, http://dx.doi.org/10.1111/j.1439-0329.1998.tb01191.x.

Jalkanen A., Mattila U., 2000: Logistic regression models for wind and snow damage in northern Finland based on the National Forest Inventory data, Forest Ecology and Management 135: 315–330, http://dx.doi.org/10.1016/S0378-1127(00)00289-9.

Jensen, M., 2014: Reliability vs Robustness, https://blogs.mentor.com/mikej/blog/2014/10/07/reliability-vs-robustness/.

Jevrejeva, S., Jackson, L.P., Riva, R.E.M., Grinsted, A., Moore, J.C., 2016: Coastal sea level rise with warming above 2 °C, Proc. Natl. Acad. Sci. 113: 13342–13347.

Jolliffe, I.T., Stephenson, D.B. (Eds.), 2011. Forecast Verification: A Practitioner's Guide, in: Atmospheric Science, 2nd ed., John Wiley & Sons, Incorporated, New York, United Kingdom.

Kaiser, J. et al., 2012: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9 (1): 527–554.

Kaufman, Y., Ichoku, C., Giglio, L., Korontzi, S., Chu, D., Hao, W., Li, R.-R., Justice, C. 2003: Fire and smoke observed from the earth observing system modis instrument-products, validation, and operational use, International Journal of Remote Sensing, 24 (8): 1765–1781.

Kundzewicz, Z.W., et al., 2013. Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal, 59, 1–28.

Liu, B., Siu, Y.L., Mitchell, G., Xu, W., 2016: The danger of mapping risk from multiple natural hazards, Natural Hazards 82(1):139-153, doi 10.1007/s11069-016-2184-5.

Lokys, H.L., Junk, J., Krein, A., 2015: Future changes in human-biometeorological Index classes in three regions of Luxembourg, Western-Central Europe, Advances in Meteorology 2015: ID 323856.

Lehtonen I., Hoppula P., Pirinen P., Gregow H., 2014: Modelling crown snow loads in Finland: a comparison of two methods, Silva Fennica 48(3), article id 1120.

Lehtonen, I., Kämäräinen, M., Gregow, H., Venäläinen, A., Peltola, H., 2016: Heavy snow loads in Finnish forests respond regionally asymmetrically to projected climate change, Natural Hazards and Earth System Sciences, 16: 2259–2271.

Marécal, V. et al., 2015: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev. 8: 2777-2813.

Martín-Alcón, S., González-Olabarria, J.R., and Coll, L., 2010: Wind and snow damage in the Pyrenees pine forests: effect of stand attributes and location, Silva Fenn., 44: 399–410, doi:10.14214/sf.138.

Martiník, A., Mauer, O., 2012: Snow damage to birch stands in Northern Moravia, J. Forest Sci., 58: 181–192.

Mason, I.B., 1982: A model for assessment of weather forecasts, Aust. Meteorol. Mag., 30, 291-303.

McKee, T.B., Doesken, N.J., Kleist, J., 1993: The relationship of drought frequency and duration to time scale, In: Proc. of 8th Conf. on Applied Climatology, Anaheim, California, 17-22 January 1993, Boston, American Meteorological Society, 179-184.

Meigh, J., Tate, E., McCartney, M., 2002: Methods for identifying and monitoring river flow drought in southern Africa. In: Van Lanen, H.A.J., Demuth,S. (Eds.), FRIEND 2002 – Regional Hydrology: Bridging the Gap between Research and Practice, IAHS Publ. no. 274, 181–188.

Metta, S., von Hardenberg, J., Ferraris, L., Rebora, N., and Provenzale, A., 2009: Precipitation nowcasting by a spectral-based non-linear stochastic model, J. Hydrometeorol., 10, 1285–1297.

Nykänen M.-L., Peltola H., Quine C.P., Kellomäki S., Broadgate M.,1997: Factors affecting snow damage of trees with particular reference to European conditions. Silva Fennica 31: 193–213.

Olliaho P. 2014: Parametric uncertainty in numerical weather prediction models http://urn.fi/URN:ISBN:978-951-697-824-9.

Orlowsky, B., Seneviratne, S.I., 2013: Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections, Hydrol. Earth Syst. Sci. 17: 1765–1781.

Palau, R.M., Hürlimann, M., Berenguer, M., Sempere-Torres, D., 2020: Influence of mapping unit for regional landslide Early Warning Systems. Comparison between pixels and polygons in Catalonia (NE Spain), Landslides (submitted November 2019).

Pappenberger, F., Jendritzky, G., Staiger, H., Dutra, E., Di Giuseppe, F., Richardson, D.S., Cloke, H.L., 2015: Global forecasting of thermal health hazards: the skill of probabilistic predictions of the Universal Thermal Climate Index (UTCI), Int. J. Biometeorol. 59(3): 311-323.

Park, S., Berenguer, M., Semper-Torres, D., 2019: Long-term analysis of gauge-adjusted radar rainfall accumulations at European scale, J. Hydrol., 573, 768-777, doi:10.1016/j.jhydrol.2019.03.093.

Park, S., Tuller, S.E., Jo, M., 2014: Application of Universal Thermal Climate Index (UTCI) for microclimatic analysis in urban thermal environments, Landsc. Urban Plan. 125: 146–155.

Petty J.A., Worrell R., 1981: Stability of coniferous tree stems in relation to damage by snow. Forestry 54: 115–128, http://dx.doi.org/10.1093/forestry/54.2.115.

Prudhomme, C., Giuntoli, I., Robinson, E.L., Clark, D.B., Arnell, N.W., Dankers, R., Fekete, B.M., Franssen, W., Gerten, D., Gosling, S.N., Hagemann, S., Hannah, D.M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., Wisser, D., 2014: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proceedings of the National Academy of Sciences 111(9): 3262–3267, doi:10.1073/pnas.1222473110.

Raïmat, C, Hürlimann, M., Corominas, J., Luis- Fonseca, R., Moya J., 2013: Reconstrucción de la frecuencia de corrientes de derrubios en el barranco de Erill (Lérida), In : Alonso, E., Corominas, J., Hürlimann, M. (Eds.), VIII Simposio Nacional sobre Taludes y Laderas Inestables, Santander, pg. 519-529.

Räisänen, J. and Ylhäisi, J.S., 2015: CO2-induced climate change in northern Europe: CMIP2 versus CMIP3 versus CMIP5, Clim. Dynam. 45: 1877–1897, doi: 10.1007/s00382-014-2440-x.

Räisänen, J., (2016): Twenty-first century changes in snowfall climate in Northern Europe in ENSEMBLES regional climate models, Clim. Dynam. 46: 339–353, doi: 10.1007/s00382-015-2587-0.

Reeves, H.D., Elmore, K.L., Ryzhkov, A., Schuur, T., Krause, J., 2014: Sources of uncertainty in precipitation-type forecasting, Weather Forecasting 29: 936–953, https://doi.org/10.1175/WAF-D-14-00007.1.

Ritter, J., Berenguer, M., Corral, C., Park, S., Sempere-Torres, D., 2019: ReAFFIRM – Real-time Assessment of Flash Flood Impacts: a Regional high-resolution Method, Environment International, (in press).

Robine, J-M., Cheung, S.L. K., Roy, S.L., van Oyen, H., Griffiths, C., Michel, J-P., Herrmann, F.R., 2008: Death toll exceed 70,000 in Europe during the summer of 2003, C. R. Biologies 331, 171-178, doi:10.1016/j.crvi.2007.12.001.

Rogers, E., Deaven, D.G., DiMego, G.J., 1995: The regional analysis system for the operational 'early' Eta model: original 80-km configuration and recent changes, Weather Forecasting, 10: 810-825.

Saltikoff, E., Haase, G., Delobbe, L., Gaussiat, N., Martet, M., Idziorek, D., Leijnse, H., Novák, P., Lukach, M., Stephan, K., 2019. OPERA the Radar Project. Atmosphere 10: 320, https://doi.org/10.3390/atmos10060320.

Schauwecker, S., Gascón, E., Park, S., Ruiz-Villanueva, V., Schwarb, M., Sempere-Torres, D., Stoffel, M., Vitolo, C., Rohrer, M., (2019): Anticipating cascading effects of extreme

precipitation with pathway schemes – three case studies from Europe, Environmental International, 127: 291-3043. doi: 10.1016/j.envint.2019.02.072

Schauwecker, S., Schwarb, M., Rohrer, M., Stoffel, M., 2019: Operational prediction of heavy precipitation with increased lead times - Challenges and implications for risk preparedness in mountain environments, Environment International (under review).

Schelhaas M.-J., Nabuurs G.-J., Schuck A., 2003: Natural disturbances in the European forests in the 19th and 20th centuries, Global Change Biology 9: 1620–1633, http://dx.doi.org/10.1046/j.1365-2486.2003.00684.x.

Schroeder, W., Prins, E., Giglio, L., Csiszar, I., Schmidt, C., Morisette, J., Morton, D., 2008: Validation of goes and modis active fire detection products using aster and etm+ data, Remote Sensing of Environment, 112 (5): 2711–2726.

Seidl R., Fernandes P.M., Fonseca T.F., Gillet F., Jönsson A.M., Merganičová K., Netherer S., ArpaciA., Bontemps J.-D., Bugmann H., González-Olabarria J.R., Lasch P., Meredieu C., Moreira F., Schelhaas M.-J., Mohren F., 2011: Modelling natural disturbances in forest ecosystems: a review, Ecological Modelling 222: 903–924, http://dx.doi.org/10.1016/j.ecolmodel.2010.09.040.

Shutts G., Pallares, A., 2014: Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction model, https://doi.org/10.1098/rsta.2013.0284.

Simpson, M.J.R., Nilsen, J.E.Ø., Ravndal, O.R., Breili, K., Sande, H., Kierulf, H.P., Steffen, H., Jansen, E., Carson, M., Vestøl, O., 2015: Sea Level Change for Norway: Past and Present Observations and Projections to 2100, Norwegian Centre for Climate Services report 1/2015, ISSN 2387-3027, Oslo, Norway.

Simpson, M., Breili, K., Kierulf, H.P., Lysaker, D., Ouassou, M., Haug, E., 2012: Estimates of Future Sea-Level Changes for Norway, Technical Report of the Norwegian Mapping Authority.

Simpson, M.J.R., Ravndal, O.R., Sande, H., Nilsen, J.E.Ø., Kierulf, H.P., Vestøl, O., Steffen, H., 2017: Projected 21st Century Sea-Level Changes, Observed Sea Level Extremes, and Sea Level Allowances for Norway, J. Mar. Sci. Eng. 5: 36. https://doi.org/10.3390/jmse5030036.

Slater, L.J., Villarini, G., 2018: Enhancing the predictability of seasonal stream flow with a statistical-dynamical approach, Geophysical Research Letters 45(13): 6504-6513.

Smith, P.J., C. Vitolo, Á. Rodríguez Ramos, X. Llort, M. Berenguer, P. Ciavola, H. Cloke, E. Gascon, F. di Giuseppe, T. Fernandez Montblanc, A-M. Harri, J. Koistinen, H.A.J. Van Lanen, C. Di Napoli, C., P. Pylkkö, N. Rebora, S.J. Sutanto, A.J. Teuling, and R. Uijlenhoet, R., 2017: Modules/tools requirements to be integrated into the MH-EWS, ANYWHERE Internal Report, ECMWF, Reading, UK.

Solantie R., 1994: Effect of weather and climatological background on snow damage of forests in Southern Finland in November 1991, Silva Fennica 28: 203–211.

Spinoni, J., Naumann, G., Vogt, J.V., Barbosa, P., 2015: The biggest drought events in Europe from 1950 to 2012, J. Hydrol. Regional Studies 3: 509-524, https://doi.org/10.1016/j.ejrh.2015.01.001.

Spinoni, J., Vogt, J.V., Naumann, G., Barbosa, P., Dosio, A., 2018: Will drought events become more frequent and severe in Europe? Int. J. Climatol. 38: 1718–1736, doi: 10.1002/joc.5291.

Sutanto, S.J., Van Lanen, H.A.J., Wetterhall, F., Llort, X., 2019a: Potential of pan-European seasonal hydro-meteorological drought forecasts obtained from a Multi-Hazard Early Warning System, Bulletin of the American Meteorological Society. https://doi.org/10.1175/BAMS-D-18-0196.1.

Sutanto, S.J., Vitolo, C., Di Napoli, C. D'Andrea, M., Van Lanen, H.A.J., 2019b: Heatwaves, droughts, and fires: exploring compound and cascading dry hazards at the pan-European scale, Environment International 134: 105276, https://doi.org/10.1016/j.envint.2019.105276.

Sutanto, S.J., van der Weert, M., Wanders, N., Blauhut, V., Van Lanen, H.A.J., 2019c: Moving from drought hazard to impact forecasts, Nature Communications, https://doi.org/10.1038/s41467-019-12840-z.

Stagge, J.H., Tallaksen, L.M., Gudmundsson, L., Van Loon, A.F., Stahl, K., 2015: Candidate distributions for climatological drought indices (SPI and SPEI), Int. J. Climatol.35: 4027–4040, doi: 10.1002/joc.4267.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acacio, V., Dias, S., Stagge, J.H., Tallaksen, L.M., Kampragou, E., Van Loon, A.F., Barker, L.J., Melsen, L.A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H.A.J., 2016: Impacts of European drought events: insights from an international database of text-based reports, Nat. Hazards Earth Syst. Sci., 16: 801–819, doi:10.5194/nhess-16-801-2016.

Syphard A.D., Sheehan T., Rustigian-Romsos H., Ferschweiler K., 2018: Mapping future fire probability under climate change: Does vegetation matter?, PLOS ONE 13(8): e0201680, https://doi.org/10.1371/journal.pone.0201680.

Termonia P., Fischer C., Bazile E., Joly A., 2018: The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, https://www.geosci-model-dev.net/11/257/2018/.

Turco, M., Jerez, S., Doblas-Reyes, F.J., AghaKouchak, A., Llasat, M.C., Provenzale, A., 2018: Skilful forecasting of global fire activity using seasonal climate predictions, Nature Communications, 9(1): p.2718.

Trambauer, P., Werner, M., Winsemius, H.C., Maskey, S., Dutra, E., Uhlenbrook, S., 2015: Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa, Hydrol. Earth Syst. Sci., 19: 1695-1711, doi:10.5194/hess-19-1695-2015.

Valinger E., Lundqvist L., 1992: The influence of thinning and nitrogen fertilization on the frequency of snow and wind induced stand damage in forests, Scottish Forestry 46: 311–320.

Van Hateren, T.C., Sutanto, S.J., Van Lanen, H.A.J., 2019: Evaluating skill and robustness of seasonal meteorological and hydrological drought forecasts at the catchment scale – Case Catalonia (Spain). Environment International. https://doi.org/10.1016/j.envint.2019.105206.

Van Huijgevoort, M.H.J., Van Lanen, H.A.J., Teuling, A.J., Uijlenhoet, R., 2014: Identification of changes in hydrological drought characteristics from a multi‐GCM driven ensemble constrained with observed discharge, Journal of Hydrology 512: 421 – 434, doi.org/10.1016/j.jhydrol.2014.02.060.

Van Lanen, H. A. J., C. Prudhomme, W. Wanders, M. H. J. Van Huijgevoort, 2018: Future of Drought. In: Iglesias, A., Assimacopoulos, D. and Van Lanen, H.A.J. (Eds.) (2018): Drought: Science and Policy, Wiley Blackwell, pg. 69-92, https://www.wiley.com/en-us/Drought%3A+Science+and+Policy-p-9781119017073

Van Lanen, H.A.J., Berenguer, M., Ciavola, P., Cloke, H., Di Napoli, C., Fernandez Montblanc, T., Gascon, E., Harri, A-M, Koistinen, J., Llort, X., Park, S., Pylkkö, P., Rebora, N., Rodríguez Ramos, Á., Salamon, P., Sempere, D., Smith, P.J., Sutanto, S.J., Teuling, A.J., Uijlenhoet, R., Vitolo, C., Vousdouskas, M., 2017: Compilation of the first version of the algorithms and products to forecast impacts of weather-induced hazards at adequate time-space resolution for different geo-climatic settings across Europe, ANYWHERE Internal Report, Wageningen, The Netherlands.

Van Lanen, H.A.J., Ballesteros Canovas, J.A., Berenguer, M., Bergman, T., Ciavola, P., Cloke, H., Corral, C., Di Napoli, C., Fernandez Montblanc, T., Gascon, E., , Koistinen, J., Llort, X., Park, S., Pignone, F., Pylkkö, P., Rebora, N., Roca Sancho, J., Rodríguez Ramos, Á., Rohrer, M., Salamon, P., Sempere, D., Smith, P.J., Stoffel, M., Sutanto, S.J., Teuling, A.J., Uijlenhoet, R., Vitolo, C., von Lerber, A., Vousdouskas, M., 2019: Final version of the code of the algorithms to be implemented in the Multi-Hazard Early Warning System (MH-EWS), ANYWHERE Internal Report, Wageningen, The Netherlands (Revised report).

Van Loon, A.F., (2015): Hydrological drought explained. WIREs Water, 2: 359–392. doi: 10.1002/wat2.1085.

Van Loon, A.F., Van Lanen, H.A.J., 2012: A process-based typology of hydrological drought, Hydrol. Earth Syst. Sci. 16 1915–1946, https://doi.org/10.5194/hess-16-1915-2012.

Van Vuuren, D.P., Edmonds, J., Kainuma, M. et al., 2011: The representative concentration pathways: an overview, Climatic Change 109: 5.

Van Wagner, C.E., Forest, P., 1987: Development and structure of the Canadian forest fire weather index system, In Can. For. Serv., Forestry Tech. Rep.

Vicente-Serrano, S.M. et al., 2010: A multi-scalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index-SPEI, Journal of Climate 23(7): 1696-1718.

Vitolo, C., Di Giuseppe, F., D'Andrea, M., 2018: Caliver: An R package for CALIbration and VERification of forest fire gridded model outputs, PloS one, 13(1): p.e0189419, https://doi.org/10.1371/journal.pone.0189419.

Vitolo, C., Di Giuseppe, F., Krzeminski, B., San-Miguel-Ayanz, J., 2019a: A 1980–2018 global fire danger re-analysis dataset for the Canadian Fire Weather Indices, Scientific Data, 6:p.190032, https://doi.org/10.1038/sdata.2019.32.

Vitolo, C., Di Napoli, C., Di Giuseppe, F., Cloke, H.L., Pappenberger, F., 2019b. Mapping combined wildfire and heat stress hazards to improve evidence-based decision making, Environment International, 127: 21-34, https://doi.org/10.1016/j.envint.2019.03.008.

Vousdoukas, M.I., Almeida, L.P., Ferreira, Ó., 2012: Beach erosion and recovery during consecutive storms at a steep-sloping, meso-tidal beach, Earth Surf. Process. Landf. 37: 583–691.

Wanders, N., Van Lanen, H.A.J., 2015: Future discharge drought across climate regions around the world modelled with a synthetic hydrological modelling approach forced by three General Circulation Models, Natural Hazard Earth Syst. Sci, 15, 487–504, doi:10.5194/nhess-15-487-2015.

Wanders, N., Wada, Y., 2015: Human and climate impacts on the 21st century hydrological drought, Journal of Hydrology 526: 208–220, http://dx.doi.org/10.1016/j.jhydrol.2014.10.047.

Wanders, N., Wada, Y., Van Lanen, H.A.J., 2015: Global hydrological droughts in the 21st century under a changing hydrological regime, Earth System Dynamics 6: 1–15, doi:10.5194/esd-6-1-2015.

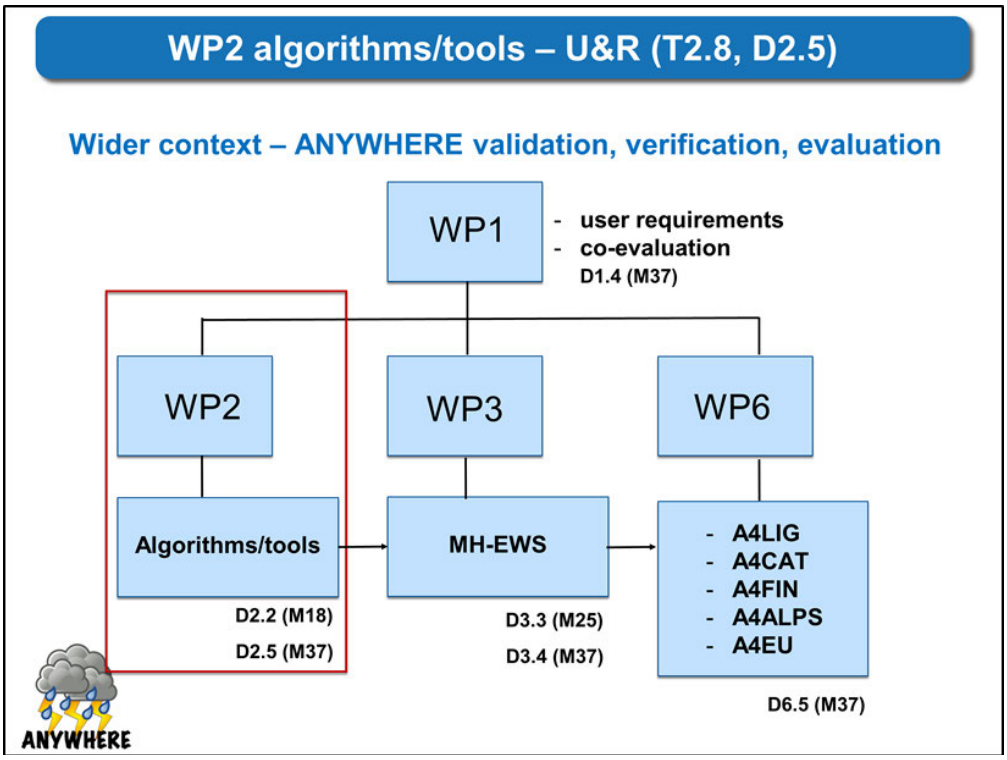
Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., Wood, E.F., 2019: Development and evaluation of a Pan-European multimodel seasonal hydrological forecasting system, J. Hydrometeorology, 20: 99-114, doi: 10.1175/JHM-D-18-0040.1.

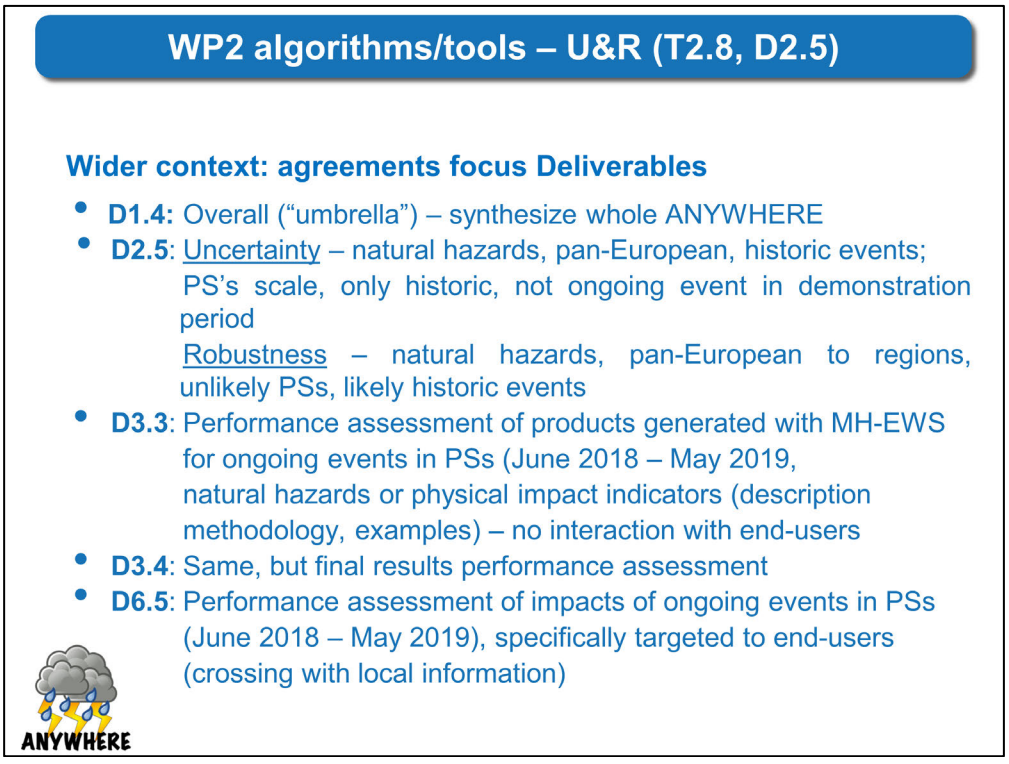
Wilson, J.W., Crook, N.A., Mueller, C.K., Sun, J., Dixon, M., (1998): Nowcasting Thunderstorms: A Status Report, Bull. Am. Meteorol. Soc. 79: 2079–2100, https://doi.org/10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2.

Wooster, M.J., Roberts, G., Perry, G., Kaufman, Y., 2005: Retrieval of biomass combustion rates and totals from fire radiative power observations: Frp derivation and calibration relationships between biomass consumption and fire radiative energy release, Journal of Geophysical Research: Atmospheres (1984–2012), 110 (D24).

Yevjevich, V., 1967: An objective approach to definition and investigations of continental hydrologic droughts, Hydrology Papers 23, Colorado State University, Fort Collins, USA.

Annex I: Validation, verification and evaluation: wider context, other WPs (top), and focus of each of WPs (bottom)





Annex II: Metrics used to assess forecasting skill

The continuous ranked probability skill score (CRPS) is a measure of the integrated squared difference between the cumulative distribution function of the forecasts and the corresponding cumulative distribution function of the proxy for observations.

The probability of detection (POD) is the fraction of observed events that is forecasted correctly The range of POD is between 0 and 1, where 1 represents the good end, and 0 means no prediction skill.

In a reliability diagram, the observed frequency is plotted against the forecast probability which is divided in to a number of bins. Forecasts are perfectly reliable if the curve lies on the diagonal. If the curve lies above the diagonal it indicates under-forecasting, and if below over-forecasting. The frequency of forecasts in each probability bin are shown in the histogram and the climatology line shows the climatological probability of the precipitation type during verification period.

The relative operating characteristic (ROC) curve (Jolliffe and Stephenson, 2011) measures the ability of a probabilistic forecast to discriminate between an event and no event (in this case precipitation). For a given intensity threshold and a range of probability thresholds between 0 and 1 for a yes/no decision, the curve is constructed by plotting the probability of detection (POD) against the false alarm ratio, or probability of false detection (POFD), for the given probability thresholds. The area under the curve (AUC) can be used as a summary statistic that measures potential skill.

Area under the ROC curve (AUC) measures the resolution, which means the ability of the model to discriminate different outcomes of forecast distributions with events and non-events. Model that forecasts always the climatological probability has perfect reliability but no resolution. For perfect forecast AUC is 1 and for random forecast it is 0.5.

The Brier skill score (BSS) is widely used in hydrometeorology for quantifying the accuracy of probabilistic predictions. It is the mean of squared difference between the paired binary observations (*o*) and probability forecasts (*f*):

$$BS = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2.$$

In the equation, probabilistic forecast (y_i) have value between 0 and 1, and observation (o_i) value 0 or 1. The BS has values from 0 to 1 and gives smaller values for better forecasts. The BS can be decomposed into the sum of three components: uncertainty, reliability, and resolution. Uncertainty is dependent only on observations, so BS is a summary score of reliability and resolution of the forecasts.

The Brier skill score measures the relative improvement of the BS of investigated forecast model compared to the BS of the reference model, usually sample climatology:

$$BSS = 1 - \frac{BS}{BSref}.$$

The BSS have a range of -∞ to 1. Positive values stand for a more accurate model than reference, and negative values indicate that model is less accurate than reference. If the BSS is 0 both models have equal skill.

The economic value of probabilistic precipitation type forecasts can be evaluated with cost/loss decision model. Relative economic value tells the relative improvement in economic value between climatological and perfect forecasts and gives information that can be used in decision making when precautionary actions are needed to prevent costly danger situations (e.g. using anti-icing to protect an aircraft against icing in-flight). In the model, taking an action causes a cost even if an event occurs or not, and a loss when action is not taken but the event occurs. If precautionary action is not taken and event does not occur the cost and loss is zero. The cost and loss are different for different users; therefore, the economic value is plotted as a function of cost-loss ratio. The lower the cost-loss ratio is, the larger the loss is in comparison to the cost.

Annex III: Background information to FMI snow-load model

Forest damage caused by snow loading on trees occurs frequently in boreal environments. On a global scale, snow-induced damage to the forest ecosystem dynamics is not considered as a major natural disturbance (e.g. Seidl et al. 2011), however, it has regional importance both in Central and Northern Europe. It is estimated, in addition to northern parts of the continent, also in the mountainous regions e.g. in the Alps and Pyrenees, the amount of timber damaged by snow during a typical year vary between one and 4 million m^3 (Nykänen et al., 1997; Schelhaas et al., 2003, Martín-Alcón et al., 2010).

Common forms of snow-induced forest damage include stem breakage and bending or leaning of stems, but trees can also be uprooted if the soil is unfrozen (Petty and Worrell 1981; Valinger et al. 1994; Nykänen et al. 1997). Tree type, positioning and stand characteristics control the resistance of trees to snow, and some tree species are thus more vulnerable to snow damage than others, e.g. conifers are often considered to be most badly affected (Nykänen et al., 1997) but, for instance, birches are vulnerable for bending (Martiník and Mauer, 2012).

The snow load accumulation on the tree canopy introduces a high risk for the trees to fall or their branches to break on the transmission lines causing possible long-lasting, and therefore life-threatening, power-cuts in rural areas with cold temperatures. This was witnessed in Finland in winter 2017-2018, where extreme snow load conditions and extensive snowfalls engaged close to 300 electricians to return the power and telecommunication connections to over 3000 households and occupied civil protection authorities with military assistance to provide help to households without electricity for several days. Economic damages to forest are approximated to be close to 50 million euros.

The concrete prevention measures include e.g. burying the transmission lines or logging the trees around transmission lines, however, there are economic and pragmatic limitations in practical implementation in rural areas. And although a warning of the threatening snow-load event can be issued beforehand, the damages to forest are impossible to prevent. Nevertheless, the warning and forecasting of snow load accumulation are found to be useful and essential for both the power and telecommunication companies and the authorities to increase the self-preparedness and planning of resources.

Finnish Meteorological Institute provides forecasts of snow load accumulation on canopy and transmission lines based on an experimental model. The model separates the four different types (rime, dry snow, wet snow, and frozen snow) of snow load, with the main forming processes being riming and wet snow accumulation. The rimed snow load accumulates when the super-cooled cloud droplets freeze to surfaces. Temperature and wind conditions affect snow density, and because of favorable environmental conditions, riming increases with height and is more common in the trees on hills, than in the valleys (Jalkanen and Konôpka 1998; Jalkanen and Mattila 2000). Wet snow load is formed when the melting snow is attached to trees.

Temperatures are typically close to 0…+0.5 degree Celsius, in addition, favorable wind direction and speed (3-6 m/s) enforces the packing of the wet snow to canopy or transmission lines. The removal of the snow load is occurring typically by melting and moderate wind conditions shake the snow off, but too strong winds can break the trees or branches weaken under the heavy load.

The FMI model on snow load has been used operationally to predict heavy crown snow loads over a decade, and the parameters of the model have evolved based on the experience of model performance in different weather situations. Hence the model parameters are empirical and based on statistics, not physics. There have been model verification studies, e.g. Lehtonen et al. (2014), where the spatial occurrence of different snow load types was studied and compared to a simple model (Gregow et al., 2008) in Finland covering years from 1961 to 2010. The climatology was built using meteorological observations at 29 locations across Finland, classified daily images of canopy snow cover at the Hyytiälä forestry field station located in the region of Pirkanmaa and with the help of two short case studies. It was concluded that the forests most prone to heavy riming are those located on tree-covered hills in northern Finland. This is despite the fact that weather conditions most favorable for riming occur most frequently in the interior of western Finland as well as in the hills and fells in the north.

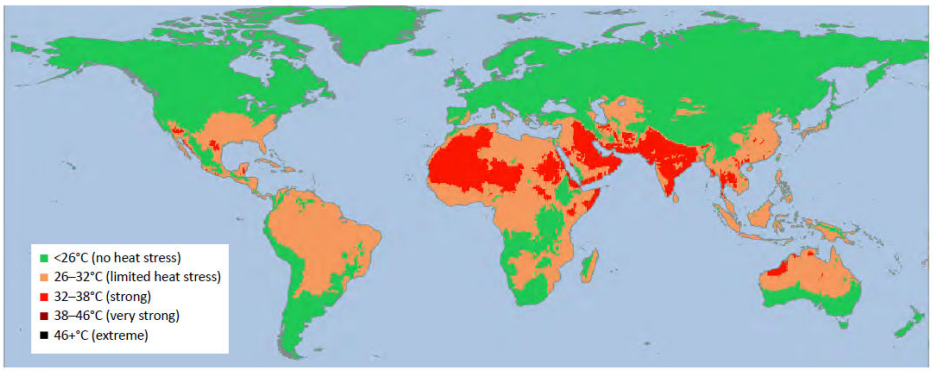
Annex IV: Projections for Universal Thermal Climate Index

Studies on future projections for UTCI have been performed both at the global and local level. At the global level, Kjellstrom et al. (2017) investigated the future UTCI bioclimate 2071-2099 (1981-2010 as baseline, ISI-MIP data, HadGEM and GFDL models) under a RCP6.0 greenhouse gas scenario. The RCP6.0 scenario, where RPC stands for representative concentration pathways[28], considers lower emissions than RCP8.5 (worst future scenario) due to the application of some mitigation strategies and technologies. CO_2 concentration is predicted to rise less rapidly but still reaching 660 ppm by 2100, and total radiative forcing to stabilise shortly after 2100 (van Vuuren et al., 2011). The result of the projection is shown in Figure IV-1.

Map 1 on the top level (Fig. IV-1) shows the areas currently experiencing no heat stress, limited heat stress, strong or more intense heat stress. Map 2 (Fig. IV-1) shows that heat stress levels are generally predicted to increase all over the world. This result is backed up by studies carried at more local level, from Brazil to the Tibetan plateau (De Souza Hacon; 2019; Chi et al. 2018). Europe will also witness an increase in heat stress conditions. In the highly populated areas of Western-Central Europe a large population will be exposed to this change. For instance, in Luxembourg - with its dense population and the large cross-border commuter flows - cold stress levels are expected to decrease significantly in the near future up to 2050, whereas the increase in heat stress will turn statistically significant in the far future up to 2100 for an emission scenario with rapid economic growth, a balanced use of energy resources, and an increasing global population until the middle of this century (Lokys et al., 2015).

[28] There are four greenhouse gas concentration trajectories adopted by the IPCC for its Fifth Assessment Report (AR5) in 2014 describing four possible climate futures: RCP2.6, RCP4.5, RCP6.0 and RCP8.5. Relative to 1850–1900, global surface temperature change for the end of the twenty-first century (2081–2100) is projected to likely exceed 1.5°C for RCP4.5, RCP6.0 and RCP8.5 (high confidence). Warming is likely to exceed 2°C for RCP6.0 and RCP8.5 (high confidence), more likely than not to exceed 2°C for RCP4.5 (medium confidence), but unlikely to exceed 2°C for RCP2.6 (medium confidence) (See WGI SPM E.1, 12.4.1, Table 12.3, available from https://ar5-syr.ipcc.ch/topic_futurechanges.php).

Map 1: Climate Research Unit (CRU) baseline data for the period 1981–2010



Map 2: ISI-MIP data, HadGEM and GFDL model mid-points of Representative Concentration Pathways* 6.0 (increase of global mean temperature between 2.6 and 3.1 degrees) for 2071–2099

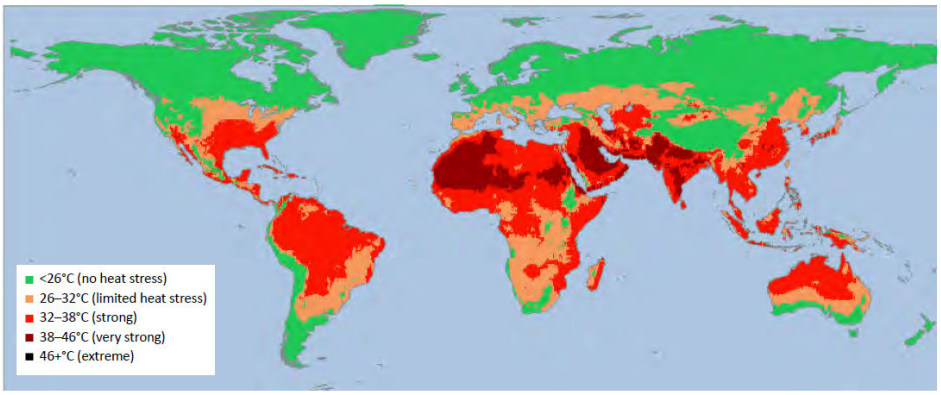


Figure IV-1: Maps of UTCI monthly mean levels in the shade in 67,000 grid cells for the hottest month (Kjellstrom et al., 2017).

Annex V: Projections of air quality

Anticipating future air quality is a major concern and it has been the focus of many atmospheric chemistry research projects over the past decades. We here report the results from two recent studies on RAQ forecasts under a future climate for two pollutants considered in **ANYWHERE**, namely ozone and particulate matter.

One study has been carried out by Colette et al. on ozone (2012). In their analysis an ensemble of air quality models covering both regional and global spatial scales were implemented in a coordinated manner for future projections of anthropogenic emissions at the 2030 horizon. The two scenarios explored were developed in the framework of the Global Energy Assessment (Riahi et al., 2012). They include identical measures for air quality legislation but different climate policies, namely one of the scenarios is a baseline (*reference*), whereas the other limits global warming to 2°C by the end of the century (*sustainable*). Colette et al. report that ozone precursors such as NO_x will drop down to 30% to 50% of their current levels, depending on the scenario. As a result, annual mean ozone will slightly increase in NO_x saturated areas, but the overall ozone burden will decrease substantially (Figures V-1 and V-2).

Future European particulate matter concentrations have also been evaluated under the influence of climate change and anthropogenic emission reductions (Lacressonnière et al. 2017). Specifically, 30-year simulations for present and future scenarios were performed with an ensemble of four regional chemical transport models and +2°C scenarios (RCP4.5) were issued from different regional climate simulations. Results showed a large reduction of PM_{10} and $PM_{2.5}$ concentrations in a +2°C climate over Europe, which can be mostly attributed to emission reduction policies. Under a current legislation scenario, annual PM_{10} could be reduced by between 1.8 and 2.9 µg m^{-3} (14.1–20.4%). If maximum technologically feasible emission reductions were implemented, further reductions of 1.4–1.9 µg m^{-3} (18.6–20.9%) are highlighted.

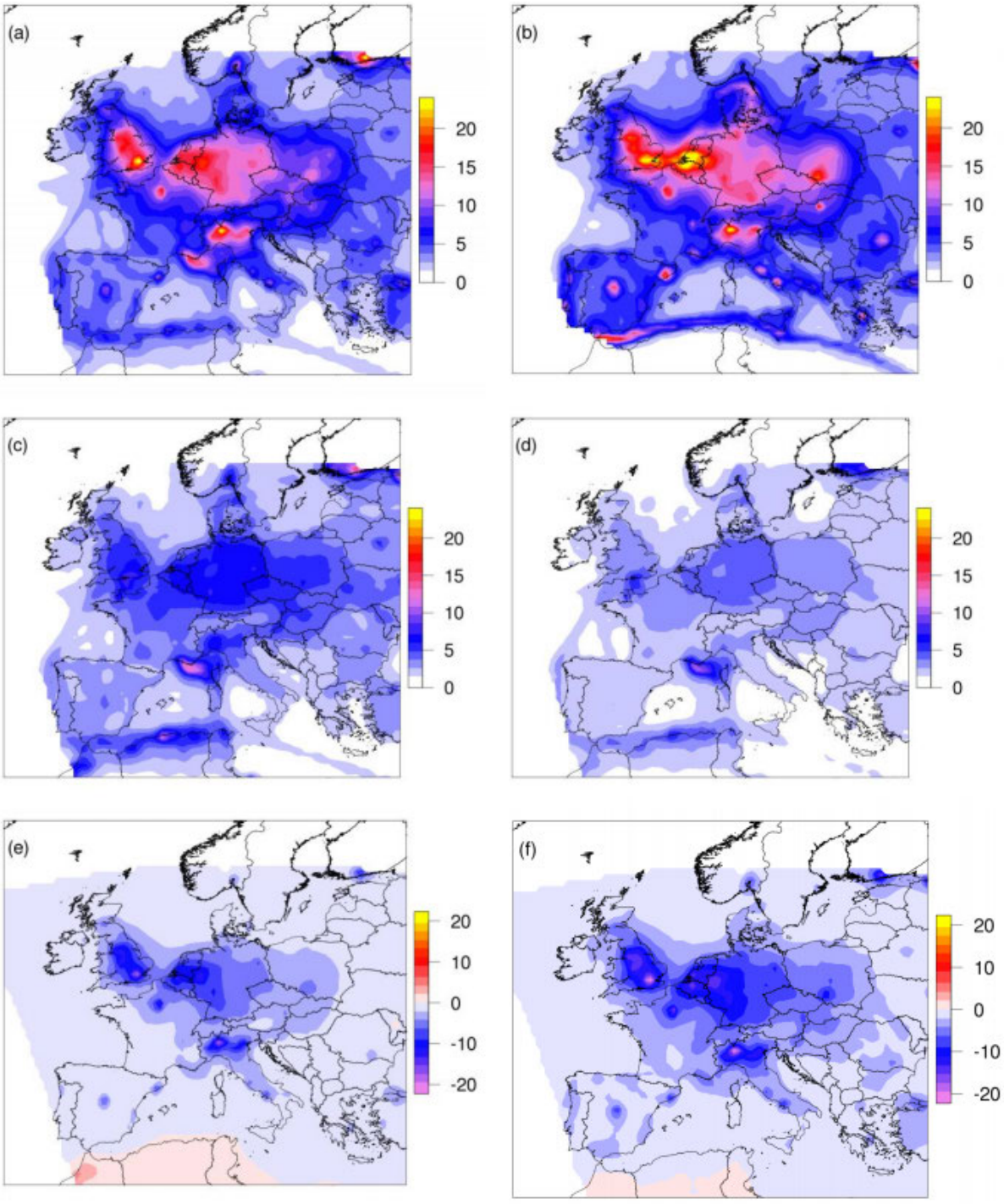


Figure V-1: Panels a, b, c, d: Ensemble median of average NO_2 concentrations (μg m^{-3}) over the 10 yr of simulation for (a) the GEA 2005 emissions, (b) the EMEP 1998–2007 emissions, (c) "reference" 2030 and (d) "sustainable" 2030. Panels e, f: the difference between "reference" 2030 and 2005 and between "sustainable" 2030 and 2005, respectively (Colette et al. 2012).

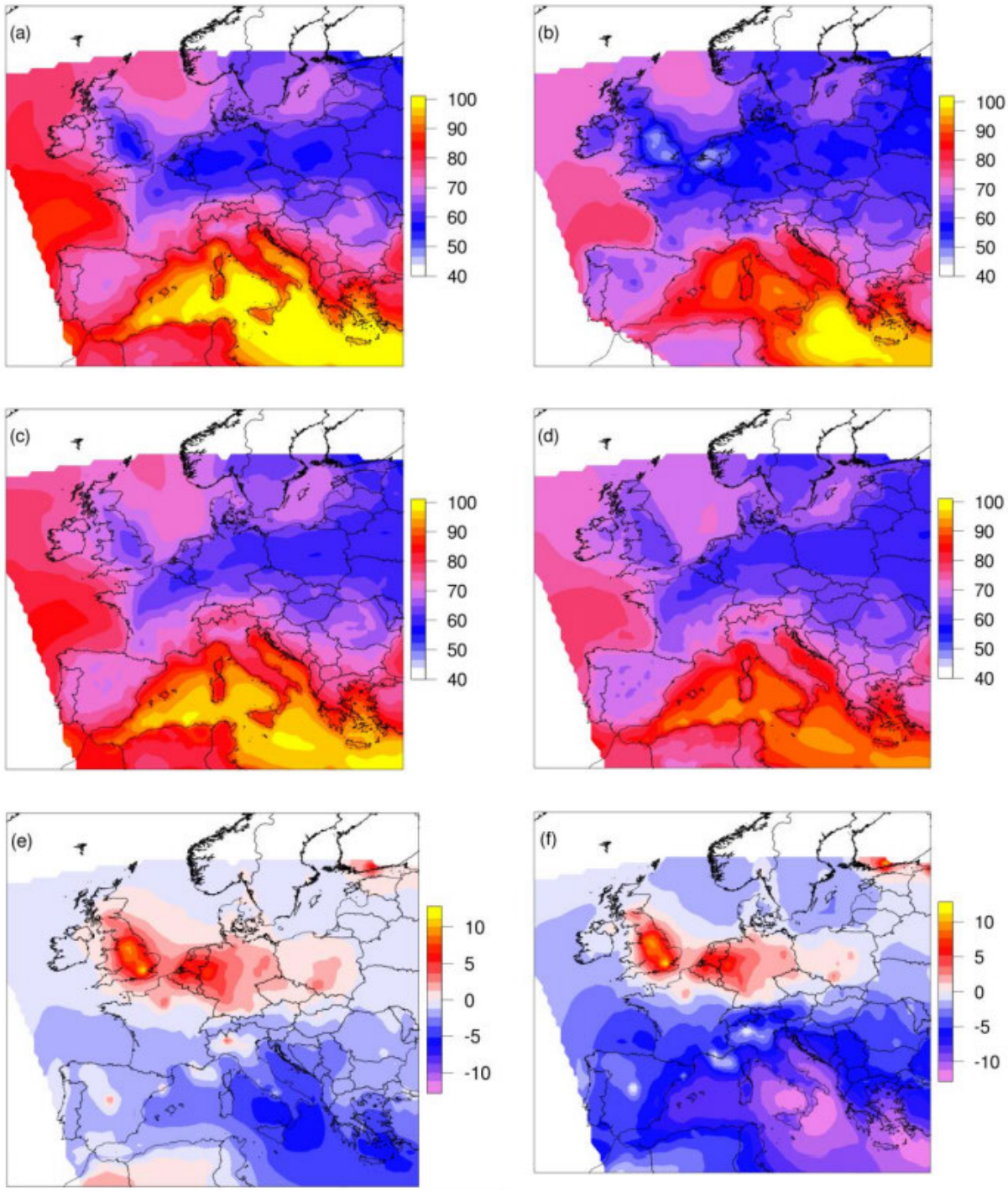


Figure V-2: Same as Figure V-1, but for O_3 ($\mu g\ m^{-3}$, Colette et al. 2012).

Annex VI: Projections of drought

One of the first large - scale studies that used early - twenty - first - century emission scenarios (IPPC's SRES) to investigate future drought characteristics was done by Arnell (2003). Previous work was based on climate scenarios from the 1990s and largely focussed on annual flow. Arnell (2003) applied a single hydrological model that was driven by the output of several climate models and emission scenarios. An important finding was that the coefficient of variation of annual runoff is projected to increase, which will lead to higher frequency of drought in runoff in 2050, particularly in parts of Europe and southern Africa. Several years later, multi - global hydrological models were introduced in the WaterMIP project to explore the impact of global warming on hydrological extremes (Harding et al., 2011; Corzo Perez et al., 2011), using the output from multi - climate models and multi - emission scenarios (SRES). In the framework of WaterMIP, uncertainty in future hydrological drought was explored by including only those models that performed reasonably in the past (Van Huijgevoort et al., 2014), see Figure VI-1.

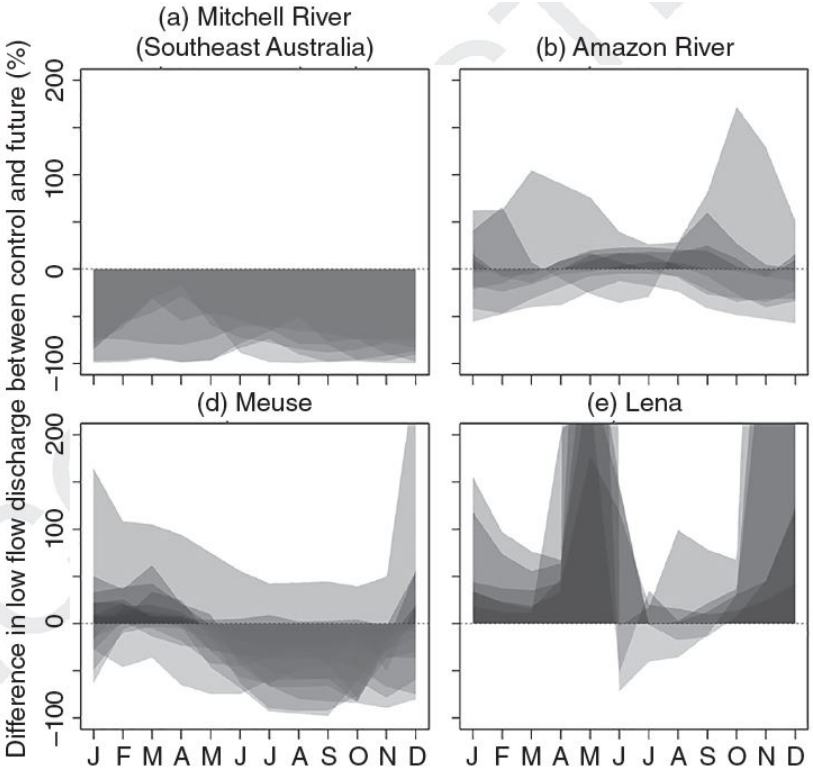


Figure VI-1: Relative change in low flows (Q80) between future period and control period for representative river basins across the globe obtained from a multi-model experiment (Van Huijgevoort et al., 2014). Positive means an increase in future low flows, whereas negative implies the opposite. Shading indicates model agreement (the darker the more models agree).

Single large - scale models were still used to investigate specific drought aspects. For example, Van Lanen et al. (2013) and Wanders and Van Lanen (2015) introduced the

similarity index, based on the bivariate probability distributions of the drought duration and deficit volume. A single hydrological model was also used to illustrate how the European river network will be affected by future drought (Forzieri et al., 2014). This model was also applied to explore the impact of human influences on future drought. The Inter - Sectoral Impact Model Intercomparison Project (ISI - MIP) was introduced as a follow - up to WaterMIP. In ISI - MIP, the most recent emission scenarios, RCPs, were used, and, in addition to the natural hazard, inter - sectoral impacts were assessed. Prudhomme et al. (2014) and Wanders et al. (2019) provide a comprehensive overview of future drought (multi - RCPs, multi - climate models, multi - hydrological models), including an uncertainty measure, and the influence of large - scale models and climate models. Single hydrological models were applied to study human impacts on future drought – for example, the influence of reservoirs (Wanders and Wada, 2015), or the adaptation to a gradually changing hydrological regime (Wanders et al., 2015).

References

Arnell, N.W., 2003: Effects of IPCC SRES* emissions scenarios on river runoff: a global perspective, Hydrology and Earth System Sciences 7: 619–641, doi:10.5194/hess-7-619-2003.

Corzo Perez, G.A., Van Lanen, H.A.J., Bertrand, N., Chen, C., Clark, D., Folwell, S., Gosling, S., Hanasaki, N., Heinke, J., and Voss, F., 2011: Drought at the Global Scale in the 21st Century, WATCH Technical Report No. 43, 117 pg., available at: http://www.eu - watch.org/publications/technical - reports/2.

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., and Bianchi, A., 2014: Ensemble projections of future streamflow droughts in Europe. Hydrology and Earth System Sciences 18: 85–108, doi:10.5194/hess-18-85-201418-85-2014.

Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat., P., Tallaksen, L.M., Warnaars, T., Wiberg, D., Weedon, G.P., van Lanen, H.A.J., Ludwig, F., and Haddeland, I., 2011: Water and Global Change (WATCH) special collection: current knowledge of the terrestrial Global Water Cycle, Journal of Hydrometeorology 12(6): 1149–1156, doi: 10.1175/JHM-D-11-024.1.

Prudhomme, C., Giuntoli, I., Robinson, E.L., Clark, D.B., Arnell, N.W., Dankers, R., Fekete, B.M., Franssen, W., Gerten, D., Gosling, S.N., Hagemann, S., Hannah, D.M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., Wisser, D., 2014: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proceedings of the National Academy of Sciences 111(9): 3262–3267, doi:10.1073/pnas.1222473110.

Van Huijgevoort, M.H.J., Van Lanen, H.A.J., Teuling, A.J., Uijlenhoet, R., 2014: Identification of changes in hydrological drought characteristics from a multi - GCM driven ensemble constrained with observed discharge, Journal of Hydrology 512: 421–434, doi.org/10.1016/j.jhydrol.2014.02.060.

Van Lanen, H.A.J., Wanders, N., Tallaksen, L.M., and Van Loon, A.F., 2013: Hydrological drought across the world: impact of climate and physical catchment structure. Hydrology and Earth System Sciences 17: 1715–1732. doi:10.5194/hess-17-1715-2013.

Wanders, N., Van Lanen, H.A.J., 2015: Future discharge drought across climate regions around the world modelled with a synthetic hydrological modelling approach forced by three General Circulation Models, Natural Hazard Earth Syst. Sci, 15, 487–504, doi:10.5194/nhess-15-487-2015.

Wanders, N., Wada, Y., 2015: Human and climate impacts on the 21st century hydrological drought, Journal of Hydrology 526: 208–220, http://dx.doi.org/10.1016/j.jhydrol.2014.10.047.

Wanders, N., Wada, Y., Van Lanen, H.A.J., 2015: Global hydrological droughts in the 21st century under a changing hydrological regime, Earth System Dynamics 6: 1–15, doi:10.5194/esd-6-1-2015.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., Wood, E.F., 2019: Development and evaluation of a Pan-European multimodel seasonal hydrological forecasting system, J. Hydrometeorology, 20: 99-114, doi: 10.1175/JHM-D-18-0040.1.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., Wood, E.F., 2019: Development and evaluation of a pan-european multimodel seasonal hydrological forecasting system, J. Hydrometeor., 20, 99–114, doi:10.1175/JHM-D-18-0040.1.

Annex VII: Future change of snow load

The risk on snow damage is strongly dependent upon weather and climatological variables. Temperature influences the moisture content, i.e. in the temperature range close to zero degrees the precipitating snow effectively attaches to tree crowns and branches (Solantie, 1994). Therefore, for impacts related from accumulated snow load, the accretion of heavy wet snow poses the greatest risk (Lehtonen et al., 2016). Moderate wind speeds enhance snow accumulation, but on the other hand, strong winds dislodge most of the snow from the tree crowns. And topography also plays an important role, largely because rime accumulation is most efficient in places located higher terrains but also the orographic effects add the precipitation.

The FMI snow load model is an empirical algorithm, based on a tuned parametrization of accumulation and removal terms of estimated snow load for four different snow load types. It has been used to examine the impacts of projected climate change on heavy snow loads on Finnish forests by Lehtonen et al. (2016). For snow load modelling, they used daily data from five global climate models under representative concentration pathway (RCP) scenarios RCP4.5 and RCP8.5, where they statistically downscaled the modelled values onto a high-resolution grid using a quantile-mapping method. The derived results suggest that the projected climate warming regionally causes asymmetric response on heavy snow loads. In eastern and northern Finland, the annual maximum snow loads on tree crowns were seen to increase during the present century, as opposed to southern and western parts of the country. The change was rather similar both for all the snow load types (heavy rime, wet snow and frozen snow), only the heaviest dry snow loads were projected to decrease over almost the whole of Finland. They concluded that the risk for snow-induced forest damage is likely to increase in the future in the eastern and northern parts of Finland. The increase is partly due to the increase in wet snow hazards but also due to more favourable conditions for rime accumulation in a future climate that is more humid but still cold enough (Lehtonen et al., 2016). An example of the projected results is shown in Figure VII-1.

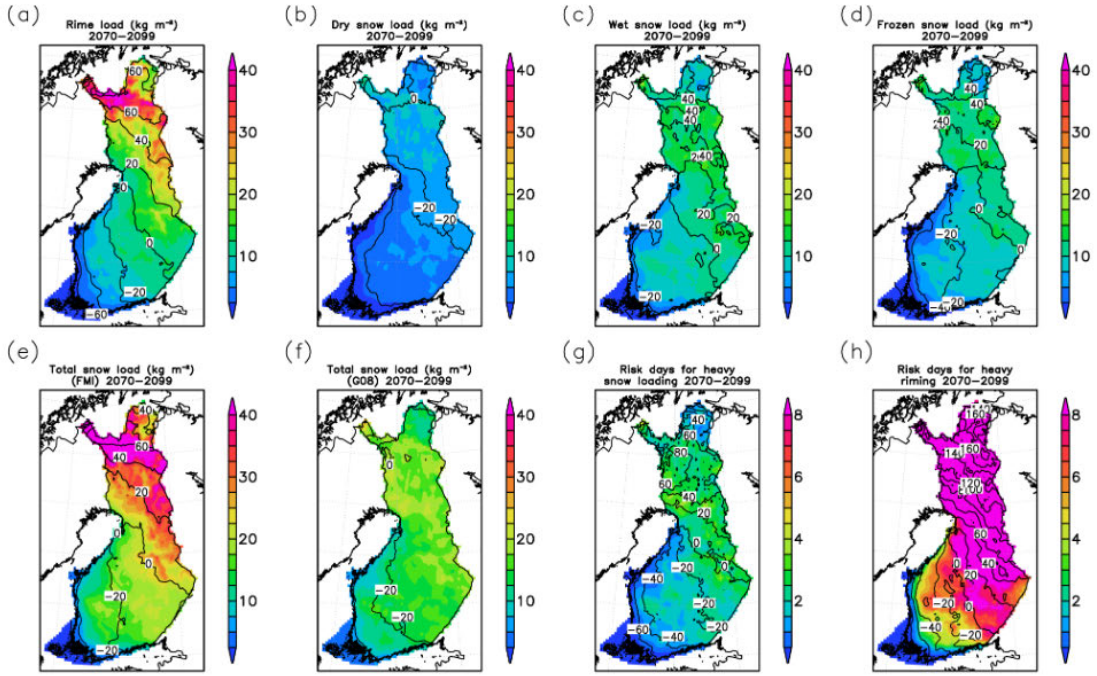


Figure VII-1: The annual maximum rime loads (a), dry snow loads (b), wet snow loads (c), frozen snow loads, and (d) total snow loads based on the FMI method and (e) total snow loads based on the (Gregow et al. 2008) method (f) for the period 2070–2099 under the RCP8.5 scenario as a multi-model mean. Contours show the multi-model mean change from 1980–2009 to 2070–2099 (Lehtonen et al. 2016).

Annex VIII: TRL Scale in Horizon 2020

The Technology Readiness Level (TRL) scale was introduced into the EU funded projects arena in 2014, as part of the Horizon 2020 framework program (European Commission, 2014). If you are planning to submit a Horizon 2020 project proposal, even more so if your project is technology oriented – this article is for you. It is important to understand the exact implications of this scale on your project and the ways it is used to present, evaluate and measure the progress of a Horizon 2020 or ERC project. In this article, we will discuss exactly that. Let's start by understanding the fundamentals of the TRL scale and how it works.

What is the TRL Scale?

The Technology Readiness Level (TRL) scale was originally defined by NASA in the 1990's as a means for measuring or indicating the maturity of a given technology. The TRL spans over nine levels as follows:

TRL 1 Basic principles observed

TRL 2 Technology concept formulated

TRL 3 Experimental proof of concept

TRL 4 Technology validated in lab

TRL 5 Technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)

TRL 6 Technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)

TRL 7 System prototype demonstration in operational environment

TRL 8 System complete and qualified

TRL 9 Actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)

Typically, many products go through the various stages of the TRL scale in their life cycle. It is possible that iterations will be needed between various TRL levels, especially during the development phase, although not limited to that. The TRL is perceived as an effective way to indicate the development stage of a given technology or product.