multi-Risk sciEnce for resilienT commUnities undeR a changiNgclimate

Codice progetto MUR: **PE00000005** – **I33C22006910006**

Deliverable title: Identification the contexts and methods for uncertainty assessment for hazard indicators

Deliverable ID: 8.5.1

Due date: 30 June 2024

Submission date: 30 June 2024

AUTHORS

Alberto Montanari (UNIBO), Warner Marzocchi (UNINA), Alice Crespi (EURAC), Elena Toth (UNIBO), Fabio Ricci Feliziani (FS), Gianpiero Cossarini (OGS), Maria Vittoria Struglia (ENEA), Marco Marani (UNIPD), Marco Reale (OGS), Monica Ponte (ITALFERR), Paolo Lazzari (OGS), Pierpaolo Alberoni (ARPAE), Stefano Salon (OGS), Tommaso Diomede (ARPAE), Mattia Neri (UNIBO)

with the contribution of the RETURN Uncertainty Interspoke Task Force

Technical references

Project Acronym	RETURN
Project Title	multi-Risk sciEnce for resilienT commUnities undeR a changiNg climate
Project Coordinator	Domenico Calcaterra UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II domcalca@unina.it
Project Duration	December 2022 – November 2025 (36 months)

Deliverable No.	DV8.5.1
Dissemination level*	PU
Spoke	DS8 - Science underpinning climate services for risk mitigation and adaptation. Note: what is presented in the deliverable is the result of a cooperation involving all the RETURN spokes.
Work Package	WP5 - Uncertainty assessment for climate and weather scenarios
Task	T 5.1 - Validation of the processed predictions and evaluation of their performance for multiple time scales T 5.2 - Definition of a statistical method to assess uncertainty of hazard indicators for the historical period
Lead beneficiary	UNIBO
Contributing beneficiary/ies	ENEA, EURAC, CIMA, OGS

* PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

Document history

Version	Date	Lead contributor	Description
0.1	21.06.2024	Alberto Montanari (UNIBO)	First draft
0.2	28.06.2024		Critical review and proofreading
0.3	29.06.2024		Edits for approval
1.0	30.06.2024		Final version

ABSTRACT

This deliverable presents the activity until the midterm of the project related to uncertainty assessment for hazard and impact indicator. Uncertainty assessment is a central theme for environmental risk estimation. During the activity of RETURN, through interspoke discussions, the topic has been identified as an ideal ground for interspoke collaboration, in view of the above general interest and the recognition of its key role for the mitigation of risk, adaptation and emergency management. Therefore, a task force has been activated to solicit interspoke collaboration on the subject. The RETURN Interspoke Task Force for Uncertainty Assessment puts together more than 40 researchers from all the RETURN Spokes, with the presence of expert in communication, philosophy, social sciences and technical emergency management. The Task Force met regularly during one year of activity and set the ground for a synergy of the interdisciplinary knowledge on the topic. In particular, the goal of the task force is to identify context sensitive methods for uncertainty assessment to compose a modular and flexible "RETURN workflow for uncertainty assessment and communication". A necessary requirement for the workflow is the capability to adapt uncertainty assessment to different operational context in terms of information availability, technical needs, target audience and social impact. The first part of this deliverable describes the setup of the RETURN workflow. The second part proposes a first operational methods for uncertainty assessment and validation, and related public domain software. Such method refers to the case of flood prediction and is part of the activity of Workpackage 5 "WP5 – Uncertainty assessment for climate and weather scenarios" of spoke DS8 "Science underpinning climate services for risk mitigation and adaptation". In particular, the activity is part of tasks "T 5.1 - Validation of the processed predictions and evaluation of their performance for multiple time scales" and "T 5.2 - Definition of a statistical method to assess uncertainty of hazard indicators for the historical period".

1. Table of contents

2. List of Figures

represent two hypothetical "true" distributions of the data, that in applications are usually unknown. The vertical error bars represent the probability of exceedance observed for the two cases (the amplitude of the bars is determined by the number of observations). In the right panel the exceedance probability distributions obtained by each model are depicted (histograms), along with the distribution of epistemic uncertainty that is obtained by fitting the histogram with a Beta distribution. The dashed green and blue curves refer to the observation of such frequencies for the cases A and B, respectively. One can see from this representation that case A is compatible with the prediction models while case B is not.

Figure 3. Return Workflow for Uncertainty Assessment

Figure 4. Workflow of Bluecat

Identification the contexts and methods for uncertainty assessment for hazard indicators

4 The RETURN Interspoke Task Force on Uncertainty Assessment

The RETURN Interspoke Task Force on Uncertainty Assessment has been formed in Fall 2023. It now includes more than 40 RETURN researchers and meets regularly. So far, 6 meetings have been convened.

The purpose of the first phase of the work has been the identification of common operational grounds and problems related to the topic. The researchers presented their research questions, the types of uncertainty they usually consider and an overview of the methods that they are currently applying. The discussion has been also extended to RETURN spoke leaders and ambassadors, thus generating a truly interdisciplinary dialogues.

The RETURN Interspoke Task Force on Uncertainty Assessment is producing a document in progress to make a synthesis of its work, that is leading to the elaboration of a RETURN workflow for uncertainty assessment with the purpose to guide scientists and end users in the selection of the most appropriate methods and strategy for communication. The task force is also producing a glossary of terms to identify a recommended taxonomy to support interdisciplinary dialogue and communication.

The task force is assuming that a prediction model is selected and used to support risk management in a multidisciplinary context. The prediction model can be used for different purposes, ranging from real time prediction of a given variable, the offline prediction of an expected value during a considered future time horizon, prediction of missing values into a data record, prediction of future statistical features for variables and so on. It is also assumed that uncertainty of predictions is to be evaluated. A brief synthesis of the work in progress is presented in what follows.

5 A note on terminology

In different disciplines a different terminology is used. The RETURN Uncertainty task force is compiling a glossary to introduce an agreed reference in RETURN.

For the sake of clarity, we would like to make clear that we use here the term "prediction" with the following meaning that is copied from the Cambridge dictionary:

> "**a statement about what you think will happen in the future**".

Wikipedia (last accessed on July 19, 2024) provides the following definition:

> "A **prediction** (Latin *præ-*, "before," and *dictum*, "something said"[1]) or **forecast** is a statement about a future event or about future data. Predictions are often, but not always, based upon experience or knowledge of forecasters. There is no universal agreement about the exact difference between "prediction" and "estimation"; different authors and disciplines ascribe different connotations.
>
> Future events are necessarily uncertain, so guaranteed accurate information about the future is impossible. Prediction can be useful to assist in making plans about possible developments."

Note that the term "forecast", in some disciplines, is used with a different meaning with respect to "prediction".

Details on the RETURN glossary – that is still under development – will be presented in a dedicated deliverable.

6 Types of uncertainties

With regard to the identification of uncertainties that are being considered, they can be summarized in the following types:

- Data uncertainty, that is in turn related to data quality, sample size and data completeness.
- Model uncertainty, which includes parameter uncertainty and model structural uncertainty. Model uncertainty is related to the estimation of the weight to be assigned to each modelling response when computing the average of an ensemble simulation.
- Uncertainty defined as the difference between observed and simulated data.
- Uncertainty in the initial conditions, which may be incorporated into model uncertainty.

The above uncertainties affect the mathematical description of the processes and the prediction of the considered variables. It is interesting to mention that the discussion also focused on the whole spectrum of uncertainties affecting decision making, therefore highlighting the relevant role of uncertainties that are not directly related to mathematical modeling, such as for instance:

- Uncertainty of the decision makes that has to select mutually exclusive alternatives.
- Uncertainty in communication, due to the use in mathematical modelling of complex models based on concepts that cannot be easily communicated, such as probability, non linearity and others.
- Uncertainty due to cognitive biases, such as those discussed by Merz et al. (2015).
- Uncertainty due to a biased perception of scientific methods by the public and decision makers, which may imply that uncertainty is interpreted in a non correct manner.
- Uncertainty in legal implications of decisions and erroneous interpretation of scientific results by responsible authorities.

While models for describing different types of environmental risk are markedly different depending on the considered threat (earthquakes, volcanic eruptions, landslides, floods, etc.) they often deal with similar uncertainties that are described in the above list. Therefore, similar methods for uncertainty assessment can be applied in different disciplines to describe different threats.

7 Methods for uncertainty assessment

Each specific discipline dealing with environmental risk adopts different predicting models, ranging from heuristic approaches, stochastic and deterministic models. Uncertainties in the mathematic description can be dominated by data or model uncertainty. However, the discussion highlighted that several similarities can be identified in uncertainty assessment methods, that often differ each other only for the solutions that have been adopted to deal with information scarcity or treatment of different statistical behaviors of predictions and observations.

In particular, in several disciplines uncertainty is considered to be better described through probabilities, while other disciplines prefer a qualitative description through the definition of a "confidence" in the results. While for given types of threat uncertainty in prediction is described in aggregated form through the definition of one single probability distribution, in other cases attempts are made to describe each source of uncertainty individually, or to classify uncertainties into groups (like "epistemic" and "aleatory" uncertainty).

Regarding the selection of prediction models and/or uncertainty assessment methods, the RETURN Interspoke Uncertainty Task Force agreed that:

- Within the context of RETURN, the selection of the most appropriate modelling solution for obtaining the desired prediction – deterministic, stochastic or empirical – should be driven by the available information and the target of providing end users with useful technical indications.
- Similarly, model evaluation and uncertainty assessment should take into account the specific target of the prediction. The goal is to minimise the risk of failure of the operational management of extreme events.
- In particular, the implications of false positives (false alarms) or false negatives (missed alarms) should carefully evaluated in cooperation with end users.
- Model testing and uncertainty assessment testing is particularly important for the operational use of the procedures.

Some models may directly provide the prediction and related uncertainty assessment in a direct manner. Furthermore, given that a "true" model is never available, in some application several predictive models may be considered, which may differ for model structure and/or parameters. In some cases, like for instance hurricane tracking, the different models are based on the same physical concepts and equations but adopt different semplifications of the model structure that are necessary to obtain a solution that provides predictions within a manageable time frame. The range of the output encompassed by the several considered models depicts an uncertainty interval that resembles the difference between the considered modeling hypotheses. Such type of uncertainty is called by some authors "epistemic uncertainty", namely, an uncertainty that refers to the uncertainty of the model (epistemology is the study of knowledge).

In some disciplines, like for instance hydrology, the use of one model only is frequent and is dictated by different theoretical assumptions or operational needs. In such cases "epistemic uncertainty" is estimated within an integrated solution by aggregating it to other uncertainty types. For instance, in some application the difference between model prediction and observed values is assumed to resemble the sum of all uncertainty types, which is sometimes called “global uncertainty”.

When multiple models are adopted, their output may be synthesized in different ways that we can summarize as follows:

1. A best model is selected and used. Such model may be selected basing on its operational efficiency.
2. A combination is used. Such combination may be obtained through a weighted average, or other solutions. Weights may be estimated through statistical testing, expert knowledge or others.
3. The prediction is expressed through the range encompassed by all models, usually represented through a probability distribution. According to such solution, all the models are taken into account when presenting the results.

An example of an application following the procedure #3 presented above is given by Figure 1 (courtesy by A. Viglione) which delivers the estimate for the annual maximum peak river flow corresponding to a return period of 500 years, computed by different candidate probability distributions.

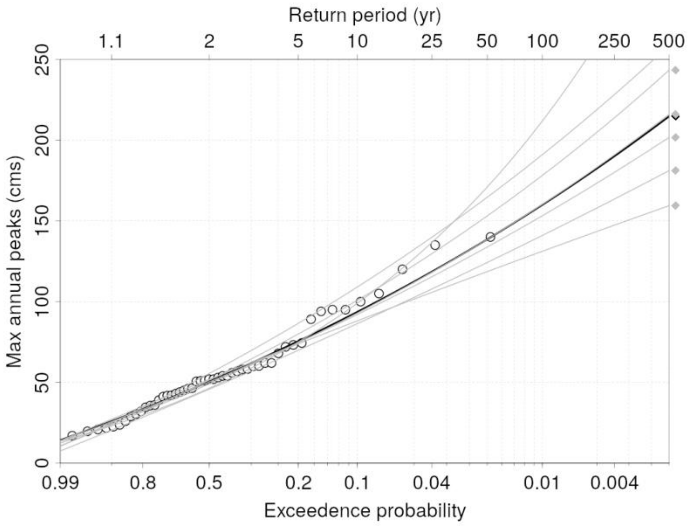

Figure 1. Estimation of annual maximum peak river flow corresponding to a return period of 500 years (probability of exceedance of 0.2%), computed by different candidate probability distributions (grey lines). Dots represent the observations that have been used to fit the distributions. Note that in this case epistemic uncertainty is given by the range of the considered probability distributions and not by any single distribution. In other words, the fact that a prediction model is probabilistic is not a sufficient condition for ensuring the model itself delivers an estimate of uncertainty associated to the prediction.

The solution that is chosen among the three identified above dictates the most appropriate procedure for model testing. If the prediction is given by a single model and uncertainty of the prediction is estimated in a probabilistic manner, several options for model and uncertainty testing are offered by the literature (see, for instance, Gneiting and Katzfuss (2014); Laio and Tamea (2007); Koutsoyiannis and Montanari (2022)).

If sevaral predictive models are considered, model testing becomes more complicated but – as argued by some authors – more informative if we suppose that the true observation lies within the prediction range encompassed by the different models. For instance, figure 2 (courtesy by W Marzocchi) show san hypothetical example illustrating that considering several models may lead to a more consistent estimate of uncertainty.

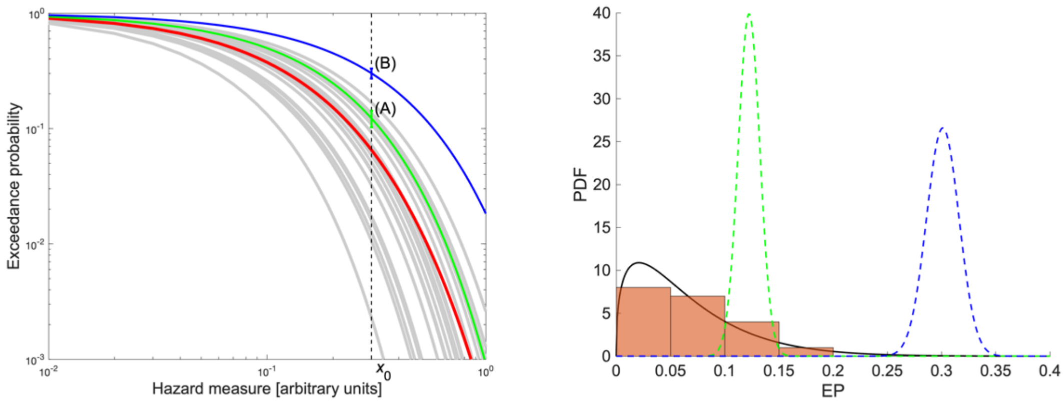

Figure 2. Schematic representation of different uncertainty types through several models and their role for model testing. In the left panel the grey curves represent different probabilistic models. The red line represents the average model. The cases A (in green) and B (in blue) illustrate two hypothetical situations: the continuous green and blue lines represent two hypothetical "true" distributions of the data, that in applications are usually unknown. The vertical error bars represent the probability of exceedance observed for the two cases (the amplitude of the bars is determined by the number of observations). In the right panel the exceedance probability distributions obtained by each model are depicted (histograms), along with the distribution of epistemic uncertainty that is obtained by fitting the histogram with a Beta distribution. The dashed green and blue curves refer to the observation of such frequencies for the cases A and B, respectively. One can see from this representation that case A is compatible with the prediction models while case B is not.

The discussion on methods for uncertainty assessment is expected to deliver a set of solutions that are currently adopted in the different disciplines, to be used within the RETURN Workflow of Uncertainty Assessment that is presented here below.

8 The RETURN Workflow for Uncertainty Assessment

The RETURN Task Interspoke Force on Uncertainty Assessment has recently presented a workflow to provide indications for uncertainty assessment in a multi environmental risk context. The workflow has been presented at the RETURN dissemination workshop in Bari (see the attached presentation).

The RETURN Workflow for Uncertainty Assessment is presented in Figure 3.

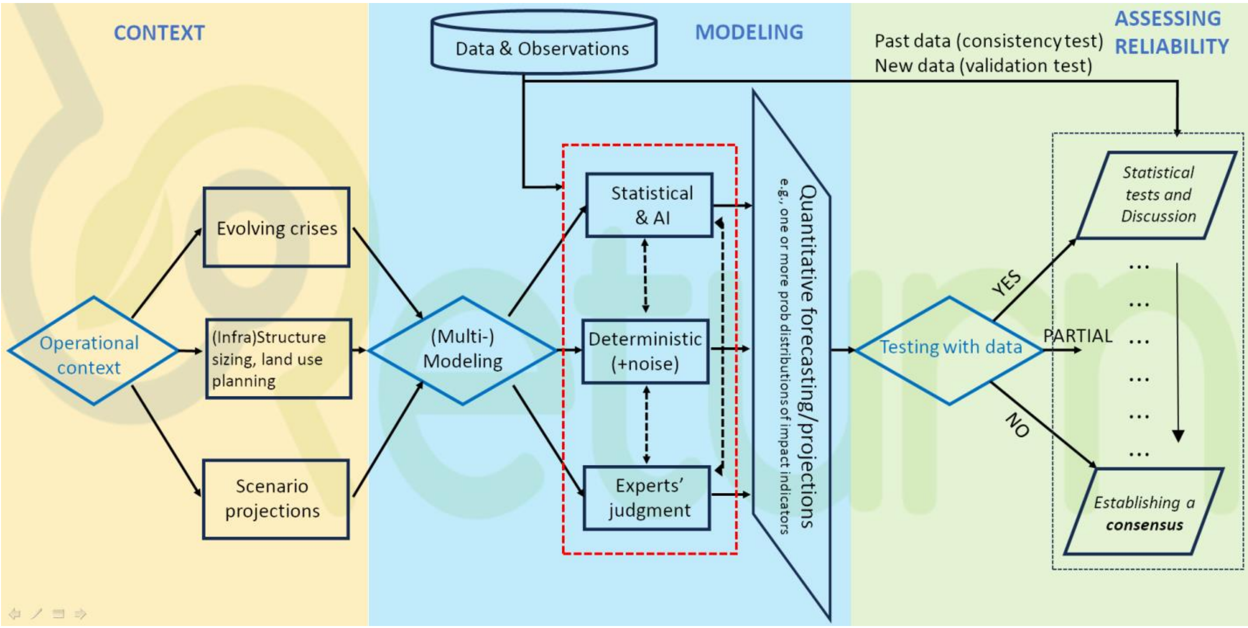

Figure 3. Return Workflow for Uncertainty Assessment

The workflow has been designed for scientists/technicians/operators that aim at producing hazard and risk estimations. It could be important for decision-makers and communicators to understand, in general terms, how models are built and how their reliability is estimated.

In figure 3 rhombuses, rectangles and parallelograms indicate decision, task and output, respectively. The workflow first identifies the operational context on the left, with the three alternatives that have been identified in RETURN. Then, the modeling part is depicted in the middle, which presents in three distinct sectors the prediction model(s), methods for uncertainty assessment and the output. Prediction model and uncertainty assessment may be incorporated into a single modeling framework, as we mentioned above. Testing is located in the third part on the right, by distinguishing between the situations of data availability and lack of data.

As mentioned above, the workflow is intended to be associated with a review of predictive models and testing models from the relevant literature, to support the selection of procedures that look more appropriate for the target of RETURN. In the next chapter an example of uncertainty assessment method along with a testing procedure is provided.

9 An example: the BLUCAT method (and software) for uncertainty assessment

We provide here a first example of a recently developed method that may be useful for the purpose of RETURN. An open software is available for the application of the method, which can be downloaded from GitHub (see below) which, however, is conceived for the specific application of river flow forecasting. A dedicated version of the software, to be applicable to the general case of the prediction of any variable, is being prepared under RETURN. It is relevant to note that the method is based on the statistical analysis of the difference between predictions and observations and therefore, strictly speaking, can be applied only when observed data – in the form of a sufficiently long sample – are available and can be compared to predictions.

The method is called Bluecat, an acronym which stands for "Brisk Local Uncertainty Estimator for Deterministic Simulations and Predictions". In essence, Bluecat is a method to transform a deterministic prediction model into a stochastic prediction model, therefore turning from a point prediction to the probability distribution of the predictand. From the latter distribution, a mean (or median) prediction is obtained along with the confidence bands. Therefore Bluecat performs two tasks:

- It updates the deterministic prediction, therefore providing a new point prediction;
- It provides confidence limits for the prediction for an assigned confidence level.

Therefore, Bluecat is not only an uncertainty assessment method: it is rather a prediction model with a stochastic structure, and eventually a physical basis that is rooted into the deterministic model. Bluecat can be applied in conjunction with any deterministic prediction model. The workflow of Bluecat is depicted in Figure 4, which refers to the case of rainfall-runoff modeling.

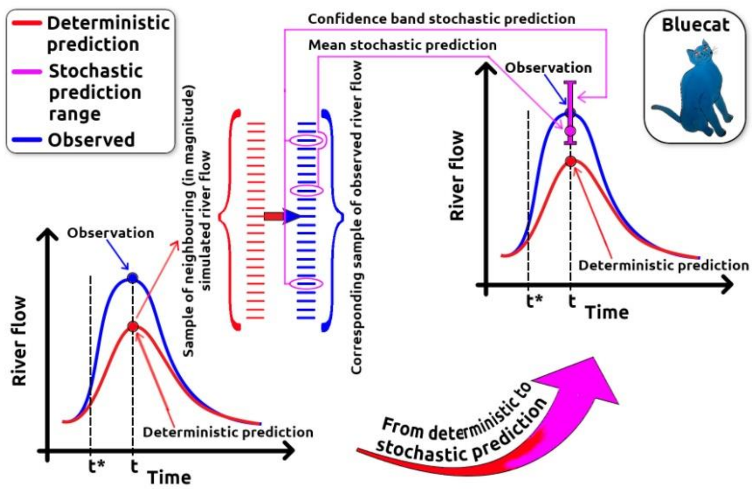

Figure 4. Workflow of Bluecat

Bluecat is rooted in the bluprint for process-based modeling of uncertain hydrological systems proposed by Montanari and Koutsoyiannis (2012). Bluecat is described in full detail By Koutsoyiannis and Montanari (2022, paper available here).

Bluecat is a simple and transparent tool to transform point predictions obtained by any deterministic model in stochastic predictions, therefore deriving the probability distribution of the predictand. In what follows, we will use the terms "D-model" and "S-model" to denote the deterministic model and its stochastic counterpart, respectively.

The information that is needed to perform the above transformation is obtained in Bluecat by building on the well established concept of comparing the D-model output with observed data; namely, the same concept that we commonly use for parameter estimation. Basing on such comparison, Bluecat estimates the probability distribution of observed data conditioned on the D-model output and therefore obtains the corresponding S-model output, along with its mean (or median) value and confidence band. It is important to make clear that the S-model prediction may be markedly different from the D-model one. In fact, the latter is not necessarily included into the confidence band of the S-model, which are displaced around the mean prediction of the S-model itself. Such possible outcome is schematically represented in Figure 4, where the concept of Bluecat is depicted.

Being based on the comparison between the D-model output and the observations, Bluecat is therefore transparent and easily understandable, while the theoretical development ensures that such interpretation of uncertainty is rigorous and asymptotically consistent in estimating global uncertainty.

Bluecat is based on the following main assumptions:

1. A single D-model is considered, with a single parameter set.
2. The stochastic processes describing the modeled variables are stationary during the calibration and application period. Non-stationarity can be accounted for by using non-stationary D-models.
3. The calibration data set is extended enough to ensure that sufficient information is available to upgrade the D-model into the S-model.

The flow chart of the procedure for applying Bluecat is as follows (see Figure 4):

1. The D-model is calibrated by using observed data;
2. At the prediction time t* the D-model is run to produce a prediction Q(t) at time t;
3. A set of predicted variables from the calibration data set, including the one with the smallest difference from Q(t) plus some lower and some greater in magnitude of it, is extracted;
4. From the obtained sample of corresponding true observations the mean (or median) prediction and the confidence band for assigned confidence level from the S-model are estimated by fitting a proper probability distribution to the identified observed data set.

Thus, the S-model performs an adjustment of the D-model to compensate its inability to fully reproduce the observed reality.

A software package was prepared to apply Bluecat in the R environment. The software applies Bluecat in conjunction with the HyMod rainfall-runoff model (Boyle, 2000). Therefore, the software includes a routine to run HyMod and optimise its parameters. Extension to any other prediction model and any other predicted variables is being prepared in RETURN. The software comes with a detailed help function, along with data sets and detailed instructions to apply Bluecat to reproduce the case studies presented in the Bluecat paper. The software is open source and fully commented. Click here to access the software in GitHub.

Further information can be retrieved from:

- A Powerpoint presentation of Bluecat.
- A Powerpoint presentation of Bluecat in the Italian language.
- A presentation of Bluecat at the EGU22 General Assembly.

A video presentation of Bluecat is available in YouTube which uses the above Powerpoint presentation as a supporting media (actually the video shows a previous version of the Powerpoint presentation.

10 Conclusions

For more details on the work of the RETURN Interspoke Task Force please refer to the attached presentation. We summarise here below the conclusions from such presentation:

- Uncertainty is always present in several different forms. **Uncertainties are many but they are not obscure**.
- Assessing and **understanding uncertainty** is **just as important as making a prediction**!
- **Uncertainty cannot be eliminated but can be understood**. We often believe we know it and how to deal with it, but more often we do not.
- **Our target**: to provide end users with the means to recognise the type and magnitude of uncertainties they are dealing with. Provide a review of methods and conditions to ensure reliability.
- **Next step**: communication of uncertainty.

RETURN is an exciting opportunity to better handle uncertainty in a multirisk context. The work that is now being developed focuses on the preparation of open source software to be applied for uncertainty assessment and testing in RETUTN. The software will be made available in GitHub.

11. References

Boyle, D. (2000). Multicriteria calibration of hydrological models (Unpublished doc- toral dissertation). Univ of Arizona, Tucson.

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*(1), 125-151.

Koutsoyiannis, D., & Montanari, A. (2022). Bluecat: A local uncertainty estimator for deterministic simulations and predictions. *Water Resources Research*, *58*(1), e2021WR031215.

Laio, F., & Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, *11*(4), 1267-1277.

Merz, B., Vorogushyn, S., Lall, U., Viglione, A., & Blöschl, G. (2015). Charting unknown waters—On the role of surprise in flood risk assessment and management. *Water Resources Research*, *51*(8), 6399-6416.

Montanari, A., & Koutsoyiannis, D. (2012). A blueprint for process-based modeling of uncertain hydrological systems. *Water Resources Research*, *48*(9).