D1.3 Impact modelling data requirements and methods to treat data gap filling and data uncertainty.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101093806. The publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.

D1.3: Impact modelling data requirements and methods to treat data gap filling and data uncertainty

Summary

This document presents ICARIA's approach to identify and approach data gaps in climate resilience projects and to keep in line with the latest advances in data-driven and AI methodologies as applied in climate-related projects.

Deliverable number	**Work package**	
D1.3	WP 1	
Deliverable lead beneficiary	**Deliverable author(s)**	**Contributor(s)**
CERTH	Anastasia Moumtzidou (CERTH) Ilias Gialampoukidis (CERTH)	Àlex de la Cruz (AQUA) Beniamino Russo (UPC) Mattia Leone (UNINA) Athanasios Sfetsos (DMKTS) Ioannis Zarikos (DMKTS) Marianne Bügelmayer (AIT) Refiz Duro (AIT) Maria Guerrero Hidalgo (CETAQUA) Alberto Couce Rodriguez (CETAQUA)
Internal reviewer(s)	**External reviewer(s)**	
Mattia Leone (UNINA)	Jose Luis Dominguez (IREC)	
Planned delivery date	**Actual delivery date**	
30/06/2024	30/06/2024	
Dissemination level	**PU = Public** PP = Restricted to other programme participants RE = Restricted to a group specified by the consortium. Please specify: __________________ CO = Confidential, only for members of the consortium	

Document history

Date	Version	Author	Comments
14/11/2023	1.0	Anastasia Moumtzidou (CERTH) Ilias Gialampoukidis (CERTH)	Completed the first version of the document

30/04/2024	2.0	Anastasia Moumtzidou (CERTH) Ilias Gialampoukidis (CERTH)	Completed the second version of the document
14/05/2024	3.0	Anastasia Moumtzidou (CERTH) Ilias Gialampoukidis (CERTH)	Completed the third version of the document
15/05/2024	4.0	Anastasia Moumtzidou (CERTH) Ilias Gialampoukidis (CERTH)	Updated third version of the document
23/06/2024	5.0	Àlex de la Cruz (AQUA) Beniamino Russo (UPC) Mattia Leone (UNINA) Athanasios Sfetsos (DMKTS) Ioannis Zarikos (DMKTS) Marianne Bügelmayer (AIT) Refiz Duro (AIT) Maria Guerrero Hidalgo (CETAQUA) Alberto Couce Rodriguez (CETAQUA)	Updated cookbook's structure and context. Comments from internal and external reviewing
29/06/2024	6.0	Àlex de la Cruz (AQUA) Beniamino Russo (UPC) Mattia Leone (UNINA) Athanasios Sfetsos (DMKTS) Ioannis Zarikos (DMKTS) Marianne Bügelmayer (AIT) Refiz Duro (AIT) Maria Guerrero Hidalgo (CETAQUA) Alberto Couce Rodriguez (CETAQUA)	Final version

Table of contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

3CM	Conceptual cognitive concept mapping
BHT	Building height
AAD	Average Annual Damage
ACF	Autocorrelation Function
AHF	Anthropogenic Heat Flux
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
ARIMA	Autoregressive integrated moving average
BAL	Bayesian Active Learning
BBB	Bayes by Backprop
BbH	Bayers by hypernet
BCI	Biophysical Composition Index
BN	Bayesian network
C3S	Copernicus Climate Change Service's
CCL	Certainty-driven consistency loss method
CESM	Community Earth System Model
CH	Climate change and Hazard data
CI	Clinical Infrastructure
CMIP6	Coupled Model Intercomparison Project 6
CNNs	Convolutional Neural Network

CNRM-CM5	Centre National de Recherches Météorologiques Climate Model 5
COICOP	Classification of Individual Consumption by Purpose
CoP	Community of practices
Cordex	Coordinated Regional Climate Downscaling Experiment
COSMO	Consortium for Small-scale modelling
CRI	Completely regularized spline
CS	Case Study
CSF	Case Study facilitator
CSFs	Case Study Facilitators
DCN	Deconvolutional Network
DCS	Depth wise-separable Convolutions
DEUP	Direct epistemic uncertainty prediction
DGPs	Deep Gaussian Processes
DL	Deep Learning
DMP	Data Management Plan
DOI	Digital Object Identifier
DSS	Decision Support System
DSVNP	Doubly Stochastic Variational Neural Process
EBBI	Enhanced Built-up and Bare land Index
EC-Earth	A European community Earth System Model
ECMWF	European Centre for Medium-Range Weather Forecasts

EDU	Positions of Emergency Detection Units
EM-DAT	Emergency Events Database
EO	Earth Observation
EPW	EnergyPlus Weather Format
ERA-40	ECMWF re-analysis of the global atmosphere and surface conditions
ERA5	Fifth generation ECMWF atmospheric reanalysis of the global climate
EUCP	European Climate Prediction
EV	Exposure and Vulnerability data
EW	Equally Weighted combinations
FAIR	Findable, Accessible, Interoperable, and Reusable
FICLIMA	FIC - Fundación para la Investigación del Clima
GBRT	Gradient Boosting Regression
GCMs	General Circulation Models
GDP	Gross Domestic Product
GIS	Geographic Information System
GLMs	General Linear Models
GP	Gaussian Processes
GPI	Global polynomial interpolation
Grad-CAM	Gradient-weighted Class Activation Mapping
GRP	Gross Regional Product
HCAI	Human-Centered AI

HK	Hurst–Kolmogorov
HMS	Harmonized monitoring scheme
HSVR	Hybrid Support Vector Regression
HVAC	Heating, Ventilation, and Air Conditioning
HW	Harmonically Weighted combinations
IBTrACS	International Best Track Archive for Climate Stewardship
ICHEC-EC-EARTH	Irish Centre for High-End Computing model
IDA	ICE IDA Indoor Climate and Energy
IDM	Intelligent Dasymetric mapping
IDW	Inverse Distance Weight
IDW	Inverse distance weighted
IMD	Maps of impervious surface fraction
IPCC DDC	IPCC Data Distribution Centre
IPCC	Intergovernmental Panel on Climate Change
IPSL-CM5A-MR	Institut Pierre Simon Laplace model
KD	Disjunctive kriging
KO	Ordinary kriging
KS	Kolmogorov-Smirnov
KU	Universal kriging
LCZ	Local Climate Zones
LPI	Local polynomial interpolation

LRM	The Land Parameter Retrieval Model
LULC	Land Use Land Cover
MKMK	Metropolis-Kuo-Mallick (MKMK) method
ML	Machine Learning
MLP	Multilayer Perceptron
MOHC-HadGEM2-ES	Met Office Hadley Centre model
MOS	Model Output Statistics
MPI-ESM-LR	Max Planck Institute model
MSARMs	(Non-homogeneous) Markov switching autoregressive models
MSMC	Markov chain Monte Carlo
MVG	Matrix Variate Gaussian
NADs	Neural Architecture Distribution Search
NCPs	Noise Contrastive priors
NDBAI	Normalized Difference BAreness Index
NDBI	the mean Normalized Difference Built Index
NDUI	Normalized Difference Urban Index
NDVI	Normalized Difference Vegetation Index (NDVI)
NDWI	Normalized Difference Water Index
NIST	National Institute of Standards and Technology
NIWA	National Institute of Water and Atmospheric Research
NN	Neural Networks

NOAA	National Oceanic and Atmospheric Administration
NRFA	National River flow archive
NUSAP	Numeral, Unit, Spread, Assessment and Pedigree
PACF	Partial Autocorrelation Function
PCA	Principal Component Analysis
PNs	Dubbed Prior Networks
PW	Performance Weighted combinations
RACMO2	Regional Atmospheric Climate Model
RCM	Regional Climate Model
RDM	Research data management
RegCM	ICTP Regional Climate Model
RPN	Residual-Predicting Network
RS	Remote Sensing
RW	Random Walk
S-BGNN	Subjective Bayesian GNN
SCF	Seasonal Climate Forecasts
SDII	Simple Daily Intensity Index
SEE	Structure Expert Elicitation
SHAP	SHapley Additive exPlanations
SI	Simple kriging
SLANG	Stochastic, Low-rank, Approximate Natural Gradient

SMB	Surface Mass Balance
SOTA	State-of-the-art
SQUAS	Stochastic Quantized Activation Distributions
SSO	Specific Sub-objectives
SSPs	Shared Socioeconomic Pathways
SVF	Sky View Factor
SVM	Support Vector Machines
SWA	Stochastic Weight Averaging
SWMM	Storm Water Management Model
TIN	Triangular Interpolation Network
TMYx	Typical Meteorological Years
TPS	Thin Plate Spline
UHI	Urban Heat Island
VSCN	Virtual Climate Station Network
WDF	The Wet Day Frequency
WISC	Windstorm Information Service
WP	Work Package
WRF	Weather Research & Forecasting Model
XAI	Explainable Artificial Intelligence

Executive summary

This document presents the ICARIA cookbook and offers a set of alternative options and advanced methodological tools for identifying and optimizing the treatment data gaps, and a list of sources for collecting any further potentially available data of relevance to conducting climate risk and resilience assessment of critical assets. As ICARIA aims to establish a robust and efficient framework for climate adaptation and resilience, such a document aims for serving as a scaffold for lab tests within the project for application and extension of the currently applied methodologies. Nonetheless, selecting an optimal or combination of various methodologies can often prove perplexing. It is far from uncommon to encounter circumstances where data scarcity emerges as a consistent challenge, especially for assets and services sensitive to the initial amount of available data and bounded by restrictions in accessibility. Consequently, the role of the existing climate data in fully characterizing the overall risk/impact assessment methodology from a single or multi-hazard perspective, supporting the selection of optimal methodologies, identifying data gaps, and ultimately guiding and contributing to the formulation of effective decision-making policies amidst the exigency of climate change, remains critical. Further, when confronted with issues in climate adaptation and resilience, it is critical to consider that data gaps introduce substantial uncertainties in impact assessments, vulnerability analysis, and the design of adaptation strategies, thereby hindering the development of effective resilience measures capable of invariance to various contexts. To support addressing the critical challenge of data gaps, within ICARIA's goals, the design of a supplementary cookbook is included, aimed at assisting practitioners operating within ICARIA in identifying alternative methodologies for data creation and augmentation in supporting the applied viable methodologies being employed in cases where the need to extend and support the currently chosen ones, remains necessary. In section 2, the cookbook attempts to include briefly some of the methodologies applied in ICARIA so far, a list of currently identified data gaps, and provide suggestions for downscaling techniques, hazard assessment, exposure analysis, vulnerability evaluation, and strategies to address data gaps, exploring alternative methods and concepts relevant to the application of climate-related methodologies in resilience and adaptation. To support any further lab tests within ICARIA, a template table for treating data gaps is proposed. Furthermore, it includes sets of additional data sources for assets and services, as examples for further application and reproduction in other areas of interest. While the cookbook covers different methodologies, the sequence is not paramount; rather, the relevance to the core resilience domains and compatibility with the ICARIA project are the primary considerations. In section 3, a domain user survey is conducted utilizing input from experts with diverse backgrounds, offering additional information on current data-driven methodologies related but not limited to climate change and adaptation, and additional resources for the practitioners of the ICARIA project. Finally, in section 4, a summary with reflections on treating data gaps is offered, summarizing in a condensed manner the current knowledge and state-of-the-art.

This deliverable is the first result of T1.3 with inputs from T1.1, T1.2, and T1.4, in WP1 and T2.1 in WP2.

1 Introduction to project ICARIA

The number of climate-related disasters has been progressively increasing in the last two decades and this trend could be drastically exacerbated in the medium- and long-term horizons according to climate change projections. It is estimated that, between 2000 and 2019, 7,348 natural hazard-related disasters have occurred worldwide, causing 2.97 trillion US$ losses and affecting 4 billion people (UNDRR, 2020). These numbers represent a sharp increase of the number of recorded disaster events in comparison with the previous twenty years. Much of this increase is due to a significant rise in the number of climate-related disasters (heatwaves, droughts, flooding, etc.), including compound events, whose frequency is dramatically increasing because of the effects of climate change and the related global warming. In the future, by mid-century, the world stands to lose around 10% of total economic value from climate change if temperature increase stays on the current trajectory, and both the Paris Agreement and 2050 net-zero emissions targets are not met.

In this framework, **Project ICARIA** has the overall objective to promote the definition and the use of a comprehensive asset level modeling framework to achieve a better understanding about climate related impacts produced by complex, compound and cascading disasters and the possible risk reduction provided by suitable, sustainable and cost-effective adaptation solutions.

This project will be especially devoted to critical assets and infrastructures that are susceptible to climate change, in a sense that its local effects can result in significant increases in cost of potential losses for unplanned outages and failures, as well as maintenance – unless an effort is undertaken in making these assets more resilient. ICARIA aims to understand how future climate might affect life-cycle costs of these assets in the coming decades and to ensure that, where possible, investments in terms of adaptation measures are made up front to face these changes.

To achieve this aim, ICARIA has identified 7 Strategic Subobjectives (SSO), each one related to one or several work packages. They have been classified according to different categories: scientific, corresponding to research activities for advances beyond the state of the art (SSO1, SSO2, SSO3, SSO4, SO5); technological, suggesting and/or developing novel solutions, integrating state-of-the art and digital advances (SSO6); societal, contributing to improved dialogue, awareness, cooperation and community engagement as highlighted by the European Climate Pact (SSO7); and related to dissemination and exploitation, aimed at sharing ICARIA results to a broader audience and number of regions and communities to maximize project impact (SSO7).

- SSO1.- Achievement of a comprehensive methodology to assess climate related risk produced by complex, cascading and compound disasters
- SSO2.- Obtaining tailored scenarios for the case studies regions
- SSO3.- Quantify uncertainty and manage data gaps through model input requirements and innovative methods

- SSO4.- Increase the knowledge on climate related disasters (including interactions between compound events and cascading effects) by developing and implementing advanced modeling for multi-hazard assessment
- SSO5.- Better assessment of holistic resilience and climate-related impacts for current and future scenarios
- SSO6.- Better decision taking for cost-efficient adaptation solutions by developing a Decision Support System (DSS) to compare adaptation solutions
- SSO7.- Ensure the use and impact of the ICARIA outputs

2 Objectives and context of Deliverable 1.3

The document presents the creation of a complementary cookbook and the results of a domain user survey conducted within the ICARIA project as a contribution to the strategic sub objective SSO3. - Quantify uncertainty and manage data gaps through model input requirements and innovative methods (WP1, WP1 - Project framework, climate scenarios, and modelling inputs). Specifically, this document corresponds to Deliverable 1.3 and includes the results of Task 1.3 - modelling input requirements and methods to treat data gaps and uncertainties.

The necessity for a cookbook and a domain survey

Climate change impacts are produced by complex interactions, and are often characterized by compound events and cascading effects that demand tailored-made, sustainable adaptation solutions. To support the ICARIA holistic asset-level modelling framework defined in D1.1 (ICARIA, (2023a)), this deliverable introduces a set of methodologies and tools to treat data gap and uncertainties with respect to modelling input requirements relevant for the case study areas, which has been designed as a "cookbook" potentially transferable to other contexts. Additionally, a dedicated domain user survey has been designed to specifically highlight critical data gaps recurring in the field of hazard and impact assessment. The necessity of both is preparatory, providing merely a scaffold for potential support and improvement of the current state-of-the-art methodologies already employed in ICARIA, which will be then tested in D1.4 (ICARIA, 2024c) for a number of selected data gaps in the case study areas, and further applied in WP4 to cover all relevant data gaps emerging from the Trial execution.

The main goals of D1.3 are summarized as follows:

- Identification of recurring data gaps in hazard/impact assessment: this ambitious task is achieved through a twofold approach. Firstly, ICARIA Case Study Facilitators focused on data gaps related to the implementation of Trials and Mini-Trials in the case-study areas. In parallel, experts were engaged via the domain survey, to provide feedback on data-driven techniques so that the practitioners can comprehensively understand the current limitations and prioritize which data gaps need to be addressed with priority.
- Creation of a list of alternative methods: to supplement the methodologies identified with respect to the key variables and datasets emerging from the impact assessment modelling architecture defined for the ICARIA case studies and based on the proposed holistic modelling framework (see D1.1, Section 3), and to investigate and identify the possibility for filling of data gaps via data augmentation/substitution techniques and/or expert knowledge collection.
- Improvement of data granularity supporting quantitative impact assessments: due to the observed lack of comprehensive datasets that can support assessments at local level (i.e., sub-regional, urban level), both concerning climate change and hazard assessment for complex events, both for exposure and vulnerability analyses, alternative routes need to be explored, including but not limited to the creation of synthetic data through statistical approaches, dynamical downscaling and/or expert elicitation methods. The latter, in particular, offers a unique opportunity to address recurring data gaps in multi-hazard

assessment (e.g., probability of occurrence of coincident compound events, probability of hazard transition in consecutive compound events and/or cascading effects scenarios), detailed quantitative exposure and/or vulnerability analysis (e.g., recurring construction typologies in a given area; health implication of different heat stress levels on specific age groups) that cannot be treated otherwise for the lack of data.
- Exploration of advanced data-driven methodologies: The enormously expanding production of data through remote sensing and research at global level, increasingly available through public data repository and open web platforms, suggests the effectiveness of approaches aimed at harnessing/handling information and addressing data gaps through data-driven methodologies supported by machine learning, AI and data fusion techniques. Similarly, consolidated statistical methods (including geospatial statistical methods) and dynamic downscaling approaches, help to expand the application potential of available data at global/EU level to produce quantitative impact assessments at local scale.

The specific objectives of D1.3 can be summarized as follows:
- to provide a list of possible data gaps in relation to Hazard (H), Exposure (E), and Vulnerability (V) as key elementary bricks of the ICARIA holistic impact assessment model
- to develop a cookbook providing a list of methodologies and technical specifications for filling data gaps in impact assessment, focusing on priorities emerging from the case studies modelling workflows identified in D1.1
- to provide templates to map data gap-filling and uncertainty treatment, including expert elicitation and user-provided data, in Trials and Mini-trials
- to conduct a domain survey about existing and emerging data-driven methodologies
- to provide the results of the domain survey as a form of recommendation for the practitioners.

Structure of the document

The document is organized as follows: in Section 3, the main data gaps in relation to the ICARIA holistic modelling framework as emerged from Case Study Facilitators assessment and the results of the domain user survey are presented. In Section 4, the main methodologies used in ICARIA to treat data gaps and uncertainties are introduced, illustrating the "*cookbook*" structure and the Jupyter book. In section 5, dedicated tables list selected literature and reference studies with respect to the identified methodologies, including those preliminarily applied in Lab Task T1.4, for their potential adoption in WP4 for Trials implementation. In section 6, the domain user survey, developed as a dedicated questionnaire for experts internal or external to the ICARIA consortium, is presented. The survey is designed to include current and emerging trends in data-driven methodologies, and the section includes the results from the testing made by ten (10) selected experts. Finally, in the Conclusions, a reflection on the importance of treating data gaps and uncertainties even beyond the "*gap-filling*" issue, but rather in relation to the correct communication of hazard/impact modelling results to inform decision makers and practitioners is introduced, aimed at supporting the presentation of impact assessments and the identification of suitable resilience measures within CoP activities in WP5. The Appendix includes the main relevant open data repositories useful to support data-gap filling in ICARIA Trials.

3 Recurring data gaps in impact assessment modelling

ICARIA modelling framework (see D1.1) is aimed at quantifying impact from complex events, implying compound hazards and cascading effects conditions, with respect to multiple assets. Such multi-hazard and multi-receptor focus increase the complexity of assessing Hazard, Exposure and Vulnerability variables in time with the adequate spatial resolution to provide quantitative impact assessment information to support resilient planning and decision-making. Therefore, addressing data gaps implies their mapping across Trials and Mini-Trials, both to fill gaps and treating uncertainties, both to acknowledge in the assessment the assumptions and limitations related to data and their elaboration through modelling.

The templates provided in Annex (introduced in D1.1 to map for each Trial main data types expected input from T1.2, WP2 and WP3) have been developed to map the relevant data required the implement the hazard characterization, the exposure and vulnerability analysis and the risk/impact assessment in Case Study and followers' regions. They also include an "event tree scenario building tool", adapted from SNOWBALL project (Zuccaro et. al., 2018), intended to provide a visual representation of the specific modelling workflow(s) adopted in the studies and useful to support data-gap filling and uncertainties treatment (e.g., to i.e., determine hazards transition probabilities through expert elicitation procedures).

The ICARIA cookbook and the Jupyter platform introduced in Section 5 have been developed to provide references and technical specifications to address data gaps, including methods tested in T1.4.

Based on the analysis from domain user survey and contributions from Case Study Facilitators (CSFs), the main critical data gaps can be grouped in two main categories:

- <u>Climate Change and Hazard data</u>, which determine the boundary conditions for hazard characterization in space and time, including aspects long term variation and seasonal trends of climate change patterns, extreme events frequency and intensity, local downscaling of hazards (i.e., with a spatial resolution higher than that derived by Regional Climate models (RCMs)), probability of occurrence of coincident compound events, probability of transition among natural hazards in consecutive compound events, probability of transition in cascading effects from a given triggering hazard impacting critical service networks (e.g., transport, energy, water distribution).
- <u>Exposure, Vulnerability, and Impact data</u>, which allow to determine expected impacts on exposed risk receptors based on specific vulnerability and impact assessment models input requirements. Even considering the diversity of data input required by different exposure, vulnerability, and impact models, recurring datasets can be identified with respect to the main hazard types.

Climate change and Hazard data

Table 1: Climate change and Hazard data gaps' table.

Data gap	Description	Data-gap filling approach	Source
Weather/Climate data not fully covering spatially the CS area considered	Thin Plate Spline (TPS) methodology	Spatial interpolation techniques for distribution of climate parameters	CSFs, and Domain user survey
Weather/Climate data presents gaps in its time series, (i.e., lack of some daily registers for the variables considered).	For each station, the observations located at a distance of less than 20 km and with a correlation of more than 0.7 are selected. In case there are less than 6 observatories that meet these requirements, the radius of proximity will be recursively extended by 5 km until there are 6 nearby observatories. Of the six (6) observatories that meet the above requirements, the three (3) with the highest number of data between 1979 and 2020 are selected. A multiple linear regression is performed with the data that the observatory to be filled and the three (3) observatories selected above have in common. With the parameters of the linear regression, the gaps of the standard observatory that are empty and whose data from the three (3) selected observatories are complete are filled in. No gaps are filled for a given day if any of the three (3) observatories with which the linear regression has been performed have no data for that day. All the previous points are repeated six (6) times with the objective that, in each iteration, more and more gaps are filled in. In the first three (3) iterations, gaps will be filled for the whole series, while in the last three (3) iterations, only gaps between the years 1979 and 2020 will be filled.	"Closest-correlated neighbor" weather/climate data gap-filling methodology	CSFs, and Domain user survey

Data gap	Description	Data-gap filling approach	Source
Weather observations are too short, either not reaching the minimum years required for its use or not having data back in time enough for some modelling.	For some procedures, the use of a temporally homogeneous weather dataset is mandatory, with data covering (without gaps) the same extension of time. For this reason, and since most of the weather observations are quite recent, a temporal extension of weather data is performed thanks to the use of a climate reanalysis, the ERA5-Land in this case. Since observations and reanalysis do not always correlate perfectly (event with improved experiments), checking and testing of correlations are done locally between ERA5-Land and a set of regional stations. To further improve the reanalysis, a set of parameters is obtained and used to correct the reanalysis against the regional weather signals. On the other hand, this reanalysis is crossed with the weather observations, and thanks to the use of several transfer functions, the ERA5-Land, in the temporal extension desired, is corrected to properly reflect the true weather signal of the point, constructing, therefore, a simulated extended observation for the desired point. Last, this corrected simulated observation is crossed and filled with the original observation in those spots available, obtaining the original checked weather observation with gaps filled with a corrected reanalysis up to the date desired.	Weather data temporal extension methodology	CFSs, and Domain user survey
Weather/Climate observations are used for statistical downscaling presenting outliers, errors, or missing data.	A two-way quality control methodology is applied generally to all weather data sources prior to their use. 1. Basic consistency. Direct rejection of self-evident wrong values: for example, negative values for precipitation. 2. Atypical values or 'outliers'. Unusual values within a data set: values that could come from different sources of data or values that could have been generated differently from the rest of the data. In this case, the theoretical difficulty of their recognition depends on our definition of "atypical". In practice, the recognition is generally referred to values whose absolute magnitude is unusually high. Also conduct statistical tests	Weather/Climate Data Quality Control	CFSs, and Domain user survey

Data gap	Description	Data-gap filling approach	Source
	(e.g., Standard Z-Score method, Modified Z-Score coupled with decomposition, the Exponential Moving Average with the Coupled Modified Z-Score and decomposition)		
Weather/Climate observations used for statistical downscaling presenting changes in trends due to modifications of instrumentations or locations, among others.	The way to proceed with the homogeneity test that we have used is based on a method that detects gaps inhomogeneities in daily data (Monjo et. al., 2013): 1. To measure how similar is data belonging to one year to data belonging to another year, it is used a distribution comparison test based on the Kolmogorov-Smirnov (KS) test. The KS test is a non-parametric statistical test (it does not presuppose distributions of the studied variable) which provides a p-value that can be used as a measurement of the similarity between two years. Values that are close to 0 show that two years have a value distribution very similar and we can infer that there is not an inhomogeneity between them. The lower the value for Log (KS), the greater the probability of inhomogeneity between two consecutive values. 2. If one year has been selected as a possible indicator of inhomogeneity, then it is subjected to another test ("Similarity between years"). Once we select the year that possibly presents an inhomogeneity and the following one, we figure out the p-value of every year of the series with respect those two years. If a jump or a break shows up between all those p-values in the years that we are considering, then we can infer that there is a true inhomogeneity for all the series.	Weather/Climate Observations Homogeneisation techniques	CSFs, and Domain user survey
Hazard downscaling at high resolution (<250m)	Climate change affects urban areas unevenly depending on local conditions. Both heat and flood extreme events' hazard intensity magnitude can greatly vary depending on specific local factors, such as urban morphology, surface materials and vegetative cover, location of critical assets and components of service networks (e.g., transport, energy, water distribution). Most hazard models adopted in ICARIA include the	Remote sensing data can be used as proxy of heat / flood hazard (e.g., Land Surface Temperature as proxy of UHI, soil imperviousness as proxy of flood	CSFs

Data gap	Description	Data-gap filling approach	Source
	assessment of such variables to quantify impacts at local level and support resilient planning and decision-making. In order to capture the effect of urban infrastructure on hazards and potential cascading effects, a Land Use Land Cover (LULC) database mapping their geospatial distribution should include information with adequate resolution to provide hazard models with the needed input parameters.	hazard) Data fusion methods combining multiple open datasets	
Probability of occurrence of coincident compound events	The occurrence of coincident compound hazards in a given area can be assessed through joint probability approaches. The analysis must be linked to historical data related to the respective hazards to identify possible correlations and supported by expert elicitation. Different methods are proposed in ICARIA (see D2.2), such as Spearman's rank or Pearson coefficient, to determine hazards correlation from past events datasets analysis. Once this step is completed, joint probability can be assessed through statistical methods such as copula models, which allow to analyze the relationship/dependence between variables, or Monte Carlo simulations, where a large number of scenarios based on their individual probability distributions and correlations can be run. The final step Analysis can then be carried out on the frequency that both these hazards occur simultaneously (or sequentially) to estimate their joint probability.	Statistical methods (joint probability), supported by open datasets (e.g., database of past events, such as EM-DAT) and expert elicitation methods	CSFs
Probability of occurrence of consecutive compound events and/or of cascading effects	The occurrence of consecutive compound hazards in a given area can be assessed through conditional probability approaches. The analysis can be linked to historical data related to the respective hazards to identify possible correlations and supported by expert elicitation. Recurring approaches include Bayesian methods to assess conditional probabilities of transition among consecutive compound events and/or cascading effects, and the use of conditional probability tables to illustrate the relationships.	Statistical methods (conditional probability), supported by open datasets (e.g., database of past events, such as EM-DAT) and expert elicitation methods	CSFs

Exposure, Vulnerability, and Impact data

Table 2: Exposure, Vulnerability, and impact data gaps' table.

Data gap	Description	Data-gap filling approach	Source
Missing information at local scale (i.e., urban area)	Detailed exposure and vulnerability analyses require high-resolution information based on specific impact models requirements. Such information can be integrated in a GIS platform with the LULC database to optimize data exchange and interoperability. Recurring data gaps by hazard type can be summarized as follows: Heat waves - Buildings: building envelope (walls + windows) S/V ratio; construction typologies; HVAC system type; Outdoor spaces (artificial and vegetated): albedo, emissivity, shading conditions, SVF, evapotranspiration, surface temperature; Population: spatial distribution; age; income. Floods (pluvial/coastal/river) - Buildings: ground/underground level permeability; Outdoor spaces (artificial and vegetated): urban watershed relative altimetry; proximity of run-off streams; drainage capacity by land use; Service networks: sewer capacity (including manholes status); Energy network: primary/secondary cabins; distribution network; Population: spatial distribution; Property (structure + content): opening ratio at ground floors; underground floors; sidewalk / steps to enter ground floors. Droughts - Land cover (natural, agricultural, urban green); vegetation inventory; local water supply sources; resistance of plant types. Forest fire - Buildings: building envelope (walls + windows); construction typologies; Outdoor spaces (artificial and vegetated): Land cover (natural, agricultural, urban green); vegetation inventory; local water supply sources; resistance of plant types; Population: spatial distribution; age; income.	Data interpolation through statistical methods based on available open data sets; Expert elicitation	CSFs

4 Methodologies for data-gap filling and uncertainties treatment

Main methodologies in ICARIA Cookbook

The ICARIA Cookbook provides a series of "*recipes*" that include references and technical specifications to address data-gaps and uncertainties with respect to the ICARIA Holistic modelling framework. The main methods used in ICARIA, widely consolidated in literature, are summarized in the following.

Statistical end dynamical methods

Climate Models (GCMs) are numerical models that represent the climate system with varying degrees of complexity and are based on the physical, chemical and biological properties of its components, their interactions and feedback processes. Each GCM represents all components of the earth system (atmosphere, cryosphere, biosphere, ocean, ice-sheets), as well as human impacts via greenhouse gas emissions and simulates possible future climate states. By representing all components, also their interactions are depicted (e.g., melting of sea ice changes the ocean's salinity and albedo, in turn affecting ocean's temperature, which then affects the atmospheric temperature). These models provide important information; however, their spatial resolution is relatively coarse (e.g., 100 km x 100 km, meaning 1 temperature, precipitation etc. value for a grid box of 10.000km^2). To address this limitation, downscaling techniques are employed. In this sense, ICARIA has tried to tackle this uncertainty by not sticking to one but considering the two main sources of generation of information at the local scale available: dynamical and statistical downscaling. ICARIA has incorporated into its procedures these two downscaling methods.

Statistical downscaling obtains empirical relationships between large-scale variables from GCMs and high-resolution (surface) variables, allowing us to obtain very local results (like in a village) with less error than the dynamical one. The statistical downscaling methodology applied in ICARIA by FIC, named FICLIMA (Ribalaygua et. al., 2013), consists of a two-step analogue/regression statistical method which has been used in national and international projects with good verification results. The first step is common for all simulated climate variables and it is based on an analogue stratification. For the second step, the procedures applied depend on the variable of interest, ranging from multiple linear regression in temperature, clustering of rainfall most analogous days for precipitation, or transfer functions between probability distributions and parametric bias corrections for wind or RH. This methodology was applied in ICARIA for the three case studies using 10 GCMs and the 4 Tier 1 SSPs (1-2.6, 2-4.5, 3-7.0 and 5-8.5).

Dynamical downscaling, on the other hand, increases the resolution of the GCMs over the region of interest with RCMs, taking into account local characteristics and altering physical processes, allowing us to obtain results in areas (like watersheds) where there are no observed data as well as a better representation of atmospheric processes. For the so-called GCM to better represent local features such as topography or land use, the output of the general circulation models is used to drive regional climate models (ARSINOE Project, 2023; Nikulin et. Al., 2018). Regional climate models represent atmospheric processes at spatial resolutions of ~1 - 12 km. This so-called dynamical downscaling was applied within ICARIA for 2 SSP scenarios (SSP126, SSP585), using two kinds of regional climate models (Weather Research and Forecasting Model (WRF) and Consortium for Small-scale modelling (COSMO) for Salzburg and South Aegean region.

On the other hand, there's the uncertainty problem about climate information. Efforts within the scientific community focus on addressing and quantifying uncertainties in climate simulations. Within climate projections two kinds of uncertainties are discussed, first the scenario uncertainty, and second the model uncertainty. The first one relates to the fact that we don't know which emission scenario will become reality until 2100, thus the temperature evolution until the end of the century is uncertain, but within the simulated ranges (low to high emissions). The latter represents the fact that for the same emission scenario, two models might yield opposing trends, which is the case for precipitation. As highly complex processes of different temporal and spatial scales are at play for precipitation, its correct representation within climate models is still subject of research. Even though we already know a lot, the approaches taken within two different models might cause opposing trends of precipitation amounts until 2100. Within ICARIA, each downscaling methodology assesses its own uncertainty with inner processes of verification through the use of different procedures and statistics, ensuring that the methodology introduces the least amount of uncertainty into the outcomes. As a result, both methods are then combined following the ensemble strategy, displaying the different outcomes and impacts for future climate states. Often the medians and quantiles are applied to gain a better knowledge and reduce uncertainty, enhancing the understanding of future climates for specific locations.

Data-driven and data fusion methodologies

Data-driven methodologies offer an additional tool for climate resilience and especially for defining, estimating, and treating data gaps (Harder et. al., 2022; Reichstein et. al., 2019; Andrychowicz et. al., 2023). While it finds application in a majority of fields, the ICARIA project and D1.3, include data-driven methods for studying data gaps for climate resilience. These methods guide us in filling data gaps in time series forecasting at large scales, in addressing weather data, in expanding hazard datasets by combining input from inventories (e.g., before and after an extreme event), in enhancing large-scale quantitative hazard assessments, etc... Further, the integration of public data with asset characteristics (e.g., buildings), helps to estimate and generate representative values for characteristics when data gaps appear, providing key input for improving critical response strategies and risk assessments from regional to national scales. Data-driven methodologies can play a key role as an auxiliary tool for data-gaps treatment while additionally providing complementary information to real-world experiments, filling geospatial data gaps about infrastructure, and expanding the output of climate models.

Expert elicitation methods

Expert elicitation is a structured procedure for obtaining uncertainty judgments from experts, measuring their individual judgment capabilities with a performance-based metric, and then applying mathematical scoring rules to combine these individual judgments into a 'rational consensus' that can inform the deliberations of policy-makers. One of the most widely adopted elicitation methods is the "*classical model*" formulated by Cooke (Cook et. al., 2018). The classical model has been developed to aggregate expert judgments based on performance measures, and is based on the scoring of expert judgment in terms of statistical accuracy and informativeness. The statistical accuracy, representing the "calibration" of expert's opinion, is tested through a number of questions for which the answer can be confronted with observed values and expressed by a probability index (the so called "seed variables", whose value is known to the analysts at the moment of the elicitation but are not known to the experts at the moment of the elicitation). Experts are thus scored according to their performance in assessing seed variables. A low value (near zero) expresses a high accuracy. The product of statistical accuracy and informativeness for each expert is their combined score, expressed as Performance Weighted (PW) combinations. Other assessment values can be derived from elicitation, such as Equally Weighted (EW) or Harmonically Weighted (HW) combinations, as well as individual expert assessments.

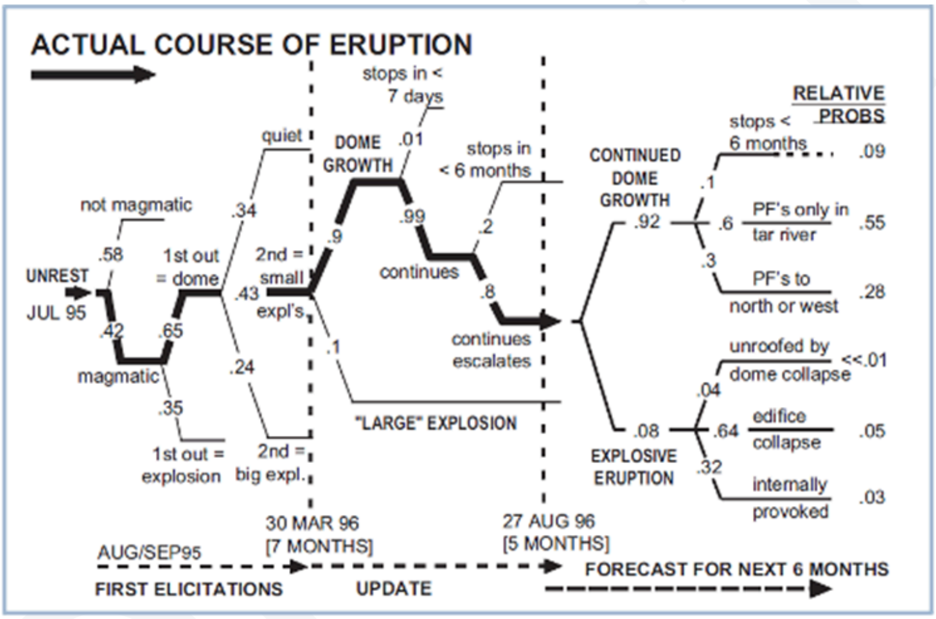


Figure 1: Simplified volcanic eruptive scenario event tree, incorporating probabilities of occurrence of different eruptive events derived from successive expert elicitations (Mader, 2016). Probabilities of hazard transitions are derived from expert elicitation.

The application of expert elicitation methods is particularly appropriate to determine target variables characterized by significant level of uncertainty, which cannot be sufficiently described using current models or field data, but for which a rational consensus among experts can be reached. Based on the Cooke's classical model, several expert elicitation approaches and supporting tools have been developed, and are characterized by the following common features:

1. Scrutability: All data and processing tools are open to peer review and results must be reproducible by competent reviewers.
2. Empirical control: Quantitative expert assessments are subjected to quality controls.
3. Neutrality: The method for evaluating expert opinions should encourage experts to state their true opinions.
4. Fairness: Expert opinions are not judged, prior to processing the results of their assessments.

Uncertainty treatment methods

The Climate-ADAPT uncertainty guidance (Zuccaro et. al., 2018) highlights the many levels of uncertainties associated to climate change impacts and adaptation: Future emissions trajectories of greenhouse gases and aerosols, which are influenced by demographic, economic, and technological developments and international climate agreements, will determine the scale and speed of future climate change. The impact of climate change on the environment and society will be shaped by the future development of non-climatic factors, including socio-economic, demographic, technological, and environmental changes. Measurement errors arise from using imperfect observational tools, such as rain gauges, and from data processing methods, like algorithms for estimating surface temperatures from satellite data. Aggregation errors occur due to incomplete temporal and spatial data coverage. Natural variability is driven by unpredictable processes within the climate system, such as atmospheric and oceanic changes, future volcanic eruptions, and dynamics within climate-sensitive environmental and social systems, like ecosystems. Model limitations in climate and impact models stem from their limited resolution, which affects the detailed representation of phenomena like cloud physics, and from an incomplete understanding of individual Earth system components, such as dynamic ice sheets, their interactions and feedbacks, like climate-carbon cycle feedbacks, and environmental or social systems under study, such as demographic changes in flood-prone areas or specific urban morphologies and features of building, open spaces and vegetative cover that affect soil drainage capacities and urban heat island conditions. Lastly, societal preferences and political priorities influence the significance placed on specific climate impacts, such as local or regional biodiversity loss

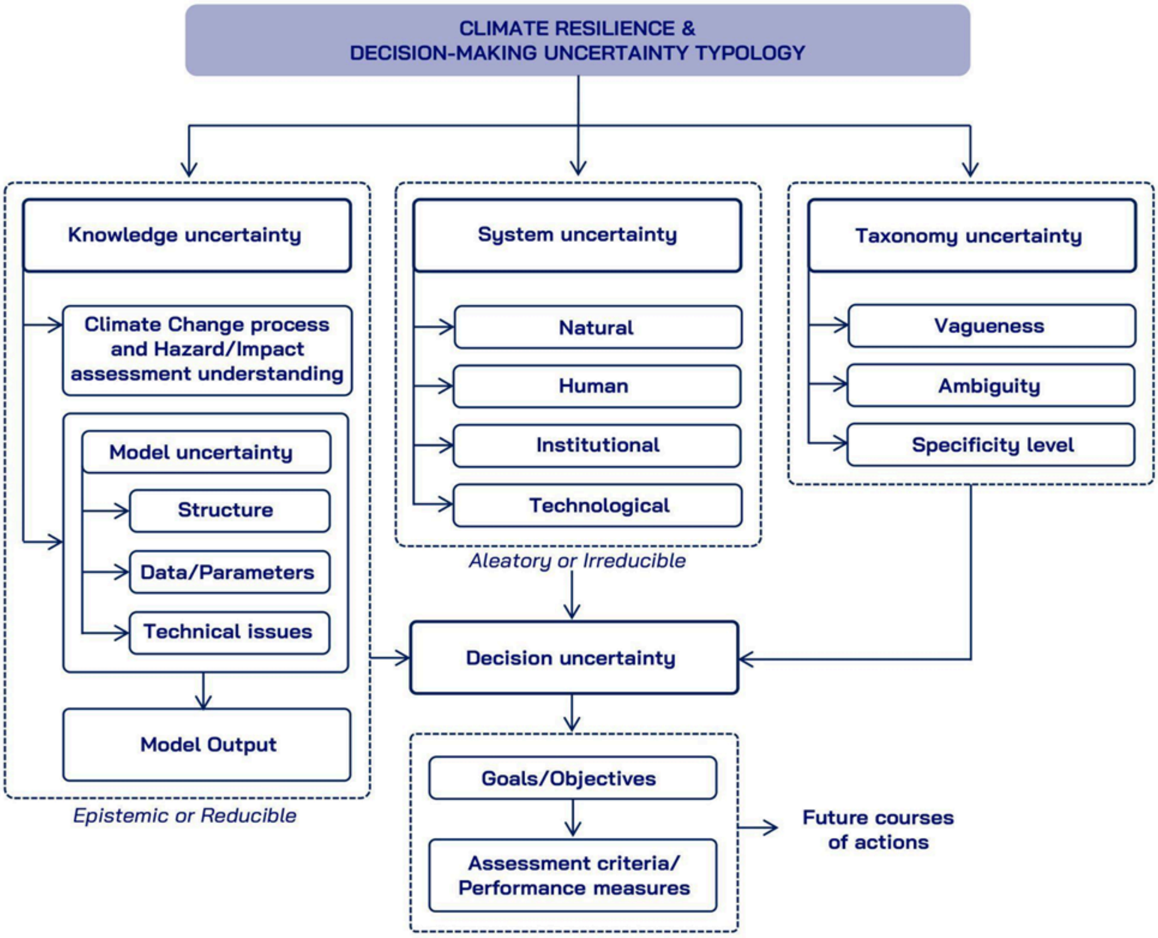


Figure 2: Climate resilience & Decision-making uncertainty Typology based on knowledge uncertainty, system uncertainty, taxonomy uncertainty, and decision uncertainty (modified after Ascough II et al., 2008).

These types of uncertainties can be connected to various areas of "*Knowledge Uncertainty*", "*Variability Uncertainty*" and "*Decision Uncertainty*", as defined by Walker et al. (2003), which lead to the necessity of considering uncertainty as essential component of decision-making for climate resilience. According to Street and Nilsson (2014), recognizing and reflecting the nature and characteristics of uncertainty in the use of evidence leads to better-informed, more relevant, and robust decisions. By acknowledging uncertainties instead of expecting clear-cut outcomes, uncertainties become more manageable, enabling the formulation of coherent decisions and policies. Furthermore, the acknowledgment of uncertainties in hazard/impact assessments contribute to minimizing the risk of maladaptation and to a more effective risk management. In particular, the focus of ICARIA on probabilistic assessment of complex events (compound coincident, compound consecutive, cascading effects, see D1.1), which requires articulated event tree analysis, whose uncertainties in terms of hazard transition probabilities, and/or in terms of likelihood of cascading effects given a certain damage threshold on critical service assets and networks1, may lead to propagation of error in the impact assessment.

Ensemble strategy for ICARIA climate information

In the generation of climate information for ICARIA, one of the primary challenges faced by climate scientists and decision-makers is the inherent uncertainty in climate data (Camps-Valls et. al., 2023; Lenton et. al., 2019). Climate Models (CMs) are numerical models that represent the climate system with varying degrees of complexity, each simulating past and future climate states uniquely, thus introducing a degree of uncertainty based on the selected CM. The climate system itself has inherent variability due to the different time scales of its components (e.g., atmospheric processes occur over days, oceanic processes over years) and their impacts on weather patterns and phenomena like ENSO or AMO. While CMs effectively simulate broad atmospheric circulation, they lack the resolution (around 100 km) for capturing smaller-scale local phenomena, necessitating downscaling techniques that further add uncertainty. Additionally, the emission scenarios (SSPs) used to drive future climate projections introduce another layer of complexity and uncertainty in interpreting and communicating model results and their local impacts. The scientific community addresses and quantifies uncertainties in climate simulations primarily through the ensemble strategy (Zuccaro et. al., 2018), which involves using different models to compute the same SSP scenario. This approach displays various outcomes and impacts for future climate states, highlighting the spread within model simulations and enhancing the understanding of future climates for specific locations. Different procedures (Wilcke et al. 2016) can further reduce uncertainty from ensemble outcomes, such as selecting different ranges of change.

ICARIA tackles this uncertainty not only through the ensemble approach but also by utilizing both dynamical and statistical downscaling methods. Each method assesses its own uncertainty through verification processes using different procedures and statistics, ensuring minimal uncertainty is introduced into the outcomes. By combining these two approaches, ICARIA gains a broader perspective, assessing uncertainties and their implications for future projections. This dual-method approach allows for a better representation of variability and possible future states while being time efficient. Consistent results from both methodologies at the same location enhance the reliability of ICARIA's climate outcomes, providing trustworthy information for case studies and other partners. Conversely, significant differences between the methodologies indicate high uncertainty in future states, dependent on the model used. Once the results from both downscaling methodologies were delivered in D1.2 (ICARIA, 2024a), ICARIA established its ensemble strategy for handling all climate information produced in WP1. This strategy addresses the primary type of uncertainty inherent in the project: the climate information itself.

The ensemble strategy used is derived from the RESCCUE (Velasco et. al., 2018) project. It goes beyond simply using SSPs for CMIP6 by incorporating an impact approach. This approach considers the expected changes for a variable from all potential future scenarios, thereby accounting for uncertainties from downscaling methods, unknown socioeconomic evolutions, and the inherent variability and divergence in climate models.

1. The first step involves analyzing projections related to main variables and hazard indicators for impact modelling. All scenarios (combinations of GCM + SSP) from both downscaling methodologies are considered as an ensemble, resulting in 48 total scenarios for SLZ and SAR (40 from statistical and 8 from dynamical), and 40 for AMB (only statistical).

2. These scenarios are ordered based on expected changes relative to the climate baseline (1981-2020) for future climate periods.
3. An impact approach is then used to identify the "*most probable scenario*" (P50, or median) and the "*worst-case scenario*" (P90). The scenarios in the ensemble strategy are traceable, allowing identification of the specific scenario selected (e.g., the most-probable scenario P50 for TMax corresponds to model MPI-ESM2-1-HR and projection SSP3-7.0).
4.

By proceeding in this manner, uncertainty in the evolution of socioeconomic scenarios is accounted for. Since it is unknown which SSP pathway humanity will follow, it is advisable not to rely on a single SSP. Instead, all SSPs are gathered, and the appropriate one is selected. This approach allows for flexibility, as humanity might not follow, for example, SSP 3-7.0 precisely, but another close scenario with similar results might be more accurate at some point. The selected scenarios will then be used for impact modelling, considering the expected evolution of changes in variables.

For compound events in ICARIA, joint probability in hazard modelling is resolved similarly. All scenarios for each variable in the compound event are considered. The same model + SSP combination for each variable is selected to maintain the inner dynamics of the climate model. Joint probabilities are obtained and sorted by their probability of occurrence, ultimately selecting the median (most likely scenario) and P90 (for uncertainty assessment).

Evaluation of uncertainty in ICARIA's compound events approach
In ICARIA, two methods for the evaluation of the uncertainties connected to the compound events and cascading effects timelines are suggested: (1) Bayesian methods, and (2) Expert Elicitation methods (see also above).

<u>Bayesian methods in uncertainty treatment</u>
Statistics comprises two main competing schools of thought: the frequentist (or classical) approach to statistical inference, which includes hypothesis testing and confidence intervals, and the Bayesian approach. The fundamental difference between these approaches lies in their definitions of probability. A frequentist sees probability as a long-run frequency. When a frequentist claims that the probability of a fair coin landing heads is 1/2, they mean that, over many tosses, the coin will land heads about half the time. On the other hand, a Bayesian, who would also state that the probability of a coin landing heads is 1/2, is expressing a belief about the likelihood of the coin landing heads, perhaps based on the symmetry of the coin suggesting no reason to favor one side over the other. This is known as subjective probability. In practice, frequentists use probability to describe the frequency of specific data types over repeated trials, whereas Bayesians use probability to represent the degree of belief in a statement about unknown quantities (Glickman et al., 2007).

At the core of Bayesian analysis lies Bayes' rule. For two events, A and B, with probabilities P(A) and P(B), respectively, the conditional probability of A given B, P(A | B), can be determined using Bayes' rule:

$$P(A \mid B) = P(A) \cdot P(B \mid A) \,/\, P(B) \tag{1}$$

Bayes' rule allows us to convert a probability like P(B | A) into one like P(A | B), meaning it translates the probability of B occurring given A has occurred into the probability of A occurring given B. In this context, P(A) is referred to as the prior probability, P(B | A) as the likelihood, and P(B) as the normalization factor. Bayes' rule can be straightforwardly extended to random variables and their distribution functions. It can be utilized to combine a prior distribution with a likelihood function to produce a posterior distribution, which can then serve as an input for risk analysis. Bayes' rule can be expressed as:

$$P(\theta | E) = P(\theta)\, P(E | \theta) / P(E) \qquad (2)$$

where P denotes probability mass (or density), θ is a value of the random variable in question (such as the magnitude of a hazard), and E denotes the evidence considered (such as a triggering event). $P(\theta)$ is the prior probability that the random variable takes on the value θ, and its integral over θ is one because it is a distribution. $P(E | \theta)$ is the conditional likelihood function, representing the probability of the evidence given a particular value of θ. Bayes' rule is applied to all values of θ to obtain $P(\theta | E)$, the posterior distribution of θ given the evidence. Both the prior and the likelihood are functions of θ, and Bayes' rule for distributions is essentially their product for each possible value of θ. The normalizing factor is a single value ensuring the resulting posterior distribution integrates to unity.

For most Bayesians, the prior distribution reflects the analyst's opinions or beliefs and represents subjective knowledge before considering specific evidence. It may stem from preconceptions, reasoning, hearsay, or a combination. The likelihood function represents a model, often based on the analyst's subjective knowledge, of what data suggests about the variable in question. The normalizing factor can be difficult to compute analytically, but using conjugate pairs can simplify the problem. If these shortcuts aren't feasible, modern software can handle the computation using intensive methods.

In summary, a typical Bayesian analysis involves the following steps (Glickman et. Al., 2007):

1. Formulate a probability model for the data
2. Decide on a prior distribution, representing the uncertainty in the unknown model parameters before observing the data
3. Observe the data and construct the likelihood function based on the data and the probability model from step 1. Combine the likelihood with the prior distribution from step 2 to determine the posterior distribution, which quantifies the uncertainty in the model parameters after observing the data
4. Summarize key features of the posterior distribution or calculate quantities of interest based on it. These constitute statistical outputs, such as point estimates and intervals.

The main goal of Bayesian statistical analysis is to obtain the posterior distribution of model parameters. The posterior distribution can be seen as a weighted average of knowledge about the parameters before data is observed (represented by the prior distribution) and the information contained in the observed data (represented by the likelihood function). From a Bayesian perspective, almost any inferential question can be addressed through an appropriate analysis of the posterior distribution. Once obtained, the posterior distribution allows for computing point and interval estimates of parameters, prediction inference for future data, and probabilistic evaluation of hypotheses.

Elicitation methods in uncertainty treatment

Expert judgment is sought when significant scientific uncertainty impacts the decision-making process. Due to this uncertainty, experts may not agree. Informally soliciting expert advice is not new, but structured expert judgment aims to apply transparent methodological rules to treat expert judgments as scientific data in a formal decision process. The scientific method itself facilitates expert agreement (Cook et al., 2004). A valid goal of structured elicitation is to quantify, not eliminate, uncertainty in the decision process.

The "*classical model*" (Cook, 1991) used in Snowball methodology is a structured procedure for obtaining uncertainty judgments from experts, measuring their individual judgment capabilities with a performance-based metric, and using mathematical scoring rules to combine these judgments into a rational consensus that informs policy deliberations. The Classical Model method employs proper scoring rules to weight and combine expert judgments based on statistical accuracy and information scores, measured on calibration variables (see Cooke, 1991). It operationalizes rational consensus principles via a performance-based linear pooling or weighted averaging model. The weights are derived from experts' calibration and information scores, measured on seed item calibration variables. Calibration variables serve a threefold purpose (Aspinall et al., 2013):

1. to quantify experts' performance as subjective probability assessors
2. to enable performance-optimized combinations of expert distributions
3. to evaluate and hopefully validate the combination of expert judgments

The name "Classical Model" comes from an analogy between calibration measurement and classical statistical hypothesis testing. In the Classical Model, performance-based weights are determined using two quantitative measures of competency: calibration and information. Calibration assesses the statistical likelihood that a set of experimental results align with the expert's assessments, while information measures how concentrated an expert's uncertainty distribution is.
The main steps of the Classical Model are as follows:

1. **Selection of Experts**: A group of experts is chosen.
2. **Assessment of Seed Items**: Experts express their views as elemental uncertainty distributions and assess a set of variables (seed items) whose true values are known or will become known later.
3. **Scoring Experts' Responses**: Experts' responses are scored based on the statistical likelihood that their distributions over the seed items match the observed or measured results. They are also scored on informativeness compared to a uniform background distribution.
4. **Combining Scores**: The calibration and information scores are combined to form a weight for each expert.

5. **Elicitation of Uncertainty Judgments**: Experts individually provide their uncertainty judgments on the questions of interest (target items).
6. **Weighted Pooling of Responses**: Performance-based or equal weight scores are applied to individual responses to obtain a weighted pooling of uncertainty distributions for each target item.

5 ICARIA cookbook: recipes for data gap filling

This chapter is dedicated to gathering a series of methodologies for data gap filling and data uncertainty methods compiled in a cookbook. It outlines data gap groups, data requirements, data collection templates, and sources, emphasizing the potential replicability of any of the methodologies in case studies or during lab tests. A series of approaches are provided to address data gaps and uncertainty, including but not limited to automated data downscaling, extrapolation, synthetic data generation, etc. focusing on data-driven methodologies and their applicability for addressing data gaps and uncertainty. Additionally, a domain users' survey based on ICARIA's internal and external network is compiled, from experts with diverse backgrounds, collecting key feedback on the current and emerging functionalities of data-driven methodologies that the experts are already using or considered to use. The survey can be treated as a recommendation tool for the CFs, promoting replication of ICARIA results beyond the case studies within the project.

Cookbook structure and general template in Jupyter book

A Jupyter Book[1] is an open-source framework designed to serve as a generator of digital documents and books by integrating Jupyter Notebooks and Markdown files. It enables the seamless unification and presentation of data, code, and narrative text, making it highly suitable for interdisciplinary research and educational purposes. For combining climate resilience methodologies with data-driven techniques, the Jupyter Book will provide a structured environment for rendering extensive datasets, documenting analytical workflows, and coherently delivering results, ensuring replication, strengthening collaboration within the project, and fostering comprehensive dissemination of findings. In ICARIA, a supporting cookbook will be developed by collecting and compiling datasets and methodologies from the literature in a Jupyter notebook. This notebook will mirror the recipes listed in the initial D1.3 document, creating a scaffold for a more rigid understanding of the data gaps, which will later inform the implementation of Trials and Mini-Trials, prioritizing which gaps tend to appear, yielding fruitful results when addressed with representative methods. More specifically the Jupyter book will be organized as follows: each section of the D1.3 document will be systematically transferred to the Jupyter book, with each section receiving its dedicated chapter. For Chapter 3, which enumerates the cookbook's recipes, distinct categories—statistical methods, dynamical downscaling methods, data-driven methodologies, expert elicitation methods, and uncertainty treatment methods—will each be allocated a dedicated subsection containing a detailed list of recipes. Representative information for each recipe, as presented in their designed tables, will be transferred to the appropriate subsection. Furthermore, the domain survey data, including the questionnaire, responses, and a summary of results, will be thoroughly documented. Finally, the chapter referring to the reflections on data gaps and supplementary information from the appendix will be incorporated to ensure a comprehensive and scientific presentation. The Jupyter book will be hosted in a GitHub repository, freely accessible, allowing for continuous updates and extensions of the content, besides easy access and modifications by the case study facilitators. The link for the Jupyter book is the following:

[1] For more information, see here: https://jupyterbook.org/en/stable/intro.html#

https://georgiti.github.io/ICARIA-book/content/recipes/introduction/introduction.html. Finally, an example figure of the landing page of the Jupyter book can be seen below.

Figure 3: Landing page of the ICARIA Jupyter book.

Cookbook "*recipes*"

The following sections contain the technical specifications and tools selected from literature, organized and taking into account the main underlying methodology with respect to those identified in Section 3. Although it is worth noting that the case studies implementing the suggested "recipes" often adopt hybrid approaches, combining multiple methodologies.

Description of Recipe and Identification:

Due to the interdisciplinary character and diverse areas of application of the methodologies, attempting a totally rigid categorization would only add additional confusion, if not, being far from realistic. Thus, the categorization of the recipes within the cookbook was tailored to align with the project's objectives. As a result, a straightforward yet effective way to distinguish each recipe while providing a meaningful description was aimed. Each recipe is labeled using the following format:

Recipe - [Data gap category] [Recipe category numbered] [Secondary Category/Example] [Additional Characterization]

An example: a recipe categorized under downscaling methodologies, listed second, focused on statistical downscaling methods and identified as a review paper The unique label would be the following:
Recipe CH-DD1-R, where:

- "*CH*" denotes Climate Charge and Hazard (or "*EV*" denotes Exposure and Vulnerability data),
- "*01*" represents the unique number,
- "*DD*" representing data-driven related methodologies ("*S*" representing Statistical downscaling related methodologies, "*D*" representing Dynamical downscaling related methodologies, "*EE*" representing Expert Elicitation related methodologies, "*U*" representing Uncertainty related methodologies, and "*HEV*" representing all Hazard, Exposure, and Vulnerability related methodologies), and
- "*R*" is appended to indicate it is a review paper.

Thus, all labels in recipes within the cookbook will follow the same manner, depending on the subsection and category they belong to.

Statistical methods

S1. Long-term daily stream temperature record for Scotland reveals spatio-temporal patterns in warming of rivers in the past and further warming in the future

Table 3: S1's recipe table.

Abbrev.	Categories and data		
CH-S1	Title		Long-term daily stream temperature record for Scotland reveals spatio-temporal patterns in warming of rivers in the past and further warming in the future (Loerke et. al, 2023). [Link]
	Summary		This recipe presents a rigid methodology to estimate long-term daily stream temperature due to the scarcity of available datasets for creating a national daily stream water temperature dataset for Scotland.
		Variables (input)	• Climatic and hydrological variables (e.g., air temperature, etc.). • Harmonized monitoring scheme (HMS) dataset. • National river flow archive (NRFA).
		Methods/Models	• CNNs, GLMs, XAIs. • CNNs to statistically downscaled max and min temperatures over SSA (CNN-R: CNN model with non-linear configuration). • "*CNN showed good skills to produce plausible projections, however, differences with RAW and GLM in the intensity of the signal were identified when non-linearity was considered (CNN-R)*" (*sic*).
		Results/Remarks	• Outputs: (i) Estimation of long-term daily stream water temperature, (ii) Mean monthly and year precipitation maps, (iii) Environmental controls maps on studied variables, (iv) Average base flow index maps, (v) Forecasting of future stream water temperature (long-term daily records, etc.), (vi) Analysis of historical trends as coarser temporal resolution and future changes at high temporal resolution, (vii) Explore the role of controls for individual catchments. • Temperature maps allowed for identification of the sites with the highest temperature

Abbrev.	Categories and data		
			increases, allowing for implementing thermal moderation measures to address this issue.
	Resources	• ERA-Interim analysis. • "*EC-Earth model simulations from the CMIP5 modelling experiment*" (*sic*). • "*Suggestion: model uncertainty and multi-GCM ensembles with prospects to be further utilized in climate change studies over the region*" (*sic*).	
	Keywords	Statistical downscaling; ML; Extreme temperature; GCMs; GLMs; CNNs.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Analyze historical trends and understand the role of environmental controls and the future changes under regional climate projections.

S2. Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach

Table 4: S2's recipe table.

Abbrev.	Categories and data		
CH-S2	Title	Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach (Doury et. al, 2023). [Link]	
	Summary	This recipe offers a hybrid downscaling methodology to extend the high-resolution RCM simulation ensembles at a reasonable cost and identify sources of uncertainty.	
		Variables (input)	• Daily near-surface temperature from EUR11-CORDEX simulations based on the CNRM-ALADIN63 regional climate model driven by the CNRM-CM5 global climate model used in CMIP5. • The historical period runs from 1951-2005. • Scenarios (2006-2100) are based on RCP4.5 and RCP8.5. • 2D: (i) Geopotential, (ii) Humidity, (iii) Temperature, (iv) Ward wind, and (v) Sea level pressure. • 1D: (i) Digital spatial means of 2D, (ii) Daily spatial standard deviation of 2D, (iii) Solar and ozone forces, and (iv) Seasonal

Abbrev.	Categories and data		
			indicators.
		Methods/Models	• The RCM-emulator is based on a fully convolutional neural network algorithm, called UNet. • Evaluation both on perfect model and GCM worlds. • Reproduces the high-resolution spatial structure and daily variability of the RCM. • Issues reproducing accurate simulations of extreme events. • Issues reproducing the complete climate change magnitude. • RCM's general functioning can be broken down into two parts: a large-scale transformation and a downscaling function. • A novel hybrid downscaling approach that emulates the downscaling function of an RCM. • The combination of both empirical statistical downscaling methods and RCMs is part of the novelty of this recipe.
		Results/Remarks	• Transformation from low resolution information to the high resolution near surface temperature. • Aiming at the feasibility to emulate the RCM complexity at high-frequency and high-resolution. • RCM-GCM inconsistencies at large scales. • A high-resolution simulation is provided by the emulator, corrected from the GCM-RCM large-scale inconsistencies.
	Resources	• Transformation from low resolution information to the high resolution near surface temperature. • Aiming at the feasibility to emulate the RCM complexity at high-frequency and high-resolution. • RCM-GCM inconsistencies at large scales. • A high-resolution simulation is provided by the emulator, corrected from the GCM-RCM large-scale inconsistencies.	
	Keywords	Emulator; Hybrid downscaling; RCMs; Statistical downscaling; Deep neural network; Machine Learning; EURO-CORDEX.	
	Tag/Type	Climate Change and Hazard data.	

Abbrev.	Categories and data		
	Application in ICARIA	Potential use in ICARIA	Emulating the downscaling function of an RCM to improve the resolution of climate change data in case study areas.

S3. Bayesian analysis of high-frequency water temperature time series through Markov switching autoregressive models

Table 5: S3's recipe table.

Abbrev.	Categories and data		
CH-S3	Title		Bayesian analysis of high-frequency water temperature time series through Markov switching autoregressive models (Spezia et. al., 2023). [Link] 2023
	Summary		This recipe offers a methodology based on autoregressive models for the estimation of water temperature time series.
		Variables (input)	• River temperature (along with covariates: flow, air temperature, rainfall, wind speed and direction, radiation, and soil temperature).
		Methods/Models	• Methods (i) Bayesian interference, (ii) Markov chain Monte Carlo (MSMC), and (iii) Metropolis-Kuo-Mallick (MKMK) method, and (iv) (Non-homogeneous) MSARMs (Markov switching autoregressive models). • MSARMs allow: (i) Discrete-time stochastic process, (ii) Modelling non-linear and non-normal time series by assuming that different autoregressions, each one depending on a hidden state, alternate according to the Markovian regime switching, and (iii) Classifying the observations into a small number of homogeneous groups, labeled as the regimes of the Markov chain. • Observed state-dependent autoregressive processes driven by an unobserved, or hidden, Markov chain, Markov switching autoregressive models.
		Results/Remarks	• Methods to reconstruct high-frequency time series. • Bayesian model to study the dynamic evolution of water temperature. • MSARMs can be improved using filly Gibbs sampling algorithms avoiding the random

Abbrev.	Categories and data		
			walk Metropolis moves. ● Data augmentation techniques can be further applied to non-homogeneous hidden Markov chains to extend the model.
	Resources	● MSARMs_Codes	
	Keywords	Time-series; Non-linearity; Stochastic variable selection; missing values.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Assess river water temperature variation as source of hazard of natural ecosystems.

S4. High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand

Table 6: S4's recipe table.

Abbrev.	Categories and data		
CH-S4	Title	High-resolution downscaling with interpretable deep learning: Rainfall extremes over New Zealand (Rampal et. al., 2022). [Link]	
	Summary	This methodology tests deep learning techniques against existing statistical approaches for downscaling historical rainfall events.	
		Variables (input)	● Gridded rainfall from Virtual Climate Station Network (VSCN). ● Variables: Consecutive available potential energy, Mean Sea level pressure, Specific humidity, Temperature, Wind, Geopotential height, and Precipitation. ● Time periods: (i) Training period: 1980-2012, (ii) Validation period: 2013–2016, and (iii) Testing period: 2017–2020.
		Methods/Models	● Linear statistical models: Principal component analysis (PCA) for geopotential height and air temperature. ● Candidate predictor variables: Geopotential height, air temperature, zonal wind, meridional wind, wind speed, and geostrophic vorticity. ● Deep learning/Evaluated models: (1) CNNs: (i) Models: Non-linear CNN Gamma, Linear CNN,

Abbrev.	Categories and data		
			Non-Linear Gamma, Linear Gamma, and Linear dense. ● Interpretable deep learning ("explainable AI"): (1) Grad-CAM, and (ii) Able to target the most relevant meteorological features for predicting extreme rainfall events.
		Results/Remarks	● Simple CNNs with traditional encoder-decoder structure provide superior results over other more recent networks. ● The deep learning framework improves rainfall downscaling, the largest for extreme rainfall events. ● The best CNN model outperforms existing statistical approaches (temporal variability and mean and extreme rainfall). ● Maps of extreme events for the Grad-CAM+ model were provided. ● Prediction of rainfall and extreme rainfall events (downscaling rainfall). ● The potential applicability of other types of networks e.g., conditional generative adversarial networks cGANs (unsupervised) is mentioned. ● Limitations: (i) Temporal relationships in data are not captured by CNNs, (ii) VCSN is known to have rainfall biases when the observation network is sparse, (iii) CNNs provide dry bias for extreme rainfall, and (iv) Due to limitations of input (e.g., historical data), these methods might not capture non-stationary processes.
	Resources	● The Grad-CAM interpretable deep learning code ● ERA5 analysis, C3S ● "The VCSN data is developed and maintained by NIWA and can be obtained through a data access agreement through correspondence with the authors" (*sic*)	
	Keywords	Statistical downscaling; Deep learning; Machine learning; Precipitation extremes; Rainfall.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Hazard characterization of extreme precipitation events.

S5. Dasymetric Mapping of Population Using Land Cover Data in JBNERR, Puerto Rico during 1990–2010

Table 7: S5's recipe table.

Abbrev.	Categories and data	
CH-S5	Title	Dasymetric Mapping of Population Using Land Cover Data in JBNERR, Puerto Rico during 1990–2010 (Cartagena-Colón et. al., 2022). [Link]
	Summary	This recipe provides a methodology to estimate the spatial population of an area and was proposed as a solution when critical data are scarce.
	Variables (input)	• Input: (i) Total population, (ii) Land cover datasets (high-resolution), and (ii) Raster data (climate parameters).
	Methods/Models	• A dasymetric mapping methodology for enhancing population spatial data by using various geospatial sources to produce a European Union-wide dataset of population variations. (i) Target zone estimation, (ii) Density estimation of ancillary class, and (iii) Error estimation. • The methodology combines widely available geospatial data like CLC, Openstreet map, and Copernicus Land Monitoring Service datasets (European Settlement Maps) with third-party datasets (Multinet, ToMTom datasets) and statistical data from EUROSTAT in a novel approach to map daytime population dynamics. • The approach can be used to enhance other data more relevant to disaster response or impact mapping. • The proposed methodology can be a general-use methodology for enhancing/improving sectorial data used in climatic scenarios.
	Results/Remarks	• Output: (i) The multi-temporal population grids for the European Union at 1 km2 resolution that have been generated during this study have been deposited in the European Commission's Joint Research Centre Data Catalog, with an identifier, and can be accessed at ENACT-POP R2020A - ENACT 2011 Population Grid, (ii) Dasymetric mapping error assessment, (iii) Maps of dasymetric population.

Abbrev.	Categories and data		
	Resources	• Uncovering temporal changes in Europe's population density patterns using a data fusion approach	
	Keywords	Intelligent Dasymetric mapping (IDM); Land use; Land cover; Census Data.	
	Tag/Type	Exposure and Vulnerability data	
	Application in ICARIA	Potential use in ICARIA	Categorizing population data with respect to land cover information.

S6. Climate change and energy performance of European residential building stocks – A comprehensive impact assessment using climate big data from the coordinated regional climate downscaling experiment

Table 8: S6's recipe table.

Abbrev.	Categories and data		
CH-EV-S6	Title	Climate change and energy performance of European residential building stocks – A comprehensive impact assessment using climate big data from the coordinated regional climate downscaling experiment (Yang et. al., 2021). [Link]	
	Summary	This recipe offers an impact assessment of climate change on the energy performance of residential building stocks considering different climate scenarios.	
		Variables (input)	• Numerical energy simulations for building stocks. • Gathering data using: (i) Tabula webtools, and (ii) "EPISCOPE". • Measure of energy performance of buildings: Cooling and heating demands estimations. • Future climate datasets. • Future climate scenarios periods: 2010–2039 (near-term or NT), 2040–2069 (medium-term or MT), and 2070–2099 (long-term or LT). • Climate data synthesized: (i) RCA4, (ii) RCPs: RCP 2.6, RCP 4.5, and RCP 8.5, (iii) Spatial resolution of 12.5 km and temporal resolution of 15 min, and (iv) GCMs: Centre National de Recherches Météorologiques Climate Model 5 (CNRM-CM5), Irish Centre for High-End Computing model (ICHEC-EC-EARTH), Institut Pierre Simon Laplace model (IPSL-CM5A-MR),

Abbrev.	Categories and data		
			Met Office Hadley Centre model (MOHC-HadGEM2-ES), and Max Planck Institute model (MPI-ESM-LR).
		Methods/Models	• Models the energy and performance in IDA Indoor Climate and Energy (IDA ICE). • It uses the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report model summary data of the HadCM3 A2 experiment ensemble which is available from the IPCC Data Distribution Centre (IPCC DDC). • The tool transforms 'present-day' EPW weather files into climate change EPW or TMY2 weather files which are compatible with most building performance simulation programs. • Future climate scenarios are simulated using GCMs. • Dynamically downscaled weather data generated by RCMs.
		Results/Remarks	• Output: (i) Annual average of heating and cooling demand, (ii) Forecasting of weather conditions (e.g., temperature distribution, etc.), (iii) Forecasting of averages for heating demands, (iv) Forecasting of averages for cooling demands, (v) Indoor thermal comfort based on different RCPs. • The framework focuses on building energy performance case studies and on solutions for reducing energy demands. • The vulnerability of cities to climate change based on the indoor thermal comfort is also considered and estimated. • Short- and long-term climate variations and extremes should be considered when assessing energy demands and building performance.
	Resources	• CCWorldWeatherGen • Climate-related risks and extreme events	

Abbrev.	Categories and data		
			• Total net greenhouse gas emission trends and projections in Europe • Thermal comfort
	Keywords	Climate change; Extreme Events; Energy performance of buildings; Thermal comfort; Assets.	
	Tag/Type	Climate Change and Hazard data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Correlate climate change trends with variation of seasonal energy needs. Incorporate in assets' data attributes the estimate of energy performance and related cost.

S7. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes

Table 9: S7's recipe table.

Abbrev.	Categories and data		
CH-S7	Title		Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes (Papacharalampous et. al., 2019). [Link]
	Summary		This recipe compares stochastic and data-driven methods for multi-step forecasting via computational experiments using simulated time series and real-world river discharge data.
		Variables (input)	• Framework for evaluating forecasting methods in hydrology; River discharge forecasting. • Simulation of time series using stochastic models. • Mean annual river discharge time series. • Hydrological variables at large time scales. • Input: (i) 12,000 simulated and 92 monthly streamflow time series, (ii) 6,000 simulated, (iii) 135 annual temperature time series, (iv) 24,000 simulated, 185 annual temperatures, and (v) 112 annual precipitation time series.
		Methods/Models	• Comparison between stochastic and data-driven methods for the forecasting of hydrological processes based on large-scale simulations. • Time series forecasting can be classified into

Abbrev.	Categories and data		
			eight categories: (1) Exponential smoothing, (2) ARIMA, (3) Seasonal models, (4) State space and structural models and the Kalmar filter, (5) Nonlinear models, (6) Long-range dependence models, e.g., the family of Autoregressive Fractionally Integrated Moving Average (ARFIMA) models, (7) Autoregressive Conditional Heteroscedastic/Generalized Autoregressive Conditional Heteroscedastic (ARCH/GARCH) models, and (8) Count data forecasting. ● Simulated processes: (i) ARMA (p,q), and (ii) ARFIMA (p.q.d). ● Real-time world series: (i) Mean annual river discharge time series, (ii) Autocorrelation Function (ACF), (iii) Partial Autocorrelation Function (PACF), and (iv) Hurst–Kolmogorov (HK). ● (a) Forecasting methods: (1) Stochastic methods: (i) Packages: arfima, Arima, auto_arima, BATS, ets, forecast, rwf, ses theta, and built-in-R functions, (ii) Naive forecasting method, and (iii) Random Walk forecasting method. (b) Data-driven methods: (i) Random Forest, (ii) NN, RF, SVM; package: ksvm, and (iii) Utilization of a single hidden-layer Multilayer Perceptron (MLP). ● Evaluation criterion for the models: (i) Type 1 accuracy: the closeness of the forecasted time series to the target time series, (ii) Type 2: the closeness of the mean of the forecasts to the mean of the target value.
		Results/Remarks	● Stochastic generation of weather data. ● Benchmark information is available for methodologies in this recipe. ● Heatmaps of the average-case performance of the forecasting methods. ● In total, ML models are more likely to outperform the stochastic methods in terms of accuracy and computational costs, remaining prone to their own limitations. ● Models applied for forecasting can be transferred for studying hydrometeorological concepts.
	Resources	● R packages: Cgwtools, Devtools, EnvStats, Forecast, Fracdiff,	

Abbrev.	Categories and data		
		Gdata, HKprocess, Knitr, Plyr, Readr, Rminer,Tidyr, hydroGOF. • No free lunch theorem – Blog source.	
	Keywords	No free lunch theorem; Random Forests; River discharge; Support Vector Machines; Time series.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Improve local climate change information through stochastic and data-driven approaches, focused on hydrological processes but transferable to other relevant variables.

S8. Downscaling probabilistic seasonal climate forecasts for decision support in agriculture: A comparison of parametric and nonparametric approach

Table 10: S8's recipe table.

Abbrev.	Categories and data		
CH-S8	Title	Downscaling probabilistic seasonal climate forecasts for decision support in agriculture: A comparison of parametric and nonparametric approach (Han et. al., 2017). [Link]	
	Summary	This recipe presents two downscaling methodologies, parametric and non-parametric, which are compared for seasonal rainfall forecasts, and their performance for stable simulations of the total rainfall distributions is explored.	
		Variables (input)	• Seasonal rainfall and its characteristics. • Downscale scenarios: (i) Frequency-only (π-only), (ii) rainfall amount Only (R_m-only), (iii) rainfall intensity (μ-only), (iv) both rainfall frequency and intensity (π-μ), (v) rainfall frequency and constraining total rainfall (R_m-μ), and (vi) rainfall intensity and constraining total rainfall (R_m -μ).
		Methods/Models	• Stochastic non-parametric temporal downscaling method – FResampler1: (i) Based on the concept of "conditional block sampling", and (ii) Disaggregate SCF (seasonal climate forecasts) to daily weather realizations. • Parametric downscaling method – predictWTD; Based on conditional stochastic

Abbrev.	Categories and data		
			weather generator.
		Results/Remarks	• Sensitive to data volume, sampling size, and number of realizations (requirement for stochastic models). • FResampler1 performs equally well to the parametric predictWTD method, captures seasonality and temporal correlation structure of data, remains sensitive to the number of realizations and to data availability. • The predictWTD remains sensitive to the length of observed data, not sensitive to number of realizations, and required longer periods of observations for rainfall amount or conditioning or intensity.
	Resources	• IRI Net Assessment Seasonal Climate Forecast	
	Keywords	Stochastic disaggregation; Probabilistic seasonal climate forecast; Parametric downscaling; Non-parametric downscaling; Rainfall.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Probabilistic non-parametric downscaling methodology for seasonal climate forecasts.

S9. An R package for daily precipitation climate series reconstruction

Table 11: S9's recipe table.

Abbrev.	Categories and data		
CH-S9	Title		An R package for daily precipitation climate series reconstruction (Serrano-Notivoli et. al., 2017). (Link)
	Summary		This recipe offers the technical characteristics of an open-source package written in R language for treating large data gaps, applied to sample precipitation datasets.
		Variables (input)	• Input: Daily precipitation (complete precipitation datasets).
		Methods/Models	• Treating large data gaps. • *"The observatories were located at less than 20 km and with a correlation of more than 0.7,*

Abbrev.	Categories and data		
			else to meet the necessary requirements, the radius of proximity was recursively extended by 5km". ● A multiple linear regression was performed with the data all observatories had in common. ● The gaps of the standard observatory that are empty were filled in with data from the other observatories.
		Results/Remarks	● "Closest-correlated neighbor" weather data gap-filling methodology". ● Data gap-filling for weather observations used by FIC in ICARIA.
	Resources	● Related projects: Junta de Andalucía SICMA Climate Change local scenarios ● [reddPrec package]()	
	Keywords	reddPrec; Daily precipitation; Quality control; Missing values.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Generation of complete daily weather series (no gaps) for improved assessment of past climate in ICARIA case study areas and development of statistical downscaling.

S10. Description and validation of a two-step analogue/regression downscaling method

Table 12: S10's recipe table.

Abbrev.	Categories and data		
CH-S10	Title	Description and validation of a two-step analogue/regression downscaling method (Ribalaygua et. al., 2013). ([Link]())	
	Summary	This recipe introduces a two-step analogue statistical downscaling method for daily temperature and precipitation, and accurately simulates the past climate on a local scale in the studying areas.	
		Variables (input)	● Surface observation Data (vector data) is combined with. ERA5-Land reanalysis

Abbrev.	Categories and data		
			datasets (raster data); Low resolution data are sourced from an observed reference dataset – ERA40 Reanalysis (atmospheric dataset). • 10 CMIP6 Global Climate Models + 4 Tier 1 SSP's. • Local downscaled climate projections at point (observation) scale.
		Methods/Models	• FICLIMA two-step analogue/regression downscaling method. • Estimates high-resolution surface meteorological fields for a day “x”, in two steps: (i) The first step is an analogue technique better adapted and improved. (ii) In the second step, high-resolution surface information is estimated differently for precipitation (using a probabilistic approach) and temperature (using multiple linear regression).
		Results/Remarks	• Problems addressed in this recipe: (i) Changeable relationships between predictors and predictands (including non-linear ones), (ii) Predictors are simulated by the GCMs. • Output: Simulation of precipitation and temperature (spatial distribution of verification metrics for the studied variables (max and min temperature and precipitation). • Advantages: (i) Low computational cost, (ii) Microclimatic features are implicit, (iii) Good verification results (e.g., full range of data variability is considered). • Limitations: (i) Historical observations of the variables are needed, (ii) Spatial and temporal inconsistencies cannot be eliminated, (iii) Non-stationary problem in the predictors-predictands relationships cannot be excluded, and (iv) Autumn precipitation in Mediterranean areas is poorly estimated.
	Resources	• ERA40- Reanalysis	
	Keywords	Statistical downscaling; Mean absolute error; Analogue techniques.	

Abbrev.	Categories and data		
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Already considered and applied for ICARIA project's demands.

S11. Weather Data Quality Control | Weather data temporal extension methodology

Table 13: S11's recipe table.

Abbrev.	Categories and data		
CH-S11	Title		A. Weather Data Quality Control | Weather Observations Homogeneisation Techniques. B. Weather data temporal extension methodology | Mainstream (traditional) gap filling methods (Gudmundsson et. al., 2012). [Link]
	Summary		These recipes offer techniques for data-gap filling in meteorological data.
		Variables (input)	• Raw weather observations (A). • Quality-treated weather observations (A). • Weather data with the presence of data gaps not covering all the time desired (B).
		Methods/Models	• (A) • A two-way quality control methodology is applied to all weather data sources prior to their use: (i) Basic consistency. Direct rejection of self-evident wrong values, and (ii) Atypical values or 'outliers. Unusual values within a data set. • The way to proceed with the homogeneity test is based on the following methodology (see ref. in "References/Useful Links" section): (i) To quantify the similarity between data across different years, a distribution comparison test based on the Kolmogorov-Smirnov (KS) test is used, (ii) The KS test is a non-parametric statistical test which provides a p-value that can be used as a measurement of the similarity between two years., (iii) The lower the value for Log (KS),

Abbrev.	Categories and data		
			the greater the probability of inhomogeneity between two consecutive values, (iv) If one year has been selected as a possible indicator of inhomogeneity, then it is subjected to another test ("*Similarity between years*"), and (v) If a jump or a break shows up between p-values in the selected years, there is a true inhomogeneity for all the series. ● (B) ● Depending on the circumstances the use of a temporally homogeneous weather dataset is mandatory. ● Temporal extension of the weather data is performed using a climate reanalysis, the ERA5-Land. ● Reanalysis is crossed with the weather observations. ● Corrected simulated observation is crossed and filled with the original observation obtaining the original checked weather observation with gaps filled. ● Voronoi polygons to determine the buildings supplied by the substations: (i) Substitution: the water distribution system layout was missing therefore the road layout was used instead assuming that the water network followed its layout, (ii) Logical rule-based reasoning: used for the burst locations (sewers overload), and (iii) Approach based on similar historical events previously within the city or in other areas with similar conditions is explored.
		Results/Remarks	● Quality checking performed by FIC in ICARIA for all the weather data gathered (A). ● Homogenization techniques performed by FIC in ICARIA for all the weather data gathered (A). ● Data gap-filling for weather observations used by FIC in ICARIA (B).
	Resources	● Detection of inhomogeneities in daily data: a test based on the Kolmogorov-Smirnov goodness-of-fit test.	

Abbrev.	Categories and data		
		• Relevant projects: RESCCUE D3.2. deliverable. Tools with updated impact assessment models (A, B).	
	Keywords	Weather observations; Kolmogorov-Smirnov test; Data gaps.	
	Tag/Type	Climate Change and Hazard Data.	
	Application in ICARIA	Potential use in ICARIA	Creation and checking of extended weather observations for improved assessment of past climate in ICARIA case study areas and development of statistical downscaling.

S12. A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations

Table 14: S12's recipe table.

Abbrev.	Categories and data		
CH-S12	Title	A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations (Wang et. al., 2012). [Link]	
	Summary	This recipe presents an efficient method for handling large spatio-temporal datasets introduced and applied to a global soil moisture product from remote sensing images.	
		Variables (input)	• Works for large spatiotemporal datasets (both spatial & temporal variability). • Global volumetric soil moisture product (satellite) with the Land Parameter Retrieval Model (LRM) (2003-2009).
		Methods/Models	• A penalized least square method based on three-dimensional discrete cosine transforms (DCT-PLS), for the purpose of filling data gaps in large spatio-temporal datasets (for example: soil moisture satellite data) is introduced. • This DCT-PLS method has some novel features with respect to other gap-filling methods. (i) It is a method of full three-dimensionality, and thus (ii) explicitly utilizes both spatial and temporal information

Abbrev.	Categories and data		
			of the dataset to derive the statistical model and (iii) Predict the missing values. Instinctively. ● This strategy is preferable for spatio-temporal datasets rather than using only spatial or temporal modelling. ● The method utilized both spatial and temporal information of the moisture dataset.
		Results/Remarks	● The statistical modelling process is completely controlled by one smoothing parameter which is easy to specify and eliminates the need for complicated model parameterizations. ● DCT-PLS provides estimation with small errors for the global moisture dataset and can be used to fill in missing values.
	Resources	● BiomeCardio (Matlab package).	
	Keywords	Remote sensing; Soil moisture; Gap filling; Penalized least square regression; Discrete cosine transform.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Improving quality of heat and flood hazard assessment introducing soil moisture as key variable.

S13. Spatial interpolation techniques for climate data in the GAP region in Turkey

Table 15: S13's recipe table.

Abbrev.	Categories and data		
CH-EV-S13	Title	Spatial interpolation techniques for climate data in the GAP region in Turkey (Apaydin et. al., 2004). [Link]	
	Summary	This recipe provides a benchmark for identifying the optimal methodology for interpolating the spatial distribution of a specified set of tested climate parameters through geostatistical interpolation techniques.	
		Variables (input)	● Climate parameters: (i) Solar radiation, (ii) Sunshine duration, (iii) Temperature, (iv)

Abbrev.	Categories and data		
			Relative humidity, and (v) Wind speed and (vi) rainfall. • Long-term yearly predicted temperature maps.
		Methods/Models	• Interpolation techniques: (1) Inverse distance weighted (IDW), (2) Global polynomial interpolation (GPI), (3) Local polynomial interpolation (LPI), (4) Completely regularized spline (CRI), (5) Cokriging, and (6) Kriging with four subtypes: (i) Ordinary kriging (KO), (ii) Simple kriging (KS), (iii) Universal kriging (KU), and (iv) Disjunctive kriging (KD).
		Results/Remarks	• Maps of long-term yearly predicted temperature in the studied area. • Spatial interpolation techniques are utilized for distribution of climate parameters.
	Resources	N/A	
	Keywords	Spatial interpolation; Inverse distance weighted; Polynomial interpolation; Kriging; Cokriging; Completely regularized spline.	
	Tag/Type	Climate Change and Hazard data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Improved mapping of local climate change hazard conditions and of critical variables for exposure and vulnerability assessment through geostatistical interpolation techniques.

Dynamical downscaling

D1. A simple hybrid statistical–dynamical downscaling method for emulating regional climate models over Western Europe. Evaluation, application, and role of added value?

Table 16: D1's recipe table.

Abbrev.	Categories and data	
CH-D1	Title	A simple hybrid statistical–dynamical downscaling method for emulating regional climate models over Western Europe. Evaluation, application, and role of added value? (Boé et. al.,

Abbrev.	Categories and data		
			2023) [Link]
	Summary		This recipe describes an emulation methodology which is based on a hybrid statistical-dynamical approach based on analogues to simulate regional climate models.
		Variables (input)	• Input: (i) Temperature, and (ii) Precipitation Data. • RCMs from Euro-CORDEX at 12km resolution. • Scenario: RCP8.5. • Regional simulations: (1) CNRM-CM5 GCM from CMIP5 downscaled using three regional simulations with the GCM/RCM mode and the RCM/RCM mode, (2) RCM/RCM mode to downscale from CMIP6, and (3) Historical and ssp5-8.5 simulations from the thirteen CMIP6 models.
		Methods/Models	• A hybrid statistical–dynamical downscaling method: (i) Statistical model based on the results of RCMs, (ii) Applied to downscale GCMs, (iii) Aims to emulate regional climate models, and (iv) It does not require the stationarity assumption of statistical downscaling. • Emulate RCM results, based on the constructed analogues approach. • Estimation method based on constructed analogues: • Analogues of large-scale predictors from a low-resolution climate projection are considered: for high-resolution precipitation (temperature), the chosen predictor is low-resolution precipitation (temperature). • Emulated models: (1) the GCM/RCM mode: "*Fine-Scale Ref is a high-resolution regional climate simulation and Coarse-scale its driving GCM*" (*sic*), and (2) the RCM/RCM mode: "*Coarse-Scale Ref is simply the Fine-Scale Ref simulation aggregated on the low-resolution grid of the model to be downscaled (Mod)*" (*sic*).
		Results/Remarks	• Novelty: the analogues are searched within an RCM and therefore both in the past and the

Abbrev.	Categories and data		
			future climate. • The hybrid method is shown to reproduce climate change signals very well and to outperform a conventional statistical downscaling method. • "*Emulation methods make it possible to downscale very large ensembles of global climate projections and therefore to fully explore the uncertainties involved in regional climate changes*" (*sic*). • "*In the RCM/RCM mode, the climate change signal at large scale of the original GCM is very well captured by the hybrid statistical downscaling method, independently of its magnitude*" (*sic*).
	Resources	• Euro-CORDEX regional climate projections: Earth System Grid Federation • Constructed analogues method links: i. Searching for analogues, how long must we wait? ii. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. iii. The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. iv. Hydrologic extremes – an intercomparison of multiple gridded statistical downscaling methods.	
	Keywords	RCMs; Hybrid statistical dynamical downscaling; Climate change; Emulation.	
	Tag/Type	Climate Change and Hazard data	
	Application in ICARIA	Potential use in ICARIA	Emulating the downscaling function of an RCM and improving resolution of climate change data.

D2. Dynamical and statistical downscaling of SSPs in AMB

Table 17: D2's recipe table.

Abbrev.	Categories and data		
CH-D2	Title		Dynamical and statistical downscaling of SSPs in AMB (ARSINOE Project). [Link] 2021
	Summary		Dynamical and statistical downscaling methods are employed to combine SSPs and RCPs with land use cover temporal series data and obtain future projection scenarios.
		Variables (input)	• Land use cover changes. • Ease of change data. • Qualitative and quantitative drivers. • Observed land uses for each time series. • Land use demands.
		Methods/Models	• iCLUE/CLUEMONDO model for future land use projections simulation. • Dynamical Downscaling combining RCPs and SSPs scenarios. • CORINE land use cover provides land use time series data (with satellite quality) to feed the projection models. • SSPs provide a framework to integrate the future socioeconomic pathways in an RCP environment to approach a more realistic projection.
		Results/Remarks	• Output: Future land use projections with integrated information about the different socioeconomic and climatic scenarios adjusted to a more accurate downscaled geographical case study. • In contrast with previous methodologies, this method provides a framework for the integration of socioeconomic scenarios to the simulation parameters.
	Resources		• Relevant projects: ARSINOE: Huber García, V., Meyer, S., Kok, K., Verweij, P., & Ludwig, R. (2018). Deriving spatially explicit water uses from land use change modelling results in four river basins across Europe. Science of The Total Environment, 628-629, 1079-1097. https://doi.org/10.1016/j.scitotenv.2018.02.051. • IDESCAT Population Data • IDESCAT GDP (province and councils) • IGN CORINE land cover (time series) • CORINE land cover CORINE land cover (metadata)

Abbrev.	Categories and data	
	Keywords	Socioeconomic projections (SSPs), Downscaling, RCPs, AMB, Land Use Cover, Climatic projections.
	Tag/Type	Climate Change and Hazard data.
	Application in ICARIA	Potential use in ICARIA: Already considered and applied for ICARIA project's demands.

D3. Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications

Table 18: D3's recipe table.

Abbrev.	Categories and data	
CH-D3	Title	Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications (Manzanas et. al., 2018). (Link)
	Summary	This recipe presents an intercomparison of dynamical and statistical downscaling methods for seasonal forecasting over Europe, based on a 15-member hindcast from the EC-EARTH global model, focusing on summer mean temperature
	Variables (input)	• E-OBS. • EC-earth (EUROPIAS project) - Global seasonal predictions. • Regional CM used: RACMO2, WRF, RegCM. • Raster data (gridded seasonal temperatures). • Vegetation maps (from ECOCLIMAP for dynamical downscaling).
	Methods/Models	• Dynamical downscaling: (1) RACMO2: hydrostatic model employing 40 hybrid coordinate full vertical levels. (2) Weather Research and Forecasting system: non-hydrostatic dynamic core, employing 30 full eta vertical levels, and (3) RegCM modelling system: hydrostatic, compressible, sigma-p, vertical coordinate model considering 18 sigma-p levels. • Statistical downscaling: relying on coarse-resolution global simulated predictors: (1) Perfect prognosis (PP), and (2) Model

Abbrev.	Categories and data		
			Output Statistics (MOS); " *PP techniques can be applied on a daily, monthly or seasonal basis, whereas MOS techniques require working at monthly or seasonal time-scales*" (*sic*). ● PP-ANA is based on the popular analogue technique, which estimates the local downscaled values corresponding to a particular atmospheric configuration from the local observations corresponding to a set of similar (or analog) atmospheric configurations within a historical catalog formed by reanalysis. ● PP-MLR is an extension of simple linear regression which attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. The fit is determined by minimizing the sum of the residuals between the regression line and the observed data. ● Dynamical downscaling is based on regional models, which run on a relatively fine grid (e.g., 10–20 km) over a limited domain (e.g., Europe) initialized and driven at the boundaries by the coarse global model outputs. ● These models can generate regional predictions for a suite of climate variables but still may suffer from significant biases which require post-processing with bias adjustment techniques before they can be used in impact applications.
		Results/Remarks	● Maps of dynamical downscaling products for representation of the mean and extreme values. ● Statistical downscaling methods typically show minimal biases and provide realistic climate information (from the mean to the extremes) when compared to global models. ● Regional information still plays a significant

Abbrev.	Categories and data		
			role in specific sector climate indices and impact models. • ROC Skill score was used as a measure of the accuracy of the probabilistic forecasts, providing maps where both dynamical and statistical downscaling methods were included; Both downscaling methods resulted in similar patterns showing low-to-moderate skill over most continent. • Used on energy performance case studies.
	Resources	• Relevant projects: (i) SPECS, and (ii) EUROPIAS • Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe. • A Description of the Advanced Research WRF Version 3 – Technical report. • RegCM4: model description and preliminary tests over multiple CORDEX domains. • Comparison of dynamically and statistically downscaled seasonal climate forecasts for the cold season over the United States. • Regional climate modelling.	
	Keywords	Dynamical downscaling; Statistical downscaling; Seasonal forecasting; Multiple linear regression; Precipitation; Heatwaves.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Improving resolution of climate change data, in particular for assessing seasonal energy needs.

D4. Dynamical and statistical downscaling of a global seasonal hindcast in eastern Africa

Table 19:D4's recipe table.

Abbrev.	Categories and data		
CH-D4	Title	Dynamical and statistical downscaling of a global seasonal hindcast in eastern Africa (Nikulin et. al., 2018). [Link]	
	Summary	Dynamical and statistical downscaling methods are combined to access seasonal forecast for impact modelling.	
		Variables (input)	• EC-Earth

Abbrev.	Categories and data		
			• Datasets: (1) Climate research unit time-series, (2) Global precipitation climatology center, (3) Tropical applications of meteorology, (4) African rainfall climatology, (5) African estimation algorithm, (7) Climate hazards group InfraRed precipitation stations, and (8) WATCH-forcing-Datra-ERA-Interim. • Regional CM used: CCLM4-8-21 (CCLM4), RCA4 (RCA4), RegCM-4-3 (RegCM4), WRF341I (WRF341), WRF381D (WRF381)
		Methods/Models	• (1) Downscale ECMWF system-4 seasonal hindcasts, (2) RCMs: a domain of configurations has been selected, (3) Empirical statistical downscaling: Two (2) ESD methods were selected to downscale the full stream of the EC-EARTH hindcast: (i) AN1: variation of the Analogue technique, and (ii) A variation of the generalized Linear Models (GLMs), (4) Selection of the subregions was based on a group of the initial datasets to the LEAP platform, and (5) Rainfall indexes were considered to evaluate seasonal forecasts: (i) the Simple Daily Intensity Index (SDII), and (ii) the Wet Day Frequency (WDF). • Verification metrics: (i) Interannual correlation, (ii) Brier skill score, and (iv) ROC Skill Score: ROCCS maps for the EC-EARTH hindercasted rainfall provide allow to detect observational uncertainties. • Dynamical Downscaling using RCMs is computationally expensive delaying the provision of the forecasts and requires much more resources than ESD (e.g., saving a wealth of driving boundary conditions from GCMs). • LEAP platform offers information about humanitarian needs and interventions.
		Results/Remarks	• Output: (i) Interannual correlation maps and maps of the global and downscaled hindcasts, (ii) ROCCS maps for the global and

Abbrev.	Categories and data		
			downscaled hindcast rainfall, (iii) Maps of rainfall tercile forecast for each model, (iv) Verification data for the rainfall indices (via maps of distribution of the ROCCSS for the global and downscaled rainfall forecasts). • In contrast to the ESD approach, RCMs can provide a larger number of variables in a physically consistent way, including regional and local feedback which can be important in seasonal forecasting. • Main results: (i) Observational uncertainties, (ii) A global forecast system, (ii) Both dynamical and statistical downscaling, (iv) Applicability of rainfall indexes and (v) The capabilities of an early warning system (LEAP platform).
	Resources	• Relevant projects: EUROPIAS • downscaleR package – GitHub repository. • Can a Regional Climate Model Improve the Ability to Forecast the North American Monsoon? • Dynamical downscaling of ECMWF Ensemble seasonal forecasts over East Africa with RegCM3. • Downscaling ECMWF seasonal precipitation forecasts in Europe using the RCA model.	
	Keywords	Seasonal forecast; Downscaling; Drought early-warning system; RCMs; Precipitation; Dynamic downscaling; Generic.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Already considered and applied for ICARIA project's demands.

Data-driven based methodologies and data fusion methods

DD1. Developing novel machine-learning-based fire weather indices

Table 20:DD1's recipe table.

Abbrev.	Categories and data		
CH-EV-DD-1	Title		Developing novel machine-learning-based fire weather indices (Shmuel et., al., 2024). [Link]
	Summary		This recipe introduces a data-driven fire weather index that outperforms current traditional fire indexes (which often fall short due to the non-linear nature of wildfire risk factors) and provides accurate wildfire risk estimations which are key for optimal forest management and firefighting.
		Variables (input)	• Global wildfire datasets. • Variables: (1) daily ignition, (2) 2m temperature, (3) humidity, (4) 10m wind speed, (5) precipitation, (6) mean slope, (7) population density, (8) NDVI, and (9) Incoming short-wave solar radiation.
		Methods/Models	• Models would benefit if could include: (i) including meteorological factors, (ii) topography, (iii) fuel loads, (iv) anthropogenic factors, (v) include 2-meter temperature, (vi) precipitation, (vii) RH, (viii) 10-meter wind velocity, based on the ERA5 hourly reanalysis data, and (ix) population density. • Fourteen (1$) indexes are used for comparison, grouped in three (3) categories as following: (1) Canadian Forest Service Fire Weather Index Rating System, (2) Australian McArthur Mark 5 Rating System, (3) U.S. Forest Service National Fire-Danger Rating System. • The variables in each category include: (1a) fire weather index, (1b) build up index, (1c) danger index, (1d) drought code, (1e) duff moisture code, (1f) initial fire spread index, (1g) fine fuel moisture code, (1h) fire daily severity rating. (2a) Keetch-Byram drought index, and (2b) fire danger index. (3a) spread component, (3b) energy release component, (3c) burning index, and (3d) ignition component. These variables were available in a 0.25◦ resolution from the Copernicus Climate Change Service (Fire Danger Indices Historical Data from the Copernicus Emergency Management Service).

Abbrev.	Categories and data		
			• Metrics: (1) Under the curve metric (AUC), (2) Operating characteristics curve (ROC), and (3) Precision-Recall Curve. • Classification models: (1) RF, (2) Extreme Gradient Boosting (XGBoost), (3) MLP, a form of Neural Network and (iv) logistic regression. • Method to study the actors in terms of wildfire risk: SHAP (SHapley Additive exPlanations) values analysis.
		Results/Remarks	• XGBoost model achieved the highest score based on the ROC-AU curves. • Traditional indices and subindices: the Duff Moisture Code and the Keetch-Byram Drought Index achieved the highest performance. • ROXC curves for wildfire ignition prediction for the ML models are provided. • Prediction accuracy maps for all ML-based index methods are provided, for the studied variables. • Maps of the two-day dependence plots for wildfire occurrence are provided. • Global maps of wildfire occurrence for all ML-based index methods are provided. • Temperature was estimated as the key feature for the prediction of wildfire occurrences. • ML-based FWIs are able to provide wildfire occurrence estimation in a daily resolution in all regions worldwide.
	Resources	N/A	
	Keywords	Machine learning, fire weather indices, forest management, wildfire risk.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	A methodology combining data-driven methodologies and fire weather indices.

DD2. PVS-GEN: Systematic Approach for Universal Synthetic Data Generation Involving Parameterization, Verification, and Segmentation

Table 21: DD2's recipe table.

Abbrev.	Categories and data		
CH-EV-DD-2	Title	PVS-GEN: Systematic Approach for Universal Synthetic Data Generation Involving Parameterization, Verification, and Segmentation (Kim et. al., 2024). [Link]	
	Summary	This recipe offers a new method that parameterizes empirical time-series data with minimal intervention.	
		Variables (input)	• Empirical data and synthetic data from general-purpose time-series data of any type. • Datasets: (i) Gas sensor array dataset, (ii) Low-Energy house dataset, (iii) EEG Alcoholism dataset, and (iv) Heterogeneity activity recognition dataset.
		Methods/Models	• Process: (i) Parameterization: utilize empirical data with ACRIMA to derive automated parameters, (ii) Verification: compare the synthetic data with the empirical data using our proposed metric, the possibility of reproducibility (RoR), and (iii) Segmentation for universal synthetic data generation: enhance the time-series consistency and regularity. • Statistical models: (i) SES, (ii) ARIMA, and (iii) GMM. • Data-driven methods: (i) SVR, and (ii) LSTM. Other alternatives: (i) GANs, (ii) VAEs, and (iii) RNNs.
		Results/Remarks	• Reduced user intervention and reduced resources for acquiring and labeling large amounts of empirical data. • A universal methodology could benefit having the following characteristics: (i) Automatic generation of time-series data for a range of sensors and data process, (ii) Parametrization of empirical data, (iii) Independent of sensor data traits, (iv) Encapsulation of the temporal dynamics of time-series data, (v) Enabling quantitative comparisons between generated and empirical data by reflecting the time-series characteristics of the data with their descriptive statistics, and (vi) Scalability.
	Resources	N/A	

Abbrev.	Categories and data		
	Keywords	Time-series sensor data; synthetic data generation; time-series synthesis; IoT data generation; possibility of reproducibility.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Multi-purpose data generation system based on the processing of empirical data and time series.

DD3. A single-building damage detection model based on multi-feature fusion: A case study in Yangbi

Table 22: DD3's recipe table.

Abbrev.	Categories and data		
EV-DD-3	Title	A single-building damage detection model based on multi-feature fusion: A case study in Yangbi (Du et. al., 2024). [Link]	
	Summary	This recipe offers a multi-fusion feature model for accurate identification and classification of building damage detection to reduce information redundancy applied to earthquake events for demonstration.	
		Variables (input)	• Input: (i) Binary map of buildings, (ii) Outlines of buildings, (iii) Satellite DOM image, (iv) UAV DOM and DSM image, and (v) Google Maps-based satellite images.
		Methods/Models	• Data from satellites and UAVs are obtained for damage building detection after a hazard event (earthquake). • An image change detection model is applied. • Methodology: (i) nDSM: extracts building contours, (ii) image segmentation: K-nearest neighbor was used for classification of the study area using spectral average grayscale value, rectangularity features, (iii) morphological closure operation: results of building contour extraction, (iv) segmentation of buildings, (vi) classification of damage types of buildings, and (vii) texture feature change analysis, image fusion, and PCA. • Statistical analysis: (i) Maximum Likelihood Classification (ML), (ii) Neural Net

Abbrev.	Categories and data		
			Classification (NN), (iii) Mahalanobis Distance Classification (MD), and (iv) Support Vector Machine Classification (SVM) machine learning models were applied.
		Results/Remarks	• Output: (i) Maps of damaged buildings, (ii) Post-and pre-hazard UAV and satellite images, (iii) Maps of extraction and distribution results of damaged buildings • Limitations: (i) Model should be able to process data from diverse resources, (ii) A larger input of damaged buildings should be considered, and (iii) The procedure of extraction and selection should be automated for optimization.
	Resources	N/A	
	Keywords	Multi-feature fusion; Damage detection model; Earthquake; Normalized Digital Surface Model (nDSM); Buildings; Structural Damage; Building impact assessment.	
	Tag/Type	Exposure and Vulnerability analysis.	
	Application in ICARIA	Potential use in ICARIA	Exposure and vulnerability analysis for single buildings, with a focus on structural damage, based on seismic impact assessment, potentially transferable to other hazards.

DD4. Assessing automated gap imputation of regional scale groundwater level data sets with typical gap patterns

Table 23: DD4's recipe table.

Abbrev.	Categories and data		
CH-DD -4	Title	Assessing automated gap imputation of regional scale groundwater level data sets with typical gap patterns (Bikše et. al., 2023). [Link]	
	Summary	This recipe introduces and compares two data imputation methodologies to simulate complex missing value patterns by mimicking typical gap patterns, tested for reproducing daily groundwater hydrographs.	
		Variables (input)	• Regional scale groundwater level data sets: (i)

Abbrev.	Categories and data		
			Rechange, (ii) Groundwater-surface, and (iii) Water interaction. ● Mimicking of: (i) Typical gap patterns, and (ii) Random gap patterns.
		Methods/Models	● Linear Interpolation. ● MissForest: (i) Non-parametric, (ii) Iterative, (iii) Missing values imputation, (iv) Random Forest algorithm, (v) Automatic, (vi) Unsupervised missing imputations, (vii) No assumptions about data distribution, (viii) No need for tuning parameters, and (ix) Performs for the infilling daily groundwater hydrographs. ● ImputePCA: (i) Multiple imputation method, (ii) Principal components method on an incomplete dataset, and (iii) Performs iteratively principal component analysis (PCA). Convergence: when the difference between two successive iterations is below a defined threshold. ● Artificial gaps: (i) Typical gap patterns: (ii) Eleven (11) distinct groups of gaps in groundwater hydrographs were performed, and (ii) Artificial gaps of time series were simulated. (ii) Random gap patterns: Artificial gaps were introduced in 109 hydrographs. ● Imputation's performance: I) imputePCA performed less accurately with more dispersed results, (ii) Typical gap patterns were demanding for all imputation algorithms, (iii) missForest outperformed both the imputePCA and the linear interpolation algorithms, and (iv) All infilled hydrographs performed poorly when imputing typical gap patterns.
		Results/Remarks	● Impact of individual gaps: (i) All methods performed poorly in estimating long continuous gaps, (ii) The accuracy of the infilling in the beginning and at the end of the hydrographs performed poorly when compared to the rest of the gaps.

Abbrev.	Categories and data		
			• Estimation of changes in the number of daily hydrographs. • Extreme hydrographs remain challenging to address. • Gap-filling methods fail around gaps that contain extremes. • Random-like gap patterns are linked with more simple imputation methods in terms of performance and accuracy. • Typical gap patterns do not offer a consistent performance on imputation, despite the gap characteristics.
	Resources	• WATERRES: EU-integrated management system of cross-border groundwater resources and anthropogenic hazards.	
	Keywords	Time series; Missing values; Gap filling; Droughts; Abstraction.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	This recipe offers alternative methodologies for data imputation.

DD5. From theory to practice: optimization of available information for landslide hazard assessment in Rome relying on official, fragmented data sources

Table 24: DD5's recipe table.

Abbrev.	Categories and data		
CH-DD-5	Title	From theory to practice: optimization of available information for landslide hazard assessment in Rome relying on official, fragmented data sources (Esposito et. al., 2023). [Link]	
	Summary	This recipe offers a landslide hazard risk management protocol for highly urbanized areas.	
		Variables (input)	• List of geological hazards. • Datasets for known landslides. • Prediction of landslide susceptibility. • Point-based landslide database is represented by: (i) 1099 LIPs (289 original and 810 synthetic), and (ii) The 67 related to the January 2014 extreme rainfalls are excluded. • Continuous map(s) of landslide initiation

Abbrev.	Categories and data		
			susceptibility based on data-driven model(s). ● Detection rate curves for the classification of the susceptibility. ● Spatial density maps of shallow landslides and earth slides.
		Methods/Models	● Objectives: (i) Dataset preparation, (ii) Susceptibility assessment, and (iii) Information about landslides. ● Evaluation of intensity and temporal probability of landslides. ● Definition of rainfall-induced landslide hazard: (i) Definition of the spatial component of the landslide hazard, (ii) Temporal component of the landslide hazard, (iii) Preliminary and large-scale quantitative hazard description, and (iv) Evaluation of the return periods of landslide trigger rainfall events.
		Results/Remarks	● Outputs: (i) A uniform, updated database is composed, (ii) The spatial component of the hazard is depicted by (a) Continuous maps of landslide initiation, and detection rate curves, (b) Classification of landslide susceptibility, (iii) The temporal component of the hazard is depicted by (a) Resulting landslide frequency estimation, and (b) Rainfall probability curves for the tested areas, and (iv) Persistent scatterer interferometry: (a) A-DInSAR velocity maps, and (b) Susceptibility hazard index maps for the tested areas. ● GIS- and ML-based methods were applied to collect and integrate open-source landslide inventories into a database. ● The reported products aid decision-makers in managing landslide risks by reporting activity status and susceptibility, supporting informed monitoring and investment prioritization for prevention and mitigation.
	Resources	● Known landsides ● Open access land use maps ● Hydro-geological Structure Plan	

Abbrev.	Categories and data	
		• Municipal land use plan • Civil protection plans • IDROGEO Platform
	Keywords	Susceptibility; Machine learning; Rainfall probability; Landside hazards; Landside inventories; Interferometry.
	Tag/Type	Climate Change and Hazard Data.
	Application in ICARIA	Potential use in ICARIA: Landslide hazard assessment in relation to rainfall intensity, supporting dynamic analysis of landslide risk depending on seasonal changes in rainfall patterns and extreme events frequency/intensity.

DD6. Modelling national residential building exposure to flooding hazards

Table 25: DD6's recipe table.

Abbrev.	Categories and data		
EV-DD-6	Title		Modelling national residential building exposure to flooding hazards (Paulik et. al., 2023). [Link]
	Summary		This recipe offers a model for flood risk assessment by studying the building characteristics for object-level replacement evaluation in flooded areas utilizing public data, and data-driven methodologies for the estimation of the hazard area exposure.
		Variables (input)	• Variables: (i) Location (e.g., Address count and units, etc.), (ii) Geometric and non-geometric characteristics (e.g., Floor area, and height, Building count, Land area, etc.). • Integration of public data and geospatial physical/non-physical building characteristics.
		Methods/Models	• Workflow to estimate physical and non-physical characteristics for object-level replacement valuation using (i) spatial data, (ii) geometric data, and (iii) data-driven methods. • Geometry building properties were estimated based on geospatial operations and open topographic data.

Abbrev.	Categories and data		
			• Building height (BHT) extraction and level (BL) enumeration were performed using a geospatial model and LIDAR point clouds. • Data-driven methods as tree-ensemble models for value imputation (supervised learning regression and classification algorithms): (i) Random Forest, and (ii) XGBOOST.
		Results/Remarks	• Estimation of residential building characteristics and replacement values. • Output: (i) Regional residential building and replacement value exposure, (ii) RF demonstrates higher overall performance compared to XGBOOST, and (iii) Object-level information for exposure in risk assessments can be critical at regional to national scales.
	Resources	• Geospatial datasets: (i) NZ building outlines (ii) NZ primary land parcels, (iii) LIDAR point clouds, (iv) NZ functional urban areas, (v) RiskScape software, (vi) CostBuilder.	
	Keywords	Floods; Residential buildings; Exposure; Monetary values; Supervised learning; Object-level modelling.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Flood hazard exposure and vulnerability analysis of buildings.

DD7. Deep Learning Regional Climate Model Emulators: A Comparison of Two Downscaling Training Frameworks

Table 26: DD7's recipe table.

Abbrev.	Categories and data	
CH-DD-7	Title	Deep Learning Regional Climate Model Emulators: A Comparison of Two Downscaling Training Frameworks (Van Der Meer et. al., 2023). [Link]
	Summary	This recipe describes a methodology where it explores the potential of using data-driven methodologies alternatively to dynamical downscaling, applied to a global climate model (GCM) to regional resolution.

Abbrev.	Categories and data		
		Variables (input)	• Input: (i) Precipitation, (ii) Downward radiation, (iii) Humidity, (iv) Temperature, and (iv) Pressure. • RCM target domain: box of 64 x 64 pixels. • RCM (from MAR(ACCESS1.3)), GCM(CMIP5). • Input features to RCM: 1D (e.g., Seasonal indicators, Spatial mead of 2D variables, etc.), and 2D (e.g., Wind, Downward radiation, Humidity, etc.). • Climate variables: precipitation, temperature, etc....
		Methods/Models	• Two Surface mass balance (SMB) emulators, a perfect and an imperfect one, were chosen to downscale a GCM. • Two ML models were applied to SMB to examine the downscaling potential: U-Net model. • U-Net (CBAM (Convolution block attention) combined with depth wise-separable convolutions (DSC)), Smart-UNet.
		Results/Remarks	• Maps of SMB predictors of the RCM emulators over a test period. • Maps of evaluation metrics on predictions from the RCM emulators over a test period. • Output: (i) Perfect model fails to reproduce SMB's extreme values, (ii) Both perfect and imperfect models succeed in reproducing complex spatial structure of the RCMs, and (iii) Inconsistencies due to the difference in resolution between large-scale and local-scale variables are not negligible and might confuse the RCM-emulator. • Limitations: (i) Temporal and spatial inconsistencies might occur if an offset is present in the RCM time series, (ii) Inconsistencies between RCM and GCM variables remain present, (iii) The typical limitations of machine learning methods are also applicable in this context.
	Resources	• ACCESS 1.3 GCM data.	

Abbrev.	Categories and data		
		• [Historical and future RCP8.5 simulations – GitHub repository]().	
	Keywords	DL; RCMs; GCMs; Downscaling; Deep Learning RCM-emulator; Dynamic Downscaling of GCMs; Surface Mass Balance (SMB).	
	Tag/Type	Climate Change and Hazard Data.	
	Application in ICARIA	Potential use in ICARIA	Emulating RCM downscaling to improve resolution of climate change data in case study areas.

DD8. Self-supervised learning for climate downscaling

Table 27: DD8's recipe table.

Abbrev.	Categories and data		
CH-DD-8	Title	Self-supervised learning for climate downscaling (Singh et. al., 2023). [Link]	
	Summary	This recipe offers a self-supervised deep-learning solution for climate downscaling that can be applied without requiring high-resolution ground truth data.	
		Variables (input)	• Low-Resolution Climate data: Synthetic LR data can also be created by degrading real High-Resolution data. • Climate variables: (i) The surface temperature, (ii) total precipitation, and (iii) topographical gradient.
		Methods/Models	• Community Earth System Model (CESM): Fully coupled global climate model. • Capabilities of neural networks to reconstruct high-resolution data from given low-resolution simulations. • Legacy low-resolution simulations can be downscaled to reconstruct high-resolution detail. • Past observations that have been taken at lower resolutions can be increased to higher resolutions, opening new analysis possibilities. • Networks: (i) Low-Frequent Data:

Abbrev.	Categories and data		
			Residual-Predicting Network (RPN), and (ii) High-Frequent Data: Deconvolutional Network (DCN). • Downscaling is performed over the surface temperature and the topographic gradient. • Deep-learning methodologies: CNN method (each model trained for 500 epochs, using Adam optimizer and LeakyReLU). • Characteristics: (i) Self-supervised deep-learning solution for climate downscaling, (ii) No high-resolution ground data input required, (iii) CNN models train on a single instance at a time, and (iv) Improvement of downscaling performance.
		Results/Remarks	• The method is compared to self-supervised models ($SSL_{SRResNet}$, and SSL_{GINE}). The two SSLs have also been used as training components on the pseudo-LR and HR data. • The method is capable of estimating HR climate data without ground truth data.
	Resources	• Repository for the work titled "Self-supervised learning for climate downscaling" - GitHub repository Data for this recipe: Ultra-high-resolution climate simulation project.	
	Keywords	Climate downscaling; self-supervised; Deep Learning; CNNs; Super-resolution; Earth system models; Climate Simulation.	
	Tag/Type	Climate Change and Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Improving resolution of climate data.

DD9. An Exploration of Interpolation - Machine Learning Model for Climate Model Downscaling Under the Limitation of Data Quantity

Table 28: DD9's recipe table.

Abbrev.	Categories and data	
CH-DD-9	Title	An Exploration of Interpolation - Machine Learning Model for Climate Model Downscaling Under the Limitation of Data

Abbrev.	Categories and data		
			Quantity (Prathom et al., 2023). [Link]
	Summary		This recipe describes a methodology to perform downscaling and address any data-gap issues introduced, by combining interpolation and data-driven methodologies.
		Variables (input)	• LR climate data (IPSL-CM6A-LR in CMIP6 with 250 km of spatial resolution are selected as the GCM output).
		Methods/Models	• IDW (Inverse Distance Weight). • TIN (Triangular Interpolation Network). • ANNs (Artificial Neural Networks). • GBRT (Gradient Boosting Regression). • GLM (Generalized Linear Model). • SVP (Support Vector Machine). • HSVR (Hybrid Support Vector Regression). • Combinations: (i) IDW-ANN, (ii) IDW-GBRT, (iii) TIN-ANN, and (iv) TIN-GBRT.
		Results/Remarks	• "*The combination of IDW-ANN becomes the proper method for downscaling the climate model output for both temperature and precipitation under the limitation of data quality*" (*sic*).
	Resources	N/A	
	Keywords	Interpolation; ML; Climate model downscaling; Data gaps; GCMs; Deep learning.	
	Tag/Type	Climate Change Hazard data.	
	Application in ICARIA	Potential use in ICARIA	Climate model downscaling through interpolation of existing data.

DD10. A 'Total' Imputation Algorithm that Fills Gaps in Time Series Measurements for ADEV and Phase Noise Characterizations of Power-law Noise Models

Table 29: DD10's table recipe.

Abbrev.	Categories and data	
CH-EV-DD-10	Title	A 'Total' Imputation Algorithm that Fills Gaps in Time Series Measurements for ADEV and Phase Noise Characterizations of Power-law Noise Models (Howe et. al., 2022). [Link]

Abbrev.	Categories and data		
	Summary		This recipe introduces an imputation algorithm for data gaps occurring in live measurements, by extending a T-length data run and enhancing long-term ADEV(τ) estimation, consistently recovering gaps across various power-law noise models.
		Variables (input)	• NIST H-maser time series measurements: (I) Clock, and (ii) oscillator phase measurements.
		Methods/Models	• The Total imputer is an effective method yet devised in filling data gaps for: (i) computations of ADEV (Allan Deviation), and (ii) phase noise levels over the fullest possible range of τ-values and Fourier-frequencies.
		Results/Remarks	• Total imputation algorithm. • Equally spaced time-series data without gaps. • Treatment of large gaps.
	Resources	• Time-series-imputation package. • Executable for Gap-filling Script for Noisy Time Series – Zenodo. • Gap-filling Script for Noisy Time Series - Zenodo. • Characterizing Frequency Stability Measurements Having Multiple Data Gaps.	
	Keywords	Time series; Noise model; Imputation Algorithm; Allan Deviation; Data models; Large data gaps.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Integration of missing modelling variables, applicable to any kind of data gap.

DD11. A data filling methodology for time series based on CNN and (Bi)LSTM neural networks

Table 30: DD11's recipe table.

Abbrev.	Categories and data	
CH-DD-11	Title	A data filling methodology for time series based on CNN and (Bi)LSTM neural networks (Tzoumpas et. al., 2022). (Link)
	Summary	This recipe develops a method combining deep Learning models such as CNNs, LSTMs, and BiLSTMs to fill data gaps in internal temperature time series from monitored apartments, using both pre- and post-gap data, correlated external temperature data,

Abbrev.	Categories and data		
			the method accurately reconstructs the target time series, outperforming baseline deep-learning architectures.
		Variables (input)	• Data from monitored apartments. • Data from sensors for moisture, humidity, temperature, CO2 concentration. energy consumption.
		Methods/Models	• CNNs, LSTMs and BiLSTMs. • CNN-LSTM (256, 128 and 64 neurons). • CNN-BiLSTM (32, and 16 neurons). • Use both networks in pre- and post-gap data.
		Results/Remarks	• Time-series reconstruction. • CNN-BiLSTM is the most promising model. • CNN-BiLSTM has the best-performing approximation of the time series. • CNN-LSTM model performs better in generalizing its predictions. • CNN-BiLSTM and CNN-LSTM models show a promising ability to generalize to unseen data. • Both models outperform purely LSTM networks.
	Resources	• Feedforward and LSTM Neural Networks, Gap filling. • Deep learning and time series forecasting. Review paper. • SINFONIA Project.	
	Keywords	Neural Networks; Data filling; Time series; Sensor data; Deep learning; High-resolution heat wave; Hazard assessment.	
	Tag/Type	Climate Change and Hazard Data.	
	Application in ICARIA	Potential use in ICARIA	Improvement of heat wave hazard characterization in urban areas based on existing monitoring devices in buildings, understanding effect of air temperature variation and urban heat island on indoor comfort, as a proxy of thermal capacity of building envelope and HVAC systems efficiency.

DD12. Increasing the detail of European land use/cover data by combining heterogeneous data sets

Table 31: DD12's recipe table.

Abbrev.	Categories and data		
EV-DD-12	Title		Increasing the detail of European land use/cover data by combining heterogeneous data sets (Rosina et. al., 2020). [Link]
	Summary		This recipe presents methods to improve the spatial resolution of land cover and land use data through the combination of different datasets available through Copernicus Land Monitoring System.
		Variables (input)	• European Settlement Map • Corine Land Cover • Copernicus High Resolution Layers • Urban Atlas • TomTom Multinet Polygons • OpenStreetMap • Local Sub-National Land use data
		Methods/Models	• Spatial refinement based on the cartographic synthesis of categorical raster data, interval raster data, and vector polygon data. • Data fusion performed using an automated chain of raster-based map algebra operations on a set of raw or pre-processed datasets. • Input vector data rasterized to the target 100m resolution beforehand, using the maximum combined area method to identify the dominant class in each cell. • At each step of the sequence, the cells either remain unchanged or are updated by the overlaid input data layer, following pre-established decision rules. • Random forest classification as machine-learning technique to predict land use classes using the derived predictor variables.
		Results/Remarks	• Urban fabric classification by use. • Vegetation classification.
	Resources	• ENACT project	
	Keywords	Data fusion; Land use; Land cover; Machine learning; Points of interest	
	Tag/Type	Exposure and Vulnerability data.	

Abbrev.	Categories and data		
	Application in ICARIA	Potential use in ICARIA	Case study tailored improvement of exposure and vulnerability information based on open datasets.

DD13. Power Network Component Vulnerability Analysis: A Machine Learning Approach

Table 32: DD13's recipe table.

Abbrev.	Categories and data		
EV-DD-13	Title		Power Network Component Vulnerability Analysis: A Machine Learning Approach (Anand et. al., 2021). (Link)
	Summary		This recipe suggests using data-driven methodologies on publicly available large-scale data to gauge power network vulnerability, elevate grid stability, minimize failure risks, and boost resilience for smart grids.
		Variables (input)	• State of components after an extreme event: (i) non-operational (outage), (ii) operational (in service). • Network components: (i) power plants, (ii) transmission lines, (iii) substations. • Extreme hazard(s): historic disruptive events data as input. Data sourced from the NOAA website. • Variables: (i) operation capacity, (ii) total number of lines, (iii) risk factor(s), (iv) vulnerability index, and (v) disruption distance.
		Methods/Models	• Data-driven methodologies: Supervised learning model: Support Vector Machines (SVMs): (i) 3216 components for training the model, (ii) 1345 operational components under thunderstorms winds, and (iii) 1309 operational components under tornadoes. Kernels: (i) linear, (ii) gaussian, (iii) polynomial, and (iv) sigmoid. • Calculation of the vulnerability index.
		Results/Remarks	• Investigation of critical components under disruptions in a network. • Consideration of an extreme event.

Abbrev.	Categories and data		
			• Maps of in-service or outrage components. • Output for: (i) the evaluation of the network stability, (ii) the understanding of the risk of cascade failure, and (iii) the improvement of the resilience of the overall network.
	Resources	N/A	
	Keywords	Power network resilience; Vulnerability analysis; Machine learning; Predictive analytics; Extreme events.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Enhance the predictability of resilience methodologies for power networks.

Expert elicitation methods

EE1. ELICIPY 1.0: A Python online tool for expert elicitation

Table 33: EE1's recipe table.

Abbrev.	Categories and data		
CH-EV-EE-1	Title		ELICIPY 1.0: A Python online tool for expert elicitation (De' Michieli Vitturi et. al., 2024). (Link)
	Summary		Python tool to perform expert elicitation sessions through a framework that covers both the questionnaire collection and the analysis parts.
		Variables (input)	• Expert input through webforms (question label and extended text for in multiple languages; units of the expected answer; scale - uniform or logarithmic; range of admissible values for the elicited percentiles; question type - "*seed*" or "*target*". • Experts' weight.
		Methods/Models	• Cooke Classical method. • Expected Relative Frequency method.
		Results/Remarks	• Experts' weights with different weighting schemes. • Itemwise graphs for seed questions (including the text of the Itemwise graphs for target questions (including the text of the questions), together with a simplified probability density function and cumulative distribution plots of the DM. • Percentiles of target questions. • Optional graphs where multiple target questions could be visualized along with their percentiles. • Probability density functions and barplots for target questions, along with the percentile values for the used weighting schemes.
	Resources	• ELICIPY 1.0 GitHub	

Abbrev.	Categories and data		
	Keywords	Expert elicitation; Uncertainty quantification.	
	Tag/Type	Climate Change and Hazard data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Assess probability of compound events and cascading effects in selected case studies scenarios; Assess variables related to exposure and vulnerability analyses for specific single-hazards.

EE2. Using expert elicitation to strengthen future regional climate information for climate services

Table 34: EE2's recipe table.

Abbrev.	Categories and data		
CH-EE-2	Title	Using expert elicitation to strengthen future regional climate information for climate services (Grainger et. al., 2022). [Link]	
	Summary	This recipe explores the use of structured expert elicitation to access uncertainties for future climate changes as an extension to the results of climate model simulations.	
		Variables (input)	• Input: (i) Temperature, and (ii) Precipitation. • CMIP5 analysis: (i) Historical data from 1975-2005, (ii) Calculated periods: 2040s and 2080s, and (iii) RCPs: RCP2.6, 4.5, 6.0 and 8.5.
		Methods/Models	• Climate model outputs from CMIP5. • Structured expert elucidation: ○ Use of structured expert elicitation (SEE) for regional climate change. ○ Expert elicitation judgment (individually or in a group): (a) Provides additional information and knowledge that is absent from modelling approaches, and (b) Builds a framework for discussion between climate experts and regional stakeholders. ○ Snowball sampling is considered. ○ Estimates of future temperature and precipitation change are provided. ○ Sources of uncertainty in estimating

Abbrev.	Categories and data		
			long-term climate changes: rank sources based on their overall contribution. • Considering all possible GHG concentration scenarios
		Results/Remarks	• Narrower uncertainty ranges for deviations in both temperature and precipitation. • Framework for supporting adaptation decisions. • "*Practices are shaped by local epistemic, institutional and political cultures*" (*sic*). • "*SEE used alongside modelling approaches, can contribute to a richer understanding of regional climate knowledge for use in climate services*" (*sic*) • "*Elicitation methods should be considered within the 'toolbox' of approaches available to climate service providers*" (*sic*)
	Resources	N/A	
	Keywords	Knowledge quality assessment; Climate change adaptation; Yangtze; China; Assessing Climate Uncertainties; Expert elicitation.	
	Tag/Type	Climate Change and Hazard Data.	
	Application in ICARIA	Potential use in ICARIA	Adapt the methodology to assess probability of compound events and cascading effects in selected case studies scenarios.

EE3. Expert Elicitation: Using the Classical Model to Validate Experts' Judgments

Table 35: EE3's recipe table.

Abbrev.	Categories and data	
CH-EV-3-R	Title	Expert Elicitation: Using the Classical Model to Validate Experts' Judgments (Colson et. al., 2018) (Link)
	Summary	Review of thirty-three professionally contracted classical model studies that were performed between 2007 and March 2015 using the EXCALIBUR software package for structured expert judgement elicitation using Cooke's Classical Model.

Abbrev.	Categories and data		
		Variables (input)	• Expert input and weighting through the EXCALIBUR tool.
		Methods/Models	• Cooke Classical model
		Results/Remarks	• Scoring of individual variables • Scoring of average probabilities
	Resources	• [EXCALIBUR Windows Tool] • [EXCALIBUR Python Tool]	
	Keywords	Expert elicitation; Uncertainty quantification.	
	Tag/Type	Climate Change and Hazard data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Assess probability of compound events and cascading effects in selected case studies scenarios; Assess variables related to exposure and vulnerability analyses for specific single-hazards.

Uncertainty treatment methods

U1. How Certain is Good Enough? Managing Data Quality and Uncertainty in Ordinal Citizen Science Data Sets for Evidence-Based Policies on Fresh Water

Table 36: U1's recipe table.

Abbrev.	Categories and data		
EV-U-1	Title	How Certain is Good Enough? Managing Data Quality and Uncertainty in Ordinal Citizen Science Data Sets for Evidence-Based Policies on Fresh Water (Stankiewicz et. al., 2023). [Link]	
	Summary	This recipe focuses on the collection of data sets for water quality involving the active contribution of citizens, offering an additional way to study and treat data gaps and uncertainties.	
		Variables (input)	• Input: (i) Temperature, (ii) Oxygen parameters, (iii) Turbidity, and (iv) pH. • Indicators concerning: (i) Aquatic plants, (ii) Water flow, (iii) Water depth, and (iv) Riverbank characteristics.

Abbrev.	Categories and data		
		Methods/Models	• Water Blitz events: instructions combined with a sampling kit. • Measuring the nitrate-nitrogen and phosphate-phosphorus concentration as collected from the field kits. • Coordinated measurements via GPS.
		Results/Remarks	• Citizen science contributes to monitoring activities. • Development of resilient ways to interact with the aquatic ecosystems. • Contribution to awareness and reflection. • Addresses an uncertainty in a sustainable government. • Limitations: (i) Role of the citizen science in the community, (ii) Individual characteristics of water systems and parameters, and (iii) Public access to environmental data from government(s).
	Resources	N/A	
	Keywords	Social-ecological systems; Water quality, Monitoring; Uncertainty.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Uncertainty treatment methodology for trials and mini-trials.

U2. Where does scientific uncertainty come from, and from whom? Mapping perspectives of natural hazards science advice

Table 37: U2's recipe table.

Abbrev.	Categories and data	
CH-RV-U-2	Title	Where does scientific uncertainty come from, and from whom? Mapping perspectives of natural hazards science advice (Doyle et. al., 2023). [Link]
	Summary	This recipe focuses on identifying sources of uncertainty associated with natural hazards using mental model mapping and a semi-structured interview protocol.

Abbrev.	Categories and data		
		Variables (input)	• The area of study was exposed frequently to a wide variety of natural hazards. • A range of participants was recruited using the snowball approach. • In total twenty-five (25) participants ranged from twenty-five (25) to seventy-five (75) years old.

		Methods/Models	• Aims: (i) Understand what a disaster risk is, (ii) Integrate technology in decision-making about risk, and (iii) Find disaster risk communication methodologies.
			• A three-face interview was constructed "to understand individual's perceptions of uncertainty associated with natural hazards" + brainstorming + indirect elicitation questions.
			  ○ A systematic review of mental model interview approaches.
			  ○ Conceptual cognitive concept mapping (3CM).
			• Mental models approach:
			  ○ Key concepts: (i) Uncertainty, (ii) Knowledge, and (iii) Science.
			  ○ Sources of uncertainty: (i) The scientists, (ii) The media, (iii) The communicators, (iv) The range of possible outcomes, (v) Human responses, and (vi) The unknown unknowns.

		Results/Remarks	• Individual's mental models to identify sources of uncertainty: (i) Actors, and (ii) Known unknowns. • Translate uncertainty in a meaningful way for the people (the public). • Creation of science-policy interfaces for effective decision-making frameworks in disaster management crisis. • Key influences for uncertainty: (i) Governance and funding, (ii) Societal factors, (iii) Outcomes, (iv) Emotions, (v) The communication landscape, and (vi)

Abbrev.	Categories and data		
			Decision-making.
	Resources	• [National Emergency Management Agency's National Disaster Resilience Strategy](). • [Resilience to Nature's Challenge is one of Aotearoa New Zealand's National Science Challenges](). • [QuakeCoRE is the NZ Centre of Research Excellence for Earthquake Resilience]()	
	Keywords	Uncertainty; Mental models; Natural hazards; Societal and economic factors.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Communicating uncertainties associated with complex events hazard/impact scenario assessment to decision makers.

U3. A review of uncertainty quantification in deep learning: Techniques, applications and challenges

Table 38: U3's recipe table.

Abbrev.	Categories and data		
CH-U-3-R	Title	A review of uncertainty quantification in deep learning: Techniques, applications and challenges (Abdar et. al., 2021). ([Link]()).	
	Summary	This recipe presents the application of Bayesian and ensemble techniques in various domains discussing the recent advancements in uncertainty methods within deep learning for optimization and decision-making processes.	
		Variables (input)	• N/A
		Methods/Models	• Quantification methods: ▪ Bayesian techniques: (i) Monte Carlo (MC) dropout, (ii) Markov chain Monte Carlo (MCMC), (iii) Variational inference (VI), (iv) Bayesian Active Learning (BAL), (v) Bayes by Backprop (BBB), (vi)

Abbrev.	Categories and data		
			Variational autoencoders, (vii) Laplacian approximations, and (viii) Uncertainty quantification in reinforcement learning. ▪ Ensemble techniques: (i) Deep NNs (DNNs), (ii) Deep ensemble Bayesian/Bayesian deep ensemble, and (iii) Uncertainty in Dirichlet deep networks. ▪ Uncertainties: (i) Two main types of uncertainty: (a) epistemic (model uncertainty) and (b) aleatoric (data uncertainty). (ii) Three (3) uncertainly models were considered: (1) the MC dropout, (2) the Bootstrap model, and (3) the GMM model. ● Others: ○ Neural Architecture Distribution Search (NADs). ○ Single model estimates for DNNs of epistemic and aleatoric uncertainty. ○ Method to find and reject distribution data points for training a deterministic deep model with a single forward pass at test time. ○ MC-DropConnect. ○ Gradient-based optimization techniques. ○ Noise contrastive priors (NCPs) to estimate consistent uncertainty. ○ Uncertainty-based class imbalance learning. ○ Variational approximation, termed Bayes by hypernet. (BbH), deducting hypernetworks as implicit

Abbrev.	Categories and data		
			distributions. ○ I Do not Know (IDK) prediction cascade approach. ○ Models inspired by the nonlinear differential equations utilized by physics-informed neural networks. ○ ProbDepthNet. ○ DNNs trained with mix-up. ○ Local interpretable model-agnostic explanations (LIME). ○ Randomized approach sampling from the hidden layers. during the DNN interference period. ○ Certainty-driven consistency loss (CCL) method. ○ Modified knowledge distillation method. ○ Models based on kernel techniques. ○ Stochastic quantized activation distributions (SQUAD). ○ Probabilistic DL method (approximate Bayesian inference + heteroscedastic noise technique). ○ Gaussian Processes (GP). ○ Stochastic, low-rank, approximate natural gradient (SLANG) technique. ○ Dubbed prior networks (PNs). ○ DVERGE. ○ Direct epistemic uncertainty prediction (DEUP). ○ Subjective Bayesian GNN (S-BGNN). ○ Doubly stochastic variational neural process (DSVNP). ○ Non-Bayesian NN models. ○ Uncertainty-aware deep Dirichlet neural networks. ○ Deep Gaussian processes (DGPs): (i) In combination with stochastic weight averaging (SWA), (ii) SWA-Gaussian, (iii) GPDNNs: a hybrid model of GP and

Abbrev.	Categories and data		
			DNNs, (iv) GPs + YOLOv3, (v) A natural gradient-based algorithm for Gaussian mean-field, (vi) Matrix variate Gaussian (MVG), and (vii) Introduction of a variety of stochastic layers. ○ A variety of other techniques uniquely specified and tailored for desired applications are listed within the last subsections of the paper.
		Results/Remarks	• Gaps and methods to approach them: (1) Fusion-based methods, (2) Ensemble methods, (3) Decision making, (4) Active learning, (5) Transfer learning, (6) Neural architecture search (NAS) methods, (7) Self-supervised learning (SSL) methods, (8) Hypernetworks, (9) Continual learning, (10) GNNs: Graph Neural Networks, (11) BO: global optimization method for optimizing time-consuming black-box objective functions, and (12) Uncertainty calibration.
	Resources	N/A	
	Keywords	Uncertainty quantification; Deep learning; Machine learning; Bayesian statistics; Ensemble learning; Review article.	
	Tag/Type	Climate Change and Hazard Data.	
	Application in ICARIA	Potential use in ICARIA	Uncertainty quantification in propagation of damage following consecutive compound events and/or cascading effects (Climate Change and Hazard data).

U4. SHELF: The Sheffield Elicitation Framework

Table 39: U4's recipe table.

Abbrev.	Categories and data	
CH-EV-4	Title	SHELF: The Sheffield Elicitation Framework (Gosling et. al., 2018). (Link)
	Summary	R-based package of documents, templates and software to carry out elicitation of probability distributions for uncertain

Abbrev.	Categories and data		
			quantities from a group of experts.
		Variables (input)	• Expert input and weighting through the SHELF package of documents, templates and software.
		Methods/Models	• Single expert • Multiple experts • Bivariate elicitation • Dirichlet elicitation • Extension method (continuous) • Extension method (discrete)
		Results/Remarks	• Probability distribution with respect to elicited variables. • Multiple visual, graph and table formats.
	Resources	• SHELF Tool	
	Keywords	Expert elicitation; Uncertainty quantification.	
	Tag/Type	Climate Change and Hazard data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Assess probability of compound events and cascading effects in selected case studies scenarios; Assess variables related to exposure and vulnerability analyses for specific single-hazards.

U5. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System

Table 40: U5's recipe table.

Abbrev.	Categories and data		
EV-U-5	Title	Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System (Van Der Sluijs et. al., 2005). [Link]	
	Summary	This recipe showcases the applicability of a system designed to combine quantitative and qualitative uncertainty measures, demonstrating its effectiveness for accessing both parameter uncertainty and model assumptions.	
		Variables (input)	• Data from environmental policy issues, for example, emissions of acidifying gases (NOx,

Abbrev.	Categories and data		
			SO_2, and NH_3).
		Methods/Models	• Treatment of multidimensional uncertainty assessment with increasing complexity. • Emission monitoring systems, complex energy models, and environmental indicators are used. • NUSAP is applied to complex models in a meaningful way. • Ability to serve as a diagnostic tool for assessing the robustness of a given knowledge base for policy-making.
		Results/Remarks	• Numeral Unit Spread Assessment Pedigree (NUSAP) system for multidimensional uncertainty assessment. • Potentially applicable for ICARIA use cases/trials, etc... • A tool for prioritizing uncertainties qualitatively and quantitatively.
	Resources	• A framework to assess quality and uncertainty in disaster loss data	
	Keywords	Uncertainty; Controversy; Value-laden assumptions; Problem frames; Diagnostic analysis.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Uncertainty treatment methodology for trials and mini-trials.

Other methodologies related to hazard, exposure, and vulnerability

HEV1. Urban pluvial flood modelling in the absence of sewer drainage network data: A physics-based approach

Table 41: HEV1's recipe table.

Abbrev.	Categories and data		
CH-EV-HEV-1	Title		Urban pluvial flood modelling in the absence of sewer drainage network data: A physics-based approach (Montalvo et. al., 2024). [Link]
	Summary		This recipe focuses on a physics-based method for assessing urban pluvial floods using a virtual sewer network generation tool when sewer network data is scarce. Comparing results from four storm events, the method effectively, and accurately represented drainage capacity and accounting for sewer overflows, confirming its robustness for urban flood modelling.
		Variables (input)	• Information from a virtual sewer network. • Information from 1D/2D models using the actual sewer network. • Stormwater flow (synthetic design storm parameters). • Wastewater flow (based on the population density, land registry, and daily per capita water capacity). • Number of rainfall events. • Network elements: manholes, outfalls, and conduits.
		Methods/Models	• Physics-based assessment of urban pluvial floods. • 1D/2D hydrodynamic dual drainage model Iber-SWMM. • 2D dynamic sewer mode: the Soil Conservation Service Curve Number was used. • 1D dynamic sewer model: (i) EPA SWMM (Storm Water Management Model), and (ii) Allows retrieval of hydraulic variables. • Creation of a representative virtual sewer network: Virtual sewer network generation and dimensioning tool: (i) Defines realistic network topology, (ii) Manholes are located where streets intersect with each other, and (iii) Manhole invert elevations are calculated. • Scenarios: (i) 1D/2D dual model using the actual network, (ii) 2D overflow model without sewer network, (iii) 2D overflow model with rainfall reduction for sewer network representation, (iv) 1D/2D dual model using a

Abbrev.	Categories and data		
			virtual sewer network, and (v) 2D overflow model using a virtual inlet layout.
		Results/Remarks	• Estimates accurately the sewer's networks drainage capacity (SNDC). during pluvial floods. • Soil Conservation Service Curve Number maps are created. • Virtual and real sewer network maps are provided. • Numerical simulated maximum inundation extent maps for all five (5) scenarios are provided. • Scenarios with the actual network and without the sewer network indicate the importance of the SNDC. • Scenario using the rainfall induction method are not efficient. • The virtual sewer scenario resulted in the most effective estimations in the absence of any sewer network information. • The SNDC of the virtual network is more extended when compared to the actual network. • Evaluation through satellite images might contribute to a optimal performance of the proposed method in this recipe.
	Resources	• Meteorological data from the MeteoGalicia agency.. • Observed sea level data from the Spanish Port system.. • Digital elevation model from the Spanish Ministry for Ecological Transition and Demographic Challenges.. • Land Registry, soil use, and population information from the Spanish General Directiorate of the Land Registry. • The Street network layout from OpenStreetMap geodatabase. • The LiDAR-derive digital elevation model, observed water elevation data form the gause station, and the sewer network layout from the regional water administration Augas de Galicia.	
	Keywords	Urban pluvial flooding, Dual models, Iber, SWMM, Data scarcity.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	

Abbrev.	Categories and data		
	Application in ICARIA	Potential use in ICARIA	A potential methodology for the creation of a virtual network might be useful for the lab tests in ICARIA project.

HEV2. Storm damage beyond wind speed – Impacts of wind characteristics and other meteorological factors on tree fall along railway lines

Table 42: HEV2's recipe table.

Abbrev.	Categories and data		
CH-EV-HEV-2	Title		Storm damage beyond wind speed – Impacts of wind characteristics and other meteorological factors on tree fall along railway lines (Lorenz et. al., 2024). [Link]
	Summary		This recipe examines the tree fall risk during hazard events, emphasizing the role of meteorological factors and conditions that influence tree falls, supporting the addition of local climatological conditions for improved risk assessment.
		Variables (input)	• Parameters: (i) Wind speeds, precipitation, soil water volume, air density, and the precipitation sum of the previous year increase tree fall risk, and snow. • Datasets: (i) Deutsche Bahn (2017-2021) and meteorological data from ERA5 reanalysis and RADOLAN radar, (ii) Tree fall events along the German railway network were derived from a data set created by the Deutsche Bahn, (iii) It contains 15311 tree fall events between 2017 and 2021, (iv) The dataset ranging from 2017 to 2021 covering the whole country and including long-term and large-scale storm damage contributes to the novelty of the recipe, (v) Hourly ERA5 data for all meteorological parameters except precipitation accessed using the ClimXtreme Central Evaluation System and (vi) Rail density index.
		Methods/Models	• -A logistic regression model predicts the risk of a tree falling on a railway line in a 31 km grid cell.

Abbrev.	Categories and data		
			• The analysis considered only the extended winter season, focusing on winter windstorms, which cause the most extreme peaks in tree fall events. • Meteorological predictors like precipitation or soil moisture are considered less often: (i) Predictors describing precipitation and soil conditions at different time scales are also considered, (ii) Wind load can be considered as a model predictor, and (iii) Interactions can reveal the combined effect of predictors and their interconnection. • Storm duration, gust factor, and air density are important factors in calculating the risk of tree fall. • Inclusion of antecedent weather situations; Index: The Standardized Precipitation-Evapotranspiration Index (SPEI), which has been used in recent research on forest disturbance. • Snow and soil frost: (i) Potentially influential variables, and (ii) Derived from ERA5, which tends to overestimate snow water equivalent in the Northern Hemisphere. • Limitations: (i) The events might be related to meteorological events not resolved by the ERA5 analysis, (ii) The wind speeds caused by heavy thunderstorms are likely to be underestimated, and (iii) Data with higher spatial resolution that include convective effects were not included (helpful in understanding the effects of the phenomena).
		Results/Remarks	• Duration of strong winds is important because trees do not fail instantly but fail with repeated swaying that fractures the root/soil system and this process can take many hours. • Future modelling might benefit from the addition of local tree wind exposure. • High and prolonged wind speeds, especially in combination with wet conditions (high

Abbrev.	Categories and data		
			precipitation and high soil moisture) and a high air density, increase the risk of tree fall. ● Predictors for daily precipitation, daily soil water volume, and daily maximum gust speed might improve the model's skill. ● Previous trees fall and forest storm damage events are restricted to a single event or a small research region. ● Wind-related parameters (e.g., gust factor, duration of strong wind speeds, air density, etc.) and predictors related to meteorology, have a significant impact on tree fall risk. ● Taking tree adaptation to the environment should be considered. ● Models that can add trees, soil, or stand data or have access to higher spatial resolution meteorological data will likely produce better model skills and be able to examine the relationships between tree fall and meteorology in more detail.
	Resources	N/A	
	Keywords	Windstorm; Railway system; Trees; Logistic regression model; Gust speed; Meteorological parameters.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability Data.	
	Application in ICARIA	Potential use in ICARIA	Improved assessment of windstorm impact on trees, support to vulnerability analysis of trees under extreme winds, potentially correlated to cascading impacts on transport and energy networks.

HEV3. OpenStreetMap for multi-faceted climate risk assessments

Table 43: HEV3's recipe table.

Abbrev.	Categories and data	
CH-EV-HEV-3	Title	OpenStreetMap for multi-faceted climate risk assessments (Mühlhofer et. al., 2024). [Link]
	Summary	This recipe presents specific approaches to exploit OSM data and tools for informing hazard/impact assessments.

Abbrev.	Categories and data		
		Variables (input)	• OpenStreetmap data, and automation: Geofabrik. • Automatic pipeline with CLIMADA. • Variables: (i) Wind: for forests, wind intensities above 38.9 m/s were considered more adequate to capture tree snapping, (ii) Nightlight intensity and population (count at 30 arcsecond resolution with LitPop), and (iii) A population exposure at the same resolution based on the SEDAC GPW v4.0 dataset. • Exposure data for heritage sites, forests, etc...
		Methods/Models	• Retrieval of geospatial exposure data from OSM. • Features can be extracted from OSM and converted into geographical tabular format: (i) By reading data directly from the Overpass API, (ii) By downloading regional data dumps as protocol buffer binary format files. • Technical details: (i) Python-based (tabular formats, geopandas, etc.), (ii) Efficiently parses large sets of OSM data based on user-specified queries from PBF data dumps within arbitrary and fully user-defined geographical boundaries, and (iii) Computational efficiency and user flexibility required to perform multi-faceted risk analyses. • Integration within the natural hazard risk assessment platform CLIMADA and perform end-to-end assessments. • OSM with CLIMADA: (i) Compute risk according to the IPCC risk definition as the product of hazard, exposure, and vulnerability, (ii) CLIMADA's engine is designed only for pointwise data, thus data must be interpolated to points before impact calculations, and (iii) Dedicated hazard and vulnerability data can be provided in the CLIMADA data API. • Risk management strategies, opening the potential for non-conventional exposure data.

Abbrev.	Categories and data		
		Results/Remarks	• For forests and general asset values, physical losses and damages are modeled. • Vulnerability curves relating to hazard intensity to damage extent were obtained for the wind-induced general asset damages. • The possibility of retrieving an even larger variety of features (ecological regions, critical infrastructure, urban assets, etc.) remains open. • Comparison with print media accounts, official records, and insurance reports, strengthens the picture of multi-faceted impacts obtained from computations. • The standard impact computations within CLIMADA capture these metrics as asset damages and affected population. • Considering probabilistic risk metrics to gauge the potential risk landscape and to adequately place the occurrence of historic events therein might be considered beneficial.
	Resources	• OSM-flex package – GitHub repository. • Polygon files extraction – GitHub repository.	
	Keywords	OpenStreetMap, open-source GIS tools, climate risk assessment, natural hazards, adaptation, Urban areas, Hazard local effect; Vulnerability analysis.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Integration of high resolution local geospatial data variables to improve hazard characterization (e.g., run-off of different land cover types), exposure and vulnerability analysis (e.g., classification of buildings, road network, open spaces) depending on specific impact models input.

HEV4. On the positioning of emergencies detection units based on geospatial data of urban response centres

Table 44: HEV4's recipe table.

Abbrev.	Categories and data	
EV-HEV-4	Title	On the positioning of emergencies detection units based on geospatial data of urban response centres (Peixoto et. al., 2023). [Link]
	Summary	This recipe proposes a data-driven methodology for the optimal placement of multi-emergency detection units in smart cities, combined with geospatial data on urban infrastructure, ultimately, to define mitigation zones and enhance urban resilience to emergencies.
	Variables (input)	• Input from mitigation response centers and their geospatial metadata (files in XML format). • Input from existing urban infrastructure. • General data: (i) Extraction area: OpenStreetMap, (ii) PoIs and roads: OpenStreetMap, (iii) Mitigation zones (MZ), and (iv) Positions of Emergency Detection Units (EDU). • Parameter for algorithms: Severity index of the zones.

	Methods/Models	• Mathematical definitions of the zone of interest, mitigation zone, point of interest, and severity index are proposed.
		• Algorithms to indicate EDUs' positions:
		  ○ Random: (i) Lowest computational cost, and (ii) No criteria for selecting the units' positions.
		  ○ Balanced: (i) Utilized after the definition of MZs, (ii) Avoid uncovered areas and EDUs overlapping, (iii) Calculates coverage radius of EDU in each mitigation level so units avoid overlapping, and (iv) No guaranteed coverage around the level's borders.
		  ○ Balanced+: (i) Improves Balanced algorithm positioning unit correctly covering around the level's borders, and (ii) No uncovered areas within the area of influence.
		  ○ Restricted: (i) Add restrictions to the optimization problem for improving

Abbrev.	Categories and data		
			the applicability of the units' deployment, (ii) Cases where actual deployment is not possible, and (iii) Most computationally demanding. ○ Computation of (i) the risk perception function, (III) the equation for a mitigation level of a zone, (iv) the total number of EDUs, (v) the total number of EDUs per mitigation level. ● Mitigation zones act as data layers of a target area. ○ Based on the indirect relation between geospatial data, and urban hazards and emergencies. ○ The resilience to untreated emergencies based on the urban infrastructure was considered.
		Results/Remarks	● The concept of mitigation zones is proposed. ● The placement of Emergency Detection Units is considered as a potential issue for macrosystems solutions for entire cities. ● Maps of mitigation zones and response centers for an area of influence. (i) Depiction of the zones with low, medium, and high mitigation risk, and (ii) Depiction of the positions of the EDUs within the zones of risk to highlight uncovered areas.
	Resources	N/A	
	Keywords	Smart cities; Sustainable and resilient cities; Emergency detection, Sensors positioning; Urban infrastructure monitoring, Vulnerability analysis of service networks.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Sensor-based model to enhance the real time monitoring of service network (e.g., electricity, transport) and improve vulnerability analysis over time through monitoring.

HEV5. Advancing building data models for the automation of high-fidelity regional loss estimations using open data

Table 45: HEV5's recipe table.

Abbrev.	Categories and data		
EV-HEV-5	Title		Advancing building data models for the automation of high-fidelity regional loss estimations using open data (Angeles et. al., 2022). [Link]
	Summary		This recipe presents a conceptual data model to integrate and query detailed building information, substantiated by a case study showing the model's effectiveness in generating models for accurate loss estimation.
		Variables (input)	• Spatial and geometric modelling data (Fragility library available).
		Methods/Models	• Building damages: (i) Assembling and managing spatial and geometric data of thousands of constructed buildings, and (ii) Building Topology Ontology (BOT). • Hazards: (i) Wind, wind-driven rain, wind-borne debris, (ii) Specification on site-specific building geometries and locations paired with DEM, and (iii) Rain Admittance Factor (RAF) to discern the intensity of horizontal rainfall. • Response: (i) Realistic loading characterizations for each actual constructed building, (ii) Supporting holistic simulation of each building's unique load path to derive consequent response quantities, and (iii) The automated development of structural analysis models to incorporate engineering demand parameters for specific components. • Damage: (i) Automated probabilistic analyses in the damage assessment, (ii) Scenarios → global fragility or vulnerability curves for each generic building model, (iii) Fragility descriptions for wind-vulnerable building elements, (iv) Component-specific fragility models, and (v) Assembly-based vulnerability (ABV) approaches. • Scenarios: (i) Scenario 1: Hazard Analysis: Given a reference building and a site, find all buildings on the site that are within a given

Abbrev.	Categories and data		
			distance from the reference building, (ii) Scenario 2: Response Analysis (Fault Trees): Given a breach in the building façade, quantify the new internal pressure per story and determine which elements require load recalculation, (iii) Scenario 3: Damage Analysis: Determine the correlation coefficient, between the damage states of the i-th and j-th elements, considering elements of the same type, and (iv) Scenario 4: Loss Analysis: For a given story within a building, calculate the total surface areas of each element sub-class and type.
		Results/Remarks	N/A
	Resources	N/A	
	Keywords	Hurricane, Commercial, Regional loss estimation, Data model, Open data.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Multi-hazard exposure and vulnerability classification and analysis for buildings structural damage assessment.

HEV6. Estimating exposure of residential assets to natural hazards in Europe using open data

Table 46: HEV6's recipe table.

Abbrev.	Categories and data		
EV-HEV-6	Title	Estimating exposure of residential assets to natural hazards in Europe using open data (Paprotny et. al., 2020). [Link]	
	Summary	This recipe offers a methodology based on a non-parametric Bayesian network model using open data to estimate residential asset exposure to natural hazards in various European capitals, providing improved national-level economic valuations of residential properties.	
		Variables (input)	• Data: (i) building footprints (OpenStreetMap), (ii) high-resolution city models, (iii) pan-European raster datasets, (iv) historical hazard events, (v) Country-level

Abbrev.	Categories and data		
			socioeconomic data, and (vi) Alternative country-level asset value estimates. • Main sources: (i) Eurostat, (ii) OpenStreetMap, (iii) HANZE database, and (iv) Copernicus Land Monitoring Service.
		Methods/Models	• Calculations: (i) Identification of residential buildings, (ii) Models for building height prediction, (iii) Calculation of floor space, (iv) Residential building stock estimation, (v) Household content stock estimation. • Application of a building-level damage model. • The estimation of building height and number of floors was based on a non-parametric Bayesian network (BN), a probabilistic model allowing multivariate dependency analysis and uncertainty distributions of the predictors. • Country-level asset validation of buildings and households: (i) Perpetual inventory method (PIM): estimate the value of a stock (e.g., stock of dwellings), and (ii) Households: memorandum items in ESA 2010 were considered. The PIM method was applied once again. Annual investment was calculated based on the Classification of Individual Consumption by Purpose (COICOP).
		Results/Remarks	• Average national-level gross replacement costs of the residential assets are computed. • Output information: (i) Building-level asset value estimates for a test study area, (iii) Country-level building/content value per square space of floor space. • Validation statistics for the building height prediction model and pan-European estimations of content value both for buildings and households. • Limitations: (i) datasets can be affected by regional methodological specifics (e.g., rural areas), (ii) Quality of expenditure data are not robust given the divergence in deflators for individual items, (iii) households stock also

Abbrev.	Categories and data		
			semi-perishables and perishables which are excluded from any wealth assessments due to lack of information, (iv) non-homogeneous datasets, (v) the BN model is configured using expert knowledge. ● This methodology might be extended: (i) to calculate past recorded damages from natural hazards, (ii) to calculate the average quality of residential buildings and households' incomes and (iii) to rescale absolute damage functions.
	Resources	● BN model code (in Matlab) is available upon request. ● UNINET Tool (free for academic purposes). ● Data retrieval and processing in formats other than GIS utilized GDAL/OGR tools. ● Flood damage in the HOWAS21 database. ● Additional resources of data for estimating residential building value can be seen in the supplementary information, Table S3. ● Additional information on Consumption expenditure categories by COICOP 3-digit codes and durable items included in those categories by COICOP 4-digit codes can be seen in the supplementary information, Table S6. ● Data for the household final consumption expenditure can be seen in the supplementary information, Table S8.	
	Keywords	Residential assets; Bayesian-based models; Buildings Exposure; Buildings Vulnerability.	
	Tag/Type	Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Classification of buildings features supporting exposure and vulnerability analysis of residential assets under different natural hazards.

HEV7. Asset exposure data for global physical risk assessment

Table 47: HEV7's recipe table.

Abbrev.	Categories and data	
EV-HEV-7	Title	Asset exposure data for global physical risk assessment (Eberenz et. al., 2020). [Link]

Abbrev.	Categories and data		
	Summary		This recipe introduces a transferable, high-resolution asset exposure dataset using the LitPop methodology combining nightlight intensity and population data to improve the spatial distribution of asset values, to enhance economic disaster risk assessments and climate change adaptation methods.
		Variables (input)	● Datasets: (i) Gridded nightlights (Lit); satellite data from Nasa Earth Observatory/Black Marble product, (ii) Gridded population (Pop); non-spatial population and cartography data, (iii) Produced capital, (iv) GDP-to-wealth ratio, (v) GDP and GRP. ● GDP does not directly measure physical assets but fills data gaps for the evaluation of the LitPop methodology.
		Methods/Models	● LitPop downscaling: Lit and Pop data produce a gridded digital number and are combined with the total asset value per country to obtain asset exposure data and to compare GDP (macroeconomic output indicator) against the GRP for evaluation of the approach. ● Total assets per country and GDP are distributed and calculated according to a function of nightlight luminosity and population count. ● Ipix; the asset value per cell grid; Itot: represents either asset value or GDP. ● Evaluation: ten combinations of Lit and Pop are assessed ($Lit^{m}Pop^{n}$). ● Limitations: (i) Asset distribution assumes physical wealth is distributed equally, (ii) Assets assumed to be located exactly where people live, (iii) Population data in many countries are coarse, (iv) Nightlight-based models are prone to saturation and blooming limitations, (v) Lack of reference asset value data on subnational level, (vi) Lack of consistent exposure data on a global scale, and (vii) The methodology does not include infrastructure type and vulnerability.
		Results/Remarks	● LitPop-based asset exposure data are accessible and usable as a basis for global

Abbrev.	Categories and data		
			comparable economic risk assessments. ● The application of the asset exposure data for local assessments in countries within low-income groups should be treated with caution. ● Lit[1]Pop[1] combination distributes favors GDP distribution to the subnational level than the other combinations of nightlight and population data accessed. ● Additional sector-specific asset inventories and auxiliary data should be considered. ● High-resolution asset exposure for LitPop combinations is estimated.
	Resources	● Asset exposure for 224 countries. ● Open-source software for adaptation: CLIMADA. ● LitPop module and Tutorial on LitPop calculations.	
	Keywords	Asset exposure; Nightlight intensity; economic assessment; risk assessment; climate change.	
	Tag/Type	Exposure and Vulnerability methodology.	
	Application in ICARIA	Potential use in ICARIA	Global high-resolution asset exposure methodology, applicable for estimating economic impact assessment under different natural hazards.

HEV8. Mapping Europe into local climate zones

Table 48: HEV8's recipe table.

Abbrev.	Categories and data		
CH-EV-HEV-8	Title	Mapping Europe into local climate zones (Demuzere et. al., 2019). [Link]	
	Summary	This recipe constructs a European database, focusing on characterizing urbanized landscapes, offering dedicated datasets for the training areas.	
		Variables (input)	● WUDAPT data. ● Products: (i) Sentinel-1, (ii) Sentinel-2, (iii) Defense Meteorological Program Operational

Abbrev.	Categories and data		
			Linescan System: Nighttime lights, and (iv) Local Climate Zone maps (LCZ) and their parameters. ● Indexes based on Earth Observation data: (i) the minimum and maximum Normalized Difference Vegetation Index (NDVI), (ii) the Biophysical Composition Index (BCI), (iii) the mean Normalized Difference BAreness Index (NDBAI, (iv) the mean Enhanced Built-up and Bare land Index (EBBI), (v) the mean Normalized Difference Water Index (NDWI), (vi) the mean Normalized Difference Built Index (NDBI), (vii) the Normalized Difference Urban Index (NDUI: the combination of the maximum values of NDVI, NDWI and NDBI indexes with the coarser resolution nighttime light imagery.
		Methods/Models	● A step-by-step methodology in creating a general-use European Local Climate Zone map that can be used for climate studies is presented. The derived datasets can be used as substitute data to cover limitations in sector-specific local data gaps. ● Evaluation of LCZ map techniques: (i) Urban land cover, (ii) Impervious surface cover (IMD) and building height (BH), (iii) Anthropogenic heat flux (AHF), and (iv) Sky view factor (SVF).
		Results/Remarks	● Map of the European LCZ classification based on the random forest classifier, binary urban maps, and assessment of urban land cover. ● Urban canopy parameters: (i) Building height, (ii) Maps of impervious surface fraction (IMD) and anthropogenic heat flux (AHF), Sky view factors map.
	Resources	● The European LCZ map is available from the official WUDAPT (World Urban Database) data portal. ● There is an LCZ generator provided at https://lcz-generator.rub.de/. ● LCZ datasets derived from the use of the generator are provided at https://lcz-generator.rub.de/submissions.	
	Keywords	European Local Climate Zone map; E-OBS data.	

Abbrev.	Categories and data		
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Improved hazard assessment based on urban/land morphology and cover. Classification of urban density, building types and land cover.

HEV9. CLIMADA v1: a global weather and climate risk assessment platform

Table 49: HEV9's recipe table.

Abbrev.	Categories and data		
CH-EV-HEV-9	Title	CLIMADA v1: a global weather and climate risk assessment platform (Aznar-Siguan et. al., 2019). (Link)	
	Summary	This recipe presents an open-source, modular multi-hazard decision support tool for assessing extreme events and socioeconomic impact by hazard, exposure, and vulnerability data, supporting scalable, parallel computations and multi-hazard probabilistic assessment.	
		Variables (input)	• Hazard events for: storms, floods, droughts, heatwaves. • Socioeconomic aspect: exposure and impact (vulnerability) functions. • Exposure: e.g., geographical distribution of people, livelihoods, infrastructure, services, etc. • Impact functions: impact of a hazard on the corresponding exposures. • Nighttime lights of NASA's Black Marble 2016. • GDP values. • International Best Track Archive for Climate Stewardship (IBTrACS) archive for tropical cyclones.
		Methods/Models	• A fully probabilistic risk assessment modelling methodology is combined with economic development and climate impact scenarios to assess adaptation measures.

Abbrev.	Categories and data		
			• Engine's risk metrics: (i) Expected annual impact (EAI), (ii) Average annual impact (AAI), (iii) Probable maximum impact (PMI), and (iv) Impact at the event. • Medium (10 km) to high (500 m) resolution.
		Results/Remarks	• Supports risk management options and adaptation measures. • Estimates: (i) the expected socioeconomic impact of weather and climate, (ii) the incremental increase from economic growth, (iii) the incremental increase due to climate change. • Characteristics: of CLIMADA (i) open source, (ii) modular, (iii) scalable.
	Resources	• CLIMADA GitHub repository and scripts for the recipe. • ETH Data Archive • Shaping climate-resilient development: a framework for decision-making	
	Keywords	Risk assessment; Socioeconomic impact; Probabilistic approaches; Damage estimation.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Open-source platform for multi-hazard calculations suitable for regional to whole-country economic impact assessment studies.

HEV10. Comparing an insurer's perspective on building damages with modelled damages from pan-European winter windstorm event sets: a case study from Zurich, Switzerland

Table 50: HEV10's recipe table.

Abbrev.	Categories and data	
CH-EV-HEV-10	Title	Comparing an insurer's perspective on building damages with modelled damages from pan-European winter windstorm event sets: a case study from Zurich, Switzerland (Welker et. al., 2019). [Link]
	Summary	This recipe highlights the benefits of a probabilistic approach for assessing rare events impacts and uncertainties, for

Abbrev.	Categories and data		
			claims-based risk assessment with test examples of the risks of winter windstorms in Europe.
		Variables (input)	• OpenStreetMap data. • Hazard input: (i) Windstorm Information Service (WISC): Windstorms, and (ii) GVZ database: damages of past events. • Other input: (i) Historic windstorm hazard set: 140 windstorm events in Europe (1940-2014); "*WISC operational*", (ii) Windstorm footprints (computed running the UK Met Office Unified Model at 4.4km resolution with ERA-20C reanalysis and ERA-Interim analysis; "*WISC synthetic*", and (iii) Insurance claims data.

		Methods/Models	• Probabilistic windstorm hazard extension; "*WISC probabilistic extension*": perturbations in the WISC historic events are introduced creating new probabilistic footprints (scenarios) of possible hazard events (windstorms in this recipe).
			  ○ The frequencies of the footprints are estimated to recreate the cumulative distribution function of generalized extreme values fitted to historic events.
			  ○ 4118 probabilistic events and 142 original events were considered.
			• Damage modelling: the damage results are based on: (i) the intensity of the hazard event, (ii) the value of the assets, and (iii) the susceptibility of the asset to damage.
			  ○ GVZ damage model (proprietary): (i) uses a dedicated building database, (ii) only buildings affected by gusts with speed more than 90km h^{-1} are considered, (iii) the value of each building is multiplied by the mean damage degree (MDD) factor (obtain from vulnerability curves) to estimate damage, and (iv) provides the probability of building affected using a stochastic approach.

Abbrev.	Categories and data		
			○ CLIMADA impact model: (i) exposure is based on public data (the total value of the physical assets, the nightlight intensity, and the population density are included), (ii) uses MDD curves for exposure evaluation. • Damage and risk assessment: risk = extent of damage x probability of damage. Metric for risk assessment in this recipe: (i) Average annual damage (AAD), (ii) Exceedance frequency curve (uncertainty of exceedance frequency curve), and (iii) Pareto pricing (defines the price of insurance contracts) and general Pareto distribution (GPD).
		Results/Remarks	• Output: (i) Maps of wind gusts for every grid cell in the tested area for WISC historic, synthetic, and probabilistic extensions, (ii) AAD is provided based on the insured damages, (iii) Exceedance frequency curves for building damages including uncertainty are provided, (iv) Normalized the insured total damages in comparison to the modeled total damages are provided, (v) Rapid damage estimation directly after a hazard event is offered, useful for damage assessment, and (vi) WISC historic event data and local exposure information enable a reliable derivation of the return period of a (rare) hazard event. • Uncertainties for damage estimation: (i) uncertainty associated with the assessment of the event, (ii) uncertainty regarding the damage model itself, (iii) Insurance claims might not report the exact time and date of damage (which introduces uncertainties), and (iv) In damage modelling estimations, movable property, damage to infrastructure, and business interruption are not included. • A probabilistic extension of the assessment of potential asset damages and risk of (extreme) hazard events are provided and associated with the evaluation of

Abbrev.	Categories and data		
			uncertainties.
	Resources	• CLIMADA GitHub repository and scripts for the recipe • ETH Data Archive • Winter windstorm model : (i) Probabilistic Windstorm Hazard Event Set for Europe, and (ii) The probabilistic hazard event set WISC probabilistic extension for each European country	
	Keywords	Hazard events; Windstorms; Probabilistic methods; Insurance claims.	
	Tag/Type	Climate Change and Hazard Data; Exposure and Vulnerability data.	
	Application in ICARIA	Potential use in ICARIA	Connection of public hazard datasets with exposure and vulnerability methodology to assess impacts and uncertainties, focused on windstorms but transferable to other hazards.

6 ICARIA's domain user survey

The main objective of the survey

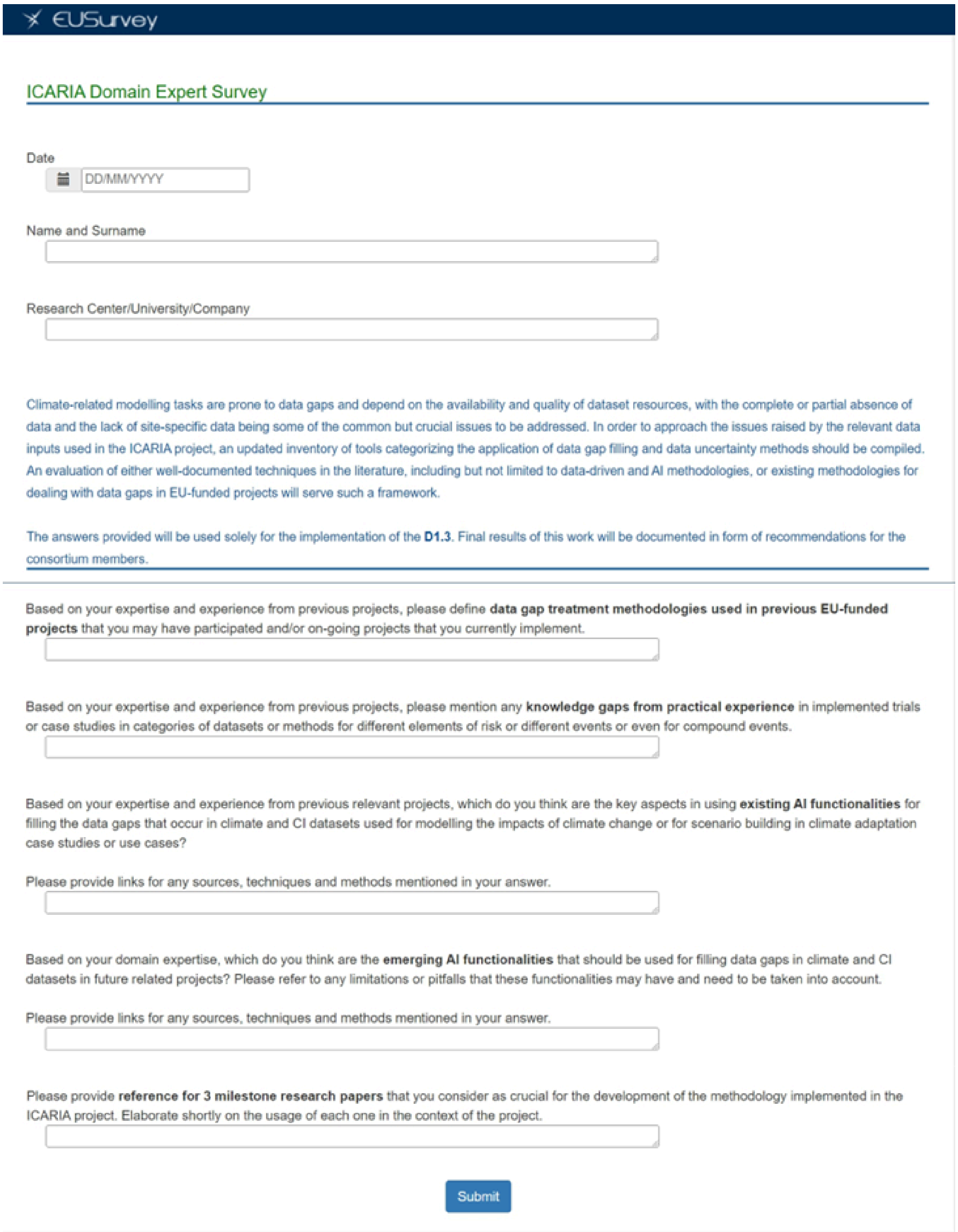
The development of ICARIA's holistic model aligns with current SOTA methodologies, focusing on identifying risk/impact assessment strategies from a multi-hazard perspective and considering all climate-hazard categories, (including heatwaves, forest fires, droughts, floods, storm surges, and storm winds), covering complex, compound, and cascading events. However, accessing the replicability and capabilities of the holistic model remains the ultimate target. To achieve this, case study areas, namely the Barcelona Metropolitan Area, South Aegean Region, and Salzburg Region were subjected to combined climate-hazard events. This practical exposure allowed experts to concurrently assess and identify modelling gaps and uncertainties during the data collection phase. Subsequently, experts could unravel the correlations between impact/risk assessment methodologies, case study areas, and modelling requirements. While this approach provides a robust foundation for creating and applying the holistic model, the complementary input from experts regarding data gaps associated with data-driven methodologies remains crucial. This input is sought to address potential expansions and modifications of the chosen methodologies. The domain user survey serves exactly this complementary role, gathering answers from a panel of ten (10) experts (internally or externally to the consortium). Their expertise and experience in local and/or EU-funded projects guide an assessment of the latest state-of-the-art methodologies present in data-gaps treatment methodologies and data-driven techniques. The survey is structured following a systematic approach, commencing with the treatment of data gaps identified in previous projects where experts actively participated. This information is key, helping in recognizing recurring patterns of data gaps that may be shared across ICARIA and related projects, guiding case facilitators towards an extensive understanding of methodologies appropriate for addressing data gaps and uncertainties. Subsequently, as a second step, experts are prompted to identify potential knowledge gaps based on their practical experience. This aspect proves valuable in the analysis of both single and compound or cascading scenarios from local authorities and case facilitators, identifying vulnerabilities within specific risk categories. Further, an additional key point of the survey is the evaluation of the functionality and applicability of existing and emerging AI methodologies, specializing in utilizing AI to treat data gaps and address uncertainties within climate and CI datasets when modelling climate adaptation studies, mirroring the use cases of ICARIA. Lastly, experts are asked to provide references to milestone papers that may or may not play a crucial role in the integration of data gap treatment methodologies with AI techniques, potentially extending to areas such as climate data. This inclusion ensures that ICARIA remains aligned with the latest state-of-the-art approaches.

A brief list of the experts and a description of the range of the topics are covered based on their background and expertise. While climate resilience remains the main object in the ICARIA project, diverse backgrounds of experts create a whole picture of the current practices for data-driven/AI-based methodologies (when applied to climate data or otherwise). The following list of eight (8) experts including their backgrounds and specialties can be found in Table 1 below.

Table 51: List of experts participated in ICARIA's domain user survey.

Name	Institution	Expertise	Inter-/External
Robert Monjo i Agut	FIC	Climate statistical downscaling and weather data	Internal
Siddharth Seshan	KWR	AI models and data fusion techniques	External
Gerasimos Antzoulatos	CERTH-ITI	Computer intelligence methods and applications	Internal
Damianos Mantsis	CERTH-ITI	Mathematics, Meteorology and Oceanography	Internal
Konstantinos Vlachos	CERTH-ITI	Geoscience & RS in data-driven methodologies	Internal
Ioannis Papoutsis	National Observatory Athens	EO and AI methodologies; Climate change & natural disasters monitoring	External
Ioannis Prapas	National Observatory Athens	DL for Earth systems, Big Data, and ML methodologies	External
Spyros Kondylatos	National Observatory Athens	DL for Earth systems; Bayesian DL, and Wildfire forecasting	External

Experts' background spans a diverse pool of subjects, encompassing methodologies including but not limited to EO, RS, statistical and dynamical downscaling techniques, along with methodologies for DL, AI applied to climate resilience, and a range of other domains. Although the initial emphasis was placed on updating internal expertise, the participation of external specialists is equally indispensable in achieving the purpose of compiling lists of state-of-the-art emerging data-driven methodologies, particularly when integrated with insights from user studies. For the survey purposes, the EU Survey portal was used to initiate, create, publish, and collect the results of the survey. This portal offers a unique, user-friendly UI that intuitively guides the users through the creation of a survey, providing a plethora of options in terms of the structure of the survey. The link to the survey is provided in this link: https://ec.europa.eu/eusurvey/runner/f96af3c2-67bc-11df-2c43-f5f6b97351cd. A figure of the survey as can be found in the EU survey's dedicated web interface can be find below:


EUSurvey

ICARIA Domain Expert Survey

Date

DD/MM/YYYY

Name and Surname

Research Center/University/Company

Climate-related modelling tasks are prone to data gaps and depend on the availability and quality of dataset resources, with the complete or partial absence of data and the lack of site-specific data being some of the common but crucial issues to be addressed. In order to approach the issues raised by the relevant data inputs used in the ICARIA project, an updated inventory of tools categorizing the application of data gap filling and data uncertainty methods should be compiled. An evaluation of either well-documented techniques in the literature, including but not limited to data-driven and AI methodologies, or existing methodologies for dealing with data gaps in EU-funded projects will serve such a framework.

The answers provided will be used solely for the implementation of the **D1.3**. Final results of this work will be documented in form of recommendations for the consortium members.

Based on your expertise and experience from previous projects, please define **data gap treatment methodologies used in previous EU-funded projects** that you may have participated and/or on-going projects that you currently implement.

Based on your expertise and experience from previous projects, please mention any **knowledge gaps from practical experience** in implemented trials or case studies in categories of datasets or methods for different elements of risk or different events or even for compound events.

Based on your expertise and experience from previous relevant projects, which do you think are the key aspects in using **existing AI functionalities** for filling the data gaps that occur in climate and CI datasets used for modelling the impacts of climate change or for scenario building in climate adaptation case studies or use cases?

Please provide links for any sources, techniques and methods mentioned in your answer.

Based on your domain expertise, which do you think are the **emerging AI functionalities** that should be used for filling data gaps in climate and CI datasets in future related projects? Please refer to any limitations or pitfalls that these functionalities may have and need to be taken into account.

Please provide links for any sources, techniques and methods mentioned in your answer.

Please provide **reference for 3 milestone research papers** that you consider as crucial for the development of the methodology implemented in the ICARIA project. Elaborate shortly on the usage of each one in the context of the project.

Submit

Figure 4: Overview of the ICARIA's domain survey questionnaire.

Overview of the summary

A summary of the results, and a brief description of the key methodologies proposed, as well as the output of the survey will be added below.

Table 52: ICARIA's domain user survey list of questions.

Nr.	Questions
1	Based on your expertise and experience from previous projects, please define data gap treatment methodologies used in previous EU-funded projects that you may have participated in and/or on-going projects that you currently implement.
2	Based on your expertise and experience from previous projects, please mention any knowledge gaps from practical experience in implementing trials or case studies in categories of datasets or methods for different elements of risk or different events or even for compound events.
3	Based on your expertise and experience from previous relevant projects, which do you think are the key aspects in using existing AI functionalities for filling the data gaps that occur in climate and CI datasets used for modelling the impacts of climate change or for scenario building in climate adaptation case studies or use cases? Please provide links for any sources, techniques and methods mentioned in your answer.
4	Based on your domain expertise, which do you think are the emerging AI functionalities that should be used for filling data gaps in climate and CI datasets in future related projects? Please refer to any limitations or pitfalls that these functionalities may have and need to be taken into account. Please provide links for any sources, techniques and methods mentioned in your answer.
5	Please provide reference(s) for three (3) milestone research papers that you consider as crucial for the development of the methodology implemented in the ICARIA project. Elaborate shortly on the usage of each one in the context of the project.

Nr.	Answers
1	Data gap treatment methodologies used in previous EU-funded projects: (1) interpolation, (2) gap filling with a specific value, (3) Kolmogorov-Smirnov-based inhomogeneity test, (4) masked modelling, (5) availability of accurate, high resolution (~ 2km) meteorological forecasts (used for training ML models, Copernicus ERA-5 reanalysis data), (6) availability of meteo forecasts of different quality standards (this distribution shift results in forecasting models to underperform), (7) lack of human-related assets (as opposed to environmental monitoring variables, that affect modelling climate change impacts; these assets are typically neglected, or proxies must be used e.g., population distribution and road network as a proxy for increased for example wildfire ignition risk), (8) simulations: atmospheric models to generate meteorological conditions in areas where observations at the desired frequency are missing. The generated data will have biases compared to the real data, and the quality will depend on the sophistication of the model simulation, (9) Multi-Stakeholder's forums, (10) Leverage crowdsourcing and social media platforms (data from heterogeneous sources), (11) – Semantically representation (the usage of Smart Data Models and standardize data formats, units and structures, enables the seamless fusion and harmonization of heterogeneous data ensuring compatibility, coherence and data sharing), (12) Specific domain ontologies (in the context of climate and critical infrastructure (CI) datasets used for modeling climate change impacts or scenario building in climate adaptation case studies), and (13) FAIR principles, data provenance.
2	Knowledge gaps from practical experience: (1) inhomogeneities, (2) outliers, (3) physical inconsistencies, (4) uncertainty of past events. Historical meteorological observations with adequate geographical coverage date back just a few decades in the past. For example, observations of sea surface temperature over the oceans or precipitation over the ocean or remote areas are absent beyond 1950. This means that records of historical extremes are not available, which can compromise the simulation of future extreme events, (5) Temporal resolution, (6) Spatial resolution, (7) Socioeconomic data, (8) Incomplete datasets, (9) Incomplete event catalogues of natural disasters, (10) Real-time data and monitoring, (11) Lack of annotated datasets, (12) Limiting effectiveness of predictive models, (13) High uncertainty in climate models, and (14) Risk mapping (creation of accurate hazard and risk maps).
3	Existing AI functionalities: (1) temporal and spatial auto-correlation, (2) climate projections & simulations, (3) domain adaptation models, (4) non-stationarity (due to climate change are key aspects that should be taken into account, especially for the evaluation), (5) use of spatiotemporal masked autoencoders for pre-training on the available data (and make them more resilient to data-gaps), (6) missing data imputation (using data-driven algorithms), (7) anomaly detection, (8) synthetic data generation, (9) heterogeneous data integration (techniques like data fusion and ensemble learning can help combine different datasets effectively), (10) interoperability (harmonizing data forms and formats), (11) digital twin (AI can generate multiple scenarios for climate adaptation by simulating various climate and socio-economic conditions in a Digital Twin environment), (12) spatial downscaling (AI techniques like convolutional neural networks (CNNs) can downscale global climate model outputs to higher resolutions needed for local impact assessments), (13) temporal downscaling (AI can refine temporal granularity of climate data, making it suitable for short-term event analysis and adaptation planning, (14) uncertainty analysis (AI can be used to quantify uncertainties in climate projections and impact assessments. Bayesian neural networks and ensemble learning methods can provide probabilistic estimates and confidence intervals, (15)

	sensitivity analysis (AI can perform sensitivity analysis, identifying the most influential variables and reducing uncertainty in model predictions), (16) trend Analysis (AI algorithms can identify long-term trends and patterns in climate and CI datasets, helping to detect gradual changes and emerging risks, (17) real-time monitoring (AI can process real-time data from sensors and IoT devices to provide immediate insights into current climate conditions and infrastructure status, and (18) dynamic adaptation (AI can be used to dynamically adjust models and scenarios based on real-time data, enhancing the responsiveness of climate adaptation strategies.
4	Emerging AI functionalities: (1) domain adaptation models, extrapolation (2) AI methods: GANs, Diffusion models, (3) self-supervised learning, (4) physics-aware models, (5) self-supervised pre-training, creation of ML-based emulators of physical purposes, (6) Reinforcement Learning (RL), (7) Spatial-Temporal Graph Neural Networks (ST-GNNs), Long Short-Term Memory (LSTM) Networks, (8) Synthetic Minority Over-sampling Technique (SMOTE).
5	Reference(s) for three (3) milestone research papers: Source for Kolmogorov-Smirnov goodness-of-fit test, (1) Example of a physics-aware model, (2) Example of DL application for Earth system science, (3) Example of causal relations from data, (4) Example of DL forecasts from sparse observations, (5) Artificial intelligence reconstructs missing climate information.

Table 53: ICARIA's domain user survey answers table.

Summary of the survey

The summary of the domain user survey for data gap treatment methodologies, knowledge gaps from practical experience, and AI functionalities can be organized in three lines: (1) Data gap treatment, (2) Existing AI functionalities, and (3) Emerging AI functionalities. The list of the experts who participated in the survey shared their extensive experience in participating in previous and current EU-funded projects and highlighted data inhomogeneity and inconsistency as the main issues to be addressed. For that reason, methodologies such as interpolation, value data gap filling, and inhomogeneity tests were proposed. Additionally, it was underlined that high-resolution meteorological forecasts and Copernicus ERA-5 reanalysis data are critical for training machine learning models, though varying quality standards can impact forecasting accuracy. This is only an example of the utilization of machine learning and AI methodologies for climate change. In terms of AI functionalities, current AI capabilities include temporal and spatial auto-correlation, climate projections and simulations, domain adaptation models, non-stationarity, and the use of spatiotemporal masked autoencoders as key candidate methodologies to improve resilience to data gaps. This is reflected in the ICARIA project and has the potential to be linked with the climate change and hazard data for cases where weather data are in scarcity (e.g., not fully covering the studied area or when weather observations are not reaching the minimum years necessary for providing accurate and robust output). Further, improving climate change methodologies using AI tools, should include a list of emerging methods including, domain adaptation models, extrapolation AI methods such as GANs, Diffusion models, self-supervised learning, physics-aware models, and self-supervised pre-training, creation of machine learning-based emulators of physical purposes. These methodologies allow for further investigation and application through ICARIA's lab tests and trials and mini-trials, allowing for inclusion of the suggestions from experts, utilizing concrete tools developed in previous EU-funded and other various related projects.

The diversity of the experts allows for creating a more complete idea of the applicability of AI tools, and how realistic such tools would be for ICARIA's purposes.

7 Conclusions

The issue of data gaps and uncertainties treatment in quantitative hazard/impact assessment is crucial to guarantee reliability of results. A variety of methods and tools can be found in literature, with robust approaches that can increasingly rely on the ever-growing availability of high-resolution information from remote sensing, in-situ monitoring networks and citizen-science tools, as well as of computing capacity enhanced by machine learning and AI. Nevertheless, the identification and communication of data gaps and uncertainties related to specific hazard/impact modelling results is important beyond the "*data gap filling problem*", to acknowledge uncertainties and limitations in risk assessments and simulations derived by the implementation of the ICARIA Holistic modelling framework in any s case study area.

This is even more crucial in the context of assessments encompassing the impacts on multiple assets determined by complex multi-hazard events (compound coincident, compound consecutive, cascading effects, see D1.1) which, compared to single-hazard assessments, highlight even more the complexity for decision makers and planners to make choices and take science-informed decisions aimed at increasing resilience. These uncertainties are related not only to the missing data concerning specific Hazard-Exposure-Vulnerability (H-E-V) variables used as input in a given impact assessment model, but intrinsic to the dynamics of compound events and cascading effects (Zuccaro et al., 2018), in which uncertainties such as the probability of transition among hazards, and the probability of triggering cascading effects following a given threshold of damage on a critical service asset or network component contribute to propagate errors in the quantitative assessment of the final scenario.

Therefore, considering how uncertainties related to climate change itself depend on a variety of aleatory factors and tipping points (Lenton et al., 2019) it is of extreme importance that the results of ICARIA probabilistic impact assessment models, while improving their reliability through data gap filling, data refinement (and associated uncertainties) with respect to the space-time variables and to H-E-V parameters involved in the areas object of the analysis, always acknowledge all data sources used as input, existing data gaps or low-resolution data used e.g., as a proxy of a missing variable This will be achieved by mapping in the Trials and Mini-Trials modelling framework the key variables and datasets (see D1.1 section 3.2) used, the data sources and the application (already implemented in ICARIA Lab Tasks or potential within WP4 Trial implementation) of the ICARIA cookbook. This will allow decision-makers to develop climate adaptation and resilience plans adequately informed by scientific evidence and existing limitations in knowledge, pursuing the achievement of encountering hazard events with robust and informative models and frameworks.

.

References

Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U. R.; Makarenkov, V.; Nahavandi, S. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. Information Fusion 2021, 76, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

Anand, H.; Darayi, M. Power Network Component Vulnerability Analysis: A Machine Learning Approach. Procedia Computer Science 2021, 185, 73–80. https://doi.org/10.1016/j.procs.2021.05.008.

Andrychowicz, M.; Espeholt, L.; Li, D.; Merchant, S.; Merose, A.; Zyda, F.; Agrawal, S.; Kalchbrenner, N. Deep Learning for Day Forecasts from Sparse Observations. arXiv 2023. https://doi.org/10.48550/ARXIV.2306.06079.

Angeles, K.; Kijewski-Correa, T. Advancing Building Data Models for the Automation of High-Fidelity Regional Loss Estimations Using Open Data. Automation in Construction 2022, 140, 104382. https://doi.org/10.1016/j.autcon.2022.104382.

Apaydin, H.; Sonmez, F.; Yildirim, Y. Spatial Interpolation Techniques for Climate Data in the GAP Region in Turkey. Clim. Res. 2004, 28, 31–40. https://doi.org/10.3354/cr028031.

ARSINOE Project. ARSINOE Project: Climate-Resilient Regions through Systemic Solutions and Innovations. 2023, https://doi.org/10.3030/101037424.

Ascough, J. C.; Maier, H. R.; Ravalico, J. K.; Strudley, M. W. Future Research Challenges for Incorporation of Uncertainty in Environmental and Ecological Decision-Making. Ecological Modelling 2008, 219 (3–4), 383–399. https://doi.org/10.1016/j.ecolmodel.2008.07.015.

Aspinall, W. P.; Cooke, R. M. Quantifying Scientific Uncertainty from Expert Judgement Elicitation. In Risk and Uncertainty Assessment for Natural Hazards; Rougier, J., Sparks, S., Hill, L. J., Eds.; Cambridge University Press, 2013; pp 64–99. https://doi.org/10.1017/CBO9781139047562.005.

Aznar-Siguan, G.; Bresch, D. N. CLIMADA v1: A Global Weather and Climate Risk Assessment Platform. Geosci. Model Dev. 2019, 12 (7), 3085–3097. https://doi.org/10.5194/gmd-12-3085-2019.

Bikše, J.; Retike, I.; Haaf, E.; Kalvāns, A. Assessing Automated Gap Imputation of Regional Scale Groundwater Level Data Sets with Typical Gap Patterns. Journal of Hydrology 2023, 620, 129424. https://doi.org/10.1016/j.jhydrol.2023.129424.

Boé, J.; Mass, A.; Deman, J. A Simple Hybrid Statistical–Dynamical Downscaling Method for Emulating Regional Climate Models over Western Europe. Evaluation, Application, and Role of Added Value? Clim Dyn 2023, 61 (1–2), 271–294. https://doi.org/10.1007/s00382-022-06552-2.

Camps-Valls, G.; Gerhardus, A.; Ninad, U.; Varando, G.; Martius, G.; Balaguer-Ballester, E.; Vinuesa, R.; Diaz, E.; Zanna, L.; Runge, J. Discovering Causal Relations and Equations from Data. arXiv 2023. https://doi.org/10.48550/ARXIV.2305.13341.

Cartagena-Colón, M.; Mattei, H.; Wang, C. Dasymetric Mapping of Population Using Land Cover Data in JBNERR, Puerto Rico during 1990–2010. Land 2022, 11 (12), 2301. https://doi.org/10.3390/land11122301.

Colson, A. R.; Cooke, R. M. Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. Review of Environmental Economics and Policy 2018, 12 (1), 113–132. https://doi.org/10.1093/reep/rex022.

Cooke, R. M. Experts in Uncertainty: Opinion and Subjective Probability in Science; Environmental ethics and science policy series; Oxford University Press: New York, 1991.

Cooke, R. M.; Goossens, L. H. J. Expert Judgement Elicitation for Risk Assessments of Critical Infrastructures. Journal of Risk Research 2004, 7 (6), 643–656. https://doi.org/10.1080/1366987042000192237.

De' Michieli Vitturi, M.; Bevilacqua, A.; Tadini, A.; Neri, A. ELICIPY 1.0: A Python Online Tool for Expert Elicitation. SoftwareX 2024, 25, 101641. https://doi.org/10.1016/j.softx.2024.101641.

Demuzere, M.; Bechtel, B.; Middel, A.; Mills, G. Mapping Europe into Local Climate Zones. PLoS ONE 2019, 14 (4), e0214474, https://doi.org/10.1371/journal.pone.0214474.

Doury, A.; Somot, S.; Gadat, S.; Ribes, A.; Corre, L. Regional Climate Model Emulator Based on Deep Learning: Concept and First Evaluation of a Novel Hybrid Downscaling Approach. Clim Dyn 2023, 60 (5–6), 1751–1779. https://doi.org/10.1007/s00382-022-06343-9.

Doyle, E. E. H.; Thompson, J.; Hill, S.; Williams, M.; Paton, D.; Harrison, S.; Bostrom, A.; Becker, J. Where Does Scientific Uncertainty Come from, and from Whom? Mapping Perspectives of Natural Hazards Science Advice. International Journal of Disaster Risk Reduction 2023, 96, 103948. https://doi.org/10.1016/j.ijdrr.2023.103948.

Du, H.; Lin, X.; Jiang, J.; Lu, Y.; Du, H.; Zhang, F.; Yu, F.; Feng, T.; Wu, X.; Peng, G.; Deng, S.; He, S.; Bai, X. A Single-Building Damage Detection Model Based on Multi-Feature Fusion: A Case Study in Yangbi. iScience 2024, 27 (1), 108586. https://doi.org/10.1016/j.isci.2023.108586.

Eberenz, S.; Stocker, D.; Röösli, T.; Bresch, D. N. Asset Exposure Data for Global Physical Risk Assessment. Earth Syst. Sci. Data 2020, 12 (2), 817–833. https://doi.org/10.5194/essd-12-817-2020.

Esposito, C.; Mastrantoni, G.; Marmoni, G. M.; Antonielli, B.; Caprari, P.; Pica, A.; Schilirò, L.; Mazzanti, P.; Bozzano, F. From Theory to Practice: Optimisation of Available Information for Landslide Hazard Assessment in Rome Relying on Official, Fragmented Data Sources. Landslides 2023. https://doi.org/10.1007/s10346-023-02095-7.

Faroux, S.; Kaptué Tchuenté, A. T.; Roujean, J.-L.; Masson, V.; Martin, E.; Le Moigne, P. ECOCLIMAP-II/Europe: A Twofold Database of Ecosystems and Surface Parameters at 1 Km Resolution Based on Satellite Information for Use in Land Surface, Meteorological and Climate Models. Geosci. Model Dev. 2013, 6 (2), 563–582. https://doi.org/10.5194/gmd-6-563-2013.

Glickman, M. E.; Dyk, D. A. Basic Bayesian Methods. In Topics in Biostatistics; Ambrosius, W. T., Ed.; Walker, J. M., Series Ed.; Methods in Molecular BiologyTM; Humana Press: Totowa, NJ, 2007; Vol. 404, pp 319–338. https://doi.org/10.1007/978-1-59745-530-5_16.

Gosling, J. P. SHELF: The Sheffield Elicitation Framework. In Elicitation; Dias, L. C., Morton, A., Quigley, J., Eds.; International Series in Operations Research & Management Science; Springer International Publishing: Cham, 2018; Vol. 261, pp 61–93. https://doi.org/10.1007/978-3-319-65052-4_4.

Grainger, S.; Dessai, S.; Daron, J.; Taylor, A.; Siu, Y. L. Using Expert Elicitation to Strengthen Future Regional Climate Information for Climate Services. Climate Services 2022, 26, 100278. https://doi.org/10.1016/j.cliser.2021.100278.

Gudmundsson, L.; Bremnes, J. B.; Haugen, J. E.; Engen-Skaugen, T. Technical Note: Downscaling RCM Precipitation to the Station Scale Using Statistical Transformations – a Comparison of Methods. Hydrol. Earth Syst. Sci. 2012, 16 (9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.

Han, E.; Ines, A. V. M. Downscaling Probabilistic Seasonal Climate Forecasts for Decision Support in Agriculture: A Comparison of Parametric and Non-Parametric Approach. Climate Risk Management 2017, 18, 51–65. https://doi.org/10.1016/j.crm.2017.09.003.

Harder, P.; Hernandez-Garcia, A.; Ramesh, V.; Yang, Q.; Sattigeri, P.; Szwarcman, D.; Watson, C.; Rolnick, D. Hard-Constrained Deep Learning for Climate Downscaling. arXiv 2022. https://doi.org/10.48550/ARXIV.2208.05424.

Howe, D. A.; Champagne, C.; Schlossberger, N. A 'Total' Imputation Algorithm That Fills Gaps in Time Series Measurements for ADEV and Phase Noise Characterizations of Power-Law Noise Models. In 2022 Joint Conference of the European Frequency and Time Forum and IEEE International Frequency Control Symposium (EFTF/IFCS); IEEE: Paris, France, 2022; pp 1–2. https://doi.org/10.1109/EFTF/IFCS54560.2022.9850921.

ICARIA (2023a). Deliverable 1.1 – CARIA holistic modelling framework. Available at: https://www.icaria-project.eu/downloads/#deliverables.

ICARIA (2024a). Deliverable 1.2 – Climate projections and hazard scenarios. Available at: https://www.icaria-project.eu/downloads/#deliverables.

ICARIA (2024b). Deliverable 1.3 – Impact modelling data requirements and methods to treat data gap filling and data uncertainty. Available at: https://www.icaria-project.eu/downloads/#deliverables.

ICARIA (2024c). Deliverable 1.4 – WP1 Testing results. Available at: https://www.icaria-project.eu/downloads/#deliverables.

Kim, K.-M.; Kwak, J. W. PVS-GEN: Systematic Approach for Universal Synthetic Data Generation Involving Parameterization, Verification, and Segmentation. Sensors 2024, 24 (1), 266. https://doi.org/10.3390/s24010266.

Lenton, T. M.; Rockström, J.; Gaffney, O.; Rahmstorf, S.; Richardson, K.; Steffen, W.; Schellnhuber, H. J. Climate Tipping Points — Too Risky to Bet Against. Nature 2019, 575 (7784), 592–595. https://doi.org/10.1038/d41586-019-03595-0.

Loerke, E.; Pohle, I.; Wilkinson, M. E.; Rivington, M.; Wardell-Johnson, D.; Geris, J. Long-Term Daily Stream Temperature Record for Scotland Reveals Spatio-Temporal Patterns in Warming of Rivers in the Past and Further Warming in the Future. Science of The Total Environment 2023, 890, 164194. https://doi.org/10.1016/j.scitotenv.2023.164194.

Lorenz, R.; Becker, N.; Gardiner, B.; Ulbrich, U.; Hanewinkel, M.; Benjamin, S. Storm Damage beyond Wind Speed – Impacts of Wind Characteristics and Other Meteorological Factors on Tree Fall along Railway Lines. February 23, 2024. https://doi.org/10.5194/egusphere-2024-120.

Mader, H. M., Statistics in Volcanology; International Association of Volcanology and Chemistry of the Earth's Interior, Geological Society of London, Eds.; Special publications of the International Association of Volcanology and Chemistry of the Earth's Interior; Geological Society: London, 2006.

Manzanas, R.; Gutiérrez, J. M.; Fernández, J.; Van Meijgaard, E.; Calmanti, S.; Magariño, M. E.; Cofiño, A. S.; Herrera, S. Dynamical and Statistical Downscaling of Seasonal Temperature Forecasts in Europe: Added Value for User Applications. Climate Services 2018, 9, 44–56. https://doi.org/10.1016/j.cliser.2017.06.004.

Monjo, R.; Pórtoles, J.; Ribalaygia, J. Detection of Inhomogeneities in Daily Data: A Test Based on the Kolmogorov-Smirnov Goodness-of-Fit Test. 2013. https://doi.org/10.5281/ZENODO.5089753.

Montalvo, C.; Reyes-Silva, J. D.; Sañudo, E.; Cea, L.; Puertas, J. Urban Pluvial Flood Modelling in the Absence of Sewer Drainage Network Data: A Physics-Based Approach. Journal of Hydrology 2024, 634, 131043. https://doi.org/10.1016/j.jhydrol.2024.131043.

Mühlhofer, E.; Kropf, C. M.; Riedel, L.; Bresch, D. N.; Koks, E. E. OpenStreetMap for Multi-Faceted Climate Risk Assessments. Environ. Res. Commun. 2024, 6 (1), 015005. https://doi.org/10.1088/2515-7620/ad15ab.

Nikulin, G.; Asharaf, S.; Magariño, M. E.; Calmanti, S.; Cardoso, R. M.; Bhend, J.; Fernández, J.; Frías, M. D.; Fröhlich, K.; Früh, B.; García, S. H.; Manzanas, R.; Gutiérrez, J. M.; Hansson, U.; Kolax, M.; Liniger, M. A.; Soares, P. M. M.; Spirig, C.; Tome, R.; Wyser, K. Dynamical and Statistical Downscaling of a Global Seasonal Hindcast in Eastern Africa. Climate Services 2018, 9, 72–85. https://doi.org/10.1016/j.cliser.2017.11.003.

Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D. Comparison of Stochastic and Machine Learning Methods for Multi-Step Ahead Forecasting of Hydrological Processes. Stoch Environ Res Risk Assess 2019, 33 (2), 481–514. https://doi.org/10.1007/s00477-018-1638-6.

Paprotny, D.; Kreibich, H.; Morales-Nápoles, O.; Terefenko, P.; Schröter, K. Estimating Exposure of Residential Assets to Natural Hazards in Europe Using Open Data. Nat. Hazards Earth Syst. Sci. 2020, 20 (1), 323–343. https://doi.org/10.5194/nhess-20-323-2020.

Paulik, R.; Zorn, C.; Wotherspoon, L.; Sturman, J. Modelling National Residential Building Exposure to Flooding Hazards. International Journal of Disaster Risk Reduction 2023, 94, 103826. https://doi.org/10.1016/j.ijdrr.2023.103826.

Peixoto, J. P. J.; Costa, D. G.; Franca Rocha, W. D. J. S. D.; Portugal, P.; Vasques, F. On the Positioning of Emergencies Detection Units Based on Geospatial Data of Urban Response Centres. Sustainable Cities and Society 2023, 97, 104713. https://doi.org/10.1016/j.scs.2023.104713.

Prathom, C.; Champrasert, P. An Exploration of Interpolation - Machine Learning Model for Climate Model Downscaling Under the Limitation of Data Quantity. In 2023 International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC); IEEE: Jeju, Korea, Republic of, 2023; pp 1–6. https://doi.org/10.1109/ITC-CSCC58803.2023.10212662.

Rampal, N.; Gibson, P. B.; Sood, A.; Stuart, S.; Fauchereau, N. C.; Brandolino, C.; Noll, B.; Meyers, T. High-Resolution Downscaling with Interpretable Deep Learning: Rainfall Extremes over New Zealand. Weather and Climate Extremes 2022, 38, 100525. https://doi.org/10.1016/j.wace.2022.100525.

Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep Learning and Process Understanding for Data-Driven Earth System Science. Nature 2019, 566 (7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Ribalaygua, J.; Torres, L.; Pórtoles, J.; Monjo, R.; Gaitán, E.; Pino, M. R. Description and Validation of a Two-Step Analogue/Regression Downscaling Method. Theor Appl Climatol 2013, 114 (1–2), 253–269. https://doi.org/10.1007/s00704-013-0836-x.

Rosina, K.; Batista E Silva, F.; Vizcaino, P.; Marín Herrera, M.; Freire, S.; Schiavina, M. Increasing the Detail of European Land Use/Cover Data by Combining Heterogeneous Data Sets. International Journal of Digital Earth 2020, 13 (5), 602–626. https://doi.org/10.1080/17538947.2018.1550119.

Serrano-Notivoli, R.; De Luis, M.; Beguería, S. An R Package for Daily Precipitation Climate Series Reconstruction. Environmental Modelling & Software 2017, 89, 190–195. https://doi.org/10.1016/j.envsoft.2016.11.005.

Shmuel, A.; Heifetz, E. Developing Novel Machine-Learning-Based Fire Weather Indices. Mach. Learn.: Sci. Technol. 2023, 4 (1), 015029. https://doi.org/10.1088/2632-2153/acc008.

Singh, K.; Jeong, C.; Park, S.; Babur, A. N.; Zeller, E.; Cha, M. Self-Supervised Learning for Climate Downscaling. In 2023 IEEE International Conference on Big Data and Smart Computing (BigComp); IEEE: Jeju, Korea, Republic of, 2023; pp 13–17. https://doi.org/10.1109/BigComp57234.2023.00012.

Spezia, L.; Gibbs, S.; Glendell, M.; Helliwell, R.; Paroli, R.; Pohle, I. Bayesian Analysis of High-Frequency Water Temperature Time Series through Markov Switching Autoregressive Models. Environmental Modelling & Software 2023, 167, 105751. https://doi.org/10.1016/j.envsoft.2023.105751.

Stankiewicz, J.; König, A.; Pickar, K.; Weiss, S. How Certain Is Good Enough? Managing Data Quality and Uncertainty in Ordinal Citizen Science Data Sets for Evidence-Based Policies on Fresh Water. Citizen Science: Theory and Practice 2023, 8 (1), 39. https://doi.org/10.5334/cstp.592.

Street, R. B.; Nilsson, C. Introduction to the Use of Uncertainties to Inform Adaptation Decisions. In Adapting to an Uncertain Climate; Capela Lourenço, T., Rovisco, A., Groot, A., Nilsson, C., Füssel, H.-M., Van Bree, L., Street, R. B., Eds.; Springer International Publishing: Cham, 2014; pp 1–16. https://doi.org/10.1007/978-3-319-04876-5_1.

Tzoumpas, K.; Estrada, A.; Miraglio, P.; Zambelli, P. A Data Filling Methodology for Time Series Based on CNN and (Bi)LSTM Neural Networks. arXiv 2022. https://doi.org/10.48550/ARXIV.2204.09994.

Van Der Meer, M.; De Roda Husman, S.; Lhermitte, S. Deep Learning Regional Climate Model Emulators: A Comparison of Two Downscaling Training Frameworks. J Adv Model Earth Syst 2023, 15 (6), e2022MS003593. https://doi.org/10.1029/2022MS003593.

Van Der Sluijs, J. P.; Craye, M.; Funtowicz, S.; Kloprogge, P.; Ravetz, J.; Risbey, J. Combining Quantitative and Qualitative Measures of Uncertainty in Model–Based Environmental Assessment: The NUSAP System. Risk Analysis 2005, 25 (2), 481–492. https://doi.org/10.1111/j.1539-6924.2005.00604.x.

Velasco, M.; Russo, B.; Martínez, M.; Malgrat, P.; Monjo, R.; Djordjevic, S.; Fontanals, I.; Vela, S.; Cardoso, M.A.; Buskute, A., Water 2018 10: 1356-1366. https://doi.org/10.3390/w10101356.

Walker, W. E.; Harremoës, P.; Rotmans, J.; Van Der Sluijs, J. P.; Van Asselt, M. B. A.; Janssen, P.; Krayer Von Krauss, M. P. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 2003, 4 (1), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.

Wang, G.; Garcia, D.; Liu, Y.; De Jeu, R.; Johannes Dolman, A. A Three-Dimensional Gap Filling Method for Large Geophysical Datasets: Application to Global Satellite Soil Moisture Observations. Environmental Modelling & Software 2012, 30, 139–142. https://doi.org/10.1016/j.envsoft.2011.10.015.

Welker, C.; Röösli, T.; Bresch, D. N. Comparing an Insurer's Perspective on Building Damages with Modelled Damages from Pan-European Winter Windstorm Event Sets: A Case Study from Zurich, Switzerland. Nat. Hazards Earth Syst. Sci. 2021, 21 (1), 279–299. https://doi.org/10.5194/nhess-21-279-2021.

Wilcke, R.A.I. and Bärring, L., 2016: Selecting regional climate scenarios for impact modelling studies. Environmental Modelling & Software 78, 191-201 http://dx.doi.org/10.1016/j.envsoft.2016.01.002.

Yang, Y.; Javanroodi, K.; Nik, V. M. Climate Change and Energy Performance of European Residential Building Stocks – A Comprehensive Impact Assessment Using Climate Big Data from the Coordinated Regional Climate Downscaling Experiment. Applied Energy 2021, 298, 117246. https://doi.org/10.1016/j.apenergy.2021.117246.

Zuccaro, G.; De Gregorio, D.; Leone, M. F. Theoretical Model for Cascading Effects Analyses. International Journal of Disaster Risk Reduction 2018, 30, 199–215. https://doi.org/10.1016/j.ijdrr.2018.04.019.

Annex A: Main sources of open data repositories for local hazard downscaling and exposure/vulnerabilities classification and analyses

Table A 1: *List of open data repositories.*

Hazard /Feature	Relevant Parameters / Indicators	Dataset	Source
Urban Climate	Local Climate Zones	World Urban Database	https://www.wudapt.org/lcz-maps/
	Land Surface Temperature	GoogleEarthEngine	https://earthengine.google.com/
Extreme Events	European Severe Weather Database	ESWD	https://eswd.eu/
Hazard, Exposure, Vulnerability, Losses	Risk Data Library	Global Facility for Disaster Reduction and Recovery (GFDRR) World Bank	https://riskdatalibrary.org/, https://www.gfdrr.org/en
Land characterization	Digital Surface Model	EU-DEM	https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model
	Land cover classification Coastal zones classification Land use Imperviousness	Copernicus Land Monitoring Service	https://land.copernicus.eu/
	High resolution(100m) maps of land cover/use, population, GDP, and fixed assets for 42 countries from 1870 to 2020	HANZE v2.0 exposure model	https://zenodo.org/records/6826536
	A dual database at 1 km resolution that includes an ecosystem classification and a coherent set of land surface parameters (Faroux et. al., 2015)	ECOCLIMAM	https://opensource.umr-cnrm.fr/projects/ecoclimap/wiki
Population	Population distribution (regional level) Domestic product (regional) Employment (regional) Labour (regional) Households (regional)	ARDECO	https://urban.jrc.ec.europa.eu/ardeco/explorer?lng=en

	Population distribution (local level)	European Settlement map	https://human-settlement.emergency.copernicus.eu/datasets.php
	Excess mortality by month	Eurostat	https://ec.europa.eu/eurostat/web/covid-19/population-health#Health
	Population projections at regional level	Eurostat	https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_projections_at_regional_level
	Population affected by hazard type Deaths by hazard type Injured by hazard type Homeless by hazard type	EM-DAT	https://public.emdat.be/data
	Population projections	CIESIN Columbia University Eurostat	https://sedac.ciesin.columbia.edu/data/collection/popdynamics/maps/services https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_projections_at_regional_level
	Spatiotemporal population and activity map	Joint Research Centre Data Catalogue	https://data.jrc.ec.europa.eu/collection/id-00155
Buildings	Residential and non-residential areas Built-up surface Building height Built-up volume	European Settlement Map	https://human-settlement.emergency.copernicus.eu/datasets.php
	Built-up impervious areas	Copernicus Land Monitoring Services	https://land.copernicus.eu/en
	Building geometries	Open Street Map	https://www.openstreetmap.org/
	Energy demand	Enerdata	https://zebra-monitoring.enerdata.net/overall-building-activities/
Transport networks	Transport network graphs	OpenStreetMap	https://www.openstreetmap.org/
	Road surfaces	Urban Atlas	https://land.copernicus.eu/
Vegetation	Burnt Area	Copernicus Land	https://land.copernicus.

	Vegetation Properties Vegetation Indices Trajectories Vegetation Phenology and Productivity Parameters	Monitoring Services	eu/en/products/vegetation
	Dominant leaf type Forest type Tree cover density Grassland Water and wetness Small woody features	Copernicus Land Monitoring Services	https://land.copernicus.eu/en
	Vegetation plots and types	European Vegetation Archive	https://euroveg.org/eva-database/obtaining-data
Energy	Energy production by source Energy intensity by fuel and use type	IEA Database Enerdata	https://www.iea.org/data-and-statistics https://www.enerdata.net/expertise/data-science.html
Greenhouse gas emissions	GHG emission by sector	Eurostat-EEA IEA Database Our World in Data Enerdata	https://ec.europa.eu/eurostat/web/environment/database https://www.iea.org/data-and-statistics https://ourworldindata.org/co2-emissions https://www.enerdata.net/expertise/data-science.html
Trees	High Resolution Tree Cover Density product at pan-European level	European Environmental Agency DataHub	https://www.eea.europa.eu/en/datahub/datahubitem-view/b8a5a51f-0c8e-44bc-bac2-2afa6b9444da
	International Tree-Ring Data Bank (ITRDB) (Paleoclimatology)	National Oceanic and Atmospheric Administration (NOAA) product	https://www.ncei.noaa.gov/products/paleoclimatology/tree-ring
	Awesome-forests – Curated list of forest datasets	GitHub repository	https://github.com/blutjens/awesome-forests
	European Forest Institute Datasets and Maps	European Forest Institute	https://efi.int/
Windstorm	Extreme Wind Storms (XWS) Catalogue	Data repository	https://www.europeanwindstorms.org/
	The Windstorm Information	Windstorm Information	https://climate.copernic

	Service (WISC) provide information about European Windstorms	Service	us.eu/windstorm-information-service
Fire	Fire Weather Index – Dataset description	Copernicus Climate Change Service	https://datastore.copernicus-climate.eu/documents/sis-european-tourism/C3S_D422_Lot2_TEC_FWI_dataset_description_v2.pdf
	Fire danger indicators for Europe from 1970 to 2098 derived from climate projections	Copernicus Climate Change Service	https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-tourism-fire-danger-indicators?tab=overview

Annex B: Templates to map data gap filling and uncertainty treatment

Table B 1: Key input/output of the hazard assessment models - CS NAME.

DATASETS	Climate Hazards							
	Hazard n. 1		Hazard n. 2		Hazard n. 3		Uncertainty treatment	Cookbook Reference
	Data	Source	Data	Source	Data	Source		
INPUT DATASETS								
Historic climatic variables								
Future climate projections								
Land use and terrain information								
			-	-				
OUTPUT DATASETS								
Hazard								

Table B 2: Key input/output of the vulnerability assessment models for the Trials - CS NAME.

DATASETS	Climate Hazards							
	Hazard n. 1		Hazard n. 2		Hazard n. 3		Uncertainty treatment	Cookbook Reference
	Data	Source	Data	Source	Data	Source		
INPUT DATASETS								
Exposure PEOPLE								

Exposure PROPERTIES								
Exposure NATURAL AREAS								
Exposure TRANSPORT								
Exposure WATER SECTOR								
Exposure ELECTRICITY SECTOR								
Exposure WASTE SECTOR								
OUTPUT DATASETS								
Vulnerability classes/functions /data								

Table B 3: Key input/output of the risk/impact assessment models for the Trial - CS NAME.

DATASETS	Climate Hazards							
	Hazard n. 1		Hazard n. 2		Hazard n. 3		Uncertainty treatment	Cookbook Reference
	Data	Source	Data	Source	Data	Source		
INPUT DATASETS								
Exposure / Vulnerability PEOPLE								
Exposure / Vulnerability PROPERTIES								
Exposure / Vulnerability NATURAL AREAS								
Exposure / Vulnerability TRANSPORT								

Exposure / Vulnerability WATER SECTOR								
Exposure / Vulnerability ELECTRICITY SECTOR								
Exposure / Vulnerability WASTE SECTOR								
OUTPUT DATASETS								
Risk/Impact								

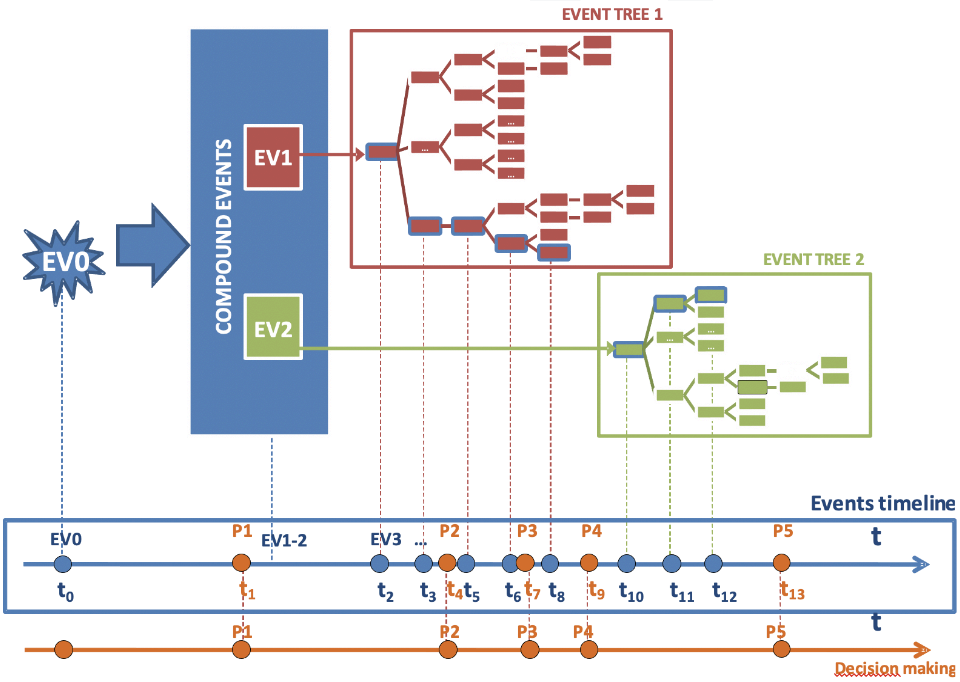


Figure 5: Event tree scenario building tool, adapted from SNOWBALL (Zuccaro et. al., 2018).

Annex C: EU Projects

1. BINGO - Bringing INnovation to onGOing water management – A better future under climate change CORDIS.
2. BRIGAID - BRIdging the GAp for Innovations in Disaster Resilience, CORDIS.
3. CLARITY - Integrated Climate Adaptation Service Tools for Improving Resilience Measure Efficiency, CORDIS, GitHub, Zenodo.
4. ClimateAdapt - Testing the limits and potential of evolution in response to climate change, CORDIS.
5. ClimateFarmDemo - a European-wide network of pilot farmers implementing and demonstrating climate-smart solutions for a carbon-neutral Europe, CORDIS.
6. ClimEMPOWER – Climate resilience in regional development, CORDIS.
7. CRISIS-ADAPT II - Climate Risk Information for SupportIng ADAptation Planning and operaTion.
8. CRISTAL Project - Climate resilient and environmentally sustainable transport infrastructure, with a focus on inland waterways, CORDIS.
9. DRIVER+ - DRiving InnoVation in crisis management for European Resilience: CORDIS.
10. ESPRESSO - Enhancing Synergies for disaster PRevention in the EurOpean Union, CORDIS.
11. EU-CIRCLE - A panEuropean framework for strengthening Critical Infrastructure resilience to climate change, CORDIS.
12. FireLogue - Cross-section dialogue for Wildfire Risk Management, CORDIS.
13. KNOWING: Framework for defining Climate Mitigation Pathways based on Understanding and Integrated Assessment of Climate Impacts, Adaptation Strategies and Social Transformation, CORDIS.
14. MAGICA – Maximizing the synergy of European research Governance and Innovation for Climate Action, CORDIS.
15. MAIA - Mapping and Assessment for Integrated Ecosystem Accounting, CORDIS.
16. NetworkNature - Advancing nature-based solutions together, CORDIS.
17. PLACARD: PLAtform for Climate Adaptation and Risk reDuction, CORDIS.
18. RAIN: RAIN will quantify the complex interactions between weather events and land-based infrastructure systems, CORDIS.
19. RECONECT - Regenarating ECOsystems with Nature-based solutions for hydro-meteorological risk rEduCTion, CORDIS.
20. RESCCUE - Resilient cities facing climate change, CORDIS.
21. Snowball - Modelling framework and tools supporting impact assessment from cascading effects, CORDIS.
22. SOCLIMPACT - DownScaling CLImate ImPACTs and decarbonization pathways in EU islands, and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and Beyond, CORDIS.
23. weADAPT – a dynamic, collaborative space for knowledge exchange on climate change adaptation issues.

Annex D: Data Management Statement

Table D 1: Data used in preparation of ICARIA Deliverable 1.3.

Dataset name	Format	Size	Owner and re-use conditions	Potential utility within and outside ICARIA	Unique ID
N/A	N/A	N/A	N/A	N/A	N/A

Table D 2:. Data produced in preparation of ICARIA Deliverable 1.3.

Dataset name	Format	Size	Owner and re-use conditions	Potential utility within and outside ICARIA	Unique ID
N/A	N/A	N/A	N/A	N/A	N/A

More info: **www.icaria-project.eu**

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101093806. The publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.