AI4Drought

Response to:

ESA/AO/1-10797/21/I-EF

AI for SCIENCE - Multi-Hazards, Compounds and Cascade events

–Proposal–

lobelia. Barcelona Supercomputing Center Centro Nacional de Supercomputación eurecat Centre Tecnològic de Catalunya	Proposal: AI4DROUGHT Ref. LOBELIA_AI4DROUGHT_PRO_31 issue: 1.0 - Date: 14/6/2021

Table of Contents

Table of Figures

List of Tables

1 TECHNICAL PROPOSAL

1.1 PROPOSED DEVELOPMENT

Droughts are among the most detrimental natural disasters due to their broad environmental, social and economical impacts (Gerber et al. 2017). It is estimated that around 20% of the total economic costs caused by global natural hazards can be identified with droughts (Wilhite 2000) and their complex interplay and cascade of possible effects (e.g. reduced precipitation, drier forests, increased risk of fires). The multidimensional nature of droughts leads to very heterogeneous representations in different disciplines (i.e. lack of rainfall in meteorology, dry soil and crop failure in agriculture or reduced river discharges and water reservoirs in hydrology) that in the end avoids to have a comprehensive characterization of these events. Moreover, under a global warming scenario, where mean annual rainfall rates are expected to decrease but extreme precipitation to increase in many regions of the world including most of Europe, the future impacts of droughts through their multifaceted relationships are only poorly understood and mostly uncertain (AR5-IPCC). Considering all these elements together, it emerges a clear and urgent need to deepen our understanding on the occurrence and the cascading effects of droughts. In this proposal, we consider seasonal climate predictions in the Iberian peninsula (Spain and Portugal), which is projected to move towards a drier climate (Vicente-Serrano et al. 2014, Donelly et al. 2017), as the perfect case study to test our ability to improve drought related phenomena predictions and measure the impact of the proposed development.

Today, Earth Observation datasets provide valuable information on drought from different perspectives. While traditionally these have been based on vegetation, notably due to the difficulty in accurately quantifying precipitation from remote sensing data, the main drawback in assessing drought through vegetation indices is that the drought is monitored when effects are already causing vegetation damage. In order to address drought in the early stages, we need to monitor it from the moment when the lack of precipitation occurs, and the advent of soil moisture dedicated missions has paved the way for drought monitoring based on soil moisture data (Escorihuela et al. 2020). Variables such as soil moisture or NDVI hold complex relationships to temperature and precipitation (e.g. by memory). Therefore it is plausible that eXplainable AI methods can describe these relationships.

Seasonal climate predictions have witnessed considerable improvements in the last two decades demonstrating that probabilistic forecasting can inform better decision making at some temporal and spatial scales (Alessandrini et al. 2013, Doblas-Reyes et al. 2013). However, despite these improvements, it is notoriously difficult to provide skilful predictions at seasonal time scales for certain climate variables. In particular, seasonal predictions of precipitation exhibit often low skill in the extra-tropics (e.g. Cohen et al. 2019; Mishra et al. 2019). Even if the chaotic behaviour of the atmosphere does not allow predicting with accuracy the changing weather beyond a few days, climate predictions for the forthcoming months or seasons are feasible because atmospheric variability on monthly/seasonal time-scales is modulated by slowly-varying boundary conditions of the atmosphere such as soil moisture, sea surface temperature, sea-ice and snow cover.

The dynamical-numerical climate models used in operational Seasonal Prediction Systems (SPS) are, however, not perfect; and exhibit persisting shortcomings. Of relevance to the prediction of climate extremes, including drought, the dominant patterns of atmospheric variability are not correctly reproduced by the atmospheric models (e.g. Walz et al 2018), and land surface models systematically overestimate drought intensity and duration in the dry season (Ukkola et al. 2016).

These types of errors lead to systematic biases that need to be corrected in post process steps. Moreover, the occurrence of extreme events are inherently tied to certain spatial resolutions (shorter than a few kilometers) that are not directly resolved by current operational SPS which present a much coarser resolution. A downscaling is therefore usually required in order to implement local impact studies and corresponding mitigation strategies. Traditionally, in-situ observations or reanalyses are commonly used as ground truth for all these post processing adjustments, although both are subject to known limitations as well. In-situ observational references often exhibit important spatial and temporal inconsistencies due to the sparse availability of historical weather records while reanalyses, being a numerical output, may present systematic biases as well.

In this sense, the use of different available satellite-based Earth Observation (EO) products (like soil moisture, NDVI, lake levels or burned area) appear to be an attractive alternative to integrate into SPS a multidimensional, high quality and high resolution source of information required to properly characterize the multifaceted nature of droughts. This alternative approach faces however the inherent challenge of combining two intimately related but different representations of the climate system; the physical drivers provided by the SPS (temperature, precipitation, humidity...) and their direct impacts on the environment as seen by the satellite missions. To bridge this gap, the recent explosion of breakthroughs and successful examples of applying Artificial Intelligence (AI), in particular the branch of deep learning, to many diverse fields (computer vision, speech recognition, natural language processing) and also to Earth science (Tsagkatakis et al. 2019, Reichstein et al., 2019) appears as an evident promising strategy. Some recent studies have already explored the potential of AI to exploit the temporal auto-correlations between EO-products alone to build skillful data-driven prediction models (Kraft et al. 2019, Foley et al. 2020). Others applied deep learning to explain in-situ environmental indicators (like soil moisture) from meteorological drivers (Cai et al. 2019). And finally, very few studies have already shown that valuable data-driven models can be built combining deep learning, meteorological data and satellite-based observations (Peng et al. 2018, Requena-Mesa et al. 2020, Klingmüller et al. 2021). However, the current approaches have only focused on describing one particular environmental impact represented by one single EO-product (i.e. soil moisture or NDVI), have been limited to explore temporal relationships of a few days or have only minimally explored the interpretability of the often inescrutable data-driven models.

In this proposal we present a comprehensive approach combining AI techniques, dynamical SPS and multiple EO-products that are appropriate for studying societal impacts (e.g. soil moisture, NDVI, lake levels or burned area) with the goal of improving our prediction capabilities and enriching our understanding regarding the causes, evolution and consequences of droughts at seasonal time-scales. We propose two lines of work (figure 1);

- The first one will aim to directly improve raw seasonal forecast outputs (temperature, precipitation, humidity…) through an AI-based bias-adjustment and downscaling post process that will include EO-data during training.
- The second one will instead provide seasonal predictions of multi EO-products making use of historical data (physical drivers and EO-products) during training.

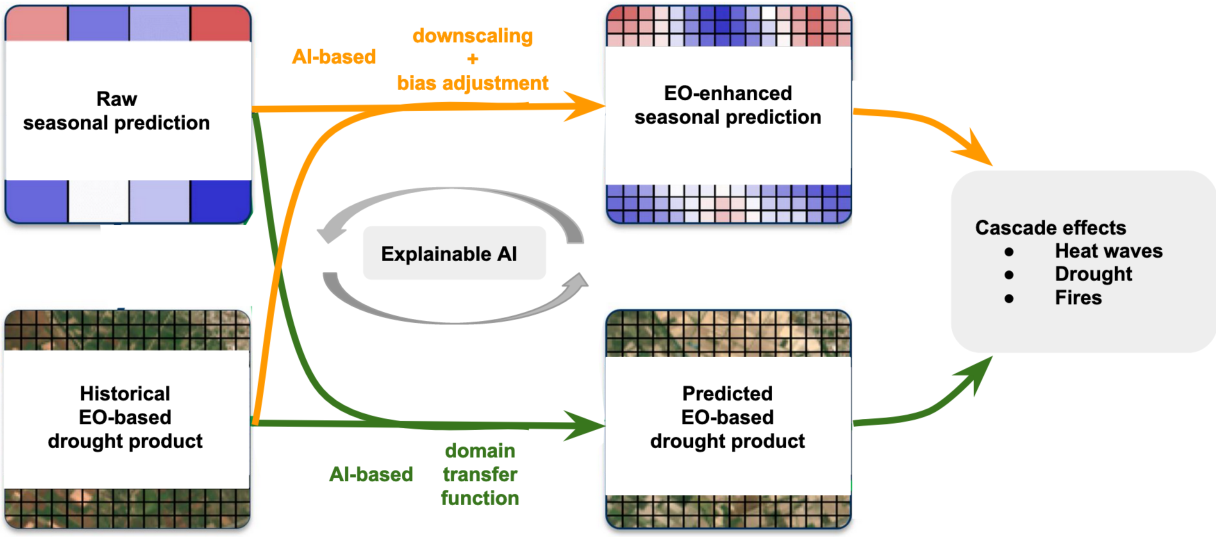


Figure 1 Overview of the proposed development in AI4DROUGHT. Adapted from Requena-Mesa et al. 2020

Both lines of work will be analysed with eXplainable AI (XAI) techniques in order to provide interpretability of the data-driven models and generate trust in the corresponding results. To unveil the potential cascade effects, a Hazards Events knowledge graph (Masmoudi et al. 2021) will be established providing causal interrelations between identified climate extreme events variables. AI4DROUGHT aims to become a milestone achievement in the generation of actionable knowledge on the relationship between Earth Observation long-term data records and modelled climate variables through innovative multi-scaled AI approaches.

The AI4DROUGHT methodology will be demonstrated in the prediction of drought climatic events over the Iberian peninsula. With end-to-end demonstration scenarios, we will assess the improvement potential of such prediction and identification, as well as the assessment of actionable knowledge generated for policy-makers and risk-managers. Finally, the proposed methodology combining numerical climate models with AI driven approaches at different temporal and spatial scales to identify multi-hazards and cascading effects will be highly scalable, replicable and transferable to other regions and applications, thanks to data driven approaches and pipelines that permit to automate and continuously store climatic experiences.

AI4DROUGHT is designed as a stepping-stone in Destination Earth, as it will provide:

- functional inputs to data lake repositories and platforms enabling connectivity and access to cloud-based modelling and multi-scale predictive tools;
- a wealth of actionable long data records on agricultural drought characterisation from Copernicus and Earth Observation sources, covering past, present, and future temporal scenarios, with a fast an easy access by application programming interfaces (APIs);
- actual simulation engines available in real/near real time combining Earth Observation pipelines and seasonal forecasting models, based on data fusion and eXplainable AI, directly linkable to Digital Twins, fully documented for their integration;
- a multi-hazard events knowledge graph around drought will store the causal interrelation between climate extreme events variables and cascading effects, under which semantic analytics and reasoning will serve to distil collateral cascading effects, demonstrating the direct impact of having a Digital Twins of the Earth system (Bauer et al. 2021).

1.2 SCIENTIFIC OR TECHNICAL OBJECTIVES

The following are the main objectives to be achieved. Under each objective, specific target points are summarised, which contribute to achieving the objectives and their demonstration.

Objective 1 tackles the scientific challenge in AI4DROUGHT and Objective 2 the technical one.

Objective 1: To enhance our knowledge on the cause and effects of drought events by the combination of the complementary climate system descriptions provided by EO-based observations and Seasonal Prediction Systems (SPS) through the implementation of AI-based algorithms

a. Improve operational SPSs raw outputs through AI-based post-processing techniques trained including EO-products to increase spatial resolution (downscaling) and to remove systematic biases leading to an enhanced predictive skill.

b. Provide seasonal EO-based drought product predictions computed from AI-based models that translate climate model-based seasonal predictions into simulated EO products.

c. Unveil and quantify the highly complex relationships that relate climate state variables typically provided by SPSs to characterize drought (e.g., precipitation, temperature, humidity) with those provided by EO-based datasets (e.g., soil moisture, NDVI, lakes level) through the use of appropriate explainable AI techniques applied to both predictive lines described above. These comprehensive results will enrich our knowledge of the physical processes that interplay in drought events providing at the same time hints to improve the physical models in the SPSs through better representations of the inter-variable relationships unveiled.

d. Explore the chain of possible effects related to drought events linking the initial physical drivers (precipitation, temperature,....) all the way to the final measurable cascade of impacts through the implementation of data-driven causality graph models. Additionally, we will test our enhanced ability to predict these cascading effects making use of both hybrid seasonal prediction pathways previously developed.

Objective 2: To design the appropriate deep learning architectures that allow us to maximize the extraction of information from both the EO-based datasets and the SPS.

This objective will be achieved by the implementation of the following steps:

a. Set-up of relevant operational big data ingestion and processing pipelines to ensure smooth running of all tasks.

b. Development of an AI-based bias adjustment and downscaling post processing model inspired in image-to-image translation deep learning architectures that will incorporate drought related information from relevant EO-products to enhance the predictive skill of raw SPS.

c. Development of an AI domain transfer function based on image-to-image translation deep learning architectures that will convert most relevant droughts' physical drivers (e.g., temperature, precipitation) provided by SPS into convenient EO-based drought products.

d. Analysis of hazard events variables to predict potential risks and also, provide a hintered link between the variables to generate automatic pipelines for annotating risks and cascading effects.

e. Elaboration of a data-driven causality graph model (called "Hazards Events Knowledge Graph") in order to annotate the different extreme events and predict/identify potential cascading effects using semantic analytics.

1.3 REQUIREMENTS TO BE ADDRESSED BY THE PROPOSED DEVELOPMENT

The following requirements are defined as expressed in the SoW, and according to the understanding of the consortium.

Consistent language will be used in the requirements wording in accordance with the following conventions:
- "shall" indicates a mandatory provision.
- "should" indicates a recommended provision.
- "will" is a declaration of purpose such as a design goal.
- "may" indicates a permission for a provision.
- "can" indicates a possibility or capability.

A first iteration of the AI4DROUGHT requirements is presented, coming from relevant SoW requirements and a list of preliminary requirements compiled by the consortium according to different nature: scientific, conceptual and design, system and software, testing, verification and validation, data, functional and non-functional, according to the proposed work. The consolidated baseline requirements will be generated as a task in the project, after a thorough state-of-the-art and technical and scientific gap analysis.

Table 1 Preliminary Requirements Baseline

ID	Name	Description	Verification method
REQ-SW-001	Explainability	Explainability should be considered in the design of physics-driven AI	Setting of a human-grounded methodology (human-in-the-loop) for the evaluation of the model results and the rule-based outcome provided by the applied XAI techniques
REQ-SW-002	Transparency	Outputs of AI based models shall be checked against transparency	Assumptions encoded in the models must be plausible and backed up by scientific works as well as the combined expertise of the 3 partners. Development of systematic data exploratory and data causality analysis to gain transparency prior to AI model training

REQ-SW-003	Testability	Outputs of AI based models shall be checked against testability	WP3000 will focus on the development of models. The outputs will be implemented as a self-standing demonstration scenario, fully without executable open code from a set on input data, which will be also be made available
REQ-SW-004	Fairness	Outputs of AI based models shall be checked against fairness, accountability and trust	Project outputs will be tested against ground truth datasets fully documented and accessible
REQ-SW-005	Interpretability	AI solutions for Earth Science should consider aspects related to interpretability such as: - inherently interpretable and/or - permit model inspection and/or interpretation and/or - Error control, error modelling, quantified uncertainty	A combination of techniques (permutation analysis, visualization of features-heatmap activations, and clustering activations) will be used to open-up the AI black box and gain trust on what the model has learnt.
REQ-SW-006	Uses Case	The contractor shall define a use case considering one or more extreme climate and/or weather events with regional to global impact, focusing on aspects related to drivers, triggers, interactions and/or impacts	The project focused on drought in the Iberian Peninsula, and the climatic variables that characterise this agricultural drought in particular, from the physical modeling and Earth Observation worlds. Drought cascading events will be analysed in the combination with heatwaves and consequent fires.
REQ-SW-007	Test and Validate	The contractor shall implement, test and validate an AI4EO methodology and algorithm incorporating explainability and demonstrate its suitability for the proposed Earth System Science use case focused on extreme events	Task 4.1 will provide with an independent benchmarking of the methods developed Task 4.2 will result in an error analysis and uncertainty assessment applied to drought cascading events in the Iberian Peninsula in the last decades
REQ-SW-08	Scalability	The contractor shall deploy the solution on a cloud infrastructure and demonstrate the solution is computationally scalable	Task 3.4 implementation and results will be uploaded to a cloud repository, together with the input data needed to execute the scalable demonstration scenarios end-to-end. A data lake will be set up in Task 3.2 which will greatly support scalability
REQ-SW-09	Workshops	The contractor shall organise collocated workshops at the Phi-week 2021/2022 and/or ESA Living Planet Symposium 2022	WP5000 addresses both conferences
REQ-	Demonstration	The proposal shall demonstrate	Performance of the proposed methods will

SW-010		the viability and performance of their proposed methods on concrete demonstration scenarios with occurrence or impact at regional to global scale.	be measured by Task 4.2 on the demonstration scenario cascading effects
REQ-SW-011	Data availability	The proposal shall describe whether and how the algorithms and data products generated through the course of the proposed research will be made available to the scientific community for replication and validation	Self-standing end-to end demosnatrtion scenarios will be implemented on a cloud repository as open source with access to the input data for its full execution. See Implementation aspects.
REQ-SW-012	Duration	The project shall be carried out within maximum 24 months	Management and contractual sections
REQ-SW-013	Open Science	The projects should adopt as much as possible the principles of Open Science	AI4DROUGHT will provide open source code and will promote publication in open access journals
REQ-TVV-001	Reliable Risk Assessment models	Demonstrate the viability and reliability of the risk assessment digital tool and AI driven models in specific scenarios of droughts with the theoretical study to transfer to other types of scenarios and casuistics	Demonstration of a combination of XAI and semantic analysis over a semantic knowledge graph and climate seasonal modelling outputs to identify hazardous events and generate actionable cascading effects pipelines.
REQ-TVV-002	Analysis of cascading impacts caused by extreme events	Combination of seasonal climate models with AI to enhance the understanding of risks and adaptation pathways.	Elaboration of a Hazards Events knowledge graph to analyse patterns related to hazardous events and potential cascading effects caused by previous and historical events.
REQ-TV-V-003	Discovery of extreme events patterns using AI	Combination of the seasonal climatic models outputs and EO / observations with AI driven tools to the identification and annotation of extreme climate events	Elaboration of two-layered AI driven tools to (i) detect patterns produced by the combination of different EO variables; and (ii) detection of patterns caused by temporal variability of the variables.
REQ-DATA-001	EO input data consistency	EO datasets need to be regionally comprehensive, global in scope and consistently formatted to be useful for evaluating and improving climate and Earth	A data preparation step will format and prepare cropped global data in its original grid, converted to Zarr, generate metadata files in STAC and make them available in a cloud-based data lake repository

		system model (National Research Council. 2012)	

1.4 INNOVATIVE ELEMENTS WITHIN THE PROPOSED DEVELOPMENT

AI4DROUGHT is highly innovative in nature, as it responds to the AI4Science initiative overall. In the context of the proposal, the following elements can be highlighted, as they represent a clear advancement in the state-of-the-art:

1) Modelling of Earth Observations from physical model outputs will allow the prediction of climate hazards such as drought, heat waves and fires

Climate change is global, although its visible impacts have a great variation at local scale. Earth's surface observations and relationships with driving climate variables remain partly unresolved by existing physical models, and in some cases even unknown. We build an AI model predicting future earth observations with inputs combining future mesoscale climate variables obtained by physical modelling (reanalysis or projections) and past memory of high resolution earth surface observations. This approach is unprecedented since it places EO data predictions as an intermediate task for climate impacts modelling. By using high resolution grountruth provided by earth observations, the proxy for climate hazards assessments (drought, heatwaves, hazards) turns into a high resolution dataset. Last December, at the 2020 NeurIPS conference, Requena-Mesa et al. 2020, presented their EarthNet 2021 analysis ready dataset, containing target spatio-temporal past and future Sentinel 2 satellite imagery at 20 m resolution, matched with high-resolution topography and mesoscale (1.28 km) future weather variables. The challenge is to predict future NDVI as seen from space, given coarse weather projections. By following this forefront path, we aim at forecasting a larger pool of EO data (soil moisture, NDVI, water bodies, burned area ..) and at a seasonal time scale.

2) Data assimilation, or the process of combining observations and the forecast output from a weather/climate prediction model, is a challenging and imperfect process with many data constraints.

AI techniques can help to fill the gap between modelled datasets and EO products by learning the complex spatio-temporal and multi-variate statistical links that exist between climate variables and satellite products. This knowledge will be extracted from observations-based records (gridded observations, reanalyses and satellite products) and transferred to seasonal predictions in order to correct its biases and to expand the portfolio of products that a climate service can offer. These statistical relationships will be encoded in AI-based transfer functions that can be trained on reanalysis and satellite products and then applied to variables from a seasonal forecasting system (Perfect prognosis) or learnt directly by fitting seasonal predictions to satellite and reanalysis products (Model Output Statistics).

3) Working with a production mindset in order to deliver operational products from state-of-the art research

Big geo data access, management, processing and visualisation define computational constraints preventing from delivering models adapted to deployment. Therefore, strong efforts in cuning state-of-the art models can hide impossibilities of applying these models as solutions to real-world problems. By building on all recent years tools, platforms, libraries, data standards, production frameworks (see 1.3 Requirements), we aim at thinking operationalisation since the first moments of conceptualisation.

4) Next generation of AI4EO: stepping up from traditional vision problems to multi-types data blending

Although DL-based techniques, namely convolution-based architectures (among the frequent frameworks of AI4EO are found CNNs, ResNets, UNETs, conv-LSTM), have been used extensively in EO and climate science (Reichstein et al. 2019), there is not a general pre-designed implementation that fits all purposes. Visions problems have been particularly tackled, since their transfer domain from natural images to remotely sensed data appears straightforward. Classification models (most recent EfficientNet, NFNets, ViT), semantic segmentation (UNET, Segnet), instance segmentation/detection (most recent Mask RCNN, Yolo V5, Transformers DETR) are being adopted by the deep learning EO community for vision at the rhythm of the release of the frameworks implementation. However, in AI4DROUGHT, we will address the challenges of cross-domain data fusion (improving over the naive approach of adding variables as channels in the input of a CNN), domain transfer (in the perfect prognosis approach), of mapping a probabilistic input (SPS) to an observational reference (in the model output statistics approach), and producing a probabilistic improved seasonal prediction (instead of a deterministic one) for uncertainty quantification purposes.

5) Actionable and Augmented information for AI modelling interpretability

One of the main challenges of AI4EO is on improving the understanding of extreme climate events and corresponding cascading effects. During recent years, there has been noticed the importance of cross-correlating and analysing past events from different spatial and temporal scales (Rudd et al. 2019, Brunner et al. 2016, Kemter et al. 2020). These studies and analysis are mostly reported in literature but some few of them have been operationalised in specific databases. Therefore, there is a lack of knowledge and data to empower decision-making in relation to the (predicted) occurrence of extreme climate events. Complementing this information, there is also a trend on generating open data spaces in several areas due to making information and knowledge available for a broadening range of disciplines including environment, water, etc. Under this perspective and trends, AI4DROUGHT will advance on the generation of knowledge graphs and databases to provide augmented information considering different EO variables interrelated at different scales and temporalities. Moreover, AI4DROUGHT will also make this data openly available to be used by other digital systems (Digital Twins) or scientists.

1.5 SCIENTIFIC OR ENGINEERING DEVELOPMENT APPROACH

1.5.1 Scientific/Technical Steps

This project will develop methodology to explain the relationship between Earth Observation long term data records and modelled climate variables allowing the improved prediction of key variables such as soil moisture in forecasting systems. The impact of this work will be measured in the scope of cascading effects in drought related phenomena, including environmental and socioeconomic factors.

AI4DROUGHT will follow a clear strategy through the following steps:

1.5.1.1 Data preparation and working environment

An object store or data lake will hold all inputs necessary for the various algorithmic components, as well as their outputs. This data lake will be exploited both during initial state-of-the-art analysis and experimentation, through the implementation of the refined prototypes and up to the final cloud-based capability demonstration.

Input Data

1. CCI and selected EO datasets providing global long-term data records appropriate for climate studies

Under ESA's Climate Change Initiative, international science teams are undertaking research to generate 21 Essential Climate Variables (ECVs). These are key indicators that describe the Earth's changing climate and are defined by the Global Climate Observing System. The 21 ECVs produced by the Climate Change Initiative (CCI) teams are ECVs that can be primarily generated from satellite data. They are validated against independent datasets, they have high levels of traceability and consistency, and include quantitative estimates of uncertainty.

Table 2 Climate-quality data records are considered for their suitability to climate research

	start	end	Spatial resolution	Temporal resolution	Sensor	Ref
Soil moisture	Nov-1978	Ongoing	0.25º	daily	multi	CCI
Burned area	Jan 1982	Dec 2018	0.25º	monthly	AVHRR	CCI
Cloud	Jan 1981	Dec 2016	0.05º	daily	AVHRR	CCI
Land cover	1992	2015 (extended to 2020	0.3 km	yearly	optical	CCI
lakes level	1992	2019	0.05º	daily	multi	CCI
NDVI	1981	ongoing	0.05º	daily	AVHRR	NOAA
LST	1979	2009	0.5	hourly	LandSat	Princeton Hydrology

2. Climate modelled datasets and in-situ observations from Copernicus Data Store:

The Copernicus Climate Change Service (C3S) provides a wealth of authoritative and quality controlled climate modelled datasets, such as Reanalyses and Seasonal Forecasts, that can be used to inform of risk of extreme events in the upcoming months.

From C3S, we will use the ECMWF SEAS5 seasonal prediction system forecast, which contains 25 members and a hindcast (i.e. retrospective forecasts) starting in 1993. As observational reference, the ERA5 reanalysis will be the benchmark employed as ground truth for bias adjustment and verification, although in this project EO products will be used as well for that purpose. Other high-resolution observational gridded datasets such as E-OBS, Iberia01 and ERA5-Land are also suitable for the purpose of increasing accuracy, resolution and skill.

Table 3 Quality controlled climate modelled datasets from the CDS

Dataset	Vars	Spatial coverage	Spatial resolution	Time coverage	Time resolution	Gaps	References

ERA5 - CDS	psl, g500, tas, ta850, hur850, hus850, hurs, huss, prlr, uas, vas, ua850, va850...	Global	0.25º	1979-2020 (preliminary version extension up to 1950)	hourly	NO	https://doi.org/10.24381/cds.bd0915c6
E-OBS v20	tas, prlr, tasmin, tasmax, humidity, radiation...	Europe	0.1º	1950-2020 Extension up to 1920	daily	Only land values	https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php
ERA5 land - CDS	uas,vas,psl, rsds,prlr,swv,tas	global	0.1° x 0.1°; Native resolution is 9 km.	January 1981 to present	Hourly	Only land values	https://doi.org/10.24381/cds.bd0915c6
Iberia01	tas,tasmin,tasmax,prlr	Iberian peninsula	0.1º regular (and 0.11º CORDEX-compliant rotated)	1971–2015	daily	No	https://digital.csic.es/handle/10261/183071
Seasonal: SEAS5-C3S	g500 g850 hus700 hus850 prlr psl rsds sfcWind ta850 tas tos	global	1º	Hcst 1993-2016 Fcst 2017-now	6-hourly/daily	NO	https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels?tab=overview

Based on the above mentioned input data, an initial data volume assessment for the Iberian Peninsula (Portugal and Spain) has been carried out and is summarised in the following table. Note that the SEAS5 seasonal forecast is assumed to be obtained directly from CDS or its BSC mirror in an operational pipeline configuration.

Table 4 Data lake volume estimation

Dataset	**Characteristics**	**Original format**	**# variables considered**	**Data volume [GB]** Global	**Data volume [GB]** Iberian Peninsula
ERA5, daily	From raw hourly data, 1979-2020	netCDF	20	638.3	1.28
ERA5-Land, daily	From raw hourly data, 1979-2020	netCDF	7	1391.5	2.79
E-OBS	1950-2020	netCDF	10	60.0	0.12
Iberia01		netCDF	4	2.8	2.80
CCI soil moisture	1978-2021	netCDF	All	12.0	0.02
CCI burned area	1982-2018	netCDF	All	17.8	0.04
CCI cloud	1981-2016	netCDF	All	64.7	0.13

CCI land cover	1992-2020	netCDF, GeoTIFF	All	67.0	0.13
CCI lake levels	1992-2019	netCDF	All	2100.0	4.22
NDVI	1981-2021	netCDF	All	780.0	1.57
LST	1979-2009	netCDF	All	52.0	0.10
TOTAL				**5186.2**	**13.21**

Data will be downloaded from their sources to in-house equipment, cloud-optimised and uploaded to the bucket. This cloud optimisation procedure will entail:

- Cropping the original grid, without regridding
- Conversion to a unified format, Zarr (file sizes very similar to the original ones above)
- Creation of multiple dataset views
- Generation of metadata files (STAC)
- Uploading of data and metadata artefacts to the data lake

Zarr format. Beyond the in-depth expertise with this format gathered recently by the partners of this consortium, Zarr has been *a priori* selected for this project due to multiple factors:

- Its **suitability for multidimensional data** such as the multivariate, spatiotemporal fields required in this Project. By contrast, Cloud-Optimised GeoTIFFs are limited to the geospatial dimensions, requiring a large number of objects to represent the same data and losing the capability to compress in non-geospatial dimensions (e.g. time, elevation).

- Its unique combination of **simplicity** (just a chunked N-dimensional matrix, which can be distributed globally through a Content Delivery Network) and **flexibility** (arbitrary selection of compression schemes and key-value stores, arbitrary metadata properties, etc.), which in practice have led to a proliferation of **interoperable software tools**, especially in its native Pangeo ecosystem but also more and more in the data visualisation domain.

- Compatibility with **current backend-agnostic tools**, e.g. xarray, dask.

Given the relatively low data volumes, it is expected that *multiple* copies (so-called *dataset views*) be stored, with different chunking configurations. These will enable various use cases, such as optimised bulk processing (large chunks, reducing overhead), and interactive exploration (small chunks in multiple dimensions, facilitating access to small subsets).

The SpatioTemporal Asset Catalog (STAC). Metadata will be stored alongside the data in the same serverless fashion, using the emerging standard STAC. Data browsing and discovery will be possible thanks to interoperable tools such as Pangeo's Intake, or downstream services which provide additional features (e.g. search).

High scalability: storage. The selection of object storage for the data repository aims to cope with the demand for high scalability. Given the serverless architecture, no infrastructure redesign is needed even if eventually the area of interest is expanded to global coverage. Storage itself follows the pay-as-you-go model, so it is virtually boundless, and high performance levels are independent from volume (typical values shown):

- Data availability: ~99.99%
- Data durability: ~99.999999999% over a given year
- Low latency, thanks to fast origins (S3 buckets), a dense CDN, and no server in the middle

Object storage features are provided not only by conventional cloud providers (AWS, GCP, Azure) but also by EO-specialised clouds (DIAS).

High scalability: containerisation. All algorithm components will be containerised, with two goals in mind: (i) facilitating portability and repeatability; and (ii) enabling parallelisation and horizontal scalability (clustering) in a large data volume scenario, also benefiting from the high throughput and parallel-access characteristics provided inherently by the data lake

1.5.1.2 Modelling approaches (applied data fusion)

Machine learning models, and specifically deep learning models, are universal approximators able to represent highly complex nonlinear spatio-temporal and multi-variable relationships. Their successful application in diverse Earth science fields have been already demonstrated in very recent years (Reichstein et al., 2019) fueling an unstoppable motivation for further new promising results. Here, our aim is make use of AI to maximize the extraction of information from the available EO-based products related with droughts in order to incorporate it to current operational SPS to generate two types of complementary and improved predictions; (1) EO-corrected seasonal predictions of common drought's physical drivers (e.g., temperature, precipitation, humidity) and (2) EO-based products (e.g., soil moisture, NDVI, lake levels) at the seasonal timescale. Models from both approaches will learn during training the relationships between physical drivers and the EO-based products but they will differ in providing distinct and complementary outputs (see figure 1 and 3).

Supervised Deep Learning (DL) architectures, such as those commonly used in computer vision (Salcedo-Sanz et al. 2020) are particularly suited to the spatial structure of the rasters embedding climate variables and satellite imageries. Common applications in computer vision encompass classification (most famous architectures are ResNets, DenseNets, EfficientNets, NFNnets among other CNN-based models), semantic segmentation (UNets, Segnets and other Encoder-Decoder types architectures), detection/instance segmentation (Yolo, Faster R-CNN/Mask R-CNN). Generative models, like Variational Autoencoders or Generative Adversarial Networks, represent very promising architectures to be used in climate science since they aim to directly learn the statistical distribution of the ground truth which nicely resonates with the inherent unpredictability of the climate system and correspondingly with the requirement of working with probability estimations instead of deterministic approaches in any type of climate prediction. Recently, Transformers are revolutionizing the DL field for vision by by-passing restricting the limited local field of view to creating connections between distant pixels thanks to the "attention mechanism". Transformers for detection and panoptic segmentation (implemented by the DETR architecture) disclose outstanding performances. Mapping architectures (segmentation/panoptic segmentation) appear to be the most suitable candidates to map these complex inter-variable Earth system relationships (Reichstein et al., 2019). However, some are easier than others to train by construction and in terms of the required amount of training data. Therefore, baselines such as well mastered UNets or ResNets should be favoured to consider for first experimentation of these multi temporal scales data fusion. Data blending should take into consideration the different input/output time scales, where blocks of convolutional layers should take in input similar time correspondances if channels are composed of different features. For instance, one convolutional block dedicated for past EO data memory, and another one for future climate variables before merging these layers into the mapping architecture (which should conserve the feature maps spatial size in output).

In climate prediction, post-processing techniques are commonly applied to raw climate model outputs in order to correct systematic biases and improve the limited spatial resolution (downscaling). This type of post-processing basically correlates past model outputs (hindcasts) with available representations of the past observed weather, usually from ground observations or from reanalyses. Two possible approaches are usually chosen for doing so (Maraun et al. 2010); Model Output Statistics (MOS) and Perfect Prognosis (PP). MOS aims at reducing the error of the model hindcast outputs by fitting them directly to observational data. Instead, the PP approach consists of two consecutive steps; a statistical model is first calibrated to correlate two independent observational references with distinct spatial scales (one with the climate model low resolution and the other with the target high resolution). And then this calibrated model is used to correct the raw climate model outputs. Both methods have specific advantages and disadvantages. MOS is more accurate by constructions since the calibration directly minimizes the model-specific systematic errors, however limited hindcast time series may lead to less robust calibrations and in turn to some degree of overfitting. On the contrary, PP may provide a more generic and thus more robust transfer function since the calibrated model is not model-specific being also trained with usually longer available time series. The drawback is that it generally provides lower performance than MOS. In AI4DROUGHT we will implement DL-based models for both PP and MOS approaches and compare them in a robust way.

1.5.1.2.1 EO-enhanced seasonal prediction

Only recently, AI-based methods have been proposed to learn the highly non-linear transfer functions needed to correct climate model outputs (Chen et al. 2020, Gronquist et al. 2021, Steininger et al. 2021). These deep learning-based approaches are promising proof-of-concept of the potential of CNN-based architectures for the task of correcting information from climate models, although it is worth pointing out that none of them have focused on improving a dynamical seasonal forecasting system and neither of them have incorporated EO-based products to enhance the predictive skill.

We will train our CNN-based models including as predictors all the relevant physical drivers fields (e.g., temperature, precipitation, humidity) at inference time-step t from the raw SPS hindcast in MOS or from the low resolution observational reference in PP. In order to incorporate EO data during training, we will also incorporate as inputs the selected EO products at time-step t=0 to provide a further complete representation of the climate system initial state. The ground truth will always be the observational reference of the physical drivers at high resolution together with the EO products both at time-step t.

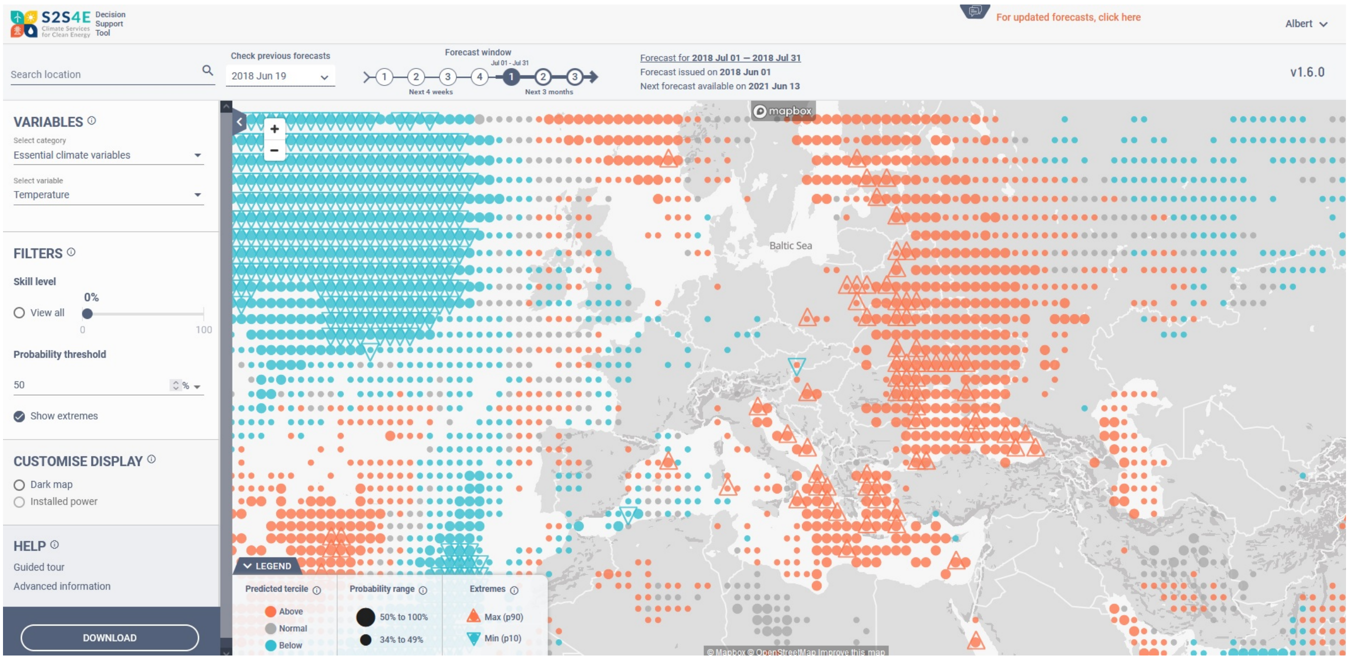


Figure 2 Seasonal forecast for temperature available in the S2S4E DST (https://s2s4e-dst.bsc.es/) for July 2019 issued in June. Only the forecasts where skill is above climatology are shown. The summer of 2019 in western-central Europe was marked by two short episodes of extreme heat.

1.5.1.2.2 EO-based drought products prediction

Again, only very recently some studies have explored the potential of predicting specific EO-based products into the future making use of AI (Cai et al. 2019, Klingmüller et al. 2021). However these studies focus on predicting one single EO field. Here, in order to extract as much information as possible from the EO world and also in order to better characterize the multidimensional nature of droughts, we propose to explore the use of multiple EO products at the same including fields like soil moisture, ndvi, lake level and burned area. This strategy may slightly reduce the predictive skill for specific fields but will lead to a more robust AI domain transfer function between the physical drivers provided by the SPS and the environmental impacts represented by the EO products since it provides more physical constraints.

Analogously to the EO-enhanced seasonal prediction, our CNN-based model will use as inputs the physical drivers (raw SPS hindcast in MOS and low resolution observational reference in PP) at inference time-step t together with the multi product EO data at time t=0 to characterize the system initial state. In this case however, the ground truth will be the multi EO-based product at time t.

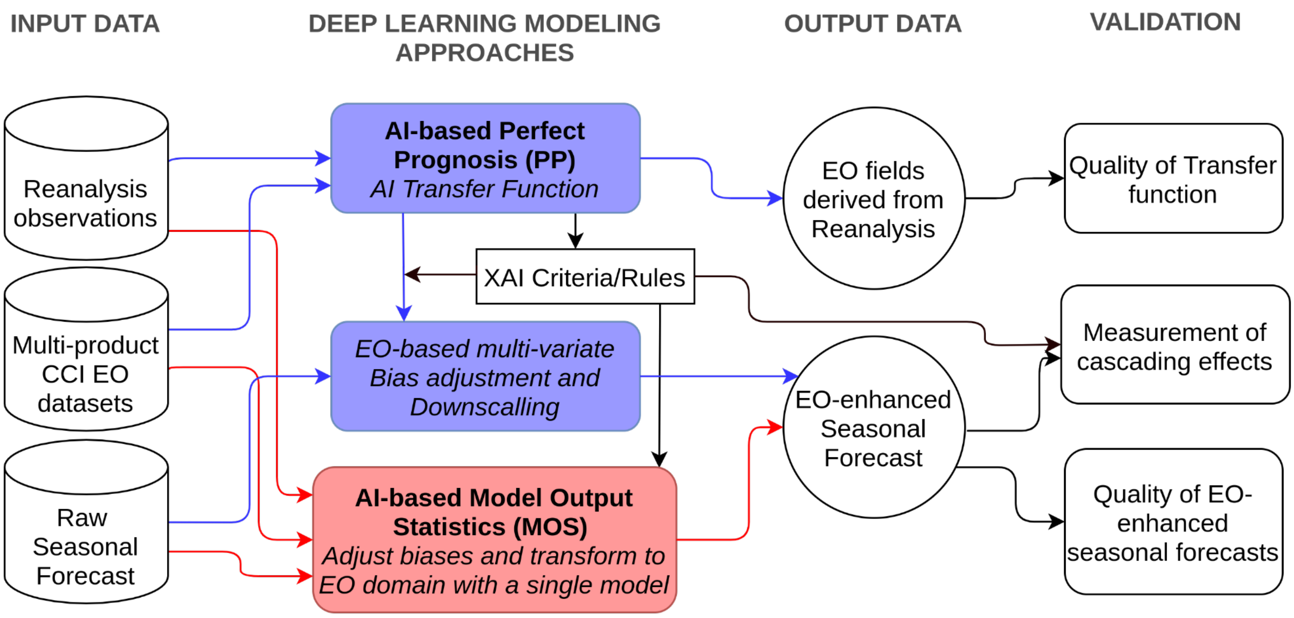


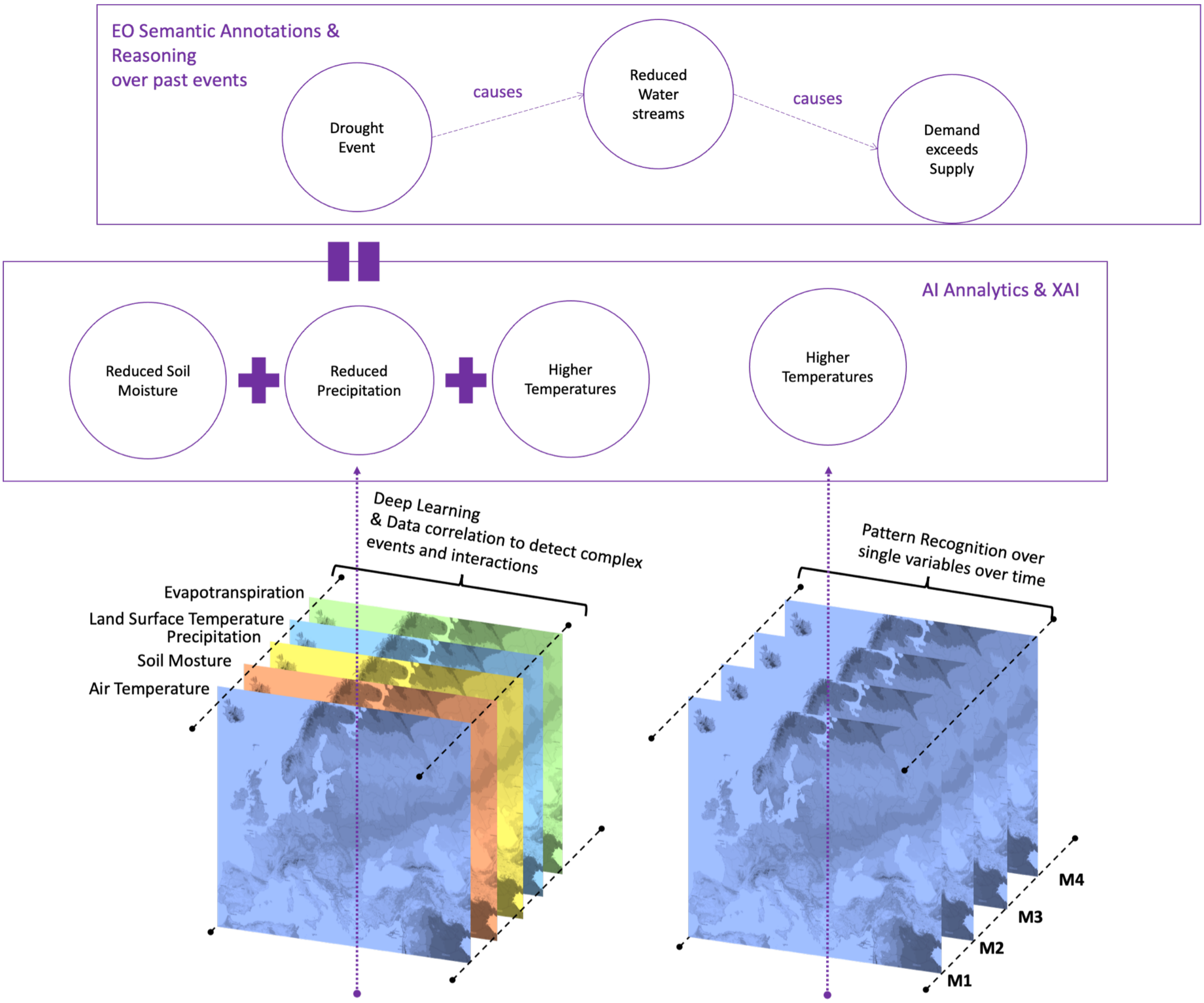
Figure 3 Simplified diagram with the general workflow proposed for the training of DL-based post-processing models.

1.5.1.2.3Cascading effects

Continuous climate change in the southern part of Europe will lead to an accumulation and intensification of different extreme events [Enenkel et al. 2021]. Drought and heat events are counting as the more impacting events in this area. Moreover, this region is also prone to continuous episodes of fires that negatively impacts biodiversity and vegetation. It is expected these events will be more frequent in the near future causing substantial impacts on agriculture, inland waterways, forestry and land use.

In the way to understand extreme climatic events over this region, existing approaches are focused on the elaboration of different climatic models to predict specific key variables for detecting and monitoring droughts (soil moisture, temperatures, precipitation to mention a few). The advancement and accurate information generated by satellite data [Lie et al. 2021, Papoutsis et al. 2021] permit accurate predictions and indeed improved monitoring of potential events. However, there is a need to understand multi-hazard risk and the associated cascading effects. Better understanding of risk assessment and subsequent cascading effects will support policy-makers and risk-managers to prioritize mitigation/adaptation actions and lately, support on the elaboration of climate adaptation pathways considering a balance in the AFOLU (agriculture, forestry and other land use) and socio-economic parameters.

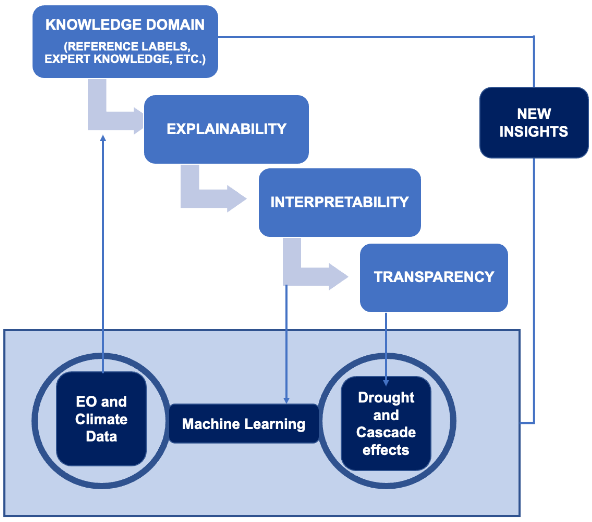
An improvement in the seasonal climate prediction taking advantage of AI will be necessary to identify extreme events at different geographic and temporal scales. However, there is a need to advance in the understanding of such risk and also the potential cascading effects that could affect different areas such as agriculture, water, energy, land cover and other relevant socio-economic aspects. Moreover, better insights of risk and cascading effects assessment will facilitate policy-makers and risk-managers to take accurate actions and design adaptation pathways at different scales. Responding to these challenges, the proposal will take advantage of a multi-scale AI driven approach to identify, categorise and analyse cascading effects of multi-hazards that could occur over different time-scales and across different observed/forecasted variables.



Indeed, the methodology we propose is based on the approaches exposed in several risk assessment articles that analyze risk at different levels [Vinuesa et al. 2020, de Brito 2021, DEEPCube (https://deepcube-h2020.eu)]. Under these methodologies, there is highlighted the importance of valuable data considering multi-scaled and temporal variables and extreme events identification and categorization. Another highlighted aspect is the analysis of the patterns and occurred extreme events to analyse hintered interconnections and interrelations between the variables to operationalise the identification of extreme events and cascading effects (through explainable AI or similar). As a response to these current challenges, the methodology proposed will combine seasonal climatic model information to analyse deeply multi-variable patterns to identify extreme events (e.g. using deep learning). Complementing this analysis, temporal analysis of the specific variables will serve to analyse potential impacts of the risks and also potential cascading effects that could occur (e.g. using time series analysis, pattern recognition mechanisms, etc). To empower the extreme climate events identification, a **Hazards Events knowledge graph** will be established considering previous experience and scientific climate information derived from existing databases such as EM-DAT [https://www.emdat.be/] and SENDAI [https://www.desinventar.net/] among others. This knowledge graph will serve to store and visualise causal interrelations between climate extreme events variables and cascading effects (link with other climate extreme events). Under this knowledge graph, semantic analytics and reasoning will serve to distill collateral cascading effects. Complementing the Hazards Events Knowledge Graph, an explainable AI mechanism will be studied and incorporated to analyse variable interrelationships and automatically annotate causal interrelations between variables and events to be stored into the knowledge graph. For that purpose, explainable AI algorithms (detailed in 1.5.1.3) will be used as a predictive risk algorithm of climatic events that permit at same time, to understand the provenance of this event.

At the end, the proposed solution will empower the generation of actionable information combining different sources of information to generate actionable knowledge through innovative multi-scaled AI driven approaches combined with seasonal climate predictions. This methodology will be demonstrated in the prediction of fire events and related impacts over the Iberian peninsula. Specifically, we will focus on the interrelation between droughts and heatwaves patterns and the interrelation on the appearance of fires in this mentioned region. With this demonstration, we will assess the improvement potential of such prediction and identification, as well as the assessment of actionable knowledge generated for policy-makers and risk-managers. Finally, the proposed methodology combining numerical climatic models with AI-driven approaches at different scales and temporalities to identify multi-hazards and cascading effects will be highly scalable, replicable and transferable thanks to data driven approaches and pipelines that permit to automate and continuously deal with climatic experiences.

1.5.1.3 eXplainable AI



Models are ultimately trained and used to improve decision making. The eXplainable AI (XAI) techniques are used intensively in several disciplines with success, but their systematization and intensive use in Earth Science is far from the rest of the disciplines [Bussman et al. 2020, Demajo et al. 2020, Sachan et al. 2020]. Moreover, the current taxonomy and classification of XAI tools is extensive and depends largely on the input data, the model used, and the extent of the explanation to be obtained. We find in the literature a multitude of methodologies that are classified under a taxonomy in which the extension of the explanation is taken into account (local methods - global methods), its integration with respect to the adjustment of the model (pre and post trained) and its specificity to the algorithmic family used [Linardatos et al. 2021]. However, there is no standardized use of these techniques in Earth Sciences that allows designing an optimal flow of techniques for the interpretability of black box models [McGovern et al. 2019]. Therefore, the objective of this section is to design a methodological scheme that allows gaining depth in the knowledge of drought events, combining observational measurement data with EO data.

To do this, this task will be developed in three main blocks :

- Generation of a specific taxonomy for earth science, specially adapted to drought prediction, which takes into account the types of interpretation that give value to this discipline.

- Selection of techniques according to defined taxonomy to open up de AI4DROUGHT models to gain understanding about their learning performance and predictions. Rule-based explanations for drought events anc cascade effects will be generated.
- Evaluation of the explanations and results generated by each technique.

The first part of the work will consist of defining the outcomes of the current methods and adapting them to the domain of Earth Science. In this sense, this taxonomy will be stratified around three dimensions:

Passive vs active approaches:

It divides the existing approaches according to whether they require changing the network architecture or the optimization process. The passive interpretation process starts from a trained network, with all the weights already learned from the training set. Thereafter, the methods try to extract logic rules or extract some understandable patterns. In contrast, active methods require some changes before the training, such as introducing extra network structures or modifying the training process. Until now, the techniques used successfully in Earth Sciences are based solely on a passive approach.

Type / format of explanation:

It refers to the type of explanation that we want to obtain and fully responds to the project's intrinsic objectives, rather than to the construction of the model itself. In this case, explanations can be obtained around the importance and weight of the attributes, make sense of hidden network layers (Hidden semantics) or obtain sets of logical rules. Current applications, which have been implemented in remote sensing image classification environments, focus on the development of techniques for obtaining hidden semantics, specifically.

Global and local interpretability:

The last dimension, from local to global interpretability (wrt the input space), has become very common in recent papers where global interpretability means being able to understand the overall decision logic of a model and local interpretability focuses on the explanations of individual predictions.

The selection of XAI techniques will be done under a stratification methodology that encompasses both model-specific and model-agnostic techniques [Kakogeorgiou et al. 2021].

On the one hand, we will work with global methods to understand how the model makes decisions, based on a holistic view of its features and each of the learned components such as weights, other parameters, and structures. Global model interpretability helps to understand the distribution of the target outcome based on the features. Different techniques will be approached to obtain indicators of feature importance, such as permutation importance or Accumulated Local Effects (ALE) to obtain measures and relationships of the different variables in the final prediction. In this case, ALE is chosen instead of PDP's (Partial Dependence Plots) due to the high correlation of the input variables. Special attention will be paid to agnostic techniques, since they are not dependent on the specific algorithmic approach (LIME, SKATER, SHAP and ANCHOR).

On the other hand, local techniques will also be applied to understand and characterize how the ESM measures behave in relation to the EO variables by adjusting the AI transfer with deep learning techniques. For this, gradient methods, deconvolutional and guided back-propagation will be used.

This point is especially relevant since the methodology presented in the proposal uses both EO data and physical model data, so the application of these XAI techniques in this context provides the opportunity to relate SPS parameters with drought events. Precisely, the use of observations and measurements will give greater reliability to the drought prediction model, and it is much more advisable to apply XAI techniques in this case, since its value is especially relevant in models with very high-performance metrics. Moreover, rule based and global explanations will be derived to gain a deeper understanding of drought and extreme climate events.

Evaluation will take as input the adjusted model and the explanations provided by the implemented methods. To this end, we will involve human evaluators to perform a human-level assessment of the explanatory properties of the tools to measure the extent to which the explanations of the models are human-friendly and understandable. In addition, the ability to explain will also be measured through the complexity of the model.

The transferability of each method, that is, the ability of the explain technique to work across multiple machine learning models, will also be considered. For example, specific-model methods have low transferability. In addition, other properties will be taken into account in the evaluation of the results:

- Stability: measures how similar the explanations are for similar instances.
- Importance Grade: measures how well the explanation reflects the importance of the features.
- Representativeness: measures the proportion of instances that are covered by the explanation method.

1.5.1.4 Benchmarking-based validation

Bias adjustment is a fundamental step to reduce forecast errors and produce usable, tailored and high-quality climate predictions (Doblas-Reyes et al., 2013). Current methods (Torralba et al., 2017) apply bias adjustment post-process to adjust the statistical properties of climate predictions to those of an observational reference. The evaluation of the forecast quality is based on the comparison of the hindcasts and past observations. This comparison of conventional bias-adjustment methods used as benchmarking, will allow us to analyse the merits and caveats of the new AI methods using EO developed in AI4DROUGHT. A benchmarking exercise will be carried out in T1.1 that will be used for comparison in T4.1.

1.5.1.5 Error analysis and uncertainty measurements

While these errors persist in state-of-the-art numerical climate models, blending the climate model-based seasonal predictions with observational data using Artificial Intelligence (AI) methods holds great promise to improve the skill of seasonal predictions. In particular, drought and heat extremes affect the state of the land surface, and therefore EO-based data representing different land surface characteristics (such as soil moisture or vegetation) can help to correct some of the existing errors when predicting those extremes.

AI and in particular the branch of deep learning has made advancements in EO using computer vision to analyse patterns over vegetation or land cover to mention some. Deep learning has also shown great potential across many different problems in EO (Tsagkatakis et al. 2019) and climate sciences (Reichstein et al., 2019) thanks to its ability to exploit large spatio-temporal and multi-variable datasets.

Modelled datasets such as seasonal predictions and reanalyses provide information on a set of climate variables that are key to simulate the ocean/atmosphere/sea ice dynamics. On the other hand, EO products are typically derived from satellite radiances, quantities that are not readily available from models. Therefore modelled datasets and EO products are sometimes difficult to compare and combine. AI techniques can help to fill the gap between modelled datasets and EO products by establishing the complex spatio-temporal and multi-variate links that exist between climate variables and satellite products.

This knowledge can be extracted from observed records (reanalyses and satellite products) and transferred to seasonal predictions to correct its biases and to expand the portfolio of products that a climate service can offer. This knowledge and relationships will be encoded in AI-based transfer functions that can be trained on reanalysis and satellite products. Lately, this AI-based transfer functions will be applied to variables from a seasonal forecasting system (Perfect prognosis) or learnt directly by fitting seasonal predictions to satellite and reanalysis products (Model Output Statistics). The explainable component of these AI methods is also very valuable to help advance the scientific understanding of key processes in the physical sciences that govern the complex interactions between modelled climate variables and observable quantities. From the point of view of building effective climate services, it is also beneficial to have an understanding of the physical aspects and causality links that can back up the predictions of extreme events. Accompanying a prediction with a storyline (i.e. the reasoning behind a specific prediction) can give confidence to decision-making as it hints on the original sources of the uncertainties.

1.5.1.6 Output preparation for transfer knowledge: Demonstration scenario

Once the methodology is fully tested, and the models are working at a known performance, an end-to end demonstration scenario will be fully documented and implemented in open source code on cloud environment, together with a subset of input data allowing the execution by any given third party demonstrating the impact of the improved solution on Earth system modelling on one side, and on the measurement of cascade events such as the occurrence of fires when drought and heat waves are recorded in the Iberian Peninsula.

1.5.2 Implementation aspects

AI4DROUGHT will require a great amount of experimentation with a wide range of inputs, as well as various techniques and tools. The design of the processing platform shall provide a high degree of flexibility to the scientific experts. It shall also ensure a clear migration path to move from a Development (experimental) stage to a Production stage with limited effort.

In the following diagram a first iteration on the overall system design is presented. Arrows apply to each of the interfaces that will be defined within the system

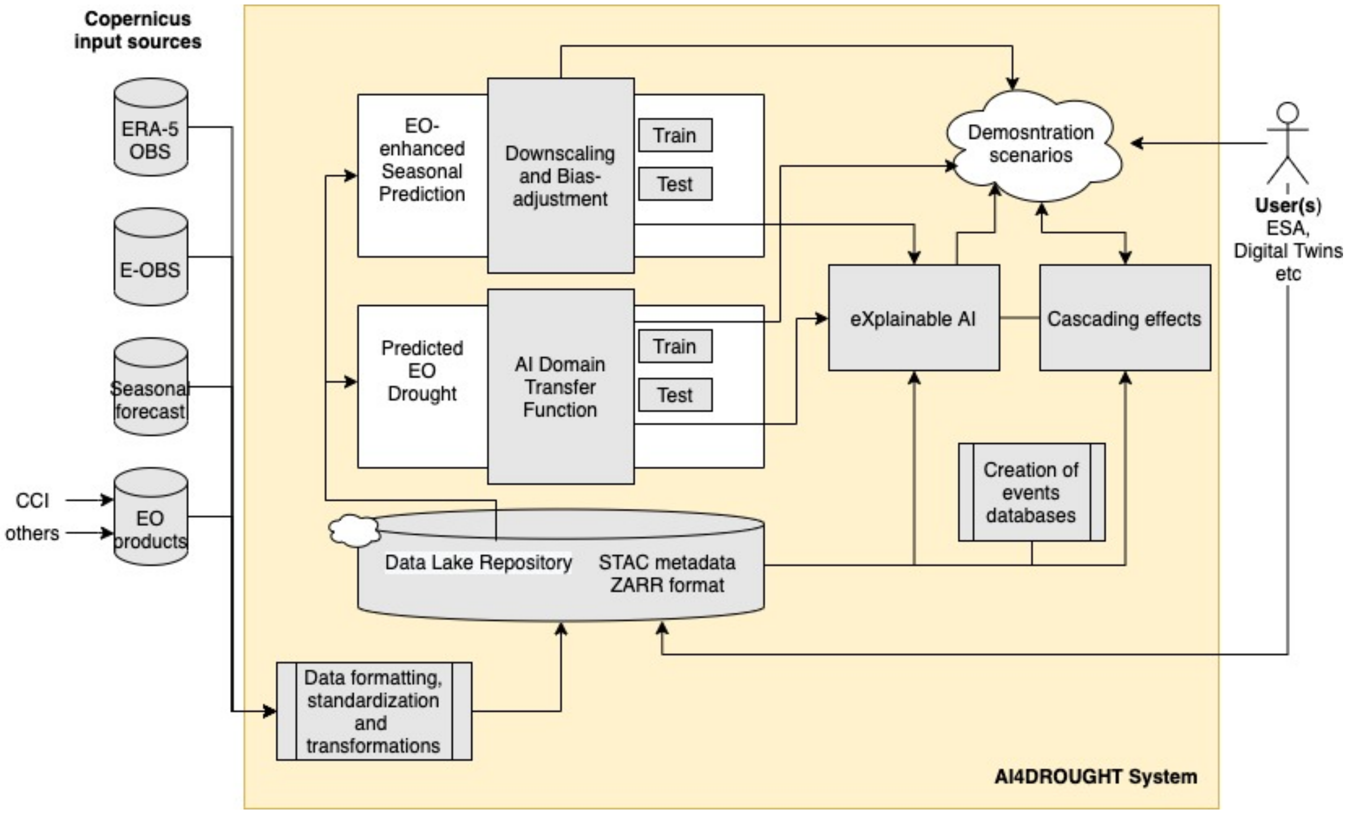


Figure 4 Baseline system design

Data consistency and harmonization (of the datasets listed in Section 1.5.1.2): In particular EO datasets will be calibrated, regridded to match the resolution and temporally resampled to match the resolutions and temporal samplings of the reanalysis and seasonal forecasting system. We may explore AI-based downscaling post-processing for coarse climate data, but will default to a conventional interpolation-based regridding when needed. A common data lake will be used when possible, at least for a subset of the EO and climate data used for training the AI post-processing functions.

Common strategy for Machine Learning training: A strategy should be implemented and agreed throughout the project, including common data formats and data structures, a data splitting strategy for training, validation and testing purposes, and techniques for efficient training of deep neural networks on big datasets. We will rely on existing open-source packages such as:

- xarray for input-output
- holoviews, hvplot, ecubevis for interactive data visualization
- scikit-learn for traditional ML tasks
- jupyter-lab for data exploration and experimentation with DL architectures
- tensorflow (keras) and pytorch for the implementation of DL models
- horovod for efficient multi-device training of DL models (either for data or model parallelism)
- xskillscore for forecast verification
- easyVerification, multiApply, startR, easyNCDF, CSTools, s2dv for climate science tasks

Tests will be conducted as part of the common Machine Learning approach and agreed by the three partners. A clear test plan and test workflow will be integrated in the Technical Report. Moreover, a benchmarking validation will be also conducted on the results of the demonstration scenario, to state the scientific baseline of improvements in drought prediction, and dedicated test results on error and uncertainty analysis will highlight the impact of the quality of the measurement in cascading effects.

Best scientific practices will need to be put in place throughout the project, including preliminary experimentation.

AI4DROUGHT will adopt the principles of Open Science, and will facilitate the transfer of science results adopting Open Source.

1.5.3 First iteration of Task 1: Problem formulation, state of the art and gap analysis

Drought is a very complex hazard with multidimensional implications which explain why there is not a common understanding about drought definition. In fact, there is a range of definitions for drought. In increasing order of severity, we can talk about: meteorological drought is associated to a lack of precipitation, agricultural drought, hydrological drought and socio-economic drought is when some supply of some goods and services such as energy, food and drinking water are reduced or threatened by changes in meteorological and hydrological conditions.

While the water management sector has routinely been using weather forecasts up to one or two weeks in advance, beyond this time horizon, climatological data (typically 5-30 year averages) are used. A common assumption in this method is that future conditions will be similar to past conditions. This assumption entails two inherent shortcomings. The first one is that past conditions can be highly variable, which can make them of limited use to forecast future ones. The second one is that climatology cannot predict events which have never happened before, e.g. extreme events, which can be particularly harmful and whose prediction is of special interest for stakeholders. Our knowledge of climatology is based on a finite sample of past events. This sample is limited in time, and is typically not fully representative of what can happen. Moreover, a climatological approach does not take into account changes in atmospheric dynamics, such as those caused by climate change. Climate change may render past conditions useless for predicting future events, as they may no longer be reproduced.

The skill of seasonal forecasting has improved continuously over the past two decades, due to model improvements and better data and forecast initialization. However, despite these improvements, it is notoriously difficult to provide skilful predictions at seasonal time scales because seasonal forecasting have major challenges such as: the complexity of the earth system dynamics and their integration in modelling systems, the proper definition of initial conditions, seasonal systems upgraded only occasionally (in general terms at intervals of four to six years due to the resources needed to complete the large re-forecast sets required for calibration), coarse resolution, and small number of ensemble members. Those shortcomings are mainly derived from model formulation deficiencies, the knowledge on the observed initial conditions and the limited computational resources available for its operational generation. Developing more skillful seasonal predictions would have, however, an enormous positive impact in a broad spectrum of sectors including fields as diverse as environment management, agriculture, energy production, health, tourism, logistics, transports, etc. A nice and comprehensive example of a recent application of state of the art sSPS is the S2S4E project (https://s2s4e.eu/; Soret et al., 2019) aiming to provide reliable source information for the management of renewable energy variability (see Figure 2).

Precipitation remains a very challenging essential climate variable (ECV) to model but also even to observe due to its complex nature. Ground-based instruments, including gauges and radars, are widely used across the globe for assessing precipitation. However, both of these observations have limitations. Rain gauges are prone to errors from wind effects and evaporation. Radar-based measurements are affected by errors due to contamination by surface backscatter, attenuation and extinction of signal, brightband effects, and uncertainty of the reflectivity–rain-rate relationship. As a result, comprehensive near-real-time coverage of the earth can only be recorded with satellite precipitation sensors [Magioni 2016]. Satellite Precipitation Products (SPP) provide quasi-globally in uncovered areas but indirect estimates of rainfall through cloud-top properties modelling introducing biases in many regions.

Datasets available for climate analysis and projections integrate a large variety of in-situ measurements but hardly integrate some EO observations of high spatio-temporal resolution. The sparsity of ground measurements can definitely be compensated by EO data which represent "the truth" of the earth system' status with no spatial gap. Hydrological forecast modelling is particularly a good candidate in assimilating EO datasets such as soil moisture for runoff and discharge variables and lots of efforts have been dedicated these last two years [Hang et al. 2019]. Challenges in data assimilation lie in quantifying and accounting for uncertainties in model inputs, parameters, and model structure (Liu et al. 2007), in improving computational efficiency (Moradkhani et al. 2005; Sun, Seidou, Nistor, & Liu, 2016), and in assimilating multiple types of observations such as streamflow, soil moisture, and snow-water-equivalent (Bergeron, Trudel, & Leconte, 2016).

One very solid piece of evidence that proves how EO data can in fact provide crucial information related to precipitation is the recent production of accurate precipitation fields directly estimated from processing remote sensing Soil Moisture (SM) data [Brocca 2016, Brocca 2017, Ciabatta 2018]. Surface soil moisture can be seen as the trace of the precipitation and, consequently, can be used for providing a way to estimate rainfall accumulation as well as a constraint to rainfall algorithms. One of the most famous forecasting operational models, Global Flood Awareness System (GloFAS) has for instance introduced the SMOS soil moisture data in 2020 (Baugh et al. 2020). Assimilation is especially beneficial for forecasts in ungauged basins or where processes are poorly understood. Other traces of precipitation, like NDVI or lake levels could be used in theory as well to provide an enriched and multifaceted stream of information to the system. Here the challenge relies on how to blend and incorporate this type of information taking into account their very complex and poorly unknown relationships with the physical drivers such as precipitation or temperature.

Climate predictions are affected by systematic errors resulting from the inability of global circulation models to reproduce all the relevant processes responsible for climate variability and the uncertainty affecting the initial conditions. Hence, current methods (Torralba et al., 2017) apply bias adjustment post-process to adjust the statistical properties of climate predictions to those of an observational reference. In AI4DROUGHT we want to go one step further blending the climate model-based seasonal predictions with EO data using AI methods to improve the forecast quality (or skill) of seasonal predictions. The AI transfer functions are trained in either a Perfect Prognosis or a Model Output Statistics setup, and are based on DL-based architectures for image-to-image translation. The estimation of the skill of both, the bias-adjusted seasonal forecast using current methods and the improved seasonal forecast by means of AI, is based on the performance of the forecasts to simulate past conditions. Moreover, the evaluation of the forecast quality is based on the comparison of the hindcasts and past observations. This comparison will allow us to analyse the merits and caveats of both methods (conventional bias-adjustment methods and AI-based correction using EO).

AI4EO has mainly been shown to be effective and deployed in production for vision problems, benefitting from the success of such methods in natural world problems (face detection, image enhancements filters etc..) However, data gaps due to missing pixels and diversity of dates of acquisitions of remote sensing data represent serious challenges specific to the EO world and a particular shortcoming for data fusion. While deep learning architectures are transferred to EO usage in a timely fashion after release (ResNet, UNET, Yolo, Mask R-CNN), recently released frameworks (EfficientNet, NFNnet, Transformers ViT and DETr) are still being under-represented though showing unprecedented performances. Going beyond traditional CNN-based encoder-decoder (e.g., UNET) or residual (i.e., ResNet) models for data fusion constitutes one way to explore designing the AI transfer functions of the system for perfect prognosis and MOS.

1.6 SCIENTIFIC/TECHNICAL FEASIBILITY, PROBLEM AREAS AND DEVELOPMENT RISK

Risk 1. Data pipeline, management and processing architectures delayed implementation. All data-driven based models require a very large volume of data to be organised and formatted in a suitable manner before actually playing with modelling. The first task of the project (Task 1) has been specially designed to define with a high level of agreements considering the entire processing chain of what will be the exact in/outs and what will be tried out with reasonable time evaluation. Mitigation: As explained in "Data preparation and work environment", particular attention is dedicated to designing the data pipeline in parallel to the experience design, following good practises and dedicated developed tools.

Risk 2. Lack of representative EO data to elaborate AI tools in relation to hazards identification and cascading effects. The lack of representative data directly impacts on the elaboration of the AI-driven models. As a mitigation measure, AI4DROUGHT will select reliable datasets to test and validate the AI driven tools.

Mitigation: In case of insufficient data, we will evaluate the generation of synthetic data using for example GANS technique.

Risk 3. The hazardous events based on historical information are limited. Similarly as the previous case, there is a need for the identification and selection of historical climate events to identify relevant patterns and cascading effects to model the knowledge graph and also train AI algorithms. As a mitigation measure, AI4DROUGHT will select reliable datasets to test and validate the Ai driven tools.

Mitigation: In case of insufficient data, we will evaluate the generation of synthetic data using generative modelling algorithms.

Risk 4. Selected AI driven models do not offer the performance and accuracy as expected. During the initial tasks of AI4DROUGHT, there will be selected the most suitable AI driven tools to (i) discover multivariate patterns to determine climate events; (ii) temporal analysis of the variables to predict trends and also detect fine-grained patterns; and (iii) explain the variables influence of climatic events to determine potential cascading effects. If some of the selected techniques do not offer a performance and accuracy as expected on the theoretical studies.

Mitigation: AI4DROUGHT will work with a combination of AI driven models or selection of newer ones to offer better results as possible.

Risk 5. Selected XAI techniques do not give clear and actionable knowledge. In parallel to the development of each AI model implemented in AI4DROUGHT, a selection of the most suitable XAI techniques will be carried out to (i) model performance understanding; (ii) gaining error control and (iii) generating rule based global explanations of drought and related climatic events. To mitigate the lack of human interpretable results provided by the techniques, systematic data exploratory and causality analysis will be performed in parallel to the AI model design and adjustment.

1.7 APPLICATION OF DEVELOPMENTS RESULTING FROM THE PROPOSED ACTIVITY

Being one of the most important sources of information towards the achievement of sustainable development in its many different angles (Earth resources management, policy making, or environmental legislation enforcing to say a few), Earth Observation remains underused in the scope of operational climate services. These so-called operational climate services are delivered in the scope of public and private adaptation plans to climate change, in vulnerability assessment studies, or in resilient investment scenarios looking for optimal climate adaptation interventions. In the face of the climate emergency and the necessary rush towards the achievement of Agenda 2030 goals, operational climate experts are devoted to the understanding of sectoral needs and the development of robust solutions in a fast-growing market, and while Earth Observation offers essential quantifiable information on the Earth past, present -and even future- natural resources which are the very target of sustainable development, the integration of Earth Observation data with predictive information remains too complex to be actionable.

Earth Observation has largely demonstrated its potential to contribute to climate resilience, providing unique, large-scale and high-definition measurements of the Earth system. Hence, projects such as AI4DROUGHT have the potential to truly represent a breakthrough in the operational use of Earth Observation in operational climate services, specially those targeted to climate adaptation and mitigation interventions, paving the way towards actionable, EO-fed predictive services.

The AI4DROUGHT consortium is greatly knowledgeable of the existing needs for the scientific, operational and commercial uptake of the outputs of the project as it is part of their expertise and their day-to-day activities with stakeholders, users and clients. the Barcelona Supercomputing center, a first class research center in Earth System science and supercomputing, and Eurecat, an innovation center that counts with more than 650 professionals and a dedicated applied artificial intelligence unit working on the industrial, energy and sustainability sectors, and Lobelia, a pioneering company in the use of Earth Observation satellites to address the climate emergency for international accounts, can guarantee the direct application, large uptake and high impact of the results of the project in their own fields of work across numerous sectors.

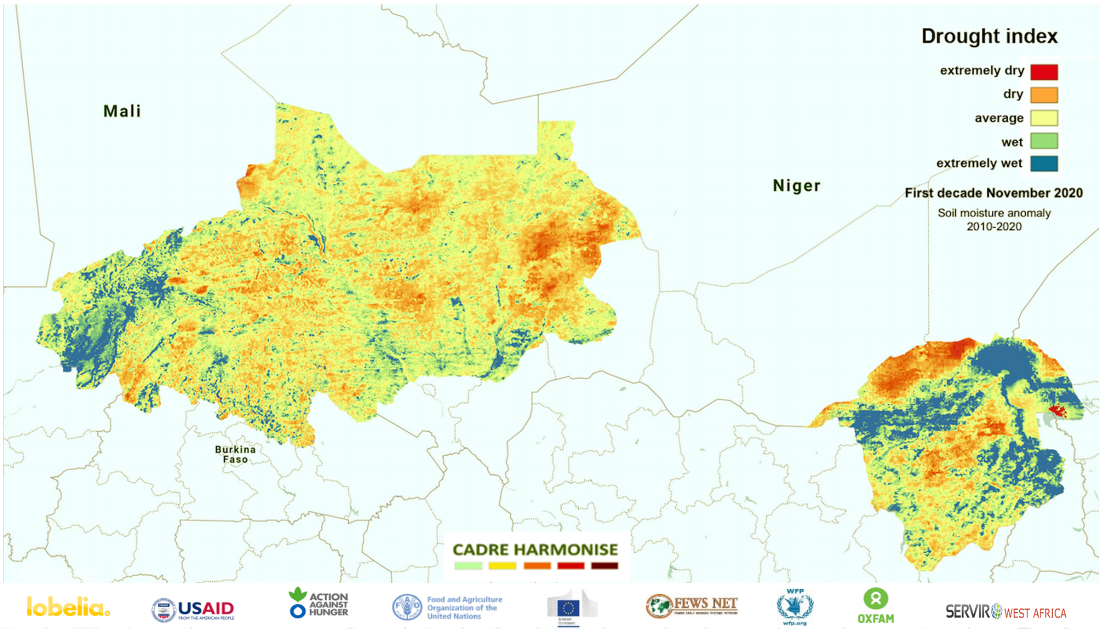


Figure 5 Lobelia provides drought monitoring operationally to the Cadre Harmonisé in West Africa every 10 days based on their daily drought product at 1km from soil moisture (Dispatch REF).

1.8 MECHANISM FOR COMMUNICATING/PROMOTING OUTPUTS OF THE PROPOSED ACTIVITY

Dissemination will include the publication of (at least) two scientific papers in international peer-review journals and attendance to major international conferences including:

- Phi Week 2021 and 2022
- Living Planet Symposium 2022

Oral presentations as well as workshops and dedicated sessions will be pursued at these events, with the idea of establishing a dynamic and open collaboration with the large network of scientists and stakeholders in the community.

It is in the interest of the consortium to communicate and promote the outputs of the project via Twitter and the partner's web sites, ESA News, and a use case publication in the form of a Story Map.

Communication will be centralised through the project web page. This webpage, on top of providing information about the objectives and state of development of the project, will be the main portal through which the dissemination and outreach activities will be published. The AI4DROUGHT web page will make use of interactive tools such as Twitter to receive feedback. It will also provide access to the AI4DROUGHT outputs.

.

1.9 TECHNICAL IMPLEMENTATION / PROGRAMME OF WORK

1.9.1 Proposed Work Logic

The rationale of the work logic kicks off with the complete and thorough formulation of the problem and a gap analysis that will establish the baseline for the study in **WP1000 State of the Art and**

Gap Analysis. The requirements baseline will be consolidated in **WP2000 Requirements Baseline** including the previous WP output, including details on the statement of work requirements as well as a full list of requirements for every component of the system. With these outputs, the system design and architecture will be established in **WP3000 Design and Development**, together with the set-up of a data lake, or centralised cloud repository containing the input data for the AI-based modelling approaches developed and implemented in this WP. WP3000 ends with the transfer of the output information as an end-to-end demonstrator scenario to a cloud repository. In **WP4000 Test and Validation**, a benchmark of the results against observational data will be conducted as a validation means, and the error and uncertainty measured in the cascading effects will be analysed also according to the demonstration scenario in WP4000. **WP500 Outreach and Communication** and **WP600 Management** will be active during the entire project.

It is important to point out that comprehensive main deliverables will be presented at the beginning of the project: the Scientific Report (SR) and the Technical Report (TR). Both reports will be living documents with contributions from the different tasks and work packages every quarter, which means that the output of tasks such as the architecture design, the test plan, and implementation details will be documented in the technical report as dedicated sections rather than separate deliverables. The Scientific Report will contain full documentation of the Demonstration Scenario, impact of the development and a scientific roadmap for the future. Other deliverables are also marked in the following diagram with a circle: the requirements baseline document will include the state of the art and the gap analysis outputs. This approach aims to reflect the Statement of Work indications and maintain the good development of the project.

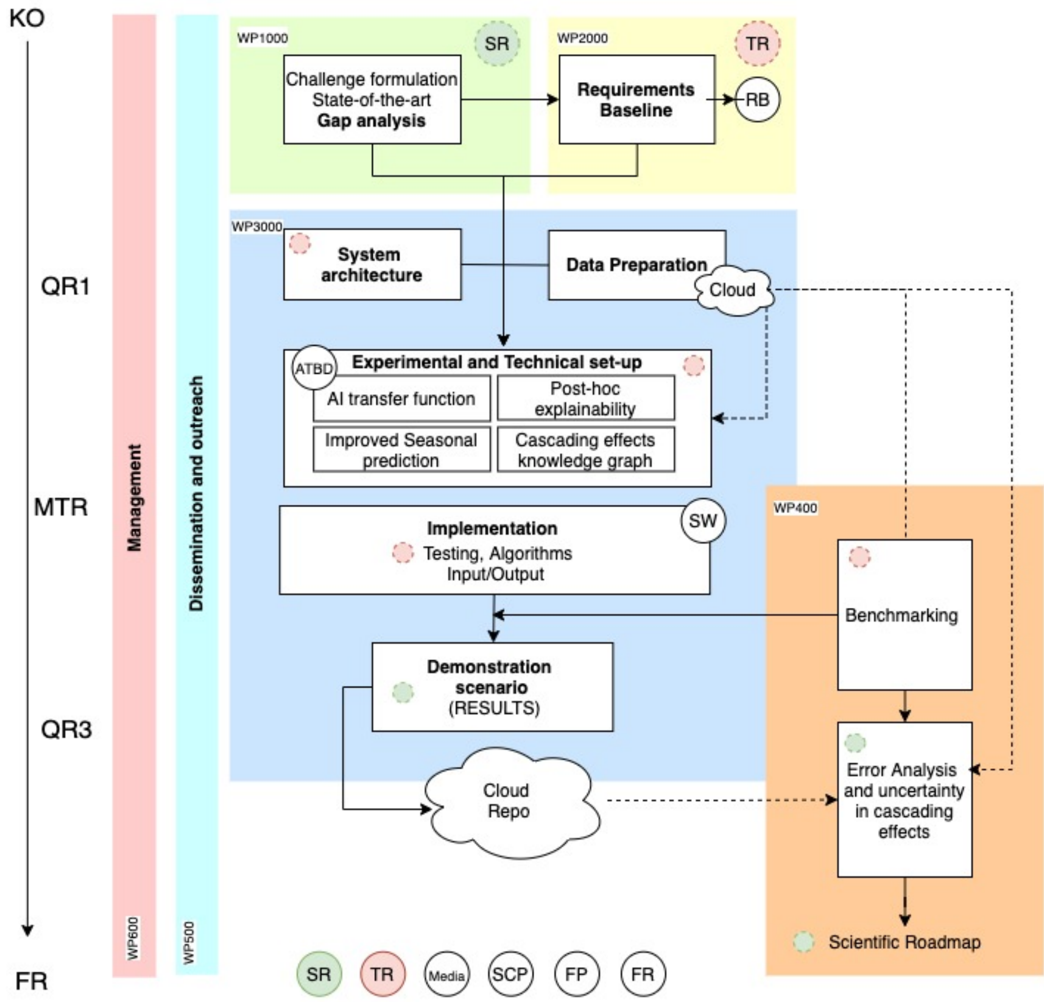


Figure 6 Proposed Work logic

1.9.2 Contents of the proposed work

1.9.2.1 Work Breakdown Structure (WBS)

Table 5 Work Breakdown Structure

WP and tasks	Title	Leader
WP100	Problem formulation, state of the art and Gap analysis	BSC
1.1	Problem formulation	BSC
1.2	State of the art and Gap analysis	LOBELIA
WP200	Requirements baseline	LOBELIA
WP300	Design and development	BSC
3.1	System architecture	EURECAT
3.2	Data preparation	LOBELIA
3.3	Methods	BSC

3.4	implementation and results	LOBELIA
WP400	Testing and validation	EURECAT
4.1	Benchmarking validation	BSC
4.2	Error analysis and uncertainty	EURECAT
WP5000	Dissemination and outreach	LOBELIA
5.1	Scientific publications	LOBELIA
5.2	Web Page	LOBELIA
5.3	Scientific dissemination	LOBELIA
WP6000	Management	LOBELIA

1.9.2.2 Work Package Description (WPD)

Project	AI4DROUGHT			WP:	1000
WP Title	Problem formulation, State-of-the art and scientific gap analysis			Sheet	1 of 1
Participants	Lead: BSC, Contributor: LOB, EUT			Issue:	1
WP Manager	Albert Soret			Date:	14/6/2021
Effort	610 hours				
Start Event	KO	Planned Date:	T0		
End Event	PR2+1	Planned Date:	T0 + 5		

Inputs:
- Statement of work
- Raw seasonal forecasts and associated hindcast from Climate Data Store (CDS)
- ERA5 reanalysis and E-OBS for ECV such as temperature and precipitation
- CCI EO datasets
- Bibliography, state-of-the art publications

Tasks:

Task 1.1. Challenge formulation: skill in seasonal forecasting can be improved with the integration of EO data (L: BSC; P: LOB) [M0-M3].

The aim of this task is two-fold. First, the formulation of a range of AI methods for blending seasonal climate predictions with EO, based on the available state-of-the-art scientific publications on the topic, to explore the frontiers of weather conditions affecting drought management for future months. Second, the establishment of a skill benchmark for seasonal climate prediction variables, based on current statistical methods, that will be used for comparison in WP4 (T4.1). Definition of the Scientific Report complete structure and fulfillment of the first section “Problem Formulation”

Task 1.2. Digital EO Services and gap analysis (L: LOB; P: BSC, EUT) [M0-M5].

This task aims to lay the foundation of the data-intensive analysis needed to investigate the relationship between the essential climate variables from EO (LOB), modeled observations (reanalyses) (LOB and BSC), seasonal forecast (BSC and LOB) and the identification and selection of relevant variables and indexes to categorise climate hazardous events at different scales and temporalities. Moreover, this task will also design the digital architecture to store and expose the predictions and AI driven model results (e.g. SOA, REST)
Outputs: - D1. Problem formulation (M2) BSC - D2. SoA and gap analysis (M5) LOB - SR - Science Report with (living document updated quarterly with input along the project)

Project	AI4DROUGHT			WP:	2000
WP Title	Requirements Baseline			Sheet	1 of 1
Participants	Lead: LOB, Contributor: BSC, EUT			Issue:	1
WP Manager	Guillermo Grau			Date:	14/6/2021
Effort	314 hours				
Start Event	PR1	Planned Date:	T0+2		
End Event	PR2+1	Planned Date:	T0 + 3		
Inputs: - Problem formulation - Architecture design and Requirements definition from proposal - State-of-the-art - Gap analysis					
Tasks: - Consolidation of scientific and technical requirements including infrastructure and their verification method, maintaining traceability to ESA AI4Science requirements - Definition of the technical REport complete structure and planned inputs along the project					
Outputs: - RB Requirements Baseline V1 - proposal, V2 - end of Task 2 - TR - Technical Report V1 (living document updated quarterly)					

Project	AI4DROUGHT			WP:	3000
WP Title	Design and Development			Sheet	1 of 1
Participants	Lead: BSC, Contributor: LOB, EUT			Issue:	1
WP Manager	Carlos Gómez			Date:	14/6/2021
Effort	3839 hours				
Start Event	PR2	Planned Date:	T0+4		
End Event	QR3	Planned Date:	T0+18		

Inputs:
- Raw seasonal forecasts and associated hindcast from Climate Data Store (CDS)
- ERA5 reanalysis and E-OBS for ECV such as temperature and precipitation
- CCI EO datasets
-D1. Problem formulation

Tasks:

Task 3.1. General System Design Architecture (ADD) (L: EUT; P: LOB, BSC) [M3-M8].

This task will define and implement the digital system architecture for putting together the seasonal climatic models outputs, AI transfer functions, Explainable AI, hazardous knowledge graphs and AI models for cascading effects. For that, this task will analyse the needs identified under WP1000 and the requirements defined under WP2000. As an output of the task, a design document will be established according to software development standards (UML).

Task 3.2. Data preparation (LOB; P: EUT, BSC) [M3-M12].

As defined in section 1.5.1.2, input data will come from different sources (CDS-C3S, CCI, etc.) and are stored in different formats, following different standards and access procedures. The data will be made accessible through a data lake centralised repository, quality controlled and formatted to be integrated in the AI4DROUGHT digital architecture established on Task 3. This task will provide an output to the AI4EU data store to drive the proposed innovations.

Task 3.3. Methods (ATBD) (L: BSC; P: EUT, LOB) [M4-M18].

This task will describe the algorithms to be implemented and evaluated in task 3.4.: AI Transfer Function (LOB), Explainable AI: causality, interpretability (EUT), Improved seasonal forecast (BSC) and cascading effects (EUT).

Task 3.4. Implementation and results (L: LOB; P: EUT, BSC) [M5-M18].

Implementation and execution of the algorithms in a self-cointained demonstration scenario will be uploaded to a cloud repository.

Outputs:
- D3. ADD (M8) EUT
- D4. Data repository (M12) LOB
- D5v1. ATBD (M12) EUT
- D5v2.ATBD update (M18) EUT
- D6. Software release (M18) LOB
- SR update (BSC)
- TR update (LOB)

<table>
<tr><td>Project</td><td colspan="3">AI4DROUGHT</td><td>WP:</td><td>4000</td></tr>
<tr><td>WP Title</td><td colspan="3">Testing and Validation</td><td>Sheet</td><td>1 of 1</td></tr>
<tr><td>Participants</td><td colspan="3">Lead: EUT, Contributor: BSC, LOB</td><td>Issue:</td><td>1</td></tr>
<tr><td>WP Manager</td><td colspan="3">Aitor Corchero</td><td>Date:</td><td>14/6/2021</td></tr>
<tr><td>Effort</td><td colspan="3">1823 hours</td><td></td><td></td></tr>
<tr><td>Start Event</td><td>MTR</td><td>Planned Date:</td><td>T0 + 12</td><td></td><td></td></tr>
<tr><td>End Event</td><td>FR</td><td>Planned Date:</td><td>T0 + 24</td><td></td><td></td></tr>
<tr><td colspan="6">Inputs:
- Improved Seasonal Forecast
- Explainable AI models
- AI transfer function
- Semantic Data Cube
- Risk Assessment and AI driven model</td></tr>
<tr><td colspan="6">Tasks:
Task 4.1. Comparison of improved forecast and observations (benchmark) (L: BSC; P: LOB; EUT)[M12-M18].
The aim of this task is two-fold. First, the obtention of the forecast quality assessment of the results obtained from seasonal forecasts in T3.4 by a set of probabilistic metrics that evaluate the performance of the predictions against reanalyses and observational gridded datasets. Second, the comparison of the skill obtained for temperature, precipitation and drought index in the previous step with the skill benchmark obtained in T1.1.
Task 4.2- Error Analysis and uncertainty in cascading effects (L: EUT; P: BSC; LOB)[M12-M24].
This task is mainly devoted to establishing a framework to analyse the errors caused by the hazardous events identification and cascading effect tool. To establish the evaluation framework, first action will be focused on the selection of relevant extreme events to be assessed. This selection will be followed up by the preparation of evaluation datasets simulating similar patterns related to such events. Second action will be aimed at elaborating a blinded cross-validation technique in order to analyse the sensitivity of the algorithm.</td></tr>
<tr><td colspan="6">Outputs:
- Comparison of improved forecasts and observations (BSC)
- Sensitivity analysis over the risk assessment and cascading effects tool (EUT)
- SR final version (BSC)
- TR final version (LOB)</td></tr>
</table>

<table>
<tr><td>Project</td><td colspan="3">AI4DROUGHT</td><td>WP:</td><td>5000</td></tr>
<tr><td>WP Title</td><td colspan="3">Scientific dissemination and outreach</td><td>Sheet</td><td>1 of 1</td></tr>
<tr><td>Participants</td><td colspan="3">Lead: LOB, Contributor: BSC, EUT</td><td>Issue:</td><td>1</td></tr>
<tr><td>WP Manager</td><td colspan="3">Maria Jose Escorihuela</td><td>Date:</td><td>14/6/2021</td></tr>
<tr><td>Effort</td><td colspan="3">439 hours</td><td></td><td></td></tr>
<tr><td>Start Event</td><td>KO</td><td>Planned Date:</td><td>T0</td><td></td><td></td></tr>
<tr><td>End Event</td><td>FR</td><td>Planned Date:</td><td>T0 + 24</td><td></td><td></td></tr>
</table>

Inputs: - RB - TR - SR
Tasks: - Design and execute an outreach plan to ensure the results of the project reach related communities - Create a project web page - Coordinate AI4DROUGHT participation in relevant conferences - Provide user-oriented training - Prepare and submit two scientific papers about the methodology and the validation of the development - Provide recommendations for new activities - Identify potential collaborations with the project - Document in detail the main achievements of the work carried in the study
Outputs: - Project Web Page - Training materials and presentations - Scientific Papers Publication

1.10 BACKGROUND OF THE BIDDERS

1.10.1 Lobelia Earth

Lobelia Earth is an SME based in Barcelona, specialised in satellite technology, computational intelligence and data visualisation for climate action. Lobelia Earth was created in November 2018 as the downstream-services unit of isardSAT Group, a consolidated science and technology SME specialising in the development of algorithms to process Earth Observation satellite data since 2006.

Lobelia recently spun-off becoming a linked third-party.

Lobelia works for the private and public sectors in climate services.

Lobelia develops applications and services for the exploitation of Earth Observation data. Droughts, floods, atmospheric pollution and climate are activity domains tackled by Lobelia, devoted to climate action (more info at lobelia.earth). Both geospatial data and infrastructure-based services are supported by the organisation, that counts with a product development team specialized in data management and visualization, with scientists, engineering developers, designers, and business developers working hand by hand to provide valuable products in a timely and reliable manner. Lobelia has extensive experience in project management and technical coordination (see isardSAT'S group projects with the involvement of the Lobelia team at https://www.isardsat.cat/en/projects/).

Lobelia has placed two products on the market: TeroMaps© technology, serving the visualisation of Copernicus and Sentinel data in the WEkEO DIAS cloud and CMEMS MyOcean Viewer, and the Lobelia Air© service, an operational air quality monitoring and forecasting prototype that has revolutionised air pollution action plans in Barcelona and Madrid, and was presented in COP25 officially in favour of sustainability and smart mobility strategies.

Lobelia is an active partner of the CCRI (Coalition for Climate Resilient Investment), a UN COP26 Flagship, where the company acts as a climate risk data provider to several end-to-end pilots addressing drought and extreme precipitation.

Lobelia is a pioneering company that develops insightful platforms grounded on scientific knowledge and radically new software technologies, and paves the way towards a zero- carbon economy in the mass economy.

Selected relevant projects:

Digital Twin of the Ocean (ESA, 2020-2021): The project aims to develop a prototype of an instance of a Digital Twin Earth applied from Digital Twin Ocean precursor demonstrating the capability to visualize, monitor and forecast natural and human activity on the planet and proposes an end-2-end architecture for the Digital Twin Earth – Ocean precursor. Lobelia Earth is responsible for the development of the Globe Story Engine, a rotating, 3D digital globe that serves as a leitmotiv for stories on Earth sciences. The web-based visualization tool will allow the creation of interactive scenarios including the Ocean as their main character.

WEkEO Data Discovery Platform (Mercator Océan and EUMETSAT, 2020): Project to develop and integrate the WEkEO Data Discovery Platform (DDP) within the WEkEO website and encompass the catalogue, viewer and subsetter functionalities. It consists of two independent but closely interrelated parts, the Back End and the Front End. It displays over 3700 layers of data, from Copernicus marine, Climate and Atmosphere services, and Sentinel 1, 2, 3 and 5P. Data can be discovered in time and space, downloaded and accessed through the Harmonised Data Access of the WEkEO service (https://wekeo.eu/)

SnapEarth. H2020-SPACE-2019
SnapEarth project is to foster the Market growth of COPERNICUS by instigating the development of new EO applications and to develop general public awareness to EO data. SnapEarth is to initiate the creation of a virtuous circle of innovation by providing to EO data users an innovative platform with leading edge EO segmented datasets, Neural Networks models and Cloud computing ecosystem. Lobelia and isardSAT are responsible for the Food security pilot development based on agricultural drought.

Thermal Stress in the Climate Crisis (ECMWF 2020-2021).
The goal of this Project is to exploit the wealth of data available in ECMWF's Climate Data Store (CDS) to communicate how thermal stress is evolving under climate change conditions. The Project proposes a Story Hub for the communication of insights related to thermal stress/comfort, relevant to different fields (urbanism, energy, health, leisure, migration), aiming at wide swaths of the population, and creating fact-based and engaging experiences.

SEEDS "Sentinel EO-Based Emission and Deposition Service" (H2020-SPACE-2020)
The main objective of SEEDS is to develop a proof-of-concept for an add-on service to CAMS on pollutant emissions and depositions that will enhance the use of satellite observations and provide new products to boost European competitiveness and sustainability actions across a diversity of economic sectors while contributing to the protection of European citizens and ecosystems.

MIREIA "Marine Litter Signatures in Synthetic Aperture Radar Images" (ESA, 2020-2021). Lobelia at isardSAT coordinates project MIREIA, is a research project aiming at demonstrating the potential of Synthetic Aperture Radar (SAR) images for marine litter detection in combination with AI techniques. The proposed solution will be tested in the Balearic Islands, a hotspot area for plastic marine debris accumulation. The team at Lobelia designs the overall system, including the characterization and exploitation of different information gathered by radar sensors, large in-situ datasets and pattern recognition models, to identify marine debris.

OceanScan (ESA ESTEC, 2020-2021) Satellite remote sensing has demonstrated great potential to become a breakthrough in the mapping of marine litter. One limiting factor for its full development is the access to reliable, extensive, and consistent ground truth of plastic occurrence in the ocean. Some of the best performing technologies for image analysis today were built using open labeled databases. In OceanScan, the team is responsible for the creation of an inclusive labeled global ocean plastic database and platform. Lobelia at isardSAT is the project coordinator and is developing a platform to explore and access matching ground truth to satellite experiments, with selectable baseline options in time, space and sensor type, and an app to input geolocalized and timestamped information, with options for pictures to facilitate the data labelling (https://www.oceanscan.org/).

TEP Hydrology, ESA, 2014-2021, Lobelia at isardAST coordinates the Hydrology Thematic Exploitation Platform and is the provider of the operational water level service based on altimetry data, with enhanced functionalities due to the integration with a hydrological model and a flood monitoring service with SAR. Through this workspace platform, isardSAT is in contact with a large user community, that is providing feedback on the service, and also offering offers support to those potential users wishing to upload their processing chains to the TEP cloud, and as well as ensures the successful execution of the water level service by pilot users.

MyOcean Viewer (Mércator Océan, 2020): Development and operation of a new central tool for the Copernicus Marine service. MyOcean is a fast, intuitive, detailed, and accurate ocean viewer aimed to be used by all communities in various sectors, from citizens to students, policymakers, data scientists, business companies and start-ups. From temperature, currents, and waves to acidity and plankton, convenient ocean maps become readily available. It is entirely developed by Lobelia Earth and has already gained international reputation, it is free-of-use and accessed by thousands of users worldwide (https://cmems.lobelia.earth/)

AirQast, H2020, 2017-2020, is a commercial platform providing operational Air Quality services using Earth Observation data. These services provide updated emissions inventories, advanced forecasting systems and decision making tools to manage air quality events in order to reduce their economic and social impact. The platform covers pollutants having a higher economic and social impact such as PM2.5, PM10, SO2, NO2 and CO. Lobelia at isardSAT is the coordinator and provider of the API and tools.

C3S_SMHI_Lot1, ECMWF, C3S (Copernicus Climate Change Service), 2017-2019. The GLORIOUS-project brings together excellence in climate sciences, impacts modelling, and technical expertise to develop a new operational service together with a global user community. The aim is to ensure user uptake of relevant high-impact climate information from the C3S Climate Data Store, addressing sectors such as food security, natural hazards, health, transport and biodiversity. isardSAT is the climate change knowledge purveyor and provides extreme events analysis with Artificial Intelligence and Machine Learning for global users (UN-HABITAT, PwC and OXFAM) within Copernicus services to ECMWF as well as its application to downscaling and bias correction methodology to generate high-resolution, locally-adapted data describing the future climate in the target cities.

EO Clinic: Rapid-Response EO-Based Solutions to Development Aid Project Requests, European Space Agency (2019-2020). The objective is to produce EO products and solutions that respond to the specified user requirements, but that can be more innovative and exploratory in nature providing information that is further down the value chain, and closer to decision-making levels. Such solutions may be experimenting with innovative methodologies, combining EO-derived information with other, non-EO data sources related to socioeconomic and environmental parameters and indicators. Lobelia supports two out of the ten thematic groups within the EO Clinic Framework, namely climate change and water resources management.

1.10.2BSC

The Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC), created in 2005, has the mission to research, develop and manage information technology to facilitate scientific progress. At BSC, more than 700 people from 50 different countries perform and facilitate research in Computer Sciences, Life Sciences, Earth Sciences, and Computational Applications in Science and Engineering. The BSC is one of the four hosting members of the European PRACE Research Infrastructure. The Center houses MareNostrum, one of the most powerful supercomputers in Europe, and has been one of the top eight Spanish centres awarded the 'Severo Ochoa Centre of Excellence' Accreditation by the Spanish Government.

The department involved in this proposal is the Earth Sciences Department of the BSC (ES-BSC) which was established to carry out research on Earth system modelling. The department focuses on multiscale (global to urban) air quality and meteorological modelling, global and regional mineral dust modelling, and global and regional climate modelling. Over the years, the department has been active in numerous European projects, including in FP7 and H2020, not only as a partner, but also as coordinator of an H2020 project (S2S4E), a FP7 (SPECS) and a COST action (inDust). It is also currently involved in several ESA services (e.g. ESA/RFQ/3-15131/18/I-SBo CMUG-3-TECHPROP, ESA AO/1-10546/20/I-NB, ESA AO/1-10548/20/I-NB, D/565/67238959), Copernicus Atmospheric Monitoring Services (e.g. CAMS_43, CAMS_50, CAMS_61 and CAMS_84) as well as in several Copernicus Climate Change Services (i.e. C3S_512, C3S_429g and C3S_34c), two of them coordinated by BSC.

ES-BSC is structured around four groups: Climate Prediction, Atmospheric Composition, Computational Earth Sciences, and Earth System Services, with more than 100 employees, including researchers and technical staff. The Earth System Services group (ESS), the Climate Prediction (CP) group and Computational Earth Sciences (CES) group are the three groups that will contribute to AI4DROUGHT.

The ESS group is specialised in the co-creation of climate and air quality services with a transdisciplinary approach where earth system modelling scientists and knowledge transfer experts engage with stakeholders to create user-driven services.

The CP group aims at developing climate prediction capabilities on seasonal, decadal and multi-decadal time scales. The CP group has a long experience in seasonal to decadal climate prediction, and in understanding changes in climate extremes and other impact-relevant climate events.

The CES group is a multidisciplinary team with different IT profiles that interact closely with all the other groups of the Earth Sciences Department. The group provides help and guidance to the scientists with the technical issues relating to their work and develops a framework for the most efficient use of HPC resources. In order to improve the use of the variety of computing resources

available at the BSC and in other HPC institutions, a solid software development, profiling and optimisation area has been created for Earth system model codes towards exascale computing, and to provide feedback on this to modellers around Europe.
Last but not least, the development of a framework to disseminate the outputs generated by the BSC among the research and service community has been established. This area takes advantage of the unique environment of the BSC where research in BigData and Artificial Intelligence is already a priority that will be extended in the next few years.

Selected relevant projects:

Coordination of the H2020 project S2S4E: Sub-seasonal to seasonal predictions for the energy sector (S2S4E-776787), which offers an innovative service to improve the management of renewable energy variability. The main result of the project is a Decision Support Tool (www.s2s4e.eu/dst), an operational climate service to enable renewable energy producers and providers, electricity network managers and policy makers to design better-informed strategies at sub-seasonal to seasonal timescales.

Coordination of the C3S-512 contract (ECMWF): The Evaluation and Quality Control (EQC) function of the Copernicus Climate Change Service (C3S) has a critical role to ensure that the service meets the needs of a range of users for high-quality data and information, and in proposing the necessary evolution of the service itself, while shaping the research agenda to attend the most important challenges detected. The contract aims at developing a solution for the EQC function to respond to the needs identified in previous contracts using a continuous user-engagement process.

ESA CMUG-CCI+ (Climate Model User Group-Climate Change Initiative+, 2018-2022): An ESA project specifically conceived to ensure that the ESA-CCI data products are developed and provided in a form most useful for climate analysis and modeling work and that they are widely promoted within the climate research community, with a special emphasis on the activities related to the Climate Model Intercomparison Project. The ES-BSC department is experienced in the use of satellite observations for climate modeling and forecasting purposes.

H2020 project ESiWACE2 - Centre of Excellence in Simulation of Weather and Climate in Europe (ESiWACE2-823988). It will link, organise and enhance Europe's excellence in weather and climate modelling to (1) enable leading Europe a weather and climate models to leverage the performance of pre-exascale systems with regard to both compute and data capacity as soon as possible and (2) prepare the weather and climate community to be able to make use of exascale systems when they become available.

H2020 project EUCP: The overarching objective of the European Climate Prediction (EUCP-77661) system is to develop an innovative European regional ensemble climate prediction system based on a new generation of improved and typically higher-resolution climate models, covering timescales from seasons to decades initialised with observations, and designed to support practical and strategic climate adaptation and mitigation decision-taking on local, national and global scales.

H2020 project PRIMAVERA: PRocess-based climate sIMulation: AdVances in high resolution modelling and European Climate Risk Assessment (PRIMAVERA-641727). The main objective is to develop a new generation of advanced and well-evaluated high-resolution global climate models capable of simulating and predicting regional climate with unprecedented fidelity, for the benefit of governments, business and society in general. PRIMAVERA developed and applied various participatory approaches, including the User Interface Platform, and tested the latest scientific knowledge in case studies co-developed with key sectoral users.

H2020 project MED-GOLD: Turning climate-related information into added value for traditional MEDiterranean Grape, OLive and Durum wheat food systems (MED-GOLD-776467), will develop novel pilot climate services focusing on three staples of the Mediterranean food system and water management.

1.10.3EURECAT

Eurecat is the leading private Research and Technology Organization (RTO) in Catalonia (Spain) and the second largest private RTO in Southern Europe. In 2020 figures, Eurecat turnover was 50M€ and staff reached 650 professionals. Eurecat is a multi technology / multisector RTO with four major technological divisions devoted to industrial, digital, sustainability and biotech and works for both private and public markets.. Currently, Eurecat is involved in more than 200 R&D projects and has a customer portfolio of over 1.600 business clients, and participating in more than 60 EU funded collaborative projects, mainly in the Horizon 2020 Program. In addition to this wide experience at European level, Eurecat is also a strong player in the various R&D programs sponsored by the Spanish administration (mainly with CDTI).. Technology transfer is also an essential activity in Eurecat, with 36 international patents and 7 technology-based companies started-up from the center's R&D activities.

Eurecat's Digital Technologies Division features 100 professionals with 20% holding a PhD degree. Data Science, Big Data technologies and Artificial Intelligence are amongst the main specialties developed in this Division, applied in multiple domains like Agriculture, Energy and Resources, Industry (I4.0), Digital Health and others. Eurecat also leads the Center for Innovation in Data Technologies and Artificial Intelligence (CIDAI, www.cidai.eu) which is a public-private association to promote and spread innovation based on AI and data technologies. BSC, one of AI4DROUGHT's partners, is also a CIDAI member.

The Project tasks will be developed by the following two units within the Digital Technologies Division:

- **Applied Artificial Intelligence (AAI)**: this unit holds a sound and proven background on researching and tailoring advanced solutions based on Machine Learning and Artificial Intelligence for the different value chains and nexus components inside the water, environment and blue economy sectors. AAI has proven experience in data value chain management (collection-information-knowledge-intelligence) by means of research, design and development of solutions (algorithms, methodologies, modules, mobile apps, platforms) based on the combination of different technologies such as AI, machine learning, data analytics, optimization, and information and knowledge management.

- **Big Data & Data Science (BD&DS)**: The unit has extensive experience in the application of Data Science and Artificial Intelligence techniques both for business applications and in the field of computational social science. Its most remarkable activity is focused on i) Natural language processing for text interpretation and classification of intends in the development of virtual assistants and ii) AI driven models of sales acceleration and the analysis and prediction of mobility models in both human field and transport areas.

Among its research areas, the most recent stand out is the development of **algorithms without bias and the explainability of black box AI models**. In addition, the unit holds specific knowledge in Big Data end-to-end process and architecture design ranging from data management and databases, distributed and parallelization systems, to real time and service orchestration.

Selected relevant projects:

SIM4NEXUS (EU H2020). Development of innovative methodologies to address these barriers, by building on well-known and scientifically established existing "thematic" models, simulating different components/"themes" of the Nexus and by developing: (a) novel complexity science methodologies and approaches for integrating the outputs of the thematic models; (b) a Geoplatform for seamless integration of public domain data and metadata for decision and policy making; (c) a Knowledge Elicitation Engine for integrating strategies at different spatial and temporal scales with top down and bottom up learning process, discovering new and emergent knowledge, in the form of unknown relations between the Nexus components and policies/strategies; (d) a web-based Serious Game for multiple users, as an enhanced interactive visualisation tool, providing an immersive experience to decision- and policy-makers.

FIWARE4WATER (EU H2020). The project intends to link the water sector to FIWARE (European open source framework for smart applications) by demonstrating its capabilities and the potential of its interoperable and standardised interfaces for both water sector end-users (cities, water utilities, water authorities, citizens and consumers), and solution providers (private utilities, SMEs, developers).

VITIGEOS (EU H2020): this project uses satellite imagery (Copernicus) and on-field sensor data to increase resolution and reliability of satellite information applied to all aspects of viticulture and specific wine-business operations.

STOP-IT (EU H2020). This project focuses on developing new technologies for protection of critical infrastructures- STOP-IT solutions are based on: a) mature technologies improved via their combination and embedment (incl. public warning systems, smart locks) and b) novel technologies whose TRL will be increased (incl. cyber threat incident services, secure wireless sensor communications modules, context-aware anomaly detection technologies; fault-tolerant control strategies for SCADA integrated sensors, high-volume real-time sensor data protection via blockchain schemes; authorization engines; irregular human detection using new computer vision methods and WiFi and efficient water contamination detection algorithms).

PATHOCERT (EU H2020). The project aims at strengthening the coordination capability of first responders in emergency events were they have to act in places with a high risk of contamination through water. Within this scope, the project develops pathogen contamination emergency response technologies, tools and guidelines to be validated by first responders, helping them to detect pathogens quickly and to better control emergency situations.

SCOREWATER (EU H2020). The project SCOREwater develops and tests three large-scale demonstration pilots cases for collecticollecting, computing and presenting various data generated in the city's sewage system. The Barcelona's pilot, currently in progress, is focused on a new domain referred as “sewage sociology”, whose purpose is to derive from these sewage data meaningful biomarkers of community-wide lifestyle habits. from sewage.

PROCEED (commissioned by Spanish CDTI). PROCEED’s objective is to implement a data-driven model to forecast SARS-COVID-2 outbreaks. The model is based on clinical data augmented with mobility, environmental, sewage and social media data related to the pandemics.

SATELIOT IoT and IA Platform: project commissioned by SATELIOT, a New Spacea company that is building a constellation to provide IoT and 5G services. Eurecat is developing a platform featuring BD, ML and DL modules to support EO use cases analysis.

1.11 RESOURCES

1.11.1 Lobelia Earth

Lobelia is located at the Barcelona Technology Research Park, with access to ultra-wide Internet access, conference facilities, and meeting rooms. Using the resources shared by the companies located at the park, Lobelia has immediate access to video conference rooms, an auditorium of 500 m2 hosting up to 240 people equipped with projector, video and audio recording, 8 spaces for press and possibility of simultaneous translation system, etc.

Lobelia owns two dedicated data processing multi-core Linux workstations, custom-designed to solve the computing needs of EO projects and web services. It also owns a 28-TB NAS server, as well as the software and licences needed to perform its research activities. A high-speed internal network provides access to services such as name resolution, user database, disk storage, application server, code versioning, dependency management, documentation management tools, system backup. VPN is available for remote access to network resources by employees. The company provides resources on commercial cloud platform providers for publicly available services. Its also owns personal workstations running macOS, Linux and Windows operating systems, several printers and photocopiers, fax machines, and several external backup hard drives.

Lobelia also owns an in-house solution for unified data access named TeroData, which allows querying and downloading heterogeneous datasets, both for prototyping and large-scale processing, with bindings for Python, JavaScript and MATLAB, TeroMaps as an integrated fast rendering and maping engine –including 3D layouts, and TeroBot for the monitoring of systems.

1.11.2 BSC

BSC is the National Supercomputing Facility of Spain and hosts a range of HPC systems including MareNostrum IV. The new supercomputer will be 12.4 times more powerful than the current MareNostrum 3 that will have a performance capacity of 13, 7 Petaflop/s. The general purpose element will have 48 racks with more than 3,400 nodes with next generation Intel Xeon processors and a central memory of 390 Terabytes. The second element of MareNostrum 4 will be formed of clusters of three different technologies that will be added and updated as they become available. These are technologies currently being developed in the US and Japan to accelerate the arrival of the new generation of pre-exascale supercomputers. The BSC is a key element of and coordinates the Spanish Supercomputing Network, which is the main framework for granting competitive HPC time to Spanish research institutions. Furthermore, BSC is one of six hosting nodes in France, Germany, Italy and Spain that form the core of the PRACE network. PRACE provides competitive computing time on world-class supercomputers to researchers in the 25 European member countries.

1.11.3 Eurecat

Eurecat holds 11 operative branches spread throughout the Catalan region. The BD&DS unit is located at Eurecat's Headquarters in Barcelona and the AAI unit is located at the PCITAL, the agrotech cientific park in the city of Lleida.

Eurecat, being a multidisciplinary RTO, holds several laboratory facilities, (e.g. plastronics pilot plant, robotics and drone lab, EV battery lab, cybersecurity lab, blockchain infrastructure, ect.). Specifically for the purpose of this project, Eurecat has an **Inference lab** used for R&D projects requiring a high computational demand (deep learning, reinforcement learning, ect.). The lab consists of a Service Platform with automatic Big Data tools provisioning, able to deliver many different Big Data technologies for storing, processing, and analyzing large volumes of data. This platform, internally named as DATURA, consists of 528 CPU and 40.960 GPU cores, respectively. with 5.5 TB RAM, and 4 PB storage distributed in 8 computational nodes plus a RedHat Open Stack for resource virtualization.

2 MANAGEMENT PROPOSAL

2.1 TEAM ORGANISATION AND PERSONNEL

2.1.1 Proposed team

The Consortium presented in the framework of this project groups together a super computing research infrastructure and a research centre (BSC), a research and technology organization (EURECAT) and a company (Lobelia Earth). The team is highly qualified and includes scientific expertise on climate modelling, artificial intelligence techniques and state-of-art in Earth Observation services

Lobelia Earth (LOB) is the prime contractor and the main interface with ESA at the contractual and project management.

2.1.1.1 Overall team composition, key personnel

Table 6 Key personnel

Key Personnel Name	Institution/ Company	Role	Dedication
Laia Romero	LOB	Project Manager	30%
Maria José Escorihuela	LOB	WP5000 Leader Earth observation for hydrology and agriculture Senior Researcher	36%
Jesús Peña Izquierdo	LOB	Task 1.2 Leader AI and climate expert Senior researcher	40%
Melissande Machefer	LOB	Task 3.2 Leader AI and EO expert Applied researcher	60%
Guillermo Grau	LOB	WP2000 Leader Task 3.4 Leader Senior Telecommunications and SW engineer	50%
Albert Soret	BSC	WP1000 Leader Task 1.1 Leader Established researcher	15%
Carlos Gómez	BSC	WP3000 Leader Task 3.3 Leader Recognised researcher	15%
Markus Donat	BSC	Leading researcher	7,5%
Nube González	BSC	Recognised researcher	7,5%
Llorenç Lledó	BSC	Recognised researcher	7,5%
Lluís Palma	BSC	Junior research engineer	50%
Lali Soler	EUT	Senior engineer eXplainable AI expert	1,2%
Xavier Domingo	EUT	Leading engineer	1,2%
Aitor Corchero	EUT	WP4000 Leader Task 4.2 Leader	50%

		Senior engineer	

Project main actors and their relationships:

- Project Coordinator: She will lead the project and will be responsible for the technical implementation of the project. She will also be the main line of communication with ESA regarding all technical, administrative, financial and contractual issues.
- WP leaders: The Project and Scientific Coordinators are also responsible for the coordination of technical and scientific activities at the work package level.
- Task leaders: They will be responsible for the execution of key tasks assigned according to the work packages description.
- Coordination Support Team: They will assist the Project on administrative, contractual and financial tasks.

2.2 CURRICULA VITAE

2.2.1 Lobelia

Name:	Laia Romero		
Role:	Project Manager	Dedication	30%
Qualifications and experience of interest for this proposal:			
Laia Romero is isardSAT Group Director of Operations and Strategy, and Managing Director of Lobelia Earth. MSc in Physical Oceanography from the Polytechnic University of Catalonia (UPC), she is responsible for systems development and operational services. Over the last 15 years she has worked extensively in Earth Observation with radar technologies, in the development and procurement of geoinformation systems, deployment and execution of services. Before joining isardSAT Group, Laia was Director of New Business and Innovation at Altamira Information (CLS Group), where she managed the division of R&D and Data Management Solutions. Prior to that, she held the position of Information Systems Manager, in which she managed the development life cycle of operational contracts such as the Copernicus Marine Service (CMEMS) and commercial international contracts involving the development, integration, verification, and operations of large information systems, in tight collaboration with dedicated technical teams. Laia has solid experience in project management and technical coordination in numerous ESA contracts, bringing together large research entities and companies in tight collaboration. She is a member of the International Ocean-Colour Coordinating Group (IOCCG) on Remote Sensing for Marine Litter and is co-charing a special issue on this topic, for which she organised a successful phi week event in 2020. She has vast experience with Copernicus data and Information, and cloud infrastructure services, such as the TEPs and the WEkEO DIAS. She conducts technical project management and coordination of strategic projects for the development of new lines of work.			
Relevant publications:			
● Savastano, S; Cester, I. and Romero, L. A first approach to the automatic detection of marine litter in SAR images using artificial intelligence. IGARSS 2021			

- Martínez-Vicente, V.; Clark, J.R.; Corradi, P.; Aliani, S.; Arias, M.; Bochow, M.; Bonnery, G.; Cole, M.; Cózar, A.; Donnelly, R.; Echevarría, F.; Galgani, F.; Garaba, S.P.; Goddijn-Murphy, L.; Lebreton, L.; Leslie, H.A.; Lindeque, P.K.; Maximenko, N.; Martin-Lauzer, F.-R.; Moller, D.; Murphy, P.; Palombi, L.; Raimondi, V.; Reisser, J.; Romero, L.; Simis, S.G.H.; Sterckx, S.; Thompson, R.C.; Topouzelis, K.N.; van Sebille, E.; Veiga, J.M.; Vethaak, A.D. Measuring Marine Plastic Debris from Space: Initial Assessment of Observation Requirements. Remote Sens. 2019, 11, 2443. https://doi.org/10.3390/rs11202443
- Romero, L. "Web application for fast visualization and advanced analysis of millions of entries from remote sensing data" ESA Living Planet Symposium, Prague, 2016
- G. Griffiths, J. Blower, A. López, I. Polo, L. Romero, T. Loubrieu, S. Brégent, "Recent innovations in using Web Map Services to display gridded and non-gridded ocean data", EGU General Assembly 2014, Geophysical Research Abstracts, Vol. 16, EGU2014-15086, 2014
- Romero, Laia & Motte, Erwan & Egido, Alejandro & Reppucci, Antonio & Fernandez, Bonifacio & Castro, Lina & Caparrini, Marco. (2012). Streamflow prediction based on satellite and in situ measurements for hydro studies in central Chile. International Journal on Hydropower and Dams. 19. 54.

Name:	Maria Jose Escorihuela		
Role:	WP5000 Scientific dissemination and Outreach leader Earth Observation Senior Scientist specialised on hydrology and agriculture	Dedication	36%

Qualifications and experience of interest for this proposal:

Maria José received an Engineering degree in electronics and telecommunications from the Universitat Politècnica de Catalunya (UPC) in 1996 and the Ph. D. degree in Environmental, Space and Universe Sciences from the 'Institut National Polytechnique' in Toulouse (France) in 2006.

Maria José joined isardSAT in January 2008. Her responsibilities include technical lead of several R&D projects with a budget of over a million euros, development and validation of models and algorithms to retrieve geophysical variables from satellite data. Her scientific fields of interest are the application of passive and active microwave remote sensing to hydrology and climate change studies. She is appointed reviewer for several peer-review journals.

Relevant publications:

- Escorihuela *et al.*, SMOS based high resolution soil moisture estimates for desert locust preventive management, Remote Sensing Applications: Society and Environment, https://doi.org/10.1016/j.rsase.2018.06.002
- Piou, C., Gay, P. E., Benahi, A. S., Babah Ebbe, M. A. O., Chihrane, J., Ghaout, S. & Escorihuela, M. J. (2019). Soil moisture from remote sensing to forecast desert locust presence. Journal of Applied Ecology, 56(4), 966-975.
- Escorihuela and Quintana-Seguí, Comparison of remote sensing and simulated soil moisture datasets in Mediterranean landscapes, Remote Sensing of Environment, DOI: 10.1016/j.rse.2016.02.046.
- Escorihuela *et al.*, Effective soil moisture sampling depth of L-band radiometry: A case study, Remote Sensing of Environment https://doi.org/10.1016/j.rse.2009.12.011

- Escorihuela *et al.*, A Simple Model of the Bare Soil Microwave Emission at L-Band, IEEE Transactions on Geoscience and Remote Sensing, DOI: 10.1109/TGRS.2007.894935.

Name:	Guillermo Grau		
Role:	Tech Lead and responsible for Task 3.2 Data preparation and data lake set-up	Dedication	50%

Qualifications and experience of interest for this proposal:

Guillermo Grau is Co-founder and Technical Lead at Lobelia. He obtained his MSc. in Electrical Engineering from the Polytechnic University of Valencia (UPV), after presenting his Master Thesis in the Leibniz Universität Hannover. He has extensive experience in various branches of the aerospace sector including navigation, communications and Earth Observation. His main expertise is in systems engineering, having participated in multiple projects for ESA, EUMETSAT and the European Commission during his stint for Indra Espacio (2001-2017) and later at Lobelia (2017-). His hands-on approach has gained him experience in multiple fields, such as digital signal processing, leading the simulation and real-time implementation of the Galileo Search and Rescue processor; remote-sensing data management and visualisation, co-leading the Tero line of products and tools; database design and management, shipping tools that are used daily in the healthcare sector. As a software engineer, Guillermo enjoys developing complex and performant platforms with flexible APIs and highly-usable front ends. He has recently created the CMEMS MyOcean Viewer, where he onboarded early feedback from the Client as well as from end users as part of the continuous improvement practice that he has always pursued. Guillermo has also created open-source software libraries and tools with more than 1 million downloads/week, and has contributed to many other open-source projects.

Relevant publications:

- CMEMS MyOcean Viewer
- WEkEO Data Discovery Platform
- Thermal Stress Story Hub
- Lobelia Air

Name:	Melissande Machefer		
Role:	EO Applied Research Scientist	Dedication	60%

Qualifications and experience of interest for this proposal:

Melissande Machefer is a double degree M.Sc. in Earth and Space Observation and Mathematical Modelling with Engineering Management. As an Earth Observation (EO) Applied Research Scientist, she is in charge of processing remote sensing imagery and ancillary environmental data and developing algorithms to extract intelligence from these data. She graduated in 2018 and has worked on several projects allying EO and Data Science for different companies (Thales Alenia Space, DHI Gras, Hummingbird Technologies). Among these, she had the opportunity to work on super resolution and semantic segmentation from optical satellite imageries and several algorithms for precision agriculture products (canopy coverage, plant counting, soil type detection, weed detection) using drone and Sentinel 2 imagery. She joined isardSAT/Lobelia in 2019 and is now in charge of scientific data science supervision and algorithms development for applications such as water level, water bodies identification, soil moisture, coastal erosion, plastic detection based on EO data.

Relevant publications:
• Machefer, M.; Lemarchand, F.; Bonnefond, V.; Hitchins, A.; Sidiropoulos, P. Mask R-CNN Refitting Strategy for Plant Counting and Sizing in UAV Imagery. Remote Sens. 2020, 12, 3015. https://doi.org/10.3390/rs12183015 • Machefer, M.; Escorihuela , M.J.; Romero, L. High resolution soil moisture retrieval from multi spatio-temporal scales Earth Observation data-driven models – 2020 Phi Week. https://www.youtube.com/watch?v=QLqDGt3mh0c • Machefer, M.; Pattle, M.; Garcia-Mondejar, A;. Romero L.; Beck A.L.; Hennen M. Large-scale coastal erosion monitoring from SAR imagery over a 25-year time span – 2020 Phi Week. https://www.youtube.com/watch?v=LS5HP7BRbuQ • Machefer, M.; Brun A.; Romero L. Situated knowledge and climate services: miscellaneous scales and levels of interpretation, using physical observations and data science – Virtual Workshop at ETH Zurich Data Science in Climate and Climate Impact Research 2020

Name:	Jesús Peña-Izquierdo		
Role:	Senior Climate Data Scientist	Dedication	40%
Qualifications and experience of interest for this proposal:			
Jesús Peña-Izquierdo is a Climate Data Scientist. Graduated in Physics, MSc and PhD in Physical Oceanography, he has 10 years of experience in climate science research having worked in 4 different international institutions (ICM, SCRIPPS-SIO, UNSW-CCRC, BSC), co-authored 10 peer-reviewed papers and 3 book chapters. He has extensive expertise working with both in-situ and numerical models data. During his early years as researcher he participated in 4 oceanographic cruises, being the physical oceanographer leader of one of them, collecting, processing and analysing different sources of experimental data. In 2015 he published a climatology of the Tropical Atlantic with more than 40,000 historical observations after developing exhaustive quality controls and statistical analysis. He has combined observational data with numerical models, specially Lagrangian simulations where the trajectories of millions of virtual floats are computed and analysed to estimate ocean circulation patterns and their corresponding transport of heat and nutrients. All this expertise working with different types of data led him to work within one of the first international attempts of establishing a standardized quality control framework for the massive and heterogeneous collection of datasets included in the Copernicus Climate Data Store. He was the leader of the evaluation team of the reanalysis datasets (which includes popular products such as ERA5 or UERRA). He has been progressively interested in applied science, so he firmly decided to extend his data analysis skills with several training in Machine Learning which he successfully applied in weather nowcasting and downscaling of seasonal forecasts. With this spirit, he joined the Lobelia Earth team in 2020 aiming to apply his scientific experience to solve real-world problems transforming data into actionable insights.			
Relevant publications			
• Orúe-Echevarría, D., Pelegrí, JL., Alonso-González, I., Benítez-Barrios, VM., De La Fuentea, P., Emelianova, M., Gassera, M., Herrero, C., Isern-Fontaneta, J., Peña-Izquierdo, J., Ramírez- Garrido, S., Rosell-Fieschi, M., Salvador, J., Saraceno, M., Valla, D., Vidale, M., (2019), Dataset on the TIC-MOC cruise onboard the R/V Hespérides, March 2015, Brazil-Malvinas Confluence. Data in Brief, 22, 185-194. doi.org/10.1016/j.dib.2018.12.004			

- Llanillo, P., Pelegrí JL., Talley L., Peña-Izquierdo J., Cordero R. (2017), Oxygen pathways and budget for the eastern South Pacific Oxygen Minimum Zone. J. Geophys. Res. Oceans, 123. doi: 10.1002/2017JC013509
- Peña-Izquierdo, J., van Sebille E., Pelegrí JL., Mason E., Sprintall J., Llanillo P., Machín F. (2015), Water mass pathways to the North Atlantic Oxygen Minimum Zone. J. Geophys. Res. Oceans, 120, 3350–3372, doi:10.1002/2014JC010557.
- Benazzouz A., Pelegrí JL., Demarcq H., Machín F.,Mason E., Orbi A., Peña-Izquierdo J., Soumaya M. (2014), On the temporal memory of coastal upwelling off NW Africa. J. Geophys. Res. Oceans, 119, doi:10.1002/ 2013JC009559.
- Peña-Izquierdo, J., Pelegrí JL., Pastor M.V., Castellanos P., Emelianov M., Gasser M., Salvador J., & Vázquez-Domínguez E. (2012), The continental slope current system between Cape Verde and the Canary Islands. Scientia Marina, 76(S1): 65-78 doi: 10.3989/scimar.03607.18C.

2.2.2 BSC

Name:	Albert Soret		
Role:	Climate services expert Leader of WP1000 Problem formulation and state of the art	Dedication:	15%

Qualifications and experience of interest for this proposal:

Dr. Albert Soret holds a PhD in Environmental Engineering from the Polytechnic University of Catalonia (Barcelona). He is the head of the Earth System Services group at the Earth Sciences Department of the BSC. The group hosts ~28 research engineers, physicists, social scientists, economists, communication experts, and air quality/climate researchers who try to bring the latest developments in Earth sciences to the society. He is a postdoc researcher with 15 years of experience in the areas of Air Quality and Climate. His main expertise includes emission, meteorological and air quality modelling, and climate services. His research facilitates technology transfer from local and national to international levels to advance sustainable development in key sectors such as urban development, infrastructure, energy, transport, health, and agriculture and water management. He is the principal investigator of the S2S4E project (EC-H2020), a member of the External Advisory Board of Clim2Power (ERA4CS), and Work Package leader within Clim4Energy (Copernicus), VISCA (H2020) and MAGIC (Copernicus). Dr Soret is also involved in several EC-FP7 and H2020 projects, and CAMS contracts: NEWA, EUPORIAS, SPECS, IMPREX, PRIMAVERA, CAMS95 and APPRAISAL.

Selected publications:

- Torralba, V., N. Gonzalez-Reviriego, N. Cortesi, A. Manrique, L. Lledó, R. Marcos, A. Soret and F.J. Doblas-Reyes (2020). Challenges in the selection of atmospheric circulation patterns for the wind energy sector International Journal of Climatology, doi:10.1002/joc.6881
- Ramon, J., Ll. Lledó, V. Torralba, A. Soret and F.J. Doblas-Reyes (2019). What global reanalysis best represents near-surface winds? Quarterly Journal of the Royal Meteorological Society, 145, 3236-3251, doi:10.1002/qj.3616
- Soret, A., V. Torralba, N. Cortesi, I. Christel, L. Palma, A. Manrique-Suñén, Ll. Lledó, N. González-Reviriego and F.J. Doblas-Reyes (2019). Sub-seasonal to

seasonal climate predictions for wind energy forecasting. Journal of Physics: Conferences Series, 1222, 012009, doi:10.1088/1742-6596/1222/1/012009

- Turco, M., A. Ceglar, C. Prodhomme, A. Soret, A. Toreti and F.J. Doblas-Reyes (2017). Summer drought predictability over Europe: empirical versus dynamical forecasts. Environmental Research Letters, 12, 84006, doi:10.1088/1748-9326/aa7859.
- The S2S4E Decision Support Tool (https://s2s4e-dst.bsc.es) is an operational climate service that assists decision-making for the energy sector by means of subseasonal and seasonal predictions tailored to user needs.

Name:	Carlos Gómez-Gonzalez		
Role:	"Deep learning for climate science" expert Leader of WP3000 Design and Development	Dedication:	15%
Qualifications and experience of interest for this proposal:			

Dr. Carlos Gomez-Gonzalez is a STARS (MSCA-COFUND) postdoctoral fellow at the Earth Sciences department of the Barcelona Supercomputing Center (BSC-ES). He holds a Ph.D. in Science from the University of Liège (Belgium) where he carried out an interdisciplinary doctoral thesis at the interface of Computer Vision and Astrophysics. Before joining the BSC-ES, he worked as a "junior research chair in Data Science for Earth and Space sciences" at the Université Grenoble Alpes (France). With his multidisciplinary background at the interface of software development, machine learning and scientific data science, he joined the Computational Earth Sciences group at the BSC-ES to establish a research line on Artificial Intelligence for Earth Sciences. This effort focuses on the development of machine and deep learning algorithms for topics, such as statistical downscaling and bias correction techniques, data-driven parameterisations, and the study of extreme climate events.

Selected publications:

- Gómez-Gonzalez, C. A., Palma Garcia, L., Lledó, L., Marcos, R., Gonzalez-Reviriego, N., Carella, G., and Soret Miravet, A.: Deep learning-based downscaling of seasonal forecasts over the Iberian Peninsula, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-12253, https://doi.org/10.5194/egusphere-egu21-12253, 2021.
- Carella, G., Esters, L., Galí Tàpias, M., Gomez Gonzalez, C., and Bernardello, R.: Estimating the air-sea gas transfer velocity from a statistical reconstruction of ocean turbulence observations, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-10045, https://doi.org/10.5194/egusphere-egu21-10045, 2021.
- Gómez-Gonzalez, C., Serradell Maronda, K., & Donat, M., (2020). Learning to simulate precipitation with supervised and generative learning models. Presented at the Virtual Event: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction, Zenodo. http://doi.org/10.5281/zenodo.4106514
- Gómez-Gonzalez, C., Aceves Soley G., Serradell Maronda, K., Guevara Vilardell, M., Identification of Wastewater CH4 Emission Sources with Computer Vision and Sentinel-2 Observations, Presented at the ESA Phi-week 2021
- Gomez-Gonzalez, C., Absil, O., and Van Droogenbroeck, M., Supervised detection of exoplanets in high-contrast imaging sequences, A&A 613 A71 (2018), DOI: 10.1051/0004-6361/201731961

Name:	Markus Donat		
Role:	Climate extremes expert	Dedication:	7,5%
Qualifications and experience of interest for this proposal:			

Dr. Markus Donat is co-leader of the Climate Prediction group at the BSC, and an internationally recognised expert in studying climate extremes and climate variability, and predictability. Markus has published more than 90 peer-reviewed journal articles since 2010, ten of these in Nature-family journals and 4 book chapters, he has contributed to the IPCC 5th Assessment Report and is contributing author to the IPCC 6th Assessment Report. Markus is an Associate Investigator with the Australian Research Council Centre of Excellence for Climate Extremes and was selected as a member of the World Meteorological Organization (WMO) Expert Team on Data Requirements for Climate Services. Based on his achievements he has been awarded the World Climate Research Program (WCRP) / Global Climate Observing System (GCOS) International Data Prize 2017. Markus has strong expertise in the analysis of climate extremes and their variability, predictability and drivers. Of particular relevance to the proposed work in this project is Markus' expertise regarding the role of land-atmosphere interactions in driving or amplifying heat and drought extremes.

5 relevant publications:

- Donat, M. G., A. J. Pitman, O. Angélil (2018), Understanding and reducing future uncertainty in midlatitude daily heat extremes via land surface feedback constraints, *Geophysical Research Letters*, 45, 10,627–10,636. https://doi.org/10.1029/2018GL079128
- Donat, M. G., A. J. Pitman, and S. I. Seneviratne (2017), Regional warming of hot extremes accelerated by surface energy fluxes, Geophysical Research Letters, 44, 7011–7019, doi:10.1002/2017GL073733.
- Ukkola, A. M., A. J. Pitman, M. G. Donat, M. G. De Kauwe, O. Angélil (2018), Evaluating the contribution of land-atmosphere coupling to heat extremes in CMIP5 models, Geophysical Research Letters, 45, 9003–9012. https://doi.org/10.1029/2018GL079102
- Donat, M. G., A. D. King, J. T. Overpeck, L. V. Alexander, I. Durre, D. J. Karoly (2016), Extraordinary heat during the 1930s US Dust Bowl and associated large-scale conditions, Climate Dynamics, 46(1), 413-426, doi: 10.1007/s00382-015-2590-5
- Donat, M. G., L. V. Alexander, N. Herold, A. J. Dittus (2016), Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations, J. Geophys. Res. Atmos., 121, 11,174–11,189, doi:10.1002/2016JD025480

Name:	Nube Gonzalez-Reviriego		
Role:	Climate service scientist	Dedication:	7,5%
Qualifications and experience of interest for this proposal:			

Dra. Nube Gonzalez-Reviriego (PhD) is the leader of the Climate Services team (consisting of 10 people) within the Earth System Services Group at Barcelona Supercomputing Center (BSC-CNS). She has 7 years of experience in the field of climate sciences and 7 years in the field of climate services research. Her expertise lies in sub-seasonal, seasonal and decadal climate predictions for the development of climate services tailored to sectoral needs of different sectors: agriculture, renewable energy, water management and retail. She is an expert

on co-development and capacity building in the context of climate services. She is the principal investigator at BSC-CNS of the European project MED-GOLD, and has participated in 10 European projects (MED-GOLD, S2S4E, FOCUS, VITIGEOSS, EUCP, VISCA, INDECIS, IMPREX, SPECS and EUPORIAS), 6 Spanish projects, 1 cooperation project and 4 private contracts (C3S and private companies). She has supervised MSc and PhD students and has experience teaching at university. More than sixty research publications and thirty conferences in the last 5 years. Of particular relevance to the proposed work in this project is Nube' expertise on predictions of drought indices such as SPEI based on seasonal and decadal predictions and their forecast quality assessment. 5 relevant publications ● Solaraju-Murali, B., Gonzalez-Reviriego, N., Caron, L. P., Ceglar, A., Toreti, A., Zampieri, M., Bretonnière, P., Samso, M. & Doblas-Reyes, F. J. (2021). Multi-annual prediction of drought and heat stress to support decision making in the wheat sector. npj Climate and Atmospheric Science, 4(1), 1-9. ● Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N., & Doblas-Reyes, F. J. (2020). Choices in the verification of S2S forecasts and their implications for climate services. Monthly Weather Review, 148(10), 3995-4008 ● Lee, D.Y., F.J. Doblas-Reyes, V. Torralba, and N. Gonzalez-Reviriego (2019) Multi-model seasonal forecasts for the wind energy sector. Climate Dynamics, doi: 10.1007/s00382-019-04654-y ● Solaraju-Murali, B., Caron, L. P., Gonzalez-Reviriego, N., & Doblas-Reyes, F. J. (2019). Multi-year prediction of European summer drought conditions for the agricultural sector. Environmental Research Letters, 14(12), 124014. ● Gonzalez-Reviriego, N., C. Rodriguez-Puebla and B. Rodriguez-Fonseca (2015) Evaluation of observed and simulated teleconnections over the Euro-Atlantic region on the basis of partial least squares regression. Climate Dynamics 44 (11-12): 2989-3014

Name:	Lluís Palma		
Role:	Climate services and machine learning scientist	Dedication:	50%
Qualifications and experience of interest for this proposal:			
Lluís Palma García is a Junior research engineer from the computational earth sciences group at the Barcelona Supercomputing Center (BSC). He holds a bachelor's degree in aerospace engineering from the Universitat Politècnica de Catalunya (UPC) and an MSc in Meteorology from the Universitat de Barcelona (UB). He joined the BSC in 2018, where he has worked applying statistical and machine learning techniques for the post-processing of sub-seasonal to seasonal (S2S) climate predictions. Following the same line, he has participated in several H2020 projects such as S2S4E, VISCA, MED-GOLD, or Vitigeoss, in which he has been mainly in charge of designing and implementing data pipelines retrieving and post-processing real-time forecasts. In addition, he holds relevant experience working with different prediction systems, applying multiple bias-adjustment techniques, and working with traditional and machine learning downscaling methods. Selected publications:			

- Palma, L., Manrique, A., Lledó, L., Nicodemou, A., Bretonnière, P.-A., Pérez-Zanón, N., Ho, A., and Soret, A.: Lessons learned from the implementation of the near real-time S2S4E Decision Support Tool, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-15537, https://doi.org/10.5194/egusphere-egu21-15537, 2021.
- Gómez-Gonzalez, C. A., Palma Garcia, L., Lledó, L., Marcos, R., Gonzalez-Reviriego, N., Carella, G., and Soret Miravet, A.: Deep learning-based downscaling of seasonal forecasts over the Iberian Peninsula, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-12253, https://doi.org/10.5194/egusphere-egu21-12253, 2021.
- Martínez Botí, A., Palma, L., Roura, F., Manrique-Suñén, A., González-Reviriego, N., Marcos, R., González, S., López, A., and Soret, A.: Climate services for the retail sectors: the Filomena's case, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-15813, https://doi.org/10.5194/egusphere-egu21-15813, 2021.
- Soret, A., Torralba, V., Cortesi, N., Christel, I., Palma, L., Manrique-Suñén, A., Lledó, L., González-Reviriego, N., & Doblas-Reyes, F. J. (2019). Sub-seasonal to seasonal climate predictions for wind energy forecasting. *Journal of Physics: Conference Series*, *1222*, 12009. https://doi.org/10.1088/1742-6596/1222/1/012009

Name:	Llorenç Lledó		
Role:	Downscaling and Bias adjustment expert	Dedication:	7,5%

Qualifications and experience of interest for this proposal:

Dr. Llorenç Lledó holds a PhD in Physics from the Universitat de Barcelona and a MsC in Meteorology from the same university. He has 5 years of experience in co-producing climate services for the energy sector and has specialized in the downscaling, bias adjustment and post-processing of S2S dynamical predictions. Before joining the academia, he worked for ten years in the private sector developing applications of numerical weather prediction for short-term wind power forecasting. He has participated in dozens of wind resource assessment studies for developing new wind farm projects. He has supervised MsC and PhD students and has received a "Personal Técnico de Apoyo" grant from the Spanish ministry of science. He has participated in many research projects such as H2020 S2S4E, FOCUS, VITIGEOSS, EUPORIAS, ERA4CS INDECIS and MEDSCOPE, Copernicus CLIM4ENERGY, Spanish Ministry RESILIENCE and other private contracts.

- Ramon, J., Lledó, L., Bretonnière, P.-A., Samsó, M., & Doblas-Reyes, F. J. (2021). A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. Environmental Research Letters, 16(5), 054010. https://doi.org/10.1088/1748-9326/abe491
- Lledó, Ll., Torralba, V., Soret, A., Ramon, J., & Doblas-Reyes, F. J. (2019). Seasonal forecasts of wind power generation. Renewable Energy, 143, 91–100. https://doi.org/10.1016/j.renene.2019.04.135
- Ramon, J., Lledó, L., Torralba, V., Soret, A., & Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? Quarterly Journal of the Royal Meteorological Society, 145(724), 3236–3251. https://doi.org/10.1002/qj.3616

- Gómez-Gonzalez, C. A., Palma Garcia, L., Lledó, L., Marcos, R., Gonzalez-Reviriego, N., Carella, G., and Soret Miravet, A.: Deep learning-based downscaling of seasonal forecasts over the Iberian Peninsula, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-12253, https://doi.org/10.5194/egusphere-egu21-12253, 2021.
- Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., Adeniyi, K., Andela, B., Arnone, E., Berg, P., Caron, L.-P., Cionni, I., Corti, S., Drost, N., Hunter, A., Lledó, L., Mohr, C. W., Paçal, A., Pérez-Zanón, N., … Eyring, V. (2021). Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP. Geoscientific Model Development, 14(6), 3159–3184. https://doi.org/10.5194/gmd-14-3159-2021

1.1.1. EURECAT

Name:	Lali Soler		
Role:	EO Data Scientist and Explainable AI	Dedication:	1,2%
Qualifications and experience of interest for this proposal:			
Director of the Big Data & Data Science unit at Eurecat. She holds a Mathematics degree from the University of Barcelona. She also holds a master's degree in computer Vision and Artificial Intelligence from the Center for Computer Vision of Catalonia and an MBA from ESADE Business School. She has more than 15 years of experience in data analysis and image processing. She led the Remote Sensing and Photogrammetry Unit at the Cartographic and Geological Institute of Catalonia (ICGC). Recently, she specialized in management of innovation in environmental business and, more specifically, in application of data mining techniques, and design, conceptualization and management of products and services. She has several publications in Earth Observation related disciplines and presented her work at the most important EO conferences such as ISPRSS and IGARSS. She is also a lecturer in Data Science in several masters and degrees at the Universitat Autònoma de Barcelona and ESADE Business School.			
Relevant publications:			
● J. Talaya, W. Kornus, R. Alamús, **E. Soler**, M. Pla, A. Ruiz Analyzing DMC Performance in a Production Environment. Commission IV, WG IV/9 ISPRS 2005 ● Kornus, W., Magariños, A., Pla, M., **Soler**, E., Pérez, F. Photogrammetric processing using ZY-3 satellite imagery, ISPRS 2016			

Name:	Xavier Domingo		
Role:	AFLOU data scientist Expert	Dedication:	1,2%
Qualifications and experience of interest for this proposal:			
Director of the Applied Artificial Intelligence (AAI) unit at the Eurecat Technology Center. He is Diploma of Advanced Studies (DEA) in Artificial Intelligence by the University of Lleida (UdL) and Computer Engineer by the Polytechnic University of Catalonia (UPC). He has extensive experience in the application of Artificial Intelligence methodologies and techniques supported by hybrid IoT architectures with analytical capabilities for small / BIG data. Among them, smart platforms with Industry 4.0 technologies, solutions for predictive maintenance, planning and optimization of resources, monitoring and control of machinery, or traceability applied to the manufacturing industry, as well as applications in other sectors			

such as resources (water and energy), aerospace, logistics, environmental, or agri-food. He collaborates as a senior consultant in companies related to workflow management, logistics optimization, fleet management, and personal and fleet security systems, with around 20 years of professional experience. He is a collaborator of ESADE and CIHEAM-IAMZ, and partial time lecturer in the Department of Informatics and Industrial Engineering (DIEI) at the University of Lleida (UdL). He has participated in multiple European research projects (FP7, H2020) such as FIWARE4WATER, VITIGEOSS, SESAME, REVAMP or SIM4NEXUS, as well as in others of national or regional scope, or technology transfer to the private sector.

Relevant publications:

- Sušnik, J., Mereu, S., Trabucco, A., Evans, B., Khoury, M., Luchner, J., **Domingo, X.**, Vamvakeridou-Lyroudia, L. S., Chew, C., Savić, D. A., Laspidou, C., & Brouwer, F. (2018). Serious gaming to explore the water-energy-food-land-climate nexus with multi-stakeholder participation: The sim4nexus approach. 1st International WDSA / CCWI 2018 Joint Conference.
- Sušnik, J., Chew, C., **Domingo, X.**, Mereu, S., Trabucco, A., Evans, B., Vamvakeridou-Lyroudia, L., Savić, D. A., Laspidou, C., & Brouwer, F. (2018). Multi-stakeholder development of a serious game to explore the water-energy-food-land-climate nexus: The SIM4NEXUS approach. Water (Switzerland), 10(2), 139. https://doi.org/10.3390/w10020139
- Laspidou, C., Witmer, M., Vamvakeridou, L. S., **Domingo, X.**, Brouwer, F., Howells, M., Susnik, J., Blanco, M., Bonazountas, M., Fournier, M., & Papadopoulus, M. P. (2017). The water-land-food-energy-climate Nexus for a resource efficient Europe The water-land-food-energy-climate Nexus for a resource efficient Europe. 15th International Conference on Environmental Science and Technology, September.
- Evans, B., Vamvakeridou-Lyroudia, L., Susnik, J., Trabucco, A., Mereu, S., **Domingo Albin, X.,** Chew, C., & Savic, D. (2018). SIM4NEXUS – Coupling a System Dynamic Model with Serious Gaming for Policy Analysis. HIC 2018, 3, 676–667. https://doi.org/10.29007/w5vl
- Brouwer, F., Giampietro, M., Anzaldi, G., Blanco, M., Bukkens, S., Castro, B., **Domingo, X.**, Fournier, M., Funtowicz, S., Kovacic, Z., Laspidou, C., Martínez, P., Matthews, K., Munaretto, S., Romanovska, L., Schmidt, G., Serrano, T., Strand, R., Vamvakeridou-Lyroudia, L., & Witmer, M. (2017). The Nexus : efficient approaches. Pan European Networks: Science & Technology, 25, 1–4. http://edition.pagesuite-professional.co.uk/html5/reader/production/default.aspx?pubname=&edid=c780812e-ef43-4ea4-bf5a-7c8f0cfe1e7d

Name:	Aitor Corchero		
Role:	Semantic and Interoperability Data Science Expert	Dedication:	50%

Qualifications and experience of interest for this proposal:

Senior researcher and R+D project manager in the Applied Artificial Intelligence R&D Group of Eurecat Technology Centre. He studied Computer Science Engineering at the University of Mondragon (MUN) and also obtained the MSc degree in computer science at University of Lleida. He has more than 10 years of experience as data scientist and semantic web. Specifically, he has experience on semantic web technologies, data analytics (machine learning/data mining and deep learning), decision support systems (rule based reasoning and case-based reasoning) and cognitive AI for a broad of domains including water management, building energy efficiency management and physical and logical security systems (Botnet detection and remediation systems). Moreover, he is involved in Water domain (OGC®, ICT4Water Cluster), semantic web (IoT Schema.org) and IoT associations (AIOTI, BDVA). Currently, he is chair of the water management action group of the AIOTI and also, chair of the "standardization and Interoperability" action group of the ICT4WATER cluster. Moreover, Aitor has been involved and leading from EUT side more than 20 EU projects covering FP7, H2020 and LIFE projects.

Relevant publications:

- E. Rubión, **A. Corchero,** X. Domingo, L. Echeverria and G. Anzaldi, "Towards an Open, Low Cost and Enhanced Standards-Based IoT Architecture for Autonomous and Smart Water Quality Control and Monitoring" in 17th International Computing & Control for the Water Industry Conference, Exeter, Sep. 2019.
- **A. Corchero**, L. Echeverria, E. Westerhof, S. Masia, G. Anzaldi, X. Domingo, E. Rubion, J. Susnik, R. García, C. Laspidou, L. Vamvakeridou-Lyroudia and F. Brouwer, "A nexus Ontology to Support the Generation of cross domain policies," in 17th International Computing & COntrol for the Water Industry Conference, Exeter, Apr. 2019.
- G. Anzaldi, E. Rubion, **A. Corchero,** and R. Sanfeliu, "Towards an enhanced knowledge-based decision support system (DSS) for integrated water resource management (IWRM)," Procedia, 2014.
- D. Sancho et al., "UrbanWater And WatERP: Decision Support Systems For Efficient And Integrated Water Resources Management," 2014,
- **A. Corchero,** X. Domingo, and R. García, "Semantic sensor web data exploration and visualization for intelligent decision support: position paper," Proceedings of the 3rd International, 2013, [Online]. Available: https://dl.acm.org/citation.cfm?id=2479831.

2.3 Management Plan

The general purpose of the project management is to coordinate the consortium to achieve the project objectives and to control progress for each work package, co-ordination of different activities and implementation of quality control mechanisms by issuing appropriate project standards. Project management will cover administrative, financial and quality.

The project will be performed under the responsibility of Lobelia. As Project Manager, TBD will be the main communication point both within the AI4DROUGHT project team, and externally with ESA.

Individual work package leaders will be responsible for controlling and monitoring the progress of their work packages while they are active. This includes the active soliciting of input from all contributors to deliverable items. The WP leader shall monitor progress against the Project Schedule. Any problems or schedule slippage shall be reported to the Project Manager at the earliest opportunity. The Project Manager and Work Package leader shall devise a recovery plan.

L.Romero will be in direct communication with the sub-contractor and they will report to each other any problems or considerations. Technical communications among the project team will be handled through the scheduled project meetings and through emails, phone calls or skype.

The following tools and procedures for managing the subcontractor will be employed:

1. Strong communication between the subcontractor and the Project manager, ensuring that customer requirements are fulfilled and progress is properly communicated.

2. Use of a project schedule that establishes schedule constraints and includes relevant project milestones.

3. Regular progress reporting on schedule, budget and risks. This will be done on a monthly basis and included in the summary progress reports.

4. An acceptance process of each of the deliverables from the subcontractor. This will be managed by TBD and will involve checking for completeness of the document, verifying coverage of the applicable technical and scientific requirements and reviewing results with the subcontractor (if applicable).

Suitable dates for all project meetings shall be booked and agreed during the Kick-Off (KO) meeting. Progress meetings (PM), will take place every two months. The coordination of these meetings will be the responsibility of the coordinator organising the meeting. The Project Coordinator shall announce exact dates of the PM at least one month in advance.

2.4 PLANNING

2.4.1 *Gantt chart*

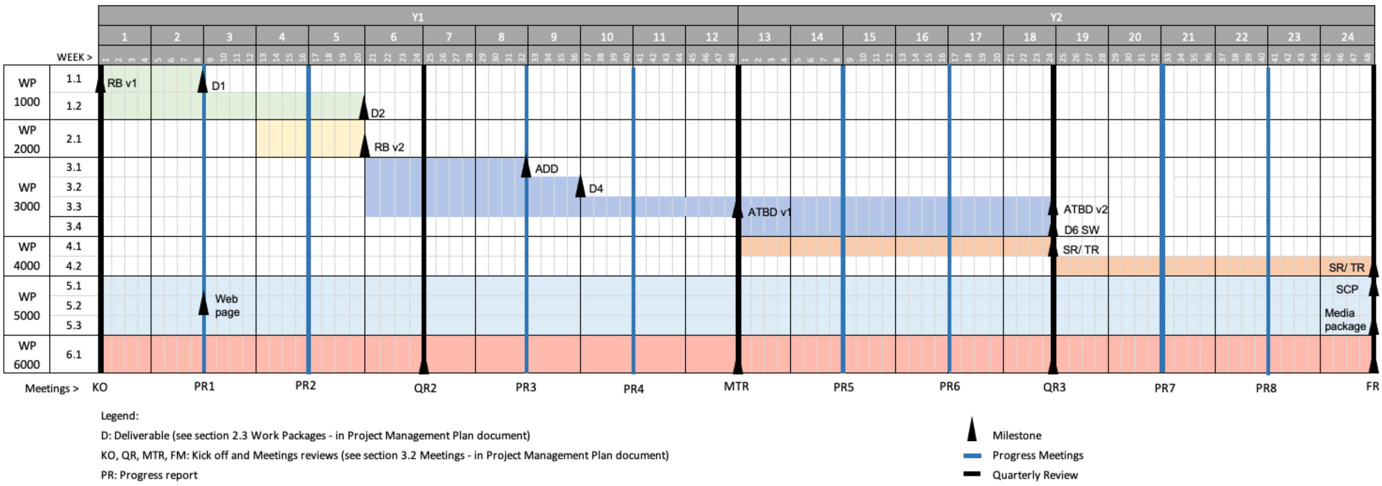


Figure 7 Gantt Chart

2.4.2 *Proposed Schedule*

We foresee a KO meeting by October 2021.
The complete meeting calendar is provided. The Final meeting could be held in October 2023.

2.4.3 Meeting and Travel Plan

Table 7 Meeting and travel plan

Meeting	Milestone acronym	Purpose	Attendees	Dates	Location	WP or Milestone
Kick-off	KO	Contract KO	All	T0	Video-conference	KO
Progress Meeting 1	PM1	Progress Review	All	T0 + 2	Video-conference	WP1
Progress Meeting 2	PM2	Progress Meeting	All	T0 +	Video-conference	WP1
First Quarterly Review	QR	Progress Review	All	T0 + 6	Video-conference	WP3
Progress Meeting 3	PM3	Progress Meeting	All	T0 + 8	Video-conference	WP1
Progress Meeting 4	PM4	Progress Meeting	All	T0 + 10	Video-conference	WP1
Mid-term Review	MTR	Progress Review	All	T0 + 12	Lobelia	MTR
Progress Meeting 5	PM5	Progress Meeting	All	T0 + 14	Video-conference	WP1
Progress Meeting 6	PM6	Progress Meeting	All	T0 + 16	Video-conference	WP1
Third Quarterly Review	QR	Progress Meeting	All	T0 + 18	Video-conference	WP1
Progress Meeting 7	PM7	Progress Meeting	All	T0 + 20	Video-conference	WP1
Progress Meeting 8	PM8	Progress Meeting	All	T0 + 22	Video-conference	WP1
Final Review	FR	Final Review	All	T0+24	ESA-ESRIN	FR

2.5 DELIVERABLE ITEMS

2.5.1 Deliverables

Table 8 Deliverables List

ID	Title	Milestone	Description
D1*	Problem formulation	end of Task 1.1	Problem formulation (D1 will be integrated as a section in RB)
D2*	SoA and gap analysis	end of Task 1.2	Statement of the art and gap analysis (D2 will be integrated as a section in RB)
D3*	ADD	end of Task 3.1	ADD (D3 will be integrated in TR) System architecture design Internal and external interfaces

			Dependencies
D4	Data repository	end of Task 3.2	Cloud repository with data bucket ready-to-use, accessible by the 3 partners
D5	ATBD	V1 at MTR V2 at QR3	Algorithm Theoretical Basis Documents for the 4 developments: AI transfer Function (Lobelia) Explainable AI: causality, interpretability (Eurecat) Improvement of seasonal forecast (BSC) Cascading effects (Eurecat)
D6	SW Release	QR3	SW Package including input and output data enabling executions of the algorithm demonstrators A user manual to such demonstrator and full documentation will be found in the Scientific Report (SR)
RB	Requirements Baseline	V1 – proposal V2 - End of Task 2	Report detailing the problem formulation, state-of-the art and gap analysis, scientific and technical requirements, including foreseen infrastructure
TR	Technical Report	V1 – End of task 2, updated quarterly	Living document, updated quarterly, including: scientific, experimental and technical setup system design system architecture test flow system testing validation
SR	Science Report	V1 – End of Task 1, updated quarterly	Living document, updated quarterly, consisting of a well-defined and self contained scientific study in a demonstration scenario, with replicable methodology and technical approaches; final version of the Science Report includes a scientific roadmap with the remaining open questions, scientific gaps and proposed ways forward.
PR	Progress Report	Monthly	Living document updated monthly, recording the project backlog. Includes progress, pending items, challenges and actions;
MTR	Mid term Review Report	Mid-term Review	Summary of main activities, progress and issues outstanding. It includes the corresponding TR and SR iterations
FR	Final Report	Final Review	Summary of main activities, results and conclusions of the project, including an executive summary
SCP	Scientific Communication Package		Publications in peer-reviewed publications, digitally accessible at the time of publication or within a reasonable time period after publication; the proposal shall include a

			preliminary identification of targeted peer-reviewed journals; this package includes full versions of publications (published or accepted for publishing)
MP	Media Package	Final Review	Communication materials including web page (with public access to datasets generated during the contract), project multimedia, scientific publications, social media (including analytics), and web story
CCD	Contract Closure Documentation	Contract Closure	
FP	Final Presentation	Final Review	

3 FINANCIAL PROPOSAL

3.1 PRICE QUOTATION FOR THE CONTEMPLATED CONTRACT

The total price for this proposal is **399,508.00 €.** The validity of this proposal is 4 months.

The type of price is Firm Fixed Price (FFP) in Euro, delivery duty paid, exclusive of import duties and value added taxes in ESA Member States, etc., in pursuance of the pricing conditions fixed in the "Draft Contract" included in the ITT.

3.2 DETAILED PRICE BREAKDOWN

3.2.1 PSS costing forms

The following PSSs are included as Annex to this proposal, for the Prime contractor (Lobelia Earth) and for the subcontractors (BSC and EURECAT) as well as the Aggregated PSSs:

- PSS A1 Company Cost Rates and Overheads
- PSS A2 Company Price Breakdown Form
- PSS A2 Exhibit B – Travel and subsistence plan
- PSS A8 Person months & Price Summary per WP
- PSS A15.1 Company price projection vs. payment plan
- PSS A2, Exhibit B and PSS A8 are included as Annex to this proposal, both for the Prime contractor (isardSAT) and for the subcontractor (IRD).

3.2.2 Milestone Payment Plan

Milestone (MS) Description	Schedule Date	Payments from ESA to (Prime) Contractor (in Euro)	Country (ISO code)
Milestone 1 (including provision for advance payment as identified hereunder): Upon successful first quarterly review and acceptance of relevant deliverables	T0+6	119,852€	ES
Milestone 2: Upon successful Mid Term review and acceptance of relevant deliverables	T0+12	79,902€	
Milestone 3: Upon successful third quarterly review and acceptance of relevant deliverables	T0+18	119,852€	
FINAL: Upon successful completion of contract and acceptance of all deliverables including contract closure documentation	T0+24	79,902€	
TOTAL		**399,508 €**	

Prime (P)	Company Name	ESA Entity Code	Country (ISO code)	Advance Payment (in Euro)	Offset against[1]	Offset by Euro	Condition for release of the Advance Payment
P	Lobelia Earth SL	1000034919	ES	44,800.00 €	100% MS-1 25% MS-4	Lobelia Earth SL	1000034919

3.2.3 Travel and subsistence plan

Travel and subsistence plan is included in Exhibit B of the PSS A2 forms from Lobelia Earth, BSC and Eurecat

4 CONTRACTUAL PROPOSAL

4.1 Intellectual Property Rights

4.1.1 Background Intellectual Property

In line with Article 6.3 of the Draft Contract, no Background Intellectual Property will be incorporated in the deliverables.

4.1.2 Foreground Intellectual Property and IP ownership

Algorithms developed in the frame of the AI4DROUGHT project, as well as software implementing those algorithms and programs later developed which include these algorithms will be open source.

4.2 Specification of all inputs to enter into blanks existing in the draft contract

All correspondence for the Contractor shall be addressed as follows:

a) for technical matters as follows:

	To	With copy to:
Name	Laia Romero	Jesús Peña Izquierdo
Telephone No.	+34 933 505 508	+34 933 505 508
e-mail address	laia.romero@lobelia.earth	jesus.pena@lobelia.earth

b) for contractual and administrative matters as follows:

c)

	To	With copy to:
Name	Lluís Vinyals	Giulia Galante
Telephone No.	+34 933 505 508	+34 933 505 508
e-mail address	administration@lobelia.earth	giulia.galante@lobelia.earth

4.3 Other remarks on the Draft Contract

The proposed contract duration is 24 months

ANNEX 1: PSS FORMS

Annex 1: Lobelia Earth SL PSSs

Annex 1: Barcelona Supercomputing Center PSSs

Annex 1: Eurecat PSSs

Annex 1: Aggregated PSSs

Annex 2. References

Alessandrini, S., Sperati, S., & Pinson, P. (2013). A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. In *Applied Energy* (Vol. 107, pp. 271–280). Elsevier Ltd. https://doi.org/10.1016/j.apenergy.2013.02.041

Bauer, P., Stevens, B., & Hazeleger, W. (2021). A digital twin of Earth for the green transition. In *Nature Climate Change* (Vol. 11, Issue 2, pp. 80–83). Nature Research. https://doi.org/10.1038/s41558-021-00986-y

Baugh, C., de Rosnay, P., Lawrence, H., Jurlina, T., Drusch, M., Zsoter, E., & Prudhomme, C. (2020). The Impact of SMOS Soil Moisture Data Assimilation within the Operational Global Flood Awareness System (GloFAS). *Remote Sensing*, *12*(9), 1490. https://doi.org/10.3390/rs12091490

Belle, V., & Papantonis, I. (2020). *Principles and Practice of Explainable Machine Learning*. http://arxiv.org/abs/2009.11698

Bergeron, J. M., Trudel, M., & Leconte, R. (2016). Combined assimilation of streamflow and snow water equivalent for mid-term ensemble streamflow forecasts in snow-dominated regions. *Hydrology and Earth System Sciences*, *20*(10), 4375–4389. https://doi.org/10.5194/hess-20-4375-2016

Brock, A., De, S., Smith, S. L., & Simonyan, K. (n.d.). *High-Performance Large-Scale Image Recognition Without Normalization*. Retrieved June 14, 2021, from https://github.com/deepmind/

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*(3), e1520. https://doi.org/10.1002/wat2.1520

Brunner, M. I., Viviroli, D., Sikorska, A. E., Vannier, O., Favre, A., & Seibert, J. (2017). Flood type specific construction of synthetic design hydrographs. *Water Resources Research*, *53*(2), 1390–1406. https://doi.org/10.1002/2016WR019535

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in Fintech Risk Management. *Frontiers in Artificial Intelligence*, *3*, 26. https://doi.org/10.3389/frai.2020.00026

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., & Xue, X. (2019). Research on soil moisture prediction model based on deep learning. *PLoS ONE*, *14*(4), e0214508. https://doi.org/10.1371/journal.pone.0214508

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). *End-to-End Object Detection with Transformers*. http://arxiv.org/abs/2005.12872

Chen, Z., Liu, H., Xu, C., Wu, X., Liang, B., Cao, J., & Chen, D. (2021). Modeling vegetation greenness and its climate sensitivity with deep-learning technology. *Ecology and Evolution*, *00*(1), 1–11. https://doi.org/10.1002/ece3.7564

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E. (2019). S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. In *Wiley Interdisciplinary Reviews: Climate Change* (Vol. 10, Issue 2, p. e00567). Wiley-Blackwell. https://doi.org/10.1002/wcc.567

Council, N. R. (2012). *A National Strategy for Advancing Climate Modeling*. National Academies Press. https://doi.org/10.17226/13430

de Brito, M. M. (2021). Compound and cascading drought impacts do not happen by chance: A proposal to quantify their relationships. *Science of the Total Environment*, *778*, 146236. https://doi.org/10.1016/j.scitotenv.2021.146236

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(4), 245–268. https://doi.org/10.1002/wcc.217

Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P., & Ludwig, F. (2017). Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level. *Climatic Change*, *143*(1–2), 13–26. https://doi.org/10.1007/s10584-017-1971-7

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929

Enenkel, M., Brown, M. E., Vogt, J. V., McCarty, J. L., Reid Bell, A., Guha-Sapir, D., Dorigo, W., Vasilaky, K., Svoboda, M., Bonifacio, R., Anderson, M., Funk, C., Osgood, D., Hain, C., & Vinck, P. (2020). Why predict climate hazards if we need to understand impacts? Putting humans back into the drought equation. *Climatic Change*, *162*(3), 1161–1176. https://doi.org/10.1007/s10584-020-02878-0

Escorihuela, M. J., Quintana-Seguí, P. Q., Stefan, V., & Gaona, J. (2020). Drought monitoring in the Ebro basin: comparison of Soil Moisture and Vegetation anomalies. *EGU General Assembly 2020*. https://doi.org/10.5194/egusphere-egu2020-11120

Foley, C. J., Vaze, S., Seddiq, M., Unagaev, A., & Efremova, N. (n.d.). *SMARTCAST: PREDICTING SOIL MOISTURE INTERPO-LATIONS INTO THE FUTURE USING EARTH OBSERVA-TION DATA IN A DEEP LEARNING FRAMEWORK*. Retrieved June 14, 2021, from www.deepplanet.ai

Gerber, N., & Mirzabaev, A. (2017). *Integrated Drought Management Programme Working Paper No. 1 Benefits of action and costs of inaction: Drought mitigation and preparedness-a literature review*. https://www.gfdrr.org/post-disaster-needs-assessments

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., & Hoefler, T. (2020). Deep Learning for Post-Processing Ensemble Weather Forecasts. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194). https://doi.org/10.1098/rsta.2020.0092

Jocher, G., Stoken, A., Borovec, J., NanoCode012, Chaurasia, A., TaoXie, Changyu, L., V, A., Laughing, tkianai, yxNONG, Hogan, A., lorenzomammana, AlexWang1900, Hajek, J., Diaconu, L., Marc, Kwon, Y., oleg, … Ingham, F. (2021). *ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations*. https://doi.org/10.5281/ZENODO.4679653

Kakogeorgiou, I., & Karantzalos, K. (2021). *Evaluating Explainable Artificial Intelligence Methods for Multi-label Deep Learning Classification Tasks in Remote Sensing*. http://arxiv.org/abs/2104.01375

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint Trends in Flood Magnitudes and Spatial Extents Across Europe. *Geophysical Research Letters*, *47*(7), e2020GL087464. https://doi.org/10.1029/2020GL087464

Klingmüller, K., & Lelieveld, J. (2021). Climate model-informed deep learning of global soil moisture distribution. *Geoscientific Model Development Discussions*, 1–17. https://doi.org/10.5194/gmd-2020-434

Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., & Reichstein, M. (2019). Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks. *Frontiers in Big Data*, *2*, 31. https://doi.org/10.3389/fdata.2019.00031

Li, Z., Zhang, Z., & Zhang, L. (2021). Improving regional wheat drought risk assessment for insurance application by integrating scenario-driven crop model, machine learning, and satellite data. *Agricultural Systems*, *191*, 103141. https://doi.org/10.1016/j.agsy.2021.103141

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2021). Explainable ai: A review of machine learning interpretability methods. In *Entropy* (Vol. 23, Issue 1, pp. 1–45). MDPI AG. https://doi.org/10.3390/e23010018

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. In *Water Resources Research* (Vol. 43, Issue 7, p. 7401). John Wiley & Sons, Ltd. https://doi.org/10.1029/2006WR005756

Marie Demajo, L., Vella, V., & Dingli, A. (2020). *EXPLAINABLE AI FOR INTERPRETABLE CREDIT SCORING*. 185–203. https://doi.org/10.5121/csit.2020.101516

Masmoudi, M., Lamine, S. B. A. Ben, Zghal, H. B., Archimede, B., & Karray, M. H. (2021). Knowledge hypergraph-based approach for data integration and querying: Application to Earth Observation. *Future Generation Computer Systems*, *115*, 720–740. https://doi.org/10.1016/j.future.2020.09.029

McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

Mishra, N., Prodhomme, C., & Guemas, V. (2019). Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. *Climate Dynamics*, *52*(7–8), 4207–4225. https://doi.org/10.1007/s00382-018-4404-z

Moradkhani, H., Hsu, K. L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*(5), 1–17. https://doi.org/10.1029/2004WR003604

Noura, H. N., Salman, O., & Sider, A. (2021). *A DEEP LEARNING OBJECT DETECTION METHOD FOR AN EFFICIENT CLUSTERS INITIALIZATION A PREPRINT*. https://arxiv.org/pdf/2104.13634.pdf

Papoutsis, I., Baglatzi, A., Touloumtzi, S., Reichstein, M., Carvalhais, N., Gans, F., Camps-Valls, G., Piles, M., Kakantousis, T., Dowling, J., Koubarakis, M., Bilidas, D., Pantazi, D.-A., Stamoulis, G., Demange, C., Journel, L.-G., Bianchi, M., Gervasi, C., Rucci, A., … Paschalis, A. (n.d.). *DEEPCUBE: EXPLAINABLE AI PIPELINES FOR BIG COPERNICUS DATA*. Retrieved June 14, 2021, from https://deepcube-h2020.eu/

Peng, B., Guan, K., Pan, M., & Li, Y. (2018). Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield. *Geophysical Research Letters*, *45*(18), 9662–9671. https://doi.org/10.1029/2018GL079291

R.K. Pachauri, L. A. M. (2014). Climate Change 2014: Synthesis Report. In *IPCC,Geneva, Switzerland* (Vol. 218).

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Requena-Mesa, C., Benson, V., Denzler, J., Runge, J., & Reichstein, M. (2020). *EarthNet2021: A novel large-scale dataset and challenge for forecasting localized climate impacts*. http://arxiv.org/abs/2012.06246

Rudd, A. C., Kay, A. L., & Bell, V. A. (2019). National-scale analysis of future river flow and soil moisture droughts: potential changes in drought characteristics. *Climatic Change*, *156*(3), 323–340. https://doi.org/10.1007/s10584-019-02528-0

Sachan, S., Yang, J. B., Xu, D. L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, *144*, 113100. https://doi.org/10.1016/j.eswa.2019.113100

Salcedo-Sanz, S., Ghamisi, P., Piles, M., Werner, M., Cuadra, L., Moreno-Martínez, A., Izquierdo-Verdiguier, E., Muñoz-Marí, J., Mosavi, A., & Camps-Valls, G. (2020). Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. *Information Fusion*, *63*, 256–272. https://doi.org/10.1016/J.INFFUS.2020.07.004

Soret, A., Torralba, V., Cortesi, N., Christel, I., Palma, L., Manrique-Suñén, A., Lledó, L., González-Reviriego, N., & Doblas-Reyes, F. J. (2019). Sub-seasonal to seasonal climate predictions for wind energy forecasting. *Journal of Physics: Conference Series*, *1222*(1), 012009. https://doi.org/10.1088/1742-6596/1222/1/012009

Thi Thanh Hang, P., & Hang, P. T. T. (2019). Applying the Distributed Hydrological Model for Tropical Monsoon Basins by Using Earth Observation Data (Case Studies: Kone and Ba River Basins). *Journal of Geoscience and Environment Protection*, *07*(01), 23–37. https://doi.org/10.4236/gep.2019.71003

Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., & Davis, M. (2017). Seasonal climate prediction: A new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, *56*(5), 1231–1247. https://doi.org/10.1175/JAMC-D-16-0204.1

Tsagkatakis, G., Aidini, A., Fotiadou, K., Giannopoulos, M., Pentari, A., & Tsakalides, P. (2019). Survey of deep-learning approaches for remote sensing observation enhancement. In *Sensors (Switzerland)* (Vol. 19, Issue 18). MDPI AG. https://doi.org/10.3390/s19183929

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., Decker, M., & Haughton, N. (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environmental Research Letters*, *11*(10), 104012. https://doi.org/10.1088/1748-9326/11/10/104012

Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., & Espejo, F. (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. *Environmental Research Letters*, *9*(4), 044001. https://doi.org/10.1088/1748-9326/9/4/044001

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. In *Nature Communications* (Vol. 11, Issue 1, pp. 1–10). Nature Research. https://doi.org/10.1038/s41467-019-14108-y

Walz, M. A., Donat, M. G., & Leckebusch, G. C. (2018). Large-Scale Drivers and Seasonal Predictability of Extreme Wind Speeds Over the North Atlantic and Europe. *Journal of Geophysical Research: Atmospheres*, *123*(20), 11,518-11,535. https://doi.org/10.1029/2017JD027958

Wang, L., & Yan, J. (2020). Stewardship and analysis of big Earth observation data. *Big Earth Data*, *4*(4), 349–352. https://doi.org/10.1080/20964471.2020.1857055

Wilhite, D. A. (2000). *Chapter 1 Drought as a Natural Hazard: Concepts and Definitions*. http://digitalcommons.unl.edu/droughtfacpubhttp://digitalcommons.unl.edu/droughtfacpub/69

Zsoter, E., Cloke, H., Stephens, E., de Rosnay, P., Muñoz-Sabater, J., Prudhomme, C., & Pappenberger, F. (2019). How Well Do Operational Numerical Weather Prediction Configurations Represent Hydrology? *Journal of Hydrometeorology*, *20*(8), 1533–1552. https://doi.org/10.1175/JHM-D-18-0086.1

[End of the document]