RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

2025 *Activity Report*

RESEARCH CENTRE: **Inria Paris Centre**

IN PARTNERSHIP WITH: **Université Versailles Saint-Quentin, Sorbonne Université, CNRS**

Project-Team

ARCHES

AI Research for Climate Change and Environmental Sustainability

In collaboration with Laboratoire Atmosphères, Observations Spatiales

Project-Team ARCHES

Creation of the Project-Team: 2025 August 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

A3.1.4. – Uncertain data

A9.2.2. – Unsupervised learning

A9.2.8. – Deep learning

A9.7. – AI algorithmics

A9.10. – Hybrid approaches for AI

A9.11. – Generative AI

A9.12. – Computer vision

A9.14. – Evaluation of AI models

A9.16. – Societal impact of AI

Other research topics and application domains

B3. – Environment and planet

B3.1. – Sustainable development

B3.2. – Climate and meteorology

B3.3. – Geosciences

B3.4.1. – Natural risks

Contents

1 Team members, visitors, external collaborators

Research Scientists

- Claire Monteleoni [Team leader, INRIA, Senior Researcher, HDR]
- Anastase Charantonis [INRIA, Chair, Professor Junior (CPJ)]
- Emmanuel De Bezenac [INRIA, ISFP]
- Julie Keisler [INRIA, Starting Research Position, from Jun 2025, (Previously PhD student in ARCHES)]
- Sarah Safieddine [CNRS, Researcher, HDR]

Faculty Members

- Laurent Barthes [UVSQ, Associate Professor, IPSL/LATMOS, HDR]
- Aymeric Chazottes [UVSQ, Associate Professor, IPSL/LATMOS]
- Cecile Mallet [UVSQ, Professor, IPSL/LATMOS, HDR]

Post-Doctoral Fellows

- Victor Enescu [UVSQ, from Nov 2025]
- Assaad Zeghina [CNRS, from Nov 2025, projet PRISME France 2030]

PhD Students

- Valerio Actis Dato Casale [Sorbonne Université, from Sep 2025]
- Nathan Chalumeau [AXA CLIMATE, CIFRE, from Apr 2025, Sorbonne]
- Graham Clyne [INRIA, Sorbonne]
- Clement Dauvilliers [INRIA, Sorbonne]
- Aymeric Delefosse [INRIA, from Jul 2025, Sorbonne (Previously Research Engineer in ARCHES)]
- Pierre Garcia [Amphitrite, CIFRE, Sorbonne with LiP6]
- Baptiste Guigal [BOWEN, CIFRE, UVSQ]
- David Landry [INRIA, Sorbonne]
- Angel Luque Lazaro [Spascia, CIFRE, from Sep 2025, Sorbonne]
- Gabriela Martinez Balbontin [Mecrator Oceans, CIFRE, from Mar 2025, Sorbonne with LiP6, and hosted in Brest]
- Matthieu Meignin [UVSQ, Académie Spatiale d'île de France]
- Renu Singh [GOOGLE, CIFRE, from Apr 2025, Sorbonne (Previously Research Engineer in ARCHES)]
- Ganglin Tian [LMD X, until Nov 2025]

Technical Staff

- Gregoire Mourre [INRIA, Engineer, from Nov 2025]
- Marc Pyolle [CNRS, Engineer, from Sep 2025]

Interns and Apprentices

- Dimitrii Drozdov [LiP6, from Apr 2025 until Sep 2025]

Administrative Assistants

- Derya Gok [INRIA]
- Anne Mathurin [INRIA]

External Collaborator

- Guillaume Couairon [Google DeepMind, (Previously SRP in ARCHES)]

2 Overall objectives

2.1 Overview and scientific context

Understanding and addressing climate change is an urgent challenge. Meanwhile, the study of climate change is an extremely data-rich field, especially considering not only the rapidly growing amount of satellite retrievals but also the massive amounts of simulation output from physics-driven climate models, providing a lens into the distant past and distant future. For over a decade, the proposing team members have been pursuing a research vision that machine learning can shed light on and help in confronting climate change, launching the interdisciplinary field of *Climate Informatics* which was recognized as key research priority in The World Economic Forum's report on AI for the Earth, in 2018.

2.1.1 Research objectives

ARCHES: AI Research for Climate Change and Environmental Sustainability is focused on using AI to address climate change, and to enable environmentally sustainable solutions. In particular, ARCHES research is focused along three axes:

1. AI for Climate Change Adaptation – Forecasting and informing near-term decisions
2. AI for Climate Change Mitigation – Forecasting and informing mid-term decisions
3. AI for Understanding Climate Change Impacts – Projecting long-term impacts

While addressing these problems, we will also continue to advance core AI research in machine learning in computer vision. Our past work has demonstrated that climate and environmental applications open new questions for the design and analysis of machine learning algorithms. We have also found that applied research can yield unorthodox twists, even on standard machine learning techniques, which in turn spark interest in the machine learning research community. Examples include our work on climate prediction via sparse matrix completion, unsupervised learning of data-driven, probabilistic definitions of diverse, multivariate extreme events, and deep unsupervised learning for anomaly detection in problems with limited labeled data.

3 Research program

To address the three axes stated above, ARCHES will start with research on the topics described below, for which we already have collaborators, and then build out additional areas as we gain new hires and new collaborations. Since climate change is a global issue, we will also pursue international collaborations, and build on our existing ones, including several in the US.

In order to study climate change on a regional or urban scale (more than 50 percent of the world's population is affected by changes in cities), global data must be supplemented by local observations to capture small-scale heterogeneities and weak signal (sensor networks, opportunistic sensors). The team's overall research approach is to develop and apply machine learning algorithms and models that combine and exploit information from a range of data sources: from Global Circulation Model (GCM) simulations (e.g., CMIP6), to *reanalysis* data, the product of *data assimilation* which processes observation data onto a geospatial grid, using physical laws (e.g., ERA5), to satellite data at various level of processing, to in-situ measurements such as ocean gliders and radar stations.

3.1 AI for Climate Change Adaptation: Forecasting Extreme Weather, Cascading Hazards

With the changing climate, many communities across the globe are being hard-hit by extreme weather events and the resulting hazards. Events such as extreme precipitation and heatwaves can result in flooding and wildfire, as well as cascading hazards resulting from multiple extreme events, such as powerful and extremely dangerous debris flows (mudslides) which can occur when there are heavy rains after drought or wildfire. ARCHES members have demonstrated that machine learning can improve the detection and forecasting of a variety of extreme events, e.g., tropical cyclones, avalanches, and extreme precipitation events.

Our ongoing work contributes to AI for weather forecasting, including the particular challenges related to precipitation and extreme events. Addressing these challenges will also enable progress on our longer-term research agenda of confronting cascading hazards. The methods we develop can be used by our collaborators at MétéoFrance and ECMWF to implement AI-driven forecasting tools to provide decision-support for communities and decision-makers.

3.2 AI for Climate Change Mitigation

Climate Change Mitigation (*"Attenuation," en français*) refers to actions that society can take in the near to mid-term in order to reduce the risks of the worst possible long-term impacts of climate change. According to IPCC (*GIECC*) global warming has been accelerated by anthropogenic emissions of carbon dioxide and other greenhouse gasses. Targeting reduced emissions of carbon dioxide, ARCHES will focus on AI approaches to accelerating the renewable energy transition, and to better modeling the effects of land-changes and land-use changes on carbon fluxes.

3.3 AI for Understanding Climate Change Impacts: Atmosphere, Ocean, and Water-cycle

The atmosphere, ocean, and processes at their intersection are critical in understanding climate change. Indeed, many of the extreme weather events discussed earlier depend on *climate modes of variability*, such as the El Niño-Southern Oscillation (ENSO), which may themselves be changing in a warming climate. Our team has worked on machine learning and causal inference approaches to better understanding which sub-regions of the Pacific Ocean will be more indicative of ENSO, in a changing climate. We have also worked with the Indian Meteorological Department (IMD), which has been incorporating AI into some of their forecast tools, on using machine learning to improve forecasts of precipitation extremes during the Indian Summer Monsoon, a phenomenon that significantly effects the GDPs of the entire Indian subcontinent. Our past work has also shown that machine learning can robustify the long-term projections of climate model ensembles, by training on both their simulations, as well as observation (reanalysis) data.

We currently have several projects on forecasting sea-level rise from satellite altimetry data and GCM simulations, under various climate forcings. We also have projects on reducing the uncertainty in future climate projections, with the use of our AI-driven climate emulators. A longer-term goal is to use AI to better model the effects of land-use on carbon emissions, with the eventual goal of studying longer-term carbon emissions and their climate change impacts

3.4 Links between research objectives

From an application perspective, all three research objectives described above are highly interdependent. For example, weather forecasting is an essential component of addressing the energy transition; the main renewable energy sources (wind, solar, hydro) are all heavily dependent on meteorological conditions and correctly predicting their energy output at a variety of time-scales is essential for the stability of the electricity grid. Similarly, the energy demand is strongly correlated with temperature, so correctly predicting extreme events such as cold waves with sufficient advanced warning is very important for energy planning. Because of this synergy among our research objectives, each machine learning method we develop will be applied to several of our proposed application objectives. For example, machine learning-based downscaling approaches will allow us to contribute both to accelerating the renewable energy transition (as discussed), as well as to studying the impacts of global-warming on sea-level.

3.5 Advancing machine learning research

ARCHES team members have already demonstrated that new machine learning research is often needed when addressing applications in climate change and sustainability. Spatiotemporal data is increasingly prevalent in a variety of applications, for example, climate science and agriculture. Moreover, incorporating additional dimensions (which need not be spatial) into time-series data addresses a range of fields, including financial monitoring over multiple markets. Our proposed environmental and climate change research objectives cannot be achieved solely by applying existing AI and machine learning methods, as there are a variety of challenges. For example,

- AI models often struggle to handle multi-source, sparse, and noisy observational data.
- AI methods are not designed to incorporate data and detect patterns at multi-temporal and spatial scales simultaneously, nor to adapt to changing regimes.
- Training AI models for under-sampled phenomena such as extreme weather events requires novel approaches.

Our research targets the various challenges inherent in *learning from spatiotemporal data* including *multi-source data* with multiple sources of *uncertainty*. Our current approaches are primarily focused on (1) *self-supervised learning* and (2) *generative models*. The team also has significant expertise in (3) learning from *non-stationary spatiotemporal data*.

4 Application domains

As the team name indicates, ARCHES primarily focuses on applications in the fields of climate change and environmental sustainability.

Team members have developed AI approaches for a variety of environmental and climate change applications, including day-ahead forecasting of tropical cyclone tracks, multiple day-ahead forecasting of precipitation extremes during the Indian Summer Monsoon, and sub-seasonal forecasting of available solar power. In collaboration with Météo-France we have worked on avalanche detection from satellite imagery. We have also worked on a variety of ill-posed inverse problems, from retrieving the state of the deep ocean from sea-surface observations, to filling satellite data gaps, to exploring how to best combine physical models and observations using deep learning and data assimilation. Our research spans a range of applications involving forecasting different aspects of the earth system at different time-scales, which is critical for planning and adapting to the effects of climate change. We have works spanning short-term "nowcasting," to exploring past climates.

Applications include but are not limited to:

- **Weather** Deterministic and probabilistic forecasting and nowcasting, post-processing, super-resolution, particular foci on precipitation and extreme events
- **Climate** Deterministic and probabilistic forecasting, climate model emulation under unseen scenarios, super-resolution, unpaired domain alignment, climate trend evaluation based on spatial observations
- **Oceans** Surface ocean currents estimation and forecasting from satellite and in-situ observations, biogeochemical vertical composition compression and forecasting
- **Remote Sensing** Domain alignment, quantitative precipitation estimation, cyclone tracking, sea-surface height inpainting from sparse data, radar rain maps, denoising and inpainting

5 Social and environmental responsibility

ARCHES is addressing climate change adaptation and mitigation and environmental sustainability, by design.

5.1 Reducing the carbon footprint of weather forecasting and climate modeling

Because of its critical importance in many domains (agriculture, logistics, energy, etc.), weather forecasting is one of the main usages of supercomputers today. For instance, Météo France operates a 20 petaflops datacenter. By developing AI methods for weather forecasting, our research demonstrates that it is possible to significantly reduce its carbon footprint. The recent machine learning-based weather models from Google DeepMind and other large teams learn to imitate physics-based models, while optimally allocating their computational budget to make the most accurate forecast, resulting in inference speedups of orders of magnitude (estimated 1000-10,000). But now, ARCHES has developed more frugal weather models that take these savings even further; ArchesWeather and ArchesWeatherGen have training times 20-100x lower than that of Google DeepMind's AI weather models, without degrading performance in weather prediction.

Adapting our models to climate model emulation yields even greater speedups. Our model, ArchesClimate, is about 400x faster compared to the climate model it emulates, in part by operating on a subset of variables, allowing for a "reduced complexity" run without sacrificing the complexity of physical processes.

6 Highlights of the year

Team launch The ARCHES project-team became official in August 2025, and the launch event was held in November 2025.

Professional honors Claire Monteleoni, invited seminar, College de France, May 2025.

6.1 A note on 2025 publications

ARCHES launched officially in August 2025. Therefore, many of our 2025 publications were not linked to ARCHES within HAL. The only way the RADAR system can add these into this report is through the section Major Publications, not through Publications of the year.

7 Latest software developments, platforms, open data

7.1 Latest software developments

7.1.1 DRAGON

Name: Directed Acyclic Graph OptimisatiON

Keywords: Automated machine learning, Neural architecture search, Neural networks

Scientific Description: The search space is made from Python objects called Variables, which can encode integers, arrays, or other elements. These variables are combined to create DAGs representing neural networks. Each variable can also have neighbor or mutation operators to explore slightly different configurations, and crossover operators allow combining multiple configurations. DRAGON includes several search algorithms: Random Search, Evolutionary Algorithm, Mutant-UCB, and HyperBand. Some algorithms (like Evolutionary Algorithm and Mutant-UCB) use the neighbor/mutation functions to explore new configurations. Algorithms also come with memory-efficient storage and an optional distributed version for running on multiple processors. Finally, performance evaluation is handled by the user: a network is built from a configuration, trained, and evaluated to return a loss, which the search algorithms aim to minimize.

Functional Description: DRAGON is an open-source Python tool that helps design and optimize deep learning models. It represents neural networks as graphs of connected operations, where each operation can be adjusted to improve performance. Unlike many "automatic" AI tools, DRAGON requires users to define the structure of the network and how it is trained, giving much more flexibility to solve different problems. Users can explore different network designs by slightly modifying existing configurations or combining multiple designs. The tool includes methods to test many options efficiently, keeping memory usage low and even allowing multiple computers to work together. Once a network is designed, it is trained and evaluated, and the results guide the next round of improvements. DRAGON has been utilized for tasks such as image recognition and predicting energy load, but it can be adapted to address many other AI problems.

URL: https://github.com/JulieKeisler/DRAGON

Contact: Julie Keisler

7.1.2 geoarches

Keywords: Machine learning, Forecasting, Climate change, Data processing

Scientific Description: geoarches is a research-friendly machine learning library for training, running, and evaluating models on geospatial data, mainly weather and climate data. Built on PyTorch, Pytorch Lightning, and Hydra, geoarches offers a clean, modular structure for developing and scaling ML pipelines. It can also be used to run the ArchesWeather and ArchesWeatherGen weather models.

Functional Description: geoarches is a machine learning library for training, running and evaluating models on weather and climate data.

Release Contributions: This is the first version.

URL: https://github.com/INRIA/geoarches

Contact: Renu Singh

7.1.3 SerpentFlow

Name: SharEd-structuRe decomPosition for gEnerative domaiN adapTation

Keywords: Deep learning, Super-resolution, Domain Adaptation

Functional Description: SerpentFlow (SharEd-structuRe decomPosition for gEnerative domaiN adapTation) is a framework for unpaired domain alignment. It separates shared low-frequency structures from domain-specific high-frequency content and uses Flow Matching for generative modeling.

Contact: Julie Keisler

7.1.4 Motif

Name: Multi-source transformer via factorized attention

Keywords: Forecasting, Deep learning, Climate change, Satellite imagery

Scientific Description: This repository implements a DL architecture adapted to learning from multiple sources. The possibilities of inputs include:

A flexible number of sources: while a large set of sources can be used as input to the model, a specific sample may contain any subset of the sources. Samples with different numbers of sources can be bathed together during training or inference. Sources of different dimensionalities and natures (0D, e.g. station measurements, 1D, e.g. vertical profile, 2D, e.g. remote sensing images). Sources misaligned in space and time, for example remote sensing images covering different geographical areas (which may even be disjoint), and at irregular time intervals. Sources of the same type with different characteristics, e.g. remote sensing images in the same band from different satellites with different exact frequencies and ground sampling distance

Functional Description: Motif is a Python package to train AI models on geospatial data from multiple sources. The code includes a data engineering pipeline and a novel neural networks architecture for data fusion.

Contact: Clement Dauvilliers

7.2 Open data

Participants: Laurent Barthès.

A dataset of annotated ground-based images for the development of contrail detection algorithms [9].

Participants: Sarah Safieddine.

Development of a Merged CO Climate Data Record from IASI and MOPITT Observations [8].

8 New results

Here we provide descriptions of our new research results in 2025 along each research objective.

8.1 Climate change adaptation

8.1.1 Climate change dependence on local time

Participants: Sarah Safieddine
.

In [14], we calculate climate trends in local time. In fact, essential Climate Variables, such as near-surface (T2m) and land surface temperatures (LST), are typically reported in Coordinated Universal Time (UTC) for global consistency. However, their diurnal variability leads to temperature trends that differ by the local hour, a factor not analyzed on the global nor regional scale. Using ECMWF ERA5-Land reanalysis data (1981–2022), we assess temperature trends by local hour and month. Our results show that the trends can change significantly during the day. LST and T2m warming or cooling trends peak in the afternoon, while showing large spatial variability across both hemispheres. Using MODIS observations, we show how the nominal Equator crossing times of TERRA and AQUA influence LST trends. These findings highlight the necessity of accounting for local time in climate assessments to improve adaptation strategies.

8.1.2 AI for weather forecasting

Participants: Guillaume Couairon, Renu Singh, Anastase Charantonis, Claire Monteleoni.

Weather forecasting plays a vital role in today's society, from agriculture and logistics to predicting the output of renewable energies, and preparing for extreme weather events. Deep learning weather forecasting models trained with the next state prediction objective on ERA5 have shown great success compared to numerical global circulation models. However, for a wide range of applications, being able to provide representative samples from the distribution of possible future weather states is critical. In [1], we propose a methodology to leverage deterministic weather models in the design of probabilistic weather models, leading to improved performance and reduced computing costs. We first introduce ArchesWeather, a transformer-based deterministic model that improves upon Pangu-Weather by removing overrestrictive inductive priors. We then design a probabilistic weather model called ArchesWeatherGen based on flow matching, a modern variant of diffusion models, that is trained to project ArchesWeather's predictions to the distribution of ERA5 weather states. ArchesWeatherGen is a true stochastic emulator of ERA5 and surpasses IFS ENS and NeuralGCM on all WeatherBench headline variables (except for NeuralGCM's geopotential). Our work also aims to democratize the use of deterministic and generative machine learning models in weather forecasting research, with academic computing resources. All models are trained at 1.5° resolution, with a training budget of approximately 9 V100 days for ArchesWeather and 45 V100 days for ArchesWeatherGen. For inference, ArchesWeatherGen generates 15-day weather trajectories at a rate of 1 minute per ensemble member on a A100 GPU card. To make our work fully reproducible, our code and models are open source, including the complete pipeline for data preparation, training, and evaluation, accessible here.

Participants: David Landry, Claire Monteleoni, Anastase Charantonis.

In [11], we propose a machine-learning-based methodology for in situ weather forecast postprocessing that is both spatially coherent and multivariate. Compared with previous work, our Flow MAtching Postprocessing (FMAP) represents the correlation structures of the observation distribution better, while also improving marginal performance at stations. FMAP generates forecasts that are not bound to what is already modeled by the underlying gridded prediction and can infer new correlation structures from data. The resulting model can generate an arbitrary number of forecasts from a limited number of numerical simulations, allowing for low-cost forecasting systems. A single training is sufficient to perform postprocessing at multiple lead times, in contrast with other methods, which use multiple trained networks at generation time. This work details our methodology, including a spatial attention transformer backbone trained within a flow-matching generative modeling framework. FMAP shows promising performance in experiments on the EUMETNET Postprocessing Benchmark (EUPPBench) dataset, forecasting surface temperature and wind-gust values at station locations in western Europe up to five-day lead times.

Participants: Cecile Mallet.

A major issue limiting the successful deployment of deep learning algorithms in geophysical applications is their inability to generalize to new contexts. Regarding the quantitative precipitation estimation (QPE) from the Global Precipitation Mission (GPM) satellite constellation, the GPM Microwave Imager (GMI) contains enough co-located brightness temperatures and rain rates data to train a deep learning inverse model to retrieve precipitation intensity. However, the difference in instrumental configurations makes it impossible to directly apply this inverse operator to another space-borne radiometric imager. A domain adaptation is thus necessary to solve the domain shift problem encountered when applying the model trained on one satellite to another satellite. The paper, [16], tests a method to map the SSMI/S data to the GMI data. In the absence of sufficient paired images between the two satellites, we applied a Cycle consistent Generative Adversarial Network (CycleGAN), which allows for an Unsupervised Domain Adaptation approach. Evaluating the quality of adapted images is a complex problem. This paper employs two tactics: a brief evaluation of adapted radiometric images and a qualitative/quantitative evaluation of rain retrieval. Over several case studies, the results show that the domain adaptation step produces adapted SSMI/S images that retain the majority of the rain structure. Next, the rain detection score and intensity bias are then compared using 847 overpasses. The same analysis is carried out over mainland France by comparing the results with rainfall products supplied by Météo-France. In both comparisons, the adapted images allow the inverse operator to provide a better score in rain detection and intensity.

8.1.3 Subseasonal wind forecasting

Participants: Ganglin Tian, Anastase Alexandre Charantonis.

In [17] to improve the spatial representation of uncertainties when regressing surface wind speeds from large-scale atmospheric predictors for sub-seasonal forecasting. Sub-seasonal forecasting often relies on large-scale atmospheric predictors such as 500 hPa geopotential height (Z500), which exhibit higher predictability than surface variables and can be downscaled to obtain more localised information. Previous work by Tian et al. (2024) demonstrated that stochastic perturbations based on model residuals can improve ensemble dispersion representation in statistical downscaling frameworks, but this method fails to represent spatial correlations and physical consistency adequately. More sophisticated approaches are needed to capture the complex relationships between large-scale predictors and local-scale predictands while maintaining physical consistency. Probabilistic deep learning models offer promising solutions for capturing complex spatial dependencies. This study evaluates three probabilistic methods with distinct uncertainty quantification mechanisms: Quantile Regression Neural Network that directly models distribution quantiles, Variational Autoencoders that leverage latent space sampling, and Diffusion Models that utilise iterative denoising. These models are trained on ERA5 reanalysis data and applied to ECMWF sub-seasonal hindcasts to regress probabilistic wind speed ensembles. Our results show that probabilistic downscaling approaches provide more realistic spatial uncertainty representations compared to simpler stochastic methods, with each probabilistic model offering different strengths in terms of ensemble dispersion, deterministic skill, and physical consistency. These findings establish probabilistic downscaling as an effective enhancement to operational sub-seasonal wind forecasts for renewable energy planning and risk assessment.

8.1.4 Oceanic data interpolation

Participants: Dmitrii Drozdov, Pierre Garcia, Anastase Alexandre Charantonis.

In [4], we propose a novel method for reconstruction of high-resolution Sea Surface Height (SSH) fields from sparse along-track satellite altimetry. We explore the usage of an observation-driven deep-learning method for inpainting, with a focus on diffusion-based generative models for spatial reconstruction and data assimilation. The study includes data preparation from reanalyses and satellite observations, the definition of evaluation metrics relevant to oceanography, and comparison with baselines. We discuss model design choices, uncertainty characterization through stochastic sampling, and limitations in real-world deployment. Effectively, in this work, we develop and validate an observation-driven prior, allowing us to sample from the ground-truth distribution of SSH. By not relying on simulation results for training, we propose a step towards observationdriven Deep-Learning analysis of SSH and its uncertainties at small scales

8.1.5 Oceanic data forecasting and assimilation

Participants: Anastase Alexandre Charantonis.

Abstract Sea Surface Height Anomaly (SLA) is a signature of the mesoscale dynamics of the upper ocean. Sea surface temperature (SST) is driven by these dynamics and can be used to improve the spatial interpolation of SLA fields. In [13] we focused on the temporal evolution of SLA fields. We explored the capacity of deep learning (DL) methods to predict short-term SLA fields using SST fields. We used simulated daily SLA and SST data from the Mercator Global Analysis and Forecasting System, with a resolution of (1/12)° in the North Atlantic Ocean (26.5-44.42°N, -64.25-41.83°E), covering the period from 1993 to 2019. Using a slightly modified image-to-image convolutional DL architecture, we demonstrated that SST is a relevant variable for controlling the SLA prediction. With a learning process inspired by the teaching-forcing method, we managed to improve the SLA forecast at 5 days by using the SST fields as additional information. We obtained predictions of 12 cm (20 cm) error of SLA evolution for scales smaller than mesoscales and at time scales of 5 days (20 days) respectively. Moreover, the information provided by the SST allows us to limit the SLA error to 16 cm at 20 days when learning the trajectory.

Participants: Pierre Garcia, Anastase Alexandre Charantonis.

In [5], we explore the capacity of recent diffusion and flow matching techniques for data assimilation of oceanic, sparsely observed fields. Providing regular and physically consistent predictions of the ocean state is critical for numerous scientific, operational, and societal needs. Observations of the ocean surface are gathered through various remote sensing and in situ instruments, and are typically assimilated into numerical models to reconstruct the ocean state. However, this often involves millions of data points, making it computationally intensive, which suggests deep learning may be a cheaper alternative. Deterministic data-driven approaches typically learn about ocean dynamics from numerical simulations or sparse observational data. However, such methods often lack physical realism in uncertain settings. Due to mode averaging, they produce non-physical or overly simplified states. Generative models offer a promising approach to generating physically realistic ocean states. We present GloFM: a Glorys Flow-Matching emulator for spatio-temporal ocean data assimilation. Our generative model produces coherent estimates of ocean surface fields. GloFM uses flow matching to assimilate observational data for nowcasting of surface currents, sea surface height (SSH), and sea surface temperature (SST). Compared to deterministic regression-based approaches, GloFM demonstrates improved realism metrics, capturing finer-scale variability and more physically plausible ocean states.

8.2 Climate change mitigation

8.2.1 New AI tools for energy forecasting

Participants: Julie Keisler.

Current technologies only allow storage by expensive and inefficient means, which makes it difficult to store electricity on a large scale. For the grid to function properly, electricity fed into the grid must match electricity used at all times. Historically, and still today, production resources are planned in advance of demand to maintain this balance. It is therefore crucial to forecast electricity consumption as accurately as possible. The integration of renewable energies, whose production is intermittent and dependent on weather conditions, is making the balance increasingly unstable. Managing this is becoming more complex, making forecasting wind and photovoltaic production now essential. Statistical learning models are used to make consumption and production forecasts. These models take past values and data from explanatory variables and use them to model the signal. To build efficient models, one must choose the input variables, the type of model, and its parameters. Given the vast number of signals to be forecasted, it would be beneficial to automate these choices to create competitive models.

Automated Machine Learning (AutoML) is the process of automating the generation of learning models optimized according to the use case. Over the last ten years, numerous AutoML tools have been developed. However, most of them focus on optimizing classification or regression models on tabular data, or on optimizing neural network architectures for image or text processing. These tools are not appropriate for optimizing electricity consumption and production forecasting models. This thesis is a progress towards automating the generation of time series forecasting models required for power system management. [10] focused on developing the DRAGON Python package, which offers a range of tools for specific yet widely used models: neural networks. DRAGON can be used to create flexible search spaces encompassing a wide variety of neural networks by simultaneity optimizing the architecture and the hyperparameters. They are encoded by Directed Acyclic Graphs (DAGs), where the nodes are operations, parameterised by various hyperparameters, and the edges are the connections between these nodes. To navigate these graph-based search spaces and optimize their structures, the package proposes various search algorithms based on meta-heuristics and bandits-approaches. This thesis details how DRAGON is used for electricity consumption and production forecasts, enabling state-of-the-art models to be generated for these two industrial use cases.

Electricity demand forecasting is key to ensuring that supply meets demand lest the grid would blackout. Reliable short-term forecasts may be obtained by combining a Generalized Additive Models (GAM) with a State-Space model, leading to an adaptive (or online) model. A GAM is an over-parameterized linear model defined by a formula and a state-space model involves hyperparameters. Both the formula and adaptation parameters have to be fixed before model training and have a huge impact on the model's predictive performance. In [2], we propose optimizing them using the DRAGON package mentioned above, originally designed for neural architecture search. This work generalizes it for automated online generalized additive model selection by defining an efficient modeling of the search space (namely, the space of the GAM formulae and adaptation parameters). Its application to short-term French electricity demand forecasting demonstrates the relevance of the approach

8.3 Projecting long-term climate change impacts

8.3.1 Climate model emulation

Participants: Graham Clyne, Guillaume Couairon, Claire Monteleoni, Anastase Charantonis.

Climate projections have uncertainties related to components of the climate system and their interactions. A typical approach to quantifying these uncertainties is to use climate models to create ensembles of repeated simulations under different initial conditions. Due to the complexity of these simulations, generating such ensembles of projections is computationally expensive. In [27], we present ArchesClimate, a deep learning-based climate model emulator that aims to reduce this cost. ArchesClimate is trained on decadal hindcasts of the IPSL-CM6A-LR climate model at a spatial resolution of approximately 2.5x1.25 degrees. We train a flow matching model following ArchesWeatherGen, which we adapt to predict near-term climate. Once trained, the model generates states at a one-month lead time and can be used to auto-regressively emulate climate model simulations of any length. We show that for up to 10 years, these generations are stable and physically consistent. We also show that for several important climate variables, ArchesClimate generates simulations that are interchangeable with the IPSL model. This work suggests that climate model emulators could significantly reduce the cost of climate model simulations.

8.4 Advancing core AI research in Machine Learning and Computer Vision

8.4.1 A novel preconditioning-inspired iterative approach to solving PDEs with neural networks

Participants: Emmanuel de Bézenac.

In [12], physics-informed deep learning often faces optimization challenges due to the complexity of solving partial differential equations (PDEs), which involve exploring large solution spaces, require numerous iterations, and can lead to unstable training. These challenges arise particularly from the ill-conditioning of the optimization problem caused by the differential terms in the loss function. To address these issues, we propose learning a solver, i.e., solving PDEs using a physics-informed iterative algorithm trained on data. Our method learns to condition a gradient descent algorithm that automatically adapts to each PDE instance, significantly accelerating and stabilizing the optimization process and enabling faster convergence of physics-aware models. Furthermore, while traditional physics-informed methods solve for a single PDE instance, our approach extends to parametric PDEs. Specifically, we integrate the physical loss gradient with PDE parameters, allowing our method to solve over a distribution of PDE parameters, including coefficients, initial conditions, and boundary conditions. We demonstrate the effectiveness of our approach through empirical experiments on multiple datasets, comparing both training and test-time optimization performance. The code is available at .

8.4.2 A new model for multi-source data fusion via self-supervised learning

Participants: Clément Dauvilliers, Claire Monteleoni.

In [3], we present a deep learning architecture that reconstructs a source of data at given spatio-temporal coordinates using other sources. The model can be applied to multiple sources in a broad sense: the number of sources may vary between samples, the sources can differ in dimensionality and sizes, and cover distinct geographical areas at irregular time intervals. The network takes as input a set of sources that each include values (e.g., the pixels for two-dimensional sources), spatio-temporal coordinates, and source characteristics. The model is based on the Vision Transformer, but separately embeds the values and coordinates and uses the embedded coordinates as relative positional embedding in the computation of the attention. To limit the cost of computing the attention between many sources, we employ a multi-source factorized attention mechanism, introducing an anchor-points-based cross-source attention block. We name the architecture MoTiF (multi-source transformer via factorized attention). We present a self-supervised setting to train the network, in which one source chosen randomly is masked and the model is tasked to reconstruct it from the other sources. We test this self-supervised task on tropical cyclone (TC) remote-sensing images, ERA5 states, and best-track data. We show that the model is able to perform TC ERA5 fields and wind intensity forecasting from multiple sources, and that using more sources leads to an improvement in forecasting accuracy.

8.4.3 A new generative domain alignment algorithm via shared-structure decomposition

Participants: Julie Keisler, Anastase Charantonis, Claire Monteleoni.

Domain alignment refers broadly to learning correspondences between data distributions from distinct domains. In this work, we focus on a setting where domains share underlying structural patterns despite differences in their specific realizations. The task is particularly challenging in the absence of paired observations, which removes direct supervision across domains. In [28], we introduce a generative framework, called SerpentFlow (SharEd-structuRe decomPosition for gEnerative domaiN adapTation), for unpaired domain alignment. SerpentFlow decomposes data within a latent space into a shared component common to both domains and a domain-specific one. By isolating the shared structure and replacing the domain-specific component with stochastic noise, we construct synthetic training pairs between shared representations and target-domain samples, thereby enabling the use of conditional generative models that are traditionally restricted to paired settings. We apply this approach to super-resolution tasks, where the shared component naturally corresponds to low-frequency content while high-frequency details capture domain-specific variability. The cutoff frequency separating low- and high-frequency components is determined automatically using a classifier-based criterion, ensuring a data-driven and domain-adaptive decomposition. By generating pseudo-pairs that preserve low-frequency structures while injecting stochastic high-frequency realizations, we learn the conditional distribution of the target domain given the shared representation. We implement SerpentFlow using Flow Matching as the generative pipeline, although the framework is compatible with other conditional generative approaches. Experiments on synthetic images, physical process simulations, and a climate downscaling task demonstrate that the method effectively reconstructs high-frequency structures consistent with underlying low-frequency patterns, supporting shared-structure decomposition as an effective strategy for unpaired domain alignment.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

Google DeepMind, Contrat de doctorat privé, Renu Singh, started April 2025

Participants: Renu Singh, Claire Monteleoni.

AXA Research, CIFRE, Nathan Chalumeau, started April 2025

Participants: Nathan Chalumeau, Emmanuel de Bézenac, Claire Monteleoni.

BOWEN, CIFRE, Baptiste Guigal, 2022-2026 (through LATMOS)

Participants: Baptiste Guigal, Laurent Barthès.

AMPHITRITE, CIFRE, Pierre Garcia, started april 2024 (through LiP6)

Participants: Pierre Garcia, Anastase Charantonis.

MERCATOR OCEANS, CIFRE, Gabriela Martinez Balbontin, started march 2025 (through LiP6)

Participants: Gabriela Martinez Balbontin, Anastase Charantonis.

9.2 Bilateral Grants with Industry

Prométhée, France 2030, PRISME, 2024-2027

Participants: Laurent Barthès, Cecile Mallet.

INRIA-EDF Défi, partially funding SRP position of Julie Keisler who started June 2025.

Participants: Julie Keisler, Anastase Charantonis, Claire Monteleoni.

10 Partnerships and cooperations

Participants: All ARCHES.

10.1 International initiatives

10.1.1 Visits of international scientists

Maike Sonnewald

Status Professor

Institution of origin: University of California Davis

Country: USA

Dates: December 18, 2025

Context of the visit: Lecture and meet with team

Mobility program/type of mobility: Lecture

Seyoung Yun

Status Professor

Institution of origin: KAIST

Country: South Korea

Dates: July 8th, 2025

Context of the visit: Lecture and meet with team

Mobility program/type of mobility: ARGO team hosted summer visit

10.2 National initiatives

ANR TSIA submission, led by Charontonis, collaborative with Sorbonne, UVSQ. Status: Pending.

10.3 Public policy support

10.3.1 Invited talks/panels at EU events

Monteleoni: European Central Bank Workshop/Conference on The Transformative Power of AI: Economic Implications and Challenges, Frankfurt, Germany, April 2025

Charantonis: “Securing European Digital Sovereignty: Evaluating Global Dependencies, Risks, and the European Response,” Brussels, October 2025

Monteleoni: EU Science for Preparedness Conference, Turin, Italy, November 2025

Monteleoni: "National Meteorological Services and the EU : provide resilience in a changing climate and fostering European innovation," Brussels, November 2025

11 Dissemination

Participants: All ARCHES.

ARCHES team members have long been committed to building a research community at the intersection of machine learning and the study of climate change and environmental sustainability. They have been working together to do so. Monteleoni co-founded the annual conference on Climate Informatics in New York City in 2011, Charantonis co-chaired its first international event (Paris, 2019), and they both continue to serve on its Steering Committee. Monteleoni and Charantonis also serve as founding editors of Cambridge University Press journal, Environmental Data Science, which we launched in December 2020.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Co-Founder

- Monteleoni: International Conference on Climate Informatics (14th annual event in 2025)

Tutorials Co-Chair

- Monteleoni: ICML 2024, ICML 2025

Steering and Advisory Committees

- Steering Committee, International Conference on Climate Informatics: Charantonis, Monteleoni

- Advisory Board, Green AI Challenge, AI Action Summit, Paris 2025: Monteleoni

11.1.2 Scientific events: selection

Member of conference program committees

- Monteleoni: Senior Area Chair: NeurIPS 2025, ICML 2025
- Monteleoni: Area Chair: AAAI 2026 (work done in 2025)

11.1.3 Journal

Editor in Chief

- Monteleoni, Founding Editor in Chief, Environmental Data Science, Cambridge University Press. 2020-2025

Member of editorial boards

- Charantonis, Editor, Environmental Data Science, Cambridge University Press. 2020-
- Safieddine, Guest Editor, Environmental Data Science, Cambridge University Press. 2025

Reviewer - reviewing activities

- Nature: Charantonis, Monteleoni, Mallet
- Nature Climate Change: Safieddine
- Atmospheric Science: Mallet

11.1.4 Invited talks

Selected Invited Talks

Emmanuel de Bézenac: Scientific Machine Learning: error control and analysis, Besancon, January 15-16 2025

Claire Monteleoni: Keynote, 25th Anniversary of the Bjerknes Centre for Climate Research, Bergen, Norway, March 2025

Claire Monteleoni: College de France, Grand Evênement : L'IA et les mathématiques pour la météorologie et la climatologie, Paris, May 2025

Claire Monteleoni: Keynote, Launch of La Maison de l'IA, Université de Versailles-Saint-Quentin-en-Yvelines, June 2025

Claire Monteleoni: Workshop on AI for the Carbon Cycle, CMCC (Euro-Mediterranean Center on Climate Change), Como, Italy, June 2025

Cécile Mallet, MétéoFrance / Inria joint workshop, Toulouse, October 2025

Emmanuel de Bézenac: Inaugural Conference of PAV-IA, University of Pavia, Italy, October 2025

Claire Monteleoni: ECCE (Expertise Center for Climate Extremes) Seminar, University of Lausanne, October 2025

Claire Monteleoni: Workshop on Uncertainty Quantification for Climate Science, Institut Henri Poincaré, November 2025

Claire Monteleoni: Keynote, EurIPS Rethinking AI workshop, Copenhagen, December 2025

11.1.5 Leadership within the scientific community

Charantonis:

- Lead, SCAI (Sorbonne Center for Artificial Intelligence) / IPSL (Institut Pierre Simon Laplace) masters internship fellowship program
- Bureau, SAMA (Statistics for Analysis, Modelling and Assimilation), IPSL
- Co-organizer, AI4Climate seminars

Monteleoni:

- External Advisory Board, ICCS (Institute of Computing for Climate Science), Cambridge University, 2023-
- U.S. National Science Foundation (NSF) Advisory Committee for Environmental Research and Education, 2021-2025
- Advisory Board, Climate Change AI, 2021-
- Global Partnership on AI (GPAI) Committee on Climate Action & Biodiversity Preservation 2021-

11.1.6 Research administration

Monteleoni: Scientific Committee, IFREMER, 2025-

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

11.2.1 Founding and leadership of degree programs

Barthès and Chazottes (and Co-Founded by Mallet), Head of Master TRIED (ML & Data Processing) Neural Networks /Statistical Analysis of Real Datasets/Applied Artificial Intelligence Univerité UVSQ-Paris-Saclay/ IPP/ CNAM TRIED

11.2.2 Teaching

Julie Keisler, Practical Work: Introduction to Deep Learning, Faculté des Sciences d'Orsay - Université Paris Saclay

11.2.3 Supervision

HDR Sarah Safieddine. February 2025 FR : Interactions entre la Température et la Composition Atmosphérique de la Surface à la Stratosphère ENG: Interactions Between Temperature and Atmospheric Composition from the Surface to the Stratosphere

PhD students All PhD students listed in the first section are supervised or co-supervised by ARCHES team members.

Thesis defenses

- Julie KEISLER Co-supervisor: Claire Monteleoni Université de Lille Discipline: Informatique Janvier 2025 Optimisation de réseaux de neurones : algorithmes et logiciel pour un système électrique durable Automated Deep Learning : algorithms and software for energy sustainability
- Ganglin TIAN Co-supervisor: Anastase Charantonis SORBONNE UNIVERSITÉ Discipline : Sciences de l'Atmosphère Novembre 2025 PRÉVISIONS MÉTÉOROLOGIQUES POUR L'ÉNERGIE AUX ÉCHÉANCES SOUS- SAISONNIÈRES Improving Sub-seasonal Weather Forecasts for Energy

New PhD students in 2025

- Aymeric Delefosse, Inria supervisor: Anastase Charantonis

- Nathan Chalumeau, CIFRE AXA Research, Inria supervisor: Emmanuel de Bezenac
- Renu Singh, Google DeepMind, Inria supervisor: Claire Monteleoni

CSI: Committe de Suivi Individuel Monteleoni:

- Amaury Lancelin, ENS
- Pierre Chapel, ENS

11.2.4 Juries

- Mallet:
  - Rapporteure, Thesis defended by Raul Carreira Rufato "Reconnaissance des arcs de défaut dans les aéronefs par apprentissage artificiel" - École doctorale : Informatique, Télécommunications et Électronique de Paris - Sorbonne université
  - Rapporteure, Thesis defended by Daria Botvynko "Lagrangian trajectories simulation on the sea surface using deep Learning" - École doctorale : Mathématiques et Sciences et technologies de l'Information et de la Communication en Bretagne Océane
  - Rapporteure, Thesis defended by Clément Bazantay "Quantification of ice crystal morphological properties in deep convective cloud systems, based on in-flight observations" Ecole doctorale des Sciences Fondamentales
- Monteleoni:
  - Tenure committee member, Tom Buecler, University of Lausanne, October 2025
  - Rapporteure, HDR, Dennis Wilson, Université Toulouse Capitole, April 2025
  - Examiner, HDR, Pierre Gaillard, Université Grenoble Alpes, July 2025
  - Chair, Thesis defense, Lawrence Stewart, ENS, June 2025
  - Examiner, Ganglin Tian (see Thesis defenses above), November 2025

11.3 Popularization

11.3.1 Productions (articles, videos, podcasts, serious games, ...)

Safieddine: Article in the Conversation France: link

11.3.2 Participation in Live events

Monteleoni: Invited panelist, Global Talent, French Future: Stories of AI Researchers in France. Vivatech, Paris, June 2025

Monteleoni: Invited panelist, L'apport de l'IA dans l'adaptation au changement climatique, round table organized by CEREMA and the Société d'Encouragement pour l'Industrie Nationale, Paris, July 2025

11.3.3 Others science outreach relevant activities

ARCHES research mentioned in the media:

- Le Monde Science & Medecine, Pour une convention nationale de l'éthique environnementale de la recherche. 29 Jan 2025 P.7
- L'Info Durable, L'IA peut-elle vraiment aider à lutter contre le réchauffement climatique ?. 12 Feb 2025
- CHOISEUL MAGAZINE L'IA, le levier qui accélérera la transition environnementale Edition : Winter 2025 P.24-27

12 Scientific production

12.1 Major publications

[1] G. Couairon, R. Singh, A. Charantonis, C. Lessig and C. Monteleoni. *ArchesWeather & ArchesWeatherGen: a deterministic and generative model for efficient ML weather forecasting*. 2024. DOI: 10.48550/arXiv.2412.12971. URL: https://inria.hal.science/hal-05473722 (cit. on p. 11).

[2] K. Das, J. Keisler, M. Brégère and A. Durand. *AutoML algorithms for online generalized additive model selection: application to electricity demand forecasting*. 4th Apr. 2025. URL: https://hal.science/hal-05020138 (cit. on p. 15).

[3] C. Dauvilliers and C. Monteleoni. 'MoTiF: a self-supervised model for multi-source forecasting with application to tropical cyclones'. In: *Environmental Data Science* 4 (23rd July 2025), e36. DOI: 10.1017/eds.2025.10014. URL: https://hal.science/hal-05447408 (cit. on p. 16).

[4] D. Drozdov, P. Garcia, D. Béréziat and A. A. Charantonis. 'Inpainting of sparse tracks image satellite using Plug and Play and learned prior'. In: VISAPP 2026 - International Conference on Computer Vision Theory and Applications. Marbella, Spain, 9th Mar. 2026. URL: https://hal.sorbonne-universite.fr/hal-05450868 (cit. on p. 13).

[5] P. Garcia, T. Archambault, D. Béréziat and A. Charantonis. 'GloFM: a GLORYS Flow-Matching emulator for spatio-temporal ocean data assimilation'. In: VISAPP 2026 - 21st International Conference on Computer Vision Theory and Applications. Marbella, Spain, 9th Mar. 2026. URL: https://hal.sorbonne-universite.fr/hal-05450893 (cit. on p. 14).

[6] P. Garcia, I. Larroche, A. Pesnec, H. Bull, T. Archambault, E. Moschos, A. Stegner, A. Charantonis and D. Béréziat. 'ORCAst: Operational High-Resolution Current Forecasts'. In: *Artificial Intelligence for the Earth Systems* (2025). URL: https://hal.sorbonne-universite.fr/hal-05052411. In press.

[7] L. Gelbart, L. Barthès, F. Mercier-Tigrine, A. Chazottes and C. Mallet. 'Enhanced quantitative precipitation estimation through the opportunistic use of Ku TV-SAT links via a dual-channel procedure'. In: *Atmospheric Measurement Techniques* 18.2 (21st Jan. 2025), pp. 351–370. DOI: 10.5194/amt-18-351-2025. URL: https://insu.hal.science/insu-04622049.

[8] M. George, C. Clerbaux, J. Hadji-Lazaro, S. Safieddine, S. Whitburn, S. Sinnathamby, D. Hurtmans, P.-F. Coheur, H. M. Worden, C. Vigouroux, B. Langerock and S. Compernolle. 'Development of a Merged CO Climate Data Record from IASI and MOPITT Observations'. In: ESA Living Planet Symposium 2025. Vienna, Austria, 23rd June 2025. URL: https://hal.science/hal-05281708 (cit. on p. 11).

[9] N. Gourgue, O. Boucher and L. Barthès. 'A dataset of annotated ground-based images for the development of contrail detection algorithms'. In: *Data in Brief* 59 (Apr. 2025), p. 111364. DOI: 10.1016/j.dib.2025.111364. URL: https://hal.science/hal-04938966 (cit. on p. 11).

[10] J. Keisler. 'Automated Deep Learning : algorithms and software for energy sustainability'. Université de Lille, 24th Jan. 2025. URL: https://theses.hal.science/tel-05097588 (cit. on p. 14).

[11] D. Landry, C. Monteleoni and A. Charantonis. 'Generating ensembles of spatially coherent in situ forecasts using flow matching'. In: *Quarterly Journal of the Royal Meteorological Society* (26th Oct. 2025). DOI: 10.1002/qj.70055. URL: https://hal.science/hal-05459405 (cit. on p. 12).

[12] L. Le Boudec, E. de Bézenac, L. Serrano, R. D. Regueiro-Espino, Y. Yin and P. Gallinari. 'Learning a neural solver for parametric PDEs to enhance physics-informed methods'. In: ICLR 2025 - Thirteenth International Conference on Learning Representations. Singapour, Singapore, 11th Feb. 2025. URL: https://hal.sorbonne-universite.fr/hal-05093943 (cit. on p. 15).

[13] L. Ollier, S. Thiria, C. Mejia, M. Crépon and A. A. Charantonis. 'Neural network approaches for sea surface height predictability using sea surface temperature'. In: *Environmental Data Science* 3 (2nd Jan. 2025), e42. DOI: 10.1017/eds.2024.33. URL: https://hal.science/hal-04943709 (cit. on p. 13).

[14] S. Safieddine, C. Clerbaux, J. Muñoz-Sabater and J.-N. Thépaut. 'Local hourly trends in near-surface and land surface temperatures'. In: *Scientific Reports* 15.1 (2025), Article number: 29915. DOI: 10.1038/s41598-025-15731-0. URL: https://hal.science/hal-05237167 (cit. on p. 11).

[15] S. Safieddine, C. Clerbaux, S. Whitburn and M. Doutriaux-Boucher. 'Earth's skin temperature: the underrated climate variable'. In: EUMETSAT Meteorological Satellite conference 2025. Lyon, France, 15th Sept. 2025. URL: https://hal.science/hal-05308359.

[16] V. Sambath, N. Dubois-Quilici, N. Viltard, A. Martini and C. Mallet. 'Unsupervised Domain Adaptation to Mitigate Out-of-Distribution Problem of Spatial Radiometer Images: Application to Quantitative Precipitation Estimation'. In: *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), p. 5301414. DOI: 10.1109/TGRS.2024.3403373. URL: https://insu.hal.science/insu-04592155 (cit. on p. 12).

[17] G. Tian, A. A. Charantonis, C. L. Coz, A. Tantet and R. Plougonven. 'Quantile Regression, Variational Autoencoders, and Diffusion Models for Uncertainty Quantification: A Spatial Analysis of Sub-seasonal Wind Speed Prediction'. In: *2018 IEEE 14th International Conference on e-Science (e-Science)* (2025), pp. 415–422. DOI: 10.48550/arXiv.2510.16958. URL: https://hal.science/hal-05471392 (cit. on p. 13).

[18] M. Zouzoua, S. Bastin, F. Lohou, M. Lothon, M. Chiriaco, M. Jome, C. Mallet, L. Barthes and G. Canut. 'Using a data-driven statistical model to better evaluate surface turbulent heat fluxes in weather and climate numerical models: a demonstration study'. In: *Geoscientific Model Development* 18.11 (2nd June 2025), pp. 3211–3239. DOI: 10.5194/gmd-18-3211-2025. URL: https://insu.hal.science/insu-05141829.

12.2 Publications of the year

International journals

[19] C. Dauvilliers and C. Monteleoni. 'MoTiF: a self-supervised model for multi-source forecasting with application to tropical cyclones'. In: *Environmental Data Science* 4 (23rd July 2025), e36. DOI: 10.1017/eds.2025.10014. URL: https://hal.science/hal-05447408.

[20] D. Landry, C. Monteleoni and A. A. Charantonis. 'Generating ensembles of spatially coherent in situ forecasts using flow matching'. In: *Quarterly Journal of the Royal Meteorological Society* (26th Oct. 2025). DOI: 10.1002/qj.70055. URL: https://hal.science/hal-05459405.

[21] G. Tian, A. A. Charantonis, C. L. Coz, A. Tantet and R. Plougonven. 'Quantile Regression, Variational Autoencoders, and Diffusion Models for Uncertainty Quantification: A Spatial Analysis of Sub-seasonal Wind Speed Prediction'. In: *Monthly Weather Review* (2025), pp. 415–422. URL: https://hal.science/hal-05471392.

[22] G. Tian, C. L. Coz, A. Tantet, A. Charantonis, N. Goutham and R. Plougonven. 'Improving Subseasonal Wind Speed Forecasts in Europe with a Nonlinear Model'. In: *Monthly Weather Review* 153.9 (Sept. 2025), pp. 1761–1779. DOI: 10.1175/MWR-D-24-0248.1. URL: https://hal.science/hal-05471373.

[23] M. Zouzoua, S. Bastin, F. Lohou, M. Lothon, M. Chiriaco, M. Jome, C. Mallet, L. Barthes and G. Canut. 'Using a data-driven statistical model to better evaluate surface turbulent heat fluxes in weather and climate numerical models: a demonstration study'. In: *Geoscientific Model Development* 18.11 (2nd June 2025), pp. 3211–3239. DOI: 10.5194/gmd-18-3211-2025. URL: https://insu.hal.science/insu-05141829.

Conferences without proceedings

[24] M. Meignin, N. Viltard, L. Barthes and C. Mallet. 'Refining Infrared-Only Rainfall Estimation with Deep Learning'. In: Climate Informatics 2025. Rio de Janeiro, Brazil, 28th Apr. 2025. URL: https://hal.science/hal-05370381.

[25] N. Viltard, V. Sambath, A. Martini, L. Barthès and C. Mallet. ‘Evolution of global rain intensities over the TRMM and GPM era using a deep-learning algorithm to assess the impact of climate change on Earth water cycle’. In: Climate Informatics 2025. Rio de Janeiro, Brazil, 28th Apr. 2025. URL: https://hal.science/hal-05280703.

Edition (books, proceedings, special issue of a journal)

[26] *AutoML algorithms for online generalized additive model selection: application to electricity demand forecasting*. AutoML 2025: International Conference on Automated Machine Learning. Vol. 293. New York (NY), United States: PMLR, 11th Nov. 2025, pp. 23/1–19. URL: https://hal.science/hal-05020138.

Reports & preprints

[27] G. Clyne, G. Couairon, G. Gastineau, C. Monteleoni and A. Charantonis. *ArchesClimate: Probabilistic Decadal Ensemble Generation With Flow Matching*. 2026. URL: https://inria.hal.science/hal-05461931 (cit. on p. 15).

[28] J. Keisler, A. A. Charantonis, Y. Goude, B. Oueslati and C. Monteleoni. *SerpentFlow: Generative Unpaired Domain Alignment via Shared-Structure Decomposition*. 19th Dec. 2025. URL: https://hal.science/hal-05425230 (cit. on p. 16).