Weather4cast at NeurIPS 2025 – Multi-task Challenges for Weather and Pollution Pattern Prediction on the Road to Hi-Res Foundation Models

Aleksandra Gruca* ALEKSANDRA.GRUCA@POLSL.PL
Silesian University of Technology, Poland

Xavier Calbet XCALBETA@AEMET.ES
Pilar Rípodas PRIPODASA@AEMET.ES
Llorenç Lliso JLLISOV@AEMET.ES
Agencia Estatal de Meteorología, AEMET, Spain

Federico Serva FEDERICO.SERVA@TERRARUM.EU
CNR-ISMAR, Italy

Levi Harris LEVLEVI@CS.UNC.EDUU
Tianlong Chen TIANLONG@CS.UNC.EDU
UNC Chapel Hill, USA

Luciano Araujo Dourado Filho LUCIANOADFILHO@GMAIL.COM
Almir Moreira da Silva Neto ALMIRNETO338@GMAIL.COM
Advanced Data Analysis and Management, University of Feira de Santana, Brazil
National Institute of Metrology, Quality and Technology, Brazil

Anthony Miyaguchi ACMIYAGUCHI@GMAIL.COM
Georgia Institute of Technology, USA

Rodrigo Pereira David RDAVIDRD@GMAIL.COM
National Institute of Metrology, Quality and Technology, Brazil

Rodrigo Tripodi Calumby RTCALUMBY@UEFS.BR
Advanced Data Analysis and Management, University of Feira de Santana, Brazil

Lukas Picek LUKASPICEK@GMAIL.COM
Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
University of West Bohemia in Pilsen, Pilsen, Czechia

Kaushik Gopalan KAUSHIK.GOPALAN@FLAME.EDU.IN
Centre for Interdisciplinary Artificial Intelligence and School of Computing and Data Sciences FLAME University, Pune, Maharashtra, India

Atharva Deshpande ATHARVA.A.DESHPANDE@FLAME.EDU.IN
Jeet Shah JEET.C.SHAH@FLAME.EDU.IN
Hrishikesh Simu HRISHIKESH.SIMU@FLAME.EDU.IN
Ansh Kushwaha ANSH.KUSHWAHA@FLAME.EDU.IN
Anushree Bhuskute ANUSHREE.BHUSKUTE@FLAME.EDU.IN
Centre for Interdisciplinary Artificial Intelligence, FLAME University, Pune, Maharashtra, India

David P. Kreil* W4C25@KREIL.ORG
Vienna, Austria

Editors: Tao Qin, Kun Zhang, Jes Frellsen

* Corresponding Author

© 2026 A. Gruca *et al.*

Abstract

The *Weather4cast* competition series aims to advance AI and machine learning algorithms through a highly topical interdisciplinary challenge (www.weather4cast.net). The core task involves predicting high-resolution rain radar movies from multi-band satellite sensors, which requires the application of data fusion across complementary signal sources, multi-channel video frame prediction, and super-resolution techniques. To evaluate the models' ability to extract relevant mechanistic patterns, the competition incorporates spatio-temporal shifts, evaluating predictions both within and across diverse Earth regions with varying climates and precipitation distributions.

Now, in its fifth year, *Weather4cast* moves to improve forecasts world-wide on an expansive data set with over a magnitude more hi-res rain radar data, allowing a move towards Foundation Models through multi-modality, multi-scale, multi-task challenges. Accurate rain predictions are becoming ever more critical for everyone, with climate change increasing the frequency of extreme precipitation events. Notably, the new models and insights will have a particular impact for the many regions on Earth where costly weather radar data are not available. As a complementary application-specific forecasting endpoint, in 2025, for the first time, we added a pollution forecasting challenge.

1. Introduction

The *Weather4cast* competition (www.weather4cast.net) seeks to improve worldwide forecasts by providing an expansive dataset containing over a magnitude more high-resolution rain radar data, in the context of varying specific competition challenges. It facilitates the transition toward Foundation Models by introducing multi-modality, multi-scale, and multi-task machine learning challenges. With climate change increasing the frequency of extreme precipitation events, accurate rain predictions are becoming increasingly critical. Furthermore, starting from last year, besides keeping data fusion and super-resolution tasks, on the road to flexible foundation models, the competition moves from basic precipitation prediction to testing generalization performance and emergent capabilities of probabilistic models. In particular, *Weather4cast* 2025 challenges included the predictions of rare extreme precipitation events, cumulative rain relevant to flooding, and pollution.

1.1. Background and Impact

The *Weather4cast* series restates the complex problem of weather forecasting as a generalized movie prediction task, allowing researchers to apply modern computer vision algorithms to exploit spatio-temporal correlations. By predicting high-resolution atmospheric events such as rainfall from lower-resolution satellite inputs, the competition creates a unique real-world domain to demonstrate advances in general computer vision, data fusion, and super-resolution. This year's challenge goes beyond the current state of the art by actively encouraging the development and validation of Foundation Models with emergent capabilities. The competition has a profound socio-economic impact: by training models to predict ground-level weather data from economical, global-scope satellite movies, it provides immense value for weather forecasting in low-income countries where expensive weather radar infrastructure is typically unavailable. Ultimately, *Weather4cast* critically accelerates the development of a new class of models dedicated to understanding and predicting complex geo-spatial processes.

1.2. Related Work

Much attention has recently focused on the power of Foundation Models, first demonstrated on textual data (Zhou et al., 2023; Brown et al., 2020; OpenAI et al., 2024) and images (Radford et al., 2021). Remarkably, these have been shown to learn powerful representations of the underlying latent structure in the data, giving rise to emergent model capabilities (Bommasani et al., 2022), including surprisingly good performance in tasks for which they have not been trained. Foundation models are typically pre-trained on vast amounts of general data employing self-supervised training and fine-tuned and evaluated using multi-task learning. While first foundation models for weather and climate last year showed promise at low resolutions and medium forecasting horizons (trained at 1.4°, hence $> 50-100$ km; forecasts for 1/4 to 14 days), extensions to now-casting and higher resolutions require additional efforts (Mukkavilli et al., 2023). In *Weather4cast*, we examine the roles of both training corpus and tasks. Our satellite training corpus features a more than ten times higher resolution for satellite data (*ca* 4–8 km), with an additional super-resolution challenge down to *ca* 2 km for the rain radar target data. Notably, this is comparable to the data underlying the latest large-scale high-resolution weather models for now-casting in the USA like MetNet-3 (Andrychowicz et al., 2023). With our continent-scale high-resolution rain radar data to augment satellite data, we explore models exploiting multiple data modalities, regional to national scales, and their effects on model generalization. Specifically, models can learn underlying structures from both modalities while allowing predictions based on only one. From an application perspective, predictions from satellite data alone support the aim of making state-of-the-art now-casting also available in less developed economies where radar data are not available.

Recent advances in foundation models for videos have identified means to reduce resource requirements in self-supervised training (Fan et al., 2023), making research on foundation models accessible to a broader scientific community. Notably, we represent weather data as multi-channel *movies*. Emergent capabilities in weather models have received little attention to date (Nguyen et al., 2023; Bodnar et al., 2024). Orienting the *Weather4cast* competitions towards questions supporting the development and benchmarking of foundation models now thus seems timely.

Specifically, we here introduce application endpoint oriented tasks to explore emergent model capabilities, such as the prediction of the likelihood of *extreme precipitation events* of unusually bad weather in the next hours or days, or accurate estimates of *cumulative rainfall* in the context of flood warnings, complementing a pollution prediction challenge.

Consequently, *Weather4cast* leads the way in providing both **unprecedented rich high-resolution weather data** and establishing an **expanding range of application-specific meteorological endpoint-related benchmarking tasks** for emergent properties of potential foundation models to the scientific community at NeurIPS.

2. The Weather4cast Competition

In 2025, we entered the fifth year of the *Weather4cast* weather movie prediction series. In the first *Weather4cast* challenge, movie channels comprised temperature, rainfall, cloud cover, and turbulence information derived from satellite data and thus provided the first high-resolution multi-channel forecasting challenge (Gruca et al., 2021; Herruzo et al., 2021).

The next *Weather4cast* competition focused on predicting 'black & white' hi-res rain radar movies from lo-res multi-channel spectral satellite images, requiring skills in data fusion, super-resolution, and models that generalize under spatio-temporal domain shifts (Gruca et al., 2022). This was extended in 2023, when *Weather4cast* focused on quantitative prediction of rain rates, moving from classification to *zero-inflated regression* tasks, while maintaining the need for data fusion, super-resolution, and generalization under spatio-temporal shifts (Gruca et al., 2026). Starting from *Weather4cast* 2024 we aim to investigate how leading models generalize to domain-specific tasks.

2.1. Competition Data

2.1.1. Core Weather Data

Similarly to the previous competitions, in *Weather4cast* 2024 and 2025 we encourage the community to create new models inspired by the best solutions from the previous years. For that reason the previous *Weather4cast* 2023 dataset (Gruca et al., 2026) has been significantly extended. Starting from 2024 we provided 36× as much hi-res rain radar data as before, covering not just the central areas of interest but the whole context regions, in order to support multi-modality and multi-scale training.

The foundational ground truth for the basic dataset is derived from OPERA (Huuskonen et al., 2014; Saltikoff et al., 2019) radar rainfall rates, utilizing 11-channel EUMETSAT Meteosat Second Generation SEVIRI data (Shcmetz et al., 2002) as the primary inputs. We selected 10 regions across Europe that represented a range of precipitation regimes. In the previous years, the challenge required participants to forecast precipitation levels over an upcoming 4 to 8-hour window, segmented into 16 to 32 intervals of 15 minutes each. These forecasts relied on an input sequence representing the immediate past hour, which was broken down into four 15-minute time slots. The core input data consisted of satellite radiance imagery spanning 11 chanels covering visible and near-infrared (VIS), thermal infrared (IR), and water vapor absorption (WV). Alongside the imagery, static geographical features—specifically latitude, longitude, and elevation—were supplied for each respective region.

Specifically, the inputs utilized satellite images of 252×252 pixel resolution representing the whole region to capture a broad contextual perimeter surrounding a specific 42×42 pixel region of interest. Ultimately, the models were tasked with generating high-resolution OPERA rain rate predictions for this targeted area. The corresponding ground truth for the region of interest was provided as a 252×252 pixel OPERA image. Thus, the final prediction comprised a sequence of up to 32 images representing probabilities of different rain intensities from ground-radar reflectivities. While predictions have the same temporal resolution of 15 minutes, super-resolution techniques need to be employed to obtain a 6× higher spatial resolution. A comprehensive description of region selection, data generation, reprojection of the OPERA data to a geostationary grid, and baseline code is provided in (Gruca et al., 2022) and (Gruca et al., 2026).

In the previous years, for the training phase, the OPERA images has been available only for a specific area of regions of interest. Starting from 2024 we provide the particpants with both low resolution satellite images and high resolution OPERA images for the whole regions. This data can be used to pre-train foundation models on 252×252 images in satellite

resolution, while the corresponding ground truth is represented by the $1,512 \times 1,512$ pixel images in OPERA resolution.

To evaluate the generalization performance and emergent capabilities of probabilistic models, the *Weather4cast* 2024 competition introduces two specialized downstream tasks: a Cumulative Rainfall prediction challenge and a Rare Extreme Weather Event prediction challenge. The foundational ground truth for both datasets is derived from OPERA radar rainfall rates, utilizing 11-channel METEOSAT radiances as the primary inputs.

2.1.2. Cumulative Rainfall

A cumulative rainfall dataset has been prepared for the task of forecasting the total accumulated precipitation in a given small area such as an urban region or around an airport at a particular time. Specifically, 60 targets (20 for each region) of relatively small size (32×32 pixels in OPERA resolution) were chosen according to their high impact on populations, prioritizing major cities and airports. The selected locations span a wide geographical area across Europe (e.g., Pamplona, Helsinki, Belgrade, Warsaw, and Barcelona) to ensure models are tested against a diverse range of regional climatic conditions. The compilation of this dataset followed a rigorous selection process. Case identification was based on the OPERA radar data and for each target location, the closest corresponding OPERA pixels were identified and then verified for inclusion within the training and validation splits. Precipitation time series were constructed for the selected pixels to isolate specific cases making sure the dataset includes situations with significant precipitation accumulation, specifically capturing instances where rain is already falling at the forecast's starting point, as well as scenarios just before rain starts. To maintain a balanced representation, events with little to no precipitation are also included.

2.1.3. Rare Extreme Weather Events

The rare extreme event prediction challenge asks models to predict the specific locations, spatio-temporal extents, and severities of extreme weather. We here describe how we (i) automated the identification of extreme weather events, (ii) tuned parameters of the algorithm for this, and (iii) what descriptors were selected to characterize events.

In the first step we applied *scikit-learn* `scipy.ndimage.find_objects`, a spatio-temporal clustering method that returns bounding boxes of connected binarized features in space and time, allowing spatially extended features to be tracked from their initial appearance to their demise. A binarization threshold of 3 was high enough to identify substantial rainfall and low enough to not fragment larger weather systems. A minimal feature size of 3×3 pixels reduced noise while maintaining high resolution. The output of this initial clustering then needs to be filtered to identify extreme weather events, where the maximum area-averaged rainfall rate exceeded a limit, and with a minimum and a maximum duration of events. For a principled choice of parameters, we employ an expert curated source of severe weather events, the European Severe Weather Database (ESWD).[1] It collects documented reports of severe weather events across Europe, such as heavy rainfall, large hailstorms, and tornadoes (Dotzek et al., 2009). We find that selecting events with a duration of 2 to 48 hours and a maximum area-average rainfall rate exceeding $15\,\mathrm{mm/h/m^2}$ gives a good recall

1. https://eswd.eu/

of ESWD reported events, exceeding 50% for all regions, with a median recall of $> 66\%$. This is also in line with the threshold of the German weather service for notable weather / strong rain in an hour (Starkregen).[2] With this rate threshold, we predict about $40\times$ more events than reported in the ESWD, which is in line with domain estimates of report coverage (AEMET, *pers. comm.*). The matching clusters were, on average, longer in duration, larger in size, and characterized by higher precipitation rates than ESWD clusters. The automated pipeline mitigates inherent limitations of the ESWD, such as that human reports are mainly concentrated in urban areas and that reported event frequency is highly uneven across different countries and over time. Our spatio-temporal clustering algorithm thus is expected to detect many severe events than are not reported in the ESWD due to selection / reporting bias. The observed regional differences in the overlaps between the clusters that we detected and the events reported in the ESWD can thus be understood as a result of both genuine regional weather differences as well as regional variation in reporting. Moderate improvements in recall from less strict thresholds would have come at the expense of unrealistically much higher cluster numbers.

2.1.4. Pollution Prediction Challenge

For a new out-of-domain pollution prediction task introduced in 2025, following Bodnar et al. (2024), fine-tuning can use about 1/2 TB of data by the Copernicus Atmosphere Monitoring Service (CAMS) dataset (ECMWF, 2023). The original analysis data has a $0.4^\circ \times 1\,\mathrm{h}$ resolution, and an additional 4 TB of data are available at a lower $0.75^\circ \times 3\,\mathrm{h}$ resolution. The dataset includes over 50 chemicals such as ozone, NO_2, CO, *etc.*, 7 aerosol types (dust, sea salt, organic matter, black carbon, sulphate, nitrate, ammonium), and additional meteorological variables. This can be related to the core *Weather4cast* data, as the total column values of pollutants (the sum of the pollutants over all vertical layers) reflect the summarial satellite sensor signals.

2.2. Competition Tasks

The prediction tasks correspond directly to real operational forecasting problems in meteorological agencies. While state-of-the-art models provide suitable short-term predictions where they have been trained, they have not been developed with a focus on generalization outside the training domain, which is essential both for covering rare extreme weather events, dealing with shifts due to climate change, and bringing high-quality now-casting also to economically less strongly developed communities, where radar data are typically not available.

Prior work by Google Research and DeepMind has demonstrated that Convolutional Neural Networks performed well for the first 4-6h at this task (Sønderby et al., 2020; Espeholt et al., 2022; Andrychowicz et al., 2023; Ravuri et al., 2021). However, the high-resolution data used to train these models has never been made available, and the model performance exponentially decreases when making predictions beyond 4–6 hours, and short-term prediction ('now-casting') remains a challenging task. Conversely, different applications have different requirements regarding the accuracy of forecasts in terms of location,

2. https://www.dwd.de/DE/service/lexikon/begriffe/S/Starkregen.html

time, and precipitation intensity. Our selection of diverse pre-training and down-stream tasks reflects this application heterogeneity.

2.2.1. Pre-training and fine-tuning for multi-modality, multi-scale, multi-task learning

The pre-training challenge tasks are to predict hi-res precipitation radar images up to 8 h (32×15 mins) into the future. Forecasts can be fine-tuned to explore ranges of 1 h to 8 h, where short-term predictions until 4 h are considered easier than longer term projections. Input data is provided for 1 h before the prediction window (4×15 mins). Input data can be from one or several modalities: OPERA hi-res rain radar data, 11-band spectral satellite image data, and static data. The multiple information sources are complementary. There was no leaderboard for the pre-training challenge as the main reason for the participants is to use this data to train foundation models which then can be fine-tuned for the specific downstream tasks.

2.2.2. Cumulative Rainfall Prediction Challenge

In this task task, participants were asked to pre-train a generic model, which is then fine-tuned to this specific challenge of the prediction of cumulative rainfall in a given small area like an urban region or around an airport at a particular time. The given input sequence consists of the 1 hour (4 time slots) of 11-channel multi-spectral satellite images for a whole-region (252×252 pixels). The ground truth for a predicted cumulative rainfall is averaged over a 32×32 pixel area of hi-res radar rain rates and then aggregated over 4h. As rain rates are per hour and there are 4 slots per hour, the values are averaged per hour and then summed over 4 hours.

2.2.3. Rare Extreme Weather Event Prediction Challenge

The second *Weather4cast* downstream task challenges participants to pre-train a generic model and finetune it for the prediction of high-precipitation events starting within a 4-hour window. Similarly like in the cumulative rainfall challenge, the given input sequence consists of the 1 hour (4 time slots) 11-channel multi-spectral satellite data for a whole-region (252×252 pixels). Based on the input sequences the participants were tasked with predicting five distinct weather events that will start in the subsequent four hours. As it is possible that in the given time slot there are more than 5 reference regions, the predictions are scored against the best matching references. To characterize these events, participants had to forecast several specific variables: maximum precipitation (as mm in per hour), event duration, event location and event size which is represented as a diagonal of the event at its mid time. All those variables must be predicted on the natural log scale. Additionally, models must predict the linear-scale x and y coordinates of the event's center at mid-time.

2.2.4. Pollution Prediction Challenge

As a complementary task to the precipitation prediction challenges, *Weather4cast* 2025 added a new challenge for forecasting atmospheric pollution. Since this challenge is based on the publicly available CAMS dataset, there is no hidden test data, and therefore we did

not introduce a public leaderboard. Instead, participants were asked to show improvements over the Aurora Baseline (Bodnar et al., 2024) using a scoring method of their choice and to share their work by submitting a scientific paper to the conference.

2.3. Metric

The competition uses the Continuous Ranked Probability Score (CRPS) (Zamo and Naveau, 2018) to evaluate predictions, a metric that supports both probabilistic forecasts and point estimates. For the cumulative rainfall prediction task, submissions are evaluated based on empirical cumulative distribution functions (eCDF) provided for each case, where participants must map cumulative rain rates to cumulative probabilities.

For the rare extreme event prediction, the CRPS reduces to the absolute error when calculating point estimates. The metric applies different error calculations depending on the specific variable being predicted. For log-scale variables which are: maximum precipitation (`maxPrec.ln`), duration (`duration.ln`), start time (`start.offset.ln`) and event size (`mid.diag.ln`), the error is calculated as the predicted value (`.pred`) minus the true value (`.true`) on the log scale, providing an unbiased symmetrical error. For the linear-scale variables which are event center coordinates (`mid.x` and `mid.y`), the absolute error is scaled by the true size of the event. After these errors are determined, the assessment calculates the Root Mean Square Error (RMSE) for a single event using the following specific formula:

$$\left[\frac{1}{6}\left((\text{maxPrec.ln.pred} - \text{maxPrec.ln.true})^2 + (\text{duration.ln.pred} - \text{duration.ln.true})^2 + (\text{start.offset.ln.pred} - \text{start.offset.ln.true})^2 + (\text{mid.diag.ln.pred} - \text{mid.diag.ln.true})^2 + \left(\frac{\text{mid.x.pred} - \text{mid.x.true}}{\text{mid.diag.true}}\right)^2 + \left(\frac{\text{mid.y.pred} - \text{mid.y.true}}{\text{mid.diag.true}}\right)^2\right)\right]^{1/2} \tag{1}$$

The final competition score for the rare extreme event prediction challenge is then computed by averaging this RMSE across all events, regions, and years.

Notably, the best model submissions made probabilistic forecasts.

3. Results, Model Highlights, and Insights

This year, with 450 registered participants resutling, we scored 550 Cumulative Rainfall prediction challenge submissions from 20 active teams, with 15 teams outperforming the baseline, a 10× improvement over the previous year, showing the accumulating advances facilitated by annually recurring competition series that these NeurIPS Competition Track Proceedings articles document and make accessible to the community on publication. In contrast, we scored fewer submissions (40) for the Rare Extreme Weather Event prediction challenge, with just a single team beating the baseline, reflecting the difficulty of this challenge in its first year.

This section presents results, model highlights, and insights from the work of a selection of winning teams.

3.1. Conditional GAN with Optical Flow

The authors proposed a two-part pipeline designed to forecast cumulative rainfall from Spinning Enhanced Visible and Infrared Imager (SEVIRI) infrared data by linking dense optical flow with a conditional Generative Adversarial Network (cGAN). During the initial preprocessing phase, the framework isolates cloudy regions using Otsu thresholding and averages four highly correlated infrared channels to reduce model complexity. Next, a Lucas-Kanade dense optical flow algorithm estimates cloud system movements to forecast satellite radiance images up to four hours into the future. Finally, a U-Net-inspired cGAN is applied to translate these predicted radiance sequences into spatial precipitation rate maps, optimizing via adversarial, pixel-wise, and perceptual losses to output cumulative rainfall at the SEVIRI spatial resolution (Deshpande et al., 2024).

3.2. SaTformer

In this work, the authors modified a transformer-based architecture developed for video understanding for precipitation nowcasting (Bertasius et al., 2021; Harris and Chen, 2026). This was achieved by discretizing the target label space into non-overlapping bins and training via a frequency-weighted cross-entropy loss. Equivalently, SaTformer predicts an empirical distribution of rainfall rates given an HRIT sequence as input. Despite its simplicity, this approach generated well-calibrated density estimates. The authors note, however, that modifications will be required to extend their architecture to larger spatial inputs or longer temporal sequences. Such changes may include implementing a sparse attention scheme (e.g., space-time interleaved) or reducing parameter count by removing transformer blocks.

3.3. Probabilistic Forecasting via Frozen DINOv3 Encoders

To effectively balance computational efficiency with accurate probabilistic nowcasting, this framework (Filho et al., 2025) integrates a frozen DINOv3 (Siméoni et al., 2025) ViT-L/16 foundation model pre-trained on satellite imagery with a trainable V-JEPA (Assran et al., 2025) (Joint-Embedding Predictive Architecture) video projector. The DINOv3 backbone functions as a domain-aligned "world-model" prior, extracting latent patch tokens from the input multispectral sequences. The narrow asymmetrical V-JEPA projector subsequently models the spatiotemporal dependencies among these latent tokens and maps them to a discrete categorical distribution representing a 4-hour rainfall empirical cumulative distribution function (eCDF). This projector is trained end-to-end using a discrete variant of the Continuous Ranked Probability Score (CRPS), focusing the model entirely on the probabilistic readout and aligning the objective strictly with calibration evaluation. This framework also benchmarks against fully learnable 3D-UNet baselines that operate directly on the spatiotemporal satellite stack through either a discrete-CRPS eCDF head or a zero-inflated Gamma (Gamma-Hurdle) probabilistic head (Filho et al., 2025).

3.4. Two-Stage Transfer Learning via ConvGRU

This methodology presents a two-stage transfer learning pipeline combining Convolutional Gated Recurrent Units (ConvGRU) with empirically derived physical transformations to

map satellite infrared data to radar-based precipitation rates. In the first stage, cascaded ConvGRU models are trained across specific staggered temporal offsets to purely forecast future sequences of SEVIRI brightness temperatures, allowing the recurrent network to isolate and learn nonlinear atmospheric dynamics directly in the satellite domain. In the subsequent stage, a calibrated nonlinear transformation translates these predicted cloud-top cooling patterns into OPERA radar rainfall rates. This sequence is then processed through a 3D connected-component labeling algorithm to locate the spatial footprints, duration, and intensity of extreme precipitation events (Bhuskute et al., 2025).

3.5. Lightweight Architectures for Particulate-Matter Nowcasting

Addressing the need for rapid regional air-quality predictions, this approach abandons large-scale foundation models in favor of lightweight, parameter-specific architectures specifically designed for 6-hour particulate matter (PM1, PM2.5, PM10) nowcasting. Utilizing Copernicus Atmosphere Monitoring Service (CAMS) analysis fields restricted to an India-centric domain, the framework tests compact U-Net, ConvGRU, and ConvLSTM models, which possess merely 1 to 4 million parameters. The architectures—particularly the U-Net featuring bilinear upsampling and the recurrent convolutional models integrating specific dropout regularizations—demonstrate that highly specialized networks can achieve low Root Mean Squared Errors (RMSE) and effectively suppress spatial overestimation bias while outperforming the 1.3-billion parameter Aurora foundation model (Kushwaha and Gopalan, 2025).

4. Discussion and Outlook

The last iteration of the *Weather4cast* Competition at NeurIPS highlights progress in the field amidst a number of non-trivial challenges remaining. Emerging Earth, Climate, or Weather Foundation Models raise the promise of universal models that can be adapted easily to diverse sets of applications in the domain (Zhu et al., 2026). This is particularly relevant for situations where curated training data is partly biased or missing in the sufficient quantities for training a dedicated model from scratch. In such situations, and as witnessed in the preparation of severe weather event data for our current competition and discussed in this manuscript, the compilation of suitable data for fine-tuning can in turn be challenging in its own right, albeit at a smaller scale. Having a diverse set of such downstream tasks, however, is critical for a systematic evaluation and further development of foundation models, and this is an area where competitions like *Weather4cast* make a critical contribution to the field.

One maybe unexpected interim outcome of the focus on foundation models this year has been the observation that foundation models are adopted by others less quickly than perhaps hoped, in stark contrast to the high citation counts of these seminal papers. This is both reflected by the models submitted for our competition as well as other published work. Most citations appear either in reviews or in Introduction and Discussion sections for background, rather than from work building on the foundation models or comparing model performances. To date, about a year after publication of the Aurora model parameters on HuggingFace[3], there are just three derivatives on HuggingFace (finetuned for weather

3. https://huggingface.co/microsoft/aurora

forecasting in Europe or ocean modeling, and a distillation of the pollution model), with not even preprints of scientific manuscripts available yet. Similarly, on github, most of the over 300 forks seem to be unchanged copies. That limited uptake may thus not only be due to computational resource demands but also due to the complexity of the models and the fact that alternative dedicated models can often be developed with reasonable ease, and often with matching or better performance than fine-tuned foundation models, as evidenced for pollution prediction in this year's submissions.

Dedicated models, however, perform well on one task. Specifically, one team comparing alternative model architectures found that although both architectures effectively captured local and temporal rainfall dynamics, one architecture (ConvGRU) proved more effective for cumulative rainfall forecasting, where faster convergence and reduced parameter complexity allowed better generalization across regions, while another (ConvLSTM) did better in event prediction, benefiting from its ability to retain long-term dependencies an generalizing well to unseen geographical regions by transfer learning.

Future work to improve foundation models will thus have to focus on further challenging and benchmarking the emerging capabilities of foundation models where end-to-end training from scratch is ineffective due to insufficient amounts of training data, which may include new downstream tasks where zero, one, or few-shot learning is required. The forecasting of extreme events, in fact, remains a domain where there is still ample room for improvement (Zhang et al., 2026).

References

Marcin Andrychowicz, Lasse Espeholt, Di Li, Samier Merchant, Alexander Merose, Fred Zyda, Shreya Agrawal, and Nal Kalchbrenner. Deep learning for day forecasts from sparse observations, 2023.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba, Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-jepa 2: Self-supervised video models enable understanding, prediction and planning, 2025. URL https://arxiv.org/abs/2506.09985.

Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding?, 2021. URL https://arxiv.org/abs/2102.05095.

Anushree Bhuskute, Kaushik Gopalan, and Jeet Shah. Computationally-efficient deep learning models for nowcasting of precipitation: A solution for the weather4cast 2025 challenge, 2025. URL https://arxiv.org/abs/2511.11197.

Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Anna Vaughan, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A. Weyn, Haiyu Dong, Jayesh K. Gupta, Kit Thambiratnam, Alexander T. Archibald, Chun-Chieh Wu, Elizabeth Heider, Max Welling, Richard E. Turner, and Paris Perdikaris. A foundation model for the earth system, 2024. URL https://arxiv.org/abs/2405.13063.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2022.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

Atharva Deshpande, Kaushik Gopalan, Jeet Shah, and Hrishikesh Simu. A conditional generative adversarial network model for the weather4cast 2024 challenge, 2024. URL https://arxiv.org/abs/2412.00451.

Nikolai Dotzek, Pieter Groenemeijer, Bernold Feuerstein, and Alois M. Holzer. Overview of essl's severe convective storms research using the european severe weather database eswd. *Atmospheric Research*, 93(1):575–586, 2009. ISSN 0169-8095. doi: https://doi.org/10.1016/j.atmosres.2008.10.020. URL https://www.sciencedirect.com/science/article/pii/S0169809508003141. 4th European Conference on Severe Storms.

ECMWF. IFS documentation CY48R1 - part VIII: Atmospheric composition. *IFS Documentation CY48R1*, (8), 06/2023 2023. doi: 10.21957/749dc09059.

Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gazen, Rob Carver, Marcin Andrychowicz, Jason Hickey, Aaron Bell, and Nal Kalchbrenner. Deep learning for twelve hour precipitation forecasts. *Nature*

Communications, 13(1):5145–, 2022. ISSN 20411723. doi: 10.1038/s41467-022-32483-x. URL https://doi.org/10.1038/s41467-022-32483-x.

David Fan, Jue Wang, Leo Liao, Yi Zhu, Vimal Bhat, Hector Santos, Rohith Mysore Vijaya Kumar, and Xinyu (Arthur) Li. Motion-guided masking for spatiotemporal representation learning. In *ICCV 2023*, 2023. URL https://www.amazon.science/publications/motion-guided-masking-for-spatiotemporal-representation-learning.

Luciano Araujo Dourado Filho, Almir Moreira da Silva Neto, Anthony Miyaguchi, Rodrigo Pereira David, Rodrigo Tripodi Calumby, and Lukáš Picek. Dinov3 as a frozen encoder for crps-oriented probabilistic rainfall nowcasting, 2025. URL https://arxiv.org/abs/2511.10894.

Aleksandra Gruca, Pedro Herruzo, Pilar Ripodas, Andrzej Kucik, Christian Briese, Michael K. Kopp, Sepp Hochreiter, Pedram Ghamisi, and David P. Kreil. *CDCEO'21 - First Workshop on Complex Data Challenges in Earth Observation*, page 4878–4879. Association for Computing Machinery, New York, NY, USA, 2021.

Aleksandra Gruca, Federico Serva, Llorenç Lliso, Pilar Rípodas, Xavier Calbet, Pedro Herruzo, Jiří Pihrt, Rudolf Raevskyi, Petr Šimánek, Matej Choma, Yang Li, Haiyu Dong, Yury Belousov, Sergey Polezhaev, Brian Pulfer, Minseok Seo, Doyi Kim, Seungheon Shin, Eunbin Kim, Sewoong Ahn, Yeji Choi, Jinyoung Park, Minseok Son, Seungju Cho, Inyoung Lee, Changick Kim, Taehyeon Kim, Shinhwan Kang, Hyeonjeong Shin, Deukryeol Yoon, Seongha Eom, Kijung Shin, Se-Young Yun, Bertrand Le Saux, Michael K Kopp, Sepp Hochreiter, and David P Kreil. Weather4cast at NeurIPS 2022: Super-resolution rain movie prediction under spatio-temporal shifts. In Marco Ciccone, Gustavo Stolovitzky, and Jacob Albrecht, editors, *Proceedings of the NeurIPS 2022 Competitions Track*, volume 220 of *Proceedings of Machine Learning Research*, pages 292–313. PMLR, 28 Nov–09 Dec 2022. URL https://proceedings.mlr.press/v220/gruca23a.html.

Aleksandra Gruca, Xavier Calbet, Pilar Pilar Rípodas, Llorenç Lliso, Federico Serva, Bertrand Le Saux, Lu Han, Xu-Yang Chen, Han-Jia Ye, De-Chuan Zhan, P. Jyoteeshkumar Reddy, Harish Baki, Sandeep Chinta, Pablo Izquierdo Ayala, Akshay Punjabi, Ajitabh Kumar, Rafael Pablos Sarabia, Assent Ira, Joachim Nyborg, Morten Birk, and David P. Kreil. Weather4cast at NeurIPS 2023 — data fusion for quantitative hi-res rain movie prediction under spatio-temporal shifts. *Proceedings of Machine Learning Research*, 2026. submitted.

Levi Harris and Tianlong Chen. A space-time transformer for precipitation nowcasting, 2026. URL https://arxiv.org/abs/2511.11090.

Pedro Herruzo, Aleksandra Gruca, Llorenç Lliso, Xavier Calbet, Pilar Rípodas, Sepp Hochreiter, Michael Kopp, and David P. Kreil. High-resolution multi-channel weather forecasting – First insights on transfer learning from the Weather4cast Competitions 2021. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 5750–5757, December 2021. doi: 10.1109/BigData52589.2021.9672063.

A. Huuskonen, E. Saltikoff, and I. Holleman. The operational weather radar network in europe. *Bulletin of the American Meteorological Society*, 95:897–907, 2014. doi: https://doi.org/10.1175/BAMS-D-12-00216.1.

Ansh Kushwaha and Kaushik Gopalan. A comparison of lightweight deep learning models for particulate-matter nowcasting in the indian subcontinent and surrounding regions, 2025. URL https://arxiv.org/abs/2511.11185.

S. Karthik Mukkavilli, Daniel Salles Civitarese, Johannes Schmude, Johannes Jakubik, Anne Jones, Nam Nguyen, Christopher Phillips, Sujit Roy, Shraddha Singh, Campbell Watson, Raghu Ganti, Hendrik Hamann, Udaysankar Nair, Rahul Ramachandran, and Kommy Weldemariam. Ai foundation models for weather and climate: Applications, design, and implementation, 2023.

Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. ClimaX: A foundation model for weather and climate. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 25904–25938. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/nguyen23a.html.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie

Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. Gpt-4 technical report, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skillful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 2021. doi: 10.1038/s41586-021-03854-z.

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

Johannes Shcmetz, Paolo Pili, Stephan Tjemkes, Johan Kerkmann, Sergio Rota, and Alain Ratier. An introduction to meteosat second generation (msg). *Bulletin of the American Meteorological Society*, 83(7):977–992, 2002.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. Dinov3, 2025. URL https://arxiv.org/abs/2508.10104.

Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv:2003.12140 [physics, stat]*, March 2020. URL http://arxiv.org/abs/2003.12140. arXiv: 2003.12140.

Michaël Zamo and Philippe Naveau. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2):209–234, 2018. ISSN 1874-8953. doi: 10.1007/s11004-017-9709-7. URL https://doi.org/10.1007/s11004-017-9709-7.

Zhongwei Zhang, Erich Fischer, Jakob Zscheischler, and Sebastian Engelke. Physics-based models outperform ai weather forecasts of record-breaking extremes. *Science Advances*, 12(18):eaec1433, 2026. doi: 10.1126/sciadv.aec1433. URL https://www.science.org/doi/abs/10.1126/sciadv.aec1433.

Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu, and Lichao Sun. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt, 2023.

Xiao Xiang Zhu, Zhitong Xiong, Yi Wang, Adam J. Stewart, Konrad Heidler, Yuanyuan Wang, Zhenghang Yuan, Thomas Dujardin, Qingsong Xu, and Yilei Shi. On the foundations of earth foundation models. *Communications Earth & Environment*, 7(1):103, 2026. ISSN 2662-4435. doi: 10.1038/s43247-025-03127-x. URL https://doi.org/10.1038/s43247-025-03127-x.

Appendix

List of the repositories with source code of the described models and trained parameters:

Team Tarheels
SaTformer - https://github.com/leharris3/satformer
Detailed description of the solution: https://arxiv.org/abs/2511.11090v1

Team luciano.dourado
DINOv3 - https://github.com/acmiyaguchi/weather4cast-2025
Detailed description of the solution: https://arxiv.org/abs/2511.10894

Team kaubega
Convolutional Gated Recurrent Units - https://github.com/flame-cai/w4c25_rainfall

Detailed description of the solution: https://arxiv.org/abs/2511.11197

Team kaubega - atharva.deshpande
Conditional GAN - https://github.com/flame-cai/Weather4Cast24_NIPS
Detailed description of the solution: https://arxiv.org/abs/2412.00451

Polution challenge
ConvGRU, ConvLSTM and UNET - https://github.com/flame-cai/w4c25_pollution
Detailed description of the solution: https://arxiv.org/abs/2511.11185