Weather4cast 2024 –
Multi-task Challenges for Rain Movie Prediction on the Road to Hi-Res Foundation Models

Aleksandra Gruca Pilar Rípodas Xavier Calbet Llorenç Lliso
Federico Serva Bertrand Le Saux David P. Kreil Sepp Hochreiter

`w4c24@weather4cast.org`

Abstract

The competition will advance modern algorithms in AI and machine learning through a highly topical interdisciplinary competition challenge: The prediction of hi-res rain radar movies from multi-band satellite sensors requires data fusion of complementary signal sources, multi-channel video frame prediction, as well as super-resolution techniques. To reward models that extract relevant mechanistic patterns reflecting the underlying complex weather systems our evaluation incorporates spatio-temporal shifts: Specifically, algorithms need to forecast several hours of ground-based hi-res precipitation radar from lo-res satellite spectral images in a unique cross-sensor prediction challenge. Models are evaluated within and across regions on Earth with diverse climate and different distributions of heavy precipitation events. Conversely, robustness over time is achieved by testing predictions on data one year after the training period.

Now, in its third year, *Weather4cast* 2024 moves to improve rain forecasts worldwide 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. Join us on www.weather4cast.net!

Keywords

video frame prediction; transfer learning; super resolution; foundation models; data fusion.

1 Competition description

1.1 Background and impact

Our competition seeks to advance algorithms in AI and machine learning for a range of tasks that can be phrased as a generalized video prediction under spatio-temporal shifts, employing data fusion and super-resolution techniques. The challenge is to predict high-resolution rainfall, restating the problem of weather forecasting as a movie prediction task. This allows for the application of modern computer vision algorithms to exploit spatio-temporal correlations. This works surprisingly well, as demonstrated by our NeurIPS competitions on urban traffic forecasts already in 2019[18], subsequently by Google Research for rainfall [1, 27], and now our *Weather4cast* competitions[16, 15].

This year's challenge goes beyond the state of the art by evaluating models for generalization capability under spatio-temporal domain shifts in order to **extract the physically most meaningful patterns and the most robust models**, a core interest of the machine learning community. Specifically, by providing over an order of magnitude more hi-res ground truth data, inputs of several modalities, and a diverse set of training tasks, this year we expressly encourage the development and validation of **Foundation Models** with emergent skills. In this, we are also building on, complementing, and significantly extending the problem setting of our*Weather4cast* NeurIPS competitions in 2022 and 2023 where we first pushed to examine model robustness under temporal and spacial domain shifts [15]. Our competition asks participants to predict hi-res rain radar data reflecting ground level weather, which is typically expensive to obtain, from lo-res satellite movies which cover large economical regions. **This brings value particularly for weather forecasts in low income countries**.

Our *Weather4cast* challenge thus also includes a super-resolution task because the target rain radar observations have a much higher spatial resolution than the satellite spectral videos. Besides this super-resolution challenge, framing the task as video prediction focusing on transfer learning provides a new real-world domain for demonstrating advances in general computer vision algorithms. We thus provide a unique cross-disciplinary opportunity for advancing data driven algorithms for the exploitation of spatial and temporal correlations in multi-sensor weather data. *Weather4cast* can thus critically accelerate the development of a new class of models for understanding and predicting geo-spatial processes in general.

In previous editions of our competition we received over 200 registrations, and 10–30 teams regularly submitted to the leaderboards, with 300–1,600 submissions scored per competition. Most participants' interests were in artificial intelligence [6, 9], reflected by a team from Alibaba Cloud team winning last year's competition [20]. In 2021, however, the winner notably was a meteorological research group [19]. We thus again expect around 50 delegates coming from both general artificial intelligence audience and domain experts to attend the conference in 2024.

1.2 Novelty

This year, we aim to expand on the prominently recurring theme of Transfer Learning in the *Weather4cast* competition series: Much attention has recently focused on the power of Foundation Models, first demonstrated on textual data [29, 8, 22] and images [23]. Remarkably, these have shown to learn powerful representations of the underlying latent structure in the data, giving rise to emergent model abilities [7], including surprisingly good performance on tasks for which 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 [21]. In *Weather4cast*, we can specifically 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 like MetNet-3 [2]. Moreover, with continent-scale high-resolution rain radar data available to augment satellite data, we can 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.

Emergent capabilities in weather models have received little to no attention to date. In *Weather4cast*, we can introduce application endpoint oriented tasks to explore emergent model capabilities, such as prediction of the likelihood of unusually bad weather in the next hours ('extreme precipitation events'), or estimates of cumulative rainfall in the context of flood warnings. Recent advances in foundation models for videos have identified means to reduce resource requirements in self-supervised training [13], making research on foundation models accessible to a broader scientific community. Notably, we represent weather data as multi-channel movies. Orienting the *Weather4cast* competitions towards questions supporting the development and benchmarking of foundation models in 2024 thus seems timely.

Weather4cast will thus lead the way in providing both unprecedented rich high-resolution weather data and establishing meteorological endpoint related benchmarking tasks for emergent properties of potential foundation models to the scientific community at NeurIPS.

In 2024, we enter the third 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. Following the successful meetings at ACM CIKM 2021 [14] and the Weather4cast Special Session at IEEE BigData 2021 [16], the competition at NeurIPS 2022 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 [15]. Despite only having to predict rain/no-rain event labels, the task proved challenging and the best models turned out to incorporate application domain knowledge into transformers or modified U-Nets. At NeurIPS 2023 we 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. Interestingly, the winning team built on and extended the best complex models from another team the year before. Therefore, we now wish to investigate whether consolidated paradigms can still generalize to domain-specific tasks.

This year, besides keeping data fusion and super-resolution tasks, on the road to flexible foundation models, we move from basic precipitation prediction to testing generalization performance and emergent capabilities of probabilistic models.

Video prediction is of course a central discipline of artificial intelligence, and typical workshops and competitions in the field include *Traffic4cast* at NeurIPS 2019/2020/2021, *Precognition: Seeing through the Future* and *Large-scale Video Object Segmentation Challenge* at CVPR 2020-2023 and *Multi-Modal Video Analysis* at ECCV 2020. Interestingly, winning algorithms from computer vision also excel on non-photorealistic images and videos.

Distribution shifts of various types were studied in our *Traffic4cast* competition at NeurIPS 2021 [11], the *Shifts Challenge* at NeurIPS 2021 (including a weather task using tabular data – representing a different data domain than our movies), and *VisDA: Visual Domain Adaptation* at NeurIPS 2021/2022.

Satellite image analysis is an active field of research including on questions of super-resolution and data fusion: *Tackling Climate Change with ML* (NeurIPS 2020–2023, ICLR 2023, ICML 2021), *EarthVision* (CVPR 2020–2023), *MultiEarth* (CVPR 2022–2023), *Perception Beyond the Visible Spectrum* (CVPR 2020–2023), *Agriculture–Vision* (CVPR 2020–2023), *Computer Vision in Plant Phenotyping and Agriculture* (ECCV, 2020, ICCV 2021 & 2023), *Computer Vision in the Ocean* (ICCV 2021, ECCV 2022), *AI for Earth Sciences* (NeurIPs 2020 and ICLR 2020 & 2022). Workshops, however, are **not focused on the specific challenges of modelling weather**, with even the 'climate' workshop including just 1–2 weather relevant contributions.

We thus build on the efforts of these related communities. Our competition will both bring these communities together and provide a rare opportunity to foster interdisciplinary work for the advancement of modern AI algorithms.

1.3 Data

We provide three types of data for the competition, with different types of data as an input and high-resolution rain radar data as target:

- **Meteosat Second Generation SEVIRI data:** 11 spectral bands (8 infrared + 3 visible bands) from the Meteosat Second Generation (MSG) satellites by EUMETSAT [26], transformed by projection and minor noise to permit public distribution.
- **OPERA radar data:** rain rates (mm/h) from ground-based radar variables provided by EUMETNET OPERA; rain rates [17] through AEMET.
- Static maps containing the location (latitude and longitude) and the elevation per pixel.

This year, we provide $36\times$ **as much hi-res rain radar data as in previous years**, covering not just the central areas of interest but the whole context regions, in order to support **multi-modality and multi-scale training**.

We have compiled a unique resource of **(11+2)-channel weather movies** spanning across Europe (see Figure 1 - Right Panel) with $\sim 4 * 10^{11}$ data points for training: 13 bands $\times$ 2 years $\times$ 330 days $\times$ 96 time-bins $\times$ 10 regions $\times$ (252×252) pixels. These cover 90% of days spanning 2 years. The remaining days are withheld to provide **independent test data**. The comprehensive full context rain radar data we provide for the first time in 2024 add another $\sim 10^{12}$ data points for (pre-)training: 2 years $\times$ 330 days $\times$ 96 time-bins $\times$ 10 regions $\times$ (252×252) pixels, for a total of almost 2×10^{12} data points.

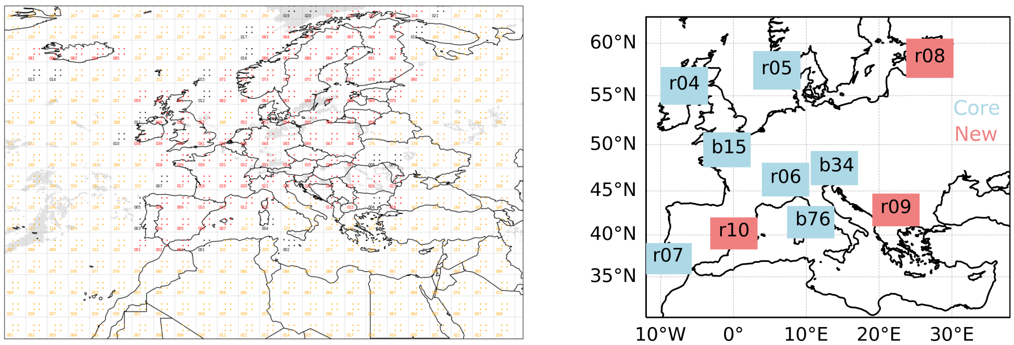


Figure 1: **Regions and data availability.** *Left:* The whole area shown is covered by satellite while we have radar data only for some regions (full coverage = red, partial cover = black, no cover = yellow). Note that radar coverage is absent to limited in less wealthy economies and countries that have not opted into the OPERA program, such as Italy. *Right:* We selected regions for core pre-training and new regions for spatial transfer learning.

The selected regions span across the whole of Europe to cover different precipitation patterns. To select regions we calculated monthly frequencies of rain events according to four rain rates classification (no rain between 0-0.1 mm/hour, low between 0.1-2.5 mm/hour, moderate for 2.5-7 mm/hour and heavy when above 7 mm/hour) from February 2019 until December 2021. Then we examined the number of rain events accumulated monthly during this time for each region focusing on regions with the highest amount of rain events. We noted that rain events belonging to moderate or heavy precipitation rate class are quite rare across the years and their monthly occurrences are highly variable over the year reflecting seasonal precipitation patterns. Finally, for the competition we picked regions as to balance between regions having more events with low rainfall rates and regions having less frequent but more intense rainfall rates.

These regions serve several specific purposes (also see Figure 2):

- *Core regions*: Seven regions where we provide extensive data across 2019 and 2020 for pre-training.
- *Regions for spatial transfer learning*: Test sets from 3 new regions across 2019 and 2020 to test the spatial generalization ability of models. No training data is provided from these regions.
- *Regions for tests under temporal shifts*: Test data from 2021 for the 7+3 regions described above. No training data is provided for 2021. With test data for 2021 from the 7 regions known from training for 2019 and 2020 we examine the temporal generalization ability of models. Test data from the 3 new regions test the joint temporal and spatial generalization performance.

Satellite images are available every 15 minutes. We consider regional areas of interest of size 42×42 pixels in satellite resolution (100–200 km), surrounded by a large context area of size 252×252 pixels ('national' scale, 600–1,200 km). Predictions comprise 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 $6\times$ higher spatial resolution, with a 252×252 pixel image in OPERA rain radar resolution corresponding to the 42×42 pixel areas of interest in satellite resolution. Correspondingly, a full context prediction at rain radar resolution spans $1,512 \times 1,512$ pixels.

The projection for raw OPERA data is Lambert Azimuthal Equal Area with a pixel size of $2km \times 2km$. This projection preserves area with respect to the earth surface. On the other hand, the MSG satellite data are in geostationary projection. In this projection areas corresponding to a pixel are bigger for pixels farther away from the point right under the satellite, increasing from 3 km × 3 km at the equator to polygonal pixels with a side size larger than 24 km over Iceland.

So in addition to forecasting, converting satellite sensor inputs to ground-radar rain rate estimates, this adds a super-resolution task due to the coarser spatial resolution of the satellite data as one pixel in the satellite resolution corresponds to six pixels in OPERA radar resolution. To ease the training of models the radar data has been re-projected to a geostationary grid. With this, both OPERA and MSG data (2D image-like data) match geographically and can also be combined easily in pre-training.

1.4 Tasks and application scenarios

Prior work by Google Research and DeepMind has demonstrated that Convolutional Neural Networks performed well for the first 4-6h at this task [27, 12, 2, 24]. 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, time, and precipitation intensity. Our selection of diverse pre-training and down-stream tasks reflects this application heterogeneity.

1.4.1 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. The eventual application goal is to learn underlying true weather patterns, with an aim of predicting weather without the more expensive radar data.

In the pre-training challenges we provide a baseline and test new models for the accurate prediction of weather, focusing on model generalization performance and prediction quality over time. This includes a **transfer-learning** aspect and a super-resolution challenge, both of which are crucial for transferring insights to the future and other parts of the world. Finally, in contrast to the MetNet studies, by making the data available to the community we establish a state-of-the-art reference benchmark for the development of advanced spatio-temporal forecast models.

The prediction tasks directly correspond 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. Consequently, to focus on generalization performance, predictions are evaluated in regions and/or in years different from the training data, and on different subsets of input modalities (see regions in Fig. 2).

1.4.2 Downstream tasks

To assess how well models develop emerging skills, we introduce complementary downstream tasks (Fig. 2). We will start with two easily introduced application scenarios outlined below and will add further tasks as they become available.

Cumulative rainfall prediction in selected flood risk areas: In many cases, rather than requiring a point by point rain movie, applications actually require a cumulative view of how much rain is likely to come down in a defined area over a given period. Typical examples of that might include cities concerned about the risk of their rain runoff sewer capacities or catchment areas of river basins. In either case, the cumulative rain determines flood risk. We will test the ability of models to predict the probability of cumulative rain amounts in selected areas and time spans.

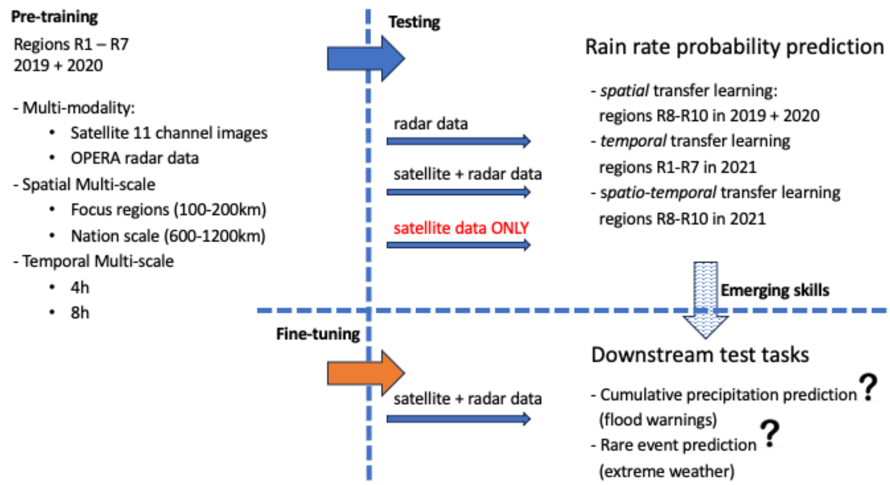


Figure 2: **The schema shows data sources and tasks for pre-training and down-stream tasks.** The success of pre-training can be tested and models fine-tuned for multi-task learning using a range of complementary rain rate probability prediction tasks, covering multiple input modalities and a range of spatial and temporal scales. Emerging skills .

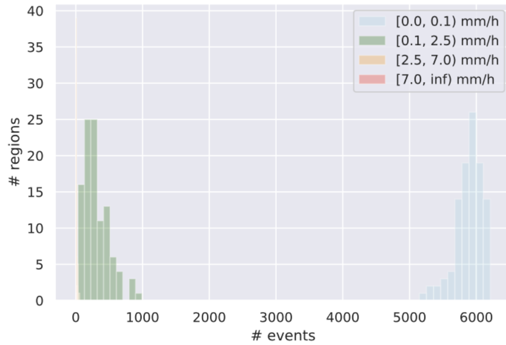
Prediction of rare extreme weather events: Extreme weather events are rare but can be high impact. With recent climate change, these extreme events are getting more frequent, yet remain difficult to predict. We will test the ability of models to predict the probabilities of extreme weather events in selected regions and time spans, focusing on properties of the event (location, start time, duration, intensity).

1.5 Metrics

To support richer models and facilitate homogenization in multi-task learning, this year in general we shift to **probabilistic forecasts**, naturally leading to **negative log-likelihood** losses or metrics. Incidentally, this turns out to feature better properties than other widely used alternatives like the Brier score [5]. For predictions of rain intensities, we ask for the prediction of the probabilities of rain intensity at a particular time and place to fall into a particular value range bin. This suitably deals with the highly skewed value distribution, where most of the time and in most places there is no detectable rain. Models can internally of course generate a richer output representation with a large number of bins, like the MetNet class of models do [27, 12, 2]. To keep the sizes of model output uploads to leaderboards manageable, however, we will assess a small number of meteorologically distinct intensity ranges (see Fig. 3). The actual bins for final leaderboards may be adapted after Stage-1 in response to feedback by participants and meteorological partners. (We note that different agencies seem to use slightly different classes and boundaries).

Accordingly, we can use the same negative log-likelihood losses or metrics for the cumulative rainfall prediction downstream task. Here, we assess the probabilities predicted for cumulative rain amounts over the forecast window at a selected region to fall into a particular value range bin.

In Stage-2, additional downstream tasks will be introduced, such as the prediction of rare extreme weather events. For these, it actually makes sense to expressly acknowledge the different costs of false positives and false negatives, such as through a Value Score curve [28]. We assess the prediction further by assessing the prediction of event properties, including the likely cumulative and peak event intensities, start, and duration (again assessed by negative log-likelihood). Spatial extent prediction can be assessed by 2D-Wasserstein distance in $O(N)$ [4] relative to a gold standard derived by spatio-temporal clustering and domain-specific thresholding from the ground truth data [3].



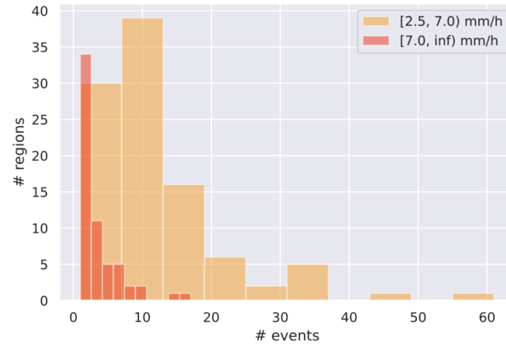


Figure 3: **Radar rain rate distribution for 2019–2021.** *Left:* most of the regions contain about 5,000 to 6,000 events of no or minimal rain (blue) and less than 1,000 events of light rain (green). *Right:* moderate (orange) and high precipitation (red) are found in much fewer events.

1.6 Baseline, code, and material provided

We will again provide baseline models like U-Nets [25] with code including data loaders and documentation on the ***Weather4cast* Github repo** to make it easy for anyone to get started. A beta phase is planned for 1 month prior to the live competition. Scientist at SUT, AEMET, and ESA will test the competition setup internally for feedback and improve the competition baseline.

1.7 Website, tutorial and documentation

The competition website will be hosted at weather4cast.net, which currently shows information about last year's competition. All content for the current competition will be updated by the time we release the data (see Section 2.3). We provide forums for discussion and scientific exchange weather4cast.net/forums/ and can be reached on `w4c24@weather4cast.org`.

2 Organizational aspects

2.1 Protocol

Participants need to register before they can download the competition data hosted. Participants submit their predictions in the Hierarchical Data Format (HDF5), for which there is good library support in many languages including Python. Our website gives detailed submission instructions, and our GitHub repo provides examples. When participants upload their predictions, the system compares them with the ground truth (see Section 1.5 re Metrics). Score are updated hourly for timely feedback.

The multi-task nature of this year's challenge already reduces the risk of model overfitting [10]. In addition we limit the number of concurrent submissions per team, and leaderboard scores are updated hourly, reducing the overall number of submissions that can be tried.

Data access required accepting the competition terms and conditions. Radar measurements are not public but provided specifically for this competition. Ground truth data is kept confidential. Participants need to provide their code and learned weights hosted in a public online repository like GitHub and outline their approach in a detailed scientific abstract.

The competition leaderboards will again be hosted on the competition website as in previous years (see, *e. g.*, Core Leaderboard of *Weather4cast* at NeurIPS 2023).

2.2 Rules and Engagement

The challenge invites researchers from diverse backgrounds, and lowers the technical, economic, and logistic barriers of entry by providing baseline code, tutorials, and lively competition forums, encouraging participation of colleagues new to the field or from underrepresented communities. Extending the data universe from previous years facilitates longer term community building. Since our challenge requires access to extensive computational resources, for participants who do not have access to such compute power, we will reach out to sponsors to provide free cloud GPU grants. For participants struggling to afford travel to the NeurIPS conference in person we will accommodate virtual presentations. Participants must accept the terms & conditions, which include the below rules:

- Participants must submit predictions to the competition leaderboards on time.
- Participants must submit an extended academic abstract describing their work and publish their code and model parameters under the Apache 2.0 open source license. This maximizes impact in the diverse research communities as desired by the competition.
- Winners need to discuss their approach in a scientific presentation at the conference.
- No one with access to the withheld ground truth data can participate in the competition. As a precaution, all employees of the institutions of the organizers are excluded from participating.

2.3 Schedule and readiness

1. **15 May 2024:** Announcement and advertising of the beta-test phase of the competition, community discussion and forum updates.
2. **17 June 2024:** STAGE1 Data Release & Leaderboard open.
3. **15 July 2024:** STAGE2 Data Release & Leaderboard open..
4. **6 October 2024:** Test dataset submission deadline.
5. **9 October 2024:** Invitation to submit conference abstracts and code.
6. **16 October 2024:** Deadline for abstract and code submissions.
7. **25 October 2024:** Acceptance notification.

2.4 Competition promotion and incentives

The competition will be heavily advertised by the main academic organizer of the challenge, AEMET as the data provider, as well as ESA, CNR, and ECMWF(TBC) in their respective academic communities. Various mailing lists and social media channels will be employed.

Apart from the core ML community, due to the interdisciplinary nature of the competition, we will also directly reach out to members of the Earth Observation, Climate and Meteorologist communities, *via* established working relationships with different research groups as well as mailing lists and other channels. Incentives to participate emphasize the academic nature of the competition:

- Post-competition joint publication
- Invitation to presenting at the conference.
- Like in the past where we have hosted leading keynotes by Piotr Mirowski of DeepMind and Nal Kalchbrenner of Google Brain, we will again invite exciting keynote speakers, which will be advertised early to raise interest in the meeting in the wider research community.

3 Resources

3.1 Organizing team

The interdisciplinary committee represents 5 countries, while 20% of the organizers are female. It brings together expertise in machine learning and artificial intelligence, complex heterogeneous data sets, and meteorology. Committee members contribute decades of experience organizing competitions and workshops at leading conferences: *Coordinators*: Gruca, Kreil; *Data providers*: Lliso, Calbet, Rípodas; *Baseline methods*: Gruca, Le Saux; *Beta tests*: Serva, Le Saux, Calbet; *Evaluators*: Kreil, Rípodas, Gruca, Le Saux, Hochreiter.

Aleksandra Gruca, Silesian University of Technology, Poland is a professor of machine learning. Her research focuses on predictive models and the integrated analysis of challenging large-scale heterogeneous datasets. Since 2010 she organizes the annual Symposium of the Polish Bioinformatics

Society. In 2009–2019 she was the organizer of the Int. Conference on Man-Machine Interactions. She also co-organized the *Traffic4cast* Special Sessions at NeurIPS 2020, the *Weather4cast* Special Session at the IEEE 2021 BigData Conference, and *Weather4cast* at NeurIPS 2022 and 2023. She was the organiser of the 1st Workshop of Complex Data Challenges in Earth Observation (*CDCEO 2021*) at ACM CIKM 2021, and 2nd *CDCEO* at IJCAI/ECAI 2022. She was also a co-chair of the IEEE BigData Cup in 2022, and a workshop chair at IEEE BigData 2023.

Llorenç Lliso, Spanish Meteorological Agency (AEMET), Spain a is Physicist who works as a meteorologist at AEMET. His work focuses on the software development and maintenance for satellite data processing, data interoperability, and OGC data diffusion.

Xavier Calbet, AEMET, Spain is the science coordinator for the EUMETSAT Nowcasting Satellite Application Facility (NWC SAF) and AEMET. He has worked with EUMETSAT in Darmstadt specializing in atmospheric profile retrievals obtained from hyper-spectral sounding instruments, such as the Infrared Atmospheric Sounding Interferometer, and planning for the future geostationary hyperspectral sounding mission MTG-IRS. He currently works as a state meteorologist at AEMET.

Pilar Rípodas, AEMET, Spain heads the EUMETSAT NWC SAF for very short range and nowcasting. She played a role in the development of the ICON prediction model at the German national weather service. She organized NWC SAF user workshops and she was an organizer of the European Nowcasting Conference 2019 and *CDCEO 2021* at ACM CIKM, as well as the *Weather4cast* competition at IEEE BigData 2021 and *Weather4cast* at NeurIPS 2022.

Peter Duben, ECMWF, United Kingdom (TBC) is a head of the Earth System Modelling Section at the European Centre for Medium Range Weather Forecasts (ECMWF). Previously he was working as a coordinator of Machine Learning and AI activities at ECMWF. His professional interests include high-resolution weather and climate simulations, high-performance computing for weather and climate models, machine learning for weather and climate predictions. He is also a coordinator of the MAELSTROM EuroHPC Joint Undertaking project which involves seven partners within Europe and a budget of 4,3 million Euro.

Federico Serva, Institute of Marine Sciences (CNR-ISMAR), Italy holds an MSc in Physics from the University of Rome Tor Vergata and a PhD in Environmental Sciences from the University of Naples Parthenope. He currently is a fixed-term research at the Italian National Research Council, and he has been a joint postdoctoral fellow at ESA Φ-lab and the Italian Space Agency. His research focuses on climate models and EO data for process studies, the retrieval of parameters from satellite observations, and the study of extreme events using different data sources.

Bertrand Le Saux, European Space Agency, Italy is a research scientist at the Φ-lab at ESA/ESRIN. His work is at the crossroads of statistics, machine learning, image processing, computer vision, and now quantum computing. He is interested in tackling practical problems that arise in Earth observation, to bring solutions to current environment and population challenges He is an associate editor of the Geoscience and Remote Sensing Letters (GRSL), co-organises the CVPR Earth Vision worskhop series (including this year at CVPR'2023) and co-organises the Humanitarian Assistance and Disaster Response (HADR) workshop at ICCV'2023.

Sepp Hochreiter, IARAI, Linz, Austria is well known pioneer of AI, who identified the vanishing gradient problem, invented long short-term memory (LSTM), as well as many other deep learning algorithms and techniques powering the field. He heads the Linz Institute of Technology AI Lab and is a director of IARAI. His current research interests include deep learning for vision, climate change, smart cities, drug design, and autonomous driving.

David Kreil, Vienna, Austria is a professor at Boku University Vienna. He runs a bioinformatics research group that focuses on analysing, calibrating, and benchmarking genome-scale quantitative assays. Since 2008, he organizes the annual Critical Assessment of Massive Data Analysis (www.camda.info), a well-recognized international contest, attracting over 100 delegates at ISMB, the largest conference in the field. He also co-organized *Traffic4cast* at NeurIPS since 2019, the ACM CIKM *CDCEO 2021*, *Weather4cast* in 2021 at IEEE BigData an *Weather4cast* at NeurIPS in 2022 and 2023.

3.2 Resources provided by organizers

We build on the success of the previous *Weather4cast* competitions [14, 16, 15].

- The Silesian University of Technology provides the infrastructure to maintain the competition website including leaderboards and servers for the evaluation of submissions.
- AEMET provides data and domain specific advice.

3.3 Support requested

The winning teams need to present their work live at the conference. It would help us a lot to know the number of available complementary tickets **at least a month before the conference**, as participants and organizers may need some time to make arrangements for registration and/or visa. We would love to take part in any travel fellowship scheme for early career researchers or colleagues from underrepresented communities or countries. Thank you for your support.

References

[1] Shreya Agrawal, Luke Barrington, Carla Bromberg, John Burge, Cenk Gazen, and Jason Hickey. Machine Learning for Precipitation Nowcasting from Radar Images. arXiv:1912.12132 [cs, stat], December 2019. arXiv: 1912.12132.

[2] 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.

[3] Mohd Yousuf Ansari, Amir Ahmad, Shehroz S. Khan, Gopal Bhushan, and Mainuddin. Spatiotemporal clustering: a review. Artificial Intelligence Review, 53:2381–2423, 2019.

[4] Federico Bassetti, Stefano Gualandi, and Marco Veneroni. On the computation of kantorovich-wasserstein distances between 2d-histograms by uncapacitated minimum cost flows, 2019.

[5] Riccardo Benedetti. Scoring rules for forecast verification. Monthly Weather Review, 138(1):203 – 211, 2010.

[6] Alabi Bojesomo, Hasan Al-Marzouqi, and Panos Liatsis. Spatiotemporal vision transformer for short time weather forecasting. In 2021 IEEE International Conference on Big Data (Big Data), pages 5741–5746, 2021.

[7] 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.

[8] 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.

[9] Sungbin Choi. Unet based future weather prediction on weather4cast 2021 stage 2. In 2021 IEEE International Conference on Big Data (Big Data), pages 5747–5749, 2021.

[10] Michael Crawshaw. Multi-task learning with deep neural networks: A survey, 2020.

[11] Christian Eichenberger, Moritz Neun, Henry Martin, Pedro Herruzo, Markus Spanring, Yichao Lu, Sungbin Choi, Vsevolod Konyakhin, Nina Lukashina, Aleksei Shpilman, Nina Wiedemann, Martin Raubal, Bo Wang, Hai L. Vu, Reza Mohajerpoor, Chen Cai, Inhi Kim, Luca Hermes, Andrew Melnik, Riza Velioglu, Markus Vieth, Malte Schilling, Alabi Bojesomo, Hasan Al Marzouqi, Panos Liatsis, Jay Santokhi, Dylan Hillier, Yiming Yang, Joned Sarwar, Anna Jordan, Emil Hewage, David Jonietz, Fei Tang, Aleksandra Gruca, Michael Kopp, David Kreil, and Sepp Hochreiter. Traffic4cast at neurips 2021 – temporal and spatial few-shot transfer learning in gridded geo-spatial processes, 2022.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Asko Huuskonen, Elena Saltikoff, and Iwan Holleman. The Operational Weather Radar Network in Europe. Bulletin of the American Meteorological Society, 95(6):897–907, June 2014. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[18] David P. Kreil, Michael K. Kopp, David Jonietz, Moritz Neun, Aleksandra Gruca, Pedro Herruzo, Henry Martin, Ali Soleymani, and Sepp Hochreiter. The surprising efficiency of framing geo-spatial time series forecasting as a video prediction task – Insights from the IARAI Traffic4cast Competition at NeurIPS 2019. In NeurIPS 2019 Competition and Demonstration Track, pages 232–241. PMLR, August 2020. ISSN: 2640-3498.

[19] Jussi Leinonen. Improvements to short-term weather prediction with recurrent-convolutional networks. In 2021 IEEE International Conference on Big Data (Big Data), pages 5764–5769, 2021.

[20] Xinzhe Li, Sun Rui, Yiming Niu, and Yao Liu. Precipitation prediction using an ensemble of lightweight learners, 2023.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs], May 2015. arXiv: 1505.04597.

[26] Johannes Schmetz, Paolo Pili, Stephen Tjemkes, Dieter Just, Jochen Kerkmann, Sergio Rota, and Alain Ratier. AN INTRODUCTION TO METEOSAT SECOND GENERATION (MSG). Bulletin of the American Meteorological Society, 83(7):977–992, July 2002. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[27] 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. arXiv: 2003.12140.

[28] D S Wilks. A skill score based on economic value for probability forecasts. Meteorological Applications, 8(2):209–219, 2001.

[29] 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.