HR-Extreme: Supplementary Materials

Nian Ran[1]**, Peng Xiao**[2]**, Yue Wang**[1]*,** Wenlei Shi**[3]**, Jiaxin Lin**[2]**, Qi Meng**[4]**, Richard Allmendinger**[5]*

[1]Zhongguancun Academy
[2]Hunan University
[3]Microsoft Research
[4]Chinese Academy of Sciences
[5]University of Manchester

CONTENTS

---

*Corresponding author,(yuewang@bjzgca.edu.cn, richard.allmendinger@manchester.ac.uk)

1 DATASET AND CODE ACCESS

Following the instructions of NeurIPS Dataset and Benchmark Track guidelines, we have uploaded our datasets to Hugging Face and code for constructing dataset to Github:

1. HR-Extreme: `https://huggingface.co/datasets/NianRan1/HR-Extreme`

2. HR-Extreme Croissant metadata: `https://huggingface.co/api/datasets/NianRan1/HR-Extreme/croissant`

3. Code: `https://github.com/HuskyNian/HR-Extreme/tree/main`

The access to NOAA Storm Event Database is `https://www.ncdc.noaa.gov/stormevents/ftp.jsp` and NOAA Storm Prediction Center is `https://www.spc.noaa.gov/climo/reports/`. Dataset and code will be available upon the acceptance of this paper.

1.1 VARIABLE LIST

There are 69 physical variables in total included in HR-Extreme, details are explained in Table 1. The description of each type of event included is in Table 4.

Variable	Definition	Unit	Range
msl	Mean Sea Level Pressure	Pa	
2t	Temperature 2 m above ground	K	
10u	U-component Wind Speed 10 m above ground	m/s	
10v	V-component Wind Speed 10 m above ground	m/s	
hgtn	Geopotential Height	gpm	
u	U-component Wind Speed	m/s	At 50, 100, 150, 200, 250, 300,
v	V-component Wind Speed	m/s	400, 500, 600, 700, 850, 925, 1000
t	Temperature	K	millibars, 13 levels in total
q	Specific Humidity	kg/kg	

Table 1: Summary of the 69 physical variables in the dataset

1.2 DATASET STATISTICS

We calculated the means and standard deviations of the HRRR (National Oceanic and Atmospheric Administration (NOAA), 2024) data to normalize the data before inputting it into deep learning models. These statistics are stored in a file accessible at `https://huggingface.co/datasets/NianRan1/HR-Extreme/tree/main/statistics`. Our dataset repository is organized into two distinct directories, *202001_202006* and *202007_202012*, which contain data from January 2020 to June 2020 and from July 2020 to December 2020, respectively. Within each directory, the dataset is stored in the WebDataset [1] format, adhering to Hugging Face's guidelines. Specifically, every 10 *npz* files are aggregated into a single *tar* file, named sequentially as *"i.tar"*, where *i* is an integer (e.g., *"0001.tar"*). Consequently, a total of 2200 files will be organized into 220 *tar* files.

1.3 CODE GUIDANCE

We provide our code on GitHub for constructing the dataset, considering different usages and potential future developments. The repository includes three Python scripts named *make_dataseti.py"*, where *i* is 1, 2, or 3. These scripts correspond to three data sources: events from the Storm Events

---

[1]`https://github.com/webdataset/webdataset`

Database (National Centers for Environmental Information (NCEI), 2024), events from the Storm Prediction Center (National Oceanic and Atmospheric Administration (NOAA), Storm Prediction Center (SPC), 2024), and excessive heat and cold events filtered manually, respectively. Users, however, need only utilize *"make_datasetall.py"*, supplying start and end dates to generate a comprehensive index file. This index file delineates the events, including their types, bounding box indices, event start and end times, and details of event spans if multiple types are involved. For instance, executing *"python make_datasetall.py 20200101 20200630"* will produce an index file named *"data_20200101_20200630_info.csv"* within the *"index_files"* directory, ready for generating bounding boxes from HRRR data.

It is important to note that users must first download a storm event details CSV file from the Storm Events Database [2], specifically the *StormEvents_details-ftp_v1.0_year_cdate.csv.gz"* file, and place it in the *"index_files"* directory before executing any of the Python scripts. Users should ensure they download the storm event details files, rather than location or other types of files, and confirm that these files cover the dates of the dataset they intend to create.

1.4 TARGET AUDIENCE

HR-Extreme is a dataset encompassing 17 types of extreme weather events for the year 2020, derived from HRRR (National Oceanic and Atmospheric Administration (NOAA), 2024) data. Users can conveniently extend this dataset using the provided code interface. The dataset is intended for a wide range of researchers in the field of weather forecasting, including those utilizing physical methods and deep learning techniques. Given the critical importance of accurately predicting extreme weather events in weather forecasting, it is imperative that researchers evaluate their models using this dataset. This evaluation will help to identify the strengths and weaknesses of their models and assess their practical applicability. While most previous studies in deep learning have focused on ERA5 (Hersbach et al., 2020) data, models can be readily adapted to work with HRRR data as well.

1.5 LICENSES

The license of NOAA HRRR data is U.S Government Work [3], which means they can be used by anyone for any purpose without seeking permission or paying royalties.

1.6 AUTHOR STATEMENT

As the authors of the HR-Extreme dataset, we bear full responsibility for any violations of rights and confirm that the dataset is released under the appropriate data license. The data has been compiled and shared in accordance with the relevant guidelines and regulations.

2 EXPERIMENT MODELS

2.1 NWP

The Numerical Weather Prediction (NWP) (Olson et al., 2022) system for predicting High-Resolution Rapid Refresh (HRRR) data is a sophisticated model that leverages high-performance computing to simulate atmospheric conditions with fine temporal and spatial resolution. HRRR is a real-time 3-km resolution, hourly updated, cloud-resolving, convection-allowing atmospheric model, initialized by 3-km grids with 3-km radar assimilation. It is designed to provide detailed forecasts over the contiguous United States for short-term predictions up to 18 hours. The HRRR system integrates a variety of data sources, including radar, satellite, and surface observations, to continuously update its predictions. It employs advanced physics parameterizations to accurately represent processes such as convection, cloud formation, and precipitation. By using a high-resolution grid, HRRR is able to capture small-scale weather phenomena, such as thunderstorms and tornadoes, with greater accuracy. The model's outputs are crucial for various applications, including aviation, re-

---

[2] https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/
[3] https://registry.opendata.aws/noaa-hrrr-pds/

newable energy, and severe weather forecasting, providing timely and precise information to support decision-making and public safety.

2.2 PANGU

The Pangu-Weather (Bi et al., 2023) model is an deep learning-based weather forecasting system that employs a three-dimensional deep neural network architecture designed specifically for Earth's unique geospatial features. The core of Pangu-Weather is the 3D Earth-specific transformer (3DEST), which incorporates height information as an additional dimension, allowing the model to effectively capture atmospheric states across various pressure levels. This architecture uses an encoder-decoder structure derived from the Swin transformer (Liu et al., 2021), with enhancements such as Earth-specific positional biases to handle geospatial relationships more accurately.

To make predictions, Pangu-Weather takes reanalysis weather data as input and processes it through a hierarchical temporal aggregation strategy. This strategy trains multiple models with varying lead times (1 hour, 3 hours, 6 hours, and 24 hours) and uses a greedy algorithm during inference to minimize the number of iterations required for medium-range forecasts. By iteratively using forecast results as new inputs, the model reduces cumulative errors and improves accuracy. The combination of 3D spatial awareness and hierarchical temporal aggregation enables Pangu-Weather to produce faster and more accurate weather forecasts, outperforming traditional numerical weather prediction systems like the ECMWF's IFS in various metrics including root mean square error (RMSE) and anomaly correlation coefficient (ACC).

2.3 FUXI

The Fuxi (Chen et al., 2023) model is another deep learning-based weather forecasting system that employs a cascaded architecture to generate 15-day global forecasts with a temporal resolution of 6 hours and a spatial resolution of $0.25°$. This model is built on the U-Transformer architecture, which effectively captures complex spatiotemporal relationships in high-dimensional weather data. The system uses a pre-trained base model, which is fine-tuned for specific forecast time windows: 0-5 days, 5-10 days, and 10-15 days. By cascading these models, Fuxi reduces error accumulation and enhances forecast accuracy over longer periods. The model was trained on 39 years of ECMWF ERA5 (Hersbach et al., 2020) reanalysis data. During inference, Fuxi utilizes an autoregressive approach where the outputs of one model serve as inputs for the next, significantly extending the skillful forecast lead time. Additionally, Fuxi employs ensemble forecasting by perturbing initial conditions and model parameters, providing a measure of forecast uncertainty and demonstrating performance comparable to the ECMWF ensemble mean for the 15-day forecast horizon.

2.4 HR-HEIM

HR-Heim The architecture of HR-Heim follows a conventional structure with an encoder, a series of transformer layers, and a decoder, inspired by the FuXi architecture (Chen et al., 2023). For the encoder, we utilize causal convolutions from MagViTv2 to capture spatial-temporal input (Yu et al., 2023). The transformer segment consists of multiple stacked SwinTransformer blocks (Liu et al., 2022). Unlike typical Vision Transformer decoders that use a simple MLP with 1×1 convolution, which can hinder resolution, our decoder progressively upscales the feature map from $\frac{H}{h} \times \frac{W}{w}$ to the target size $H \times W$ through a series of steps. Each step resolves details at its specific resolution level, incorporating convolutional layers and upsampling operations to enhance prediction quality.

Decoder Grouping by Variable Types In the context of a multi-task formulation, the optimal approach for performance would typically involve training a separate model for each subtask. However, this becomes inefficient and scales poorly during training and inference when dealing with hundreds of physical variables. To address this, we propose a shared backbone for feature extraction and temporal modeling, which we assume to be universally applicable for predicting any physical variables. We then separate multiple decoding heads for different variables.

To further enhance efficiency, we categorize the subtasks into several groups, assigning a decoding head to each. The grouping principle aims to minimize negative transfer within each group. For example, during training, the gradients of subtask A and subtask B often point in conflicting directions.

Following previous work, we train a simple model that shares all parameters and assesses conflicts and alignment between each pair of subtasks during training. Specifically, we compute the cosine similarity of the gradients of each pair of subtasks to obtain an affinity matrix.

$$\hat{Z}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \frac{g_t^i \cdot g_t^j}{\left\| g_t^i \right\|_2 \left\| g_t^j \right\|_2} \tag{1}$$

Here, g_t^i represents the flattened gradient from the loss of the ith subtask. Interestingly, we find that variables of the same type but at consecutive pressure levels (e.g., t_100 and t_120) align better than those of the same pressure level but different types (e.g., $t_1 00$ and $h_1 00$). Consequently, each group contains only one type of variable, and we determine which pressure levels to group together by maximizing the inner group affinity.

3 MODEL EVALUATIONS

3.1 METRICS

The Root Mean Square Error (RMSE) is a commonly used metric to measure the difference between values predicted by a model and the values actually observed. It is defined as the square root of the mean of the squared differences between the predicted and observed values.

Mathematically, RMSE is expressed as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where n is the number of observations, y_i is the observed value, $\hat{y}_i$ is the predicted value. The RMSE value provides a measure of how well a model predicts the outcome variable. Lower RMSE values indicate better fit, meaning the predicted values are closer to the observed values.

3.2 RESULTS

Our full results are included in Table 2 and Table 3, showing the RMSE for each variable predicted by each model.

4 DATASHEET

4.1 MOTIVATION

1. **For what purpose was the dataset created?** *We propose HR-Extreme, a dataset with high-resolution feature maps of physical variables for evaluating cutting-edge models in extreme weather prediction. This crucial aspect has been overlooked and undervalued in previous work, and there is a lack of dedicated high-resolution datasets for extreme weather forecasting.*

2. **Who created the dataset and on behalf of which entity?** *The dataset was developed by a consortium of ML researchers and climate scientists listed in the author list.*

3. **Who funded the creation of the dataset?** *Self-funded.*

4.2 DISTRIBUTION

1. **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** *Yes, it is open the public.*

2. **How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?** *The dataset will be distributed in Hugging Face, and code for constructing the dataset held in Github.*

3. **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** *No.*

4. **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** *No.*

4.3 MAINTENANCE

1. **Who will be supporting/hosting/maintaining the dataset?** *The authors of this paper.*

2. **How can the owner/curator/manager of the dataset be contacted(e.g., email address)?** *The owner/curator/manager of the dataset can be contacted by Nian Ran (r992988188@gmail.com).*

3. **Is there an erratum** *No. If errors are found in the future, we will release errata on the Github page:* `https://github.com/HuskyNian/HR-Extreme/tree/main`.

4. **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** *Yes, we will update our dataset whenever necessary to ensure accuracy, and announcements will be made accordingly. The updates will be shown in our Github page:* `https://github.com/HuskyNian/HR-Extreme/tree/main`.

5. **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted?)** *N/A*

6. **Will older version of the dataset continue to supported/hosted/maintained?** *Yes, they will be still available and easy to reproduce by our provided code.*

7. **If others want to extend/augment/build on/contribute to the dataset, is there a mechanisms for them to do so?** *Yes, they can use our provided code on (*`https://github.com/HuskyNian/HR-Extreme/tree/main`*), we have provided full instructions in our paper and this material.*

4.4 COMPOSITION

1. **What do the instance that comprise the dataset represent (e.g., documents, photos, people, countries?)** *Each instance is the feature maps with 320 in width and 320 in height, and 69 channels, which representing 69 physical variables as shown in Table 1. This areas are cropped from HRRR (National Oceanic and Atmospheric Administration (NOAA), 2024) data.*

2. **How many instances are there in total (of each type, if appropriate)?** *There are 22,774 files accounting for the period of July 2020 to December 2020, and 34,196 files for the period of January 2020 to June 2020.*

3. **Does the dataset contain all possible instances or is it a sample of instances from a larger set?** *Our dataset contains all possible instances in entire 2020. Users can use our code to extend to more years depending on their needs.*

4. **Is there a label or target associated with each instance?** *Yes, after loading a file by **Numpy** library, users can use keys "inputs" to retrieve inputs and "targets" to retrieve targets and "mask" to retrieve mask for that case.*

5. **. Is any information missing from individual instances?** *No.*

6. **Are there recommended data splits (e.g., training, development/validation, testing)?** *Considering the computational and memory requirements, we recommend using the data from July 2020 to December 2020 for evaluation. Our evaluation results are based on this subset.*

7. **Are there any errors, sources of noise, or redundancies in the dataset?** *Because our dataset is primarily based on user reports, there may be noise and errors in the range and span of each extreme event case. To mitigate these errors, we have increased the range of each case.*

8. **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** *The dataset is self-contained.*

9. **Does the dataset contain data that might be considered confidential?** *No.*

10. **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** *No.*

4.5 COLLECTION PROCESS

1. **How was the data associated with each instance acquired?** *Each instance is cropped from HRRR (National Oceanic and Atmospheric Administration (NOAA), 2024) by a prepared index file, the index file is constructed by data from NOAA Storm Event Database (`https://www.ncdc.noaa.gov/stormevents/ftp.jsp`) and NOAA Storm Prediction Center (`https://www.spc.noaa.gov/climo/reports/`).*

2. **What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?** *We use many CPU nodes to to process the data with Python libraries to access HRRR data. Users are recommended to first download ground true HRRR data in a storage for easy and frequent access.*

3. **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?** *Regular students and employees are involved, no crowdworkers are involved.*

4. **Does the dataset relate to people?** *No.*

5. **Did you collect the data from the individuals in questions directly, or obtain it via third parties or other sources (e.g., websites)?** *We obtained the dataset from open resource HRRR (National Oceanic and Atmospheric Administration (NOAA), 2024) data.*

4.6 USES

1. **Has the dataset been used for any tasks already?** *No, this dataset has only been evaluated on some models explained in the paper.*

2. **What (other) tasks could be the dataset be used for?** *Apart from evaluation, it can be used a finetune dataset for deep learning models.*

3. **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** *The varibale choices can impact the future use. But so far the variables chosen are diverse and sufficient for evaluating extreme weather.*

4. **Are there tasks for which the dataset should not be used?** *No.*

5 ADDITIONAL FIGURES AND TABLES
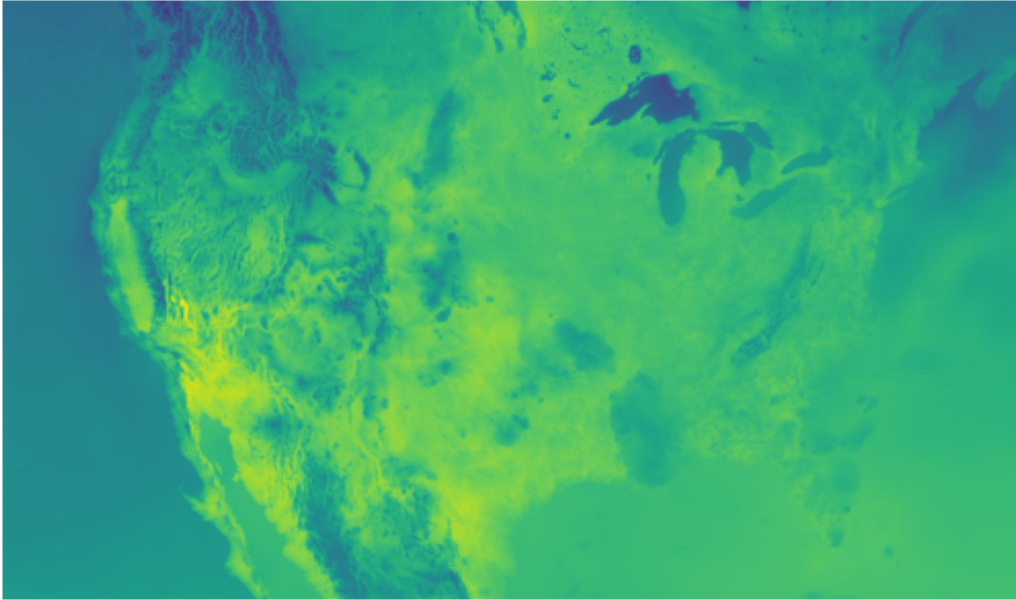
Figure 1: An example plot of the variable "temeprature 2 meters above ground" of HRRR data on U.S.
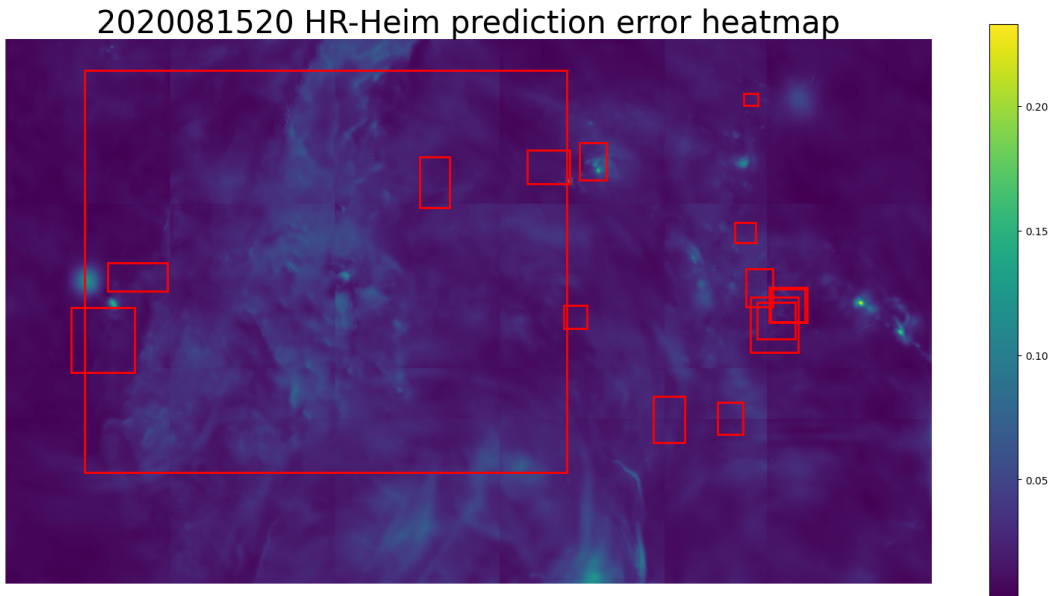
Figure 2: More examples: the mean error heatmap of all variables predicted by HR-Heim on entire U.S. on 8 p.m., 15 August. Bright area means larger avaergae loss. They are captured by the bounding boxes representing the extreme weather that is happening.

REFERENCES

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06185-3. URL https://doi.org/10.1038/s41586-023-06185-3.

Figure 3: More examples: the mean error heatmap of all variables predicted by NWP on entire U.S. on 10 p.m., 2 July. Bright area means larger avaergae loss. They are captured by the bounding boxes representing the extreme weather that is happening.

Lei Chen, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li. Fuxi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj C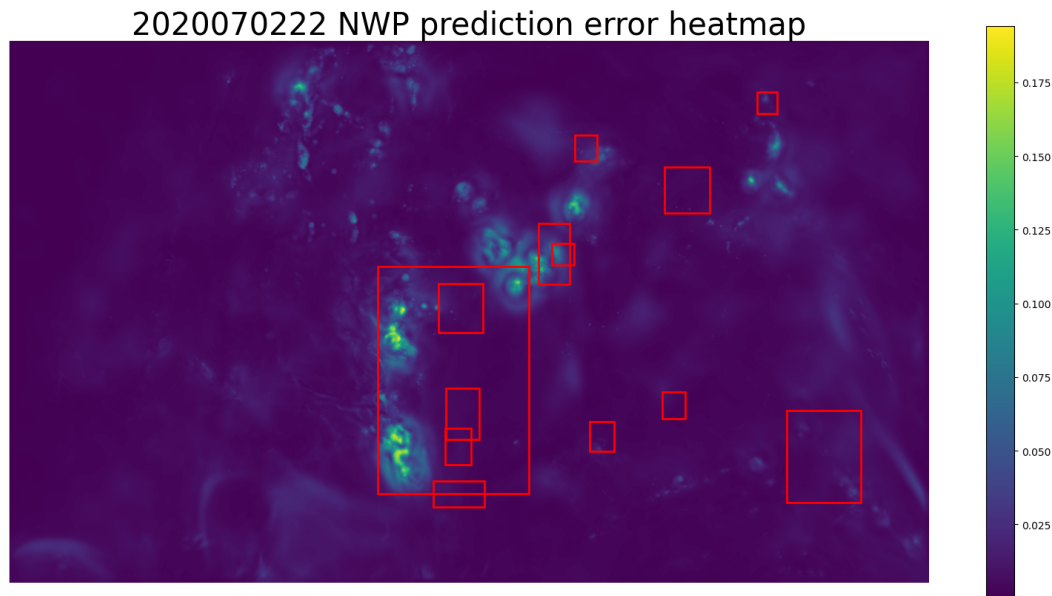limate and Atmospheric Science*, 6(1):190, November 2023. ISSN 2397-3722. doi: 10.1038/s41612-023-00512-1. URL https://doi.org/10.1038/s41612-023-00512-1.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, and J.-N Thépaut. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 06 2020. doi: 10.1002/qj.3803.

Z. Liu, H. Hu, Y. Lin, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12009–12019, 2022.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows, 2021.

National Centers for Environmental Information (NCEI). NOAA Storm Events Database, 2024. URL https://www.ncdc.noaa.gov/stormevents/. Accessed: 2024-05-20.

National Oceanic and Atmospheric Administration (NOAA). High-resolution rapid refresh (hrrr), 2024. URL https://rapidrefresh.noaa.gov/hrrr/. Accessed: 2024-05-18.

National Oceanic and Atmospheric Administration (NOAA), Storm Prediction Center (SPC). NOAA Storm Prediction Center Severe Weather Event Reports, 2024. URL https://www.spc.noaa.gov/climo/reports/. Accessed: 2024-05-20.

Joseph B. Olson, Ming Hu, Tatiana G. Smirnova, Terra Ladwig, Jaymes S. Kenyon, Ravan Ahmadov, David D. Turner, Jeffrey D. Duda, and Trevor I. Alcott. The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part i: Motivation and system description. *Weather and Forecasting*, 37(8):1371–1395, 2022. doi: 10.1175/WAF-D-21-0151.1. URL https://doi.org/10.1175/WAF-D-21-0151.1.

L. Yu, J. Lezama, N. B. Gundavarapu, et al. Language model beats diffusion–tokenizer is key to visual generation. *arXiv preprint arXiv:2310.05737*, 2023.

Variable	HR-Heim General Loss	HR-Heim HR-Extreme Loss	Pangu General Loss	Pangu HR-Extreme Loss	FuXi General Loss	FuXi HR-Extreme Loss	NWP General Loss	NWP HR-Extreme Loss
msl	**27.00**	**27.88**	37.09	36.06	40.48	78.61	60.30	80.07
2t	**0.40**	**0.47**	0.81	0.74	0.86	1.63	0.52	1.19
10u	**0.59**	**0.75**	0.81	0.90	0.78	1.09	0.89	1.32
10v	**0.59**	**0.77**	0.82	0.93	0.78	1.12	0.89	1.36
hgtn_5000	**6.48**	**6.80**	8.35	8.53	10.25	25.09	8.05	10.58
hgtn_10000	**5.32**	**5.93**	6.87	7.39	9.54	28.76	7.04	8.56
hgtn_15000	**5.14**	**5.80**	6.55	7.15	10.70	35.26	6.83	8.22
hgtn_20000	**4.77**	**5.54**	6.77	7.53	10.54	36.72	6.50	8.12
hgtn_25000	**4.32**	**5.10**	5.76	272.94	9.84	34.94	6.12	7.80
hgtn_30000	**3.95**	**4.69**	5.19	14.26	9.01	30.41	5.82	7.38
hgtn_40000	**3.30**	**3.93**	4.31	17.91	7.38	22.87	5.37	6.84
hgtn_50000	**2.85**	**3.39**	4.06	13.11	5.95	18.27	5.18	6.73
hgtn_60000	**2.58**	**3.02**	3.35	10.66	4.98	14.14	5.11	6.64
hgtn_70000	**2.40**	**2.69**	3.05	83.58	4.21	10.82	5.06	6.42
hgtn_85000	**2.27**	**2.44**	2.88	28.27	3.45	7.05	5.06	6.51
hgtn_92500	**2.31**	**2.44**	3.24	39.93	3.40	6.46	5.15	6.86
hgtn_100000	**2.26**	**2.37**	46.46	44.29	3.41	6.46	5.10	6.90
u_5000	0.49	**0.52**	0.71	0.69	0.75	1.32	**0.48**	0.93
u_10000	**0.47**	**0.55**	0.69	0.74	0.76	1.69	0.49	1.03
u_15000	**0.67**	**0.83**	0.90	1.06	0.97	2.34	0.73	1.45
u_20000	**0.96**	**1.20**	1.27	1.84	1.35	2.95	1.08	1.96
u_25000	**0.92**	**1.14**	1.22	13.53	1.35	3.02	1.03	1.89
u_30000	**0.83**	**1.04**	1.09	3.14	1.24	2.71	0.91	1.67
u_40000	**0.69**	**0.93**	0.91	3.55	1.03	2.38	0.75	1.42
u_50000	**0.62**	**0.93**	0.89	2.94	0.89	2.20	0.70	1.42
u_60000	**0.58**	**0.89**	0.80	2.75	0.82	1.87	0.68	1.39
u_70000	**0.59**	**0.88**	0.79	6.60	0.80	1.70	0.69	1.36
u_85000	**0.59**	**0.89**	0.78	4.37	0.79	1.57	0.73	1.40
u_92500	**0.55**	**0.86**	1.28	4.15	0.73	1.44	0.72	1.42
u_100000	**0.45**	**0.60**	3.12	2.64	0.57	0.96	0.63	1.04
v_5000	0.52	**0.53**	0.71	0.68	0.65	0.86	**0.47**	0.95
v_10000	**0.47**	**0.54**	0.67	0.70	0.70	1.21	0.47	1.00
v_15000	**0.68**	**0.82**	0.89	1.03	0.96	1.97	0.72	1.43
v_20000	**0.98**	**1.19**	1.28	1.90	1.34	2.69	1.07	1.95
v_25000	**0.94**	**1.13**	1.25	11.74	1.36	2.63	1.04	1.92
v_30000	**0.84**	**1.03**	1.11	3.10	1.25	2.38	0.92	1.72
v_40000	**0.70**	**0.93**	0.93	3.64	1.02	2.06	0.76	1.48

Table 2: Losses of each variable for each model

Variable	HR-Heim General Loss	HR-Heim HR-Extreme Loss	Pangu General Loss	Pangu HR-Extreme Loss	FuXi General Loss	FuXi HR-Extreme Loss	NWP General Loss	NWP HR-Extreme Loss
v_50000	**0.63**	**0.94**	0.90	3.06	0.90	1.92	0.71	1.48
v_60000	**0.59**	**0.91**	0.81	2.82	0.82	1.77	0.68	1.45
v_70000	**0.60**	**0.90**	0.81	6.52	0.80	1.60	0.69	1.40
v_85000	**0.60**	**0.91**	0.80	4.78	0.80	1.60	0.74	1.43
v_92500	**0.57**	**0.88**	1.36	4.80	0.76	1.52	0.74	1.47
v_100000	**0.46**	**0.62**	4.04	3.34	0.60	1.01	0.65	1.08
t_5000	0.32	**0.35**	0.43	0.43	0.51	0.63	**0.29**	0.63
t_10000	**0.24**	**0.29**	0.35	0.39	0.41	0.90	0.25	0.57
t_15000	**0.23**	**0.29**	0.33	0.36	0.36	0.82	0.26	0.49
t_20000	**0.31**	**0.34**	0.41	0.46	0.43	0.63	0.35	0.43
t_25000	**0.25**	**0.28**	0.34	5.42	0.38	0.83	0.29	0.38
t_30000	**0.21**	**0.26**	0.28	1.29	0.34	0.92	0.25	0.39
t_40000	**0.21**	**0.28**	0.28	1.57	0.35	0.96	0.25	0.43
t_50000	**0.19**	**0.27**	0.31	1.43	0.35	0.89	0.24	0.44
t_60000	**0.18**	**0.26**	0.28	1.47	0.33	0.87	0.22	0.43
t_70000	**0.21**	**0.30**	0.30	7.29	0.38	0.98	0.26	0.51
t_85000	**0.28**	**0.35**	0.44	2.64	0.49	1.15	0.38	0.65
t_92500	**0.31**	**0.37**	1.05	3.02	0.53	1.20	0.43	0.75
t_100000	**0.29**	**0.36**	9.79	3.11	0.52	1.19	0.48	0.95
q_5000	6.30e-08	5.61e-08	7.99e-08	6.99e-08	8.89e-08	1.49e-07	**5.03e-08**	**5.51e-08**
q_10000	**1.87e-07**	3.86e-07	2.42e-07	**3.66e-07**	4.20e-07	6.99e-07	4.45e-07	6.63e-07
q_15000	**6.13e-07**	**1.17e-06**	7.38e-07	1.18e-06	8.85e-07	1.66e-06	9.66e-07	2.13e-06
q_20000	**3.02e-06**	**5.68e-06**	3.86e-06	7.74e-06	3.94e-06	8.34e-06	4.73e-06	1.05e-05
q_25000	**9.16e-06**	**1.72e-05**	1.17e-05	9.99e-05	1.21e-05	2.72e-05	1.47e-05	3.31e-05
q_30000	**2.03e-05**	**3.85e-05**	2.58e-05	1.02e-04	2.70e-05	6.43e-05	3.17e-05	7.52e-05
q_40000	**5.82e-05**	**1.13e-04**	7.41e-05	3.92e-04	7.71e-05	1.93e-04	8.62e-05	2.15e-04
q_50000	**1.13e-04**	**2.16e-04**	1.53e-04	7.44e-04	1.46e-04	3.74e-04	1.56e-04	3.72e-04
q_60000	**1.79e-04**	**3.34e-04**	2.23e-04	1.05e-03	2.26e-04	5.60e-04	2.32e-04	5.10e-04
q_70000	**2.56e-04**	**4.76e-04**	3.14e-04	3.60e-03	3.20e-04	7.35e-04	3.22e-04	6.76e-04
q_85000	**4.13e-04**	**6.60e-04**	5.04e-04	3.27e-03	5.02e-04	9.85e-04	4.83e-04	9.04e-04
q_92500	**4.57e-04**	**5.77e-04**	1.28e-03	3.37e-03	5.78e-04	1.03e-03	5.44e-04	8.06e-04
q_100000	**3.59e-04**	**4.32e-04**	5.13e-03	2.44e-03	4.97e-04	9.81e-04	4.99e-04	8.01e-04

Table 3: Losses of each variable for each model, follow Table 2

11

Variable	Definition
Debris Flow	May be triggered by intense rain after wildfires. A slurry of loose soil, rock, organic matter, and water, similar to wet concrete, which is capable of holding particles larger than gravel in suspension.
Flash Flood	A life-threatening, rapid rise of water into a normally dry area beginning within minutes to multiple hours of the causative event (e.g., intense rainfall, dam failure, ice jam).
Flood	Any high flow, overflow, or inundation by water which causes damage. In general, this would mean the inundation of a normally dry area caused by an increased water level in an established watercourse, or ponding of water, associated with heavy rainfall
Funnel Cloud	A rotating, visible extension of a cloud pendant from a convective cloud with circulation not reaching the ground. It is a precursor to more severe weather events. The wind shear caused by it (sudden changes in wind speed and direction) will endanger aviation.
Hail	Frozen precipitation in the form of balls or irregular lumps of ice.
Heavy Rain	Unusually large amount of rain which does not cause a Flash Flood or Flood event, but causes damage
Lightning	A sudden electrical discharge from a thunderstorm, resulting in a fatality, injury, and/or damage
Marine Hail	Hail 3/4 of an inch in diameter or larger, occurring over the waters and bays of the ocean, Great Lakes, and other lakes with assigned specific Marine Forecast Zones
Marine High Wind	Non-convective, sustained winds or frequent gusts of 48 knots (55 mph) or more, resulting in a fatality, injury, or damage, over the waters and bays of the ocean, Great Lakes, and other lakes with assigned specific Marine Forecast Zones.
Marine Strong Wind	Non-convective, sustained winds or frequent gusts up to 47 knots (54 mph), resulting in a fatality, injury, or damage, occurring over the waters and bays of the ocean, Great Lakes, and other lakes with assigned specific Marine Forecast Zones.
Marine Thunderstorm Wind	Winds, associated with thunderstorms, occurring over the waters and bays of the ocean, Great Lakes, and other lakes with assigned specific Marine Forecast Zones with speeds of at least 34 knots (39 mph) for 2 hours or less.
Thunderstorm Wind	Winds, arising from convection (occurring within 30 minutes of lightning being observed or detected), with speeds of at least 50 knots (58 mph), or winds of any speed (non-severe thunderstorm winds below 50 knots)
Tornado	A violently rotating column of air, extending to or from a cumuliform cloud or underneath a cumuliform cloud, to the ground, and often (but not always) visible as a condensation funnel.
Waterspout	A rotating column of air, pendant from a convective cloud, with its circulation extending from cloud base to the water surface of bays and waters of the Great Lakes, and other lakes with assigned Marine Forecast Zones.
Wind	Severe thunderstorm wind, strong wind that causes damage, which is reported and recorded by NOAA Storm Prediction Center (National Oceanic and Atmospheric Administration (NOAA), Storm Prediction Center (SPC), 2024).
Heat	Large area of excessive heat above 37 degrees celsius in US during daytime, manually filtered.
Cold	Large area of excessive cold below -29 degress celsius in US during nighttime, manually filtered.

Table 4: Explaination of each type of extreme weather included in HR-Extreme