# Long-Term Hail Risk Assessment with Deep Neural Networks

Ivan Lukyanenko

Mikhail Mozikov

Yury Maximov

Ilya Makarov

arXiv:2209.01191v1 [physics.ao-ph] 31 Aug 2022

***Abstract*—Hail risk assessment is necessary to estimate and reduce damage to crops, orchards, and infrastructure. Also, it helps to estimate and reduce consequent losses for businesses and, particularly, insurance companies. But hail forecasting is challenging. Data used for designing models for this purpose are tree-dimensional geospatial time series. Hail is a very local event with respect to the resolution of available datasets. Also, hail events are rare - only 1% of targets in observations are marked as "hail".**

**Models for nowcasting and short-term hail forecasts are improving. Introducing machine learning models to the meteorology field is not new. There are also various climate models reflecting possible scenarios of climate change in the future. But there are no machine learning models for data-driven forecasting of changes in hail frequency for a given area.**

**The first possible approach for the latter task is to ignore spatial and temporal structure and develop a model capable of classifying a given vertical profile of meteorological variables as favorable to hail formation or not. Although such an approach certainly neglects important information, it is very light weighted and easily scalable because it treats observations as independent from each other. The more advanced approach is to design a neural network capable to process geospatial data. Our idea here is to combine convolutional layers responsible for the processing of spatial data with recurrent neural network blocks capable to work with temporal structure.**

**This study compares two approaches and introduces a model suitable for the task of forecasting changes in hail frequency for ongoing decades.**

***Index Terms*—climate, climate modeling, hail forecast, machine learning, time series**

## I. Introduction

According to Verisk's 2021 report [15], year losses due to hail in 2020 reached $14.2 billion in the USA. Insurance companies are especially vulnerable to hail events. Urban sprawl and population growth in large cities such as Dallas/Fort Worth, Texas; St. Louis, Missouri; Chicago, Illinois; and Denver, Colorado, have made large amounts of property damage from hail events more likely.

According to [13], hailstones are formed when raindrops are carried upward by thunderstorm updrafts into extremely cold areas of the atmosphere and freeze. Hailstones then grow by colliding with liquid water drops that freeze onto the hailstone's surface. The hail falls when the thunderstorm's updraft can no longer support the weight of the hailstone, which can occur if the stone becomes large enough or the updraft weakens. In most studies, it is noted that most of the hail's growth occurs at a temperature of approximately -10°C to -25°C.

An important ingredient for creating large hailstones is time. Appreciable growth is only attainable if particles remain in an environment conducive to growth for an extended period of time. Some studies suggest that large hailstones spend as much as 10–15 min or more in growth regions of storms.

The hail is an extreme event. The main complexity comes from the fact that hail is very local both in time and space in comparison with the resolution of available climate models. Hail/no hail ratio in the experiment setting is around 1%.

For nowcasting (forecasts on a period of up to 2 hours) and short-term (up to 24 hours) hail forecasts, a lot of methods exist. They are based on various meteorological models and a combination of the latter with machine learning approaches. In [2], authors add a machine learning model on top of the convection-allowing ensemble system and produce a 24-hour forecast as a result. But models for short-term forecasts differ from models for estimating hail frequencies on a climatological scale. The main reason - is differences in available data.

Next important in the context of the current research model [14] based on a statistical approach to the problem. This model works with joint distribution of atmosphere indices and provides a map of current hail probability distribution around the globe. Although this approach was quite robust and reliable across plain terrains of Europe and the USA, it lacks the ability to take into account landscape influence. Current research aims to develop a machine learning model capable to work with

various climatic models and scenarios of climate change. This model should be able to produce an estimation of changes in hail probability for a given area.

Our main approach is to use a combination of CNN and LSTM neural networks. It allows the model to catch both spatial and temporal structure of data.

The baseline approach to compare is to assume an absence of structure due to the sparsity of the target variable and apply gradient boosting.

The task is to identify favorable to hail development conditions (classification) and then evaluate the frequency of hail events in a given area relying on climate models included in different Coupled Model Intercomparison Project Phases (CMIP5/CMIP6).

To summarize, our main contributions are as follows:

- We propose a special architecture of a neural network designed to forecast changes in hail frequency for ongoing decades based on CMIP data.
- We provide a result of experiments and a comparison with the baseline model. These experiments and comparison show us that our model provides a good approximation of the real annual cycle of hail frequencies (averaged over 2010-2015) and outperforms the baseline approach.

The rest of the work is organized as follows. Section 2 is elaborating on the experiment setting. In this section, we will describe the data we used, quality metrics, and learning settings and formulate a problem statement. In Section 3 we describe the neural network approach to hail forecasting. Here we will introduce our neural network - HailNet. We will show its architecture and explain why we chose this particular type. Section 3 consists of experiment results. In this part, we will create a baseline solution based on gradient boosting and after that compare HailNet performance with the baseline and model introduced in [14].

## II. Related work

In this section, we will first overview existing approaches to meteorological and climate forecasts. Also, we will overview methods to work with geospatial data in general.

Machine learning approaches steadily integrate into the meteorological field starting from the operational part [4] and reaching narrower domains like agro-meteorology [10] or specific weather phenomena like heavy rains [6]. In agro-meteorology case machine learning is to estimate crop water demand using on-farm sensors and public weather information.

A separate direction is climate studies. The main focus here is on designing or improving global climate models with machine learning techniques involved [9], [1]. Agro-meteorology is also concerned [3] but on a different scale, here satellite and climate data were integrated to predict wheat yield in Australia using machine learning approaches.

The important idea of why the field of weather forecasts (meteorology) and the field of climate forecasts which seems closely related are yet very different when it comes to machine learning usage is well described in [16].

Let us consider the part of the meteorological field which works with hail prediction. First interesting study here is [2]. The model designed by the authors processes an output of a convection-allowing ensemble system and produces a 24-hour forecast as a result. The possible complication here is that convection-allowing ensembles are not available for almost all countries except for a few. The next study is [14]. It is based on joint distributions of atmosphere parameters and produces a map of current hail probability distribution around the globe. It is shown in the study that the model has a good generalization ability in Europe and the USA except for mountain areas.

Our study is related to geospatial time series analysis. Some approaches to visualize such kind of data are described in [11]. In the initial steps of research, visualization can help geoscientists to gain a better understanding of geospatial time series. The article [7] elaborates on AI-based techniques for 3D point clouds and geospatial data as generic components of geospatial AI.

## III. Experiment setting

In the first part of this section, we list all sources of data we use and describe preprocessing step. Next, we formulate a problem statement. In the last part, we describe metrics used for evaluating model performance.

### A. Data

Data required to perform hail forecasts arse geospatial time series. Such data are a series of values of a quantity obtained at successive times with equal intervals between them with spatial relations involved. Worth noting here that hailstones formation is associated with cumulonimbus clouds which unlike others have extended vertical structure exceeding in some cases 18 km [12]. It is important to address this fact by using data from different isobaric surfaces. So in this way data structure of geospatial time series expands to three dimensions.

The dataset for model training and evaluation includes three sources. The ERA5 hourly data on single levels by European Centre for Medium-Range Weather Forecasts (ECMWF) were used to derive surface climate variables. To address the third dimension of data structure variables from ERA5 hourly data on pressure levels were collected. The train data example is presented in Figure 1.

The time span is from 2010 to 2021, the time resolution is one hour, model's spatial resolution is $0.25°$x$0.25°$ latitude and longitude. The target value is binary, either 1: "hail" or 0: "no hail". The National Oceanic and Atmospheric Administration's (NOAA) Storm Events Database was used to collect this data. This database contains information about the occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce in the USA. The data concerning the start, end time, and coordinates of hail events were extracted. It is important to note here that on a large timescale the Storm Events Database is not homogeneous due to changes in observation techniques and differences in sources of data. But on the scale of the last decade, this

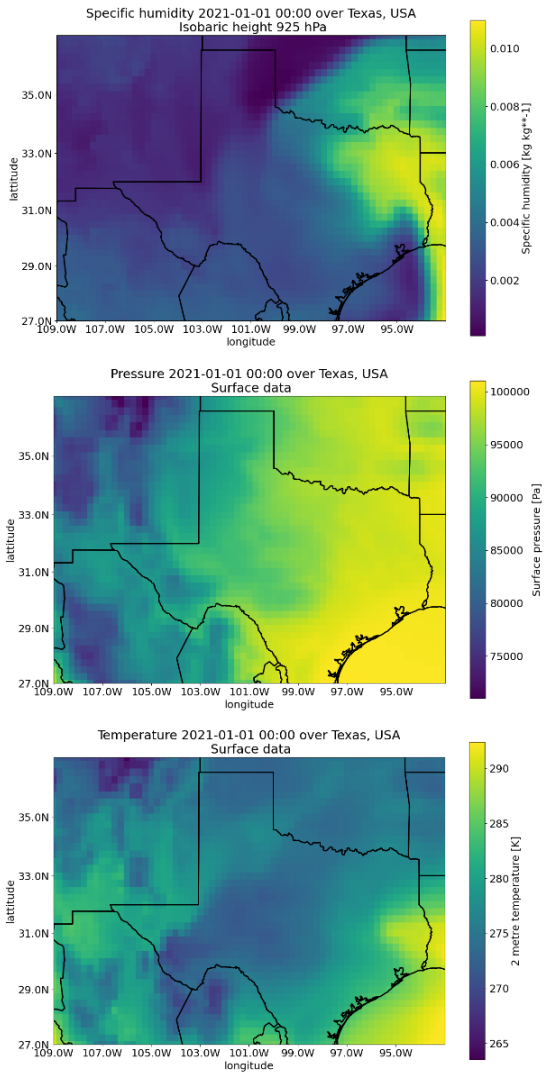



Fig. 1. The example of ERA5 data used for training an evaluation of our model. Timestamp: 01.01.2021 00:00. From top to bottom: specific humidity at 925 hPa isobaric surface, pressure at 925 hPa isobaric surface, temperature 2 meters above surface level.

inhomogeneity is almost negligible. Full list of variables is presented in Table I.

Coupled Model Intercomparison Project (CMIP) includes climate models introduced by different universities developed under a list of assumptions about climate change scenarios. Scenario Representative Concentration Pathway 8.5 (RCP8.5) is selected. It reflects the modeled climate changes for the case when emissions continue to rise throughout the 21st century. This scenario is a "worst case" option in CMIP5. As follows from the properties of hail events data as dense as possible are needed here. The suitable dataset is Meteorological Research Institutes' MRI-CGCM3. The spatial resolution of the model is 1°x1°. The parameters required for the forecast are available in 6 hours time resolution.

TABLE I
DATA DESCRIPTION

| Dataset | variables | Time span |
|---|---|---|
| ERA5 on pressure levels | U-component of wind | 2010-2021 |
| | V-component of wind | |
| | Specific humidity (calculated) | |
| | Temperature | |
| ERA5 on single levels | Runoff | 2010-2021 |
| | Surface pressure | |
| Storm Events Database | Hail event | 2010-2021 |
| MRI-CGCM3 RCP8.5 | U-component of wind | 2022-2050 |
| | V-component of wind | |
| | Specific humidity | |
| | Temperature | |
| | Runoff | |
| | Surface pressure | |

Variables describing humidity differ in ERA5 datasets and CMIP data. To coincide with CMIP data, specific humidity was calculated using the MetPy library. Also, there is a mismatch both in spatial and temporal dimensions. It was decided to preprocess all data to the resolutions of the least dense dataset - CMIP5 MRI-CGCM3. The final spatial resolution of all data is 1°x1°, temporal - 6 hours.

Also, during preprocessing step it was needed to deal with the problem of imbalanced target data. To address this problem SMOTE technique [5] is applied. After this step ratio of hail/no hail events has increased from 1/100 to 1/1.

### B. Problem statement

Let us give some designations and formalize our problem:

$n \in \mathbb{N}$ – number of climatic variables;

long $\in \mathbb{N}$ – longitude of considered area;

lat $\in \mathbb{N}$ – latitude of considered area;

$X_{i,j,k} \in \mathbb{R}^{\mathrm{n} \times \mathrm{lat} \times \mathrm{long}}$ – a tensor of climatic variables at one timestamp;

$\tilde{X}_{t,i,j,k}$ – a time-series of climatic variables corresponding to one day;

$f(w, X)$ – predicting model.

We are solving the time series classification problem for every point on the grid of the considered region. Our objects are time series of tensors $X_{i,j,k}$ corresponding to every 6-period. Our targets are lat $\times$ long matrices with zeros (no hail) and ones (hail) corresponding to hail events.

The output of the $f(w, x)$ is 2-dimensional probability pseudo-distribution. Every output value range from 0 to 1, but sum of every matrix value is not equal to 1, because favorable conditions for hail formation can occur on the same day in many points at the same day and vice versa.

Mathematically we set the optimization problem of minimizing the Mean Squared Error between output pseudo-distributions and target grids for finding the best parameters for our model.

$$w^* = \arg\min_w \frac{1}{N} \sum_{t=1}^{N} \|f(w, X_t) - Y_t\|^2 \qquad (1)$$

### C. Quality metrics

We will use two basic (precision (2) and recall (3)) and one domain-specific metric (4).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Where:

$TP$ - true positive rate of model predictions;

$FP$ - false positive rate of model predictions;

$FN$ - false negative of model predictions.

Talking about the domain-specific metric, it is highly important to ensure that our model is capable to catch the annually-periodic structure of hail occurrence. It is possible with a special metric firstly introduced in [14]. The metric is a simple root-mean-squared error ($RMSE_{om}$) between the normalized observed and modeled annual cycle of monthly mean hail frequency over the target area.

$$RMSE_{om} = \sqrt{\sum_{i=1}^{12}(o_m - m_m)^2} \tag{4}$$

Where:

$o_m$ - normalized observed annual cycle of monthly mean hail frequency;

$m_m$ - normalized modeled annual cycle of monthly mean hail frequency;

Normalization is needed due to the fact that the dataset with hail observation is partially inhomogeneous. Although this inhomogeneity is small during the 2010-2021 years, we still preserve normalization to increase further stability of metric. Also, this approach coincides with the original paper, so we can compare models' performance.

Let us denote classes of target data:

- 1: "hail"
- 0: "no hail"

In the current setting, it makes sense to pay more attention to the recall metric of our model. The cost of false negatives (the case when we underestimated hail frequency) is expected to be higher than of false positives.

## IV. Neural network approach (HailNet)

The next step is to build a model capable to catch the spatial and temporal structure of climate variables. It's possible by introducing a neural network with an architecture presented in Figure 2. Further in the paper, it will be referred to as HailNet. The neural network has two steps of feature extraction:

- Convolutional (Conv2d) + dense layers
- Recurrent layers (LSTM)

The first step allows us to catch the spatial structure of data. After that, the concatenation of outputs from Dense and Convolutional layers maps every climate tensor of time series to embedding space. Objects from embedding space have fewer dimensions and are specified with important spatial information extracted from raw climate tensors. Further, we observe time series of elements from embedding space. Using the LSTM network (second step) HailNet extracts temporal features from time series. The output of the last LSTM unit proceeds to the Dense layers with a sigmoid activation function on the output. As an output, we get a 2-dimensional grid with shape lat $\times$ long with values in a range from 0 to 1. This grid is interpreted as a 2-dimensional probability pseudo-distribution.

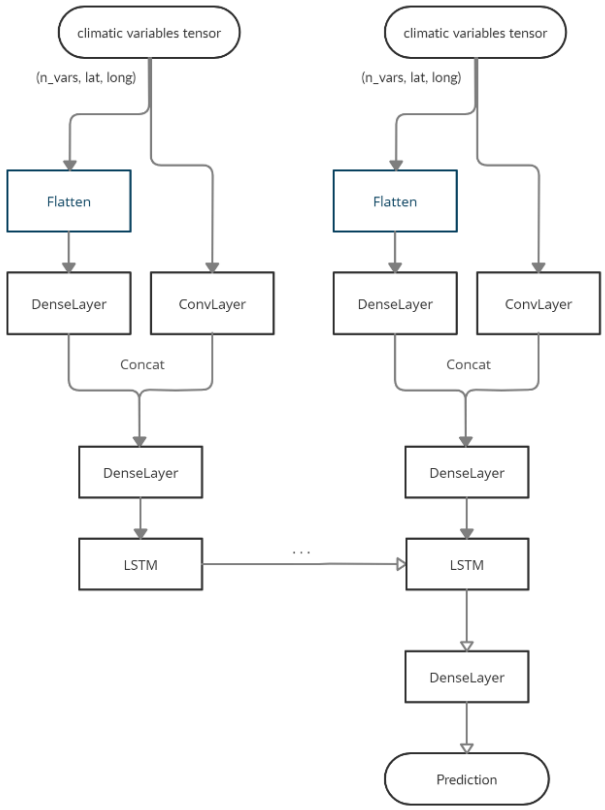



Fig. 2. Architecture of HailNet neural network

## V. Experiment results

In this section, we describe a baseline approach to the problem of hail forecasting. Later, we compare the performance of the HailNet to the baseline model and to a model introduced in [14]. In the last part, we present a sample of hail frequency forecast based on the CMIP5 MRI-CGCM3 (RCP85) model for Texas, part of New Mexico, and Oklahoma.

### A. Baseline approach description

At the beginning of our experiment, we decided to create a simple baseline solution to refer to. This solution to the problem is to group all features in bounds of one day and assume an absence of temporal and spatial structure in this timescale. The assumption is based on the fact that hail events are too local. It is enough to observe 24 hours to have all features corresponding to a single hail event.

Worth noting that ignoring spatial structure prevents the model to have any information about associated synoptic

processes, precisely, atmospheric fronts. But hail-producing thunderstorms are quite often associated with cold atmospheric fronts. Still, it is not obvious whether data with a resolution of 1°x1° still preserve spatial features corresponding to processes above the synoptic scale properly. And baseline assumption here - it is not. To train the baseline model CatBoost classifier [8] is used.

The parameters of the model are the following:

- iterations=1000
- learning_rate=0.01

The results of the baseline approach will be compared with HailNet in the next part of this section.

### B. HailNet

The first set of metrics we want to refer to is precision and recall. Values of those metrics both for the baseline CatBoost solution and the HailNet are presented in Table II, but values for the model from [14] are unavailable. As was stated before, we aimed to put more value to recall maximization, but still, it is not possible to ignore the precision metric.

Although we reached high values of metrics in focus for class 0: "No hail" the real interest is in 1: "Hail", because this class was initially underrepresented and was synthetically sampled with SMOTE. We observe good recall values for this class both for the baseline solution and the HailNet. This means that a great portion of real hail events was correctly labeled by our classifiers and the number of false negatives is small. Yet the baseline solution has remarkably small precision. This means that the model tries to be "overcautious" and labels a lot of actual 0: "No hail" events as 1: "Hail".

The HailNet shows a gain in precision with respect to the baseline model. It reaches 0.2 which is still not a high value of the metric. But in the setting of the experiment, this is acceptable.

TABLE II
PRECISION AND RECALL METRICS OF HAILNET AND BASELINE MODEL

| | Precision | Recall |
|---|---|---|
| HailNet | | |
| No hail: 0 | 0.98 | 0.76 |
| Hail: 1 | 0.2 | 0.79 |
| CatBoost baseline | | |
| No hail: 0 | 0.99 | 0.68 |
| Hail: 1 | 0.02 | 0.77 |

To ensure that our model is capable to catch the annually-periodic structure of hail occurrence RMSE between the normalized observed and modeled annual cycle of monthly mean hail frequency over the target area was calculated. The actual shapes of corresponding curves are depicted in Figure 3. We see that the HailNet approximates all months except June quite well. The June gap is a matter of further research.

The baseline model scored 0.19 for $RMSE_{om}$ which is already close to the values stated in [14]. Scores from the reference are approximately 0.18 in the best case and 0.20 for the model which is considered optimal with respect to a number of features involved. The HailNet achieved 0.16 and slightly outperformed [14] model in the $RMSE_{om}$ metric.

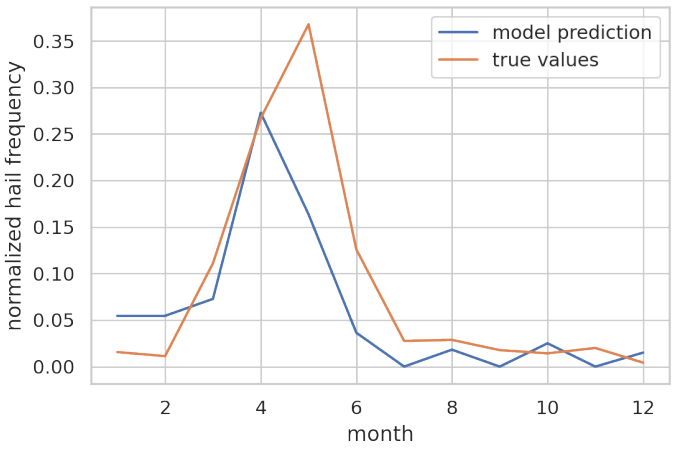



Fig. 3. Normalized observed and modeled annual cycle of monthly mean hail frequency

### C. CMIP-based forecast

Using data from CMIP5 MRI-CGCM3 model (scenario RCP8.5) we created a forecast of hail frequency. The sample of it for Texas, part of New Mexico, and Oklahoma for 2025 and 2030 years is presented in Figure 5. The hail frequency heat maps exhibit a clear persistent peak in Northern Texas. This peak is probably associated with the leeward side of mountains located to the West. The extent and direction of this influence are a matter of further research.

In the plain area, the pattern is more complicated. The hail frequency at some points may be comparable to the Northern counterpart, but there is no persistent geographical location for it.

In Figure 5 which depicts the difference in hail frequency between 2030 and 2025 years, we observe an increasing trend in the leeward area and high variability in frequency difference in plain area.

## VI. CONCLUSION

In this paper, we introduced a new model - HailNet, designed to produce climatological hail forecasts based on CMIP data. It slightly outperformed the model, designed for a close purpose [14].

In the final part of the paper, we presented a sample of hail forecast for Texas, part of New Mexico, and Oklahoma and performed a brief analysis of it.

Hail forecasting is a challenging task. One of the main reasons is the sparsity of the target. And the complexity even grows when scientists do not have a proper dataset of hail observations. It is the case for almost all countries except the USA and some parts of the European Union. But even with those available datasets, one should work with caution for a number of reasons. Single dataset (as with Storm Events

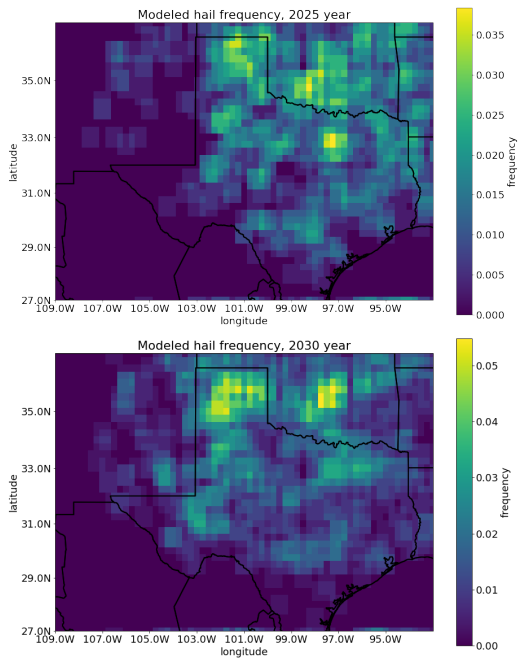



Fig. 4. Modeled hail frequency for 2025 and 2030 years based on MRI-CGCM3 (RCP8.5)

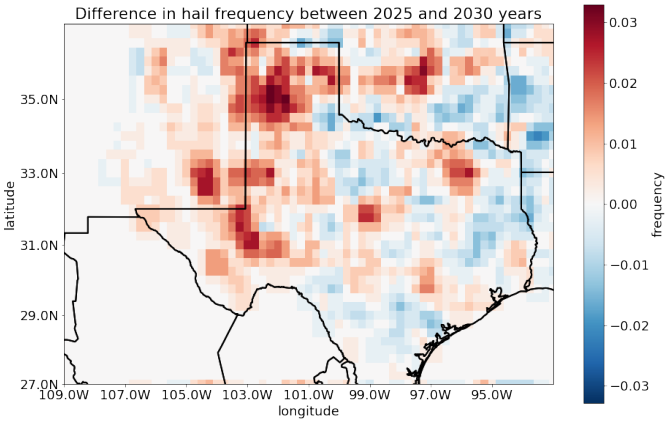



Fig. 5. Difference in hail frequency between 2030 and 2025 years

Database) may be compiled from different sources. This fact introduces inhomogeneity to data which in particular cases (e.g. fast expansion of weather station coverage) may be quite severe. Observation techniques and quality of observations, in general, improve through time. It creates an artificial increasing trend. All these factors should be treated while working with historical data.

But with side effects aside, improvements in meteorological stations' equipment, and expansion of station coverage will certainly yield a great benefit to hail forecasting in particular and meteorological forecasts in general.

## References


[1] Noah D Brenowitz, Brian Henn, Jeremy McGibbon, Spencer K Clark, Anna Kwa, W Andre Perkins, Oliver Watt-Meyer, and Christopher S Bretherton. Machine learning climate model dynamics: Offline versus online performance. *arXiv preprint arXiv:2011.03081*, 2020.

[2] Amanda Burke, Nathan Snook, David John Gagne II, Sarah McCorkle, and Amy McGovern. Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Weather and Forecasting*, 35(1):149–168, 2020.

[3] Yaping Cai, Kaiyu Guan, David Lobell, Andries B Potgieter, Shaowen Wang, Jian Peng, Tianfang Xu, Senthold Asseng, Yongguang Zhang, Liangzhi You, et al. Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches. *Agricultural and forest meteorology*, 274:144–159, 2019.

[4] Randy J Chase, David R Harrison, Amanda Burke, Gary M Lackmann, and Amy McGovern. A machine learning tutorial for operational meteorology, part i: Traditional machine learning. *arXiv preprint arXiv:2204.07492*, 2022.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[6] Changhyun Choi, Jeonghwan Kim, Jongsung Kim, Donghyun Kim, Younghye Bae, and Hung Soo Kim. Development of heavy rain damage prediction model using machine learning based on big data. *Advances in meteorology*, 2018, 2018.

[7] Jürgen Döllner. Geospatial artificial intelligence: Potentials of machine learning for 3d point clouds and geospatial digital twins. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1):15–24, 2020.

[8] Anna Veronika Dorogush, Vasily Ershov, and Andrey Gulin. Catboost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*, 2018.

[9] Peter D Dueben and Peter Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.

[10] Jason Kelley, Dalyn McCauley, G Aaron Alexander, Wilton F Gray, Rylie Siegfried, Holly J Oldroyd, et al. Using machine learning to integrate on-farm sensors and agro-meteorology networks into site-specific decision support. *Transactions of the ASABE*, 63(5):1427–1439, 2020.

[11] Patrick Köthur, Mike Sips, Andrea Unger, Julian Kuhlmann, and Doris Dransch. Interactive visual summaries for detection and assessment of spatiotemporal patterns in geospatial time series. *Information Visualization*, 13(3):283–298, 2014.

[12] T. W. Krauss, A. A. Sinkevich, N. E. Veremey, Yu. A. Dovgalyuk, and V. D. Stepanenko. Study of the development of an extremely high cumulonimbus cloud (andhra pradesh, india, september 28, 2004). *Russian Meteorology and Hydrology*, 32(1):19–27, Jan 2007.

[13] NSSL. Nssl's hail research, 2020.

[14] Andreas F. Prein and Greg J. Holland. Global estimates of damaging hail hazard. *Weather and Climate Extremes*, 22:10–23, 2018.

[15] Verisk. The hail hazard and its impact on property insurance. 2020.

[16] Duncan Watson-Parris. Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A*, 379(2194):20200098, 2021.