Big Data Research 23 (2021) 100178

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Deep Learning-Based Weather Prediction: A Survey

Xiaoli Ren [a,b], Xiaoyong Li [a,b,*], Kaijun Ren [a,b,c], Junqiang Song [a,b], Zichen Xu [d], Kefeng Deng [a,b], Xiang Wang [a,b]

[a] College of Meteorology and Oceanography, National University of Defense Technology, Changsha, Hunan, 410073, China
[b] College of Computer Science and Technology, National University of Defense Technology, Changsha, Hunan, 410073, China
[c] Laboratory of Software Engineering for Complex Systems, National University of Defense Technology, Changsha, Hunan, 410073, China
[d] College of Computer Science and Technology, Nanchang University, Nanchang, Jiangxi, 330031, China

ARTICLE INFO

Article history:
Received 27 July 2020
Received in revised form 28 October 2020
Accepted 15 November 2020
Available online 16 December 2020

Keywords:
Deep learning
Big meteorological data
Weather forecasting
Spatio-temporal feature
Time series

ABSTRACT

Weather forecasting plays a fundamental role in the early warning of weather impacts on various aspects of human livelihood. For instance, weather forecasting provides decision making support for autonomous vehicles to reduce traffic accidents and congestions, which completely depend on the sensing and predicting of external environmental factors such as rainfall, air visibility and so on. Accurate and timely weather prediction has always been the goal of meteorological scientists. However, the conventional theory-driven numerical weather prediction (NWP) methods face many challenges, such as incomplete understanding of physical mechanisms, difficulties in obtaining useful knowledge from the deluge of observation data, and the requirement of powerful computing resources. With the successful application of data-driven deep learning method in various fields, such as computer vision, speech recognition, and time series prediction, it has been proven that deep learning method can effectively mine the temporal and spatial features from the spatio-temporal data. Meteorological data is a typical big geospatial data. Deep learning-based weather prediction (DLWP) is expected to be a strong supplement to the conventional method. At present, many researchers have tried to introduce data-driven deep learning into weather forecasting, and have achieved some preliminary results. In this paper, we survey the state-of-the-art studies of deep learning-based weather forecasting, in the aspects of the design of neural network (NN) architectures, spatial and temporal scales, as well as the datasets and benchmarks. Then we analyze the advantages and disadvantages of DLWP by comparing it with the conventional NWP, and summarize the potential future research topics of DLWP.

1. Introduction

Since the beginning of human's history, people have always striven to predict and understand the world, and the ability to make better predictions has given competitive advantages in diverse contexts, such as weather [1].

Weather conditions such as temperature, humidity and wind, profoundly affect many aspects of human livelihood. Weather forecasting provide analytical support for issues related to intelligent transportation such as traffic flow prediction, air visibility analysis and so on [2,3]. Autonomous vehicles rely entirely on sensing and predicting the external environmental factors. As we know, traffic accidents and congestions are more likely to occur in severe weather conditions such as heavy rain, heavy fog, *etc.* Accurate and timely weather predictions play a fundamental role in the early warning of weather disasters. For example, nowcasting, especially quantitative precipitation nowcasting (QPN), which refers to forecast at the high spatio-temporal resolution (60-600 s, 100-1000 m) and short lead times of a few hours [4], and is particularly important to mitigate the impacts of heavy convective rainfall events to autonomous vehicles such as flooding, or sewage overflow in urban areas.

At present, almost all operational weather forecasting are based on Numerical Weather Prediction (NWP) [5], which is essentially a set of nonlinear equations, known as the primitive equations [6].

However, there are still several challenges in NWP [7]. Firstly, due to the chaotic nature of atmosphere [8], the small differences in initial conditions have a huge impact on model results. The measurement of initial conditions that input to the model, data as-

* Corresponding author at: College of Meteorology and Oceanography, National University of Defense Technology, Changsha, Hunan, 410073, China.
E-mail addresses: renxiaoli18@nudt.edu.cn (X. Ren), sayingxmu@nudt.edu.cn (X. Li), renkaijun@nudt.edu.cn (K. Ren), junqiang@nudt.edu.cn (J. Song), xuz@ncu.edu.cn (Z. Xu), dengkefeng@nudt.edu.cn (K. Deng), xiangwangcn@nudt.edu.cn (X. Wang).

https://doi.org/10.1016/j.bdr.2020.100178

similation, and incomplete understanding of the atmospheric physical processes inevitably introduce errors, which means that as the difference between the current time and the predicted time increases, the accuracy of the prediction will decrease. Secondly, a deluge of meteorological data has become available. On the one hand, the sensors and autonomous observing platforms have acquired PBs of meteorological observation data [9]. On the other hand, NWP models generate about TBs of simulation results each day. Furthermore, the diversity and many forms of uncertainties exist in the datasets, and the spatio-temporal correlations between datasets present unprecedented challenges to NWP. Last but not least, the numerical approach to solve the theory-based nonlinear equations is computational expensive, which strongly relies on the capability of supercomputers.

The artificial neural network (ANN) is a powerful data modeling tool that can capture and represent complex relationships between inputs and outputs, which is developed by the motivation of implementing artificial systems that can perform intelligent tasks similar to those performed by the human brain. In general, ANN is able to approximate any nonlinear function [10].

The deep neural network (DNN) is a kind of ANN composed of multilayers architecture, which can reconstruct the raw data sets from the original feature space into a learned feature space. In other words, they can "learn" features by neural networks (NNs) instead of selecting features manually [11], and achieve higher accuracy and better generalization with the learned features. Deep learning (DL) has achieved encouraging results in many areas, such as computer vision [12], speech recognition [13–15], natural linguistic programming [16], as well as in scientific fields in physics [17–19], chemistry [20] and bioinformation [21].

In recent years, DL has been applied to study time series problems [22], where the correlation between features is obvious, but hard to identify. Specifically, when the system behavior is dominated by spatial or temporal context (such as weather system), traditional machine learning (ML) approaches may not be optimal, whereas DL approach, which is able to extract spatio-temporal features automatically, is better to gain further understanding of the systems [1]. If the correlation is clearly analyzed and the features are correctly represented, then the predictive accuracy will be improved. Therefore, DL has been recognized as a reasonable and suitable tool for analyzing the characteristics of time series. Based on this, many scholars have applied DL to weather prediction, which is a typical multi-dimensional time series problem. It is expected that some of the conventional difficulties in weather forecasting will be addressed by the data-driven methods.

DL-based weather prediction (DLWP) has attracted attention in many fields, such as the authoritative meteorological research institution European Centre for Medium-Range Weather Forecasts (ECMWF) [23], academic journal Nature [1] and enterprises, *e.g.* Alibaba Group [24] and Google Research [25,26]. This paper surveys the state-of-the-art studies of DLWP. To the best of our knowledge, this is the first paper that surveys DL methods for weather prediction tasks. In particular, we not only focus on the NN architectures for various types of meteorological data, but also make comparative analysis from the perspectives of spatio-temporal scales, the datasets and benchmarks.

The rest of the paper is structured as follows: Section 2 describes the traditional weather prediction, and typical DNN models that suitable for meteorological data with different characteristics. Sections 3, 4, and 5 survey and analyze the state-of-the-art studies of DLWP from the perspective of architectures design for specific weather prediction tasks, the temporal and spatial scales of forecasting, as well as the datasets and benchmarks for evaluation. We compare DLWP with NWP, and analyze their merits and limitations in Section 6. Section 7 summarizes the paper and discusses the potential future research topics of DLWP.

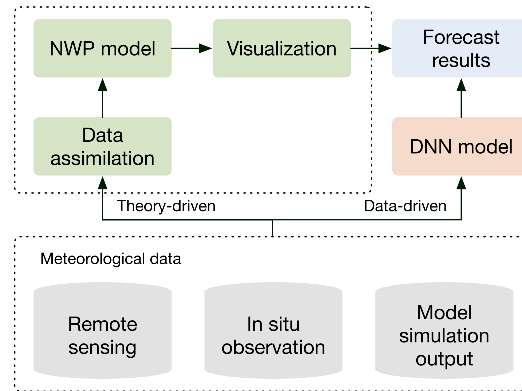


Fig. 1. Two paradigms for weather prediction: theory-driven and data-driven.

2. Weather prediction

For decades, weather prediction has been regarded as a physical theory problem, and meteorological scientists have been committed to improving the accuracy of forecasts through the understanding of physical mechanisms, which is a theory-driven approach. With the explosive growth of multi-source, multi-dimensional and multi-scale meteorological data, it has become a typical big spatio-temporal data. In recent years, data scientists have tried to apply data-driven computing paradigms to mine complex spatial and temporal relationships between meteorological data elements, and DLWP has become a hot research topic and is expected to be able to cope with the data challenges faced by traditional theory-driven approach. In this section, we briefly describe the two computing paradigms, which are illustrated in Fig. 1.

2.1. Numerical weather prediction

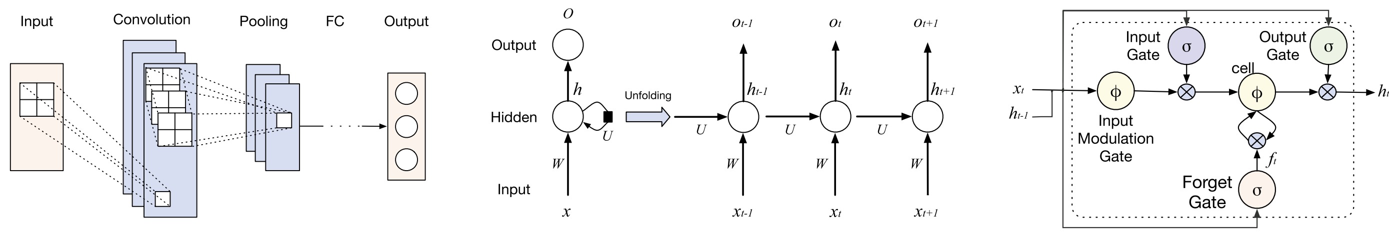
Traditionally, the weather states in the future are determined by integrating the governing partial differential equations (PDEs) based on the current weather states [7]. The nonlinear PDEs describe the dynamic, thermodynamic, radiative and chemical processes of atmosphere. Due to the continuity of the physical process, the PDEs of NWP have to be solved numerically by using spatial and temporal discretization.

NWP models usually forecast the meteorological elements such as temperature, wind speed, precipitation, average sea level pressure, *etc.* Generally, the main steps of NWP can be summarized as follows: *i*) obtaining the original observation datasets, including remote sensing data, in situ observation data, and the simulation results of the NWP models at last time point; *ii*) preprocessing the raw datasets through the data assimilation analysis system, including data quality control and quality assurance; *iii*) inputting the preprocessed datasets into the atmospheric dynamic model equations for prediction; and *iv*) post-processing and visualizing the results of prediction.

2.2. Deep learning-based weather prediction

2.2.1. Selecting of architectures towards data characteristic

DLWP is a data-driven approach. The original datasets are input to the DNN models, which aim to find the underlying laws or relationships from the input data, and capture the feature of weather changes through a large amount of data. According to the characteristics of the meteorological data, we analyze the most suitable DNN models for the corresponding data types, which is summarized in Table 1.



(a) *CNN* (b) *Unfolded RNN* (c) *LSTM block*

Fig. 2. Illustration of basic DNN architectures: (a) a CNN with a convolutional, a pooling and fully connected (FC) layers, (b) an unfolded RNN, and (c) a typical LSTM block.

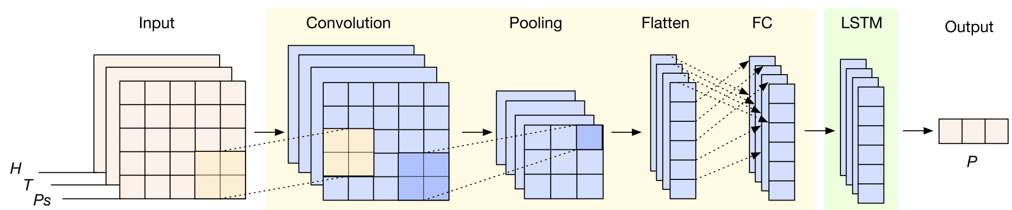


Fig. 3. Illustration of a CNN-LSTM hybrid architecture for DLWP. The CNN part is highlighted in yellow, and LSTM part is highlighted in green. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Table 1
Selection of DNN models.

Data characteristics	*Potential DNN architectures*
High-dimensional real-type	Autoencoder-based DNNs
Image	CNN-based DNNs
Long time sequence	RNN-, LSTM-based DNNs

(1) Multi-dimensional real-type data

The meteorological element data from in situ observation and model simulation outputs are multi-resolution and multi-dimensional (one to four), *e.g.*, the temperature of a region at a certain time can be represented by an array of the form $t(longitude, latitude, level, time)$, which is a four dimensional array with real-type values. The typical application of Autoencoder is to reduce data dimensions, and it is suitable for data of real-type [27,28]. Therefore, Autoencoder and its variants have a unique advantage in processing high-dimensional, real-type meteorological data collections.

(2) Satellite image data

At present, the remote sensing data like images from meteorological satellites exceed 100s of TBs per day, which are the important data for weather prediction, especially for the detection of extreme weather.

Convolutional neural network (CNN) [29,30] (as shown in Fig. 2(a)) is a kind of DNN based on feature representation, which is exclusively designed for image processing. The *weights* of the convolutional kernels are shared between neurons, and the pooling function both effectively reduce the number of hyper-parameters in each hidden layer, which avoid overfitting and falling into a local optimum. Therefore, CNNs are extensively applied for weather forecasting, especially for extreme weather detection through images.

(3) Long time sequence data

Meteorological data, especially climate datasets, are large amount of long time sequence data of decades or even hundreds of years, and there are temporal relationships between data elements.

Recurrent neural networks (RNNs) have been widely used for time series prediction problems because of the memory function [31]. Large datasets that can not be shown at once can be processed step by step using RNNs [32] (as shown in Fig. 2(b)). However, standard RNNs are difficult to train due to the well-known gradient vanishing and exploding [33], especially when they are used to problems that contain long-term dependencies. Comparing with the deep CNNs which could be over 100 layers [34], most of RNNs only consist of 2 or 3 layers [35,36].

Aiming at addressing the gradient problems of standard RNNs, long-short term memory (LSTM) network has been proposed as RNN variant [37–39]. Each LSTM block (as shown in Fig. 2(c)) contains one or more recurrently connected memory cells and three multiplicative units [40,41]: *i*) the input gate, which receives the current x_t and controls whether the LSTM considers its current input; *ii*) the forget gate, allows the LSTM to forget its previous memory c_{t-1}, which is the key to solve the gradient problems; and *iii*) the output gate, decides how much of the memory to transfer to the hidden state h_t, *i.e.*, what to output. This design makes LSTM quite effective for capturing long-term sequences, which is needed for long-term weather forecasting and climate simulation.

2.2.2. Modeling for weather prediction

Generally speaking, for a certain physical field x, single variable time series regression is usually used to find the relationship between its future state at time t and past states at time $t-1, t-2, \cdots, t-n$, and the future state is estimated by [42,43]:

$$x_t = f(x_{t-1}, x_{t-2}, \cdots, x_{t-n}) \tag{1}$$

where f can be obtained by training different NN models, and x_{t-1}, x_{t-2}, $\cdots$, x_{t-n} are the states of the past time points, which as the input of the NN models.

Fig. 3 illustrates a hybrid DNN architecture of DLWP that consists of three components.

1. Input layer. Available samples $\boldsymbol{x}$ are input to the DNN, *e.g.*, the meteorological attributes air pressure (P_s), temperature (T), and relative humidity (H) are used as the input of the network, *i.e.*, $\boldsymbol{x} = \{P_s, T, H\}$.
2. Hidden layers. The CNN part is used for capturing spatial features in the input data, while LSTM part is used for extracting temporal features from the time sequence of the input data,

Table 2
State-of-the-art hybrid DNN architectures for weather prediction tasks.

Typical tasks	*Modeled as*	*Typical hybrid DNN models*
Weather state prediction, *e.g.* precipitation prediction	Spatio-temporal sequence prediction	STConvS2S [50], ConvLSTM [51], TrajGRU [52], PredRNN [53,54], MetNet [25]
Extreme weather detection, *e.g.* typhoon	Classification, pattern recognition	Hybrid CNN-LSTM [55], multi-channel spatio-temporal convolutional encoder-decoder [56]

as well as the relationship between meteorological attributes. The number of hidden layers and neurons in each layer can be adjusted according to the performance of the network during training. Although increasing the number of hidden layers enables the network to capture more complex features, it also increases the training overhead.
3. Output layer. The output corresponding to the meteorological attribute that desired to forecast, *e.g.*, the rate of precipitation (P).

The typical training process for DNN is proposed in the revolutionary works by Hinton et al. [44], Bengio et al. [45], and Ranzato et al. [46]. The performance of the DNN is evaluated by the *loss* of the output, and its optimization can depend completely on the data features recognized by the network, or with the help of theory-guided methods, which are described in Section 3.

3. DNN models for weather prediction

From the perspective of DNN architectures for DLWP, the models can be grouped into three categories:

1. Architectures that based on the basic DNN models, such as those based on Autoencoders [47,48], CNNs [24], and LSTM networks [49];
2. Hybrid architectures composed of the basic DNN models to capture more complex temporal and spatial features, which are completely data-driven;
3. Coupling architectures of DNN and NWP models, which are not only data-driven, but also theory-guided. It is a novel research topic aims at improving the performance of forecasting.

In this paper, we mainly survey and analyze the last two types of DLWP models.

3.1. Completely data-driven hybrid architectures

Meteorological data exhibit both spatial and temporal structures, and data-driven weather prediction can be modeled as a sequence problem to exploit spatial and temporal features in the meteorological datasets [50]. We survey some of the state-of-the-art hybrid DNN models for DLWP and summarize them in Table 2.

3.1.1. Weather state prediction

A convolutional LSTM (ConvLSTM) network for precipitation nowcasting is proposed by Shi et al. [51], which has convolutional structures in both the input-to-state and state-to-state transitions. It consists of an encoding network and a forecasting network. Compared with the fully connected LSTM, the encoding part of ConvLSTM encodes the spatio-temporal relationships of meteorological data, which contributes to the improvements of the forecasting accuracy. However, the convolutional recurrence structure of convLSTM is location-invariant. To actively learn the location-variant structure in natural motion and transformation, the trajectory gated recurrent unit (TrajGRU) model is proposed in [52], which uses the current input and previous state to dynamically generate the local neighborhood set for each location at each timestamp. TrajGRU is efficient in capturing the spatio-temporal correlations in meteorological data.

Aiming at study a more general forecasting framework, Wang et al. [53] propose a well-performed predictive RNN (PredRNN), which adds extra connections between adjacent time steps in a core stacked spatio-temporal LSTM (ST-LSTM). It leverages a dual memory mechanism to extract and memorize both spatial and temporal variations of the sequences in a unified memory pool simultaneously. Furthermore, the improved predRNN (PredRNN++) is presented towards a resolution of the spatio-temporal predictive learning dilemma between deep-in-time structures and vanishing gradients [54]. PredRNN++ leverages a new recurrent structure (Causal LSTM) with cascaded dual memories to make the network deeper, and propose a gradient highway unit (GHU) to alleviate the gradient propagation difficulties. PredRNN and its variants are general frameworks and have been extended to precipitation nowcasting successfully [57].

Recently, Google Research has successively proposed DLWP models for high-resolution precipitation nowcasting to forecast rates of precipitation [25,26]. In [26], a ubiquitous U-Net CNN is leveraged, and the forecasting is treated as an image-to-image translation problem. Then in [25], a improved neural weather model (NWM) called MetNet is presented, which uses axial self-attention mechanisms. The convolutional LSTM block in [51] is used to process the downsampled time slices in the direction of time. MetNet is the first DLWP model that outperforms NWP at a certain spatial and temporal scale.

3.1.2. Extreme weather detection

Extreme weather (*e.g.*, typhoon) detection is crucial for disaster prevention and emergency decision-making. The data-driven methods can provide predictions within minutes of receiving new data, which may better suit the needs of highly responsive prediction service than traditional theory-driven NWP. On the other hand, supervised and semi-supervised DNNs are able to break through the limitations of threshold-based conventional detection approaches.

Toolkit for extreme climate analysis (TECA) is an application of large-scale pattern detection on climate data using heuristic methods [58,59]. Based on the output of TECA analysis, Liu et al. [60] apply deep CNN to predict the class label for two extreme weather event types, by considering the binary classification task on centered, cropped patches from 2-dimensional (2D) multi-channel images. In order to address multi-class detection and localization extreme weather, such as tropical cyclones, extra-tropical cyclones, tropical depressions and atmospheric rivers, Racah et al. [56] present a 3-dimensional (3D) (*i.e.*, height, width, time) multi-channel spatio-temporal convolutional encoder-decoder. This study is the first use of deep Autoencoding architecture for bounding-box regression using semi-supervised learning, by training the Autoencoder with reconstruction for unlabeled data, which effectively overcomes the difficulty of labeling meteorological datasets.

To alleviate the huge loss of life and property that may caused by typhoon, the prediction of its formation and intensity has become imperative. Chen et al. propose a hybrid CNN-LSTM model [55] for forecasting both typhoon formation and intensity.

Considering that typhoon is a strong convective weather process which is affected by several correlated meteorological attributes, the hybrid model is designed to capture the complex spatial and temporal features by three components: *i*) a 2-dimensional CNN (2DCNN), which extracts features from the local neighborhood of the previous feature map and is used to analyze 2D sea surface variables, *e.g.*, sea surface temperature (SST); *ii*) a CNN with 3-dimensional filter (3DCNN) [61], which is leveraged to capture the spatial correlations between 3D atmospheric variables, *e.g.*, wind and air pressure; and *iii*) a LSTM, which is designed to capture the temporal correlations. The performance of the hybrid model is superior to the previous methods of typhoon forecasting, including its formation and intensity.

3.2. Theory-guided coupling architectures

The paradigm of theory-guided data science [62] attempts to explore the continuum between theory-driven and data-driven models by integrating scientific knowledge in data science models. As far as we know, there are several ways to integrate numerical simulation and DL, either to enhance the interpretability of DLWP, or to improve the forecasts accuracy and timeliness of NWP.

3.2.1. Enhancing DNN models with theoretical knowledge

As described in Section 2.1, the classical numerical models are described by a set of primitive equations, which are based on the theoretical physical principles including Newton's second law of motion, the law of conservation of mass, the first law of thermodynamics, the ideal gas law and hydrostatic law. Therefore, NWP models can not only capture the spatio-temporal dynamics of multiple meteorological elements simultaneously, but also take into account the correlation between the diverse variables. Many important aspects of observed weather and climate can be reproduced by numerical models. Nevertheless, the completely data-driven DNN models are far from causal discovery from observational data [63] and model the whole complex weather system. Using scientific prior knowledge as constraints to train DNN may greatly help to ensure physical consistency and enhance the interpretability of DNN models.

(1) Training DNN with knowledge from numerical models

Based on LSTM Autoencoder [64], Wang et al. [65] cast weather forecasting problem as an end-to-end DL problem, and propose an effective information fusion mechanism to learn from historical data that incorporates prior knowledge from NWP, which enables to forecast multiple meteorological variables. With the novel negative log-likelihood error loss function, the proposed approach simultaneously implements single-value forecasting and uncertainty quantification. This is the first method of combining DNN with NWP to weather forecasting. Aiming at processing real-time series, the authors of [66] design a real-time NN architecture to predict meteorological states, which is a multi-layer perceptron NN. Based on the ECMWF data collection and the parameters of the selected sensors, the application can deliver outputs in real-time.

(2) Refining output of DNN by theoretical domain knowledge

Considering the fact that there are spatio-temporal dependencies among meteorological attributes, Grover et al. [67] propose a hybrid architecture that consists of three components: *i*) a set of bottom-up predictor for each individual attribute that are trained using historical data; *ii*) constraints of physical laws that constraining the output of the separate predictors to be spatially smooth; *iii*) a top-down deep belief network (DBN) consists of layers of stacked Restricted Boltzmann Machines (RBM), which models the joint statistical relations. The approach has the advantage of imposing long-range dependencies across space, which enables optimization of the predictive model in accordance with large-scale phenomena.

3.2.2. Improving numerical models by deep learning

With the rapid growing of meteorological data amount and the continuous improvement of models' resolution, the computing requirement of NWP has increased dramatically, which leads to the inefficiency of NWP. DNN models can "learn" the complex behavior of the system and model the process quickly, thereby avoiding solving the complex PDEs. The trained DNN can be used to emulate a module or process of NWP models to improve the accuracy or timeliness.

(1) Emulating the physics or dynamics of numerical models by DNN

Instead of using DNN to extract information from numerical models, Scher et al. design a deep CNN to completely emulate a simple global circulation model (GCM), which is a numerical model for climate forecasting. The deep CNN is trained on the GCM. It takes the complete model states of the GCM as its input, and learns to emulate the dynamics of the GCM, and then predicts the next model state [68]. Similarly, the authors of [69] use a DNN to efficiently model the motion of water in the ocean and predict SST. The general background knowledge gained from the physics is used as a guideline for designing the deep leaning methods. Specifically, the motion field is learned by a DNN, which is used to update SST via physically modeling its movement.

(2) Improving parameterizations of numerical models using DNN

The theory-driven numerical models require parameters, however, many of those cannot be easily derived from the physical principles due to the complexity of the systems, *e.g.*, significant uncertainty persist in the parameterization of clouds [70], of which the horizontal resolution can be as small as a few hundred meters, while the grid resolution of a GCM may be around 50 to 100 kilometers. This is why subgrid processes are introduced to GCMs, and errors are introduced at the same time. To improve the accuracy of subgrid parameterizations, the authors of [71] train a DNN to represent all atmospheric subgrid processes in the GCM by learning from a multi-scale model. Then the trained DNN replaces the traditional subgrid parameterizations in the GCM. It has been demonstrates that the trained DNN works fast and accurate.

4. Temporal and spatial scales

NWP models are usually run with certain time scales and spatial resolutions. The investigated domain (local, regional or global) is transformed into a grid in numerical models. Spatial horizontal resolution specifies how large the grid cells are [72]. Generally, the higher the resolution, the higher the prediction accuracy, but at the same time, the larger the calculation amount and the lower the timeliness. In general, the temporal scales include nowcasting (within 12 hours), short-term (within 3 days), medium-term (4 to 10 days), and long-term (more than 10 days).

In this section, we first analyze the limitations of NWP at different temporal and spatial scales, which is described in Table 3, and then we survey the typical DLWP studies which aim to address these limitations.

Table 3
Analysis of temporal and spatial scales.

Scales	*Key problems*	*Limitations of NWP*	*Solutions of DLWP*
Small-scale	High forecast accuracy requires high resolution	Computational limits due to resolution	Precise forecasting with high resolution
Large-scale (for extreme weather detection)	Big datasets inputs, thresholds setting	Dependent on physical model, subjective thresholds, long simulation time	Capturing spatio-temporal features from data
Multi-scale	Multi-resolution datasets	Computational limits, dependent on completely different physical and dynamical models	Learning mapping between variables of different scales

4.1. Precise forecasting of small-scale

In the aspect of spatial resolution, for some small cities or towns, usually only two or three grid cells of NWP models describe the area, which cannot approximate the local phenomenon adequately, but urban meteorological disasters seriously affect the lives of residents. In view of this, Frnda et al. [66] present a real-time DNN architecture to predict atmospheric states of urban areas for the next 3 days. The main objective of the model is to enhance the prediction in main cities.

In terms of temporal scale, very-short-term (within 1 hour) local weather forecasting is getting more and more anticipated in industrial field. Due to the lack of surface weather data and limitation of computational resources, the short-term weather forecasting within 3 hours is difficult. On the one hand, the atmospheric data provided by NWP or radars might not be accurate because of the larger mesh or other uncertain factors. Therefore, dense surface weather data is essential. On the other hand, powerful computational resources are needed to analyze the large amount of surface weather data generated by the dense weather stations or sensors. To this end, a deep learning method is proposed in [73] to predict weather element in a very-short-term from densely observed data, which can analyze the large amount of surface weather data effectively and timely. MetNet [25] is a typical study on high spatial and temporal resolution forecasting.

4.2. Climate modeling of large-scale

As described in [60], only satellites obtain 10s of TBs of global data each year to record the evolution of the climate system. Besides, high resolution numerical climate models generate 100s of TBs of data from multi-decadal run. The large amount of climate datasets presents an unprecedented challenge to the post-processing and quantitative assessment for climate research. Although climate scientists use basic ML techniques to analyze climate data, such as principal component analysis (PCA) for dimensionality reduction [74], and k-means analysis for clusterings [75], they ignore the important spatially and temporally resolved extreme weather events.

Identifying extreme climate events like tropical cyclones in large-scale climate simulations is a major challenge for climate science. Existing extreme climate events detection methods are based on evaluating of relevant spatial and temporal variables on subjective thresholds. However, the thresholds may not be universally accepted [76]. In this context, Liu et al. [60] propose the methodology of detecting extreme events by deep CNN, which learns a class of patterns from complex multi-variable climate data and avoids subjective threshold. This is the first study of applying deep CNN to tackle climate pattern recognition problems, and it is improved by Racah et al. [56], which focuses on extreme event detection on planetary-scale data that consisting of global simulations over many years.

In addition to extreme weather, parameterizations, which is a kind of physical approximations, have been heuristically developed to represent the effects of sub-grid processes on the resolved scales [77]. However, there are imperfections in the parameterizations due to the sheer complexity of the underlying physical system, which have impeded progress toward more accurate climate predictions for decades [78,79]. In this respect, Rasp et al. [71] present an objective, data-driven approach of using high-resolution modeling data for climate model parameterization, which is a general methodology that not limited to the atmosphere but can equally as well be applied to other components of the Earth system and beyond.

4.3. Multi-scale forecasting

Different from the theory-driven numerical models, data-driven DL can “learn” from existing data, and can work in multiple scales, *e.g.*, in [80], the authors generate a unique multi-resolution perturbed parameter ensemble of a global climate model, and use the basic ML method random forests to train a statistical model on the ensemble to make high-resolution predictions by only using a small number of high-resolution runs, but supplemented with low-resolution runs, which is much cheaper and greatly reducing the computational expense.

Scher et al. [68] propose an approach using DL to approximate a simple GCM, the Portable University Model of the Atmosphere (PUMA), which directly emulates the complete physics and dynamics of the GCM, and generates a deep CNN that takes the complete GCM model state as its input and then predicts the next model state. The trained DNN can successfully predict the model several time steps ahead, which is making weather forecasting. In addition, after initializing the DNN with a random initial from the climate model, and repeatedly feeding back the prediction into the input of the network, the DNN can create a complete climate run. The unique advantage of this method is that the network is stable even for very long runs (1,000 years), which outperforms the other ML-based methods for multi-resolution [23,80]. This study shows that it is possible to let a NN learn the time evolution and the dynamics of the GCM.

4.4. Conclusion for temporal and spatial scales

Based on the analysis in Sections 4.1 to 4.3, we compare the temporal and spatial scales, where NWP and DLWP perform best respectively, (in terms of forecast accuracy, timeliness, and stability of the models), and the results are illustrated in Fig. 4.

As shown in Fig. 4, DLWP covers the area of short-term and small-scale weather prediction, which means in this area, forecast accuracy and timeliness of DLWP is higher than NWP, especially for local weather prediction, however, the stability of DLWP in long-term and large-area weather prediction is lower than NWP.

5. Evaluation of DLWP

As a real-world problem that has a profound impact on our daily life, the accuracy of weather forecasting is crucial. For DLWP,

Table 4
Typical datasets and benchmarks for DLWP.

Datasets	Observation	IGRA[1] CHIRPS[2] Station Observations
	Reanalysis	CFSR[3] ERA-Interim reanalysis[4] 20 century reanalysis[5] NCEP-NCAR reanalysis[6]
	Simulations	CAM run Published by ECMWF
	Hybrid	Hybrid of observation, reanalysis, simulation or datasets generated by data mining
Benchmarks	Numerical models	ECMWF (global model) ALADIN (regional model) HRRR (high resolution model)
	Traditional ML methods	SVM, LR, *etc.*
	Specific for DLWP	rainymotion[7] ExtremeWeather[8,9]

[1] https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive.
[2] https://chc.ucsb.edu/data/chirps.
[3] https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr.
[4] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim.
[5] https://rda.ucar.edu/datasets/ds131.3/.
[6] https://rda.ucar.edu/datasets/ds090.0/.
[7] https://github.com/hydrogo/rainymotion.
[8] extremeweatherdataset.github.io.
[9] https://github.com/eracah/hur-detect.

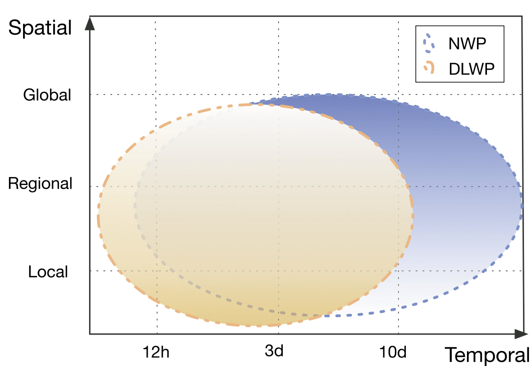


Fig. 4. Performance comparison between DLWP and NWP at different temporal and spatial scales. In the area covered by each approach, the darker the color, the better the performance.

the accuracy is determined by multiple factors including the models and the input datasets. Due to the weak interpretability of DLWP, researchers are implicitly required to compare the results with reliable benchmarks or real datasets to verify the correctness. In this section, we survey the typical datasets and benchmarks for weather forecasting, and summarize them in Table 4.

5.1. Datasets

We have mentioned that NWP is quite sensitive to the initial value, so is DLWP. It can be said that the quality of the input datasets determines the accuracy of the forecast results to a certain extent. As we know, the sensors and autonomous observing platforms such as ocean-based, ground-based, air-based and space-based have obtained PBs of meteorological observation data. In addition, NWP models generate about TBs of simulation results each day. According to the typical DLWP studies, we summarize the commonly used datasets of DLWP into four categories: *i)* observation datasets; *ii)* reanalysis datasets (obtained by assimilating disparate observational data into numerical models); *iii)* simulation output datasets and *iv)* hybrid datasets.

5.1.1. Observation datasets

As the growing availability of meteorological data, most researchers evaluate their methods with experiments on real-world datasets to highlight the promise of the approaches. Grover et al. [67] extract five years of historical data (2009-2014) from the Integrated Global Radiosonde Archive (IGRA) dataset [81]. The extracted dataset consists balloon observations recorded at 60 locations across the continental US. Nascimentoin et al. [50] use a subset of the satellite imagery and in situ station data from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) to evaluate the accuracy of rainfall forecast of their model.

5.1.2. Reanalysis datasets

Roesch et al. [82] use an ERA-Interim reanalysis [83] of the ECMWF for training (data from 1990 to 1999) and testing (data from 1990 to 2016) their models. The data has a spatial resolution of 0.75 / 0.75 degrees and the time step is 6 hours. In addition to using the observation data, Nascimentoin et al. [50] also use the air temperature from the Climate Forecast System Reanalysis (CFSR) to evaluate their approach. The data has a spatial resolution of 0.5 / 0.5 degrees and a frequency of 6 hours.

5.1.3. Simulation output datasets

Most studies of climate forecasting and extreme events detection using simulations products as data sources to evaluation their methods. The work [56] focuses on extreme event detection on planetary-scale data. The authors analyze output from the climate model's simulation results to understand the earth climate changes by the year 2100.

5.1.4. Hybrid datasets

Frnda et al. [66] take full advantages of NWP, using parameters of the selected sensors, along with the standard ECMWF output data as the input parameter of their multi-layer NN model. Specifically, instead of feeding raw observation meteorological data, the authors take the following three kinds of elements as the input of the multilayer NN model: *i*) the prediction results of temperature and precipitation by the regional numerical model ALADIN and the global numerical model ECMWF (temperature and precipitation are

Table 5
Summary of the typical DLWP studies.

Model type	*Literature*	*Specific model*	*Scale (temporal / spatial)*
Basic DNNs	[47,48]	Autoencoders	Short-term / local
	[49]	LSTMs	
	[24]	CNNs	Short-term / regional
	[60]		Extreme weather
	[73]	Deep fully connected NN	Nowcasting / local
Hybrid DNNs	[51,52]	CNN + LSTM	Short-term / local
	[25]		Nowcasting / regional
	[82]		Short-term / regional
	[55]		Extreme weather
	[56]	CNN + Encoder-decoder	
	[50]		Short-term / local
DNN and NWP Coupling models	[65,66]	Training DNN with NWP knowledge	
	[67]	Refining DNN output by domain knowledge	Short-term / regional
	[68]	Emulating the physics or dynamics of numerical models by DNN	Short or long-term / global
	[71]	Improving parameterizations of numerical models by DNN	Long-term / global

the two attributes to be predicted); *ii*) the key features that affecting the prediction attributes, which are identified by data mining technologies from NWP data; and *iii*) the day of forecast (1-3). Liu et al. [60] use both climate simulations and reanalysis products to evaluate the multi-classification of extreme events, where the simulations are generated by the Community Atmosphere Model 5.1 (CAM5.1) historical run, the reanalysis products consist of ERA-Interim reanalysis [83], 20 century reanalysis, and NCEP-NCAR reanalysis.

5.2. Benchmarks for comparison

Generally speaking, the benchmarks for NWP models to compare are the real datasets or the more authoritative numerical models internationally recognized, which are also applied to DLWP models. For example, the Coupled Model Intercomparison Project (CMIP) has successfully provided the climate community with a rich collection of simulation output from Earth system models [84]. Besides, there are already studies that provide benchmarks particularly for DLWP, and a few DLWP studies use traditional ML methods as benchmarks, such as SVM, LR, *etc.* [47,65,73]. We mainly survey the most reliable numerical models and the specific benchmarks in the following.

5.2.1. Results from numerical models

NWP models are the most commonly used benchmarks. Qiu et al. [24] evaluate the performance of their model with the public rainfall prediction system ECMWF, which is the operational system in the whole Europe. In [66], the authors show the accuracy comparison of their systems with the corresponding forecast obtained from ECMWF and ALADIN, respectively. The authors of [67] compare their method with the Winds Aloft forecast, which is state-of-the-art in wind prediction and is released by the National Oceanic and Atmospheric Administration (NOAA). The latest high spatial and temporal resolution MetNet [25] is compared with the High Resolution Rapid Refresh (HRRR) system, which is the current best operational NWP available from NOAA.

5.2.2. Specific benchmarks

The “rainymotion” software library is developed by Ayzel et al. [4], which consists of a stack of models based on different optical flow algorithms, and has been demonstrated that these models provide skillful predictions comparable with or even superior to the state-of-the-art operational system. The developers have opened it for further development in QPN models and could be served as a benchmark. For extreme events detection, the datasets *ExtremeWeather* and the codes are shared with researchers, which can be used freely as benchmarks [56].

6. Discussion of DLWP

Based on the above sections, we summarize the typical DLWP studies in Table 5. It has been shown that DLWP behaves as well as NWP, or even outperforms NWP in certain conditions [24,25,32,66]. There are discussions on whether DLWP can substitute NWP [23,85]. In this paper, we analyze DLWP and NWP in the aspects of computability, forecast timeliness, comprehensiveness, spatio-temporal scale and interpretability, which are the basic and actual requirements on weather forecasting.

(1) Computability

In order to solve the PDEs of NWP at each time step with large amount of input data, high-performance computing (HPC) centers must be deployed. Doubling the resolution requires approximately 8 times more computation [7]. NWP depends heavily on the computing capability of high-performance computers and parallel computing technology [72,86,87]. Whereas for DLWP, HPC centers are not necessary [66]. At present, most NNs are trained with GPUs, TPUs [88] or special artificial intelligence hardware [89], such as PlumeNet [90], the forecasting of large-scale air quality can be built in a few minutes on a standard GPU and updated frequently. Certainly, supercomputers can also be used to improve training efficiency [60,91]. In this respect, DLWP lowers the threshold for researchers to conduct weather forecast study, especially for those in industry fields that are obviously affected by weather conditions, such as retailing business and transportation.

(2) Timeliness

Once the DNN models are trained, they can predict orders of magnitude faster than the NWP models at the same resolution [1]. A typical example is MetNet [25], the latency of which is in the order of seconds, whereas that of NWP is the order of tens of minutes to hours under the same condition. In addition, the performance of data-driven DLWP models are not directly tied to the resolution. While for NWP, the higher the resolution, the larger the amount of computation, and the longer latency correspondingly. In this sense, DLWP is able to outperform NWP for emergency managements, such as precipitation nowcasting and extreme weather detection, which require timely decision.

(3) Comprehensiveness

Based on the law of conservation of mass, thermodynamics, and the gas law, the primitive equations represent the full set of prognostic equations upon which the change in space and time of wind, pressure, density and temperature is described [7,92]. Generally, the numerical methods simulate all these fields and output the future conditions. To the best of our knowledge, most of the proposed DLWP models are specially designed or trained for forecasting certain meteorological attributes or weather phenomenon. Consequently, the generalization ability of most DLWP models is limited, although exception exist [65]. But on the other hand, the predictive accuracy of specially designed DLWP may be higher than the numerical methods for the specific variables [25] or events [60].

(4) Scales

The scales refer to spatial and temporal scales: *i*) Spatial scale, as described in Section 4, DLWP can be designed particularly for forecasting precisely the weather condition of a small region (*e.g.* a town) with high spatial resolution (1 km × 1 km), whereas it will cost a lot for NWP due to the large amount computation at the same resolution. However, it is difficult to train a DNN for long-range or global forecasting, it would require a serious level of complexity as many as or even more than the traditional models, due to the large amount of tunable hyper-parameters [23]. *ii*) Temporal scale, most of the DLWP models are trained for short-term prediction (from minutes to hours), except for a few DLWP models that coupled with numerical models for long-term prediction [68].

(5) Interpretability

NWP models predict the future atmospheric behavior based on the current state and physics principles, which offer the potential of extrapolation beyond observed conditions [1,66]. However, DNN models are very difficult to be interpreted due to the hidden parameters. Interpretability has been identified as a potential weakness of DL [93]. Credibility and reproducibility of the simulation results are crucial, especially for applications such as weather disaster warning. Based on the perception, researchers have tried to couple the physical models with ML, both in the aspects of approaches [68,94,95] and evaluation benchmarking [4,84].

Based on the above analysis, we illustrate the comparison of the two approaches in Fig. 5. Notwithstanding that the scientific paradigm of the theory-driven NWP and data-driven DLWP is very different, they are complementary overall. The advantage of DLWP is that they are highly flexible in adapting to data, and are amendable to find potential patterns that have not been found due to insufficient understanding of physical mechanisms.

7. Summary

As a scientific problem that needs to process a large amount of observational data, weather forecasting can be implemented through data-driven deep learning approach (DLWP), which is expected to complement and enrich the conventional theory-driven

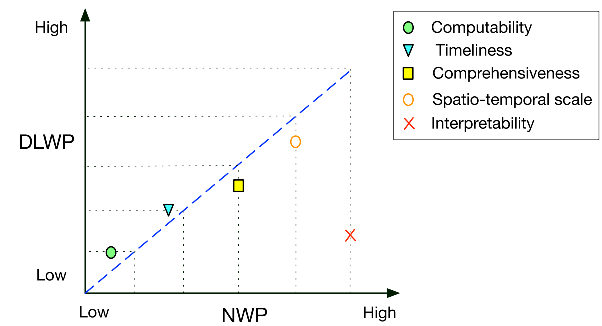


Fig. 5. DLWP *vs.* NWP. The area above the blue line indicates that DLWP is better than NWP, otherwise, NWP is better than DLWP.

method (NWP). This paper surveys the application of deep learning in weather forecasting from different perspectives. By comparing the existing studies of DLWP with the conventional NWP, we suggest that the future study of DLWP may focus on the following aspects: the interpretability of the results, prediction of more meteorological attributes with consideration of their correlation, and long-term and large-range prediction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the National Key R&D Program of China (Grant No. 2018YFB0203801), the National Natural Science Foundation of China (Grant Nos. 61702529 and 61802424), China National Special Fund for Public Welfare (Grant No. GYHY201306003).

References

[1] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, et al., Deep learning and process understanding for data-driven Earth system science, Nature 566 (7743) (2019) 195–204.
[2] T. Gneiting, A.E. Raftery, Weather forecasting with ensemble methods, Science 310 (5746) (2005) 248–249.
[3] Y. Zheng, L. Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications, ACM Trans. Intell. Syst. Technol. 5 (3) (2014) 1–55.
[4] G. Ayzel, M. Heistermann, T. Winterrath, Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1), Geosci. Mod. Dev. 12 (2019) 1387–1402.
[5] G. Marchuk, Numerical Methods in Weather Prediction, Elsevier, 2012.
[6] B. Golding, Mesoscale meteorological modelling, Q. J. R. Meteorol. Soc. 128 (585) (2002) 2557–2558.
[7] P. Bauer, A. Thorpe, G. Brunet, The quiet revolution of numerical weather prediction, Nature 525 (7567) (2015) 47–55.
[8] A. Abraham, N.S. Philip, B. Nath, P. Saratchandran, Performance analysis of connectionist paradigms for modeling chaotic behavior of stock indices, in: Second International Workshop on Intelligent Systems Design and Applications, Computational Intelligence and Applications, Dynamic Publishers Inc., USA, 2002, pp. 181–186.
[9] A. Agapiou, Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth engine© applications, Int. J. Digit. Earth 10 (1) (2017) 85–102.
[10] M.A. Nielsen, Neural Networks and Deep Learning, vol. 2018, Determination Press, San Francisco, CA, USA, 2015.
[11] X. Wang, Q. He, Enhancing generalization capability of svm classifiers with feature weight adjustment, in: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer, 2004, pp. 1037–1043.
[12] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[13] A.-r. Mohamed, G.E. Dahl, G. Hinton, et al., Acoustic modeling using deep belief networks, IEEE Trans. Audio Speech Lang. Process. 20 (1) (2012) 14–22.

[14] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, X-vectors: robust dnn embeddings for speaker recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2018, pp. 5329–5333.
[15] O. Novotnỳ, O. Plchot, O. Glembek, L. Burget, et al., Analysis of dnn speech signal enhancement for robust speaker recognition, Comput. Speech Lang. 58 (2019) 403–421.
[16] Y. Bengio, Deep learning of representations for unsupervised and transfer learning, in: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 2012, pp. 17–36.
[17] P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nat. Commun. 5 (2014) 4308.
[18] K.G. Lore, D. Stoecklein, M. Davies, B. Ganapathysubramanian, S. Sarkar, Hierarchical feature extraction for efficient design of microfluidic flow patterns, in: Feature Extraction: Modern Questions and Challenges, 2015, pp. 213–225.
[19] W. Bhimji, S.A. Farrell, T. Kurth, M. Paganini, E. Racah, et al., Deep neural networks for physics analysis on low-level whole-detector data at the lhc, arXiv preprint, arXiv:1711.03573.
[20] K.T. Schütt, F. Arbabzadah, S. Chmiela, K.R. Müller, A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, Nat. Commun. 8 (1) (2017) 1–8.
[21] Z. Chen, M. Pang, Z. Zhao, S. Li, R. Miao, Y. Zhang, X. Feng, X. Feng, Y. Zhang, M. Duan, et al., Feature selection may improve deep neural networks for the bioinformatics problems, Bioinformatics 36 (5) (2020) 1542–1552.
[22] J.N. Liu, Y. Hu, Application of feature-weighted support vector regression using grey correlation degree to stock price forecasting, Neural Comput. Appl. 22 (1) (2013) 143–152.
[23] P.D. Dueben, P. Bauer, Challenges and design choices for global weather and climate models based on machine learning, Geosci. Model Dev. 11 (10) (2018) 3999–4009.
[24] M. Qiu, P. Zhao, K. Zhang, J. Huang, X. Shi, X. Wang, W. Chu, A short-term rainfall prediction model using multi-task convolutional neural networks, in: Data Mining (ICDM), IEEE International Conference on, IEEE, 2017, pp. 395–404.
[25] C.K. Sønderby, L. Espeholt, J. Heek, M. Dehghani, A. Oliver, T. Salimans, S. Agrawal, J. Hickey, N. Kalchbrenner, Metnet: a neural weather model for precipitation forecasting, arXiv preprint, arXiv:2003.12140.
[26] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, J. Hickey, Machine learning for precipitation nowcasting from radar images, arXiv preprint, arXiv:1912.12132.
[27] P. Baldi, Autoencoders, unsupervised learning, and deep architectures, in: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 2012, pp. 37–49.
[28] H. Kamyshanska, R. Memisevic, The potential energy of an autoencoder, IEEE Trans. Pattern Anal. Mach. Intell. 37 (6) (2014) 1261–1273.
[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 1–9.
[30] G. Cheng, C. Yang, X. Yao, L. Guo, J. Han, When deep learning meets metric learning: remote sensing image scene classification via learning discriminative cnns, IEEE Trans. Geosci. Remote Sens. 56 (5) (2018) 2811–2821.
[31] S. Li, W. Li, C. Cook, C. Zhu, Y. Gao, Independently recurrent neural network (indrnn): building a longer and deeper rnn, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 5457–5466.
[32] I. Roesch, T. Günther, Visualization of neural network predictions for weather forecasting, in: Computer Graphics Forum, vol. 38, Wiley Online Library, 2019, pp. 209–220.
[33] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw. 5 (2) (1994) 157–166.
[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[35] Q.V. Le, N. Jaitly, G.E. Hinton, A simple way to initialize recurrent networks of rectified linear units, arXiv preprint, arXiv:1504.00941.
[36] A. Shahroudy, J. Liu, T.-T. Ng, G. Wang, Ntu rgb+ d: a large scale dataset for 3d human activity analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 1010–1019.
[37] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[38] R. Jozefowicz, W. Zaremba, I. Sutskever, An empirical exploration of recurrent network architectures, in: International Conference on Machine Learning, ICML, 2015, pp. 2342–2350.
[39] K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, Lstm: a search space Odyssey, IEEE Trans. Neural Netw. Learn. Syst. 28 (10) (2016) 2222–2232.
[40] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, K. Saenko, Translating videos to natural language using deep recurrent neural networks, arXiv preprint, arXiv:1412.4729.
[41] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 2625–2634.
[42] J.N. Liu, Y. Hu, J.J. You, P.W. Chan, Deep neural network based feature representation for weather forecasting, in: Proceedings on the International Conference on Artificial Intelligence (ICAI), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014, p. 1.
[43] J.N. Liu, Y. Hu, Y. He, P.W. Chan, L. Lai, Deep neural network modeling for big data weather forecasting, in: Information Granularity, Big Data, and Computational Intelligence, Springer, 2015, pp. 389–408.
[44] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[45] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, in: Advances in Neural Information Processing Systems, NIPS, 2007, pp. 153–160.
[46] M. Ranzato, C. Poultney, S. Chopra, Y.L. Cun, Efficient learning of sparse representations with an energy-based model, in: Advances in Neural Information Processing Systems, NIPS, 2007, pp. 1137–1144.
[47] M. Hossain, B. Rekabdar, S.J. Louis, S. Dascalu, Forecasting the weather of Nevada: a deep learning approach, in: International Joint Conference on Neural Networks, IJCNN, IEEE, 2015, pp. 1–6.
[48] S.-Y. Lin, C.-C. Chiang, J.-B. Li, Z.-S. Hung, K.-M. Chao, Dynamic fine-tuning stacked auto-encoder neural network for weather forecast, Future Gener. Comput. Syst. 89 (2018) 446–454.
[49] Z. Karevan, J.A. Suykens, Spatio-temporal stacked lstm for temperature prediction in weather forecasting, arXiv preprint, arXiv:1811.06341.
[50] R.C. Nascimento, Y.M. Souto, E. Ogasawara, F. Porto, E. Bezerra, Stconvs2s: spatiotemporal convolutional sequence to sequence network for weather forecasting, arXiv preprint, arXiv:1912.00134.
[51] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, Convolutional lstm network: a machine learning approach for precipitation nowcasting, in: Advances in Neural Information Processing Systems, NIPS, 2015, pp. 802–810.
[52] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, W.-c. Woo, Deep learning for precipitation nowcasting: a benchmark and a new model, in: Advances in Neural Information Processing Systems, NIPS, 2017, pp. 5617–5627.
[53] Y. Wang, M. Long, J. Wang, Z. Gao, S.Y. Philip, Predrnn: recurrent neural networks for predictive learning using spatiotemporal lstms, in: Advances in Neural Information Processing Systems, NIPS, 2017, pp. 879–888.
[54] Y. Wang, Z. Gao, M. Long, J. Wang, P.S. Yu, Predrnn++: towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning, in: ICML, in: Proceedings of Machine Learning Research, PMLR, vol. 80, 2018, pp. 5110–5119.
[55] R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, C. Yang, A hybrid cnn-lstm model for typhoon formation forecasting, Geoinformatica 23 (3) (2019) 375–396.
[56] E. Racah, C. Beckham, T. Maharaj, S.E. Kahou, M. Prabhat, C. Pal, Extremeweather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events, in: Advances in Neural Information Processing Systems, NIPS, 2017, pp. 3402–3413.
[57] C. Wang, Y. Hong, Application of spatiotemporal predictive learning in precipitation nowcasting, in: AGU Fall Meeting Abstracts, 2018.
[58] Prabhat, O. Rübel, S. Byna, K. Wu, F. Li, M.F. Wehner, W. Bethel, TECA: a parallel toolkit for extreme climate analysis, in: ICCS, in: Procedia Computer Science, vol. 9, Elsevier, 2012, pp. 866–876.
[59] Prabhat, S. Byna, V. Vishwanath, E. Dart, M.F. Wehner, W.D. Collins, TECA: petascale pattern recognition for climate science, in: CAIP (2), in: Lecture Notes in Computer Science, vol. 9257, Springer, 2015, pp. 426–436.
[60] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al., Application of deep convolutional neural networks for detecting extreme weather in climate datasets, arXiv preprint, arXiv:1605.01156.
[61] D. Tran, L.D. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3d convolutional networks, in: ICCV, IEEE Computer Society, 2015, pp. 4489–4497.
[62] A. Karpatne, G. Atluri, J.H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, V. Kumar, Theory-guided data science: a new paradigm for scientific discovery from data, IEEE Trans. Knowl. Data Eng. 29 (10) (2017) 2318–2331.
[63] J. Runge, V. Petoukhov, J.F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, J. Kurths, Identifying causal gateways and mediators in complex spatio-temporal systems, Nat. Commun. 6 (1) (2015) 1–10.
[64] N. Srivastava, E. Mansimov, R. Salakhudinov, Unsupervised learning of video representations using lstms, in: International Conference on Machine Learning, ICML, 2015, pp. 843–852.
[65] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, G. Zhang, Deep uncertainty quantification: a machine learning approach for weather forecasting, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2087–2095.
[66] J. Frnda, M. Durica, J. Nedoma, S. Zabka, R. Martinek, M. Kostelansky, A weather forecast model accuracy analysis and ecmwf enhancement proposal by neural network, Sensors 19 (23) (2019) 5144.
[67] A. Grover, A. Kapoor, E. Horvitz, A deep hybrid model for weather forecasting, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 379–386.

[68] S. Scher, Toward data-driven weather and climate forecasting: approximating a simple general circulation model with deep learning, Geophys. Res. Lett. 45 (22) (2018) 12–616.

[69] E. de Bezenac, A. Pajot, P. Gallinari, Deep learning for physical processes: incorporating prior scientific knowledge, J. Stat. Mech. Theory Exp. 2019 (12) (2019) 124009.

[70] S. Schirber, D. Klocke, R. Pincus, J. Quaas, J.L. Anderson, Parameter estimation using data assimilation in an atmospheric general circulation model: from a perfect toward the real world, J. Adv. Model. Earth Syst. 5 (1) (2013) 58–70.

[71] S. Rasp, M.S. Pritchard, P. Gentine, Deep learning to represent subgrid processes in climate models, Proc. Natl. Acad. Sci. 115 (39) (2018) 9684–9689.

[72] X. Ren, J. Zhao, X. Li, K. Ren, J. Song, D. Sun, Pagcm: a scalable parallel spectral-based atmospheric general circulation model, Concurr. Comput. 31 (20) (2019) e5290.

[73] K. Yonekura, H. Hattori, T. Suzuki, Short-term local weather forecast using dense weather station by deep neural network, in: IEEE International Conference on Big Data, Big Data, IEEE, 2018, pp. 1683–1690.

[74] A.H. Monahan, J.C. Fyfe, M.H. Ambaum, D.B. Stephenson, G.R. North, Empirical orthogonal functions: the medium is the message, J. Climate 22 (24) (2009) 6501–6514.

[75] K. Steinhaeuser, N.V. Chawla, A.R. Ganguly, Comparing predictive power in climate data: clustering matters, in: International Symposium on Spatial and Temporal Databases, Springer, 2011, pp. 39–55.

[76] D.S. Nolan, M.G. McGauley, Tropical Cyclogenesis in Wind Shear: Climatological Relationships and Physical Processes, 2012.

[77] F. Hourdin, T. Mauritsen, A. Gettelman, J.-C. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, et al., The art and science of climate model tuning, Bull. Am. Meteorol. Soc. 98 (3) (2017) 589–602.

[78] B. Stevens, S. Bony, What are climate models missing?, Science 340 (6136) (2013) 1053–1054.

[79] S. Bony, B. Stevens, D.M. Frierson, C. Jakob, M. Kageyama, R. Pincus, T.G. Shepherd, S.C. Sherwood, A.P. Siebesma, A.H. Sobel, et al., Clouds, circulation and climate sensitivity, Nat. Geosci. 8 (4) (2015) 261–268.

[80] G.J. Anderson, D.D. Lucas, Machine learning predictions of a multiresolution climate model ensemble, Geophys. Res. Lett. 45 (9) (2018) 4273–4280.

[81] I. Durre, R.S. Vose, D.B. Wuertz, Overview of the integrated global radiosonde archive, J. Climate 19 (1) (2006) 53–68.

[82] I. Roesch, T. Günther, Visualization of neural network predictions for weather forecasting, in: Computer Graphics Forum, Wiley Online Library, 2017.

[83] D.P. Dee, S.M. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d.P. Bauer, et al., The era-interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc. 137 (656) (2011) 553–597.

[84] V. Eyring, P.J. Gleckler, C. Heinze, R.J. Stouffer, K.E. Taylor, V. Balaji, E. Guilyardi, S. Joussaume, S. Kindermann, B.N. Lawrence, et al., Towards improved and more routine Earth system model evaluation in cmip, Earth Syst. Dyn. 7 (4) (2016) 813–830.

[85] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, G. Yacalis, Could machine learning break the convection parameterization deadlock?, Geophys. Res. Lett. 45 (11) (2018) 5742–5751.

[86] W. Xue, C. Yang, H. Fu, X. Wang, Y. Xu, L. Gan, Y. Lu, X. Zhu, Enabling and scaling a global shallow-water atmospheric model on tianhe-2, in: IEEE 28th International Parallel and Distributed Processing Symposium, IPDPS, IEEE, 2014, pp. 745–754.

[87] H. Fu, J. Liao, N. Ding, X. Duan, L. Gan, Y. Liang, X. Wang, J. Yang, Y. Zheng, W. Liu, et al., Redesigning cam-se for peta-scale climate modeling performance and ultra-high resolution on sunway taihulight, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC, 2017, pp. 1–12.

[88] N.P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, et al., In-datacenter performance analysis of a tensor processing unit, in: Proceedings of the 44th Annual International Symposium on Computer Architecture, ISCA, 2017, pp. 1–12.

[89] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, et al., Towards artificial general intelligence with hybrid tianjic chip architecture, Nature 572 (7767) (2019) 106–111.

[90] A. Alléon, G. Jauvion, B. Quennehen, D. Lissmyr, Plumenet: large-scale air quality forecasting using a convolutional lstm network, arXiv preprint, arXiv:2006.09204.

[91] J. Fang, H. Fu, W. Zhao, B. Chen, W. Zheng, G. Yang, swdnn: A library for accelerating deep learning applications on sunway taihulight, in: IEEE International Parallel and Distributed Processing Symposium, IPDPS, IEEE, 2017, pp. 615–624.

[92] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.

[93] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. 73 (2018) 1–15.

[94] P.M. Cox, D. Pearson, B.B. Booth, P. Friedlingstein, C. Huntingford, C.D. Jones, C.M. Luke, Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability, Nature 494 (7437) (2013) 341–344.

[95] P.M. Caldwell, C.S. Bretherton, M.D. Zelinka, S.A. Klein, B.D. Santer, B.M. Sanderson, Statistical significance of climate sensitivity predictors obtained by data mining, Geophys. Res. Lett. 41 (5) (2014) 1803–1808.