Journal of Geography and Cartography (2023) Volume 6 Issue 1
doi:10.24294/jgc.v6i1.2065

ORIGINAL RESEARCH ARTICLE

Extreme low apparent temperature forecasting for early warning of mortality by data-driven deep learning

Lei Xu[1*], Hongchu Yu[2], Xihao Zhang[1], Yuan Gan[3]

[1] *National Engineering Research Center for Geographic Information System, China University of Geosciences (Wuhan), Wuhan 430000, China. E-mail: xulei10@cug.edu.cn*

[2] *School of Navigation, Wuhan University of Technology, Wuhan 430000, China*

[3] *State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430000, China*

ABSTRACT

Abrupt changes in environmental temperature, wind and humidity can lead to great threats to human life safety. The Gansu marathon disaster of China highlights the importance of early warning of hypothermia from extremely low apparent temperature (AT). Here a deep convolutional neural network model together with a statistical downscaling framework is developed to forecast environmental factors for 1 to 12 h in advance to evaluate the effectiveness of deep learning for AT prediction at 1 km resolution. The experiments use data for temperature, wind speed and relative humidity in ERA-5 and the results show that the developed deep learning model can predict the upcoming extreme low temperature AT event in the Gansu marathon region several hours in advance with better accuracy than climatological and persistence forecasting methods. The hypothermia time estimated by the deep learning method with a heat loss model agrees well with the observed estimation at 3-hour lead. Therefore, the developed deep learning forecasting method is effective for short-term AT prediction and hypothermia warnings at local areas.

Keywords: Apparent Temperature Forecasting; Deep Learning; Neural Network; Spatiotemporal Forecasting; AI Applications

ARTICLE INFO

Received: 27 April 2023
Accepted: 29 May 2023
Available online: 7 June 2023

1. Introduction

Extremely low apparent temperature forecasting is of great importance to protect people's health and safety. Previous studies showed that more than five million deaths are related to non-optimum temperature per year, accounting for ~9.4% of global mortality, of which 8.52% are cold-related and 0.91% are heat-related[1,2]. The cold-related extreme events are a great threat to people's lives and could bring potential mortality risk when the ambient temperature goes down quickly[3]. On 22 May 2021, 21 professional runners died from hypothermia out of 172 participants in a 100 km marathon race in Jingtai County, Gansu Province, China[4], which is one of the most severe disasters in marathon history. The Gansu marathon disaster is attributed to the abrupt decline of air temperature, strong wind, high humidity and the unawareness of the extreme event and the delayed weather forecasting service accounts partly for the disaster[5]. Therefore, timely and accurate forecasting of the instantaneous cold event can help reduce socioeconomic losses[6]. The human-perceived temperature, also called apparent temperature (AT), measures the

combined effect of air temperature, wind and relative humidity on human body feeling[7]. Hypothermia occurs when the heat supply cannot satisfy the heat loss from low AT conditions for a continuous period, leading to a drop of body core temperature below 35 ℃ and potential symptoms such as shivering, mental confusion, cardiorespiratory failure or even death[8]. The Gansu marathon disaster occurred during the middle of the race (between the second and third checkpoint, out of the nine checkpoints) where the altitude is continuously increasing in this area[5]. The meteorological center of the Gansu government issued a notice in the previous night that there is a drop in temperature with wind gusts and potential precipitation on 21–22 May, but could not predict the timing and spatial locations of the cold event[5]. Therefore, the disaster-related early warning information in space and time is lacking for the marathon organizer and runners to prepare for the potential risk.

Data-driven forecasting using learning methods provides a way for prediction in near real time[9–11]. The data-driven methods predict the target variable for the future time period based on the learned relationships from historical data, which is appealing for its simplicity and usability[12,13]. The deep learning related forecasting methods have shown some potential for hydrometeorological prediction, such as sea surface temperature and precipitation[10,14,15]. The flexible model structures and the nonlinear modeling of deep learning methods provide powerful predictive ability by elaborate training processes[14,16,17]. Compared with numerical models, the data-driven deep learning methods focus on the mathematical relationships between predictors and predictands, while numerical models pay more attention to the physical mechanisms of system dynamics[11,13]. The data-driven deep learning methods can complement in physical process understanding through exploratory data modeling[18,19].

As mentioned above, the Gansu marathon disaster is the compound effects of numerous relevant factors while the abrupt decline of temperature, strong wind gusts and high humidity are the direct causes[4,5]. Therefore, successful forecasting of these weather-related factors is the key of early warning of hypothermia and mortality caused by extreme AT. The numerical prediction is known to have a relatively coarse spatial resolution and has uncertainties in parameterization and initialization[20–22]. The deep learning methods have shown some potential for short-term precipitation nowcasting, seasonal forecasting and multiyear oscillation prediction relative to numerical models[10,13,15,16]. For short-term weather forecasting or nowcasting, the deep learning models can learn the propagation and movement of weather patterns from spatial maps of historical observations[15,16], which is theoretically reasonable for extrapolating the weather process. The physics of weather processes are reflected in historical observations and are modeled by data-driven artificial intelligence models. The intelligence models learn the potential physics from data and predict future processes using the learned rules[23–25]. The forecasting processes are similar between intelligence models and numerical models from an application perspective[11]. Although the intelligence models may not beat numerical models[26], the former can be a strong supporter of the latter for earth system understanding and applications[11,13].

This study aims to develop a deep learning model and a statistical downscaling framework to predict AT conditions at 1 km resolution for 1 to 12 h in advance. We show that the developed deep learning forecasting framework together with the heat loss model of human body is able to provide timely and useful disaster warning information for extremely low AT conditions for a few hours in advance. The developed framework can be used as a short-term forecasting tool to assist extreme low AT disaster prevention and warning at a relatively high spatiotemporal resolution for local areas. The rest of this study is organized as follows. Section 2 describes the methods and materials used. Section 3 presents the results and discussions. Section 4 concludes this study.

2. Study area and methodology

2.1 Study area

The study area is the Gansu marathon area in Jingtai County, Baiyin City, Gansu Province, China (**Figure 1**). The Jingtai County features a temperate

arid continental climate, with large temperature differences between the four seasons, low precipitation and concentration in the summer, low relative humidity and high winds due to the relatively flat terrain. The annual average temperature is 8.2 °C, rainfall is 185 mm and the relative humidity is 56%. The marathon route starts from the bottom right corner of the study area in an anticlockwise direction. The digital elevation model (DEM) data is obtained from the Global 30 Arc-Second Elevation (GTOPO30)[27] at 1 km resolution, and the resolution is consistent with the resolution of the experimental data after statistical downscaling for better data visualization and presentation. GTOPO30, whose full name is Global 30 Arc-Second Elevation Data Set, is a global digital elevation model dataset produced and published by the United States Geological Survey (USGS). It covers the elevation data of the global surface and can be used for Geographic Information System (GIS) analysis and applications on a global scale. The Gansu marathon disaster occurred between the checkpoint 2 (cp2) to cp3, where the elevation increases continuously by approximately 1 km during the race. The marathon route from cp2 to cp3 is filled with hoodoos, which is one of the severest parts during the 100 km ultramarathon. The extreme weather exhibits different spatial patterns over Gansu Province in recent years, with increasing and decreasing precipitation extremes in the north and south of Gansu, respectively[28,29]. Although the temperature extremes may not increase in Gansu Province over the years, the compound weather conditions are fickle in this area. The abrupt change of one weather factor (e.g., temperature) may have limited effects on human body, while the abrupt changes of compound weather factors (e.g., temperature and wind speed) may cause great disasters such as the Gansu marathon catastrophe.

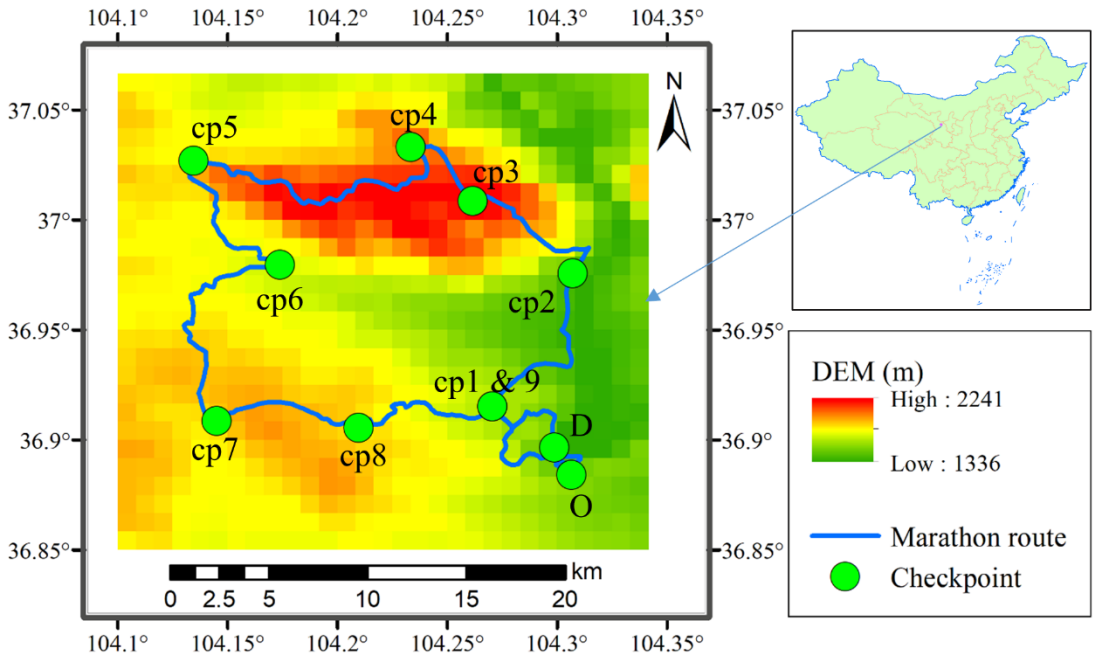

Figure 1. A demonstration of the marathon region in Baiyin City, Gansu Province, China. The checkpoints from 1 to 9 are shown in the marathon route. "O" is the origin and "D" is the destination of the marathon route.

2.2 The deep learning forecasting framework

Figure 2 demonstrates the developed deep learning framework for hourly AT prediction at 1 km resolution. Here the air temperature, wind gust and relative humidity variables are forecasted based on the historical observations in the previous three hours. The wind gust is the instantaneous 10 m wind gust over 3-second intervals from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA-5) data[30], which is selected to highlight the potential mortality risk at an upper limit. The temporal weather memory is considered by including the data in the recent three hours and the spatial dependences are modeled by the spatial convolution and pooling layers in the deep learn-

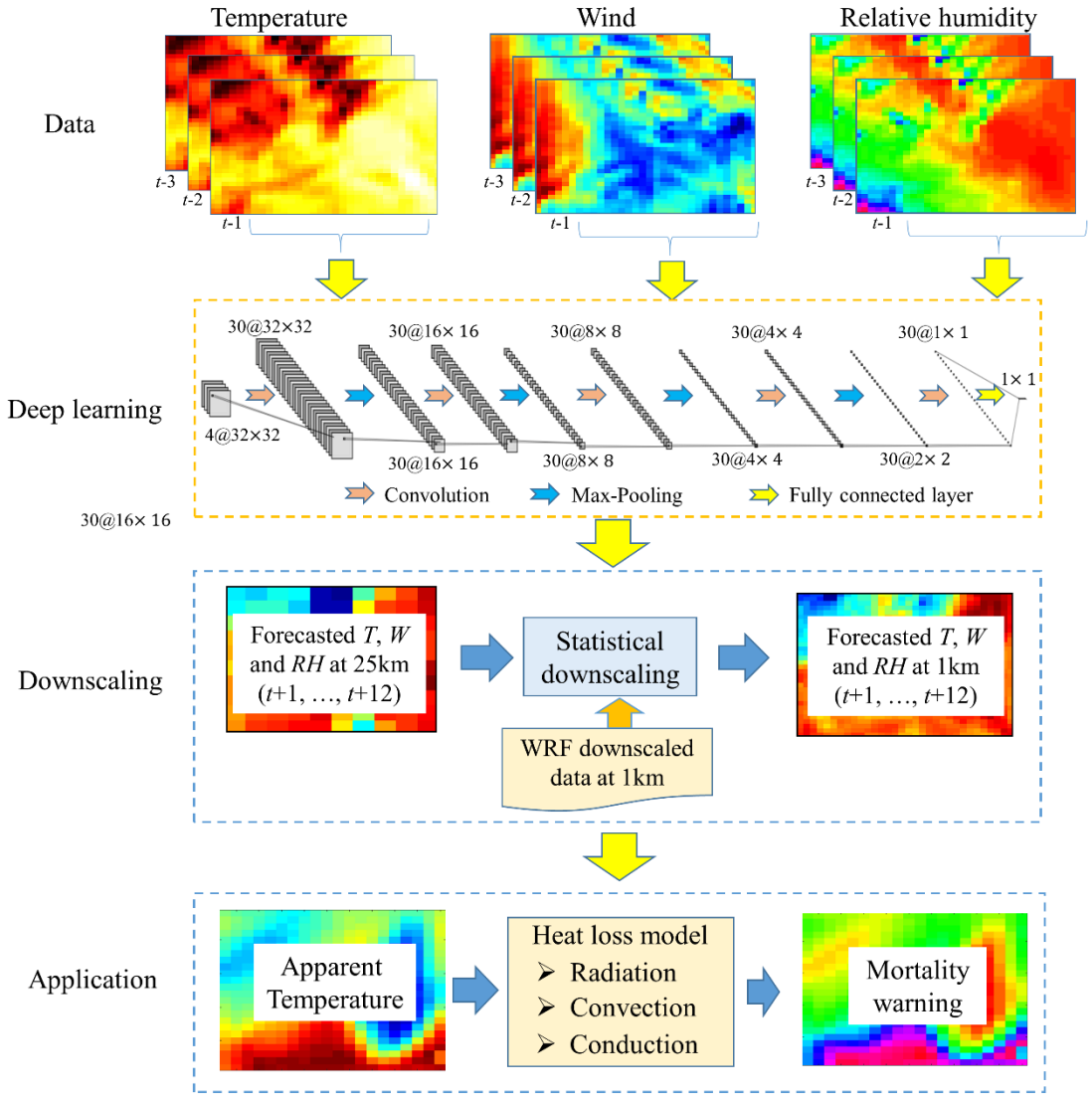

Figure 2. The developed deep learning framework for apparent temperature forecasting.

ing model[15,16]. The 2-dimensional convolution modulates different predictors in space by multiple kernels, resulting in a series of feature maps used for subsequent processing. The max pooling layer extracts local typical features and reduces spatial dimensionality. The fully connected layer maps the input features to output data through artificial neural networks.

In the convolution process, the value of the j^{th} feature in the i^{th} convolutional layer for a specific pixel (x,y) (denoted as $z_{i,j}^{x,y}$) is calculated as follows:

$$z_{i,j}^{x,y} = relu\left(\sum_{k=1}^{K_{i-1}}\sum_{l=1}^{L_i}\sum_{m=1}^{M_i} w_{i,j,k}^{l,m} z_{(i-1),k}^{\left(x+l-\frac{L_i}{2},y+m-\frac{Q_i}{2}\right)} + b_{i,j}\right) \quad (1)$$

where L_i and M_i are the horizontal and vertical dimensions of the convolution filter for the i^{th} layer, respectively. A rectified linear unit (*relu*) function is used as the activation function. The convolution filter size is set at 2×2 for all the convolutional layers in **Figure 2**. K_{i-1} represents the number of features in the $(i-1)^{th}$ layer. $w_{i,j,k}^{l,m}$ is the weight at point (x,y) for the convolution filter to connect the k^{th} feature in the $(i-1)^{th}$ layer to the j^{th} feature in the i^{th} layer.

$z_{(i-1),k}^{(x+l-\frac{L_i}{2},y+m-\frac{Q_i}{2})}$ is the value of the k^{th} feature in the $(i-1)^{\text{th}}$ layer for the point $(x+l-\frac{L_i}{2}, y+m-\frac{Q_i}{2})$.

In the training process, the number of filters is set to 20, 30, 40 and 50 to consider the uncertainty of parameters to produce an ensemble of prediction. The ensemble mean is regarded as the forecasting result and the standard deviation is the predictive spread. The parameters are trained with an iteration of 1,000 epochs and an initial learning rate of 0.001. The Adam algorithm[31] is used to optimize the parameters with the mean square error loss function. The deep learning model is trained for 1–12 h lead times separately. The ERA-5 data during 2015–2018, 2019–2020 and 2021 are adopted as the training data, validation data and testing data, respectively.

The deep learning model is compared with two baseline forecasting methods, namely the climatology forecast and the persistence forecast[32]. The climatology forecast is to perform prediction using multiyear averaged values over the same hourly period. For example, the climatology forecast for the 8:00 temperature prediction on 21 May 2021 is estimated by the averaged 8:00 temperature on 21 May from 2015 to 2020. The persistence forecast uses the weather memory in the previous day to forecast the weather today. The persistence forecast for the 8:00 temperature prediction on 21 May 2021 is obtained using the observed temperature at 8:00 on 20 May 2021. The two baselines are often used as references to measure the forecasting skill of new models[33]. The Pearson's correlation coefficient and the root mean square error (RMSE) are used to assess the performance of prediction.

2.3 The statistical downscaling scheme

The ERA-5 data is used to produce a preliminary forecasting result at a resolution of 25 km for the next 1 to 12 h. A statistical downscaling method is utilized to downscale the coarse-resolution forecasts to a fine resolution of 1 km. The downscaling method establishes a statistical relationship between the two spatial scales through a linear regression modeling[34]. The fine-resolution data is generated by a three-layer nested Weather Research and Forecasting (WRF) model[35] with the Kain–Fritsch convective parameterization[36] and the Lin microphysics scheme[37] using the 0.25° National Centers for Environmental Prediction (NCEP) Final Analysis (FNL) data[38] and is then used for regression. The statistical downscaling method assumes the stationarity between large and small-scale dynamics and ignores the complex interactions of predictors. The temperature variable is relatively continuous in space and is dependent on elevation, season, latitude and large-scale weather conditions. The regression-based downscaling process could reflect the temperature variations with elevation, which is a key step for estimating temperature distinction in mountainous areas[39]. Based on statistical regression, the downscaling of wind gusts and relative humidity can maintain the coarse-resolution variations and large-scale dynamics.

The apparent temperature is calculated based on the following formula[7].

$$AT = 1.04 \times T_a + 0.2 \times e - 0.65 \times v - 2.7 \tag{2}$$

$$e = \frac{rh}{100} \times 6.105 \times exp\left(\frac{17.27 \times T_a}{237.7 + T_a}\right) \tag{3}$$

where T_a is the air temperature (°C); e is the water vapour pressure (hPa); v is the wind gust (m/s) and rh is the relative humidity (%).

2.4 The heat loss model

The heat loss model is constructed by the consideration of radiation, convection and conduction processes[7,40–42]. In the radiation, core body tissues transfer heat in blood vessels by emitting infrared rays from the skin surface to lose heat. Convection is a heat transfer process from the skin to the surroundings by the movement of the fluid. Conduction is a process of the heat transfer from the hot area of an object to the cool area. In low temperature conditions, the heat loss due to evaporation is small and the evaporation process is neglected here. The heat loss formula is expressed as follows.

$$L_{total} = L_{radiation} + L_{convection} + L_{conduction} \tag{4}$$

$$L_{radiation} = \epsilon \cdot A_u \sigma\left(T_{skin}^4 - T_a^4\right) \tag{5}$$

$$L_{convection} = \left(12.12 - 1.16v + 11.6\sqrt{v}\right) \cdot A_u \cdot (T_{skin} - T_a) \tag{6}$$

$$L_{conduction} = \frac{K_{cond} \cdot A_c}{L} \cdot (T_{skin} - T_a) = \frac{A_c}{I} \cdot (T_{skin} - T_a) \quad (7)$$

where L_{total} is the total heat loss; $L_{radiation}$, $L_{convection}$ and $L_{conduction}$ are the heat loss due to radiation, convection and conduction, respectively; ϵ is the emissivity of the skin and is set to 1 here[43]; A_u denotes the uncovered body area and is assumed as 50% of the body area in marathon race; σ is the Stefan-Boltzmann constant, 5.67×10^{-8} W/m^2/K^4; v is the wind speed with a unit of m/s; The formula $(12.12 - 1.16v + 11.6\sqrt{v})$ is an empirical estimate of the wind cooling effect on human body and is valid between 2 m/s and 20 m/s[44,45]; K_{cond} is the thermal conductivity of the clothing material; L is the thickness of the clothing material; $I = L/K_{cond}$ is the insulation value of the clothes in the unit of *clo*; A long-sleeve dress shirt has an insulation value of 0.25 *clo* and a suit jacket has an insulation of 1.0 *clo*[46]; here the insulation value in the marathon race is set to 0.5 *clo* for a reference.

The heat supply is mainly provided by body metabolism during the marathon race. The basal metabolic rate (BMR) is the basic rate of energy expenditure when maintaining the body's life-sustaining function. When the body is at active during activity, additional heat is generated compared to the resting condition and the total metabolic rate (TMR) will be 2–20 times of the BMR[47]. The BMR is calculated based on the Mifflin-St Jeor equation[48] with a height of 1.7 m, a weight of 70 kg and an age of 30 years old. The TMR is set as 10 times of the BMR during the Gansu marathon race, which is relatively a high rate as the limit of the top runners may be up to 16 times of the BMR[47]. The hypothermia time from heat loss is estimated by calculating the time of heat loss from the normal body temperature (37 ℃) to the moderate hypothermia temperature (32 ℃). The coordinated universal time (UTC) is used throughout the study, while the Beijing time is eight hours ahead of UTC (UTC + 8:00).

3. Results and discussions

3.1 Overall performance of temperature, wind and relative humidity forecasting

The predictive skills of temperature, wind gust and relative humidity by deep learning, climatology and persistence forecast methods are shown in **Figure 3**. The correlation of hourly temperature prediction is close to 0.9 even at 12-hour lead for the deep learning method, much higher than the climatology and persistence approaches which have a correlation below 0.7 (**Figure 3a**). The average RMSE of hourly temperature prediction in May 2021 is about 3 ℃ for the deep learning model at 12-hour lead, lower than the reference methods (**Figure 3b**). The performance of wind gust prediction for deep learning method is better than climatology forecast for 1–7 h leads and may not contrast with climatology at longer lead (**Figure 3c**–**d**). As for the relative humidity, the predictive performance for deep learning outperforms persistence forecast at 1–8 h leads in terms of the correlation and at 1–12 h leads in terms of RMSE (**Figure 3e**–**f**).

Figure 4 exhibits the predictions of temperature, wind gusts and relative humidity by deep learning for 1-hour and 3-hour leads. The predicted and observed time series of air temperature are consistent with each other, with a correlation of 0.98 and 0.97 and a RMSE of 1.19 ℃ and 1.56 ℃ for 1- and 3-hour leads, respectively. The wind gust prediction is similar to observations, both for mild winds and strong gusts, with an average of RMSE of 1.78 m/s and 2.60 m/s, respectively at 1- and 3-hour leads. As for the relative humidity, a correlation coefficient of 0.96 (0.91) and a RMSE of 7.08% (9.38%) are expected for 1- and 3-hour prediction, respectively. These predictions suggest overall consistency with observations over the Gansu marathon region, indicating the efficiency for early warning of extreme weather events by the deep learning model.

3.2. The forecasting of apparent temperature by different methods

Different results were obtained using different methods for predicting AT in the Gansu marathon region (**Figure 5**). The deep learning prediction exhibits good consistency with observation for 1-hour and 3-hour AT forecasts, better than the climatology and persistence methods (**Figure 5a**–**b**). The clima-

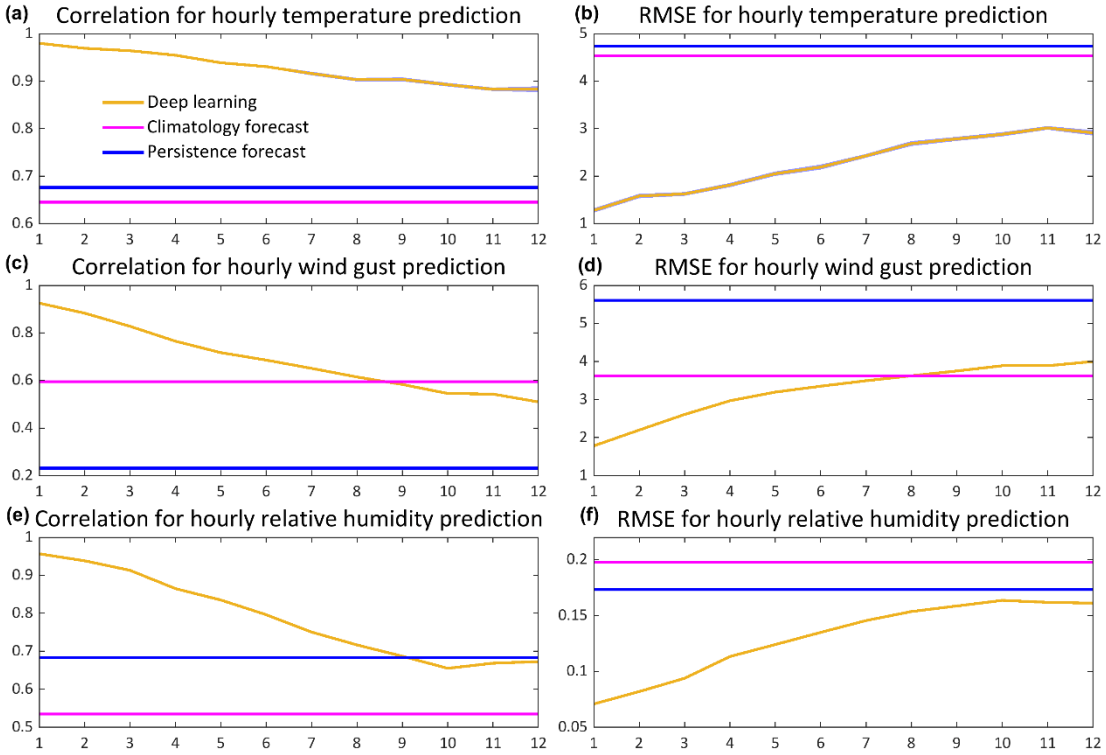

Figure 3. The performance of temperature, wind and relative humidity prediction by different forecasting methods.

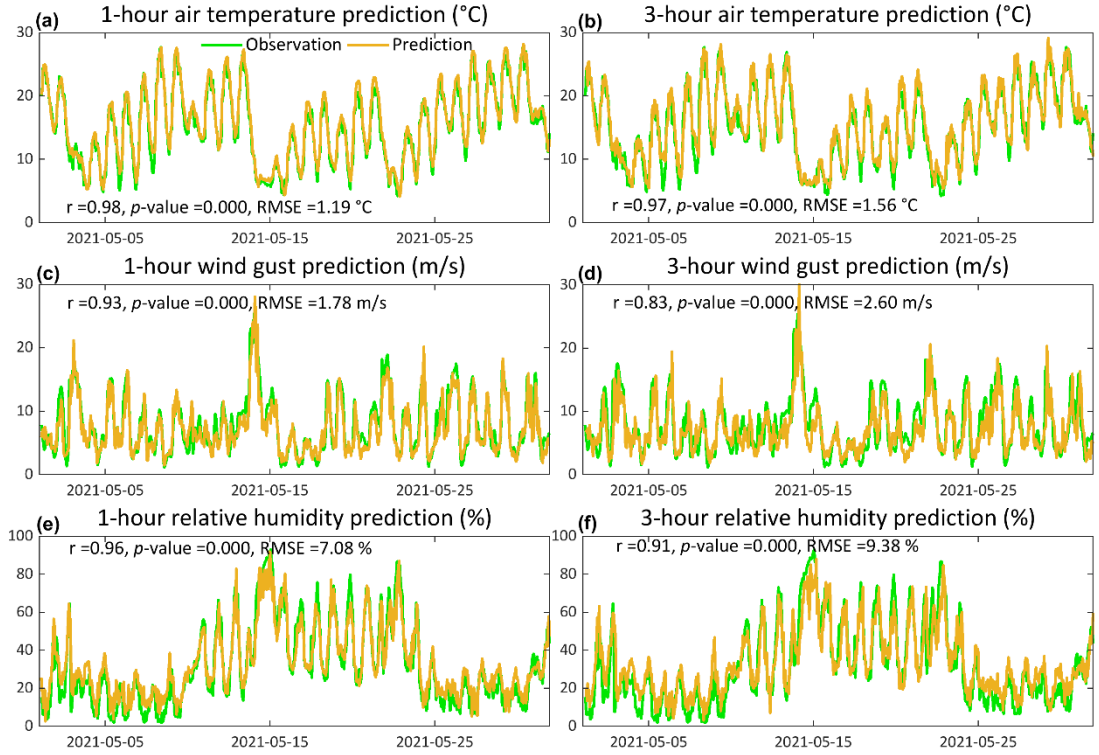

Figure 4. The observed and predicted temperature, wind gust and relative humidity prediction by deep learning for 1-hour and 3-hour leads over the Gansu marathon region.

tology forecast attempts to average the historical observations and cannot predict the anomalous high and low values[32]. The persistence forecast utilizes the weather memory and is not suitable for regions where the weather changes rapidly, especially for extreme events. For example, the persistence method predicts an extremely low AT on 15 May 2021, which overestimates the AT severity considerably.

The deep learning model learns the relationship between large-scale weather dynamics and the weather in the target region, such as the spatial propagation and movement[15,16,49].

The correlation coefficient of AT prediction decreases with the increase of lead time (**Figure 5c**), with a value of 0.97 at 1-hour lead and 0.76 at 12-hour lead, much higher than the climatology (0.56) and persistence methods (0.49), consistent with the decreasing trend with lead in previous studies[10,16,50]. The RMSE of AT prediction by deep learning is 2.0 ℃ and 4.4 ℃ at 1-hour and 12-hour leads, respectively, much lower than the climatology (5.2 ℃) and persistence (6.3 ℃) forecasts (**Figure 5d**). As for the spatial patterns, the deep learning prediction exhibits the similar spatial distribution of AT with observation at 3:00, 21 May 2021 (**Figure 5e–f**), while the climatology and persistence methods cannot capture the abrupt decline of AT this time (**Figure 5g–h**).

The predictive uncertainty of deep learning is represented by the spread of the forecasting from different kernel parameters. The predictive uncertainty by deep learning from parameters is indicated in **Figure 5a–d** and is very small, suggesting that the number of kernel filters in the deep learning model has a very limited effect on the forecasting performance. It should be noted that other uncertainties exist in the prediction, including data, model structure and parameters[11,51]. The data uncertainty from ERA-5 is not estimated here as the in-situ measurements are not available. Different structures of deep learning models may lead to differences of forecasting accuracy. The deep convolutional neural network (CNN) model developed here considers the spatiotemporal interactions of multiple variables. The spatial convolution weights the neighboring values to produce new features and the fully connected layer maps the input features into the outputs by neural networks[12,16]. The stacked convolution layers simulate nonlinear relationships of multivariate coupling. Previous studies have demonstrated that the stacked CNN models could model the spatial propagation of meteorological elements[10,14–16,49]. The important parameters in CNN models include the number of kernels, kernel size and training rules. The variant kernels can reflect the parametric uncertainty and the heuristic training ensures the optimal determination of model parameters[52].

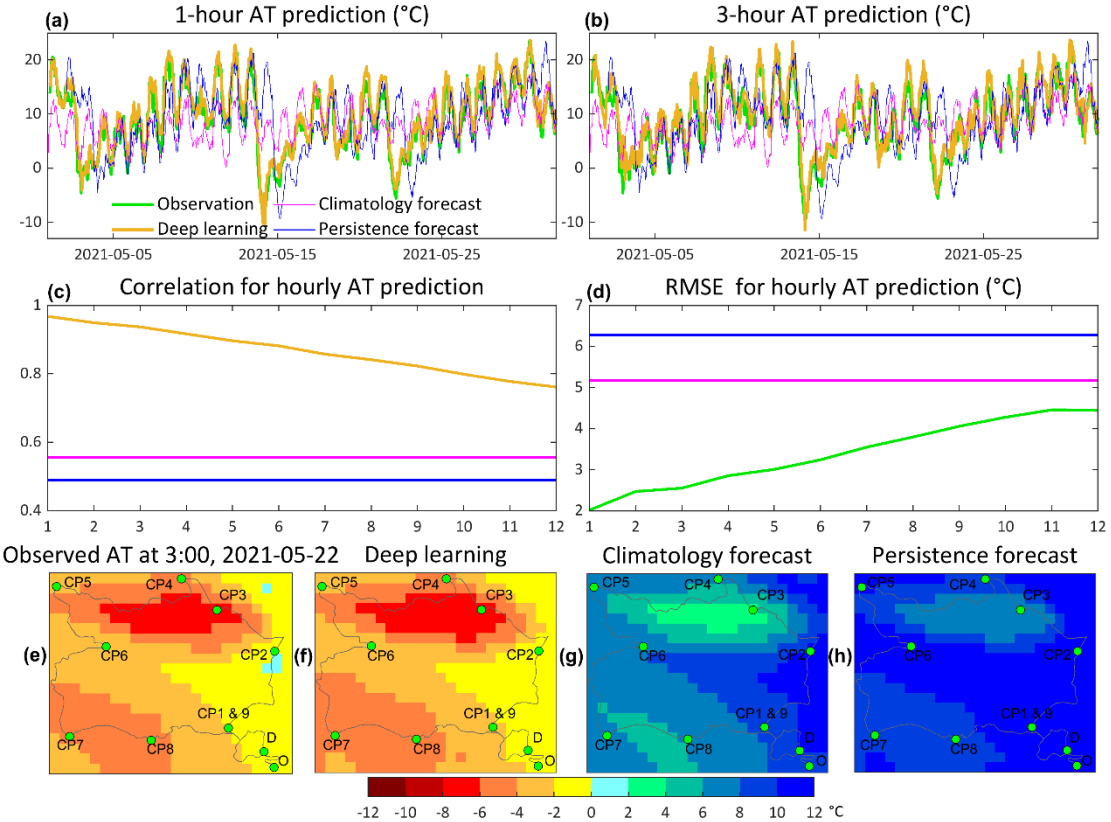

Figure 5. The predictive performance of AT by different approaches in the Gansu marathon region.

3.3 Hypothermia time estimation and early warning of mortality

Figure 6 demonstrates the estimated and predicted hypothermia time based on the method in section 2 over the Gansu marathon region. According to observations, the estimated hypothermia time exhibits an abrupt decline on 22 May 22 2021 relative to 21 May. At 6:00 on 21 May, the hypothermia time is estimated to be about 6 h (**Figure 6a**), while the hypothermia time is approximately 30 min at 6:00 (14:00 for Beijing time) on 22 May (**Figure 6b**). The 3-hour prediction by deep learning can generally capture the temporal changes of hypothermia estimation, while the climatology and persistence methods cannot (**Figure 6a**–**b**). The climatology and persistence forecasts estimate a higher hypothermia time relative to observation and deep learning methods. Although the climatology and persistence approaches predict a similar hypothermia time in the morning of 22 May (2:00 for UTC), while the hypothermia time is severely overestimated in the noon (5:00 for UTC).

Figure 6c exhibits the monthly averaged hypothermia time from observed data. The averaged hypothermia time from checkpoint 2 (cp2) to cp5 is overall smaller than other parts of the marathon race due to high elevation. The deep learning prediction demonstrates a similar spatial distribution of hypothermia time with observation (**Figure 6d**), while the climatology method overestimates the hypothermia severity (**Figure 6e**). The persistence prediction agrees well with observation on a monthly average (**Figure 6f**) because the majority of hourly predicted values are historical observations. For the day of the Gansu marathon disaster, the runners may lose the temperature quickly from cp2 to cp3 and may have a hypothermia risk within half an hour (**Figure 6g**). The deep learning model is capable of detecting the hypothermia risk versus observations for 3 hours in advance (**Figure 6h**), while the climatology and persistence approaches considerably underestimate the hypothermia severity (**Figure 6i**–**j**).

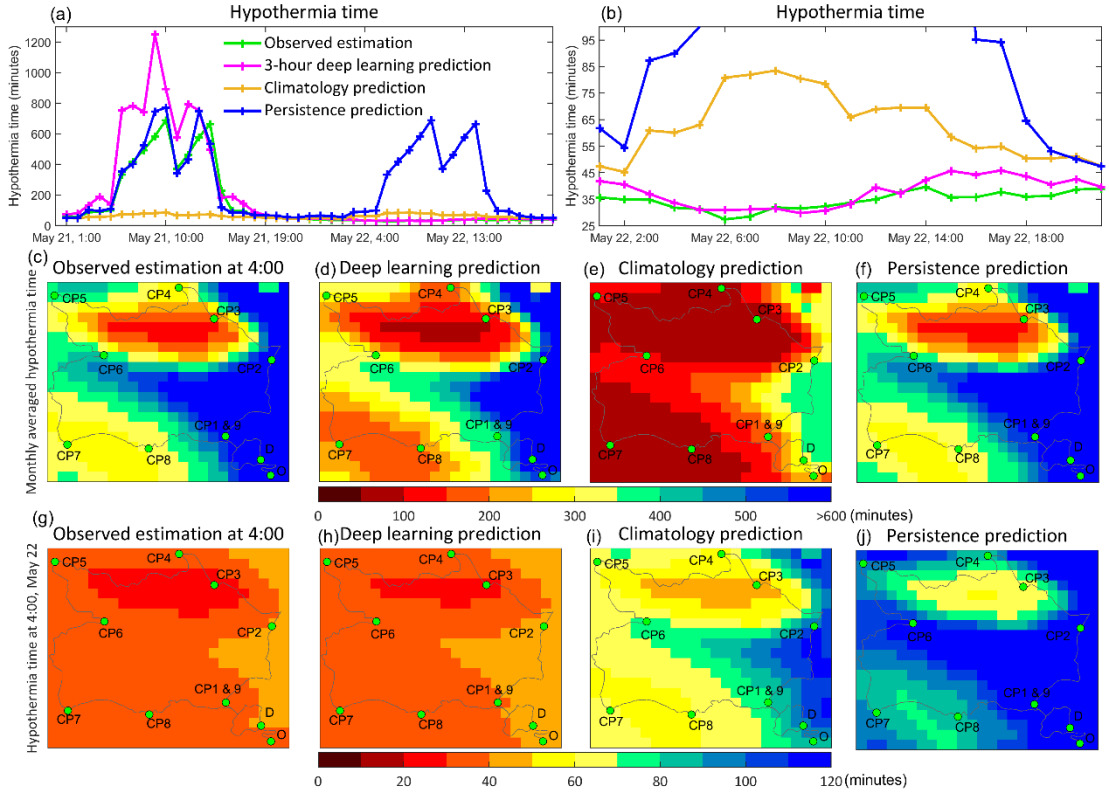

Figure 6. The estimated and predicted hypothermia time by different forecasting methods.

Figure 7 demonstrates the correlation between the predicted hypothermia time and the observed estimation of the Gansu marathon region for the testing period. The developed deep learning model achieves a correlation coefficient of 0.95 with observations at 1-hour lead and the prediction agrees well with observation, especially for hypothermia time less than 150 min, which is very useful for short-term early warning of low-AT related mortality. The correlation coefficient for deep learning generally decreases with the increase in lead time, with a correlation of 0.79 at 12-hour lead. The climatology and persistence predictions are not comparable to deep learning results even at a 12-hour lead time.

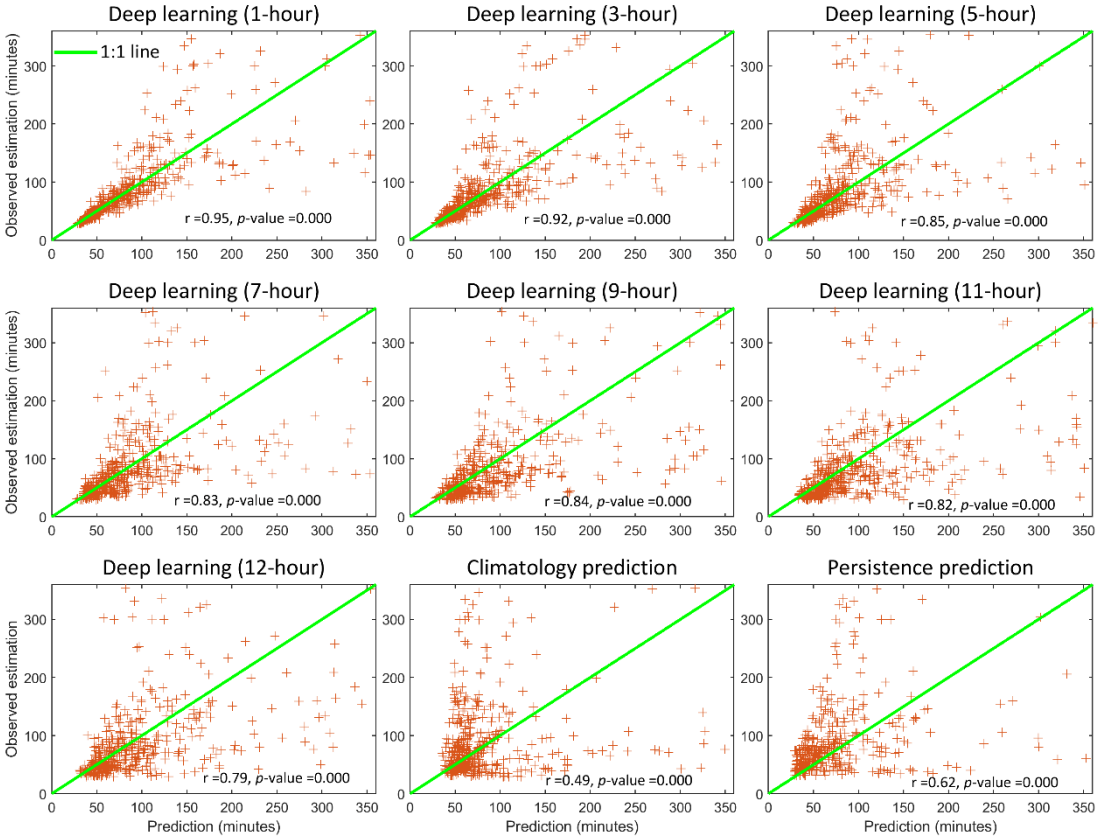

Figure 7. Comparison of hypothermia time forecasting by different methods.

Compared with existing forecasting studies[53–55], the developed deep learning forecasting method is a data-driven end-to-end approach to model the spatiotemporal relationships between large-scale weather dynamics and the dynamics at local areas. The spatiotemporal relationships, including the movement of cold air, moisture transport and wind flow, are inherently simulated by neural network layers[9,10,13,56,57]. However, the simulations of these relationships by deep learning cannot be explicitly expressed as differential equations like that of physical models, which is the major shortage of data-driven artificial intelligence methods. The interpretability of deep learning models can be improved by efforts in the model structure transparency, physical explanations and understandable tools[18,19,58]. As the deep learning model learns the end-to-end relationship between predictors and predictands, the uncertainties of parameterization schemes in physical models are not reflected in deep learning models[11]. The deep learning based artificial intelligence models may not replace the physical models but can complement the physical process understanding and predictive modeling[11,13,16].

3.4 The uncertainty of the forecasting skill

The uncertainty of modeling results is related to the uncertainty of data processing and deep learning model parameters. Specifically, the data processing uncertainty comes from the uncertainty arising from the process of downscaling the data. Since the downscaling method of statistical regression may

have problems, such as overfitting and underfitting, the data may produce problems such as feature errors, which may lead to inaccurate prediction results of downscaled data and thus generate uncertainty. The uncertainty of deep learning model parameters comes from the uncertainty of CNN parameters, where factors such as model parameter initialization and optimization algorithms may lead to different parameters of the model, thus generating uncertainty[11]. Therefore, data processing and the setting of model parameters are crucial to the performance and stability of modeling results. For example, here we consider the uncertainty of model parameters by combining the prediction results of different parameters, thus generating an ensemble of predictions, and treating the ensemble mean as the prediction result.

4. Conclusion

In this study, a deep convolutional neural network model is developed to forecast the extreme AT over the Gansu marathon region. A heat loss model is used to calculate the hypothermia time from compound effects of temperature, wind and relative humidity to estimate the hypothermia risk. The deep learning model can capture the spatiotemporal patterns of AT evolution and outperforms climatology and persistence forecasting methods up to 12-hour lead time. The abrupt decline of AT over the Gansu marathon region can be predicted by deep learning for at least 3 h in advance, which is effective for the timely warning of the potential disaster.

The AT prediction at 1 km resolution is achieved through a statistical downscaling process. The high-resolution temperature forecasts through downscaling are elevation dependent and can reflect the spatial variations of temperature due to altitude changes, which is especially useful for mountainous areas. The direct forecasting at 1 km resolution without the downscaling process is very computationally intensive and the numerical prediction at 1 km resolution by nested weather model requires great computation capacity[59]. Therefore, the simple statistical downscaling is a cost-effective tool to retrieve high-resolution temperature prediction. The static statistical downscaling suffers from uncertainty in high-resolution wind gust and relative humidity simulations and can be improved by complex spatial downscaling approaches.

The heat loss model includes the modeling of radiation, convection and conduction processes to calculate the hypothermia time based on assumptions of metabolic rate of a normal person, clothing insulation and uncovered body area. In practical cases, these assumptions vary person by person and the real hypothermia time is dependent on numerous factors. Apart from the AT prediction, the estimated hypothermia time can be regarded as a reference for a preliminary inference of how long the hypothermia symptoms may happen. The heat loss model can be easily adapted to a specific person according to clothing and biological conditions to provide a timely warning of hypothermia risk.

The developed deep learning model is cost-effective for the early warning of low-AT related hypothermia and mortality. As soon as the deep learning model is trained using historical observations, the forecasting process is very fast and can be done in several seconds, which shows an advantage over numerical weather prediction. The deep learning model can be continuously fine-tuned using observations in near real-time for operational forecasting systems in order to persistently improve the predictive performance. The training process for the developed deep learning model over large spatial scales may be computationally-intensive as that of numerical models. However, the instant prediction of deep learning model is promising for near-real-time forecasting applications in earth system science.

Code availability

The deep learning codes are written in Python language and are partly available on Github: https://github.com/xuleihuanying/apparent_tem_forecasting. The code and datasets can be obtained by contacting Lei Xu at xulei10@cug.edu.cn.

Author contributions

Conceptualization, investigation, methodology, software, writing—original draft, writing—review & editing, LX; writing—review & editing, HY; writing—review & editing, ZH; writing—review & editing, YG.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (42201509, 42101429), the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (162301212687) and the fellowship of China Postdoctoral Science Foundation (2022M722930). The used ERA-5 data in this study is available on the public website (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5).

Conflict of interest

The authors declare no competing financial interests.

References

1. Gasparrini A, Guo Y, Hashizume M, *et al*. Mortality risk attributable to high and low ambient temperature: A multicountry observational study. The Lancet 2015; 386(9991): 369–375. doi: 10.1016/S0140-6736(14)62114-0.
2. Zhao Q, Guo Y, Ye T, *et al.* Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: A three-stage modelling study. The Lancet Planetary Health 2021; 5(7): e415–e425. doi: 10.1016/S2542-5196(21)00081-4.
3. Cannistraro G, Cannistraro M. Hypothermia risk, monitoring and environment control in operating rooms. International Journal of Heat & Technology 2016; 34(2): 165–171. doi: 10.18280/ijht.340202.
4. Bieler D. 21 die in Chinese ultramarathon suddenly struck by plunging temperatures, wind and hail [Report]. BBC News; 2021.
5. Taskforce G.P.J.I. Investigation report on public safety responsibility incidents of Baiyin Jingtai's "5.22" Yellow River Stone Forest 100 km cross-country race [Report]. People's Government of Gansu Province; 2021.
6. Chiu CH, Vagi SJ, Wolkin AF, *et al.* Evaluation of the National Weather Service extreme cold warning experiment in North Dakota. Weather, Climate, and Society 2014; 6: 22–31. doi: 10.1175/WCAS-D-13-00023.1.
7. Steadman RG. A universal scale of apparent temperature. Journal of Applied Meteorology and Climatology 1984; 23(12): 1674–1687. doi: 10.1175/1520-0450(1984)023<1674:AUSOAT>2.0.CO;2.
8. Brown DJ, Brugger H, Boyd J, Paal P. Accidental hypothermia. New England Journal of Medicine 2012; 367: 1930–1938. doi: 10.1056/NEJMra1114208.
9. Salman AG, Kanigoro B, Heryadi Y. Weather forecasting using deep learning techniques. In: 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS); 2015 Oct 10–11; Depok. New York: IEEE; 2016. p. 281–285.
10. Shi X, Gao Z, Lausen L, *et al.* Deep learning for precipitation nowcasting: A benchmark and a new model. Advances in Neural Information Processing Systems 2017; 30: 5622–5632. doi: 10.48550/arXiv.1706.03458.
11. Xu L, Chen N, Chen Z, *et al.* Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. Earth-Science Reviews 2021; 222: 103828. doi: 10.1016/j.earscirev.2021.103828.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–444. doi: 10.1038/nature14539.
13. Reichstein M, Camps-Valls G, Stevens B, *et al.* Deep learning and process understanding for data-driven Earth system science. Nature 2019; 566: 195–204. doi: 10.1038/s41586-019-0912-1.
14. Xiao C, Chen N, Hu C, *et al.* A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environmental Modelling & Software 2019; 120: 104502. doi: 10.1016/j.envsoft.2019.104502.
15. Zheng G, Li X, Zhang RH, Liu B. Purely satellite data–driven deep learning forecast of complicated tropical instability waves. Science Advances 2020; 6(29): eaba1482. doi: 10.1126/sciadv.aba1482.
16. Ham YG, Kim JH, Luo JJ. Deep learning for multi-year ENSO forecasts. Nature 2019; 573: 568–572. doi: 10.1038/s41586-019-0912-1.
17. Xu L, Chen N, Zhang X, Chen Z. A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale. Climate Dynamics 2020; 54: 3355–3374. doi: 10.1007/s00382-020-05173-x.
18. Hagras H. Toward human-understandable, explainable AI. Computer 2018; 51(9): 28–36. doi: 10.1109/MC.2018.3620965.
19. Zhang Q, Zhu S. Visual interpretability for deep learning: A survey. Frontiers of Information Technology & Electronic Engineering 2018; 19(1): 27–39. doi: 10.1631/FITEE.1700808.
20. Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature 2015; 525: 47–55. doi: 10.1038/nature14956.
21. Slingo J, Palmer T. Uncertainty in weather and climate prediction. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2011; 369: 4751–4767. doi: 10.1098/rsta.2011.0161.
22. Wahl S. Uncertainty in mesoscale numerical weather prediction: Probabilistic forecasting of precipitation [PhD thesis]. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn; 2015.
23. Ayzel G, Scheffer T, Heistermann M. RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting. Geoscientific Model Development 2020; 13(6): 2631–2644. doi: 10.5194/gmd-

13-2631-2020.
24. Ravuri S, Lenc K, Willson M, *et al*. Skilful precipitation nowcasting using deep generative models of radar. Nature 2021; 597: 672–677. doi: 10.1038/s41586-021-03854-z.
25. Frame JM, Kratzert F, Klotz D, *et al.* Deep learning rainfall–runoff predictions of extreme events. Hydrology and Earth System Sciences 2022; 26: 3377–3392. doi: 10.5194/hess-26-3377-2022.
26. Schltz MG, Betancourt C, Gong B, *et al.* Can deep learning beat numerical weather prediction? Philosophical Transactions of the Royal Society A 2021; 379(2194): 20200097. doi: 10.1098/rsta.2020.0097.
27. USGS 30 ARC-second global elevation data, GTOPO30 [Report]. USGS Earth Resources Observation and Science (EROS) Center; 1996. doi: 10.5065/A1Z4-EE71.
28. An D, Du Y, Berndtsson R, *et al.* Evidence of climate shift for temperature and precipitation extremes across Gansu Province in China. Theoretical and Applied Climatology 2020; 139(5): 1137–1149. doi: 10.1007/s00704-019-03041-1.
29. Li C, Wang R. Recent changes of precipitation in Gansu, Northwest China: An index-based analysis. Theoretical and Applied Climatology 2017; 129: 397–412. doi: 10.1007/s00704-016-1783-0.
30. Hersbach H, Bell B, Berrisford P, *et al.* The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 2020; 146(730): 1999–2049. doi: 10.1002/qj.3803.
31. Kingma DP, Ba J. Adam: A method for stochastic optimization. ArXiv 2014; 1412.6980. doi: 10.48550/arXiv.1412.6980.
32. Murphy AH. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather and Forecasting 1992; 7(4): 692–698. doi: 10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.
33. Wilks DS. Statistical methods in the atmospheric sciences. Cambridge: Academic press; 2011.
34. Schoof JT, Pryor SC. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. International Journal of Climatology 2001; 21(7): 773–790. doi: 10.1002/joc.655.
35. Skamarock WC, Klemp JB, Dudhia J, *et al*. A description of the advanced research WRF version 2 [Report]. National Center for Atmospheric Research Boulder; 2005. doi: 10.5065/D68S4MVH.
36. Kain JS. The Kain–Fritsch convective parameterization: An update. Journal of Applied Meteorology 2004; 43(1): 170–181. doi: 10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
37. Chen S, Sun W. A one-dimensional time dependent cloud model. Journal of the Meteorological Society of Japan. Ser. II 2002; 80(1): 99–118. doi: 10.2151/jmsj.80.99.
38. NCEP FNL operational model global tropospheric analyses, continuing from July 1999 [Report]. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory; 2000. doi: 10.5065/D6M043C6.
39. Huth R. Statistical downscaling of daily temperature in central Europe. Journal of Climate 2002; 15(13): 1731–1742. doi: 10.1175/1520-0442(2002)015<1731:SDODTI>2.0.CO;2.
40. Cena K, Monteith JL. Transfer processes in animal coats. II. Conduction and convection. Proceedings of the Royal Society of London Series B Biological Sciences 1975; 188(1093): 395–411. doi: 10.1098/rspb.1975.0027.
41. Hanania NA, Zimmerman JL. Accidental hypothermia. Critical Care Clinics 1999; 15(2): 235–249. doi: 10.1016/S0749-0704(05)70052-X.
42. Kang J, Rong Y. Modeling and simulation of load heating in heat treatment furnaces. Journal of Materials Processing Technology 2006; 174(1–3): 109–114. doi: 10.1016/j.jmatprotec.2005.03.037.
43. Bernard V, Staffa E, Mornstein V, Bourek A. Infrared camera assessment of skin surface temperature—Effect of emissivity. Physica Medica 2013; 29(6): 583–591. doi: 10.1016/j.ejmp.2012.09.003.
44. Osczevski RJ. The basis of wind chill. Arctic 1995; 48(4): 372–382. doi: 10.14430/arctic1262.
45. Woodson WE, Tillman B, Tillman P. Human factors design handbook: Information and guidelines for the design of systems, facilities, equipment, and products for human use. New York: McGraw-Hill; 1992. p. 861.
46. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. Thermal comfort. In: 2005 ASHRAE handbook: Fundamentals. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.; 2005.
47. Shi R, Zhang J, Fang B, *et al.* Runners' metabolomic changes following marathon. Nutrition & Metabolism 2020; 17: 19. doi: 10.1186/s12986-020-00436-0.
48. Frankenfield D, Roth-Yousey L, Compher C, Group EAW. Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: A systematic review. Journal of the American Dietetic Association 2005; 105(5): 775–789. doi: 10.1016/j.jada.2005.02.005.
49. Shi X, Chen Z, Wang H, *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems; 2015 Dec 7–12; Montreal, Canada. Cambridge: MIT Press; 2015. p. 802–810.
50. Xu L, Chen N, Zhang X, Chen Z. An evaluation of statistical, NMME and hybrid models for drought prediction in China. Journal of Hydrology 2018; 566: 235–249. doi: 10.1016/j.jhydrol.2018.09.020.
51. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015; 521: 452–459. doi: 10.1038/nature14541.
52. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: 33rd International Conference on Machine Learning: PMLR; 2016 Jun 19–24; New York. Cambridge: JMLR; 2016. p. 1050–1059.

53. Fente DN, Singh DK. Weather forecasting using artificial neural network. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT); 2018 Apr 20–21; Coimbatore. New York: IEEE; 2018. p. 1757–1761.
54. Kumar R. Decision tree for the weather forecasting. International Journal of Computer Applications 2013; 76(2): 31–34. doi: 10.5120/13220-0620.
55. Vitart F, Ardilouze C, Bonet A, *et al.* The subseasonal to seasonal (S2S) prediction project database. Bulletin of the American Meteorological Society 2017; 98(1): 163–173. doi: 10.1175/BAMS-D-16-0017.1.
56. Hewage P, Trovati M, Pereira E, Behera A. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications 2021; 24(4): 343–366. doi: 10.1007/s10044-020-00898-1.
57. Hossain M, Rekabdar B, Louis SJ, Dascalu S. Forecasting the weather of Nevada: A deep learning approach. In: 2015 International Joint Conference on Neural Networks (IJCNN); 2015 Jul 12–17; Killarney. New York: IEEE; 2015. p. 1–6.
58. Chakraborty S, Tomsett R, Raghavendra R, *et al.* Interpretability of deep learning models: A survey of results. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation; 2017 Aug 4–8; San Francisco. New York: IEEE; 2017. p. 1–6.
59. Wedi NP, Polichtchouk I, Dueben P, *et al.* A baseline for global weather and climate simulations at 1 km resolution. Journal of Advances in Modeling Earth Systems 2020; 12(11): e2020M–e2192M. doi: 10.1029/2020MS002192.