We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX — CLARIVATE ANALYTICS — INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Thunderstorm Predictions Using Artificial Neural Networks

Waylon G. Collins and Philippe Tissot

Additional information is available at the end of the chapter

Abstract

Artificial neural network (ANN) model classifiers were developed to generate ≤ 15 h predictions of thunderstorms within three 400-km² domains. The feed-forward, multi-layer perceptron and single hidden layer network topology, scaled conjugate gradient learning algorithm, and the sigmoid (linear) transfer function in the hidden (output) layer were used. The optimal number of neurons in the hidden layer was determined iteratively based on training set performance. Three sets of nine ANN models were developed: two sets based on predictors chosen from feature selection (FS) techniques and one set with all 36 predictors. The predictors were based on output from a numerical weather prediction (NWP) model. This study amends an earlier study and involves the increase in available training data by two orders of magnitude. ANN model performance was compared to corresponding performances of operational forecasters and multi-linear regression (MLR) models. Results revealed improvement relative to ANN models from the previous study. Comparative results between the three sets of classifiers, NDFD, and MLR models for this study were mixed—the best performers were a function of prediction hour, domain, and FS technique. Boosting the fraction of total positive target data (lightning strikes) in the training set did not improve generalization.

Keywords: thunderstorm prediction, artificial neural networks, correlation-based feature selection, minimum redundancy maximum relevance, multi-linear regression

1. Introduction

A *thunderstorm or convective storm* is a cumulonimbus cloud that produces the electric discharge known as lightning (which produces thunder) and typically generates heavy rainfall, gusty

surface wind, and possibly hail [1]. Meteorologists use the term convection/convective to refer to the vertical component of convective heat transfer owing to buoyancy [2–4]. The term *deep moist convection* (DMC), which refers to the overturning of approximately the entire troposphere due to convective motions and involving condensation of water vapor associated with rising parcels [5], is part of the literature and includes both thunderstorms and moist convection not involving thunder [4]. The terms *thunderstorm, convective storm*, and *convection* are used interchangeably in this chapter to refer to thunderstorms. Thunderstorms adversely affect humans and infrastructure. An estimated 24,000 deaths and 240,000 injuries worldwide are attributable to lightning [6, 7]. In the USA, lightning is the third leading cause of storm-related deaths based on averages during the period 1985–2014 [8]. Additional hazards that can occur include large hail, flash flooding associated with heavy rainfall, and damage from wind from tornadoes and/or non-rotational (straight-line) wind. During the period 1980–2014, 70 severe thunderstorm events each totaling ≥ 1 billion US dollar damage occurred in the USA, which totaled to 156.3 billion US dollars (adjusted for inflation to 2014 dollars) [9]. Further, convection exacts an economic cost on aviation in terms of delays [10]. Given the adverse socioeconomic impact associated with thunderstorms, there is motivation to predict thunderstorm occurrence and location to inform the public with sufficient lead time.

However, the complexity of thunderstorm generation (hereafter *convective initiation*, or CI), given the myriad of processes (operating on different scales) that influence the vertical thermodynamic structure of the atmosphere (that directly influences the thunderstorm development) and the CI itself [11], the characteristic Eulerian (with respect to a fixed point at the surface) time and linear space scales of individual thunderstorms [12], and the inherent predictability limitations of atmospheric phenomena on the scale of individual thunderstorms [13, 14], renders the skillful prediction of thunderstorm occurrence, timing, and location very difficult. This chapter begins by explaining the thunderstorm development process in order to help the reader understand the predictor variables used in the artificial neural network (ANN) models. Next, the variety of methods used to predict thunderstorms are presented in order to acquaint the reader with the relative utility of the ANN model option. The main section starts with an account of the previous methods developed by the authors and presents a new approach to the development of ANN models to predict thunderstorms with high temporal and spatial resolution. Finally, concluding remarks, including a discussion of future research.

2. Thunderstorm development

To properly understand the process of thunderstorm development, it is essential to define the terms *parcel* and *environment*. The environment refers to the ambient atmospheric conditions. In general, a parcel is an imaginary volume of air that can be assigned various properties [1]. A parcel in this discussion is infinitesimal in dimension and is assumed to be thermally insulated from the surrounding environment which allows for adiabatic temperature changes owing to vertical displacement, and it has a pressure that immediately adjusts to the environmental pressure [15]. One method used by meteorologists to assess the potential for the development of thunderstorms is the *parcel method* [3]. This method is used to assess atmos-

pheric stability and involves the finite vertical displacement of a parcel from hydrostatic equilibrium (balance between vertical pressure force and gravity) while the environment remains unchanged. After the displacement, the temperature contrast between the parcel and the environment at the same level results in buoyancy forces that determine stability.

Consider an environment of depth H with a temperature lapse rate (decrease in temperature with height) Γ satisfying the condition $\Gamma_m < \Gamma < \Gamma_d$, where $\Gamma_d = 9.8°Ckm^{-1}$ and $\Gamma_m = 6.5°Ckm^{-1}$ are the dry and moist adiabatic/pseudoadiabatic lapse rates, respectively. Now consider a *surface-based parcel*, defined as a parcel originating in the convective *planetary boundary layer* (PBL) (also called *mixing layer*) that eventually contribute to the primary thunderstorm updraft if thunderstorms develop [16]. The convective PBL refers to the bottom layer of the atmosphere in contact with the earth surface with a diurnal depth varying between tens of meters (near sunrise) to 1–4 km (near sunset) [1]. Let us begin with an unsaturated parcel located at a particular position $(x, y, z = h)$ within the PBL and at the same temperature, density, and pressure as the environment (hydrostatic equilibrium.) Consider an upward vertical displacement of this parcel and apply the parcel method. Since the parcel is unsaturated, it will initially cool at the dry adiabatic lapse rate. Since $\Gamma < \Gamma_d$, the parcel will become cooler than its surrounding environment. Applying the ideal gas law and the assumption that the parcel's pressure instantly adjusts to the environmental pressure, the parcel's density becomes greater than environmental air density. Thus, the parcel is negatively buoyant, a condition known as *positive static stability* [15]. If this parcel is released, it will return to its original height h with negative buoyancy acting as the restoring force. However, let us assume that the parcel overcomes this negative buoyancy via certain upward-directed external force. The parcel eventually reaches the *lifted condensation level* (LCL), whereby the parcel becomes saturated followed by condensation. (The condensation of parcels is manifested by the development of cumulus clouds.) Due to the associated latent heat release, the parcel cools at the pseudoadiabatic rate during subsequent lift. Since $\Gamma_m < \Gamma$, the parcel now cools at a lesser rate than the environmental rate and (with the help of the external force) will eventually reach the environmental temperature at the level referred to as the *level of free convection*. Afterward, the parcel becomes warmer than the environment and thus positively buoyant. If the parcel is released, it will continue to rise without the aid of an external force, a condition known as *static instability*. Thus, a parcel with sufficient vertical displacement within an environment with lapse rates following the $\Gamma_m < \Gamma < \Gamma_d$ constraint may become positively buoyant. This condition is known as *conditional instability*.

The parcel will remain positively buoyant until it reaches the *equilibrium level* (EL). The magnitude of the energy available to a given parcel for convection is the *convective available potential energy* (CAPE) [3] which is the integrated effect of the parcel's positive buoyancy between its original height h and the EL:

$$CAPE_h = \int_{p_{EL}}^{p_h} R_d \left(T_{vp} - T_{ve} \right) dlnp \tag{1}$$

The variables T_{vp}, T_{ve}, R_d, and p refer to the virtual temperatures of the parcel and environment, specific gas constant for dry air, and air pressure, respectively. Recall that before the surface-based parcel reached the LFC, an external force was needed. The amount of energy required of the external force to lift the parcel to its LFC is known as *convective inhibition* (CIN) [1], represented as follows:

$$CIN_h = - \int_{p_h}^{p_{LFC}} R_d \left(T_{vp} - T_{ve} \right) dlnp \tag{2}$$

Now, consider a separate case whereby the environment is *absolutely stable* with respect to a surface-based parcel ($\Gamma_m > \Gamma$). This condition is characterized by negative buoyancy during the parcel's entire vertical path of depth H within the troposphere. Hence, the parcel method would suggest that convection is not possible. However, consider a layer of depth $l \leq H$ within this environment where water vapor content decreases rapidly with height. Owing to this moisture profile, if the entire layer l is lifted, parcels at the bottom of the layer will reach their LCL before parcels at the top of the layer. The differential lapse rates within this layer resulting from continued lifting will transform the layer from absolutely stable to conditionally unstable. This condition is known as *convectively (or potential) instability* [3, 15]. A convectively unstable layer is identified as one that satisfies:

$$\frac{\partial \theta_e}{\partial z} < 0 \tag{3}$$

The symbols θ_e and z refer to *equivalent potential temperature* and geometric height, respectively. It must be emphasized that CAPE is necessary for the potential for convection [3]. Recall that a convectively unstable layer is not necessarily unstable. In this example, the environment is absolutely unstable, devoid of positive buoyancy, and thus CAPE = 0. A mechanism is required to lift the convectively unstable layer to one characterized by conditional instability.

Air parcels extending above the LFC accelerate upward owing to positive buoyancy and draw energy for acceleration from CAPE. The relationship between maximum updraft velocity (w_{max}) and CAPE (parcel theory) is as follows:

$$w_{max} = \left(2CAPE \right)^{1/2} \tag{4}$$

The moist updraft is integral to the development of a supersaturated condition that results in excess water vapor that (with the aid of hygroscopic aerosols that serve as cloud condensation nuclei or CNN) condenses to form water in liquid and solid form (condensate) manifested as the development of a *cumulus cloud*, followed by the transition to *cumulus congestus*, then ultimately to *cumulonimbus*. With respect to the production of rain during convection, the

stochastic *collision-coalescence* mechanism is likely the predominant process that transforms cloud droplets (with broad drop-size distribution) to liquid hydrometeors large enough (diameter 500 μm) to combine with gravity and fall as rain on convective time scales [17].

As saturated parcels rise to the region with environmental temperatures colder than −4°C, the likelihood that ice crystals will develop within the cloud increases. Further, a fraction of water remains in liquid form (supercooled water) until around −35°C [15]. Thus, a region character‐ized by water in all three phases (vapor, liquid, and solid) develops. The development of the solid hydrometeors known as *ice crystals* and *graupel* within this mixed phase region contribute to the development of lightning. In particular, ice-graupel collisions contribute to the transfer of negative (positive) charge to the larger graupel (smaller ice crystal) particles with charge separation caused by gravity, resulting in a large-scale positive dipole within the cumulonim‐bus [18]. A smaller positively charged region exists near the cloud base. *Intracloud* (IC) lightning occurs in response to the foregoing dipole, *cloud-to-ground* (CTG) lightning involves a transfer of negative charge from the dipole to the earth surface, and the less common *cloud-to-air* (CTA) lightning variety links the large-scale negative charge with the smaller positive charge near cloud base [18]. The temperatures in the air channel through which lightning occurs exceed the surface of the sun and result in a shockwave followed by a series of sound waves recognized as thunder that is heard generally 25 km away from the lightning occurrence [1, 15].

Straight-line thunderstorm surface winds develop as negative buoyancy (owing to cooling associated with evaporation of water/melting of ice), condensate loading (weight of precipi‐tation dragging air initially downward before negative buoyancy effects commence), and/or downward-directed pressure gradient force (associated with convection developing within strong environmental vertical wind shear) contribute to the generation of the *convective downdrafts* which are manifested as *gust fronts* (also called *outflow boundaries*) after contact with the earth surface [19]. Hail associated with a thunderstorm involves a complex process whereby graupel and frozen raindrops serve as pre-existing hail embryos and transition to hailstones by traveling along optimal trajectories favorable for rapid growth within the region of the cumulonimbus composed of supercooled water [20].

Given the foregoing thunderstorm development process, the simultaneous occurrence of three conditions are necessary for CI: sufficient atmospheric moisture, CAPE, and a lifting/triggering mechanism. Moisture is necessary for the development of the cumulonimbus cloud condensate which serves as a source material for the development of hydrometeors rain, ice crystals, graupel, and hail. The environmental moisture profile contributes to the development conditional and convective instability. CAPE provides the energy for updraft to heights necessary for the development of the cumulonimbus cloud and the associated mixed phase region that contributes to lightning via charge separation. A mechanism is necessary to lift surface-based parcels through the energy barrier to their LFC, and to lift convectively unstable layers necessary for the development of conditional instability. A myriad of phenomena can provide lift, including fronts, dry lines, sea breezes, gravity waves, PBL horizontal convective rolls, orography, and circulations associated with local soil moisture/vegetation gradients [11, 21].

A myriad of synoptic scale [12] patterns/processes can alter the thermodynamic structure of the environment at a particular location to one favorable for the development of CAPE or convective instability [22]. One scenario in the USA involves the advection of lower-level moist air toward the north across the Southern Plains from the Gulf of Mexico in advance of an upper-level disturbance approaching from the west and advecting midtropospheric dry air originating from the desserts of northern Mexico. The thermodynamic profile over a location influenced by those air masses (e.g., Oklahoma City, Oklahoma) would become one characterized by both conditional and convective instabilities owing to the dry air mass moving over the moist air mass [3].

The foregoing discussion is not exhaustive with respect to thunderstorms. The transition to *severe* convective storms (defined as thunderstorms which generate large hail, damaging straight-line wind, and/or tornadoes), flash flooding, and convective storm mode (squall lines, single cells, multi-cells, supercells, etc.) are not relevant to the development of *non-severe* (also called *ordinary*) thunderstorms in general and are not discussed. Further, *slantwise convection* owing to *conditional symmetric instability* due to the combination of gravitational and centrifugal forces [1] is not considered.

3. Thunderstorm prediction methods

We classify thunderstorm prediction based on the following methods: (1) numerical weather prediction (NWP) models, (2) post-processing of NWP model ensemble output, (3) the post-processing of single deterministic NWP model output via statistical and artificial intelligence (AI)/machine learning (ML), and (4) classical statistical, AI/ML techniques.

3.1. Secondary output variables/parameters from Numerical Weather Prediction (NWP) models

NWP models are based on the concept of determinism which posits that future states of a system evolve from earlier states in accordance with physical laws [23]. Meteorologists describe atmospheric motion by a set of nonlinear partial differential conservation equations —derived from Newton's second law of motion for a fluid, the continuity equation, the equation of state, and the thermodynamic energy equation—that describe atmospheric heat, momentum, water, and mass referred to as *primitive equations, Euler equations*, or *equations of motion* [24–26]. These equations cannot be solved analytically and are thus solved numerically. Further, the earth's atmosphere is a continuous fluid with 10^{44} molecules (Appendix A) which the state-of-the-art NWP models cannot resolve. Thus, NWP model developers undertake the process known as *discretization*, which involves the representation of the atmosphere as a three-dimensional (3D) spatial grid (which divides the atmosphere into volumes or grid cells), the representation of time as finite increments, and the substitution of the primitive equations with corresponding numerical approximations known as *finite difference equations* solved at the grid points [26]. Atmospheric processes resolved by the NWP equations are termed *model dynamics* while unresolved processes are *parameterized* via a series of equations collectively known as

model physics [25]. Parameterization involves using a set of equations on the resolved scale to implicitly represent the unresolved process. These unresolved (sub-grid-scale) processes include solar/infrared radiation and microphysics (which occur on the molecular scale), cumulus convection, earth surface/atmosphere interactions, and planetary boundary layer/ turbulence [27]. If these primary unresolved processes are not taken into account, the quality of NWP output would deteriorate in less than 1 h when simulating the atmosphere at hori‐ zontal grid scales of 1–10 km [25]. The NWP model prediction process is an *initial value problem*. A process known as *data assimilation* is used to provide the requisite initial values. A predominate data assimilation technique involves the use of balanced (theoretical and observed winds in phase) short-term output from an earlier NWP model run to serve as a first guess, followed by the incorporation of meteorological observations to create a balanced *analysis* which serve as the initial condition for the NWP model [26]. Next, the finite difference equations are solved forward in time. The primary output variables include temperature, wind, pressure/height, mixing ratio, and precipitation. At the completion of the NWP model run, *post-processing* is performed which includes the calculation of secondary variables/parameters (CAPE, relative humidity, etc.) and the development of techniques to remove model biases [25, 26, 28].

The state-of-the-art high-resolution NWP models have the ability to explicitly predict/simulate individual thunderstorm cells rather than parameterize the effects of sub-grid-scale convec‐ tion [29]. NWP output identified as thunderstorm activity involves assessment of NWP output parameter/secondary variable known as *radar reflectivity* defined as the efficiency of a radar target to intercept and return of energy from radio waves [1]. Operational meteorologists in the NWS use radar technology to diagnose/analyze thunderstorms. With respect to hydrome‐ teor targets (rain, snow, sleet, hail, graupel, etc.), radar reflectivity is a function of hydrometeor size, number per volume, phase, shape, and is proportional to six times the effective diameter of the hydrometeor [1]. Radar reflectivity magnitudes ≥ 35dB at the −10°C level are generally regarded as a proxy for CI and for the initial CG lightning flash [30, 31]. Further, the increase in the reflectivity within the mixed-phase region of cumulonimbus clouds correlates strongly with lightning rate [32]. An example of a high-resolution (≤4-km) NWP model that can simulate/predict *radar reflectivity* is version 1.0.4 of the 3-km High-Resolution Rapid Refresh (HRRR) Model, developed by National Oceanic and Atmospheric Administration (NOAA)/ Oceanic and Atmospheric Research (OAR)/Earth Systems Research Laboratory implemented by the National Weather Service (NWS) National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) on 30 September 2014 to support NWS operations [33]. The specific model dynamical core, physics, and other components are detailed in Appendix B. Yet, we emphasize here that the HRRR does not incorporate cumulus/ convective parameterization (CP), thus allowing for the explicit prediction of thunderstorm activity. One of the output parameters relevant to thunderstorm prediction is *simulated radar reflectivity 1-km (dBZ)*, which is an estimate of the *radar reflectivity* at the constant 1-km level.

Despite the utility of NWP models, there exist fundamental limitations. In particular, the atmosphere is chaotic—a property of the class of deterministic systems characterized by sensitive dependence on the system's initial condition [13, 14, 23]. Thus, minor errors between

the initial atmospheric state and the NWP model representation of the initial state can result in a future NWP solution divergent from the future true atmospheric state. Unfortunately, a true, exact, and complete representation of the initial state of the atmosphere using the state-of-the-art NWP models is not possible. Even if the NWP model could perfectly represent the initial atmospheric state, errors associated with imperfections inherent in model formulation and time integration would grow. Model discretization and physics parameterizations introduce errors. Further, the gradient terms in the finite difference equations are approximated using a Taylor series expansion of only a few orders [26, 34], thus introducing *truncation* error. Errors associated with the initial condition, discretization, truncation, and parameterization limits predictability; an intrinsic finite range of predictability exists that is positively correlated to spatial scale [14]. A high-resolution NWP simulation of a tornadic thunderstorm can result in inherent predictability with lead times as short as 3–6 h [35].

Accurately predicting the exact *time* and *location* of individual convective storms is extremely difficult [36]. High-resolution (≤4-km) NWP models can accurately simulate/predict the *occurrence* and *mode* of convection (e.g., whether a less common left-moving and devastating *supercell thunderstorm* will develop), yet have difficulty with regard to the *time* and *location* (exactly when and where will the *supercell* occur) [37, 38]. Even very high-resolution (≤1-km) NWP models that can resolve and predict individual convective cells [29] will not necessarily provide greater accuracy and skill relative to coarser resolution NWP models [39].

3.2. Post processing of NWP model ensembles

Methods exist to generate more optimal or skillful thunderstorm predictions/forecasts when using NWP models. One such method is known as ensemble forecasting [40], which is essentially a Monte Carlo approximation to *stochastic dynamical forecasting* [28, 41]. Stochastic dynamical forecasting is an attempt to account for the uncertainty regarding the true initial atmospheric state. The idea is to run an NWP model on a probability distribution (PD) that describes initial atmospheric state uncertainty. Due to the impracticability of this technique, a technique was proposed whereby the modeler chooses a small random sample of the PD describing initial state uncertainty [41]; the members are collectively referred to as ensemble of initial conditions. The modeler then conducts an NWP model run on each member of the ensemble, hence the term ensemble forecasting [28]. In practice, each member of the ensemble represents a unique combination of model initial condition, dynamics, and/or physics [27, 42]. An advantage of ensemble forecasting over prediction with single deterministic NWP output is the ability to assess the level of forecast uncertainty. One method to assess this uncertainty is to assume a positive correlation between uncertainty and the divergence/spread in the ensemble members [28]. Prediction probabilities can be generated by post-processing the ensemble. Applying ensembles to thunderstorm forecasting, the NWS Environmental Modeling Center developed the *short-range ensemble forecast (SREF),* a collection of selected output from an ensemble of 21 mesoscale (16-km) NWP model runs. The NWS Storm Prediction Center (SPC) post-processes SREF output to create a quasi-real-time suite of products that includes the *calibrated probabilistic prediction of thunderstorms* [42]. There exist utility in the use

of ensembles in thunderstorm forecasting. According to [16], the timing of CI within selected mesoscale regions can be predicted accurately using an ensemble-based approach.

One limitation of NWP ensembles to support operational forecasters is the tremendous computational cost necessary to run since each ensemble member is a separate NWP model run. Another limitation is the realization that the true PD of the initial condition uncertainty is unknown and changes daily [28].

3.3. Post-processing of single deterministic NWP model output using other statistical and artificial intelligence/machine learning techniques

Statistical methods can be utilized to post-process NWP output to correct for certain systematic NWP model biases and to quantify the level of uncertainty in single NWP deterministic output [28, 43]. Statistical post-processing methods include *model output statistics* (MOS) [44] and *logistic regression* [28].

MOS involves the development of data-driven models to predict the future state of a target based on a data set of past NWP output (features/predictors) and the corresponding target/predictand. Following [28], a regression function f_{MOS} is developed to fit target Y at future time t to a set of predictors/features (from NWP output known at $t = 0$) represented by vector x. The development and implementation is as follows:

$$Y_t = f_{MOS}(x_t) \tag{5}$$

One limitation of the MOS technique involves NWP model changes made by the developers. If NWP model adjustments alter *systematic errors*, new MOS equations should be formulated [28]. Model changes can occur somewhat frequently. Consider the North American Mesoscale (NAM), which is the placeholder for the official NWS operational mesoscale NWP model for the North American domain. On 20 June 2006, the NAM representative model switched from the *hydrostatic Eta* to the *Weather Research and Forecasting (WRF)-Non-hydrostatic Mesoscale Model (NMM)*, a change in both the model dynamical formulation and modeling framework. Then, on 1 October 2011, the NAM representative model switched to *NOAA Environmental Modeling System Non-hydrostatic Multiscale Model on the B-grid (NEMS-NMMB)* resulting in changes to the model framework (WRF to NEMS) and model discretization (change from Arakawa E to B grid) (Appendix C).

The NWS uses multiple linear regression (MLR) (with forward selection) applied to operational NWP model predictors and corresponding weather observations (target) to derive MOS equations to support forecast operations [43]. The NWS provides high-resolution gridded MOS products which include a 3-h probability of thunderstorms [45] and thunderstorm probability forecasts as part of the NWS *Localized Aviation Model Output Statistics Program* (LAMP) [46].

Logistic regression is a method to relate the predicted probability p_j of one member of a binary target to the jth set of n predictors/features $(x_1, x_2, ..., x_n)$ to the following nonlinear equation:

$$p_j = \frac{1}{1 + exp\left(-b_0 - b_1 x_1 - b_2 x_2 - \ldots - b_n x_n\right)} \qquad (6)$$

The regression parameters are determined via the method of *maximum likelihood* [28].

Logistic regression models were used by [47] to develop MOS equations to generate probability of thunderstorms and the conditional probability of severe thunderstorms in twelve 7200 km^2 regions at 6-h projections out to 48-h in the Netherlands. The NWP output was provided by both the High-Resolution Limited-Area Model (HIRLAM) and the European Centre for Medium-Range Weather Forecasts (ECMWF) NWP model. The Surveillance et d'Alerte Foudre par Interférométrie Radioélectrique (SAFIR) lightning network provided the target data. Verification results suggest that the prediction system possessed good skills.

Artificial intelligence (AI) involves the use of computer software to reproduce human cognitive processes such as learning and decision making [1, 48]. More recently, the term *machine learning* (ML) is used to describe the development of computer systems that improve with experience [1, 49]. Specific AI/ML techniques involving the post-processing of NWP output include *expert systems* [50], *adaptive boosting* [51], *artificial neural networks* [52, 53], and *random forests* [31].

A random forest [54] is a classifier resulting from an ensemble/forest of tree-structured classifiers, whereby each tree is developed from independent random subsamples of the data set (including a random selection of features from which the optimum individual tree predictors are selected) drawn from the original training set. The generalization error (which depends on individual tree strength and tree-tree correlations) convergences to a minimum as the number of trees becomes large, yet overfitting does not occur owing to the law of large numbers. After training, the classifier prediction is the result of a synthesis of the votes of each tree.

In one study, selected thermodynamic and kinematic output from an Australian NWP model served as input for an expert system using the decision tree method to assess the likelihood of thunderstorms and severe thunderstorms [50]. Further, an artificial neural network model to predict significant thunderstorms that require the issuance of the Convective SIGMET product (called WST), issued by the National Weather Service's Aviation Weather Center, for the 3–7 h period after 1800 UTC; the model demonstrated skill, including the ability to narrow the WST outlook region while still capturing the subsequent WST issuance region [52]. Logistic regression and random forests were used to develop models to predict convective initiation (CI) ≤1 h in advance. The features/predictors included NWP output and selected Geostationary Operational Environmental Satellite (GOES)-R data. The performance of these models was an improvement of an earlier model developed based on GOES-R data alone [31].

3.4. Classical statistical and artificial intelligence/machine learning techniques

Statistical methods not involving the use of NWP model output (classical statistical) that have been used to predict thunderstorm occurrence include *multiple discriminate analysis (MDA),*

scatter diagrams, and *multiple regression*. Corresponding AI/ML techniques include *expert systems*, *artificial neural networks*, and *logistic regression*.

MDA is essentially a form of multiple linear regression used to predict an event. In particular, a discriminant function relates a nonnumerical predictand to a set of predictors; the value of the function that distinguishes between event groups [1, 55]. An MDA was used to obtain 12 h prediction functions to distinguish between the following two or three member groups within selected domains in portions of the Central and Eastern USA: thunderstorm/no-thunderstorm, thunderstorm/severe thunderstorm, and thunderstorm/severe thunderstorm/no-thunderstorm. The verification domains were within 1° latitude radius relative to the position of the data source that provided the predictor variables. The MDA prediction system provided skill in the 0-12 h period [55].

In one study, the utility of both a graphical method and multiple regression was tested to predict thunderstorms [56]. The graphical method involved scatter diagrams that were used to analyze multiple pairs of atmospheric stability index parameters in order to discover any diagram(s) whereby the majority of thunderstorm occurrence cases were clustered within a zone while the majority of the non-thunderstorm occurrence cases were outside of the zone. Two such diagrams were found—scatter diagrams of Showalter index versus Total Totals index, and Jefferson's modified index versus the George index. The multiple regression technique involved the stepwise screening of 274 potential predictors to 9 remaining. Both prediction models provided thunderstorm predictions in probabilistic terms. Objective techniques were used to convert probabilistic predictions to binary predictions for the purpose of verification. Results indicate that the multiple regression model performed better.

An expert system is a form of artificial intelligence that attempts to mimic the performance of a human expert when making decisions. The expert system includes a knowledge base and an inference engine. The knowledge base contains the combination of human knowledge and experience. Once the system is developed, questions are given to the system and the inference engine uses the knowledge base and renders a decision [57]. An expert system was developed using the decision tree method to forecast the development of thunderstorms and severe thunderstorms [58]; the tree was based on physical reasoning using the observation of meteorological parameters considered essential for convective development. An expert system named *Thunderstorm Intelligence Prediction System (TIPS)* was developed to predict thunderstorm occurrence [59]. Selected thermodynamic sounding data from 1200 UTC and forecaster assessment of likely convective triggering mechanisms were used to forecast thunderstorm occurrence for the subsequent 1500–0300 UTC period. Critical values of five (5) separate atmospheric stability parameters for thunderstorm occurrence and corresponding consensus rules served as the knowledge base. The forecaster answer regarding trigger mechanism and the values of the stability parameters served as input to the inference engine, which interrogated the knowledge base and provided an answer regarding future thunder-storm occurrence. Verification of TIPS revealed utility.

The artificial neural network (ANN) has been used to predict thunderstorms. Thermodynamic data from Udine Italy rawinsondes, surface observations, and lightning data were used to train/optimize an ANN model to predict thunderstorms 6 h in advance over a 5000-km^2 domain

in the Friuli Venezia Giulia region [60]. An ANN model was developed (also using thermo-dynamic data) to predict severe thunderstorms over Kolkata India during the pre-monsoon season (April–May) [61].

Logistic regression was used to develop a binary classifier to predict thunderstorms 6–12 h in advance within a 6825–km^2 region in Spain. A total of 15 predictors (combination of stability indices and other parameters) were chosen to adequately describe the pre-convective environment. The classifier generated satisfactory results on the novel data set [62].

A limitation of classical statistics to weather prediction is that utility is bimodal in time; confined to very short time periods (less than a few hours) or long time periods (10 days) [28]. The utility of AI/ML techniques without NWP may not be as restrictive. Convective storm prediction accuracies associated with expert systems may be similar to that of NWP models [36].

4. The utility of post-processing NWP model output with artificial neural networks to predict thunderstorm occurrence, timing, and location

The rapid drop in prediction/forecast accuracy per unit time for classical statistical techniques renders it less than optimal for thunderstorm predictions over time scales greater than nowcasting (≤ 3 h).

Predicting thunderstorms with the use of deterministic NWP models allows for the prediction of atmospheric variables/parameters in the ambient environment that directly influence thunderstorm development. As discussed previously, limitations of NWP models include predictability limitations owing to a chaotic atmosphere and inherent model error growth [13, 14, 23]. NWP model ensembles attempt to account for the uncertainty in the NWP model's initial condition which contributes to predictability limitations in NWP models [40]. The post-processing of NWP model ensembles can generate useful probabilistic thunderstorm output [42]. However, there is a much greater computational cost to generate an ensemble of deterministic runs relative to a single run. The post-processing of output from a single deterministic NWP model can improve skill by minimizing certain systematic model biases, yet predictability limitations associated with single deterministic NWP model output remain.

The authors explored the use of the artificial neural network (ANN) to post-process output from a single deterministic NWP model in an effort to improve thunderstorm predictive skill, based on an adjustment to the author's previous research [53]. As mentioned earlier, this approach involves a much lower computational cost relative to model ensembles. In particular, the single NWP model used in the development of the thunderstorm ANN (TANN) models discussed in this chapter is the 12-km NAM (North American Mesoscale), which refers to either the Eta, NEMS-NMMB, or WRF-NMM model (only one model used at a time; see Appendix C). The model integration cost to run the 21-member SREF ensemble would be of the order of 20 times the cost required to run the 12-km NAM (Israel Jirak 2016, personal communication.)

Given that ANN was developed to capture the parallel distributed processing thought to occur in the human brain and has tremendous pattern recognition capabilities [63–65], we posit that the ANN will learn the limitations/errors associated with a single NWP deterministic model solution to generate skillful forecasts, notwithstanding NWP predictability limitations. Thus, AI/ML rather than NWP model ensembles would be utilized to deal with atmospheric chaos. The remainder of this chapter focuses on the development of ANN models to predict thunderstorms (TANN) that are part of the ongoing research.

5. Artificial neural network models to predict thunderstorms within three South Texas 400-km² domains based solely on a data set within the same domain of each

In an earlier study [53], a thunderstorm ANN (TANN) was developed to predict thunderstorm occurrence in three separate 400-km² square (box) domains in South Texas (USA), 9, 12, and 15 h (+/−2 h) in advance, by post-processing 12-km NWP model output from a single deterministic NWP model (Appendix C) and 4-km sub-grid-scale output from soil moisture magnitude and heterogeneity estimates. A framework was established to predict thunderstorms in 286 box domains (**Figure 1**), yet predictions were only performed for three. The three box regions were strategically located to access model performance within the *Western Gulf Coastal Plain* region (boxes 103 and 238), the more arid *Southern Texas Plains* region (box 73), and within the region with the greatest amount of positive target data/thunderstorm occurrences based on CTG lightning density (box 238.) A skillful model was envisioned based on the combination of the deterministic future state of ambient meteorological variables/parameters related to thunderstorm development and sub-grid-scale data related to CI. TANN models were trained with the foregoing NWP model output and sub-grid data as predictors and corresponding CTG lightning data as the target. Each TANN model was trained using the feed-forward multi-layer perceptron topology with one hidden layer, the log-sigmoid and linear transfer functions in the hidden and output layers, respectively, and the scaled conjugate gradient (SCG) learning algorithm. The SCG algorithm makes implicit use of the second-order terms of the Hessian, searches the weight space/error surface in conjugate directions, avoids problems associated with line searches, and converges more efficiently than gradient descent [66]. The nomenclature X-Y-Z was used to describe the topology, where X, Y, and Z are the number of neurons in the input, hidden, and output layers, respectively. The data set consisted of a training/validation set (2004–2006; 2009–2012) and a testing set (2007–2008) for final performance and comparison with human forecasters and MLR. For each TANN model variant, the number of neurons in the hidden layer was determined iteratively For Y hidden layers, where $Y = \{1 - 10, 12, 15, 17, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100\}$, the TANN model was trained 50 times. After each training, receiver operating characteristic (ROC) curves were generated to determine the optimal threshold, defined as the case where the Peirce Skill Score (PSS) was greatest [67]. The trained ANN with optimal threshold was then used to make predictions on the testing set (2007–2008); the corresponding performance results were calculated. The mean PSS of the 50 results from both the training and testing sets, for each Y,

was archived. The chosen Y corresponded to the least number of hidden layer neurons with a PSS standard error which overlaps the standard error associated with the maximum PSS. Once Y is chosen, the performance results of the TANN are simply the mean of the 50 runs previously archived that corresponds to the chosen Y.
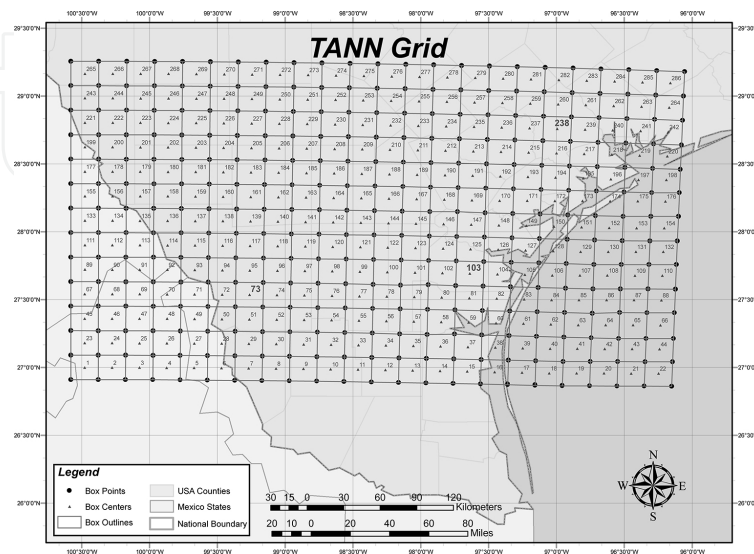

Figure 1. TANN and TANN2 grid domain. This grid is represented as 286 20 km× 20 km regions (boxes). The label inside each box is the identification number (ID). The boxes with ID values in larger font (73, 103, and 238) are the subject of this study. Image from Collins and Tissot [53].

Eighteen (18) TANN classifiers were developed with half based on a reduced set of predictors/ features due to feature selection (FS) and the other half using the full set of 43 potential predictors. FS involves a determination of the subset of potential predictors that describes much of the variability of the target/predictand. This is very important for very large dimensional models which may suffer from the *curse of dimensionality*, which posits that the amount of data needed to develop a skillful data-driven model increases exponentially as model dimension increases linearly [68]. When the features in the subset generated by the FS method are used to train an ANN model, it is important that such features are both relevant and non-redundant. Irrelevant features can adversely affect ANN model learning by introducing noise, and redundant features can result in reduced ANN model predictive skill by increasing the likelihood of convergence on local minima on the error surface during training [69]. The FS technique used was *correlation-based feature selection* (CFS) [70–72]. CFS is a filtering-based FS technique, meaning that statistical associations between the features and the target are assessed outside of the model. Specifically, CFS uses an information theory-based heuristic known as *symmetric uncertainty* to assess feature-target and feature-feature correlations to assess relevance and redundancy, respectively. The search strategy is the *Best First search* with a stopping criterion of five consecutive non-improving feature subsets.

The optimized TANN binary classifiers were evaluated on a novel data set (2007–2008), and performance was compared to MLR-based binary classifiers, and to operational forecasts from the NWS (National Digital Forecast Database, NDFD, Appendix D).

The MLR models were developed via stepwise forward linear regression (SFLR). For each MLR model, the SFLR process began with an empty set of features (constant value $y = \beta_0$. At each subsequent forward step, a predictor is added (from the list of 36 predictors in **Tables 2** and **3**, less the predictors already chosen) based on the change in the value of the Akaike information criterion (AIC) [73], while also considering removal of previously selected predictors based on the same criterion. The optimal MLR model chosen has the smallest AIC which is essentially a trade-off between model size and accuracy based on the training data. The MATLAB® function *stepwiselm* was used to perform SFLR to determine the regression equation coefficients. The resultant MLR models in this study are of the form:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \tag{7}$$

where y_i is the *i*th predictand response, β_j is the *j*th regression coefficient, β_0 is a constant, x_{ij} is the *i*th observation for the *j*th predictor, for $j = 1, \ldots, k$. Finally, ε_i represents error. Each MLR model was transformed into a binary classifier using the same method used in ANN classifier development, except that each MLR model was calibrated on the entire training sample to determine the coefficients, unlike ANN calibration which involved the splitting of the training sample into training and validation data sets. (However, one could argue that the ANN is also using the entire training sample as the validation set is inherently necessary to properly calibrate a model.)

The results were mixed—the TANN, MLR, and human forecasters performed better than the other two depending on the domain, prediction hour, and performance metric used. Results revealed the utility of an automated TANN model to support forecast operations with the limitation that a larger number of individual ANNs must be calibrated in order to generate operational predictions over a large area. Further, the utility of sub-grid-scale soil moisture data appeared limited given the fact that only 1/9 of the TANN models with reduced features retained any of the sub-grid parameters as a feature. The NWP model convective precipitation (CP) feature was retained in all the nine feature selection TANN models, suggesting that CP adequately accounted for the initiation of sub-grid-scale convection. This result is consistent with another study [47] which found that model CP was the most relevant predictor of thunderstorm activity.

6. Artificial neural network models to predict thunderstorms within three South Texas 400-km² domains based on data set from two-hundred and eighty-six 400-km² domains

With respect to TANN skill, the endeavor to predict thunderstorms in a small domain relative to domains used in other studies restricted the amount of CG lightning cases. A large number of thunderstorm cases would be beneficial to the model calibration and verification process; the amount of target data in [53] may have been insufficient to train this data-driven model

with sufficient skill owing to the curse of dimensionality [67]. Further, the method of training/ optimizing TANN models for each 400-km² domain limits operational applicability since it would require the development of literally thousands of TANN models at a country scale. In order to retain thunderstorm predictions in 400-km² domains while increasing predictive skill, a new approach was developed (TANN2), whereby for each prediction hour (9, 12, 15), a single TANN model is trained over all two-hundred and eighty six 400-km² continuous domains. This approach dramatically increased the amount of positive target data (thunderstorm cases) and total cases/instances. **Table 1** depicts the quantity of data utilized in this project. The total number of cases over the study period was 1,148,576 with 939,510 cases used for model calibration and 209,066 cases contained in the 2007–2008 testing set.

Prediction hour	Total instances (training sample)	Positive target data	Percent positive target
9	663,519	22,139	3.3
12	646,073	16,904	2.6
15	659,801	12,682	1.9

Table 1. Quantity of data available to train TANN2 models.

Relative to the previous study [53], only the NWP model and Julian date predictor variables (features) were retained (resulting in 36 potential features used in this study; see **Tables 2** and **3**.) Given that in [53], only one sub-grid parameter was chosen for only 1/9 of the box/prediction hour combinations, and that the model including this sub-grid-scale parameter did not result in classifier performance improvement, the utility of the sub-grid-scale data appeared very limited. As mentioned in [53], the NWP model CP parameter was a ubiquitous output of the FS technique and thus considered a skillful predictor of convection. Physically, it was surmised that CP adequately accounted for the effects of sub-grid-scale convection. The use of FS was retained in order to eliminate irrelevant and redundant features to improve model skill and to reduce model size. The reduction in model size/dimension (owing to FS) and the increase in both the training data set and the amount of target CG lightning cases are expected to result in a more accurate/skillful model when considering the curse of dimensionality.

Abbreviation	Description (Units)	Justification as thunderstorm predictor
PWAT	Total precipitable water (mm)	Atmospheric moisture proxy
MR_{850}	Mixing ratio at 850 hPa (g kg⁻¹)	Lower level moisture necessary for convective cell to reach horizontal scale ≥4 km in order to overcome dissipative effects [84]
RH_{850}	Relative humidity at 850 hPa (%)	When combined with CAPE, predictor of subsequent thunderstorm location independent of synoptic pattern [85]
CAPE	Surface-based convective available potential energy (J kg⁻¹)	Instability proxy; the quantity $(2CAPE)^{0.5}$ is the theoretical limit of thunderstorm updraft velocity [11]

Abbreviation	Description (Units)	Justification as thunderstorm predictor
CIN	Convective inhibition (J kg^{-1})	Surface-based convective updraft magnitude must exceed $(CIN)^{0.5}$ for parcels to reach level of free convection [11]
LI	Lifted index (K)	Atmospheric instability proxy; utility in thunderstorm prediction [86]
U_{LEVEL}, V_{LEVEL}	U,V wind components at surface, 850 hPa (LEVEL = surface, 850 hPa) (ms^{-1})	Strong wind can modulate or preclude surface heterogeneity induced mesoscale circulations [87, 88]
VV_{LEVEL}	Vertical velocity at 925, 700, and 500 hPa (LEVEL = 925, 700, 500 hPa) (Pa s^{-1})	Account for mesoscale and synoptic scale thunderstorm triggering mechanisms (sea breezes, fronts, upper level disturbances) that are resolved by the NAM
DROPOFF$_{PROXY}$	Potential temperature dropoff proxy (K)	Atmospheric instability proxy; highly sensitive to CI [89]
LCL	Lifted condensation level (m)	Proxy for cloud base height; positive correlation between cloud base height and CAPE to convective updraft conversion efficiency [90]
T_LCL	Temperature at the LCL (K)	T_LCL ≥ −10°C essential for presence of supercooled water in convective cloud essential for lightning via graupel-ice crystal collision mechanism [91]
CP	Convective precipitation (kg m^{-2})	By-product of the *Betts-Miller-Janjic* convective parameterization scheme [92] when triggered; proxy for when the NAM anticipates existence of sub-grid-scale convection
VSHEARS8	Vertical wind shear: 10 m to 800 hPa layer (×10^{-3} s^{-1})	The combination of horizontal vorticity (associated with ambient 0–2 km vertical shear), and density current (e.g., gust front) generated horizontal vorticity (associated with 0–2 km vertical shear of opposite sign than that of ambient shear can trigger new convection [93]
VSHEAR86	Vertical wind shear: 800–600 hPa layer (×10^{-3} s^{-1})	Convective updraft must exceed vertical shear immediately above the boundary layer for successful thunderstorm development [58, 89]

Table 2. Description of NAM predictor variables/parameters used in TANN and TANN2 (from [53]).

Abbreviation	Description (units)	Justification as thunderstorm predictor
U_{LEVEL}, V_{LEVEL}	U,V wind at the surface, 900, 800, 700, 600, 500 hPa levels (LEVEL= surface, 900, 800, 700, 600, 500) (ms^{-1})	Thermodynamic profile modification owing to veering of wind (warming) or backing of wind (cooling); backing (veering) of wind in the lowest 300 hPa can suppress (enhance) convective development [94]
HI$_{LOW}$	Humidity index (°C)	Both a constraint on afternoon convection and an atmospheric control on the interaction between soil

Abbreviation	Description (units)	Justification as thunderstorm predictor
		moisture and convection [94]
CTP Proxy	Proxy for convective triggering potential (dimensionless)	Both a constraint on afternoon convection and an atmospheric control on the interaction between soil moisture and convection [95]
VSHEARS7	Vertical wind shear: surface to 700 hPa layer ($\times 10^{-3}$ s^{-1})	Strong vertical shear in the lowest 300 hPa can suppress convective development [94]
VSHEAR75	Vertical wind shear: 700 to 500 hPa layer ($\times 10^{-3}$ s^{-1})	Convective updraft must exceed vertical shear immediately above the boundary layer for successful thunderstorm development [58, 89]

Table 3. Description of NAM initialization variables/parameters used in TANN and TANN2 (from [53]).

With respect to model training, validating, optimizing, and testing, the same strategy was utilized as in [53], with two differences. First, when determining the optimal number of hidden layer neurons, the range of neurons was extended to Y = {1−10, 12, 15, 17, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200}. Second, before each split of the training sample into training and validation components, a training set of the same size of the total training data available was drawn randomly from the training set with replacement. This technique allows for an exploration of training data set variability.
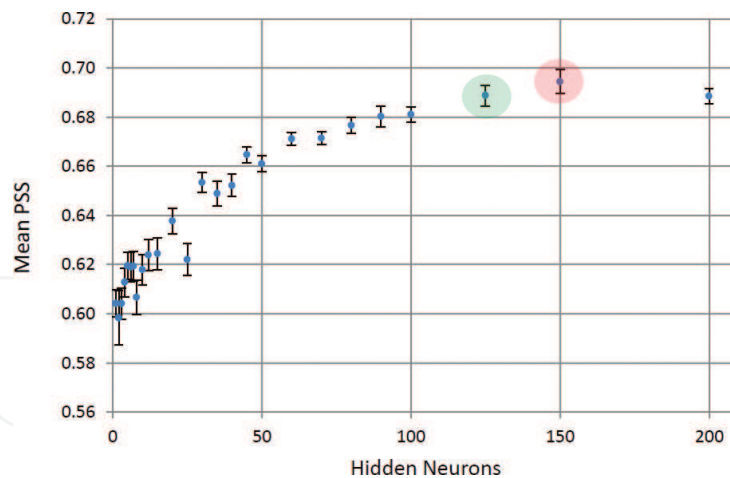

Figure 2. Determination of the optimal number of hidden layer neurons (Y) for the 12 h TANN2 36-Y-1 model (all 36 potential predictors used). Each point and corresponding error bar represents the mean Peirce Skill Score (PSS) and standard error resulting from 50 iterations with results computed based on the training portions of the training sample including the generation of ROC curves based on the same data; the selected thresholds/models correspond to the maximum PSS along the respective ROC curve. The number of hidden layer neurons chosen is the least number of hidden layer neurons with a PSS standard error which overlaps the standard error corresponding to the maximum PSS. Thus, in this example, 150 hidden layer neurons corresponds to the maximum mean PSS = 0.694 ± 0.005 (within the red circle.) The performance of the TANN2 model with 125 hidden neurons, PSS = 0.689 ± 0.005, overlapped with the 150 hidden neuron confidence interval and hence was selected as the optimum number of hidden neurons, thus TANN 36-125-1.

Figure 2 depicts an example of how the optimal Y is chosen. The light red highlight identifies the number of hidden neurons leading to the largest mean PSS while the green highlight indicates the number of hidden neurons selected as the two cases standard errors overlap. **Table 4** depicts the optimal topologies for the TANN2 X-Y-1 and 36-Y-1 models.

With respect to FS, an exhaustive search involving the 36 potential features, although ideal, would have been computationally unrealistic. The FS methods used for this work were filter based, information theoretic, and designed to choose feature subsets relevant to the corresponding target while non-redundant to each feature in the subset. The methods used were multi-variate in the sense that feature-feature relationships were also considered, rather than the univariate strategy of assessing only feature-target relationships sequentially. The methods used are CFS (described earlier) and *minimum Redundancy Maximum Relevance* (mRMR) [74]. The *mRMR classic* function from the **mRMRe** package as part of the **R** programming language [75] was used to calculate mRMR. The following is an explanation of the mRMR technique in the context of mRMR classic as described in [76]: Consider $r(x, y)$ the correlation coefficient between features x and y. The mRMR technique uses the information-theoretic parameter known as *mutual information* (MI), defined as

$$1 - r(x,y)^2 \; MI(x,y) = \frac{-1}{2} ln \tag{8}$$

Let t be the target/predictand and $X = \{x_1, ..., x_n\}$ represent the set of n features. We desire to rank X such that we maximize relevance (maximize MI with t) and minimize redundancy (minimize mean MI with all previously selected features.) First, we selected x_i, the feature with the greatest MI with t:

$$x_i = \underset{x_i \in X}{arg \; max} MI(x_i, t) \tag{9}$$

Thus, we initialize the set of selected features S with x_i. For each subsequent step j, features are added to S by choosing the feature with the greatest relevance with t and lowest redundancy with previously selected features to maximize score Q:

$$Q_j = MI(x_j, t) - \frac{1}{|S|} \sum_{x_k \in S} MI(x_j, x_k) \tag{10}$$

The mRMR classic function requires the user to select the size of S, which we chose to equal the maximum number of features. We then choose the subset by choosing only those features corresponding to $Q_j > 0$.

Table 4 depicts the reduced set of features chosen via CFS and mRMR. **Table 5** summarizes the resulting TANN2 topologies.

Prediction hour	CFS	mRMR
9 h	PWAT,	$U_{800}(0)$, HI_{LOW}, CP, VV_{925}, VV_{500} RH_{850}, CIN, LCL, T_LCL, CAPE,
	CP	VSHEAR86, $V_{600}(0)$, PWAT, CTP_PROXY, VSHEAR75(0)
12 h	CP	CP, VV_{500} , VV_{925}, PWAT, RH_{850}, CTP_PROXY, CAPE, VSHEAR86,
		HI_{LOW}, $DROPOFF_{PROXY}$, $U_{800}(0)$, VV_{700}
15 h	CP	CP, LI, CAPE,VV_{925}, PWAT,
		RH_{850}, CTP_PROXY, VV_{700} $V_{500}(0)$, $U_{SFC}(0)$, VV_{500}, HI_{LOW}

Table 4. Variables/parameters chosen by the CFS and mRMR feature selection techniques; variables followed by zero depict NAM initialization variables.

Prediction hour	Optimal 36-Y-1 topology	Optimal X-Y-1 topology (CFS)	Optimal X-Y-1 topology (mRMR)
9	36-150-1	2-60-1	15-100-1
12	36-125-1	1-1-1	12-90-1
15	36-70-1	1-1-1	12-90-1

Table 5. Optimal TANN2 topologies based on iterative method to determine the optimal number of hidden layer neurons.

Forecast	Observed		
	Yes	No	Total
Yes	*a (hit)*	*b(false alarm)*	$a + b$
No	*c(miss)*	*d(correct rejection)*	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d = n$

Table 6. Contingency Matrix from which scalar performance metrics (**Table 7**) were derived (from [53]).

Performance metric (value range)	Symbol	Equation
Probability of detection [0,1]	POD	$a/(a + c)$
False alarm rate [0,1]	F	$b/(b + d)$
False alarm ratio [0,1]	FAR	$b/(a + b)$
Critical success index [0,1]	CSI	$a/(a + b + c)$
Peirce skill score [−1,1]	PSS	$(ad - bc)/(b + d)(a + c)$
Heidke skill score [−1,1]	HSS	$2(ad - bc)/[(a + c)(c + d) + (a + b)(b + d)]$
Yule's Q (odds ratio skill score) [−1,1]	ORSS	$(ad - bc)/(ad + bc)$
Clayton skill score [−1,1]	CSS	$(ad - bc)/(a + b)(c + d)$
Gilbert skill score [−1/3,1]	GSS	$(a - a_r)/(a + b + c - a_r)$; $a_r = (a + b)(a + c)/n$

Table 7. Relevant scalar performance metrics for binary classifiers used to evaluate TANN2 and MLR Models and NDFD (from [53]).

Tables 6 and **7** depict the contingency matrix and the corresponding performance metrics for binary classifiers used in this study.

Tables 8–11 depict the performance results of the TANN2 models, trained over all 286 boxes, and applied to boxes 73, 103, 238, and 1–286 (all boxes.) For each skill-based performance metric (PSS, CSI, HSS, ORSS, CSS, GSS), the *Wilcoxon Sign Rank* Test was used to determine whether TANN2 median performance (based on the 50 runs on the 2007–2008 testing set corresponding to the optimal number of hidden neurons) was statistically significantly different (5% level) than the corresponding MLR model and human forecaster performances. The relevant human forecasters were operational forecasters from the National Weather Service (NWS) (Appendix D). A summary of the results follow.

There is a significant improvement in the value of selected performance metrics for TANN2 over TANN in absolute terms. For example, with respect to the TANN models developed without FS, the PSS metric for the TANN2 models increased over the corresponding TANN models, by approximately 10–70%, 55–74%, and 10–120%, respectively, for boxes 238, 103, and 73.

When comparing TANN2 model performance relative to the operational forecasters (NDFD), and defining superior performance as statistically significant superior performance with respect to at least one skill-based performance metric (PSS, CSI, HSS, ORSS, CSS, and GSS), the results are as follows: For *box 238*, at least one of the TANN2 model performances exceeded that of the forecasters (NDFD), for all three prediction hours; all three TANN2 models (TANN 36-150-1, TANN 2-60-1 CFS, and TANN 15-100-1 mRMR) performed superior to NDFD for prediction hour 9, both TANN 36-150-1 and TANN 2-60-1 CFS performed better for prediction hour 12, and only TANN 2-6-1 CFS performed better for prediction hour 15. With respect to *box 103*, results were mixed. None of the TANN2 models performance was superior to NDFD for prediction hour 9, the TANN 36-125-1 and TANN 12-90-1 mRMR performed better for prediction hour 12, and only the TANN 1-1-1 CFS performed superior to the forecasters for prediction hour 15. Results were again mixed with regard to *box 73*. TANN 36-150-1 and TANN 15-100-1 mRMR performed superior to NDFD for prediction hour 9, none of the TANN2 models performed better than NDFD for prediction hour 12, and only TANN 36-70-1 performed superior to NDFD for prediction hour 15.

Conducting the same analysis with respect to TANN2 model compared to MLR, namely assessing statistically significant superior performance with respect to at least one skill-based performance metric, the results are as follows: For *box 238*, all three TANN2 models performed better than MLR at 9 h, none of the TANN2 models performed better at 12 h, and only the TANN 1-1-1 CFS model performed superior to MLR at 15 h. Regarding *box 103*, the TANN 36-125-1 and TANN 15-100-1 mRMR performed better than MLR for both 9 h and 12 h, while TANN 36-70-1 and TANN 1-1-1 CFS performed better than MLR for 15 h. For *box 73*, only TANN 36-150-1 performs better than MLR at 9 h, only TANN 1-1-1 CFS performs better at both 12 h and 15 h.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 36-150-1	0.94	0.80	0.30	0.63	0.20	0.24	0.94	0.19	0.13
TANN 2-60-1 CFS	0.98	0.81	0.33	0.65	0.19	0.22	0.98	0.19	0.13
TANN 15-100-1 mRMR	0.91	0.80	0.30	0.63	0.20	0.23	0.93	0.19	0.13
MLR	0.96	0.80	0.32	0.64	0.19	0.23	0.96	0.19	0.13
9 h Operational public forecasts									
NDFD	0.94	0.81	0.35	0.59	0.19	0.21	0.93	0.18	0.12
12 h Model predictions									
TANN 36-125-1	0.81	0.93	0.28	0.53	0.07	0.09	0.83	0.06	0.04
TANN 1-1-1 CFS	0.56	0.93	0.20	0.36	0.06	0.08	0.67	0.05	0.04
TANN 12-90-1 mRMR	0.75	0.94	0.32	0.42	0.05	0.06	0.71	0.05	0.03
MLR	0.88	0.93	0.30	0.57	0.07	0.09	0.88	0.07	0.05
12 h Operational public forecasts									
NDFD	0.67	0.92	0.28	0.39	0.07	0.08	0.67	0.06	0.04
15 h Model predictions									
TANN 36-70-1	0.64	0.95	0.23	0.41	0.05	0.06	0.71	0.04	0.03
TANN 1-1-1 CFS	0.45	0.91	0.08	0.37	0.08	0.13	0.81	0.08	0.07
TANN 12-90-1 mRMR	0.64	0.96	0.28	0.36	0.04	0.04	0.63	0.03	0.02
MLR	0.73	0.95	0.27	0.46	0.05	0.06	0.76	0.04	0.03
15 h Operational public forecasts									
NDFD	0.92	0.92	0.23	0.69	0.08	0.11	0.95	0.07	0.06

Values corresponding to each TANN X-Y-Z model represent the median of 50 separate trial runs of the model. Yellow (blue) denotes TANN X-Y-Z median values of skill-based metrics (PSS, CSI, HSS, ORSS, CSS, and GSS only) NOT statistically significantly different (based on the Wilcoxon Sign Rank Tests, 2 sided, 1 sample, 5% significant level) from the NDFD (MLR) values.

Table 8. Performance results of TANN X-Y-Z models for box 238 for the 2007–2008 independent data set and corresponding comparisons to the WFO CRP forecasters (NDFD), multi-linear regression (MLR) models.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 36-150-1	0.93	0.87	0.31	0.62	0.13	0.15	0.93	0.12	0.08
TANN 2-60-1 CFS	0.90	0.88	0.33	0.57	0.12	0.14	0.89	0.11	0.07
TANN 15-100-1 mRMR	0.87	0.87	0.29	0.56	0.13	0.15	0.87	0.12	0.08
MLR	0.97	0.88	0.36	0.61	0.12	0.14	0.96	0.12	0.08
9 h Operational public forecasts									
NDFD	1.00	0.85	0.31	0.69	0.15	0.19	1.00	0.15	0.10

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
12 h Model predictions									
TANN 36-125-1	0.80	0.95	0.23	0.58	0.05	0.07	0.87	0.05	0.04
TANN 1-1-1 CFS	0.50	0.96	0.18	0.32	0.04	0.05	0.64	0.03	0.03
TANN 12-90-1 mRMR	0.90	0.95	0.27	0.63	0.05	0.07	0.92	0.05	0.04
MLR	0.80	0.95	0.25	0.55	0.05	0.06	0.85	0.05	0.03
12 h Operational public forecasts									
NDFD	0.80	0.94	0.24	0.56	0.06	0.08	0.86	0.06	0.04
15 h Model predictions									
TANN 36-70-1	0.83	0.96	0.21	0.62	0.04	0.05	0.90	0.03	0.03
TANN 1-1-1 CFS	0.50	0.92	0.06	0.44	0.07	0.12	0.89	0.07	0.06
TANN 12-90-1 mRMR	0.83	0.97	0.29	0.55	0.03	0.04	0.86	0.03	0.02
MLR	0.83	0.97	0.24	0.60	0.03	0.05	0.88	0.03	0.02
15 h Operational public forecasts									
NDFD	1.00	0.97	0.19	0.81	0.04	0.07	1.00	0.04	0.04

Values corresponding to each TANN X–Y–Z model represent the median of 50 separate trial runs of the model. Yellow (blue) denote TANN X–Y–Z median values of skill-based metrics (PSS, CSI, HSS, ORSS, CSS, and GSS only) NOT statistically significantly different (based on the Wilcoxon Sign Rank Tests, 2 sided, 1 sample, 5% significant level) from the corresponding NDFD (MLR).

Table 9. Performance results of TANN X-Y-Z models for box 103 for the 2007–2008 independent data set and corresponding comparisons to the WFO CRP forecasters (NDFD), multi-linear regression (MLR) models.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 36-150-1	0.94	0.81	0.22	0.71	0.19	0.25	0.96	0.19	0.14
TANN 2-60-1 CFS	0.94	0.85	0.30	0.64	0.14	0.18	0.95	0.14	0.10
TANN 15-100-1 mRMR	0.97	0.83	0.26	0.70	0.17	0.22	0.98	0.17	0.12
MLR	0.97	0.82	0.25	0.72	0.18	0.23	0.98	0.17	0.13
9 h Operational public forecasts									
NDFD	0.91	0.83	0.26	0.65	0.16	0.21	0.93	0.16	0.12
12 h Model predictions									
TANN 36-125-1	0.86	0.90	0.29	0.57	0.10	0.12	0.88	0.09	0.07
TANN 1-1-1 CFS	0.77	0.88	0.21	0.56	0.11	0.15	0.85	0.11	0.08
TANN 12-90-1 mRMR	0.91	0.91	0.35	0.56	0.08	0.10	0.90	0.08	0.05
MLR				0.68	0.10	0.13	1.00	0.10	0.07

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
12 h Operational public forecasts									
NDFD	0.91	0.86	0.23	0.68	0.14	0.19	0.94	0.14	0.11
15 h Model predictions									
TANN 36-70-1	0.92	0.93	0.25	0.68	0.07	0.10	0.95	0.07	0.05
TANN 1-1-1 CFS	0.33	0.96	0.15	0.18	0.04	0.04	0.47	0.03	0.02
TANN 12-90-1 mRMR	0.83	0.96	0.34	0.47	0.04	0.05	0.79	0.04	0.02
MLR	0.92	0.94	0.30	0.62	0.06	0.07	0.93	0.05	0.04
15 h Operational public forecasts									
NDFD	0.85	0.93	0.24	0.61	0.07	0.10	0.89	0.07	0.05

Values corresponding to each TANN X-Y-Z model represent the median of 50 separate trial runs of the model. Yellow (blue) denote TANN X-Y-Z median values of skill-based metrics (PSS, CSI, HSS, ORSS, CSS, and GSS only) NOT statistically significantly different (based on the Wilcoxon Sign Rank Tests, 2 sided, 1 sample, 5% significant level) from the corresponding NDFD (MLR) value.

Table 10. Performance results of TANN X–Y–Z models for box 73 for the 2007–2008 independent data set and corresponding comparisons to the WFO CRP forecasters (NDFD), multi-linear regression (MLR) models.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 36-150-1	0.92	0.86	0.24	0.68	0.14	0.18	0.95	0.13	0.10
TANN 2-60-1 CFS	0.93	0.89	0.31	0.62	0.11	0.14	0.93	0.11	0.07
TANN 15-100-1 mRMR	0.93	0.87	0.27	0.66	0.13	0.17	0.94	0.12	0.09
MLR	0.92	0.87	0.27	0.65	0.13	0.16	0.94	0.12	0.09
12 h Model predictions									
TANN 36-125-1	0.85	0.93	0.26	0.59	0.07	0.10	0.89	0.07	0.05
TANN 1-1-1 CFS	0.68	0.92	0.19	0.49	0.07	0.10	0.80	0.07	0.05
TANN 12-90-1 mRMR	0.88	0.94	0.31	0.58	0.06	0.08	0.89	0.06	0.04
MLR	0.88	0.93	0.27	0.61	0.07	0.09	0.90	0.07	0.05
15 h Model predictions									
TANN 36-70-1	0.83	0.94	0.24	0.59	0.06	0.08	0.88	0.06	0.04
TANN 1-1-1 CFS	0.53	0.92	0.11	0.42	0.08	0.11	0.80	0.07	0.06
TANN 12-90-1 mRMR	0.81	0.95	0.30	0.51	0.05	0.06	0.82	0.04	0.03
MLR	0.88	0.95	0.28	0.59	0.05	0.07	0.89	0.05	0.04

Table 11. Performance results of TANN X–Y–Z models for all 286 boxes for the 2007–2008 independent data set and corresponding comparisons to the NWS forecasters (NDFD), multi-linear regression (MLR) models.

	9-h Prediction	12-h Prediction	15-h Prediction
ORSS			
Box 73		NDFD	TANN 36-70-1
Box 103	NDFD		NDFD
Box 238	TANN 2-60-1	TANN 36-125-1	NDFD
PSS			
Box 73	TANN 36-150-1	NDFD	TANN 36-70-1
Box 103	NDFD		NDFD
Box 238	TANN 2-60-1	TANN 36-125-1	NDFD
HSS			
Box 73	36-150-1	NDFD	NDFD
Box 103	NDFD	NDFD	TANN 1-1-1
Box 238			TANN 1-1-1

Gray boxes indicate that a single performer did not distinguish itself.

Table 12. Best performers (based on the equitable ORSS, PSS, and HSS performance metrics) between the TANN2 models and NWS forecasters (NDFD) using the combination of Wilcoxon Sign Rank Tests (2 sided, 1 sample, 5% significant level) to compare each TANN model to NDFD, and the Nemenyi post-hoc analyses of pairwise combination of TANN2 models (5% significance level).

An alternative analysis was performed to determine the single best-performing classifiers for each box and prediction hour based only on performance metrics PSS, HSS, and ORSS. HSS and PSS are truly equitable and ORSS is asymptotically equitable (approaches equitability as size of the data set approaches infinity) and truly equitable for the condition $\frac{(a+c)}{n} = 0.5$ (see contingency matrix.) [77]. Equitability includes the desirable condition whereby the performance metric is zero for random or constant forecasts. For each of the three performance metrics and for each box and prediction hour combination, comparisons were made between the TANN2 models and NDFD. Such a comparison was performed separately between the TANN2 models and MLR. The best performers were determined in the following manner: First, the output from the *Wilcoxon sign rank test* was used to determine statistically significant differences between the TANN2 models and NDFD or MLR (**Tables 8–10**). Next, the *Friedman rank sum test* was used to determine whether significant differences existed only between the three TANN2 models. If differences existed, the *Nemenyi post-hoc test* [78] was performed to determine statistical significance between pairwise combinations of TANN2 models. The best performer was based on a synthesis of the Wilcoxon and Friedman/Nemenyi results (see Appendix E). The Friedman and Nemenyi post-hoc tests were performed using the *friedman.test* and *posthoc.friedman.nemenyi.test* functions from the Pairwise Multiple Comparison of Mean Rank (*PMCMR*) package in the **R** programming language [74]. Results are depicted in **Tables 12** and **13**. The results are mixed. With respect to comparisons between the TANN2 models and NDFD, the best performer is a function of the performance metric, box number,

and prediction hour. The same is true for the TANN2 model-MLR comparisons. However, it is noteworthy to mention that none of the TANN2 models based on the mRMR FS method were determined to be the single best performer for 15 h (**Tables 12** and **13**).

	9-h Prediction	12-h Prediction	15-h Prediction
ORSS			
Box 73	MLR	MLR	TANN 36-70-1
Box 103	MLR		
Box 238	TANN 2-60-1	MLR	TANN 1-1-1
PSS			
Box 73		MLR	TANN 36-70-1
Box 103			
Box 238	TANN 2-60-1	MLR	MLR
HSS			
Box 73	TANN 36-150-1	TAN 1-1-1	TAN 36-70-1
Box 103			TANN 1-1-1
Box 238			TANN 1-1-1

Gray boxes indicate that a single performer did not distinguish itself.

Table 13. Best performers (based on the equitable ORSS, PSS, and HSS performance metrics) between the TANN2 models and MLR using the combination of Wilcoxon Sign Rank Tests (2 sided, 1 sample, 5% significant level) to compare each TANN model to MLR, and the Nemenyi post-hoc analyses of pairwise combination of TANN2 models (5% significance level).

The increase in data as compared to the previous study [53] likely contributed to performance enhancements. **Figure 3** depicts the change in performance of the 36-Y-1 12 h model as a function of training data used. Note that performance improvement was positively correlated with the quantity of training data. This adds credence to the argument that the amount of data in [53] may have been insufficient. Further, note that for ≤1% of available data (~6000 instances), performance decreased after the number of neurons in the hidden layer (Y) exceeded 100, possibly due to the curse of dimensionality. **Figure 4** depicts the relationship between an overfitting metric, defined as $\left(AUC_{val} - AUC_{train}\right)/AUC_{train}$, and the number of hidden layer neurons. Note that overfitting generally increases with the number of hidden layer neurons when the training set is reduced to ≤ 10 of the total available training data, and particularly apparent when only 1 of available, or about 6000 cases. In [53], only a few thousand cases were available. This partially explains why the original TANN models are relatively smaller ($Y \leq 35$) than the TANN2 models. The ability to obtain similar performance while training on a smaller portion of the data set would have allowed substantial gains in computational efficiency. ANNs were trained using MATLAB® R2015b on several multi-core personal computers (PCs) and on a computer cluster with compute nodes with two Xeon E5 with 10 core processors each and

256 gigabytes (GB) of memory. For the full data set (**Table 1**), training times varied from less than 1 h to more than two days per batch of 50 ANNs when increasing the number of hidden layer neurons from 1 to 200. When reducing the size of the training set to 1%, training times decreased to less than 3 min for each Y case (not shown.) evaluated. Reducing the size of the training set to 10% of the full set decreased training times by about one order of magnitude for each Y case.
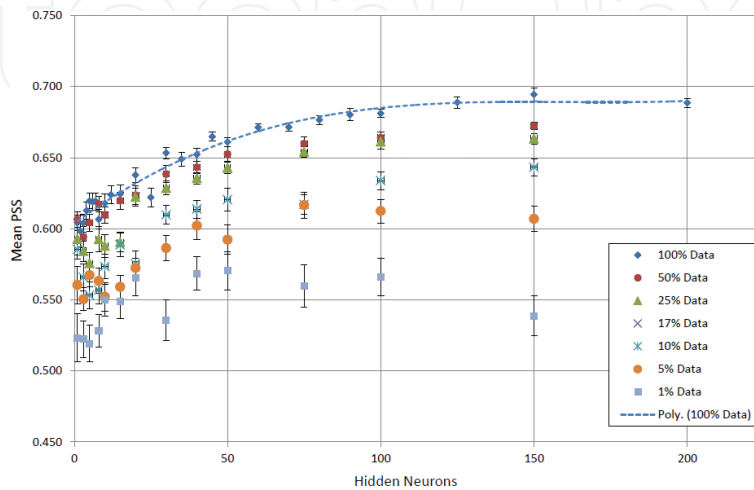

Figure 3. 12 h 36-Y-1 TANN performance versus hidden layer neuron quantity (*Y*) as a function of the percentage of total training data available.
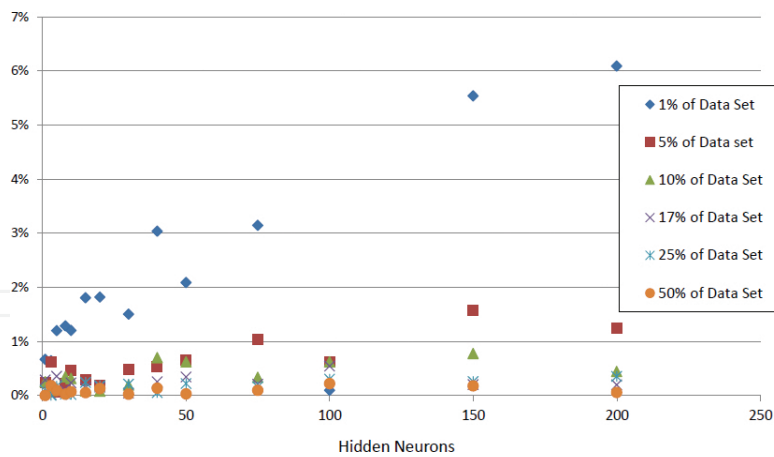
Figure 4. 12 h 36-Y-1 TANN overfitting index versus hidden layer neuron quantity (see text).

With regard to data set size and composition, it was hypothesized that performance gains may be obtained when artificially increasing the proportion of positive targets (CTG lightning strikes) in the training set. All possible inputs were included and the total number of training cases was maintained constant with positive and negative cases randomly selected (with replacement) to create target vectors with 5, 10, 25, and 50% of CTG lightning strikes. Substantial increases in performance were obtained for the training sets for all prediction lead

times; **Figure 5** depicts the 12 h prediction example. Maximum PSS increased progressively while increasing proportion of lightning strikes in the data set. However, as the percent of positive targets was raised, the performance over the 2007–2008 independent testing decreased. Efforts are continuing to further modify the training of the TANN to improve performance.
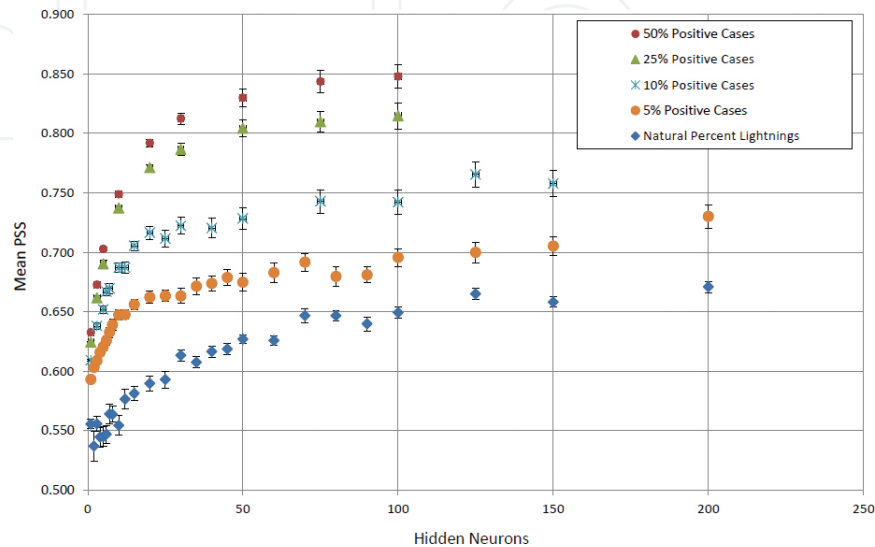

Figure 5. 12 h 36-Y-1 TANN performance versus hidden neuron quantity (*Y*) as a function of the proportion of positive target (CG lightning) data in the training set.

7. Conclusion

We presented here the results of an ANN approach to the post processing of single deterministic NWP model output for the prediction of thunderstorms at high spatial resolution, 9, 12, and 15 h in advance (TANN2.) ANNs were selected to take advantage of a large data set of over 1 million cases with multiple predictors and attempt to capture the complex relationships between the predictors and the generation of thunderstorms. This study represents an adjustment to a previous ANN model framework, resulting in the generation of a significantly larger data set. The larger data set allowed for more complex ANN models (by increasing the number of neurons in the hidden layer). Three groups of TANN2 model variants were generated based on two filtering-based feature selection methods (designed to retain only relevant and non-redundant features) and one group based on models calibrated with all predictors.

The skills of these TANN2 models within each of the three 400-km² boxes were substantially improved over previous work with the improvements attributed to the increase of the size of the data set. TANN2 model performance was compared to that of NWS operational forecasters and to MLR models. Results regarding the best-performing classifiers per prediction hour and box were mixed. Several attempts were made to further improve model performance or

decrease training time. Training the models using a small fraction of the data set reduced model calibration time yet resulted in lower performance skill. Altering the target by artificially boosting the proportion of positive outcomes (lightning strikes) resulted in substantial performance improvements over the training sets but did not lead to substantial improvements of performance on the independent 2007–2008 cases.

Given that the atmosphere is chaotic, or deterministic with highly sensitive dependence on the initial condition, one future research plan includes the prediction of thunderstorms by the post-processing of single deterministic NWP model output using ANN models that account for chaotic systems [79]. Such a strategy would be an alternative to the state of the art practice of using NWP model ensembles to account for the sensitive dependence on the initial condition. In addition, another plan involves the development of ensemble ANN models [80]. Specifically, an optimal TANN prediction can be developed by integrating output from 50 unique TANN models.

Appendix A

The total mean mass of the atmosphere: 5.1480×10^{21}g [81]

Total mol of dry air: 5.1480×10^{21}g $\times \dfrac{1}{28.97\text{gmol}^{-1}} = 1.777 \times 10^{20}$mol

Total molecules of dry air: 1.777×10^{20}mol $\times 6.02214 \times 10^{23}mol^{-1} = 1.0701 \times 10^{44}$molecules

(Mass of dry air: 28.97gmol^{-1}; Avogadro's number: 6.02214×10^{23}mol^{-1})

Appendix B

Version 1.04 of the HRRR uses the Advanced Research WRF (ARW) dynamic core within the WRF modeling framework (version 3.4.1 WRF-ARW). The HRRR uses GSI 3D-VAR data assimilation. With respect to parameterizations, the RRTM longwave radiation, Goddard shortwave radiation, Thompson microphysics (version 3.4.1), no cumulus/convective parameterization, MYNN planetary boundary layer, and the rapid update cycle (RUC) land surface model [33].

Appendix C

At any given time, only one NWP model was utilized in the TANN. Yet, three different modeling systems were used, each during a unique time period—the hydrostatic Eta model [82] (1 March 2004 to 19 June 2006), the Weather Research and Forecasting Non-hydrostatic

Mesoscale Model (WRF-NMM) [83] (20 June 2006 to 30 September 2011), and the NOAA Environmental Modeling System Non-hydrostatic Multiscale Model (NEMS-NMMB) (October 2011 to December 2013.)

Appendix D

NWS operational forecasts were obtained from the NWS *National Digital Forecast Database* [96] (NDFD) [96], a database of archived NWS forecasts; the forecasts are written to a 5-km coterminous USA (CONUS) grid (or to 16 pre-defined grid sub-sectors) and provided to the general public in Gridded Binary Version 2 (GRIB2) format [97]. The forecasts for most of the 286 boxes (**Figure 1**) originated from the NWS Weather Forecast Office (WFO) in Corpus Christi, Texas (CRP) in the USA.

Appendix E

The following are three examples to explain how the "best performers" where determined in **Tables 12** and **13**.

Example 1: Table 12 (Determining the best performers between the three TANN model variants and NDFD) Box 238 Prediction Hour 9 ORSS: Step 1: Based on the Wilcoxon Sign Rank Test, the performer with the largest ORSS value was TANN 2-60-1, and that value (0.98) was statistically significantly larger than the corresponding NDFD value (0.93) (**Table 8**). Step 2: Based on the Nemenyi post-hoc analysis of the pairwise combination of the three TANN model variants, the TANN 2-60-1 ORSS value was statistically significantly different than the corresponding ORSS values from the TANN 36-150-1 and TANN 15-100-1 models. *Thus, the best performer is TANN 2-60-1.*

Example 2: Table 12 (Determining the best performers between the three TANN model variants and NDFD) Box 103 Prediction Hour 12 PSS: Step 1: Based on the Wilcoxon Sign Rank test, the performer with the largest PSS value was TANN 12-90-1, and that value (0.63) was statistically significantly larger than the corresponding NDFD value (0.56) (**Table 9**). Step 2: Based on the Nemenyi post-hoc analysis of the pairwise combination of the three TANN model variants, there was no statistically significant difference between the PSS values for TANN 12-90-1 (0.63) and TANN 36-125-1 (0.58). *Thus, there is no single best performer.*

Example 3: Table 13 (Determining the best performers between the three TANN model variants and MLR) Box 73 Prediction Hour 12 PSS: Step 1: Based on the Wilcoxon Sign Rank Test, the performer with the largest PSS value was MLR, and that value (0.68) was statistically significantly greater than the corresponding PSS values from each of the three TANN variants (**Table 10.**) *Thus, the best performer is MLR. No additional steps are required.*

Author details

Waylon G. Collins[1*] and Philippe Tissot[2]

*Address all correspondence to: Waylon.Collins@noaa.gov

1 National Weather Service, Weather Forecast Office Corpus Christi, TX, USA

2 Texas A&M University—Corpus Christi, Conrad Blucher Institute, Corpus Christi, TX, USA

References

[1] Glickman TS. Glossary of Meteorology. 2nd ed. Boston: American Meteorological Society; 2011. 1132 p.

[2] Byers HR. General Meteorology, 3rd ed. New York: McGraw Hill Book Company; 1959. 540 p.

[3] Emanuel KA. Atmospheric Convection. New York: Oxford University Press; 1994. 580 p.

[4] Doswell CA, editor. Severe Convective Storms (Meteorological Monographs, Volume 28, Number 50). Boston: American Meteorological Society; 2001. 561 p.

[5] Groenemeijer, P. Sounding-derived parameters associated with severe convective storms in the Netherlands [thesis]. Utrecht: Utrecht University; 2005.

[6] Holle RL. Annual rates of lightning fatalities by country. In: 20th International Lightning Detection Conference; 21–23 April 2008; Tucson.

[7] Holle RL, Lopez RE. A comparison of current lightning death rates in the U.S. with other locations and times. In: International Conference on Lightning and Static Electricity; 2003; Blackpool. p. 103–134.

[8] NWS. Natural Hazards Statistics National Weather Service Office of Climate, Water, and Weather Services [Internet]. 2015. Available from: http://www.nws.noaa.gov/om/hazstats.shtml [Accessed 2015-10-26].

[9] NOAA. Billion-Dollar Weather and Climate Disasters: Summary Stats, National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information [Internet]. 2016. Available from: https://www.ncdc.noaa.gov/billions/summary-stats [Accessed 2016-01-28].

[10] Wolfson MM, Clark DA. Advanced Aviation Weather Forecasts. Lincoln Lab. J. 2006; 16: 31–58.

[11] Trier SB. Convective storms – Convective Initiation. In: Holton Jr, editor. Encyclopedia of Atmospheric Sciences. New York: Oxford University Press; 2003. p. 560–570.

[12] Orlanski I. A Rational Subdivision of Scales for Atmospheric Processes. Bull. Am. Meteorol. Soc. 1975; 56: 527–530.

[13] Lorenz EN. Deterministic nonperiodic flow. J. Atmos. Sci. 1963; 20: 130–141.

[14] Lorenz EN. The predictability of a flow which possesses many scales of motion. Tellus. 1969; 3: 1–19.

[15] Wallace JM, Hobbs PV. Atmospheric Science: An Introductory Survey. New York: Academic Press; 1977. 466 p.

[16] Kain JS, Coniglio MC, Correia J, Clark AJ, Marsh PT, Ziegler CL, Lakshmanan V, Miller SD, Dembek SR, Weiss SJ, Kong F, Xue M, Sobash RA, Dean AR, Jirak IL, CJ Melick. A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. Bull. Amer. Meteor. Soc. 2013; 94: 1213–1225.

[17] Lamb D. Rain Production in Convective Storms. In: Doswell CA, editor. Severe Convective Storms. Boston: American Meteorological Society; 2001. p. 299–321.

[18] Williams ER. The Electrification of Severe Storms. In: Doswell CA, editor. Severe Convective Storms. Boston: American Meteorological Society; 2001. p. 527–561.

[19] Wakimoto RM. Convectively Driven High Wind Events. In: Doswell CA, editor. Severe Convective Storms. Boston: American Meteorological Society; 2001. p. 255–298.

[20] Knight CA, Knight NC. Hailstorms. In: Doswell CA, editor. Severe Convective Storms. Boston: American Meteorological Society; 2001. p. 223–254.

[21] Taylor CM, Gounou A, Guichard F, Harris PP, Ellis RJ, Couvreux F. Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns. Nat. Geosci. 2011; 4: 430–433.

[22] Doswell CA, Bosart LF. 2001. Extratropical synoptic-scale processes and severe convection. In: Doswell CA, editor. Severe Convective Storms. Boston: American Meteorological Society; 2001. p. 27–70.

[23] Lorenz, EN. The Essence of Chaos. Seattle: University of Washington Press; 1993. 227 p.

[24] Bjerknes V. Das Problem der Wettervorhersage, betrachtet vom Stanpunkt der Mechanik und der Physik. Meteor. Zeits. 1904; 21: 1–7.

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

[26] Stull R. Meteorology for Scientists and Engineers. 3rd ed. Pacific Grove: Brooks Cole; 2015.

[27] Stensrud DJ. Parameterization Schemes, Keys to Understanding Numerical Weather Prediction Models. New York: Cambridge University Press; 2007. 459 p.

[28] Wilks DS. Statistical Methods in the Atmospheric Sciences. 2nd ed. Oxford: Elsevier; 2006.

[29] Bryan GH, Wyngaard JC, Fritsch JM. Resolution Requirements for the Simulation of Deep Moist Convection. Mon. Wea. Rev. 2003; 131: 2394–2416.

[30] Hodanish S, Holle RL, Lindsey DT. A Small Updraft Producing a Fatal Lightning Flash. Wea. Forecasting. 2004; 19: 627–632.

[31] Mecikalski JR., Williams JK, Jewett CP, Ahijevych D, LeRoy A, Walker JR. Probabilistic 0–1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. J. Appl. Meteor. Climatol. 2015; 54: 1039–1059.

[32] Pessi AT, Businger S. Relationships among lightning, precipitation, and hydrometeor characteristic over the North Pacific Ocean. J. Appl. Meteorol. Climatol. 2009; 48: 833–848.

[33] Earth System Research Laboratory. The High-Resolution Rapid Refresh [Internet]. 2016. Available from: http://rapidrefresh.noaa.gov/hrrr [Accessed: 2016-01-14]

[34] Pielke RA. Mesoscale Meteorological Modeling. International Geophysics Series. San Diego: Academic Press; 2002. 676 p.

[35] Zhang Y, Zhang R, Stensrud DJ, Zhiyong M. Intrinsic Predictability of the 20 May 2013 Tornadic Thunderstorm Event in Oklahoma at Storm Scales. Mon. Wea. Rev. 2016; 144: 1273–1298

[36] Wilson JW, Crook NA, Mueller CK, Sun J, Dixon M. Nowcasting Thunderstorms: A Status Report. Bull. Am. Meteorol. Soc. 1998; 79: 2079–2099.

[37] Fowle MA, Roebber PJ. Short-range (0–48 h) Numerical predictions of convective occurrence, model and location. Wea. Forecasting. 2003; 18: 782–794.

[38] Weisman ML, Davis C, Wang W, Manning KW, Klemp JB. Experiences with 0-36-h explicit convective forecasts with the WRF-ARW model. Wea. Forecasting. 2008; 23: 407–437.

[39] Elmore KL, Stensrud DJ, Crawford KC. Explicit cloud-scale models for operational forecasts: a note of caution. Wea. Forecasting. 2002; 17: 873–884.

[40] Leith C.E. Theoretical skill of Monte Carlo forecasts. Mon. Wea. Rev. 1974; 102: 409–418.

[41] Epstein ES. Stochastic dynamic prediction. Tellus. 1969; 21: 739–759.

[42] Storm Prediction Center. Short Range Ensemble Forecast (SREF) Products [Internet]. 2015. Available from: URL http://www.spc.noaa.gov/exper/sref [Accessed: 2015-10-23].

[43] Meteorological Development Laboratory. Model Output Statistics (MOS) [Internet.]. 2016. Available from http://www.weather.gov/mdl/mos_home [Accessed: 2016-03-01].

[44] Glahn HR, Lowry DA. The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteorol. 1972; 11: 1203–1211.

[45] Glahn B, Gilbert K, Cosgrove R, Ruth DP, Sheets K. The gridding of MOS. Wea. Forecast. 2009; 24: 520–529.

[46] Ghirardelli JE, Glahn B. The Meteorological Development Laboratory's aviation weather prediction system. Wea. Forecast. 2010; 25: 1027–1051.

[47] Schmeits MJ, Kok KJ, Vogelezang DHP. Probabilistic forecasting of (severe) thunderstorms in the Netherlands using model output statistics. Wea. Forecast. 2005; 20: 134–148.

[48] Costello RB. Random House Webster's College Dictionary. New York: Random House; 1992.

[49] Mitchell, T. The discipline of machine learning. Carnegie Mellon University Technical Report, CMU-ML-06-108, 2006.

[50] Mills GA, Colquhoun JR. Objective prediction of severe thunderstorm environments: preliminary results linking a decision tree with an operational regional NWP model. Wea. Forecast. 1998; 13: 1078–1092.

[51] Perler D, Marchand O. A study in weather model output postprocessing: using the boosting method for thunderstorm detection. Wea. Forecast. 2009; 24: 211–222.

[52] McCann DW. A neural network short-term forecast of significant thunderstorms. Wea. Forecast. 1992; 7: 525–534.

[53] Collins W, Tissot P. An artificial neural network model to predict thunderstorms within 400km^2 South Texas domains Meteorol. Appl. 2015; 22: 650–665.

[54] Breiman L. Random forests. Machine Learn. 2001; 45: 5–32.

[55] McNulty RP. A statistical approach to short-term thunderstorm outlooks. J. Appl. Meteorol. 1981; 20: 765–771.

[56] Ravi N. Forecasting of thunderstorms in the pre-monsoon season at Delhi. Meteorol. Appl. 1999; 6: 29–38.

[57] de Silva CW. Intelligent Machines: Myths and Realities. Boca Raton: CRC Press LLC; 2000.

[58] Colquhoun JR. A decision true method of forecasting thunderstorms serve a thunderstorms and tornados. Wea. Forecast. 1987; 2: 337–345.

[59] Lee RR, Passner JE. The development and verification of TIPS: an expert system to forecast thunderstorm occurrence. Wea. Forecast. 1993; 8: 271–280.

[60] Manzato A. The use of sounding-derived indices for a neural network short-term thunderstorm forecast. Wea. Forecast. 2005; 20: 896–917.

[61] Chaudhuri S. Convective energies in forecasting severe thunderstorms with one hidden layer neural net and variable learning rate back propagation algorithm. Asia Pac. J. Atmos. Sci. 2010; 46: 173–183.

[62] Sanchez JL, Ortega EG, Marcos JL. Construction and assessment of a logistic regression model applied to short-term forecasting of thunderstorms in Leon (Spain). Atmos. Res. 2001; 56: 57–71.

[63] Rumelhart DE, McClelland JL. Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations. Vol 1. Cambridge: MIT Press; 1986.

[64] Haykin S. Neural Networks, A Comprehensive Foundation. 2nd ed. New Jersey: Prentice Hall; 1999.

[65] Bishop CM. Neural Networks for Pattern Recognition. New York: Oxford University Press; 2005.

[66] Moller M. A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. 1993; 6: 525–533.

[67] Manzato A. A note on the maximum peirce skill score. Wea. Forecast. 2007; 22: 1148–1154.

[68] Bellman R. Adaptive Control Processes: A Guided Tour. Princeton: Princeton University Press; 1961.

[69] May R., Dandy G, Maier H. Review of input variable selection methods for artificial neural networks. In: Suzuki K, editor. Artificial Neural Networks- Methodological Advances and Biomedical Applications. Rijeka: InTech Europe; 2011. P. 19–44.

[70] Hall MA. Correlation-based feature selection for machine learning [thesis]. Hamilton: University of Waikato; 1999.

[71] Hall MA, Smith LA. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference; 1–5 May 1999; Orlando.

[72] Arizona State University. Feature Selection at Arizona State University [Internet]. 2016. Available from: http://featureselection.asu.edu/index.php# [Accessed: 2016-02-16].

[73] Akaike H. Information theory and an extension of the maximum likelihood principle. In: 2nd International Symposium on Information Theory; 1973; Budapest: Akademiai Kiado. p. 267–281.

[74] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 2005; 3: 185–205.

[75] A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [Internet]. Available from: https://www.r-project.org [Accessed: 2005-12-14].

[76] De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains. mRMRe: An R package for parallelized mrmr ensemble feature selection. Bioinformatics. 2013; 29: 2365–2368.

[77] Hogan RJ, Ferro CAT, Jolliffe IT, Stephenson DB. Equitability revisited: why the "equitable threat score" is not equitable. Wea. Forecast. 2010; 25: 710–726.

[78] Nemenyi P. Distribution-free multiple comparisons [thesis]. Princeton: Princeton University; 1963.

[79] Pan S-T, Lai C-C. Identification of chaotic systems by neural network with hybrid learning algorithm. Chaos Soliton Fract. 2008; 37: 233–244.

[80] Maqsood I, Khan MR, Abraham A. An ensemble of neural networks for weather forecasting. Neural Comput. Appl. 2004; 13: 112–122.

[81] Trenberth KE, Smith L. The Mass of the Atmosphere: A Constraint on Global Analyses. J. Climate. 2005; 18: 864–875.

[82] Rogers E, Black TL, Deaven DG, DiMego GJ. Changes to the operational "early" eta analysis/forecast system at the National Centers for Environmental Prediction. Wea. Forecast. 1996; 11: 391–413.

[83] Janjic ZI, Gerrity JP, Nickovic S. An alternative approach to nonhydrostatic modeling. Mon. Wea. Rev. 2001; 129: 1164–1178.

[84] Khairoutdinov M, Randall D. High-resolution simulation of shallow-to-deep convection transition over land. J. Atmos. Sci. 2006; 63: 3421–3436.

[85] Ducrocq V, Tzanos D, Sénési S. Diagnostic tools using a mesoscale NWP model for the early warning of convection. Meteorol. Appl. 1998; 5: 329–349.

[86] Haklander AJ, Van Delden A. Thunderstorm predictors and their forecast skill for the Netherlands. Atmos. Res. 2003; 67–68: 273–299.

[87] Dalu GA, Pielke RA, Baldi M, Zeng X. Heat and momentum fluxes induced by thermal inhomogeneities with and without large-scale flow. J. Atmos. Sci. 1996; 53: 3286–3302.

[88] Wang JR, Bras L, Eltahir EAB. A stochastic linear theory of mesoscale circulation induced by thermal heterogeneity of the land surface. J. Atmos. Sci. 1996; 53: 3349–3366.

[89] Crook NA. Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. Mon. Wea. Rev. 1996; 124: 1767–1785.

[90] Williams ER, Mushtak V, Rosenfeld D, Goodman S, Boccippio D. Thermodynamic conditions favorable to superlative thunderstorm updraft, mixed phase microphysics and lightning flash rate. Atmos. Res. 2005; 76: 288–306.

[91] Saunders CPR. A review of thunderstorm electrification processes. J. Appl. Meteorol. 1993; 32: 642–655.

[92] Janjić ZI. The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. Mon. Weather Rev. 1994; 122: 927–945.

[93] Rotunno RJ, Klemp B, Weisman ML. A theory for strong, long-lived squall lines. J. Atmos. Sci. 1998; 45: 464–485.

[94] Findell KL, Eltahir EAB. Atmospheric controls on soil moisture-boundary layer interactions: three-dimensional wind effects. J. Geophys. Res. 2003b; 108 (D8), 8385.

[95] Findell KL, Eltahir EAB. Atmospheric controls on soil moisture-boundary layer interactions: Part I: Framework development. J. Hydrometeorol. 2003a; 4: 552–569.

[96] National Weather Service National Digital Forecast Database [Internet]. 2016. Available from: http://www.nws.noaa.gov/ndfd/ [Accessed: 2016-03-23]

[97] World Meteorological Organization. WMO Manual on Codes, WMO Publication No. 306, Vol. 1, Part B. Geneva: WMO; 2001.