METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **22**: 650–665 (2015)
Published online 28 April 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1499

RMetS
Royal Meteorological Society

An artificial neural network model to predict thunderstorms within 400 km^2 South Texas domains

Waylon Collins[a,*] and Philippe Tissot[b]
[a] *National Weather Service, Corpus Christi, TX, USA*
[b] *Conrad Blucher Institute for Surveying and Science, Texas A&M University, Corpus Christi, TX, USA*

ABSTRACT: Artificial neural network (ANN) models were developed to predict thunderstorm occurrence within three separate 400 km^2 regions, 9, 12 and 15 h (±2 h) in advance. The predictors include output from deterministic Numerical Weather Prediction models and from sub-grid scale soil moisture magnitude and heterogeneity estimates. The feed-forward multi-layer perceptron ANN topology, with one hidden layer and one neuron in the output layer, was chosen. Two sets of nine ANN models each were developed; one set was developed after a filtering-based feature selection technique was used to determine the predictor subset from 43 potential predictors. The other models were developed based on all 43 predictors. For each of the 18 models, a wrapper technique was used to determine the optimal number of neurons in the hidden layer. Thunderstorm artificial neural network (TANN) model performance was compared to that of multi-linear regression (MLR) models, and to human forecasters (NDFD), based on a novel data set. Results reveal that for several of the nine box/prediction hour combinations with respect to at least one skill-based performance metric, the TANN model's performance exceeded that of the MLR models and NDFD. Yet, the performance of both the MLR models and NDFD were superior to that of the corresponding TANN models in several other cases. Results indicate that the TANN models can provide automated predictions with skills similar to that of operational forecasters. Comparisons of the two sets of TANN models reveal utility in the use of feature selection.

KEY WORDS artificial neural networks; multi-linear regression; thunderstorm prediction; feature selection

Received 1 September 2013; Revised 6 December 2014; Accepted 17 December 2014

1. Introduction

The literature on atmospheric science is replete with studies on the development of thunderstorm prediction models, using a variety of numerical techniques, that include statistics, artificial intelligence (AI), and Numerical Weather Prediction (NWP). Statistical methods to predict thunderstorms include the application of logistic regression (e.g. Sanchez *et al.*, 2001), model output statistics (MOS; Reap and Foster, 1979; Schmeits *et al.*, 2005) and multiple discriminant analysis (e.g. McNulty, 1981). AI involves the use of a device (e.g. computer) to reproduce human cognitive processes, including learning and decision making (Costello, 1992; Glickman, 2000). Specific AI techniques used to predict thunderstorms include the application of adaptive boosting (e.g. Perler and Marchand, 2009), artificial neural networks (ANNs) (McCann, 1992; Manzato, 2005; Chaudhuri, 2010), and expert systems (Colquhoun, 1987; Lee and Passner, 1993). These statistical and AI models have demonstrated skills or utility. NWP models provide deterministic predictions of the future atmospheric state by integrating the conservation equations of atmospheric water, motion, mass and heat. With respect to the gridpoint NWP model variety, the different terms in these equations are approximated *via* Taylor series expansions and then integrated forward in time (e.g. Pielke, 2002). Atmospheric processes resolved by the NWP model grid scale are termed 'model dynamics'. Corresponding sub-grid scale processes (convection, radiation, boundary layer, hydrological, land/ocean) are parameterized in terms of the grid scale. Such parameterizations are termed 'model physics' (e.g. Kalnay, 2003). With regard to convective processes, convective parameterization (CP) involves the formulation of the implicit statistical effects of sub-grid-scale convection (e.g. Arakawa, 2004). However, if the horizontal grid spacing of the model is reduced to ≤4 km, an NWP model can explicitly simulate/predict (rather than simply emulate/parameterize the effects of) the structure/evolution of mesoscale convective phenomena such as squall lines (Weisman *et al.*, 1997; Weisman *et al.*, 2008). Although Fowle and Roebber (2003) and Weisman *et al.* (2008) demonstrated the efficacy of using high resolution NWP models to forecast convective occurrence/mode explicitly, yet they had limited success with regard to thunderstorm location/timing.

This study details the development of an ANN model to predict thunderstorm occurrence, within 400 km^2 South Texas domains, 9, 12 and 15 h in advance, and with a temporal resolution of 4 h (hereafter referred as thunderstorm artificial neural network or TANN). The objective is to incorporate AI to improve thunderstorm prediction skill on these scales. In the United States, lightning, on average, is responsible for more deaths than tornadoes, high wind and hurricanes (e.g. Curran *et al.*, 2000), and, therefore, serves as a motivation for this study. Given the constraint to predict with high temporal and spatial resolution, one cannot consider only the synoptic and/or meso-α scale moisture/instability and familiar triggering mechanisms (e.g. cold fronts, sea breeze boundaries, upper level disturbances). Rather,

* Correspondence: W. Collins, National Weather Service, Corpus Christi, TX, USA. E-mail: waylon.collins@noaa.gov

one must also know the processes responsible for convective initiation (CI) on the scale of an individual thunderstorm (∼10 km; Orlanski, 1975). Several studies exist, which provide guidance in this regard. Wilson and Schreiber (1986) found that CI can occur in association with radar-identified micro-α and meso-γ scale (0.5–5 km in width) boundaries, originating as earlier thunderstorm gust fronts, and possibility originating from topographic and thermal gradients. According to Fabry (2006), storm development is extremely sensitive to updraft strength and moisture variability on the meso-γ scale. Further, land surface heterogeneity (variations in soil moisture, vegetation, soil type with a length scale ≥10 km contributes to differential surface heating, which can result in a mesoscale solenoidal circulation pattern in the boundary layer forced by baroclinicity (e.g. Trier, 2003), especially during weak synoptic conditions (e.g. Frye and Mote, 2010). The associated low-level convergence can force surface-based parcels to their level of free convection (LFC), resulting in CI (Avissar and Liu, 1996; Avissar and Schmidt, 1998; Emori, 1998; Lynn *et al.*, 2001; Trier, 2003; Taylor *et al.*, 2007, 2011); this mechanism has a negative feedback (convective development over formerly dry soil). Yet, Taylor and Lebel (1998) demonstrated a positive feedback whereby evaporation over wet soil increases the local boundary layer moist static energy and subsequent convection. Taylor *et al.* (2011) attempted to reconcile both the positive and negative feedback by suggesting that the positive feedback occurs with regard to convection on the order of 100 km, whereas the negative feedback is associated with convective cells on the order of 10 km. Further, Findell and Eltahir (2003a) provide additional information with regard to the prediction of CI owing to soil moisture heterogeneity; they developed a framework using early morning lower-level moisture and 1–3 km layer thermodynamic structure that allows for analysing atmospheric controls on the interaction between soil moisture and the boundary layer. In particular, they distinguished classes of thermodynamic soundings conducive to subsequent convection over dry or wet soils, and found that soil moisture variations can only influence CI when certain thermodynamic structures and lower-level atmospheric moisture value ranges occur. These results reveal the complexity involved in predicting the exact location of CI in response to soil moisture content (SMC) and variations. Further, Khairoutdinov and Randall (2006) suggest that high convective available potential energy (CAPE) and low convective inhibition (CIN) are necessary, yet not sufficient, conditions for surface-based convective development. The horizontal scale of convective clouds must increase to a threshold of around 4 km in order to overcome the dissipative effects of dry air entrainment, and cloud growth requires sufficient low-level moisture (high moist static energy). These studies demonstrate the immense complexity involved in CI and strongly suggests that our thunderstorm prediction model must incorporate relevant predictors on the storm scale. The thunderstorm prediction models discussed earlier contain two serious limitations. First, several of the studies include predictors acquired solely from a rawinsonde at a particular point in time and space (McNulty, 1981; Lee and Passner, 1993; Manzato, 2005; Chaudhuri, 2010), or from NWP analysis only (Perler and Marchand, 2009). When predicting beyond 3 h, it is important to include NWP model prediction output or some other means to predict the future state of the atmosphere (e.g. Wilson *et al.*, 1998). Second, the spatial scale of the predictor variables was too coarse to account for processes that actually trigger individual convective cells (McCann, 1992; Schmeits *et al.*, 2005).The goal of this work is to develop an ANN model that incorporates NWP output and that accounts for storm scale processes that can actually trigger convection. Alternatively, one could explicitly predict individual convective cells simply by increasing the resolution of NWP models to ≤1 km (e.g. Bryan *et al.*, 2003). However, in addition to the limitations mentioned earlier with regard to the studies of Fowle and Roebber (2003) and Weisman *et al.* (2008), Elmore *et al.* (2002) have shown that the explicit forecasting of convection with a high resolution NWP model may not provide additional skill. The problems include the extreme sensitivity of NWP model solution to initial conditions (chaos), and the extreme sensitivity of model solution to parameterization trigger thresholds (bifurcation).

Notwithstanding predictability limitations inherent in NWP models, the authors posit that the ANN AI approach can be used to 'learn' these limitations and generate skillful predictions. Thus, this study involves the development of an ANN model by post-processing NWP output (12 km grid spacing) and sub-grid scale data (4 km grid spacing). Thus, the model combines a deterministic prediction of the future state of the mesoscale environment with data sufficient to capture processes that can trigger individual convection cells. TANN models were developed for prediction times 9, 12 and 15 h ±2 h, relative to 1200 UTC. Each model was trained (with validation) on a randomized data set spanning most of the period 2004–2006 and 2009–2013. ANN binary classifiers were developed by using receiver operating characteristic (ROC) curves to threshold ANN continuous output. The performance of the optimized TANN models was evaluated on a novel data set (2007–2008). TANN model performances were compared to the corresponding performances of multi-linear regression (MLR) models, and to selected public and aviation forecasts from the US National Weather Service.

Methodology and model development are discussed in Sections 2 and 3, respectively. TANN model performance evaluation is presented in Section 4. A discussion and conclusion are presented in Section 5.

2. Methodology

2.1. Data

2.1.1. TANN domain

The TANN domain is slightly larger than the County Warning and Forecast Area (CWFA) of the U. S. National Weather Service (NWS) Weather Forecast Office (WFO) in Corpus Christi, Texas (CRP). The latitude/longitude pair of the SW, NW, NE and SE corners of the domain are 26.91000 ° N/100.58000 ° W, 29.25613 ° N/100.58000 ° W, 29.17955 ° N/96.05539 ° W, and 26.86122 ° N/96.15182 ° W, respectively. For this study, the TANN domain is represented as a grid of 13 × 22 equidistant points with a horizontal grid spacing of 20 km. To create the foregoing domain, the position of each grid point was determined *via* the Inverse and Forward computer software programs provided by the US National Geodetic Survey (National Geodetic Survey, 2006) and based on the equations in Vincenty (1975). These points serve as boundaries/centres for 286 20 km × 20 km (400 km^2) square regions (hereafter referred as 'boxes') (Figure 1). MATLAB® (MathWorks, 2014) software was used to establish a framework to import the necessary data, train/validate, test, and further optimize an ANN in each of these 286 boxes. Yet, for this study, TANN models for only three boxes were trained, which were chosen based on the amount of target data (discussed below) and the desire to reflect the diversity of thunderstorm triggering mechanisms, and was based on the proximity to Terminal Aerodrome Forecast (TAF) locations

(Appendix S3, Supporting Information). The first box chosen was located in the northeastern sector of the domain (Box 238), a region with a relatively high frequency of cloud-to-ground (CG) lightning strikes, in order to maximize the amount of target data. Another box chosen was located in the far southwest sector adjacent to the Rio Grande River that borders Mexico (Box 73), which climatologically has a much smaller frequency of CG lightning than the corresponding frequency over the northeastern section of the TANN domain; thus, the utility of the model with limited target data could be assessed. The third box chosen was located near the Gulf of Mexico just west of Corpus Christi (Box 103). Boxes 103 and 238 were located in the Western Gulf Coastal Plain region of Texas (Texas Parks and Wildlife, 2011). Much of the thunderstorm activity initiated within the Coastal Plain is likely to be triggered by coastal processes (e.g. sea breezes). A fraction of thunderstorms near Box 73 develops over the Sierra Madre Oriental mountain chain in Mexico and propagate into the box. Further, South Texas experiences thunderstorm development in association with synoptic scale features (fronts, upper level disturbances). However, on several occasions, even during synoptically favorable dynamics, thunderstorm development will occur over Box 238, yet not over Boxes 73 and 103, due to the development of an elevated stable layer, possibly in response to the advection of warmer air (possibly due to compressional heating) from approximately southwest to northeast downwind from the Sierra Madre Oriental. Thus, the box regions chosen reflect the diversity with respect to thunderstorm mechanism and frequency.

The strategy is to develop a binary classifier to predict the occurrence or non-occurrence of thunderstorms in the selected boxes, based solely on NWP predictions and sub-grid scale conditions within each box. Thus, with regard to the 4 km data used to assess soil moisture heterogeneity mentioned earlier, Taylor *et al.* (2007) and Taylor *et al.* (2011) used observational data to demonstrate that the heterogeneity necessary to generate circulation patterns should possess a length scale of at least 10–20 km. At length scales less than 10 km it is likely that turbulent mixing magnitudes exceed the pressure gradient force that is essential for the inducement of heterogeneity-induced mesoscale circulations, as discussed in Avissar and Liu (1996). Thus, the 400 km^2 box regions used in this study appear sufficiently large to support soil moisture heterogeneities of a scale large enough to generate mesoscale circulations. Further, the spatial resolution of the predictions in this study is greater than that of the thunderstorm models featured in the Introduction (Table 1).

2.1.2. TANN target data

In this study, the target refers to the existence, or non-existence, of thunderstorms (categorical). CG lightning was chosen as the proxy for thunderstorm occurrence. The CG source data originated from the National Lightning Detection Network (NLDN) (e.g. Orville, 2008) (see Appendix S3 for additional information regarding this data set). The source data are discrete (number of CG lightning strikes *per* unit time and location) and are transformed to binary (Section 2.3).

Figure 2 depicts the histogram (mean CG lightning $box^{-1}\,day^{-1}$ *versus* hour in UTC) of CG lightning for the entire TANN domain (Figure 1) for the period 1 January 2003 through 31 December 2011. Note that the majority of thunderstorm activity for the domain typically occurs during the afternoon and early evening hours. These results influenced the decision to predict thunderstorms only for the 2100–0300 UTC (±2 h) period.

2.1.3. TANN input data: potential predictor variables

Three primary categories of input data were selected for the TANN. The first category represents 35 output variables/parameters from (or derived based on output from) the 1200 UTC cycle of the hydrostatic *Eta* mesoscale NWP model (e.g. Rogers *et al.*, 1996) (1 March 2004 to 19 June 2006), the Weather Research and Forecasting Non-hydrostatic Mesoscale Model (WRF-NMM) Janjic *et al.*, 2001 (20 June 2006 to 30 September 2011), and from the NOAA Environmental Modeling System Non-hydrostatic Multiscale Model (NEMS-NMMB). (October 2011 to December 2012) (Tables 2 and 3). These NWP models are identified in this study as NAM (North American Mesoscale) (see Appendix S3).

The choice of NAM input variables/parameters is based on the recognition that a thunderstorm requires sufficient moisture to generate necessary hydrometeors, atmospheric instability to generate updrafts sufficient to carry parcels to heights necessary to produce ice crystals, supercooled water and graupel sufficient for the graupel-ice crystal charging mechanism (e.g. Saunders, 1993), and (unless the atmosphere is absolutely unstable) a mechanism to lift parcels to the LFC. Furthermore, various dissipative effects must be overcome, such as dry air entrainment, unfavorable microphysics, and vertical wind shear immediately above the boundary layer.

The second category of data includes seven parameters that represent meso-γ scale data (not explicitly accounted for by the 12 km NAM and, thus, hereafter referred to as sub-grid scale data) that contribute to CI (Table 5). As discussed in the Introduction Section, it is essential to account for sub-grid scale land surface heterogeneity and soil moisture. In this study, our proxy for surface heterogeneity is restricted to the soil moisture pattern and variability. The calculation of soil moisture used in this study is described in Appendix S1. With regard to the SMC *pattern*, the spatial autocorrelation (SA) metric Moran's *I* is used; a detailed description of this variable can be found in Appendix S4. Soil moisture variability is accounted for by SMC gradients and standard deviations based on calculations applied to the 4 km grid in each box.

It should be noted that soil moisture has been assimilated into the NAM *via* the NOAH land surface model since 31 January 1996 (Mitchell, 2005). However, the 12 km grid spacing of the NAM cannot resolve the soil moisture patterns and the resultant mesoscale circulations that develop in the 400 km^2 box regions used in this study. As mentioned earlier, previous studies have demonstrated that soil moisture heterogeneity, at length scales of at least 10–20 km, can influence the wind field. At least four grid points are needed to adequately resolve a feature (Ross, 1986; Lewis and Toth, 2011). Thus, an NWP model needs to have a grid spacing of ≤6.7 km to resolve a heterogeneity with a length scale of 20 km. Yet, the NAM has a 12 km grid spacing. The SMC gridded data used in this study has a grid spacing of 4 km, sufficient enough to resolve SMC patterns/resultant circulations within the 20 km × 20 km box regions.

Figure 4 illustrates the relative position of all data points, relative to Box 238, for which all predictor variables are extracted for input to the TANN. (Corresponding information from Boxes 73 and 103 is not shown.)

The third category includes the miscellaneous parameter Julian day (JD), which is based on the reasoning that TANN's performance can be improved with knowledge of thunderstorm frequency as a function of season (Table 4).

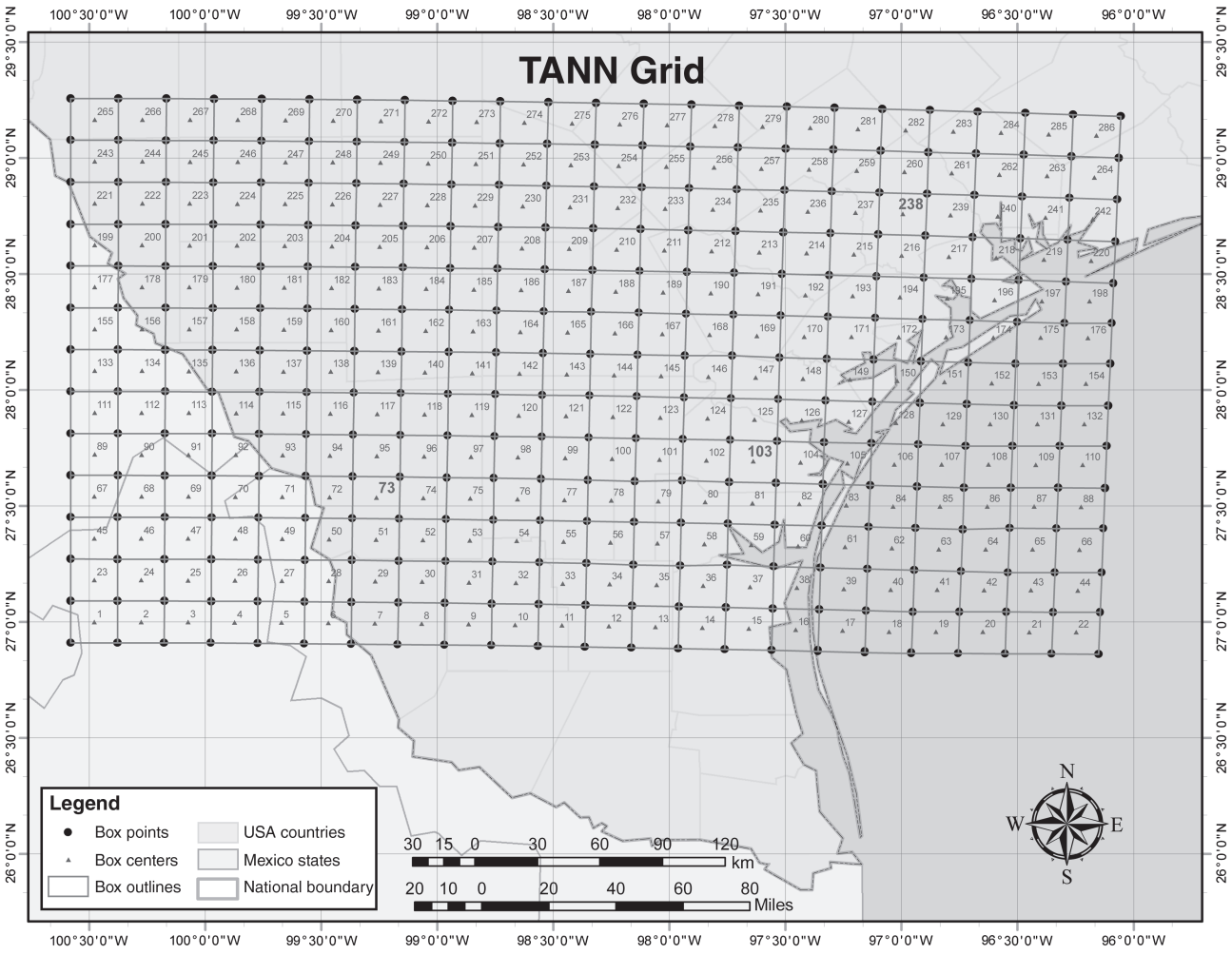

Figure 1. TANN grid domain. This grid is defined by 286 20 km × 20 km square regions ('boxes'). Note that each box is assigned an identification number (labelled inside each box). The authors established a framework to predict whether a thunderstorm will occur within each box 9–15 h in advance. Yet, for this study the authors developed ANN models to predict thunderstorm occurrence only in Boxes 73, 103 and 238 (identification numbers in larger font). Figure created by Rick Smith with modifications by Anthony Reisinger (Texas A&M University – Corpus Christi).

Table 1. Comparison of selected thunderstorm models: prediction time and space resolutions.

Peer-reviewed study	Verification domain	Prediction period (h)	Model type
McNulty (1981)	1° radius	12	Discriminate analysis
McCann (1992)	1° grid	3–7	ANN
Lee and Passner (1993)	100 km radius	12	Expert system
Sanchez *et al.* (2001)	6825 km^2	6–12	Logistic regression
Schmeits *et al.* (2005)	7200 km^2	48	MOS
Manzato (2005)	5000 km^2	6	ANN
Perler and Marchand (2009)	729 km^2	24	Adaptive boosting
Current study	400 km^2	9–15	ANN

2.1.4. TANN input data: dimensionality reduction

The authors seek a subset from the list of predictor candidates from Section 2.1.3, which will potentially improve model performance beyond that of models that include all predictors. In other words, the desire is to minimize the size of the input dimension while retaining variables sufficient to describe the behaviour of the target (feature selection). Attempts to decrease the number of predictor variables is important in part due to the curse of dimensionality (Bellman, 1961), which suggests that the amount of sample data required to develop a skillful model increases exponentially as the model dimension increases linearly. Thus, an attempt to train the TANN using all 43 inputs will increase the risk that the predictor model may require more model training data than is available for skillful model performance. Further, predictor variables that are strong/relevant (contain sufficient information regarding the behaviour of the output) are preferred. In addition, predictor variables that are irrelevant or redundant can result in a model with subpar performance (inability to generalize); redundant variables can increase the likelihood of convergence on local minima on the error surface during model training, while irrelevant data can add noise to the model and

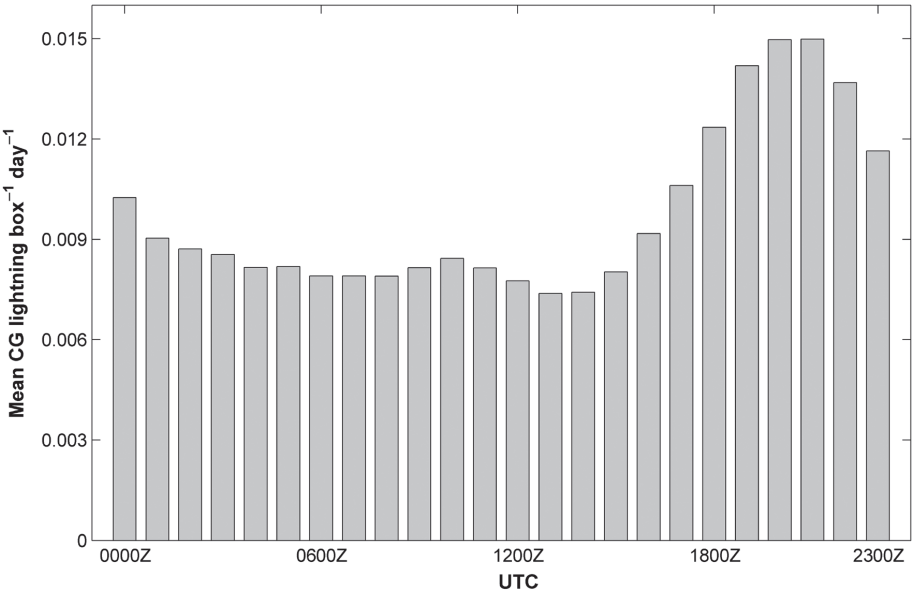

Figure 2. Histogram (mean CG Lightning box^{-1} day^{-1} *versus* hour in UTC) of CG lightning strikes for the TANN domain during the period 1 January 2003 through 31 December 2011. Note that the greatest frequency of CG strikes occurs during the afternoon and evening hours.

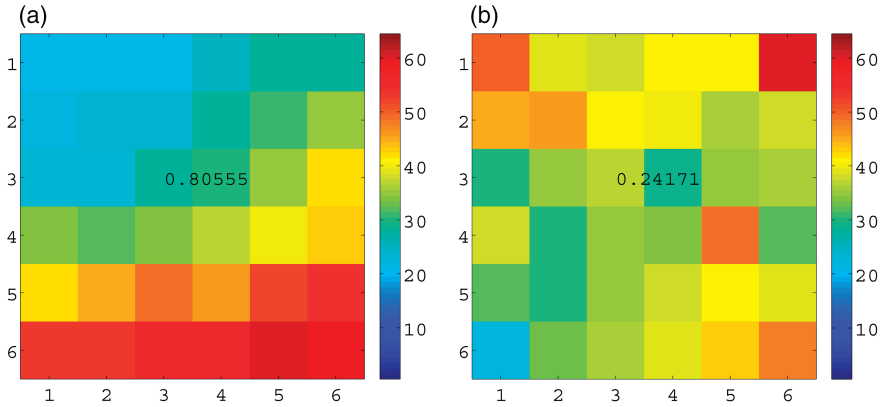

Figure 3. Soil moisture content (SMC) (rescaled to 0–60 kg m^{-2} range for visualization purposes only) patterns for 15 June 2004 (a) and 25 October 2008 (b) for Box 238. The value in the centre of each box is the corresponding GMORANI value. The pattern to the left (GMORANI = 0.80555) is more conducive to the development of heterogeneity-induced circulation and subsequent convective initiation *via* negative feedback, than the pattern on the right (GMORANI = 0.24171).

adversely affect the neural network's learning process (e.g. May *et al.*, 2011).

The determination of the subset of the 43 predictors that provide the greatest information about thunderstorm occurrence was made by the filtering-based feature selection technique of correlation-based feature selection (CFS) (Hall and Smith, 1999) applied to the entire training sample and corresponding target. CFS uses an heuristic method to access the worth of features. In particular, CFS uses symmetric uncertainty (information theory-based metric) to calculate both feature–target and feature–feature correlations, to access relevance and redundancy, respectively. These correlations are then used to search the feature subset domain. The search begins with an empty feature subset then features are added in accordance with a forward best first search with a stopping criterion of five consecutive non-improving subsets. The feature subset that maximizes the metric known as heuristic merit (to achieve maximum relevance and minimum redundancy) is chosen. CFS is a nonlinear technique in the sense that it accounts for nonlinear relationships between the features and the target. The CFS technique was run using a MATLAB® version provided by the Arizona State University Feature Selection Repository (Arizona State University, 2014).

Table 6 depicts the predictor variables chosen *via* CFS as a function of box (73, 103 and 238) and prediction hour (9, 12 and 15 h). Note that the predictors chosen and the number of predictors in the subset vary for the nine cases; the number of predictors chosen varied in the range [2,12]. The fact that convection requires adequate moisture, instability and a trigger mechanism is reflected in the selected predictors. With regard to moisture, the total precipitable water (PWAT) is common to nearly all cases. Note that $\mathrm{DROPOFF}_{\mathrm{PROXY}}$, $\mathrm{CTP}_{\mathrm{PROXY}}$, LI or CAPE is represented as an instability predictor. Further, vertical velocities VV_{700}, VV_{925} or VV_{500} is represented in five of the nine cases and likely reflects the various trigger mechanisms such as sea breezes and upper level disturbances. It is noteworthy that the NAM output predictor variable convective precipitation (CP) is common to all cases. CP is a consequence of the Betts-Miller-Janjic (BMJ)

Table 2. Description NAM predictor variables/parameters used in TANN for day = *D*.

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 a horizontal scale of ≥4 km in order to overcome dissipative effects (Khairoutdinov and Randall, 2006)
RH_{850}	Relative humidity at 850 hPa (%)	When combined with CAPE, predictor of subsequent thunderstorm location independent of synoptic pattern (Ducrocq *et al.*, 1998)
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 (e.g. Trier, 2003)
CIN	Convective inhibition (J kg^{-1})	Surface-based convective updraft magnitude must exceed $(CIN)^{1/2}$ for parcels to reach level of free convection (e.g. Trier, 2003)
LI	Lifted index (K)	Atmospheric instability proxy; utility in thunderstorm prediction Haklander and Van Delden, 2003
U_{LEVEL},V_{LEVEL}	U,V wind components at surface, 850 hPa [LEVEL = surface, 850 hPa] (m s^{-1})	Strong wind can modulate or preclude surface heterogeneity-induced mesoscale circulations (Dalu *et al.*, 1996; Wang *et al.*, 1996)
VV_{LEVEL}	Vertical velocity at 925, 700, 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 (Crook, 1996)
LCL	Lifted condensation level (m)	Proxy for cloud base height; positive correlation between cloud base height and CAPE to convective updraft conversion efficiency (Williams *et al.*, 2005)
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 collisional mechanism (Saunders, 1993)
CP	Convective precipitation (kg m^{-2})	Byproduct of the Betts-Miller-Janjic convective parameterization scheme (Janjić, 1994), when triggered; proxy for when the NAM anticipates existence of sub-grid scale convection
VSHEARS8	Vertical wind shear: 10 m to 800 hPa layer ($\times 10^{-3}$ 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 (Rotunno *et al.*, 1988)
VSHEAR86	Vertical wind shear: 800–600 hPa layer ($\times 10^{-3}$ s^{-1})	Convective updraft must exceed vertical shear immediately above the boundary layer for successful thunderstorm development (Colquhoun, 1987; Crook, 1996)

Table 3. Description of NAM initialization predictor variables/parameters used in TANN for day = *D*.

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] (m s^{-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 (Findell and Eltahir, 2003b)
HI_{LOW}	Humidity index (°C)	Both a constraint on afternoon convection and an atmospheric control on the interaction between soil moisture and convection (Findell and Eltahir, 2003b)
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 (Findell and Eltahir, 2003a)
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 (Findell and Eltahir, 2003b)
VSHEAR75	Vertical wind shear: 700–500 hPa layer ($\times 10^{-3}$ s^{-1})	Convective updraft must exceed vertical shear immediately above the boundary layer for successful thunderstorm development (Colquhoun, 1987; Crook, 1996)

CP scheme. When triggered, the BMJ scheme emulates sub-grid scale convection and removes excess instability by adjusting the model thermodynamic sounding to a climatologically-derived reference profile; the corresponding reduction in model PWAT results in CP as a byproduct (Bua and Jascourt, 2009). The ubiquity of CP as a survivor of the feature selection process suggests that the triggering of the BMJ scheme within the NAM is highly correlated with the target. In addition, the selection by CFS of U_{800} (0) and U_{600} (0) for Box 73, U_{800} (0) for Box 103 and V_{SFC} for Box 238 (when considering the curvature of the South Texas coast), likely reflects the advection of moist air from the Gulf of Mexico, which can increase the likelihood for convection). Further, note that predictor VV_{925} was selected for 9 h predictions at Boxes 103 and 238, which likely reflects the sea

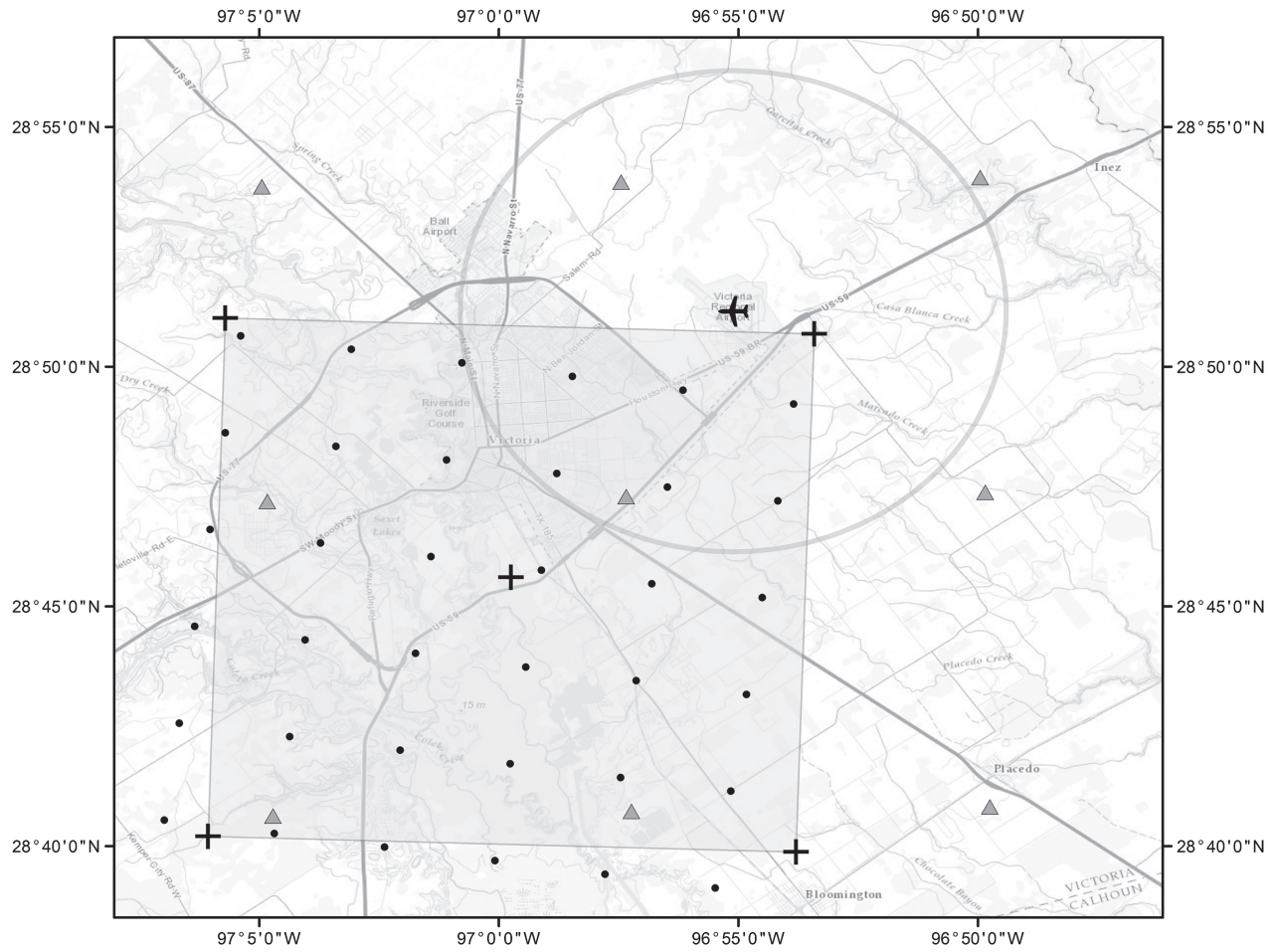

Figure 4. Box 238 domain and vicinity. Smaller box region corresponds to the Box 238 domain; plus symbols (+) mark the centre and four corners of the box. Locations of NAM grid points (grid spacing 12 km) indicated by triangles (▲); the values of the NAM parameters/variables used in this study are based on bi-linear interpolation at the centre of the box. Small circles (●) depict the MPE grid points (4 km grid spacing). The airplane symbol depicts the location of the METAR station at the Victoria Regional Airport while the circle illustrates a radius of 9.26 km (5.00 nm) relative to the METAR station.

Table 4. Miscellaneous variables/parameters used in TANN for day = *D*.

Abbreviation	Description (units)	Justification as thunderstorm predictor
JD	Julian day (day)	Provide memory to the TANN regarding thunderstorm occurrence as a function of season

breeze which is a primary driver for summer convection around 9 h within the Western Gulf Coast Coastal Plain region (which includes Boxes 103 and 238). Finally, note that only one of the nine box/prediction hour combinations (Box 103 9 h prediction) contained any sub-grid scale features (Table 5) chosen during the feature selection process. This result suggests that the contribution of soil moisture and surface heterogeneity to convective development was limited; an alternative explanation is that, although the sub-grid scale features likely influenced convection (considering the fact that SMC_MEAN and GMORANI were chosen as features by CFS for the Box 103 9 h case), for nearly all cases, the BMJ CP scheme more effectively accounted for sub-grid scale processes that relate to CI than the sub-grid scale features.

2.2. The artificial neural network

An ANN is a parallel distributed processor (e.g. Haykin, 1999). Parallel distributed processing is a framework developed to explain the theory of human cognition known as connectionism (e.g. Rumelhart and McClelland, 1986). This theory assumes that the brain is composed of a network of highly inter-connected processing units (neurons), and that cognitive processes are manifested as interactions between these units that inhibit/excite each other in parallel. Thus, knowledge is stored in the form of connection strengths between neuron pairs that are distributed across this biological neural network. There is a view that the ability of humans to rapidly solve very complex pattern recognition problems is related to the foregoing highly inter-connected and parallel structure within the human brain (e.g. Beale and Jackson, 1990; Hagan *et al.*, 1996).The ANN is used to process information in a parallel distributed fashion in order to solve complex problems. The two major types of problems that are solved by ANNs are functional approximation (estimating the functional relationship that describes the nonlinear input–output mapping) and classification (the assignment of an input pattern to a particular category or class).

One can characterize an ANN based on its topology, transfer function, and learning algorithm (e.g. de Silva, 2000). The topology refers to the manner in which the foregoing inter-connections

Table 5. Sub-grid scale (4 km) predictor variables/parameters used in TANN for day = $D - 1$ (previous day).

Abbreviation	Description (units)	Justification as thunderstorm predictor
SMC_MAXGRAD	Maximum soil moisture content (SMC) gradient ($kg\,m^{-2}\,km^{-1}$)	Soil moisture gradient of length scale $\geq$10 km can generate a lower-level mesoscale thermally direct circulation pattern; associated surface convergence can result in CI (Taylor *et al.*, 2007; Taylor *et al.*, 2011)
SMC_MEANGRAD	Mean SMC gradient ($kg\,m^{-2}\,km^{-1}$)	Soil moisture gradient of length scale $\geq$10 km can generate a lower-level mesoscale thermally direct circulation pattern; associated surface convergence can result in CI (Taylor *et al.*, 2007; Taylor *et al.*, 2011)
SMC_SD	SMC standard deviation ($kg\,m^{-2}$)	Probability of CI increases as land surface temperature anomaly standard deviations (correlated with SMC_SD) increase (horizontal scales $\leq$40 km) (Taylor *et al.*, 2011)
SMC_MEAN	Mean SMC ($kg\,m^{-2}$)	Evaporation of soil moisture increases local moist static energy in the PBL (Madden and Robitaille, 1970; Betts and Ball, 1998)
SMC_MAX	Maximum SMC ($kg\,m^{-2}$)	Evaporation of soil moisture increases local moist static energy in the PBL (Madden and Robitaille, 1970; Betts and Ball, 1998)
GMORANI	Global Moran's I of SMC (unit less); proxy for spatial autocorrelation of SMC	Pattern of wet soil (continuous region covering 50% of 400 km^2 box) adjacent to dry soil (covering the remaining 50%) is most conducive to thermally direct circulation pattern and will result in a GMORANI value of 0.857 (Longley *et al.*, 2005)
NDRY	Number of dry days [MPE = 0] since the previous rain (days)	For NDRY $\neq$ 0, evaporation of soil moisture, which can increase local moist static energy (Taylor and Lebel, 1998); proxy for indirect measure of SMC without use of the API model

between the neurons are arranged and how the data flows through the system. The transfer function is an algebraic function that facilitates the transfer of information between neurons; in particular, this function receives input values from neighboring neurons, evaluates the input, and then transfers the function output to subsequent neurons. The objective of the learning algorithm is to train the ANN to learn a particular task, which is achieved by adjusting the connection strengths/weights between neurons within the network in order to optimize performance. The process of adjusting weights and biases is consistent with learning; there is evidence that when humans learn, the coupling or connection weights between specific neurons are adjusted (Beale and Jackson, 1990; Hagan *et al.*, 1996). The topology, transfer function and learning rule are chosen in accordance with the type of problem. The development of an ANN model involves the formulation of the problem, data processing, training and evaluation (e.g. Zhang *et al.*, 2003). The modeler first needs to designate the problem the model is being developed to solve, and then identify the input parameters/variables that affect the output. Next, during the data processing phase, the modeler would select the proper range/sample distribution of data, organize the data into subsets for training, validation and testing and then pre-process the data (scaling). After data processing, the ANN is trained and validated. Finally, the trained ANN is evaluated by assessing model performance on novel data (the testing set). Subpar performance on the testing set suggests possible over-fitting whereas good performance similar to that of the training set indicates ability for the ANN to generalize.

The ANN has advantages over the use of traditional statistical techniques, such as nonlinear regression and discriminant analysis. Nonlinear regression models require assumptions regarding the form of the functional relationship between the predictors and the predictand (e.g. polynomial regression assumes that the relationship can be described by polymonials). Discriminate analysis assumes that the probability density functions of the predictors satisfy a Gaussian distribution. In contrast, ANNs do not require knowledge or assumptions of the form of the functional relationship between the predictors and predictand. By adjusting the number of hidden layers and associated neurons, the optimum function can be determined. Thus, ANNs are essentially universal approximators (e.g. Hornik *et al.*, 1989).

2.3. TANN architecture and training algorithm

With regard to TANN topology, the feed-forward multi-layer perceptron (MLP) structure was chosen. Within MLP, the neurons are arranged in layers. In particular, there is an inter-connection between the input layer, hidden layer(s), and an output layer, wherein each layer can have any number of neurons. The notation X-{Y_1, Y_2, … Y_N}-Z is used. The variables X, Y and Z refer to the number of neurons in the input, hidden and output layers, respectively. N refers to the number of hidden layers. The TANN consists of one hidden layer ($N = 1$), which is sufficient (Section 2.2). Hence, the final notation X-Y-Z is used. It is known from Section 2.1.4 that X varies within the range [2, 43]. The output layer consists of only one neuron ($Z = 1$). The determination of the optimal number of neurons in the hidden layer is explained in Section 3.

Feed-forward refers to the one-way transfer of information from the inputs to the output layer. The feed-forward processing from input to hidden to output layer within the MLP can be described as follows: Consider transfer function f, and vectors $\mathbf{x}$, $\mathbf{W}$ and $\mathbf{b}$, corresponding to input, weight and bias vectors, respectively. Output $\mathbf{a}$ from the MLP, with a single hidden layer, following forward propagation of data through the network is defined as follows:

$$\mathbf{a}_2 = f_2\left(\mathbf{W}_2 f_1\left(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1\right) + \mathbf{b}_2\right) \quad (1)$$

where the subscripts 1 and 2 represent the hidden and output layers, respectively. The transfer functions used in this study are the log-sigmoid and linear equations used in the hidden and output layers, respectively. The log-sigmoid function (S) is defined as

$$S(x, k) = \frac{1}{1 + e^{-kx}} \quad (2)$$

with x and k as continuous value and slope parameter, respectively. For $x \to -\infty$ *to* $+\infty$, $S(x,k) \to [0,1]$. It has been shown that an MLP with one hidden layer and with the log-sigmoid and linear transfer functions used to activate neurons in the hidden and output layers, respectively, can approximate any continuous functional relationship (e.g. Hornik *et al.*, 1989), provided a sufficient number of neurons exist in the hidden layer (e.g. Hagan *et al.*, 1996). The training in this study is supervised. Supervised training involves the process of providing a set of N training sample examples $\left\{\left(\boldsymbol{x}_i, \boldsymbol{t}_i\right)\right\}_{i=1}^{N}$ where $\boldsymbol{x}_i$ represents an input vector and $\boldsymbol{t}_i$ depicts the corresponding desired output (known as the target). In this study, $\boldsymbol{x}_i$ refers to the X predictor variables and $\boldsymbol{t}_i$ is actually a scalar (t_i) that represents the thunderstorm binary output as follows:

$$t_i = \begin{cases} 0, & l = 0 \\ 1, & l \neq 0 \end{cases} \tag{3}$$

where l represents the number of CG lightning strikes *per* hour within each 400 km^2 box. With regard to the training algorithm/learning rule, recall that ANN learning involves adjusting network connection weights to optimize performance; optimization occurs when the global minimum of the error in weight space (error surface) is reached. The search for the global minimum begins with an initial position $\boldsymbol{X}_0$ (generally chosen randomly) in weight space. Subsequent search directions are updated iteratively in accordance with the relationship of the form $\Delta \boldsymbol{X}_k = (\boldsymbol{X}_{k+1} - \boldsymbol{X}_k) = \eta_k \boldsymbol{d}_k$,where the positive scalar parameter η_k is the learning rate and $\boldsymbol{d}_k$ is the search direction; η_k determines the incremental distance in the search direction (e.g. Hagan *et al.*, 1996). Various numerical optimization techniques exist representing different search strategies. First-order techniques assume that the error surface is represented as a first-order Taylor series expansion about position $\boldsymbol{X}_k$ in weight space, while second-order methods assume the error surface around $\boldsymbol{X}_k$ is represented as a second-order Taylor series (quadratic). A first-order method where the search direction $\boldsymbol{d}_k$ is given by the negative of the local gradients is known as gradient descent. However, the local gradient and choice of η may result in an oscillation of the search in weight space thus resulting in an inefficient search path (e.g. Bishop, 2005; Hagan *et al.*, 1996). A second-order method known as Newton's method involves the generation of a local quadratic approximation of the error surface (*via* second-order Taylor series expansion) followed by a line search towards the location of the minimum of the quadratic approximation. However, this method requires the calculation of the Hessian matrix, which may become impractical for nonlinear ANNs due to high computational cost. In this study, a second-order method known as scaled conjugate gradient (SCG) is used, which searches the weight space along conjugate directions, makes implicit use of the second-order terms provided by the local Hessian matrix, converges more efficiently than gradient descent and avoids problems associated with line searches (Moller, 1993).

The ANN training, validating and testing were performed *via* the MATLAB® Neural Network Toolbox software version R2014a (MathWorks, 2014). All input data was scaled to the range [−1, 1] before serving as input to the network. The initial values in the weight vector **W** are randomly selected prior to training *via* SCG.

3. TANN model development

This study involves the development of 18 TANN models to predict thunderstorm occurrence within three 400 km^2 box domains (Boxes 73, 103 and 238) and for three prediction hours (9, 12 and 15 h), consisting of two sets of nine models each. In one set, the predictor variables were chosen *via* CFS (final selection depicted in Table 6), hereafter referred to as the TANN *X*-*Y*-1 models (recall *X*-*Y*-*Z* topology convection from Section 2.3). The second set used all 43 predictors (hereafter referred as TANN 43-*Y*-1 models). This division was generated in part to assess the utility of feature selection (discussed in Section 2.1.4).

The development of the TANN models began with the determination of ANN topology. Note from Section 2.1.4 that the number of predictor variables (X) determined for TANN *X*-*Y*-1 models varies within the range [2, 12]; $X = 43$ for the nine TANN 43-*Y*-1 models. Both sets of models possess only one neuron in the output layer ($Z = 1$). Before discussing the method used to determine the number of hidden neurons in each model, it is necessary to clarify the definition of the training, validation and testing sets used in this study. The terminology is compressed to training sample and independent testing sample. The training sample refers to the data for the periods 2004–2006 and 2009–2013. This training sample is used to calibrate the models, including the selection of the threshold to transform continuous model output to a binary classifier. The independent testing sample refers to the data for the period 2007–2008. The independent testing sample is used to evaluate the performances of the calibrated TANN models, the corresponding MLR models and human forecasters, which provide a basis for comparison. The determination of the optimal number of hidden layers (Y) for each model was determined as follows. For each model, only the training sample is used. The training sample is divided into training and validation data sets at fractions of 0.8 and 0.2, respectively, with random attribution of the training and validation cases. The training sample was used to train and validate the model. The output of the trained model is continuous. As the goal is to develop an ANN binary classifier, a receiver operating characteristic (ROC) curve (e.g. Jolliffe and Stephenson, 2003) was created on the training component of the training sample. The ROC curve is a graphical measure of skill for binary classifiers and is based on signal detection theory (e.g. Swets, 1973). The ROC curve was created by adjusting the decision threshold at an iteration of 1/10 000$^{\text{th}}$ of the ANN output range and calculating the POD and F at each iteration. The threshold chosen corresponds to the point on the ROC curve where Peirce skill score (PSS) performance metric is maximized (Manzato, 2007). Figure 6 depicts one of the set of ROC curves associated with the development and evaluation of the TANN 43-20-1 model (Box 73 15 h). The foregoing process from training/validating to testing and determining the maximum PSS within the training sample is repeated 50 times and a mean PSS and standard error PSS were calculated. This process was repeated for the [1,100] range of hidden layer neurons to determine the optimum number, defined as the minimum number of hidden neurons corresponding to a PSS standard error that overlaps the PSS standard error associated with the maximum PSS. Figure 5 depicts the Box 73 15 h model development example. Tables 7 and 8 depict the chosen topologies, including the number of hidden layer neurons (Y), as a function of box and prediction hour for the 18 models. Note that 72, 56 and 28% of the 18 TANN models contain greater than 5, 10, and 20 neurons in the hidden layer, respectively. The greater number of hidden layer neurons allow the TANN to explore highly nonlinear relationships between the predictors and corresponding target.

Table 6. Variables (Tables 2–5) chosen by the CFS feature selection technique; variables followed by zero depict NAM initialization variables.

Prediction Hour	Box 73	Box 103	Box 238
9 h	PWAT, CTP_PROXY, LI, CIN, DROPOFF_PROXY, $U_{800}(0)$, $U_{600}(0)$, CP, T_LCL, RH_{850}	$U_{800}(0)$, HI_{LOW} , CP, VV_{925}, RH_{850}, MR_{850} , LI, CIN, SMC_MEAN, PWAT, CTP_PROXY, GMORANI	CTP_PROXY, CP, VV_{925}, V_{SFC}, RH_{850}, CAPE, PWAT, DROPOFF_PROXY
12 h	CTP_PROXY, CP, CAPE, PWAT, T_LCL	CP, VV_{500} ,VV_{925}, PWAT, RH_{850}, CTP_PROXY,	CP, RH_{850}, PWAT, LI
15 h	CP, VV_{700}	CP, VV_{925}, PWAT	CP, PWAT

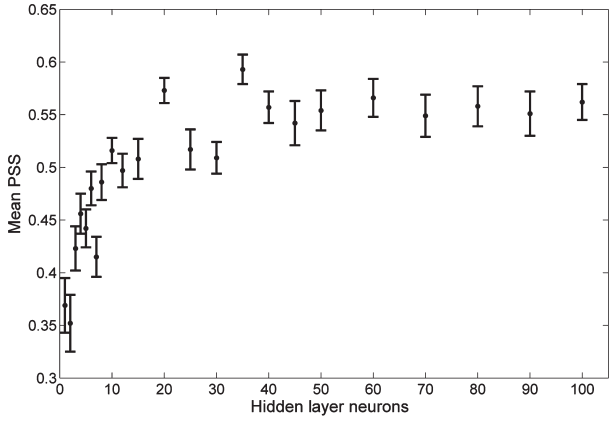

Figure 5. Determination of the optimal number of hidden layer neurons for the Box 73 15 h prediction TANN 43-20-1 model. Each point and corresponding error bar represents the mean Peirce skill score (PSS) and standard error resulting from 50 iterations with results computed based on the training portions 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, 35 hidden layer neurons correspond to the maximum mean PSS; yet, 20 hidden layer neurons was chosen as optimal.

Table 7. Topology (*X*-*Y*-*Z*) of TANN *X*-*Y*-1 models.

Prediction hour	Box 73	Box 103	Box 238
9	10-25-1	12-30-1	8-7-1
12	5-1-1	6-8-1	4-15-1
15	2-5-1	3-12-1	2-2-1

Table 8. Topology (*X*-*Y*-*Z*) of TANN 43-*Y*-1 models.

Prediction hour	Box 73	Box 103	Box 238
9	43-25-1	43-8-1	43-1-1
12	43-5-1	43-30-1	43-15-1
15	43-20-1	43-35-1	43-20-1

4. TANN model performance evaluation

Once the neural network topology determination was complete (Tables 7 and 8), each of the TANN models, with corresponding threshold, were applied to the independent data set (2007–2008) and nine performance metrics are computed. Figure 6 depicts one of the sets of ROC curves associated with the development and evaluation of the TANN 43-3-1 model (Box 73 15 h). This procedure is repeated 50 times to generate median performance statistics. The confusion matrix convention chosen and the corresponding equations of the specific performance metrics (common for binary classifiers) used in this study are listed in Tables 9 and 10, respectively. The skill-based metrics used are PSS, critical success index (CSI), Heidke skill score (HSS), odds-ratio skill score (ORSS) (otherwise known as Yule Q), Clayton skill score (CSS) and Gilbert skill score (GSS). Hogan *et al.* (2010) demonstrated that the performance metrics HSS and PSS are truly equitable. Equitability includes the requirement that the performance metric equals zero for random or constant forecasts; this is desirable because a true skill score should equal zero for forecasts/predictions that require no skill (random or constant forecasts). According to Hogan *et al.* (2010), ORSS is asymptotically equitable (trend towards equitability as the data sample approaches infinity) and truly equitable for the special case of the observed frequency $[(a+c)/n$, where $n =$ data sample size] of the condition predicted equals 0.5. See Wilks (2006) and Jolliffe and Stephenson (2003) for a more detailed description of the utility of these parameters. TANN model performance was compared to the performances of MLR models and to the performance of thunderstorm forecasts generated by WFO CRP forecasters, embodied within aviation and public forecasts. The MLR models were developed *via* stepwise forward linear regression (SFLR) applied to the training sample and corresponding target. For each MLR model, the SFLR process began with a constant value ($y = \beta_0$.) Next, at each forward step in the SFLR process, a predictor is added (from the list of 43 predictors in Tables 2–5, less the predictors already chosen) based on the change in the value of the Akaike information criterion (AIC) (Akaike, 1973); this process continues until the AIC is no longer decreased. As a final step, the method checks if the removal of one or more of the selected predictors further decreases the AIC and, if so, removes the predictor. AIC is a commonly used and implemented model selection process that accounts for the number of predictors and favours parsimonious models. Following Akaike (1973), the model chosen has the smallest AIC and the process results in the selection of the predictors and a trade-off between model size and accuracy as measured 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:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i \tag{4}$$

where y_i is the i^{th} predictand response, β_j is the j^{th} regression coefficient; β_0 is a constant, and 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

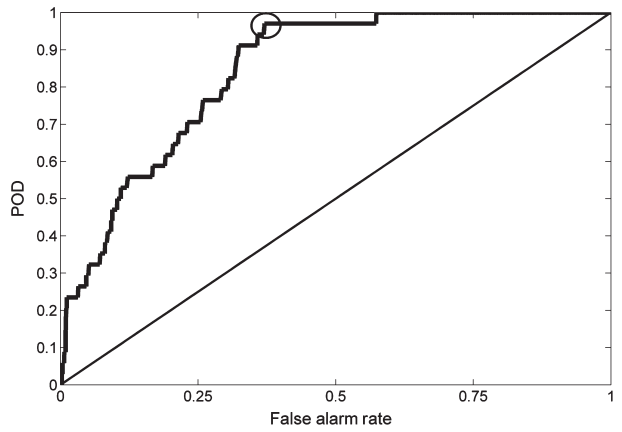

Figure 6. Determination of model threshold example. The curve depicts one of the 50 sets of ROC curves generated, and corresponding threshold determination, on the training sample in connection with the development and evaluation of the TANN model with 43-20-1 topology (Box 73 15 hour prediction). After the TANN 43-20-1 model was trained and validated, ROC curves are generated. The solid curve is the ROC curve based on the training portion of the training sample; the centre of the circle depicts the point on the ROC curve where PSS is maximized. The threshold chosen corresponds to the point on the ROC curve where PSS is maximized. The threshold is combined with the trained model to make predictions on the 2007–2008 independent testing set.

Table 9. Contingency matrix from which scalar performance metrics (Table 10) are derived.

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 10. Scalar performance metrics to evaluate TANN, TANN-MOS, TAFs and NDFD.

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	$(a-a_r)/(a+b+c-a_r)$; $a_r=(a+b)(a+c)/n$

which involved the splitting of the training sample into training and validation data sets. Hence, MLR model calibration had the advantage of a greater amount of training data than used in ANN model calibration. The aviation forecaster outputs are from selected Terminal Aerodrome Forecast (TAF) products issued by WFO CRP. TAFs are specialized 24 or 36 h forecasts (at select airport locations) of weather variables/parameters of critical importance to the aviation community (e.g. Aviation Weather Center, 2011).The public forecasts are obtained from the NWS national digital forecast database (NDFD) (Meteorological Development Laboratory, 2012).

The performance evaluation for the TANN models, the TAFs (aviation forecasts) and the NDFD (public forecasts) for the period 2007–2008 is summarized in Tables 11–14. Tables 11–13 depict the performance results with respect to the TANN and MLR models, and the forecasters; each TANN model is compared to the corresponding MLR model and NDFD (same box and prediction hour). For each skill metric, the Wilcoxon signed rank test (single sample, two sided) is performed to determine whether the median of the 50 trial runs is statistically significantly different (5% significance level) from the corresponding MLR or NDFD value. Table 14 compares the performance between the 18 TANN models. For each skill metric, the two-sided two-sample and no-pairing Wilcoxon rank sum test (Mann–Whitney) was applied to test whether the location of the distributions between corresponding TANN models (same box and prediction hour) differ by zero (5% significance level). The Wilcoxon tests were performed *via* the wilcox.exact function within the exactRankTests package as part of the R programming language (R Core Team, 2014). The results are mixed and are summarized below.

Examination of the Tables 11–14 inclusive, shows that several key observations are noteworthy. For all three box regions, the 15 h NDFD forecasts were vastly superior to that of the corresponding MLR and TANN model predictions; with respect to the corresponding differences for the 9 and 12 h predictions, there is no clear trend. Nevertheless, in 4 of the 9 box h^{-1} case combinations with respect to at least one skill performance metric, the TANN *X*-*Y*-1 models performed superior to the MLR models (both Box 73 12 h and Box 238 9 and 12 h with respect to ORSS, and Box 73 9 h with respect to all skill metrics), and in 2/9 of the cases, the TANN *X*-*Y*-1 models exceeded NDFD skill (Box 103 12 h with respect to ORSS, and Box 238 12 h with respect to PSS and ORSS). In 2/9 of the cases, the TANN 43-*Y*-1 models were more skillful than the MLR (Box 73 15 h with respect to all skill metrics, and Box 238 12 h with respect to HSS and ORSS), and in only one case (1/9), a TANN 43-*Y*-1 model was more skillful than the NDFD (Box 238 12 h with respect to PSS and ORSS). Note that there is no performance enhancement of the TANN models beyond the MLR with respect to Box 103; the same is true for the NDFD, except for case Box 103 12 h. Results are mixed with respect to the comparison between the TANN model variants. In particular, in 1/3 of the nine box/prediction hour case combinations with respect to at least one skill performance metric, the TANN *X*-*Y*-1 models performance was superior to that of the TANN 43-*Y*-1 models (Box 73 9 and Box 103 12 h cases, both with respect to all skill-based performance metrics, and Box 238 9 h with respect to ORSS), while in 1/3 of the cases, the opposite is also true (Box 73 15 h with respect to all skill performance metrics, and Box 238 both 12 h and 15 h with respect to CSI, HSS, and GSS). Further, note that for $X < 4$ with respect to the TANN *X*-*Y*-1 models, performance is generally worse than that of the corresponding (same box and prediction hour) TANN 43-*Y*-1 models; otherwise, performance of the corresponding TANN *X*-*Y*-1 models is either statistically similar or superior to that of the TANN 43-*Y*-1 models. When considering the influence of the sub-grid scale features, it must be noted that the Box 103 9 h case was the only case wherein the features chosen *via* CFS for the TANN *X*-*Y*-1 models included sub-grid scale parameters. Yet, for the Box 103 9 h case, the TANN models performed worse than that of the MLR models and NDFD. However, as noted in Section 2.1.4, the influence of sub-grid

Table 11. Performance results of TANN *X-Y-Z* models (Section 3; Tables 7 and 8) for *Box 238* for the 2007–2008 independent data set and corresponding comparisons to the WFO CRP forecasters (NDFD), multi-linear regression (MLR) models (Section 2.1.4), and to the Terminal Aerodrome Forecasts (TAF) for the Victoria Regional Airport (VCT) issued by WFO CRP forecasters.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 8-7-1	0.96	0.82	0.38	0.57[c]	0.18[a]	0.20[c]	0.95[a]	0.18[a]	0.11[c]
TANN 43-1-1	0.91	0.81	0.35	0.56[c]	0.19[a]	0.21[a]	0.90[c]	0.18[a]	0.12[a]
MLR	0.91	0.79	0.32	0.60	0.20	0.23	0.92	0.20	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
9–12 h Operational aviation forecasts									
TAF	0.38	0.83	0.07	0.32	0.14	0.20	0.80	0.15	0.11
12 h Model predictions									
TANN 4-15-1	0.81	0.93	0.39	0.43[b]	0.07[c]	0.07[b]	0.76[c]	0.06	0.04[c]
TANN 43-15-1	0.81	0.93	0.35	0.43[b]	0.07	0.08[a]	0.75[c]	0.06	0.04
MLR	0.75	0.93	0.33	0.42	0.07	0.07	0.71	0.06	0.04
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 2-2-1	0.73	0.97	0.52	0.21[c]	0.03[c]	0.02[c]	0.43[c]	0.02[c]	0.01[c]
TANN 43-20-1	0.55	0.97	0.35	0.24[c]	0.03[c]	0.03[c]	0.48[c]	0.02[c]	0.01[c]
MLR	0.82	0.96	0.41	0.41	0.04	0.04	0.73	0.03	0.02
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 (Section 4). The superscript values a, b, and c denote TANN *X-Y-Z* median values of skill-based metrics (PSS, CSI, HSS, ORSS, CSS and GSS only) statistically significantly different (based on the Wilcoxon sign rank tests, two-sided, one sample, 5% significant level) from the corresponding MLR, NDFD, or both (MLR and NDFD) values, respectively.

Table 12. Same as Table 11 except for *Box 103* and for Terminal Aerodrome Forecasts (TAF) issued for the Corpus Christi International Airport (CRP).

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 12-30-1	0.70	0.89	0.32	0.39[c]	0.11[c]	0.11[c]	0.68[c]	0.09[c]	0.06[c]
TANN 43-8-1	0.70	0.88	0.29	0.40[c]	0.11[c]	0.12[c]	0.70[c]	0.10[c]	0.07[c]
MLR	0.73	0.86	0.25	0.48	0.13	0.16	0.78	0.12	0.09
9 h Operational public forecasts									
NDFD	1.00	0.85	0.31	0.69	0.15	0.19	1.00	0.15	0.10
9–12 h Operational aviation forecasts									
TAF	0.19	0.93	0.04	0.15	0.05	0.08	0.69	0.06	0.04
12 h Model predictions									
TANN 6-8-1	0.90	0.95	0.34	0.52[c]	0.05[c]	0.06[c]	0.87[c]	0.05[c]	0.03[c]
TANN 43-30-1	0.60	0.95	0.28	0.35[c]	0.04[c]	0.05[c]	0.65[c]	0.03[c]	0.02[c]
MLR	0.90	0.93	0.27	0.63[c]	0.07[c]	0.09[c]	0.92[c]	0.06[c]	0.05[c]
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 3-12-1	0.67	0.97	0.24	0.40[c]	0.03[c]	0.03[c]	0.71[c]	0.02[c]	0.02[c]
TANN 43-35-1	0.50	0.97	0.20	0.37[c]	0.03[c]	0.03[c]	0.69[c]	0.02[c]	0.02[c]
MLR	1.00	0.96	0.27	0.73	0.04	0.05	1.00	0.04	0.03
15 h Operational public forecasts									
NDFD	1.00	0.96	0.19	0.81	0.04	0.07	1.00	0.04	0.04

scale parameters on the target is already explained by CP. The performance of a portion of the TANN models may be explained by the number of lightning cases available; the vast majority of the cases involved cases without lightning, suggesting that the ability to train TANN models that generalize well require a sufficient number of lightning cases. Table 15 depicts an estimate of the number of CG lightning cases available both for training the TANN 43-*Y*-1 models and for evaluation of the trained models on the novel data set. Note that for the Box 103 15 h case, only a paucity of lightning cases were available for model training which may account for the sub-par performance of the TANN models relative to the MLR model and NDFD.

5. Summary and conclusions

The authors developed a feed-forward multi-layer perceptron artificial neural network (MLP ANN) to predict thunderstorm occurrence – within three 400 km^2 domains – 9, 12 and 15 h in advance and with temporal accuracies of 4 h. The model framework involves the use of predictors that account for both

Table 13. Same as Table 11 except for *Box 73* and for Terminal Aerodrome Forecasts (TAF) for the Laredo International Airport (LRD).

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
9 h Model predictions									
TANN 10-25-1	0.91	0.83	0.28	0.62[c]	0.16[a]	0.20[c]	0.92[a]	0.16[a]	0.11[c]
TANN 43-25-1	0.86	0.84	0.28	0.57[b]	0.16[b]	0.19[b]	0.89[b]	0.15[b]	0.11[b]
MLR	0.88	0.85	0.30	0.58	0.15	0.18	0.89	0.14	0.10
9 h Operational public forecasts									
NDFD	0.91	0.83	0.26	0.65	0.16	0.21	0.93	0.16	0.12
9-12 h Operational aviation forecasts									
TAF	0.47	0.96	0.09	0.39	0.04	0.06	0.81	0.04	0.03
12 h Model predictions									
TANN 5-1-1	0.91	0.90	0.39	0.52[c]	0.10[c]	0.11[c]	0.88[c]	0.09[c]	0.06[c]
TANN 43-5-1	0.91	0.90	0.36	0.52[c]	0.10[c]	0.11[c]	0.87[b]	0.09[c]	0.06[c]
MLR	0.86	0.88	0.30	0.56	0.12	0.14	0.87	0.11	0.08
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 2-5-1	0.42	0.97	0.34	0.11[c]	0.03[b]	0.01[c]	0.23[c]	0.01[c]	0.01[b]
TANN 43-20-1	0.67	0.96	0.33	0.31[c]	0.04[c]	0.04[c]	0.59[c]	0.03[c]	0.02[c]
MLR	0.42	0.96	0.25	0.17	0.03	0.03	0.37	0.02	0.01
15 h Operational public forecasts									
NDFD	0.85	0.93	0.24	0.61	0.07	0.10	0.89	0.07	0.05

Table 14. Comparison of optimized TANN *X-Y-Z* model performances (see Section 2.3 for meaning of *X, Y* and *Z*) for the 2007–2008 novel independent testing set.

	POD	FAR	F	PSS	CSI	HSS	ORSS	CSS	GSS
Box 238: 9 h predictions									
TANN 8-7-1	0.96	0.82	0.38	0.57	0.18	0.20	**0.95**	0.18	0.11
TANN 43-1-1	0.91	0.81	0.35	0.56	0.19	0.21	0.90	0.18	0.12
Box 238: 12 h predictions									
TANN 4-15-1	0.81	0.93	0.39	0.43	0.07	0.07	0.76	0.06	0.04
TANN 43-15-1	0.81	0.93	0.35	0.43	**0.07**	**0.08**	0.75	0.06	**0.04**
Box 238: 15 h predictions									
TANN 2-2-1	0.73	0.97	0.52	0.21	0.03	0.02	0.43	0.02	0.01
TANN 43-20-1	0.55	0.97	0.35	0.24	**0.03**	**0.03**	0.48	0.02	**0.01**
Box 103: 9 h predictions									
TANN 12-30-1	0.70	0.89	0.32	0.39	0.11	0.11	0.68	0.09	0.06
TANN 43-8-1	0.70	0.88	0.29	0.40	0.11	0.12	0.70	0.10	0.07
Box 103: 12 h predictions									
TANN 6-8-1	0.90	0.95	0.34	**0.52**	**0.05**	**0.06**	**0.87**	**0.05**	**0.03**
TANN 43-30-1	0.60	0.95	0.28	0.35	0.04	0.05	0.65	0.03	0.02
Box 103: 15 h predictions									
TANN 3-12-1	0.67	0.97	0.24	0.40	0.03	0.03	0.71	0.02	0.02
TANN 43-35-1	0.50	0.97	0.20	0.37	0.03	0.03	0.69	0.02	0.02
Box 73: 9 h predictions									
TANN 10-25-1	0.91	0.83	0.28	**0.62**	**0.16**	**0.20**	**0.92**	**0.16**	**0.11**
TANN 43-25-1	0.86	0.84	0.28	0.57	0.16	0.19	0.89	0.15	0.11
Box 73: 12 h predictions									
TANN 5-1-1	0.91	0.90	0.39	0.52	0.10	0.11	0.88	0.09	0.06
TANN 43-5-1	0.91	0.90	0.36	0.52	0.10	0.11	0.87	0.09	0.06
Box 73: 15 h predictions									
TANN 2-5-1	0.42	0.97	0.34	0.11	0.03	0.01	0.23	0.01	0.01
TANN 43-20-1	0.67	0.96	0.33	**0.31**	**0.04**	**0.04**	**0.59**	**0.03**	**0.02**

Each value represents the median of 50 separate model runs on the 2007–2008 independent testing set. The TANN *X-Y-Z* median values (PSS, CSI, HSS, ORSS, CSS and GSS only) both statistically significantly different (two-sided, two-sample, no pairing Wilcoxon rank sum/Mann–Whitney test, 5% significance level) and greater than its counterpart for the same box/prediction hour are denoted in boldface.

the mesoscale environmental conditions conducive to convective development, and high resolution soil moisture magnitude and heterogeneity that can trigger individual convective cells, especially within a synoptically benign environment. Two sets of nine TANN models each were calibrated, one set using the entire 43 variable predictor set, while each model from the other set was developed based on a unique subset of predictor inputs determined based on correlation-based feature selection (CFS) as a filtering-based feature selection technique. An assessment of TANN model performance was conducted by comparisons to operational forecasters, multi-linear regression (MLR) models, and by inter-comparison amongst the TANN models. For

Table 15. Number of CG lightning cases used in training (80% of training sample) of the TANN 43-*Y*-1 models; corresponding number of CG lightning cases used computing performance on the 2007–2008 novel data set in parentheses.

Prediction hour	Box 73	Box 103	Box 238
9	74 (33)	109 (30)	138 (47)
12	72 (22)	53 (10)	77 (16)
15	50 (12)	30 (6)	40 (11)

the majority of the 9 box per prediction hour combination with respect to at least one skill-based performance metric, the MLR and NDFD were superior to that of the TANN models. In only one case was a TANN model more skillful than the corresponding NDFD, suggesting that the TANN models did not overcome the expertise of the human forecasters. Nevertheless, the greater skill of the TANN models in a minority of cases, and similar performance for several other cases, demonstrates at least some utility and the potential for an automated thunderstorm predictions method. The utility of the sub-grid scale parameters to thunderstorm prediction may have been muted by Betts-Miller-Janjic (BMJ) convective parameterization physics option in the North American Mesoscale (NAM). Comparison amongst the TANN *X*-*Y*-1 and TANN 43-*Y*-1 models suggest utility in the use of the CFS feature selection technique. Yet, the benefit of feature selection became apparent only when ≥4 features were chosen, suggesting that CFS may be too aggressive at times in reducing the number of features. Further, the number of lightning cases in the model training and evaluation data sets may have been insufficient in some cases.

These mixed TANN model performance results reflect the difficulty faced with regard to the development of a skillful data driven model. Recall that a major constraint was placed on the TANN classifiers, to predict thunderstorm occurrence in a binary fashion 9, 12 and 15 h in advance within a somewhat small region of 400 km^2. The desire to predict thunderstorms on these spatial and temporal scales is ambitious given the constraints of a chaotic atmosphere, which limits predictability. Nevertheless, adjustments to the design of this study may improve TANN model performance. In particular, recall that the MLR models had the advantage of being trained on more data. Although the MLR was calibrated on the entire training sample, recall from Section 3 that the TANN models were trained on 80% of the training sample. Further, although the 43 potential predictor variables have relevance to the convective initiation (CI) problem, an expansion to the potential predictor set may be warranted. For example, it may be beneficial to account for atmospheric conditions in adjacent boxes because convective development in adjacent boxes can propagate into a given box in question. Also, the authors attempted to incorporate aerosol optical depth (AOD) as a predictor given the relationship between aerosol concentration and convective updraft magnitude (van den Heever *et al.*, 2006) and density of cloud-to-ground (CG) lightning strikes (Steiger *et al.*, 2002). Yet, the sparseness of the data precluded inclusion.

With respect to operational implementation of the TANN, a given operational forecast entity can generate a forecast grid (similar to Figure 1) using Geographic Information System (GIS) techniques. The NAM output is available to both NWS forecasters and to the general public with a lag time of ∼3 h. Both archived and quasi-real time Multi-sensor precipitation estimator (MPE) data can be obtained from WGRFC in order to perform the soil moisture content (SMC) calculations (spin-up and daily adjustments) *via* the antecedent precipitation index (API) model (Appendix S1) followed by calculations of the other sub-grid scale variables; recall that the SMC calculations are based on the previous day (Table 5) and, hence, would be available. Operational forecast offices should determine their own set of potential predictors and then use feature selection to determine the unique set of high relevance and low redundant predictors. MATLAB® software can be used to perform feature selection, and then train, validate, test and implement an operational TANN.

Research is ongoing, including the plan to greatly enhance the amount of target data by altering the TANN framework to develop TANN models that will be trained across all 286 boxes rather than develop one TANN model *per* box. The authors speculate that the greater the amount of data, especially lightning cases, *per* TANN model will allow for more complex ANN topologies, or simply decrease the influence of the curse of dimensionality, to improve TANN model skill.

Acknowledgements

This study represents the outgrowth of a project initially funded by a 2005 United States Department of Commerce *Pioneer Grant*. A number of individuals and organizations provided invaluable assistance. We sincerely thank Texas A&M University-Corpus Christi (TAMUCC) Professor Rick Smith for the application of GIS (Geographical Information Systems) expertise to generate Figure 1 with minimal error. Anthony Reisinger (TAMUCC) provided enhancements to Figure 1. We recognize the contribution of Sergey Reid and Julien Clifford (TAMUCC) who created Figure 4. We are grateful to Rick Hay (TAMUCC Center for Water Supply Studies), Arthur Taylor (NOAA), Ingo Bethke and Niall Durham (TAMUCC) for providing the computer software necessary for much of the data processing. Further, Valery Dagostaro (NOAA/NWS), Irv Watson (NOAA/NWS), Robert Rozumalski (NOAA), Dan Swank (NCDC/NOMADS) provided selected source data sets. Finally, we thank Whitney Rutledge (TAMUCC), Alok Sahoo (Land Surface Hydrology Group – Princeton University), Matthew Grantham (NOAA/NWS), and Arthur Taylor for assistance with regard to the quality control of select calculations.

Supporting information

The following material is available as part of the online article:

Appendix S1. Soil moisture content calculation.
Appendix S2. Calculation of parameters derived from NAM output.
Appendix S3. Relevant information regarding NLDN, NWP Models, NDFD and TANN performance comparisons with NDFD and aviation forecasts.
Appendix S4. Calculation of global Moran's I spatial autocorrelation proxy.
Figure S1. Comparison of the mean 4 km soil moisture content (SMC) from the API model (solid black line) to that from the mean 1/8th resolution LSMEM SMC (grey), for Box 238.
Figure S2. Same as in Figure S1, except for Box 73.

References

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory: 267-81*, Petrov BN, Csaki F (eds). Akademiai Kiado: Budapest.
Arakawa A. 2004. The cumulus parameterization problem: past, present, and future. *J. Clim.* **17**: 2493–2525.

Arizona State University. 2014. Feature selection at Arizona State University. Arizona State University. http://featureselection.asu.edu (accessed 26 June 2014).
Aviation Weather Center. 2011. TAF Decoder. DOC/NOAA/NWS/AWC. http://aviationweather.gov/static/help/taf-decode.php (accessed 16 June 2013).
Avissar R, Liu Y. 1996. Three-dimensional numerical study of shallow convective clouds and precipitation induced by land surface forcing. *J. Geophys. Res.* **101**: 7499–7518.
Avissar R, Schmidt T. 1998. An evaluation of the scale at which ground-surface heat flux patchiness affects the convective boundary layer using large-eddy simulation. *J. Atmos. Sci.* **55**: 2666–2689.
Beale R, Jackson T. 1990. *Neural Computing: An Introduction*. Institute of Physics Publishing: London.
Bellman R. 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press: Princeton, NJ.
Betts AK, Ball JH. 1998. FIFE surface climate and site-average dataset 1987–89. *J. Atmos. Sci.* **55**: 1091–1108.
Bishop CM. 2005. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc.: New York, NY.
Bryan GH, Wyngaard JC, Fritsch JM. 2003. Resolution requirements for the simulation of deep moist convection. *Mon. Weather Rev.* **131**: 2394–2416.
Bua B, Jascourt S. 2009. University Corporation for Atmospheric Research/Cooperative Program for Meteorological Education and Training: How Models Produce Precipitation and Clouds – Version 2. http://www.meted.ucar.edu (accessed 1 September 2013).
Chaudhuri S. 2010. Convective energies in forecasting severe thunderstorms with one hidden layer neural net and variable learning rate back propagation algorithm. *Asia Pac. J. Atmos. Sci.* **46**: 173–183.
Colquhoun JR. 1987. A decision true method of forecasting thunderstorms serve a thunderstorms and tornados. *Weather Forecast.* **2**: 337–345.
Costello RB. 1992. *Random House Webster's College Dictionary*. Random House, Inc.: New York, NY.
Crook NA. 1996. Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Weather Rev.* **124**: 1767–1785.
Curran EB, Holle RL, López RE. 2000. Lightning casualties and damages in the United States from 1959 to 1994. *J. Clim.* **13**: 3448–3464.
Dalu GA, Pielke RA, Baldi M, Zeng X. 1996. Heat and momentum fluxes induced by thermal inhomogeneities with and without large-scale flow. *J. Atmos. Sci.* **53**: 3286–3302.
de Silva CW. 2000. *Intelligent Machines: Myths and Realities*. CRC Press LLC: Boca Raton, FL.
Dey CH. 1998: *GRIB: The WMO Format for the Storage of Weather Product Information and the Exchange of Weather Product Messages in Gridded Binary Form as Used by NCEP Central Operations*. NWS Office Note 388. NOAA/NWS, Washington, DC.
Ducrocq V, Tzanos D, Sénési S. 1998. Diagnostic tools using a mesoscale NWP model for the early warning of convection. *Meteorol. Appl.* **5**: 329–349.
Elmore KL, Stensrud DJ, Crawford KC. 2002. Explicit cloud-scale models for operational forecasts: a note of caution. *Weather Forecast.* **17**: 873–884.
Emori S. 1998. The interaction of cumulus convection with soil moisture distribution: an idealized simulation. *J. Geophys. Res.* **103**(D8): 8873–8884.
Fabry F. 2006. The spatial variability of moisture in the boundary layer and its effect on convective initiation: project-long characterization. *Mon. Weather Rev.* **134**: 79–91.
Findell KL, Eltahir EAB. 2003a. Atmospheric controls on soil moisture-boundary layer interactions: Part I: Framework development. *J. Hydrometeorol.* **4**: 552–569.
Findell KL, Eltahir EAB. 2003b. Atmospheric controls on soil moisture-boundary layer interactions: three-dimensional wind effects. *J. Geophys. Res.* **108**(D8): 8385, DOI: 10.1029/2001JD001515.
Fowle MA, Roebber PJ. 2003. Short-range (0–48 h) numerical predictions of convective occurrence, model and location. *Weather Forecast.* **18**: 782–794.
Frye JD, Mote TL. 2010. Convection initiation along soil moisture boundaries in the Southern Great Plains. *Mon. Weather Rev.* **138**: 1140–1151.
Glickman TS. 2000. *Glossary of Meteorology*, 2nd edn. American Meteorological Society: Boston, MA; 855.
Hagan MT, Demuth HB, M B. 1996. *Neural Network Design*. International Thomson Publishing Inc: Boston, MA.
Haklander AJ, Van Delden A. 2003. Thunderstorm predictors and their forecast skill for the Netherlands. *Atmos. Res.* **67–68**: 273–299.
Hall MA, Smith LA. 1999. 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, AIII Press: Orlando, FL.
Haykin S. 1999. *Neural Networks, A Comprehensive Foundation*, 2nd edn. Prentice Hall: Upper Saddle River, NJ.
Hogan RJ, Ferro CAT, Jolliffe IT, Stephenson DB. 2010. Equitability revisited: why the "equitable threat score" is not equitable. *Weather Forecast.* **25**: 710–726.
Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**: 359–366.
Janjić ZI. 1994. The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* **122**: 927–945.
Janjic ZI, Gerrity JP, Nickovic S. 2001. An alternative approach to nonhydrostatic modeling. *Mon. Weather Rev.* **129**: 1164–1178.
Jolliffe IT, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons Ltd: West Sussex; 240.
Kalnay E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press: Cambridge, UK.
Khairoutdinov M, Randall D. 2006. High-resolution simulation of shallow-to-deep convection transition over land. *J. Atmos. Sci.* **63**: 3421–3436.
Lee RR, Passner JE. 1993. The development and verification of TIPS: an expert system to forecast thunderstorm occurrence. *Weather Forecast.* **8**: 271–280.
Lewis P, Toth G. 2011. University Corporation for Atmospheric Research/Cooperative Program for Meteorological Education and Training: Ten Common NWP Misconceptions. http://meted.ucar.edu/norlat/tencom (accessed 9 December 2011).
Longley PA, Goodchild MF, Maguire DJ, Rhind DW. 2005. *Geographic Information Systems and Science*, 2nd edn. John Wiley & Sons, Ltd: West Sussex; 517.
Lynn BH, Tao W, Abramopoulos F. 2001. A parameterization for the triggering of landscape-generated moist convection. Part I: Analysis of high-resolution model results. *J. Atmos. Sci.* **58**: 575–592.
McCann DW. 1992. A neural network short-term forecast of significant thunderstorms. *Weather Forecast.* **7**: 525–534.
McNulty RP. 1981. A statistical approach to short-term thunderstorm outlooks. *J. Appl. Meteorol.* **20**: 765–771.
Madden RA, Robitaille FE. 1970. A comparison of the equivalent potential temperature and the static energy. *J. Atmos. Sci.* **27**: 327–329.
Manzato A. 2005. The use of sounding-derived indices for a neural network short-term thunderstorm forecast. *Weather Forecast.* **20**: 896–917.
Manzato A. 2007. A note on the maximum Peirce skill score. *Weather Forecast.* **22**: 1148–1154.
MathWorks. 2014. *Matlab: The Language of Technical Computing*. The Mathworks Inc.: Natick, MA.
May R, Dandy G, Maier H. 2011. Review of input variable selection methods for artificial neural networks. In *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, Suzuki K (ed). InTech Europe: Rijeka; 19–44.
Mitchell K. 2005. The Community NOAH Land-Surface Model (LSM), DOC/NOAA/NWS/NCEP/EMC. http://www.emc.ncep.noaa.gov/mmb/gcp/noahlsm/Noah_LSM_USERGUIDE_2.7.1.htm (accessed 15 October 2012).
Moller M. 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **6**: 525–533.
Orlanski I. 1975. A rational subdivision of scales for atmospheric processes. *Bull. Am. Meteorol. Soc.* **56**: 527–530.
Orville RE. 2008. Development of the national lightning detection network. *Bull. Am. Meteorol. Soc.* **89**: 180–190.
National Geodetic Survey. 2006. Inverse/Forward/Inverse3D/Forward3D Computation Utilities, DOC/NOAA/NOS/NGS. http://www.ngs.noaa.gov/TOOLS/Inv_Fwd/Inv_Fwd.html (accessed 9 September 2005).
Perler D, Marchand O. 2009. A study in weather model output postprocessing: using the boosting method for thunderstorm detection. *Weather Forecast.* **24**: 211–222.
Pielke RA. 2002. *Mesoscale Meteorological Modeling. International Geophysics Series*, Vol. **78**. Academic Press: San Diego, CA; 676.
R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org (accessed 4 December 2014).
Reap RM, Foster DS. 1979. Automated 12–36 hour probability forecasts of thunderstorms and severe local storms. *J. Appl. Meteorol.* **18**: 1304–1315.

Rogers E, Black TL, Deaven DG, DiMego GJ. 1996. Changes to the operational "early" Eta analysis/forecast system at the National Centers for Environmental Prediction. *Weather Forecast.* **11**: 391–413.

Ross BB. 1986. An overview of numerical weather prediction. In *Mesoscale Meteorollogy and Forecasting*. American Meteorological Society: Boston, MA; 793.

Rotunno RJ, Klemp B, Weisman ML. 1988. A theory for strong, long-lived squall lines. *J. Atmos. Sci.* **45**: 464–485.

Rumelhart DE, McClelland JL. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, Vol. **1**. MIT Press: Cambridge, MA.

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

Saunders CPR. 1993. A review of thunderstorm electrification processes. *J. Appl. Meteorol.* **32**: 642–655.

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

Steiger SM, Orville RE, Huffines G. 2002. Cloud to-ground lightning characteristics over Houston, Texas: 1989–2000. *J. Geophys. Res.* **107**: D11, DOI: 10.1029/2001JD001142.

Swets JA. 1973. The relative operating characteristic in psychology. *Science* **182**: 990–1000.

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

Taylor CM, Lebel T. 1998. Observational evidence of persistent convective-scale rainfall patterns. *Mon. Weather Rev.* **126**: 1597–1607.

Taylor CM, Parker DJ, Harris PP. 2007. An observational case study of mesoscale atmospheric circulations induced by soil moisture. *Geophys. Res. Lett.* **34**: L15801, DOI: 10.1029/2007GL030572.

Texas Parks & Wildlife. 2011. Level III Ecoregions of Texas. Texas Parks & Wildlife Department GIS Lab. http://www.tpwd.state.tx.us/publications/pwdpubs/media/pwd_mp_e0100_1070z_08.pdf (accessed 13 April 2013).

Trier SB. 2003. Convective storms - convective initiation. In *Encyclopedia of Atmospheric Sciences*, Holton JR (ed). Academic Press: Oxford University Press: New York, NY; 560–570.

Van den Heever SC, Carrió GG, Cotton WR, DeMott PJ, Prenni AJ. 2006. Impacts of nucleating aerosol on Florida storms. Part I: Mesoscale simulations. *J. Atmos. Sci.* **63**: 1752–1775.

Vincenty T. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Surv. Rev.* **22**: 88–93.

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

Weisman ML, Davis C, Wang W, Manning KW, Klemp JB. 2008. Experiences with 0-36-h explicit convective forecasts with the WRF-ARW model. *Weather Forecast.* **23**: 407–437.

Weisman LM, Skamarock WC, Klemp JB. 1997. The resolution dependence of explicitly modeled convective systems. *Mon. Weather Rev.* **125**: 527–548.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2 edn. Elsevier: Oxford.

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

Wilson JW, Crook NA, Mueller CK, Sun J, Dixon M. 1998. Nowcasting thunderstorms: a status report. *Bull. Am. Meteorol. Soc.* **79**: 2079–2099.

Wilson JW, Schreiber WE. 1986. Initiation of convective storms at radar-observed boundary-layer convergence lines. *Mon. Weather Rev.* **114**: 2516–2536.

Zhang QJ, Gupta KC, Devabhaktuni VK. 2003. Artificial neural networks for RF and microwave design: from theory to practice. *IEEE Trans. Microw. Theory Tech.* **51**: 1339–1350.

