Tuning AutoNowCaster Automatically

Valliappa Lakshmanan[1,2], John Crockett[3,5], Kenneth Sperow[3,4],

Mamoudou Ba[3], Lingyan Xin[3*]

*Corresponding author: V Lakshmanan, 120 David L. Boren Blvd, Norman OK 73072; lakshman@ou.edu

[1]Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma; [2]National Oceanic and Atmospheric Administration / National Severe Storms Laboratory; [3]National Weather Service Meteorological Development Laboratory; [4] Cooperative Institute for Research in the Atmosphere, Colorado State University; [5] Wyle Information Systems

ABSTRACT

AutoNowCaster (ANC) is an automated system that nowcasts thunderstorms, including thunderstorm initiation. However, its parameters have to be tuned to regional environments, a process that is time-consuming, labor-intensive and quite subjective. When the National Weather Service decided to explore using ANC in forecast operations, a faster, less labor-intensive and objective mechanism to tune the parameters for all the forecast offices was sought.

In this paper, a genetic algorithm approach to tuning ANC is described. The process consisted of choosing data sets, employing an objective forecast verification technique and devising a fitness function. ANC was modified to create nowcasts offline using weights iteratively generated by the genetic algorithm. The weights were generated by probabilistically combining weights with good fitness, leading to better and better weights as the tuning process proceeded.

The nowcasts created by ANC using the automatically determined weights are compared with the nowcasts created by ANC using weights that were the result of manual tuning. It is shown that nowcasts created using the automatically tuned weights are as skilled as the ones created through manual tuning. In addition, automated tuning can be done in a fraction of the time that it takes experts to analyze the data and tune the weights.

1. Introduction

High quality nowcasts of thunderstorms have the potential to be a tremendous benefit to the general public. Properly integrated as a critical impact factor into management of the National Air Space, they could help reduce the lengthy air traffic delays commonly experienced during the spring and summer in the United States. According to economic statistics published by NOAA[1], 70 percent of air traffic delays are attributed to weather.

Since 2006, the National Weather Service (NWS), in collaboration with the National Center for Atmospheric Research (NCAR), has been testing an automated system for nowcasting thunderstorms at the Dallas/Fort Worth Weather Forecast Office (WFO), hereafter referred to as FWD. The system, known as "AutoNowCaster" or ANC (Mueller et al. 2003; Wilson and Mueller 1993; Wilson et al. 1998), was also recently installed at the Melbourne, FL WFO.

ANC uses a fuzzy logic algorithm based on conceptual models of storm initiation, storm growth, and storm dissipation. The system assimilates a variety of datasets to analyze characteristic features of the atmosphere associated with pre-storm environments, and to produce 60-minute nowcasts of storm initiation, growth and dissipation. These analyses include evaluation of convective instability, moisture convergence, and trigger mechanisms to produce interest fields (Table 1) that are used as inputs to the fuzzy logic algorithm. The interest fields used in the storm initiation algorithm are converted into dimensionless likelihood fields using fuzzy membership functions. These likelihood fields have a dynamic range from -1 to 1 with increasing positive values used to indicate regions of increasing likelihood of storm initiation. The various likelihood fields are weighted using values determined by human experts, and the weighted likelihood fields are summed to produce a combined likelihood field that is then filtered and smoothed. Regions with values greater than 0.7 in the combined filtered and smoothed likelihood field are indicative of storm initiation in the next 60 minutes. An

[1] www.weather.gov/com/2004_economic_statistics1.pdf

example of an ANC nowcast of convective initiation is shown in Figure 2c. For a detailed description of the AutoNowCaster system, the reader is directed to Mueller et al. (2003); Wilson and Mueller (1993).

Forecasters guide ANC by drawing boundaries along which convection is anticipated and by selecting a convective regime such as cold front or dry line. For each regime it is possible to have a different set of predictor fields, membership functions and weights.

ANC uses an idealized conceptual framework to tune the weights of the various interest fields used to nowcast storm initiation for each convective regime. The tuning is done manually using a limited set of cases by visually examining results for each regime to determine if the predictor fields are applicable or in need of adjustment. The time required for an expert to visually inspect the nowcasts and up to 17 different predictor fields for a representative sample of data for each regime makes manually tuning ANC impractical. Additionally, it is unrealistic to assume that an expert will be able to come up with the optimum set of weights because convective initiation is fairly complex and representative datasets are relatively large.

The NWS is currently exploring a concept of operations for ANC with the goal of providing valuable thunderstorm nowcasts to the aviation community and the public. To enable deployment of ANC nationwide, an automatic way of tuning ANC is necessary. In the rest of this paper, we describe the development of an automated tuning mechanism for ANC. This approach described here may be applicable to the tuning of other, multi-parameter, complex systems for operational uses.

a. Genetic Algorithm

A genetic algorithm (Goldberg 1989) is a search-and-optimization technique that is built to mimic the process of Darwinian evolution by modeling processes such as inheritance, mutation, selection and crossover. Genetic algorithms have been widely used in meteorology to find breakpoints of fuzzy functions (Lakshmanan 2000), to validate dispersion models (Haupt et al. 2006), to optimize mesoscale models (O'Steen and Werth 2009) and to find consensus forecasts from ensembles (Roebber 2010; Bakhshaii and Stull 2009).

Where genetic algorithms excel over traditional optimization techniques is in their ability to handle non-differentiable error functions. Optimization techniques based on gradient descent, for example, require that the error function be a differentiable function of the parameters to be tuned. Thus, gradient descent is a workable solution for backpropagation single-layer neural networks where the error function is often a least squares error and each prediction is a weighted sum of transformed inputs with the transformation being an easily differentiated function such as a logistic exponential function. Genetic algorithms have no such restriction. They can easily handle error functions that are non-linear, non-continuous, and even completely unknown functions of their inputs. This makes genetic algorithms a particularly good choice to tune a "black box" system that only exposes a few tuneable parameters.

In a genetic algorithm, the search is carried out in parallel, with a fixed number of potential solutions evaluated at each step. These potential solutions are termed chromosomes, the iterations are called generations, and the group of chromosomes at a generation is called a population. Thus, a genetic algorithm consists of finding better and better populations of chromosomes after each generation. The chromosomes themselves consist of "genes," which are the tuneable parameters. Instead of an error function being minimized, the formulation is in terms of the chromosome's "fitness" being maximized.

In our genetic algorithm, the weights (in the range 0 to 1) of the different ANC predictor variables are the genes. All the weights of all the predictor variables together form the chromosome. Although it is conceivable that the GA can also be used to tune the breakpoints of the fuzzy membership functions, we did not do so. We retained ANC's "factory" settings for the membership functions, and changed only the weights of the predictor variables in order to attain good nowcasts for all the training cases. Also, although it is possible to tune the weights for all the nowcasts produced by ANC, we concentrate in this paper on the nowcast of convective initiation.

A genetic algorithm improves the fitness of a population by applying Darwinian selection principles to create the population at the next generation. The population at the next generation consists of chromosomes that are formed mostly by crossing over a pair of chromosomes at the current generation. Since the chromosomes are essentially just a list of tuneable parameters, crossover involves merely choosing some parameters from the first chromosome and the remaining parameters from the second one. This choice of parameters is done randomly so that different children of the same pair of chromosomes could be different. Optimization occurs because of how the pair of chromosomes is chosen: the best fit individuals are chosen probabilistically, i.e., if we imagine a pie divided into slices, one for each chromosome, the size of the slices is directly proportional to the fitness of the chromosome (See Figure 1a). Thus, when a pair of chromosomes are randomly chosen, higher-fit individuals are more likely to be chosen, with the likelihood given by the fitness of the chromosome relative to the average fitness of its generation. These two chromosomes are then crossed over, i.e., some genes selected from one chromosome and others from the other chromosome (See Figure 1b), to yield a new chromosome whose fitness can be calculated.

Although crossover is the main mechanism by which the next population of chromosomes is formed, a few other evolutionary principles are added because it has been shown (Goldberg 1989) that inheritance, mutation and diversity improve the performance of a genetic algorithm. Some individuals are formed not by crossing over a pair of chromosomes from the previous generation, but by simply copying over a well-fit individual from the previous generation ("inheritance"). As a special case of inheritance, the best fit member of a population is always retained in the next generation, so as not to lose the best parameters discovered during the search. Mutation is carried out by taking a crossed-over or inherited chromosome and slightly modifying some of its genes (See Figure 1c). This enables the search space to be locally expanded. The average fitness of a population converges rapidly when successive populations are carried out using the evolutionary paradigm. However, there is no guarantee that this convergence is to a global optimum. Therefore, it is usually worthwhile to keep the search space as wide as possible, to take advantage of the parallel local search afforded by a genetic algorithm. Thus, in addition to choosing chromosomes probabilistically based on fitness, a diversity penalty is added so that the size of the slice in the probability pie decreases once a chromosome has been chosen from it. Finally, because the genetic algorithm is guaranteed to converge, but not even to the local maxima, we periodically carried out simulated annealing (Metropolis et al. 1953), a local search and optimization technique, around the population to push each member of the population to its local maximum. Because of this, in Figure 4, one sees all the chromosomes in the population pushed to the local maximum at any generation where simulated annealing is carried out.

In our genetic algorithm implementation for this study, we initialized the population by randomly generating the chromosomes. Each population consisted of 200 chromosomes. The crossover probability was 0.7, the mutation probability was 0.005, and 75% of the population was randomized after simulated annealing, which was carried out every 10 generations. The genetic algorithm process was carried out until the fitness improvement was below 0.001 for 30 generations or for 100 generations.

2. Method

Over the last few years, scientists at NCAR, in close collaboration with FWD, have collected case studies to be used for tuning ANC. The case studies cover the primary convective regimes that affect North Texas during the course of a normal convective season. Some of these cases, as well as the data collected during a five-week Intensive Operations Period conducted at FWD by the NWS's Meteorological Development Laboratory (MDL) between April 19 and May 23, 2010, were used in this study. The cases studied are classified per convective regime and human involvement with ANC as shown in Table 2.

Prior to the implementation of ANC at FWD, the system was running with only one set of fuzzy logic rulesets. The system was modified to allow the forecaster to select one of the multiple logic rulesets that are tailored to different synoptic regimes typically experienced in Texas. Currently, the system is implemented with six convective regimes: the default regime referred to as the mixed regime, cold front, dryline, stationary/warm front, pulse storm, and advecting MCS. The mixed regime served as the basis for the development of the rulesets for all the other regimes. The mixed regime is selected when the forcing of convection is unclear or a variety of forcing is expected within the domain.

a. Forecast Verification

A critical element to iteratively tuning ANC is to determine whether, with a new set of weights, ANC's nowcasts are improved. This is determined by running ANC with the new set of weights and comparing the resulting nowcast fields with ground truth, i.e., with what actually happened 60 minutes later.

Comparing the forecast field with ground truth is known as forecast verification and is an extensively researched issue in meteorology. Sophisticated methods of forecast verification are needed because straightforward pixel-to-pixel measures of error suffer from a double-penalty issue whereby forecasts are unduly penalized for displacement errors. Many of the methods of forecast verification that have been proposed in the literature can be broadly categorized (Gilleland et al. 2009) into filtering-based methods that operate on the neighborhood of pixels (e.g., Ebert (2009)), displacement-based methods that rely on features (e.g., Davis et al. (2006)) and displacement methods that rely on field deformation (e.g., Keil and Craig (2007)). Newer methods such as that of Lakshmanan and Kain (2010) blur these categories somewhat as does the verification method described in this paper.

We need to determine whether ANC's nowcast of convective initiation over the next 60 minutes is correct. This is harder than verifying, say, precipitation forecasts because there is no direct observation of initiation. What we do have are radar reflectivity images that cover ANC's domain (See Figure 2). Images 60 minutes apart have to be examined to find where new convection has happened. We do this by warping the past observation to best align it with the current observation, using a cross-correlation optical flow method (Barron et al. 1994; Wolfson et al. 1999) to determine the warp. Essentially, this involves finding a smooth motion field based on the two images and then advecting the second grid backwards to align it with the first. Once the two images have been aligned, pixels that were below the convective threshold (we used 35 dBZ to fit with ANC's definition of convection) that are now greater than the convective threshold are considered to be convective initiation. However, such a direct pixel-to-pixel match would lead to too many pixels on the boundaries of storms being marked as having initiated. Therefore, we searched in a 5x5 neighborhood (approximately 5km x 5km) and considered a pixel above the convective threshold as having initiated only if there was no above-threshold pixel in the 5x5 neighborhood of this pixel. Using such a distance threshold provides some leeway for small errors in the motion estimate. Thus, the formulation of our truth field involved both warping and neighborhood processing.

After aligning the pair of images, we classified each pixel of the radar image into one of these categories:

- *New Convection*: The pixel in the second image is above the convective threshold and there is no pixel in a 5x5 neighborhood of this pixel in the (aligned) first image that is above the convective threshold.

- *Ongoing Convection*: The pixel in the second image is above the convective threshold but there is at least one pixel in a 5x5 neighborhood of this pixel in the (aligned) first image that is above the convective threshold.

- *Not Convective*: The pixel in the second image is not above the convective threshold.

In Figure 3c, we show four categories, but Decayed convection is lumped along with Not Convective for purposes of verification since our scoring will treat them the same.

This field created by aligning the pair of images and classifying the pixels is termed the verification field.

Once the verification field has been created from pairs of radar reflectivity images spaced 60 minutes apart, a nowcast of convective initiation can be compared against the verification field valid for the time of the nowcast. To do this, we create a contingency table (Wilks 1995) considering every grid point of the nowcast field and classify each pixel in the domain into one of these categories:

- *Don't Care*: A pixel that is Ongoing Convection in the verification field is classified as being one that we don't care about. ANC was neither penalized nor rewarded for categorizing ongoing convection as Convective Initiation (CI).

- *Hit*: The nowcast pixel is CI and there is New Convection in the verification field within a 5x5 neighborhood.

- *False Alarm*: The nowcast pixel is CI but there is no New Convection in the verification field within a 5x5 neighborhood.

- *Miss*: The pixel in the verification field is New Convection and no nowcast pixel in a 5x5 neighborhood is CI.

- *Null:* None of the above categories.

Once the hits, misses, false alarms and nulls are determined, the contingency table is complete and can be used to compute a skill score.

b. Fitness Function

The best ANC weights are those weights that produce good nowcast skill across a diverse set of training cases. Consequently, the skill score was computed in two ways: on a single nowcast basis and on the training set as a whole. The Critical Success Index (CSI Donaldson et al. (1975)), for example, was computed in two ways: by taking into account the hits, misses and false alarms for all the pixels in a single nowcast and by taking the hits, misses and false alarms for all the pixels in all the nowcasts.

The fitness score was defined as:

$$f = 0.3CSI_{all} + 0.3CSI_{avg} + 0.3CSI_{min} + 0.1HSS_{all} \quad (1)$$

where

$$CSI = \frac{hits}{hits + misses + false\_alarms} \quad (2)$$

and CSI_{all} is computed by considering the hits over all the pixels in all the training cases while the CSI_{avg} and CSI_{min} refer to the average and minimum of the CSI computed for each of the nowcasts used for training. While it is possible for the CSI_{all} to be high just by getting a few of the training cases right, the use of CSI_{avg} and, especially, CSI_{min} rewards the genetic algorithm for choosing weights that work well on all the training cases. The CSI is used even though its shortcomings are well-known (See Marzban (1998) for a discussion) because it was the measure of performance used in earlier validation studies of ANC. The CSI by itself is inappropriate for genetic algorithm training because chromosomes with extremely bad parameters will result in CSIs of zero (all of which have no hits) and hence it is not possible to rank the chromosomes at the beginning of the training cycle (when we start with a population of random chromosomes). Therefore, we incorporated the Heidke Skill Score (HSS Heidke (1926)) into our fitness function. The HSS is defined as

$$HSS = \frac{2 * (nulls * hits - misses * false\_alarms)}{(false\_alarms + hits) * (false\_alarms + nulls) + (nulls + misses) * (misses + hits)} \tag{3}$$

and is used so that the nulls play a small role in the verification measure. Since the CSIs are weighted significantly higher, the HSS contributes mostly when CSI is near zero. In such situations, the relative number of false alarms and nulls play a factor in the ranking of chromosomes because of the HSS.

c. Data

The data used to tune ANC came from an archival store of ANC-generated nowcast interest fields kept at NCAR. The geographical area covered by the data includes FWD; the range of dates spanned by the data is 27 August 2006 through 14 May 2010. This domain was chosen because it provides a convenient way to compare the results of automated tuning with the performance of ANC tuned by human experts (Nelson et al. 2005).

Part of the role of the forecaster in ANC is to choose the convective weather regime and to indicate where boundaries are present. ANC convective weather regimes were selected during the days of the forecasts by the forecasters at FWD. The convective weather regimes which ANC allows for selection are cold front (CF), dry line (DL), mesoscale convective system (MC), mixed (MX), pulse storm (PS), and stationary/warm front (SW). It is important to note that the MX regime is meant to be selected when a forecaster either cannot identify the mode of convection or cannot select a clearly dominant mode of convection from multiple modes which are present.

Each of ANC's convective weather regimes is parameterized by the set of nowcast interest fields which are thought to play the greatest roles in the development of convective initiation in that regime. Table 1 provides a description of these interest fields and shows in which regimes each field is a parameter.

Because the input fields differ in spatial resolution, every field is interpolated to the smallest, highest resolution grid among the set of fields, i.e., to the common area covered by all the inputs (1km resolution covering FWD's domain). Temporally, ANC allows for the fact that fields may not be calculated at their expected frequency. For example, failed connectivity to the requisite raw data servers could prevent model output from being available for more than one hour. In such circumstances, ANC uses the closest-in-time prior fields that it can retrieve within allowed, field-dependent maxima of time. Table 1 shows these time-retrieval maxima. If it so happens that a field is both not currently available and not retrievable within the maximum allowed time in the past, no nowcast is generated.

Exploratory tuning scenarios were investigated in order to determine the optimal methodology to use for conducting this study. The final methodology is as follows.

i. The dates of interest were divided into groups; each group represented those dates for which one and only one ANC convective weather regime was considered to be present.

ii. For each date within a group, nowcast times were chosen solely from the time frame of convective initiation or, if that time frame wasn't available, from the time frame of active weather.

iii. From the selected time frame, the nowcast times that both (a) obeyed the relation

$HH + 15\ min <= nowcast\ time < HH + 30\ min$, where HH indicates the top of the hour of the nowcast, and (b) were the earliest times at which all of the fields were available, and none of the fields needed to be retrieved from a prior hour, were chosen. Thus, for any given HH, only a single nowcast time was selected (if at all). The above restrictions are considered valid because multiple exploratory tuning scenarios which differed solely with respect to the number and time distribution of nowcasts in an hour yielded sufficient evidence to conclude that, for the purpose of tuning ANC automatically, only one nowcast per hour was necessary, and nowcasting at or shortly after the first quarter-hour would allow for the latency at which real-time model nowcasts are available.

It is important to note that methodology elements one through three guarantee neither an equal number of nowcast dates per convective weather regime nor an equal number of nowcast times per date. The latter consequence is considered unimportant. However, the former consequence bore on the need to have as consistent a methodology as possible across regimes. The process of tuning requires not only nowcast dates and times with which to tune but also nowcast dates and times to use for independent testing of the tuning's results. Thus, two further restrictions were placed.

iv. For each regime, only three nowcast dates were used for tuning.

v. For each regime, three tuning scenarios were run. Each scenario used two of the three nowcast dates for tuning; the third date was used as the independent control.

The final sets of nowcast dates and times per ANC convective weather regime are shown in Table 2.

For verification purposes, the radar reflectivity field closest in time to the nominal time of the nowcast, i.e., 60 minutes from the nowcast time, was used. The time discrepancy between the nowcast field and the verification was never more than 7 minutes.

The interest fields and verification fields were provided to the genetic algorithm which ran ANC to create a variety of nowcasts, one for each chromosome in the population. Based on the fitness of each set of weights (chromosomes), the next population was created through an evolutionary algorithm. At each generation, the population increased in fitness, as shown in Figure 4. The fittest chromosome after the genetic algorithm converged was chosen as the final set of weights, herein after referred to as the automatically tuned weights.

3. Results

Table 3 summarizes the results from the three tuning scenarios for each convective weather regime. For each scenario, both the nowcast dates used for tuning and the corresponding control date are shown. Alongside these is recorded the final overall fitness of the tuning dates' nowcasts generated by the best-fit regime weights calculated by the genetic algorithm software. For the purpose of comparing the prior, subjectively-tuned, regime-specific weights with the objectively-tuned, regime-specific weights output by the genetic algorithm software, the final overall fitness of the control date's nowcasts was calculated using both sets of weights. The results of these calculations are also shown in the table. The run time of each scenario is also noted. All of the tuning scenarios were run on a Dell PowerEdge R710 server with 32 GB of memory, two 2.93 GHz, hyper-threaded, dual-core Intel Xeon X5570 CPUs and running the CentOS operating system.

From Table 3, it is clear that, for the CF, DL and MC regimes, the objectively-tuned weights yield better nowcasts than do the subjectively-tuned weights. As measured by the fitness values, the range of improvement is between 5 and 110 percent. The MX, PS and SW regimes yielded mixed results. For the MX regime, only one-third of the scenarios appear to yield better nowcasts using the objectively-tuned weights. Also, compared to the fitness of the tuning dates' nowcasts, there is a huge drop-off in the fitness of the second scenario's control date's nowcasts for both the subjectively- and objectively-tuned weights. Such a drop-off is an indication, however, that the control date's data are rather unique and, as such, should be included in the training dates, thus increasing the size of the dataset. It appears that using just two training cases is not enough for the mixed mode, since this category captures a wide variety of "unclassifiable" modes. The PS and SW regimes are similarly constrained; more training cases are needed to capture the full diversity of weather scenarios in these regimes.

A goal of this study is to investigate the sensitivity of the MX regime's weights to the modes of convective initiation, i.e., to determine whether or not a properly tuned MX regime's weights could be used to generate statistically good enough nowcasts for every regime rather than needing to have specific weights for each regime. A driving force behind this avenue of investigation is the idea that, were ANC to be deployed for nationwide use, being able to use a single set of convective initiation weights would be a welcome simplification. Noting again that the selection of the MX regime by a nowcaster indicates either an undetermined (singular) mode of convection or an indeterminate dominant mode of convection among multiple modes, it is to be understood that, unlike the other regimes considered in this study, the MX regime isn't pure. Rather, it represents an amalgamation of the other regimes and, as such, cannot necessarily be tuned in the same manner. With reference to Table 3's first MX regime scenario, it is entirely possible, for example, that the dominant mode of convection on the tuning dates was a dry line, whereas the dominant mode of convection on the control date was a cold front. It could be the case, then, that the objectively-tuned MX weights for this scenario would not nowcast the control date's environment as well as the CF regime's objectively-tuned weights would, because the tuning dates' data would contain no CF-related signal. This leads to the hypothesis that, by using a combination of the pure regimes' data and the MX regime's data to tune the MX regime, the resulting set of weights will nowcast the environments of the pure regimes well enough that we would not need those separate regimes' sets of weights. To test this hypothesis, additional tuning scenarios were created and run.

The first such scenario used all of the MX regime's nowcast dates, the second and third of the CF regime's nowcast dates, the first and second of the DL regime's nowcast dates, the first and third of the MC regime's nowcast dates, and the first and third of the SW regime's nowcast dates in order to tune the MX regime's weights. In this manner, a control date remained for all of the "pure" regimes. Those control dates' nowcasts were then generated using 1) the subjectively-tuned MX weights, 2) the objectively-tuned MX weights previously found by using MX-only data, and 3) the objectively-tuned MX weights found by using the aforementioned combination of MX and "pure" regime data. The final overall fitness of these nowcasts was then calculated. The overall magnitude of the CSI[2] is quite low, but this is a limitation not of the tuning method, but of ANC itself. As noted in Wilson et al. (2010), present-day nowcasting systems do not possess a sufficient level of accuracy to disseminate

[2]The fitness function is dominated by the CSI.

the nowcast to users without human oversight. Instead, they are meant to be employed as decision aids.

From Table 1 it may be noted that the MX regime does not incorporate two of the nowcast interest fields which are used in some of the “pure” regimes. Because the MX regime could be used, however, at times in which such fields might play a role, a second additional tuning scenario was run. This scenario was set up exactly as the first, except that the MX regime was tuned using *all* of the nowcast interest fields. As before, the control dates’ nowcasts were generated using the resulting weights, and the corresponding overall fitness value calculated.

In manual tuning, no interest field is allowed to be weighted less than 0.08. The same criterion was applied to the automated tuning as well. A third additional training scenario was thus run, in which no field’s contribution was allowed to fall below 0.08.

The results of these three additional scenarios are summarized in Table 4. In general, the automatically tuned MX weights generated by including all the “pure” regimes, using MX-only nowcast fields and allowing for weights less than 0.08 result in the best “pure” regime-specific nowcasts. The exception is the CF regime where the manually tuned MX weights perform marginally better.

Comparing the results in Table 3 with those in Table 4 (details behind the CSI are listed in Table 5), it is clear that the regime-specific weights in Table 3 are not always better than the one-size-fits-all weights created using the MX regime. Indeed, it could be argued that by making it unnecessary for the forecaster to choose a regime, always using only the MX regime makes ANC easier to use.

The weights of the different interest fields when manually tuned, and as obtained from the automated tuning system using the MX weights, are shown in Table 6. We wish to caution that the relative values of the weights of interest fields are poor proxies for the importance of any interest field, since these interest fields are highly correlated. One way to determine the relative importance of a field is to leave it out, tune the system and check if there has been any resulting decrease in performance. We did not do this, so it is not clear how important any of these fields are. The zero weights indicate that, in the training data set, the information content provided by an interest field was probably already present in some of the other fields. Experts tuning ANC often attempt to have non-zero weights for each of the fields; such a constraint is one that we will experiment with in future work. It should also be realized that these weights are a result of training using data from the warm season; adding training cases from the cold season will presumably also affect the applicability of these weights.

a. Summary

From Table 3, it can be determined that the average run time for the regime-specific tuning scenarios is on the order of 24 hours, completely unattended. Thus, the amount of time, labor and cost required to create objectively-tuned, regime-specific weights is substantially less than the amount of time (several weeks) which is needed to create subjectively-tuned, regime-specific weights. In addition, these objectively-tuned weights outperform subjective tuning by human experts in nearly all cases and can easily be rerun, once datasets and membership functions are identified, for new interest fields. Following the objective tuning mechanism followed in this paper will, thus, enable the easy rollout of ANC to the large number of forecast offices envisioned by the US National Weather Service.

We wish to emphasize that the automation is purely in terms of the one-time tuning of ANC weights. Forecast input is critical in choosing the training cases for automated tuning. In routine operation of ANC, forecaster input is critical in that forecaster-drawn boundaries are a key interest field for ANC.

Also, in this paper, we limited the tuning to optimizing the weights of the various interest fields used by ANC. The interest fields themselves are created by applying a fuzzy membership function to model-derived or observed meterological variables. Forecasters should examine the membership functions to ensure that they are reasonable for the dominant weather modes in their region. Lin et al. (2012) suggest that the fuzzy membership functions themselves can be designed objectively using univariate conditional probabilities obtained from a long-term archive of data. Forecasters should also consider incorporating other predictor variables if these variables could help diagnose thunderstorm initiation.

We suggest that operational forecasters use this process to customize ANC to their forecast area:

i. Verify that the ANC predictor variables and membership functions are reasonable for the predominant weather modes in their region.

ii. Choose a set of cases that illustrate the weather scenarios where gridded nowcast guidance would be helpful.

iii. Draw boundaries to guide ANC (forecaster-drawn boundaries are a key interest field for ANC).

iv. Use the automated system described in this paper to tune ANC weights in MX mode.

If it is found that ANC does not perform well on some scenario, we suggest that the forecaster add that case to the training dataset and retune ANC. We strongly caution against manually tuning ANC's weights. An automated algorithm will be better able to balance the predictor field weights so as to obtain good performance on all the situations used in training. Finally, we suggest that there is little incentive to separate convective regimes, because the one-size-fits-all MX weights perform just as well as the regime-specific weights.

Acknowledgments

Funding for the lead author was provided under NOAA-OU Cooperative Agreement NA17RJ1227. We would like to thank Rita Roberts, Amanda Anderson, Dan Megenhardt and Eric Nelson of NCAR's Research Application Laboratory for their kind and timely assistance with archival data and for providing us the extensive information necessary to remap the nowcast interest fields. We also wish to thank Stephan Smith of MDL, Jim Wilson of NCAR and two anonymous reviewers for their suggestions on improving the manuscript.

REFERENCES

Bakhshaii, A. and R. Stull, 2009: Deterministic ensemble forecasts using gene-expression programming*. *Wea. Forecasting*, **24 (5)**, 1431–1451, doi:10.1175/2009WAF2222192.

1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2009WAF2222192.1`, `http://journals.ametsoc.org/doi/pdf/10.1175/2009WAF2222192.1`.

Barron, J., D. Fleet, and S. Beauchemin, 1994: Performance of optical flow techniques. *Int'l J. Comp. Vis.*, **12 (1)**, 43–77.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Monthly Wea. Review*, **134 (7)**, 1772–1784.

Donaldson, R., R. Dyer, and M. Kraus, 1975: An objective evaluator of techniques for predicting severe wea. events. *Preprints, Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 321–326.

Ebert, E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24 (6)**, 1498–1510.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24 (5)**, 1416–1430, doi:10.1175/2009WAF2222269.1, `http://journals.ametsoc.org/doi/pdf/10.1175/2009WAF2222269.1`.

Goldberg, D., 1989: *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Publishing Company, Inc., 432 pp.

Haupt, S. E., G. S. Young, and C. T. Allen, 2006: Validation of a receptor-dispersion model coupled with a genetic algorithm using synthetic data. *J. of Applied Meteorology and*

Climatology, **45 (3)**, 476–490, doi:10.1175/JAM2359.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JAM2359.1, http://journals.ametsoc.org/doi/pdf/10.1175/JAM2359.1.

Heidke, P., 1926: Berechnung des erfolges und der gute der windstarkvorhersagen im sturmwarnungsdienst. *Geogr. Ann.*, **8**, 301–349.

Keil, C. and G. Craig, 2007: A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Wea. Rev.*, **135**, 3248–3259.

Lakshmanan, V., 2000: Using a genetic algorithm to tune a bounded weak echo region detection algorithm. *J. of Applied Meteorology*, **39 (2)**, 222–230.

Lakshmanan, V. and J. Kain, 2010: A Gaussian mixture model approach to forecast verification. *Wea. Forecasting*, **25 (3)**, 908–920, doi:10.1175/2010WAF2222355.1.

Lin, P., P. Chang, B. Jou, J. Wilson, and R. Roberts, 2012: Objective prediction of warm season afternoon thunderstorms in northern taiwan using a fuzzy logic approach. *Wea. Forecasting*, **1 (1)**, accepted.

Marzban, C., 1998: Scalar measures of performance in rare-event situations. *Wea. Forecasting*, **13**, 753–763.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, 1953: Combinatorial minimization. *J. Chem. Phys.*, **21**, 1087–1092.

Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and H. Yee, 2003: NCAR auto-nowcaster system. *Wea. Forecasting*, **18**, 545–561.

Nelson, E., et al., 2005: Evaluation of the NCAR Auto-nowcaster during the NWS Ft. Worth operational demonstration. *32nd Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., 6R.1.

O'Steen, L. and D. Werth, 2009: The application of an evolutionary algorithm to the optimization of a mesoscale meteorological model. *J. of Applied Meteorology and Climatology*, **48 (2)**, 317–329, doi:10.1175/2008JAMC1967.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2008JAMC1967.1`, `http://journals.ametsoc.org/doi/pdf/10.1175/2008JAMC1967.1`.

Roebber, P. J., 2010: Seeking consensus: A new approach. *Monthly Wea. Review*, **138 (12)**, 4402–4415, doi:10.1175/2010MWR3508.1, URL `http://journals.ametsoc.org/doi/abs/10.1175/2010MWR3508.1`, `http://journals.ametsoc.org/doi/pdf/10.1175/2010MWR3508.1`.

Wilks, D., 1995: *Statistical Methods in Atmospheric Sciences*. Academic Press.

Wilson, J., N. A. Crook, C. K. Mueller, J. Z. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079–2099.

Wilson, J. and C. Mueller, 1993: Nowcast of thunderstorm initiation and evolution. *Wea. Forecasting*, **8**, 113–131.

Wilson, J., F. Yerong, M. Chen, and R. Roberts, 2010: Nowcasting challenges during the Beijing Olympics: successes, failures, and implications for future nowcasting systems. *Wea. Forecasting*, **25**, 1691–1714.

Wolfson, M., B. Forman, R. Hallowell, and M. Moore, 1999: The growth and decay storm tracker. *8th Conf. on Aviation*, Dallas, TX, Amer. Meteor. Soc., 58–62.

List of Tables

Interest Field	Resolution, Grid Size, and Timing						Convective Weather Regime					
	Δx	Δy	N_x	N_y	Δt	T_{max}	CF	DL	MC	MX	PS	SW
CAPE	20	20	55	55	60	195	✓	✓	✓	✓		✓
CIN	20	20	55	55	60	195	✓	✓	✓	✓	✓	✓
Likelihood of frontal zone	20	20	55	55	60	195	✓	✓	✓	✓	✓	✓
Relative humidity	20	20	55	55	60	195	✓	✓	✓	✓		✓
Gradient of θ_ϵ	20	20	55	55	60	195		✓				
Instability $1000-700mb$	20	20	55	55	60	195	✓	✓	✓	✓		✓
Vertical velocity $700mb$	20	20	55	55	60	195	✓	✓	✓	✓		✓
Surface mass convergence	10	10	400	400	5	20	✓	✓	✓	✓	✓	✓
Lifting Index	10	10	400	400	5	20	✓	✓	✓	✓	✓	✓
Areas along human-denoted boundaries	2	2	360	330	6	14	✓	✓	✓	✓		✓
Lifting area (colliding boundaries)	2	2	360	330	6	14	✓	✓	✓	✓	✓	✓
Vertical motion along boundaries	2	2	360	330	6	14	✓	✓				
Steering flow relative to boundary	2	2	360	330	6	14	✓	✓	✓	✓	✓	✓
Cloud top temperature	1	1	1100	820	15	70	✓	✓	✓	✓	✓	✓
Cloud-free areas	1	1	1100	820	15	70	✓	✓	✓	✓	✓	✓
Areas with cumulus and congestus clouds	1	1	1100	820	15	70	✓	✓	✓	✓	✓	✓

TABLE 1. Descriptions of the different ANC nowcast interest fields, their spatial resolutions (Δx and Δy in kilometers), sizes ($N_x \times N_y$ pixels), the update rate (Δt in minutes), the maximum amount of time in the past from which they can be retrieved to generate a nowcast (T_{max} in minutes). The ✓s indicate ANC convective weather regimes in which the parameter is an input. The weather regimes are cold front (CF), dry line (DL), mesoscale convective system (MC), mixed (MX), pulse storm (PS), and stationary/warm front (SW). It can be noted that four of the interest fields (areas along human-denoted boundaries, lifting area associated with colliding boundaries, vertical motion along boundaries, and steering flow relative to boundaries) are directly related to forecaster-drawn boundaries: these tend to be very important.

Regime	Date	Time
CF	Aug. 27, 2006	14-21 Z
	July 13, 2008	19-22 Z
	July 15, 2008	19-23 Z
DL	Apr. 24, 2007	16-23 Z
	Mar. 31, 2008	17-23 Z
	May 14, 2010	04-18 Z
MC	Apr. 30, 2007	04-23Z
	July 30, 2008	20-23 Z
	July 31, 2008	19-23 Z
MX	July 21, 2007	07-23 Z
	Aug. 1, 2007	17-23 Z
	Oct. 8, 2007	13-23 Z
PS	May 12, 2007	17-23 Z
	Sep. 3, 2007	13-23 Z
	July 8, 2008	17-23 Z
SW	May 2, 2007	01-23 Z
	May 6, 2007	10-23 Z
	May 27, 2007	01-23 Z

TABLE 2. The nowcast dates and times used in this study, subdivided by ANC convective weather regime.

Regime	Tuning Dates	Control Date	Fitness (Train)	Fitness of Control Date Human	GA	Δf (%)	Time (hr)
CF	20060827, 20080713	20080715	0.195	0.106	0.112	5	20
	20080713, 20060715	20060827	0.177	0.149	0.156	5	18
	20060827, 20080715	20080713	0.168	0.122	0.185	51	23
DL	20070424, 20080331	20100514	0.184	0.112	0.144	28	21
	20080331, 20100514	20070424	0.179	0.142	0.155	9	25
	20070424, 20100514	20080331	0.168	0.155	0.184	19	21
MC	20070430, 20080730	20080731	0.144	0.045	0.092	105	25
	20080730, 20080731	20070430	0.188	0.035	0.070	99	18
	20070430, 20080731	20080730	0.135	0.088	0.185	111	25
MX	20070721, 20070801	20071008	0.153	0.130	0.139	7	27
	20070801, 20071008	20070721	0.174	0.063	0.031	-51	24
	20070721, 20071008	20070801	0.141	0.174	0.149	-14	27
PS	20070512, 20070903	20080708	0.139	0.159	0.164	3	17
	20070903, 20080708	20070512	0.149	0.164	0.130	-21	17
	20070512, 20080708	20070903	0.176	0.076	0.075	-1	16
SW	20070502, 20070506	20070527	0.142	0.046	0.083	82	27
	20070506, 20070527	20070502	0.147	0.034	0.046	36	42
	20070502, 20070527	20070506	0.137	0.111	0.099	-11	38

TABLE 3. The results of the three tuning scenarios for each ANC convective weather regime. Each scenario is characterized by the nowcast dates (in YYYYMMDD format) used for tuning, the corresponding control date, the final overall fitness of the nowcasts used for tuning, the final overall fitness of the corresponding control date's nowcasts using both the subjectively-tuned and objectively-tuned regime-specific weights, the improvement seen by using the genetic algorithm (GA), and the time taken by the GA to tune.

Regime	Control Date	Automated tuning scenario			Fitness	CSI
		Tuning Data	Interest Fields	Weights		
		Tuning Data	Interest Fields	≥ 0.08?		
CF	20060827	MX-only	MX-only	No	0.129	0.075
		All	MX-only	No	0.130	0.082
		All	All	No	0.144	0.097
		All	MX-only	Yes	0.126	0.077
		Manual tuning			**0.146**	0.103
DL	20100514	MX-only	MX-only	No	0.123	0.056
		All	MX-only	No	**0.155**	0.099
		All	All	No	0.125	0.052
		All	MX-only	Yes	0.130	0.054
		Manual tuning			0.116	0.031
MC	20080730	MX-only	MX-only	No	0.169	0.098
		All	MX-only	No	**0.185**	0.110
		All	All	No	0.182	0.111
		All	MX-only	Yes	0.176	0.015
		Manual tuning			0.096	0.018
PS	20070512	MX-only	MX-only	No	0.171	0.091
		All	MX-only	No	**0.191**	0.126
		All	All	No	0.184	0.107
		All	MX-only	Yes	0.180	0.107
		Manual tuning			0.166	0.102
SW	20070506	MX-only	MX-only	No	0.104	0.017
		All	MX-only	No	**0.117**	0.037
		All	All	No	0.103	0.020
		All	MX-only	Yes	0.098	0.016
		Manual tuning			0.109	0.024

TABLE 4. Different MX regime tuning scenarios applied to ANC's pure weather regimes on the control dates. The final overall fitness value of the control dates' nowcasts using both manual tuning and automatic tuning four different ways are shown. The best method of tuning is highlighted.

Regime	Control Date	MX weights				Regime-specific weights			
		Fitness	CSI	POD	FAR	Fitness	CSI	POD	FAR
CF	20060827	0.130	0.082	0.787	0.916	0.156	0.108	0.729	0.888
DL	20100514	0.155	0.099	0.381	0.883	0.144	0.084	0.15	0.843
MC	20080730	0.185	0.110	0.872	0.888	0.185	0.116	0.825	0.881
PS	20070512	0.191	0.126	0.742	0.868	0.130	0.063	0.882	0.936
SW	20070506	0.117	0.037	0.856	0.963	0.099	0.013	0.722	0.987

TABLE 5. Using regime-specific weights sometimes improves the performance of ANC over always using the MX weights, but it is not clear-cut. Hence, it is possible that WFOs might choose to let ANC always operate in MX mode.

Interest Field	Manually-tuned weight	Auto-tuned weight
CAPE	0.20	0.12
CIN	0.12	1.00
Convergence	0.10	0.72
Likelihood of frontal zone	0.22	0.00
Areas along human-denoted boundaries	0.20	0.32
Cloud top temperature	0.10	0.23
Lifting Index	0.20	0.28
Lifting area (colliding boundaries)	0.12	0.35
Relative humidity	0.18	0.50
Cloud-free areas	0.40	0.51
Areas with Cu and CuC clouds	0.12	0.00
Boundary-relative steering flow	0.18	0.00
Instability 1000-700mb	0.12	0.00
Vertical velocity 700 mb	0.08	1.00

TABLE 6. A comparison of the weights obtained as a result of manual tuning and as a result of automated tuning. The auto-tuned weight is the result of tuning on data consisting of all the regimes and using the MX regime, i.e., it is not regime-specific.

List of Figures

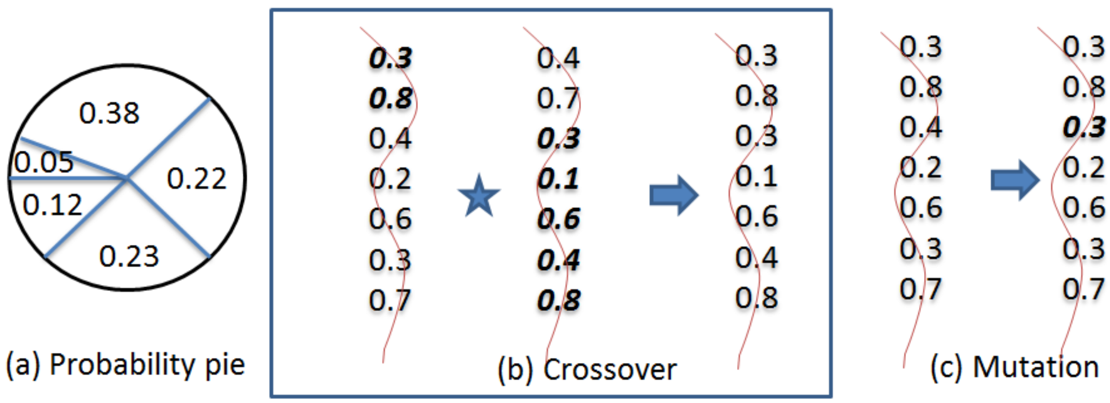

FIG. 1. (a) In a genetic algorithm, chromosomes are chosen probabilistically, with better fit chromosomes more likely to be chosen. The numbers represent the fitness of the chromosome corresponding to the slice. (b) Crossover involves creating a new chromosome that contains the first part of the chromosome of one parent and the second part of the chromosome from another parent. The split point is chosen randomly. The numbers here represent the parameters being tuned (in our case, the weights of each of the interest fields). (c) Mutation involves creating a new chromosome by modifying one of the genes of a parent chromosome.

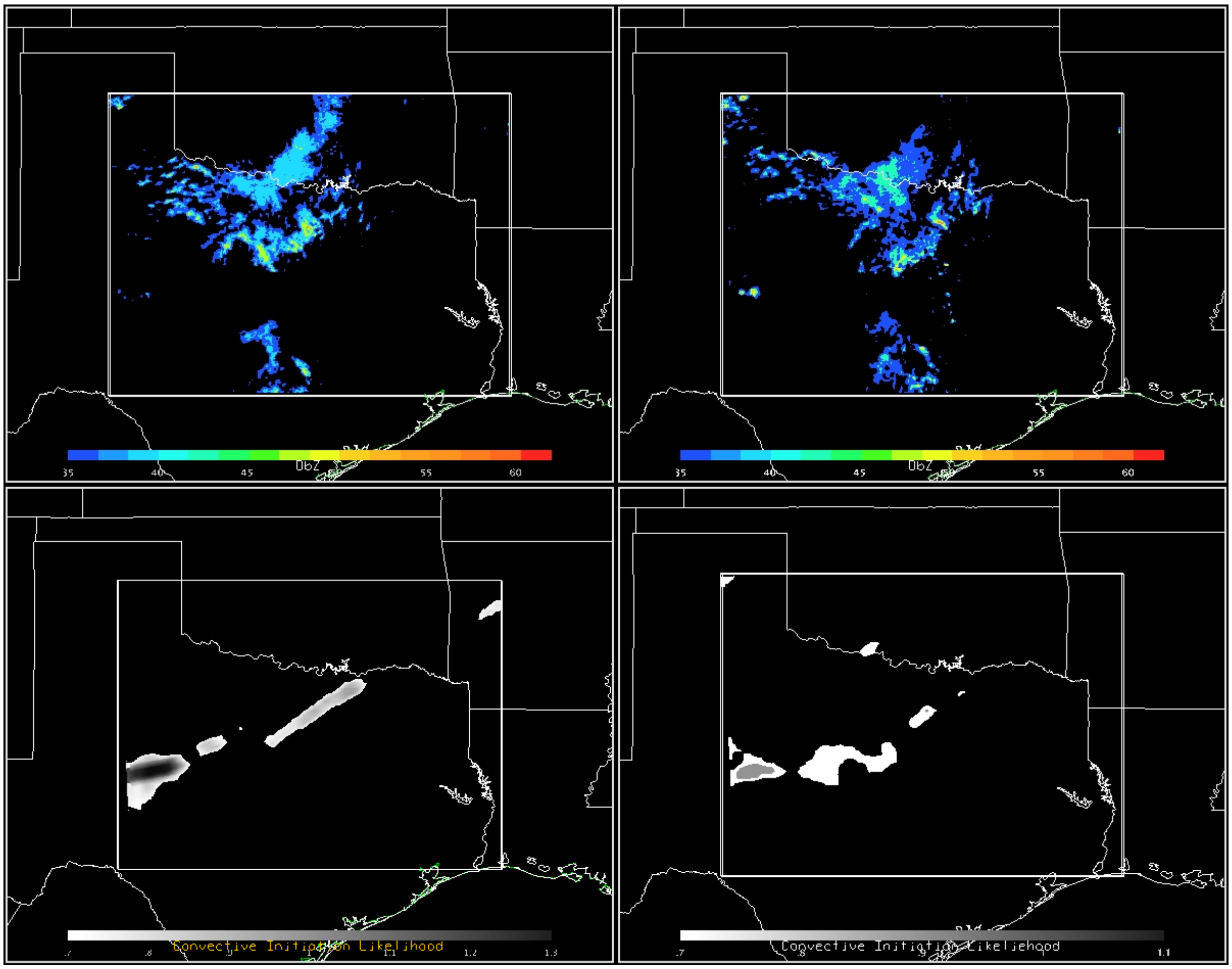

FIG. 2. The top row shows (a) Radar observation on May 14, 2010 at 17:15 UTC. The domain is centered on the Dallas-Fort Worth WSR-88D (KFWS) and (b) the radar observation an hour later. The bottom row shows (c) 60-minute initiation nowcast at 17:15 UTC by manually tuned ANC (d) 60-minute initiation nowcast at 17:15 UTC by auto-tuned ANC.

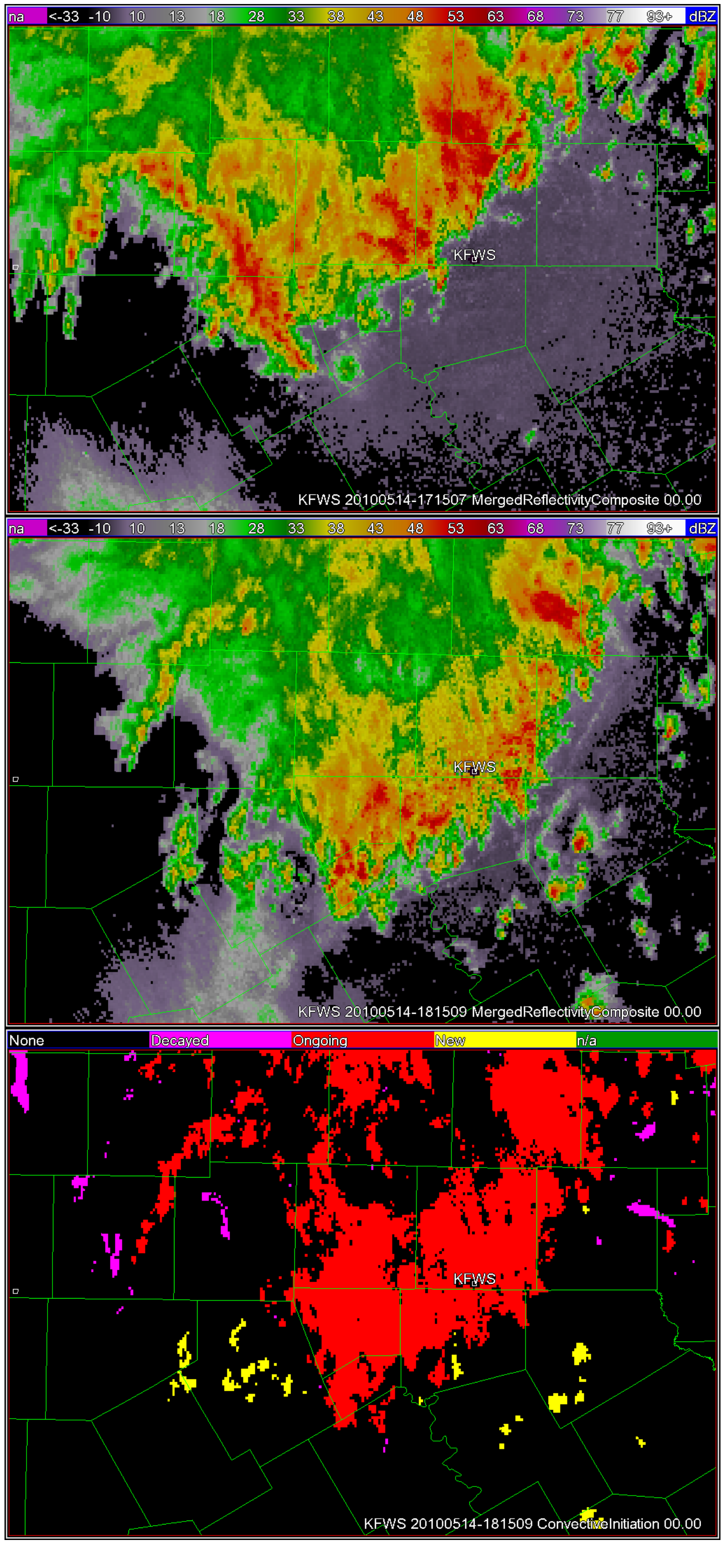

FIG. 3. Top-to-bottom: (a) Radar observation on May 14, 2010 at 17:15 UTC (detail from Figure 2a; the location of the radar is marked as KFWS). (b) Radar observation an hour later. (c) Verification field created by warping the image at 17:15 and looking for new convection. The purples are decayed convection, reds are ongoing convection while the yellow is new convection.

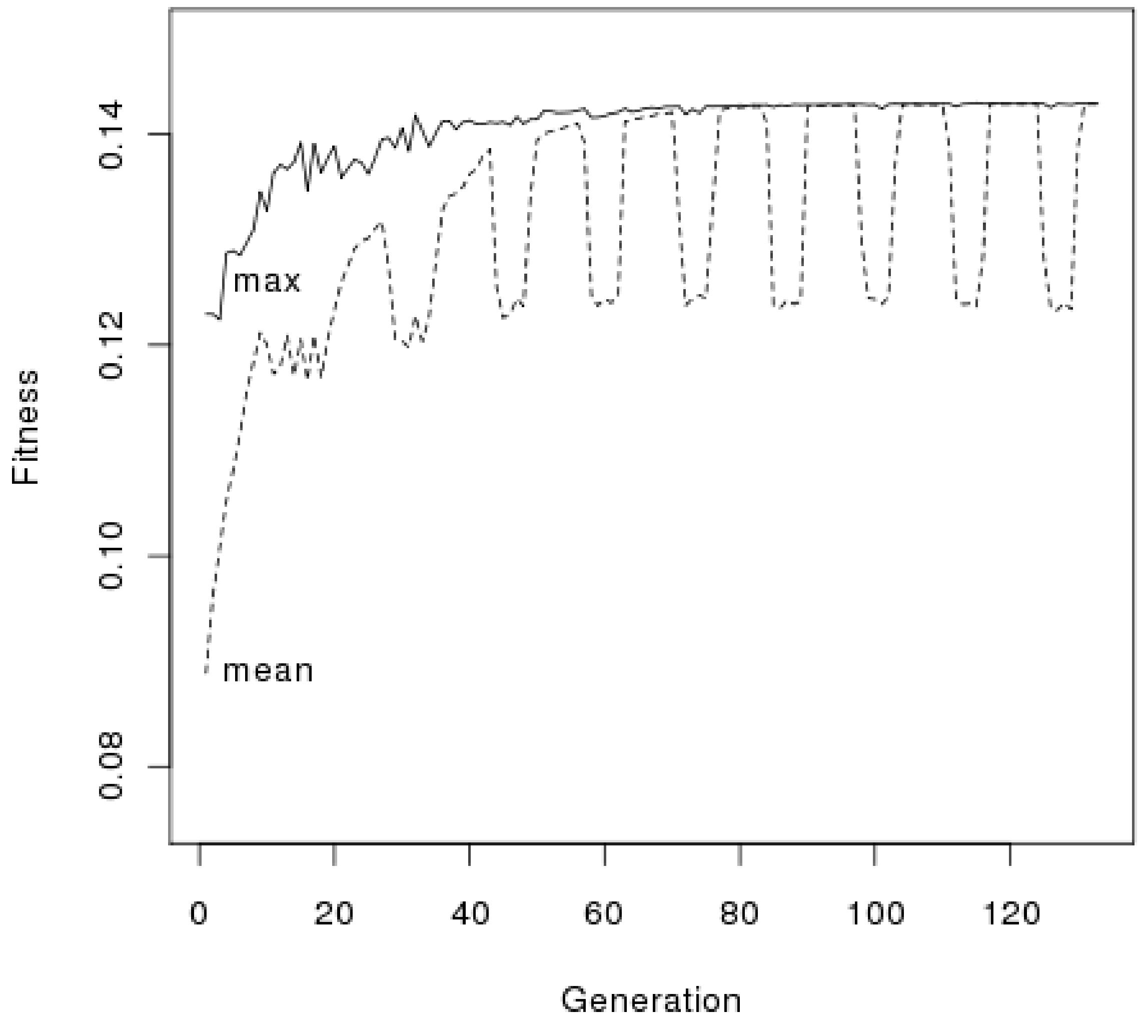

Fig. 4. The genetic algorithm iteratively tries different weights to improve the fitness. The solid line shows the fitness of the best member at each generation while the dotted line shows the means of the fitness values at each generation as training progresses. The sawtooth nature of the graphs is because of the periodic use of simulated annealing to perform a local search around each member.