Adaptive Evolutionary Programming

PAUL J. ROEBBER

*Atmospheric Science Group, Department of Mathematical Sciences, and School of Freshwater Sciences,
University of Wisconsin–Milwaukee, Milwaukee, Wisconsin*

(Manuscript received 28 March 2014, in final form 14 October 2014)

ABSTRACT

Previous work has shown that evolutionary programming is an effective method for constructing skillful forecast ensembles. Here, two prototype adaptive methods are developed and tested, using minimum temperature forecast data for Chicago, Illinois, to determine whether they are capable of incorporating improvements to forecast inputs (as might occur with changes to operational forecast models and data assimilation methods) and to account for short-term changes in predictability (as might occur for particular flow regimes). Of the two methods, the mixed-mode approach, which uses a slow mode to evolve the overall ensemble structure and a fast mode to adjust coefficients, produces the best results. When presented with better operational guidance, the mixed-mode method shows a reduction of 0.57°F in root-mean-square error relative to a fixed evolutionary program ensemble. Several future investigations are needed, including the optimization of training intervals based on flow regime and improvements to the adjustment of fast-mode coefficients. Some remarks on the appropriateness of this method for other ensemble forecast problems are also provided.

1. Introduction

A drawback common to statistical forecast approaches is the difficulty of accommodating operational improvements to the numerical models and data assimilation. While the forecast inputs may improve, the statistical model remains fixed to its training dataset. To counter this, updating methods have been developed for model output statistics (MOS), such as Kalman filters (e.g., Persson 1991; Simonsen 1991; Homleid 1995; Crochet 2004), updateable MOS (Ross 1987; Wilson and Vallée 2002), N-day running mean bias removal (e.g., Cheng and Steenburgh 2007), and weighted MOS (e.g., Baars and Mass 2005). Other bias removal methods include incorporating observations made after the time of forecast issuance but prior to verification (e.g., Nipen et al. 2011; Huang et al. 2012), and the use of analogs (Delle Monache et al. 2011).

Evolutionary programs [EP; see Fogel (1999) for a historical overview of the technique] have been shown in recent studies to provide deterministic and probabilistic temperature forecasts that are superior to those obtained from operational ensembles and MOS (Roebber 2010, 2015a,b). EP, when they are used in this way, can be thought of as a form of statistical weather forecasting, and also suffer from this updating limitation. The ensemble form of EP (Roebber 2015a; see also section 2) produces a population of solutions through simulated natural and sexual selection, mutation, and gene flow (through migration of solutions between ecological niches). These primary factors are modified by niche-specific cultural practices (i.e., the use of inputs and mating behaviors; see section 2) and disease, which also affect reproductive fitness and thereby the evolution of the population. While updateable methods for EP do not yet exist, given these evolutionary attributes, one might expect that the method is uniquely suited to adaptation.

The construction and demonstration of two adaptive methods for EP ensembles is the subject of this paper, which is organized as follows. In section 2, a brief overview of the ensemble EP method is provided, along with a presentation of two distinct adaptive approaches. Readers interested in complete details concerning the ensemble form of EP should consult Roebber (2015a). In section 3, the forecast dataset to which the adaptive method is applied is described. Section 4 provides the

Corresponding author address: Paul J. Roebber, Department of Mathematical Sciences, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, WI 53201-0413.
E-mail: roebber@uwm.edu

results and discussion for these two methods, while section 5 briefly summarizes the paper.

2. The ensemble EP method and adaptation

In this study, we will use the ensemble form of EP developed by Roebber (2015a). The algorithms are composed of a series of up to 10 IF-THEN equations involving linear and nonlinear combinations of predictors. Each of the IF-THEN statements can be considered a gene (EP-gene), such that the ith algorithmic genome has the form

$$F_i = \sum_{j=1}^{10} \partial_{ij}, \qquad (1)$$

where

IF$(V_{1ij} O_{Rij} V_{2ij})$THEN

$$\partial_{ij} = (C_{1ij} V_{3ij}) O_{1ij} (C_{2ij} V_{4ij}) O_{2ij} (C_{3ij} V_{5ij})$$

and V_{1ij}, V_{2ij}, V_{3ij}, V_{4ij}, and V_{5ij} can be any of the input variables; C_{1ij}, C_{2ij}, and C_{3ij} are real-valued multiplicative constants in the range $[-1, 1]$; O_{Rij} is a relational operator ($\leq$, $>$); and O_{1ij}, O_{2ij} are either addition or multiplication operators. The input variables and the forecast variable are normalized to $[0, 1]$ based on the minimum and maximum of the training data.

The first generation of algorithms is initialized with random groups of variables and operators, and random values for the coefficients. Successive generations are produced through a combination of qualified algorithms. Roebber (2015a) found that the preferential combination of the most successful algorithms—that is, a simulated sexual selection while invoking "disease" to limit saturation of the "gene pool" by a small number of individuals (see, e.g., Hillis 1990)—accelerates improved algorithm performance. Algorithmic innovation is introduced through mutation, that is, random changes of variables, operators, or coefficients, or through gene transposition (copy error), accomplished by selecting one of three EP-gene segments—$(V_{1ij} O_{Rij} V_{2ij})$, or $(C_{1ij} V_{3ij}) O_{1ij} (C_{2ij} V_{4ij}) O_{2ij}$, or $C_3 V_{5ij}$—from one of the 10 lines of a child algorithm and copying that segment to a different line. Roebber (2015a) found that mutations were generally most effective at earlier stages of evolution.

The development of independence between algorithms is promoted by establishing 20 developmental groupings (ecological niches), with each grouping allowing different combinations of inputs, disease susceptibility, and "mating" practices (Table 1). Successful

TABLE 1. EP developmental groupings (niches) used in training. The inputs column denotes the selection of the 14 input variables used to train within that niche, where model refers to the GFS ensemble MOS and MEX variables, obs refers to snow on the ground and the three upstream 24-h average observations, and see text for MLR. Disease occurs in only some niches, and in those niches, it can cause either mutation or the "death" of that algorithm, as indicated. The mating variables include the number of female partners per qualifying male algorithm, the number of children produced per partner, whether the female partner is selected from only inside the niche or from any niche, and whether female partners are qualified by MSE.

Niche	Inputs	Disease	Mating Partner	Child	Select	Qualify
1	Model	None	2	5	Inside	All
2	Model	None	10	1	All	MSE
3	Model	Mutate	2	5	Inside	All
4	Model	Death	10	1	All	MSE
5	Obs	None	2	5	Inside	All
6	Obs	None	10	1	All	MSE
7	Obs	Mutate	2	5	Inside	All
8	Obs	Death	10	1	All	MSE
9	Model, MLR	None	2	5	Inside	All
10	Model, MLR	None	10	1	All	MSE
11	Model, MLR	Mutate	2	5	Inside	All
12	Model, MLR	Death	10	1	All	MSE
13	Obs, MLR	None	2	5	Inside	All
14	Obs, MLR	None	10	1	All	MSE
15	Obs, MLR	Mutate	2	5	Inside	All
16	Obs, MLR	Death	10	1	All	MSE
17	Model, MLR, obs	None	2	5	Inside	All
18	Model, MLR, obs	None	10	1	All	MSE
19	Model, MLR, obs	Mutate	2	5	Inside	All
20	Model, MLR, obs	Death	10	1	All	MSE

approaches are shared through the migration of algorithms between niches (when populations dwindle) and cross-niche partner selection. An overarching natural selection pressure is provided through a gradually tightening mean-square-error (MSE) threshold during training, whereby only those qualifying algorithms pass their genetic structures to subsequent generations. A further qualification is provided by a check of the correspondence between an algorithm's cumulative distribution function of forecast temperature and that of the observations for the training data. Succeeding generations are subjected to these evolutionary forces until convergence criteria based upon performance on the training data are satisfied.

We construct adaptive forms of EP in two ways, given below. The first is informed by our understanding of adaptation in nature (section 2a), while the second is suggested by results from the model selection and

Bayesian model combination procedures applied in Roebber (2015a,b).

a. Mixed-mode adaptation

Notably, in response to rapid climate change, heritable genetic changes have occurred in animal populations such as birds, squirrels, and mosquitoes over relatively short periods (Bradshaw and Holzapfel 2006). As noted by those authors, animals with short life cycles are more favorably positioned to adapt under such conditions. We introduce a similar effect here by performing evolution in both "slow" and "fast" modes (mixed mode).

First, we allow the evolution of the overall IF-THEN genetic framework using a moving window of M cases set to the size of the training data (e.g., for 50 training cases, the forecast for case 51 would be based on training for cases 1–50, the forecast for case 52 would be based on training for cases 2–51, etc.). This slow evolution produces one new generation at each time step.

Second, given a particular overall framework produced over the past M cases by the slow mode, the coefficients of that framework are evolved based on the past N cases (where $N << M$, i.e., the fast mode). Note that the choice of N is not clear, and that it could likely be optimized based on the flow regime and time of year. En route to developing a regime-dependent technique, for example, Greybush et al. (2008) evaluated a range of days for weighting ensemble model consensus forecasts in the Pacific Northwest and found an optimum for $N = 7$. Tests for this study (not shown) established a small overall improvement in MSE when increasing N from 5 to 7 but also revealed substantial seasonal differences in performance between ensembles using the two N values, with the largest variation occurring during transitional periods in the spring and fall; thus, there is a future opportunity to define regime-dependent fast-mode periods that would likely improve forecast skill.

Fast-mode coefficients are produced by randomly selecting two partners (call them A and B) from any but the best-performing niche, and either swapping coefficients between A and B, averaging the A and B coefficients, or randomly selecting new coefficients for A and B. In any instance, one of these three possibilities is selected at random with equal probability. This process is repeated until every algorithm not within the best-performing niche is selected at least one time, and either the MSE of the ensemble mean formed by all of these algorithms over the past seven cases is better than that of the best-performing niche or the number of such iterations equals a critical value (set to 100). The above-mentioned choices are somewhat ad hoc but are a reasonable means to sample a wide variety of coefficient combinations and to determine whether such combinations are capable of producing forecast improvements. Since this is a demonstration of the approach, no attempt to optimize these choices has yet been made.

The mixed mode allows adjustment to changes in the predictive capability of the inputs (as with a model improvement) but also can implicitly account for short-term changes in guidance bias associated with particular flow regimes and seasonal factors—such a variation in performance has long been recognized by forecasters who attempt to account for this feature by selecting a "model of the day" (e.g., Hoffman et al. 2006).

The forecast for case $M + 1$ (i.e., the independent forecast of interest) is produced from this mixed-mode training using a grand ensemble consisting of all those individual algorithms that outperform either the best-performing niche ensemble or the overall ensemble performance (based on the past seven cases), whichever threshold is lower. The window of M cases is then shifted forward one step and the mixed-mode process is repeated for the next forecast date. Computational requirements for this procedure are minimal—for the data considered here, one full mixed-mode forecast step requires 12 s of wall time on a dual-core 1.6-GHz Intel system running nonoptimized FORTRAN code.

b. Model selection and Bayesian model combination adaptation

Roebber (2015a,b) showed that using ensemble member subselection in conjunction with a technique known as Bayesian model combination (BMC) can produce further gains in performance for a pooled ensemble consisting of evolutionary program members and the 21-member Global Forecast System ensemble MOS (GFS ensemble MOS; see www.nws.noaa.gov/mdl/synop/enstxt.php). BMC is conceptually similar to the more commonly used Bayesian model averaging (BMA; e.g., Raftery et al. 2005) except that rather than seeking to select the best member, one instead seeks to select the best *combination* of model members. This slightly different approach appears to confer some predictive advantage compared to BMA for identical datasets (see Monteith et al. 2011).

We use these concepts here to produce adaptive ensembles in another way. As in section 2a above, we employ the slow mode of evolution with a moving window of M cases. Next, we identify the 10 best-performing ensemble members over the prior seven days, selecting from a combined pool of EP and GFS ensemble MOS members. Best is defined by first comparing the forecast variance between every possible pair of ensemble members over the past seven days, and if that variance falls below a critical threshold (here, 0.25°F), eliminating the member of the pair with the

higher MSE over that same period. After this screening, the top 10 of all remaining ensemble members are selected based on the MSE over the past seven days. Thus, "improved information" is immediately incorporated at the model selection stage if it is provided as a separate model and if that performance advantage is demonstrated during that particular interval. Other improvements will be incorporated slowly, through the slow evolution of the overall EP architecture (e.g., if data are used as an input to the EP training).

BMC with linear weights (e.g., Monteith et al. 2011) is then applied and the selected ensemble members and weights are used to produce the next day's forecast. As in Roebber (2015a,b), we evaluate four possible weights for each ensemble member, resulting in the evaluation of 4^{10} or 1 048 576 possible combinations, with normalized weights ranging from 1/37 to 4/13. The best ensemble combination is selected by defining the posterior probability for the combination (e) given the training data (D) as

$$p(e \mid D) \propto \frac{1}{4^{10}}(1 - \in)^r \in^{n-r}, \qquad (2)$$

where $\in$ is estimated using the average error rate of the ensemble combination on the training data, r is the number of correct predictions, and n is the total number of training cases. Here, we consider a model combination to be correct if the weighted forecast is within 5° of the observed value. The selected combination is the one that maximizes the logarithm of (2). This process is then repeated for the next forecast case after sliding the M window forward one case. Although BMC is more computationally intensive, a single forecast can be produced using this method in less than one minute using FORTRAN code running on a dual-core 1.6-GHz Intel system.

From the above, it should be apparent that this technique, while straightforward to implement, is strongly constrained by the model selection step. If, for example, none of the improved GFS ensemble MOS members is selected for the 10-member set, then the adaptive forecast for that case will approximate EP with BMC—the difference being that the EP members are undergoing a slow evolution rather than remaining with a fixed architecture established by the original training with the inferior GFS ensemble MOS. Roebber (2015b) found that EP with BMC provided similar or slightly degraded performance based on RMSE compared to the standard EP method, when improved guidance was not an issue. Thus, we should expect better performance here than that of a fixed EP, but how much better will depend on the degree that the improved information content of the GFS ensemble

TABLE 2. Forecast cues for minimum temperature forecasts for Chicago (ORD). Cues based on observations are shown in boldface. Brackets indicate that the cue is derived from the GFS MOS ensemble. For the initial EP, the MEX and ensemble values are provided by the 60-h forecast data, while for the adaptive cases, these are replaced by the 36-h forecast data, simulating an operational model forecast improvement.

Variable	Description
SNOW	Snow on the ground (at time of forecast issuance)
$[CC]_{min}$	Minimum forecast cloud cover over all GFS ensemble MOS members
Sin (JD)	Sine of the Julian day
Cos (JD)	Cosine of the Julian day
WS	MEX wind speed
CC_{GRB}	MEX cloud cover at Green Bay
PP_{GRB}	MEX probability of precipitation at Green Bay
WS_{DVN}	MEX wind speed at Davenport
PP_{DVN}	MEX probability of precipitation at Davenport
DSM	Average temperature at Des Moines in the 24 h prior to forecast issuance
MSP	Average temperature at Minneapolis in the 24 h prior to forecast issuance
STL	Average temperature at St. Louis in the 24 h prior to forecast issuance
MLR	Multiple linear regression using the 60-h forecast data
$[T]$	GFS ensemble MOS mean temperature

MOS can be incorporated through model selection. In section 4, we will interpret the test results in light of this understanding.

3. Chicago dataset

For testing the two adaptive methods, we return to the minimum temperature forecast dataset for Chicago, Illinois (ORD), used in Roebber (2015a,b). For the present study, these data include observed and forecast data for April 2008–March 2012, the latter obtained from the GFS ensemble MOS issued at 0000 UTC. These forecasts are supplemented with the 0000 UTC extended-range GFS-based MOS (MEX) guidance for Davenport, Iowa (DVN), and Green Bay, Wisconsin (GRB). The set of forecast inputs (Table 2) were the minimum forecast cloud cover at ORD over all the GFS ensemble MOS members, the MEX wind speed at ORD, the MEX cloud cover and the probability of precipitation at GRB, the MEX wind speed and the probability of precipitation at DVN, and the GFS ensemble MOS mean temperature at ORD. Observational inputs (Table 2) included snow on the ground at the time of forecast issuance, the sine and cosine of the Julian day, and the last daily average temperature (through 0000 UTC of the forecast issue time) at three "upstream" sites: Des Moines, Iowa (DSM); Minneapolis, Minnesota (MSP); and St. Louis, Missouri (STL). The set of 14 overall

inputs were used both for the initial EP, which is developed with 60-h forecast data, and the adaptive EP, which evolves after the initial training using 36-h forecasts in place of the 60-h forecasts to simulate operational model improvement.

A multiple linear regression (MLR) forecast is also produced, based on a stepwise regression procedure on a more complete set of inputs than used in the EP training [see Roebber (2015b) for a complete description of these inputs]. The selected MLR predictors were the persistence minimum temperature (i.e., the minimum temperature 12–24 h prior to forecast issuance), the observed precipitation amount category for the 24 h prior to forecast issuance (0 if less than 1 mm; 1 if at least 1 mm and less than 5 mm; 2 if at least 5 mm), the minimum forecast temperature and the minimum precipitation probability across all members of the GFS ensemble MOS, the GFS ensemble MOS mean cloud cover, the cosine of the Julian day, and the MEX forecast dewpoint temperature and precipitation probability at DVN. Roebber (2010) showed that MLR can provide a competitive forecast for temperature. Thus, the MLR constructed in this way provides a quasi-independent reference forecast that can be held fixed during the adaptive phase and whose weighting can then be compared to another, improved base input (the GFS ensemble MOS).

Training is based on 690 cases (i.e., $M = 690$) obtained from the forecasts and observations for the period 8 April 2008–12 March 2010, while the 715 adaptive cases are composed of forecasts and observations for the period 13 March 2010–29 March 2012. Note that the GFS ensemble MOS mean RMSE for this test period is 4.27°F (3.72°F) at 60 h (36 h). This substantial transition in forecast accuracy between 60 and 36 h ensures that the effect of the improved information, if successfully incorporated, will be readily apparent.

4. Results and discussion

As noted previously, rather than establishing specific levels of performance between various forms of guidance (i.e., a "bake-off"), our interest is to determine whether the adaptive approaches proposed here can effectively incorporate improved forecast information. In that regard, we find that the ordered sequence of RMSE for the reduced test data period (see below) are 60-h GFS ensemble MOS mean (4.30°F), fixed EP based on 60-h guidance (4.18°F), 36-h GFS ensemble MOS mean (3.73°F), BMC adaptive EP (3.61°F), and mixed-mode adaptive EP (3.54°F) (Table 3). Thus, we find that both forms of adaptation do indeed successfully incorporate the improved information represented by the

TABLE 3. RMSE (°F) for the test data beginning 24 May 2010 and ending 29 Mar 2012 for GFS ensemble MOS at the 60-h forecast range, the normal EP trained with 60-h forecast data, the GFS ensemble MOS at the 36-h forecast range, and the adaptive EP incorporating 36-h forecast data after 13 Mar 2010 using the BMC method and the mixed mode.

Forecast	RMSE (°F)
GFS ensemble MOS (60 h)	4.30
EP with fixed training (60 h)	4.18
GFS ensemble MOS (36 h)	3.73
Adaptive EP (BMC)	3.61
Adaptive EP (mixed mode)	3.54

36-h GFS ensemble MOS mean, but that the mixed-mode approach is the more effective of the two. We also note that the separation in error between the fixed and the adaptive EP occurs within approximately 50 days for these data (Fig. 1), as measured by the first crossing of the 30-day moving average RMSE between the two forecasts. When comparing the mixed-mode adaptive RMSE to that of the 36-h GFS ensemble MOS, we find this crossing occurs at about 24 May 2010, about 72 days into the adaptive phase (this date is used as the beginning of the test period for the comparisons shown in Table 3). This is sufficiently fast that the method could be operationally useful.

Since the mixed-mode adaptive method appears to be the more effective of the two approaches, it is of interest to understand these forecasts in more detail. First, if the slow adaptation of the mixed mode is working properly, then one would expect that the weighting of MLR, which is fixed to the same inputs and coefficients during the adaptive phase as during training, would decrease substantially compared to that of the GFS ensemble MOS (noting that this input has improved through reference to the 36-h forecast instead of the 60-h forecast). We measure this weighting by considering the mean coefficient of all invoked EP-genes multiplying MLR versus all those multiplying the GFS ensemble MOS, at the beginning (13 March 2010) and again at the end (29 March 2012) of the full test period. We find this weighting for the MLR falls from 0.504 to 0.059, while that of the GFS ensemble MOS rises from 0.045 to 0.761 through the period.

Since both the mixed-mode and BMC adaptive EPs use the slow mode, while only the former uses the fast mode, one can think of their differing performance as a rough measure of the importance of the fast mode; in this instance, 0.07°F or about a 2% reduction in error across all cases. Similarly, the fixed EP does not evolve after training, while the BMC adaptive EP evolves through the slow mode, so the differing performance between these two approaches can give a sense of the
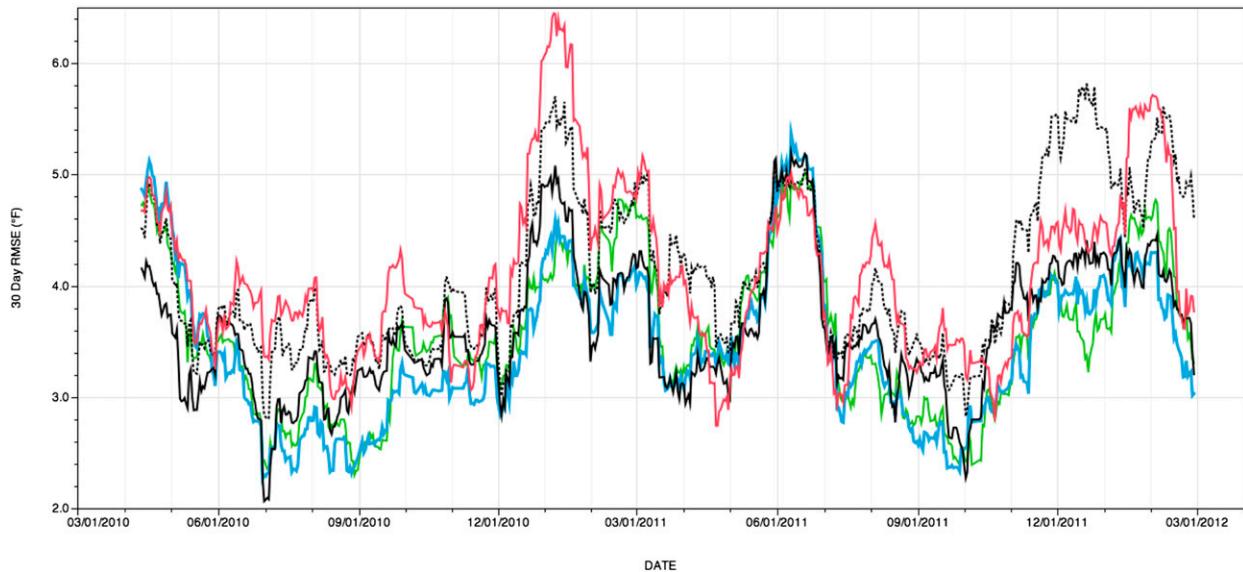
FIG. 1. Time series of the 30-day moving average of the RMSE (°F) for the fixed (red), BMC (green), and mixed-mode adaptive (blue) EP. Also shown are the 60-h GFS (dotted black) and the 36-h GFS ensemble MOS (solid black).

relative effectiveness of the slow mode—here, 0.57°F or 14% reduction in error. However, as noted in section 2, the effectiveness of the BMC adaptive method is dependent on its ability to incorporate the improved GFS model in the selection stage. Although at times during the test period as many as 8 of the 10 selected members were from the GFS ensemble MOS, 85% of the BMC adaptive forecasts did not use any GFS members as part of the 10-member ensemble. This renders the above-mentioned comparisons less precise than otherwise. Thus, model selection remains an important constraint on operational implementation of BMC as an adaptive tool, as was also found by Roebber (2015a,b) from the standpoint of EP calibration.

Roebber (2015a) established that EP diversity is beneficial to probabilistic forecast performance. One measure of genome diversity is the frequency of a specific gene in a population. We employ this measure, counting each of the 10 lines of every algorithm (provided that the logic implies that the gene will at least sometimes be invoked), and retain only those genes for which at least 1000 instances occur for at least one time step (Table 4; Fig. 2). Three of these 54 genes were dominant in the population at the beginning of the adaptive phase, one involving MLR, one involving upstream temperature at MSP, and one involving the combined influence of those two variables. Of these, only the EP-gene involving MSP was retained at the end of the adaptive phase, however, while three "new" EP-genes, which first appeared at various points throughout this evolution, are present at the end. These new EP-genes

involved a combination of the GFS ensemble MOS mean temperature and MSP, a complex and conditional combination of time of year, snow cover, and MSP, and another conditional combination involving forecast cloud cover, time of year, and precipitation probability at DVN. The data (Fig. 2; Table 4) suggest increased use of conditionals compared to the nonadaptive EP, and also that of transient clusters of EP-genes (e.g., the clusters in August 2010, January 2011, July 2011, and November–March 2012) that might help to account for seasonal characteristics (e.g., August 2010 EP-genes feature cloud cover, while January 2011 feature combinations of cloud cover and snow).

Another means to examine the effectiveness of the fast mode of the mixed-mode EP adaptive method is to consider the adjustment of coefficients during a short period in which a substantial transition is occurring in the accuracy of particular inputs. An example of this adjustment is provided by the period 7–13 April 2011, which featured the development and passage of a surface cyclone and trailing cold front through the region (not shown). The 7-day moving average RMSE for the GFS ensemble MOS at the beginning of the period was 3.63°F, and the associated "weight" of that input was 0.39. The representation of the cold frontal passage by the GFS was less successful, however, so that by the end of the period, the 7-day RMSE had risen to 4.42°F and the corresponding weight fell to 0.35. Note that while the MLR RMSE improved from 4.54° to 3.46°F during this period, the weighting of that variable remained relatively constant and lower than the GFS at 0.06,

TABLE 4. List of the 54 EP-genes from the mixed-mode adaptive EP that appeared in a minimum of 1000 instances for at least one particular time step during the test period (see Fig. 2 for a time series of these occurrences). The translation of the EP-gene to an IF-THEN statement is provided. See Table 2 for variable definitions.

| Code ID	Translation
1	$0.76677 \times \text{MLR} + 0.35094 \times \text{MSP}$
2	$0.09407 - 0.26072 \times \text{MSP}$
3	$0.02057 \times \text{MLR}^2$
4	IF STL > MLR THEN $-0.07756 \times \text{SNOW} \times \cos(\text{JD})$
5	IF $\cos(\text{JD})$ > WS THEN $0.07952 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
6	$0.02057 \times \text{MLR}$
7	$0.76677 \times [T] + 0.35094 \times \text{MSP}$
8	$0.03958 \times \text{MLR}$
9	$0.76677 \times [T] + 0.33547 \times \text{MSP}$
10	$0.04170 \times \text{MLR}$
11	IF STL > MLR THEN $-0.08524 \times \text{MSP} \times \text{SNOW} \times \cos(\text{JD})$
12	$0.06172 \times \text{MLR}$
13	$-0.01954 \times \text{CC}_{\text{GRB}} \times \text{WS}$
14	IF STL > MLR THEN $-0.05010 \times \text{MSP} \times \text{SNOW} \times \cos(\text{JD})$
15	$-0.01954 \times \text{WS} \times [\text{CC}_{\text{min}}]$
16	IF STL > MLR THEN $-0.05010 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
17	IF $\cos(\text{JD})$ > WS THEN $0.08689 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
18	IF $\text{PP}_{\text{GRB}} \leq [\text{CC}_{\text{min}}]$ THEN $-0.01954 \times \text{CC}_{\text{GRB}} \times \text{WS}$
19	IF $\cos(\text{JD})$ > WS THEN $0.08939 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
20	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.08939 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
21	IF WS $\leq [\text{CC}_{\text{min}}]$ THEN $-0.01954 \times \text{PP}_{\text{GRB}} \times [\text{CC}_{\text{min}}]$
22	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.08689 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
23	IF STL > MLR THEN $-0.06140 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
24	IF $\text{PP}_{\text{GRB}} \leq [\text{CC}_{\text{min}}]$ THEN $-0.01954 \times \text{PP}_{\text{GRB}} \times [\text{CC}_{\text{min}}]$
25	$0.03319 \times \text{MSP} \times [\text{CC}_{\text{min}}] \times \text{SNOW}$
26	IF STL > MLR THEN $-0.07116 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
27	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.08737 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{MLR}$
28	IF $\text{CC}_{\text{GRB}} \leq [\text{CC}_{\text{min}}]$ THEN $-0.01954 \times \text{PP}_{\text{GRB}} \times [\text{CC}_{\text{min}}]$
29	$0.05089 \times \text{MSP} \times [\text{CC}_{\text{min}}] \times \text{SNOW}$
30	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.07932 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{PP}_{\text{DVN}}$
31	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.07932 \times \cos(\text{JD}) \times \text{PP}_{\text{DVN}}$
32	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.07952 \times \cos(\text{JD}) \times \text{PP}_{\text{DVN}}$
33	IF DSM $\leq [\text{CC}_{\text{min}}]$ THEN $-0.01954 \times \text{PP}_{\text{GRB}}$
34	IF MSP > MLR THEN $-0.04641 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
35	$0.02565 \times \sin(\text{JD}) \times [\text{CC}_{\text{min}}] \times \text{SNOW}$
36	$0.01621 \times \sin(\text{JD}) \times [\text{CC}_{\text{min}}] \times \text{SNOW}$
37	IF MSP $\leq$ DSM THEN 0.00844
38	IF $\cos(\text{JD})$ > WS_{DVN} THEN $0.06399 \times [\text{CC}_{\text{min}}] \times \cos(\text{JD}) \times \text{PP}_{\text{DVN}}$
39	IF MSP $\leq$ DSM THEN 0.01154
40	IF MSP > MLR THEN $-0.04654 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
41	IF CCGRB $\leq [\text{CC}_{\text{min}}]$ THEN $0.02593 \times \sin(\text{JD}) \times [\text{CC}_{\text{min}}] \times \text{SNOW}$
42	IF $\text{WS}_{\text{DVN}} \leq [\text{CC}_{\text{min}}]$ THEN $-0.02801 \times \sin(\text{JD}) \times \text{PP}_{\text{DVN}}$
43	$-0.02801 \times \sin(\text{JD}) \times \text{PP}_{\text{DVN}}$
44	IF $\text{CC}_{\text{GRB}} \leq [\text{CC}_{\text{min}}]$ THEN $0.02593 \times \sin(\text{JD}) \times \text{CC}_{\text{GRB}} \times \text{SNOW}$
45	$-0.03085 \times \sin(\text{JD}) \times \text{CC}_{\text{GRB}} \times \text{PP}_{\text{DVN}}$
46	IF MSP > MLR THEN $-0.05541 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
47	$-0.03085 \times \sin(\text{JD}) \times \text{PP}_{\text{DVN}}$
48	IF MSP > MLR THEN $-0.05260 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
49	IF MSP $\leq$ DSM THEN 0.01151
50	IF MSP > MLR THEN $-0.07559 \times \text{WS}_{\text{DVN}} \times \text{SNOW} \times \cos(\text{JD})$
51	IF STL $\leq \text{CC}_{\text{GRB}}$ THEN $-0.03085 \times \sin(\text{JD}) \times \text{PP}_{\text{DVN}}$
52	IF $\text{CC}_{\text{GRB}} \leq [\text{CC}_{\text{min}}]$ THEN $0.02985 \times \sin(\text{JD}) \times \text{CC}_{\text{GRB}} \times \text{SNOW}$
53	IF MSP $\leq$ DSM THEN 0.01057
54	IF MSP > MLR THEN $-0.05541 \times \text{SNOW} \times \cos(\text{JD})$
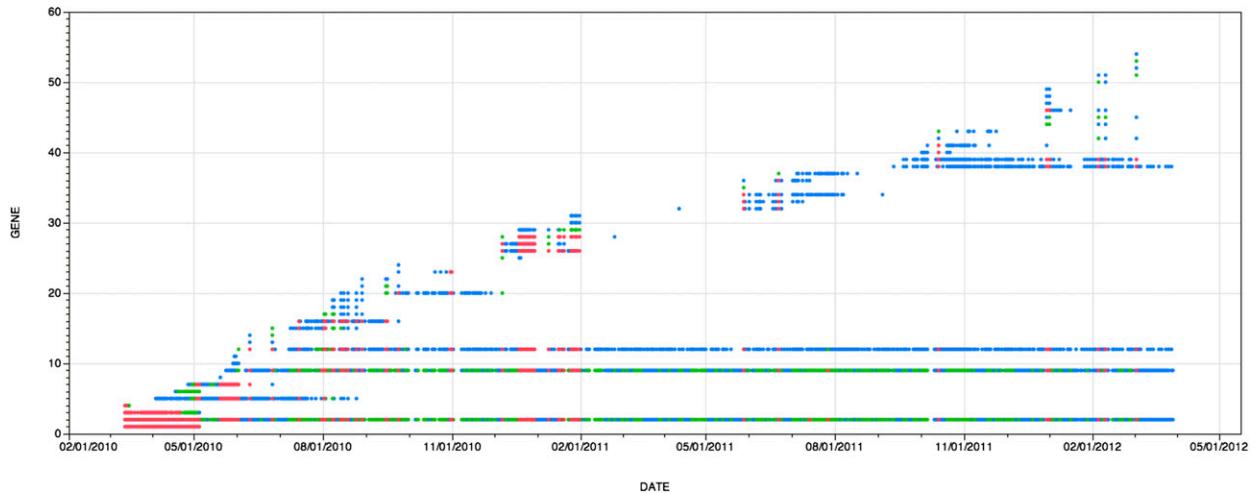
FIG. 2. Time series of the frequency of the most common EP-genes over the test period (counts: blue from 1000 to 2000, green from 2000 to 3000, red > 3000). See Table 4 for identification of the specific EP-genes and their functional translation. Note that EP-genes whose maximum occurrence at any forecast time was less than 1000 are not shown.

consistent with the overall transition to increased weighting of the GFS compared to the MLR over the entire test period. In essence, the fast mode shifted some of the weight from the GFS to variables other than the MLR during this problematic forecast period.

Finally, although the focus of this paper has not been on EP probabilistic performance per se, it is worth noting that the Brier skill score, computed based on standard deviations from the monthly climatic normal as described in Roebber (2015b), is 0.079 for the adaptive EP compared to 0.071 for the fixed EP and 0.049 for the 36-h GFS ensemble MOS for the test data.[1] Thus, the relative probabilistic value of the EP method compared to the operational guidance is retained and improved using the adaptive procedure. As noted in Roebber (2015a,b), however, calibration using BMC improves the GFS probabilistic performance such that the largest advantages of EP compared to calibrated guidance is in providing better deterministic (probabilistic) forecasts at all (longer) ranges.

5. Summary

Two adaptive forms of evolutionary program ensembles have been developed in this paper. The mixed-mode method is developed in a way that allows slow and fast

modes of adaptation, such that adjustments to changes in forecast inputs (as might occur with the introduction of new numerical models and/or improved data assimilation) and short-term regime changes in predictability are both accommodated. These capabilities are demonstrated using minimum temperature forecasts from the GFS ensemble MOS for Chicago, Illinois, for the period 8 April 2008–29 March 2012. A second form, using the slow mode only, but followed by ensemble member subselection and Bayesian model combination, also is able to effectively incorporate improved input information. The mixed-mode adaptive evolutionary program ensemble is shown to be the most effective of the two methods, when compared to the operational MOS ensemble.

Since the purpose of this work has been a proof of concept, however, a number of issues remain to be investigated related to optimizing the method. The most promising of these are developing regime-based training intervals and determining the best means of adjusting fast-mode coefficients. Roebber (2015a,b) show that small improvements in forecasts, including the GFS ensemble MOS, can be produced through application of bias correction. Thus, a further easy improvement would be to first bias correct forecast inputs before providing them to the mixed-mode adaptive EP.

The approach outlined here is easily extended to other forecast variables. The basic requirements of the method are straightforward: define a set of predictors for a problem with a quantifiable measure of success, and for which solutions can be produced through the application of mathematical operations. In that regard, the algorithmic form is similar to but more generalizable

---

[1] Each forecast and matching observation is standardized by the climatological mean and standard deviation for each month. Thirty-three bins of width 0.25 standard deviations from the monthly climatic mean are specified, covering plus/minus four standard deviations, and the probability forecast is made for each bin.

than MOS equations, which are widely used in weather prediction. Roebber (2013, 2015a) extended EP to provide ensembles of forecasts, which is the form used here, so the adaptive method is readily applicable to both deterministic and probabilistic forecasting. Like any nonlinear method, however, choices need to be made that are guided less by theory and more by empirical understanding of the problem at hand. For example, details concerning the choice of inputs and the length of the training period are sure to vary depending on the forecast of interest. At present, it is not clear whether ''hard'' problems like a quantitative precipitation forecast (QPF) or severe weather prediction are amenable to EP, although recent success with long-range temperature forecasts (156 h) is suggestive that the method also can provide utility for forecast problems with a weaker predictive signal. Along those lines, current investigations using the adaptive EP method include studies of regional QPF and of wind power forecasting at specific wind farms.

REFERENCES

Baars, J. A., and C. F. Mass, 2005: Performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *Wea. Forecasting,* **20,** 1034–1047, doi:10.1175/WAF896.1.

Bradshaw, W. E., and C. M. Holzapfel, 2006: Evolutionary response to rapid climate change. *Science,* **312,** 1477–1478, doi:10.1126/science.1127000.

Cheng, W. Y. Y., and W. J. Steenburgh, 2007: Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Wea. Forecasting,* **22,** 1304–1318, doi:10.1175/2007WAF2006084.1.

Crochet, P., 2004: Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland. *Meteor. Appl.,* **11,** 173–187, doi:10.1017/S1350482704001252.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.,* **139,** 3554–3570, doi:10.1175/2011MWR3653.1.

Fogel, L. J., 1999: *Intelligence through Simulated Evolution: Forty Years of Evolutionary Programming.* Wiley Series on Intelligent Systems, John Wiley, 162 pp.

Greybush, S. J., S. E. Haupt, and G. S. Young, 2008: The regime dependence of optimally weighted ensemble model consensus forecasts of surface temperature. *Wea. Forecasting,* **23,** 1146–1161, doi:10.1175/2008WAF2007078.1.

Hillis, W. D., 1990: Co-evolving parasites improve simulated evolution as an optimization procedure. *Physica D,* **42,** 228–234, doi:10.1016/0167-2789(90)90076-2.

Hoffman, R. R., J. W. Coffey, K. M. Ford, and J. D. Novak, 2006: A method for eliciting, preserving, and sharing the knowledge of forecasters. *Wea. Forecasting,* **21,** 416–428, doi:10.1175/WAF927.1.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting,* **10,** 689–707, doi:10.1175/1520-0434(1995)010<0689:DCOSTS>2.0.CO;2.

Huang, L. X., G. A. Isaac, and G. Sheng, 2012: Integrating NWP forecasts and observation data to improve nowcasting accuracy. *Wea. Forecasting,* **27,** 938–953, doi:10.1175/WAF-D-11-00125.1.

Monteith, K., J. Carroll, K. Seppi, and T. Martinez, 2011: Turning Bayesian model averaging into Bayesian model combination. *Proc. 2011 Int. Joint Conf. on Neural Networks,* San Jose, CA, IEEE, 2657–2663.

Nipen, T. N., G. West, and R. B. Stull, 2011: Updating short-term probabilistic weather forecasts of continuous variables using recent observations. *Wea. Forecasting,* **26,** 564–571, doi:10.1175/WAF-D-11-00022.1.

Persson, A., 1991: Kalman filtering—A new approach to adaptive statistical interpretation of numerical meteorological forecasts. Lectures presented at the WMO Training Workshop on the Interpretation of NWP Products in Terms of Local Weather Phenomena and Their Verification, H. R. Glahn et al., Eds., WMO/TD 421, WMO PSMP Rep. Series 34, XX-27–XX-32. [Available online at http://library.wmo.int/pmb_ged/wmo-td_421.pdf.]

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174, doi:10.1175/MWR2906.1.

Roebber, P. J., 2010: Seeking consensus: A new approach. *Mon. Wea. Rev.,* **138,** 4402–4415, doi:10.1175/2010MWR3508.1.

——, 2013: Using evolutionary programming to generate skillful extreme value probabilistic forecasts. *Mon. Wea. Rev.,* **141,** 3170–3185, doi:10.1175/MWR-D-12-00285.1.

——, 2015a: Evolving ensembles. *Mon. Wea. Rev.,* **143,** 471–490, doi:10.1175/MWR-D-14-00058.1.

——, 2015b: Using evolutionary programs to maximize minimum temperature forecast skill. *Mon. Wea. Rev.,* **143,** 1506–1516, doi:10.1175/MWR-D-14-00096.1.

Ross, G. H., 1987: An updateable model output statistics scheme. Extended abstracts of papers presented at the WMO Workshop on Significant Weather Elements Prediction and Objective Interpretation Methods, WMO PSMP Rep. Series 25, 25–28.

Simonsen, C., 1991: Self adaptive model output statistics based on Kalman filtering. Lectures presented at the WMO Training Workshop on the Interpretation of NWP Products in Terms of Local Weather Phenomena and Their Verification, H. R. Glahn et al., Eds., WMO/TD 421, WMO PSMP Rep. Series 34, XX-33–XX-37.

Wilson, L. J., and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) system: Design and development tests. *Wea. Forecasting,* **17,** 206–222, doi:10.1175/1520-0434(2002)017<0206:TCUMOS>2.0.CO;2.