

유전 프로그래밍 기반 단기 기온 예보의 보정 기법

Genetic Programming Based Compensation Technique for Short-range Temperature Prediction

현 병 용* · 현 수 환** · 이 용 희*** · 서 기 성†

(Byeongyong Hyeon · Soohwan Hyun · Yonghee Lee · Kisung Seo)

Abstract - This paper introduces a GP(Genetic Programming) based robust technique for temperature compensation in short-range prediction. Development of an efficient MOS(Model Output Statistics) is necessary to correct systematic errors of the model, because forecast models do not reliably determine weather conditions. Most of MOS use a linear regression to compensate a prediction model, therefore it is hard to manage an irregular nature of prediction. In order to solve the problem, a nonlinear and symbolic regression method using GP is suggested. The purpose of this study is to evaluate the accuracy of the estimation by a GP based nonlinear MOS for 3 days temperatures in Korean regions. This method is then compared to the UM model and has shown superior results. The training period of 2007-2009 summer is used, and the data of 2010 summer is adopted for verification.

Key Words : Temperature prediction, MOS(Model Output Statistics), Genetic programming, Compensation

1. 서 론

기상은 불확실성과 혼돈성이 큰 자연계의 현상이므로, 이를 예보하는 것은 매우 어려운 일이나, 연산 능력의 향상과 수치 모델 및 예측 기법의 발달로 인해 기상 예측 능력이 크게 향상되어 왔다. 특히, 다양한 종류 및 대규모의 기상 데이터를 체계적으로 처리할 수 있는 여러 예측 모델이 개발되어 현업에 사용되고 있다[1]. 이들 중 UM(Unified Model)은 영국에서 개발되어 전 세계적으로 널리 사용되는 모델로 현재 기상청에서도 운영되고 있다[2]. 그러나, 기상 시스템에 대한 정확한 모델링이 어렵기 때문에 기상예보 모델만으로는 예보의 정확성에 한계가 있다.

예보 모델의 오차를 줄이기 위하여, 수치예보모델 자료의 통계적 분석을 통해 수치예보 모델이 가지는 체계적인 오차를 보정하는 MOS(Model Output Statistics)[1] 기법이 사용되고 있으며, 관련 연구들이 다음과 같이 진행되어 왔다.

Crochet은 적응형 칼만 필터를 이용하여 예보 온도(2m)와 예보 풍속(10m)의 계통오차 감소 및 정확도 향상에 관한 연구를 수행하였으며[3], Carvalho 등은 칼만 필터를 이용하여 PRECIS(Providing REgional Climates for Impacts Studies) 모델로부터 예보된 온도의 계통 오차를 줄이는 연구를 수행하였다[4]. Ustaoglu 등은 ANN(Artificial Neural

Network)를 이용하여 일 평균, 최소, 최대 온도를 예보하는 연구를 수행하였다[5]. 또한 20여개의 주요 예보인자 선택하여 적용한 UMOS(Updatable MOS) 연구와, 고해상도의 격자형 MOS 개발 연구등이 수행되었다[6,7]. 그러나, 이러한 연구들은 예보인자들의 탐색 영역이 제한적이거나, 보정 수식이 간단하다는 단점이 있다. 또한, 현재 기상청의 동네예보 단기예보 기온 가이드스에서 사용되는 MOS에도 1차 선형식으로 구성된 보정법이 사용되고 있다.

그러나 이러한 방식들은 기온 예보 시스템이 가지고 있는 비선형적 특성을 고려하지 못한다. 또한 예보의 보정이 다변수 문제를 포함하며, 탐색 공간이 다봉인(multimodal) 특성을 가지고 있기 때문에, 기존의 수식적인 최적화 방법의 적용도 쉽지 않다.

따라서 예보 모델의 복잡한 비선형성과 다변수 및 다봉(multimodal) 함수 문제에 대처할 수 있는 최적화 방법이 요구된다[8]. 본 논문에서는 상기의 문제점들을 해결하기 위해서 유전 프로그래밍(GP, Genetic Programming[9,10])을 이용하여 잠재적 예보 인자들의 비선형 결합에 의한 보정 수식 생성 기법을 제안한다. 이 기법은 차수와 계수가 열린(open-ended) 공간의 탐색에 적합하므로, 기온 예측에 대해서 제한되지 않는 비선형 회귀분석 방식의 접근이 가능하다. 이를 통해 기존의 차수가 고정된 선형적인 회귀분석 접근법의 단점인 탐색의 한계성을 해결할 수 있다.

부가적으로 삼각함수 등의 주기적인 함수를 회귀분석 식 생성의 구성요소로 사용할 수 있기 때문에 기온 현상 등에 나타나는 변형된 주기성 등을 표현하는데 우수하다고 볼 수 있다. 우리나라 전역에 대해서 제안된 GP 기반 보정기법의 실험 결과를 UM 모델과 비교시 평균적으로 23.5%에서 최대 47%까지 오차에 대한 개선효과를 보여주고 있다.

* 비 회 원 : 서경대학교 전자공학과 석사과정

** 비 회 원 : 현대중공업 기전연구소 연구원

*** 비 회 원 : 국립기상연구소 기상연구관

† 교신저자, 정회원 : 서경대학교 전자공학과 교수

E-mail : ksseo@skuniv.ac.kr

접수일자 : 2012년 8월 9일

최종완료 : 2012년 10월 20일

2. UM과 MOS

동네 기상 예보는 5 km 해상도를 가진 149 × 253 개의 격자에 대하여 수행되며, 육상 격자는 총 9,291 개로, 한반도 뿐만 아니라 제주도, 울릉도, 독도 등을 모두 포함한다. 동네 기상 예보는 한번에 +06h부터 +66h 까지 3시간 간격의 예보 인자를 생성하며, 하루에 두 번(00 UTC, 12 UTC) 수행된다.

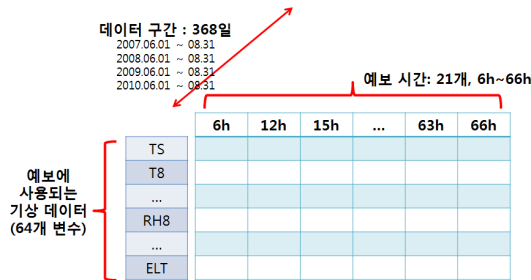


그림 1 기존 UM 의 예보 모델 자료 구조
 Fig. 1 Data structure of existing UM model

그림 1과 같이 UM에 기초한 기존 예보 모델은 향후 66 시간(3시간 간격, 21 개)의 기상 데이터들을 예측한다. 기존 예보 모델로부터 예측되는 기상 데이터는 총 64개로, TS는 지상의 온도를 의미한다. 이러한 예측 데이터들은 하루에 2 번 생성되는데, 각각 00UTC, 12UTC에 생성된다.

표 1은 현재 기상청에서 사용되는 수치 예보 모델인 UM(Unified Model)의 예보인자와 해당 코드를 나타낸 것이다. 각각의 예보 인자들은 예보 고도에 따라서 1~4개로 구분된다.

- 예보 대상 위치 : 9,291 지점
- 예보 시간 : +06h ~ +66h (21 개)
- 예보 생성 시점 : 00 UTC, 12 UTC(2 개)

MOS(Model Output Statistics)는 수치예보모델의 오차를 줄이는 예보기술로서, MOS는 수치예보모델의 계통오차를 설명할 뿐만 아니라, 예보시간대마다 적합한 예보인자를 선택하여 표본의 평균에 가까운 예보를 생산할 수 있고, 다양한 예보요소에 대한 접근성이 용이하다.

기존의 MOS에서 사용되는 지상 기온 예보 보정 기법은 식 (1)과 같이 잠재적 예보인자들과 계수의 곱합을 통한 선형적인 수식으로 구성되어 있다. 여기서 TS'은 보정된 지상 예보 온도이고, VAR_i는 표 1의 잠재적 예보인자들 중 선택된 인자들로서, 실험등을 통해 지상 온도와 상관관계가 높은 인자들로 구성될 수 있으나, 이들 사이의 상관관계가 명확히 알려져 있지 않다. 그리고 a는 각 예보인자의 계수이다.

$$TS' = a_1 VAR_1 + a_2 VAR_2 + \dots + a_N VAR_N \quad (1)$$

표 1 잠재적 예보인자

Table 1 Potential predictors

Types	Potential Predictors
Air Temperature	TS, T8, T7, T5
Thickness	DZ18, DZ17, DZ85
Dew-point	TDD8, TDD7, TDD5
Specific humidity	QS, Q8, Q7, Q5
Difference between specific humidity and saturated specific humidity at 500 hPa	DQ85, DQ75
Relative humidity	RH8, RH7, RH5
Layer averaged RH	MRH17, MRH15, MRH85
Zonal wind	US, U8, U7, U5
Meridional wind	VS, V8, V7, V5
Wind speed	WSS, WS8, WS7, WS5
Wind direction	WDS, WD8, WD7, WD5
North-westerly wind speed	NWS, NW8, NW7, NW5
North-easterly wind speed	NES, NES8, NES7, NE5
Lapse rate	LR87, LR85
Total rain amount (3hr accumulated)	PCP

3. 유전 프로그래밍 기반 기온 예보의 보정 기법

3.1 유전 프로그래밍

본 연구에 사용된 최적화 알고리즘은 유전 프로그래밍(GP, Genetic Programming)[9,10]으로서, 유전 알고리즘(GA, Genetic Algorithm)과 달리 개체를 트리 형태로 표현하는 특징을 가지고, 유전자 크기가 가변이며, 구조적으로 열린 공간의 탐색 문제에 적합하다. 대표적으로 널리 적용되는 분야는 회귀분석으로 구성요소가 제한되지 않는 다항식을 생성할 수 있다.

GP의 수행과정은 다음과 같다. 사전에 유전자를(트리) 구성하는 함수와 터미널을 정의한다. 함수와 터미널은 해를 조합적으로 구현할 수 있는 요소들을 선택한다.

초기에 트리로 구성된 각 개체들을 임의로 생성한다. 그리고, 각 개체 트리를 해석하여 구한 후보 해를 적합도 함수로 평가한다. 이후, 주어진 선택 방법에 의해 유전 연산에 참여할 개체들을 선택한다. 선택된 개체들을 대상으로 유전 연산(교배, 돌연변이) 수행한다. 그리고 이 전체과정을 종료조건이 만족될 때 까지 반복한다. GP에서의 교배와 변이 연산자는 아래 (그림 2와 3처럼 선택점에서 서브트리를 교체하는 방식으로 수행된다.

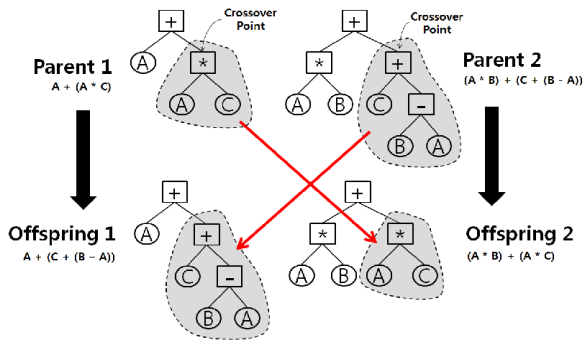


그림 2 GP의 Crossover 연산
Fig. 2 Crossover operation of GP

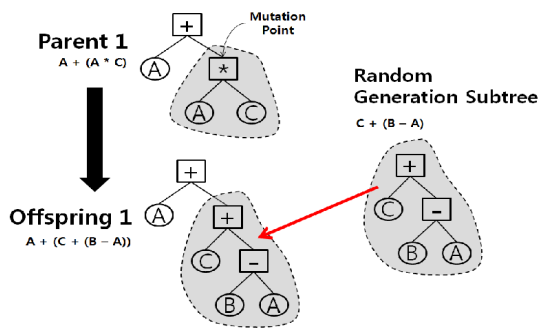


그림 3 GP의 Mutation 연산
Fig. 3 Mutation operation of GP

3.2 유전프로그래밍 기반 기온 예보 보정 기법

본 연구에서 제안된 유전 프로그래밍(이하 GP) 기반의 지상 기온 예보의 보정 기법의 개념도가 그림 4에 나와 있다.

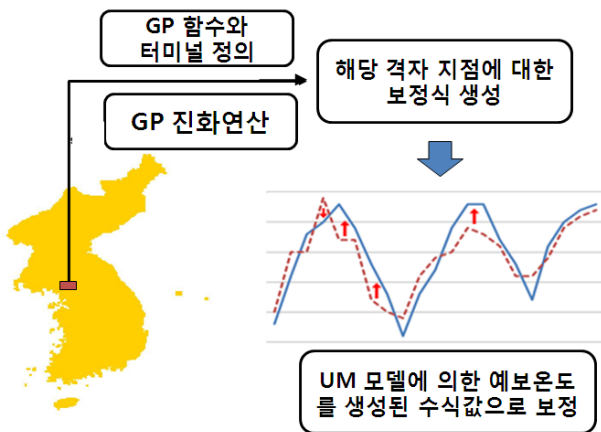


그림 4 GP 기반 예보온도 보정 기법
Fig. 4 GP based compensation technique for temperature forecast

사전 과정으로 GP 트리 생성을 위한 터미널과 함수는 다음과 같이 정의된다. 먼저 GP의 터미널은 표 1의 49개 예보

인자와 실수 ERC로 이루어지며, 각 예보 인자에 대해서는 정규화가 수행된다. 함수는 사칙연산(+, -, *, /)과 삼각함수(cosine, sine)가 사용된다.

3.1 절에서 설명한 진화연산의 초기 생성, 평가, 선택 그리고 GP 유전연산의 반복 수행을 거쳐서 각 예보 지점마다 적합도가 가장 우수한 개체를 구한다. 얻어진 개체는 트리의 형태를 가지고 있기 때문에 이를 깊이우선으로 해석하여 보정식을 생성한다. UM 을 통해 예측된 TS(기온)에 구해진 보정 수식값을 추가하여 식 (2)와 같이 최종 기온 TS'을 예보한다.

$$TS' = TS + \text{보정식} \tag{2}$$

이해를 돕기 위해서 그림 5에 보정식의 가상적인 한 예를 표현하였다. 실험에서 결과해로 생성된 GP 트리는 크기가 방대하기 때문에 설명을 위하여 깊이가 4인 임의의 트리를 예로 들었다. 이 트리를 깊이우선으로 해석하여 보정식을 구한다. 그림 4에서 볼 수 있듯이 트리의 각 노드에 사칙연산 및 초월함수등을 포함할 수 있어서 선형식으로 구성된 보정식에 비해서 다양한 비선형성 및 주기적인 양상을 나타낼 수 있다.

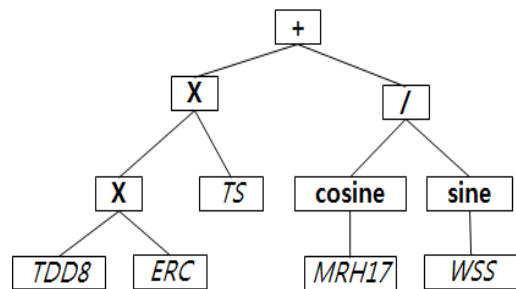


그림 5 예보온도 보정식의 설명 예
Fig. 5 Illustrative example of forecast for temperature compensation formula

GP 기반의 지상 온도 보정 수식의 생성에 대한 최적화는 식 (3)과 같은 적합도 함수를 정의하여 사용하여 평가하였다. 여기에서 KLAPS(Korea Local Analysis and Prediction System)는 실험 지상 온도, GP는 식 (2)를 통해 보정된 지상 예보 온도, Days는 평가 기간이다. 적합도는 각 날짜에 대해서 KLAPS(실험 온도) 값과 GP 진화연산에 의해 얻어진 보정 온도 값의(식 (2)의 TS'에 해당) RMSE 오차를 누적하여 평균을 구한 것이다.

$$Fitness = \sqrt{\frac{\sum_{i=0}^{Days} (KLAPS_i - GP_i)^2}{Days}} \tag{3}$$

학습 데이터의 평가 기간은 2007년~2009년의 여름(6~8월)으로, 총 276일(00 UTC는 274일)이다. 검증 데이터의 평가 기간은 2010년 6~8월로, 총 92일이다.

4. 실험 및 결과

4.1 유전프로그래밍 파라미터 설정

GP 파라미터는 다음과 같이 설정하였으며, 각각의 실험은 최대 20회 반복 수행하였다. GP기반 비선형 MOS 수식은 총 390,222(9,291×21×2)개다.

- Population size: 300
- Generation: 300
- Initial Tree Depth: 2-5
- Initial Tree Method: Half and Half
- Max Depth: 9
- Crossover Rate: 0.9
- Mutation Rate: 0.1
- Selection: Tournament(size = 7)

4.2 GP 기반의 예보 온도 보정 실험

GP 기반의 지상 온도 보정 실험은 여름에 해당하는 학습 데이터(2007~2009년의 6~8월)와 검증 데이터(2010년 6~8월)를 이용하여 수행하였다.

표 2는 GP 기반의 지상 온도 보정 수행 결과이다. GP 최적해는 전체 390,222개의 검증 실험 대상 중 98%에 해당하는 382,437개 실험 대상에서 UM보다 성능 우위를 보였다. 또한, RMSE가 가장 크게 나타난 실험에서 GP는 UM의 RMSE 대비 약 51.6%의 개선 효과를 보였다. 평균적으로 GP는 UM의 RMSE 대비 00 UTC에서 약 24%, 12 UTC에서 약 23%의 개선 효과를 보였다.

표 2 UM과 GP의 지상 예보 온도 RMSE 비교

Table 2 Comparison of forecast temperature RMSE using UM and GP

		학습 RMSE(K) (’07.6~8, ’08.6~8, ’09.6~8)			검증 RMSE(K) (’10.6~8)		
		최소	평균	최대	최소	평균	최대
00 UTC	UM	0.863	2.035	7.822	0.718	2.059	7.416
	GP	0.732	1.512	3.514	0.673	1.565	3.344
12 UTC	UM	0.830	2.049	7.775	0.659	2.066	7.403
	GP	0.591	1.547	3.705	0.643	1.591	3.585

그림 5는 00 UTC와 12 UTC에 대한 검증 실험의 평균 RMSE를 예보 시간별로 나타낸 것이다. 00 UTC의 실험 결과 다른 예보 시간대에 비해 낮 시간대(+06h, +27h~30h, +51h~54h)에 RMSE가 높게 나타났다. 검증 실험에서의 평균 RMSE 고저차(RMSE 최댓값 - RMSE 최솟값)는 UM: 1.435, GP: 1.104으로, 약 0.331 정도 개선된 것을 확인하였다.

12 UTC에 대한 검증 실험 결과, 평균 RMSE 고저차는 UM : 1.656, GP : 1.33으로, 약 0.326의 개선효과를 보였다. GP의 RMSE 향상률은 00 UTC가 크게 나타난 반면, RMSE 고저차의 개선 효과는 12 UTC가 크게 나타났다.

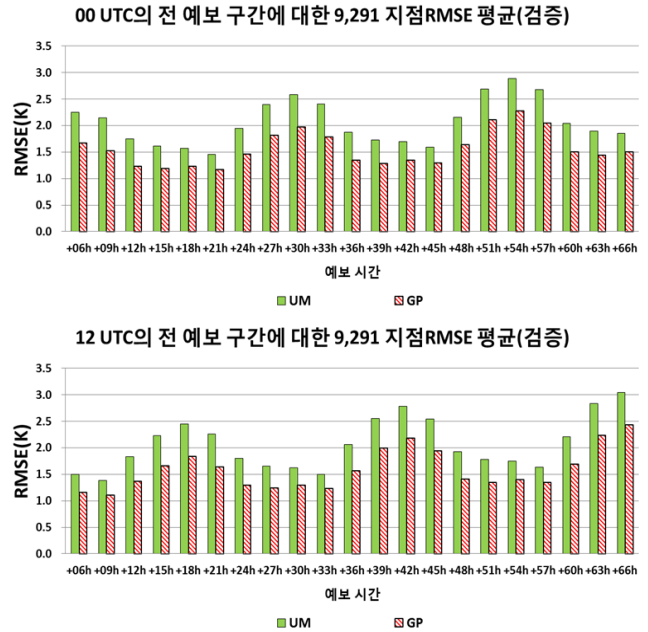


그림 5 00UTC(위), 12UTC(아래)에서의 평균 RMSE 비교
Fig. 5 Comparison of average RMSE in verification for 00UTC(top) and 12UTC(bottom)

4.3 예보 온도 보정 수식의 예

아래의 수식 (4)는 051.107 지점의 +09h(00UTC) 예보 온도 보정에 사용되는 GP 트리이다. 각각의 기상데이터는 정규화를 수행하였다. 최종 예보 온도는 TS + GP_TREE 이다. 051.107 지점의 +09h(00UTC) GP 트리는 63개의 노드와 9의 깊이를 가지며, sin 함수와 +, -, * 연산이 사용되었다. cos 함수와 / 연산은 GP 진화과정에서 선택되지 않았다. 또한, 평균습도인 MRH 계열과 기온감률인 LR 계열의 터미널이 다수 선택된 특징이 나타난다.

$$\begin{aligned}
 \text{GP_TREE} : & \{ \sin(\sin(((\text{WS8}+\text{WS8}-\text{TS}) * (\text{LR87}-\text{RH7})) + \\
 & \sin(\text{KI}-\sin(\text{TS}))) + \sin(\text{US}+\text{DZ18}-\sin(\text{V8}))) \} + \\
 & \{ \sin(\sin(\sin(\text{LR87}-\text{MRH85}+\text{LR87}-\text{MRH15}))) \} + \\
 & \{ \sin(\sin(\text{LR87}-\text{MRH85}+\text{DZ17}-\text{MRH15}+\text{DZ17}-\text{MRH15}))) \} \\
 & + \{ \sin(\sin(\sin(\sin(\sin(\text{V7})))))) \} - \sin(\sin(\text{TS})) - \text{U7}
 \end{aligned}
 \tag{4}$$

그림 6은 앞에서 소개한 051.107 +09h(00UTC)의 GP 트리를 이용하여 전체 실험 데이터('07. 6 ~ '10. 8)의 051.107 +09h(00UTC) 지상 온도(TS)를 보정한 결과와 보정 온도의 오차를 표현한 것이다. 학습 데이터('07. 6 ~ '09. 8)의 RMSE는 UM : 3.587, GP : 1.478이며, 이는 UM 대비 58.8% 향상된 결과이다. 검증 데이터('10. 6 ~ '10. 8)의 RMSE는 UM : 4.358, GP : 1.22으로, UM 대비 72.01 % 향상된 결과이다.

그림 7은 전체 실험 데이터에서의 GP 트리에 의한 051.107 +09h(00UTC)의 예보 온도와 KLAPS 실험 온도의 오차, UM과 KLAPS 실험 온도의 오차를 보여준다. 전체 실험 데이터에 대하여 UM의 절대값 오차는 최소 : 0.0062, 최대 : 10.1048, 평균 : 3.224 이었으며, GP 트리는 최소 :

0.00097, 최대 : 5.2596, 평균 : 1.1164의 오차를 보였다.

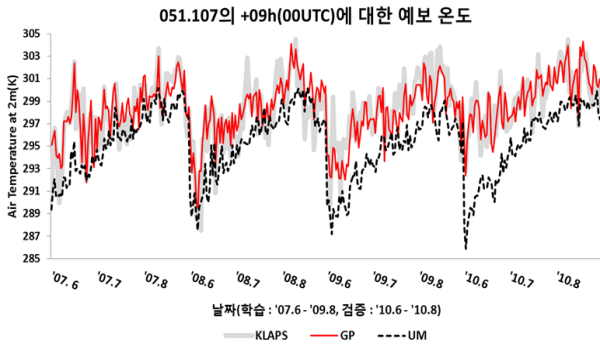


그림 6 051.107(+09h, 00UTC)의 예보온도
Fig. 6 Forecast temperature in 051.107(+09h, 00UTC)

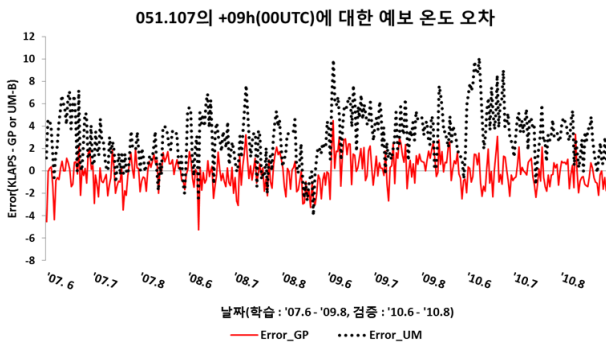


그림 7 051.107(+09h, 00UTC)의 예보온도 오차
Fig. 7 Error of forecast temperature in 051.107(+09h, 00UTC)

수식 (5)는 045.058 지점의 +33h(12UTC) 예보 온도 보정식에 사용된 GP 트리이며, 69개의 노드와 9의 깊이를 갖는다. 선택된 함수는 cos, sin 함수, +, *, - 연산으로 구성되어 있고, 터미널로 특정고도의 온도인 T7 과 북서풍 NW가 상당수 포함되어 있다.

<p>GP_TREE :</p> $\begin{aligned} & \{ (LR87 * T7 * T7 * T7 * T7 * LR87 * T7 * T7 * T7 * T7) \\ & + WD8 - T5 + (LR87 * T7 * T7 * T7) + DQ85 - T5 \\ & + (-0.29985) - T5 - TS + NW7 \} * \\ & \{ \cos(\sin(\{ \sin(Q7) + \cos(TS + NW7) + \sin(Q7) - TS + NW7 \} \\ & * \{ \cos(WD8 - \cos(\cos(T7 + WSS) \} \} \} \} \end{aligned}$	(5)
---	-----

그림 8은 앞에서 소개한 045.058 지점의 +33h(12UTC) GP 트리를 이용하여 전체 실험 데이터('07. 6 ~ '10. 8)의 051.107 +09h(00UTC) 지상 온도(TS)를 보정한 결과이다. 학습 데이터('07. 6 ~ '09. 8)의 RMSE는 UM : 1.787, GP : 1.293이며, 이는 UM 대비 27.64 % 향상된 결과이다. 테스트 데이터('10. 6 ~ '10. 8)의 RMSE는 UM : 1.171, UM-B : 1.111, GP : 1.061으로, UM 대비 5.12 % 향상된 결과이다.

학습에 의한 RMSE 감소율과 테스트 데이터에서의 RMSE 감소율에는 큰 차이가 나타나는 것을 확인하였다.

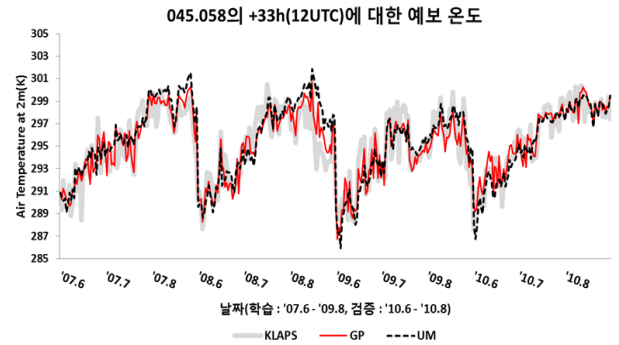


그림 8 045.058(+33h, 12UTC)의 예보온도
Fig. 8 Forecast temperature in 045.058(+33h, 12UTC)

그림 9는 전체 실험 데이터에서의 GP 트리에 의한 045.058 +33(12UTC)의 예보 온도와 KLAPS 실험 온도의 오차, UM과 KLAPS 실험 온도의 오차를 보여준다. 전체 실험 데이터에 대하여 UM의 절대값 오차는 최소 : 0.0067, 최대 : 6.5456, 평균 : 1.2997 이었으며, GP 트리는 최소 : 0.00368, 최대 : 4.1657, 평균 : 0.9871의 오차를 보였다.

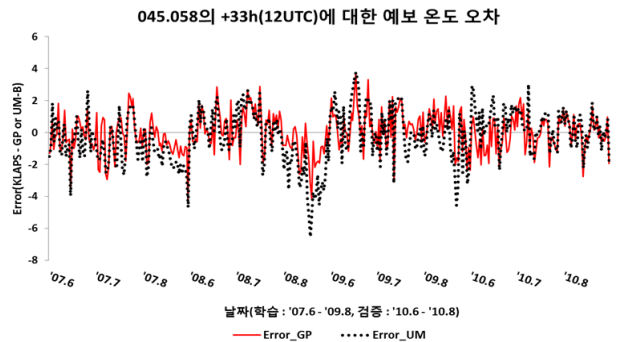


그림 9 045.058(+33h, 12UTC)의 예보온도 오차
Fig. 9 Error of forecast temperature in 045.058(+33h, 12UTC)

4.4 RMSE 분포도

표 3과 그림 10은 3개 예보 시간(00 UTC)의 검증 데이터에 대한 RMSE와 RMSE 분포도이다. 4.2 절의 실험결과가 06시부터 66시까지 3시간 간격으로 21시점에 대한 예보 결과를 보여준 데 반하여, 아래의 분포도는 21시, 30시, 54시의 3개 시점에 대한 RMSE 오차를 한반도 영역에 대해서 색정보를 그래픽으로 나타낸 것이다.

+21h에 대한 실험은 UM과 GP 모두 00UTC의 다른 예보 시간에 대한 실험 결과보다 RMSE가 가장 좋은 것으로 나타났다. +21h에 대한 UM의 평균 RMSE는 1.4546이고, GP 트리의 평균 RMSE는 1.1711이다. 전체 격자의 RMSE 표준편차는 UM : 0.4159, GP : 0.2309로, GP가 더욱 안정적인 성능을 보였다. +30h와 +54h는 오후 3시(+30h는 하루 뒤 오후 3시, +54h는 이틀 뒤 오후 3시)를 의미하며, 두 시간대의 지상 예보 온도 RMSE는 다른 예보 시간대의 실험 결과보

다 높게 나타났다. 한 격자점에서 UM의 최대 RMSE는 각각 7.2630, 7.4028로, 전체 21개 예보 시간 오직 이 예보 시간에서만 최대 RMSE가 7이상으로 나타났다. 이는 +30h, +54h의 RMSE 분포도에서 빨간색으로 표시되며, 주로 서해안 지역에서 발생하는 것을 확인할 수 있다. 그러나 GP는 최대 RMSE가 각각 3.1785, 3.5849로 UM의 최대 RMSE보다 많이 개선되었으며, UM의 RMSE 분포도에서 나타난 빨간색 영역들이 GP의 RMSE 분포도에서는 나타나지 않았다.

표 3 00 UTC(+21h, +30h, +54h)의 전체 격자에 대한 예보 온도 RMSE 비교

Table 3 Comparison of forecast temperature RMSE for 00UTC(+21h, +30h, +54h)

00UTC		RMSE			
		min	avg	max	dev
+21h	UM	0.7185	1.4546	4.7964	0.4159
	GP	0.6734	1.1711	2.1943	0.2309
+30h	UM	1.3442	2.5870	7.1925	0.7619
	GP	1.0115	1.9758	3.0114	0.2094
+54h	UM	1.4640	2.8892	7.4158	0.7189
	GP	1.1990	2.2755	3.3441	0.2559

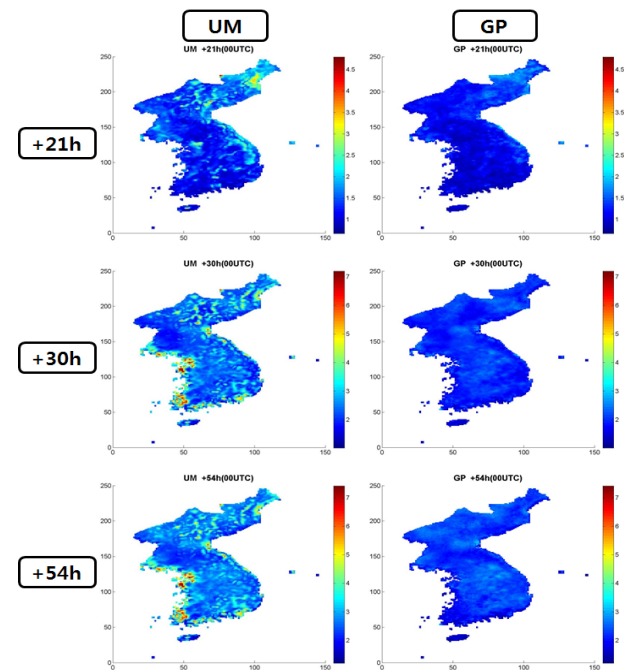


그림 10 UM(좌)과 GP(우)의 RMSE 분포도 00UTC (+21h, +30h, +54h)

Fig. 10 RMSE distribution of UM(left) and GP(right) for 00UTC(+21h, +30h, +54h)

표 4와 그림 11은 12 UTC의 예보 시간 중 평균 RMSE가 가장 작게 나타난 +09h와 높게 나타난 +42h(이틀 뒤 오후 3시), +66h(3일 뒤 오후 3시)에 대한 RMSE와 RMSE 분포도이다. +09h에 대한 UM의 평균 RMSE는 4.9043이고,

GP의 평균 RMSE는 2.1027이다. 전체 격자의 RMSE 표준편차는 UM : 0.4548, GP : 0.2416으로, GP가 더욱 안정적인 성능을 보였다. 한 격자점에서 UM의 최대 RMSE는 각각 7.7345, 7.256으로 +30h, +54h의 RMSE 분포도에서 빨간색으로 표시되며, 그 범위가 그림 7에서의 +30h와 +54h의 것과 비슷하게 서해안 지역에서 발생하는 것을 확인할 수 있다. GP의 최대 RMSE는 각각 3.1785, 3.5949으로 나타났다.

표 4 12 UTC(+09h, +42h, +66h)의 전체 격자에 대한 예보 온도 RMSE 비교

Table 4 Comparison of forecast temperature RMSE for 12UTC(+09h, +42h, +66h)

12UTC		RMSE			
		min	avg	max	dev
+09h	UM	0.6588	1.3892	4.9043	0.4548
	GP	0.6429	1.1032	2.1027	0.2416
+42h	UM	1.4281	2.7824	7.2630	0.7345
	GP	1.0533	2.1871	3.1785	0.2474
+66h	UM	1.5080	3.0448	7.4028	0.7256
	GP	1.1476	2.4333	3.5849	0.3506

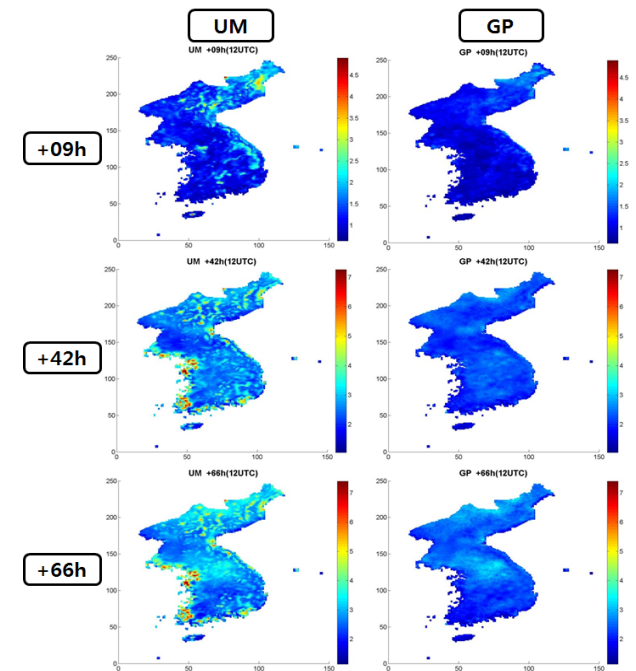


그림 11 UM(좌)과 GP(우)의 RMSE 분포도 12UTC(+09h, +42h, +66h)

Fig. 11 RMSE distribution of UM(left) and GP(right) for 12UTC(+09h, +42h, +66h)

기상 연구의 특성상, 기존 MOS에서 사용된 인자들이 구체적으로 주어지지 않기 때문에, 기존 결과와 직접적인 비교가 불가능 하였다. 따라서 기존의 최신 연구결과와[5] 수치상으로 비교를 하였는데, 같은 여름 기간에 대한 평균 RMSE 값에서 UMOS 결과는 2.115, MOS_R 결과는 1.77

로 나온 것에 비해, GP 는 평균 1.578 값을 보이고 있다. 따라서 기존 UMOS 결과 대비 약 25%, MOS_R 대비 11%의 개선 효과를 보였다.

5. 결 론

선형적인 기존 MOS를 개선하기 위해서, 유전 프로그래밍(GP) 기반의 회귀분석기법을 응용한 잠재적 예보 인자들의 비선형적인 수식 생성 기법을 제안하였다. GP 진화연산을 통해 얻어진 예보 보정식들은 적합도가 높은 인자의 선택과 초월함수와의 다양한 결합으로 구성됨으로써, 우수한 예보 보정 기능의 구현이 가능하다.

개발된 기법은 전국 9,291개의 격자점에 대해 00 UTC와 12 UTC 두 예보 생성 시점 별로 각각 +6~+66 시간까지의 21개 예보 시간에 대해 적용되었다. 학습에는 2007~2009년의 KLAPS 데이터를 사용하고, 검증은 2010년도의 데이터를 통해 이루어졌다. 제안된 GP 기반 보정기법을 UM 모델과 비교시 평균적으로 23.5%에서 최대 47%까지 오차에 대한 개선효과를 보여주고 있다.

참 고 문 헌

[1] Korean Meteorological Society, Introduction to Atmospheric Science, Sigma Press, 2009.

[2] United Kingdom Met Office's website : <http://www.metoffice.gov.uk>

[3] P. Crochet, "Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland," Meteorological Applications, vol. 11, no. 2, pp. 173-187, 2004.

[4] J. R. P. Carvalho, E. D. Assad, H. S. Pinto, "Kalman filter and correction of the temperatures estimated by PRECIS model," Atmospheric Research, vol. 102, no. 1-2, pp. 218-226, 2011.

[5] B. Ustaoglu, H. K. Cigizoglu, M. Karaca, "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods," Meteorological Applications, vol. 15, no. 4, pp. 431-445, 2008.

[6] J. Kang, M. Suh, K. Hong, C. Kim, "Development of updateable Model Output Statistics (UMOS) System for Air Tmperature over South Korea," Asia-Pacific Journal of Atmospheric Sciences, vol. 47, no. 2, pp. 199-211, 2011.

[7] X. Yu, SK. Park, YH. Lee, KD. Ahn, SH. Choo, "The gridding of MOS for high resolution forecasting," The fifth Korea-Japan-China Joint Conference on Meteorology, pp. 18-21, 2011.10.

[8] B. Hyeon, S. Hyun, K. Seo, Y. Lee, "Compensation Technique in Short-range Temperature Prediction using Genetic Programming," 2012 Proceedings of Information and Control Symposium, pp. 157-158,

2012. 4.

[9] Koza, J. R., Genetic Programming: On the Programming of Computers by Means of Natural Selection, The MIT Press, 1992.

[10] W. B. Langdon, R. Poli Foundations of Genetic Programming, Springer, 2002.

감사의 글

이 논문은 2012년도 국립기상연구소 예보기술지원 및 활용연구의 지원을 받아 수행된 연구임. (NIMR 2012-B-1)

저 자 소 개



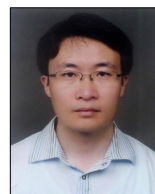
현 병 용 (玄柄龍)

2011년 서경대학교 전자공학과 졸업. 2012년~현재 동 대학원 전자공학과 석사과정. 관심분야는 유전 알고리즘, 유전 프로그래밍, GPGPU, 영상인식.



현 수 환 (玄守桓)

2010년 서경대학교 전자공학과 졸업. 2012년 동대학원 전자공학과 졸업(석사). 2012년~현재 현대중공업 기전연구소 연구원. 관심분야는 진화연산, 지능로봇.



이 용 희 (李龍熙)

1997년 경북대학교 천문대기과학과 졸업(석사). 1998~2007년 기상연구소 기상연구사. 2007년~현재 국립기상연구소 기상연구관. 관심분야는 기상수치예보모델, 자료동화, 초단기 기상예보.



서 기 성 (徐奇成)

1993년 연세대학교 대학원 전기공학과 졸업(박사). 1999~2003년 Michigan State University Research Associate. 1993년~현재 서경대학교 전자공학과 부교수. 관심분야는 진화연산, 지능로봇.