УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ" - ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ISSN 1857- 8691

ГОДИШЕН ЗБОРНИК 2012

YEARBOOK 2012

ГОДИНА 1 VOLUME I

GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ“ – ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ГОДИШЕН ЗБОРНИК
2012
YEARBOOK
2012

ГОДИНА 1 **МАРТ, 2013** **VOLUME I**

GOCE DELCEV UNIVERSITY – STIP
FACULTY OF COMPUTER SCIENCE

ГОДИШЕН ЗБОРНИК
ФАКУЛТЕТ ЗА ИНФОРМАТИКА
YEARBOOK
FACULTY OF COMPUTER SCIENCE

Редакција и администрација	**Address of the editorial office**
Универзитет „Гоце Делчев“-Штип	Goce Delcev University – Stip
Факултет за информатика	Faculty of Computer Science
ул. „Крсте Мисирков“ 10-А	Krste Misirkov 10-A
п. фах 201, 2000 Штип	PO box 201, 2000 Štip,
Р. Македонија	R. of Macedonia

СОДРЖИНА
CONTENT

DEVELOPING CLOUD COMPUTING'S NOVEL COMPUTATIONAL METHODS FOR IMPROVING LONG-TERM WEATHER GLOBAL FORECAST

Zubov Dmytro
University for Information Science and Technology "St. Paul the Apostle"
dmytro.zubov@uist.edu.mk , dzubovcs@yahoo.com

Abstract:
Weather data mining methods and forecast algorithms have been of long standing interest. Recent research based on the global satellite data and special synergetic methods showed possibility of the long-term (up to half a year ahead) forecast with up to 10 % average mistake (standard is 20 %). Particularly, the average daily air temperature forecast's mistake is up to 6.5 % for Skopje Airport (half a year ahead). This approach is characterized by the final linear difference equations' simplicity and the high computational complexity of the above equations reasoning. The cloud computing web-site's prototype was developed (weatherforecast.tk). Main research proposals: improving the user interface based on 3D or/and ubiquitous computing technologies; developing new synergetic methods for the appropriate realization in the multithread cloud application, including the code and data parallelization; increase of the forecast parameters' quantity (e.g., precipitation). This paper main results are: precipitation's long-term (up to half a year ahead) forecast has very low quality now, and, therefore, it is not recommended for practice; the forecasting places' quantity is changed modifying the text file in the cloud application's package; the web-site http://weatherforecast.tk user interface was enhanced using 3D Chart diagram.

Keywords: cloud computing, inductive modeling, long-term average daily air temperature's forecast.

1 Introduction

Weather data mining methods and forecast algorithms have been of long standing interest because of high importance for noosphere. Recent research on the basis of the global satellite data and special synergetic methods showed possibility of the long-term (up to half a year ahead) forecast with up to 10 % average mistake (standard is 20 %). Particularly, the average daily air temperature forecast's mistake is up to 6.5 % for Skopje Airport (half year ahead). This approach is characterized by the final linear difference equations' simplicity and the high computational complexity of the above equations reasoning. The cloud computing web-site's prototype was developed (URI – http://weatherforecast.tk/) on the basis of this know-how. Some details of the above approach can be found in the papers which can be downloaded free against http://weatherforecast.tk/.

1.1 Previous results' analysis

There is currently no completed long-term weather forecast system. Several approaches for long term and short term weather forecasts have been suggested in [1, 2, 4]. However, precise forecast of the weather conforming to localised environment conditions is fraught with difficulties due to computational complexity of long-term forecasts requiring computing grid and/or clouds to compute the forecast values in reference points using the classical inductive analogue method for more than 80 % forecast quality at the average [3]. The problem focusing on enhancement in quality of the long- and short-term forecasts of weather processes, therefore, has drawn considerable attention [1]-[11]. As may be seen in [1]-[3] and Internet resources [7]-[9] that rise to this situation can be attributed to weather processes and noosphere interplay. It is well known that classical hydrodynamic equations are used for continuous modelling of these events. But, this approach depends on the variables variation (the theory catastrophe's known effect). In addition, discrete modelling is applied on the basis of analogue complexing algorithm [4]. Yet, experts note that the forecasts' quality (up to 80 % at the average) and forward (up to two weeks mainly) are not enough nowadays. A feature-specific forecasting method for high-impact weather events that takes advantage of high-resolution numerical weather prediction models and spatial forecast verification methodology was proposed in [11]. An application of this method to the prediction of a severe convective storm event was given only. But, the idea of the high-impact weather events' usage is considered is very effective. Furthermore, the standard software and hardware combined solutions with user-friendly interface have not been developed as yet.

1.2 Organization of the paper

Paper includes three main parts: average daily air temperature’s long-term forecast model’s synthesis, software’s brief description, daily precipitation’s correlation analysis.

This paper’s structure is based on the main prospects of the developing cloud computing’s novel computational methods for improving of the long-term weather global forecast:

1. Improving the user interface of the web-site http://weatherforecast.tk. Particularly, 3D user interface or/and ubiquitous computing technology are planned for realization.
2. Developing new synergetic methods for the appropriate realization in the multithread cloud application, including the code and data parallelization for the continuous updating of the mathematical models’ structures, increase of the forecasting places’ quantity, and up-to-date weather forecast.
3. Increase of the forecast parameters’ quantity. As may be seen in http://weatherforecast.tk/, the average daily air temperature’s long-term forecast was realized only. Precipitation is next very important factor for forecasting.

The research plan can be formulated as (1^{st} stage can be done in parallel):

1.a. Developing the user interface on the basis of 3D or/and ubiquitous computing technology.

1.b. Air temperature and precipitation’ global data mining and classification.

2. Increase of the forecast places’ quantity (the average daily air temperature parameter). Developing cloud computing multithread application’s prototype which includes the separate treads for the computationally complicated models (3) reasoning (models (3) are discussed below).
3. Developing the forecasting models for the precipitation’ long-term forecast.
4. Developing cloud computing multithread application (web-role) on the basis of Microsoft Windows Azure for long-term weather global forecast (the precipitation and the average daily air temperature parameters).

In addition, the scientific problem of the two-level hierarchical criterial system’s creation for the inductive forecasting method is proposed for solution optionally (as may be seen below, one criterion (4) is in use only).

The above mentioned cloud computing technology’s usage is grounded on the basis of Microsoft Windows Azure Educator Grant which one was received by the University for Information Science and Technology "St. Paul the Apostle".

Expected outcomes of this research include:

1. New up-to-date user interface of the web-site http://weatherforecast.tk on the basis of 3D or/and ubiquitous computing technology.
2. Advanced cloud software based on the code and data parallelization, which one allows to generate new, more effective global forecasting models on-line.
3. New forecasting models for different meteorological data (precipitation at least).

2 The average daily air temperature's long-term forecast model's synthesis

67 places took part in the correlation analysis initially (names were written according to the www7.ncdc.noaa.gov; places were chosen with 2 criterion: one country – one representative place; time series have to be continuous from 1st January, 1973): Nwso Agana, Aarhus Lufthavn, Abbeville, Aeropuerto Pettiros, Amman Airport, Amsterdam AP Schiph, Annaba, Ashgabat Keshi, Athinai al Helliniko, Auckland Airport, Bangkok Metropolis, Beijing, Ben-Guron International Airport, Beograd-Surcin, Bogota-Eldorado, Brasilia-Aeroporto, Bratislava-Letisko, Bruxelles National, Bucuresti INMH-Bane, Budapest-Ferihegy I, Busan, Cairo Airport, Canberra Airport, Caracas-Maiquetia, Damascus International Airport, Geneve-Cointrin, Gibraltar, Guernsey Airport, Helsinki-Vantaa, Hengchun, Jersey Airport, Kiev, Kingston-Norman Man, Kisinev, Kwajalein-Bucholza, La Paz-Alto, Lima-Callao Airport, Lisbon, London, Luqa, Luxembourg, Minsk, Moscow, Nassau Airport New, New Delhi-Safdarjun, Noumea-Nlle-Calledo, Nuuk, Oslo-Gardermoen, Paphos Airport, Praha-Libus, Rabat-Sale, Rarotonga, Reykjavik, Riga, Roma-Ciampino, Skopje Airport, Tallin-Harku, Tashkent, Tokyo, Torshavn, Tripoli, Tunis-Carthage, Ulaanbaatar, Vaduz, Warszawa-Okecie, Washington National, Wien-Hohe Warte. Data was downloaded in as the text files against www7.ncdc.noaa.gov (USA National Environmental Satellite, Data, and Information Services).

20 time series were selected as the most correlated to Skopje Airport with half-year (approximately) delay (correlation function is normalized and centralized):

0. Skopje Airport (delay 181, autocorrelation function's value -0.8327425 97188605).

1. Aeropuerto Pettiros (delay 187, correlation function's value 0.6230456 98552278).
2. Ashgabat Keshi (delay 184, correlation function's value -0.8516412838 20963).

3. Auckland Airport (delay 175, correlation function’s value 0.725715828 489775).
4. Canberra Airport (delay 182, correlation function’s value 0.7988013098 32135).
5. Gibraltar (delay 164, correlation function’s value -0.823452379725064).
6. Guernsey Airport (delay 163, correlation function’s value -0.7828445577 22741).
7. Jersey Airport (delay 167, correlation function’s value -0.78219805049 1036).
8. Lisbon (delay 165, correlation function’s value -0.761637258660198).
9. London (delay 174, correlation function’s value -0.781332533443968).
10. Nassau Airport New (delay 165, correlation function’s value -0.7676340 36673984).
11. New Delhi-Safdarjun (delay 199, correlation function’s value -0.84031 2570729328).
12. Noumea-Nlle-Calledo (delay 166, correlation function’s value 0.782559778633274).
13. Nuuk (delay 168, correlation function’s value -0.740870261967304).
14. Paphos Airport (delay 166, correlation function’s value -0.858819 233109924).
15. Rabat-Sale (delay 165, correlation function’s value -0.7843430 51359234).
16. Reykjavik (delay 172, correlation function’s value -0,73488389 2104582).
17. Tashkent (delay 183, correlation function’s value -0.833431861928644).
18. Tokyo (delay 168, correlation function’s value -0.855582219493774).
19. Washington National (delay 179, correlation function’s value -0.837 538640603375).

Average daily air temperature’s long-term forecast model has next linear structures:

$$\frac{X[i]}{\max\{X[i]\}} = k_0 + k_1 \frac{X_{j_1}[i]}{\max\{X_{j_1}[i]\}}, \quad (1)$$

$$\frac{X[i]}{\max\{X[i]\}} = k_0 + k_1 \frac{X_{j_1}[i]}{\max\{X_{j_1}[i]\}} + k_2 \left.\frac{X_{j_2}[i]}{\max\{X_{j_2}[i]\}}\right|_{j_2 \neq j_1}, \quad (2)$$

$$\frac{X[i]}{\max\{X[i]\}} = k_0 + k_1 \frac{X_{j_1}[i]}{\max\{X_{j_1}[i]\}} + k_2 \left.\frac{X_{j_2}[i]}{\max\{X_{j_2}[i]\}}\right|_{j_2 \neq j_1} + k_3 \left.\frac{X_{j_3}[i]}{\max\{X_{j_3}[i]\}}\right|_{\substack{j_3 \neq j_1 \\ j_3 \neq j_2}}, \quad (3)$$

where $X[i]$ – air temperature data; i – data position's number in time series, i=1, 2, 3, ..., 14198 (July19, 1973 – June 1, 2012); $X_{j_1}[i], X_{j_2}[i], X_{j_3}[i]$ – biased (with appropriate delay) time series for the appropriate places; k_0, k_1, k_2, k_3 = [-2;+2] – weighting coefficients (this range allows to find model with physical meaning); j_1, j_2, j_3=0,1,2,...,20 – number of the place in the above list. The main task is to find weighting coefficients k_0, k_1, k_2, k_3. We will use combinatorial (step is equal to 0.01) inductive modelling with next criterion (minimum of the regularity plus displacement):

$$\alpha_1 \frac{\sum_{\substack{i=1 \\ i\in B_1}}^{13680} \left| X^*[i] - X[i] \right|}{6840 \max\{X[i]\}} + \alpha_2 \frac{\sum_{\substack{i=1 \\ i\in B_2}}^{13680} \left| X^*[i] - X[i] \right|}{6840 \max\{X[i]\}} + \alpha_3 \frac{\left| \sum_{\substack{i=1 \\ i\in B_1}}^{13680} \left| X^*[i] - X[i] \right| - \sum_{\substack{i=1 \\ i\in B_2}}^{13680} \left| X^*[i] - X[i] \right| \right|}{6840 \max\{X[i]\}} \to \min$$

, (4)

where |.| – absolute value, B_1 – first learning sample (odd numbers); B_2 – second learning sample (even numbers); $X^*[i]$ – forecast values; $\alpha_1 = \alpha_2 = 1, \alpha_3 = 10$ – criteria's weighting coefficients.

Thus, a formula (1) has next view (temperature measures Fahrenheit degrees):

$$X^*[i] = 92.6\left(1.08 - 0.8\frac{X_2[i]}{100.8}\right). \quad (5)$$

A criterion (4) has next view on the learning sample (i=1, 2, ..., 13680):

0.0726133559941771 + 0.0726451595651796 +
+ 10 · | 0.0726133559941771 – 0.0726451595651796 | = 0,14557655 1269382

A criterion (4) has next view on the training sample (i=13681, 13682, ..., 14198):

0.0696785289719569 + 0.0671852780183542 +
+ 10 · | 0.0696785289719569 – 0.0671852780183542| = 0.16179631 6526338

A formula (2) has next view:

$$X^*[i] = 92.6\left(1.29 - 0.49\frac{X_2[i]}{100.8} - 0.54\frac{X_{11}[i]}{103.7}\right). \quad (6)$$

A criterion (4) has next view on the learning sample:

0.0648588883227404 + 0.0649056479360714 +
+ 10 · | 0.0648588883227404 – 0.0649056479360714 | = 0.13023213 2392122

A criterion (4) has next view on the training sample (*i*=13681, 13682, ..., 14198):

0.0638972281191006 + 0.0630384066284932 +
+ 10 · | 0.0638972281191006 - 0.0630384066284932 | = 0.13552384 9653668

It is clear that criteria's values are identical. Results' analysis shows that forecast model (6) is more precise than (5). I.e., we have the average daily air temperature forecast's mistake up to 6.5 % for Skopje Airport (half year ahead).
Model (3) is not discussed because of the high computational complexity (the main development perspectives in the multithread cloud application). E.g., model (2) calculation time is 2 days approximately on the basis of Intel® Core™ i5 processor.
The similar results were achieved for Beijing, China (mistake is up to 6.2 %), Kiev, Ukraine (up to 9.5 %), Moscow, Russia (up to 8.7 %), Tokyo, Japan (up to 5.7 %), and Washington National Airport, USA (up to 7 %). The forecasting places' quantity can be increased modifying the text file "CF.txt" in the cloud application's package.

3 Software's brief description

Web-site http://www.weatherforecast.tk/ was developed on the basis of ASP.NET technology initially, and, then, was moved to Microsoft Windows Azure web-role (C# is programming language). Web-site's screenshot is shown in Figure 1. User can choose place, date, and degree regime (Fahrenheit or Celsius) for forecast.

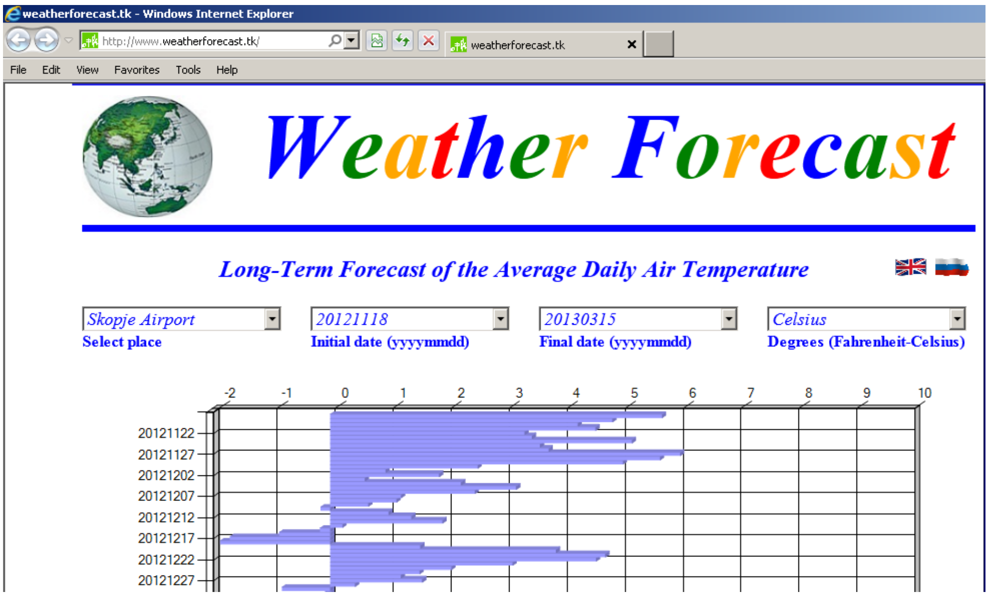

Figure 1 Web-site http://www.weatherforecast.tk/ screenshot

In praesenti, web-site code was improved. Firstly, it was speeded up using the partial-page update (UpdatePanel Control). Secondly, Chart Control's XmlDataSource file exception was handled using the set of the one-type files – file's title is changed in a loop. Thirdly, user interface was enhanced using 3D Chart diagram.

4 Skopje Airport daily precipitation's correlation analysis

Precipitation is a second important parameter in long-term weather forecast. Unfortunately, Skopje Airport daily precipitation's correlation analysis shows the practical inexpediency because of the low value of the correlation functions (up to 0.1 %; the pairs precipitation – precipitation and precipitation – temperature were considered). The similar results were achieved when the decade precipitation's correlation analysis is discussed [5]. I.e., the precipitation's long-term (up to half a year ahead) forecast has very low quality now, and, therefore, it is not recommended for practice.

5 Conclusion

In this paper, the cloud computing's novel computational methods for improving long-term weather global forecast were developed:

1. Precipitation's long-term (up to half a year ahead) forecast has very low quality now, and, therefore, it is not recommended for practice.

2. The forecasting places' quantity is changed modifying the text file "CF.txt" in the cloud application's package.
3. The web-site http://weatherforecast.tk user interface was enhanced using 3D Chart diagram.

At the same time, the main prospects for future research are following: the user interface based on the ubiquitous computing technology; developing new synergetic methods for the appropriate realization in the multithread cloud application, including the data parallelization for the continuous updating of the mathematical models' structures.

References

1. G.C. Onwubolu, P. Buryan, S. Garimella, V. Ramachandran, V. Buadromo, A. Abraham (2007): *Self-organizing Data Mining for Weather Forecasting*. IADIS European Conference Data Ming, pp. 81-88.
2. A.S. Cofino, J.M. Gutierrez (2003): *Implementation of Data Mining Techniques for Meteorological Applications, Realizing Teracomputing.* Ed. by W. Zwieflhofer and N. Kreitz, World Scientific Publishing Company, pp. 215-240.
3. D. Zubov, Y. Vlasov, M. Grigorenko (2008): *Method of the Decade Air's Temperature Long-Range Prognosis with Robust Inductive Models and Analogue Principle*. 2nd Int. Conf. on Inductive Modeling, Kyiv, Ukraine, Sept. 15-19, 2008, pp. 263-266.
4. G. Ivakhnenko (2008): *Short-Term Process Forecasting by Analogues Complexing GMDH Algorithm*. 2nd Int. Conf. on Inductive Modelling, Kyiv, Ukraine, Sept. 15-19, 2008, pp. 241-245.
5. D.A. Zubov, Y.N. Vlasov (2004): *Long-term Forecasting of the Air Average Temperature and Atmospheric Precipitations on the Linear Auto Regression Model and Maximal Error's Minimisation Criterion Basis*. Proceedings of Scientists of East Ukrainian National University named V.Dahl, Lugansk, Ukraine, pp. 40-49.
6. D. Zubov (2011): *Development of Web Application Structure for Weather Inductive Forecasting*. 4th International Workshop on Inductive Modelling (ICIM'2011), Kyiv, Ukraine, July 4-11, 2011, Kyiv, 2011, pp.123-127.
7. *Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency*. It was read on November 18, 2012. http://wind.geophys.tohoku.ac.jp/~sakai/lab/JMA.pdf
8. Handbook – *Use a Radio Spectrum for Meteorology*. [Online]. It was read on November 18, 2012. http://www.docstoc. com/docs/24344415/ HANDBOOK---Use-of-Radio-Spectrum-for-Meteorology

9. *Basic principles for the Weather Forecast Estimations and Calculations*. It was read on November 18, 2012. www.dvgu.ru/meteo/book /Synoptic/Glava_16.pdf
10. *WeatherXML Developers Kit*. It was read on November 18, 2012. http://www.weather.ua/services/xml/
11. Jacob R. Carley, Benjamin R. J. Schwedler, Michael E. Baldwin, Robert J. Trapp, John Kwiatkowski, Jeffrey Logsdon, Steven J. Weiss (2011): *A Proposed Model-Based Methodology for Feature-Specific Prediction for High-Impact Weather*. Weather and Forecasting Journal 26:2, pp. 243-249.