Use of genetic algorithms in numerical weather prediction

Liviu Oană, Adrian Spătaru
Department of Computer Science
West University of Timişoara
Timişoara, Romania
liviu.oana88@e-uvt.ro, adrian.spataru@e-uvt.ro

***Abstract*—This paper investigates the use of genetic algorithms in conjunction with the WRF - Weather Research and Forecast numerical weather prediction system in order to optimize the physical parametrization configuration and to improve the forecast of two important atmospheric parameters: 2 meter temperature and relative humidity. The research conducted reveals good results in improving the average prediction error in limited amount of iterations. This proves helpful in building GA optimized forecast ensembles, especially when focusing on specific atmospheric parameters. The optimization process performed well in finding optimal physical configurations for humidity prediction, but showed poor results for temperature forecast; more experiments need to be conducted in order to have a clear view over the utility of using GA techniques for physical parametrization optimization.**

***Keywords*-numerical weather prediction; genetic algorithms; WRF model; physical configuration;**

I. Introduction

Numerical weather prediction models have been used for the past decades to forecast and research atmospheric circulation, at first at large scales with many limitations. With the advance of technology and computer science, meteorology became a quantitative science and forecast models became more accessible to scientists, forecasters and interested persons. In recent years, many numerical weather prediction systems were developed, some open-source and some proprietary and, with increasing computing power, regional or limited area models emerged. Today, weather prediction is focused more on particular forecasts rather than general, playing an important role in society with many activities becoming more dependent on weather modeling development.

Numerical prediction models take into account a large number of variables in the form of micro-physics parameters, convective schemes, soil-surface interactions and radiation behaviour. It has been observed that a small change in these parameters can significantly change the outcome of the prediction model. A closer look at the parameter value selection problem led to us to consider a configuration of parameters as a chromosome with a one-to-one mapping between genes and parameter values. This allows us to *breed* well suited configurations with the incentive that it will lead to a better configuration.

The rest of the paper is structured as follows: Section II presents related efforts conducted in the direction of parameter optimization; Section IV discusses the parameters that influence numerical weather prediction and their relation with the outcome of the forecast; In Section IV we present a genetic algorithm approach for physical optimization for the given parameters and in Section V and compare its behaviour within 6 optimization procedures. Finally, in Section VI we present our conclusions and provide manners in which our approach can be improved.

II. Related Work

Many studies that research the physical parametrization sensitivity regarding errors evaluating different weather situations and conditions suggests that for each weather type or event, a different set of configurations could be more suitable. A study in particular suggests that some planetary boundary layer schemes did not performed very well under stable atmospheric conditions and over-predicted the minimum temperature inside a valley [1], and other studies point out the necessity of altering planetary boundary layer (PBL) schemes in order to produce best results [2], [3]. All of them point that surface weather parameters are very sensitive to PBL schemes. Other research aimed to compare the sensitivity of different weather parameters to parametrization schemes in WRF concluded that they are sensitive to many physical schemes, pointing out the sensitivity of the temperature to land-surface schemes. Under different atmospheric conditions the Mean Sea Level Pressure, which is not dependent on any parametrization schemes in cold season, shows some degree of dependency to long wave radiation during summer. Also this study pointed out that for different regions, different combinations of the four schemes evaluated perform better and it is important to make tough studies, prior to any WRF application [4]. All studies mentioned above express that the schemes available in the WRF Model are dependent to the atmospheric profile, and some physics configurations work better for a certain type of weather condition or region. The WRF Model is also dependent on the initial and lateral boundary information provided by the coarser model. The parametrization schemes lose sensitivity to weather parameters over time [5], because the errors induced by the main model will surpass

2470-881X/16 $31.00 © 2016 IEEE
DOI 10.1109/SYNASC.2016.69

the errors generated by the parametrization schemes. By choosing a module from each category, one may create a physical configuration, that may be suitable for different weather scenarios or specific weather parameters (i.e., 2 meter temperature, relative humidity, atmospheric pressure, or even heavy precipitation events).

Genetic Algorithms can tackle different problems inside the WRF Model. Modifying coefficients from the Kain-Frisch convective scheme in the model can improve the precipitation forecast in a tropical cyclone [6]. Wind speed forecasts can be improved using GP to perform a symbolic regression from a set of past forecasts obtained from the WRF-ARW grid [7]. Venkadesh et al. [8] use genetic algorithms to determine the optimal duration and resolution of prior data for weather variables that was considered a potential input for an ANN model. A study done Yan et al. uses GA techniques to optimize the automatic identification of weather systems [9]. Three other studies done by Ihshaish and his team [10], [11] and [12] developed a numerical weather prediction ensemble (G-Ensemble), based on genetic programming approach. They used the WRF model in order to initiate a calibration phase in which a short genetic algorithm was started, and the final ensemble members were chosen by the GA skill. The information encoded in the chromosomes was defined by closure parameters from land use and soil data integrated into the WRF model. Then with the best performing members selected, they started the actual ensemble forecast which was greatly improved comparing to ordinary forecasts.

Our approach in this study is to build a short iteration genetic algorithm suitable for optimizing the physical configuration for the WRF-ARW model and to see if this method can be used for generating optimized ensemble forecasts, similar to what Ihshaish and his team done, but focusing on physical parametrization options, rather than closure parameters from land-use data.

III. Weather modeling physical parametrization

For our experiments, we used the open-source WRF - ARW (Weather Research and Forecast - Advanced Research WRF) numerical weather prediction system, which is currently developed and maintained at NOAA, National Centers for Environmental Prediction department. Because it is a state-of-the-art open-source model, it relies on additional parametrization modules.

A numerical weather prediction model, in essence, takes as input atmospheric conditions from periodic weather observations and global models, and static geographic data like digital elevation model, slope, land use, land cover, soil type and many other parameters. It interpolates the data on a grid and integrates the model in time, running primitive equations in order to compute future states of the atmosphere at any given time step. This approach worked very well with old global circulation models in which mesoscale or local phenomena are insignificant. Today almost all models use additional physical parametrization modules in order to better approximate the atmosphere state. These additional parametrization sub-models include:

- Microphysical parametrization, which solves or estimates microphysical interactions in the atmosphere and their output influence, like latent heat and heat fluxes;
- Radiation parametrization that includes in the model the radiation effects from the sun, scattered radiation from clouds, or even absorbed and re-emitted radiation from the ground;
- Convective parametrization that can calculate convective processes such as cumulus or cumulonimbus clouds that otherwise could not be estimated using only the primitive equations.
- Soil and soil-surface processes which can estimate soil physics and what happens at the interaction between soil and surface layer.
- Planetary boundary layer physics (first layer of the troposphere) and calculates the physical processes that occur between soil, surface layer and upper troposphere.
- Urban physics that estimates the influence of urban areas on the atmosphere, which can include heat reflected by roofs or emitted by air conditioning systems.

Because the WRF model is an open-source software, multiple modules were introduced for each category and with newer versions being continuously developed, many more will be added in the future.

IV. Parameter optimization

Our approach inspects the use of genetic algorithms to explore the search space based on the assumption that some combinations of low-error configurations of the model will yield a configuration which is more accurate. We designed a genetic algorithm to optimize the physical configuration of the WRF model, namely *WRF-GA*. The process of fitness evaluation implies running the WRF model times the generation size each iteration, which can be time consuming depending on the grid-point resolution and the model domain extent used; for our experiments we used a 110x70 grid with a 10 kilometer resolution, centered on western Romania as seen in Figure 1. Each generation is composed of 14 chromosomes on which we encoded the following physical configuration options: microphysics, cumulus physics, longwave radiation physics, shortwave radiation physics, surface physics, PBL physics and soil physics. After a randomly generated initial population, an evaluation is made based on the absolute average error from 5 weather stations. The chromosomes are ascendingly sorted based on the forecast error. After this step, the first chromosome will become elitistic and it will not change with future iterations. For the random point crossover operation, the first 11 chromosomes are selected and they will reproduce two by two, each pair generating two new offspring, this way the generation size is maintained constant. The mutation operation is somewhat different from usual, with the last two chromosomes (with the largest errors) being completely replaced by randomly generated ones with a 10 percent probability in each iteration, making sure that new elements have high chances of being added in this short-lived genetic algorithm. After the mutation, the process is repeated until 30th generation. We considered that because each iteration is time consuming (for our experiments each iteration lasted on average 10 minutes, but can be longer depending on the grid size and resolution), to stop after 30 iterations.

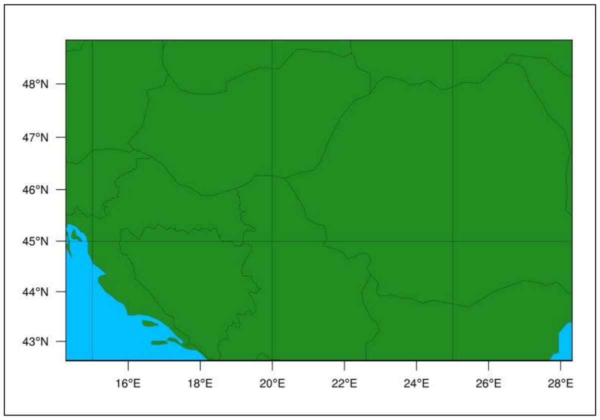

Figure 1. WRF model domain extent used in simulations

V. EXPERIMENTS

A. Data Set

For our experiment we chose to analyze a number of three cases, each one with different atmospheric conditions:

- Case 1: 18 june 2016, a day with variable weather, with clear sky in the first part, and convective thunderstorms in the second part due to a passing frontal system.
- Case 2: 28 june 2016, a rainy day with lower than average temperatures and heavy precipitations in western parts of Romania
- Case 3: 22 july 2016, a day with variable weather, in which scattered convective clouds were present most of the time in northwestern parts of Romania and with temperatures close to normal.

We evaluated the behaviour of WRF-GA, in optimizing the forecast of two important atmospheric parameters in each case: 2-meter temperature and relative humidity which are closely connected and can influence the development of severe convective weather phenomena. The WRF-ARW Model used in simulations was integrated for a period of 12 hours, except for the 2-meter temperature in Case 1, where we tested the optimization and capture of associated phenomena for a frontal system and was integrated for 18 hours.

Input data for our WRF model initiations was provided by the GFS (Global Forecast System) atmospheric model, developed and maintained at NOAA-NCEP. For prediction evaluation, we used the observed data from 5 Romanian WMO weather stations, namely Arad, Deva, Cluj-Napoca, Caransebes and Craiova.

We record the maximum, average and minimum error, and how they evolved in during each generation, and we count how many chromosomes had errors below a certain threshold (which varies from case to case) in each generation, in order to better represent the entire generation forecast skill.

B. Results

The experiment yielded positive results in all three case studies and showed that genetic algorithms can be used in conjunction with WRF model in order to optimize the physical configuration for specific parameters:

Figure 2. Evolution of maximum, average and minimum temperature error in relation with passed generations for the 18 june 2016 case.

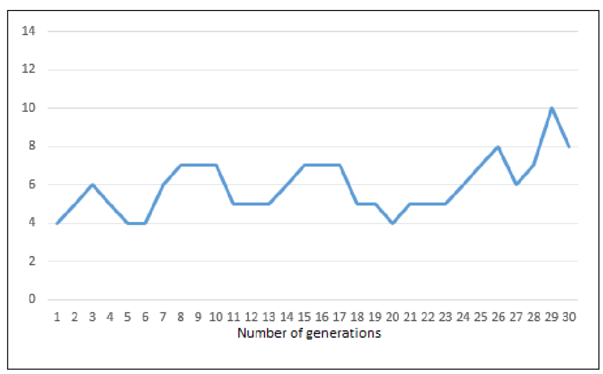

Figure 3. Evolution of the number of chromosomes in each generation with temperature errors smaller than 3.5 degrees celsius for the 18 june 2016 case.

In Case 1 we notice that the temperature forecast is not well optimized within the 30 generations, the average error remaining close to 4 degrees Celsius as seen in Figure 2. However, at the end of the GA run it can be observed that the

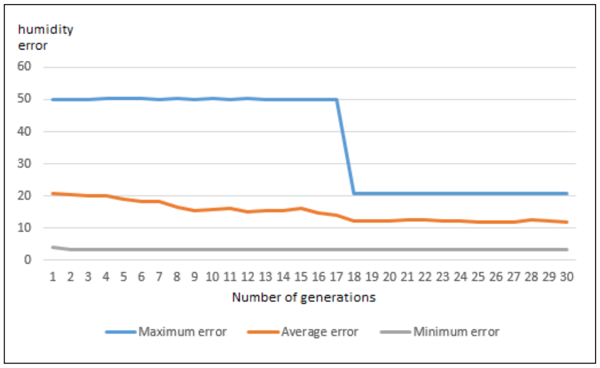

Figure 4. Evolution of maximum, average and minimum relative humidity error in relation with passed generations for the 18 june 2016 case.

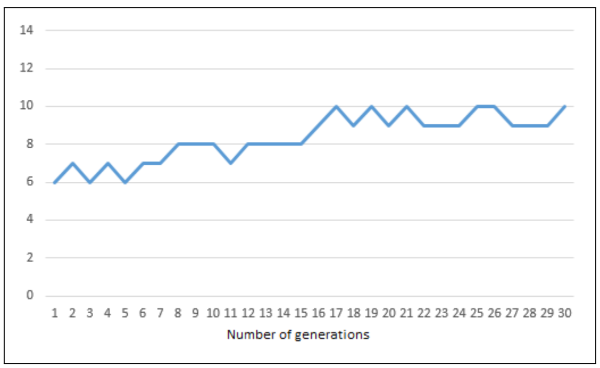

Figure 5. Evolution of the number of chromosomes in each generation with relative humidity errors smaller than 15 percent celsius for the 18 june 2016 case.

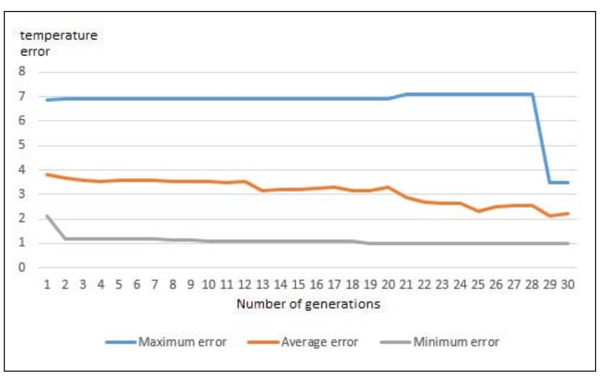

Figure 6. Evolution of maximum, average and minimum temperature error in relation with passed generations for the 28 june 2016 case.

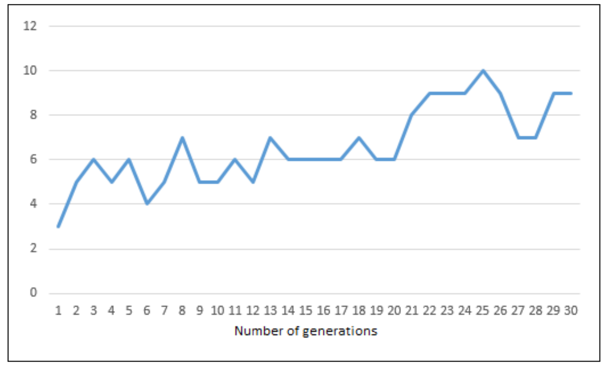

Figure 7. Evolution of the number of chromosomes in each generation with temperature errors smaller than 2.5 degrees celsius for the 28 june 2016 case.

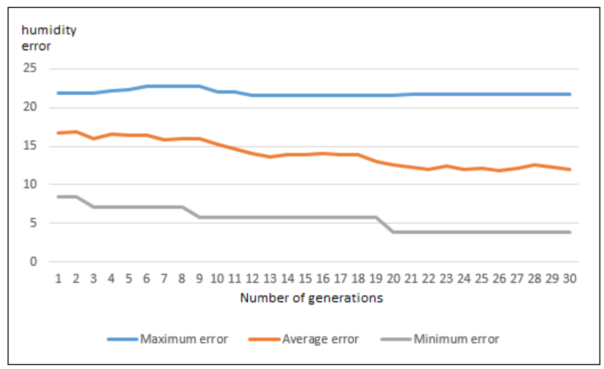

Figure 8. Evolution of maximum, average and minimum relative humidity error in relation with passed generations for the 28 june 2016 case.

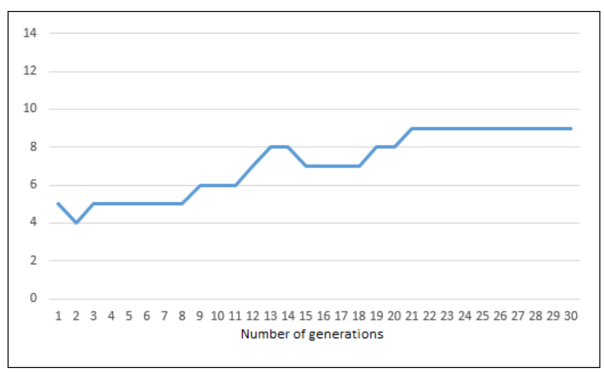

Figure 9. Evolution of the number of chromosomes in each generation with relative humidity errors smaller than 15 percent celsius for the 28 june 2016 case.

Figure 10. Evolution of maximum, average and minimum temperature error in relation with passed generations for the 22 july 2016 case.

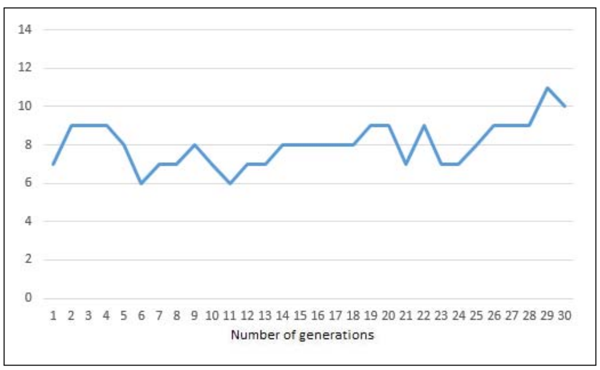

Figure 11. Evolution of the number of chromosomes in each generation with temperature errors smaller than 1.5 degrees celsius for the 22 july 2016 case.

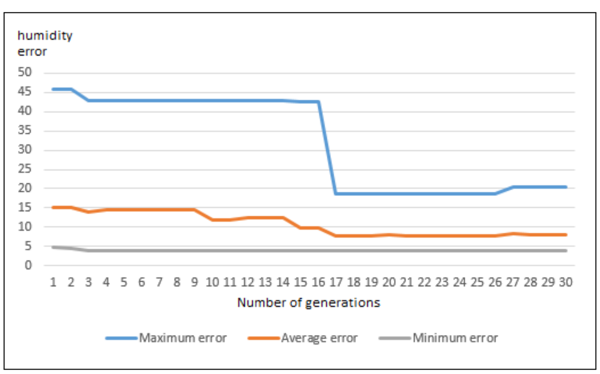

Figure 12. Evolution of maximum, average and minimum relative humidity error in relation with passed generations for the 22 july 2016 case.

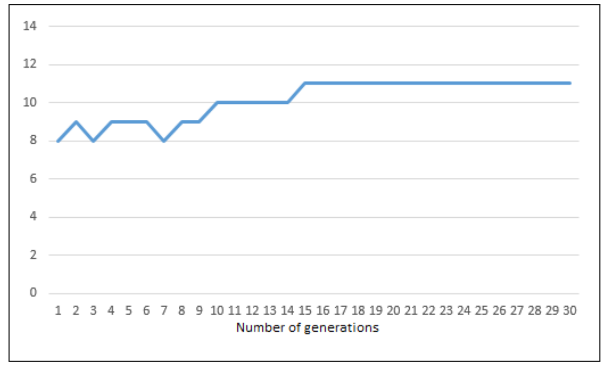

Figure 13. Evolution of the number of chromosomes in each generation with relative humidity errors smaller than 10 percent celsius for the 22 july 2016 case.

error decreases and the number of chromosomes with errors smaller than 3.5 degrees increases in number, according to Figure 3. For the relative humidity we observe a decline in average error (from 20% to 12%) in the first 18 generations, followed by steady evolution seen in Figure 4. Figure 5 shows that the number of chromosomes in each generation with errors smaller than 15% increases in the first part of the GA run (from 6-7 to 9-10), indicating that the algorithm was successful in finding good performing configurations within 30 iterations.

In the second case examined, the average temperature error decreases, more prominently in the second part of the GA run from almost 4 degrees, up to a little over 2 degrees and also we noticed a somewhat steady increase for the number of chromosomes with errors smaller than 2.5 degrees as seen in Figures 6 and 7. For the relative humidity also we noticed a decrease in error from 17% to 12% according to figure 8 and the entire generation forecast skill improved, as can be seen in the evolution of the number of chromosomes with errors smaller than 15% (Figure 9).

The last case examined showed no significant improvement for temperature prediction as presented in Figures 10 and 11, but we noticed an improvement in relative humidity error from 15% to a little under 10% in the first half of the GA run, that was also noticeable in the evolution of the number of chromosomes with errors smaller than 10% as seen in Figures 12 and 13.

VI. Conclusion and Future Work

This paper has investigated an automatic approach using genetic algorithms for physical parametrization configuration in WRF model. We have shown that the average error for a generation can be improved by initiating a short-iteration genetic algorithm (30 generations) and finding suitable configurations for specific atmospheric parameter forecast. Our results concluded that algorithms initiated to optimize the relative humidity forecast performed good in all three cases analyzed, while the ones initiated for the temperature forecast optimization showed good, but not significant results.

More experiments need to be conducted in order to better evaluate the utility of using genetic algorithms in conjunction with the WRF model, and as for the physical configuration, we need to find relations between different parametrization modules in order to automatically exclude conflicting configurations and to improve the number of iterations needed to find optimal solutions, as well to further improve the algorithm itself. Another experiment will inspect the cross-over of the most suitable configurations for specific atmospheric parameters and if this will decrease the overall error of the system.

VII. Acknowledgement

The authors acknowledge the West University of Timioara for providing the computing resources on the InfraGRID cluster and BlueGene supercomputer, without which the conducted research would not have been possible.

References

[1] R. Dimitrova, Z. Silver, H. Fernando, L. Leo, S. Di Sabatino, C. Hocut, and T. Zsedrovits, "16.3 intercomparison between different pbl options in wrf model: Modification of two pbl schemes for stable conditions," 2014.

[2] X.-M. Hu, J. W. Nielsen-Gammon, and F. Zhang, "Evaluation of three planetary boundary layer schemes in the wrf model," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 9, pp. 1831–1844, 2010.

[3] J. Jin, N. L. Miller, and N. Schlegel, "Sensitivity study of four land surface schemes in the wrf model," *Advances in Meteorology*, vol. 2010, 2010.

[4] P. Mooney, F. Mulligan, and R. Fealy, "Evaluation of the sensitivity of the weather research and forecasting model to parameterization schemes for regional climates of europe over the period 1990–95," *Journal of Climate*, vol. 26, no. 3, pp. 1002–1017, 2013.

[5] H. Klug and L. Oana, "A multi-purpose weather forecast model for the mondsee catchment," *GI_Forum*, vol. 2015, pp. 600–609, 2015.

[6] X. Yu, S. K. Park, Y. H. Lee, and Y. S. Choi, "Quantitative precipitation forecast of a tropical cyclone through optimal parameter estimation in a convective parameterization," *SOLA*, vol. 9, no. 0, pp. 36–39, 2013.

[7] G. Martınez-Arellano and L. Nolle, "Short-term wind power forecasting with wrf-arw model and genetic programming," in *Proceedings of the 19th International Conference on Soft Computing MENDEL*, 2013.

[8] S. Venkadesh, G. Hoogenboom, W. Potter, and R. McClendon, "A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks," *Applied Soft Computing*, vol. 13, no. 5, pp. 2253–2260, 2013.

[9] K. Y. Wong, C. L. Yip, and P. W. Li, "Automatic identification of weather systems from numerical weather prediction data using genetic algorithm," *Expert systems with Applications*, vol. 35, no. 1, pp. 542–555, 2008.

[10] H. Ihshaish, A. Cortés, and M. A. Senar, "Towards improving numerical weather predictions by evolutionary computing techniques," *Procedia Computer Science*, vol. 9, pp. 1056–1063, 2012.

[11] ——, "Parallel multi-level genetic ensemble for numerical weather prediction enhancement," *Procedia Computer Science*, vol. 9, pp. 276–285, 2012.

[12] H. Ihshaish, A. Cortés, and M. Senar, "Genetic ensemble (g-ensemble) for meteorological prediction enhancement," 2011.