Computational Acquisition of Meteorological Data for Applications in Electric Power Systems

Nigel Woodhouse*
Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
nwoodhou@ualberta.ca

Petr Musilek
Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
pmusilek@ualberta.ca

***Abstract*—Humans heavily rely on mother nature for hospitable living conditions, plentiful harvests, and energy generation, yet have no control. The best humans can do is plan and predict. Climatological statistics and forecasts provided by public weather services serve as traditional methods for obtaining meteorological information. However, through Numerical Weather Prediction models, one can simulate climate fluctuations with high spatial resolution over long periods. Uses for Numerical Weather Prediction models include analyzing the energy flux of smart homes, smart grid technology, impact on power transmission infrastructure, and energy production through wind and photovoltaic farms. The efficiency of these technologies is dependant on the surrounding weather phenomena. The optimization of these systems to the environment upon which they exist can both reduce wasted resources and the economic impact on consumers and organizations. This paper outlines the methods used for the acquisition of weather data through computer simulations at a spatial resolution of 1.2 km in 15-minute intervals with an accuracy of 2%.**

***Index Terms*—Numerical Weather Prediction, Weather Research and Forecasting, Smart Home Technology**

I. INTRODUCTION

Historically, collecting meteorological information for research purposes relied heavily on forecasts from public weather services and climatological statistics. This method remains viable and widely used but can lack the spatial and time resolutions required for detailed inquiry. Numerical Weather Prediction (NWP) models allow an individual or organization to simulate, analyze, revise, and anticipate various weather phenomena in great detail. NWP models reproduce the governing equations that dictate the atmosphere's behaviour and is an invaluable tool for detailed weather forecasting and research purposes.

Potential areas of inquiry that can benefit from the use of NWP's include the weather's impact on power transmission infrastructure [1], dynamic thermal rating calculations [2], power management [3], environmental monitoring [4], energy storage systems [5], wind power generation [6], and photovoltaic power generation [7].

Work described in this article is supported by Natural Sciences and Engineering Research Council (NSERC) of Canada grant CRDPJ 500547-16, and by Alberta Electric System Operator (AESO), Altalink LP and ATCO Electric.

The study described in this article relies on *eGauge*, a service that records the energy flux of smart homes and provides real-time updates on current energy generation and consumption [8]. A primary energy consumption activity is the regulation of the homes' internal temperature, done through heating in the winter and air conditioning in the summer. Meanwhile, solar photovoltaic energy generation is highly dependant on the time of day and season. One can, therefore, deduce that the energy flux experienced by the home (the net of energy generating and consuming activities) is a function of external environmental factors. Optimizing the energy flux of the home reduces both the cost to the homeowner and resource consumption. This insight provides motivation for the ensuing investigation.

The following sections summarize the methodologies to simulate and collect detailed weather data to further understand the interplay between the energy flux of smart homes and meteorological trends. Data used in the study has been acquired to study the Seattle, Washington and Edmonton, Alberta regions. The NWP simulates six years of weather data.

The remainder of the paper is organized as follows: Section II outlines methods for data acquisition, software setup and data extraction. Section III organizes results. Finally, Section IV discusses the research outcomes.

II. METHODS

A. Weather Data Acquisition

Many resources offer free archives of historical weather data. Noteworthy sources include the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Atmospheric Research (NCAR) [9] [10]. Both NOAA and NCAR offer the North American Regional Reanalysis (NARR) dataset produced by the National Centers for Environmental Prediction (NCEP) [11]. In the dataset, variables include air temperature, air pressure, wind speed, solar radiation flux, precipitation, and humidity. NARR logs data at a geographical resolution of 32 km at three-hour increments, with records stemming back to January 1st, 1979. Due to its extensive historical archive, the expanse of its geographic range, and the magnitude of accessible variables, NARR continues to be a widely used dataset for detailed climate modelling today.

NCAR Research Data Archive serves as the resource for downloading the NARR dataset [12]. The data of interest for this study spans from 2014 to 2020.

B. Software Overview

Weather Research and Forecasting (WRF) is a Linux based software that encapsulates the governing physics equations that regulate the atmosphere's behaviour. WRF is an industry-standard NWP, used by environmental agencies and universities globally [13]. The software allows for the input of real-world meteorological observations and interpolates spatial and time domains to export precise climatological results.

The WRF Pre-Processing System (WPS) acts as the initial point for the construction of spatial and time aspects of the domain's under analysis, while also interpreting the real-world weather input. WPS acts to "define the model horizontal domain … horizontally interpolate static data to the model domain," decode data from the desired weather archive, extract variables, and "ingest static data and raw meteorological fields [14]." All information "output conforms to the WRF I/O API [14]." Section II-C elaborates upon the setup and details.

WRF then uses the WPS output files, where the software integrates the weather data across the specified spatial and time domains. WRF supports the functionality to edit the underlying physics parameters to engage in more niche areas of study, such as cloud formation, fire modelling, and oceanic behaviour [15]. As no exploration of unique meteorological phenomena occurs in this line of inquiry, the use of default physics parameters is sufficient.

Simulations described in this article use WPS version 3.9.1 and WRF version 3.9.1.1.

C. Domain Construction Methods

The WRF software package allows for the constructions of multiple nested domains; that being, one large, main (or parent) domain with smaller domains scatter within its boundaries. The advantage of this attribute allows for both large scale meteorological trends and small-scale fluctuations to be studied simultaneously by both the parent and nested domains. Nested domains are imperative to the high-resolution analysis of the region under investigation. Merely running a single, small, high spatial resolution domain may invoke errors along its boundaries due to the limited information and the software's interpolation abilities.

WPS sets up necessary parameters for domain creation, including the number of domains, the time range of analysis, and the position and size of the domain area. Performance is best when the center of the first domain aligns with the center of the area under investigation or the center of mass of the data points. The domain also requires the definition of bounding latitude lines to curve its north and south edges as it conforms to the Lambert conformal conic projection map.

When setting up nested domains, the user should be cognisant of the desired level of spatial resolution required to investigate their question. Each nested domain's resolution is a ratio of its parents' resolution. Additionally, the position of a nested domain inside the parent is specified by two Cartesian coordinates in intervals of the parent's spatial resolution.

WRF extends the domain construction to satisfy the time-step requirements. The stepping method in the software is Runge-Kutta third order, and the time-step is equal to six times that of the first domain's spacing (in kilometres). Nested domains follow a ratio of the parent domain regarding both time and space parameters. It is easy to identify the trade-off between spatial resolution and length of the computation. That being said, if one wishes to improve the spatial resolution by a factor of two, the time-step decreases six-fold, thus increasing computation time significantly.

The efficiency of the simulations and the accuracy of the results depend upon the quality of the domains. The following methods for good domain construction derive from work of Pytlak and Musilek [1]. When constructing nested domains of increasing spatial resolution, the user should implement a systematic refinement strategy.

Ideally, the parent domain should encompass an area significantly larger than the domain under analysis. Using a scaling factor of four between domains in the following analysis allows for the formation and observation of macro climatological trends that influence small areas.

Along similar lines of thinking, the nested domains should be larger than the areas of interest. This method allows for a smooth transition between domain resolutions by making use of the software's nested domain feedback functionality to reduce errors that incur through boundary conditions, ultimately improving the accuracy of the results.

Instead of using one high-resolution domain, one could set up multiple, small, high-resolution domains. The purpose is that the multiple small domains can hone in upon important points and exclude areas of little interest under the study. By simulating less area (or less area under high spatial resolution parameters), it can reduce the computation time and resources required by the simulation. One can begin investigating the construction of multiple small domains through clustering algorithms (K-means, for example).

D. Edmonton and Seattle Domain Parameters

The application of Section II-C's methods can construct nested domains over the Edmonton, Alberta and Seattle, Washington areas. These areas are of interest as they contain the locations of smart homes using the *eGauge* system [8]. There exist 15 and 52 smart homes in the Edmonton and Seattle regions, respectively, serving as the main points of interest.

Three nested domains with increasing spatial resolution analyze the climate patterns over the two regions independently. Each nested domain has a spatial grid resolution one-third of its parents' domain while encapsulating one-fourth of the area. The grid resolution values for the domains derive from the spatial resolution of the NARR weather dataset.

All Edmonton domains center on 53.518°N, -113.488°W, while Seattle domains center on 47.514°N, -122.267°W. The

TABLE I
EDMONTON NESTED DOMAIN PARAMETERS

Domain	D01	D02	D03
Grid resolution [km]	10.8	3.6	1.2
Grid dimensions	22 x 22	34 x 34	49 x 49
Side length [km]	237.6	122.4	58.8
Area [km^2]	56453.8	14981.8	3457.4
Area ratio	-	3.77	4.33

TABLE II
SEATTLE NESTED DOMAIN PARAMETERS

Domain	D01	D02	D03
Grid resolution [km]	10.8	3.6	1.2
Grid dimensions	39 x 39	58 x 58	85 x 85
Side length [km]	421.2	208.8	102
Area [km^2]	177409.4	43597.4	10404
Area ratio	-	4.07	4.19

centers of these domains coordinate with the midrange of the smart homes' geographical locations.

Edmonton domains simulate weather data between January 1st, 2015, to June 30th, 2020, whereas Seattle domains begin one year earlier. These date ranges are chosen based upon the earliest *eGauge* data and the most recent available NARR data at the time of the study.

Table I and Table II summarize the parameters for the Edmonton and Seattle WRF domains. The area ratio compares the area of a nested domain to its parent.

E. Experimental Setup and Data Extraction

All simulations run on Compute Canada supercomputers: Beluga, Cedar, and Graham [16]. Sixteen instances of the WRF software suite have been installed on each cluster to run several weather simulations independently. Queueing of simulations on the Compute Canada clusters is organized by the Slurm system [17].

Edmonton weather data is split into monthly intervals and simulated, whereas Seattle data is split into weekly intervals and simulated, as to manage simulation duration. Simulations run with a 64.8-second time-step. As the Seattle domains are approximately three times larger than their Edmonton counterpart, this leads to a proportional increase in computation time for the same time window, hence the splitting methodology. While feasible to run Seattle weather data in monthly portions, the increase in required resources keeps simulations in the Slurm queue system for extended periods and is undesirable.

Simulations run using 2 CPUs, running 32 threads each, with a default of 250 GB of allocated memory. Under this setup, it requires 12-18 hours to compute a single weather simulation, regardless of region. Overall, an estimated 234 days of computational effort is required to simulate all meteorological data.

WRF outputs data in the form of netCDF files. Custom Python 3.7 scripts written by the research team borrow from the netCDF4 Python library to extract variables of interest. These variables include surface temperature, surface pressure, wind speed, solar radiation flux, precipitation, and humidity, correlated to their time and geographical location.

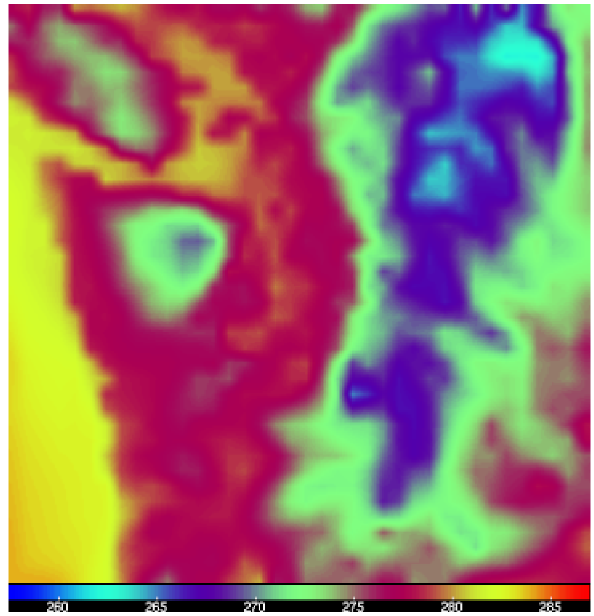

Fig. 1. Seattle D01 domain showing two-meter elevation temperature [K] on January 4, 2020, 12:15. The Pacific Ocean [left], Olympic National Park [center-left], Puget Sound [center], and Rocky Mountains [right] are easily identifiable.

Other team-written Python scripts borrow from the NumPy and Pandas Python libraries to organize and save the parameters from D03 domains to a usable file format, after which filtering for data points with the closest geographical locations to the smart homes occurs. The filtering mechanism borrows from the GeoPy and SciPy Python libraries.

Ncview [18] (a Linux based software) can produce visualizations of the simulated variables across a domain, as shown in Fig. 1.

F. Accuracy Verification

Simulated temperature, wind speed, and surface pressure are compared to weather archives to verify the accuracy of the results. The comparison involves 2550 points of weather data archived by the Edmonton Blatchford weather station and simulated data from the nearest data point [19]. Table III in Section III summarizes the results of the accuracy analysis.

III. RESULTS

The use of the Weather Research and Forecasting (WRF) software and the North American Regional Reanalysis (NARR) weather archive produces the following results. Using the WRF parameters and methods described, Edmonton's D03 domain outputs 2304 data points per 15-minute period, of which 15 are of interest due to their proximity to the smart homes under study. Similarly, Seattle's D03 domain generates 3136 data points per 15-minute period, of which 52 are of interest. For Edmonton homes, the average geographical distance variation between the home and the closest data point

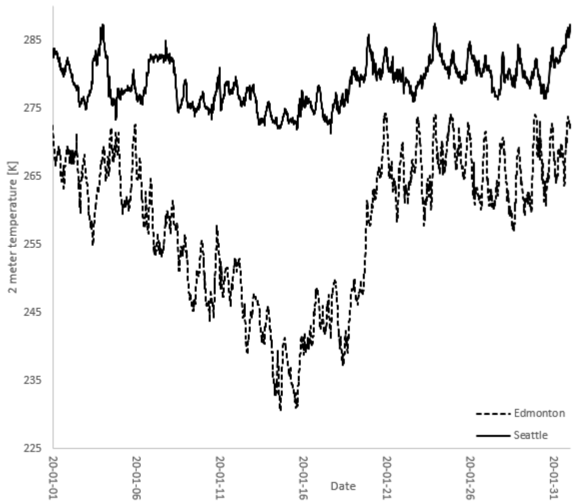

Fig. 2. Temperature fluctuations at two-meter elevation experienced by a smart home in Seattle and Edmonton between January 1st, 2020, and February 1st, 2020, as displayed through 15-minute increments.

is 509.6 meters (SD = 170.3). Similarly, for Seattle homes, the average geographical variation is 421.2 meters (SD = 162.4).

The parameters of interest extracted from the WRF output netCDF file are the temperature at two-meter elevation [K], surface pressure [kPa], wind speed at 10-meter elevation divided into its x and y components [m/s], shortwave radiation flux up and down [W/m2], longwave radiation flux up and down [W/m2], precipitation [mm], and humidity [kg/kg], due to their relevance to the study. The research team expects that these parameters will show a significant correlation to the energy flux of the smart homes under analysis in future research.

The simulation of Edmonton weather data from January 1st, 2015 to June 31st, 2020, at 15-minute increments produce 1.92×10^5 data points per home or 2.89×10^6 data points across all Edmonton locations. Similarly, the Seattle weather data simulations from January 1st, 2014 to June 31st, 2020, at 15-minute increments produce 2.27×10^5 data points per home or 1.18×10^7 data points across all Seattle locations. In total, 1.47×10^7 data points are available for further study.

Fig. 2 plots one month of two-meter elevation temperature for a single home in Edmonton and Seattle.

Comparison of 2550 temperature, wind speed, and surface pressure measurements from the Edmonton Blatchford weather station to a nearby data point tests the accuracy of the WRF simulation over the Edmonton area [19]. The time frame for the verification spans from March 16th, 2020, to June 30th, 2020. The data point selected from the D03 domain is 346 meters away from the Blatchford weather station. This point serves as the best representation of the site. Table III summarizes the results.

TABLE III
WRF EDMONTON ACCURACY

	Mean Absolute Error	Standard Deviation
Temperature [K]	5.20	3.74
Wind Speed [m/s]	1.94	1.61
Surface Pressure [kPa]	0.22	0.17

IV. DISCUSSION

NARR is a robust dataset due to the number of weather phenomena it can support, its extensive historical archive, vast geographical expanse, and reasonable time and spatial resolutions. WRF simulates NARR data between 2014 and 2020 on a 64.8-second time-step, as limited by the D01 domain spatial resolution. The small time-step, in conjunction with the refined domains, allows for rapid data production, hence the generation of 2304 Edmonton data points and 3136 Seattle data points per 15-minute period.

The methodology undertaken to construct the WRF domains follows from Pytlak and Musilek [1]. Applying a one-third rule to nested domains regarding their resolution allows for a smooth transition in data, reducing errors along the boundaries while taking advantage of WRF's domain feedback feature. Creating concentric square domains with a consistent area ratio aids in the WRF software's ability to capture both large and small-scale meteorological phenomena over the desired area. All domains center amongst the midrange of the smart homes' geographical locations. The smallest domain is large enough to encompass all the smart home locations.

Note that the methods require large domains when only a small fraction of data points are necessary for this form of research; this is one of the software's pitfalls—retaining 1.66 percent of Seattle's D03 domain and 0.65 percent of Edmonton's D03 domain means that many resources have gone to waste in its production. Pytlak's methods are more beneficial to his application as his interests explored the entire domain simulated, in contrast to individual points [1].

The creation of both the Edmonton and Seattle domains does not use a clustering algorithm in their construction as the points of interest form a loose cluster; its implementation would interfere with other domain design considerations, and possible improvements that could derive from a clustering approach are minimal.

The advent of increased computational prowess allows the production of large datasets for various areas of research. The access to three supercomputer clusters across Canada through Compute Canada vastly speeds up the analysis process; what could take nearly two-thirds of a year to compute can be done in a matter of weeks. The parameters selected to run the WRF jobs attempt to balance both queue and run times. An increase in requested resources can dramatically increase the simulation's queue time.

Custom Python 3.7 scripts written by the research team find the nearest data points between the smart homes and simulated domains. The average distance between data points is approximately one third that of the resolution of the D03 domain. While increased resolution would be desirable, it may be unnecessary as the variations to other nearby data points are minimal. The possible advantage is servery outweighed by the expected computational resources.

Table III displays the accuracy of simulated Edmonton data based upon 2550 data points. Seattle is expected to have comparable accuracy due to the number of data points available.

The mean absolute error in temperature and wind speed is, is 2.78 and 1.10 times greater, respectively, when compared to the results compiled by Pytlak [1]. The wind speed is error is close; however, the discrepancy in temperature is sizeable between the two findings, even though methods are similar. The variation in temperature results could be due to two factors.

The first is the scope of the data; Pytlak analyzed a general area, whereas we used point samples [1]. Another potential source of variation is from the weather archive chosen, as there is limited information regarding the parameter details [19]. For example, the archive does not specify the altitude of temperature measurements, or if averaging of measurements occur. Without this information, it is challenging to make one-to-one comparisons. Regardless, temperature differences between Blatchford measurements and simulated data are only about 2%.

Further inquiry into understanding the variation in results is to be conducted by the research team in the near future using intelligent systems.

V. Conclusion

The use of Numerical Weather Prediction (NWP) models provides a method of data acquisition for many areas of meteorological research applications. Weather Research and Forecasting (WRF) software extracts 1.47×10^7 points of weather data across 67 locations in the Edmonton, Alberta and Seattle, Washington areas, spanning from 2014 to 2020. Three nested domains with spatial resolutions of 10.8 km, 3.6 km, and 1.2 km, with a 15 minute output time resolution, generate the dataset. The extracted meteorological parameters include temperature, pressure, wind speed, precipitation, humidity, and solar radiation flux. WRF can produce data with temperatures within average errors of 5.20 K, 1.94 m/s wind speed, and 0.22 kPa air pressure, compared to archived meteorological observations, keeping in line with previous inquiry in this topic.

The presented methods can investigate multiple processes influenced by atmospheric variation. The data collected will explore the correlation between how meteorological fluctuations influence the energy generation and consumption trends of smart homes. This investigation's motivation is to reduce consumed resources and the financial impact on homeowners. The obtained data will investigate the NWP's accuracy through the use of intelligent methods.

References

[1] P. Pytlak and P. Musilek, "An intelligent weather-based system to support optimal routing of power transmission lines," *2010 IEEE Electrical Power & Energy Conference*, Halifax, NS, 2010, pp. 1-6, doi: 10.1109/EPEC.2010.5697231. doi:10.1109/epec.2010.5697231, pp.43, 71-72, 2010.

[2] P. Pytlak, P. Musilek and J. Doucet, "Using Dynamic Thermal Rating systems to reduce power generation emissions," *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 2011, pp. 1-7, doi: 10.1109/PES.2011.6039401.

[3] D. Pimentel and P. Musilek, "Power management with energy harvesting devices," *2010 Canadian Conference on Electrical and Computer Engineering CCECE2010*, Calgary, AB, 2010, pp. 1-4, doi: 10.1109/CCECE.2010.5575118.

[4] P. Musilek, M. Prauzek, P. Krömer, J. Rodway, and T. Bartoň, "Intelligent Energy Management for Environmental Monitoring Systems," *Smart Sensors Networks*, pp. 67–94, 2017.

[5] R. Martins, H. Hesse, J. Jungbauer, T. Vorbuchner, and P. Musilek, "Optimal Component Sizing for Peak Shaving in Battery Energy Storage System for Industrial Applications," *Energies*, vol. 11, no. 8, p. 2048, 2018.

[6] E. Pelikán et al., "Wind power forecasting by an empirical model using NWP outputs," *2010 9th International Conference on Environment and Electrical Engineering*, Prague, 2010, pp. 45-48, doi: 10.1109/EEEIC.2010.5490019.

[7] M. Brabec, E. Pelikán, P. Krč, K. Eben and P. Musilek, "Statistical modeling of energy production by photovoltaic farms," *2010 IEEE Electrical Power & Energy Conference*, Halifax, NS, 2010, pp. 1-6, doi: 10.1109/EPEC.2010.5697249.

[8] "Monitor Electricity on Every Circuit with Precision and Accuracy," eGauge. [Online]. Available: https://www.egauge.net/. [Accessed: 10-Aug-2020]

[9] National Oceanic and Atmospheric Administration. [Online]. Available: https://www.noaa.gov/. [Accessed: 11-May-2020].

[10] "NCAR," National Center for Atmospheric Research. [Online]. Available: https://ncar.ucar.edu/. [Accessed: 11-May-2020].

[11] "NCEP NARR," NCEP NARR — NCAR - Climate Data Guide. [Online]. Available: https://climatedataguide.ucar.edu/climate-data/ncep-narr. [Accessed: 11-May-2020].

[12] National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, "NCEP North American Regional Reanalysis (NARR)," 30-Aug-2005. [Online]. Available: https://rda.ucar.edu/datasets/ds608.0/. [Accessed: 27-Jun-2020]

[13] "Models," MMM. [Online]. Available: https://www.mmm.ucar.edu/weather-research-and-forecasting-model. [Accessed: 11-May-2020].

[14] Wrf-Model, "wrf-model/WPS," GitHub. [Online]. Available: https://github.com/wrf-model/WPS. [Accessed: 20-Aug-2020]

[15] M. M. M. Webmaster, WRF Model Physics References. [Online]. Available: https://www2.mmm.ucar.edu/wrf/users/phys_references.html. [Accessed: 30-Aug-2020].

[16] "Home," Compute Canada. [Online]. Available: https://www.computecanada.ca/home/. [Accessed: 03-May-2020]

[17] "Slurm Workload Manager," Slurm Workload Manager - Documentation. [Online]. Available: https://slurm.schedmd.com/documentation.html. [Accessed: 01-Jul-2020].

[18] D. W. Pierce, "Ncview: a netCDF visual browser," Ncview. [Online]. Available: http://meteora.ucsd.edu/ pierce/ncview_home_page.html. [Accessed: 11-Jun-2020]

[19] "Data Download for Edmonton," Amateur Weather Statistics for Edmonton, Alberta. [Online]. Available: https://edmonton.weatherstats.ca/download.html. [Accessed: 20-Jul-2020].