Updating Australia's Reference Meteorological Years with the Addition of Hourly Precipitation Data

Chithral Kodagoda, Masoume Mahmoodi[1], Nihal Abdul-Hameed, Trevor Lee

Exemplary Energy, 32 Fihelly Street, Fadden, Canberra, Australia

[1] *College of Engineering and Computer Science, Australian National University, Canberra, Australia*

Abstract

Since July 2019, updating of the weather and climate files published for practitioner use has been prevented by the suspension of processing of satellite-derived solar radiation data by the Bureau of Meteorology (BoM) despite widespread concern based on a changing climate needing to be accurately represented by the data sets used in simulation activities. The foreshadowed release of subsequent data (Jain, et al., 2021) allows the updating of those Reference Meteorological Years (RMYs) files for over 230 Australian locations by the end of this year. It is proposed, however, to not just update the files from 1990 to the end of 2020 but to enhance them by the addition of hourly precipitation data at the same time.

A major difficulty in updating and enhancing weather and climate files is that hourly precipitation data is often only available for recent years. This paper reports on our methodology and synthesis of hourly precipitation data derived from the earlier years of daily precipitation data from 1990.

The availability of higher temporal resolution precipitation (mostly rainfall) data is important for a wide range of engineering and modelling applications such as urban hydrological applications, design of flash flood control structures, etc. (Poduje & Haberlandt, 2018). While daily precipitation data are usually available, unfortunately, long series of recorded rainfall with hourly temporal resolution does not exist for all locations. In this paper, we develop a probabilistic precipitation model to generate long synthetic series with hourly resolution from early daily resolution data to concatenate with recent measured hourly data for the same location to produce three decades of hourly data.

An important application of the resultant hourly rainfall data is their usage as input for an anticipated version of NatHERS software used for modelling the energy performance of a dwelling[1]. This work will allow accurate prediction of condensation issues for enhanced building healthiness and durability. (Tanaka & Zirkelbach, 2016) note that *"concerning the hygrothermal performance evaluation of building components, the local climate influence can be crucial"*. Accordingly, updated energy modelling input files incorporating precipitation data can also demonstrate full compliance with the National Construction Code (NCC, 2019) condensation requirements. Australian Climate Data Bank[2] (ACDB) weather and climate files are prepared for NatHERS software applications as part of the process of generating new Reference Meteorological Year (RMY) files for each of ~250 locations. For maximum utility and compliance demonstration purposes, ACDB files might soon need to include data on hourly precipitation.

The algorithm that we develop will also be used in the ClimateCypher[3] software to disaggregate available daily precipitation data into an hourly sequence. This precipitation addition to

---

[1] The Nationwide House Energy Rating Scheme (NatHERS) is administered by the Commonwealth on behalf of all Australian governments.

[2] Files in the same file format as the Australian Climate Data Bank (ACDB), jointly developed by BoM with CSIRO.

[3] ClimateCypher is a proprietary software of Exemplary Energy Partners that is used to synthesise raw weather data obtained from the Bureau of Meteorology into ACDB and TMY weather formats.

ClimateCypher will allow the generation of more detailed weather files in the TMY3 (Wilcox & Marion, 2008) and enhanced ACDB formats.

In continuation with the work carried out by (Anderson, et al., 2020), we perform correlation analysis to determine the correlation coefficients between hourly precipitation and different weather parameters for Canberra between the years 2010-2019 without differentiation by the time of year. Figure 1 shows the hourly precipitation data for these years. The correlation coefficients between precipitation and nine different weather elements and the precipitation value immediately preceding it in the time-series sequence are represented in Figure 2.

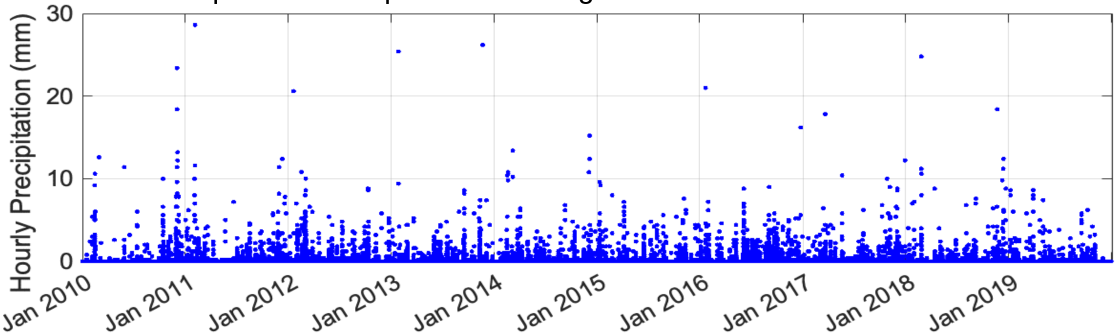

Figure 1: Time-series plot of hourly precipitation values for Canberra

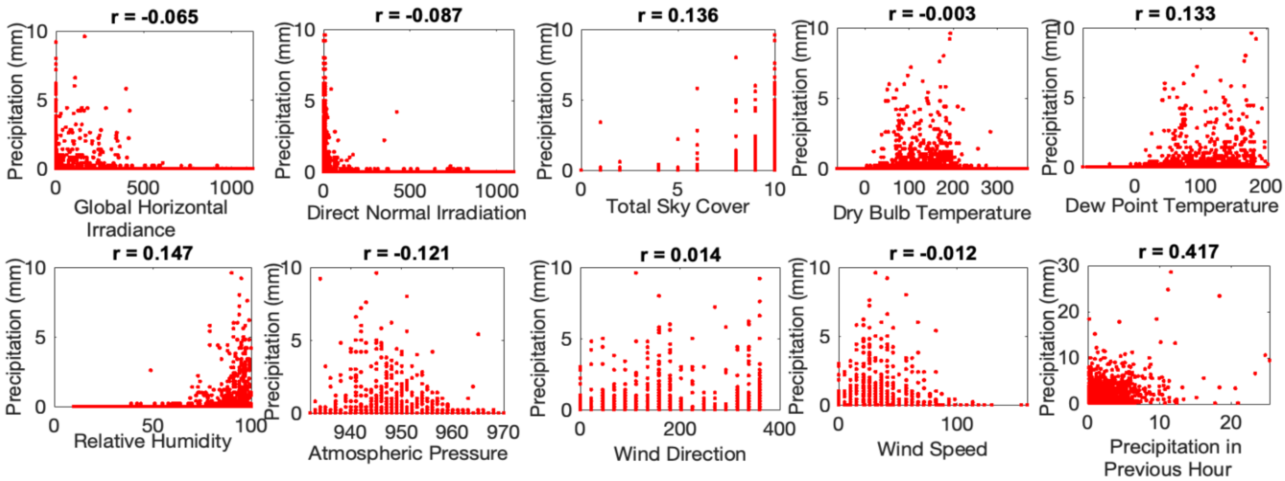

Figure 2: Correlation Coefficients between hourly precipitation and selected weather elements for 2010-2019

As can be seen from Figure 2, there is a relatively high correlation (r = 0.417) between a given precipitation value and the precipitation value immediately preceding it in the time-series sequence.

One approach to deal with these kinds of time-series models, where each state of the system is correlated with the previous states, is to use the Markov-based time series modelling framework (Thyer & Kuczera, 1999). A Markov chain includes different condition states of a system moving from one state to another over time. The order of Markov-based models depends on the number of time steps in the past influencing the probability distribution of the present state. (Thyer & Kuczera, 1999) used a first-order Markov chain model to generate annual rainfall data. In addition, (Thyer & Kuczera, 1999) and (Shamshad, et al., 2005) developed first and second-order Markov chain models to produce hourly wind speed data for a north-western region of Turkey and two stations in Malasia, respectively.

Markov chain behaviour can be described using a Markov transition matrix (MTM). The MTM gives the probabilities of transitioning from one state to another at each time. To construct an MTM, we first determine the total number of states which shows the dimension of the matrix. To determine the total number of states, we divide the range of available hourly precipitation values into **n** equal intervals that correspond to the states. Then we calculate the probability of a state transitioning from time **t** to time **t+1**. Note that the sum of every row of the MTM equals one. The MTM is a square matrix (**n x n**) given by:

$$\text{Markov transition matrix,} \qquad \mathrm{MTM} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n.1} & p_{n.2} & \cdots & p_{n,n} \end{pmatrix}$$

Where the rows represent the current state (precipitation value at hour **t**), and columns represent the next state (precipitation value at hour **t+1**) and $0 \le p_{i,j} \le 1$ represents the probability of transitioning from state **i** to state **j**. After constructing the MTM matrix, we calculate the cumulative density function for each row of the MTM matrix as follows:

$$CDF = \sum_{1}^{j} p_{i,j}$$

To create the MTM, we divide our data into two categories: training and test data. Since we have hourly precipitation data for 10 years (2010-2019), we utilise 80% of the data for training and the remaining 20% for testing. Therefore, we use the data for the first eight years to obtain the MTM matrix. After obtaining the MTM using training data, we use this matrix to estimate synthetic hourly precipitation values for years 2018-2019, which is the testing data.

After calculating MTM, the next step is to generate the synthetic hourly precipitation values from the daily precipitation data. For each day, the initial point will be the 23rd hour (last hour) of the previous day. For the first reading and days which do not have consecutive records of daily precipitation, the initial point will be 0. Then for each hour, a random number between 0 and 1 is generated, and we cumulatively add the probabilities of the MTM until it exceeds the randomly generated number. This process repeats until all 24 hours have a synthetic precipitation value.

Following this, the accuracy of the generated values is tested by determining the error between the known daily precipitation and the sum of the hourly generated precipitation values.

$$\left| Daily\ Precipitation - \sum_{0}^{23} Hourly\ Precipitation \right| < Tolerance$$

If the error is less than the tolerance specified, then the synthetic values of precipitation are saved, and we move on to the next day. If the condition fails, the whole process of generating the hourly synthetic values repeats until the condition passes.

Results

So far, we have worked with data for Canberra, but results for at least the eight capital cities will be presented at the conference.

From the results generated, we get a similar stochastic profile as shown in Figure 3, which is promising. For quantification, we use two different metrics to evaluate the performance of our approach; *i)* the estimated number of rain hours compared to observed rain hours, *ii)* the root mean square error (RMSE) measure which estimates how accurately the model determines the response. The total number of estimated rain hours using our approach is 1078 hours, while the total number of observed rain hours is 1030. Note that the hours do not necessarily match and therefore, we calculated the RMSE to evaluate the error of hourly values estimation. Our results yielded an RMSE value of 0.8778 which is the mean difference between our estimated and observed values. Note that there is room for further improvement required to allow synthesized precipitation data to be included in weather and climate files offered for practitioner use.

For our model so far, we have only considered the occurrence of precipitation without considering other weather elements. These weather elements will be considered through data classification and clustering. For instance, when cloud cover is incorporated, clear sky hours will be clustered separately to overcast hours, each with its MTM. The project aims to determine the optimum number of parameters that will achieve a low RSME score. We will present these findings at the conference in December.

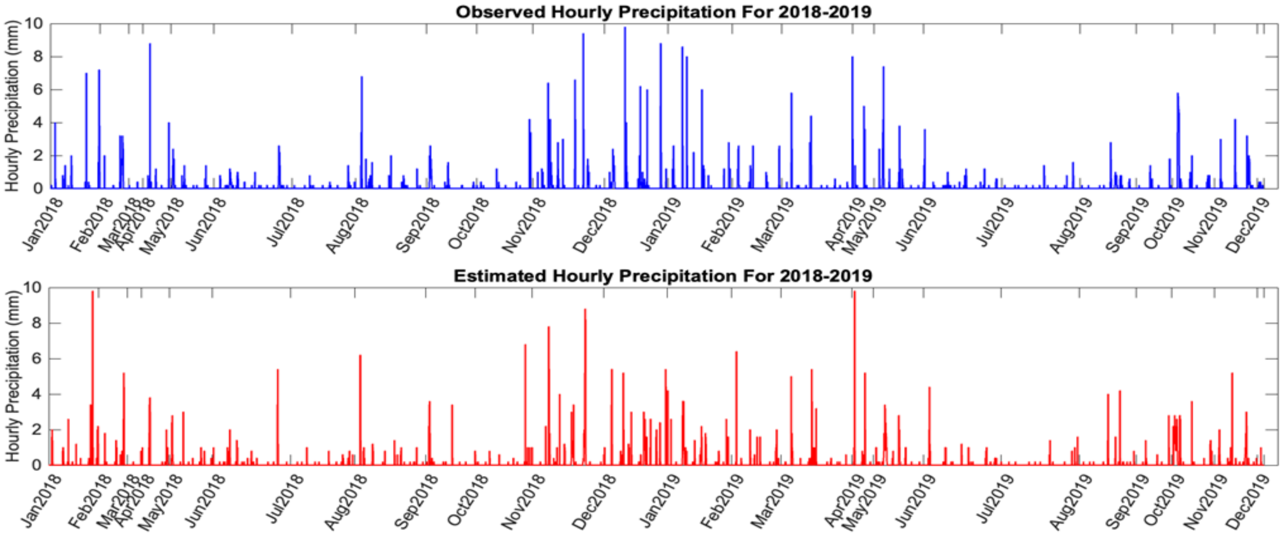

Figure 4: Observed (above) vs Generated (below) values of hourly precipitation

Acknowledgements

The authors acknowledge the Australian Bureau of Meteorology (BOM) as the source of all the raw weather data. We also acknowledge the other Exemplary people who worked on this project: Graham Anderson, Yoke Fung, ZhongRan Deng and Chun Yin Wu. Dr Dong Chen, Senior Principal Research Scientist, CSIRO Energy, is also thanked for his collaboration in designing the enhancement of the ACDB format to incorporate hourly precipitation data with a time stamp convention identical to its convention for solar radiation: 30 minutes either side of the time stamp.

References

Anderson, G. et al., 2020. *Updating Australia's Reference Meteorological Years (RMYs) with the addition of Hourly Precipitation Data.* Melbourne.

Jain, N., Lee, T. A. H. N. & Ferrari, D., 2021. *Extending Real Time Year weather data services with Bureau of Meteorology data.* Melbourne.

NCC, 2019. *Condensation and water vapor management.* [Online]
Available at: https://ncc.abcb.gov.au/editions-national-construction-code
[Accessed 8 2021].

Poduje, A. C. C. & Haberlandt, U., 2018. Spatio-Temporal Synthesis of Continuous Precipitation Series Using Vine Copulas. *Advances in Multivariate Analysis of Environmental Phenomena: Celebrating the 15th Anniversary of Copulas in Hydrology.*

Shamshad, A. et al., 2005. First and second order Markov chain models for synthetic generation of wind speed time series. *Energy,* pp. 693-708.

Tanaka, E. & Zirkelbach, D., 2016. *Local Climate Models for Hygrothermal Building Component Simulations.* s.l., Fraunhofer Institute for Building Physics, Germany.

Thyer, M. & Kuczera, G., 1999. Modeling long-term persistence in hydroclimatic time series using a hidden state Markov Model. *Water Resources Research ,* Volume 36, pp. 3301-3310.

Wilcox, S. & Marion, W., 2008. *Users Manual for TMY3 Data Sets.* Golden, CO: National Renewable Energy Laboratories.