Disaggregating Daily Precipitation Data 1990 to 2022 into Half-Hourly Intervals Using LSTM Models

Asia-Pacific Solar Research Conference 2024
5th December 2024

exemplary ENERGY

www.exemplary.com.au

Presenter: Harrison Oates (harrison@harrisonoates.com)

Authors: Harrison Oates, Nayan Arora, Hong Gic Oh and Trevor Lee

Outline

1. Background
2. Data Selection and Preparation
3. Model Architecture and Training
4. Results
5. Conclusions and future directions

The need for Half-Hourly Data for Built Environment Simulation

- At least thirty years of weather data is required to define climate norms and extremes[1]. Shorter periods may not produce reliable statistics
- Australia has only measured half-hourly precipitation from the late 1990s. Hourly or sub-hourly data is essential for reliable built environment modelling.
- Half-hourly precipitation necessary to ensure consistency between EnergyPlus Weather (EPW) and Australian Climate Data Bank (ACDB) timestamp conventions.

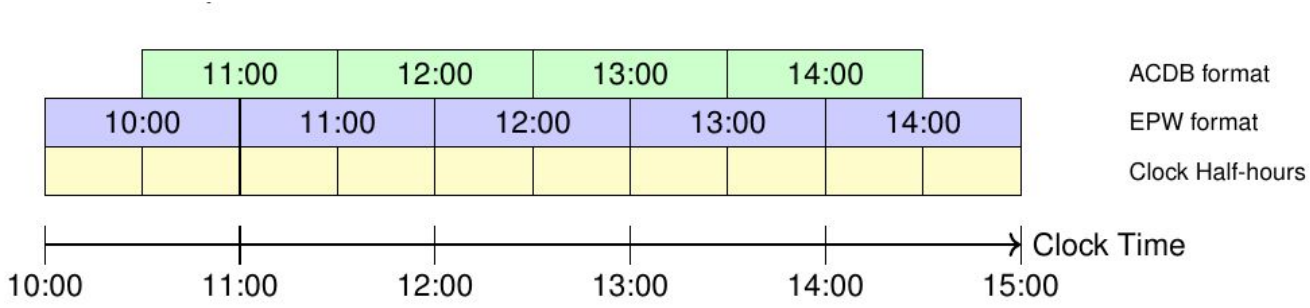

Figure 1: Timestamp conventions for ACDB and EPW formats

[1]World Meteorological Organization, 2023. 'Guide to Climatological Practices, 3rd edn.

exemplary ENERGY
www.exemplary.com.au

Data Selection and Preparation

- **Input features:** dew point temperature, dry bulb temperature, atmospheric pressure, and relative humidity
- **Preprocessing procedure for each station:**
  a. Linearly interpolate hourly non-precipitation weather records to half-hourly.
  b. Inner join linearly interpolated data with half-hourly precipitation records.
  c. Data Splits:
     - Test: 2020 – 2022
     - Validation: 2018 – 2019
     - Train: all remaining data

Table 1: Stations used in the study

Location	Climate Zone[2]	Precipitation Half-hourly Record Start
Brisbane	Climate Zone 2	2000-03
Sydney	Climate Zone 5	1998-12
Melbourne	Climate Zone 6	1997-10
Canberra	Climate Zone 7	2000-04

[2]Australian Building Codes Board, 2024, 'Climate zone map' Available at https://www.acdb.gov.au/resources/climate-zone-map. Accessed 5th July 2024

exemplary ENERGY
www.exemplary.com.au

Model Architecture

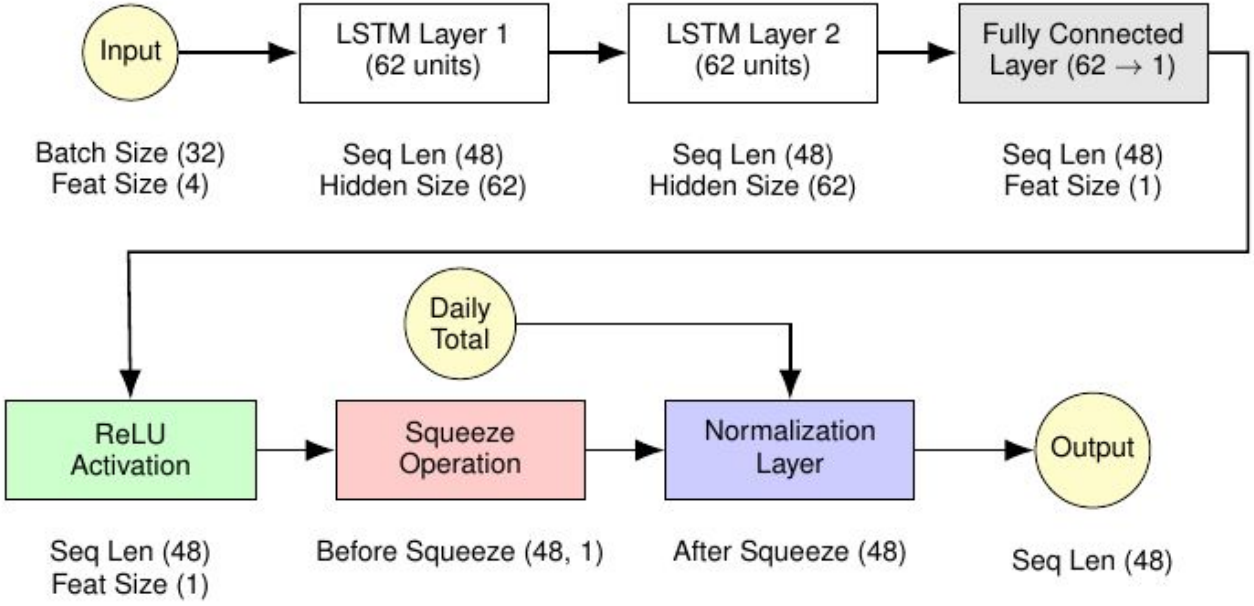

Figure 2: Model Architecture

exemplary ENERGY
www.exemplary.com.au

Model Training Details

- Loss function for a predicted tensor p and target tensor q is:

$$\ell(p,q) = \underbrace{\mathrm{MSE}(p,q)}_{\text{Mean squared error}} + \underbrace{KL(\sigma(p),\sigma(q))}_{\text{Kullback-Liebler Divergence}} + \underbrace{|V(p) - V(q)|}_{\Delta \text{ variance}} \qquad (1)$$

- Adam optimizer with initial learning rate of 10^{-3}
- Batch size of 32
- Learning rate scheduling with reduction on plateau strategy
- Initial run of 140 epochs → select epoch with lowest validation loss, then train for further 50 epochs at learning rate of 5×10^{-6}
- Trained on single Nvidia 4070ti Super GPU, ~2.8sec / epoch

www.exemplary.com.au

Results: Visualizations

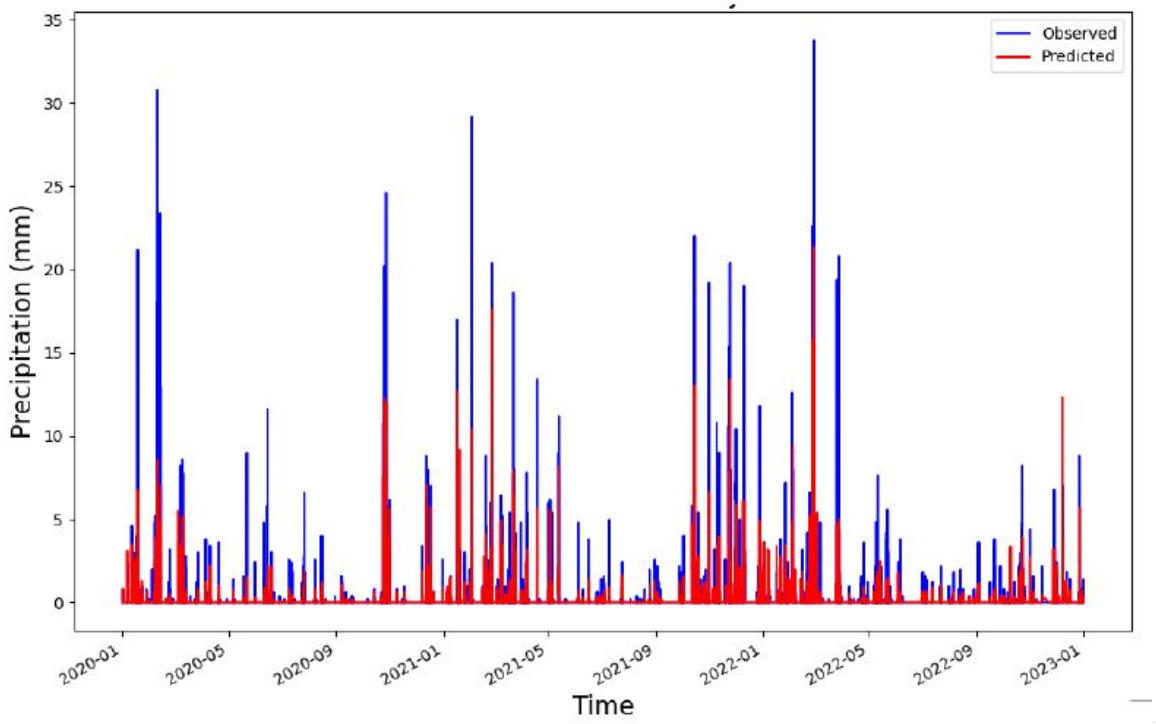

Figure 3: Half-Hourly series for Brisbane

exemplary ENERGY
www.exemplary.com.au

Results: Visualizations

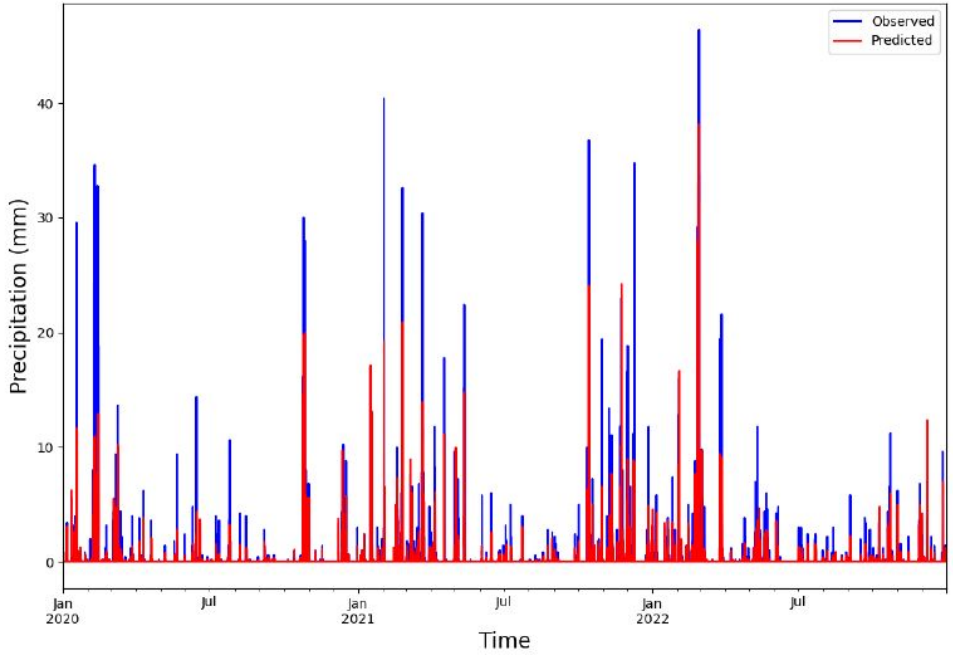

Figure 4: Hourly series for Brisbane

exemplary ENERGY

www.exemplary.com.au

Quantitative Results

Table 2: Model Results

Location	RMSE (mm)	Relative error in total number of rainfall half-hours	Proportion of correctly detected rainfall half-hours
Brisbane	0.57	8.36%	61.63%
Sydney	0.51	13.15%	69.71%
Melbourne	0.23	8.02%	56.76%
Canberra	0.29	22.35%	67.08%
Mean	0.40	12.97%	63.80%

Table 3: Comparison of re-aggregated results with Ferrari, et al.[3]

Model	RMSE (mm)	Relative error in total number of rainfall hours	Proportion of correctly detected rainfall hours
LSTM (average)	0.45	12.57%	69.04%
Markov Chain Monte Carlo	0.65	~7%	20%

[3]Ferrari, D., Mahmoodi, M., Kodagoda, C., Hameed, N.A., Lee, T., and Anderson, G., 2022, 'Disaggregation of precipitation data applicable for climate-aware planning in built environments'. Australian Building Simulation 2022 Conference Proceedings, p24-27

exemplary ENERGY
www.exemplary.com.au

Conclusions

- Model generally captures temporal patterns of rainfall
- Underestimates magnitude of extreme events
  - Smoothing of LSTM models
  - Relative scarcity of extreme rainfall instances in training data. For Brisbane:
    - 6.13% of wet days features half-hourly precipitation ≥ 10mm
    - 1.5% of wet half-hours ≥ 10mm
- LSTM yields lower error rates and improved detection of rainfall hours compared to Markov Chain Monte Carlo
- Tradeoff of increased error in total number of rainfall hours

exemplary ENERGY

www.exemplary.com.au

Further Directions

- Select architecture per climate zone to address performance variations
- Enhance model's ability to capture fine-grained precipitation variations by incorporating additional meteorological variables
- Refine model architecture to better handle rainfall intermittency
- Apply model to more locations and conduct more performance evaluations
- Train model on entire climate zone instead of individual stations – could improve model performance due to increase data availability

These refinements could further increase the model's accuracy under various climate contexts. The generated series can then be used to define a climate normal, ensuring that precipitation can be reliably used for modelling and simulation of built environments

exemplary ENERGY

www.exemplary.com.au

Thank You & Questions

Full manuscript submitted and under peer review - see https://tinyurl.com/APSRC2024Precipitation

- **Website:** exemplary.com.au
- **Enquiries:**
  - exemplary.energy@exemplary.com.au
  - harrison@harrisonoates.com

Harrison Oates, Nayan Arora, Hong Gic Oh, Trevor Lee

exemplary ENERGY

www.exemplary.com.au