ISBN for HWRS 2025 is 978-1-925627-81-7

Of fragments and cubes, a century of continuous rainfall in the Darling Downs

Lindsay Millard[1*], Michael Batchelor[1], Henry Hamer[1,] Luke Gericke[2]

[1]WRM Water & Environment [2] Department of Transport and Main Roads

[*]*Lindsay.millard@wrmwater.com.au*

ABSTRACT

This paper presents a method for developing 136 years of hourly climate inputs for continuous simulation modelling (CSM) in the URBS runoff-routing model. The authors devised an innovative approach for generating input datasets for design event modelling and CSM. The approach overcame a previous limitation, which required different modelling software and parameters, enabling a consistent model and parameters to confidently predict flood quantiles, average annual time of submergence (AATOS), and the economic measures required for Detailed Business Cases. Previous CSM methods used bespoke models and generated sub-daily rainfall inputs based on 1) repetitive sampling from a limited set of spatio-temporal patterns or 2) generating rainfall depths entirely stochastically.

The authors present an innovative methodology for creating a climatic 'data cube' of scale and detail as inputs to hydrologic models. The climatic data cube covered 450 km x 475 km with a 2.5 km grid cell resolution centred on the Darling Downs region. Each cell contained an hourly rainfall and evaporation value from 1 January 1889 to the present, meticulously produced from the best available data sources, drawn from 381 sub-daily and 1,556 daily rainfall gauges. The data cube was augmented with the available SILO and BARRA-R2 datasets. The paper outlines a comprehensive methodology for processing the rainfall datasets into a unified data cube.

The paper introduces a regionalised method that generates hourly rainfall from a pool of fragments from observed events. The point measures of rainfall are used to condition BARRA-R2 to preserve observed depths. The data cube was reconciled against Bureau of Meteorology intensity-frequency-duration (IFD) statistics. The data cube was then input into nine URBS basin models. Key techniques discussed in the paper include:

1) generating hourly datasets from available sub-daily donor rainfall fragments,

2) conditioning disaggregated daily depths with expected at-site IFDs,

3) constraining and modifying reanalysis grids to observed rainfall depths,

4) combining all data cubes into a final reconciled data cube to extract subarea inputs.

ISBN: 978-1-925627-81-7

INTRODUCTION

Overview

The Darling Downs is traversed by critical sections of the National Land Transport Network and seven state highways, comprising a total network spanning over 1,307 km and featuring 1,549 highway culverts and 57 bridges.

In 2024, the Queensland Department of Transport and Main Roads (TMR) undertook the Darling Downs Flood Study (DDFS) to help prioritise infrastructure upgrades across the network, as shown in Figure 1. The study focused on whole-of-link closure characteristics by developing continuous historical water level sequences at each major bridge and culvert reflecting the actual historical interdependence and relative timing of floods in adjacent catchments. The DDFS appraised the probability and duration of 16,384 potential network closure scenarios. Accurately assessing the frequency of food inundation in the network required a robust method to generate inputs for long-term, subdaily, continuous simulation modelling (LTCS).

Hourly streamflow datasets were generated by long-term continuous simulation modelling (CSM) (1889 to 2024) using URBS and gridded hourly rainfall and evapotranspiration datasets created for the project's 116,410 km^2 catchment area. This approach implicitly accounted for the complex interactions between antecedent conditions and the temporal and spatial variation of rainfall.

This paper presents the methodology for creating the 136-year hourly rainfall sequence for LTCS in the DDFS. This sequence, created using Python in the *rainsplitter.py* package, maximises the use of observed data for model calibration. The sequence was created by splicing 45 years of observed hourly rainfall data with 90 years of synthetic hourly rainfall. This approach aimed to maximise the use of observed data for model calibration. The method also involved the development of *rainfitter.py*, which constrained the stochastically generated hourly rainfall to match BoM's IFDs throughout the study area. This ensured consistency with established design rainfall standards and allowed for climate change impact assessments by scaling sequences to match future IFD relationships. The constraining step was essential to overcome the limitation of using gridded data that may not perfectly capture localised intense rainfall events. Constraining rainfalls to match IFD relationships required a solution for the challenge of inter-dependence between higher and lower duration rainfall extremes. Scaling at one duration can unintentionally alter another, requiring careful recursive adjustment.

ISBN: 978-1-925627-81-7

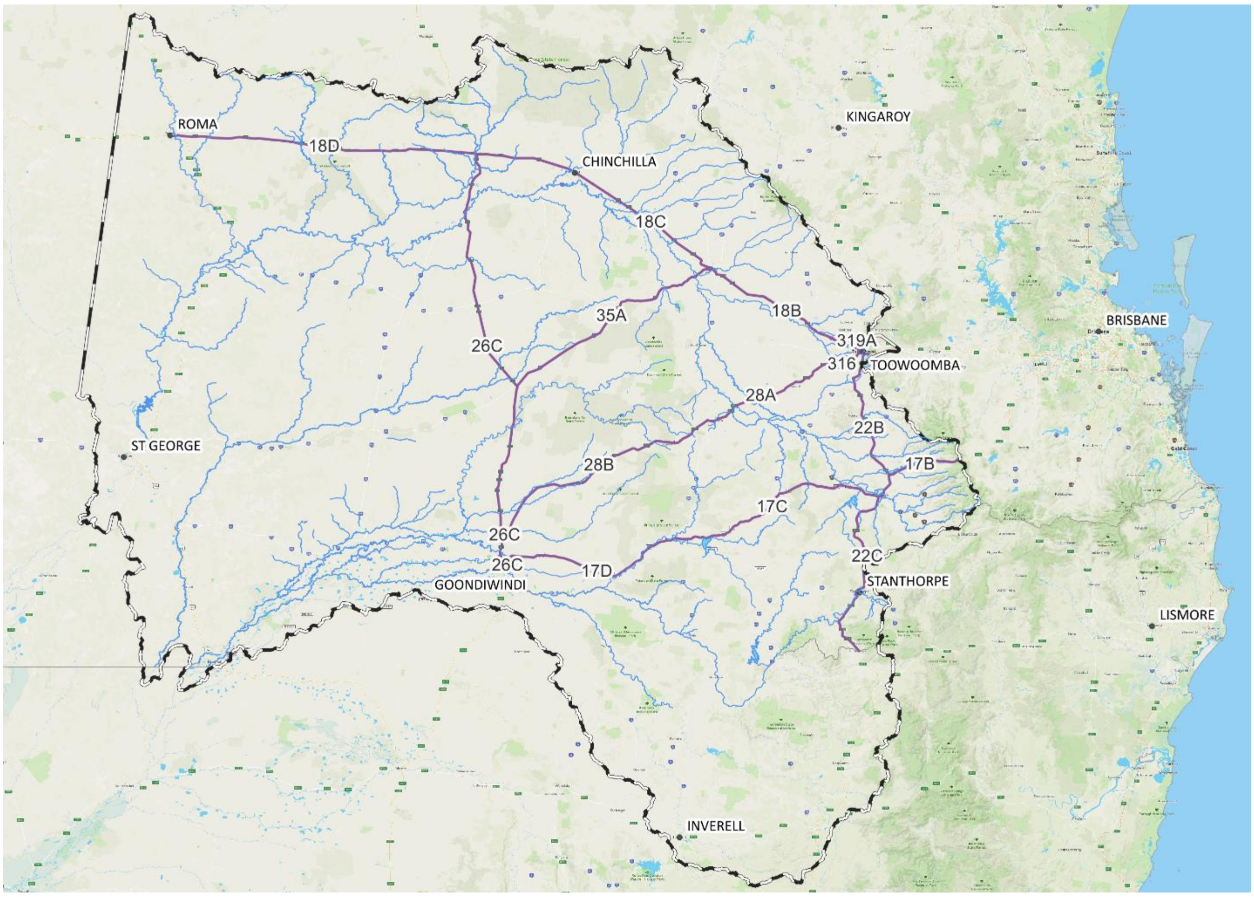

Figure 1: Project Locality

Defining the Problem

Unlike conventional design flood estimation, which focuses on individual crossings, continuous simulation with disaggregated rainfall enabled the assessment of interdependence and relative timing of flood impacts across 1,600 crossings throughout the 1,300 km network of seven highways. This allowed for more accurate calculations of critical statistics, such as Average Annual Time of Closure (AATOC) and Average Duration of Closure (ADC), for entire links, which are vital for economic analysis and prioritisation of upgrades.

Continuous simulation is a powerful method that can implicitly account for complex interactions such as event rainfalls, antecedent conditions and the behaviour of distributed systems sensitive to varying temporal and spatial scales of storms. Daily timestep continuous simulation (CSM) models are widely used to simulate the behaviour of water management systems. However, TMR required a disaggregation approach to generate a long, continuous time series of sub-daily (hourly or 6-minute) rainfall data. This was essential because observed sub-daily rainfall records are often too short or sparse for the required LTCS.

Previous approaches by others developed rainfall inputs primarily based on gridded daily rainfall data from the Australian Water Availability Project (AWAP). Daily rainfall totals from AWAP were then disaggregated into hourly values based on observed hourly patterns or within-day patterns randomly sampled from other recorded events. For the DDFS, we opted for a higher fidelity approach by

ISBN: 978-1-925627-81-7

generating a longer 136-year hourly rainfall sequence. A key difference was the adoption of layering the highest quality datasets. This involved splicing a dataset of available observations with 45 years of hourly rainfall data (from 1979-2024) with 90 years of hourly rainfall generated for the data-sparse earlier period (1889-1979). This maximised the use of observed data for model calibration.

METHODOLOGY

The following methodology outlines our approach to generating hourly datasets from long-term daily records. The strategy is based on sub-daily rainfall patterns extracted from surrounding 'donor' rain gauges to enhance the rainfall record and overcome data limitations, such as short or sparse observed sub-daily rainfall records. In this way, our approach provides the necessary hydrologic input to enable the modelling of long-term processes governed by rainfall variability.

Utilising as much observed rainfall data as possible facilitated the straightforward calibration of the URBS model with the observed streamflow data. The regionalised method of fragments (Westra, 2012 Boughton, 1999) was employed to disaggregate daily rainfall into smaller intervals, which allowed for confirmation of the disaggregated rainfall's Intensity-Frequency-Duration (IFD) characteristics against observed sub-daily rainfall records and traditional Bureau of Meteorology (BOM) IFDs. This approach ensured that the URBS models produced representative runoff sequences across all locations of interest in the study, regardless of the catchment size.

Conceptual overview

Figure 2 below illustrates the procedure for generating subdaily rainfall data. The process is divided into four steps and is briefly explained here:

(1) **Hourly rainfall database:** All available rainfall pluviographic stations were collected, filtered and gridded to generate a sparse data cube* representing the best possible observational data record. This sparse data cube was supplemented by backfilling the gaps that arose when the pluviograph had a missing temporal representation or the spatial extent between pluviometric stations was too wide. These gaps were filled by combining gridded daily rainfall totals with reanalysis spatiotemporal patterns, as described in (2).

(2) **Filling data gaps:** The rainfall database was infilled by combining the gridded daily rainfall totals and Bureau of Meteorology Australian Region Re-Analysis (BARRA-R2) hourly temporal patterns. This step is intended to use the BARRA-R2 data to provide an hourly temporal distribution of the recorded daily rainfall station totals, ensuring that the rainfall totals match the daily totals. As described in (1), where daily totals were missing or spatial coverage was inadequate, the SILO gridded rainfall data were combined with BARRA-R2.

(3) **Disaggregating recorded daily rainfall:** This step takes a daily rainfall record at the point of interest (POI) (as a long-term subdaily record is often unavailable) along with subdaily records at nearby gauges, and stochastically generates a subdaily rainfall record matching the observed daily totals. (Westra, 2012; Dykman, 2024) Section 7.4.2.2 of Book 2 ARR v4.2 (Ball et al, 2019) explains the Regionalised Method of Fragments. The approach follows the state-based approach outlined in Westra et al. (2012) and overcame the discontinuities between consecutive days in Boughton's 1999 Method of Fragments approach to disaggregate daily rainfall totals.

(4) **Match to IFDs:** The data cubes were then reconciled to match the BoM IFD and the IFD generated by pluviometric stations. This step checks and constrains the data cube to match the regionalised statistics. This process involved rigorous validation and explicit "constraining" of the generated synthetic hourly rainfall sequences to match BoM's IFD grids.

* spatial grids of rainfall depths with a time dimension, after Lewis, 2016

ISBN: 978-1-925627-81-7

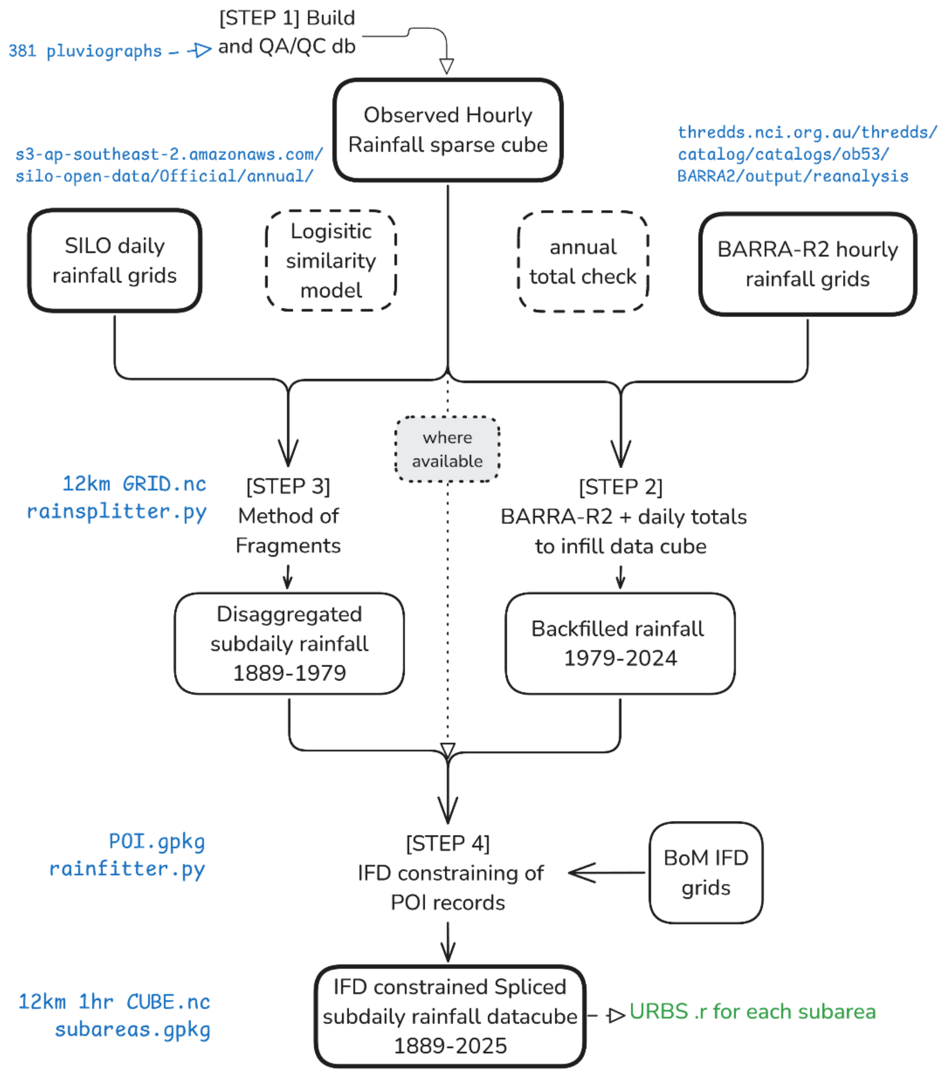

Figure 2: Process of generating rainfall data cube

Figure 2 illustrates the steps, detailed and explained below. The input pluviometric datasets described had records that ranged from 2 to 71 years, with a mean length of 28 years. Approximately 45% of the pluviometric data was missing. Point daily rainfall was obtained from the BoM's climate data service download. Gridded daily rainfall totals for DDFS were obtained from the Qld Government's SILO Datadrill service. SILO is a gridded daily rainfall surface derived either by splining or kriging the observational data at a 5 km grid resolution.

Step 1: Building the observed hourly rainfall database

The 381 pluviographs of varying record length were used as the input point data for the observed hourly rainfall database. The database generation process required a thorough quality control to generate the sparse data cube. Erroneous station data was removed, cumulative data was converted to hourly incremental data, and multi-day totals were spread over the preceding days based on records from nearby stations. The observed rainfall from each pluviometric station was transformed into IFD and compared with the BoM's 2016 IFDs. The data was resampled to an hourly step, and where it was

ISBN: 978-1-925627-81-7

sufficiently close to other stations, spatially interpolated to generate a sparse data cube. The spatio-temporal data cube gaps were filled by combining gridded daily rainfall totals with reanalysis spatiotemporal patterns, as described in Step 2.

Step 2: Infilling the sparse data cube using the BARRA-R2 and daily record

Point daily rainfall

The observational database developed in step 1 had spatial and temporal gaps in the data cube that required backfilling. The daily rainfall totals observed at points within the grid, were disaggregated across the hourly pattern of each day in BARRA-R2. This effectively converted all daily rainfall stations into sub-daily rainfall stations. Note that any sub-daily and hourly rainfall station records (outside of errors or missing data) were maintained in the database and not modified by this step.

Missing data from all point-source rainfall stations (daily and sub-daily) were then filled by creating a continuous hourly record at each station. The missing hourly values are interpolated using inverse distance weighting (IDW) from surrounding stations (within a 25 km radius). The layer of point hourly rainfall records was then interpolated and smoothed to create an hourly rainfall surface covering the study area for each hour from 1979 to 2024.

Gridded daily rainfall

Where no point daily rainfall were available for IDW, SILO's daily rainfall total was disaggregated using BARRA-R2 hourly temporal distribution. BARRA-R2 gridded hourly precipitation is available on THREDDS for 1979 – 2024. The BARRA-R2 dataset is not observed rainfall data but a reanalysis of climate model outputs on a 12 km grid. The BARRA-R2 dataset is not expected to estimate rainfall depths accurately; however, it does indicate the temporal and spatial distribution of rainfall throughout the study area at an hourly resolution. This generated a grid that could backfill any missing hourly values too far from observational data. This step uses the BARRA-R2 data to provide an hourly temporal rainfall distribution, with the rainfall totals conditioned to preserve the daily totals.

Step 3: Disaggregating the daily rainfall record using the Method of Fragments

Very few pluviometric stations were operational prior to 1979, and there is no BARRA-R2 coverage. To generate hourly rainfall prior to 1979, an approach known as the "regionalised" Method of Fragments (MoF) was required. This approach addresses data sparsity by sampling fragments from the full record of any nearby stations. Rather than randomly generating rainfall patterns, the MoF concept is "state-based," and compares the previous and next-day wetness states at the target location. The 'state' concept sets a boolean based on whether it was dry or wet. The MoF will select similar 'states' from pluviometric stations donating fragments to the target. In this way, the MoF approach explicitly disaggregates daily rainfall to hourly intervals from a pooled 500-year record of fragments of observed storm events at nearby "donor" pluviograph stations. From this pooled record, they were sampled to generate subdaily temporal patterns, which were combined with the daily rainfall total for each grid cell.

Pluviometric data and additional metadata were required for the MoF step. The metadata for each station defined its latitude, longitude, elevation, and distance to the nearest coastline. Observational pluviograph records were gathered from all available stations throughout southern QLD and categorised. These datasets are referred to as the 'donor' stations.

Disaggregated data was generated from the two inputs using *rainsplitter.py*, an implementation of the Regionalised Method of Fragments (Westra, 2012). This process is more detailed in Westra (2012) or Book 2 of ARR 2019 (Ball et al., 2019). Below is a high-level overview of this process:

ISBN: 978-1-925627-81-7

- Donor stations were ranked by their 'similarity' to the target station, based on a multivariate logistic regression model. The regression model considered latitude, longitude, the product of latitude and longitude, elevation and distance to the nearest coast. Book 2 ARR 2019 provides the logistic regression coefficient values in Table 2.7.7 (Ball et al, 2019)
- Data from donor stations was pooled to create a 50-year sampling dataset of rainfall 'fragments' – hourly rainfall extracts from the donor station data, each 24 hours in length. Fragments were preferentially drawn from donor stations in order of their similarity ranking, and fragments with missing data were discarded. The sampling of fragments continues searching until sufficient donor locations are assembled to define the pool of fragments available in the database.
- For each daily total at the target station, a group of candidate fragments would be found that had a similar daily total, fell in the same part of the year and had similar rainfall levels on immediately preceding and succeeding days.
- These candidate fragments were sampled based on a harmonic distribution determined by the inverse of their similarity ranking. The sampled fragment is then scaled to match the daily total observed at the target station and becomes the subdaily disaggregation for that day.

Step 4: Splice, condition and constrain

The previous steps generated a MoF-based data cube that covered the period from 1889 to 2024. The second data cube, based on BARRA-R2 and pluviometric data, covered the period from 1979 to 2024.

The two data cubes were constrained to the IFD datasets by recursively rescaling annual maximum and non-annual maximum rainfall to minimise biases. The Python *rainsplitter.py* package, also used in the DDFS, incorporates novel improvements for constraining timeseries to a target IFD. This has practical applications when assessing future climate scenarios, as it allows conditioning based on climate covariates (e.g., temperature), or to constrain the generated sequences to match future climate seasonality adjustments and the IFD relationships, as required in ARR v4.2. This approach is discussed in detail by Batchelor et al, (2025).

The hourly rainfall sequences were then analysed at every Point of Interest (POI) to check that they have IFD characteristics similar to those of both the sub-daily rainfall records and those retrieved from the BoM IFD grids. The IFDs underwent a constraining process, based on a modified implementation of the iterative IFD fitting algorithm proposed by Woldemeskel et al. (2016). The annual maximum rainfall rescaling factor is calculated for a priority duration (see Section 2.1 of Woldemeskel et al., 2016). To reduce the magnitude of the adjustment, this factor is then damped with a user-defined parameter. The process described in Batchelor et al, (2025) is repeated to ensure that subdaily events are statistically similar to the BoM IFD record.

The process for the creation of the 'spliced' 136-year rainfall period is outlined as follows:

- The synthetic hourly temporal pattern generated for each of the 11 target locations was assigned to all daily rainfall grid cells within the study via a Thiessen polygon analysis, creating a synthetic hourly disaggregation of daily rainfall at each grid cell;
- This synthetic record was combined with the 45-year observed hourly rainfall database record at each cell to generate a 136-year rainfall sequence at each URBS model subarea for continuous simulation modelling.

RESULTS

Figure 3 illustrates the steps involved in disaggregating daily rainfall and subsequently constraining the final dataset to the IFD. The IFD derived from the final hourly rainfall series was compared with the observed pluviograph IFD and the BoM2016 IFD. The 136-year spliced rainfall IFDs demonstrate an

ISBN: 978-1-925627-81-7

improved match against the observed rainfall IFDs (compared to the gridded rainfall disaggregation IFD) for storm durations less than 24 hours, particularly between 50% AEP and 1% AEP. The 136-year spliced rainfall IFDs show that the input rainfall sequences for the URBS continuous simulation model, as discussed in Cullen et al (2025), generally align with the BoM 2016 IFDs and observed pluviography IFD data. This consistency increases confidence that the flood quantiles estimated by the continuous simulation will closely match the results of Monte Carlo verification and any future traditional event modelling using the URBS models.

The IFD for the MoF disaggregated daily gridded rainfalls is generally consistent with the observed pluviograph IFD at each gauge for durations of 24 hours or longer. However, for shorter storm durations, the MoF disaggregated daily gridded rainfall IFD is lower than the observed pluviography IFD. The underestimation of short-duration rainfalls by the method of fragments disaggregation is caused by:

- The use of gridded daily rainfalls as an input, which are regionalised, and may not pick up the most intense rainfall recorded at each of the target stations; and
- The daily gridded rainfall is being disaggregated based on a temporal fragment from a potentially distant donor station,

The above issues mean that the disaggregated record is unlikely to produce short-duration rainfall bursts with the same intensities as the observed pluviographs.

DISCUSSION

The resulting spliced data cube could generate hourly rainfalls for all sub-area centroids within the study area. The hourly rainfall sequence was extracted and input into URBS hydrologic modelling software. The output from URBS flood hydrographs produced an excellent representation of flood frequency and average annual time of closure curves.

The approach outlined was regionalised, and sub-daily rainfall "fragments" were randomly sampled from nearby pluviograph stations only if they exhibited statistically similar daily-to-sub-daily scaling characteristics—the inclusion of a "state-based" logic. Fragments were selected conditionally based on daily rainfall amounts and whether the previous and following days were wet or dry. This significantly improved the continuity of the resampled sub-daily fragments and was enabled by the larger sample size from multiple nearby stations. This overcame the limitation of sparse local sub-daily records by leveraging data from a broader spatial domain. The logistic regression models were employed to predict the probability that daily-to-sub-daily scaling relationships were statistically similar between two locations. This was based on differences in physiographic properties (e.g., latitude, longitude, elevation, distance to coast) and specific metrics of sub-daily rainfall behaviour (e.g., maximum 6-minute intensity as a fraction of total daily rainfall, fraction of zero sub-daily timesteps, time of day for peak intensity.

While effective, MoF disaggregation of daily gridded rainfall sometimes resulted in lower Intensity-Frequency-Duration (IFD) characteristics for short durations compared to observed pluviograph data. This was attributed to the regionalised gridded input not always capturing the most intense local rainfall bursts, or the use of temporal fragments from potentially distant donor stations.

ISBN: 978-1-925627-81-7

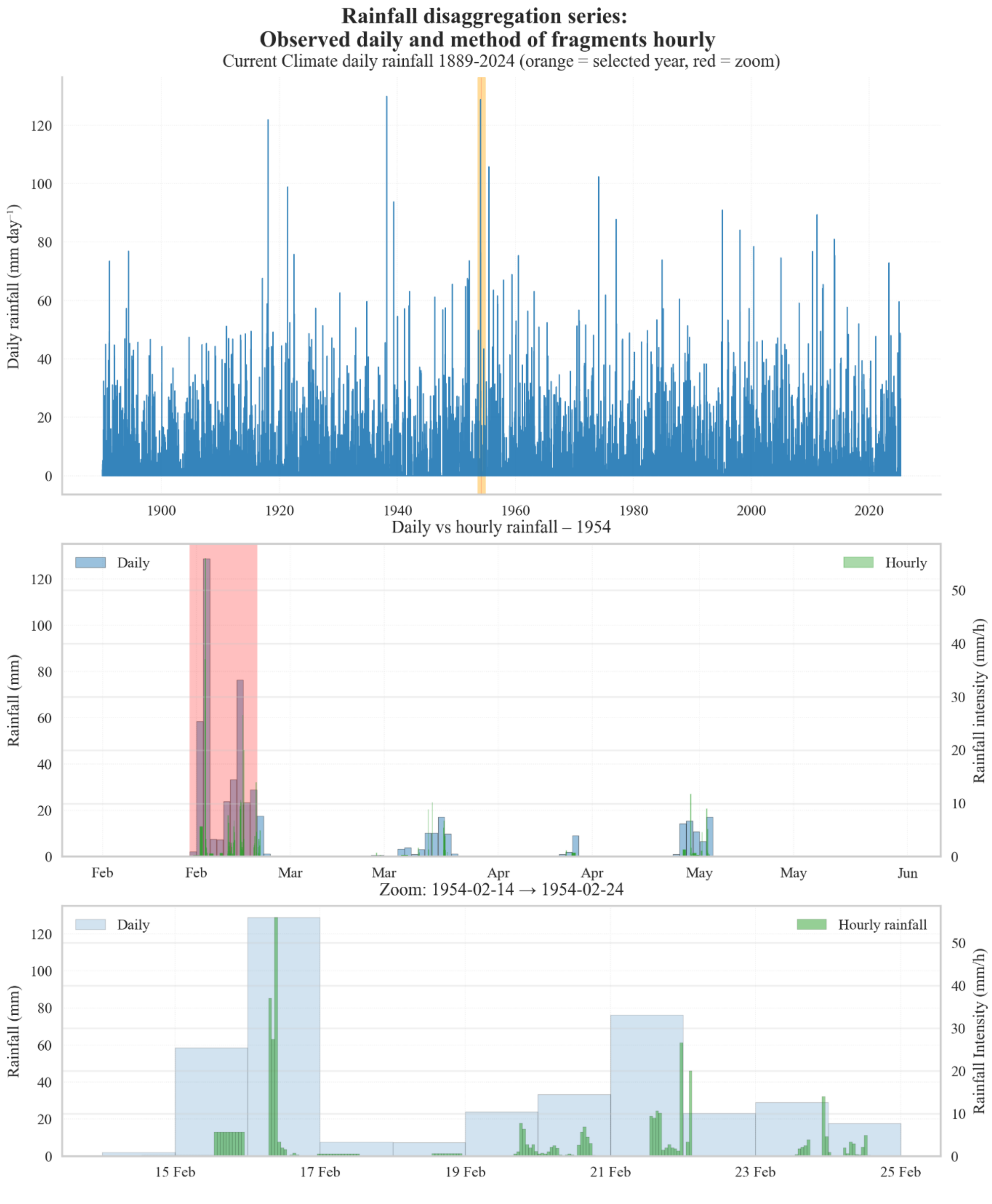

Figure 3: Hourly rainfall dataset disaggregated using MoF

ISBN: 978-1-925627-81-7

CONCLUSION

The culmination of this study was the development of *pysplitter* and *pyfitter.py, which* enabled the disaggregation and creation of long-term, continuous hourly rainfall sequences for each URBS model sub-catchment. For the DDFS, this resulted in a gridded 136-year hourly rainfall sequence covering 116,410 km² by combining a 45-year observed hourly rainfall dataset with 90 years of hourly rainfall disaggregated from daily data and available donor pluviometric stations. These sequences served as the primary input for the URBS rainfall-runoff models, enabling the estimation of flood hydrographs and flood closure statistics for the highway network. The application of the rainfall sequences and fitting of flow is discussed in Cullen et al (2025). While stochastic generation techniques have been shown to create reliable synthetic sequences, the study also reinforces the value and need to base simulations on modified actual sequences of rainfall.

ACKNOWLEDGEMENTS

The authors are grateful for the collaborative efforts of the extended project team, including Corey Baldwin, Trinity Graham and Greg Rogencamp and TMR's project sponsors in the Southern Queensland Region's Emma Kirkby and Hydraulics Branch's Chris Russell and Jagath Abeynayake for their partnership.

Code and data availability. The code described in this study was Python 3.11 and was adapted from the method described in Dykman et al (2024). The data presented in this paper can be made available at the request of the corresponding author. The dataset of the Australian rainfalls were purchased from the Australian Bureau of Meteorology. Daily rainfalls and Design rainfall IFDs are available at www.bom.gov.au/climate/data/ and www.bom.gov.au/water/designRainfalls/revised-ifd/

REFERENCES

Australian Bureau of Meteorology, 2016. Design Rainfall Data System. http://www.bom.gov.au/water/designRainfalls/revised-ifd/.

Batchelor, M. Millard, L. Hamer, H (2025) Scaling of historical sub-daily rainfall datasets to represent future climate conditions Hydrology and Water Resources Symposium

Boughton, W.C. (1999). A daily rainfall generating model for water yield and flood studies. Rep. Report 99/9, 21pp, CRC for Catchment Hydrology, Monash University, Melbourne

Boughton, W., & O. Droop (2003). Continuous simulation for design flood estimation—a review, Environ. Model. Software, 18(4), 309–318

Cullen, R. Millard, L. Guse, H. Carroll, D. and Abeynayake, J. (2025) Beyond the event-based horizon: URBS models for long-term continuous simulation in the Darling Downs

Dykman C, Sharma A, Wasko C, Nathan R (2024). Pyraingen: A Python package for constrained continuous rainfall generation https://pyraingen.readthedocs.io/en/latest/index.html https://doi.org/10.1016/j.envsoft.2024.105984

Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data, Environmental Modelling and Software, Vol 16/4, pp 309–330. DOI: 10.1016/S1364-8152(01)00008-1

Lewis, A., Lymburner, L., Purss, M.B., Brooke, B., Evans, B., Ip, A., Dekker, A.G., Irons, J.R., Minchin, S., Mueller, N. and Oliver, S., 2016. Rapid, high-resolution detection of environmental change over continental scales from satellite data–the Earth Observation Data Cube. International Journal of Digital Earth, 9(1), pp.106-111. https://doi.org/10.1080/17538947.2015.1111952

ISBN: 978-1-925627-81-7

Stocker, T.F., et al., 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.IPCC, Cambridge, United Kingdom and New York, NY, USA.

Westra S, Mehrotra, R Sharma, A & Srikanthan R (2012). Continuous rainfall simulation: 1. A regionalised subdaily Disaggregation approach WATER RESOURCES RESEARCH, VOL. 48, W01535, doi:10.1029/2011WR010489, 2012

Woldemeskel, F., Mehrotra, R., Sharma, A., Westra, S. (2015). PROJECT 4 Continuous Rainfall Sequences at a Point. Australian Rainfall & Runoff.

Woldemeskel, F.M., Sharma, A., Mehrotra, R., Westra, S. (2016). Constraining continuous rainfall simulations for derived design flood estimation. J. Hydrol. 542, 581–588.https://doi.org/10.1016/j.jhydrol.2016.09.028.

BIOGRAPHY

First Author is Lindsay Millard. Working with a strong focus on hydrologic analysis. His expertise extends to flood and water resource modelling for natural resource management projects, the transport and mining sectors, and flood risk assessments for local Councils. He has led hydrologic and hydraulic modelling on multiple design flood studies and has extensive experience using industry software. Lindsay has played a pivotal role in flood studies and floodplain management plans, overseeing their delivery and ensuring their technical quality. His projects have involved the calibration of historic events, assessment of mitigation options, flood damage assessments, consequence and cost-benefit analysis. https://au.linkedin.com/in/lindsaymillard

Second Author is Michael Batchelor. With thirty years' experience, Michael has contributed to the planning, design and delivery of major infrastructure projects across Australia. He has undertaken numerous flood studies, design assessments and water resource yield studies for major road, mining and water resource projects. Michael has also authored the water resources chapters of a number of environmental impact statements. He has designed water supply dams, water distribution systems, stream diversion works, sedimentation dams and associated drainage systems for diverse locations, especially across northern Australia.

ISBN: 978-1-925627-81-7