USING SURFACES OF BIG DATA TO UNDERPIN CONTINUOUS SIMULATION IN SYSTEMS ANALYSIS

Peter J Coombes¹, Michael E Barry²

ABSTRACT: The systems framework was designed to explore the multiple scale impacts of water, energy and environmental decisions using a bottom up approach to spatial systems analysis. This approach utilises big data layers of local information to create sub-daily surfaces of climate, stormwater runoff, streamflow and urban water demand. These surfaces of information and a multi-scaled framework of processes permit the generation of multi-site synthetic rainfall and associated continuous simulation of urban processes. This approach underpins the understanding of the multiple scale systems dynamics of water and energy systems. This paper presents the methodology for simulating multi-site rainfall and water demands based on surfaces of climate, demographic, socioeconomic, topography and environmental information. The approach satisfactorily reproduces observed daily, monthly, annual statistics of rainfall and water use at multiple scales. This demonstrates that the method is able to capture the inter-annual persistence and spatial variability of water use and rainfall that exists within the Greater Melbourne and Sydney regions. The systems method was able to adequately estimate regional water demand including the day to day variation, distributions and strong seasonal patterns. The bottom up construct of the Systems Framework can provide robust evaluation of the regional responses of local interventions such as Water Sensitive Urban Design.

KEYWORDS: Big Data, Synthetic Rainfall, Scales, Continuous Simulation, Water Demand, Stormwater

¹ Peter J Coombes, Urban Water Cycle Solutions, Newcastle, Australia. Email: peter@uwcs.com.au
² Michael Barry, BMT WBM, Brisbane, Australia. Email: michael.barry@bmtwbm.com.au
1 INTRODUCTION

The systems framework\(^1\) was designed to explore the multiple scale impacts of water, energy and environmental decisions using a bottom up approach to spatial systems analysis. This approach utilises big data layers of local information to create sub-daily surfaces of climate, stormwater runoff, streamflow and urban water demand. These surfaces of information and a multi-scaled framework of processes permit the generation of multi-site synthetic rainfall and associated continuous simulation of urban processes. This approach underpins the understanding of the multiple scale systems dynamics of water and energy systems.

This paper presents the methodology for simulating multi-site rainfall and water demands based on surfaces of climate, demographic, socioeconomic, topography and environmental information. The approach extends the previous work of Coombes et al., (2003).\(^2\) Climate series generated by the multi-site method were used with non-parametric matching sourced from molecular sciences to incorporate water balances in households in a regional water demand model to estimate daily water demand in the Sydney and Melbourne regions. This bottom up method was evaluated using observed spatial variability of climate and water use within the Greater Melbourne and Sydney urban regions by comparison to observed daily, monthly, annual statistics of rainfall and water use at multiple sites.

2 INCLUDING SPATIAL CLIMATE

Urban areas are subject to considerable spatial and temporal variation in climate (rainfall and temperature) that is a driver of spatially variable stormwater runoff and water use behaviours. The Systems Framework processes account for this variability of urban systems by incorporating bottom up inputs throughout urban regions in continuous simulation.\(^3\) This process involves methods to transform typically fragmented data from inconsistent time periods (such as from the Bureau of Meteorology) into temporally and spatially continuous surfaces of information from the same time period.

The non-parametric nearest neighbourhood schemes outlined by Coombes et al., (2002; 2004) are utilised to develop spatial surfaces of daily rainfall, days with rainfall (a measure of rainfall frequency), and minimum and maximum temperatures.\(^4,5\) This process is described for constructing surfaces of daily rainfall for a given region as follows:

1. Define the boundaries of the required region (such as the Greater Melbourne and Greater Sydney water supply regions) and subdivide into zones required for analysis (such as local government areas);
2. Select all observed daily rainfall records within or near the region and prepare records for analysis;
3. Evaluate all rainfall records to determine lengths and completeness of data;
4. Determine the distances from the centroid of each selected zone to the locations of all rainfall records;
5. Select the time horizon of the new daily rainfall surface (such as 1913 to 2013) that occurs within the time period of the observed rainfall data;
6. Construct a complete daily rainfall record for each zone using the nearest complete sequences of observed rainfall that matches the time period of missing data;
7. Provide statistics about the development of the rainfall record for each zone such as main source of observations (BOM file number) and distance from the zone, and details of observed rainfall sites and associated distances from zones used for infill of missing data.

This process to produce daily rainfall surfaces is enacted using computer software within the Big Data part of the Systems Framework.

2.1 RAINFALL DEPTH AND FREQUENCY

Development of surfaces of daily rainfall and associated frequency of rainfall (annual rain days) for Greater Melbourne and Sydney regions for the period 1913 to 2013 is presented in this section. These results were used to construct the spatial plots of average annual rainfall and rainfall days for the Greater Melbourne region shown in Figures 1 and 2. These Figures demonstrate that the Greater Melbourne region experiences a high level of spatial variation in rainfall depth and frequency with higher values in the east to lower values in the west. Development of a daily rainfall for each of the 36 zones in Greater Melbourne involved analysis of 494 daily rainfall records from the BOM.
The outputs from this process include diagnostic information and statistics from each zone. For example, construction of daily rainfall for the Banyule zone involved selection of adjacent rainfall records from minimum, mean and maximum distances of 1.3 km, 4.32 km and 7.2 km. Average monthly statistics for average daily rainfall and rain days are also produced.

Daily rainfall was also constructed for 46 zones across the Greater Sydney region. The resultant surfaces of average annual rainfall and rainfall days are shown in Figures 3 and 4. These Figures show the spatial variation in rainfall depth and frequency that varies with distance from the coast. The method utilised 493 observed rainfall records to reproduce this variability across 46 local government zones.

Construction of daily rainfall for the Ashfield zone, for example, involved selection of adjacent rainfall records from minimum, mean and maximum distances of 0.49 km, 0.85 km and 6.4 km. Creation of a spatial surface of daily rainfall records of the same length has reproduced the natural spatial variation of rainfall and permits relative analysis of the solutions throughout the region.

2.2 SURFACES OF CONTINUOUS RAINFALL

The Systems Framework utilises the software engine from the Probabilistic Urban Rainwater and wastewater Reuse Simulator (PURRS) to continuously simulate local water balances at sub-daily time steps (typically 6 minutes). This detailed local process is necessary to capture the distributed local behaviours of people, buildings and land uses that impact on infrastructure and
ecosystems, and drive the performance of the entire system. Continuous rainfall (6 minute time steps) records of equal length are required for each local zone within a region to facilitate simulation of local behaviours. Synthetic continuous (6 minute time step) rainfall records are derived for each zone using the local synthetic daily rainfall within a non-parametric nearest neighbourhood scheme. Data from nearest observed pluviograph rainfall records is utilised to disaggregate daily rainfall into a synthetic continuous rainfall record. The concept is illustrated in Figure 5.

Figure 5: Diagram of the non-parametric nearest neighbourhood scheme for development of synthetic continuous rainfall records

Figure 5 shows the concept that is enacted for each day in the daily rainfall record to select a day of pluviograph rainfall (6 minute intervals) using climate and seasonal parameters, and a ranking scheme. The non-parametric scheme matches climate and seasonal parameters (daily rainfall depth, month, count of days since last rain event) at the daily rainfall and at the nearby pluviograph rainfall sites to select a day of pluviograph rainfall from the most appropriate nearby pluviograph record.

Nearby pluviograph records are ranked on the basis of proximity to the location of the daily rainfall record, similarity of seasonal rainfall depths, topography and distance from the coast. This allows disaggregation of daily rainfall records into a series of storm events and dry periods that constitute a synthetic continuous rainfall record. The pluviograph rainfall records from the BOM used to make synthetic continuous rainfall for Banyule (for example) are summarised in Table 1.

<table>
<thead>
<tr>
<th>Location</th>
<th>Number</th>
<th>Period (yrs)</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bundoora</td>
<td>86351</td>
<td>1984 - 2012</td>
<td>4.1</td>
</tr>
<tr>
<td>Preston Res</td>
<td>86096</td>
<td>1929 - 1974</td>
<td>7.2</td>
</tr>
<tr>
<td>Melbourne RO</td>
<td>86071</td>
<td>1873 – 2011</td>
<td>13.1</td>
</tr>
<tr>
<td>Mitcham</td>
<td>86074</td>
<td>1939 - 1977</td>
<td>13.4</td>
</tr>
</tbody>
</table>

3 INCLUDING SPATIAL RESIDENTIAL WATER USE

The use of average water demands and household sizes to simulate the performance of urban water strategies produces considerable errors and uncertainty. Annual average household water demands were derived for local government areas using water utility billing records for Greater Melbourne (2004 to 2005) and for Greater Sydney (1996 to 2003) as shown in Figures 6 and 7. It is important to note that these results for annual average water use are the average of all dwellings types and households sizes within each local government area. Nevertheless, Figures 6 and 7 reveal that average household water use for local government areas is subject to significant variation across the regions. This variation is influenced by a range of factors including the distribution of dwelling types, household sizes, climate and household income. It is a key issue that the average household water use for each zone is not the actual water use in each dwelling.

Figure 6: Average annual residential water demand across Greater Melbourne (2004 -2005)

The performance of urban water strategies is primarily dependent on water use behaviour at each household, building and land use.
Information about the distribution of household sizes and dwelling types was available for local government areas from the Australian Bureau of Statistics.\cite{8} Water demands at any location are dependent on the distribution of household sizes and dwelling types. As demonstrated in Figure 8 for the Melbourne zone and in Figure 9 for the Wyndham zone, distributions of dwelling types and household sizes are vastly different across a region. These local variations of household sizes and dwelling types are included in the Systems Framework to overcome these differences that skew average water use values.

As shown in Figure 8 and 9, for example, the dwelling stock in each zone comprises different dwelling types that also generate different behaviours that will influence the characteristics of household water use. A detached dwelling may allow opportunity for significant outdoor water use whilst a unit dwelling is unlikely to provide opportunities for significant outdoor use. The known distributions of household sizes and dwelling types provide an opportunity to disaggregate average household water demands sourced from a water utility for a local government area into the likely water demands in each dwelling. This task also requires an estimate of the proportion of water demand that is used outdoors.

The variability of outdoor water use for various household types in different climate zones is not usually measured. A unique study of household water use analysed indoor and outdoor water use in 192 houses across 5 climate zones, 14 demographic regions and 12 years in the Hunter region of New South Wales and derived relationships for estimating monthly average daily outdoor (OutDem: Equation 1) and indoor (InDem: Equation 2) use as follows:\cite{4,9}

\[
\text{OutDem} = 7.5M - 11.3\text{AvexR} - 0.025\text{Inc} - 0.82\text{Rdays} + 24.44G + 19.1\text{AvexT} - 251
\]  
\[
\text{InDem} = 27.8 + 145.7P - 0.4M - 10.6\text{AvexR} + 6.7\text{Rdays} - 0.16\text{Inc} - 12.3G + 0.49\text{AvexT}
\]
where P is household size, M is a seasonal index with values from 1 to 6 (January and December = 1; June and July = 6), Inc is the average income of people in the household, AveR is average monthly daily rainfall, G is annual population growth, Rdays is the number of rain days in each month and AveT is the average monthly daily maximum temperature.

The relationships for monthly average daily indoor and outdoor use were derived using the climate outputs derived using processes described in Section 2 and using socioeconomic data from the Australian Bureau of Statistics (ABS). These water use values are used as central boundary conditions in the climatic and economic dependent water use simulations of dwellings within the Systems Framework. The process of developing calibrated household water use models is summarised as follows:
1. Determine monthly average daily indoor and outdoor water use for each zone in a given region using Equations 1 and 2;
2. Use monthly average indoor and outdoor water use boundary conditions with synthetic continuous rainfall, climate and demographic statistics in the PURRS model. Outdoor use in a semi-detached or unit dwelling is 10% or 5% of outdoor use in a detached dwelling. Outdoor use is independent of household size. Simulate the performance of 5 household sizes in each of 3 dwelling types (for example);
3. Combine distributions of household sizes and dwelling types with results from the preliminary simulations of water use at each dwelling to develop an average water use for each zone;
4. Adjust observed water use values to account for take up rates of water efficient appliances and rainwater harvesting. This information was sourced from the data underpinning the ABS Environmental Series publications[10] as discussed in Coombes et al., (2018)[11]. This process develops the base water use of households for use in calibration. Additional dwellings with water efficient appliances and rainwater harvesting is added to simulations after calibration to ensure accurate responses;
5. Compare observed average water use (from a water utility) to the simulated average water use for the time period and intervals of available observations. For example, the observed time period for Sydney was 1996 to 2003 and the interval of observations was quarterly. Calibrate the local scale models to reproduce the observed average water use.

This process provides calibrated indoor and outdoor use values for each household size in dwelling type for each zone (see Figure 10).

Figure 10 demonstrates the calibration of observed and predicted average household water use for the Bankstown local government area using observed data from Sydney Water Corporation. The calibrated average daily indoor and outdoor uses for the area are also shown in Figure 10. The residential land uses were combined with non-residential land uses including agricultural, commerce, industry, education, medical, forests, irrigated parks and transport. Local scale continuous simulations are completed for each dwelling type with different levels of known water efficient behaviours and for land use in each zone at time steps ranging from one second to six minutes using the local sequences of rainfall data. Outputs from the local scale analysis include sequences of water demands, wastewater discharges, stormwater runoff, energy demand, water quality, soil moisture and finances. These results are combined with climate data and passed to the Transition Framework as reference files.

3.1 A FRAMEWORK TO MAKE WATER FLOWS AT A ZONE

The sequences of water use, wastewater flows, stormwater runoff, financial transactions and energy use from the local scale analysis were combined in each zone using town planning projections and replicates of daily spatial climate sequences as shown in Figure 11. A transition framework is used to generate daily water cycle responses for each zone. Sequences of daily water and energy balance, and financial results from local scale are linked using seasonal information and historical climate data (including daily rain depths, cumulative days without rainfall and average daily
maximum ambient air temperature) to create resource files of water demand, wastewater generation, stormwater runoff, energy use and economic transactions.

This method of non-parametric aggregation (Coombes et al., 2002) generates daily outputs from each zone using the historical resource files and climate replicates generated for the simulation of the regional system. Climate replicates are multiple sequences of equally likely future climate drivers (such as rainfall, temperature) that are generated using Monte Carlo processes.

### 3.2 REGIONAL SCALE PROCESSES

The Systems Framework combines water, wastewater and stormwater infrastructure networks with catchments and inlets in an integrated network. Spatial and temporal information generated by the lot scale simulations are combined by the zone scale transition as inputs to the network analysis. Details of systems analysis of the Melbourne and Sydney water supply systems are provided by Coombes (2012; 2005). The regional scale of the Framework includes water sources from ground water, surface water sourced from regional river basins, shared surface water with other river basins, wastewater reuse and stormwater harvesting. The linked analysis utilises stream flows, reservoir storage volumes, wastewater discharges, information about the operation of water systems and data from global climate model as inputs. The simulations also include operating rules and regional policies such as water restriction triggers. The behaviour of the System Framework is verified at the regional scale using a hindcasting process (described below) that compares predicted and observed behaviours for key processes within historical time periods.

### 4 RESULTS

The results of predicted rainfall and water demands are compared to observed values in this Section. Recognised two sample statistical tests (t Test and Z scores) are also used to understand the results.

#### 4.1 RAINFALL

The efficacy of the synthetic daily rainfall process was evaluated by comparison to the two longest observed daily rainfall records at Observatory Hill in the Sydney zone and at Melbourne RO in the Melbourne zone as shown in Figure 12. Note that the Observatory Hill daily rainfall record was not used to create the synthetic daily rainfall for the Sydney zone.

Figure 12 shows that the synthetic daily rainfall was similar to the observed rainfall across the entire distribution of rainfall depths in Melbourne and Sydney. The mean synthetic daily rainfall for the Melbourne zone was the same as the observed mean rainfall with 99.99% level of certainty. The coefficient of determination $R^2$ indicated that the synthetic rainfall described 99% of the variation in the observed rainfall. The predicted 150 average annual rain days were similar to the observed 149 average annual rain days. This result was expected because the Melbourne RO rainfall was included in the process of estimating rainfall for the Melbourne zone.
The mean synthetic daily rainfall for Sydney zone was the same as the mean observed rainfall with a 95% level of certainty. The $R^2$ value indicated that the synthetic rainfall described 96% of observed rainfall. The observed annual average rain depth of 1,221 mm was similar to the predicted rainfall depth of 1,215 mm. The predicted 119 average annual raindays was 15% less that the observed 139 raindays. These results are excellent given that the local Observatory Hill observations were not used to develop the synthetic rainfall for the Sydney zone.

Local scale processes in the Systems Framework rely on synthetic continuous rainfall records to simulate the performance of households and land uses in each zone. The distributions of hourly rainfall totals from Synthetic continuous rainfall and observed pluviograph rainfall was evaluated by comparison to the two longest pluviograph rainfall records at Observatory Hill in the Sydney zone and at Melbourne RO in the Melbourne zone as shown in Figure 13. Note that the lengths of the observed pluviographs was different to the lengths synthetic records.

Figure 13 demonstrates that the predicted continuous rainfall recorded similar distributions of hourly rainfall as the observed records. The small difference between the distributions at the Sydney zone at less than 1 mm hourly rain depth may be caused by the different lengths of the predicted and observed records. These results indicate that use of the synthetic continuous rainfall in local simulations will produce realistic patterns of sub-daily water balance responses such as stormwater runoff, rainwater harvesting and impacts on wastewater networks.

4.2 REGIONAL WATER DEMAND

The “bottom up” process of generating local water use from dwellings and land uses in each zone was evaluated by comparison of historical observed water use to predicted water for the entire Greater Melbourne and Sydney regions. The predicted daily and monthly water demands for Greater Melbourne are compared to the historical observations in Figure 14.

Figure 14 demonstrates that the predicted water demands are consistent with the seasonal patterns of the observed demands for the Greater Sydney region. The similarity between predicted and observed water demands in the later portion of Figure 14 also show that the Systems Framework processes has also reproduced the reduced water use due to increased household water efficiency and responses to water restrictions during the period 2005 to 2009.

The mean predicted daily demands for Greater Melbourne are the same as the observed daily demands with 99% level of certainty. The $R^2$ value indicates that predicted daily demands described 66% of the variation in observed water demands.
The mean of predicted monthly demands are the same as the mean of observed monthly demands with a 95% level of certainty. The R² value indicates that the predicted monthly demands describe 95% of the variation in observed monthly demands. These results show that the mean across observation horizon and much of variation of observed water demands for Greater Melbourne were successfully predicted using the framework of bottom up water use from across the region.

The predicted daily and monthly water demands for Greater Sydney are compared to historical observations in Figure 15 for the period 1997 to 2005.

Figure 15: Observed and predicted daily and monthly water demands for Greater Sydney (1997 – 2005)

Figure 15 shows that the predicted water demands have reproduced the daily and monthly patterns of regional water demands for Greater Sydney. The predicted and observed mean daily demands are the same with 99% level of certainty and the R² value suggests that predicted daily demands describe 62% of the variation in daily observed demands. The mean of predicted monthly demands was the same as the mean of observed monthly demands with a 95% level of certainty. Predicted monthly demands describe 68% of the variation in observed monthly demands as indicated by the R² value. Distributions of daily water demands for Greater Sydney (top pane) and Greater Melbourne (bottom pane) are compared to observed demands in Figure 16 for the period of limited water restrictions from 1997 to 2005.

Figure 16: Distributions of predicted and observed daily water demands for Sydney (top pane) and Melbourne (bottom pane) for 1997 to 2005.

Figure 16 demonstrates that the distributions of predicted daily demands is similar to the distributions of observed observed daily demands for Greater Sydney and Melbourne. Distributions of predicted and observed daily demands are compared for the period of water restrictions (2005 – 2009) in Greater Melbourne in Figure 17.

Figure 17: Predicted and observed daily water demands before water restrictions (top pane) and during water restrictions (bottom pane) for Greater Melbourne

Figure 17 demonstrates that the bottom up process of predicting water demands was able to produce a high level of agreement with the observed
demands. This indicates the Systems Framework processes were about the predicted the changes in household water saving behaviours and water restrictions on regional water demands.

5 DISCUSSION

This study has highlighted that the Greater Melbourne and Sydney regions are subject to strong spatial variation in household water use, rainfall depths and frequency. Observed average household water demands for a local government area do not represent the average water use in the different dwellings and household sizes that are the components of the area. Continuous simulation of the performance of different household sizes in a variety of dwelling types can be used to deconstruct average water use into calibrated water use for different households.

Demographic, socioeconomic, climate and water use information was successfully transformed into surfaces of local data, and household and non-residential water balances. This structure was then used in a systems framework to reproduce the patterns and magnitudes of rainfall and water use throughout the regions. The bottom up structure of this method will enable more robust investigations of regional responses of local interventions including Water Sensitive Urban Design.

6 CONCLUSION

The Systems Framework processes utilises local rainfall and household water balances processes to satisfactorily reproduce observed daily, monthly, annual statistics of rainfall and water use at multiple sites. This demonstrates that the method is able to capture the inter-annual persistence and spatial variability of rainfall and water use that exists throughout the Greater Melbourne and Sydney regions. This demand method was able to adequately estimate regional water demand including the day to day variation and strong seasonal trends of water demand for the regions. This bottom up continuous simulation method provides an opportunity to understand the benefits of local solutions such as Water Sensitive Urban Design on regional systems.

REFERENCES


