An R Package for Open, Reproducible Analysis of Urban Water Systems, With Application to Chicago

Laura E Erban¹, Stephen B Balogh¹, Daniel E Campbell¹, Henry A Walker¹

¹USEPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, Narragansett, RI 02882

ABSTRACT

Urban water systems consist of natural and engineered flows of water interacting in complex ways. System complexity can be understood via mass conservative models that account for the interrelationships among all major flows and storages. We have developed a generic urban water system model in the R package CityWaterBalance. CityWaterBalance provides a reproducible workflow for studying urban water systems by facilitating automated retrievals of open data and post-processing with open source R functions. It allows the user to 1) rapidly assemble a quantitative, comprehensive assessment of flows thorough an urban area, and 2) easily change the spatial and temporal boundaries of analysis. We use CityWaterBalance to evaluate the water system in the Chicago metropolitan area on a monthly basis for water years 2001-2010. Results are used to consider 1) impacts of management decisions aimed at reducing stormwater and combined sewer overflows and 2) the significance of future changes in precipitation.

Keywords
urban systems, urban water, reproducible research, R, Chicago

1. Introduction

Water flows through tightly coupled engineered and environmental systems in cities. In Chicago waterways, up to 70% of flow at a given time is wastewater [Duncker and Johnson, 2015]. This wastewater constitutes 25-65% of cooling water used by downstream thermoelectric facilities, leading to significant de facto reuse [Barker and Stillwell, 2016]. Downstream municipalities also reuse it, though most in the Chicago area are permitted to take water from Lake Michigan. Permittees seek to achieve distribution losses of under 8% [CMAP/CNT, 2014]. Losses leak to groundwater, which in turn infiltrates sewer lines [Abrams et al., 2015; Mills et al., 2014]. Most municipal users are connected to a combined sewer system that also receives stormwater and periodically overflows, sending an average 58 billion gallons of untreated sewage per year [MWRD, 2016] to Chicago’s waterways.

The complexity of an urban water system like Chicago’s remains difficult to evaluate. A variety of hydraulic and hydrologic models, too numerous to discuss exhaustively, have been developed for particular applications. Urban runoff, sewerage and drainage are commonly studied using US Environmental Protection Agency’s (EPA) Stormwater Management Model (SWMM; Rossmann and Huber, 2015]. SWMM is in the public domain, but is often run with commercial software packages (e.g., PCSWMM, MIKE URBAN), which can include additional functional-
ity for urban water supply planning and management analyses. Usage of these tools requires an upfront investment of time to learn for a user with a specialized technical background, as well as significant time to apply to a given study area. Some of these obstacles are addressed in tools like Colorado State University’s Integrated Urban Water Management (IUWM, see: www.erams.com/iuwm) model, which focuses on urban water demand forecasting and conservation strategies. IUWM can acquire and process a selection of data streams in the cloud. This approach works well when the desired inputs and outputs for a study are aligned with IUWM. For any given city, however, the best data sources and water-related concerns vary enough to warrant a more flexible analysis framework.

We have developed an adaptable workflow for analyzing the urban water system as the package CityWaterBalance for R. R is an open source programming language and software environment with many active users in the statistical, scientific, and broader analytical communities. R packages, which include bundled functions and documentation specific to a task, are constantly being developed and updated, making the inclusion of new methods in a user’s analysis both straightforward and transparent. CityWaterBalance, which may be installed from CRAN or https://github.com/USEPA/CityWaterBalance (development version), builds upon existing R packages and web services to automate much of the process of quantifying urban water flows. These data are used as input to a simple mass balance model of the networked flows and storages in the urban water system, which is the basis of CityWaterBalance. The model tracks the major pathways for urban water and estimates values for components of the system lacking data. It is readily understood by a user without a background in hydrology. Inputs can be quickly assembled or updated for a particular application. Outputs of the model provide a comprehensive, quantitative and reproducible assessment of the urban water system.

In this paper, we introduce CityWaterBalance and discuss its application to the Chicago metro area. This area, defined for our study by the seven counties in the Chicago Metropolitan Agency for Planning (CMAP) region, is home to more than 8 million people. Water supply for the region is constrained by a Supreme Court Decree limiting Lake Michigan withdrawals (Wisconsin v. Illinois 449 US 48), contamination of rivers and shallow groundwater, and historic overdraft of deep groundwater [Abrams et al., 2015]. Overabundance of water in the form of flooding and combined sewer overflows (CSO) are also chronic issues, particularly in Chicago [Changnon and Westcott, 2002; Villarini et al., 2013a]. We evaluated changes in the major flows and storages of water in the CMAP region on a monthly basis for water years 2001-2010. This period was chosen because it has the highest abundance of single-source data sets whose retrieval can be automated. Automated data retrieval and handling, to the greatest extent practicable, were design criteria for this work, in order to test the current limitations. There is no reason, however, that CityWaterBalance cannot be applied using datasets from earlier periods or with greater temporal resolution that are otherwise available to the user. Similarly, though the current version of CityWaterBalance leverages US-based web services, our framework is flexible enough to apply to cities outside of the US.

2.0 Methods

The CityWaterBalance package is a set of functions for retrieving and preprocessing data, passing that data to a model of the urban water system (i.e., mass conservative network of flows and storages), assessing the data and parameter uncertainty, and visualizing results. We do not go into detail on the execution of these functions here, as they are best explored using the package vignette. The vignette comes installed with the package or can be viewed on Github (see: https://github.com/USEPA/CityWaterBalance). In this paper, we focus on our process of building the generic urban water system model and evaluating it for the Chicago metropolitan area, along with the lessons learned.

2.1 Model development

Our model of the generic urban water system is shown in Figure 1. The diagram was drawn according to the Energy Systems Language (or ESL; Odum, [1994]). In addition to providing a conceptual overview of the system, the diagram specifies a system of 1st-order differential equations, i.e., a mathematical solution. The equations represent the changes in storage for all state variables as a balance of inflows and outflows. They may be solved for steady state or transient conditions, according to the availability of input data and the desired model output.

According to the standard ESL notation, a large rectangular box marks the physical boundaries of the system. To the left and right of this box are circles showing
the sources and sinks, respectively, of water. Sources include precipitation (PRCP), streamflow (INFLOW) and imports (IMPORTS) from outside the urban boundary. Sinks include evapotranspiration (ET) and streamflow out (OUTFLOW) of the urban boundary. Within the box are symbols labeling the state variables. State variables in our model are categorized as storages, “producers” and “consumers”. Terms in quotations deviate from their ecological definitions as described presently. Storages include surface water bodies (SW), shallow groundwater (SGW, i.e., unconfined aquifers, including the unsaturated zone), deep groundwater (DGW, i.e., confined aquifers), and the combined sewer system (CSS). Producers include water treatment plants (PUR), which produce potable water, wastewater treatment plants (WTP) and industrial facilities (IND), which produce non-potable effluent discharged to surface water bodies, including thermoelectric cooling water discharges. Consumers, or end-users, are grouped according to their use of potable (POT) or non-potable (NPOT) water. Flows between sources, storages, producers, consumers and sinks, are represented using arrows, with the direction of flow indicated by the arrowhead. Each flow is labeled, and the corresponding processes are given in Table. 1.

The equations specified by our ESD (Figure 1.) are coded in the CityWaterBalance package for R, in a function of the same name, and may be solved for any

Figure 1. Energy systems diagram for an urban water system. Sources and sinks of water are the global flows. Internal flows between system components are labeled and described in Table 1. Flows indicated in gray (lateral groundwater flows lacking direct data) are not explicitly modeled, but considered through scenario analysis. Arrows indicate the likely net flow direction, but in some cases, e.g., sewer infiltration (#12), the actual direction may depend on spatial and temporal conditions.
Table 1. Flows in the Urban System Model and Method of Evaluation by CityWaterBalance.

<table>
<thead>
<tr>
<th>Flow #</th>
<th>Process(^a)</th>
<th>Evaluation(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interception</td>
<td>PET(^c) x interc</td>
</tr>
<tr>
<td>2</td>
<td>Runoff</td>
<td>RUNOFF x (1-run_css)</td>
</tr>
<tr>
<td>3</td>
<td>Runoff to sewers</td>
<td>RUNOFF x run_css</td>
</tr>
<tr>
<td>4</td>
<td>Infiltration</td>
<td>PRCP(^c) - flow1 - flow2 - flow3</td>
</tr>
<tr>
<td>5</td>
<td>River inflow</td>
<td>INFLOW(^c)</td>
</tr>
<tr>
<td>6</td>
<td>Imports to surface water</td>
<td>ETC_IMPORTS</td>
</tr>
<tr>
<td>7</td>
<td>Imports for potable use</td>
<td>POT_IMPORTS</td>
</tr>
<tr>
<td>8</td>
<td>Surface water evaporation</td>
<td>PET(^c) x open_water</td>
</tr>
<tr>
<td>9</td>
<td>Industrial withdrawals (SW)</td>
<td>SW_IND(^c)</td>
</tr>
<tr>
<td>10</td>
<td>Potable withdrawals (SW)</td>
<td>SW_POT(^c)</td>
</tr>
<tr>
<td>11</td>
<td>Non-potable withdrawals (SW)</td>
<td>SW_NPOT(^c)</td>
</tr>
<tr>
<td>12</td>
<td>Sewer infiltration</td>
<td>flow30 x leak_css</td>
</tr>
<tr>
<td>13</td>
<td>Baseflow</td>
<td>BASEFLOW</td>
</tr>
<tr>
<td>14</td>
<td>Industrial withdrawals (SGW)</td>
<td>GW_IND(^c) x (1 - dgw)</td>
</tr>
<tr>
<td>15</td>
<td>Potable withdrawals (SGW)</td>
<td>GW_POT(^c) x (1 - dgw)</td>
</tr>
<tr>
<td>16</td>
<td>Evapotranspiration</td>
<td>ET(^c)</td>
</tr>
<tr>
<td>17</td>
<td>Non-potable withdrawals (SGW)</td>
<td>GW_NPOT(^c) x (1 - dgw)</td>
</tr>
<tr>
<td>18</td>
<td>Deep groundwater recharge</td>
<td>DGR</td>
</tr>
<tr>
<td>19</td>
<td>Industrial withdrawals (DGW)</td>
<td>GW_IND(^c) x dgw</td>
</tr>
<tr>
<td>20</td>
<td>Potable withdrawals (DGW)</td>
<td>GW_POT(^c) x dgw</td>
</tr>
<tr>
<td>21</td>
<td>Non-potable withdrawals (DGW)</td>
<td>GW_NPOT(^c) x dgw</td>
</tr>
<tr>
<td>22</td>
<td>Evaporation of industrial water</td>
<td>(flow9 + flow14 + flow19) x ind_evap</td>
</tr>
<tr>
<td>23</td>
<td>Discharge of industrial water</td>
<td>flow9 + flow14 - flow19 - flow22</td>
</tr>
<tr>
<td>24</td>
<td>Conveyance of potable water</td>
<td>(flow10 + flow11 + flow20) x (1 - nonrev)</td>
</tr>
<tr>
<td>25</td>
<td>Leakage of potable water</td>
<td>(flow10 + flow11 + flow20) x nonrev</td>
</tr>
<tr>
<td>26</td>
<td>Evaporation of potable water</td>
<td>flow24 x pot_atm</td>
</tr>
<tr>
<td>27</td>
<td>Wastewater generation</td>
<td>flow24 x wast_gen</td>
</tr>
<tr>
<td>28</td>
<td>Infiltration of potable water</td>
<td>flow24 x (1 - pot_atm - wast_gen)</td>
</tr>
<tr>
<td>29</td>
<td>Evaporation of non-potable water</td>
<td>(flow11 + flow16 + flow21) x (1 - npot_infilt)</td>
</tr>
<tr>
<td>30</td>
<td>Conveyance of wastewater</td>
<td>flow32 x (1 - slud_evap)</td>
</tr>
<tr>
<td>31</td>
<td>Evaporation of sludge</td>
<td>flow30 - flow32</td>
</tr>
<tr>
<td>32</td>
<td>Wastewater discharge</td>
<td>WTPE</td>
</tr>
<tr>
<td>33</td>
<td>Infiltration of non-potable water</td>
<td>(flow11 + flow16 + flow21) x npot_infilt</td>
</tr>
<tr>
<td>34</td>
<td>Combined sewer overflows</td>
<td>CSO</td>
</tr>
<tr>
<td>35</td>
<td>River outflow</td>
<td>OUTFLOW(^c)</td>
</tr>
</tbody>
</table>

\(^a\) Acronyms correspond to labeled storage terms in Fig. 1

\(^b\) All-caps variable names are data inputs to CityWaterBalance. Most are named after the corresponding process. Other data inputs include: potential evapotranspiration (PET) and precipitation (PRCP). Italicized variable names are parameter inputs. See Tab. 3 for parameter details.

\(^c\) Indicates input data can be acquired via web services using CityWaterBalance functions
urban system. The global water balance, or change in total system storage volume \( \frac{dS}{dt} \) is given by the difference between sources and sinks:

\[
\frac{dS}{dt} = PRCP + INFLOW + IMPORTS - ET - OUTFLOW \quad \text{Eq. 1}
\]

where right hand side terms are precipitation (PRCP), streamflow in (INFLOW), imports (IMPORTS), streamflow out (OUTFLOW) and evapotranspiration (ET).

Similarly, the change in storage of each state variable is written according to the numbered flows. For example, the equation for the change in combined sewer system storage \( \frac{dCSS}{dt} \) is:

\[
\frac{dCSS}{dt} = R_{sewer} + I_{sewer} + W_{generation} - W_{treatment} - CSO \quad \text{Eq. 2}
\]

where \( R_{sewer} \) is runoff to sewers (flow #3), \( I_{sewer} \) is sewer infiltration (#12), \( W_{generation} \) is wastewater generation (#27) and \( CSO \) is combined sewer overflows (#34).

After establishing this system of differential equations, unknowns include the 1) magnitude and temporal variability of flows and 2) change in the storage of state variables. The solution is highly underdetermined, but also highly constrained by the relationships among flows and the availability of data (at least in the US) with which to estimate them.

### 2.2 Model solution

The CityWaterBalance model can be applied to a particular study area in a sequence of four steps. First, the user assembles times series of the requisite input data with or without the use of package functions that can retrieve them from web services (see Table. 1). Second, the user determines minimum and maximum values for parameters that are used to estimate unmeasured flows (see Tables. 2,3). Defaults are provided, though location-specific literature may help to refine parameter values. Third, the user determines acceptable value

---

**Figure 2** a. Map of study area, waterways, and watershed boundaries. The study area is defined by the seven counties in the Chicago Metropolitan Agency for Planning (CMAP) region. The stream gauges used for this analysis are indicated, and rivers discussed in the text are labeled. The hashed gray area drains to Lake Michigan. The rest of the mapped area is part of the Upper Mississippi watershed. b. Land covers in the CMAP region [Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N.D. Herod, J. D. Wickham, and K. Megown 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. Photogram Eng and Rem Sens, 81, 345-354.]
ranges for each storage term (i.e., the solution criteria). Solution criteria may be left at default values or determined from ancillary observations (e.g., groundwater levels). Finally, the user runs a solver function to search for solutions. The solver generates parameter sets from the ranges provided using Latin hypercubes [McKay et al., 1979] built with R package tgp [Gramacy, 2007]. Latin hypercubes are commonly used to efficiently sample the full parameter space of a model and assess uncertainty [Thiele et al., 2014]. The solver runs CityWaterBalance with different parameter combinations until the set of solutions is stable with respect to mean flow values.

2.3 Model application to Chicago

Having introduced the development of CityWaterBalance for studying urban water systems generally, we now describe its application to the Chicago metro area. This area could be defined by one of several sets of boundaries. We used the seven Illinois counties in the Chicago Metropolitan Agency for Planning (CMAP) region. CMAP is the official regional planning organization for northeast Illinois. Analysis at the CMAP scale facilitates 1) use of USGS water use data, which are available through web services at the county level and 2) potential integration of our results with planning activities. The CMAP boundary is shown superimposed on watershed, county and state boundaries in Figure 2. Most of the CMAP region drains to the Upper Mississippi watershed, with only ~2% now draining to Lake Michigan after considerable engineering to redirect flow. We limited the analysis period to 2000–2010 based on the availability and overlap of single-source continuous datasets (i.e., only one source of data was needed to create a complete time-series for a given variable), considering complete water years. Since a water year begins in October of the prior calendar year, we used data for the period 10/1/2000–9/30/2010, or 10 water years (WY 2001–2010). During this period, precipitation in the CMAP region averaged 962 mm/yr.

2.3.1 Data on measured flows

After establishing boundaries, the next step in an urban water systems analysis using CityWaterBalance is to determine the external sources and sinks of water. Sources for the CMAP region include precipitation, streamflow of the Fox, Des Plaines, Little Calumet and Kankakee Rivers, lateral groundwater flow, and imports from Lake Michigan. Lake Michigan water is imported for 1) public water supply 2) lockage and leakage throughout lakefront control structures, and 3) dilution of effluent and navigation make-up flow in the Chicago Area Waterway System (CAWS). Sinks of water for the region include evapotranspiration (ET), lateral groundwater flow, Lake Michigan, and streamflow of the Kishwaukee, Fox, Du Page and Des Plaines and Kankakee rivers. We exclude the Kankakee from our analysis, as the available gauge data is insufficient and the river makes only a short traverse of our study area (i.e., we assume inflow = outflow).

For most of these flows, data retrieval can be largely automated, for US urban areas, using CityWaterBalance and dependencies. CityWaterBalance is written to access two sources of gridded atmospheric data that are hosted by the USGS Geo Data Portal (GDP). It extracts monthly, areal-averaged time series using the GDP’s Area Grid Statistics (weighted) algorithm and R package geoknife [Read et al., 2016]. Area boundaries can be manually uploaded to the GDP as a shapefile, as we did for the CMAP region, or GDP-hosted geometries may be used. CityWaterBalance retrieves precipitation, minimum and maximum temperatures from PRISM model output [Daly et al., 1994] and actual evapotranspiration of vegetated areas from the SSEBop model [Senay et al., 2013]. Potential evapotranspiration (PET) is then estimated using the PRISM temperatures and the Priestley-Taylor method as in R package EcoHydRology [Fuka et al., 2014]. CityWaterBalance can also retrieve daily average streamflow data from the USGS National Water Information System (NWIS) using R package dataRetrieval [Hirsch and De Cicco, 2015]. Stream gauges may be selected using NWIS mapper (at https://maps.waterdata.usgs.gov/mapper/index.html). To estimate total river inflow to the CMAP area, we used data at three gauges (#s 05545750, 05527800, 05536195). For outflow, we used five gauges to estimate flow at four locations (#s 05438500, 0551540, 0552500, 05540500, 05537980), averaging data from two Fox River gauges to estimate flow at the CMAP boundary, located approximately midway between them (see Fig. 2). Ungauged streams drain ~1.4% of the CMAP area to Lake Michigan (see hatched area of Fig. 2). To account for this run-off, we supplemented total river outflow by 1.4% of precipitation (equivalent to 2.4% of gauged outflow).

Data for the remaining source and sink terms cannot be retrieved in this way. There is also no direct source of data for estimating lateral groundwater flow, and we did not explicitly evaluate its components.
(i.e., source and sink to shallow and deep aquifers, see Fig. 1). Instead, we used hydraulic head time series and interpolated maps from the Illinois State Water Survey (ISWS) [Abrams et al., 2015; ISWS, 2015] to make informed assessments, discussed further below, about the significance of these flows. We extracted data on Lake Michigan imports from the Lake Michigan Diversion Accounting tables [USACE, 2016], which are available online but not accessible via a web service.

Internal modeled flows are represented in large part by different water uses. Water use data are compiled by the USGS under the National Water-Use Information Program and served by NWIS at the county-level at 5-yr reporting intervals (average withdrawal rate during the reporting year). They are categorized by purpose of use, water quality (fresh or saline) and water source (surface or groundwater). Using R packages dataRetrieval and CityWaterBalance, we retrieved freshwater withdrawals for years 1995, 2000, 2005, and 2010 for all seven CMAP counties. We aggregated these withdrawals according to nine of our modeled flows (Fig. 1, Tab. 1). The nine flows represent withdrawals from three distinct storages of water (surface water, shallow and deep groundwater) and three categories of use: industrial (mainly thermoelectric cooling, flow #s 9, 14, 19), potable (#s 10, 15, 20) and non-potable (#s 11, 17, 21). To summarize potable withdrawals, we included NWIS data categories in total effluent discharged to surface water bodies. MWRD also compiles data on CSO volumes, for quarterly reporting (unpublished) to the Illinois Environmental Protection Agency (IEPA). We did not consider overflows outside of the MWRD service area. The service area includes Chicago and 128 suburban communities.

2.3.2 Estimation of unmeasured flows

The data sources just discussed cover many of the required inputs for CityWaterBalance (see Table 1). Other inputs and parameter values (see Table 2) for estimating the magnitudes of unmeasured flows should be determined from location specific information. In a large area of northeast Illinois, of which CMAP is a sub-region, a water balance study estimated runoff and baseflow as a fraction of monthly precipitation [Yeh and Famiglietti, 2009]. We used these estimates as a starting point for calibrating our model (details below). We assumed a constant rate of deep groundwater recharge (1.1 mm/yr), based on analysis of the deep sandstone aquifer system and overlying Maquoketa shale [Walton, 1965]. We determined the areas of impervious surface and open water from the National Land Cover Dataset for years 2001 and 2011 [Homer et al., 2007; Homer et al., 2015]. We used the area of impervious surface to validate the ET data from SSEBop (see Results) and the area of open water to estimate surface water evaporation (see Table 1). Ideally these coverages would also be
quantified through an automated procedure, and future versions of CityWaterBalance may have such functionality. It could be done in an open source software like QGIS, but for this work we used ArcGIS (v 10.3.1).

### 2.3.3 Additional assumptions

To evaluate the CityWaterBalance model for the CMAP region, we made the following additional assumptions. First, we assumed no changes in storage for state variables that are classified as producers or consumers (see Figure 1). That is to say, there is no accumulation of water at producer facilities and all water withdrawals are used by consumers on the

Table 2. CityWaterBalance Parameter Values for the CMAP Case.

<table>
<thead>
<tr>
<th>#</th>
<th>Parameters</th>
<th>Description</th>
<th>Values [-]</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>1</td>
<td>interc</td>
<td>Fraction of PET lost to interception</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>et_mult</td>
<td>Multiplier for ET</td>
<td>1</td>
<td>1.14</td>
</tr>
<tr>
<td>3</td>
<td>flow_mult</td>
<td>Multiplier for outflow</td>
<td>1.02</td>
<td>1.12</td>
</tr>
<tr>
<td>4</td>
<td>open_wat</td>
<td>Area of open water</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>run_mult</td>
<td>Multiplier for runoff</td>
<td>1</td>
<td>4.5</td>
</tr>
<tr>
<td>6</td>
<td>run_css</td>
<td>Fraction of runoff diverted to sewers</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>bf_mult</td>
<td>Multiplier for baseflow</td>
<td>0.5</td>
<td>1.25</td>
</tr>
<tr>
<td>8</td>
<td>nonrev</td>
<td>Fraction of treated water lost to leaks (non-revenue)</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td>pow_evap</td>
<td>Fraction of cooling water that evaporates</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>wast_gen</td>
<td>Fraction of potable use that returns to sewers</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>11</td>
<td>pot_atm</td>
<td>Fraction of potable use that evaporates</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td>12</td>
<td>npot_infilt</td>
<td>Fraction on nonpotable use that infiltrates</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>13</td>
<td>slud_evap</td>
<td>Fraction of wastewater that evaporates from sludge</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>leak_css</td>
<td>Fraction of wastewater effluent from gw infiltration</td>
<td>0.05</td>
<td>0.24</td>
</tr>
<tr>
<td>15</td>
<td>dgw</td>
<td>Fraction of gw from deep, confined aquifers</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>16</td>
<td>dgw_rep</td>
<td>Multiplier for deep gw pumping replacement</td>
<td>0.08</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. CityWaterBalance Parameter Values for the CMAP Case.

Where no reference is given, reasonable values were chosen based on the data.

* indicates min/max values are 50/150% of reference value
monthly timescale considered here. Changes in storage only occur for state variables classified as storages. We assumed lateral flows of shallow groundwater into and out of our system are much less significant than vertical flows. For deep confined aquifers, largely disconnected from vertical flows by thick and extensive shale, we used a time series of hydraulic head maps to assess the degree to which lateral flow may be compensating for past and present groundwater pumping.

3.0 Results

3.1 Water balance for the Chicago region (CMAP)

Of the major flows of water in the CMAP region, precipitation is the largest source and streamflow is the largest sink for water, followed closely by evapotranspiration (ET). ET peaks in the summer months when much of the year’s precipitation is also received. Streamflow into the region is considerably less than water imports from Lake Michigan. Imports from the lake are the largest source of water for potable uses in Chicago and its sub-

Figure 3. Monthly time-series of water flows in the Chicago region. In our analysis, the Chicago region is defined by the seven CMAP counties. Only flows that are directly measured (i.e., not modeled) are shown in these plots. a. Global flows (i.e., sources and sinks). b. Internal flows. Dashed lines indicate monthly values have been interpolated from data reported (by USGS) at 5-yr intervals.
urbs, though rivers and groundwater increase in importance for municipalities towards the west and southern boundaries of our study area [Abrams et al., 2015].

The global water balance (Eq. 1) for the CMAP region can take on a limited range of values within the uncertainty of these data. Using our initial estimates of sources and sinks (see Section 2.3.1 for details), we found a global imbalance of 45 mm/yr, on average over the 10-yr period. This potential surplus of water is equivalent to 4.7% of precipitation or 30% of Lake Michigan imports. Surface water bodies could not have gained this much water; the depth of which would reach more than 11m even if spread over twice the area of open water (~2% of CMAP region). The actual area of open water remained near-constant in NLCD land use classifications over the 2001-2011 period. The global imbalance can therefore only be accounted for by 1) an increase in storage elsewhere within the study area or 2) error or misspecification of sources or sinks.

The largest storage within the CMAP region is the groundwater system. Shallow monitoring wells [ISWS, 2015], indicate a modest rise in water level among wells (n = 3). This rise, estimated by the Sen slope [Komsta, 2013; Sen, 1968], averages 52 mm/yr (range: 30-79 mm/yr). Only ~25% of the net increase can be attributed to water filling void spaces in the unconsolidated glacial deposits that make up the surficial aquifer, or ~13 mm/yr, averaged over the full study period. However, these are noisy data, with multi-year periods of net rising and falling water levels that do not correspond particularly well to the global water imbalance. At the same time, withdrawals from deep aquifers also averaged ~13 mm/yr [Meyer et al., 2012; USGS, 2016], and may be underreported. Water levels in the deep aquifers indicate areas of both rebound and overdraft, with some desaturation occurring in the southwestern part of our study area [Abrams et al., 2015]. These observations suggest that total ground-
water storage (shallow and deep aquifers) did not increase by an amount comparable to the global surplus, and it may even have decreased over this period.

Rather, the ancillary data suggest an error in the values of sources or sink terms. To validate precipitation, we compare gauge data (daily mean of 9 gauges in the study area) from the National Climate Data Center (NCDC) against the PRISM-based estimates. The two sources track each other well, agreeing to within 2%, cumulatively, with PRISM estimates lower than those from NCDC. This difference is not trivial, since precipitation is the largest component of the water budget. In fact, it is larger than all reported non-potable water withdrawals in the study area. But it is unlikely that the lower-valued PRISM estimates of precipitation are overestimating this source of water to our system. To validate ET estimates from SSEBop, we use two independent studies of ET measured in northeast Illinois, which conclude that ET losses amount to 70-75% of annual precipitation [Jones, 1966; Yeh and Famiglietti, 2009]. If we assume a reduction in ET proportionate to the coverage of impervious surface (21%, per NLCD), this range could be reduced to 55-59% of precipitation. However, SSEBop indicates ET losses amounting to only 51-52% of annual precipitation (PRISM and NCDC-based, respectively). Thus actual ET from non-impervious cover in our study area may be 7-16% higher than indicated by SSEBop. Evaporation of intercepted precipitation, neglected in our initial set-up, is another sink. We do not have a source for regional-scale estimates of interception, but if it alone is responsible for the global surplus, interception could be as high as 4.7% of precipitation. For streamflow, there is no alternative to the USGS stream gauge network for comprehensive data. Streamflow is both a source and sink, and is subject to errors related to measurement, misalignment of gauges with our study area boundaries (see Figure 2), and omissions for ungauged streams. Outflow is likely underestimated, as most of the gauges with sufficient data to quantify streamflow out of the study area are located upstream of the boundary, and tributaries to the Kankakee flow within it are not included (not gauged). If errors in streamflow alone are responsible for the lack of global water balance closure, net gauged outflow must be increased by close to 12%.

3.2 Solutions for the complete water system

Uncertainty in the flow data, parameter values, and ancillary data on changes in storages makes for a range of possible solutions. We searched for solutions for the CMAP case using the parameter ranges
and calibration targets given in Table 2 until additional solutions changed the mean flow estimate differed by less than 1 mm (cumulative over model run) for any flow. The search performed 253,400 runs of CityWaterBalance, finding 196 solutions. These solutions are summarized in boxplots in Figure 4. Despite the various uncertainties, the modeled flow values are reasonably well separated and constrained.

In the discussion that follows, the mean flow estimate of all model runs is given as a percent of precipitation. Precipitation was not varied among runs. Of the precipitation falling on this region, most infiltrates the subsurface (80%), where it is either used by vegetation (54%), or recharges shallow groundwater (26%). Most recharge is later discharged as baseflow (24%) to rivers and streams. A small amount of recharge is diverted by pumping wells (1.4%). Precipitation that does not infiltrate is lost to interception (1.8%) or diverted to runoff (18%).

Of the flows of water appropriated for human use within the region, withdrawals for industrial processes, primarily thermoelectric cooling, are the largest (42%). Most cooling water is withdrawn from and discharged to surface waters—a very small amount evaporates (<1%). Wastewater discharges are the second largest manmade flow (22%), followed by withdrawals for potable uses (16%). These withdrawals are mainly taken from Lake Michigan (12.4%), with the remainder from inland surface water bodies (1.6%) and groundwater (1.1%). Most used potable water is discharged to the sewer system as wastewater, which also contains stormwater runoff (9.6%). Stormwater diverted to the sewer system in excess of its capacity causes CSO events (2.7%). Figure 5 uses these results to show the general urban water system model (Figure 1) evaluated for the CMAP region.

**4.0 Discussion**

The CMAP case study demonstrates how CityWaterBalance can be used to study an urban water system in a unified way. All major water flows through CMAP are represented in a single quantified portrait (Fig. 5), an assessment not otherwise available despite an extensive literature on water in this region. With the network of water flows established and evaluated, management alternatives are readily compared. For example, stormwater management and mitigation of CSO events is a top priority for the city of Chicago and the MWRD, leading to the installation of a mix of grey and green infrastructures. As of 2014, Chicago had 5.5 million sq ft of green roof and 330,000 sq ft of permeable pavement, with a combined capacity to capture ~90 million gallons of stormwater annually [CoC, 2014]. These green infrastructures detain ~0.16% of the annual average CSO volume, according to the CSO data for 2007-2010. Detained water must then be evapotranspired in order to leave the system, or else it infiltrates to groundwater, where it can reinfiltretur sewer pipes. Green infrastructures may be locally important for mitigating floods, contaminants, and urban blight, among other benefits, but they are clearly not a scalable solution for mitigating the volume of stormwater produced by Chicago.

Another perspective on the CSO problem, made visible through our analysis, considers the significance of wastewater in the combined sewer system. Used potable water, most of which is discharged to sewers, exceeds annual CSO volumes by more than a factor of 4. Put another way, a 25% reduction in wastewater discharged to the sewer system would equal the volume of all CSOs. Such a reduction would not eliminate CSOs, which are also a function of timing and conveyance capacity. However, this context underscores the importance of water conservation and reuse efforts that could reduce wastewater flows. These efforts tend to get more attention for other co-benefits, including consumer savings of money and energy, but the case for promoting them is made stronger by a systems analysis that also considers stormwater.

Changes in urban water management at the city scale occur within the context of climate change, which alters the larger precipitation and temperature regimes. Climate models project a complicated suite of changes to seasonal precipitation totals and extreme weather events in the Chicago area. Annual precipitation is expected to increase, mostly in winter and spring, by as much as 20% by mid-century [CCAP, 2008; Hayhoe et al., 2010b], an amount that is approximately double our estimates of stormwater runoff to sewers today. The intensity of rainfall delivered by summer convective storms is already increasing in the larger region, exacerbating urban flooding [Villarini et al., 2013b; Yang et al., 2013]. Additional future precipitation is not likely to be removed by increased evapotranspiration, projected to increase a modest 1% with rising temperatures [CCAP, 2008]. Temperatures in Chicago are expected to rise by 2-3°C, on an average annual basis by mid-century, with a significant increase in the number of hot (>90°F) summer days [Hayhoe et al., 2010a,b]. A significant change in the frequency of hot days will undoubtedly affect water demand in ways beyond the scope of this analysis.
Although much of what we have presented focuses on the largest flows of water, smaller flows are also significant. Groundwater makes up a relatively small portion of potable water withdrawals, but this source is crucial to municipalities that are not permitted to tap Lake Michigan, or in locations where river water quality is poor. Pumping even small amounts can reduce water levels considerably in the confined aquifers this region relies on [Abrams et al., 2015]. Although non-potable withdrawals are also minor in comparison to others, their timing may be especially important. Agriculture in the CMAP region is largely rainfed [USDS/NASS, 2012], but withdrawals from a surface or groundwater source may save crops during a drought. Leakage of water into and out of pipes are individually low flows, but they add up, not only by volume, but also in terms of pumping and treatment costs. These costs divert energy and financial resources from other urban functions.

Our analysis of the CMAP region highlights a number of present limitations. Our approach prioritizes accurate accounting of water flows at the city-scale over precise resolution of related properties. Locations vulnerable to flooding, or inundation depth, for example, are important concerns that must be determined using other models. The functions available in CityWaterBalance can be used to efficiently generate input datasets for such models, as well as to change the spatial and temporal boundaries of analysis. However, many important data sources remain difficult to obtain, even for an urban area of this size in the US. Data related to wastewater, CSO events and sewer coverage were not readily accessible. Data needed to resolve flows at sub-county scales remain a major challenge to assemble, as these are typically held offline by a variety of agencies. As city Data Portals and government web services expand their content, these data streams can be incorporated into the CityWaterBalance workflow. We anticipate that in the near future, the kind of analysis we have done will become easier, more detailed, and more extensible to other urban areas.

5.0 Conclusions

We have built an open and reproducible workflow, CityWaterBalance for R, for quantifying the natural and manmade flows of water in an urban system. CityWaterBalance assembles the complete picture, including the relative magnitudes and temporal variability, of the major urban water flows and the relationships among them. It keeps the whole system in context while evaluating individual components that can be studied in greater depth through more spatially detailed or process-based modeling. Application of the package to analysis of the Chicago metro area highlights the relative significance of systemic factors controlling stormwater and sewer overflows. It also demonstrates that many of the data sources for analyzing any urban area in the US can be automatically acquired using our process. We anticipate that more of the requisite input data will be available through web services in the near future and can be incorporated into this workflow. Our approach can also be expanded to include related sectors of the urban environment that are essential to its sustainability, including but not limited to energy, materials, food, and financial systems.

6.0 Acknowledgments

We gratefully acknowledge Marty Jaffe for helpful conversations on water in Chicago, and Fay Costa for providing reported CSO data not available online. We thank Jeff Hollister for guidance on R package development. We appreciate the comments from internal and external reviewers. This work was funded by the USEPA’s Office of Research and Development under the Sustainable and Healthy Communities research program project SHC 4.61.3: Community Sustainability Assessment and Management.

7.0 Disclaimer

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the EPA. The EPA does not endorse any commercial products, services, or enterprises.

8.0 Software Availability

CityWaterBalance version 0.1.0 for R available (as of 6/16/2017) on CRAN: https://cran.r-project.org/web/packages/CityWaterBalance/index.html

Development version is available at: https://github.com/USEPA/CityWaterBalance

Developer: Laura Erban.
Open source (license: CC0), platform independent, requires R version 3.0.0 or higher.

Data for this paper is included with the package. Size of package archive is 1.7 MB.

9.0 References


CoC. (2014), Green Stormwater Infrastructure Strategy (pp. 48). Chicago, IL: City of Chicago.


USEPA (1977), Tunnel Component of the Tunnel and Reservoir Plan Proposed by the Metropolitan Sanitary District of Greater Chicago, Environmental Impact Statement. United States Environmental Protection Agency Region 5 Water Division, Chicago IL.


