Journal of Hydrology 614 (2022) 128629

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

# How much water is stolen by sewers? Estimating watershed-level inflow and infiltration throughout a metropolitan area

Jeremy E. Diem [a,*], Luke A. Pangle [a], Richard A. Milligan [a], Ellis A. Adams [b]

[a] Department of Geosciences, Georgia State University, Atlanta, GA 30302, United States
[b] Keough School of Global Affairs, University of Notre Dame, Notre Dame, IN 46556, United States

ARTICLE INFO

This manuscript was handled by Nandita Basu, Editor-in-Chief, with the assistance of Claire J. Oswald, Associate Editor.

*Keywords:*
Water budget
I&I
Actual evapotranspiration
Piped flows
Sewer overflow
Southeastern United States

ABSTRACT

Human activities can have substantial impacts on watersheds, yet a major but understudied impact on many urban watersheds is the inflow and infiltration (I&I) of water into sewage infrastructure. I&I is important, because it is a major cause of sewer overflows, increases wastewater treatment plant costs, and reduces base flows of urban streams. Unfortunately, I&I information is deficient at the watershed scale. Therefore, this study uses a water-budget approach to quantify the magnitude of I&I for 90 watersheds in the Atlanta, Georgia USA region during 2013–2020. I&I for each watershed is calculated by subtracting outflows (i.e., stream discharge, water withdrawn by public water systems, and actual evapotranspiration (AET)) from inflows (i.e., precipitation, water-supply pipe leakage, and non-I&I effluent from wastewater treatment plants). Included in the AET estimates is irrigation water from public water systems and water withdrawals for agriculture. Results show that I&I is a major contributor to the total outflow in urban watersheds. The mean annual I&I total for the 15 most urbanized watersheds is 138 mm, constituting 25% of stream discharge. The mean annual I&I total from the watersheds with the five highest totals is 216 mm, which is 40% of stream discharge. These annual I&I totals align well with totals calculated for urban catchments in Europe, where most of the previous research has been conducted. Regression analyses show that the density of older housing, which is a proxy for deteriorating sewage infrastructure, is the most important predictor of I&I across the Atlanta region. Despite the uncertainties in estimating annual totals for all components, especially AET for urban watersheds, of the water budget, we conclude that estimating I&I using the water-budget approach is a useful initial approach to estimating I&I throughout a region.

## 1. Introduction

Human activities can have substantial impacts on water budgets. The budget for reference watersheds (i.e., watersheds with minimal anthropogenic modifications) over a robust number of years (i.e., enough years to minimize the change in water storage) that have negligible net groundwater exchange has three components: precipitation as an inflow, and actual evapotranspiration (AET) and stream discharge as outflows (Healy et al., 2007). Human alterations of the landscape can alter precipitation (Shepherd, 2005), AET (Grimmond and Oke, 1999), and stream discharge (Booth and Jackson, 1997) in urban watersheds. Anthropogenic water inputs to watersheds include inter-basin transfers of water in the form of leaking water-supply pipes, irrigation water, and effluent from wastewater treatment plants (WWTPs) (Townsend-Small et al., 2013). Anthropogenic outputs include the withdrawal of surface water and groundwater (Kampf et al., 2020).

One important anthropogenic outflow that has largely been overlooked in watershed analyses is inflow and infiltration (I&I). Inflow is rainwater and surface water that enters a sewer system from above ground through direct connections (e.g., downspouts, yard drains, sump pumps, manhole covers, etc.) (Cahoon and Hanke, 2017; Pawlowski et al., 2014), while infiltration is groundwater seeping into the sewers, typically through defective pipes, pipe joints, connections, or manholes (Kracht et al., 2007; U.S. EPA, 2014). Infiltration is the dominant process in humid regions (Karpf and Krebs, 2011; Pangle et al., 2022; Staufer et al., 2012; Zhang and Parolari, 2022). Sewer systems deteriorate over time (Malek Mohammadi et al., 2020); therefore, older systems have been shown to have more I&I than newer systems (Prigiobbe and Giulianelli, 2009).

I&I is important because it is a major cause of sewer overflows,

* Corresponding author.
*E-mail address:* jdiem@gsu.edu (J.E. Diem).

https://doi.org/10.1016/j.jhydrol.2022.128629
Received 7 June 2022; Received in revised form 9 September 2022; Accepted 17 September 2022
Available online 29 October 2022

increases WWTP costs, and impacts water budgets. In previous I&I studies (Bareš et al., 2012; Diem et al., 2021; Dirckx et al., 2019; Kracht et al., 2007; Pangle et al., 2022; Prigiobbe and Giulianelli, 2009; Rödel et al., 2017; Shehab and Moselhi, 2005; Weiβ et al., 2002), the percentage of sewage water that is I&I ranges from 14% to 70%. I&I increases during precipitation events (Cahoon and Hanke, 2017) and it can overload sewage systems, thereby causing combined-sewer overflows (CSOs) and sanitary-sewer overflows (SSOs) and subsequent pathogen contamination (Arnone and Perdek Walling, 2007). In the United States for example, there are tens of thousands of SSOs each year, discharging pathogens into waterways (U.S. EPA, 2022). I&I increases pumping costs at the treatment plant and collection system and likely reduces treatment efficiency (Dirckx et al., 2016). Finally, I&I can be a major outflow in urban watersheds (Bhaskar and Welty, 2012), thereby making it a major contributor to reduced baseflows in urban streams (Bhaskar et al., 2016; Diem et al., 2021).

Despite the importance of I&I, especially its environmental impacts, few studies have provided sufficient I&I information to determine the magnitude of I&I with respect to the water budget. Much I&I information pertains to infiltration rates into sewer pipes (e.g., Karpf and Krebs, 2013; Schulz et al., 2005) and the aforementioned percentage of wastewater flow that is I&I. What is known regarding I&I as a water-budget component is shown below for multiple studies, and all studies examined either wastewater flow within individual pipes or effluent from WWTPs. Using variations in pollutant load and continuous water quality and quantity monitoring, Bareš et al. (2009, 2012) quantifies I&I for six catchments in Prague, Czech Republic over a variety of months-long periods during multiple years, and the annual-equivalent I&I total is approximately 204 mm. Staufer et al. (2012) uses hydrograph separation to estimate I&I at two catchments in western Germany over 103 weeks, and the annual mean I&I total is approximately 232 mm. Bhaskar and Welty (2012) uses results from wastewater-flow analyses for a 10-year period to estimate 530 mm as the mean annual I&I for the two heavily urbanized watersheds in Baltimore, MD USA. A similar value is found when solving for I&I using the water-budget approach (Bhaskar and Welty, 2012). Diem et al. (2021) uses monthly WWTP effluent totals over seven years to estimate a mean annual I&I of 37 mm for three suburban watersheds near Atlanta, GA USA. For the three watersheds in Diem et al. (2021) and an additional nearby watershed, Pangle et al (2022) uses precipitation and sewer-flow data for one year to estimate a mean annual I&I of 41 mm.

From the small number of studies that have examined multiple catchments in a region, it is known that older or unrehabilitated sewer systems have increased I&I totals. In Rome, Italy, it has been shown that a catchment with a 14-yr old sewer system has negligible infiltration, while a catchment with a 40-year-old system has 50% infiltration (Prigiobbe and Giulianelli, 2009). A rehabilitated sewer system in a catchment in Germany has 24% less groundwater infiltration than a system in a catchment with similar pipe density and land cover to the rehabilitated catchment (Staufer et al., 2012).

Based on the information in the preceding paragraphs, there is a strong need for an examination of I&I in a region across many catchments/watersheds with varying sewer-infrastructure and land-cover characteristics. Although most previous studies have quantified I&I using measured flows within pipes or entering or leaving WWTPs, these measurements are not widely available and therefore cannot support a broad-scale comparative analysis of many different watersheds. An alternative approach that may be more widely applicable is to infer magnitudes of I&I based on water-budget calculations, where I&I is the residual unknown term in the water-budget equation. While the water-budget approach has disadvantages due to uncertainties in the terms (Kampf et al., 2020), if the magnitude of I&I is as large as previous studies show, then using the water-budget approach to estimate I&I is a worthwhile endeavor.

Our overarching research question is as follows: How do the magnitudes of watershed-level I&I vary across a metropolitan area? The major objectives are to: (1) estimate all inflows and non-I&I outflows for a wide variety of watersheds; (2) determine the degree to which land-cover, population-density, and housing-density variables impact I&I; and (3) explore how uncertainties in water-budget terms impact I&I estimates.

## 2. Material and methods

The methods involved producing estimates for all components of the water budget except for I&I that is exported out of the watershed (i.e., the unknown outflow term), and then estimating and analyzing I&I. Since eight years of data are used in this study, it is assumed that changes in water storage over time are negligible. Performing the analyses with only a few years might violate the assumption (Aulenbach and Peters, 2018). The equation below, which is based on the water-budget equation in Bhaskar and Welty (2012), was used to estimate I&I:

$$I\&I = P + L + Q_E + I_P - W - Q_T - AET_T, \tag{1}$$

where $P$ is precipitation, $L$ is water-supply pipe leakage, $Q_E$ is non-I&I effluent (i.e., treated potable water) discharged by WWTPs, $I_P$ is irrigation water from public water systems, $W$ is water withdrawn by public water systems and transported out of the watershed, $Q_T$ is stream discharge, and $AET_T$ is total actual evapotranspiration. In this paper, $AET_T$ is calculated as follows:

$$AET_T = AET_B + E + AET_I, \tag{2}$$

where $AET_B$ is biome AET, $E$ is evaporation from impervious surfaces, and $AET_I$ is additional AET from irrigation water. $AET_B$ is calculated using models in Fang et al. (2016). $AET_I$ is calculated as follows:

$$AET_I = I_P + I_A \tag{3}$$

where $I_P$, which was presented earlier as an inflow, is irrigation from public water systems, and $I_A$ is irrigation water withdrawn from the watershed for agriculture, which includes golf courses and farms. It was assumed that all irrigation water went towards AET. This assumption is supported by the lack of excessive irrigation by households in the Atlanta region (DeOreo et al., 2016).

### 2.1. Study region

The chosen study region, the Atlanta–Sandy Springs–Gainesville combined statistical area in the southeastern United States, was an ideal locale for using water budgets to examine I&I (Fig. 1). The Atlanta region had 90 suitable gaged watersheds with data for 2013–2020. Of the 90 watersheds used in Diem et al. (2021), 89 were used in this study. Among the 90 watersheds were three reference watersheds (R1, R3, and R4) (Fig. 1). One more reference watershed (R2) was added for the AET component of this study. All watersheds had a humid-subtropical climate (Cfa), which is characterized by hot, humid summers and no seasonal differences in precipitation (Trewartha and Horn, 1980). Typical annual precipitation and potential evapotranspiration totals in the Atlanta region are approximately 1248 mm and 1183 mm, respectively (Aulenbach and Peters, 2018). In addition, all watersheds existed entirely within the Piedmont physiographic province, which is characterized by rolling hills underlain by metamorphic and igneous rocks (Miller, 1990). The Atlanta region is served almost entirely by sanitary sewers (i.e., separate pipes for sewage and stormwater), with combined-sewer pipes (i.e., pipes that transport both sewage and stormwater) existing only in the City of Atlanta and those pipes representing only 4% of the city's sewage-system piping (City of Atlanta, 2011). I&I has been recognized as a serious environmental problem in the region: the water diverted by I&I has contributed to both SSOs and combined sewer overflows in the City of Atlanta and SSOs in DeKalb County that have ultimately led to federal consent decrees aimed at improving sewage

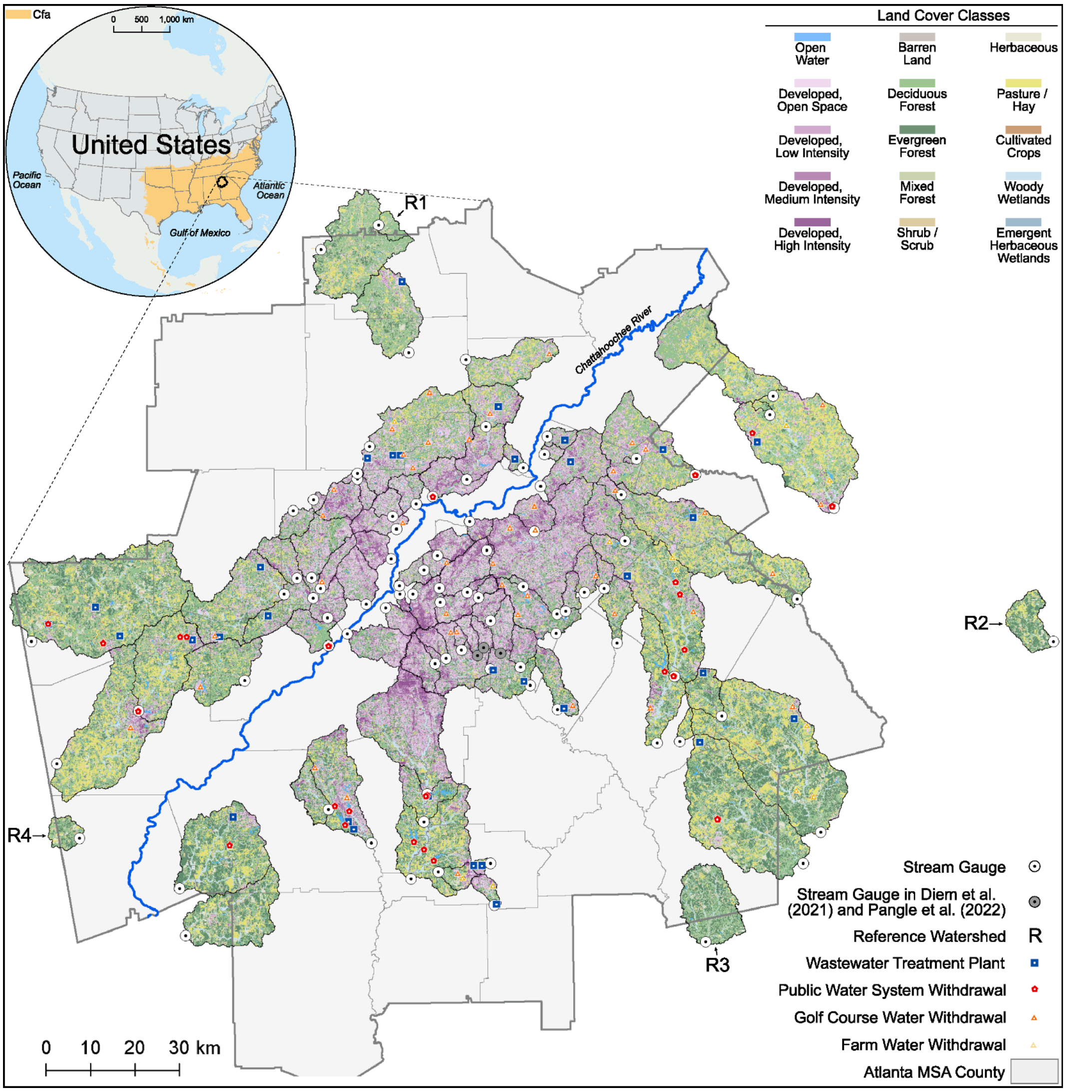


**Fig. 1.** Locations of the 91 gauged watersheds used in this study. All watersheds are within and proximate to the Atlanta metropolitan statistical area (MSA). The four reference watersheds are identified as are the three watersheds examined in both Diem et al. (2021) and Pangle et al. (2022). Shown within the watersheds are land cover and locations of wastewater treatment plants and water withdrawals for public water systems, golf courses, and farms. The inset map shows the location of the Atlanta MSA within the Cfa climate type, which covers most of the southeastern United States.

infrastructure in those municipalities (U.S. District Court, 1998, 1999, 2011). Finally, as is shown in the next section, there was an abundance of data in the Atlanta region for estimating water-budget components.

### 2.2. Land cover

Land cover and associated variables were essential for the estimation of most water-budget terms. High-resolution, gridded land-cover, imperviousness, and tree-canopy data were obtained from the National Land Cover Database of the Multi-Resolution Land Characteristics Consortium. All three datasets had a spatial resolution of 30 m. The land-cover product had 15 classes for the Atlanta region (Fig. 1), the imperviousness product had for each grid cell the percentage of developed surface that was impervious surfaces, and the tree-canopy product had for each grid cell the percentage that was tree canopy. Over the 2013–2020 period of this study, land-cover and imperviousness data existed for 2013, 2016, and 2019, while tree-canopy data only existed for 2016. Land-cover proportions for each watershed (i.e., the proportional coverage of each land-cover class) for 2014, 2015, 2017, 2018, and 2020 were estimated using a weighted interpolation of land-cover proportions from 2013, 2016, and 2019. All three 2016 datasets were combined to determine the percent imperviousness and percent tree canopy for all land-cover classes. Forest was assumed to be the tree canopy for all classes but shrub/scrub and woody wetlands: shrub was the canopy cover for shrub/scrub, and a mixture of forest and shrub was the canopy cover for woody wetlands. The following assumptions were made for the land-cover classes regarding the remaining portion of grid cells that were not impervious or forest: (1) water for open water, woody wetlands, and emergent herbaceous woodlands; (2) grass for the developed classes, barren, grassland, herbaceous, and pasture/hay; and (3) crops for cultivated crops. In addition, mixed forest was assumed to be the forest cover for all land-cover classes that were not deciduous forest or evergreen forest.

### 2.3. Population and housing

Population data were used to estimate values for some of the water-budget terms, and both population and housing data were used in the analyses of estimated I&I totals. Both population and housing data at the block-group level were obtained from the U.S. Census Bureau's American Community Survey 5-year estimates for 2010–2014 and 2015–2019. While only total population was used in this analysis, the following categories of housing totals (i.e., number of units), which denote when a structure was built, were used: ≥2014, 2010–2013, 2000–2009, 1990–1999, 1980–1989, 1970–1979, 1960–1969, 1950–1959, 1940–1949, and ≤ 1939. Population and housing totals were converted to density values, and weighted means (based on area of a block group within a watershed) were used to estimate population and housing densities within the watersheds. The density estimates from the 2010–2014 data were used for 2013 and 2014, while the density estimates from the 2015–2019 data were used for the subsequent years.

### 2.4. Precipitation (P)

Gridded daily precipitation data from 2013 to 2020 were obtained from the PRISM (Parameter-elevation Regressions on Independent Slopes Model) Climate Group. PRISM calculates a climate-elevation regression for each grid cell of a digital elevation model, and precipitation gauges entering the regressions are assigned weights based primarily on the physiographic similarity of the gauge to the grid cell (Daly et al., 2008). Radar data also are used to improve the estimates (Daly et al., 2021). These serially-complete data had a 4-km spatial resolution. Precipitation totals for the watersheds were weighted means of grid-cell values, with the weights derived from the areas of the grid cells within the watersheds. PRISM precipitation totals have been shown to be 3.25% higher than those measured at U.S. climate-monitoring stations (Buban et al., 2020); therefore, all monthly totals for the 90 watersheds were divided by 1.0325.

### 2.5. Pipe leakage (L)

Estimating water inflows via pipe leakage relied on annual water-loss audits from 65 public water systems serving populations within the 90 watersheds. Those audits include estimates of water supplied, population served, and real losses (i.e., pipe leakage) (Georgia Environmental Protection Division, 2022). Water-supply rates for each water system for each of the eight years were calculated by dividing water supplied by the population served. Pipe-leakage rates for each water system for each of the eight years were calculated by dividing real losses by water supplied. Mean pipe-leakage and water-supply rates over the eight-year period were calculated for each system. If multiple water systems existed in a watershed, the water-supply and pipe-leakage rates used for the watershed were a weighted average of all the rates applicable to the watershed. The weighting factor was the area of the water-supplying municipality within each watershed. Water use was calculated by multiplying the per capita water use by the population total, and then the total pipe leakage was water use multiplied by the pipe-leakage rate. Frequency distributions of per-capita water use and leakage rates for the 65 public water systems are shown in Fig. S1.

### 2.6. WWTP effluent ($Q_E$)

Monthly effluent data from 2013 to 2020 were acquired for 32 WWTPs that discharged effluent into one or more of the 90 watersheds. These data were extracted from the U.S. Environmental Protection Agency's ICIS-NPDES Permit Limit and Discharge Monitoring Report Data Sets. From previous studies (Bareš et al., 2012; Diem et al., 2021; Dirckx et al., 2019; Kracht et al., 2007; Pangle et al., 2022; Prigiobbe and Giulianelli, 2009; Rödel et al., 2017; Shehab and Moselhi, 2005; Weiβ et al., 2002), the mean percentage of sewage water that that is not precipitation or groundwater is approximately 30%; therefore, to approximate the portion of WWTP effluent totals that is an inflow of water from public water systems, effluent totals in this study were multiplied by 0.70.

### 2.7. Landscape water from public water systems ($I_P$)

Monthly water-use data from the City of Atlanta and DeKalb County were the main data used to estimate water inflows via irrigation with municipal water (Fig. 2). Those two municipalities are in the center of the study region and intersect with 30 of the 90 watersheds. It was assumed that the intra-annual variability of water use in those municipalities was representative of the intra-annual variability in the remaining watersheds. The City of Atlanta provided monthly water-use data for customers within the city limits. DeKalb County provided daily outflow data from its filtration plant, and those values were multiplied by the proportion of water not lost through pipe leakage (i.e., one minus the leakage rate) to yield estimates of water use by customers. Mean daily water-use values for both the City of Atlanta and DeKalb County were calculated for each month and the mean daily values from the two water systems were summed.

Irrigation water ($I_P$) was calculated by applying outdoor water use to grass in developed areas. The following major assumptions were made: (1) all outdoor water was used for irrigation; and (2) there was negligible groundwater recharge from irrigation water. Using mean daily water-use data per month, the average of the three lowest-use months was assumed to represent indoor use and all differences in water use between the other months and the mean value from the lowest-use months was considered to be outdoor use (Gleick et al., 2003). The daily differences were converted to monthly totals, and the monthly totals were summed to yield an annual irrigation total. The irrigation rate (i.e., amount of water applied annually to grass) was calculated as follows:

$$R_I = 1000^*\left(\frac{T_I}{A_G}\right), \tag{4}$$

where $R_I$ is the irrigation rate in mm yr$^{-1}$, $T_I$ is total irrigation water in m$^3$, and $A_G$ is the area (in m$^2$) of grass in low-intensity, medium-intensity, and high-intensity developed areas in the City of Atlanta and DeKalb County. The annual irrigation rate was then multiplied by the proportion of developed-land grass (i.e., the proportion of watershed that was grass in low-intensity, medium-intensity, and high-intensity developed areas) in each watershed to yield watershed-specific estimates of annual irrigation totals.

### 2.8. Stream discharge ($Q_T$)

Daily mean streamflow from 2013 to 2020 was acquired from the United States Geological Survey (USGS) for 91 gauges. The entire streamflow dataset for the 91 gauges was missing only 0.01% of daily values. Only 15 of the gauges were missing values, and the percentage of missing values for those gauges ranged from 0.03% to 0.21%. Predicted values from linear regression equations were used to replace missing values. The independent variables were daily values from a gauge with a time series that had the strongest correlation with the time series of the gauge having the missing values. The serially complete daily streamflow values were normalized by watershed area to produce discharge values in millimeters.

### 2.9. Public water system withdrawals (W)

Estimates of surface-water withdrawal within the 90 watersheds by public water systems were made using lists of non-agricultural permits along with water-supplied data from the aforementioned annual water audits. There were 25 intake locations within the 90 watersheds (Fig. 1), and these intakes were identified from an analysis of a geospatial database provided by the Southern Environmental Law Center (Southern Environmental Law Center, 2014). Water-withdrawal totals for each intake were estimated by multiplying total water supplied by the water system of the intake by the ratio of the permitted total for the intake to the permitted total for the water system.

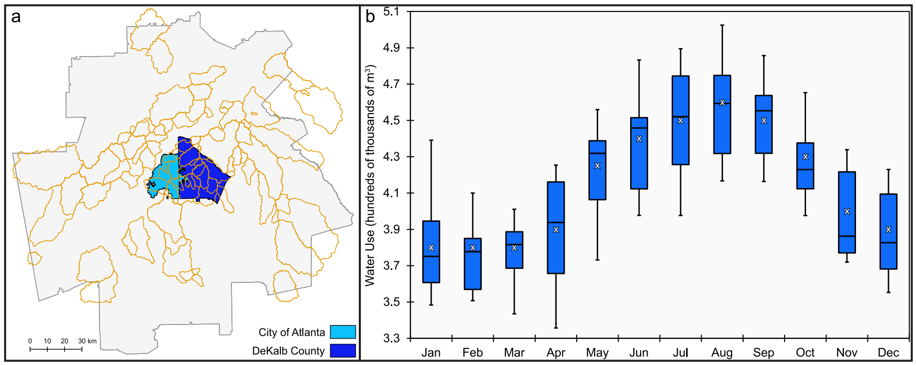


**Fig. 2.** (a) Location of the City of Atlanta and DeKalb County with respect to the 90 watersheds and (b) daily water use per month for both municipalities during 2013–2020. January, February, and March had the three lowest mean daily water-use values.

### 2.10. Agricultural water withdrawals ($I_A$)

Estimates of water withdrawal for agricultural use were made using lists of permits along with irrigation-prediction information from the Environmental Protection Division of the Georgia Department of Natural Resources. Relevant information for each permit included the permit holder, county of the permitted location, and area of land to be irrigated. The locations of all permitted locations that had information for permit holders that were not individuals were determined. Those permits typically had names of golf courses or farms as permit holders. There were at least 44 golf courses and 11 farms with water-withdrawal permits within the 90 watersheds (Fig. 1). The only geographic information for permits that only had individuals as permit holders were counties, and all irrigated land from those permits was summed for each county. Using the Environmental Protections Division's golf course prediction and estimation worksheet (Georgia Environmental Protection Division, 2014) along with monthly reference ET from the TerraClimate dataset (Abatzoglou et al., 2018) and precipitation totals from PRISM, irrigation rates for the April-November period were calculated separately for golf courses and farms. Golf courses were assumed to be 54% rough, 43% fairways, and 3% greens (Golf Course Superintendents Association of America, 2007), which have separate water requirements. Farms were assumed to have vegetative cover similar to golf-course rough. Irrigation rates were multiplied by areas to yield total annual water withdrawals, and those withdrawals were assigned to the watershed within which the permitted golf course or farm was located. For the remaining permits for which only county location was known, total annual water withdrawals were calculated at the county level (i.e., irrigation rate multiplied by irrigated area) and the proportion of each watershed within a county was multiplied by the county total to obtain an additional annual water withdrawal value for that watershed. Therefore, a watershed could have irrigation totals from specific golf courses, specific farms, and estimated totals from farms that were not geo-located.

### 2.11. AET

Mean monthly $AET_B$ totals (i.e., AET that did not include impervious surfaces or account for irrigation) for shrubland, cropland, deciduous forest, evergreen forest, and mixed forest were estimated for the 91 watersheds (i.e., the 90 watersheds and the additional reference watershed) over the 2013–2020 period using regression models in Fang et al. (2016), which uses precipitation, potential evapotranspiration (PET), and leaf area index (LAI) to predict AET totals derived from eddy-covariance measurements (Table S1). Water was not a land cover in that study; therefore, for this study, monthly AET for water was assumed to be equal to PET. Monthly $AET_B$ totals for a watershed were a weighted mean of the land-cover AET values, with the weight for a land cover being the proportion of the watershed that was that specific land cover. The precipitation data used in the AET regression models were the aforementioned PRISM data. Monthly PET was calculated using the Hamon method (Hamon, 1963), which is what Fang et al. (2016) used when developing the regression equations. This PET-estimation method also is recommended for the southeastern United States (Lu et al., 2005). The Hamon equation predicts daily PET and is as follows:

$$PET = k*0.165*216.7*N*\left(\frac{e_s}{T+273.3}\right), \tag{5}$$

where $k$ is the calibration coefficient (unitless), $N$ is the daytime length in multiples of 12 h, $e_s$ is the saturation vapor pressure (hPa), and $T$ is mean monthly temperature (°C). Fang et al. (2016) uses a calibration coefficient of 1.0 (i.e., no calibration); therefore, the same coefficient was used in this study. Monthly temperature and dew-point temperature from PRISM and daily sunlight duration were used to calculate PET. LAI data, which were 8-day composites at 500-m resolution, were extracted from the MODIS MCD15A2H Leaf Area Index product (Myneni et al. 2015) for large (≥1 km$^2$), homogeneous areas of deciduous forest, evergreen forest, shrubland, and combined herbaceous and pasture/hay areas in the Atlanta region. The last group was used for grassland and crops.

Annual $AET_B$ estimates (i.e., the sum of the 12 monthly values) were adjusted based on comparisons between observed and estimated AET totals at the four reference watersheds. The reference watersheds were located in or near the northern, eastern, southern, and western portions of the Atlanta MSA (Fig. 1). Observed AET was calculated using the water-budget approach: discharge was subtracted from precipitation. An adjustment factor was calculated for each reference watershed, and these factors were the ratio of observed $AET_B$ to estimated $AET_B$. These factors were used to adjust $AET_B$ totals at all 90 watersheds, with the factor used at a specific watershed being a weighted mean of the adjustment factors at the four reference watersheds. It was assumed that there was strong positive spatial autocorrelation in $AET_B$; therefore, inverse-distance weighting was used to produce the adjustment factors specific to the 90 watersheds, with the distance being the distance from a watershed centroid to the centroids of the four reference watersheds.

Annual impervious evaporation ($E$) was estimated using evaporation rates found in the literature (Table S2). For this study, it was initially assumed that 20% of precipitation that fell on impervious surfaces was evaporated. This value was the mean evaporation rate from five studies (Cohard et al., 2018; Ragab et al., 2003a; Ragab et al., 2003b; Ramier et al., 2004; Ramier et al., 2011) that measured evaporation rates for paved surfaces, plates, and roofs in temperate oceanic climates. Since the rates had a relatively large range of 10% to 38%, evaporation rates of 10% and 30% also were used to calculate evaporation totals from impervious surfaces.

$AET_T$ was the sum of biome AET ($AET_B$), irrigation AET ($AET_I$), and evaporation from impervious surfaces ($E$). $AET_B$ (i.e., Fang et al. (2016) models) and $E$ are discussed in the preceding paragraphs. As shown in Eq. (3), $AET_I$ was the sum of landscape water from public water systems and agricultural water withdrawn from the watershed. It was assumed that all irrigation water went towards AET (i.e., a negligible amount of irrigation water went towards groundwater recharge). This assumption is verified by the lack of excessive irrigation by households in the Atlanta region (DeOreo et al., 2016). Finally, unless noted otherwise, all $AET_T$ values presented in the paper are based on the 20% evaporation rate from impervious surfaces.

### 2.12. Watershed groups

To better elucidate differences in inflow and outflow terms for various types of watersheds, the 90 watersheds were divided into six groups of 15 watersheds based on the degree of urbanization. An urbanization score was developed for each watershed by using principal component analysis (PCA) to produce a component score from the following multi-collinear variables: percent developed land, percent imperviousness, population density, and housing density. General watershed characteristics (i.e., land cover, population density, and housing density) as well as values of inflow terms, outflow terms, and urbanization variables were determined for each group.

### 2.13. Modeling of I&I

PCA also was used to screen 13 land-cover, population, and housing variables that were potential predictor variables in regression models predicting I&I and $I\&I_{\%Q}$ (I&I as a percentage of stream discharge). The screening was performed to ensure that the regression models did not have multicollinearity (Kachigan, 1991). The land-cover variables were percent imperviousness, percent all developed land, percent low-intensity to high-intensity developed land, and percent medium-intensity and high-intensity developed land. The population and housing variables were population density, total housing-unit density, and density of housing units built prior to the following years: 2000, 1990, 1980, 1970, 1960, 1950, and 1940. Each of the 13 variables was associated with the component on which it loaded the highest.

Multiple-linear regression models were developed using all combinations of variables, with only variables that loaded highly on different components being included in a combination. Jackknife cross-validation was used to produce a Nash-Sutcliffe $E$ value (hereafter shown as $NSE$) for each model. $NSE$ values range from $-\infty$ to 1, with higher values indicating better agreement between observations and estimates (Nash and Sutcliffe, 1970). An $NSE \geq 0.5$ was used as the criterion for satisfactorily accurate model (Moriasi et al., 2007), and only those models were retained.

Residuals from the models were examined to ensure that the models were stable and to better understand controls of I&I. Observed vs predicted scatter plots provided insights into the severity of overpredictions and underpredictions. Spatial-autocorrelation analyses of residuals via Moran's $I$ tests were used to determine if important spatial variables were missing in the regression models.

### 2.14. Analysis of impacts of uncertainties of water-budget terms on I&I totals

All water-budget terms have uncertainties, and those uncertainties contribute to the uncertainty of I&I estimates; therefore, the impacts on I&I totals from uncertainties in the terms were calculated. For each watershed group, each water-budget term was adjusted by 10% and new I&I totals were generated. The absolute difference between the new I&I total and the original I&I total represented the impact of the water-budget term on I&I.

## 3. Results

### 3.1. Composition of land-cover classes

Most land-cover types were comprised of a mixture of forest and grass, while impervious surfaces had substantial coverage in the developed classes (Fig. S2). Forest was the majority vegetative cover in most of the non-developed classes, except for herbaceous and pasture/hay. Among the developed classes, open-space developed land and low-intensity developed land had substantial vegetative cover. Open-space developed land was only 8% impervious surfaces. Low-intensity developed had more grass than forest, compared to open-space developed land, and was 33% impervious surfaces. Impervious surfaces accounted for 63% and 91% of medium-intensity developed land and high-intensity developed land, respectively. For both those classes, grass was more prevalent than forest.

The six groups of watersheds ranged from strongly rural to strongly urban (Fig. 3). The least urbanized watersheds (i.e., Group A) were located on the periphery of the study region. Developed land and impervious surfaces constituted approximately 9% and 2%, respectively, of those watersheds. Population and housing densities also were very low, with mean values of 40 persons $km^{-2}$ and 16 units $km^{-2}$, respectively. The most urbanized watersheds (i.e., Group F) were located in the center of the study region. Developed land and impervious surfaces constituted approximately 85% and 39%, respectively, of those watersheds. Forest cover declined with increasing urbanization: forest was 62% of the least urbanized group and 32% of the most urbanized group. Grass remained relatively constant with increasing urbanization. Finally, there was negligible coverage by shrub, water, and crops (not shown in Fig. 3) of any watershed group.

### 3.2. Precipitation

Precipitation not only was the largest inflow variable but also the largest water-budget variable for all watersheds (Fig. 4a). Annual precipitation totals ranged from 1,257 mm to 1,609 mm; thus, precipitation was immensely larger than all the anthropogenic inflows combined. For example, even for the watershed with the largest total of anthropogenic inflows, the precipitation total was ten times larger. There was little difference among the watershed groups with respect to precipitation totals. The least urbanized group and the most urbanized group had the largest and smallest variability, respectively, due to, as noted earlier, the least urbanized watersheds located farthest from each other on the periphery of the study region and the most urbanized watersheds clustered in the center of the region (Fig. 3).

### 3.3. Anthropogenic inflows ($L$, $I_P$, and $Q_E$)

Pipe leakage ($L$) was reported by all 65 public water systems (Fig. S1), but the magnitude of water added to watersheds via leaking pipes was extremely small compared to precipitation (Fig. 4b). Annual totals of leaking-pipe water ranged from 0 mm to 53 mm. The mean values for the two most urbanized watershed groups were 20 mm and 32 mm, respectively. There was considerable variability for the most urbanized group, since water inflow from pipe leakage depended on leakage rates, per-capita consumption, and population density, and was less strongly linked to imperviousness and developed land.

Similar to pipe leakage, inflows via irrigation from public water ($I_P$) was relatively small compared to precipitation (Fig. 4c). The annual irrigation rate applied to grass in low-intensity, medium-intensity, and high-intensity developed lands was 106 mm. The irrigation rate for lawns in Clayton County, which is in the south-central portion of the Atlanta region, is 94 mm (DeOreo et al. 2016); therefore, the rate of 106 mm is probably appropriate for most of the watersheds. Annual totals of public-water irrigation for the 90 watersheds ranged from 0 mm to 20 mm. There was a steady increase with increased urbanization of the

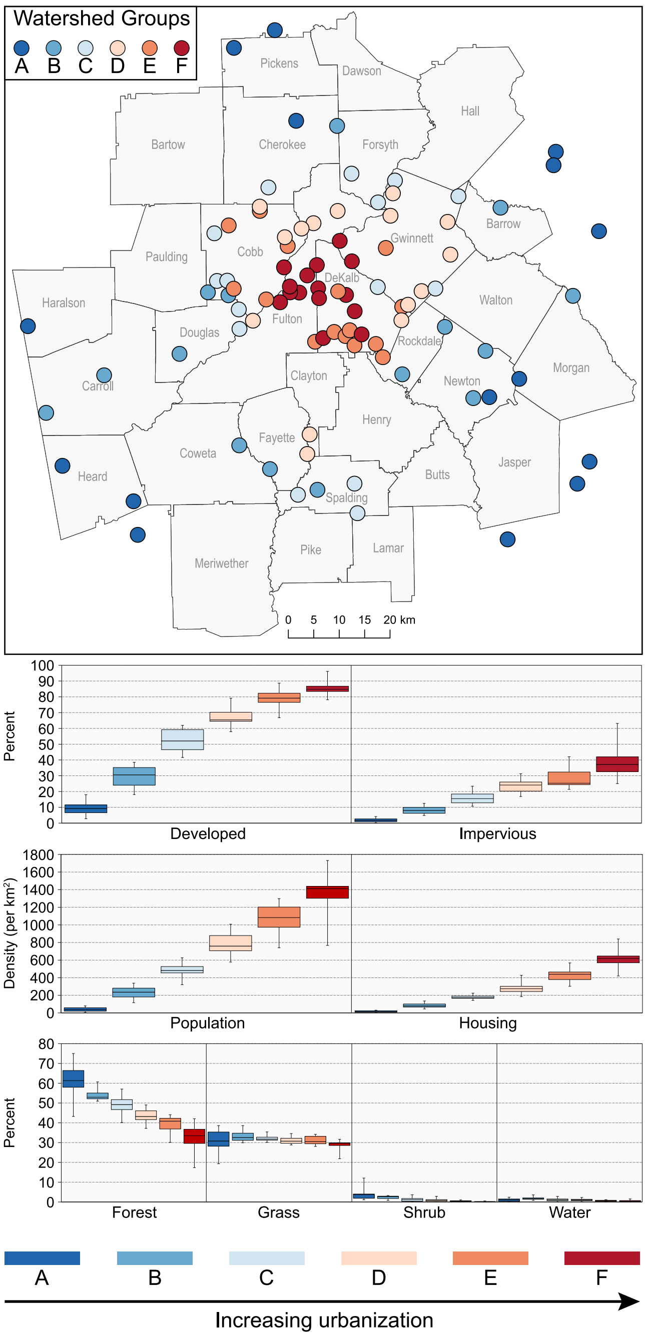


**Fig. 3.** Locations, land-cover, and population and housing-unit densities of the six groups of 15 watersheds.

watersheds, with the mean values for the three most urbanized groups approximately 16 mm.

Only 33 of the 90 watersheds received WWTP effluent ($Q_E$), and this low number of receiving watersheds caused the over-all magnitude of effluent for the 90 watersheds to be small (Fig. 4d). WWTP effluent did not exist for the most urbanized watersheds, since treated wastewater from many urbanized watersheds is released into the Chattahoochee River (Fig. 1). The two largest annual totals of effluent among the 90 watersheds were 70 mm and 110 mm. Therefore, for those watersheds, effluent was a substantial inflow, especially when compared to pipe leakage and irrigation using public water.

### 3.4. AET

$AET_B$ estimates from the Fang et al. (2016) models were upwardly adjusted, and those totals for the watersheds decreased with increasing urbanization (Fig. 4e). Adjustment factors that were applied to each of the 90 watersheds were weighted means of the adjustment factors at the reference watersheds, and those factors ranged from 1.089 to 1.201 (Table 1).

Evaporation totals from impervious surfaces ($E$) were negligible for the least urbanized watersheds and substantial for the most urbanized watersheds (Fig. 4f). Annual totals of evaporation from impervious surfaces ranged from 1 mm to 180 mm. The mean annual evaporation total for Group F, which as noted earlier had a mean imperviousness of 39%, was 115 mm.

The remaining AET term, $AET_I$, was considerably smaller than $AET_B$ for all watersheds and smaller than $E$ for the more urbanized watersheds (Fig. 4g). $AET_I$ was the sum of water from public-water systems (Fig. 4c) and irrigation water withdrawn from the watershed for agriculture, which includes golf courses. Annual totals of $AET_I$ ranged from 0 mm to 36 mm.

While $AET_T$ decreased with increasing urbanization of the watersheds, it was still the largest outflow variable for nearly all watersheds (Fig. 4 and S3). $AET_T$ among the 90 watersheds ranged from 584 mm to 992 mm. Group F watersheds had approximately 37% less $AET_T$ than Group A watersheds. When increasing evaporation rates to 30%, there was minimal change in $AET_T$ for Group A watersheds, while $AET_T$ for Group F watersheds increased from 741 mm to 796 mm (Fig. S3b,c). Evaporation rates of 10% led to $AET_T$ of 686 mm for Group F watersheds (Fig. S3a).

### 3.5. Water withdrawal (W)

The mean values of water withdrawals by public water systems were low for the watershed groups, since only 14 of the 90 watersheds had withdrawals (Fig. 4h). The maximum value for a watershed was 52 mm. Similar to WTTP effluent, water withdrawals did not exist in the most urbanized watersheds.

### 3.6. Stream discharge ($Q_T$)

Annual stream discharge ($Q_T$) generally increased with urbanization (Fig. 4i). Values for the 90 watersheds ranged from 259 mm to 863 mm. For the least urbanized watersheds as whole (i.e., Group A), $Q_T$ was less than half of $AET_T$. Conversely, four of the most urbanized watersheds (i.e., Group F) had $Q_T$ larger than $AET_T$.

### 3.7. I&I

I&I, which as noted in Section 2 was calculated by subtracting the sum of the other outflow totals from the sum of the inflow totals, was a major contributor to outflow in urbanized watersheds (Fig. 4j and 5a,c). Group A and Group F watersheds had mean I&I totals of −2 mm and 138 mm, respectively. The histogram of I&I totals for the 90 watersheds was slightly positively skewed, with a majority of the watersheds having values within 40 mm of zero (Fig. 5c). All watersheds with I&I totals exceeding 100 mm were urbanized watersheds located in the center of the study region (Fig. 5a). More specifically, those watersheds encompassed parts of the City of Atlanta and central and western DeKalb County. This is not unexpected, since, as noted in Section 1, both the City of Atlanta and DeKalb County have been under consent decrees to reduce sewer overflows. The evaporation rate used for impervious surfaces had a large effect on I&I totals of the most urbanized watersheds: the I&I total for Group F watersheds increased from 138 mm to 194 mm when the 10% evaporation rate was used and decreased from 138 mm to 83 mm when the 30% evaporation rate was used (Fig. S3d,e,f).

The spatial pattern and frequency distribution of $I\&I_{\%Q}$ (i.e., I&I as a

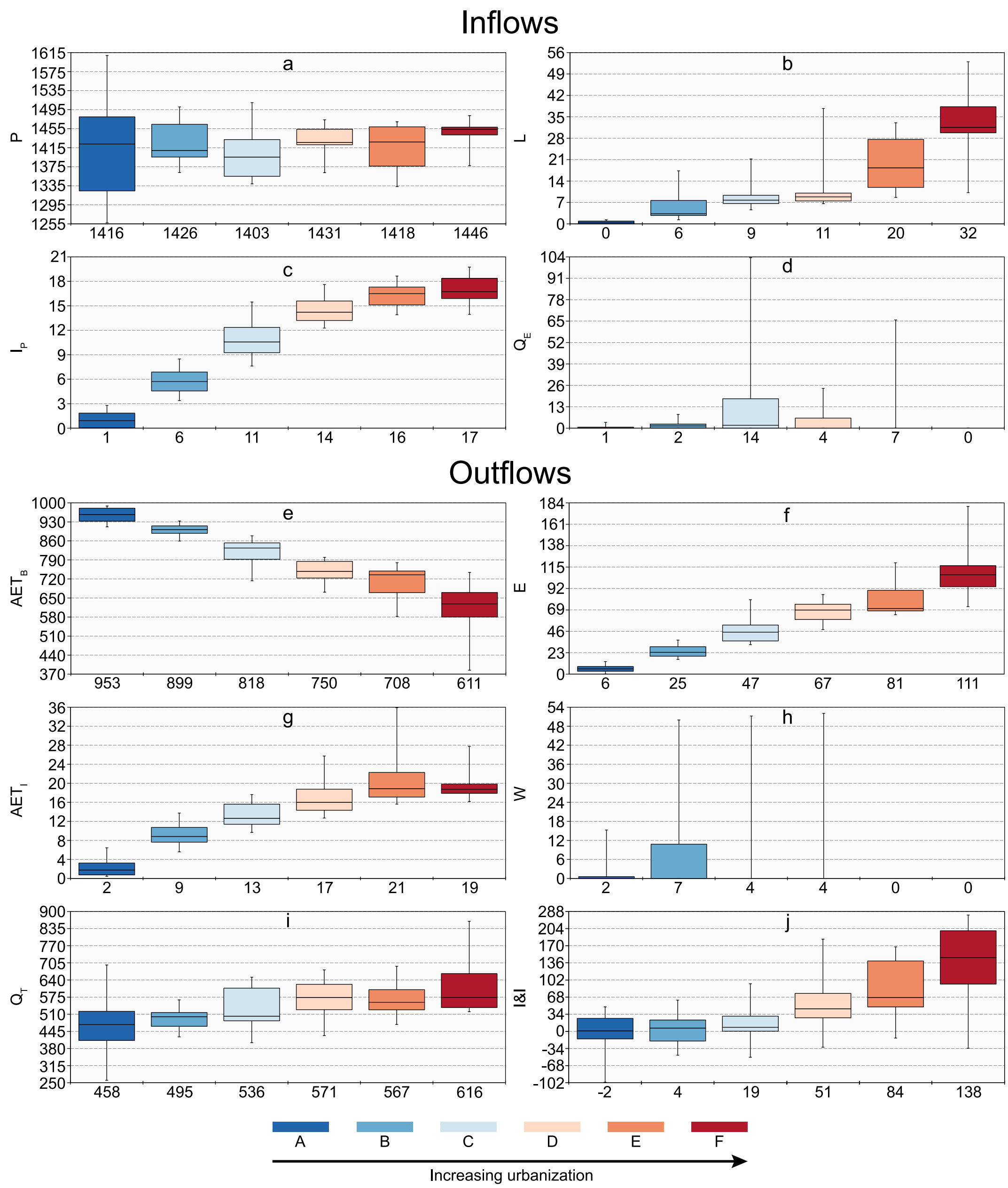


**Fig. 4.** Box-and-whisker plots for the six groups of 15 watersheds for (a) precipitation (P), (b) pipe leakage (L), (c) irrigation with public water ($I_P$), (d) effluent rom wastewater treatment plants ($Q_E$), (e) biome AET ($AET_B$), (f) evaporation from impervious surfaces (E), (g) additional AET from irrigation ($AET_I$), (h) water withdrawal by public systems (W), (i) stream discharge ($Q_T$), and (j) inflow and infiltration (I&I). Mean values are provided on the x-axes. All units are in mm.

**Table 1**
Observed AET ($AET_O$) and predicted AET ($AET_P$) at the four reference watersheds and resulting adjustment factors. ID is the USGS identification number for the stream gauge. Refer to Fig. 1 for the locations of the reference watersheds.

| Watershed | ID | $AET_O$ | $AET_P$ | Adj. Factor |
|---|---|---|---|---|
| R1 | 02381600 | 912 | 759 | 1.201 |
| R2 | 02193340 | 1028 | 891 | 1.153 |
| R3 | 02212600 | 988 | 907 | 1.089 |
| R4 | 02338523 | 935 | 838 | 1.115 |

percentage of stream discharge) closely matched those for I&I totals (Fig. 5b,d). The values for the 90 watersheds ranged from −21% to 44%. Urbanized watersheds in the center of the region had large $I\&I_{\%Q}$ values: four watersheds there had values larger than 40%.

Two components were produced from the PCA of the 13 potential predictor variables for regression modeling, and all variables were strongly and significantly ($\alpha = 0.01$; one-tailed) correlated with both I&I variables (Table S3). The first component represented urbanization in general, and the correlations ranged from 0.57 to 0.71 for I&I and from 0.52 to 0.70 for $I\&I_{\%Q}$ Total housing-unit density had the largest correlations. The second component represented density of older housing, which was a proxy for infrastructure age, and the correlations ranged from 0.58 to 0.77 for I&I and from 0.56 to 0.76 for $I\&I_{\%Q}$. Density of housing units built prior to either 1980 or 1970 had the largest correlations.

Out of the 40 models developed to predict I&I, there were 33 satisfactorily accurate models (i.e., $NSE \geq 0.50$) and those models had the anticipated overprediction of small I&I totals and underprediction of large I&I totals (Fig. 6). There were 27 satisfactorily accurate $I\&I_{\%Q}$ models, and the scatterplot and residuals map for those models were nearly identical to those for I&I and thus are not shown. The largest

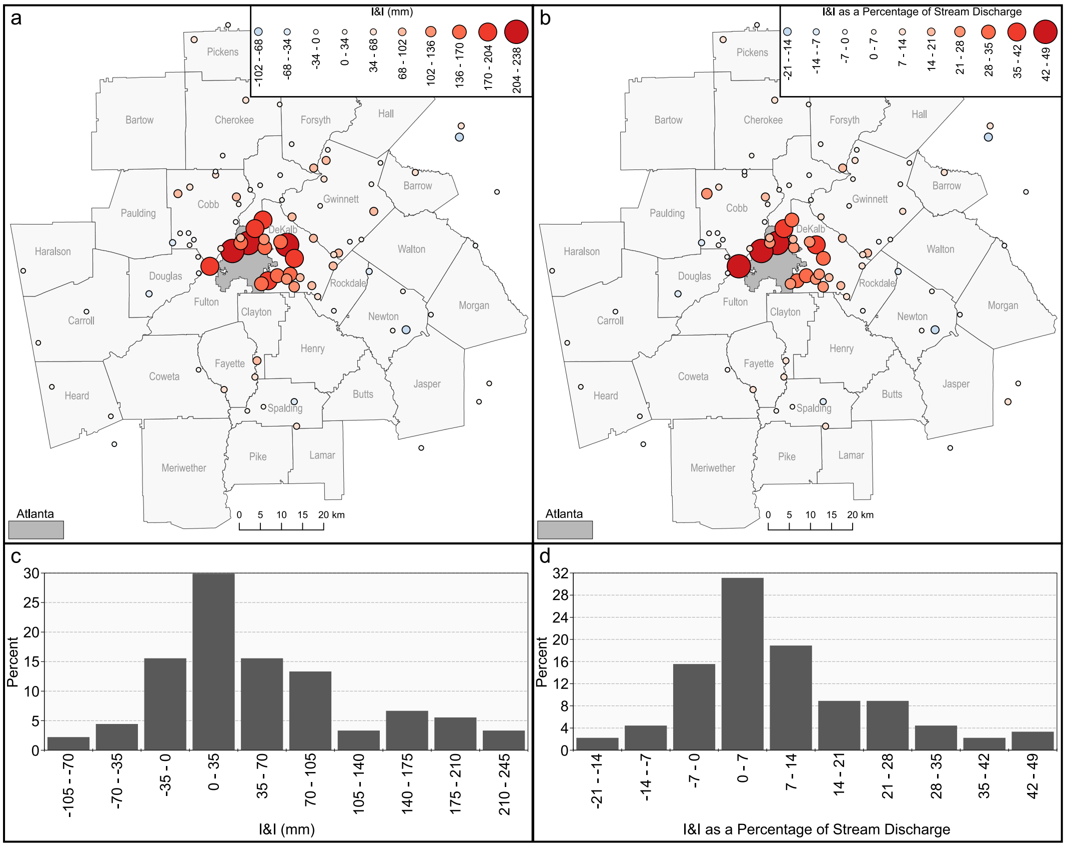

**Fig. 5.** (a) I&I and (b) I&I as a percentage of stream discharge at the 90 watersheds in the Atlanta region. Histograms of values for the 90 watersheds for (c) I&I and (d) I&I as a percentage of stream discharge.

absolute residual occurred at Rottenwood Creek (USGS 02335910) in Cobb County, and this watershed was in the most urban group. Rottenwood Creek had the largest overprediction for all models. An important spatial variable was not excluded from the models, since there was no significant spatial autocorrrelation among the residuals (Fig. 6b).

Infrastructure-age variables produced the most accurate models for I&I (Table 2). The top three models in terms of *NSE* values had *NSE* values of either 0.59 or 0.60 and $r^2$ values of either 0.61 or 0.62. The predictor variables were $HD_{70}$ (i.e., density of housing units built prior to 1970) along with either population density, density of all housing units, or percent imperviousness. Based on the standardized slope coefficients, $HD_{70}$ was the more impactful predictor in all models.

The most accurate models for $I\&I_{\%Q}$ were similar to the models for I&I (Table 2). *NSE* values for the three models were either 0.57 or 0.58 and $r^2$ values were either 0.59 or 0.60. The predictor variables were nearly the same as those in the I&I models, and $HD_{70}$ was again the more impactful predictor.

### 3.8. *Impacts of uncertainties of water-budget terms on I&I totals*

There was considerable variation among the water-budget terms with respect to impacts of uncertainties on I&I totals (Table S4). *P* was the largest term in the water budget, and, if *P* were not connected to $AET_B$, a 10% change in precipitation could produce extremely large changes (±140 mm) in I&I. However, this is a greatly exaggerated impact on I&I: *P* was used to estimate values of two outflow terms, $AET_B$ and *E*; therefore, the impacts of a systematic overestimation or underestimation of *P* were buffered by corresponding changes to $AET_B$ and *E*. In addition, *P* uncertainty was more likely 5% rather than 10% (McMillan et al., 2012). The two other large terms, besides *P*, were $AET_B$ and $Q_T$, and, similar to *P*, small changes in those terms would have large impacts on I&I. Those two terms most likely had uncertainties of at least 10% (Baffaut et al., 2020; McMillan et al., 2012). *E* definitely had an uncertainty larger than 10% due to uncertainties in evaporation rates from impervious surfaces; thus, the impact of *E* on I&I totals was much larger than what is shown in Table S4. The remaining terms ($L$, $I_P$, $Q_E$, $I_A$, and $W$) had either no impacts or minimal impacts on I&I for a watershed group.

## 4. Discussion

### 4.1. *Urban-watershed I&I totals in this study are realistic*

Annual I&I totals for urban watersheds in this study are similar to I&I totals found for urban catchments in Europe, where much of the existing I&I analyses have been conducted. The comparisons with results from European studies are few, since – as noted in Section 1 – most studies do not provide enough information for calculating area-normalized depths of I&I. The mean I&I total from the watersheds in this study with the five largest totals (i.e., top 6%) is 216 mm, which is 40% of stream discharge. These I&I totals are comparable to what has been found for monitored urban catchments in Prague, Czech Republic and western Germany, where annual I&I totals are between 200 mm and 250 mm (Bareš et al., 2009; Bareš et al., 2012; Staufer et al., 2012). These are the only known published European studies that provide all the information needed to calculate I&I as an area-normalized depth. I&I in the six catchments in Bareš et al. (2009, 2012) are only measured for several months at a time over several years, and after upscaling each of the six sub-annual I&I totals to an annual total, the mean total for those catchments is 206 mm. The mean I&I total for the two catchments in western Germany, measured over 73 months prior to the rehabilitation of one of the sewer networks, is 246 mm (Staufer et al., 2012).

Differences in population and housing density seem to explain why I&I totals for urban watersheds in the Atlanta region are substantially

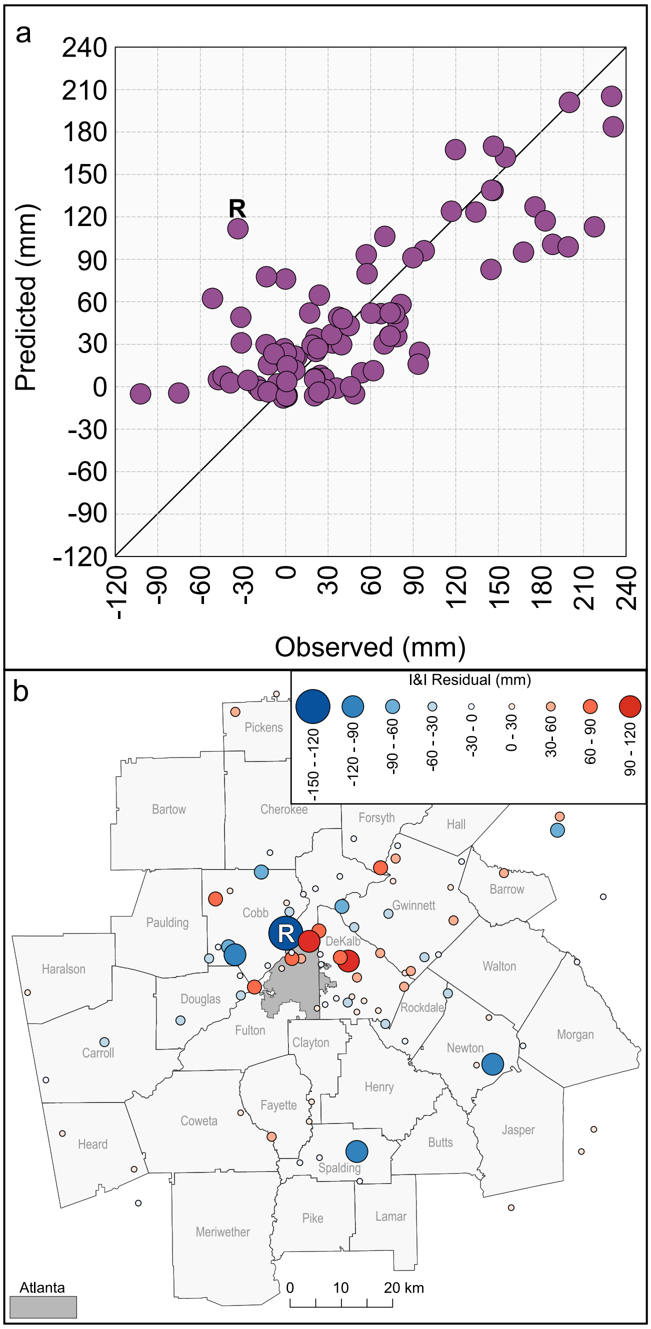


**Fig. 6.** (a) Scatterplot of observed I&I totals vs predicted I&I totals for the 90 watersheds and (b) map of residuals. Mean predicted I&I totals from the 33 satisfactorily accurate models were used. “R” (Rottenwood Creek) is the largest overprediction (i.e., most negative residual).

less than totals for two urban watersheds in Baltimore. Baltimore is a good comparison with Atlanta, because both locales have Cfa climates and are located wholly or partially in the Piedmont physiographic province of the eastern United States. The urban watersheds in Bhaskar and Welty (2012) have a mean annual I&I total of 565 mm, which is 120% of stream discharge. The Baltimore I&I total, which is the largest value in the literature, is 2.4 times larger than the largest I&I total in this study. That total was at Peachtree Creek (USGS 02336300). The difference in I&I magnitude might be explained by the much higher degree of urbanization – and thus increased sewer-pipe density – in the Baltimore watersheds: the mean population and housing densities of the Baltimore watersheds were twice that of Peachtree Creek.

**Table 2**
The three most accurate multiple linear regression models used to predict I&I and $I\&I_{\%Q}$ (i.e., I&I as a percentage of stream discharge). Provided to the right of the equations are coefficients of determination ($r^2$) and the Nash-Sutcliffe *E (NSE)*. The variables in the models are described below. $D_A$ is percent all developed land. $HD_{70}$ is density of housing units built prior to 1970. $HD_T$ is density of all housing units. I is percent imperviousness. PD is population density. Housing density is units $km^{-2}$. Population density is persons $km^{-2}$.

| Variable | Equation | *NSE* | $r^2$ |
|---|---|---|---|
| I&I | $-10.140 + 0.365 * HD_{70} + 1.052 * PD$ (0.501) | 0.60 (0.354) | 0.62 |
| I&I | $-8.274 + 0.315 * HD_{70} + 0.137 * HD_T$ (0.432) | 0.60 (0.405) | 0.62 |
| I&I | $-8.486 + 0.440 * HD_{70} + 1.44 * I$ (0.604) | 0.59 (0.274) | 0.61 |
| $I\&I_{\%Q}$ | $-1.548 + 0.072 * HD_{70} + 0.009 * PD$ (0.525) | 0.58 (0.312) | 0.60 |
| $I\&I_{\%Q}$ | $-1.170 + 0.064 * HD_{70} + 0.022 * HD_T$ (0.469) | 0.57 (0.350) | 0.60 |
| $I\&I_{\%Q}$ | $-2.831 + 0.082 * HD_{70} + 0.119 * D_A$ (0.598) | 0.57 (0.245) | 0.59 |

A comparison of results with previous I&I studies involving the same watersheds in the Atlanta region indicates that this study might overestimate I&I in suburban watersheds. The same three watersheds are examined in Diem et al. (2021), Pangle et al. (2022), and in this study, and those watersheds exist within the South River Watershed (Fig. 1). The mean I&I totals for those watersheds from those three studies are 37 mm, 45 mm, and 121 mm, respectively. It is possible that Diem et al. (2021) has a conservative estimate of I&I totals, because it relies on monthly WWTP-effluent data and the method used has the assumption that the month with the lowest effluent total has no I&I. It also is possible that I&I estimates reported by Pangle et al. (2022) may be lower than average considering the results are reported for a calendar year that was relatively dry: the precipitation total for the year was greater than 100 mm less than the climatological mean (Aulenbach and Peters, 2018; Diem et al., 2018). Pangle et al. (2022) also included an underlying assumption that groundwater infiltration becomes negligibly small during the driest time of the year.

The modeling results in this study support previous findings that aging infrastructure is a major control of I&I. The most accurate regression models all have density of housing units built prior to 1970 as the most important predictor of I&I among the 90 watersheds. This variable, which is a proxy for sewage-infrastructure age, has a standardized slope coefficient that is larger than the coefficient for the other variable in the models. The other variable is either imperviousness, percent all developed land, population density, or total housing density. Sewage infrastructure that is 40 years old has been shown to have much higher groundwater infiltration than infrastructure that is several decades newer (Prigiobbe and Giulianelli, 2009).

The regression modeling may have discovered significant sewage-system repair in an urban watershed. Rottenwood Creek in eastern Cobb County appears to have negligible I&I and thus has the largest I&I overprediction of all the watersheds. Sewer replacement is a likely reason for the negligible I&I: approximately 1.7 km of interceptor-sewer pipe that was constructed over 60 years ago was replaced and relocated (Gilman, 2019).

### 4.2. Uncertainties in water-budget terms are in issue for both rural and urban watersheds

Except for large imbalances in the water budget that occurred for several watersheds, the water budgets for rural watersheds in general are almost fully closed. There was strong agreement between inflow totals and outflow totals, since the mean I&I totals for the most rural watersheds is approximately zero, which is the expected I&I total. This agreement is enabled by the adjustment of $AET_B$ totals based on $AET_B$ totals at reference watersheds that are derived from precipitation and stream discharge. However, even with the adjustment of $AET_B$, there are large negative I&I totals (i.e., $<-40$ mm) – which are unrealistic – at several watersheds. It is likely that those large negative totals are caused by one or more of the following: underestimated precipitation, overestimated stream discharge, and overestimated AET. For example, 5% adjustments in any combination of terms at those watersheds would increase I&I totals to close to zero.

While considerable effort is involved in estimating anthropogenic flows such as pipe leakage rates and irrigation water use in urban watersheds, those terms have minor impacts on I&I in the Atlanta region; rather, AET probably has the largest magnitude of uncertainty – and thus the largest impact on I&I totals in urban watersheds – among all water-budget components. The accuracy of the urban AET estimates in this study are unknown: no previous studies have estimated urban AET totals in the southeastern United States and there is a large range (350 mm to 613 mm) in published AET estimates for urban areas in general (Bhaskar and Welty, 2012; Claessens et al., 2006; Fang et al., 2020; Grimmond and Oke, 1986; Jia et al., 2001; Kokkonen et al., 2018; Litvak et al., 2017; Mitchell et al., 2003; Sloto and Buxton, 2005). Urban AET totals in this study, which are initially estimated using a 20% evaporation rate for impervious surfaces, are higher than the totals above: the minimum AET total in this study is 584 mm and the mean value for the most urbanized group of watersheds (i.e., Group F) is 741 mm. AET totals should be relatively high in this study for the following reasons: (1) the southeastern United States has high annual totals of both precipitation (Hijmans et al., 2005) and potential evapotranspiration (Weiß and Menzel, 2008); (2) many urban watersheds in the Atlanta region have substantial vegetative cover (see Section 3.1); and (3) this study accounted for evaporation from impervious surfaces. In the Atlanta region and other places with wet climates, the choice of evaporation rate can change AET greatly and in turn I&I. If a 30% evaporation rate is used instead of a 10% evaporation rate, AET for the Group F watersheds increases from 686 mm to 796 mm and I&I totals thus decrease from 194 mm to 84 mm (i.e., a 57% decrease).

### 4.3. I&I can be a major anthropogenic flow in humid climates

I&I is among the largest anthropogenic flows occurring in urban watersheds in the Atlanta region and probably most other urban watersheds in the eastern United States. In the Atlanta region, annual I&I totals in urban watersheds (i.e., Group F in this study) are four and eight times larger than water-system pipe leakage and irrigation with municipal water, respectively. I&I also is much larger than both pipe leakage and irrigation in urban watersheds in Baltimore (Bhaskar and Welty, 2012). The two other relevant piped flows, WWTP effluent and water withdrawal by water systems, do not occur in urban watersheds in the Atlanta region. Water use, which increases with irrigation and is a control of pipe leakage, is relatively low in the Atlanta region: the mean per capita rate for 65 water systems in this study is 160 $m^3\ y^{-1}$, while the mean rate for the United States as a whole is 190 $m^3\ y^{-1}$ (Dieter et al., 2018). The Atlanta region has an irrigation rate of approximately 100 mm per lawn, while urban areas with arid, semi-arid, and Mediterranean climates in the western United States and Canada can have annual irrigation rates approaching 2000 mm (DeOreo et al., 2016).

## 5. Conclusions

Through the calculation of water budgets for 90 watersheds in the Atlanta region, this study has shown that I&I is an important outflow component for urban watersheds. The magnitude of I&I in the watersheds is comparable to findings for urbanized catchments in Germany and the Czech Republic. As should be expected, the largest I&I totals in the Atlanta region occur for the most part in municipalities that are under consent decrees to improve their sewer infrastructure. There are several suburban watersheds in this study that have I&I estimated in previous studies, and the I&I estimates in this study are substantially larger. This suggests that I&I totals for suburban watersheds in this study might be overestimates. This study shows that I&I is controlled primarily by the density of older housing, a proxy for older sewer infrastructure. I&I also is strongly correlated with imperviousness, the presence of developed land, population density, and housing density in general. Nevertheless, watersheds with similar land-cover, population densities, and housing densities can have widely varying I&I totals. For example, recently replaced sewage infrastructure in an urbanized watershed greatly reduces I&I. The I&I totals in this study should be treated as approximations, since estimating I&I using the water-budget approach involves uncertainties for all water-budget components. Accurate estimates of AET are crucial to obtaining valid estimates of I&I; however, obtaining accurate AET estimates for a large number of watersheds is difficult. One way this study attempts to minimize AET errors is by using results from reference watersheds to produce adjustment factors for the other watersheds. However, the choice of evaporation rate for impervious surface greatly impacts AET totals for urban watersheds. Despite the uncertainties involved with using the water-budget approach to estimate I&I, this study shows that is a suitable exploratory procedure for a large number of watersheds. Major drawbacks of the water-budget approach are the aforementioned uncertainties and the data-extensive nature (e.g., many geo-spatial data sets) of the procedure. Nevertheless, annual totals for anthropogenic components (e.g., water-supply pipe leakage), which do require involved methods, have small impacts on the water budget compared to precipitation, AET, and stream discharge. Estimating I&I using the water-budget approach is a useful initial approach to estimating I&I throughout a region.

## 6. Data statement

Land-cover, population, housing, and water-budget data for the 90 watersheds are available at https://data.mendeley.com/datasets/j7kkgm52xm.


### CRediT authorship contribution statement

**Jeremy E. Diem:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Luke A. Pangle:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Richard A. Milligan:** Writing – original draft, Writing – review & editing. **Ellis A. Adams:** Writing – original draft, Writing – review & editing.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


### Data availability

Land-cover, population, housing, and water-budget data for the 90 watersheds are available at https://data.mendeley.com/datasets/j7kkgm52xm.

## Acknowledgements

This work was supported by the National Science Foundation [grant number 1853809].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2022.128629.

## References

Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci. Data 5, 170191. https://doi.org/10.1038/sdata.2017.191.

Arnone, R.D., Perdek Walling, J., 2007. Waterborne pathogens in urban watersheds. J. Water Health 5, 149–162. https://doi.org/10.2166/wh.2006.001.

Aulenbach, B.T., Peters, N.E., 2018. Quantifying climate-related interactions in shallow and deep storage and evapotranspiration in a forested, seasonally water-limited watershed in the southeastern United States. Water Resour. Res. 54, 3037–3061. https://doi.org/10.1002/2017WR020964.

Baffaut, C., Baker, J.M., Biederman, J.A., Bosch, D.D., Brooks, E.S., Buda, A.R., Demaria, E.M., Elias, E.H., Flerchinger, G.N., Goodrich, D.C., Hamilton, S.K., Hardegree, S.P., Harmel, R.D., Hoover, D.L., King, K.W., Kleinman, P.J., Liebig, M.A., McCarty, G.W., Moglen, G.E., Moorman, T.B., Moriasi, D.N., Okalebo, J., Pierson, F.B., Russell, E.S., Saliendra, N.Z., Saha, A.K., Smith, D.R., Yasarer, L.M.W., 2020. Comparative analysis of water budgets across the U.S. long-term agroecosystem research network. Journal of Hydrology 588, 125021. 10.1016/j.jhydrol.2020.125021.

Bareš, V., Stránský, D., Sýkora, P., 2009. Sewer infiltration/inflow: long-term monitoring based on diurnal variation of pollutant mass flux. Water Sci. Technol. 60, 1–7. https://doi.org/10.2166/wst.2009.280.

Bareš, V., Stránský, D., Sýkora, P., 2012. Evaluation of sewer infiltration/inflow using COD mass flux method: case study in Prague. Water Sci. Technol. 66, 673–680. https://doi.org/10.2166/wst.2012.229.

Bhaskar, A.S., Beesley, L., Burns, M.J., Fletcher, T.D., Hamel, P., Oldham, C.E., Roy, A.H., 2016. Will it rise or will it fall? Managing the complex effects of urbanization on base flow. Freshwater Science 35, 293–310. https://doi.org/10.1086/685084.

Bhaskar, A.S., Welty, C., 2012. Water balances along an urban-to-rural gradient of metropolitan Baltimore, 2001–2009. Environ. Eng. Geosci. 18, 37–50.

Booth, D.B., Jackson, C.R., 1997. Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. J. Am. Water Resour. Assoc. 33, 1077–1090. https://doi.org/10.1111/j.1752-1688.1997.tb04126.x.

Buban, M.S., Lee, T.R., Baker, C.B., 2020. A Comparison of the U.S. Climate Reference Network Precipitation Data to the Parameter-Elevation Regressions on Independent Slopes Model (PRISM). J. Hydrometeorol. 21, 2391–2400. https://doi.org/10.1175/JHM-D-19-0232.1.

Cahoon, L.B., Hanke, M.H., 2017. Rainfall effects on inflow and infiltration in wastewater treatment systems in a coastal plain region. Water Sci. Technol. 75, 1909–1921. https://doi.org/10.2166/wst.2017.072.

City of Atlanta, 2011. 2011 Comprehensive Development Plan. accessed 8 September 2022. https://dca.ga.gov/sites/default/files/atlanta_ci_cdp_community_assessment_2011_part_i.pdf.

Claessens, L., Hopkinson, C., Rastetter, E., Vallino, J., 2006. Effect of historical changes in land use and climate on the water budget of an urbanizing watershed. Water Resour. Res. 42 https://doi.org/10.1029/2005WR004131.

Cohard, J.-M., Rosant, J.-M., Rodriguez, F., Andrieu, H., Mestayer, P.G., Guillevic, P., 2018. Energy and water budgets of asphalt concrete pavement under simulated rain events. Urban Clim. 24, 675–691. https://doi.org/10.1016/j.uclim.2017.08.009.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064. https://doi.org/10.1002/joc.1688.

Daly, C., Doggett, M.K., Smith, J.I., Olson, K.V., Halbleib, M.D., Dimcovic, Z., Keon, D., Loiselle, R.A., Steinberg, B., Ryan, A.D., Pancake, C.M., Kaspar, E.M., 2021. Challenges in observation-based mapping of daily precipitation across the conterminous United States. J. Atmos. Oceanic Technol. 38 (11), 1979–1992. https://doi.org/10.1175/JTECH-D-21-0054.1.

DeOreo, W.B., Mayer, P.W., Dziegielewski, B., Kiefer, J., 2016. Residential End Uses of Water, Version 2. Water Research Foundation, Denver CO.

Diem, J.E., Hill, T.C., Milligan, R.A., 2018. Diverse multi-decadal changes in streamflow within a rapidly urbanizing region. J. Hydrol. 556, 61–71. https://doi.org/10.1016/j.jhydrol.2017.10.026.

Diem, J.E., Pangle, L.A., Milligan, R.A., Adams, E.A., 2021. Intra-annual variability of urban effects on streamflow. Hydrol. Process. 35 https://doi.org/10.1002/hyp.14371.

Dieter, C.A., Maupin, M.A., Caldwell, R.R., Harris, M.A., Ivahnenko, T.I., Lovelace, J.K., Barber, N.L., Linsey, K.S., 2018. Estimated Use of Water in the United States in 2015: U.S. Geological Survey Circular 1441, 65 p. https://doi.org/10.3133/cir1441.

Dirckx, G., Van Daele, S., Hellinck, N., 2016. Groundwater Infiltration Potential (GWIP) as an aid to determining the cause of dilution of waste water. J. Hydrol. 542, 474–486. https://doi.org/10.1016/j.jhydrol.2016.09.020.

Dirckx, G., Fenu, A., Wambecq, T., Kroll, S., Weemaes, M., 2019. Dilution of sewage: Is it, after all, really worth the bother? J. Hydrol. 571, 437–447. https://doi.org/10.1016/j.jhydrol.2019.01.065.

Fang, D., Hao, L., Cao, Z., Huang, X., Qin, M., Hu, J., Liu, Y., Sun, G., 2020. Combined effects of urbanization and climate change on watershed evapotranspiration at multiple spatial scales. J. Hydrol. 587, 124869 https://doi.org/10.1016/j.jhydrol.2020.124869.

Fang, Y., Sun, G., Caldwell, P., McNulty, S.G., Noormets, A., Domec, J.-C., King, J., Zhang, Z., Zhang, X., Lin, G., Zhou, G., Xiao, J., Chen, J., 2016. Monthly land cover-specific evapotranspiration models derived from global eddy flux measurements and remote sensing data: Monthly ET Models. Ecohydrology 9, 248–266. https://doi.org/10.1002/eco.1629.

Georgia Environmental Protection Division, 2014. Golf Course Irrigation Prediction and Estimation Worksheet (WWW Document). accessed 12 May 2022. https://www.epd.georgia.gov/document/publication/golfcourseirrigationpredictionandestimationworksheet070914xls/download.

Georgia Environmental Protection Division, 2022. Water Efficiency and Water Loss audits (WWW Document). accessed 12 May 2022. https://epd.georgia.gov/watershed-protection-branch/water-efficiency-and-water-loss-audits.

Gilman, J., 2019. Marietta Water: Rottenwood Creek Sewer Interceptor Project (WWW Document). https://storymaps.arcgis.com/stories/9f43bb7dd4834f42bd491cc1ed4ffe62 (accessed 12 May 2022).

Gleick, P.H., Haasz, D., Henges-Jeck, C., Srinivasan, V., Wolff, G., Cushing, K.K., Mann, A., 2003. Waste Not. The Potential for Urban Water Conservation in California. Pacific Institute, Want Not https://pacinst.org/wp-content/uploads/2003/11/waste_not_want_not_full_report.pdf.

Golf Course Superintendents Association of America, 2007. Golf Course Environmental Profile. Golf Course Superintendents Association of America, Lawrence, KS.

Grimmond, C.S.B., Oke, T.R., 1999. Evapotranspiration rates in urban areas. Proceedings of IUGG 99 Symposium HS5, 259, 235–243.

Grimmond, C.S.B., Oke, T.R., 1986. Urban water balance: 2. Results from a suburb of Vancouver. British Columbia. Water Resources Research 22, 1404–1412. https://doi.org/10.1029/WR022i010p01404.

Hamon, R.W., 1963. Estimating potential evapotranspiration. Transactions of the American Society of Civil Engineers 128, 324–338. https://doi.org/10.1061/TACEAT.0008673.

Healy, R.W., Winter, T.C., LaBaugh, J.W., Franke, O.L., 2007. Water Budgets: Foundations for Effective Water-Resources and Environmental Management. United States Geological Survey. Circular 308. https://pubs.usgs.gov/circ/2007/1308/pdf/C1308_508.pdf.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. https://doi.org/10.1002/joc.1276.

Jia, Y., Ni, G., Kawahara, Y., Suetsugi, T., 2001. Development of WEP model and its application to an urban watershed. Hydrol. Process. 15, 2175–2194. https://doi.org/10.1002/hyp.275.

Kachigan, S.K., 1991. Multivariate Statistical Analysis: A Conceptual Introduction. Radius, New York.

Kampf, S.K., Burges, S.J., Hammond, J.C., Bhaskar, A., Covino, T.P., Eurich, A., Harrison, H., Lefsky, M., Martin, C., McGrath, D., Puntenney-Desmond, K., Willi, K., 2020. The Case for an Open Water Balance: Re-envisioning Network Design and Data Analysis for a Complex. Uncertain World. Water Resources Research 56. https://doi.org/10.1029/2019WR026699.

Karpf, C., Krebs, P., 2011. Quantification of groundwater infiltration and surface water inflows in urban sewer networks based on a multiple model approach. Water Res. 45, 3129–3136. https://doi.org/10.1016/j.watres.2011.03.022.

Karpf, C., Krebs, P., 2013. Modelling of groundwater infiltration into sewer systems. Urban Water J. 10, 221–229. https://doi.org/10.1080/1573062X.2012.724077.

Kokkonen, T.V., Grimmond, C.S.B., Christen, A., Oke, T.R., Järvi, L., 2018. Changes to the water balance over a century of urban development in two neighborhoods: Vancouver, Canada. Water Resour. Res. 54, 6625–6642. https://doi.org/10.1029/2017WR022445.

Kracht, O., Gresch, M., Gujer, W., 2007. A stable isotope approach for the quantification of sewer infiltration. Environ. Sci. Technol. 41, 5839–5845. https://doi.org/10.1021/es062960c.

Litvak, E., Manago, K.F., Hogue, T.S., Pataki, D.E., 2017. Evapotranspiration of urban landscapes in Los Angeles, California at the municipal scale. Water Resour. Res. 53, 4236–4252. https://doi.org/10.1002/2016WR020254.

Lu, J., Sun, G., McNulty, S.G., Amatya, D.M., 2005. A comparison of six potential evapotranspiration methods for regional use in the southeastern United States. J. Am. Water Resour. Assoc. 41, 621–633. https://doi.org/10.1111/j.1752-1688.2005.tb03759.x.

Malek Mohammadi, M., Najafi, M., Kermanshachi, S., Kaushal, V., Serajiantehrani, R., 2020. Factors influencing the condition of sewer pipes: State-of-the-art review. J. Pipeline Syst. Eng. Pract. 11, 03120002. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000483.

McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26, 4078–4111. https://doi.org/10.1002/hyp.9384.

Miller, J.A., 1990. Hydrologic Investigations Atlas, Segment 6, Alabama, Florida, Georgia, and South Carolina. (Report No. 730G; Hydrologic Atlas). United States Geological Survey https://pubs.usgs.gov/ha/730g/report.pdf.

Mitchell, V.G., McMahon, T.A., Mein, R.G., 2003. Components of the total water balance of an urban catchment. Environ. Manage. 32, 735–746. https://doi.org/10.1007/s00267-003-2062-2.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.
Myneni, R., Knyazikhin, Y., Park, T., 2015. MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006 . NASA EOSDIS Land Processes DAAC. Accessed 2022-03-18 from 10.5067/MODIS/MCD15A2H.006.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Pangle, L.A., Diem, J.E., Milligan, R., Adams, E., Murray, A., 2022. Contextualizing inflow and infiltration within the streamflow regime of urban watersheds. Water Resour. Res. 58 https://doi.org/10.1029/2021WR030406.
Pawlowski, C.W., Rhea, L., Shuster, W.D., Barden, G., 2014. Some factors affecting inflow and infiltration from residential sources in a core urban area: Case study in a Columbus, Ohio, neighborhood. J. Hydraul. Eng. 140, 105–114. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000799.
Prigiobbe, V., Giulianelli, M., 2009. Quantification of sewer system infiltration using δ18O hydrograph separation. Water Sci. Technol. 60, 727–735. https://doi.org/10.2166/wst.2009.399.
Ragab, R., Bromley, J., Rosier, P., Cooper, J.D., Gash, J.H.C., 2003a. Experimental study of water fluxes in a residential area: 1. Rainfall, roof runoff and evaporation: the effect of slope and aspect. Hydrol. Process. 17, 2409–2422. https://doi.org/10.1002/hyp.1250.
Ragab, R., Rosier, P., Dixon, A., Bromley, J., Cooper, J.D., 2003b. Experimental study of water fluxes in a residential area: 2. Road infiltration, runoff and evaporation. Hydrol. Process. 17, 2423–2437. https://doi.org/10.1002/hyp.1251.
Ramier, D., Berthier, E., Andrieu, H., 2004. An urban lysimeter to assess runoff losses on asphalt concrete plates. Physics and Chemistry of the Earth, Parts A/B/C 29, 839–847. https://doi.org/10.1016/j.pce.2004.05.011.
Ramier, D., Berthier, E., Andrieu, H., 2011. The hydrological behaviour of urban streets: long-term observations and modelling of runoff losses and rainfall-runoff transformation. Hydrol. Process. 25, 2161–2178. https://doi.org/10.1002/hyp.7968.
Rödel, S., Günthert, F.W., Brüggemann, T., 2017. Investigating the impacts of extraneous water on wastewater treatment plants. Water Sci. Technol. 75, 847–855. https://doi.org/10.2166/wst.2016.570.
Schulz, N., Baur, R., Krebs, P., 2005. Integrated modelling for the evaluation of infiltration effects. Water Sci. Technol. 52, 215–223. https://doi.org/10.2166/wst.2005.0136.
Southern Environmental Law Center, 2014. SELC Drinking Water Database. Chapel Hill, NC.
Shehab, T., Moselhi, O., 2005. Automated detection and classification of infiltration in sewer pipes. J. Infrastruct. Syst. 11, 165–171. https://doi.org/10.1061/(ASCE)1076-0342(2005)11:3(165).
Shepherd, J.M., 2005. A review of current investigations of urban-induced rainfall and recommendations for the future. Earth Interact 9, 1–27. https://doi.org/10.1175/EI156.1.
Sloto, R.A., Buxton, D.E., 2005. Water Budgets for Selected Watersheds in the Delaware River Basin, Eastern Pennsylvania and Western New Jersey (Report No. 2005–5113), Scientific Investigations Report. United States Geological Survey. 10.3133/sir20055113.
Staufer, P., Scheidegger, A., Rieckermann, J., 2012. Assessing the performance of sewer rehabilitation on the reduction of infiltration and inflow. Water Res. 46, 5185–5196. https://doi.org/10.1016/j.watres.2012.07.001.
Townsend-Small, A., Pataki, D.E., Liu, H., Li, Z., Wu, Q., Thomas, B., 2013. Increasing summer river discharge in southern California, USA, linked to urbanization. Geophys. Res. Lett. 40, 4643–4647. https://doi.org/10.1002/grl.50921.
Trewartha, G.T., Horn, L.H., 1980. An Introduction to Climate. McGraw-Hill.
U.S. District Court, 1998. Upper Chattahoochee Riverkeeper Fund, Inc. et al. v. The City of Atlanta. Civil Action File NO.l:95-CV-2550-TWT. United States District Court for the Northern District of Georgia, Atlanta Division. https://www.epa.gov/sites/production/files/2013-10/documents/atlanta1998-cd.pdf.
U.S. District Court, 1999. The United States of America and the State of Georgia v. The City of Atlanta. First Amended Consent Decree. Civil Action File NO.l:98-CV-1956-TWT. United States District Court for the Northern District of Georgia, Atlanta Division. https://www.epa.gov/sites/production/files/documents/atlanta1999-cd.pdf.
U.S. District Court, 2011. The United States of America and the State of Georgia v. DeKalb County, Georgia. Consent Decree. United States District Court for the Northern District of Georgia, Atlanta Division. https://www.epa.gov/sites/production/files/documents/dekalb-cd.pdf.
U.S. EPA, 2014. Quick Guide for Estimating Infiltration and Inflow for Region 1 NPDES Annual Reporting. https://www3.epa.gov/region1/sso/pdfs/QuickGuide4EstimatingInfiltrationInflow.pdf.
U.S. EPA, 2022. Sanitary Sewer Overflows (SSOs) (WWW Document). https://www.epa.gov/npdes/sanitary-sewer-overflows-ssos (accessed 12 May 2022).
Weiß, G., Brombach, H., Haller, B., 2002. Infiltration and inflow in combined sewer systems: long-term analysis. Water Sci. Technol. 45, 11–19. https://doi.org/10.2166/wst.2002.0112.
Weiß, M., Menzel, L., 2008. A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. Adv. Geosci. 18, 15–23. https://doi.org/10.5194/adgeo-18-15-2008.
Zhang, K., Parolari, A.J., 2022. Impact of stormwater infiltration on rainfall-derived inflow and infiltration: A physically based surface–subsurface urban hydrologic model. J. Hydrol. 610, 127938 https://doi.org/10.1016/j.jhydrol.2022.127938.