Incorporating spatial heterogeneity into evapotranspiration estimates for bioretention basins

Joshua S. Caplan[1*], Martin Bouda[2†], Allyson B. Salisbury[1††], Michael Alonzo[3], Jonathan E. Nyquist[4], Laura Toran[4], Sasha W. Eisenman[1]

[1] Department of Environmental Design, Temple University, Ambler, PA 19002, USA
[2] Botanical Institute of the Czech Academy of Sciences, Průhonice, Czechia
[3] Department of Environmental Science, American University, Washington, DC, USA
[4] Department of Earth and Environmental Science, Temple University, Philadelphia, PA, USA

[†] Current address: Department of Plant Ecophysiology, University of Hohenheim, 70599 Hohenheim, Germany
[††] Current address: Department of Ecology, Evolution and Natural Resources, Rutgers – the State University of New Jersey, New Brunswick, NJ, USA

Correspondence to: Joshua S. Caplan (jcaplan@temple.edu)

Abstract. Green stormwater infrastructure (GSI) systems such as bioretention basins are frequently used in urban settings to reduce the amount of stormwater runoff entering combined sewer systems, thus protecting downstream waterbodies. Retaining stormwater in GSI allows it to infiltrate into the soil or return to the atmosphere via evapotranspiration (ET). While infiltration rates can be quantified with reasonable accuracy, methods of quantifying ET typically rely on models designed for homogeneous landcover like agricultural fields; the high spatial variation in factors including vegetation, light, and soil moisture renders estimates of ET from bioretention basins highly uncertain. To assess the influence of such variation on basin-scale ET and evaluate means of correcting for it, we quantified ET for a bioretention basin in Philadelphia, USA using three approaches: (1) an empirically-based model that incorporated chamber flux measurements of ET and accounted for heterogeneity in plant size, light conditions, and microtopography, (2) an empirical estimate of ET based on changes in soil moisture at a single location, and (3) a set of six conventional ET models that did not account for spatial heterogeneity. We further evaluated three methods of adjusting conventionally-modeled ET estimates to better align with those from the modeling approach based on chamber flux data. Our empirically-based model indicated that basin-scale, daily ET ranged from 0-6 mm d^{-1}, with temporal variation depending on weather conditions and time of year. A sensitivity analysis demonstrated that the spatial composition of plant height and shade strongly influenced basin-scale estimates. The soil moisture-based method found daily values to range from 0-4 mm d^{-1}, which more closely matched empirically based estimates for the sensor location than basin-scale estimates. Most conventional models overpredicted ET on average, though three models (Granger-Gray, Hargreaves-Samani, and Matt-Shuttleworth) were less sensitive to variation in atmospheric conditions and thus overpredicted ET at the low to middle part of the range but underpredicted ET at the upper end of the range. The muted response to atmospheric conditions limited the ability of additive or multiplicative adjustments (i.e., landscape coefficients) to improve agreement. In contrast, additive and multiplicative adjustments, as well as corrections

accounting for shade, substantially improved agreement for the three models more sensitive to atmospheric conditions (Penman-Monteith ASCE, Penman-Monteith FAO, and Priestley-Taylor), with the strongest agreement resulting from additive adjustment of Penman-Monteith-derived estimates. Further analyses indicated that discrepancies between Penman-Monteith and empirical estimates could be attributed, in part, to differences in their treatment of wind speed, whereas accounting for soil-water limitation had little effect on these discrepancies. Our results highlight the importance of implicitly or explicitly accounting for spatial heterogeneity when quantifying ET, especially with respect to vegetation height and shade. For basins similar to our focal basin, this can be accomplished through the provided adjustments to conventional models. Additional calibration is required otherwise, but the growing availability of required data makes this increasingly viable.

1 Introduction

Green stormwater infrastructure (GSI) is a widely-used approach for managing stormwater in urban areas globally (McPhillips and Matsler, 2018; Li et al., 2019). Bioretention systems, such as rain gardens and bioswales, are a common type of GSI that incorporate both plants and soil media. These basins perform an array of hydrological functions such as allowing stormwater to infiltrate into the subsurface or return to the atmosphere via evapotranspiration (ET); these processes reduce peak flows as well as the volume of stormwater entering municipal stormwater systems (Pennino et al., 2016; Daniels and Yeakley, 2024). Reducing stormwater inputs to combined sewer systems can ultimately reduce the volume of combined sewer overflows into receiving water bodies (Hung and Harman, 2020; Joshi et al., 2021). An important contrast between infiltration and ET is that infiltration occurs predominantly during and immediately following storm events, whereas ET is much slower and acts throughout the period between events (Hess et al., 2021). As a result, ET more strongly determines if a basin's soil media is able to return to baseline moisture levels before the subsequent event (Johnston et al., 2020; Nasrollahpour et al., 2022).

The overall importance of ET to the removal of stormwater in bioretention basins is highly variable. In their review of ET in GSI, Ebrahimian et al. (2019) found that studies reported ET removing 19-84% of influent stormwater, though a modeling-based analysis estimated ET below 5% (Brown and Hunt, 2011), widening the range further. More recently, Huang et al. (2025) reported stormwater removal via ET ranging from near zero to almost 90% across bioretention systems. Although this broad range encompasses systems operating under widely differing conditions, some of the variation can be attributed to well-established effects of climate, hydrology, vegetation, and system design. For example, removal via ET tends to decline with increasing hydraulic loading ratio (Huang et al., 2025) and also varies with vegetation and media characteristics (Nasrollahpour et al., 2022). However, the factors underlying this variation are often inadequately represented in ET estimates, particularly when heterogeneity occurs within individual bioretention systems. When bioretention basins are designed or evaluated (e.g., for regulatory compliance), ET is typically either disregarded or estimated using models (Traver and Ebrahimian, 2017). While modeling approaches are presumably less error-prone than ignoring ET altogether, conventional models were designed to inform agricultural irrigation practices and large-scale hydrologic budgets (Shuttleworth, 2007). They may therefore be inaccurate when used in urban settings with notably high heterogeneity in light, (micro)topography, and vegetation, all of which are fundamental to the energy and mass transfer equations on which ET models are based (Xue et al., 2025). Notably, standard implementations of energy-based models do not account for variation in plant size, which can be considerable at the scale of a bioretention basin. Depending on plant size and structure, the total leaf area overlying a portion of the basin's ground surface (i.e., leaf area index; LAI) may span several orders of magnitude. Although methods exist to help circumvent this limitation, such as multiplying by a landscape coefficient (K_L; Costello et al., 2000), they have not been evaluated in bioretention settings. A further challenge is that the most commonly-used models omit reductions in ET imposed by soil water limitation (McMahon et al., 2013), which are likely substantial in bioretention basins due to their fast-draining soil media, especially for periods between storms and in portions of basins with higher topographic positions. Solar radiation can also be variable at the basin-scale in densely urbanized environments, where buildings and other structures can cast significant shade. Given that weather and plant function drive ET, approaches that can accommodate the fine temporal scale resolution of ET may offer distinct advantages, though they often require larger quantities of input data (Ebrahimian et al., 2019; Skorobogatov et al., 2020).

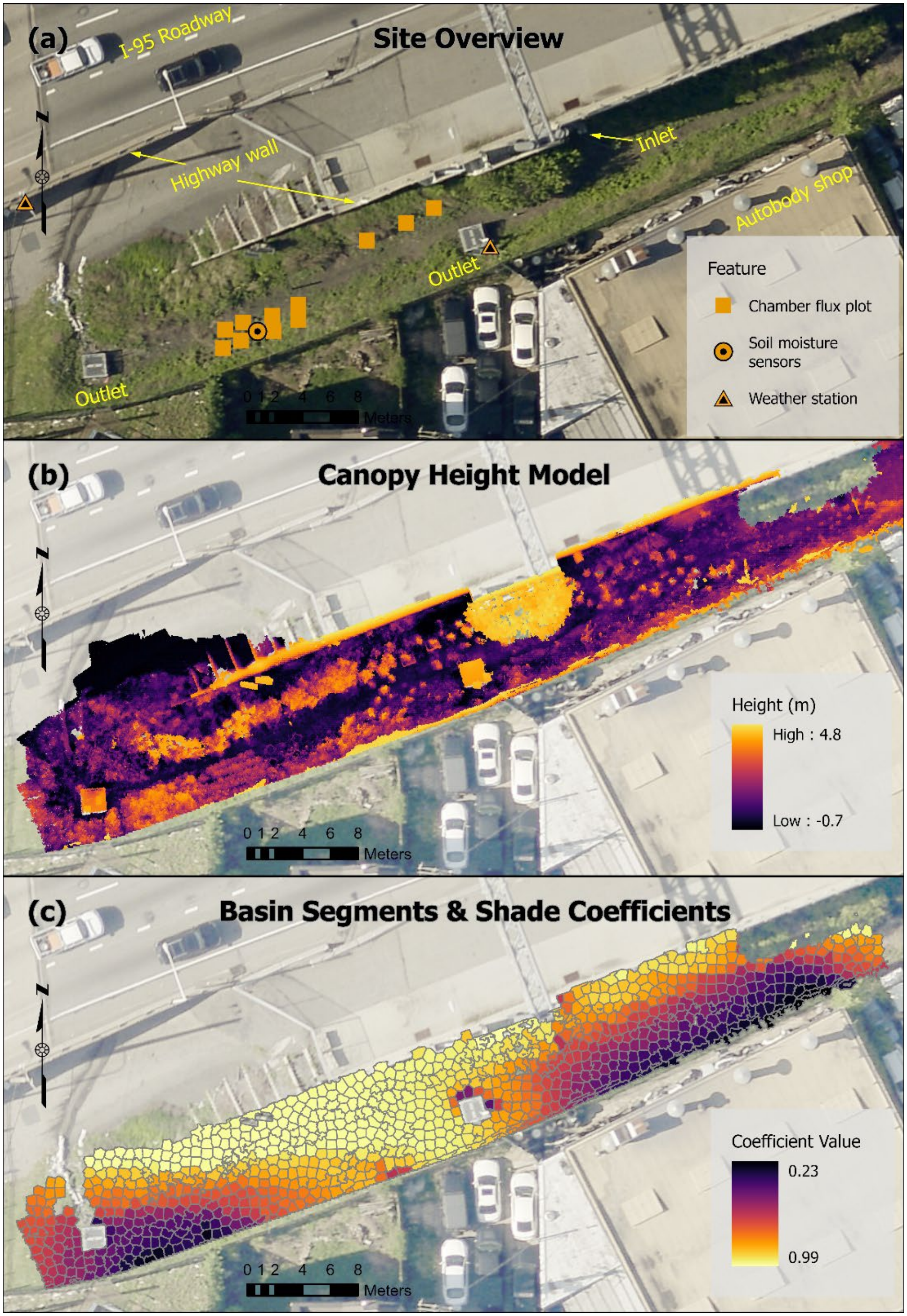
Greater understanding of the accuracy and limitations of model-based ET estimates for use in bioretention basins would ideally come through comparison with measured or empirically-based estimates of ET. Empirical methods for quantifying ET exist but are rarely used in bioretention systems, either for research or in practice. This can be attributed to the fact that ET measurement systems require specialized equipment and can be labor intensive to operate. It is also possible to estimate ET from a water budget in which all other major fluxes are known (inflow, infiltration, and surface outflow; Li et al., 2009). However, methods such as static chambers, weighing lysimeters, and sap flow techniques are advantageous because they directly measure ET (or transpiration alone, in the case of sap flow) at fine spatial scales (Kool et al., 2014; Hamel et al., 2015). There are several examples of researchers applying these techniques to bioretention basins and other GSI (e.g., Denich and Bradford, 2010; Hess et al., 2017; Scharenbroch et al., 2016). However, few studies, if any, have examined the spatial and temporal variation of ET to evaluate the ability of conventional ET models to accurately estimate bioretention basin ET.

In this study, we applied plant physiological and geospatial techniques to a bioretention basin in Philadelphia, USA to answer two primary questions: (1) how can we incorporate the heterogeneous characteristics of bioretention basins into predictions of basin ET? and (2) how can we use our empirical observations to improve the accuracy of ET models that might be used when designing and assessing bioretention basins? To address the first question, we collected ET measurements across a gradient of zones within the basin over the course of a growing season, characterized the distribution of plants, shading, and topography in the basin, and modeled ET through the growing season using the empirical data. We also assessed the sensitivity of this empirical model to several input parameters to examine the implication of excluding or oversimplifying spatial variability in bioretention ET modeling. For the second question, we compared our results to conventional ET estimation methods (e.g., Penman-Monteith) and calculated coefficients to tailor those estimates to bioretention basins. Empirically-based ET models tailored for bioretention basins can both be used by designers to better apportion credit for ET-based volume reduction and adjust design choices such as the location of larger versus smaller plants to maximize ET.

2 Materials and Methods

2.1 Study site

This study focused on a bioretention basin in Philadelphia, Pennsylvania, USA (39.96662° N, 75.13404° W). It is designated GR2 SMP A (hereafter, BASIN A) and was constructed in 2016 as part of an effort to reduce combined sewer overflows into the Delaware River. It is situated in a mixed-use neighborhood (residential and commercial) but is adjacent to an elevated portion of the Interstate 95 highway (I-95). The annual average daily traffic volume of this portion of I-95 is ~198,000 vehicles per day (PennDOT, 2022). An approximately 2695 m^2 portion of the roadway and shoulder drains into the basin via two inlet pipes, though some runoff also enters via a sloped structure between two segments of the highway wall (Fig. 1a). The basin is approximately linear (148 m in length and 5 m in width) and has a trapezoidal cross section (~1.5 m depth). The middle 44 m of the basin's length contains a rock-filled gabion that causes the basin sections at either end to remain hydrologically separated except during extremely large storms. For the purposes of this study, we focused only on the portion of the basin to the southwest of the gabion (Fig. 1a). During construction, urban fill was replaced with an engineered soil medium with a gravelly loamy fine sand texture (as per USDA) to a depth of approximately 0.5 m. Prior to planting, the basin's flanks (i.e., sloped sides) were covered with landscape fabric. Vegetation is predominantly composed of graminoids, herbaceous perennials, and shrubs. Annual and perennial weeds are also present but are removed several times during each growing season. Most vegetation reaches <1 m in height when mature but some plants are taller, including a cluster of shrubs growing along the highway wall near the inlet (~4 m; Fig. 1b). The vast majority of plants are deciduous and therefore leafless from late fall through early spring. Additional site characteristics were provided by Ampomah et al. (2023), Caplan et al. (2024), Pope et al. (2025), and Shakya et al. (2023).



125 **Figure 1. (a) Aerial view of the lower portion of Basin A, which served as the focal site for this study. Symbols indicate where ET, weather, and soil moisture data were measured. (b) Distribution of plant canopy heights; color scale is non-linear. (c) Distribution of shade coefficients for an example day (day of year = 180) at the segment scale. The aerial photograph was taken in 2022 by the City of Philadelphia.**

2.2 Chamber flux measurements

We measured ET rates at 11 fixed locations (i.e., plots) within the lower portion of BASIN A. Eight of the plots were established such that half were at high or low topographic positions (to capture soil moisture variability within the basin), while half from each group were either in a cluster of small stature vegetation or in a minimally-vegetated area (Fig. 1a). Three additional plots were in an area with larger vegetation. The eight plots containing small or large vegetation were centered on a single, dominant plant (Table 1, Fig. 1a, Fig. S1), though small weeds were also typically present. The minimally vegetated plots contained bare ground along with small weeds.

Table 1. Characteristics of the plots in the study basin that were used for ET measurements

Plant Stature	Primary Taxon	Number in Elevation Class	
		High	Low
Large	*Calamagrostis ×acutiflora* 'Karl Foerster' (feather reed grass)	0	2
Large	*Itea virginica* Little Henry® (Virginia sweetspire)	1	0
Small	*Hemerocallis* 'Happy returns' (daylily)	2	2
Minimal	None	2	2

Atmospheric conditions and soil moisture were intentionally variable across the seven measurement dates, which spanned June through October 2019. Plants were full-sized through this period, eliminating the influence of seasonal canopy size variation on our data. In particular, we targeted several dates with little antecedent rainfall and other dates that followed storm events. All measurement dates were rain-free to ensure intercepted rainwater evaporating from leaf surfaces would not confound ET measurements. Air temperature and humidity also varied considerably across the growing season, though we also captured variation in these parameters by taking measurements multiple times through each day (see below).

ET measurements were made using the closed dynamic chamber method, in which all vegetation and soil within a plot are temporarily enclosed in a fixed-volume chamber to measure the accumulation of water vapor. Although more commonly used to measure carbon fluxes (e.g., Bansal et al., 2023; Geoghegan et al., 2018; Norman et al., 1997), the method has been successfully applied to water flux measurements as well (Garcia et al., 2008; Hamel et al., 2015; Luo et al., 2018). Custom-made collars were set into the soil at the beginning of the field season, demarcating plot boundaries (Fig. 1a). The upper faces of collars were square and approximately parallel to the basin's bottom. Plots with minimal and small-stature vegetation had collars with upper surfaces of the same dimensions (52 × 52 cm; n = 8), while those used for larger plants were substantially greater (92 × 92 cm; n = 3). We measured ET at each plot five to six times per day, typically between 800 and 1900 h, at approximately 90-minute intervals. For each measurement, we first attached a transparent chamber and lid to the abovementioned collar (Fig. S1). We then measured water vapor in the enclosure once per second using an ultra-portable greenhouse gas analyzer (UGGA; Los Gatos Research, San Jose, CA, USA) that was connected to the chamber using a pair of flexible tubes (4.3 m long). Calculating ET required isolating the linear portions of time series recorded while chambers were closed and performing linear regression to determine the rate of water vapor accumulation. After excluding cases with insufficient data quality (~17%), our dataset contained 302 usable ET measurements. Additional details on chamber flux measurements and associated calculations are provided in Supplement 1.

2.3 Environmental data

Weather data were measured at two locations (Fig. 1a). One was elevated above the highway and provided air pressure and total solar radiation for the entire study period at 5-minute intervals. The second was located within BASIN A and provided an incomplete set of temperature, relative humidity, and wind speed data, also at 5-min intervals. To infill missing temperature and relative humidity data, we obtained estimated values for BASIN A from the DarkSky weather data service (www.darksky.net), which uses spatial interpolation to generate hourly time series at a specified location. DarkSky values were interpolated to the 5-min interval using the Stineman (1980) method (via R library *imputeTS*). The resulting time series were well-correlated with the partial datasets measured in the basin itself ($R^2_{Temp} = 0.98$, $R^2_{Rh} = 0.95$).

Volumetric soil water content (VWC) was measured using TEROS 12 sensors (Meter Environment, Pullman, WA, USA) installed on the south flank of BASIN A, approximately 40 cm above the basin floor (Fig. 1a). The sensors were a subset of those used in a separate study (Pope et al., 2025) and were within 1 m of each other. Sensors were positioned at 5, 10, 30, and 60 cm depth, with the latter three arranged in a vertical profile. For empirical modeling of ET (see Sect. 2.5), we used the mean VWC from 5- and 10-cm depths since these measurements most directly depicted near-surface moisture conditions. Missing values in the 10-cm time series were first interpolated from concurrent measurements recorded by nearby sensors, yielding complete data for 137 days (8 June through 22 October). The sensors at 10, 30, and 60 cm provided a soil moisture time series spanning the soil profile and were used in analyses requiring data from multiple depths (see Sects. 2.7 and 2.8). Profile data were missing for three of the days included in the empirical-model time series as they could not be accurately interpolated (n = 134 days).

Because the air's influence on ET derives from the combined effects of air temperature and relative humidity, we computed vapor pressure deficit (VPD) to describe evaporative demand with a single metric. The calculation of VPD was made using $VPD = VP_{sat}\ (1\text{-}Rh)$, where VP_{sat} is the saturated vapor pressure of the air (a function of its temperature; we used the equation provided by Wagner and Pruss, 1993) and Rh is the relative humidity.

2.4 Basin spatial characterization

We characterized the basin's topography, as well as its spatial heterogeneity in vegetation height, using a combination of terrestrial LiDAR and drone-based photogrammetry. First, we generated a 3D point cloud with Trimble TX5 3D terrestrial laser scanner (Trimble Inc., Sunnyvale, CA, USA), which was georeferenced and edited to yield a bare earth cloud model;

this was subsequently converted to a grid (the final DEM). In addition, we generated a digital surface model (DSM) from drone-based photographs using structure-from-motion (methods followed Alonzo et al., 2020). Vegetation height was determined by subtracting the DEM from the DSM (Fig. 1b). The DEM data further enabled us to create a set of polygons demarcating topographic position within the basin; we classified areas as either "lower" (the bottom where stormwater was intended to flow and infiltrate) or "upper" (the basin's flanks). Additional details are provided in Supplement 1.

We characterized spatial and temporal variation in solar radiation by adjusting the empirical time series of unshaded irradiance (from the sensor above the highway) with correction factors that accounted for how the site was shaded by surrounding features. The primary tool used to generate these correction factors (which we call shade coefficients) was the Area Solar Radiation calculator in ArcMap 10.8 (ESRI, Redlands, CA, USA). This tool uses a DEM as input to quantify solar radiation for each pixel at hourly or coarser intervals. Because we were interested in the shading effects of features such as the highway wall, nearby buildings, and a billboard, we generated a DEM from an aerial LiDAR dataset collected over the Philadelphia region in 2018 (City of Philadelphia 2018); it had a 30 cm resolution. We worked with data covering an area large enough to include all features that could potentially shade the study site (238 × 167 m). For each pixel in this DEM, the Area Solar Radiation calculator determined total direct and diffuse solar radiation for each day of the study period but did so in the absence of clouds. These were used to compute shade coefficients ($C_{p,t}$) for each pixel (p) on each day (t) throughout the study period:

$$C_{p,t} = \frac{Direct_{p,t} + Diffuse_{p,t}}{Direct_{u,t} + Diffuse_{p,t}}, \tag{1}$$

where u indicates the pixel containing the unshaded sensor. Shade coefficients could reach as high as one if there was no shading from nearby structures but typically ranged from 0.75 to 0.95; they almost never fell below 0.3 due to diffuse radiation (Fig. 1c). The matrix of shade coefficients was multiplied by the empirical time series from the unshaded sensor to yield location-specific daily time series of solar radiation that included the shading effects of both clouds and structures surrounding the site.

In order to make the computation of ET tractable across BASIN A, we used object-oriented image analysis to aggregate predictor variables to roughly the plant scale (Blaschke, 2010). This process segmented the basin into polygons (n = 967) with relatively homogenous vegetation height, specifically using the simple linear iterative clustering algorithm (Achanta et al., 2012; details in Supplement 1). We then aggregated pixel-scale values of the shade coefficient (Fig. 1c), topographic position, and canopy height to the segment scale using the zonal median; segment-scale values were used for all subsequent analyses.

2.5 Empirical model of ET

We developed a statistical model of ET to make it possible to estimate ET while incorporating temporal and/or spatial variation in plant height and the environmental conditions described above. The response variable was ET as measured by chamber flux; high-quality data from all 11 plots and 7 dates were included (n = 302). Predictor variables included topographic position (categorical, where 0 = lower basin and 1 = upper basin) and the following continuous variables: plant height, vapor pressure deficit, solar radiation, and mean soil moisture from 5- and 10-cm depths. Values of these variables corresponding to the locations and times of chamber flux measurements were extracted from larger datasets for use in the statistical model. The model also included all possible two- and three-way interaction terms among the five predictor variables as well as random intercepts for plot identity, measurement date, and their interaction. Including topographic position as a binary variable allowed the influence of continuous variables to differ between the basin's bottom vs. flanks through interaction terms; this was expected to be particularly important for soil moisture. Higher-order interaction terms were not included in the final model since they had poor explanatory power and would have been nearly uninterpretable. A square-root transformation was applied to the response variable (ET) to improve the normality and homoscedasticity of residuals (Fig. S2a,b).

We fit the model using restricted maximum likelihood estimation. This was done with functions from the *lme4* library (Bates et al., 2015) in R v4.4.2 (R Project for Statistical Computing, Vienna, Austria). Continuous predictor variables were standardized (centered and divided by their standard deviations) prior to model fitting, allowing the magnitudes of the resulting coefficients (β) to indicate the relative effect sizes of each term. Model fit was assessed using marginal and conditional coefficients of determination (R^2_m and R^2_c, respectively; Nakagawa et al., 2017). In this context, R^2_m represents variance explained only by the fixed effects while R^2_c represents variance explained by the fixed and random effects together. The difference between R^2_m and R^2_c therefore reflects the variation accounted for by random effects. Type II Wald χ^2 tests were used to determine which coefficients were statistically different from zero (using the *car* package; Fox and Weisberg, 2019). We focused our interpretation on significant terms ($p < 0.05$) with relatively large effects, although we retained all terms in the model to preserve model hierarchy.

We used the parameterized ET model to predict ET for each basin segment; these had unique topographic positions, plant heights, and shade coefficients but all followed the same time series of solar radiation, vapor pressure deficit, and soil water content. ET was estimated at 5-minute intervals at the segment level and then summed for each day. Predictions spanned 8 June to 22 October (n = 137 days), corresponding to the period with complete near-surface soil moisture data. Daily ET for each segment was multiplied by segment area, summed across all segments, and divided by total basin area to obtain daily, basin-scale ET. Basin-scale ET was converted from mmol m^{-2} d^{-1} to mm d^{-1} using the molar mass and density of water (18 g mol^{-1} and 1 g cm^{-3}, respectively). Night was excluded since our empirical model was parameterized only under daylight conditions; this means that values are slightly lower than they would be if ET estimates spanned 24-hr periods (likely ~5%; Groh et al., 2019).

2.6 Sensitivity analysis

We assessed the importance of accounting for spatial heterogeneity when determining basin-scale ET by comparing initial model estimates to those from additional calculations that did not account for it. Specifically, we determined the sensitivity of the whole-basin empirical model to plant height, topographic position, and the shade coefficient, calculating basin-scale ET under eight scenarios in which each of those variables took on fixed values (Table 2). We then summed daily ET values across the study period for each scenario and used total ET through the 137-day focal period as the basis of comparison.

Table 2. Variables used to test the sensitivity of the whole-basin empirical model

Variable	Scenario	Values Used
Vegetation height	None	Plant height = 0 cm
	Mean Ht	Plant height = 36 cm (area-weighted mean)
	100 cm	Plant height = 100 cm
Topographic position	Lower	All basin segments assigned to lower position (0 in regression model)
	Upper	All basin segments assigned to upper position (1 in regression model)
Shade	Coef = 0.25	25% of incoming solar radiation
	Coef = 0.50	50% of incoming solar radiation
	Coef = 1.00	100% of incoming solar radiation

2.7 Comparison to soil moisture-based estimates

To provide an empirical point of comparison for our ET model, we calculated ET from a soil moisture time series. This is a relatively simple method previously developed for estimating ET in bioretention systems (Hess et al., 2021), though it does not account for spatial variability unless measurements are made at multiple locations. Data came from the set of soil moisture sensors at 10, 30, and 60 cm depth. For each storm event during the study period, we first identified the period of rapid decline in soil moisture values corresponding to gravimetric flow draining the soil. We isolated data after these periods, namely when soil moisture was held by capillary tension and continued to decline, though at a slower rate. This phase of soil moisture loss can be attributed exclusively to ET in most cases (Bowman and King, 1965; Hess et al., 2021). To calculate

ET, we assumed that volumetric water content in the top 10 cm equaled the sensor reading at 10 cm, and linearly interpolated between the other sensor depths to estimate total water content in the upper 60 cm. We then calculated the rate of ET as the net change in total water content over time, isolating each day during the study period with complete data (n = 68).

We compared soil moisture-based ET with both basin-scale empirical-model estimates and local empirical-model estimates for the four spatial segments near the soil moisture sensors. Daily ET rates for these segments were converted from mmol m^{-2} d^{-1} to mm d^{-1} as described above and the area-weighted average was computed. Agreement was quantified using concordance correlation coefficients (CCC; Lin, 1989). CCC evaluates not only how two datasets align with each other but how close they come to perfect agreement (i.e., a 1:1 line). Like other correlation coefficients, CCC ranges from -1 to 1, but here 1 indicates perfect concordance, 0 indicates no concordance, and -1 indicates perfect discordance. We additionally calculated root mean square errors (RMSE) and compared the paired absolute differences between soil-moisture-based ET and both local and basin-scale estimates using a Wilcoxon signed-rank test.

2.8 Comparison to conventional models

We estimated daily ET for the focal section of Basin A using meteorological data applied to six commonly-used ET models (i.e., conventional models) through the R library *evapotranspiration* (Guo et al., 2016; details in Table S1). These included models describing potential ET (Priestley-Taylor [PT]), reference ET (ET_0; Hargreaves-Samani [HS], Matt-Shuttleworth [MS], as well as both the FAO (Food and Agriculture Organization of the United Nations) and ASCE (American Society of Civil Engineers) variants of Penman-Monteith [PM]), and “actual” ET (Granger-Gray [GG]). Using these six conventional models, we estimated ET at a daily timescale from 8 June to 22 October. Table S1 summarizes the input variables and constants used in these model runs.

Our primary model estimates used the raw time series of solar radiation, with the implication that they did not include the influence of shade from nearby structures, unlike our empirical-modeling approach. We therefore additionally generated estimates using a solar radiation time series adjusted for the effect of shade from structures. This required calculating shade coefficients that could be applied to the basin as a whole for each day of the study period. To do this, we first calculated the area-weighted mean of shade coefficient values across all basin segments; this was done for each day of the study period. Then we multiplied the total solar radiation value for each day by the corresponding shade coefficient, yielding the adjusted time series. CCC values were the primary means by which we compared ET estimates from conventional models to those from our whole-basin empirical model.

We also evaluated whether wind speed or soil-water limitation contributed to differences between conventional and empirical ET estimates. For wind speed, we used linear regression to determine if the magnitude of these differences (i.e., bias) exhibited a systematic relationship with wind speed. If so, it would indicate that not explicitly accounting for variation in wind speed in our empirically-based model contributed to discrepancies with one or more of the conventional models. For soil-water limitation, we adjusted conventional modeling estimates of ET with an FAO-56-style water-stress coefficient (K_S; Allen et al., 1998) and evaluated the effects of incorporating soil moisture stress on bias (specifically, conventional minus empirical ET) and concordance between conventional and empirical ET estimates. K_S values were derived from the profile-based soil moisture time series and water retention characteristics (Shakya et al., 2023), using three plausible effective rooting depths (20, 30, and 40 cm; additional details in Supplement 1).

2.9 Landscape coefficients

Landscape coefficients (K_L) are used to adjust reference evapotranspiration rates (ET_0) so they better reflect ET in ornamental or landscaped settings (Costello et al., 2000). Like crop coefficients used in agricultural settings, K_L values are multiplicative and can be considered a form of bias correction. K_L has been determined traditionally by observing the type of species, planting density, and microclimate conditions at a site and applying guidelines from Costello et al. (2000). More recently, K_L values have been determined by comparing measurements of actual ET with reference ET (Sun et al., 2012), facilitating their determination for a wide array of landscaped settings and plant combinations.

We used the latter approach by comparing our daily-scale, basin-wide empirical ET calculations (ET_{emp}) with the ET estimates from the six conventional models described above (ET_{ref}). Specifically, we fit the following equation using a trust-region-reflective algorithm implemented via the Matlab function *fit* (Coleman and Li, 1994; Coleman et al., 1996):

$$ET_{emp} = K_L \times ET_{ref} \tag{2}$$

where ET_{emp} and ET_{ref} were both calculated at a daily interval using weather data from 8 June to 22 October. We then multiplied K_L by ET_{ref} to generate corrected, daily-scale ET estimates, which we compared to ET_{emp} using the CCC.

2.10 Additive correction factors

We also tested additive correction, a simple approach that can be used to adjust for model bias when it is consistent across a sufficiently wide range of values. It has been applied to estimates of various meteorological parameters, including ET (Paredes et al., 2018), although it has the potential to yield negative values when the correction exceeds the uncorrected estimate. Such values would have little influence on estimates aggregated over longer periods as long as they are small and infrequent. To determine the correction factors, we first regressed each set of ET_{ref} estimates against ET_{emp} and then subtracted the fitted intercept from each ET_{ref} estimate when the intercept was positive. As with the landscape coefficient adjustments, we used the CCC to assess how additive corrections improved ET_{ref}.

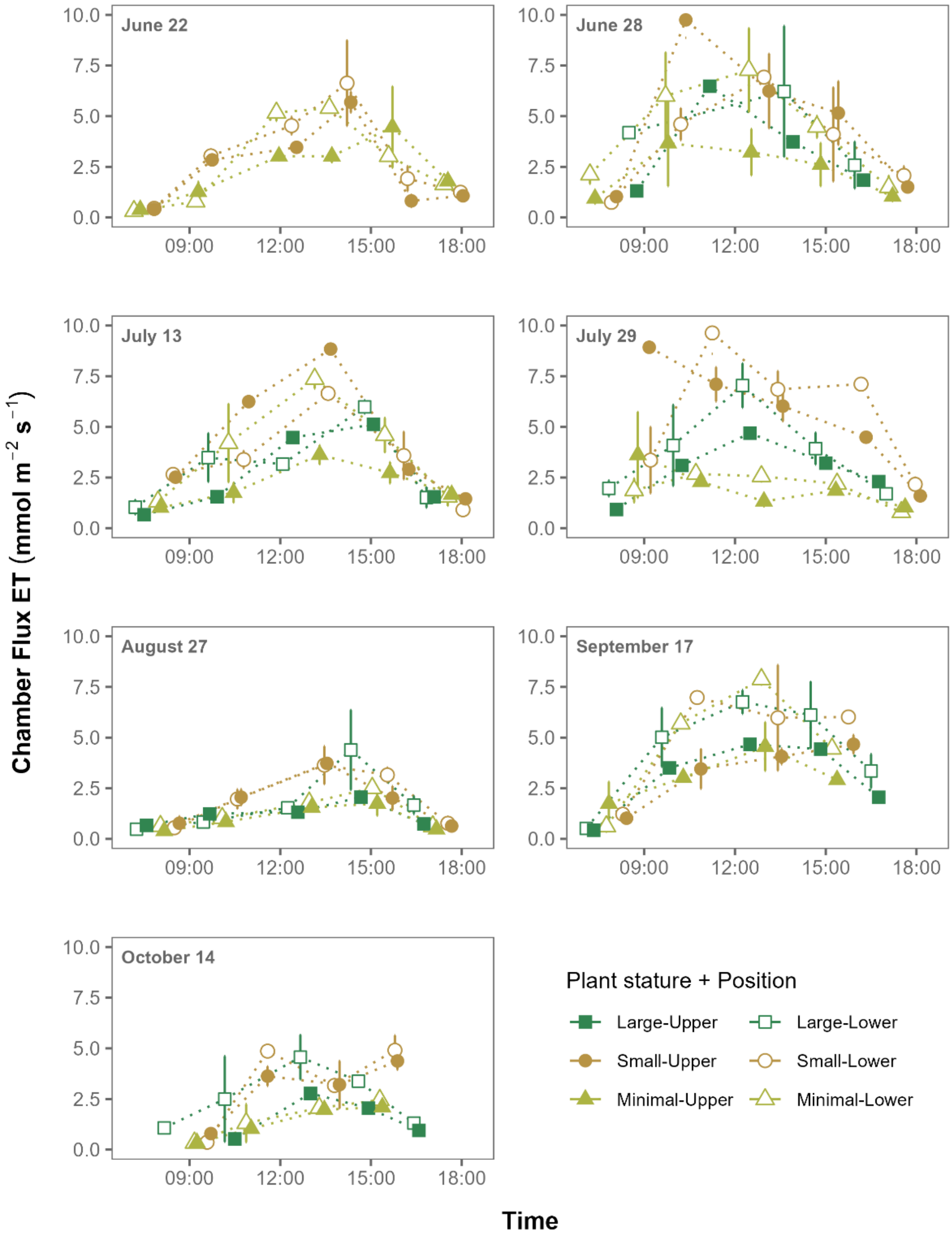
3 Results & Discussion

3.1 Empirical ET patterns and model

On all seven measurement dates, ET followed the expected diurnal pattern: it was near-zero in the early morning, peaked in the mid-afternoon, and declined in the evening (Fig. 2). However, the amplitude of this oscillation varied considerably from day to day. It was notably muted for all plots on 27 August, when temperature and VPD were elevated and there had been little antecedent precipitation (Fig. 3). The diurnal amplitude was also muted (though to a lesser degree) on 14 October, likely due to the dry September that preceded it, as well as temperature and solar radiation declining towards the end of the season.

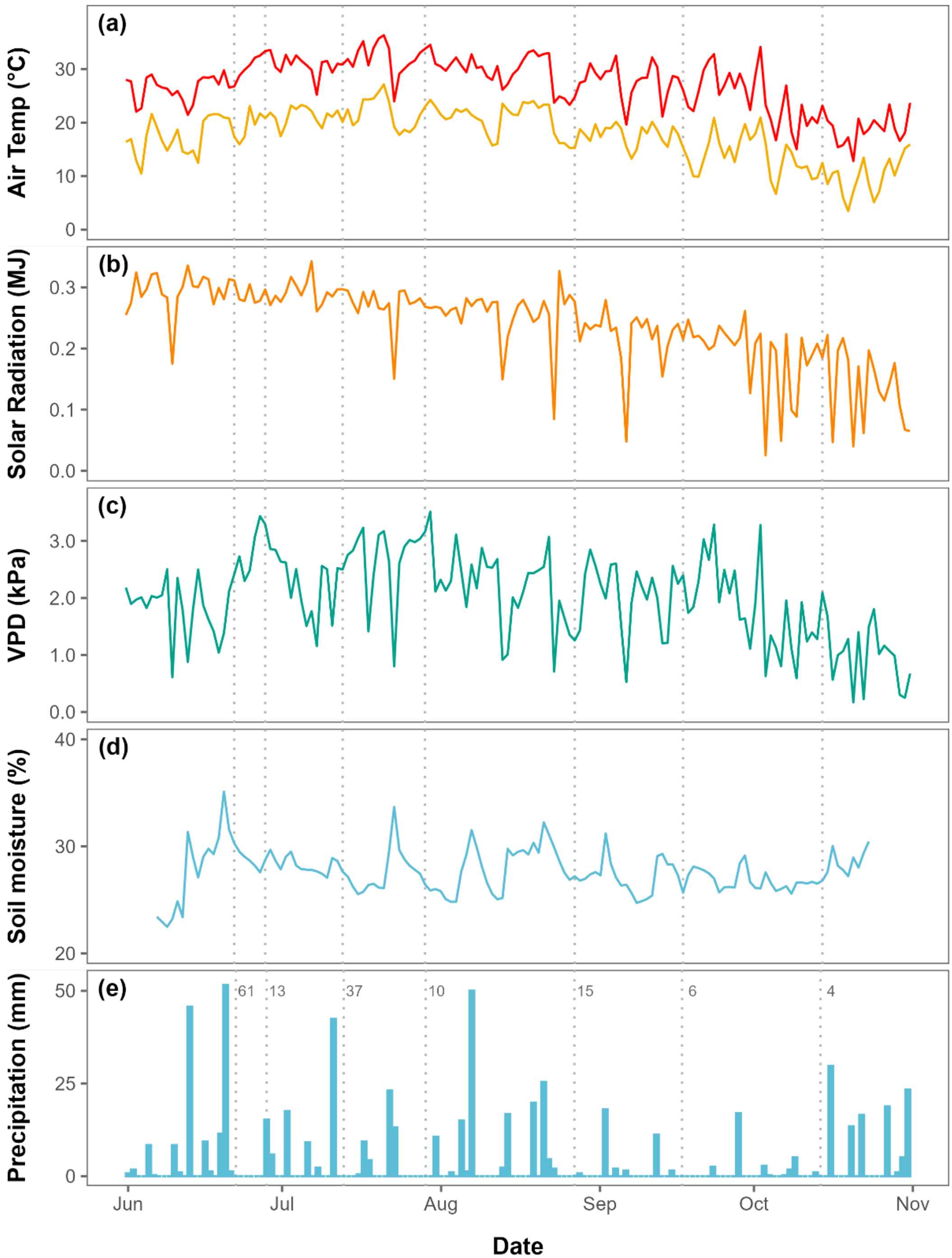
The statistical model was able to explain most of the observed variation in ET across measurement dates and plots (R^2_m = 0.65, R^2_c = 0.78; Fig. S2c,d), with its largest coefficients (β) realistically encapsulating environmental and plant processes (Table 3). For one, ET increased particularly strongly with solar radiation and VPD (β = 0.393 and β = 0.375, respectively). Also, ET was lower in the upper part of the basin (β = –0.158), likely because flooding by stormwater runoff was less frequent higher on basin flanks. Further, the effect of VPD on ET intensified when soil media was wetter (β = 0.181), reflecting the greater ability of evaporative demand to drive ET when unsaturated soil contains more water. Less intuitively, the effect of solar radiation on ET appeared to weaken when soils were wetter (β = –0.089), though this can be explained by reduced evaporative efficiency and/or plants limiting transpiration under (near-)saturated conditions. However, this relationship differed by topographic position (β = 0.112); on basin flanks, greater moisture enhanced the ability of solar radiation to promote ET, whereas in low-lying areas, greater moisture levels dampened this response, likely due to (near-)saturation restricting water loss from plants and soil.

ET was also greater in locations where plants were taller (β = 0.093), reflecting the ability of greater leaf cover per unit ground area (a.k.a. leaf area index, or LAI) to raise ET (Steduto and Hsiao, 1998). The fact that this effect was not larger was likely due to available energy constraining the extent to which greater LAI could increase ET, as well as larger plants shading the soil surface more completely, inducing a tradeoff between soil evaporation and transpiration. Another factor may have been the fact that larger plants tended to have small or no other plants around them, diminishing LAI at the plot level. Previous observations of similar leaf-level transpiration rates among the three focal species (Toran et al., 2017) suggest that differences in physiological rates had a minimal influence on the relationship between plant size and ET.



345

Figure 2. Plot-scale evapotranspiration rates in Basin A measured using flux chambers. Mean rates (±SE) for each combination of plant stature and position are shown. Points without error bars either represent single points or their SEs are hidden by symbols.



350 **Figure 3. Environmental conditions at the field site through the study period. Shown are daily maximum and minimum air temperatures, daily maximum solar radiation and vapor pressure deficit (VPD), mean soil moisture at 5 cm depth, and total precipitation. Vertical dotted lines indicate dates of ET measurements while numbers in (e) are antecedent precipitation index values (Kohler and Linsley, 1951) calculated using a 10-day window.**

Table 3. Coefficients (β) and Type II Wald χ^2 tests for the empirical model of ET from chamber flux measurements. The response variable (ET) was square-root transformed and continuous predictor variables were standardized prior to fitting; topographic position was coded as a binary variable and not standardized. Topo = topographic position, where 0 = lower basin and 1 = upper basin; Plant Size = plant height; VPD = vapor pressure deficit; Solar Rad = solar radiation; VWC = volumetric water content

Variable	β	Wald χ^2	p
Topo	-0.158	1.6	0.200
Plant Size	0.093	8.2	0.004
VPD	0.375	114.6	<0.001
Solar Rad	0.393	247.8	<0.001
VWC	0.060	0.2	0.637
Topo × Plant Size	0.068	0.5	0.464
Topo × VPD	-0.030	0.2	0.639
Topo × Solar Rad	-0.049	1.7	0.198
Topo × VWC	-0.025	<0.1	0.824
Plant Size × VPD	-0.023	0.5	0.481
Plant Size × Solar Rad	0.010	2.7	0.103
Plant Size × VWC	-0.025	3	0.083
VPD × Solar Rad	-0.019	0.3	0.570
VPD × VWC	0.181	25.3	<0.001
Solar Rad × VWC	-0.089	0.5	0.494
Topo × Plant Size × VPD	0.066	1.6	0.212
Topo × Plant Size × Solar Rad	0.072	2.1	0.145
Topo × Plant Size × VWC	-0.054	0.7	0.387
Topo × VPD × Solar Rad	0.029	0.3	0.556
Topo × VPD × VWC	-0.050	0.8	0.382
Topo × Solar Rad × VWC	0.112	5.4	0.020
Plant Size × VPD × Solar Rad	-0.020	0.7	0.405
Plant Size × VPD × VWC	-0.023	0.6	0.445
Plant Size × Solar Rad × VWC	-0.015	0.3	0.561
VPD × Solar Rad × VWC	0.044	1.6	0.213

3.2 Basin-scale empirical ET model

At the basin scale, peak ET exhibited the expected seasonal decline, progressing from about 6 mm d^{-1} in early-and mid-summer to around 1 mm d^{-1} by mid-autumn (Fig. 4a). However, ET was well below peak values on most days, likely due to limitations associated with weather (especially reduced solar radiation on cloudy days and reduced VPD on cool or humid days). The clearest spatial pattern was that ET was elevated in areas with the tallest plants. This was most notable for a cluster of shrubs adjacent to the highway wall (Fig. 4b). However, plants in other areas were relatively tall (i.e., > 100 cm), and although they exhibited modest ET, values were smaller than expected. This was likely due to the effects of shade from two buildings and a billboard immediately to the south of the basin. Given that the highway wall was to the north of the basin, its shading effects were minimal. There was also relatively little variation in ET associated with topographic position, likely because plants tended to be shorter in the low-lying portion of the basin but taller on the flanks and surrounding areas.

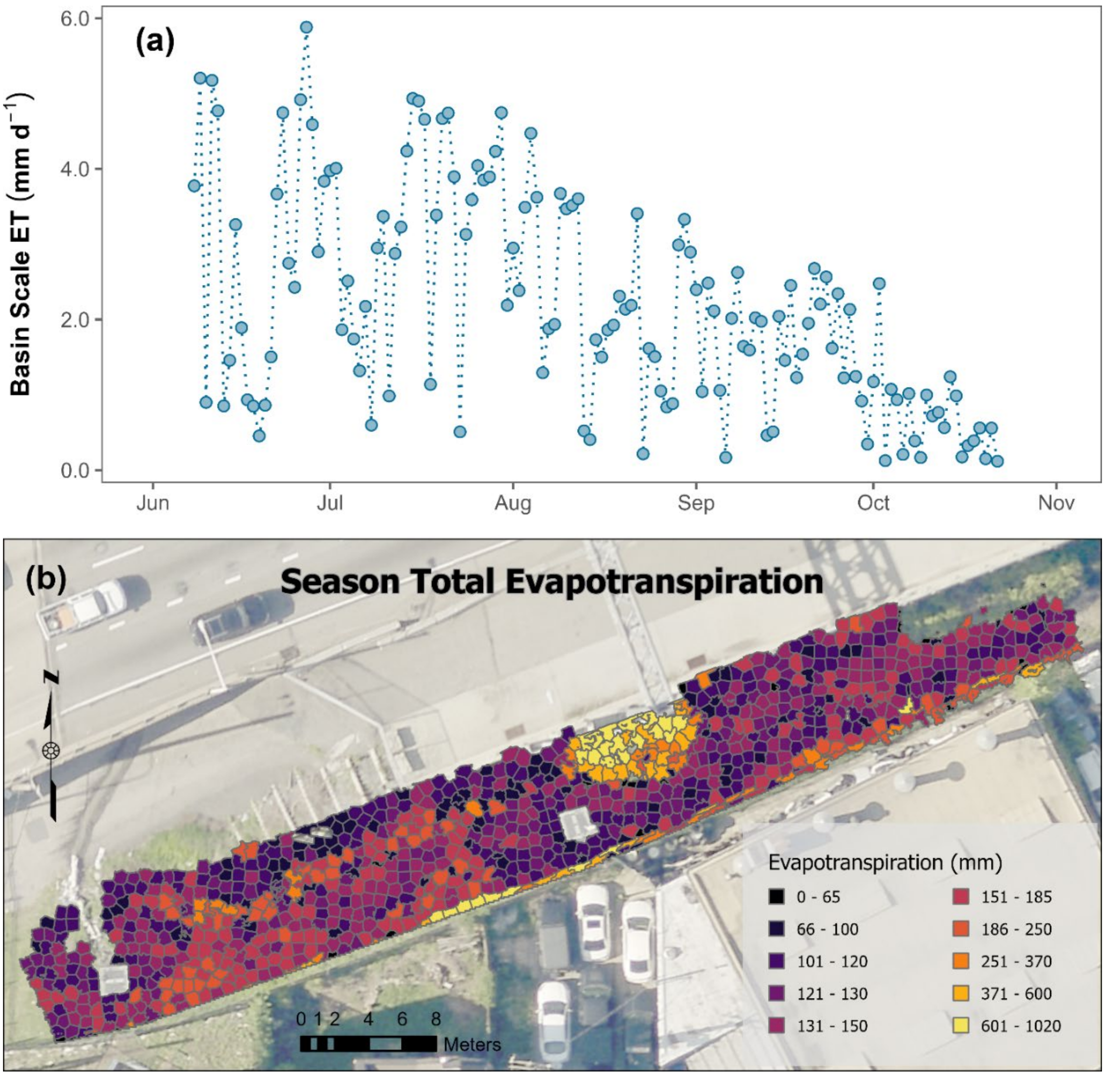


Figure 4. (a) Daily ET estimated by the empirically based model for the whole basin spanning the focal period (8 June to 22 October) and (b) season-wide ET estimated by the model for each basin segment. The aerial photograph was taken in 2022 by the City of Philadelphia.

3.3 Sensitivity analysis

Total ET through the study period, when calculated at the basin scale, was most sensitive to plant height. It fell well below the baseline (-41%) with all plant heights set to zero but rose to +60% with all plant heights set to 100 cm (Fig. 5). When we applied the area-weighted mean height to the entire basin, the season-wide estimate of ET was still reduced, at -13%. Treating all segments as having upper or lower topographic positions had little influence on seasonal ET, changing it by -3% and +0.5%, respectively (Fig. 5). In contrast, quantifying ET under the scenario that no structures shaded the basin yielded an 11% increase in season-wide ET. This decreased to -48% when we applied a shade coefficient of 0.25 (i.e., 25% of incoming solar radiation reaching the basin) and -32% with a shade coefficient of 0.5.

A study of ET in experimental bioretention cells (Nocco et al., 2016) likewise found that ET varied strongly with vegetation type, specifically with prairie plants (forbs and graminoids) > shrubs > turfgrass > bare soil. In that study, prairie plants had similar heights but a much higher density than shrubs (13.6 plants m^{-2} vs. 1.1 plants m^{-2}), likely meaning that LAI primarily drove the differences in ET.

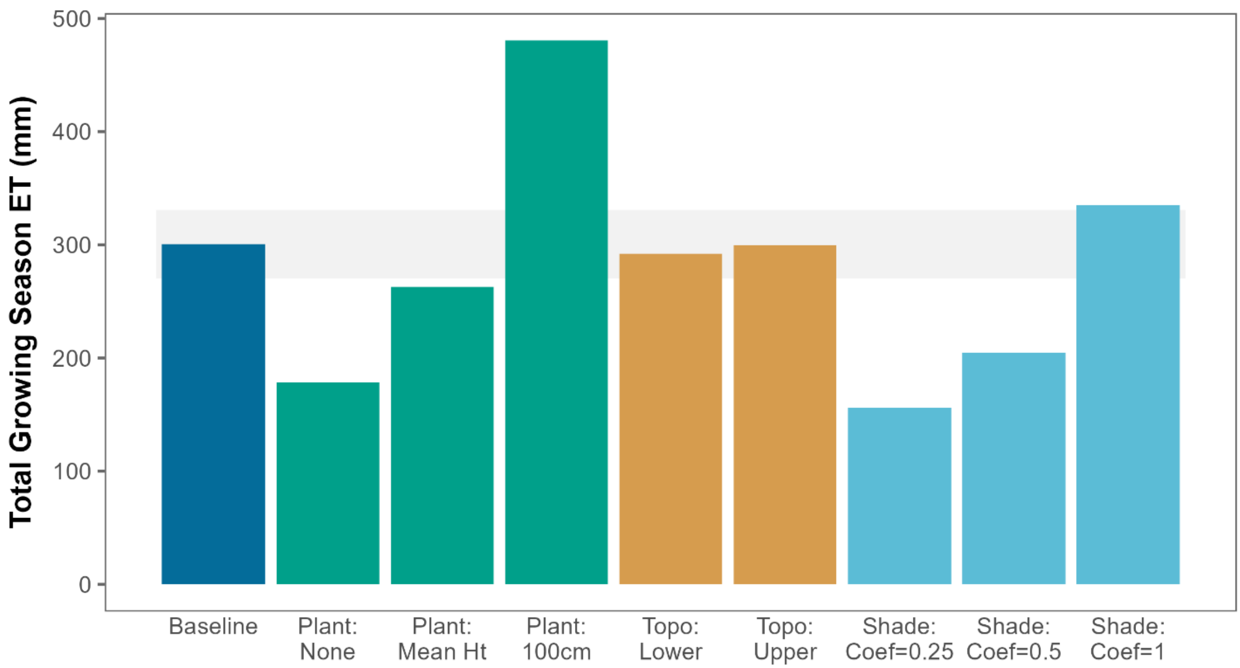


Figure 5. Comparison of whole-basin empirical ET estimates for the sensitivity scenarios described in Table 2. Total growing season ET includes all dates from 8 June through 22 October. The horizontal grey bar indicates ±10% of the baseline model results (original basin conditions). Plant, Topo, and Shade refer to scenarios in which all basin segments were assigned the same vegetation height, topographic position, and incoming solar radiation values, respectively (details in Table 2).

3.4 Comparison to soil moisture-based ET estimates

Soil moisture-based ET estimates agreed moderately well with basin-scale, empirically-modeled ET (CCC = 0.63; Fig. 6a). Further, the two approaches yielded ET rates that spanned a similar and reasonable range, suggesting that the empirical model produced reasonable estimates of ET. Nonetheless, when ET exceeded 2 mm d^{-1}, soil moisture-based ET often yielded lower ET rates. However, there were smaller differences between soil moisture-based ET and empirical-model estimates for the segments near the sensors (Fig. 6b). Although CCC was also 0.63, RMSE decreased from 1.15 to 0.92 mm d^{-1}, and absolute differences were significantly lower for the local estimates (Wilcoxon signed-rank test, $p = 0.009$).

While the soil moisture-based method has been validated using weighing lysimeters set up to mimic stormwater cells (Hess et al., 2021), comparisons among observational methods have demonstrated that ET quantification in GSI systems has substantial uncertainty (Ouédraogo et al., 2023). It is therefore not possible to definitively determine which approach more accurately reflected true, basin-scale ET. However, given that the soil moisture sensors used here were installed on the flank of the basin rather than the bottom, the soil moisture-based approach could be expected to underestimate basin-scale ET. The reduction in RMSE for the local empirical-model estimates supports this interpretation and provides additional evidence that the empirical model was capable of capturing spatial variation in ET within the basin. Given that the soil moisture-based approach can be influenced by the placement of the soil moisture sensors, especially in basins with high spatial heterogeneity, combining data from multiple sets of sensors would likely improve its accuracy considerably.

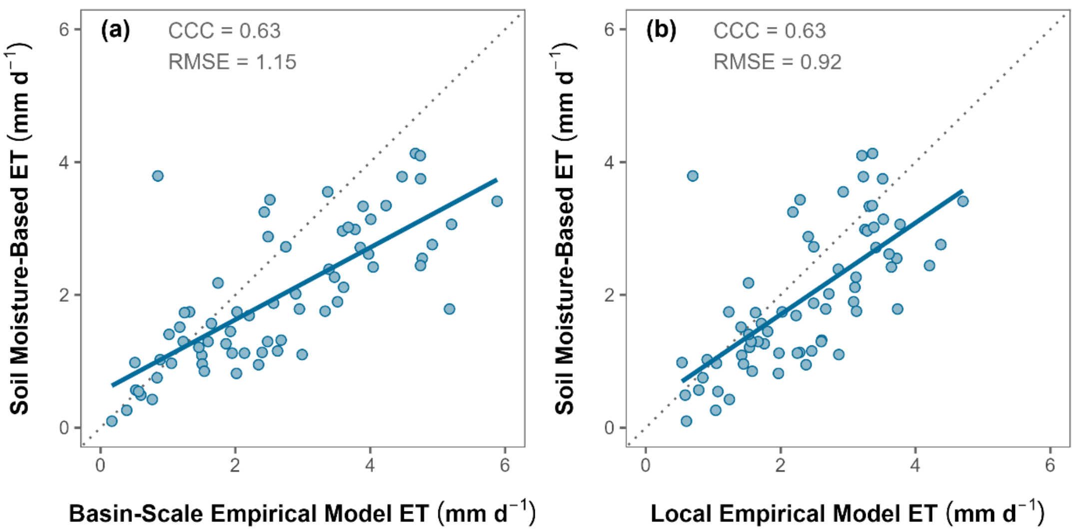


Figure 6. Daily soil moisture-based ET compared with empirical-model estimates (a) at the basin scale and (b) in the area near the soil moisture sensors. Dotted lines depict 1:1 relationships. CCC = concordance correlation coefficient; RMSE = root mean square error (*n* = 68).

3.5 Comparison to conventional ET models

In the absence of correction factors, there was moderately strong agreement between the basin-scale values generated by our empirical model and those from conventional ET models. The MS and GG models had the highest agreement (CCC_{orig} = 0.80 and 0.78, respectively; Fig. 7); however, both overestimated lower ET rates and underestimated higher ET rates. Further, the PM-ASCE, PM-FAO, and PT models consistently overestimated ET relative to our empirical model, leading to an offset that lowered CCC values (CCC_{orig} = 0.41, 0.57, and 0.60, respectively; Fig. 7). The HS model also overestimated ET and had the poorest agreement of the six models (CCC_{orig} = 0.30), although it produced estimates close to those from the empirical model for ET rates above 4 mm d^{-1}.

Differences in conventional model assumptions and parameterizations likely explain their variable performance. The MS equation was developed for arid and semi-arid climates, and the GG model explicitly assumes non-saturated soil (McMahon et al., 2013), which may explain why they underestimated relatively high ET rates. Similarly, the relatively weak sensitivity of HS to atmospheric conditions may explain why it overestimated ET at lower rates but underestimated ET at higher rates. However, model parameters that more closely matched conditions in BASIN A did not necessarily result in better performance. For example, the mean vegetation height in BASIN A was 36 cm, yet PM-FAO, which assumes a 12 cm plant height, had better agreement than PM-ASCE, which assumes a 50 cm plant height.

The strong linear relationships between our empirical ET estimates and those from PM models likely reflect their shared sensitivity to the atmospheric variables that control ET, particularly solar radiation and vapor pressure deficit. However, the two approaches use very different methods to account for surface and aerodynamic factors. The empirical model was parameterized directly from measurements within BASIN A and incorporated spatial variation in plant height, radiation, and topographic position explicitly, but other factors (like wind speed) implicitly. In contrast, the PM formulations account for vegetation and aerodynamic factors only via their inbuilt parameters but explicitly incorporate wind speed. Wind speed did appear to influence the discrepancy between PM and empirical ET estimates, as differences increased significantly with wind speed for both PM formulations (p < 0.001; Fig. S3). Further, while wind speed explained 39% and 12% of the variation in these differences for PM-ASCE and PM-FAO models, respectively, it explained ≤5% for the other conventional models, suggesting that the treatment of wind speed contributed to PM estimates being greater than the empirical estimates. Although wind speed was not an explicit predictor in the empirical model, airflow generated by the chamber mixing fans influenced the measured fluxes and was therefore implicitly represented in model calibration. However, the magnitude of this influence is uncertain because wind speed within chambers was not measured. Moreover, wind speed data used in the PM models may not have fully represented conditions within the basin, since surrounding and internal structures could have reduced and/or accelerated airflow.

Further, soil-water limitation explained only a modest amount of the discrepancy between conventional and empirical ET estimates. Across the range of effective rooting depths we evaluated (20–40 cm), mean daily water-stress coefficients (K_S) ranged from 0.969 to 0.987, indicating that meaningful water limitation was infrequent during the model-comparison period. There were periods when K_S fell below 0.75, but these accounted for only 6%, <1%, and 0% of days when using 20-, 30-, and 40-cm rooting depths, respectively. Nonetheless, incorporating K_S reduced the overestimation by the PM models by 5-7% (with a 30-cm rooting depth), primarily by decreasing ET at the upper end of its range, but it also reduced concordance between empirical ET and all six conventional models (Fig. S4). Thus, although not accounting for soil-water limitation contributed slightly to the higher ET estimates from the PM models, it is unlikely to explain their systematic offset from empirical ET. Other differences between the approaches, including their representation of vegetation characteristics and surface resistance, likely account for much of the remaining systematic offset.

Other researchers have similarly observed variability in the prediction capabilities of conventional ET models applied to bioretention basins. For example, PM-ASCE closely tracked measured ET in a mesocosm with unrestricted drainage but underestimated ET for a mesocosm with an internal water storage device (Hess et al., 2017). The PT model consistently underestimated ET in a bioretention mesocosm with mixed, short-stature vegetation (Wadzuk et al., 2015) and prairie plants (Nocco et al., 2016), whereas it was fairly accurate for experimental basins planted with shrubs or turfgrass (Nocco et al., 2016). Based on a review of these and other studies, Ebrahimian et al. (2019) recommend using the HS equation, arguing that, while it typically overestimates bioretention ET, this is outweighed by the fact that it only requires temperature and location data as inputs. Given that HS had the lowest CCC among the six models we evaluated, we recommend using other models if the required data are available, but we acknowledge that HS may be the only viable option if they are not.

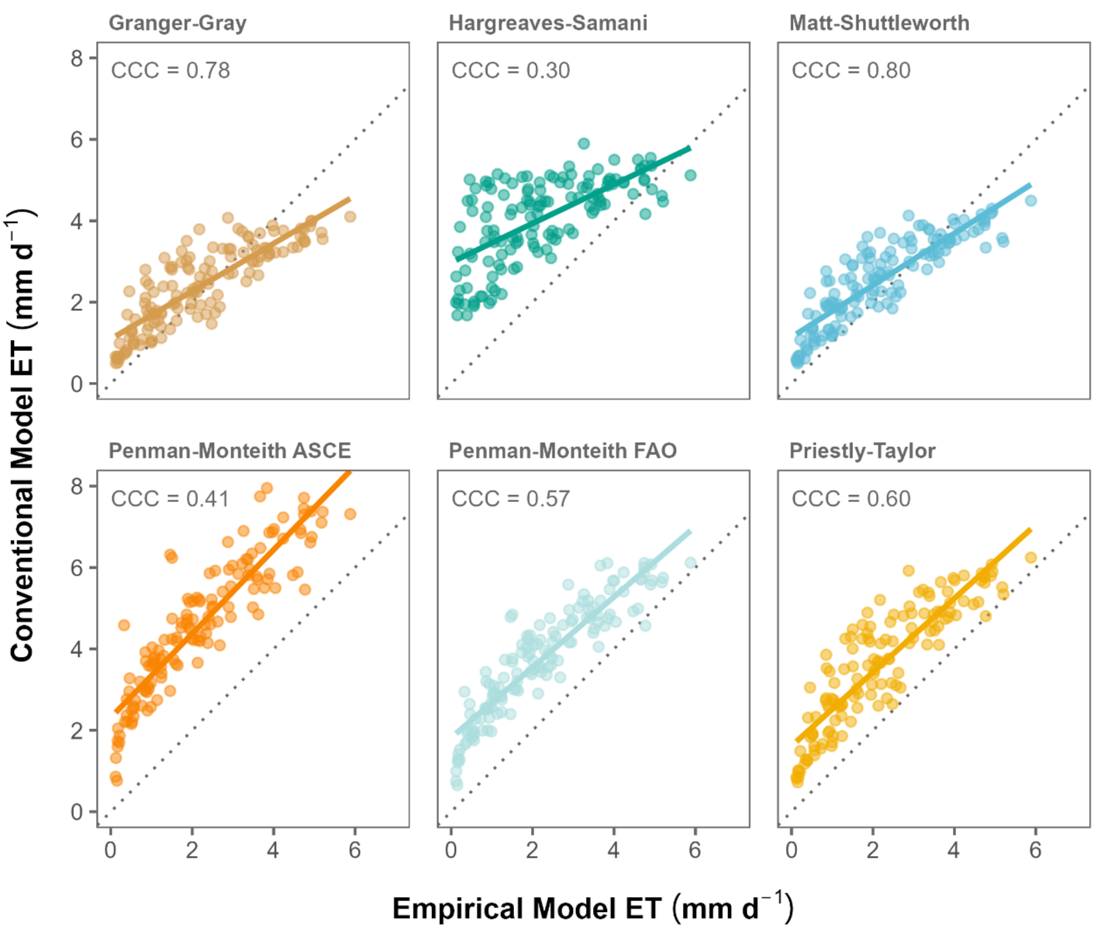


Figure 7. Daily ET calculated from our basin-scale, empirically-based model vs. ET from six conventional models. Dotted lines depict 1:1 relationships. CCC = concordance correlation coefficient.

3.6 Estimation using shade coefficients

Application of the shade coefficient to the input solar radiation data yielded improved agreement when using the PM-ASCE, PM-FAO, and PT models (CCC_{shade} = 0.44, 0.63, and 0.68, respectively) but made a negligible difference when using the other models (Fig. S5). This improvement was smaller than initially expected, likely because BASIN A was oriented east-west, with the highway wall to the north, meaning the sun was less often blocked by the wall than it would be at many other sites. Given that the data needed to generate shade coefficients are frequently available (i.e., high-resolution DEM, software to model direct and diffuse radiation, and a solar radiation sensor in an unshaded location), we recommend that shade coefficients be determined and used in ET modeling whenever possible, especially for basins subject to shading from tall structures.

3.7 Estimation using landscape coefficients

The landscape coefficients (K_L) we calculated (approximately 0.5 to 1.0; Table 4) were in the range of K_L values previously determined for park-like landscapes with mixtures of plant types (Sun et al., 2012; Pannkuk et al., 2010; Nouri et al., 2013). Applying these K_L values to the ET estimates from the majority of conventional models (PM-ASCE, PM-FAO, PT, and HS) improved agreement with those from our empirically-based approach (Table 4; Fig. S6). A mathematical byproduct of applying multiplicative corrections when initial estimates were greater than ours is that it made conventional models less responsive to changing environmental conditions.

The K_L correction yielded the greatest improvement in agreement for HS estimations (CCC_{KL} = 0.46), though this model far underperformed all other models even with the correction. MS and GG stood out in having relatively large K_L (>0.90), meaning corrections were smaller than for other models. The PM-ASCE-based K_L (0.52 ± 0.03) was slightly lower than the PM-FAO-based K_L (0.64 ± 0.03) (Table 4), corresponding to the greater plant height assumed by PM-ASCE (50 cm) vs. PM-FAO (12 cm).

The K_L approach to ET estimation is useful because it can be tailored to specific environments and conditions (e.g., Nouri et al., 2013). K_L values can be adjusted through the growing season to reflect shifts in plant size and the associated changes in the actual ET vs. ET_0 relationship through time (Pannkuk et al., 2010). However, the method may be difficult to apply to larger woody vegetation since the relationship between ET_0 and actual ET can be non-linear (Litvak et al., 2017). Nonetheless, our results indicate that K_L-based adjustments can provide reasonable estimates of actual ET in bioretention basins when used in basins with small- to medium-stature woody vegetation. This approach was particularly effective for models whose ET_0 estimates had approximately linear relationships with our empirically-determined ET values, such as PM-ASCE and PM-FAO.

Table 4. Landscape coefficients (K_L), 95% confidence intervals (CI), and concordance correlation coefficients (CCC) for the six conventional ET models. CCC were calculated in comparison to the whole-basin empirical model. Original = no adjustment to conventional model. Shade = model included shade coefficient adjusted solar radiation. K_L = model correction based on landscape

coefficient. Additive = model correction based on linear regression intercept. Numbers in bold highlight the highest concordance correlation coefficient for each model

Model	**CCC_{orig}**	**CCC_{shade}**	**CCC_{K_L}**	**CCC_{add}**	**K_L**	**K_L 95% CI**	**Additive Correction**	**Additive 95% CI**
Granger-Gray [GG]	**0.78**	0.77	**0.78**	0.62	0.97	0.91-1.02	1.10	0.9-1.3
Hargreaves-Samani [HS]	0.3	0.3	**0.46**	0.44	0.57	0.52-0.62	2.98	2.7-3.2
Matt-Shuttleworth [MS]	**0.8**	**0.8**	**0.8**	0.69	0.91	0.87-0.96	1.14	1.0-1.3
Penman-Monteith ASCE [PM-ASCE]	0.41	0.44	0.78	**0.89**	0.52	0.49-0.54	2.31	2.1-2.5
Penman-Monteith FAO [PM-FAO]	0.57	0.63	0.8	**0.89**	0.64	0.60-0.67	1.82	1.6-2.0
Priestley-Taylor [PT]	0.6	0.68	0.81	**0.88**	0.65	0.61-0.68	1.61	1.4-1.8

3.8 Estimation using additive correction factors

Applying an additive correction factor substantially increased CCCs for three of the conventional models (PM-ASCE, PM-FAO, and PT), indicating that their corrected estimates (CCC_{add} = 0.89, 0.89, and 0.88, respectively; Fig. S7, Table 4) more closely matched our empirically-derived ET values than did the uncorrected estimates. Additive corrections were more effective than landscape coefficients since these three models systematically overpredicted ET relative to our empirical approach. However, additive correction also produced negative daily ET estimates when uncorrected values were small, highlighting a limitation of this approach. Nonetheless, for these three models, additive correction yielded the greatest agreement with our empirical estimates.

In contrast, additive corrections yielded poorer agreement with our estimates for the MS and GG models (CCC_{add} = 0.69 and 0.62, respectively). In fact, none of the correction methods we tried improved agreement for either of these models with the empirically-derived values. This may be because there was already relatively good agreement (CCC ≈ 0.8), potentially due to the fact that MS and GG incorporate soil moisture limitation. MS does so implicitly, since it is essentially a variant of the PM model tailored for (semi-)arid climates (Shuttleworth and Wallace 2009). GG explicitly accounts for unsaturated soil conditions by including a coefficient for surface resistance into its estimation of ET (Granger and Gray, 1989).

3.9 Implications for bioretention design

The findings of this study have several implications for ET quantification and plant placement during the basin design process. Most importantly, several conventional ET models have the potential to accurately estimate ET in bioretention basins, despite their highly variable environments. Our findings therefore provide a potential means of improving the representation of ET in water balances used during basin design. For those with similar designs and in similar contexts to Basin A, the PM-ASCE, PM-FAO, or PT models could be used along with the correction factors we derived for Basin A, although the transferability of those corrections should be carefully evaluated. When the assumption of similarity to Basin A cannot be made, locally appropriate correction factors would need to be developed (McEvoy et al., 2022). One approach would be to develop locally appropriate K_L values through measurements of existing bioretention systems, with the resulting values applied to systems of the same type within a given region. Alternatively, K_L values could be estimated from species, vegetation density, and microclimate characteristics following the procedure outlined by Costello et al. (2000). The resulting ET estimates could improve predictions of stormwater volume reduction for proposed basins. Future measurement campaigns spanning a broader range of bioretention designs and environmental settings could establish the extent to which correction factors can be transferred among systems.

The HS model produced comparatively poor predictions for our basin. This model requires the least amount of climate input data since it estimates solar radiation based on latitude and time of year and makes assumptions about other meteorological variables (Guo et al., 2016). Considering the increasing availability of weather and solar radiation data, the simplifications of the HS model may become increasingly unnecessary. Instead, designers may be able to capitalize on other, more accurate models.

Additionally, our method of computing shade coefficients may be a simple yet useful means of adjusting solar radiation time series to increase the accuracy of ET estimates from conventional models, especially for sites with substantial shade from nearby infrastructure. The only requirements beyond a solar radiation time series are a high-resolution DEM of the site and software to compute direct and diffuse solar radiation from the DEM (such as ArcGIS).

Finally, engineers and landscape architects designing bioretention basins can potentially raise ET by placing plants that will grow larger (or otherwise produce greater leaf area) in appropriate locations within the basin. This might include basin flanks that receive little shade; locations that are frequently flooded may or may not be appropriate depending on the duration of soil saturation and soil salinity levels (Muerdter et al, 2018; Caplan et al., 2024). Larger plants may have utility in bioretention basins with internal water storage capacity (Hess et al., 2017). A potential tradeoff of increasing the size and extent of plants in a bioretention basin is the need to ensure access to adequate soil moisture through dry periods, especially since soil media is typically coarse-textured and can dry out substantially between rain events (Houdeshel et al., 2012; Nocco et al., 2016). This is a particular challenge in drier climates (Orr, 2013). We note that plants with deeper root systems (including large shrubs and trees) can access water over a broader depth range, lessening this limitation (Caplan et al., 2019).

3.10 Study limitations

Our empirical data came from a single bioretention basin, limiting the generalizability of our statistical model and basin-scale ET estimates. Nonetheless, the conclusions are likely applicable to other sites given that BASIN A contains many design features common in bioretention basins. These include the structure (a central flow path flanked by sloped sides), partial shading from surrounding structures, taxa common in GSI in the Mid-Atlantic region (Myers et al., 2026), and sandy loam soils.

Second, ET estimates for the largest plants in the basin required extrapolation from measurements from only modestly large plants, since our flux chambers were ~100 cm tall. Although plants taller than the chambers covered only ~10% of the basin area, they did have a disproportionally-large influence on basin-scale ET rates.

Third, although spatial variation in solar radiation was explicitly represented in our ET quantification, other microclimatic factors were assumed to be uniform. This is likely justified because air (and thus its temperature and humidity) tends to be well mixed at the basin scale, whereas solar radiation can be extremely variable over short distances due to shading. Nonetheless, VPD and wind speed could vary spatially within bioretention systems, particularly where surrounding structures and vegetation alter microclimate and airflow. Future studies could incorporate spatial measurements of these variables to determine whether accounting for this additional heterogeneity improves basin-scale ET estimates.

Finally, we could not calculate ET's contribution to stormwater volume reduction since we lack data on inflow into BASIN A. Empirical measurement was attempted but unsuccessful. Direct measurement did not capture low flows and approximating it using precipitation records in combination with directly-connected impervious area (DCIA) was highly unreliable because DCIA appeared to change through time depending on highway construction activities and other factors.

4 Conclusions

Bioretention basins and other types of green stormwater infrastructure play an important role in offsetting the negative hydrologic impacts of impervious surfaces. To design these systems optimally, we need methods of accurately quantifying a basin's ability to reduce stormwater volume not only through infiltration but also through evapotranspiration (ET). However, ET involves multiple soil, plant, and meteorological processes, all of which are highly variable in space and time, and some of which are challenging to measure directly.

We parameterized a spatially-explicit ET model for a bioretention basin based on measurements of water flux from soil and plants, plant size, soil moisture, weather, and shade cast by nearby buildings and other structures. This empirical whole-basin model explained most of the observed variation in ET. The model was particularly sensitive to plant height and shade, both of which should be carefully measured or estimated when modeling bioretention basin ET. They are also factors that designers can influence, meaning that ET can likewise be moderated through design.

Quantifying ET at fine spatial scales is typically unnecessary for bioretention basin design or system evaluation. While instructive for elucidating sources of ET variation within a basin, directly measuring ET within bioretention basins is a laborious process requiring specialized equipment not commonly part of stormwater practitioners' toolkits. Penman-Monteith (both ASCE and FAO) and Priestley-Taylor models produced estimates that consistently overestimated ET in comparison to our empirical model. Depending on the accuracy required, these model estimates could be used without correction, with the understanding that they are likely to overestimate ET, or they could be adjusted using additive, multiplicative, or shade-based corrections to improve agreement with empirical ET. We note that additive correction is highly context-specific and may produce negative estimates at low ET rates and should therefore be used with caution. Matt-Shuttleworth and Granger-Gray also had good agreement with our empirical model, though they tended to underpredict high ET rates and overpredict low ET. The Hargreaves-Samani model had the poorest agreement regardless of correction factors used, likely because it only requires temperature data and estimates solar radiation based on location rather than measurements. When sufficient meteorological data are available, the Penman-Monteith and Priestley-Taylor models are particularly suitable for bioretention applications because they are sufficiently responsive to variation in atmospheric conditions and their systematic biases can be reduced or removed using locally appropriate corrections.

Data and code availability

All of the code and data needed to reproduce the results of this study are available from HydroShare: http://www.hydroshare.org/resource/ba487d9c6e4b473f88d8298499ee1c6d

Author contributions

JSC: conceptualization, methodology, formal analysis, investigation, data curation, writing - original draft, writing - review & editing, visualization, supervision, project administration. MB: conceptualization, formal analysis, writing - review & editing. AS: data curation, writing - original draft, writing - review & editing, visualization. MGA: conceptualization, formal analysis, writing - review & editing. JEN: methodology, investigation, resources, writing - review & editing. LT: conceptualization, project administration, funding acquisition, writing - review & editing. SWE: conceptualization, methodology, supervision, funding acquisition.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We are grateful to Anastasia Rozovski for contributions to data cleaning and analysis, Andrew Kurzweil and the Villanova Center for Resilient Water Systems for sharing microclimate data, Sam Chattin for assistance with field data collection, Thomas Mozdzer for loaning his gas analyzer, Marilyn Can for assistance with the shade analysis, and Ro Klinger for assistance with map generation.

Financial support

This research was funded by the Pennsylvania Department of Transportation under the SR 0095 Section GIR project through a subcontract from AECOM.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell., 34, 2274–2282, https://doi.org/10.1109/TPAMI.2012.120, 2012.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper No. 56, Food and Agriculture Organization of the United Nations, Rome, 1998.

Alonzo, M., Dial, R. J., Schulz, B. K., Andersen, H.-E., Lewis-Clark, E., Cook, B. D., and Morton, D. C.: Mapping tall shrub biomass in Alaska at landscape scale using structure-from-motion photogrammetry and lidar, Remote Sens. Environ., 245, 111841, https://doi.org/10.1016/j.rse.2020.111841, 2020.

Ampomah, R., Holt, D., Smith, C., Smith, V., Sample-Lord, K., Nyquist, J.: Modeling bioinfiltration surface dynamics through a hybrid geomorphic-infiltration model. Blue-Green Sys., 5, 152–168. https://doi.org/10.2166/bgs.2023.027, 2023.

Bansal, S., Creed, I. F., Tangen, B. A., et al.: Practical guide to measuring wetland carbon pools and fluxes, Wetlands, 43, 105, https://doi.org/10.1007/s13157-023-01722-2, 2023.

Bates, D., Mächler, M., Bolker, B., and Walker, S.: Fitting linear mixed-effects models using lme4, J. Stat. Softw., 67, 1–48, https://doi.org/10.18637/jss.v067.i01, 2015.

Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65, 2–16, https://doi.org/10.1016/j.isprsjprs.2009.06.004, 2010.

Bowman, D. H. and King, K. M.: Determination of evapotranspiration using the neutron scattering method, Can. J. Soil Sci., 45, 117–126, 1965.

Brown, R. A. and Hunt, W. F.: Impacts of media depth on effluent water quality and hydrologic performance of undersized bioretention cells, J. Irrig. Drain. Eng., 137, 132–143, https://doi.org/10.1061/(ASCE)IR.1943-4774.0000167, 2011.

Caplan, J. S., Galanti, R. C., Olshevski, S., and Eisenman, S. W.: Water relations of street trees in green infrastructure tree trench systems, Urban For. Urban Green., 41, 170–178, https://doi.org/10.1016/j.ufug.2019.03.016, 2019.

Caplan, J. S., Salisbury, A. B., McKenzie, E. R., Behbahani, A., and Eisenman, S. W.: Spatial, temporal, and biological factors influencing plant responses to deicing salt in roadside bioinfiltration basins, J. Environ. Manage., 359, 120761, https://doi.org/10.1016/j.jenvman.2024.120761, 2024.

City of Philadelphia: Philadelphia LiDAR – LAS files 2018, Pennsylvania Spatial Data Access, available at: https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=2021, 2018.

Coleman, T. F. and Li, Y.: On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds, Math. Program., 67, 189–224, https://doi.org/10.1007/BF01582221, 1994.

Coleman, T. F. and Li, Y.: An interior trust region approach for nonlinear minimization subject to bounds, SIAM J. Optim., 6, 418–445, https://doi.org/10.1137/0806023, 1996.

Costello, L. R., Matheny, N. P., and Clark, J. R.: A guide to estimating irrigation water needs of landscape plantings in California: the landscape coefficient method and WUCOLS III, California Dep. Water Resour., available at: https://cimis.water.ca.gov/Content/PDF/wucols00.pdf, 2000.

Daniels, B. J. and Yeakley, J. A.: Catchment-scale hydrologic effectiveness of residential rain gardens: A case study in Columbia, Maryland, USA, Water, 16, 1304, 10.3390/w16091304, 2024.

Denich, C. and Bradford, A.: Estimation of evapotranspiration from bioretention areas using weighing lysimeters, Journal of Hydrologic Engineering, 15, 522–530, doi:10.1061/(ASCE)HE.1943-5584.0000134, 2010.

Ebrahimian, A., Wadzuk, B., and Traver, R.: Evapotranspiration in green stormwater infrastructure systems, Sci. Total Environ., 688, 797–810, https://doi.org/10.1016/j.scitotenv.2019.06.256, 2019.

Fox, J., and Weisberg, S.: An R Companion to Applied Regression, 3rd ed., Sage, Thousand Oaks, CA, 2019.

Garcia, A., Johnson, M. J., Andraski, B. J., Halford, K. J., and Myers, C. J.: Portable chamber measurements of evapotranspiration at the Amargosa Desert Research Site near Beatty, Nevada, 2003–06, USGS Sci. Investig. Rep. 2008–5135, Reston, VA, 2008.

Geoghegan, E. K., Caplan, J. S., Leech, F. N., Weber, P. E., Bauer, C. E., and Mozdzer, T. J.: Nitrogen enrichment alters carbon fluxes in a New England salt marsh, Ecosyst. Health Sustain., 4, 277–287, https://doi.org/10.1080/20964129.2018.1532772, 2018.

Granger, R. J. and Gray, D. M.: Evaporation from natural nonsaturated surfaces, J. Hydrol., 111, 21–29, https://doi.org/10.1016/0022-1694(89)90249-7, 1989.

Groh, J., Pütz, T., Gerke, H. H., Vanderborght, J., and Vereecken, H.: Quantification and prediction of nighttime evapotranspiration for two distinct grassland ecosystems, Water Resour. Res., 55, 2961–2975, https://doi.org/10.1029/2018WR024072, 2019.

Guo, D., Westra, S., and Maier, H. R.: An R package for modelling actual, potential and reference evapotranspiration, Environ. Model. Softw., 78, 216–224, https://doi.org/10.1016/j.envsoft.2015.12.019, 2016.

Hamel, P., McHugh, I., Coutts, A., Daly, E., Beringer, J., and Fletcher, T. D.: Automated chamber system to measure field evapotranspiration rates, J. Hydrol. Eng., 20, 04014037, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001006, 2015.

Hess, A., Wadzuk, B. M., and Welker, A. L.: Evapotranspiration in rain gardens using weighing lysimeters, J. Irrig. Drain. Eng., 143, 04017004, https://doi.org/10.1061/(ASCE)IR.1943-4774.0001157, 2017.

Hess, A., Wadzuk, B., and Welker, A.: Evapotranspiration estimation in rain gardens using soil moisture sensors, Vadose Zone J., 20, e20100, https://doi.org/10.1002/vzj2.20100, 2021.

Huang, T., Sage, J., Técher, D., and Gromaire, M.-C.: Hydrological performance of bioretention in field experiments and models: A review from the perspective of design characteristics and local contexts, Sci. Total Environ., 965, 178684, https://doi.org/10.1016/j.scitotenv.2025.178684, 2025.

Hung, F., Harman, C. J., Hobbs, B. F., and Sivapalan, M.: Assessment of climate, sizing, and location controls on green infrastructure efficacy: A timescale framework, Water Resources Research, 56, e2019WR026141, 10.1029/2019WR026141, 2020.

Houdeshel, C. D., Pomeroy, C. A., and Hultine, K. R.: Bioretention design for xeric climates based on ecological principles, J. Am. Water Resour. Assoc., 48, 1178–1190, https://doi.org/10.1111/j.1752-1688.2012.00678.x, 2012.

Johnston, M. R., Balster, N. J., and Thompson, A. M.: Vegetation alters soil water drainage and retention of replicate rain gardens, Water, 12, 3151, 10.3390/w12113151, 2020.

Joshi, P., Leitão, J. P., Maurer, M., and Bach, P. M.: Not all SuDS are created equal: Impact of different approaches on combined sewer overflows, Water Research, 191, 116780, 10.1016/j.watres.2020.116780, 2021.

Kool, D., Agam, N., Lazarovitch, N., Heitman, J. L., Sauer, T. J., and Ben-Gal, A.: A review of approaches for evapotranspiration partitioning, Agric. For. Meteorol., 184, 56–70, https://doi.org/10.1016/j.agrformet.2013.09.003, 2014.

Kohler, M. A. and Linsley, R. K.: Predicting the runoff from storm rainfall, U.S. Weather Bureau Research Paper No. 34, 41 pp., 1951.

Kuehler, E., Hathaway, J., and Tirpak, A.: Quantifying the benefits of urban forest systems as a component of the green infrastructure stormwater treatment network, Ecohydrology, 10, e1813, https://doi.org/10.1002/eco.1813, 2017.

Li, C., Peng, C., Chiang, P.-C., Cai, Y., Wang, X., and Yang, Z.: Mechanisms and applications of green infrastructure practices for stormwater control: A review, Journal of Hydrology, 568, 626–637, 10.1016/j.jhydrol.2018.10.074, 2019.

Li, H., Sharkey, L. J., Hunt, W. F., and Davis, A. P.: Mitigation of impervious surface hydrology using bioretention in North Carolina and Maryland, J. Hydrol. Eng., 14, 407–415, https://doi.org/10.1061/(ASCE)1084-0699(2009)14:4(407), 2009.

Lin, L. I.-K.: A concordance correlation coefficient to evaluate reproducibility, Biometrics, 45, 255–268, https://doi.org/10.2307/2532051, 1989.

Litvak, E., McCarthy, H. R., and Pataki, D. E.: A method for estimating transpiration of irrigated urban trees in California, Landsc. Urban Plan., 158, 48–61, https://doi.org/10.1016/j.landurbplan.2016.09.021, 2017.

Luo, C., Wang, Z., Sauer, T. J., Helmers, M. J., and Horton, R.: Portable canopy chamber measurements of evapotranspiration in corn, soybean, and reconstructed prairie, Agric. Water Manag., 198, 1–9, https://doi.org/10.1016/j.agwat.2017.11.024, 2018.

McEvoy, D. J., Roj, S., Dunkerly, C., McGraw, D., Huntington, J. L., Hobbins, M. T., and Ott, T.: Validation and bias correction of forecast reference evapotranspiration for agricultural applications in Nevada, Journal of Water Resources Planning and Management, 148, 04022057, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001595, 2022.

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., and McVicar, T. R.: Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis, Hydrol. Earth Syst. Sci., 17, 1331–1363, https://doi.org/10.5194/hess-17-1331-2013, 2013.

McPhillips, L. E. and Matsler, A. M.: Temporal evolution of green stormwater infrastructure strategies in three US cities, Front. Built Environ., 4, 26, https://doi.org/10.3389/fbuil.2018.00026, 2018.

Muerdter, C. P., Wong, C. K., and LeFevre, G. H.: Emerging investigator series: The role of vegetation in bioretention for stormwater treatment in the built environment: pollutant removal, hydrologic function, and ancillary benefits, Environ. Sci.: Water Res. Technol., 4, 592–612, https://doi.org/10.1039/C7EW00511C, 2018.

Myers, D., Wadzuk, B., Hess, A., and Caplan, J. S.: An integration of multiple bioretention design manuals and data sources to improve plant selection and outcomes, Journal of Sustainable Water in the Built Environment, 12, 04026007, https://doi.org/10.1061/JSWBAY.SWENG-702, 2026.

Nakagawa, S., Johnson, P. C. D., and Schielzeth, H.: The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded, J. R. Soc. Interface, 14, 20170213, https://doi.org/10.1098/rsif.2017.0213, 2017.

Nasrollahpour, R., Skorobogatov, A., He, J., Valeo, C., Chu, A., and van Duin, B.: The impact of vegetation and media on evapotranspiration in bioretention systems, Urban Forestry & Urban Greening, 74, 127680, https://doi.org/10.1016/j.ufug.2022.127680, 2022.

Nocco, M. A., Rouse, S. E., and Balster, N. J.: Vegetation type alters water and nitrogen budgets in a controlled, replicated experiment on residential-sized rain gardens planted with prairie, shrub, and turfgrass, Urban Ecosyst., 19, 1665–1691, https://doi.org/10.1007/s11252-016-0568-7, 2016.

Norman, J. M., Kucharik, C. J., Gower, S. T., Baldocchi, D. D., Crill, P. M., Rayment, M., Savage, K., and Striegl, R. G.: A comparison of six methods for measuring soil-surface carbon dioxide fluxes, J. Geophys. Res., 102, 28771–28777, https://doi.org/10.1029/97JD01440, 1997.

Nouri, H., Beecham, S., Hassanli, A. M., and Kazemi, F.: Water requirements of urban landscape plants: a comparison of three factor-based approaches, Ecol. Eng., 57, 276–284, https://doi.org/10.1016/j.ecoleng.2013.04.025, 2013.

Ouédraogo, A. A., Berthier, E., Ramier, D., Tan, Y., and Gromaire, M.-C.: Quantifying evapotranspiration fluxes on green roofs: A comparative analysis of observational methods, Sci Total Environ, 902, 166135, https://doi.org/10.1016/j.scitotenv.2023.166135, 2023.

Pannkuk, T. R., White, R. H., Steinke, K., Aitkenhead-Peterson, J. A., Chalmers, D. R., and Thomas, J. C.: Landscape coefficients for single- and mixed-species landscapes, HortScience, 45, 1529–1533, https://doi.org/10.21273/HORTSCI.45.10.1529, 2010.

Paredes, P., Martins, D. S., Pereira, L. S., Cadima, J., and Pires, C.: Accuracy of daily estimation of grass reference evapotranspiration using ERA-Interim reanalysis products with assessment of alternative bias correction schemes, Agricultural Water Management, 210, 340–353, https://doi.org/10.1016/j.agwat.2018.08.003, 2018.

[PennDOT] Pennsylvania Department of Transportation: Traffic volume maps, available at: https://www.penndot.pa.gov/ProjectAndPrograms/Planning/Maps/Pages/Traffic-Volume.aspx, last access: 22 June 2022, 2022.

Pennino, M. J., McDonald, R. I., and Jaffe, P. R.: Watershed-scale impacts of stormwater green infrastructure on hydrology, nutrient fluxes, and combined sewer overflows in the mid-Atlantic region, Sci Total Environ, 565, 1044–1053, https://doi.org/10.1016/j.scitotenv.2016.05.101, 2016.

Pope, G. G., Toran, L., Caplan, J. S., and Nyquist, J.: Spatiotemporal dynamics of road salt in a highway bioswale: a comparison of point and continuous monitoring methods, Water Air Soil Pollut., 236, 654, https://doi.org/10.1007/s11270-025-08281-8, 2025.

Shakya, M., Hess, A., Wadzuk, B. M., and Traver, R. G.: A soil moisture profile conceptual framework to identify water availability and recovery in green stormwater infrastructure, Hydrology, 10, 197, 2023.

Scharenbroch, B. C., Morgenroth, J., and Maule, B.: Tree species suitability to bioswales and impact on the urban water budget, J. Environ. Qual., 45, 199–206, https://doi.org/10.2134/jeq2015.01.0060, 2016.

Shuttleworth, W. J.: Putting the "vap" into evaporation, Hydrol. Earth Syst. Sci., 11, 210–244, https://doi.org/10.5194/hess-11-210-2007, 2007.

Shuttleworth, W. J. and Wallace, J. S.: Calculating the water requirements of irrigated crops in Australia using the Matt–Shuttleworth approach, Trans. ASABE, 52, 1895–1906, https://doi.org/10.13031/2013.29217, 2009.

Skorobogatov, A., He, J., Chu, A., Valeo, C., and van Duin, B.: The impact of media, plants and their interactions on bioretention performance: a review, Sci. Total Environ., 715, 136918, https://doi.org/10.1016/j.scitotenv.2020.136918, 2020.

Steduto, P. and Hsiao, T. C.: Maize canopies under two soil water regimes: II. Seasonal trends of evapotranspiration, carbon dioxide assimilation and canopy conductance, and as related to leaf area index, Agric. For. Meteorol., 89, 185–200, https://doi.org/10.1016/S0168-1923(97)00084-1, 1998.

Sun, H., Kopp, K., and Kjelgren, R.: Water-efficient urban landscapes: integrating different water use categorizations and plant types, HortScience, 47, 254–263, https://doi.org/10.21273/HORTSCI.47.2.254, 2012.

Toran, L., Eisenman, S., Caplan, J., van Aken, B., McKenzie, E., Nyquist, J., Ryan, R., Traver, R., Tu, M.-C., Schmidt, N., and Calt, E.: Stormwater control management and monitoring, report prepared for the Pennsylvania Department of Transportation, 2017.

Traver, R. G. and Ebrahimian, A.: Dynamic design of green stormwater infrastructure, Front. Environ. Sci. Eng., 11, 15, https://doi.org/10.1007/s11783-017-0973-z, 2017.

Wadzuk, B. M., Hickman, J. M., and Traver, R. G.: Understanding the role of evapotranspiration in bioretention: mesocosm study, J. Sustain. Water Built Environ., 1, 04014002, https://doi.org/10.1061/JSWBAY.0000794, 2015.

Wagner, W. and Pruss, A.: International equations for the saturation properties of ordinary water substance. Revised according to the international temperature scale of 1990, J. Phys. Chem. Ref. Data, 22, 783–787, https://doi.org/10.1063/1.555926, 1993.

Xue, Y., Zhang, Z., Li, X., Liang, H., and Yin, L.: A review of evapotranspiration estimation models: advances and future development, Water Resour. Manage., 39, 3641–3657, https://doi.org/10.1007/s11269-025-04191-w, 2025.

Zhang, K. and Parolari, A. J.: 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, 2022.