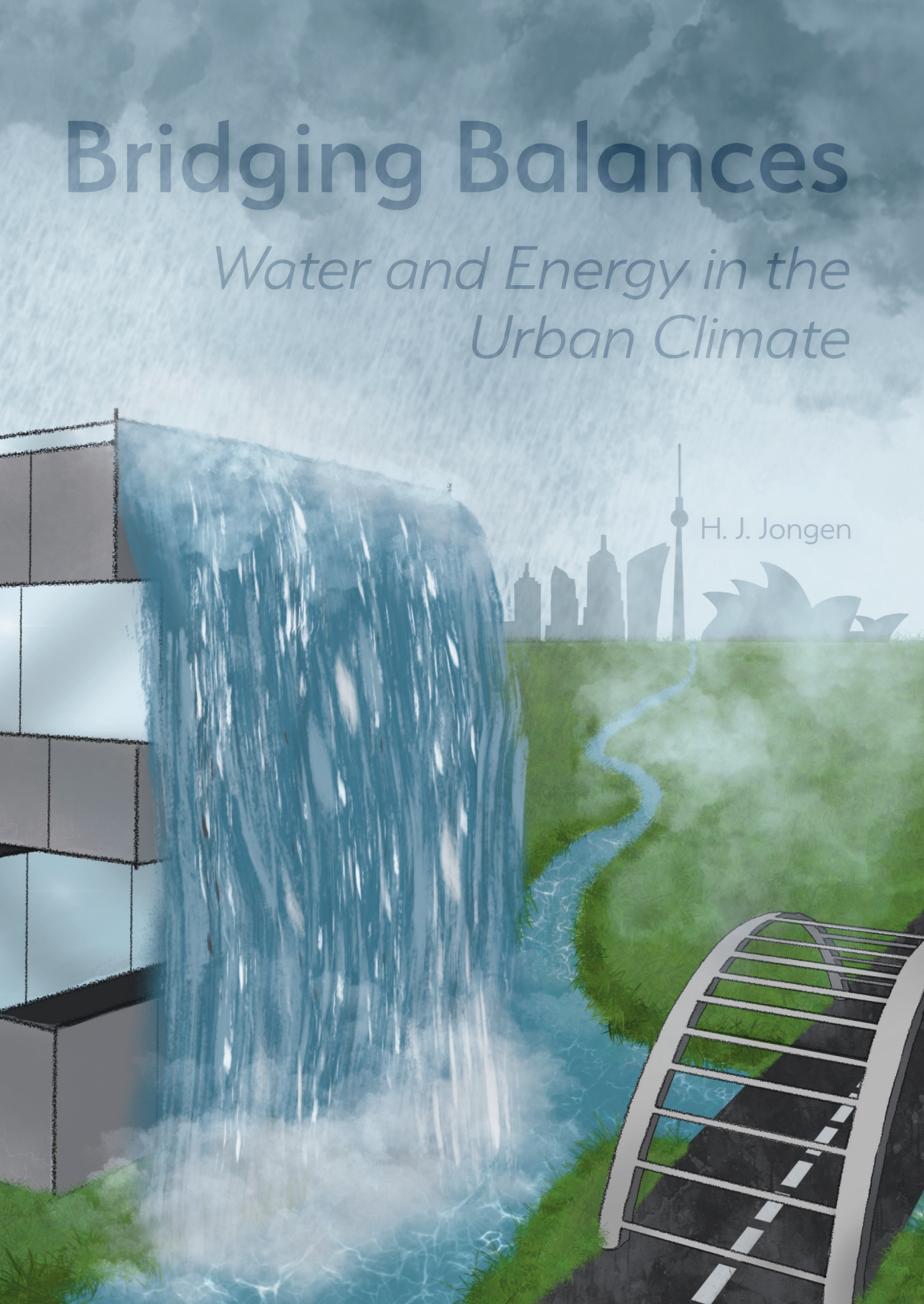
Bridging Balances

Water and Energy in the Urban Climate



H. J. Jongen

Propositions

1. Urban climate studies only considering temperature miss the (dew) point.
   (this thesis)

2. Urban complexity disguises dependencies between site and climate.
   (this thesis)

3. An independent scientist is dysfunctional.

4. The absence of recognition for acknowledgements promotes long author lists.

5. Discussions on correct scientific terminology are irrelevant.

6. Recruiting all-round scientists diminishes diversity in academia.

7. Rewarding work performed outside working hours disturbs the work-life balance.

8. The vanishing partner dancing culture is a loss for Dutch society.

Propositions belonging to the thesis, entitled

Bridging Balances: Water and Energy in the Urban Climate

Harro J. Jongen
Wageningen,
3 April 2025

Bridging Balances

Water and Energy in the Urban Climate

Harro J. Jongen

Thesis committee

Promotors:
Dr A.J. Teuling
Associate professor, Hydrology and Environmental Hydraulics Group
Wageningen University & Research

Dr G.-J. Steeneveld
Associate professor, Meteorology and Air Quality Group
Wageningen University & Research

Other members:
Prof. Dr S. Lenzholzer, Wageningen University & Research
Prof. Dr B.J.H. van de Wiel, Delft University of Technology
Dr S. Kotthaus, Institut Polytechnique de Paris, France
Prof. Dr B. Verbeiren, Vrije Universiteit Brussels, Belgium

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

Bridging Balances

Water and Energy in the Urban Climate

Harro J. Jongen

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 3 April 2025
at 3:30 p.m. in the Omnia Auditorium.

Harro J. Jongen
Bridging balances: Water and Energy in the Urban Climate
XVIII+208 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2025)
With references, with summary in English and Dutch

DOI 10.18174/681322

Contents

Summary

People change their surroundings, which is evident in urban areas with hardly anything left untouched. These changes include increased imperviousness, less vegetation, and more structures. Together they cause urban areas to experience higher air temperatures than their surroundings. These temperatures exacerbate the heat stress experienced by urban inhabitants. Strategies to mitigate this heat stress often rely on increasing evapotranspiration to divert energy from heating the air. Evapotranspiration is part of the energy and the water balance making both balances critical to the success of heat mitigation measures. Yet, the energy balance is mostly studied in isolation (**Chapter 1**). Bridging these balances can progress our understanding of the dynamics of urban climate.

This thesis aims to uncover how the water balance influences the energy balance focusing on their shared flux, evapotranspiration. To fulfill this aim, this thesis employs observations and model results from street to neighborhood scale from cities worldwide. The balances are bridged by quantifying the water storage capacity and its recession rate together governing water availability for evapotranspiration (Chapter 2); evaluating and improving the water balance representation in urban land surface models and the water balance impact on the model's performance in simulating evapotranspiration (Chapter 3 and 4); linking evapotranspiration at the neighborhood scale to the surface cover (Chapter 5); and identifying the impact of changing surface cover in the source area of eddy-covariance observations (Chapter 6).

In **Chapter 2**, a novel method is developed to quantify urban water storage using the recession of evapotranspiration observed with the eddy-covariance technique. The eddy-covariance observations provide evapotranspiration on the neighborhood level. Thus, this method bypasses the need to measure every element contributing to the storage. During dry periods, no water enters the storage, and evapotranspiration is the only flux leaving it. The storage capacity can thus be estimated from the initial evapotranspiration and the recession rate. The recession timescale corresponds to the period after which 37% of the initial evapotranspiration is reached. The approach was applied in fourteen sites with varying local climate zones and background climates. The recession timescale was found to vary between 1.8–20.1 days corresponding to storage capacities between 1.3–28.4 mm.

Given these values, the urban water storage capacity is at least five times smaller than in natural environments.

Chapter 3 evaluates the water balance in 19 urban land surface models from the Urban-PLUMBER project run for 20 sites. This evaluation is also used to study the influence of the water balance on simulated heat exchange between the atmosphere and urban land surface. Since observations for most of the water fluxes are unavailable, seven indicators compose the urban water balance representation score, which reflects how well the water balance is represented by examining the water balance closure, flux timing, and magnitude. No model manages to reach a perfect score. The water balance remains unclosed in 57% of the model-site combinations with a mismatch between the annual total incoming fluxes (precipitation and irrigation) and the outgoing (all other) fluxes higher than 3% of the outgoing fluxes. The models capture the timing of fluxes better than their magnitude with runoff as the most poorly captured flux. The hypothesis was that models with a higher score would outperform those with a lower score for the turbulent heat fluxes, but, surprisingly, no relation was found. Yet, the results demonstrate quantitative model evaluation without observations can guide model development. Models could be improved by explicitly verifying the water balance closure and revising the runoff parameterizations.

Chapter 4 takes the next step and deepens the surface runoff evaluation to unravel why urban land surface models poorly capture surface runoff. The evaluation examines the same models as the previous chapter omitting the model not separating surface and subsurface runoff. Surface runoff is generated when rainfall exceeds the infiltration, saturation, or interception capacity. Ten models do not include all these processes and seven do not produce more surface runoff when the impervious fraction increases. Large (over 20 mm) and intense (exceeding 50 mm h^{-1}) rainfall events do not lead to surface runoff in all models. For these events, models produce 43% less runoff on average than the curve number method, which estimates the surface runoff per rainfall event based on the site characteristics. The underestimation is higher for models missing one or more runoff processes than for those including all. This suggests model performance would increase if all runoff generation processes are represented.

In **Chapter 5**, the models make way for eddy-covariance observations from two sites in Berlin to link evapotranspiration at the neighborhood scale to the patch-scale surface cover. The heterogeneous urban surface is categorized into four cover types: impervious surfaces, low vegetation, high vegetation, and open water. For the four types, the evapotranspiration dynamics are estimated with patch-scale observations and conceptual models and scaled to the neighborhood scale (bottom-up). The other way around, eddy-covariance observations are attributed to the four surface cover types by solving a system of linear equations (top-down). Both approaches need to know the surface cover composition in the source area of the eddy-covariance system. The source area is referred to as the footprint and is modeled

analytically. Vegetation is responsible for more evapotranspiration than proportional to its surface fraction in the footprint. The contribution of impervious surfaces cannot be neglected, even though it is less than their surface fraction in the footprint. Their low contribution is the consequence of water availability limiting evapotranspiration within hours after rainfall, while vegetation and open water sustain evapotranspiration for longer periods.

Chapter 6 further focuses on the eddy-covariance observations in Berlin to quantify the influence of the heterogeneous footprint on the observed fluxes. The footprint is influenced by wind conditions, atmospheric stability, and the sensor location and height. The time-dependence of the footprint is quantified keeping constant values for the evapotranspiration per surface cover type scaled to the neighborhood scale with the analytically modeled surface cover fractions in the footprint. The surface cover fractions are modeled according to observed meteorological conditions. The footprint composition affects observed evapotranspiration as much due to variations in the meteorological conditions within a day as in a year, even when the day has relatively constant meteorological conditions. While analytically derived footprints reveal the temporal sensitivity of footprints, the spatial sensitivity is beyond their capacity. The analytical model neglects the irregular urban morphology. Large-eddy simulations resolve the urban flow field capturing the turbulent transport of moisture incorporating the footprint, which enables virtually mimicking eddy-covariance systems. With a grid of virtual eddy-covariance systems, it is possible to quantify the effect of the sensor location. The virtual eddy-covariance systems de-correlate within 280 m indicating how sensitive the observed fluxes are to the system's specific location and how non-representative eddy-covariance is for a neighborhood. In conclusion, eddy-covariance observations should be interpreted considering the footprints, as these footprints strongly vary in time and space.

Chapter 7 brings together all preceding chapters concluding that 1) evapotranspiration is water-limited in urban areas, 2) urban land surface models fail to accurately capture the water balance, 3) missing runoff processes in these models explain the runoff underestimation, 4) the surface cover is crucial to evapotranspiration dynamics, and 5) footprints of eddy-covariance systems are highly variable in space and time.

When studying the urban climate, one faces the "urban challenge" due to heterogeneity within urban environments, human activity in urban areas, and diversity among cities. When modeling this challenging environment, urban climate models should be guided by their purpose without employing resources only to reach higher spatial resolutions. The development of these and other models would benefit from evaluations including quantitative, logical checks next to a comparison with observations.

The urban challenge prevails when insights are translated towards application. Urban climate exists within a dynamic and interconnected system studied by many scientific disciplines. Water embodies these dynamics and interconnections raising humidity, facilitating

evaporative cooling, and keeping vegetation alive. Livable cities are created by changing our surroundings keeping these dynamics and this interconnectedness in mind.

Samenvatting

Mensen veranderen hun omgeving, wat maar al te duidelijk is in stedelijke gebieden waar praktisch niets onaangeroerd blijft. Deze veranderingen zijn onder andere meer verhard oppervlakte, minder vegetatie en meer gebouwen. Samen veroorzaken zij hogere temperaturen in de stad dan in het omliggende gebied. De hoge temperaturen verergeren de hittestress die mensen in de stad ervaren. Strategieën om de hittestress te verminderen zijn vaak gericht op het verhogen van de verdamping zodat deze energie de lucht niet kan verhitten. Verdamping is onderdeel van zowel de energie- als de waterbalans wat beide balansen cruciaal maakt voor het succes van hitte-bestrijdende maatregelen. Toch wordt de energiebalans meestal afzonderlijk onderzocht (**Hoofdstuk 1**). Een brug tussen deze balansen kan ons begrip van het stedelijk klimaat verbeteren.

Dit proefschrift heeft als doel te begrijpen hoe de waterbalans de energiebalans beïnvloedt met een focus op hun gedeelde flux: verdamping. Om dit doel te behalen worden in dit proefschrift zowel observaties als modellen gebruikt voor steden over de hele wereld van straat- tot buurtschaal. De brug tussen de balansen wordt gebouwd door de stedelijke wateropslagcapaciteit en de bijbehorende recessiesnelheid te kwantificeren samen bepalend voor de waterbeschikbaarheid voor verdamping (Hoofdstuk 2); de waterbalansimplementatie in stedelijke landoppervlaktemodellen te evalueren en verbeteren en de impact van de waterbalans te bepalen om de capaciteit van de modellen om verdamping te simuleren (Hoofdstuk 3 en 4); de verdamping op buurtschaal te koppelen aan het landoppervlakte (Hoofdstuk 5) en de impact van het veranderen van de samenstelling van het landoppervlakte in het brongebied van eddy-covariantie observaties in kaart te brengen (Hoofdstuk 6).

In **Hoofdstuk 2** wordt een nieuwe methode ontwikkeld om de stedelijke wateropslagcapaciteit te kwantificeren op basis van de afname van verdamping waargenomen met de eddy-covariantie techniek. De eddy-covariantie observaties zijn op buurtschaal. Daarom is het niet noodzakelijk om elk onderdeel van de wateropslag apart te meten. Tijdens droge periodes wordt de wateropslag niet aangevuld en is verdamping de enige flux uit de opslag. De opslagcapaciteit kan dus geschat worden aan de hand van de initiële verdamping en de snelheid waarmee die afneemt. De recessietijdschaal beschrijft de tijd wanneer de verdamping op 37% van het originele niveau is. Deze methode is toegepast in veertien steden

met verschillende stedelijke eigenschappen en achtergrondklimaten. De recessietijdschaal bleek te variëren tussen 1.8–20.1 dagen met opslagcapaciteiten tussen 1.3–28.4 mm. Deze getallen betekenen dat de stedelijke wateropslagcapaciteit minstens vijf keer kleiner is dan die in natuurlijke omgevingen.

Hoofdstuk 3 evalueert de waterbalans in 19 stedelijke landoppervlaktemodellen van het Urban-PLUMBER project toegepast op 20 locaties. Hiermee wordt ook gekeken wat de invloed is van de waterbalans op de gesimuleerde warmte-uitwisseling tussen de atmosfeer en het stedelijke landoppervlakte. Omdat observaties niet beschikbaar zijn voor de meeste waterfluxen, vormen zeven indicatoren de stedelijke waterbalansimplementatiescore. Deze score geeft aan hoe goed de waterbalans in de modellen is gerepresenteerd wordt door de waterbalanssluiting en de fluxtiming en -grootte te beoordelen. Geen model haalt een perfecte score. De waterbalans is niet gesloten in 57% van de model-locatiecombinaties met een verschil tussen de jaarlijkse inkomende fluxen (neerslag en irrigatie) en de uitgaande (alle andere) fluxen hoger dan 3% van de inkomende fluxen. De modellen zijn beter in het beschrijven van de timing dan de grootte van de fluxen en afvoer is de slechtst beschreven flux. De hypothese was dat modellen met een hogere waterbalansimplementatiescore beter zouden presteren voor de turbulente warmtefluxen dan die met een lagere score, maar hiertussen blijkt geen verband te zijn. Toch laten de resultaten zien dat kwantitatieve modelevaluatie mogelijk is zonder observaties en tot modelverbeteringen kan leiden. De modellen zouden verbeterd kunnen worden door expliciet te controleren of de waterbalans sluit en de afvoerparameterisatie te herzien.

Hoofdstuk 4 zet de volgende stap en kijkt verder naar de oppervlakte-afvoerparameterisatie om te diagnosticeren waarom stedelijke landoppervlaktemodellen de oppervlakteafvoer niet accuraat beschrijven. De evaluatie beoordeelt dezelfde modellen als het vorige hoofdstuk met uitzondering van het model dat geen onderscheid maakt tussen oppervlakteafvoer en ondergrondse afvoer. Oppervlakteafvoer wordt gegenereerd als neerslag de infiltratie-, verzadigings-, of interceptiecapaciteit overschrijdt. Tien modellen bevatten niet al deze drie processen en zeven modellen genereren niet meer oppervlakteafvoer wanneer de verharde fractie toeneemt. Grote (meer dan 20 mm) of intense (meer dan 50 mm u^{-1}) regenbuien leiden niet in alle modellen tot oppervlakteafvoer. Voor deze buien schatten de modellen de oppervlakteafvoer 43% lager in dan de zgn. curve-nummermethode, die de verwachte oppervlakteafvoer schat aan de hand van de eigenschappen van de omgeving. Deze onderschatting is hoger voor modellen die één of meer afvoerprocessen missen dan voor de modellen die alle processen beschrijven. Dit duidt aan dat de modelprestaties verbeterd kunnen worden door alle afvoerprocessen mee te nemen.

In **Hoofdstuk 5** maken de modellen plaats voor eddy-covariantie observaties van twee locaties in Berlijn om de verdamping op buurtschaal te koppelen aan het landoppervlakte op plotschaal. Het heterogene, stedelijke oppervlakte wordt in vier categorieën verdeeld:

verharde oppervlaktes, lage vegetatie, hoge vegetatie en open water. Voor elk van de vier categorieën wordt de verdampingsdynamiek geschat met observaties en conceptuele modellen op plotschaal om hieruit vervolgens de buurtschaal te reconstrueren (van onder naar boven). Andersom worden de eddy-covariantie observaties toegeschreven aan de oppervlaktetypes door een lineair stelsel van vergelijkingen op te lossen (van boven naar onder). Beide methodes gebruiken de oppervlaktesamenstelling in het brongebied van de eddy-covariantie van om de plot- en buurtschalen te verbinden. Dit brongebied wordt geschat met een analytisch model. Vegetatie is verantwoordelijk voor een groter aandeel in de verdamping dan proportioneel is tot het aandeel in het brongebied. De bijdrage van verharde oppervlaktes kan niet verwaarloosd worden, ondanks dat de bijdrage relatief laag is ten opzichte van de aanwezigheid in het brongebied. De lage bijdrage van de verharde oppervlaktes is een gevolg van de lage waterbeschikbaarheid die verdamping beperkt tot binnen enkele uren na neerslag, terwijl vegetatie en open water voor langere tijd blijven verdampen.

Hoofdstuk 6 gebruikt dezelfde eddy-covariantie observaties in Berlijn om de invloed van het heterogene brongebied op de waargenomen fluxen te bepalen. Het brongebied hangt af van de windcondities, de atmosferische stabiliteit en de sensorlocatie en -hoogte. De tijdsafhankelijkheid van het brongebied wordt gekwantificeerd door dezelfde verdampingswaardes per oppervlaktetype te vertalen naar de buurtschaal op basis van analytisch geschatte brongebieden. De brongebieden worden geschat op basis van meteorologische observaties per halfuur. De wisselende samenstelling van het brongebied leidt binnen een dag tot dezelfde variatie in verdampingsobservaties als in een jaar, zelfs als de weercondities gedurende de dag relatief constant zijn. Hoewel de analytisch geschatte brongebieden de tijdsafhankelijkheid van de observaties onthullen, ligt het bepalen van de ruimteafhankelijkheid buiten de mogelijkheden van deze aanpak. De analytische beschrijvingen verwaarlozen de onregelmatige stedelijke vorm. Numerieke, atmosferische simulaties op hoge resolutie beschrijven het stedelijke windpatroon en daarmee de totstandkoming van de brongebieden, wat het mogelijk maakt eddy-covarianties virtueel na te bootsen. Met een rooster van virtuele eddy-covariantie systemen is het mogelijk de invloed van de sensorlocatie te bepalen. De virtuele eddy-covariantie systemen verliezen hun correlatie binnen 280 m. Dit illustreert hoe specifiek eddy-covariantie metingen zijn voor de exacte locatie van de sensor en hoe ze niet representatief zijn voor een hele buurt. Daarom moet het brongebied meegenomen worden in de interpretatie van eddy-covariantie observaties, aangezien deze brongebieden sterk afhankelijk zijn van tijd en plaats.

Hoofdstuk 7 brengt alle voorgaande hoofdstukken samen met de conclusies dat 1) verdamping in stedelijke gebieden in sterke mate beperkt wordt door waterbeschikbaarheid, 2) stedelijke landoppervlaktemodellen de waterbalans niet accuraat beschrijven, 3) missende afvoerprocessen in deze modellen leiden tot onderschattingen in de afvoer, 4) het landoppervlakte cruciaal is voor verdampingsdynamiek en 5) het brongebied van eddy-covariantie systemen sterk variëren in tijd en ruimte.

Wanneer men het stedelijk klimaat onderzoekt, wordt men geconfronteerd met de “stedelijke uitdaging” die ontstaat door de heterogeniteit in stedelijke omgevingen, menselijke activiteit in stedelijke gebieden en diversiteit tussen steden. Wanneer modellen worden ingezet in deze uitdagende omgeving, zou het doel leidend moeten zijn voor hun toepassing zonder dat middelen ingezet worden alleen om hogere ruimtelijke resolutie te bereiken. De ontwikkeling van stedelijke en andere modellen zou ervan profiteren als kwantitatieve controles op fysieke limieten en consistentie een prominente rol krijgen in modelevaluaties naast een vergelijking met observaties.

De stedelijke uitdaging blijft bestaan als inzichten worden vertaald richting de praktijk. Stedelijk klimaat is onderdeel van een dynamisch en verweven systeem dat door veel wetenschappelijke disciplines wordt bestudeerd. Water symboliseert deze dynamiek en deze verwevingen omdat het de luchtvochtigheid laat stijgen, koeling door verdamping faciliteert en vegetatie in leven houdt. Leefbare steden worden gebouwd door onze omgeving te veranderen bewust van deze dynamiek en verwevenheid.

List of symbols and abbreviations

α	Albedo (-)
ΔS	Change in the water storage (mm d^{-1})
ΔS_{soil}	Change in the soil moisture storage (mm d^{-1})
$\Delta S_{intercept}$	Change in the intercepted water storage (mm d^{-1})
ΔS_{snow}	Change in the snow layer storage (mm d^{-1})
ΔQ_s	Change in the stored heat (W m^{-2})
δ_m	Maximum wet/evaporative fraction (-)
ϵ	Emissivity (-)
θ_v	Virtual air temperature (K or °C)
κ	Von Kármán constant (= 0.4)
λ	e-folding timescale (d)
ρ_a	Air density (kg m^{-3})
σ	Stefan-Boltzmann constant (= $5.67{\cdot}10^{-8}$ W m^{-2} K^{-4})
σ_{norm}	Normalized standard deviation
B	Bowen ratio (-)
c	Proportionality constant (d^{-1})
CN	Curve number (-)
c_p	Specific heat of dry air (J kg^{-1} K^{-1})
d	Water storage depth (mm)
d_m	Maximum water storage depth (mm)
DNS	Direct Numerical Simulations
DTS	Distributed Temperature Sensing
E	Evaporation (mm d^{-1})
E_p	Potential open water evaporation (mm d^{-1})
EC	Eddy-Covariance
ECMWF	European Centre for Medium-Range Weather Forecasts
ET	Evapotranspiration (mm d^{-1})
ET_0	Initial evapotranspiration after rainfall (mm d^{-1})
e	Water vapor pressure (Pa)
e_s	Saturated water vapor pressure (Pa)
f	Footprint-weighted fraction (-)

F_v	Vegetation fraction
F_{imp}	Impervious surface fraction
g	Gravitational acceleration (= 9.81 m s^{-2})
I_a	Initial abstraction (mm)
ICOS	Integrated Carbon Observation System
L	Obukhov length (m)
L_v	Latent heat of vaporization (J kg^{-1})
LCZ	Local Climate Zone
LES	Large-Eddy Simulations
LIDAR	Laser Imaging, Detection, And Ranging
$LW\downarrow$	Downward longwave radiation (W m^{-2})
MBE	Mean Bias Error
MAE_s	Systematic mean absolute error
MAE_u	Systematic mean absolute error
P	Precipitation (mm d^{-1})
P_e	Event precipitation (mm)
PILPS	Project for Intercomparison of Land Surface Parameterization Schemes
q	Specific humidity (kg kg^{-1})
Q_E	Latent heat flux (W m^{-2})
Q_f	Anthropogenic heat flux (W m^{-2})
Q_H	Sensible heat flux (W m^{-2})
Q^*	Net radiation (W m^{-2})
R	Runoff (mm d^{-1})
R^2	Coefficient of determination
RANS	Reynolds-Averaged Navier-Stokes
R_e	Event runoff (mm)
R_s	Surface runoff (mm d^{-1})
R_{ss}	Subsurface runoff (mm d^{-1})
r_0	Rainfall intensity (mm s^{-1})
ROTH	Rothenburgstraße
S	Dynamic water storage in Chapter 2 (mm) and potential maximum retention (mm) in Chapters 3 and 4
S_0	Dynamic water storage capacity (mm)
SEBAL	Surface Energy Balance Algorithm for Land
$SW_\downarrow$	Downward shortwave radiation (W m^{-2})
T_a	Air temperature (K or °C)
$t_{\frac{1}{2}}$	Half-life (d)
T_{skin}	Surface skin temperature (K or °C)
TUCC	Technical University Campus Charlottenburg
u	Horizontal wind speed (m s^{-1})
u_*	Friction velocity (m s^{-1})

UHI	Urban Heat Island
ULSM	Urban Land Surface Model
UWBR	Urban Water Balance Representation
w	Vertical wind speed (m s^{-1})
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting
z_H	Mean building height
z_s	Sensor-height above ground level

Chapter 1

Introduction

1.1 Urbanization

Everywhere humans have settled, they have changed their surroundings. People built houses to live, offices to work, stores to shop, and roads to get to all of these. Nowhere are the changes as pronounced as in cities, where more people mean more change. The changes entail more surfaces being sealed with human-made materials. At the same time, the surface diversifies bringing about a heterogeneous mosaic. These changed, urban areas are expanding worldwide providing a home for an increasing number of people (United Nations, 2018). This expansion means more areas experience a skyrocketing population density and this changes the Earth's surface almost beyond recognition.

1.2 Surface energy balance

The changed, urban surface interacts differently with the atmosphere than a more natural surface (Oke et al., 2017). To explain these differences, the surface energy balance is a good starting point. Before jumping to the urban surface energy balance, let us examine the natural situation. The natural and the urban balances are illustrated in Figure 1.1. The radiation balance is directly linked to the energy balance, as the radiation balance describes the available energy. The radiation balance and the general surface energy balance read:

$$Q^* = (1-\alpha)SW_{\downarrow} + \epsilon LW_{\downarrow} - \epsilon\sigma T_{skin}^4 = Q_H + Q_E + \Delta Q_s \tag{1.1}$$

where Q^* is the net radiation. Q^* can be derived from the incoming and outgoing radiation components using the surface albedo (α), downward shortwave radiation ($SW_{\downarrow}$), the surface emissivity (ϵ), downward longwave radiation ($LW_{\downarrow}$), the Stefan Boltzmann constant (σ, $5.67{\cdot}10^{-8}$ W m^{-2} K^{-4}), and the surface skin temperature (T_{skin}). The derived Q^* is divided over the energy components being the sensible heat flux (Q_H), the latent heat flux (Q_E), and the change in the stored heat (ΔQ_s). Normally, the advected heat flux is also included. This flux is assumed to be minimal compared to the other heat fluxes and is thus not addressed in this thesis. The difference with the urban energy balance may seem minimal at first glance:

$$Q^* + Q_f = Q_H + Q_E + \Delta Q_s \tag{1.2}$$

where only the anthropogenic heat flux (Q_f) is added. All terms in Eq. 1.1 and 1.2 are expressed in W m^{-2}. The fluxes on the right side of both equations are positive when leaving the surface. The additional Q_f is mainly important in areas with a high population density, heavy traffic, or high energy consumption (Allen et al., 2011). In these areas, Q_f estimates are as high as 400 W m^{-2} (Ichinose et al., 1999). In most places, Q_f is lower and rarely rises above 30 W m^{-2} (Chow et al., 2014a; Bonifacio-Bautista et al., 2022). Apart from this additional flux, the altered surface affects the fluxes occurring in urban

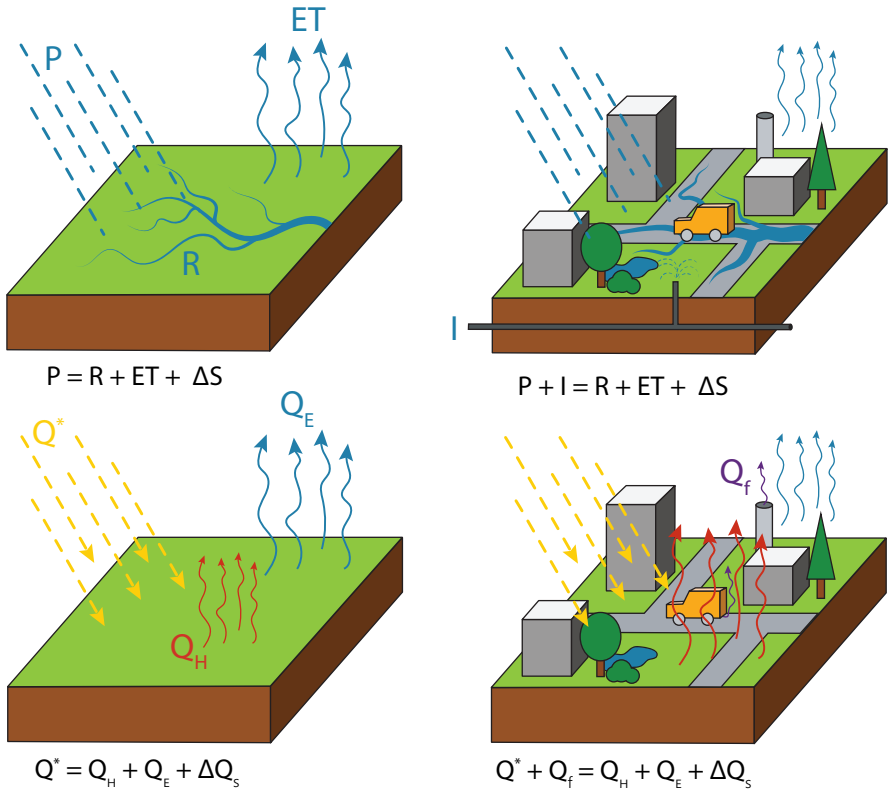


Figure 1.1: Conceptual figure showing how the water and energy balance differ between natural and urban areas. The storage is not visualized and the advected heat flux is omitted from the energy balance.

and natural situations. The modified fluxes include both the radiation components and the partitioning of Q^* over the energy fluxes Q_H, Q_E, and ΔQ_s.

The urban surface influences both incoming and outgoing radiation. Urban areas are more polluted than their surroundings. The resulting high atmospheric concentration of particulate matter may prevent shortwave radiation from reaching the surface (Oke et al., 2017). From the shortwave radiation reaching the surface, less is reflected into the atmosphere because of the low urban α. Street canyons trap radiation as it is not directly reflected towards the atmosphere. Instead, the radiation interacts with multiple surfaces before it is reflected out of the canyon. With each interaction, the urban fabric absorbs part of the radiation leading to a lower α of the urban fabric. This α is further lowered by the low α of the human-made materials and the canyon structure of the urban fabric. The radiation trapping is also relevant for longwave radiation. The longwave radiation is

further affected by the different ϵ of urban and natural materials. Long- and shortwave radiation trapping is affected by the sky view factor, i.e. how much of sky the surface can see. This determines how much shortwave radiation directly enters and how much longwave radiation escapes the canyon. Altogether, the altered radiation components result in a different Q^*.

Like the magnitude, the partitioning of Q^* depends on the surface even though mostly the same energy fluxes (Q_H, Q_E, and ΔQ_s) are present in natural and urban settings. The partitioning between Q_H and Q_E is often expressed in the Bowen ratio (β), defined as Q_H over Q_E. In natural areas, Q_E is generally higher than Q_H meaning β is below 1. Depending on the urban morphology and hydrometeorological conditions, β varies between 0.2 and 0.8. The lion's share of Q_E comes from vegetation. Hence, the partitioning in a natural environment depends on vegetation characteristics, such as the level of stomatal control and its leaf area, and the weather conditions, mainly $SW_{\downarrow}$ and the water availability.

In contrast to natural areas, Q_E is typically smaller than Q_H over urban surfaces resulting in a β higher than 1. Impervious surfaces like concrete and asphalt limit Q_E as they seal the surface. The remaining vegetation is still responsible for a large part of Q_E, while impervious surfaces contribute to evaporation after rainfall. Less Q_E leaves more energy for heating the atmosphere through Q_H.

Next to Q_H, energy not going to Q_E can be stored in the urban fabric (ΔQ_s). While ΔQ_s is relatively small in natural landscapes, 17-58% of Q^* may be stored in the city as a result of the high heat capacities and thermal conductivities of building materials (Grimmond and Oke, 1999). ΔQ_s is positive during the day when energy is stored, which reverses when the sun goes down. The release of stored heat prevents effective cooling in the evening and night keeping temperatures high after the sun has set.

So far, the focus has been on the modified surface energy balance. Buildings in urban areas also change the roughness of the surface and trap heat by preventing ventilation. The lack of ventilation and the changed energy balance together are responsible for higher air temperatures at pedestrian levels in the city than in its surroundings: the Urban Heat Island (UHI; Oke, 1982).

The UHI is more than a theoretical difference between the city and its surroundings; it affects the life and health of people in the city. The most severe impact is excess mortality (Gasparrini and Armstrong, 2011; Huang et al., 2023), which is most profound during heatwaves. The European heatwave in 2003 is a striking example with excess mortality surpassing 80.000 in Europe (Robine et al., 2008) and 15.000 in France alone (Fouillet et al., 2006; Laaidi et al., 2012). Although less extreme than increased mortality, heat stress also leads to declining labor productivity. The high nighttime temperatures prevent recovery from daytime heat and incite sleep deprivation. Sleep deprivation magnifies the heat stress-related health issues (Thommen, 2005) and further harms productivity

(Rosekind et al., 2010). The higher temperatures also increase energy consumption for cooling (Santamouris et al., 2015; Li et al., 2019b).

Solution frameworks have been proposed to mitigate heat stress and improve the quality of life in cities. The frameworks have varying names, such as low-impact development (Qin et al., 2013), sponge cities (Gaines, 2016), and sustainable drainage systems (Zhou, 2014). Despite the varying names, all frameworks revolve around the principle of increasing water storage to reduce runoff and promote Q_E. To achieve this, measures are proposed that vary from infiltration ponds to green roofs. In a broad sense, they often include nature-based solutions and vegetation as a pivotal part of their strategy. The frameworks aim to divert energy from Q_H to Q_E. However, models do not accurately reflect the cooling potential of vegetation, especially during the health-deteriorating heat waves (Gao et al., 2024). To formulate more effective plans, efforts to increase Q_E should not consider solely the energy balance but also account for the water balance.

1.3 Water balance

With only the energy balance, the dynamics of Q_E cannot be understood. After all, Q_E appears in the water balance as evaporation indicating water affects Q_E as well. Let us compare the water balance in natural and urban areas that are both shown in Figure 1.1. The general water balance reads:

$$P = R + \mathrm{ET} + \Delta S \tag{1.3}$$

where P is precipitation expressed in mm d^{-1}. All terms are expressed in the same unit. R is runoff, which can be separated into surface and subsurface flow. ET is evapotranspiration and the water equivalent of Q_E. ΔS is the change in the water storage, which includes intercepted water, open water, and subsurface moisture. Like with the energy balance, only one extra flux gives the urban water balance:

$$P + I = R + \mathrm{ET} + \Delta S \tag{1.4}$$

where I, irrigation, is the extra flux. In addition, this flux may represent any anthropogenic moisture input. Anthropogenic inputs could include subsurface flow from leaking pipes (Price, 2011), and interbasin sewer and drinking water fluxes (Hopkins et al., 2014). I describes only irrigation in this thesis.

As with the energy fluxes, the urban surface alters the partitioning of the water fluxes (Oke et al., 2017). Impervious surfaces start to increase runoff when they cover as little as 10% of the surface (Oudin et al., 2018). This increase leads to larger runoff volumes and shorter response times (Jacobson, 2011; McGrane, 2016). Higher runoff volumes leave less water available for ET, while water availability is directly relevant to the effectiveness of the earlier-mentioned solution strategies.

On top of the link with the energy balance, the urban water balance is relevant on its own. The larger runoff volumes and relatively short response times lead to urban flooding (Mignot et al., 2019) and rapid transport of contaminants (Deletic and Maksimovic, 1998; Huang et al., 2010). Flooding is a disaster that impacts people, the environment, and the economy (Hammond et al., 2015). More people will be exposed to flooding with continued urbanization, especially in vulnerable places like floodplains (Cao et al., 2022).

1.4 Evaporation

Evaporation is the bridge between the water and energy balance and the suggested road towards cooler cities. All energy used for evaporation can after all no longer heat the air. The connecting role warrants a more detailed description of the process. Evaporation is one flux but consists of transpiration from vegetation and direct evaporation from intercepted water, bare soil, and open water. Evaporation of water requires three drivers to be present. Evaporation's presence in the water and energy balance gives away the first two drivers. Next to water and energy, evaporated water vapor has to be transported away from the surface. Transport prevents the air close to the surface from saturating. The absence of water, energy, or transport limits evaporation, or, in other words, their availability drives evaporation. How these drivers affect evaporation depends on the surface cover.

How the relative importance of the drivers depends on the specific conditions is pictured in Figure 1.2. The most intuitive example is probably that open water is completely insensitive to water availability, as water is abundantly available. Open water also does not respond directly to energy availability as solar radiation can penetrate the water surface and this energy is stored in the water column. The lack of energy balance closure at the surface makes open water respond strongly to transport availability. Peatland swamps are very wet and in that way similar to open water. As a result, these swamps have the same insensitivity to water availability. However, the soil and land vegetation make this surface more responsive to energy availability.

The abundance of water makes open water and swamps uniquely insensitive to water availability. How are vegetated, non-soaked surface covers different? Logically, water availability becomes relevant. Yet, energy availability is the strongest driver of their evaporation. From grasslands to croplands and forests, the vegetated surface covers shift from following energy to following water availability.

Urban landscapes are peculiar as they may contain all the previously discussed surface covers and impervious surfaces. As a result of the impervious surfaces, urban areas are more quickly limited by water availability. Transport also plays an important role, while energy availability explains much less evaporation variation.

Having discussed the drivers of evaporation, I want to make a short note on the terminology around evaporation. So far, I have consistently called the water equivalent of Q_E

evaporation rather than evapotranspiration (ET) except in the water balance (Eq. 1.1 and 1.4). In the remainder of this thesis, ET is used when evaporative fluxes are made up of transpiration and at least one of interception, soil, and open-water evaporation. Fluxes not including transpiration are called evaporation and fluxes solely consisting of transpiration are, logically, called transpiration.

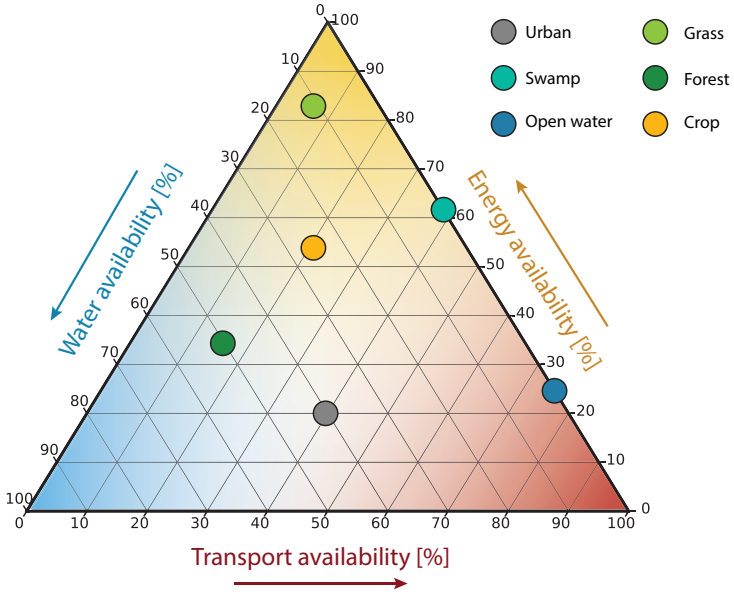


Figure 1.2: The drivers of evapotranspiration for different surface cover types based on observations from The Netherlands. The shading indicates which driver is dominant: blue for water availability, yellow for energy availability, and red for transport availability. The figure is adapted from Jansen et al., 2023.

1.5 Observing evaporation

The strategies to mitigate urban heat stress hinge on ET. To further the understanding of ET, one may analyze observations of ET. Different techniques are available to measure ET at a range of spatial and temporal scales. This section gives a concise overview starting with techniques suitable for small spatial scales and gradually moving larger. At the smallest scale, the available techniques are the evaporation pan and lysimeter.

Evaporation pans rely on measuring water loss from a pan filled with water (McMahon et al., 2016). Evaporation is inferred by periodically measuring the water level and correcting for precipitation. This makes the evaporation pan a straightforward and cheap observation

method requiring limited additional data. The pan captures the effect of the available energy and transport, but water availability is by definition not accounted for (Kahler and Brutsaert, 2006). Another disadvantage is that the heat storage dynamics of a pan are incomparable to larger water bodies with much more thermal inertia. Moreover, water may leave the pan by wind instead of evaporation (Sumner and Jacobs, 2005; Masoner and Stannard, 2010). This technique has occasionally been applied in urban areas (Ostos and Luyando, 1998; Pearlmutter et al., 2009).

Lysimeters are in a way similar to evaporation pans. They track the water content of a set volume (Payero and Irmak, 2008; Fank, 2011). Instead of a pan filled with open water, lysimeters are cylinders filled with soil. Lysimeters vary in size, shape, and (vegetation) cover. Two methods exist to keep track of the water in the cylinder. The first method weighs the entire cylinder, while the other determines the residual of the in- and outgoing water fluxes of the soil water balance (precipitation and drainage). Even though the soil is disturbed during the installation of a lysimeter, the weighing version is regarded as one of the best ways to measure ET on a small scale. In the urban context, lysimeters are mainly applied to study small-scale interventions (Denich and Bradford, 2010; Ouédraogo et al., 2022; Zhang et al., 2023).

More recently, Distributed Temperature Sensing (DTS) has been developed to measure temperature along fiber-optic cables. These temperature observations allow the estimation of turbulent heat fluxes including ET when they provide the dry and wet-bulb temperatures in a vertical temperature profile (Euser et al., 2014; Schilperoort et al., 2018). These observations can be achieved by installing two cables next to each other and wrapping one in a wet cloth. The Bowen ratio can be calculated from the temperature and water vapor gradients derived from the temperature profiles. In the end, the turbulent heat fluxes are obtained by constraining the energy balance with this Bowen ratio. The design of the cable configuration determines the exact scale of DTS observations. DTS has been applied at scales starting at a few meters (Van Emmerik et al., 2013) up to several kilometers (Dai et al., 2023), although observations have until now been limited to dry temperature at the larger scales.

Ground- and satellite-based remote sensing utilize light to estimate ET at larger scales. The ground-based remote-sensing technique is called scintillometry and consists of a light-transmitting and a light-receiving device (Jacobs et al., 2015; Ward, 2017). The transmitted beam is refracted by small optical variations caused by turbulent eddies along its path. The detected variations in the beam thus characterize the turbulent airflow in the atmosphere. The turbulence transports water vapour and energy, i.e. the sensible and latent heat flux. Scintillometry allows to estimate both when the variations in two wavelengths, optical and microwave, are observed simultaneously by the same or two synchronized scintillometers. Given the path lengths of hundreds to thousands of

meters, the fluxes represent area averages of several square kilometers (Hartogensis et al., 2003).

Satellite-based remote sensing operates on a larger spatial scale enabling studies of ET on a regional to global scale (Zhang et al., 2016). Similar to scintillometers, satellites measure light rather than ET directly. However, they rely on passively sensing the radiation reflected or emitted by the surface instead of measuring actively transmitted beams. Even though this allows ET at vast scales, satellites cannot provide observations when clouds block the visible and thermal spectrum. These spectra are needed for the different methods that translate the satellite-observed radiation to ET. The methods take different starting points. The earliest and still commonly used method is based on the energy balance. The latent heat flux is estimated as the residual of the energy balance after the sensible heat flux is calculated from the temperature derived from the infrared radiation measurement. A well-known algorithm of this type is SEBAL (Surface Energy Balance Algorithm for Land), which includes empirical relations to account for hydrometeorological heterogeneity (Bastiaanssen et al., 1998). This algorithm has been extended with for example automatic calibration with surface-based ET observations (Allen et al., 2007). Other starting points are the ET equations from Penman-Monteith and Priestley-Taylor. To estimate ET from these equations, the necessary input variables and parameters are derived from satellite products when possible and otherwise modeled (e.g. Mu et al., 2011). Therefore, the main advantage of the Priestley-Taylor equation is its limited number of inputs (Miralles et al., 2011). Finally, ET can be estimated with satellite remote sensing from other starting points such as the water balance (Wan et al., 2015), the link between ET and carbon (Yang et al., 2013), and empirical relations (Wang et al., 2007).

While progressing through the scales, I have omitted the one method essential for this thesis: eddy covariance (EC). EC observations are at the neighborhood scale of more or less 1 km^2. Arguably, EC provides the most direct observation of the turbulent latent and sensible heat fluxes by measuring temperature, humidity, wind speed, and direction at a high frequency (e.g. 10 Hz) (Aubinet et al., 2012). The turbulent heat fluxes are calculated as the covariance of the vertical wind (w) and the transported scalar, potential air temperature (T_a) for the sensible and specific humidity (q) for the latent heat flux:

$$Q_E = \rho_a L_v \overline{w'q'} \tag{1.5}$$

$$Q_H = \rho_a c_p \overline{w'T_a'} \tag{1.6}$$

where ρ_a is the air density. Multiplied with L_v, the latent heat of vaporization, ρ_a represents the energy required to evaporate one cubic meter of water. Multiplied with c_p, the specific heat of dry air, ρ_a represents the energy required to heat one cubic meter of air by one degree. w is the vertical wind speed and the primes indicate fluctuations from the mean in time. The averaging period is typically 30 or 60 minutes. The EC system has to be installed in the well-mixed layer to measure a representative flux. Practically this means

sensors have to be installed above 2–5 times the average building height (Feigenwinter et al., 2012).

Despite the elegance of almost directly measuring the turbulent heat fluxes, the EC technique has limitations and drawbacks. Firstly, wet EC systems are unreliable; thus, measurements performed during or shortly after rainfall must be discarded. Secondly, EC systems fail to close the energy balance leaving part of the available energy (Q^* and the ground heat flux when negative) unexplained (Mauder and Foken, 2006; Franssen et al., 2010). Thirdly, the EC observations require the right site. Selecting a suitable site for EC in urban areas is especially challenging. Logistics and access often limit the options to existing buildings high enough to reach the well-mixed layer, but they may distort the airflow themselves (Feigenwinter et al., 2012).

An important attribute of EC observation is their source area or their footprint. The footprint depends mainly on the wind speed and direction, atmospheric stability, and measurement height (Rannik et al., 2012), as is depicted in Figure 1.3. As the wind and atmospheric stability change continuously, footprints change as well. The dynamic footprints of EC systems present both a challenge and an opportunity in the heterogeneous urban landscape. As a result of the dynamic footprints and heterogeneous surface, the observations reflect an ever-changing combination of surface covers hampering analysis. At the same time, analytical functions and numerical models allow for footprint estimation (Vesala et al., 2008b). These estimated footprints unlock the possibility of leveraging the information contained in the footprint (Vulova et al., 2021).

1.6 Urban climate models

Evaporation observations represent one instrument to understand the urban system. However, their availability is limited, especially in urban areas (Lipson et al., 2022b). Moreover, these observations represent only their (direct) surroundings and not the city as a whole. Models may partly fill this gap having no logistic constraints and thus allowing for more flexibility. Apart from removing practical constraints, urban models provide the means to forecast urban weather and warn early for urban heat which is crucial to help mitigate the adverse health effects of heat mentioned earlier (Martinez et al., 2019). On top of that, models can assist in building more resilient cities by testing how planned changes to the urban fabric affect the micro-climate before they are implemented.

The existing models cover a wide range of complexity illustrated in Figure 1.4. The geometry of a city may be simplified to a flat sheet on one extreme of the spectrum. On the other side, every building could be described separately. Many processes can be excluded or included with varying degrees of detail. These processes may be physical processes such as turbulence or human activities such as irrigation. Below a synopsis of

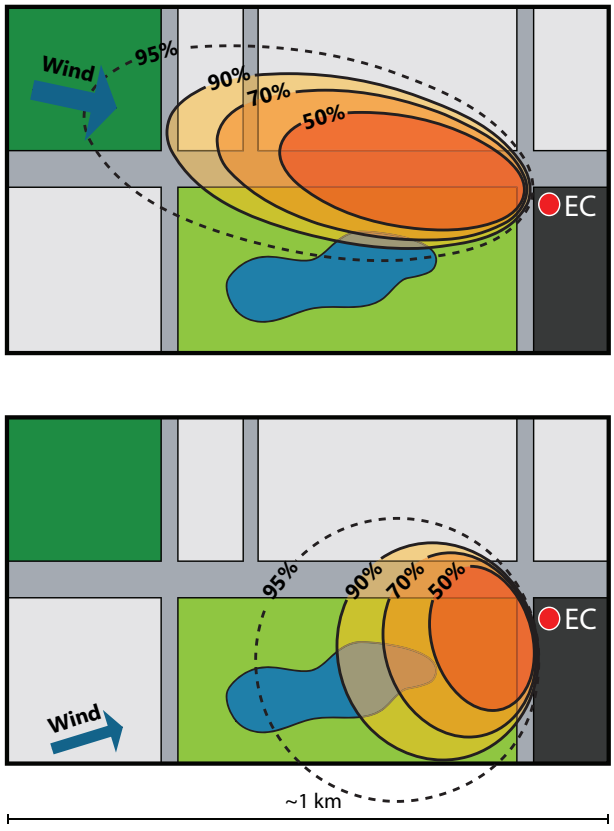


Figure 1.3: Illustration demonstrating the influence of wind speed and direction on the footprint of EC observations for two contrasting wind speeds and two different wind directions. The ellipses indicate the source area of the respective percentage of the observed flux. It is not possible to draw the 100%-contour, as the footprints are probabilistic and the analytical footprint functions are asymptotic.

the existing model types and their variations is given starting at 0- and building up to 3-dimensional models.

1.6.1 Urban land surface models

Urban land surface models (ULSMs) simulate the exchange of energy, water, and momentum between the urban surface and the overlying atmosphere. They are forced with wind, temperature, humidity, downward radiation, and precipitation from a reference height above the urban canopy. A ULSM can be run independently, i.e. offline. Offline ULSMs are useful for evaluating the schemes themselves and simulating the fluxes within the urban canopy. Alternatively, ULSMs are coupled to an atmospheric model serving as the lower boundary condition. Their significance in such online settings is increasing with

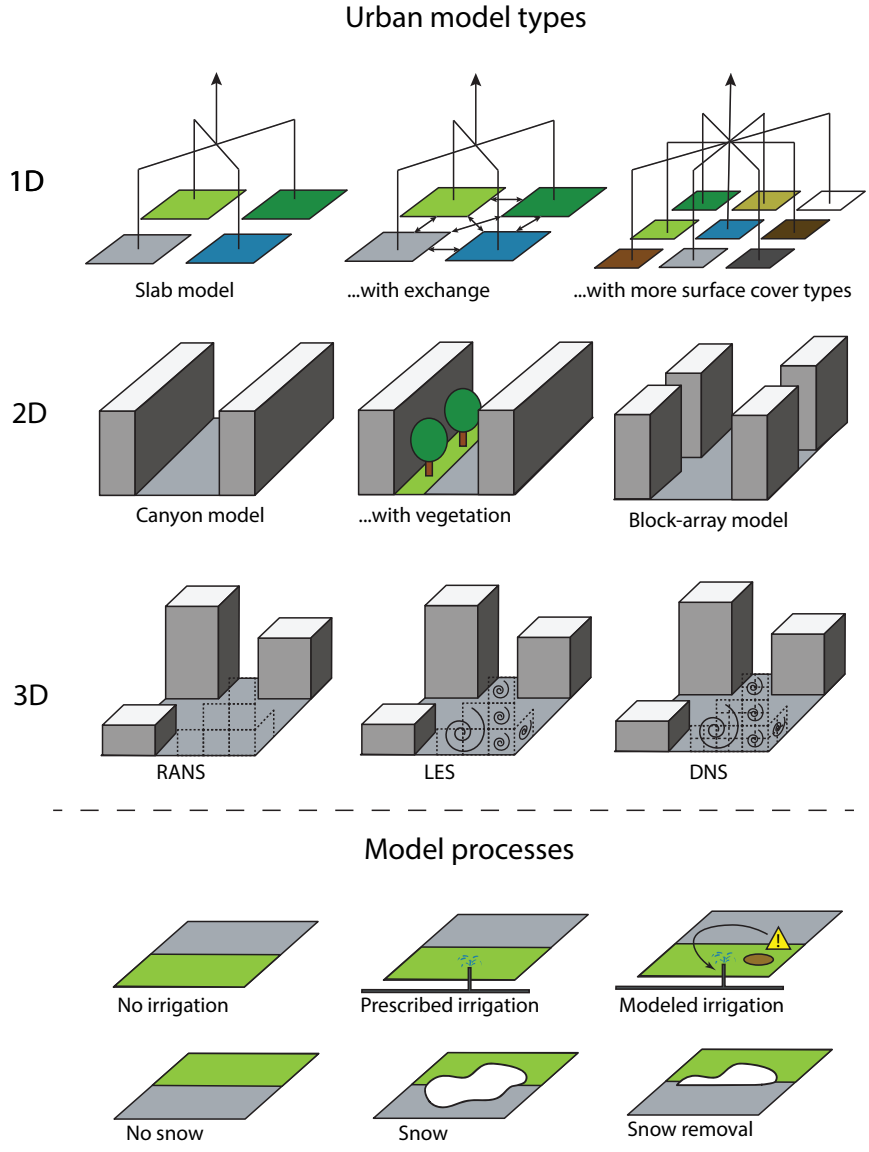


Figure 1.4: Overview of the different urban model types with increasing levels of complexity and examples of how detailed processes are potentially represented in the models. Above the dashed line, different model types are shown with increasing geometric complexity from top to bottom. The 3D models are Reynolds-Averaged Navier-Stokes (RANS), Large-Eddy Simulations (LES), and Direct Numerical Simulations (DNS). Below the dashed line, the varied complexity of specific processes is illustrated for irrigation and snow.

the evolution of higher-resolution atmospheric models, as urban areas are more often the dominant surface cover within a grid cell.

Many schemes have been developed to describe the exchanges between the urban surface and the atmosphere in the past decades (Grimmond et al., 2011; Lipson et al., 2024). These schemes are typically classified by how they represent urban geometry. In reality, these schemes cover a continuous scale of complexity representing different physical processes at different levels of detail.

Geometrically, the simplest ULSMs mimic the urban surface as a set of tiles and are called slab models. Each tile represents a certain surface cover type (e.g. grassland, deciduous forest, or bare soil). Impervious surfaces are represented by one tile only or by two tiles distinguishing roads and roofs. Depending on the model, tiles interact and exchange fluxes, or their fluxes are directly aggregated to an average flux. Slab models require bulk surface parameters for each tile such as the surface albedo and the roughness length. As ULSMs of this type consider no spatial dimensions, they do not capture the urban canyon's 2-dimensional effects like radiation trapping.

To evade this limitation, canyon models approximate the urban canyon as an infinite 2D street. The second dimension captures the interaction between urban facets (road, roof, wall). The radiation and energy balances are solved for each facet and their interaction is thus resolved capturing radiation trapping. The extra dimension adds new parameters to the model related to the canyon geometry such as building height and canyon width and related to the roof, wall, and street facets such as albedo and heat capacity. Canyon models can still include tiles next to the 2D canyon to account for the non-urban surfaces.

Although more detailed than slab models, canyon models also miss interactions. Canyon models with one atmospheric layer in the canyon do not consider the vertical gradients of wind, temperature, and humidity and their effect on turbulent transport. To resolve this, multilayer canyon models divide the atmosphere. Another simplification violating reality is the endless buildings, as nowhere in the world infinite buildings exist. Parameterizing the effect of gaps in the buildings allows to relax this assumption without moving to three full dimensions. The downsides are the higher computational burden, the increased need for spatial information to estimate parameters (e.g. material properties), and the increased need for observations to estimate the initial conditions.

Until here, I followed the convention of dividing models into geometric categories. However, countless options exist within the slab, canyon, and block models to represent relevant processes generally or in detail. A full overview of all processes and their options is better suited in a reference book on urban climate modeling than this introduction. Yet, the palette is sketched to foster understanding and appreciation for the variety found in ULSMs. For this illustration, three examples are highlighted to demonstrate how vegetation, hydrological processes, and anthropogenic fluxes may vary in complexity within the geometric categories. The three examples are canyon vegetation, snow hydrology, and

irrigation. Many more aspects are part of the ULSM pallette such as soil hydrology and the anthropogenic heat flux, but are not detailed here.

As mentioned before, urban canyon models may employ separate tiles for the non-urban vegetated surfaces. When on a separate tile, the vegetation does not interact with the 2D canyon. To capture this interaction, (part of) the vegetation can be included in the canyon. Canyon vegetation consists of both low and high vegetation. Not all vegetation is planted on the canyon floor. Green roofs and facades are suggested to combat heat stress (see Sec. 1.2). For this reason, modelers have built these into ULSMs (Meili et al., 2020; Wang et al., 2021).

Snow does not occur everywhere on our planet and models are developed in and for regions without snow. Therefore, models exist without snow hydrology. When a model accumulates snow, this can be on the urban surfaces, the vegetated surfaces, or both. The accumulated snow layer is modeled with different levels of detail in heat transfer, snow melt, and refreezing. Although not yet implemented, models could include human snow management with road salt and snow removal.

Human behavior is mimicked for irrigation. The most straightforward implementation is a prescribed irrigation flux. This prescribed flux is directly provided to the model but may be based on another model or surveys. Another option is to model the irrigation depending on the variables within the model (e.g. soil wetness).

1.6.2 Large-eddy simulations

When we add an extra dimension to canyon models, we enter the realm of urban 3D models. Three different 3D model categories are distinguished by how much turbulence is resolved. Direct Numerical Simulations (DNS) resolve all turbulent flow. To resolve all turbulence, DNS needs to cover all turbulent scales from generation to dissipation. The resulting downside is that the grid spacing has to be in the order of 1 mm. The corresponding timescales are also very short creating a massive computational burden. Currently, DNS can therefore not be applied in urban climate research. Instead of resolving all turbulence, Reynolds-Averaged Navier-Stokes (RANS) models parameterize it completely. Numerical Weather Prediction (NWP) and climate models belong in this category. While this allows much larger domains, the effect of buildings on the airflow cannot be captured properly.

Large-Eddy Simulations (LES) cover the middle ground between DNS and RANS. The net effects of the smallest turbulent scales are parameterized, while larger turbulent motions are resolved. This way turbulence is modeled at a much lower cost than DNS, but turbulence is mostly explicitly resolved. For urban areas, LES explicitly simulates the majority of the turbulent motions with a spatial resolution in the order of 1 meter. At these resolutions, LES remain restricted to several kilometers and a few days. Yet, much more spatial

detail is included than in ULSMs. In idealized simulations, the model domain is designed to reveal the effect of certain parameters on the flow or study fundamental processes through simple cases. Whereas for realistic simulations, unique buildings and urban configurations can be added throughout the model domain. This requires information on the building configuration and its properties (e.g. albedo and heat capacity) at the high model resolution. The boundary conditions of the LES domain can be generated with an NWP or academic weather model capable of simulating the large synoptic forcing.

1.7 Thesis objectives and overview

People in urban areas are exposed to more heat stress than the people living outside these areas. Strategies to mitigate this heat stress seek to cool cities and one way they aim to achieve this is promoting evaporation. Thus, water availability is critical for their success. However, this link between water and energy has rarely been the focus of urban climate studies. Thus, this thesis aims to unravel how the water balance affects the urban energy balance by studying the bridge between these balances. To achieve this aim, all fluxes of the water balance are studied with a combination of observations and models. The thesis zooms to the neighborhood-scale fluxes and their smaller-scale constituents.

This thesis contributes to the grand challenges in urban climate research formulated by the urban climate community. González et al. (2021) categorized grand challenges in urban climate in four themes: a) urban climate modeling, b) observations of the urban environment, c) cyber informatics and data management, and d) knowledge transfer and applications. The first two themes have been introduced above as they are central to the research in this thesis. In the next two paragraphs, I discuss the challenges related to these categories that play a role in this thesis. While the third theme is beyond the scope of this thesis, the thesis leverages the wealth of existing data from observations and models. In this way, it may showcase how opportunistic data use benefits the impact of existing data. The last theme summarizes the challenges of integrating the subdisciplines studying the urban climate that prohibit the effective application of generated knowledge. This describes what hampers the World Meteorological Organization (WMO) initiative *integrated urban hydrometeorological, climate, and environmental services* to translate science into services that will help to create and maintain safe, healthy, resilient, and climate-friendly cities (Grimmond et al., 2020). I will reflect on the potential of this thesis to contribute to this theme in the final chapter.

The themes of urban climate modeling and observations of the urban environment contain the current scientific challenges and developments most relevant to this thesis. Within urban climate modeling, the first challenge is presented by gaps in our understanding of how the urban fabric influences turbulent transport responsible energy, water, and momentum. These gaps hinder the cross-scale transfer of the physical fluxes and our conceptual understanding. LES is suggested as a suitable technique to fill (a part of)

these gaps. To correctly capture transport, the sources and dynamics of the transported variable need to be understood. Therefore, the second challenge lies at the bottom of the urban fabric. No answer is readily available as to how the surface should be represented. The answer to this question is obscured by the third challenge: model evaluation. While hampered by a lack of observations, model evaluations require purposeful metrics and preferably benchmarks against which models are tested. Detailed high-resolution models such as LES could substitute observations and serve as benchmarks. The last challenge concerns all of the previous three, as the models need to consider processes from many (sub)disciplines to allow predictions of high-impact events like floods and heat waves. Current research often remains limited to one discipline's processes.

Observations in the urban environment are challenged practically by logistical restraints. These restraints include, for example, permission for sensor placement and securing often costly sensors. Once installed and running sensors and their observations have to be quality controlled. Urban settings present potential problems such as obstacles blocking radiation or locally affecting wind patterns. All these challenges relate to performing observations, which cannot be addressed when opportunistically using existing data like in this thesis. Yet, reusing data is ideally suited to contribute to other challenges. Reusing data urges us to explore the potential of existing public datasets, which often remain unused. All these challenges would be more than enough, but I have saved the perhaps most pressing challenge till the last. The key question of this challenge is what an urban observation represents especially spatially. This spatial dependence is important given the urban heterogeneity. Next to the spatial mosaic, changing meteorological conditions induce temporal variations as was explained for the EC footprints earlier. How can we meaningfully interpret such temperamental observations?

In their vision paper, Gao et al. (2024) make two broad recommendations overarching these challenges. The first recommendation calls for knowledge to be generated on smaller spatial scales, i.e. neighborhood and smaller. The other recommendation states that including the nexus of urbanization, water, and energy in modeling has to mature beyond the current early stage. This thesis aligns with both ideas, as the recommended spatial scales match those studied here and the bridge between water and energy is a key interaction in the described nexus.

Below I will narrate how these challenges and developments inspired the research questions. Interwoven in this story is an outline of how these questions are answered in this thesis.

Urban water storage

Water availability has been stressed multiple times as crucial for the success of heat mitigation strategies. Here, the water-energy balance link not only lies in the shared ET but also in the water storage. The main water source is precipitation, but how much precipitation can evaporate depends on its availability (long) after rainfall has seized. In

the meantime, the water has to be stored to prevent it from being lost as runoff. The water storage capacity determines how much water can be stored. Together with its depletion rate, the capacity determines how much water is available longer after rainfall. Water availability regulates energy partitioning, which is relevant as urbanization affects this water availability.

Understanding the water storage dynamics is crucial for estimating the energy partitioning between sensible and latent heat during dry periods. Therefore, the urban water storage capacity and its depletion rate have not been quantified so far. Therefore, **Chapter 2** aims to answer the following question:

What is the urban water storage capacity? How quickly does it empty through evaporation?

Directly observing water storage is complex due to the fragmented character of the urban fabric. Storage is scattered throughout the urban surface in, for example, soil moisture, intercepted precipitation, and open water of any size. To circumvent the need to measure all these individual storages, the water storage capacity can be estimated from the neighborhood ET observed using the EC technique. ET is the only flux leaving or entering the storage in the first days after rainfall. Therefore, the water storage capacity can be estimated from how fast ET declines in this period.

Water balance in urban land surface models

Cities accommodate the majority of the world's population, but their size is modest compared to the entire globe. For a long time, their size was thus well below the grid size of climate and weather models. As a result, these models could neglect the impact of urban areas on the surface-atmosphere interaction. However, increasing computational resources resulted in higher-resolution models. No longer can the urban effect be neglected calling for the application of dedicated urban schemes for the urban grid cells. ULSMs are ready to fill this vacancy. After a global intercomparison at the beginning of the 2010s, they have been improved by including more hydrological and vegetation processes. A decade later, a new intercomparison project started to evaluate the improvements, Urban-PLUMBER.

Both intercomparisons focused on the energy balance, as ULSMs are developed from a meteorological perspective. On top of that, a lack of observations may explain why the water balance representation in ULSMs has eluded evaluation. Next to the urban water storage discussed in the previous section, other water fluxes have proven difficult to measure. These fluxes are complex to measure for the same given area. ET observations with EC have a time-varying footprint, which makes measuring runoff for the same area unfeasible. Hence, previous model intercomparisons focused on the energy balance performance. This motivates the following questions that are addressed in **Chapter 3 and 4**:

How well does the water balance representation perform in current ULSMs?

What are the effects of the water balance representation on the model performance for the surface energy balance?

How can the water balance representation in ULSMs be improved?

The answer to these questions, I examine the ULSMs that were part of the most recent ULSM intercomparison: Urban-PLUMBER. Over half of the models in this project simulate the entire water balance. These models are suitable to map the current status of the water balance representation in ULSMs. Yet, the limited observation availability hampers a straightforward evaluation with quantitative error metrics. A newly developed score is presented in **Chapter 3**, which leverages indicators of a good water balance to evaluate ULSMs in the absence of observations. **Chapter 4** follows the guidance of the previous chapter and demonstrates how the water balance representation can be improved.

Evapotranspiration from heterogeneous urban areas

Where ULSMs simulate the interaction between the surface and the atmosphere, EC systems partly observe this interaction. These observations can help to unravel the link between ET and the surface. Understanding this link is required for sustainable urban water management. Only when water is properly managed, sufficient water is available is for vegetation to mitigate the heat stress in urban areas. Water availability may become limited more frequently for more urban areas as the climate changes.

Interpretation of EC observations is hampered by the constantly changing source area. The source area of the observed ET changes dynamically because of changing wind and atmospheric conditions. Combined with the heterogenous urban fabric, the dynamic footprint leads to a non-constant composition of surface cover types. While this restricts the direct interpretability of the observations, the changing surface composition presents chances to study the effect of the surface cover. Using modeled footprints, studies have been able to map CO_2 sources in urban areas. In this endeavor, the interpretation is aided by the contrasting characteristics of CO_2 sources. While ET has spatially more continuous sources, the same idea holds the potential to further our understanding of ET at neighborhood scale. Therefore, **Chapter 5** poses the question:

How can the heterogeneous surface cover be linked to neighborhood ET?

To answer this question, an observation-driven approach is taken to disentangle the different sources of ET. The developed approach is applied in Berlin (Germany). This city offers a uniquely rich set of observations including multiple EC systems at two sites. The neighborhood-scale EC observations are linked to the surface cover through analytical footprint modeling. The connection is made from the bottom up by scaling patch-scale observations and conceptual models to the neighborhood scale. This scaling is proportional to the surface fractions in the footprint. From the top down, the EC observations are

distributed over the surfaces using a system of linear equations. In this system, each equation describes the composition of the footprint and the resulting neighborhood ET. Quantifying neighborhood ET and its components is critical for informed urban water management.

The footprint composition changes constantly through the wind and atmospheric stability. Next to these dynamic conditions, the placement of the EC system is critical given the high surface variability at small scales. The placement of EC systems is often dictated by logistic limitations, especially in cities. Many urban EC systems are installed on high buildings and otherwise need a high tower structure to observe high enough to measure a representative flux limiting the location options. This raises the question central to **Chapter 6**:

To what extent are EC observations sensitive to sensor placement in an urban setting?

The costs and logistics make it infeasible to install enough EC to study this question with observations. Therefore, an LES domain is configured and simulated for one of the same sites in Berlin as in the previous chapter. In this LES, a grid of virtual EC systems is placed in a grid around the actual observation tower. By comparing the virtual and real observations the ET in the LES can be examined. The grid shows how EC observations vary over distances between 2 and 200 meters.

While chapters 2–6 aim to answer these questions, **Chapter 7** discusses and synthesizes their main conclusions highlighting the connections between the chapters. This final chapter continues by reflecting on the lessons learned in the previous chapters interwoven with perspectives for future research opportunities to advance our understanding of the balances that bridge water and energy in the city and connect the disciplines of hydrology and meteorology.


LANDFILL

Chapter 2

Urban water storage

This chapter has originally been published as:

H. J. Jongen, G.-J. Steeneveld, J. Beringer, A. Christen, N. Chrysoulakis, K. Fortuniak, J. Hong, J.-W. Hong, C. M. Jacobs, L. Järvi, et al. (2022). "Urban water storage capacity inferred from observed evapotranspiration recession". *Geophysical Research Letters* 49 (3), e2021GL096069

Abstract

Water storage plays an important role in mitigating heat and flooding in urban areas. Assessment of the water storage capacity of cities remains challenging due to the inherent heterogeneity of the urban surface. Traditionally, effective storage has been estimated from runoff. Here, we present a novel approach to estimate effective water storage capacity from recession rates of observed evaporation during precipitation-free periods. We test this approach for cities at neighborhood scale with eddy-covariance based latent heat flux observations from fourteen contrasting sites with different local climate zones, vegetation cover and characteristics, and climates. Based on analysis of 583 drydowns, we find storage capacities to vary between 1.3–28.4 mm, corresponding to e-folding timescales of 1.8–20.1 days. This makes the urban storage capacity at least five times smaller than all the observed values for natural ecosystems, reflecting an evaporation regime characterised by extreme water limitation.

2.1 Introduction

With a large and growing share of the world population living in cities (United Nations, 2018), the impact of weather-related risks magnified by climate change, such as heatwaves and flooding (Wilby, 2007), also increases. In cities, air temperatures are typically higher than in the rural surroundings due to the Urban Heat Island effect (UHI, Oke, 1982; Santamouris, 2014; Oke et al., 2017). The UHI originates from the difference between the rural and urban energy balances due to lower albedo, radiation trapping, less vegetation, higher heat storage capacity and anthropogenic heat release (Oke, 1982). Because of its positive effect on evaporative cooling that is complemented by shading, urban vegetation is often given a central role in attempts to improve thermal comfort (Ennos, 2010). Indeed, higher vegetation fractions are associated with lower urban air and canopy temperatures (e.g. Gallo et al., 1993; Weng et al., 2004; Theeuwes et al., 2017), although in specific situations vegetation can cause higher temperatures (Meili et al., 2021). Wei and Shu (2020) showed that expanding the vegetation fraction as part of urban renewal can improve thermal comfort. However, vegetation-mediated cooling strongly depends on water availability for evapotranspiration (ET, Avissar, 1992; Manoli et al., 2020).

The generally low ET over urban areas also reflects a different water balance that makes cities more prone to flooding. A high impervious surface fraction promotes stormwater runoff, which can accumulate relatively fast (Arnold Jr and Gibbons, 1996; Fletcher et al., 2013). Consequently, high runoff ratios decrease water availability for ET, and thus indirectly contribute to the UHI (Taha, 1997; Zhao et al., 2014). Heavy rainfall in cities can lead to flood volumes that are 2–9 times higher than in rural areas (Paul and Meyer, 2001; Hamdi et al., 2011; Zhou et al., 2019), often causing considerable damage (Tingsanchali, 2012). Solutions to problems related to the urban water and energy balance have been proposed under various names such as Water Sensitive Urban Design (Wong, 2006), Low Impact Development (Qin et al., 2013), Sustainable Drainage Systems (Zhou, 2014), Sponge Cities (Gaines, 2016), and Nature Based Solutions (Somarakis et al., 2019). All these concepts promote increasing infiltration and effective storage capacity, of which the latter is crucial for their performance (Graham et al., 2004; Qin et al., 2013). Therefore, methods to assess effective storage in cities at urban landscape scale are needed.

Estimation of the urban water storage capacity is challenged by the heterogeneity of sources for ET (Sailor, 2011). Previous studies have mainly focused on ET from individual sources (e.g. Gash et al., 2008; Starke et al., 2010; Pataki et al., 2011; Ramamurthy and Bou-Zeid, 2014), as well as on their combined behavior at street or neighborhood scale (e.g. Christen and Vogt, 2004; Jacobs et al., 2015; Meili et al., 2020; Meili et al., 2021). In order to study the ET on a neighborhood scale (order of hundreds of meters to 1–2 kilometers), flux measurements through eddy covariance (EC) or scintillometry are becoming increasingly popular. Due to relatively large footprints, urban EC measurements often reflect a myriad of sources including impervious surfaces, vegetation, open water, and all other sources

of ET. Hence, in this paper, an urban surface is defined as the entire urban landscape found within the footprint, rather than impervious surface only. This is in line with many studies on urban ET from an EC perspective, since the ET sources cannot be separated (e.g. Coutts et al., 2007b; Vulova et al., 2021). In contrast, modeling-oriented studies are able to make this separation and thus often use urban and impervious interchangeably (e.g. Masson, 2000; Wouters et al., 2015). Examples of cities for which EC measurements have been studied are Arnhem (Jacobs et al., 2015), Basel (Christen and Vogt, 2004), Helsinki (Vesala et al., 2008a), Melbourne (Coutts et al., 2007b), Seoul (Hong et al., 2019) and Singapore (Roth et al., 2017). Under water-limited conditions, ET observations contain information on storage (Teuling et al., 2006). In one of the few studies directly linking urban ET and storage, Wouters et al. (2015) applied this principle to validate a new parametrization for the impervious contribution to urban water storage in Toulouse. However, the link between ET and footprint-scale urban water storage remains largely unexplored.

Recession analysis can be used to link eddy-covariance flux observations and storage properties. From the 1970s, discharge recession analysis has been extensively used in groundwater and hillslope hydrology (e.g. Brutsaert and Nieber, 1977; Kirchner, 2009; Troch et al., 2013). Similarly, daily ET values can be linked to water storage during a drydown, a period without precipitation creating water-limited conditions. Assuming that the ET decay is exponential, the e-folding time, or the timescale over which ET declines by 63%, reflects the available storage and resilience to droughts (Wetzel and Chang, 1987; Salvucci, 2001; Saleem and Salvucci, 2002). Since the storage is inferred directly from ET observations, this water storage is defined as the dynamic water storage capacity available to the atmosphere for ET, which includes soil moisture, intercepted precipitation, groundwater and open water varying from lakes to puddles. As a result of plant-physiological processes, this storage is not necessarily constant (Dardanelli et al., 2004). In studies using daily ET over natural ecosystems, Teuling et al. (2006) and Boese et al. (2019) found timescales ranging from 15 days for short vegetation to 35 days for forest ecosystems, and corresponding storage capacities of 30–200 mm, with most sites in the range of 50–100 mm. A global-scale analysis of surface soil moisture recession by McColl et al. (2017) found timescales ranging from 2 to 20 days. Although valuable insight can be obtained from a comparison of urban and rural ET dynamics, recession analysis has not yet been applied to urban ET.

This study extends the methodology developed by Teuling et al. (2006) to estimate footprint-scale water storage capacity directly from EC observations of daily ET in cities without modeling ET itself. The methodology is applied to a new, unique collection of urban ET data containing cities in a range of climate conditions and with different urban land cover and structure. This allows for a first assessment of urban storage capacity across cities, an evaluation of how site characteristics (e.g. vegetation fraction) affect water storage, and a comparison of urban water storage to that of natural ecosystems.

2.2 Data and Methods

We analyze latent heat fluxes and auxiliary meteorological observations from eddy covariance flux towers at fourteen sites in twelve cities to estimate water storage. Table 2.1 lists a number of important site characteristics, including key references. In these references, all observation sites and measurement details are fully described. The sites were selected based on the length of the data record (minimum of a year), flux footprints representing typical urban neighborhoods without other land covers, and the availability of observed precipitation and latent heat fluxes. All sites are located in reasonably flat terrain. Most sites were located in mid-latitude climates, except Mexico City with a subtropical climate, Singapore with a tropical climate, and Helsinki, Łódź and Seoul with a continental climate. Vegetation fractions in the associated footprints vary between 6–56%.

Observations were reported in averaging periods of 10–30 min depending on the measurement protocol of each site. We used hourly averages to determine the timing of rainfall and 24-hour averages for the recession analysis. For all sites, the quality control of the observed heat fluxes was performed by individual researchers responsible for their ET flux observation site. Although the exact methodology of the quality control differs per site, all fluxes have been properly tested in accordance with procedures published in literature (Aubinet et al., 2012).

During multi-day drydowns in urban areas without rainfall, runoff is typically minimal after a steep peak shortly after rainfall (Walsh et al., 2005b; Fletcher et al., 2013). Therefore, the evolution in landscape-scale dynamic storage (S) over the whole drydown can be simplified as:

$$\frac{dS(t)}{dt} = -\mathrm{ET}(t) \tag{2.1}$$

Under water-limitation, daily ET becomes a function of storage. For impervious surfaces in cities, the storage dynamics have been described by a $\frac{2}{3}$-power function resulting in depletion within a few hours of daytime (Masson, 2000; Ramamurthy and Bou-Zeid, 2014). ET from other sources will likely show different behavior (Granger and Hedstrom, 2011; Nordbo et al., 2011), with ET from (urban) vegetation behaving more as a linear reservoir (Williams and Albertson, 2004; Dardanelli et al., 2004; Peters et al., 2011). Since impervious surfaces are typically quickly depleted, open water is constant and vegetation behaves more linear, we assume the flux footprint reflecting a mixture of different ET sources to effectively behave as a linear reservoir:

$$\mathrm{ET}(t) = f(S(t)) = cS(t) \tag{2.2}$$

Table 2.1: Site characteristics. The climate statistics are long-term means (1999–2019). (LCZ (Stewart and Oke, 2012): 1 = compact high-rise, 2 = compact mid-rise, 3 = compact low-rise, 5 = open mid-rise, 6 = open low-rise, F_v: surface fraction covered by vegetation in a 500 m radius around the measurement site, z_s: height of sensors above ground level, z_H: mean building height, ET_0: initial evapotranspiration.

City	Lat. (deg)	Lon. (deg)	Köppen-Geiger climate	Avg. temp. (deg C)	Ann. prec. (mm)	LCZ	F_v (%)	z_s	z_H (m)	Start	End	Source
Amsterdam	52.37	4.89	Cfb	9.2	805	2	15	40	14	05-2018	10-2020	Ronda et al. (2017) Steeneveld et al. (2020)
Arnhem	51.98	5.92	Cfb	9.4	778	2	12	23	11	05-2012	12-2016	Jacobs et al. (2015)
Basel (AESC)	47.55	7.6	Cfb	10	778	2	27	39	17	06-2009	12-2020	Lietzke et al. (2015)
Basel (KLIN)	47.56	7.58	Cfb	10	778	2	27	41	17	05-2004	12-2020	Schmutz et al. (2016)
Berlin (ROTH)	13.32	52.46	Cfb	9.1	570	6	56	40	17	06-2018	09-2020	Vulova et al. (2021)
Berlin (TUCC)	13.33	52.51	Cfb	9.1	570	5	31	56	20	07-2014	09-2020	Jin et al. (2021) Vulova et al. (2021)
Helsinki	60.33	24.96	Dfb	5.1	650	6	54	31	20	01-2006	12-2018	Vesala et al. (2008a) Karsisto et al. (2016)
Heraklion (HECKOR)	35.34	25.13	Csa	17.8	464	3	12	27	11.3	Nov-16	May-21	Stagakis et al. (2019)
Łódź	51.76	19.45	Dfb	7.9	564	5	31	37	11	07-2006	09-2015	Fortuniak et al. (2013)
Melbourne (Preston)	-37.73	145.01	Cfb	14.8	666	5	38	40	6	08-2003	11-2004	Coutts et al. (2007b) Coutts et al. (2007a)
Mexico City	19.4	-99.18	Cwb	15.9	625	2	6	37	9.7	06-2011	09-2012	Velasco et al. (2011) Velasco et al. (2014)
Seoul	37.54	127.04	Dwa	11.9	1373	1	40	30	20	03-2015	02-2016	Hong et al. (2019) Hong et al. (2020b)
Singapore	1.31	103.91	Af	26.8	2378	3	15	24	10	03-2013	03-2014	Velasco et al. (2013) Roth et al. (2017) Harshan et al. (2017)
Vancouver	49.23	-123.08	Csb	9.9	1283	6	35	28	5	05-2008	07-2017	Christen et al. (2011)

in which $c = 1/\lambda$ is a proportionality constant. Combining Eq. 2.1 and Eq. 2.2 and solving the differential equation leads to an exponential response of ET:

$$\mathrm{ET}(t) = \mathrm{ET}_0 \exp\left(-\frac{t - t_0}{\lambda}\right) \tag{2.3}$$

where λ is the e-folding timescale, and ET_0 the initial ET. With these parameters the total dynamic storage volume S_0 in mm that would be depleted during a complete dry down ($t \to \infty$) is given by:

$$S_0 = \int_{t_0}^{\infty} \mathrm{ET}(t)\mathrm{d}t = \lambda \mathrm{ET}_0 \tag{2.4}$$

so that S_0 can be estimated by fitting observed ET in time during a drydown, without modeling the flux. Essentially, the storage capacity reflects the sum of water leaving the system as ET. Because of this direct inference without an imposed model structure, the shape of the fit has minimal influence on the results. To further tailor this concept to urban environments, the anthropogenic moisture flux can be included. This flux can contribute substantially to ET, in particular during long, dry periods (Grimmond and Oke, 1986; Moriwaki et al., 2008; Miao and Chen, 2014), and includes processes like transport, heating, cooling (indoor), human metabolism and irrigation, which do not directly depend on rainfall. Variation in the daily averages of these processes, except for irrigation, can be expected to be negligible over the course of one drydown. Thus, to account for these processes we added a constant base term to Equation 2.3. Since this yields parameters in compliance with the requirements explained below for only one drydown, we conclude that including this part of the anthropogenic moisture flux does not improve the physical representation of the city. As mentioned earlier, irrigation cannot be expected to be constant, while in some cities (e.g. Vancouver (Grimmond and Oke, 1986; Järvi et al., 2011) and Melbourne (Barker et al., 2011)) its contribution to ET can be considerable during long dry periods. We include two steps to prevent irrigation affecting the results. First we exclude irrigation by limiting drydowns to the first 10 days. This also reduces the influence of the smaller signal-to-noise ratio in the tail of the drydown on ET_0. Second we require an R^2 >0.3, in order to ensure a decreasing ET tendency reflecting storage as a main control on ET dynamics. The results converge until $R^2 \approx 0.3$ (not shown), which shows drydowns with a lower R^2 are less reliable.

To estimate the parameters λ and ET_0, we identified all periods without precipitation for at least three continuous days, the minimum requirement for an exponential fit (Figure 2.1). In order to preserve the information in ET during the first hours after rainfall (in case of low λ), we start the 24-hour averaging bins directly after the rainfall event, regardless of its magnitude. The bin-average is assigned to the middle of the day (e.g. the first bin is assigned to 0.5 day since rainfall). We exclude hours with an average shortwave

incoming radiation below 10 $\mathrm{W\,m^{-2}}$ (i.e. nighttime), since nighttime ET tends to be low. No gap-filling was applied, and only bins with at least 70% of data for daytime hours were analyzed. For the longest time series (Basel (KLIN)), requiring 70% instead of 100% increased the sample size by 48% respectively, while the median of the water storage capacities only changed by 25%. Further lowering the threshold did not increase data availability. Given the minimal effect on the results and potential to increase the sample size, 70% provides more information especially regarding cities with a shorter measurement period without compromising the results.

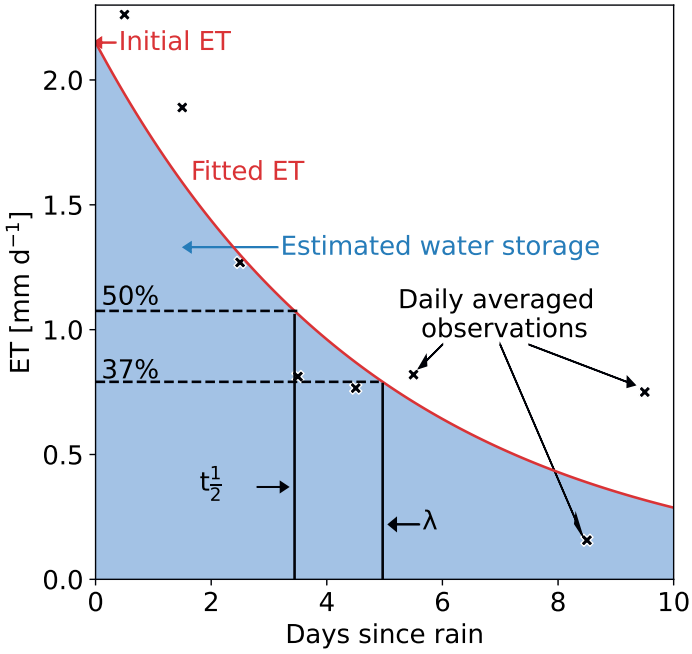


Figure 2.1: Illustration of the recession analysis. 24-hour aggregated ET versus the number of days following the last hour of precipitation for an example drydown from the Seoul data set with the fitted recession curve. Note that the fit was obtained by a linear fit on log-transformed data (see Data and Methods). In the figure, the parameters are indicated.

To allow for a variable timescale caused by a (seasonally) changing energy availability, we estimate λ and $\mathrm{ET_0}$ for every individual drydown. The parameter estimates result from linear fits (method of least squares) through the log-transformed ET observations effectively applying Equation 2.3. In addition, the parameters are required to be physically plausible meaning positive λ and $\mathrm{ET_0}$, but below 35 days (maximum found by Teuling et al. (2006)) respectively 10 $\mathrm{mm\,d^{-1}}$. The maximum timescale prevents estimation of timescales much longer than the maximum drydown duration and storage estimates based on a limited dynamical range in ET. Given this filtering only excludes 10 cases, it does

not influence our conclusions. Also, the average temperature during a drydown needs to exceed 0°C to exclude snow conditions, which is strict enough, confirmed by a check against snow records. To quantify the uncertainty of the estimated parameters, we applied bootstrapping using 5000 re-samples containing 90% of the estimates. The confidence interval is defined as the 5^{th} and 95^{th} percentile of the median distribution from the re-samples.

With λ and ET_0 the storage capacity is calculated according to Equation 2.4 (shaded area in Figure 2.1), as we assume the storage to be completely filled after every rainfall event. This assumption is supported by the absence of a dependency between the parameters and pre-drydown rainfall. Drydowns from all seasons are included and analyzed for a seasonal effect, since the water storage available to the atmosphere may change due to for example leaf phenology. Since it is not feasible to measure the water storage capacity in a complete urban footprint, this methodology offers the most direct estimation of the urban water storage. To investigate the possible impact of day-to-day variation or change in energy availability on the results, we repeated the recession analysis based on evaporative fraction (Gentine et al., 2007) multiplied by the average available energy over the drydown, which we included in the supplementary information (Table A.2.1 and Figure A.2.1 and A.2.2).

2.3 Results

In Figure 2.2, the individual drydowns (in grey) show a good resemblance of the characteristic behaviour of the recession confirming the exponential behaviour. In general, ET is quickly decaying within days after rainfall in all LCZ's represented in our sample, indicating urban ET is generally strongly limited by water availability even on the first day after rainfall. As all cities respond approximately similarly, this confirms the qualitative, decaying relation during a drydown. At some sites (e.g. Amsterdam), ET sometimes rises after 6–7 days, which is most likely due to higher ET rates during the fewer events of a duration longer than 6–7 days. The spread of the observations is higher than the uncertainty, which is the result of a seasonal dependency. The uncertainty is visibly higher in cities with shorter measurement periods, since shorter periods inevitably mean smaller samples of drydowns. For Arnhem, Basel (both), Berlin (both), Helsinki, Łódź and Vancouver, observations are available for more than two full years resulting in narrow uncertainty bands. Conversely, the uncertainty bands for the sites with records shorter than two years (Amsterdam, Melbourne, Mexico City, Seoul and Singapore) are as wide as the range of observations. In some panels (e.g. Amsterdam and Helsinki), we observe two groups of curves with distinct slopes, for which we found no explanation in seasonality, energy availability, temperature and pre-drydown rainfall (amount and timing).

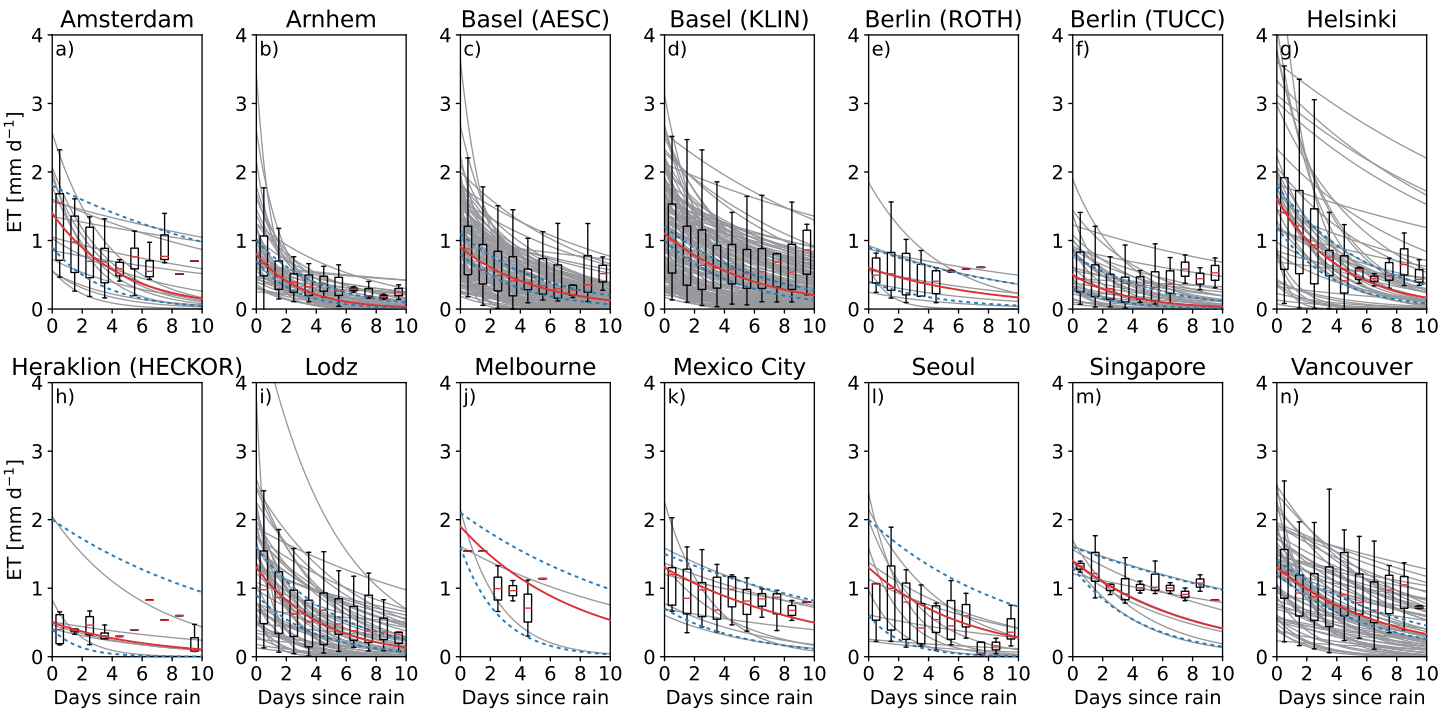


Figure 2.2: Daily average ET versus the day since the last precipitation with in red (continuous) the recession curve using the median parameter values, in blue (dotted) the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile of the median distribution from the bootstrapping re-samples, and in grey all individual drydowns. The boxplots show the spread of the observations. The parameters of the fitted curves are shown in Table 2.2. Since the parameters are based on individual drydowns, they do not necessarily follow the trend of the distributions.

Table 2.2: Summary of regression analysis. The indicated ranges for the parameters are the 5^{th} and 95^{th} percentile of the median distribution from the bootstrapping re-samples with in brackets the median itself. (ET_0: initial evapotranspiration, λ: e-folding timescale, $t_{\frac{1}{2}}$: half-life, S_0: effective, dynamic water storage capacity), R^2: median goodness-of-fit.

City	Dry-down	Days	ET_0 (mm d^{-1})	λ (day)	$t_{\frac{1}{2}}$ (day)	S_0 (mm)	R^2
Amsterdam	15	61	0.9 – 1.8 (1.4)	3.4 – 16.4 (4.5)	2.4 – 11.3 (3.1)	5.0 – 17.0 (7.3)	0.66
Arnhem	46	183	0.7 – 1.0 (0.8)	2.5 – 4.2 (3.0)	1.8 – 2.9 (2.1)	2.3 – 3.8 (3.0)	0.72
Basel (AESC)	120	500	0.8 – 1.0 (0.9)	4.2 – 5.6 (5.1)	2.9 – 4.0 (3.5)	3.6 – 4.9 (4.4)	0.75
Basel (KLIN)	158	661	1.0 – 1.2 (1.1)	4.9 – 6.8 (5.9)	3.4 – 4.7 (4.1)	5.4 – 7.8 (6.5)	0.72
Berlin (ROTH)	7	33	0.4 – 0.9 (0.6)	4.8 – 11.0 (7.9)	3.3 – 7.6 (5.5)	1.3 – 9.9 (6.3)	0.67
Berlin (TUCC)	36	149	0.3 – 0.8 (0.5)	3.0 – 5.2 (3.7)	2.1 – 3.6 (2.6)	1.4 – 3.6 (3.0)	0.75
Helsinki	45	202	1.2 – 1.8 (1.6)	3.7 – 6.1 (4.4)	2.5 – 4.2 (3.1)	6.0 – 11.0 (8.5)	0.78
Heraklion (HECKOR)	5	24	0.4 – 2.0 (0.5)	1.8 – 13.3 (6.5)	1.3 – 9.2 (4.5)	1.5 – 13.2 (2.8)	0.51
Łódź	57	261	0.9 – 1.6 (1.3)	4.0 – 5.4 (4.4)	2.8 – 3.7 (3.1)	3.8 – 6.9 (5.8)	0.66
Melbourne (Preston)	2	9	1.6 – 2.1 (1.9)	2.6 – 13.2 (7.9)	1.8 – 9.2 (5.5)	5.5 – 21.3 (13.4)	0.69
Mexico City	8	49	0.7 – 1.5 (1.3)	5.5 – 16.5 (10.4)	3.8 – 11.5 (7.2)	5.8 – 21.9 (9.5)	0.65
Seoul	10	59	0.6 – 2.0 (1.3)	2.3 – 9.9 (6.5)	1.6 – 6.9 (4.5)	3.3 – 10.7 (6.1)	0.56
Singapore	7	40	1.3 – 1.6 (1.4)	4.6 – 20.1 (8.2)	3.2 – 14.0 (5.7)	7.7 – 28.4 (11.3)	0.81
Vancouver	67	308	1.2 – 1.4 (1.3)	6.5 – 8.9 (7.3)	4.5 – 6.2 (5.1)	7.1 – 9.5 (8.3)	0.54

In Table 2.2, an overview of the parameters is given for the 583 drydowns that complied with all criteria. Of the total number of 1606 drydowns, 102 are excluded because of potential snow conditions. All drydowns had a positive ET_0, and only three exceeded 10 mm d^{-1}. 671 additional drydowns did not meet the minimum R^2 of 0.3. Finally, a negative λ led to excluding 237 drydowns and λ above 35 days to 10 more. The remaining drydowns have an R^2 of 0.69 and yielded initial evapotranspiration between 0.3–2.1 mm d^{-1} and e-folding timescales between 1.8–20.1 days with the majority below 10.4 days, corresponding to half-lives of 1.3–14.0 and 7.2 days. The related storage capacities appear to be between 1.3–28.4 mm with the majority below 13.4 mm. As mentioned before, the length of the measurement period determines the magnitude of the uncertainty, which for S_0 varies from 1.2 mm in Basel (AESC) to 20.7 mm in Singapore.

For all sites, we find a considerable spread in the ET observations (Figure 2.2), which recurs in the estimated S_0 values. In Figure 2.3, S_0 is plotted against the month of the drydown, showing a very distinct seasonal dependency explaining why the spread in observations exceeds the uncertainty. Both ET_0 and λ, on which S_0 is based, show similar behaviour (not shown). Melbourne is shifted to fit the seasonality, as it is situated on the southern hemisphere. We expect that the enhanced effective storage capacity in summer is caused by increased vegetation activity. Since Singapore is close to the equator and its vegetation is evergreen, it is not expected to show seasonal effect, which is confirmed in Figure 2.3. Any connection between S_0 and the site characteristics in Table 2.1 and climatic

variables among which precipitation regime is overshadowed by the seasonal dependency covering the full range of S_0 (Table 2.2), as we illustrate in Figure A.2.3 and A.2.4. It is unfortunately not possible to eliminate the influence of this dependency by focusing on one season due to the steep slope, and not by focusing on one month due to the low data density. Only after omitting half of the cities based on the number of drydowns, a relation between S_0 and site characteristics is visible (Figure A.2.5).

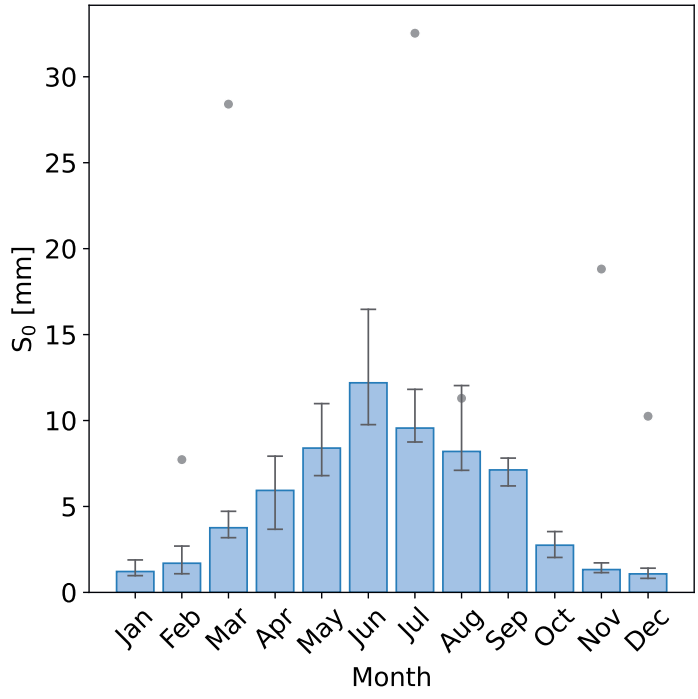


Figure 2.3: The seasonal dependency of the median S_0 for the sites on the northern hemisphere (Melbourne is included shifted by half a year) in blue and for Singapore as grey dots. The uncertainty is determined similarly as in Figure 2.2.

2.4 Discussion

In contrast to the results presented here for urban areas, Teuling et al. (2006) found timescales ranging from 15–35 days and storage varying between 30 and 150 mm for forests and grassland following a similar methodology. When compared to the urban parameter values (1.8–20.1 days and 1.3–28.4 mm), it is clear that both the timescales and storage capacities are much higher in rural areas. McColl et al. (2017) have analyzed soil moisture drydowns in a global study using satellite data with a resolution too coarse to explicitly resolve individual cities, thus resembling rural values. Although their timescales with

values from 2–20 days are closer to ours, it must be noted the temporal resolution is one in every three days and their observations only regard the first few centimeters instead of the root zone. Also, the satellite product in their research is known to underestimate the timescales compared to in-situ observations (Rondinelli et al., 2015; Shellito et al., 2016). When compared to storage values found for impervious surfaces by Wouters et al. (2015) (1.1–1.5 mm), the values in this study are higher as a result of the footprint scale analysis that includes natural in addition to impervious surfaces. Hence, the results show that both λ and S_0 are at least five times smaller in all cities than in natural ecosystems indicating shorter timescales and lower storage capacities in urban areas regardless of their climate and vegetation fraction.

Since our method is based on direct inference from observations, the footprint of observation determines the area for which the storage is estimated and the reliability of the measurements is essential to the quality of our estimates. Since the fluxes are observed at neighbourhood level, it is impossible to separate the (storage) source of ET. Further research could distinguish the different storage reservoirs by applying additional techniques like isotope analysis (Kuhlemann et al., 2021). The measurement reliability is ensured by carefully selecting locations and applying quality control (Velasco and Roth, 2010; Feigenwinter et al., 2012; Järvi et al., 2018). All sites have an observation height well above the mean building height (see Table 2.1), and measure in the inertial sublayer. This reduces the variability in flux measurements in response to the heterogeneity of the monitored footprint, which is induced by the many, unevenly distributed surfaces with different characteristics and water storage capacities in the urban landscape. The only site in this research that includes a non-homogeneous footprint is Seoul. The observations are filtered by wind direction to exclude a nearby forest. A relatively small variability between our estimates for each site suggests the observations are accurate enough for our application.

The methodology assumes that at the start of a drydown the storage capacity is completely full. A partly empty storage capacity would lead to an underestimation of the capacity, as less water is available for ET. We have compared the magnitude of the rain event before a drydown with the resulting parameters and found no correlation. Since the storage can be refilled by a series of events separated by dry days, we regressed the storage parameters against the Antecedent Precipitation Index (API) (Fedora and Beschta, 1989). The API takes into account rainfall occurring during preceding days (here limited to 20), but its observed values show no correlations with the λ and S_0. Therefore, the assumption of a completely filled storage is tangible and no selection has been performed based on rainfall event size. The evaporation directly after rainfall consists largely of interception ET from various surfaces (e.g. Grimmond and Oke, 1991; Gerrits, 2010; Oke et al., 2017). By calibrating an impervious-storage parameterization, (Wouters et al., 2015) estimated this storage to be between 1 and 1.5 mm for a site in Toulouse with little vegetation cover

(8%), suggesting interception ET is an important component of urban ET also in more diverse and greener urban landscapes included in this study.

2.5 Conclusions

The timescales of ET recession observed through eddy covariance in urban environments appear to be considerably shorter than in rural environments. This is related to the storage capacity, which is also found to be lower. Based on 583 drydowns, we find recession timescales of cities within 1.8–20.1 days with the majority below 10.4 days and storage capacities between 1.3–28.4 mm with the majority below 13.4 mm. The timescales and storage capacities are inferred for the entire footprint (including all ET sources) and do not translate to impervious surfaces. All values found in urban areas are at least five times smaller than found in rural areas. We were unable to analyze differences between cities to vegetation fraction, local climate zone or climate for two reasons. Firstly, the seasonal dependency in the storage capacities is as large as the total observed variation. Secondly, the number of sites is limited, and half of them contain data records shorter than one year. When provided with more data, the presented water storage capacity method has the potential to establish robust empirical relations explaining the differences between cities, in particular when complemented with soil moisture observations and/or Earth observation.


nab
7 Melbourne

Chapter 3

Water balance in urban land surface models

This chapter has originally been published as:

H. J. Jongen, M. Lipson, A. J. Teuling, S. Grimmond, J.-J. Baik, M. Best, M. Demuzere, K. Fortuniak, Y. Huang, M. G. De Kauwe, R. Li, J. McNorton, N. Meili, K. Oleson, S.-B. Park, T. Sun, A. Tsiringakis, M. Varentsov, C. Wang, Z.-H. Wang, and G. J. Steeneveld (2024a). "The Water Balance Representation in Urban-PLUMBER Land Surface Models". *Journal of Advances in Modeling Earth Systems* 16 (10), e2024MS004231

Abstract

Urban Land Surface Models (ULSMs) simulate energy and water exchanges between the urban surface and atmosphere. However, earlier systematic ULSM comparison projects assessed the energy balance but ignored the water balance which is coupled to the energy balance. Here, we analyze the water balance representation in 19 ULSMs participating in the Urban-PLUMBER project using results for 20 sites spread across a range of climates and urban form characteristics. As observations for most water fluxes are unavailable, we examine the water balance closure, flux timing, and magnitude with a score derived from seven indicators expecting better scoring models to capture the latent heat flux more accurately. We find that the water budget is only closed in 57% of the model-site combinations assuming closure when annual total incoming fluxes (precipitation and irrigation) fluxes are within 3% of the outgoing (all other) fluxes. Results show the timing is better captured than magnitude. No ULSM has passed all water balance indicators for any site. Models passing more indicators do not capture the latent heat flux more accurately refuting our hypothesis. While output reporting inconsistencies may have negatively affected model performance, our results indicate models could be improved by explicitly verifying water balance closure and revising runoff parameterizations. By expanding ULSM evaluation to the water balance and related to latent heat flux performance, we demonstrate the benefits of evaluating processes with direct feedback mechanisms to the processes of interest.

3.1 Introduction

The impact of urbanization on the local climate and hydrology has sparked scientists' interest and inspired research for centuries (e.g. Howard, 1833; Oke, 1982; Fletcher et al., 2013; Hamdi et al., 2020). With the increasing population in cities (United Nations, 2018) more people are impacted by increased heat stress and flooding (Heaviside et al., 2016; Gasparrini et al., 2017; Zhou et al., 2019; Botzen et al., 2020). Spatial morphological heterogeneity and human interactions make understanding the urban climate challenging (Kotthaus and Grimmond, 2014b; Sun et al., 2018; Koopmans et al., 2020; Demuzere et al., 2022), but weather and climate models need to include the effects of urban areas, as they locally exacerbate extreme events (Oleson et al., 2008; Ronda et al., 2017; Hertwig et al., 2020). Examples are increased flooding due to high impervious fractions (Zhou et al., 2019) and increased heat stress during heat waves resulting from reduced evaporation (Lemonsu et al., 2015; Li et al., 2019a). Therefore, models need to capture the impact of urban areas on their climate.

Researchers have developed, evaluated, and improved Urban Land Surface Models (ULSMs) simulating the interaction of the urban surface with the atmosphere. Coupled with a numerical weather prediction or climate model, ULSMs serve as a lower boundary condition and improve the model performance for urban environments (Tewari et al., 2007). ULSMs make different simplifying assumptions regarding urban geometry: a single homogeneous, impervious slab; multiple, individually homogeneous slabs; two-dimensional canyons; or 3D streets with individual buildings (Grimmond et al., 2009). These models also differ in whether and how they include physical processes like anthropogenic heat, irrigation, and snow processes (Lipson et al., 2024). To evaluate their performance, individual models are compared with observations (e.g. Ross and Oke, 1988; Grimmond and Oke, 2002; Hamdi and Schayes, 2007; Krayenhoff and Voogt, 2007; Porson et al., 2010). Although these individual evaluations were sometimes based on the same observations (Grimmond et al., 2009), the lack of a systematic approach prevented consistent comparison of the schemes. To compare the wide variety of models, two successive comparison projects applied a systematic approach. The first systematic comparison of ULSMs generally followed the PILPS protocol (project for intercomparison of land surface parameterization schemes, Henderson-Sellers et al. (1996)), hence PILPS-Urban (Grimmond et al., 2010; Grimmond et al., 2011). Individual modelers received meteorological input and surface characteristics to enable them to run their models. In total, 32 models completed simulations for a site in Vancouver and one in Melbourne. Grimmond et al. (2011) concluded that increased model complexity did not necessarily benefit model performance.

The second intercomparison, Urban-PLUMBER (Lipson et al., 2024), assesses 30 models initially at the PILPS-Urban Melbourne site and adopts benchmarks following the PLUMBER project (Best et al., 2015). Benchmarks serve as a relative reference, to which models are compared to assess whether a cohort performs better (or not) than the

benchmark and if input information is utilized effectively. Urban-PLUMBER is extended to the 20 sites presented by Lipson et al. (2022b) in the second phase (Lipson et al., 2023). The Urban-PLUMBER models outperform the PILPS-Urban ones for the sensible and latent heat flux. Some models representing two-dimensional canyons now perform nearly as well as one and two-tile models after efforts to improve hydrology and vegetation representation. However, models with complex urban geometry often still have relatively simple hydrology and vegetation and perform less well overall suggesting the representation of hydrology and vegetation requires more attention (Lipson et al., 2024).

Although PILPS-Urban and Urban-PLUMBER conclude vegetation and hydrology are important for model performance, neither project evaluates the water balance explicitly. The water balance satisfies the conservation of mass (Lavoisier, 1789) in the same way the energy balance satisfies the conservation of energy (Châtelet, 1740). The conservation of energy is forced in many ULSMs to prevent the energetic state of the model from drifting and the consequential, long-term bias in the modeled surface fluxes (Grimmond et al., 2010). Closure is achieved by either updating the surface temperatures based on the residual energy or restricting the turbulent heat flows to the available energy (Grimmond et al., 2010). Both PILPS-Urban and Urban-PLUMBER test whether models close the energy balance, but have not verified the numerical closure of the water balance. Similar to the energy balance, an unclosed water balance can result in model biases and consequential drifting. These biases may in turn affect the energy balance, as the energy and water balance are linked through evapotranspiration (ET), the mass counterpart of the latent heat flux (Q_E). This direct link implies errors and/or biases in one balance will affect the model's skill for the other balance. Recently, Yu et al. (2022) showed the hydrology in a coupled ULSM has the potential to improve the Q_E, humidity, and air temperature with impacts up into the boundary layer ($\sim$1 km). ET/Q_E has been amongst the most challenging fluxes for ULSMs from the first assessment (Ross and Oke, 1988) until now (Lipson et al., 2024). Given the link to the energy balance, we hypothesize closing the water balance will improve model performance for the energy balance fluxes.

However, the water balance cannot be directly assessed because of a lack of observations at the appropriate spatiotemporal scales at this time. While precipitation is measured routinely in many urban locations with rain gauges and rain radars, runoff, irrigation, and changes in water storage are not. Q_E (ET) observations from eddy-covariance systems have substantial gaps introduced in the quality control process (Feigenwinter et al., 2012) that rejects more data close to rain events (Grimmond, 2006b). Runoff is occasionally measured in urban catchments (Berthier et al., 1999; Walsh et al., 2005a), but a challenge is posed by the difference in the source area of observations for runoff and eddy-covariance techniques (Grimmond and Oke, 1986; Grimmond and Oke, 1991; Hellsten et al., 2015). External water use, often irrigation, further complicates the water balance in cities, as it mainly occurs at the micro-scale (e.g. garden irrigation). This scale can only be inferred from neighborhood piped water supply observations and water use surveys or estimated from

weather, vegetation, and soil type (Grimmond and Oke, 1986; Mitchell et al., 2001; Zeisl et al., 2018; Kokkonen et al., 2018). Tree roots penetrate (sewer) pipes causing damage (Randrup et al., 2001) and simultaneously taking out water, which is an unobserved term. Lastly, measuring the water storage change is logistically difficult, as this requires the state of each individual element contributing to water storage in the city, such as soil moisture, interception, groundwater, and surface water. Thus, a direct comparison of a full set of water balance observations is extremely challenging and an alternative approach is needed.

Here, we develop an alternative approach to evaluate the representation and dynamics of the water balance in ULSMs. To examine the water balance closure, we propose an UWBR (urban water balance representation) score. The score combines seven indicators assessing: water balance closure (1 indicator), ET (2), water storage dynamics (2), and surface runoff (2). The UWBR score is applied, given a lack of observations, to rank models' capability to accurately capture different aspects of the water balance. Assessing the score of 19 Urban-PLUMBER ULSMs with a complete water balance representation helps to identify model improvement possibilities. The water balance representation is compared with the turbulent heat fluxes model skill since we expect a better water balance representation should improve simulated latent heat fluxes.

3.2 Methods

3.2.1 Urban water balance representation (UWBR) score

The UWBR score is a linear sum of seven indicators of a good water balance, which are assigned a value of 1 if a specified threshold is passed (Table 3.1), except the $I_{S,m}$ indicator, for which both sub-metrics are assigned 0.5 if passed. No weights are assigned, as these cannot be determined objectively. The UWBR score is compared with the model performance for the latent heat flux assessed with metrics capturing different characteristics (Willmott, 1982) that are not entirely independent:

- Absolute mean bias error (|MBE|) assesses the bias providing insight into how well the quantities of the latent heat flux are modeled.
- Coefficient of determination (R^2) captures the consistency of the timing as R^2 decreases with a shift in a quasiperiodic signal like the latent heat flux.
- Normalized standard deviation (σ_{norm}, σ_{model} divided by $\sigma_{observations}$) compares the variability, which is dominated by the daily cycle in the case of the latent heat flux.
- Systematic Mean Absolute Error (MAE_s) indicates the average error. The systematic error is separated from the unsystematic error similarly to the approach presented

by Willmott (1982) for the root mean square error. This separation allows us to distinguish between systematic and random errors.

- Unsystematic Mean Absolute Error (MAE_u) assesses how well the erratic behaviour is captured.

Before the individual indicators are introduced, we define two ways to calculate water storage from the model output based on either the water storage term (explicit) or the other terms of the water balance combined (implicit). Assuming that the net change in water stored in a "catchment" or a model grid (ΔS) can be derived from the difference between the incoming and outgoing water fluxes, then the implicit water storage is:

$$\Delta S = P + I - (R + ET) \tag{3.1}$$

where P is precipitation, I irrigation, and R runoff. R represents both the surface (R_s) and the subsurface (R_{ss}) runoff. When ΔS is calculated from the fluxes on the right-hand side of Eq. 3.1, we refer to this as the implicit water storage. The second approach determines the net storage change (ΔS) based on the modeled storage components following the urban water balance (Grimmond and Oke, 1986). The storage components should account for the water storage above and below ground, such as the interception, water bodies, and groundwater. The components included depend on the model conceptualization. Here, we refer to the storage represented in the model as the explicit water storage (ΔS_{model}):

$$\Delta S_{model} = \Delta S_{soil} + \Delta S_{intercept} + \Delta S_{snow} \tag{3.2}$$

where ΔS_{soil} is storage change in the soil moisture, $\Delta S_{intercept}$ storage change in the interception storage, and ΔS_{snow} storage change in the snow cover. Depending on the model, ΔS_{soil} considers soil moisture below the impervious and pervious fraction. In the case a model does not consider soil moisture below the impervious fraction, ΔS_{soil} is adjusted accordingly. When we refer to annual timescales, the analysis is performed on all time intervals of a year in the time series, i.e. a new annual period starts at every timestep, after which a full year is modeled (e.g. NL-Amsterdam: 2018-05-01 19:00 - 2019-05-01 19:00, 2018-05-01 20:00 - 2019-05-01 20:00, etc.). Within this year, no gaps in the model data allow all timesteps to be used. This method maximizes the use of available data and eliminates the influence of choosing a specific annual period like the calendar or hydrological year.

Water balance closure

Water balance closure assumes that all fluxes add up to zero for the time and space under consideration (here $\sim$1 km^2 and one year):

$$P + I - (R + \mathrm{ET} + \Delta S_{model}) = 0 \tag{3.3}$$

where ΔS corresponds to the explicit water storage in the model (Eq. 3.2) to prevent closure resulting from calculating the storage change based on the fluxes. Three models

Table 3.1: Overview of the seven indicators that are linearly combined in the UWBR score, which is used to evaluate the urban water balance representation in ULSMs. The criterion indicates what needs to be achieved to assign a value of 1 to the indicator or 0.5 per test in the case of $I_{S,m}$. The uncertainty criteria (*) are discussed in section 3.2.1. The notation in the equations is defined in the corresponding subsections of section 3.2.1. The details on all indicators can be found in section 3.2.1.

Water balance flux	**Indicator**	**Description**	**Timescale**	**Criterion**	**Equation**
All	I_A	Closure of the annual water balance assessed relative to the precipitation plus irrigation	Annual	< 0.03	$\left\lvert \frac{P+I-(R+\mathrm{ET}+\Delta S)}{P+I} \right\rvert$
ET	$I_{\mathrm{ET},m}$	Modeled cumulative ET normalized by the benchmark ET (ET_{bench}) over the whole model period	Modeled period	Within benchmark uncertainty*	$\frac{\mathrm{ET}_{model}}{\mathrm{ET}_{bench}}$
	$I_{\mathrm{ET},t}$	Similarity of ET recession timescale distribution between model and observations from the whole model run	Modeled period	$p < 0.05$	Kolmogorov-Smirnov test (Chakravarti et al., 1967)
ΔS	$I_{S,m}$	Range over the whole model run in stored water for both the modeled explicit and implicit water storage compared to water storage capacity	Modeled period	$<$ (50% of soil volume + 3 mm interception)	$\Delta S_{model,max} - \Delta S_{model,min}$ (Eq. 3.2) and $\Delta S_{max} - \Delta S_{min}$ (Eq. 3.1)
	$I_{S,t}$	Coefficient of determination (R^2) between changes in explicit and implicit modeled water storage over the whole model period	Modeled period	>0.9	$1\text{-}\frac{\sum_{i=1}^{n}(\Delta S_i - \Delta \hat{S}_{model,i})}{\sum_{i=1}^{n}\Delta S_i - \Delta S_{model,i}}$
R_s	$I_{R,m}$	Curve number (CN) from modeled runoff events and from site characteristics	Event	Within CN uncertainty*	$\mathrm{CN} = \frac{1000}{S-10}$ (section 3.2.1)
	$I_{R,t}$	Mean lag (hours) between center of mass from precipitation and surface runoff of all events	Event	$<$ 1 hour	$R_{s,centroid} - P_{centroid}$

(8, 16, and 17) model groundwater interaction, which is not included in the model output. We examine the annual water balance closure with the annual total fluxes. Closure should also occur at every timestep, however, we were unable to undertake this more stringent check because interception storage was modeled but mostly unreported by modelers (all 19 models modeled, only 3 reported). Assuming an interception storage capacity of over 0.5 and up to 3 mm (Klaassen et al., 1998; Wouters et al., 2015; Carlyle-Moses et al., 2020), this storage can be filled in a single (half-)hourly timestep. At a single (half-)hourly timestep, 0.5 mm is a non-negligible lack of closure but it is less critical at the annual scale. Eq. 3.3 is normalized by annual precipitation plus irrigation to enable comparison between sites with a range of precipitation regimes.

The water balance closure indicator (I_A, Table 3.1) assesses if the total sum of all fluxes (including storage) is less than 3% from $P + I$. The 3% threshold allows for non-closure due to interception storage data not being provided in the model output, errors arising in latent heat flux unit conversion, or numerical model errors. According to the literature, interception storage amounts to 0.5-3 mm explaining a non-closure of up to 0.5% when it is not provided (Klaassen et al., 1998; Wouters et al., 2015; Carlyle-Moses et al., 2020). Converting the latent heat flux to ET can result in variations up to 2% depending on temperature and snow effects (Bringfelt, 1986; Petrucci et al., 2010). Not all models correct for these effects. To account for numerical model errors arising from discretization and time stepping (MacKay et al., 2022), we allow deviations of up to 0.5%.

Evapotranspiration (ET)

The two ET indicators address the magnitude and timing. The non-randomly distributed gaps in ET observations prevent direct comparison of total modeled ET (ET_{model}) over a model period. Thus, we use one of the Lipson et al. (2024) benchmark models. This allows a total ET to be obtained without gaps. The Lipson et al. (2024) benchmark model (ET_{bench}) is derived using multivariate ordinary least squares regressions with a K-means clustering approach. The K-means clustering approach is trained in-sample using 81 clusters on four variables: incoming shortwave radiation, air temperature, relative humidity, and wind speed (KM4-IS-SWdown-Tair-RH-Wind in Lipson et al., 2024). To reduce the hourly MBE, wind speed is omitted at both Helsinki sites. At all sites, the MBE is below 1 $W\ m^{-2}$ and at most sites below 0.1 $W\ m^{-2}$ evaluated against available data.

Therefore, ET_{bench} is assumed to provide a reasonable estimate of the total ET flux over the model run for the $I_{\mathrm{ET},m}$ indicator (Table 3.1). We compare in Q_E units rather than ET eliminating unit conversions and calculate the cumulative ET flux uncertainty from the benchmark based on (1) the benchmark MBE multiplied by the run duration, and (2) lack of energy balance closure associated with eddy-covariance observations (Franssen et al., 2010; Foken et al., 2012; Mauder et al., 2020). The lack of energy closure is calculated by

the net all-wave radiation minus the sum of the turbulent heat fluxes. The storage and anthropogenic heat fluxes are not observed, which prevents constraining the turbulent heat fluxes with energy balance closure. If a lack of closure occurs, the unexplained energy over the whole model run is split between Q_E and the sensible heat flux (Q_H) according to the Bowen ratio based on the benchmark fluxes (Twine et al., 2000; Hirschi et al., 2017; Mauder et al., 2020):

$$Q_{E,uncertainty} = \frac{1}{1+B}(Q^* - Q_E - Q_H) \text{ with } B = Q_H/Q_E \tag{3.4}$$

where B is the Bowen ratio and Q^* the net radiation. To this $Q_{E,uncertainty}$, the benchmark uncertainty is added. The benchmark uncertainty is the MBE of the benchmark multiplied by the run duration. A model run passes $I_{\mathrm{ET},m}$ when ET_{model} falls within the uncertainty of ET_{bench}.

The timing of modeled ET is assessed assuming exponential ET recession after rainfall based on the recession timescale estimated following the Jongen et al. (2022) methodology. This methodology considers only the first ten days to exclude the influence of longer dry periods and irrigation. A daily-timescale analysis circumvents observational gaps. Model and observations are assessed if they have the same distribution for the recession timescale with a Kolmogorov-Smirnov test (Chakravarti et al., 1967). The $I_{\mathrm{ET},t}$ indicator is assigned a value of 1 when the p-value is below 0.05.

Water storage

Indicator $I_{S,m}$ evaluates the water storage by comparing the modeled explicit and implicit water storage ranges (Section 3.2.1) over the analysis period with respect to the estimated water storage capacity. According to the literature, soil water storage capacity is maximally half the soil depth for all soil types (Saxton et al., 1986). The maximum is set as a storage capacity that models should not exceed rather than a realistic value. As urban soils are frequently disturbed making them spatially heterogeneous, reliable maps are rarely available (Van de Vijver et al., 2020). As the modeled soil depth depends on the model run, the soil water storage capacity is calculated for each separately. To account for interception storage, 3 mm is added to the estimated water storage capacity based on tree and impervious interception observations (Klaassen et al., 1998; Wouters et al., 2015; Carlyle-Moses et al., 2020). The two models not including soil moisture do not pass the first check of this indicator and are only evaluated based on the implicit water storage (Table 3.2). Other models receive 0.5 score when either the modeled explicit or implicit water storage range falls within the estimated water storage capacity (or 1 for both).

Indicator $I_{S,t}$ quantifies the internal temporal consistency between the change in explicit (Eq. 3.2) and the implicit (Eq. 3.1), which should be indicating the same flux. The coefficient of determination R^2 (Willmott, 1982) is calculated using storage changes using 30-min (or 60-min) model output depending on the site forcing data. This metric equals 1

if the timing between two fluxes is similar ($R^2 > 0.9$) independent of the flux bias, unlike other indicators (e.g. I_A). The two models without soil moisture output are assigned a value of 0 for $I_{S,t}$ as their performance could not be evaluated.

Surface runoff (R_s)

Indicator $I_{R,m}$ assesses the R_s magnitude relating total event precipitation to R_s (Figure 3.1a). Without runoff observations, curve numbers (CN) are derived to evaluate modeled total event R_s (Cronshey et al., 1985) based on the relation between the total event precipitation (P_e) and the total event R_s (R_e):

$$R_e = \frac{(P_e - 0.2S)^2}{P_e + 0.8S} \text{ with } S = \frac{1000}{\text{CN}} - 10 \tag{3.5}$$

where S is the potential maximum retention. To determine when precipitation events are independent, the auto-correlation of precipitation events is examined. A dry period of five hours (Figure A.3.1) is assumed across all sites, which is consistent with Wenzel Jr and Voorhees (1981). This dry period is several hours longer than the expected runoff response time preventing events from influencing each other (Morin et al., 2001; Berne et al., 2004; Yao et al., 2016). To exclude snow events, the analysis includes only events with a minimum air temperature above 0°C. For each model run, ordinary least squares is used with the Re and Pe data to estimate S (Figure 3.1b)from which the CN is derived (Eq. 3.5). During this process, the variance and standard deviation of S are calculated from the variance in the data points. The standard deviation of CN follows from this and is used as the uncertainty estimate from the models.

For each site, the CN is estimated using a linear interpolation of a look-up table considering the impervious fraction within the eddy-covariance footprint (Cronshey et al., 1985). Given soil texture influences CN, sand fraction (Brakensiek and Rawls, 1983; Nachtergaele, 2001) obtained from a global data set (OpenLandMap, (Hengl, 2018)) is used to constrain CN. Given the uncertainty of urban soil maps, using sand fraction is a repeatable way to assign the most uncertainty to the CN look-up tables, assuming a one-third change of CN from a one-level change in soil texture in either direction. If the site CN, including its uncertainty, overlaps with the model CN including its uncertainty, $I_{R,m}$ is assigned a value of 1.

Indicator $I_{R,t}$ addresses the rainfall-R_s response times (Leopold, 1968). The lag time is calculated as the difference between centroids of rainfall ($P_{centroid}$) and R_s ($R_{centroid}$) for the same events as the CN calculations (Figure 3.1a). Long-tail rainfall events are excluded when the $R_{centroid}$ comes before the $P_{centroid}$. As eddy-covariance systems have a footprint on the sub-square-kilometer scale (Feigenwinter et al., 2012), lag time is expected to be much faster than 30-60 minutes (Morin et al., 2001; Berne et al., 2004; Yao et al., 2016), which is the model output resolution (Lipson et al., 2024). Therefore, the mean lag time needs to be less than one hour. The mean is preferred over the median to also pinpoint models that occasionally have long lag times that would not affect the median.

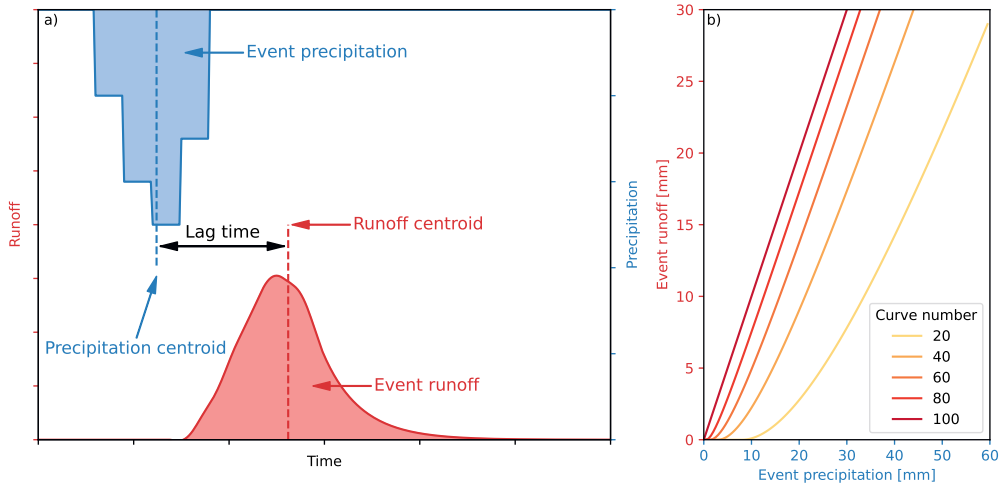


Figure 3.1: Illustration of surface runoff indicators ($I_{R,m}$ and $I_{R,t}$) showing (a) lag time between the precipitation centroid and surface runoff centroid for an illustrative event, and (b) CN values (Eq. 3.5) derived from total event precipitation and surface runoff.

Lag times of intermittent precipitation-runoff events will only decrease, as storages are already (partly) filled by earlier precipitation. Dry periods of less than five hours should also have lag times of less than one hour.

3.2.2 Models

The present study anonymously analyzes the water balance outputs from 19 Urban-PLUMBER ULSMs (Table 3.2). Other Urban-PLUMBER ULSMs did not submit the necessary outputs to allow for a water balance assessment. The outputs are for 20 sites covering a range of climates, impervious fractions, and observational periods (Table 3.3). As two models did not run all sites, 377 runs are analyzed.

For each site, modelers were provided with the site characteristics and meteorological forcing with 10-year spin-up data (Lipson et al., 2022b). The spin-up period required to reach equilibrium varies per model, with some requiring many years to come to hydrological equilibrium with the forcing meteorology (Yang et al., 1995; Best and Grimmond, 2016). The 10 years of spin-up before the evaluation observations allowed the soil moisture stores to equilibrate with local conditions prior to analysis. ERA5 reanalysis data (Hersbach et al., 2020) are used to derive hourly forcing with bias-correction including diurnal and seasonal effects for each site (Lipson et al., 2022b).

Depending on site data, evaluation is undertaken with 30- or 60-minute fluxes for periods varying between 148 and 1827 days (average 912 days, Table 3.3). Similar to the Urban-PLUMBER protocol, to minimize human errors, modelers received a preliminary analysis

Table 3.2: Overview of the 19 urban land surface models in the water balance analysis based on Lipson et al. (2024). Two models did not provide soil moisture output ([1]). Three models capable of simulating irrigation did not include it in their Urban-PLUMBER runs ([2]). Two models have an internal water balance closure check which does not hold in conditions with snowfall ([3]). Two models have an internal water balance closure check that was not used for the Urban-PLUMBER runs ([4]).

Model	Urban geometry	Vegetation	Soil hydrology	Snow accumulation	Irrigation	Water balance closure check	Reference
ASLUMv2.0	Canyon	Grass	Multi-layer	No	No[2]	No[3]	Wang et al. (2013) and Wang et al. (2021)
ASLUMv3.1	Canyon	Grass+trees	Multi-layer	No	No[2]	No[3]	Wang et al. (2013) and Wang et al. (2021)
CABLE	Non-urban	Separate tiles	Multi-layer	Veg.	No	Yes	Kowalczyk et al. (2006) and Wang et al. (2011)
ECLand	Non-urban	Separate tiles	Multi-layer	Veg.	No	No[4]	Boussetta et al. (2021)
ECLand-U	Two-tile	Separate tiles	Multi-layer	Veg.+urban	No	No[4]	McNorton et al. (2021) and Boussetta et al. (2021)
CLMU5	Canyon	Grass+shrubs	Multi-layer	Urban	No	Yes	Oleson and Feddema (2020)
JULES 1T	One-tile	Separate tiles	Multi-layer	Veg.+urban	No	Yes	Best et al. (2011)
JULES 2T	Two-tile	Separate tiles	Multi-layer	Veg.+urban	No	Yes	Best et al. (2011)
JULES MOR	Two-tile	Separate tiles	Multi-layer	Veg.+urban	No	Yes	Best et al. (2011)
Lodz-SUEB	One-tile	Lumped with urban	Multi-layer[1]	Veg.+urban	No	No	Fortuniak (2003)
Manabe 1T	One-tile	Manabe bucket	One-layer	Veg.+urban	No	No	Best et al. (2011) and Manabe (1969)
Manabe 2T	Two-tile	Manabe bucket	One-layer	Veg.+urban	No	No	Best et al. (2011) and Manabe (1969)
NOAH-SLAB	One-tile	Separate tiles	Multi-layer	Veg.+urban	No	No	Kusaka et al. (2001) and Ek et al. (2003)
NOAH-SLUCM	Canyon	Separate tiles	Multi-layer	Veg.+urban	No	No	Kusaka et al. (2001) and Ek et al. (2003)
SNUUCM	Canyon	Separate tiles	Multi-layer[1]	Veg.	No	No	Ryu et al. (2011) and Ek et al. (2003)
SUEWS	Two-tile	Separate tiles	One-layer	Veg.+urban	No[2]	Yes	Järvi et al. (2011) and Ward et al. (2016)
TERRA 4.11	One-tile	Separate tiles	Multi-layer	Veg.	No	No	Wouters et al. (2015) and Schulz and Vogel (2020)
UCLEM	Canyon	Grass+shrubs	One-layer	Veg.+urban	Yes	No	Thatcher and Hurley (2012) and Lipson et al. (2018)
UT&C	Canyon	Grass+shrubs+trees	Multi-layer	No	Yes	Yes	Meili et al. (2020)

Table 3.3: Model (Table 3.2) outputs are analyzed for 20 sites (Lipson et al., 2022b). Only wind directions coming from urban areas are included for the Minneapolis site. Characteristics include the local climate zone (LCZ, Stewart and Oke (2012), where 2 is compact mid-rise, 3 compact low-rise, 5 open mid-rise, and 6 open low-rise), impervious surface fraction (F_{imp}), displacement height (z_d), and eddy-covariance sensor height above ground level (z_s).

Country	City (site)	Name	Lat. (°)	Lon. (°)	Observed period (days)	Köppen-Geiger climate	LCZ	F_{imp}	z_d (m)	z_s (m)	Reference
Australia	Melbourne (Preston)	AU-Preston	-37.73	145.01	475	Cfb	6	0.62	8	40	Coutts et al. (2007a) and Coutts et al. (2007b)
Australia	Melbourne (Surrey Hills)	AU-SurreyHills	-37.83	145.10	148	Cfb	6	0.54	8	38	Coutts et al. (2007a) and Coutts et al. (2007b)
Canada	Vancouver (Sunset)	CA-Sunset	49.23	-123.08	1827	Csb	6	0.68	3	25	Christen et al. (2011) and Crawford and Christen (2015)
Finland	Helsinki (Kumpula)	FI-Kumpula	60.20	24.96	1096	Dfb	mix	0.46	6	31	Karsisto et al. (2016)
Finland	Helsinki (Torni)	FI-Torni	60.17	24.94	1096	Dfb	2	0.77	15	60	Nordbo et al. (2013) and Järvi et al. (2018)
France	Toulouse (Capitole)	FR-Capitole	43.60	1.45	375	Cfa	2	0.90	11	48	Masson et al. (2008) and Goret et al. (2019)
Greece	Heraklion	GR-HECKOR	35.34	25.13	367	Csa	3	0.92	17	27	Stagakis et al. (2019)
Japan	Tokyo (Yoyogi)	JP-Yoyogi	35.66	139.68	1461	Cfa	2	0.92	28	52	Hirano et al. (2015) and Ishidoya et al. (2020)
South Korea	Seoul (Jungnang)	KR-Jungnang	37.59	127.08	825	Dwa	3	0.97	15	42	Hong et al. (2020a) and Hong et al. (2023)
South Korea	Cheongju (Ochang)	KR-Ochang	36.72	127.43	780	Dwa	5	0.47	4	19	Hong et al. (2019) and Hong et al. (2020a)
Mexico	Mexico City (Escandon)	MX-Escandon	19.40	-99.18	470	Cwb	2	0.94	8	37	Velasco et al. (2011) and Velasco et al. (2014)
Netherlands	Amsterdam	NL-Amsterdam	52.37	4.89	652	Cfb	2	0.68	10	40	Steeneveld et al. (2020)
Poland	Łódź (Lipowa)	PL-Lipowa	51.76	19.45	1827	Dfb	2	0.76	7	37	Pawlak et al. (2011) and Fortuniak et al. (2013)
Poland	Łódź (Narutowicza)	PL-Narutowicza	51.77	19.48	1827	Dfb	2	0.65	11	42	Fortuniak et al. (2006) and Fortuniak et al. (2013)
Singapore	Singapore (Telok Kurau)	SG-TelokKurau	1.31	103.91	366	Af	3	0.85	7	24	Roth et al. (2017)
UK	London (King's college)	UK-KingsCollege	51.51	-0.12	638	Cfb	2	0.79	15	50	Kotthaus and Grimmond (2014b); Kotthaus and Grimmond (2014a) and Bjorkegren et al. (2015)
UK	Swindon	UK-Swindon	51.58	-1.80	715	Cfb	6	0.49	4	13	Ward et al. (2013)
USA	Baltimore (Cub hill)	US-Baltimore	39.41	-76.52	1826	Cfa	6	0.31	4	37	Crawford et al. (2011)
USA	Minneapolis	US-Minneapolis1	45.00	-93.19	1093	Dfa	6	0.21	3	40	Peters et al. (2011) and Menzer and McFadden (2017)
USA	Phoenix (West)	US-WestPhoenix	33.48	-112.14	382	Bwh	6	0.48	3	22	Chow et al. (2014b) and Chow (2017)

of the water balance to help identify major issues and were encouraged to update their results. This eliminated unit errors, added missing variables, and removed inactive soil moisture layers.

For this study, we harmonize the hydrological model output. If a model only provided Q_E (unit: $[W\ m^{-2}]$), it is converted to ET (unit: $[mm\ d^{-1}]$) using latent heat of vaporization accounting for air temperature (Bringfelt, 1986). When snow is present the latent heat of fusion is added to the latent heat of vaporization to acquire the latent heat of sublimation (Petrucci et al., 2010). In the forcing, precipitation is split into snowfall and rainfall. At only 30% of the sites, snowfall amounts to more than 10% of the precipitation. It is added as rainfall for one model without snow hydrology, while the two others do not account for this input. Irrigation is simulated in two models. For all other models, irrigation is assumed to be zero.

3.3 Results

The 19 ULSMs show a wide spread in the average yearly water fluxes at all 20 sites based on all 377 model runs (Figure 3.2). Overall, the model spread (whiskers, Figure 3.2) is often wider than the modeled ensemble mean flux (bars, Figure 3.2). Models show more variation in ET than in runoff. Sites with higher annual water input have more variability in model output fluxes, for example, the relatively high fluxes in KR-Jungnang and SG-TelokKurau compared to the lower yearly fluxes in PL-Lipowa and US-WestPhoenix.

3.3.1 Water balance closure

Although the annual mean model ensemble almost closes the water balance at most sites (Figure 3.2), most individual models do not close the water balance (Figure 3.3). Here, closure is assumed when the sum of all fluxes (Eq. 3.3) is less than 3% of P+I. This occurs in 57% of the model runs (I_A, Figure 3.4). In 25% of the model runs, non-closure exceeds 10% of P+I. Closure is model-related as the bias is similar across sites for each model (Figure 3.3). Five models close the water balance in all runs, whereas four models account for 48% of unclosed model runs. Three models pass their internal water balance closure check but do not always pass this closure check possibly due to unreported, modeled water fluxes or inconsistencies in the way fluxes were reported. To assess the impact of model run length, the analysis is repeated with sites with more than two years of observations yielding similar results.

3.3.2 Evapotranspiration (ET)

Comparison of the modeled mean diurnal cycle of the ET (Figure 3.5) shows the highest inter-model spread at the peak of the diurnal cycle, with a range of 10-600% of the model ensemble-mean flux. Along three sites with contrasting precipitation regimes (US-

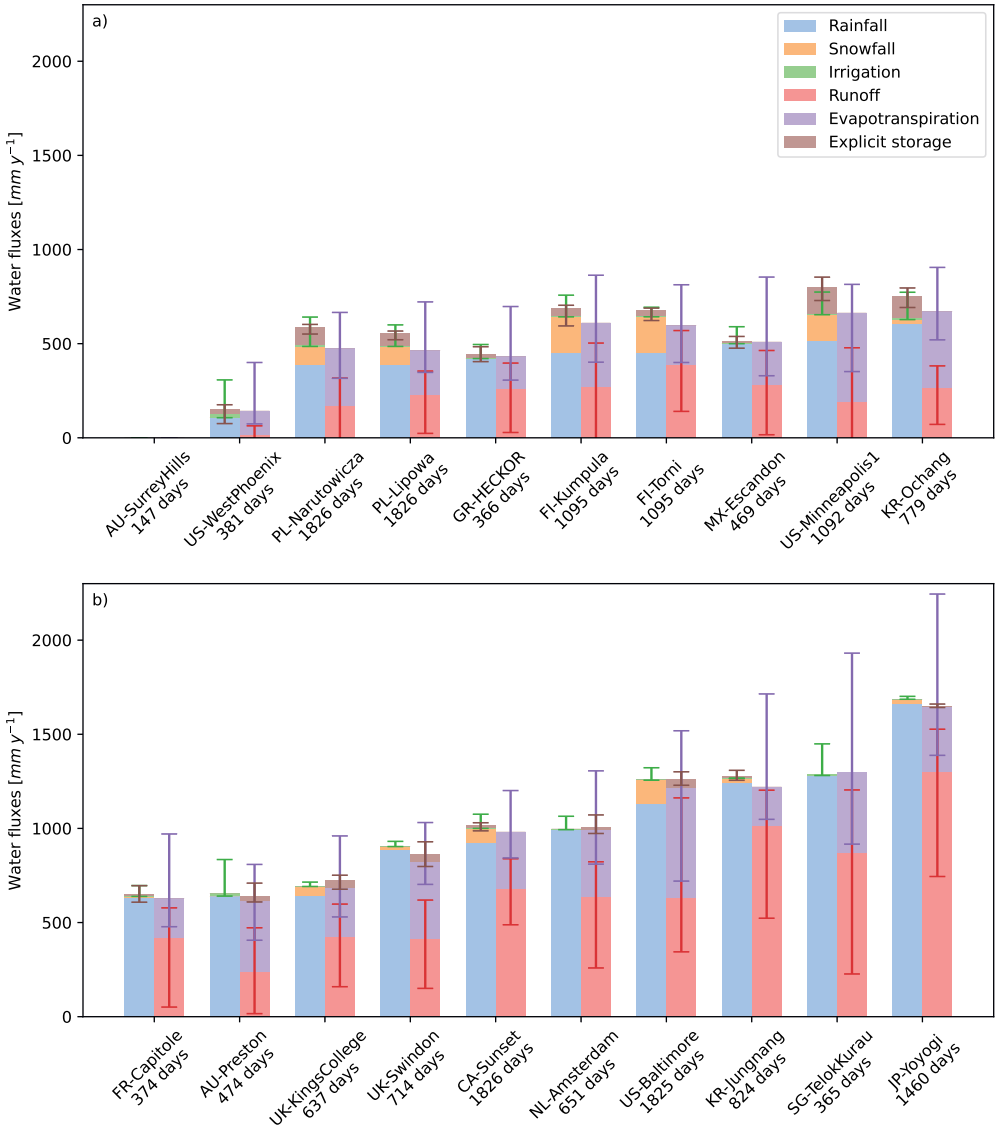


Figure 3.2: Ensemble mean (bars) and full range (minimum to maximum, whiskers) of the modeled annual water fluxes for all 20 sites ordered by increasing average annual precipitation. Explicitly modeled storage flux (Eq. 3.2, brown) appears on the left if a net input and right if a net loss. Values are means of all complete years in a data set (e.g. NL-Amsterdam: 2018-05-01 19:00 - 2019-05-01 19:00, 2018-05-01 20:00 - 2019-05-01 20:00, etc.). AU-Surreyhills has less than a year of observations.

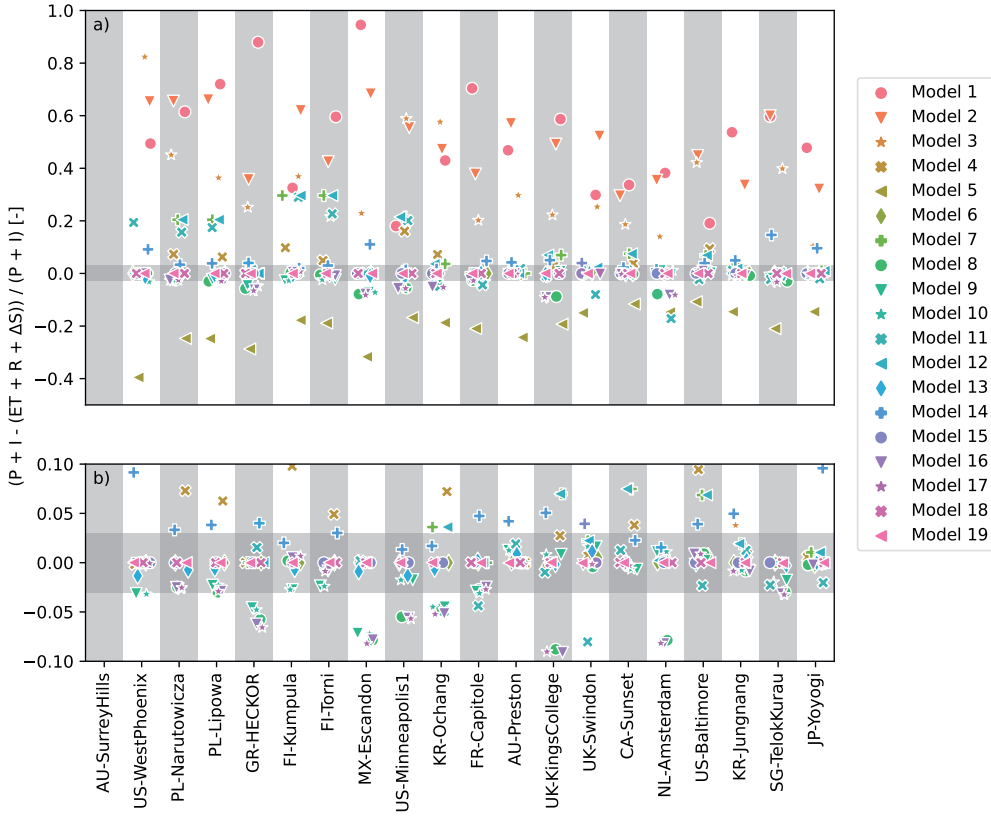


Figure 3.3: Annual water balance closure (Eq. 3.3) per model (marker) at 20 sites (by increasing average annual precipitation). Models with indicator $I_A = 1$ (Table 3.1, horizontal shading) are shown in more detail in the lower panel (b).

WestPhoenix, AU-Preston, and SG-TelokKurau), ET increases as expected at wetter sites. At US-WestPhoenix, all models but one underestimate peak ET. This underestimation likely results from the absence of irrigation in nearly all models, while irrigation is common at US-WestPhoenix (Templeton et al., 2018). The one overestimating model does not include irrigation. At the other two sites, around half the models underestimate ET (Figure 3.5). Although for these sites the model medians are better, the difficulty of capturing the correct flux magnitude is evident, as $I_{\mathrm{ET},m}$ is passed by only 26% of the model runs (Figure 3.4). No model passes this indicator at more than half of the sites.

After different rainfall events, daily ET decreases with varying timescales in both the observations and the models (Figure 3.6). The variation is higher amongst the modeled than the observed drydowns. In contrast with the ET magnitude, the recession timescale

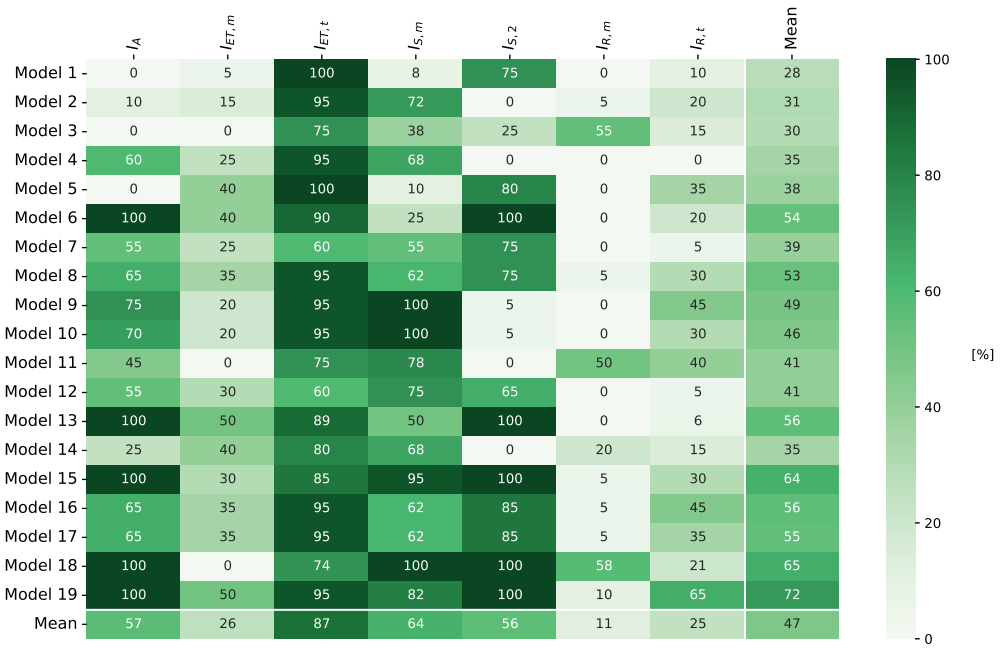


Figure 3.4: Overview of the indicators of the urban water balance representation (UWBR) score and constituent indicators (Table 3.1) over all sites. Means are corrected for missing model runs.

shows no link with annual precipitation. $I_{\mathrm{ET},t}$ shows the ET recession timescale is captured correctly in 87% of the cases (Figure 3.4).

3.3.3 Water storage

Not all models have explicit water storage values (Eq. 3.2) that are equal to the implicit values (Eq. 3.1, Figure 3.7), which is seen across all sites (not shown). However, the explicit water storage should reflect the implicit storage, as the explicit storage change is equal to the net of all water fluxes. For five models, the explicit storage change is equal to the implicit storage change at all sites. Minor differences occur in six models and large differences in six others. Two models have no differences at sites without snowfall (e.g. AU-Preston) but large differences at sites with snowfall (e.g. CA-Sunset). As these models do not account for the snowfall in the input we see an increasing difference between the explicit and implicit water storage. The models with larger differences follow a seasonal cycle likely caused by non-restricted implicit water storage combined with restricted explicit water storage by soil storage capacity.

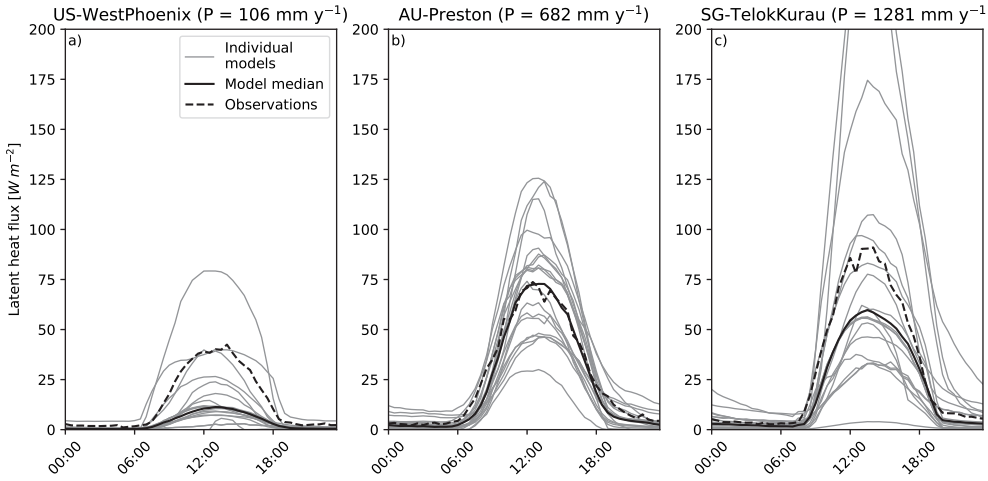


Figure 3.5: Illustration of modeled and observed (dashed) mean diurnal cycle of ET at three sites with contrasting annual precipitation increasing from left to right: (a) US-WestPhoenix, (b) AU-Preston, and (c) SG-TelokKurau. Note that the observations are direct latent heat flux observations from eddy-covariance systems and do not refer to ET_{bench}. Local time is used for all sites.

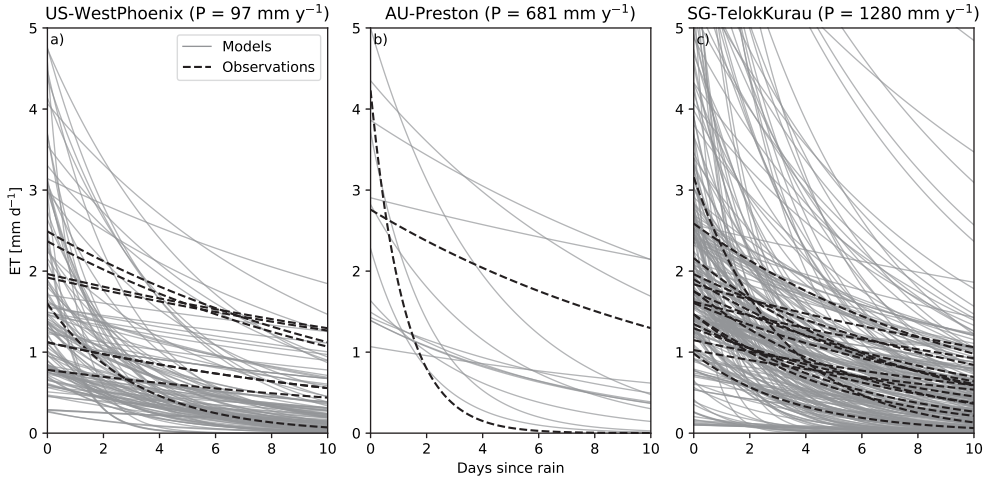


Figure 3.6: As Figure 3.5, but modeled (grey) and observed (black) daily ET following separate, individual rainfall events. Drydown events are selected based on their duration and data availability (see Jongen et al., 2022). Note that the observations are direct latent heat flux observations from eddy-covariance systems and do not refer to ET_{bench}.

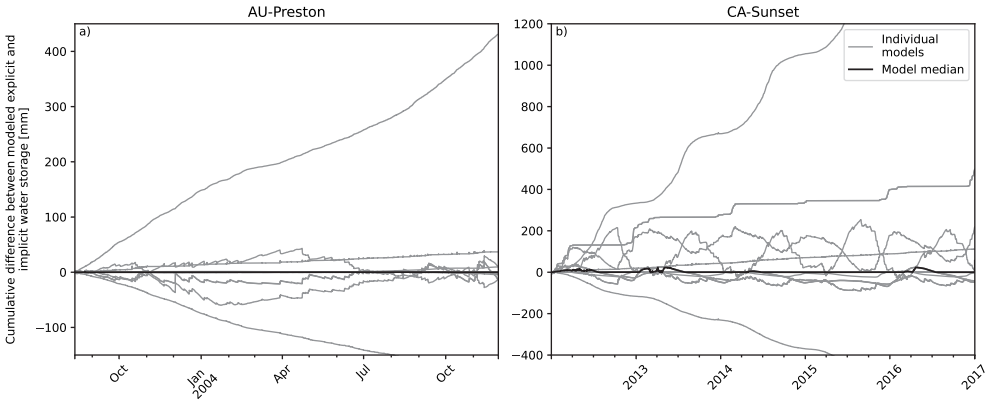


Figure 3.7: Cumulative difference between the explicit (Eq. 3.2) and implicit water storage (Eq. 3.1) at two representative sites for the entire model period for all models. Snowfall occurs at CA-Sunset, but not at AU-Preston. Some models are not visible as they are close to zero.

The range of modeled water storage exceeds the estimated site water storage capacity ($I_{S,m}$) in 64% of cases (Figure 3.4). Models 1 and 5 have the lowest score for this indicator, because they have an inconsistency between the inputs and outputs (Eq. 3.3) causing non-closure of the water balance at nearly all sites. Three models never exceed the estimated water storage capacity.

How explicit relates to implicit water storage is linked to the individual models given the consistent results across sites (Figure 3.8). With magnitude represented by water balance closure, we focus on the timing by assessing the explicit relative to the implicit water storage (Figure 3.9a-c). Model runs can have comparable directions but different patterns, e.g. model 11 (Figure 3.9a), comparable patterns but different magnitudes of change, e.g. model 9 (Figure 3.9b), or virtually no differences (e.g. model 18, Figure 3.9c). The explicit and implicit water storage changes (Figure 3.9d-f) emphasize the difference in timing, which is why the indicator uses the R^2 of these derivatives. Only five models have virtually no differences and thus an R^2 of 1 (Figure 3.4). Over half of the models have R^2 greater than 0.9 indicating timing consistency ($I_{S,t}$, Figure 3.4).

3.3.4 Surface runoff (R_s)

All models have surface runoff triggered by precipitation, but the precipitation event size causing R_s events differs between models (Figure 3.10). The model rather than the site seems to explain triggering event size despite the variation amongst sites in impervious fractions and precipitation regimes. This suggests that surface runoff parameterization may be critical. Thus, we find a large inter-model spread in the cumulative modeled R_s (Figure 3.2). One model is excluded as it does not output R_s separately from R_{sub}. Ten models

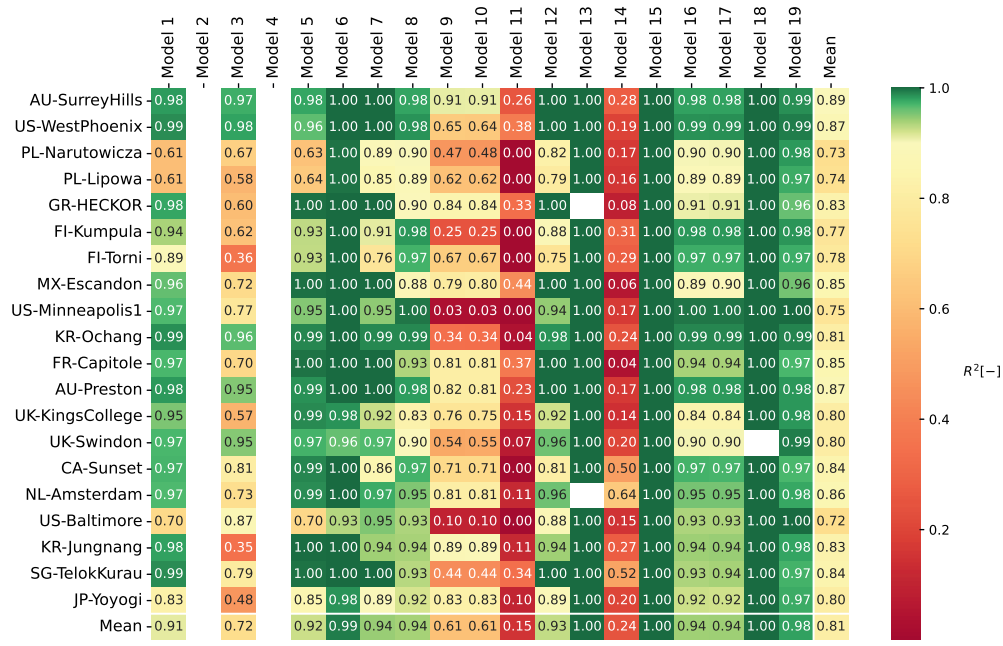


Figure 3.8: Coefficient of determination (R^2) between (half-)hourly explicit (Eq. 3.2) and implicit water storage change (Eq. 3.1) by model and site. Green indicates the 0.9 $I_{S,t}$ threshold (Table 3.1). Missing results are shown as white (i.e. cannot calculate explicit or implicit water storage change). Figure 3.9 may aid interpretation of R^2 values.

show the expected increase of cumulative R_s with increasing site impervious fraction (p >0.05, Wald test (Wald, 1943)), whereas nine models do not (Figure A.3.2).

Only in 43 of the 337 model runs, the CN (curve number: Section 3.2.1) is captured correctly, passing $I_{R,m}$ (Figure 3.4), so all other model runs have no overlap with the site estimates (see Section 3.2.1). Three models capture the CN correctly for at least half of their model runs and are responsible for 32 of the successful model runs. Most models do not match event precipitation and R_s relation. Most models underestimate the CN relative to the site estimate (Figure A.3.3). Underestimating the CN indicates a model is overestimating surface interception and/or soil infiltration, reducing R_s (Equation 3.5).

One in four model runs accurately captures the fast R_s response in the lag time (Figure 3.4) with $I_{R,t}$ passed by 25% of the model runs. With very short lag times expected, only overestimates are simulated. Most lag times averaged per model run are less than five hours, but exceptionally they are over 100 hours. Average lag times per model run are shown in Figure A.3.4.

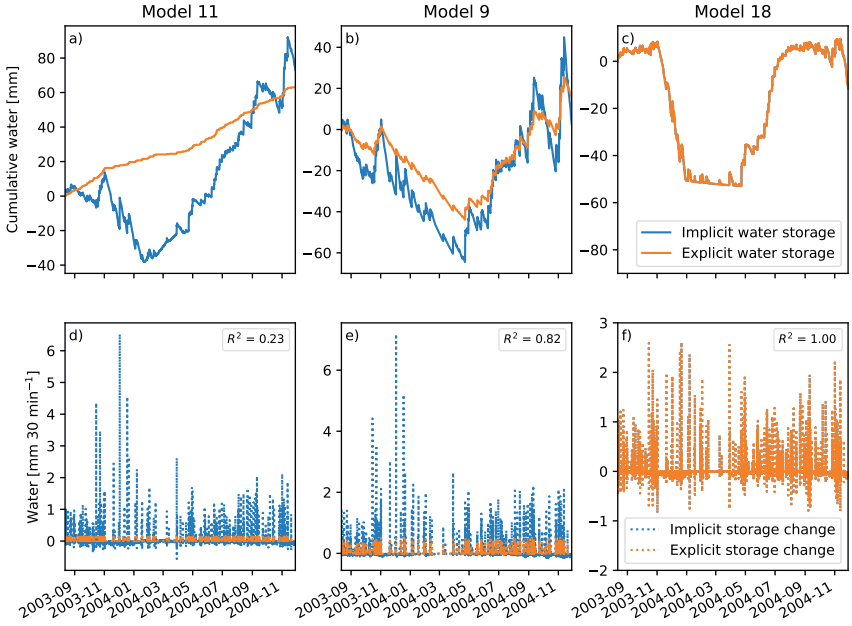


Figure 3.9: Illustration of (a-c) the hourly explicit (Eq. 3.2) and implicit water storage (Eq. 3.1) and (d-f) their derivatives both for 475 days at AU-Preston for three models with increasing coefficient of determination (R^2) of the explicit and implicit water storage change determined at (half-)hourly resolution.

3.3.5 Urban water balance representation (UWBR) score

Across all model runs, the mean UWBR score amounts to 3.3 out of the possible 7 (Figure 3.4). Although the overall pass rate across all indicators and models is 47%, pass rates strongly vary per indicator. Notably, 87% passes $I_{\mathrm{ET},t}$, while only 11% passes $I_{R,m}$. Pass rates also differ among models from 28% to 72%. Only one model run passes all indicators, while 10 model runs have a score of 6 out of 7. Model 19 accounts for five of these eleven high-scoring runs. If a model closes the water balance (I_A), it generally scores better on both storage indicators. In contrast, models with a high passing percentage for one ET indicator do not systematically score better for the other ET indicator. Overall, the ET timing ($I_{\mathrm{ET},t}$) is captured better than its cumulative magnitude ($I_{\mathrm{ET},m}$). A similar pattern is seen in the R_s indicators with the timing ($I_{R,t}$) captured slightly better than magnitude ($I_{R,m}$).

Generally, pass rates per indicator show a dependence on the model (Figure 3.4). This dependence is not found for sites (Figure A.3.5). There is no relation evident between

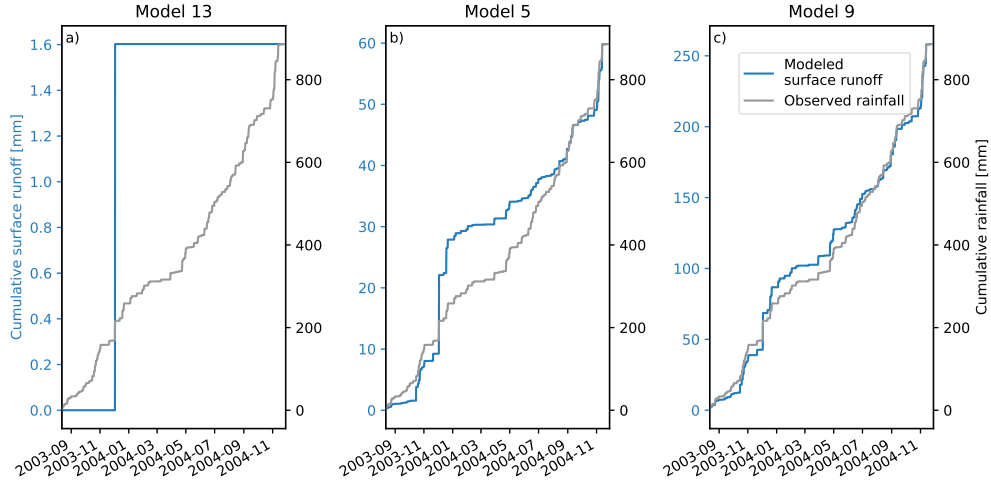


Figure 3.10: Illustration of surface runoff triggered for different AU-Preston precipitation events by three models (a) 13, (b) 5, and (c) 9. Note, the left-hand y-axis (surface runoff) increases (a→c), whereas the right-hand side y-axis (precipitation) is the same for all.

UWBR score and model approach (e.g. built surface, soil hydrology, Table 3.2), but the model is more influential than the site on UWBR score. As the Lipson et al. (2024) classification (Table 3.2) was not developed with the water balance representation as its original goal, further work would be needed to identify what model attributes are key to better UWBR score.

3.3.6 Linking the water and energy balance

Surprisingly, models do not appear to capture any aspect of the latent heat flux more accurately if their UWBR score is higher. The UWBR score does not significantly correlate with better ranking on any of the four metrics evaluating the (half-)hourly modeled Q_E: the R^2, σ_{norm}, MAE_s, and MAE_u (p >0.05, Wald test, Figure A.3.6). These correlations remain absent if one of the indicators is omitted from the analysis. The lack of correlation may be the result of the low number (11) of runs with a UWBR score higher than 5 (Figure 3.4) effectively reducing the UWBR score range. Given the lack of relations between the UWBR score and Q_E metrics, the Q_E is not better captured in model runs that pass more indicators of a realistic water balance representation, thus refuting our hypothesis that the urban water balance skill positively impacts simulated energy fluxes.

3.4 Discussion and conclusions

This study assesses the water balance representation in 19 ULSMs from the Urban-PLUMBER project. It appears the water balance is not closed (within 3%) in 57% of the model-site runs. The considerable spread in water fluxes is as wide as the absolute flux magnitude at all sites. For both ET and R_s, the timing is captured better than the flux magnitude. Modeled explicit water storage dynamics (Eq. 3.2) are inconsistent with the implicit water storage (Eq. 3.1) in 44% of the models. Refuting our hypothesis, a better water balance representation does not result in more accurate latent heat fluxes. However, it is clear that the urban water balance is imperfectly incorporated into ULSMs and more proper physically-based representations are required.

Five models close the water balance at all sites (Models 6, 13, 15, 18, and 19), while three never reach closure (Models 1, 3, and 5). The other models close the water balance at some sites. For several non-closing models, we identify the causes. One model implicitly assumes an infinite source or sink of soil moisture by adapting the modeled soil moisture when it exceeds hard-coded limits adding or removing water to remain within these limits (Model 11). Two other models do not fully couple all processes, such as runoff and evaporation calculations occurring without water availability feedback between processes (Models 1 and 5). Such uncoupled processes may also explain inconsistent water storage dynamics. Three models pass their internal water balance closure check but do not provide the modeled groundwater flux in the model output (Models 8, 16, and 17). We call on the modeling community to include all fluxes required to diagnose water balance closure in the model output. Three models without a snow module disregarded all snowfall creating a mismatch between real and modeled input (Model 2, 7, and 12). For one model, we suspect a very shallow soil layer causes large numerical errors resulting in an unclosed water balance (Model 4). Fortunately, model improvements should be able to eliminate these issues for most models.

Evidence is found that the models would benefit from reevaluating their runoff parameterizations. The runoff volumes are poorly captured, resulting in $I_{R,m}$ having the poorest overall pass rate (Figure 3.4). Runoff has not been evaluated in previous ULSM comparisons and suffers here from a lack of direct observations and small areas being modeled ($<$1 km^2). The lack of correlation between modeled cumulative R_s and the impervious fraction is worrying given the well-documented relation (Shuster et al., 2005; Jacobson, 2011). However, many models use relatively simple approaches, such as a constant fraction of rainfall that runs off independent of site characteristics, rainfall intensity, or soil moisture state. Others use poorly constrained parameters, such as how much water is routed between sub-grid tiles. Future work could help to constrain such parameters, while the simple approaches could be improved relatively straightforwardly.

Despite the lack of evidence showing a link between the UWBR score and Q_E performance, the incomplete representation of the water balance may contribute to the poor latent heat flux performance of the ULSMs. The design of the UWBR score may not be successful in revealing an existing link between the UWBR score and Q_E performance, as the UWBR score indicators assess the water balance based on physical realism and expectations derived from the literature. While a higher UWBR score indicates a more physically consistent water balance, it may still be an incorrect simulation. The opposite is also true, as, without physical constraints, machine learning approaches show good results for Q_E (Vulova et al., 2021). Apart from that, a potential link between the water balance representation and the Q_E performance may be hidden by other elements affecting Q_E performance. These elements could be other components of the model (e.g. the energy balance representation) or human errors (e.g. erroneous parameters, assuming northern-hemisphere vegetation, and results reported in wrong units). Yet, we do find a poor performance for Q_E consistent with the literature showing Q_E is among the most challenging fluxes to model (Grimmond et al., 2011; Lipson et al., 2024). As the energy and water balance are directly connected, we hypothesize potential errors in the water balance are causing, and not being caused by, the poor performance of Q_E, as the short runoff timescales in urban areas on a neighborhood scale dictate the water availability for Q_E and not the other way around. Hence, good model performance for the latent and sensible heat flux cannot be achieved without properly representing both balances. Thus, we believe an improved representation of the water balance will assist in latent heat flux simulation and other energy fluxes.

This first systematic analysis of urban water balance modeling is an opportunistic study taking advantage of model outputs, model characterizations, and observations gathered for the Urban PLUMBER project (Lipson et al., 2022b; Lipson et al., 2024). The Urban-PLUMBER setup affects this study via (1) the diversity of model outputs linked to their range of modeling approaches, and (2) a lack of observations for all the water balance terms. Intentionally, a wide range of modeling approaches are analyzed with both default parameters and provided parameters implemented by modelers (Lipson et al., 2024), impacting the model results and performance. For example, numerical discretization of soil layers can cause a flawed, reduced moisture drydown linked to irregular soil layer depths that enhance evaporation (MacKay et al., 2022). Ongoing land surface model developments to capture and link more processes increase both their scope and complexity, but the number of differing aspects complicates a systematic analysis aiming to attribute performance to certain aspects (Fisher and Koven, 2020; Blyth et al., 2021). To minimize human error, Urban-PLUMBER allowed resubmission of model outputs after web-based and manual checks. As these checks did not address the water balance, we provided an additional basic analysis of the water balance results to catch other human errors with encouragement to resubmit updated outputs. Unfortunately, resubmission reduces but does not eliminate human errors. All differences other than the water balance representation hinder the attribution of the model performance to the water balance concept as they

explain the large variety in model performance amongst models that capture the water balance equally accurately. Ideally, these differences would be eliminated by developing a multi-model framework in the future (Sadegh et al., 2019) and characterizing model types based on water balance approaches. Such a characterization could allow for teasing out more detailed strengths and weaknesses of water balance representations.

Lack of observations (e.g. runoff, soil moisture) prevents direct assessment for many water balance terms. These observations are challenging as both energy and water balance closure need to be considered, so observations need to cover a relatively large uniform area that also constrains the natural and anthropogenic water flows (Grimmond and Oke, 1986; Grimmond and Oke, 1991). A large uniform area is needed as eddy-covariance footprints vary continuously (Grimmond and Oke, 1991; Feigenwinter et al., 2012), while catchment boundaries are static. Hence, we develop a new alternative using quantitative indicators. Each indicator addresses a water balance process and checks whether it complies with physical limits, the model itself, or previous research. We refrain from weighting the indicators to minimize the score subjectivity and prevent one indicator from controlling the outcome. The systematic removal of one of the seven indicators allows us to confirm the UWBR score is not driven by one indicator.

Here, we show ULSMs produce a wide range of water balance results but often do not realistically represent important hydrological processes. Output reporting errors may cause part of the low performance. Although our results are for offline ULSMs, we expect the identified issues will persist in a coupled setting on any scale (e.g., with mesoscale and global models). ULSMs could be improved by ensuring they close the water balance and updating runoff parameterizations. Ideally, future energy-water–carbon studies will try to gather both a wider range of observations but also modeled processes. This will aid improvement of model processes and their feedbacks. However, the complexity of the urban landscape (e.g. different definitions between eddy covariance footprints, and runoff catchments) will require nested model runs and observations to ensure consistency of all. We recommend routine assessment of water balance closure in ULSM development phase applying the indicators of the UWBR score. In a broader context, both model evaluations and comparisons should extend beyond the target variables of the model to all processes that directly influence these variables. This will benefit the broader delivery of integrated urban services (WMO, 2019) and facilitate urban resilience across time scales.

Acknowledgments

We acknowledge the Urban-PLUMBER project team and all observation and modeling participants providing the data set for this research. We would like to thank Judith Boekee, Andrew Frost, and Valentina Marchionni for the fruitful discussions.

Chapter 4

Surface runoff in urban land surface models

This chapter is based on:

H. J. Jongen, A. J. Teuling, M. Lipson, S. Grimmond, M. Best, and G.-J. Steeneveld (2024d). "Do Urban-PLUMBER land surface models underestimate surface runoff?" *In preparation*

Abstract

If rainfall runs off, this water will not cool urban areas by evaporating. With urban runoff typically exceeding that of natural areas, it is a key flux in urban land surface models impacting the simulated exchange of energy and water between surface and atmosphere. Therefore, we explore the surface runoff parameterization in 18 Urban-PLUMBER models to see whether and, if so, why they underestimate surface runoff. The three processes responsible for surface runoff generation are excess water from infiltration, saturation, and interception capacity. Ten models do not represent all processes, and seven do not simulate more surface runoff with an increasing impervious fraction. Not all models produce surface runoff during large (over 20 mm) or intense (exceeding 50 mm h^{-1}) rainfall events. Models produce 43% less surface runoff on average than the curve number method. Models missing one or more runoff processes have greater surface runoff underestimates than models including all. Model differences from curve number predictions are greater during low, rather than high-intensity rainfall events. A better estimate of the surface runoff will improve the estimated water availability in turn benefiting the simulation of evaporation. These improvements would make the models more applicable to plan cooling strategies for cities.

4.1 Introduction

People living in urban areas are exposed to higher temperatures deteriorating sleep quality and even increasing mortality (Sarangi et al., 2021; Fan et al., 2022; Ho et al., 2023). While evaporation can cool urban areas by using on average roughly half of the available energy (Trenberth et al., 2009; Qiu et al., 2023), urban areas are known to produce more runoff than natural areas for a large part due to sealing by impervious surfaces (Leopold, 1968; Paul and Meyer, 2001; Shuster et al., 2005; Gurnell et al., 2007; Jacobson, 2011; McGrane, 2016)leading to lower water availability for evaporation. Vegetation's cooling potential is low in dry conditions (Kraemer and Kabisch, 2022; Nimac et al., 2022). In the future, the need for cooling and dry conditions will coincide more frequently, as compound drought and heatwaves will occur more often (Mukherjee and Mishra, 2021). From an analysis of evaporation recession over cities, Jongen et al. (2022) found that water storage in cities is considerably lower. The low water storage capacity poses a challenge in tackling heat stress amongst urban citizens, especially during these compound events. Solutions like greening are suggested to minimize flooding and heat stress (Schifman et al., 2017). To design effective solutions, it is crucial to understand how much water runs off since this determines the amount of water that remains available for evaporative cooling.

Currently, models simulating land-atmosphere interaction in the urban environment contain relatively simple runoff schemes. These urban land surface models (ULSMs) focus on modeling the energy balance that affects the local urban climate and the overlying atmosphere (Grimmond et al., 2011; Lipson et al., 2024), but runoff is relevant for these models given the physical connection between the water and energy balance through evaporation. Recently, a large set of 30 ULSMs was evaluated at 21 sites in the Urban-PLUMBER project (Lipson et al., 2024). As part of this project, Jongen et al. (2024a) evaluated the water balance based on indicators of a good water balance, as observations of the water fluxes are unavailable. The indicators verify whether the water balance closes, is consistent within the model, and water fluxes fall within the expected ranges. The indicators show that ULSMs generally poorly capture runoff illustrated by the missing relation between surface runoff and the impervious fraction in almost half of the ULSMs in that study (Jongen et al., 2024a). At the same time, Oswald et al. (2023) already urge to study the influence of urbanization on the water balance beyond the impervious surfaces. The ULSMs not capturing this relation emphasize the need to improve these models' surface runoff parameterization. Therefore, we identify how ULSMs represent runoff processes and how this could be improved.

Evaluating the surface runoff parameterization requires considering the runoff processes at play. Traditionally, surface runoff in natural environments is believed to originate from either infiltration excess (Horton, 1933) and saturation excess (Dunne and Black, 1970). Impervious surfaces produce surface runoff by preventing infiltration, which could be seen as infiltration excess. At the same time, they provide small storage that quickly

overflows, which could be seen as saturation excess. Therefore, we use a different term to describe this combination: interception excess. The surface runoff from impervious surfaces may partly drain to pervious areas where it is either buffered or causes additional saturation excess (Hopkins et al., 2015). Moreover, pervious areas themselves are affected by urbanization as soil layers are removed (Herrmann et al., 2018) and soils are compacted (Shuster et al., 2015). These changes result in more surface runoff through infiltration excess from urban than rural pervious surfaces. The increase in surface runoff is generated due to the combination of interception, infiltration, or saturation excess (Stewart et al., 2019). All three processes should therefore be considered in modeling. Yet, not all models include all three processes in their conceptualization (McDonnell, 2013).

The surface runoff evaluation is hampered by lacking appropriate observations. Even though runoff is occasionally measured in urban catchments (Berthier et al., 1999; Walsh et al., 2005a), these observations do not reflect the spatial scales considered in ULSMs (Grimmond and Oke, 1986; Grimmond et al., 2011; Lipson et al., 2024). When surface runoff observations are not available, the Curve Number (CN) method is a widely applied tool to estimate surface runoff based on the site characteristics (Cronshey et al., 1985; NRCS, 2004a). As it relies on commonly available site characteristics, it is generally applicable and CN look-up tables are based on observations for many sites. When the impervious fraction is considered, the CN method is suitable for and often used in urban areas (e.g. Bhaduri et al., 2000; Xu et al., 2020; Chin, 2022; Wu et al., 2024). Thus, the CN method provides a reference against which ULSMs can be compared.

Here we examine the modeled surface runoff from 18 ULSMs from the Urban-PLUMBER project (Lipson et al., 2024). These models contain a range of different parameterizations for surface runoff. The paper aims to provide insight into how runoff processes are captured in the current models and how this affects modeled surface runoff. Specifically, we investigate the hypothesis that ULSMs underestimate surface runoff because their parameterizations might not reflect runoff processes in urban environments. The study starts by analyzing the effect of impervious fraction and precipitation characteristics using model runs from 21 sites worldwide. Thereafter, we focus on two sites, one with a low and one with a high impervious fraction, for one low and one high-intensity precipitation event per site. For all four events, the total modeled event surface runoff is compared with estimates based on the CN method. Based on the high-intensity event at the most impervious site, we demonstrate the impact of including or excluding runoff processes from the models.

4.2 Methods

The present study analyzes the surface runoff parameterization in 18 Urban-PLUMBER USLMs outputting the surface runoff (Table 4.1). The models represent an area in the order of magnitude of 1 km^2. No observations were provided to the participating

modelers preventing calibration. The model outputs cover 21 sites with a range of climates, impervious fractions, and observational periods described in detail by Lipson et al. (2022b). Depending on site data availability, surface runoff is evaluated at 30- or 60-minute resolution for periods of 148 to 1827 days (average 912 and median 748 days). Most models calculate the surface runoff at the forcing resolution without sub-timesteps, while a minority divides the timestep into substeps for the calculations.

To start, the relation between site characteristics and surface runoff is considered across sites with varying total precipitation given differences in observational periods and local climate. This relation is examined over the whole model run based on the accumulated surface runoff and precipitation. Water fluxes are expressed as ratios of the total inflow (e.g. surface runoff ratio), defined by the left-hand side of the urban water balance (Grimmond and Oke, 1986):

$$P + I = R_s + R_{ss} + ET + \Delta S \tag{4.1}$$

where P is precipitation, I irrigation, R_s surface and R_{ss} subsurface runoff, and ΔS water storage change. In this study, we focus on surface runoff in the models. As no rivers or streams are present in these models, surface runoff is interpreted as overland flow leaving the model domain.

While general metrics, such as the surface runoff ratio, help to compare model behavior, analyzing specific events reveals more differences between models (Boer-Euser et al., 2016). Surface runoff is studied at (half-)hourly timesteps to distinguish all three runoff generation processes, infiltration, saturation, and interception excess. We select a long low- and high-intensity precipitation event at two sites, one with a low and one with a high impervious fraction. While long low-intensity rain is likely to trigger saturation and interception excess, it is unlikely to trigger infiltration excess. Whereas, a high-intensity precipitation event will generate surface runoff through all three processes. Differences in prescribed impervious fractions assist in unraveling how models including infiltration and interception runoff handle both. We select precipitation events without dry periods longer than one hour with total precipitation, for the low-intensity events, exceeding 20 mm with an average rate of maximally 2 mm h^{-1} and, for the high-intensity events, exceeding 50 mm with an average rate above 10 mm h^{-1}. The selected events are described in Table 4.2.

To compare the modeled surface runoff with expected values without observations, we estimate the expected total event surface runoff with the Curve Number (CN) method (Cronshey et al., 1985). This widely applied method provides a robust surface runoff estimation in response to a given rainfall event based on the CN of a site. The CN summarizes the site characteristics relevant for runoff generation and can be found in look-up tables that have been established based on numerous sites (Rawls et al., 1981). CN-based surface runoff estimates do not reflect subsurface information beyond the soil type meaning soil depth and bedrock permeability are ignored (Weedon et al., 2023).

Table 4.1: Characteristics of the 18 urban land surface models used for the surface runoff analysis. All models are part of the Urban-PLUMBER project (Lipson et al., 2024). Two models did not provide soil moisture output ([1]). Three models capable of simulating irrigation did not include it in their Urban-PLUMBER runs ([2]).

Model	Urban geometry	Vegetation	Soil hydrology	Snow accumulation	Infiltration excess runoff	Saturation excess runoff	Interception excess runoff	Irrigation	Reference
ASLUMv2.0	Canyon	Grass	Multi-layer[1]	No	Yes	Yes	Yes	No[2]	Wang et al. (2013) Wang et al. (2021)
ASLUMv3.1	Canyon	Grass+trees	Multi-layer[1]	No	Yes	Yes	Yes	No[2]	Wang et al. (2013) Wang et al. (2021)
CABLE	Non-urban	Separate tile	Multi-layer	Veg.	No	Yes	No	No	Kowalczyk et al. (2006) Wang et al. (2011)
ECLand	Non-urban	Separate tile	Multi-layer	Veg.	Yes	No	No	No	Boussetta et al. (2021)
ECLand-U	Two-tile	Separate tile	Multi-layer	Veg.+urban	Yes	No	No	No	McNorton et al. (2021) Boussetta et al. (2021)
CLMU5	Canyon	Grass+shrubs	Multi-layer	Urban	Yes	Yes	Yes	No	Oleson and Feddema (2020)
JULES 1T	One-tile	Separate tile	Multi-layer	Veg.+urban	Yes	Yes	Yes	No	Best et al. (2011)
JULES 2T	Two-tile	Separate tile	Multi-layer	Veg.+urban	Yes	Yes	Yes	No	Best et al. (2011)
JULES MOR	Two-tile	Separate tile	Multi-layer	Veg.+urban	Yes	Yes	Yes	No	Best et al. (2011)
Manabe 1T	One-tile	Manabe bucket	One-layer	Veg.+urban	No	Yes	Yes	No	Best et al. (2011) Manabe (1969)
Manabe 2T	Two-tile	Manabe bucket	One-layer	Veg.+urban	No	Yes	Yes	No	Best et al. (2011) Manabe (1969)
NOAH-SLAB	One-tile	Separate tile	Multi-layer	Veg.+urban	Yes	No	Yes	No	Kusaka et al. (2001) Ek et al. (2003)
NOAH-SLUCM	Canyon	Separate tile	Multi-layer	Veg.+urban	Yes	No	Yes	No	Kusaka et al. (2001) Ek et al. (2003)
SNUUCM	Canyon	Seperate tile	Multi-layer	Veg.	Yes	No	Yes	No	Ryu et al. (2011) Ek et al. (2003)
SUEWS	Two-tile	Separate tile	One-layer	Veg.+urban	Yes	Yes	Yes	No[2]	Järvi et al. (2011) Ward et al. (2016)
TERRA 4.11	One-tile	LSM	Multi-layer	Veg.	No	Yes	Yes	No	Wouters et al. (2015) Schulz and Vogel (2020)
UCLEM	Canyon	Grass+shrubs	One-layer	Veg.+urban	No	Yes	Yes	Yes	Thatcher and Hurley (2012) Lipson et al. (2018)
UT&C	Canyon	Grass+shrubs+trees	Multi-layer	No	Yes	Yes	Yes	Yes	Meili et al. (2020)

Table 4.2: Overview of the two low- and two high-intensity events at two sites, one with a low and one with a high impervious fraction,. The CN is estimated according to NRCS (2004b). The average precipitation rate is calculated considering timesteps with non-zero precipitation.

Site	***Impervious fraction [-]***	***CN [-]***	**Event intensity**	***Tot. precip. [mm]***	***Avg. precip. rate [mm h^{-1}]***
US-Baltimore	0.31	58	Low	28	2.0
			High	74	12.4
KR-Jungnang	0.96	97	Low	22	1.9
			High	59	23.4

These factors determine how runoff is divided over surface and subsurface flow and how fast (urban) catchments respond to rainfall (Weedon et al., 2015). In this study, CN look-up table values are linearly interpolated using the impervious fraction giving composite CNs (NRCS, 2004b). The water storage is assumed to be empty at the beginning of the event. As the CNs cannot be calibrated, we test the sensitivity of the results by varying the CNs within the CN uncertainty due to the soil composition given the impervious fraction. The relation between total event surface runoff (R_e) and CN is:

$$R_e = \frac{(P_e - 0.2S)^2}{P_e + 0.8S} \text{ with } S = \frac{1000}{\text{CN}} - 10 \text{ assuming } I_a = 0.2S \tag{4.2}$$

where P_e is the total event precipitation, S the potential maximum retention, and I_a the initial abstraction. The total surface runoff is compared with the surface runoff produced by the models. The comparison expresses the difference between the modeled surface runoff and CN estimate in percent of the CN estimate.

4.3 Results

With more surface runoff expected from impervious than pervious surfaces, models likely underestimate surface runoff when impervious surfaces are not correctly captured. Most models seem to correctly capture impervious surface runoff with surface runoff ratios rising with the impervious fraction (Figure 4.1). This is consistent with the observations having less evaporation at more impervious sites (Figure 4.1a). Two models generate surface runoff ratios with an almost perfect correlation with the impervious fraction (0.95 and 0.98, Figure 4.1g,s) and a surface runoff ratio near zero at the site with the low impervious fraction. This suggests these two models only generate surface runoff from impervious surfaces. Whereas, seven models have no significant correlation between modeled surface runoff ratios and impervious fraction (p >0.05, Wald (1943) test , Figure 4.1b,d,f,m,n,o,p). Of these seven, four models produce little surface runoff regardless of impervious fraction (Figure 4.1d,f,m,n). Two of these four models are not developed for urban areas, instead representing the impervious areas as bare soil (Figure 4.1d/f).

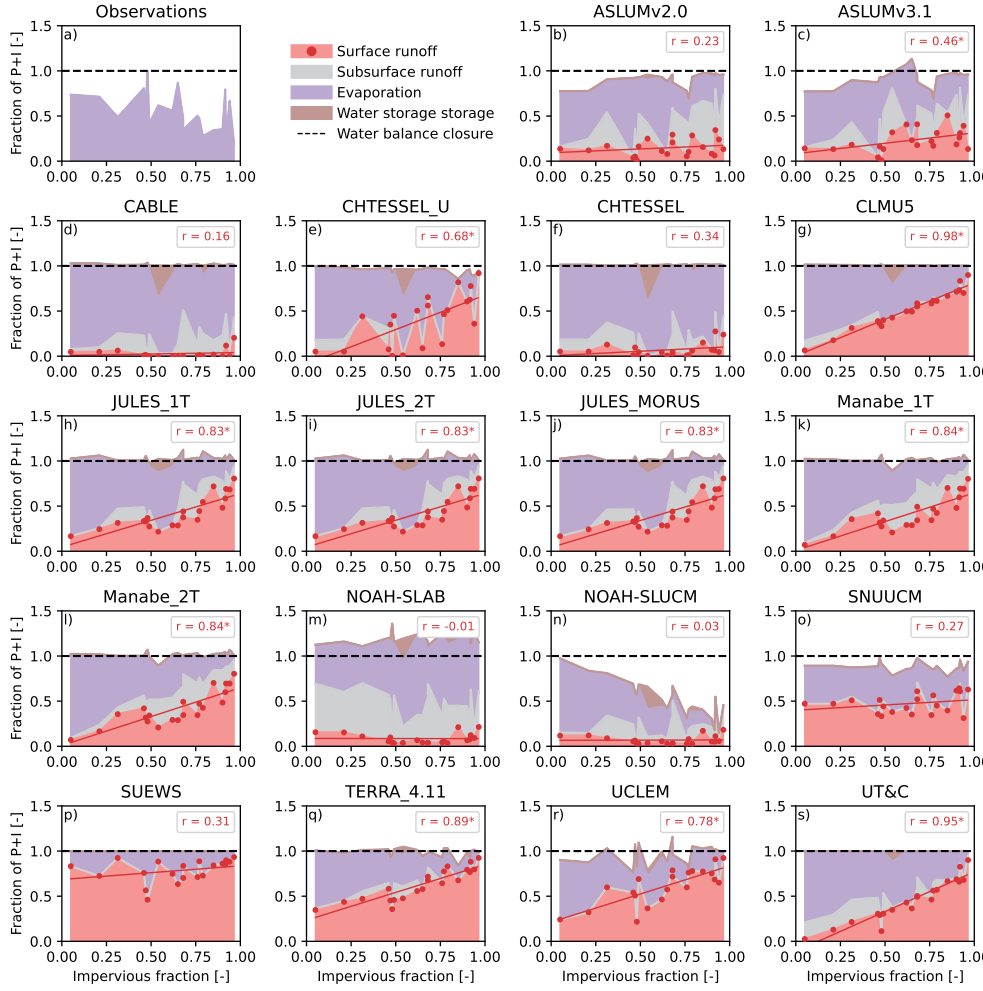


Figure 4.1: Modeled average water balance partitioning (shading) for all 21 Urban- PLUMBER sites sorted by impervious fraction for (a) observations and (b-s) the 18 models. Observations are only available for evapotranspiration. The red lines show linear regression between surface runoff and the impervious fraction. Significant correlations are indicated with an asterisk ($p > 0.05$, Wald (1943) test).

One model fails to partition all the incoming water (Figure 4.1n). This issue seems related to the impervious scheme as the unexplained fraction increases with impervious fraction. Rainfall intensity does not correlate significantly with the surface runoff ratio except in two models (not shown). No significant correlation was visible with the two other tested site characteristics (annual rainfall in mm and soil texture in % sand, not shown). Misrepresented impervious surfaces seem to cause underestimation of surface runoff, as not

all models produce the correlation between modeled surface runoff ratios and impervious surfaces.

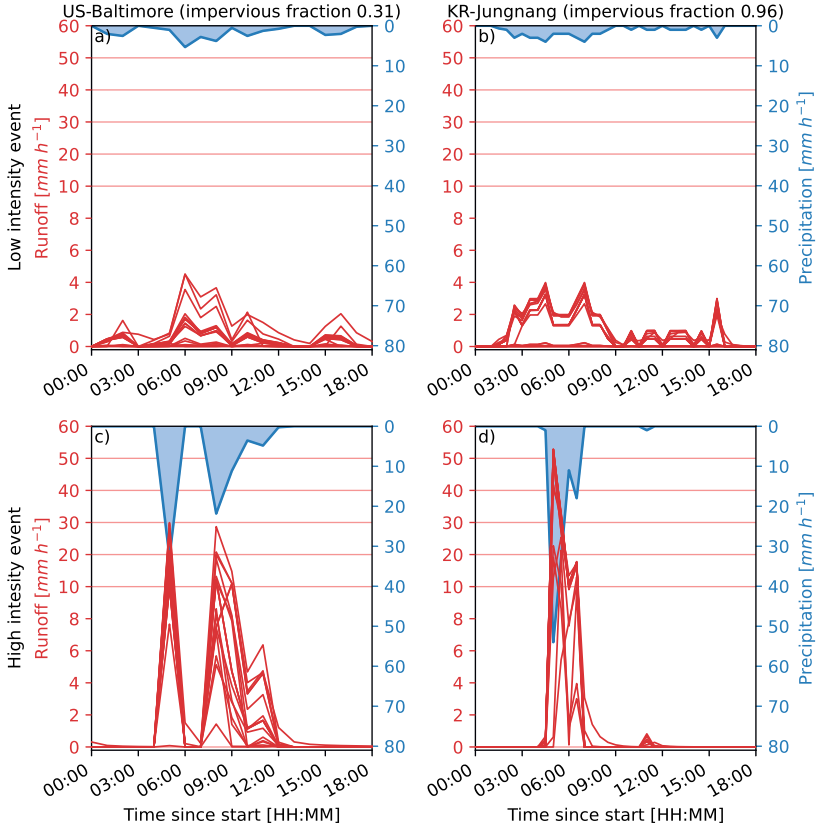


Figure 4.2: Modeled surface runoff (one line per model) for (a,b) low- and (c,d) high-intensity precipitation events (shading, Table 4.2) at sites with (a,c) low (US-Baltimore) and (b,d) high (KR-Jungnang) impervious fractions.

Next, we consider specific events expected to produce surface runoff to assess the model response to different rainfall intensities at sites with different impervious fractions (Figure 4.2). In this analysis, it is important to consider that some models calculate the runoff at shorter intervals than the model output of 30 or 60 minutes, while others only have one model timestep in this period. The latter acts to make the effective rainfall intensity greater as peaks are not averaged out (Grimmond and Oke, 1991). Modeled surface runoff closely mimics rainfall dynamics in most models during all four events. However, the amount of modeled surface runoff differs among the ULSMs ranging from no surface runoff to almost all precipitation for the low-intensity event at the less impervious site (Figure

4.2a). At the more impervious site, model results are more clustered with six models yielding no or negligible surface runoff and twelve following the rainfall intensity pattern (Figure 4.2b). During the high-intensity events, modeled surface runoff directly responds to rainfall in all but two models, yet the peak surface runoff rates differ by a factor of four (Figure 4.2c-d). Surprisingly, at the more impervious site modeled surface runoff is zero for some models during two hours with precipitation rates above 10 mm h^{-1} despite an impervious fraction of 0.96. These four events produce a large range of surface runoff and many models fail to produce surface runoff when expected.

To evaluate their realism, the ULSMs are compared to CN surface runoff estimates. The CN estimates are based on total precipitation during the event and do thus not predict runoff dependent on rainfall intensity. Generally, the ULSMs yield lower event surface runoff for all four events than the CN method (Figure 4.3) with an average underestimation of 44%. The underestimation is higher at the less impervious site (53%, cf. more 34%) and lower during high-intensity events (32%, cf. low 55%).

For low-intensity events, the major difference is whether any surface runoff is generated. For high-intensity events, surface runoff is generated by all models, but the totals vary widely. One model without modeled irrigation even produces more surface runoff than the total rainfall. Five models predict relatively low surface runoff for the high-intensity event at the more impervious site. These models also showed no correlation between impervious fraction and modeled surface runoff ratio. Models including all runoff processes are closer to the CN method estimates (35% lower) than models missing processes (50% lower). The results are not sensitive to the exact CN assigned to the sites (US-Baltimore=58 and KR-Jungnang=97). To assess the sensitivity of the results to the estimated CN, the CN range based on site impervious fractions is from 46 to 86 for US-Baltimore, and 92 and 98 for KR-Jungnang. Using these CN ranges, the average underestimation changes minimally varying between 40% and 46%. The comparison with the CN method suggests ULSMs underestimate surface runoff.

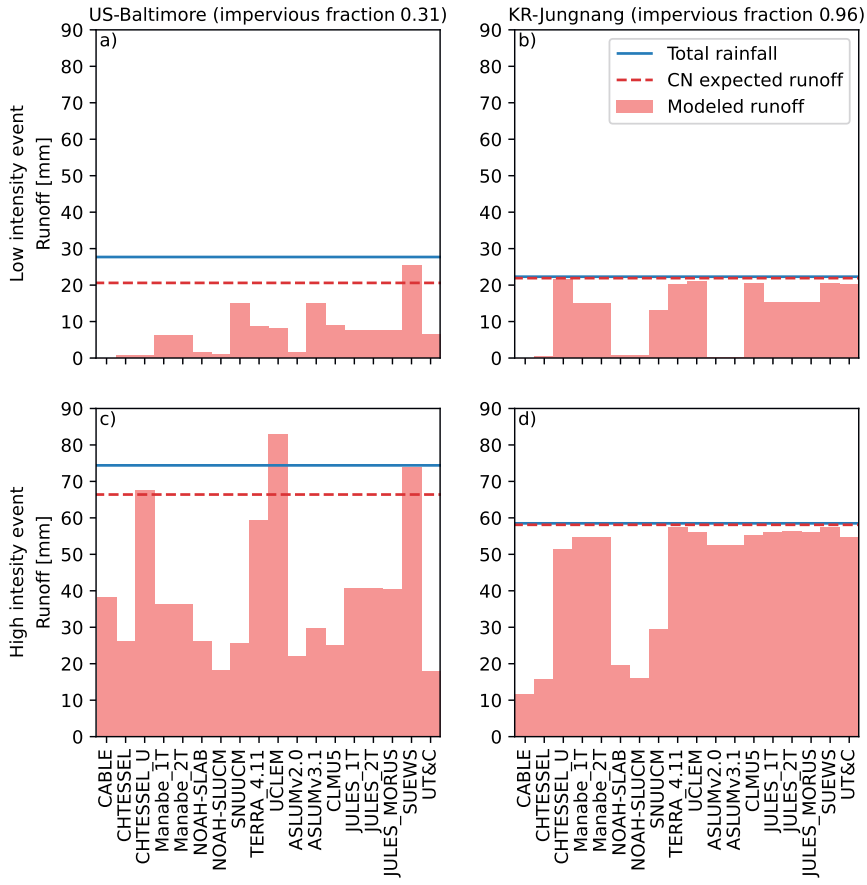


Figure 4.3: Comparison of modeled surface runoff with the corresponding Curve Number reference (dashed line) for (a,b) low- and (c,d) high-intensity precipitation events (Table 4.2) at sites with (a,c) low (US-Baltimore) and (b,d) high (KR-Jungnang) impervious fractions with total event precipitation (solid line). Models are sorted by the number of runoff processes included (increasing from left to right).

To reveal the effect of the runoff processes, we examine the differences between models that include or exclude infiltration, saturation, and interception excess. Missing runoff processes cause unlikely hydrographs with delayed or low surface runoff peaks (Figure 4.4). Models missing infiltration excess do not show very distinct behavior, as the site is mainly impervious. However, one model includes neither infiltration nor interception excess, only generating surface runoff through saturation excess. Consequently, this model has a delay before surface runoff starts even with precipitation rates of more than 50 mm h^{-1}. All models missing saturation excess simulate lower peak surface runoff than

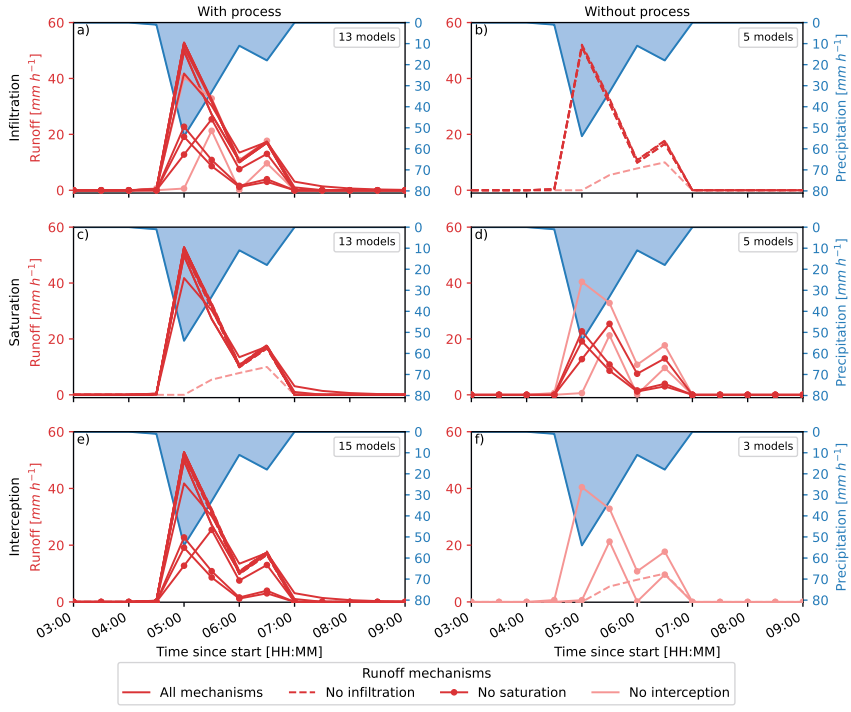


Figure 4.4: Dependence of modeled surface runoff response to model runoff parameterization illustrated with high-intensity precipitation at a highly impervious site (0.96, KR-Jungnang) for 18 Urban-PLUMBER models including (a) infiltration, (c) saturation, and (e) interception excess; and missing (b) infiltration, (d) saturation, and (f) interception excess. The indicated time since start corresponds with Figure 4.2. The number of models is indicated in each panel.

models with saturation excess (except the model with only saturation excess). The three models missing interception excess produce less and delayed surface runoff, as they do not account for imperviousness. The models missing runoff processes illustrate all processes are necessary to capture surface runoff. Missing interception causes delayed surface runoff and missing saturation causes lower peak surface runoff.

4.4 Discussion and conclusions

By examining the surface runoff parameterization of 18 Urban-PLUMBER ULSMs, we find not all surface runoff generation processes are currently included. Our findings suggest most ULSMs underestimate surface runoff. Compared to the Curve Number (CN) method, models produce 44% less surface runoff on average. The underestimation is higher at less impervious sites and during low-intensity rainfall. Models including infiltration, saturation, and interception excess generate higher surface runoff ratios and surface runoff peaks

indicating the underestimation may result from the missing runoff processes. Additionally, seven models fail to capture the increase in surface runoff due to higher impervious fractions. In these models, we suspect the impervious surface parameterization is inadequate.

Two factors relevant to surface runoff have not been analyzed, as none of the 18 models consider them. First, the slope of the topography is a key factor in surface runoff generation (Hudson, 1985). Even a gentle slope increases surface runoff; a slope of 1% can double surface runoff while a slope of 2.5% can triple surface runoff (Haggard et al., 2005). While topography from grid cell to grid cell is part of numerical weather prediction and climate models, the influence of sub-grid topography is parameterized for other processes like radiation (Helbig and Löwe, 2012) and drag (Sützl et al., 2021) but not for runoff. Inspiration to include the sub-grid slope effect could be gained from TOPURBAN (Valeo and Moin, 2000), a variable source area model adapted for the urban environment based on TOPMODEL (Beven et al., 1995). TOPURBAN uses the topographic index combining the slope with the upstream drainage area. The TOPURBAN approach includes subgrid-slope effects but is likely to change the subsurface more strongly than the surface runoff.

Second, most models do not separate drainage and piped water flows from surface runoff. At most, they provide a first-order estimate. The separation did not affect this project, as in Urban-PLUMBER the models were run standalone and not spatially distributed or coupled to atmospheric models. Surface runoff is partly routed to drainage and streamflow. In this paper, we focus on runoff generation making it less relevant whether water leaves the model as surface runoff, streamflow, or drainage. When these models are spatially distributed or coupled to atmospheric models, it would be useful to separate these fluxes. For this separation, crucial information is the connectivity of impervious surfaces to streams and drainage systems (Baruch et al., 2018).

The surface runoff and evaporation shown here are calculated at the temporal model resolution without sub-timestep calculations in most evaluated models. This lack of sub-timestep calculations means the models in this study calculate the surface runoff at relatively long (half-)hourly timesteps. Long timesteps and the loss of detail in the rainfall intensities change the direct model runoff response to rainfall having prolonged effects through the moisture conditions during dry periods (Ward et al., 2018). Moreover, the long timesteps make the models sensitive to the order of the runoff and evaporation parameterizations. Sub-timesteps have been introduced to resolve this issue but are not implemented in most of the included models. They could be included without rigorously changing the modeling concepts, also allowing the use of forcings with a higher time resolution.

Currently, ULSMs miss runoff generation processes leading to underestimation of surface runoff. Additionally, seven models do not have a significant correlation between impervious fraction and surface runoff ratio according to the Wald (1943) test ($p > 0.05$). Improvements could be made by at least including all three runoff generation processes

and potentially extending the models to account for slope and drainage. By improving the surface runoff parameterization, the models will in turn better capture the water available for evapotranspiration. As evapotranspiration from vegetation is a strategy to lower temperature in overheated cities, the ULSMs will be better suited to predict the achieved benefits from vegetation toward more livable and resilient cities.

Acknowledgments

We acknowledge the Urban-PLUMBER project team and all observation and modeling participants providing the data set for this research.

Chapter 5

Linking surface cover and evapotranspiration

This chapter has originally been published as:

H. J. Jongen, S. Vulova, F. Meier, G. J. Steeneveld, F. A. Jansen, D. Tetzlaff, B. Kleinschmit, N. Haacke, and A. J. Teuling (2024b). "Attributing urban evapotranspiration from eddy-covariance to surface cover: Bottom-up versus top-down". *Water Resources Research* 60 (9), e2024WR037508

Abstract

Evapotranspiration (ET) is a key process in the hydrological cycle that can help mitigate urban heat. ET depends on the surface cover, as the surface affects the partitioning of precipitation between runoff and evapotranspiration. In urban neighborhoods, this surface cover is highly heterogeneous. The resulting neighborhood-scale ET can be observed with eddy-covariance systems. However, these observations represent the signal from wind- and stability-dependent footprints resulting in a continuously changing contribution of surface cover types to the observation. This continuous change prevents quantifying the contribution of the surface cover types to neighborhood ET and their hourly dynamics. Here, we disentangle this neighborhood-scale ET at two sites in Berlin attributing the patch-scale ET dynamics to the four major surface cover types in the footprint: impervious surfaces, low vegetation, high vegetation, and open water. From the bottom-up, we reconstruct neighborhood ET based on patch-scale observations and conceptual models. Alternatively, we start top-down and attribute neighborhood ET to the surface cover types solving a system of equations for three eddy-covariance systems. Although data requirements for the bottom-up approach are met more frequently, both approaches indicate that vegetation is responsible for more ET than proportional to its surface fraction in the footprint related to the large evaporating surface compared to the ground surface. Evaporation from impervious surfaces cannot be neglected, although it is less than from vegetation due to limited water availability. The limited water availability causes impervious surfaces to cease evaporation hours after rainfall, while vegetation and open water sustain ET for extended periods.

5.1 Introduction

How precipitation is partitioned between runoff and evapotranspiration (ET) plays an important role in the urban climate and is governed by the surface cover composition (Paul and Meyer, 2001; Oke et al., 2017). In cities, the abundant impervious surfaces prevent infiltration and promote surface runoff leaving less water available for ET than pervious areas (Fletcher et al., 2013; McGrane, 2016; Jongen et al., 2022). On the other hand, urban vegetation has the opposite effect increasing ET (Peters et al., 2011; Gunawardena et al., 2017). While all vegetation favors ET compared to impervious surfaces, an isotope-based study revealed the vegetation type also affects infiltration and ET patterns (Kuhlemann et al., 2021). The composition of the surface cover thus controls the water partitioning and consequently ET dynamics.

Promoting green surface covers by planting vegetation over impervious surfaces can increase ET using more of the available energy (Wang and Shu, 2020). Like vegetation, open water is suggested to potentially cool its surroundings by evaporation when implemented appropriately (Solcerova et al., 2019; Jacobs et al., 2020), although warming can also occur due to the high thermal inertia (Theeuwes et al., 2013; Steeneveld et al., 2014). The energy needed for the additional ET cannot heat the air, which thus reduces temperatures and mitigates heat stress and the associated health risks (Oke, 1982; Heaviside et al., 2017; Ward and Grimmond, 2017). However, how ET at the patch level ($\sim 10^1 - 10^2$ m of a single surface cover type, ET_{patch}) translates to the neighborhood scale ($\sim 10^3$ m, $ET_{neighborhood}$) is largely unknown until now. As ET is both an energy and water balance flux, we need to quantify how surface cover impacts the partitioning of incoming water fluxes (Bonneau et al., 2018) and how this affects the partitioning on the larger, neighborhood scale. Ultimately, the neighborhood scale is where the effect of the surface cover types on ET needs to be understood. In time, this understanding will support the management of the cooling benefits and urban water demands.

At the neighborhood scale, eddy-covariance (EC) systems observe the ET of the combined surface cover types in the footprint at a given moment (Feigenwinter et al., 2012). Even though the heterogeneous urban surface results in spatially variable ET (Qin et al., 2022), the observed ET represents the weighted average flux in the footprint, as the EC systems are typically installed at a height where the heterogeneous surface flux sources are blended (Oke et al., 2017). The footprint varies temporally depending on the observation height, wind speed/direction, and atmospheric stability (Kljun et al., 2015). Previous research demonstrated it is possible to upscale ET_{patch} observations to the neighborhood-scale EC observations weighed by the contribution of a surface cover type to the footprint climatology at a relatively homogeneous urban site (Peters et al., 2011). However, hour-to-hour variation in the footprint contains useful information to quantify the timing of cooling benefits from ET. This time-dependent surface information has been applied to improve the model performance of urban $ET_{neighborhood}$ machine learning models (Vulova

et al., 2021). Thus, for heterogeneous urban sites, the footprint is crucial information to disentangle $ET_{neighborhood}$ and attribute it to the different surface cover types.

EC footprints can be estimated with a variety of models. Large-eddy simulations (LES) or Lagrangian stochastic particle dispersion models (LPD) fully model the airflow to find the source area (LES: Leclerc et al., 1997; Wang and Davis, 2008; LPD: Kljun et al., 2002; Hsieh et al., 2003; LES and LPD combined: Hellsten et al., 2015; Auvinen et al., 2017). These models are both labor-intensive and computationally expensive, which limits their applicability to relatively short-duration case studies (Vesala et al., 2008b). To analyze longer time series, faster footprint models have been developed with an analytical approach relying on the surface-layer theory e.g. Schuepp et al., 1990; Schmid and Oke, 1990; Kormann and Meixner, 2001. Their validity is restricted to certain turbulence intensities or stratifications. More recently, Kljun et al. (2015) developed a two-dimensional footprint parameterization that takes away these limitations. Their model yields robust results for most boundary layer conditions at any observation height within the surface layer. This model enables the identification of the flux's source area for a long time series with a wide range of atmospheric conditions. Therefore, this model is suitable to study the influence of the changing footprints on $ET_{neighborhood}$.

To study the influence of surface cover on $ET_{neighborhood}$, Peters et al. (2011) have described the seasonal patterns in urban ET_{patch} from major plant-functional types (trees and turf grass). Vegetation and open water were observed with EC systems and sap flow sensors, while impervious surface evaporation was assumed negligible. The two vegetation types explain the majority of $ET_{neighborhood}$ variation. They also found that the surface fraction of a vegetation type is the most important factor determining its contribution to total $ET_{neighborhood}$ underlining the importance of the EC footprint. Other studies challenge their assumption that impervious surface evaporation can be neglected (Ramamurthy and Bou-Zeid, 2014; Chen et al., 2023). Below, we will test the assumption by including evaporation from impervious surfaces. Moreover, while their analysis is focused on the seasonal timescale, we will consider the hourly timescale. The hourly ET dynamics play a key role in the urban climate experienced by urban citizens. As a verification, Peters et al. (2011) compared the sum of their observed ET_{patch} against EC observations, in essence reconstructing the $ET_{neighborhood}$ signal from the bottom up.

While the evaporation dynamics from a single surface cover type has been investigated previously, few studies have addressed these issues across a range of surface cover types. These studies show that surface cover types have very different evaporation dynamics. Four main surface cover types can be distinguished: impervious surface, low vegetation, high vegetation, and open water. Impervious surfaces only evaporate when wet directly after rainfall resulting in highly dynamic evaporation (Wouters et al., 2015). In contrast, vegetation can draw water from the soil sustaining ET long after rainfall (Teuling et al., 2006; Boese et al., 2019). Amongst vegetation, differences are seen with higher average ET

for higher vegetation with its higher leaf area density than for lower vegetation (Gillefalk et al., 2021; Kuhlemann et al., 2021; Gillefalk et al., 2022). Sufficiently deep open water has more constant evaporation given the abundant water and high heat storage capacity that can provide energy in the absence of solar radiation (Jansen et al., 2022). The term ET is used for vegetation because it contains the signals from transpiration, interception, and soil evaporation, and for EC-observed ET because it contains the combined signal of the present surface cover types including evaporation, transpiration, and anthropogenic fluxes (e.g. combustion and human metabolism). Over impervious and open water surfaces, only evaporation occurs so the term evaporation is used. We hypothesize these behaviors are combined at the neighborhood scale, as observed with EC, dependent on their relative contribution to the surface.

Very few cities have observations of all relevant hydrometeorological states and fluxes across a range of surface cover types. Berlin is a notable exception. In Berlin, meteorological observations are performed as part of two observatories: the Urban Climate Observatory operated by the Chair of Climatology at the Technische Universität Berlin (https://uco.berlin/en, Scherer et al., 2019) and the Steglitz Urban Ecohydrological Observatory from the IGB Leibniz-Institute of Freshwater Ecology and Inland Fisheries (Kuhlemann et al., 2020; Kuhlemann et al., 2021). Additionally, campaigns have added to this observation infrastructure, for example, with drone-based observations (Vulova et al., 2019) or with ground-based remote sensing (Zeeman et al., 2023). The elaborate observation infrastructure has resulted in numerous studies focusing on Berlin (e.g. Meier and Scherer, 2012; Fenner et al., 2014; Fenner et al., 2023), of which many focused on ET. A modeling study applied a physics-based model to study hourly ET (Duarte Rocha et al., 2022), which after validation was combined with remotely-sensed vegetation characteristics to map ET for all of Berlin (Rocha et al., 2022). Vulova et al. (2021) achieved similar modeling skill with machine learning trained on meteorological and remote sensing data. Because of the research infrastructure and the extensive literature, Berlin offers a unique setting to study the link between the surface cover and ET.

In this study, we aim to estimate the ET contribution of different surface cover types in the footprint profiting from the diverse observations in Berlin. With this, we will show how the footprint varies over time, how ET behaves for each surface cover type, the relation between the surface cover and neighborhood ET, and the contribution of each surface cover type to neighborhood ET. In this paper, the ET contribution is always weighted by the footprint. To study the contribution of each surface cover type to ET, we take both a bottom-up and a top-down approach to attribute the EC-observed ET to the four dominant surface cover types. For the bottom-up approach, we reconstruct the EC signal by summing the estimated ET contribution of each surface cover type weighed by its contribution to the footprint. In this approach, the ET contribution of each type is estimated with conceptual models and small-scale observations. These patch-scale models and observations can be verified with the EC observations, as the EC footprint is

modeled for every observation. The top-down approach is based on a system of equations, in which each equation describes the surface cover composition of one EC system and its resulting flux. The resulting flux can be attributed to the surface cover types by solving the system of equations. We aim to reveal how the surface cover type influences neighborhood ET behavior and to indicate how altering surface cover may affect urban climate. Understanding the relationship between urban surface cover and ET can inform future climate-resilient urban design.

5.2 Study sites

This study examines observations from the capital and largest city of Germany, Berlin, which has a population of 3.7 million spread over 891 km^2 (Amt für Statistik Berlin-Brandenburg, 2019). Situated in the east of Germany, the climate is temperate oceanic (Cfb) (Kottek et al., 2006). The closest weather station from the German Weather Service (DWD, Berlin-Tempelhof) recorded a long-term (1991-2020) mean annual rainfall of 585 mm and mean air temperature of 10.2 °C (DWD, 2021). Here, we study the warm months (April until October) of the relatively dry year of 2019 with 492 mm of precipitation, in which an intense observation campaign was organized (Vulova et al., 2019). The warm months are studied as most ET occurs during this time.

Two sites in Berlin are studied here: a suburban one and one close to the city center. The first, suburban site is an urban research garden located in the southwest of the city at the Rothenburgstraße (ROTH, 52.457°N, 13.315°E, Figure 5.1a, (Vulova et al., 2021)). This site is an ICOS (Integrated Carbon Observation System) Associated Ecosystem Station (DE-BeR). The ROTH tower is located in a fairly green and residential neighborhood, with its surroundings within 1 km consisting of 47% impervious surface, 19% low vegetation, 34% high vegetation, and no open water (see Sec. 5.3.1). High vegetation is defined as exceeding 0.5 m.

The ROTH tower is mainly surrounded by local climate zones 6 ("Open low-rise"), 2 ("Compact midrise"), and 5 ("Open midrise"). Furthermore, the mean building height, vegetation height, and sky view factor (SVF) are 14 m, 7.6 m, and 0.91, respectively, within 1 km of the ROTH tower. The research garden within which the tower is situated consists mainly of grassland (16%), shrub (7%), trees (39%), and semi-permeable or sealed surfaces (16%) (Kuhlemann et al., 2021). The low vegetation is mostly meadow which is mowed 1-2 times per year, with nearly no bare soil in the source area. Most trees at the ROTH site are deciduous and broadleaved, with key species being Acer platanoides, A. pseudoplatanus, A. campestre, Fagus sylvatica, Populus nigra, Platanus acerifolia, Platanus x hybrida, Quercus robur, and Tilia spp. (Meier and Scherer, 2012; Gillefalk et al., 2021). However, evergreen needleleaf conifers are also present in the area (e.g. Taxus baccata, Pinus sylvestris, Abies procera), as well as a deciduous conifer species (Larix

decidua). For a more detailed description of the vegetation, we refer to Meier and Scherer (2012), Gillefalk et al. (2021), and Kuhlemann et al. (2021).

At ROTH, a 40-meter tower holds three EC systems (IRGASON, Campbell Scientific) at 2, 30, and 40 meters. For all EC systems in this study, the resolution is 30 minutes. The observations are quality controlled according to the literature and only high-quality data (flag 0) is used (Foken et al., 2004). Additionally, sap flow was observed at six trees with FLGS-TDP XM1000 sap velocity logger systems (Dynamax Inc, Houston, USA), and soil moisture content was measured in two locations below high vegetation at three depths: 10-15, 40-50, and 90-100 cm (CS650 reflectometers, Campbell Scientific) (Kuhlemann et al., 2020). Finally, the leaf area index was measured monthly over three transects through high vegetation (LAI-2200, LI-COR, Lincoln, USA) (Vulova et al., 2019). Along each transect, leaf area index measurements were conducted at 1-meter intervals to capture the canopy variability, while walking in the same direction each time for standardization. A tripod on a balcony served as a reference for the above-canopy light conditions measuring every 10 seconds.

The second site is close to the city center at the TU Berlin Campus Charlottenburg (TUCC, 52.512°N, 13.328°E, Figure 5.1b, (Vulova et al., 2019; Jin et al., 2021)). Its surroundings within 1 km are more impervious than at ROTH: 62% impervious surface, 8% low vegetation, 26% high vegetation, and 3% open water (see Sec. 5.3.1). The TUCC site is in a central built-up area mainly occupied by commercial and university buildings beside the nearby Tiergarten Park. It is dominated by local climate zones 2 ("Compact midrise") and 5 ("Open midrise"). Around the TUCC tower (1-km radius), the mean building height, vegetation height, and SVF are 17.4 m, 9.7 m, and 0.87, respectively. Low vegetation in the TUCC site mainly comprises turfgrasses, with minimal bare soil in the area. Most trees in the TUCC area are also broadleaved and deciduous with the main tree species being Tilia and Acer spp. Irrigation occasionally occurs at both sites (more frequently in ROTH than TUCC) but only during very dry conditions. On the roof of TU Berlin (building height 46 m), observations are made with a ceilometer (Lufft CHM 15k) and an EC system (IRGASON, Campbell Scientific). The EC system is attached to a 10-meter tower reaching 56 meters above ground level.

5.3 Methods

5.3.1 Surface cover classification

The surface cover needs to be classified to link the surface in the EC footprint to the neighborhood-scale ET observed with the EC system. Given the surface fraction in the footprint covered by each surface cover type, the ET can be reconstructed from the evaporation dynamics of the different surface cover types (bottom-up, Figure 5.2a) or attributed to the surface cover types by linear decomposition (top-down, Figure 5.2b).

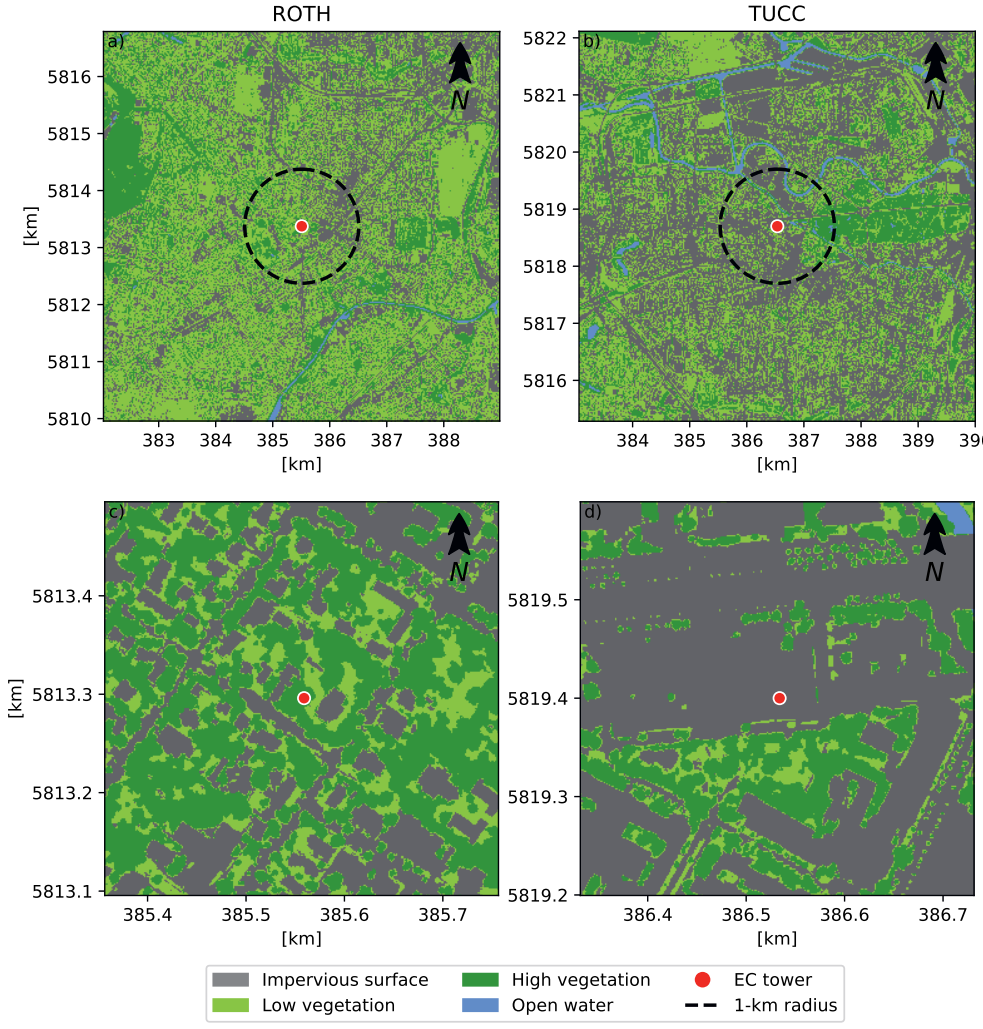


Figure 5.1: Map of Berlin indicating the location of the (a,c) ROTH and (b,d) TUCC study sites with their surroundings classified in the four surface cover types distinguished in this study with the 1-km radius (dashed black line) around the EC towers (red dots). Panel b and d are zoomed in around the EC towers to show the effect of resampling the 1-m resolution to visualize the full maps (a,b). The coordinate reference system is WGS 84 UTM/33N EPSG: 32633.

For this study, we classify the surface into four different surface cover types: impervious surface, low vegetation, high vegetation, and open water. For this purpose, we combine information from four geospatial datasets from Berlin Open Data and the Berlin Digital Environmental Atlas:

- *Building height*: raster dataset at a 1-meter spatial resolution of all buildings in Berlin (Senate Department for Urban Development, Building and Housing, 2012)
- *Vegetation height*: raster dataset at a 1-meter spatial resolution of all vegetation including trees, bushes, and grass in Berlin (Senate Department for Urban Development, Building and Housing, 2012)
- *Biotope types*: vector dataset describing the biotope type of all vegetation in Berlin according to the 7483 biotope types described by Zimmermann et al. (2015) (Senate Department for Urban Development, Building and Housing, 2013)
- *Streets*: vector dataset with all road segments in Berlin (Senate Department for the Environment, Mobility, Consumer and Climate Protection Berlin, 2014)

Around each EC tower, we classified the surface cover within a buffer of 0.025 degrees latitude and 0.05 degrees longitude in both directions, equivalent to 2.8 and 3.4 kilometers. In total, this gives an area of 5.6 by 6.8 kilometers. We selected this buffer, as for 90% of the footprints this area includes the entire footprint calculated in this study (see Sec. 5.3.2). For only 0.5% of the time, the buffer contains less than 80% of the footprint. All datasets are clipped to this area. Vector datasets are resampled to rasters at a 1-meter resolution to ensure compatibility with the raster datasets.

At the start of the classification, all described vegetated land biotopes are assigned to vegetation and all water biotopes to open water. The impervious surface is determined based on all areas in the street data and all areas that have an assigned building height. The vegetation is split into low and high vegetation depending on the height with a threshold of 0.5 meters following Kuhlemann et al. (2021). The exact threshold has minimal influence as only a negligible part of the vegetation has a height between 0.3 and 1.0 meters.

5.3.2 Footprint modeling

Footprints were calculated to determine the source area of the turbulent fluxes for all timesteps. We selected the flux footprint model from Kljun et al. (2015), which is frequently applied in urban environments e.g. Stagakis et al., 2019; Nicolini et al., 2022; Karl et al., 2023. This footprint model provides two-dimensional grids outlining the footprints and quantifying the relative weight of each pixel in the footprint. The model requires the measurement height, friction velocity, boundary-layer height, Obukhov length, horizontal wind direction, and mean and standard deviation of the horizontal wind speed. For all wind variables, EC observations are used, while the boundary-layer height is derived from

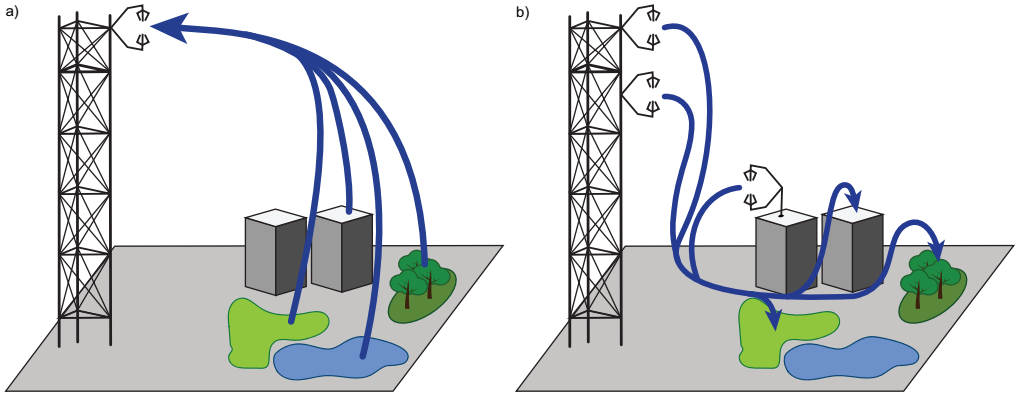


Figure 5.2: Conceptual drawing of the bottom-up (a) and top-down (b) approach. The arrows start at the data sources and end at the results of the approaches. Footprints determine the contribution for each surface cover type (not shown).

ceilometer observations at the TUCC site. The Obukhov length in m (L) is calculated according to:

$$L = -\frac{u_*^3 \bar{\theta}_v}{\kappa g (\overline{w'\theta'_v})_s} \tag{5.1}$$

where u_* the surface friction velocity in m s^{-1}, $\bar{\theta}_v$ the mean virtual potential temperature in K, κ the von Kármán constant of 0.4, g the gravitational acceleration of 9.81 m s^{-2}, and $(\overline{w'\theta'_v})_s$ the kinematic virtual potential temperature flux in K m s^{-1} at the observation height.

As the model results in contours per 10% and the 100%-contribution contour is infinite, the resulting footprint grids are limited to the 90%-contribution contour. Part of the footprint is not taken into account when the footprint extends beyond the classified area (Section 5.3.1). This last step had minimal influence, as the classified area is considerably larger than the typically considered representative area within a radius of either 0.5 or 1 kilometer (Lipson et al., 2022b). In the end, the contribution of each surface cover type to the footprint is calculated taking into account the relative weight of each pixel resulting from the footprint model.

5.3.3 Bottom-up

The bottom-up approach attributes ET to the different surface cover types by determining evaporation dynamics for each type (Figure 5.2a). Therefore, it is hypothesized that EC

observations can be accurately reconstructed by considering the ET dynamics of the surface cover types and their respective contributions to the footprint area (see Sec. 5.3.4, Eq. 5.5). For the impervious surfaces, open water, and high vegetation interception, evaporation dynamics are estimated based on conceptual models. For the low vegetation and the high vegetation transpiration, observations capture the dynamics. We assume the evaporation dynamics per surface cover type to be similar for ROTH and TUCC, as previous research found their forcing is comparable and can be used interchangeably to predict ET with the same accuracy (Duarte Rocha et al., 2022). Negative ET observations are omitted, as the conceptual models are not capable of predicting negative fluxes. This filter has a very limited impact on the results, as it excludes only 384 of the 17780 30-minute time intervals. The results are evaluated against EC-observed $\mathrm{ET}_{neighborhood}$ at two timescales, midday and daily, as these consider different aspects of ET. Midday is defined from 11:00 until 15:00 local (10:00-14:00 UTC) time with every half hour considered separately. During these hours, incoming shortwave radiation driving ET is highest. Considering multiple hours minimizes the sampling noise due to the stochastic nature of turbulence even at half-hourly timescales. The daily timescale is relevant for water resources management.

Impervious surface

Evaporation from impervious surfaces is modeled according to Wouters et al. (2015). Their parameterization includes two processes to mimic the water on an impervious surface: rainfall and evaporation. The impervious surface is characterized by the maximum water storage depth (d_m) in mm m^{-2} and the maximum wet/evaporative fraction (δ_m). These parameters were determined for Berlin based on 3D-LIDAR scans and found to be 1.03 mm m^{-2} and 13.53% (Haacke, 2022). The evaporative fraction decreases following a power law with an exponent of $-\frac{2}{3}$ depending on the water storage, which follows from the assumption that interception storage capacity linearly depends on the storage depth. Water gain from rainfall is reduced in efficiency when closer to the maximum water storage capacity described by:

$$d(t+\Delta t) = d_m \left(1 - \ln(1 - (1 - e^{(1-\frac{d(t)}{d_m})})e^{-\frac{r_0 \Delta t}{d_m}})\right) \tag{5.2}$$

where, d is the water storage depth in mm, t time in s, Δt length of the time step in s, and r_0 the rainfall intensity in mm s^{-1}. The formulation assumes constant rainfall during a time step. The evaporation is described by:

$$d(t+\Delta t) = \left(d(t)^{\frac{1}{3}} - \frac{\delta_m E_p \Delta t}{3 d_m^{\frac{2}{3}}}\right)^3 \tag{5.3}$$

where E_p is the potential evaporation. The E_p is calculated according to Penman (1956), further described in Eq. 5.4 and the open water section. Eq. 5.2 and 5.3 are calculated

consecutively for each timestep to get the new water height. The meteorological forcing has a resolution of 30 minutes, but the conceptual model is run numerically at a 30-second time step to ensure a numerically robust solution with linearly interpolated meteorological forcing.

High vegetation

The ET from high vegetation consists of transpiration and evaporation of water intercepted by the canopy. The transpiration is derived from observations of the soil moisture content and sap flow as described in Kuhlemann et al. (2021). Soil moisture content observations are used from both the "Trees" and "Shrubs" plots for the transpiration estimation from high vegetation. The soil moisture content reflects the evaporated water volume, but root water uptake does not correlate directly with transpiration apparent from the lag between the two. Therefore, we scale daily soil moisture loss with hourly sap flow observations. This method takes advantage of the correctly timed temporal variation in sap flow observations and the accurate water volume estimate of the soil moisture content observations. Soil moisture loss due to drainage is assumed to be negligible, as the deepest soil moisture observations at 95 cm depth showed no soil moisture loss indicating a drainage flux. Furthermore, soil moisture loss in the lowest layer of observations is not added to the evaporation.

The canopy interception and its evaporation are modeled with the Rutter model that allows for sub-daily resolution (Rutter et al., 1975; Valente et al., 1997). The model partitions rainfall between evaporation from the canopy and trunk, throughfall, and stem flow. Two storages are part of the model: the canopy and the trunk. Both storages evaporate at the potential rate calculated according to the Penman (1956) equation (Eq. 5.4) described in the open water section. Canopy storage capacity depends on the tree species ranging between 0.1 and 3 mm (e.g. Aston, 1979; Klaassen et al., 1998; Baptista et al., 2018; Ramírez et al., 2018), although in exceptional tropical canopies capacities up to 8 mm have been observed (Herwitz, 1985). We assume the canopy storage capacity is linearly related to the leaf area index with a storage of 0.2 mm per unit leaf area (Huang et al., 2017). Leaf area index observations at monthly intervals are interpolated with a univariate spline with four degrees of freedom. The modeled interception appears to be relatively insensitive to the other parameters set at 0.2 mm for the trunk water storage capacity, 0.015 for the fraction of exceedance of the canopy storage capacity partitioned to stem flow instead of throughfall, and 0.02 for the fraction of evaporation from the stem flow. All of these parameters concern the stem flow, which, on average, accounts for only 2% of the precipitation exceeding the canopy storage capacity (Rutter et al., 1975). The modeled interception evaporation is added to the transpiration to obtain the ET from the high vegetation.

An alternative for both vegetation types would be to model the ET_{patch} with the Penman-Monteith equation (Monteith, 1965). Grimmond and Oke (1991) have adapted this equation to urban environments and included the effect of water limitation. As a preliminary analysis, the adapted Penman-Monteith equation was used to represent vegetation in the bottom-up approach. This analysis showed that the adapted Penman-Monteith equation overestimates ET_{patch}. The estimate is twice as high as the precipitation in the same period.

Low vegetation

Low vegetation is directly represented by an EC system installed at 2 meters directly above the grass at ROTH. In their similar study, Peters et al. (2011) installed an EC system close to the surface of a golf court to estimate the ET from low vegetation. Within a forest, a comparable low EC set-up helped to differentiate the ET components in an environment with more obstacles (Paul-Limoges et al., 2020).

The quality-controlled ET is a direct observation of the low vegetation dynamics when the wind comes from between east (90°) and southwest (230°). Fluxes were considered when suitable for process-focused studies and general analysis (quality flag "0" and "1" according to Foken et al. (2004)).

Open water

Open water evaporation is estimated with a parameterization of the Penman (1956) equation (De Bruin, 1979):

$$E_p = 37 + 40\overline{u}_{2m}(e_{s,2m} - e_{40m}) \tag{5.4}$$

where $\overline{u}_{2m}$ is the mean wind speed at 2 meters (m s^{-1}), $e_{s,2m}$ the saturated vapor pressure at 2 meters (Pa), and e_{40m} the vapor pressure at 40 meters (Pa). Eq. 5.4 approximates the evaporation from an infinitely thin water layer ignoring heat storage in the water column. The lack of water temperature observations prevented the use of equations that capture the effect of the heat storage in water on ET (Jansen and Teuling, 2020). Open water is assumed to evaporate at the potential rate. In the case of a negative E_p, evaporation is set to 0.

5.3.4 Top-down

The top-down approach takes the neighborhood-scale EC observations and attributes the flux to the different surface cover types by solving a system of equations (Figure 5.2b). The system consists of three equations related to three EC systems (ROTH: 40 and 30 m; TUCC 56 m). Each equation describes how the ET_{patch} from the surface cover types is combined according to the footprint to yield the EC observation of one system. The evaporation for the three surface cover types results in three unknowns, as the evaporation per surface cover type is assumed similar for all EC systems. Open water is not considered

as it covers the least area and only three EC systems are suitable for the analysis. To filter out any timesteps affected by open water, we remove all timesteps with more than 5% open water in the footprint of any of the EC systems. The linear system can be solved, as it has an equal number of equations and unknowns.

$$f_{im,1}E_{im} + f_{lv,1}\mathrm{ET}_{lv} + f_{hv,1}\mathrm{ET}_{hv} + f_{ow,1}E_{ow} = \mathrm{ET}_{EC,1} \tag{5.5}$$

$$f_{im,2}E_{im} + f_{lv,2}\mathrm{ET}_{lv} + f_{hv,2}\mathrm{ET}_{hv} + f_{ow,2}E_{ow} = \mathrm{ET}_{EC,2} \tag{5.6}$$

$$f_{im,3}E_{im} + f_{lv,3}\mathrm{ET}_{lv} + f_{hv,3}\mathrm{ET}_{hv} + f_{ow,3}E_{ow} = \mathrm{ET}_{EC,3} \tag{5.7}$$

where f is the footprint-weighted fraction with the subscripts describing the impervious surface (im), low vegetation (lv), high vegetation (hv), and open water (ow). The numbers indicate the different EC systems. The evaporation from each surface can be determined given the footprint-weighted fractions derived from the footprints and the EC observations. We exclude solutions with estimated evaporation below 0 mm d^{-1} for one of the surface cover types, as these solutions likely have negative evaporation rates for one surface cover type that are balanced by positive evaporation rates for another type.

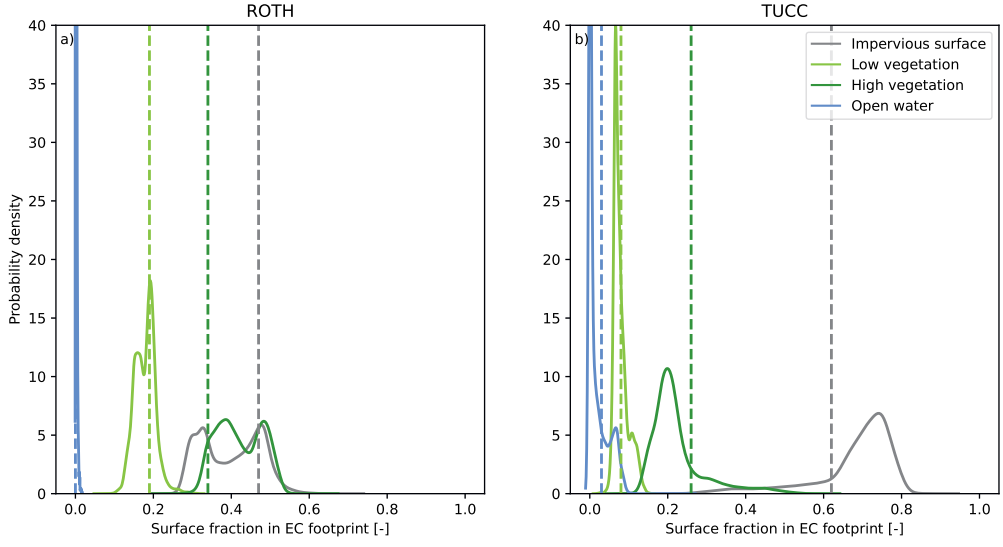


Figure 5.3: Probability density of the time-dependent footprint-weighted surface fractions in the EC footprint calculated according to Kljun et al. (2015) over the study period (April-October 2019). The dashed vertical lines indicate the average surface cover fraction within a 1-km radius of the EC (see Figure 5.1).

5.4 Results

5.4.1 Footprint variation

A high variation in footprint composition highlights the heterogeneity of the urban surface (Figure 5.3). The wide, non-normal distributions cause the footprint-weighted surface fractions in the footprint to differ substantially from the surface cover fractions within a 1-km radius of the EC system (vertical lines) at most times. The 1-km radius estimation and the footprint-weighted fraction are only similar for open water, as this covers a limited surface. For the impervious surface and high vegetation at ROTH, the bi-modality of the distribution demonstrates that a single value will not be able to capture the footprint-weighted surface fractions. Additionally, footprint-weighted surface cover fractions can vary within a wide range as seen at TUCC where the footprint-weighted impervious surface fraction varies from 0.2 up to 0.8. The high variation necessitates that the time-dependent footprint composition is considered to understand ET dynamics.

5.4.2 Surface cover composition impact on evapotranspiration

Combining the footprint variation from both sites with the EC-observed $ET_{neighborhood}$ reveals the influence of the surface cover composition on ET (Figure 5.4). We find lower ET values with more impervious surface and higher ET values with more vegetation (high and low). Open water shows a less clear relation between its surface fraction and ET, as the range in open water fraction in the footprint is minimal given the limited open water in the proximity of the EC systems (Fig. 5.1). Although the surface cover is relevant, the variation in the ET indicates meteorological conditions affect ET as well, illustrated by the ordering of the points by available energy quantified as the net radiation. While the surface cover composition in the footprint varies at one site, the two sites together reveal an evident influence of the surface on ET.

5.4.3 Evapotranspiration attribution to the surface cover

EC-observed $ET_{neighborhood}$ is approximated by $ET_{neighborhood}$ reconstructed by a weighted average of surface cover type evaporation dynamics (Figure 5.5 and Table 5.1). The negative mean bias error indicates an underestimation of total ET. The data gaps due to quality control of the 2-m EC system explain why the number of evaluated data points is lower than the duration of the study period. In one ROTH case, ET is highly overestimated when a rainfall event coincides with a high impervious fraction in the footprint and high potential evaporation (Figure 5.5a), for which the conceptual model for impervious surfaces is responsible.

Impervious surfaces contribute proportionally less to ET than their footprint-weighted surface fraction according to the bottom-up approach (Figure 5.6a). In contrast, high vegetation contributes significantly more. The relative ET contribution of low vegetation

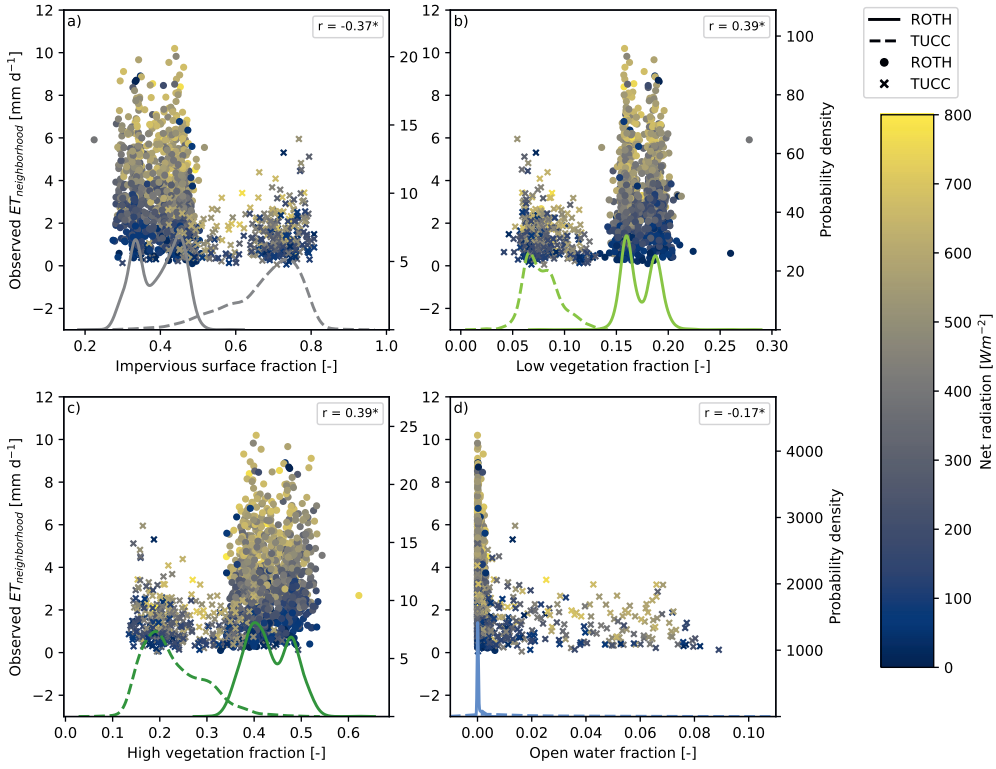


Figure 5.4: Relation of the EC-observed $ET_{neighborhood}$ and the footprint-weighted surface cover fraction for each surface cover type (a: impervious surface, b: low vegetation, c: high vegetation, d: open water) with the color indicating the net radiation. The probability density curves (right axis) describe the footprint compositions for both ROTH (solid) and TUCC (dashed). The correlation is given in the top right of each panel.

Table 5.1: Overview of the performance of the bottom-up approach compared with EC ET observations per 30 minutes as shown in Figure 5.5.

		Rothenburgstraße		**TU Berlin campus**	
		Midday	*Daily*	*Midday*	*Daily*
Figure 5.5	panel	a	b	c	d
Data points	[-]	609	100	152	30
Observed mean ET	[mm d^{-1}]	3.8	3.0	1.7	1.1
Modeled mean ET	[mm d^{-1}]	2.3	1.9	1.4	1.0
Mean bias error	[mm d^{-1}]	-1.5	-1.2	-0.3	-0.1
Mean absolute error	[mm d^{-1}]	1.8	1.2	0.8	0.6
Pearson's r	[-]	0.51	0.71	0.38	0.57

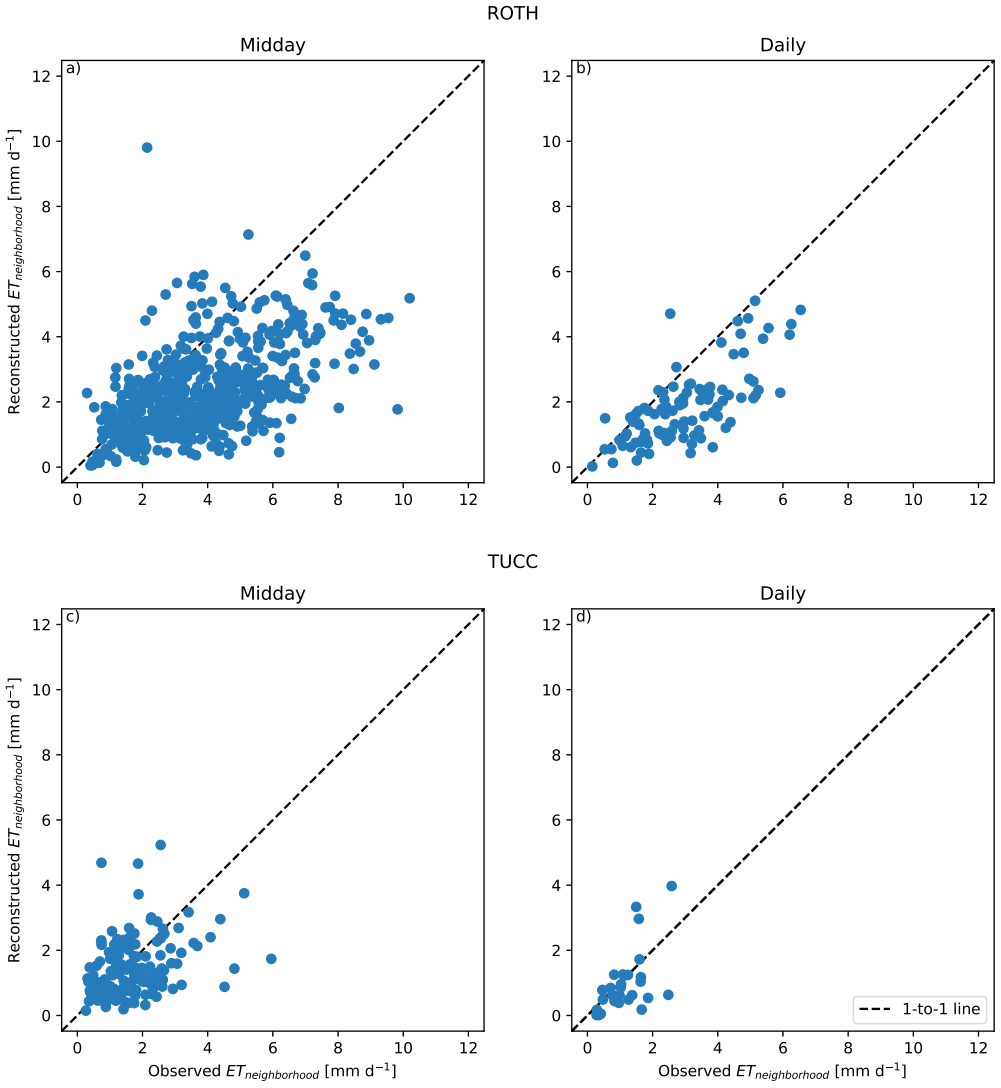


Figure 5.5: Comparison of the ET observed with EC against the ET reconstructed with small-scale observations and conceptual models at (a-b) Rothenburgstraße (40 m only) and (c-d) TU Berlin campus for (a,c) midday hours per half hour and (b,d) daily means. Midday hours are between 11:00 and 15:00 local time (10:00-14:00 UTC). Table 5.1 gives an overview of the statistics.

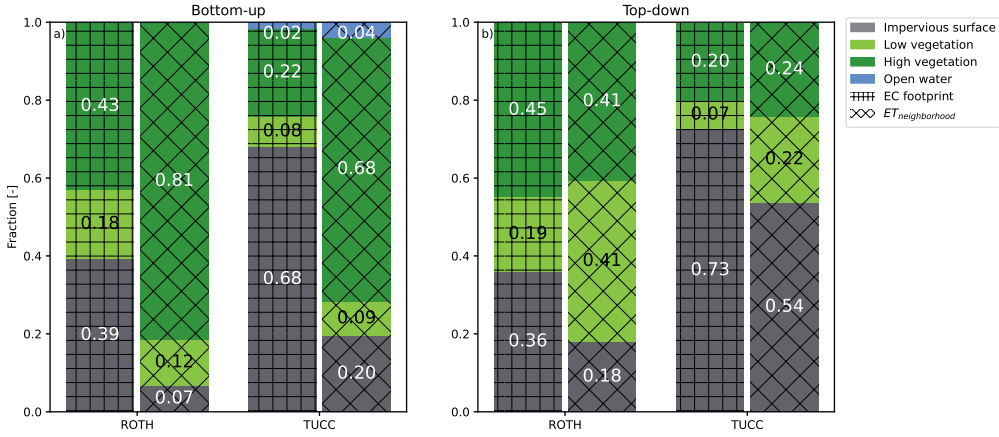


Figure 5.6: Relative contribution of the surface cover types to the EC footprint (vertical/horizontal hatch) and $ET_{neighborhood}$ (diagonal hatch, calculated as f ET_{patch}) for all available 30-minute intervals. Surface cover fractions differ between the two methods at the same site at different times and thus footprints are included in the analysis due to data availability.

varies depending on whether the remaining surface fraction in the footprint is dominantly high vegetation or impervious surface, as ET_{patch} from low vegetation is higher than from high vegetation than from impervious surface. Despite open water covering only a small part of the surface, the TUCC results indicate that the $ET_{neighborhood}$ fraction can exceed the footprint-weighted surface fraction. The relative contributions are constant throughout the months, although exact values vary mostly around 0.02 with exceptions up to 0.13. Throughout the study period, the footprint-weighted surface fraction has the same qualitative relation to ET contribution.

The top-down approach yields mostly similar relative contributions to the surface cover and ET as the bottom-up approach (Figure 5.6b). However, the $ET_{neighborhood}$ fractions are more similar to the surface fractions than the bottom-up approach indicates. Unfortunately, the open water surfaces could not be considered, as only three EC systems had observations for sufficient timesteps. Still, data gaps cause the top-down approach to yield results for 10652 timesteps. Subsequently, 9860 timesteps are excluded from the analysis as negative evaporation rates artificially enhanced the evaporation from the other surfaces. This artificial enhancement is an artifact of the linear system of equations (Eq. 5.5). Next to the $ET_{neighborhood}$ fractions, the footprint-weighted surface fractions differ slightly from the bottom-up approach because different timesteps are considered. Over the months, relative ET contributions are less constant than the bottom-up approach typically differing about 0.15 and maximally 0.28. Unlike the bottom-up approach, no direct comparison

with observations can be made, as the method gets the EC observations as input, and no observations are available at the patch scale.

5.4.4 Evaporation dynamics per surface cover type

The distinct evaporation dynamics of each surface cover type are visible when zooming in on one drydown (Figure 5.7). These dynamics can be derived from the bottom-up approach given its good performance and the high number of timesteps with attributed fluxes. The impervious surface has a unique pattern with a sharp peak after rainfall and no evaporation once the surface has dried. Meanwhile, the other three surface cover types all show a daily cycle. Low vegetation and open water show comparable changes over time without a response to the time since the last precipitation but following energy availability and transport efficiency. On the other hand, high vegetation limits ET within days after rainfall. These responses are seen in all other dyrdowns except for the last drydowns during the warm season (see Figure A.5.1). At this time, the soil moisture is more depleted triggering low vegetation to limit ET, while open water maintains the same response.

5.5 Discussion

5.5.1 Surface cover type contributions to evapotranspiration

Our study revealed that the four distinct surface cover types do not contribute to ET proportional to their footprint-weighted surface fraction. To disentangle these contributions, the ET was attributed to the surface cover types with both a bottom-up and top-down approach. Both approaches find similar ET contributions compared to the footprint-weighted surface fraction; impervious surfaces evaporate less than their footprint-weighted surface fraction, while high vegetation and open water evaporate more. The higher ET values for the (high) vegetation are likely explained by the high leaf area (Liu et al., 2016; Bian et al., 2019), which effectively increases the evaporating surface over the same ground surface. The low ET values for the impervious surface are likely linked to the low water availability (Jongen et al., 2022; Jansen et al., 2023).

For high vegetation, an isotope-evaluated model study found similar ratios between average surface fraction (∼30%) and ET contribution (∼80%) at ROTH (Gillefalk et al., 2022). From this ET, evaporation of interception accounts for 17% of the total precipitation over the study period from April to October. This is comparable to some studies finding values between 14-27% (Bryant et al., 2005; Xiao and McPherson, 2011), while others show higher interception evaporation between 45 and 77% (Asadian and Weiler, 2009; Anys and Weiler, 2024) or lower between 5-6% (Paul-Limoges et al., 2020). Although our interception evaporation is lower than most observed values, together with transpiration, it exceeds the precipitation during the study period. Soil moisture reserves supply the additional water.

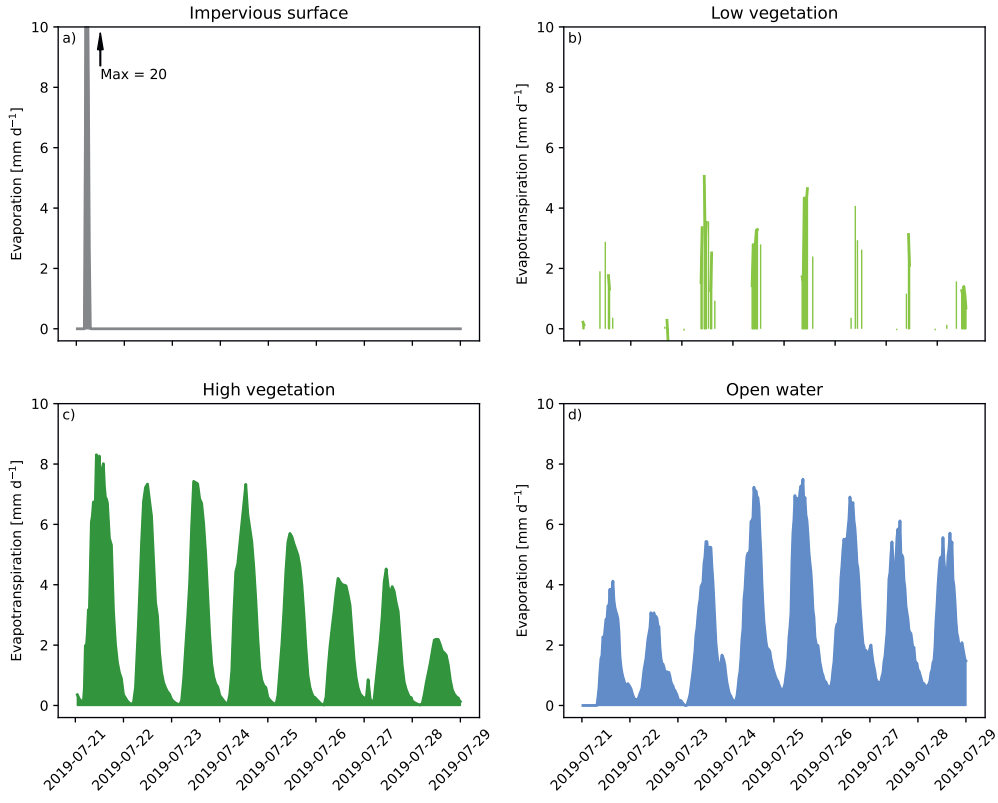


Figure 5.7: Illustration of ET dynamics during a drydown starting 30 minutes after rainfall ceased determined with the bottom-up approach for the four surface cover types (a: impervious surface, b: low vegetation, c: high vegetation, d: open water). This drydown occurred between 21-07-2019 and 29-07-2019. The gaps in the low vegetation time series are explained by the quality control of the 2-m EC measuring low vegetation.

The found ET contributions are largely in line with findings by Peters et al. (2011), who did a similar exercise for a more homogeneous neighborhood. Still, we challenge their assumption that the impervious surface did not contribute anything to ET, as we find 7% (18% top-down) of ET may come from impervious surfaces in a suburban setting (ROTH, 36/39% impervious in footprint). In the more impervious city center (TUCC, 68/73% impervious in the footprint), we find a contribution of 20% (54% top-down). Ramamurthy and Bou-Zeid (2014) and Chen et al. (2023) found ET from impervious surfaces contributed between 11 and 18%.

5.5.2 Evaporation dynamics per surface cover type

Apart from the different ET contributions compared to the surface fraction, evaporation evolves differently for each surface cover type after rainfall. In line with Ramamurthy and Bou-Zeid (2014), we find impervious surfaces evaporate all water quickly after rainfall. In contrast, open water sustains evaporation for a longer time. The open water evaporation shows a strong diurnal trend reaching zero during the night, which is incorrect according to the literature. The effect of heat storage could not be considered due to missing water temperature observations. Previous research shows the large heat capacity of water dampens the diurnal trend preventing it from going down to zero (Jansen et al., 2023). In this study, the diurnal trend results from the Penman equation (Jansen and Teuling, 2020), which was applied given the unavailability of water temperatures. High and low vegetation show different behavior from each other with the high vegetation having a higher initial ET. While high vegetation decreases peak ET within days after the last precipitation, low vegetation sustains high ET rates until soil moisture availability is limiting. This soil moisture limitation only occurred towards the end of the summer, even though our study year 2019 was relatively dry. The same responses were found in other studies (Teuling et al., 2010; Van Dijke et al., 2023). High vegetation has a stronger stomatal control that enables it to limit transpiration with sufficient available moisture, while low vegetation has lower stomatal control that causes it to limit transpiration less than high vegetation until it lacks water. In the urban setting, low vegetation is also frequently shaded possibly removing the necessity to limit transpiration.

5.5.3 Bottom-op versus top-down approach

Even though the ET contribution was similar for the bottom-up and the top-down approach, these methodologies also showed two interesting differences. Given these two differences, we think the bottom-up approach has the most potential to contribute to our understanding of the link between patch- and neighborhood-scale ET. The first difference is the need for observations from multiple EC systems simultaneously. Every EC system has gaps in the observations that do not overlap with the other EC systems' gaps. The gaps result from the quality control and are numerous because of the challenging urban environment (Feigenwinter et al., 2012; Oke et al., 2017). The top-down approach requires as many EC systems as surface cover types. In the present study, this was circumvented by excluding open water. Because a fourth EC system was unnecessary, the number of timesteps with a successful outcome rose from 44 to 792. This is comparable with the 751 for bottom-up, for which the availability is mainly limited by the low 2-m EC system with data for only 17% of the time. The maximum number is given by the study period of 244 days equal to 11,712 half hours, of which 2,196 are during the midday hours. While data availability is a challenge for both approaches, the top-down approach relies more heavily on EC observations leading to more overlapping data gaps.

The second difference is that the bottom-up approach is confined by patch-scale observations and models, while the top-down approach is not. The confined bottom-up approach provides insight into the ET contributions of the surface cover types but still has a mismatch with the observed ET. Also using a bottom-up approach, Salmond et al. (2012) reconstructed the neighborhood-scale sensible heat flux observed with an EC system with smaller-scale observations from two scintellometers. They found a mismatch of 25%, which can partly be explained by three reasons that also apply here. Firstly, even when EC systems are installed directly next to each other, the observations differ, up to 15% in the case of ET (Mauder et al., 2006; Mauder et al., 2013). These differences are partly due to large turbulent structures that are not resolved at (sub-)hourly timescales. This makes time-averaged EC observations not by definition representative of the spatial average over heterogeneous surfaces. As these structures may resolve at daily timescales, it may explain the better performance of the bottom-up approach at the daily timescale. Secondly, the footprints are calculated with an analytical model that does not account for surface heterogeneity and 3-dimensional surfaces (more in Section 5.5.4). Lastly, the patch-scale observations are not necessarily representative of the whole neighborhood scale. In our case, for example, sap flow was measured at six trees that cannot capture the diversity of the trees in the EC footprint. Another example is the low vegetation that experiences shading depending on the location within the urban canyon.

Still, the bottom-up approach yields errors comparable to urban land surface models from a decade ago and only slightly higher than more recent models (Grimmond et al., 2011; Lipson et al., 2024). These urban land surface models share our assumption that the neighborhood flux is the sum of ET_{patch} from the separate surface cover types. They do not incorporate dynamic footprint information in the forcing but use the direct surroundings or an average footprint (Lipson et al., 2024). As our approach reduces complexity and requires fewer inputs, we demonstrate the relevance of these footprints in scaling patch fluxes to the neighborhood level. The added complexity of urban land surface models allows for prediction and scenario studies, which our approach would not be suitable for. Still, the found agreement underscores the potential for utilizing surface-specific contributions to decipher ET dynamics.

In contrast, the top-down approach yields highly unlikely results as the linear system follows is not confined by patch-scale observations. The linear system counteracted high negative fluxes with high positive fluxes giving results as extreme as -2.0×10^{17} and 1.1×10^{6} mm d^{-1}. These effects were omitted from the analysis by excluding negative fluxes, which omitted the high fluxes as well due to the linear relations in the equations. Due to these direct links, the ET contributions contain the errors from the EC observations. However, these random errors will cancel out against each other, as we only look at aggregated results from the top-down results.

5.5.4 Footprint variability and modeling

Given these differences in evaporation behavior between surface cover types, the surface composition in the footprint influences the EC observations. This changing footprint has to be accounted for to understand ET dynamics, as the footprint contribution of a particular surface cover type may vary as much as 50%. Previously, the relevance of footprints for ET was illustrated by the improved performance when the footprint-weighed surface cover fraction was supplied to machine learning models in addition to meteorological observations (Vulova et al., 2021; Vulova et al., 2023). This is the first process-based study that reveals how ET dynamics differ from hour to hour per surface cover type. For other fluxes, such as CO_2, footprint modeling has also been shown to help understand flux dynamics (e.g. Velasco et al., 2009; Conte and Contini, 2019; Wu et al., 2022).

CO_2 sources including directly from humans have been identified and quantified by looking at the relation between the CO_2 flux and the surface cover composition equivalent to Figure 5.4. One example is Menzer and McFadden (2017) who study the relative importance of vegetation and anthropogenic fluxes of CO_2. Another example is Stagakis et al. (2019) who find that traffic is an important CO_2 source and human respiration accounts for 19% of the CO_2 flux. Human respiration and perspiration are unlikely to affect our results. In the center of Beijing, the water fluxes from these processes are so small they would account for only 3% of ET in Berlin (Liu et al., 2022). Given the lower population density of our sites, human respiration and perspiration are even lower. Thus, these small water fluxes from these sources are much smaller than ET and do not influence the results.

Footprint modeling is the key that connects the surface to the ET in this study. The key is however limited by the simplifications of the footprint model. Here, we applied the analytical model by Kljun et al. (2015), which generates perfectly symmetrical footprints. The model does not account for the complexity and heterogeneity of the urban morphology. More detailed footprint modeling would provide footprints depending on urban morphology, but this would also require more computational resources and thus limit the length of the period that can be studied.

5.5.5 Generalizability

Here, we studied ET in one city during the warm months of a single year, 2019, which was a relatively dry year in Berlin. While the climate and year-to-year variability may affect some aspects of the ET dynamics, others are likely to be more constant. The main aspect we expect to be relatively constant is the evolution of ET over a drydown. The impervious surface will evaporate with a short intense peak, open water will evaporate more constantly, and vegetation will respond to soil moisture. While open water contributes little to the surface cover, we included this surface cover type in our analysis. It cannot be assigned to any of the other surface cover types and its inclusion improves the transferability of our methodology. The general patterns may be the same, but the dynamics are altered

by site characteristics such as plant species, building materials, and water depth. The urban morphology is another relevant site characteristic that affects the incoming energy by providing more or less shading related to the height-width ratio. Still, we anticipate these effects to be smaller than the differences between the four surface cover types. Apart from site characteristics, weather conditions control how much each of the surface cover types contributes to the ET (Jansen et al., 2023). The weather conditions determine the water availability (number and length of drydowns), energy availability (radiation and temperature), and exchange efficiency (wind and vapor pressure deficit). These conditions will lead to changed ET dynamics dependent on the season, the climate, and the year-to-year variability.

The unique 2019 dataset from Berlin allowed us to reconstruct the ET signal from EC systems. Although relatively common observations are required for the conceptual models of the open water and impervious surfaces, the data needed to estimate the evaporation dynamics of both vegetated surfaces is more specialized. These observations included low-level EC observations, tree sap flow, and multiple, continuous soil moisture sensors. In most cities, this will not be available. Instead, the vegetated surfaces could be modeled with the Penman-Monteith equation (Monteith, 1965). However, our preliminary analysis showed a severe overestimation of ET from vegetation. Without further adaptation, $\mathrm{ET}_{neighborhood}$ cannot be reconstructed with the Penman-Monteith equation using less specialized observations.

5.6 Conclusion

This study explores the link of neighborhood-scale ET to the surface cover at two sites in Berlin to estimate the contribution of each surface cover type to ET. This link is made starting from the ET dynamic from the surface cover types reconstructing the neighborhood-scale flux (bottom-up) and from four neighborhood-scale fluxes partitioned over the surface cover types through a linear system of equations (top-down). The bottom-up approach demonstrates that patch-scale dynamics can reconstruct the neighborhood scale when the EC footprint is considered.

Impervious surfaces contribute substantially to ET after rainfall mounting up on long timescales and are thus not negligible as assumed in earlier studies. That is why they should also not be ignored in urban water management. In line with previous studies, we find most ET originates from vegetation with especially high vegetation evaporating disproportionately more than its footprint-weighted surface fraction. While both approaches support these conclusions, the bottom-up approach proved to be more successful than the top-down approach in linking the surface cover types at the patch scale to the observations at the neighborhood scale due to its lower EC data requirement.

We stress the importance of time-dependent EC footprints to understand ET dynamics. Based on these dynamics, urban land surface models and their evaluation could be improved by accounting for the changing footprint. With footprint information, parameters could be dependent on the situation in the current source area. In this way, the models would more directly represent what the EC system observes making for a more fair and better evaluation.

Understanding ET is crucial in urban water management, for example, to determine appropriate vegetation species and irrigation requirements. At the same time, ET plays a role in the energy balance and can contribute to the mitigation of heat stress. Therefore, the gained insights can support design decisions in city landscapes and urban water management to improve the living environment of urban inhabitants.

Acknowledgments

Special thanks to Judith Jongen-Boekee, Studio Scientist, for the design of the conceptual figure, her tireless support, and her eye for detail.

Chapter 6

Eddy-covariance in large-eddy simulations

This chapter is based on:

H. J. Jongen, A. J. Teuling, D. Li, E. Akinlabi, S. Gadde, and G.-J. Steeneveld (2024c). "Modeling urban evapotranspiration with large-eddy simulations: The spatial and temporal influence of eddy-covariance footprints". *In preparation*

Abstract

Vegetation is suggested to mitigate urban heat by shading and increasing evapotranspiration (ET). Understanding ET is crucial for effective urban water resources management and optimized cooling from vegetation. The most direct method to observe and study ET over larger areas is eddy-covariance (EC). EC systems measure the ET from a time-dependent source area called the footprint. The footprint depends on the meteorological conditions including wind conditions and atmospheric stability, and the sensor location and height. Although the time-dependent footprint can be modeled analytically, these analytical footprint descriptions neglect the heterogeneous urban flow field. The Large-Eddy Simulations (LES) technique explicitly resolves turbulent transport incorporating footprint dynamics in the modeled ET. Here we explore the temporal and spatial sensitivity of EC observations to the footprint at two sites in Berlin. For this purpose, we apply analytically derived footprints to analyze the footprints over a longer time and LES to capture the detailed footprints. The LES results show realistic ET rates while differences with the observations seem to be driven by the mesoscale forcing. Our findings suggest that within a day the footprint's time dependency causes as much variation in observed ET as within a year, even when the day has relatively constant meteorological conditions. Spatially, virtual EC systems in the LES are decorrelated within 280 m demonstrating how EC observations capture their specific location. Our study shows the urgency of considering footprints when analyzing instantaneous ET observations from EC systems and the ineffectiveness of representing an EC system as measuring its average surroundings.

6.1 Introduction

An ever-increasing number of urban inhabitants face heat stress due to a combination of continued urbanization (United Nations, 2018), climate change (Oleson et al., 2015), and water scarcity (Liu and Jensen, 2018). Vegetation is often suggested as an effective measure to combat heat stress (Wong et al., 2021), as it provides shading and consumes energy through evapotranspiration (ET) that would otherwise heat the urban surface and air (Qiu et al., 2013). The cooling potential of vegetation hinges on the water available for ET (Kraemer and Kabisch, 2022; Nimac et al., 2022). Thus, urban water resources must be managed to optimize the effectiveness of the vegetation and its impact on urban climate.

Understanding ET dynamics is key for water management. ET is a key flux in the urban water balance and equals 19-87% of precipitation according to observations from 20 cities around the world (Lipson et al., 2022a). ET can be nearly directly observed with eddy-covariance (EC) systems at the neighborhood level (Burba and Anderson, 2010). To observe ET, EC systems measure the vertical wind speed and the specific humidity at high frequency in the order of 10 Hz. The turbulent moisture transport, ET, is estimated from the covariance of these variables. Many studies have exploited EC observations to improve our understanding of urban ET dynamics (Ward et al., 2013; Liu et al., 2017; Jongen et al., 2022). However, EC systems measure moisture from a source area, i.e. the footprint, that continuously changes depending on atmospheric conditions such as wind speed, wind direction, and atmospheric stability (Mauder et al., 2021). The dynamic footprint hampers linking the observed ET to the surface in heterogeneous urban settings. Yet, other studies have shown that considering the footprint opens doors to new analyses (Peters et al., 2011; Vulova et al., 2021; Jongen et al., 2024b).

All these studies share a common approach: they estimate the EC footprint using an analytical algorithm. These algorithms parameterize the contribution of the source area to the total flux (Vesala et al., 2008b). The algorithms for footprint estimation have been continuously developed improving their computational efficiency and expanding the range of atmospheric conditions, in which they can be applied (Kljun et al., 2004; Kormann and Meixner, 2001; Kljun et al., 2015). The recent models allow footprints to be estimated for entire years, as the previously mentioned studies capitalize on.

Large-Eddy Simulations (LES) is a numerical modeling technique that resolves dynamically important large-scale turbulent eddies and only parameterizes small-scale, sub-grid scale eddies. Thus, LES captures turbulent transport and can mimic EC observations opening the door to study EC without observational constraints such as equipment costs, site selection, and access (Kanda et al., 2004). LES is one of the other methodologies allowing footprint estimation (Vesala et al., 2008b). Because LES resolves most turbulence, it captures the footprint in greater detail than analytical descriptions (Hellsten et al., 2015;

Auvinen et al., 2017). Auvinen et al. (2017) demonstrate that the symmetry of analytical footprint models is not valid in urban areas and that the near-field contribution is higher than predicted with analytical models. As a result, the surface composition of the footprint is substantially changed. Apart from these insights, they demonstrate the computational costs, as their method requires a steady-state simulation lasting 4.5 modeled hours for one footprint. While this makes explicitly modeling footprints for longer periods unfeasible, LES implicitly simulates these footprints by resolving the turbulent transport.

Comparing LES with observations requires EC observations, but few cities worldwide host an EC system. In Berlin, however, they have been installed at two sites (Jin et al., 2021; Vulova et al., 2019; Vulova et al., 2021). The first site is a residential neighborhood, where a 40-meter tower is equipped with EC systems at multiple heights. The other site is close to the city center where the EC system is placed on top of a building. These sites together allow us to study the effect of footprints in two contrasting neighborhoods as described in more detail below.

The objectives of this paper are 1) to evaluate the capacity of LES to simulate the turbulent transport of water vapor and 2) to assess the importance of the temporal and spatial footprint dynamics for EC observations. We see how much footprint variations over different periods affect EC observations for the temporal dynamics. For the spatial dynamics, we determine the distance at which EC observations become independent. Therefore, we start by analyzing the influence of the footprint on the observed evaporation using analytical footprints to assess how dynamics of the footprint over time. Subsequently, these analytical footprints inform the design of the LES domain. With LES, we model the turbulent transport of water vapor. In the introduction, we referred to this transport as ET, but from here onwards we will use the energy equivalent: the latent heat flux (Q_E). The simulated Q_E is compared against observations from three EC systems at two sites. Finally, a grid of virtual EC systems is extracted from the LES to reveal the effect of the sensor placement on the footprint.

6.2 Data and Methods

6.2.1 Study sites and case study

The case study presented here focuses on the turbulent transport of water vapor during one day at two sites in Berlin. Berlin is the German capital located in the eastern part of the country boasting 3.7 million people spread over 891 km^2 (Amt für Statistik Berlin-Brandenburg, 2019). The temperate oceanic climate (Cfb) results in a long-term (1991-2020) mean annual rainfall of 585 mm and mean air temperature of 10.2 °C (Kottek et al., 2006; DWD, 2021). Within this city, two sites host EC systems: one in the suburbs and one on the TU Berlin campus close to the city center.

The suburban research garden is located in Steglitz at the Rothenburgstraße (ROTH, 52.457°N, 13.315°E, (Vulova et al., 2021)). A residential neighborhood surrounds the site consisting within 1 km of 47% impervious surface, 19% vegetation below 0.5 m, 34% vegetation above 0.5 m, and no open water (Jongen et al., 2024b). This results form a mix of local climate zones 6 ("Open low-rise"), 2 ("Compact midrise"), and 5 ("Open midrise"). The mean building and vegetation heights are 14 and 7.6 m respectively with a Sky View Factor (SVF) of 0.91. The EC systems are mounted on a 40-m standalone tower at 30 and 40 m (IRGASON, Campbell Scientific), which are registered as an ICOS (Integrated Carbon Observation System) Associated Ecosystem Station (DE-BeR).

The second site, TU Berlin Campus Charlottenburg, is situated close to the city center (TUCC, 52.512°N, 13.328°E, (Vulova et al., 2019; Jin et al., 2021)). Commercial and university buildings dominate the neighborhood, which are interrupted by the Tiergarten Park and the Spree River. Compared to ROTH, the surroundings are more densely built with slightly higher buildings and vegetation (mean 17.4 m and 9.7 m) and lower SVF (0.87) best described by local climate zones 2 ("Compact midrise") and 5 ("Open midrise"). The EC system (IRGASON, Campbell Scientific) is mounted on a 10-m tower placed on the roof of the 46-m TU Berlin building. Within 1 km of the building, the surface consists of 62% impervious surface, 8% low vegetation, 26% high vegetation, and 3% open water (Jongen et al., 2024b).

At these sites, an intensive observation campaign was organized during 2019, a relatively dry year with 492 mm of precipitation (Vulova et al., 2019). From this year, we searched for a day with no precipitation or clouds. The modeled area had to cover the footprint of the EC systems. Smaller footprints would thus require a smaller model domain. Therefore, the wind speed had to be low, while the atmosphere had to be convective. Moreover, the wind direction had to be relatively constant in time, so the domain had to be extended in only one direction. The day best fulfilling the requirement of this study was 29 June 2019. This day was a relatively warm, cloud-free day. The simulation period was chosen to include the moments with the highest ET and lasted from 9:00 until 18:00 local time. During this period, observations at 40 m indicated the average temperature was 27 °C, the average relative humidity 31 %, and the average wind speed 1.7 m s^{-1}. At 2 m, the average temperature was 27 °C, the average relative humidity 30 %, and the average wind speed 0.6 m s^{-1}. The average wind direction was easterly and varied between 204 and 326°.

For the selected day, the footprints were estimated for every half-hour according to Kljun et al. (2015). The average footprint served as a basis to define the required domain size. We aimed to include at least 80% of the footprint for all timesteps.

Before diving into LES, we use analytically estimated footprints to illustrate how sensitive Q_E is to the changing footprint. Because of the computational efficiency, the footprints can be derived for a full year (2019). The surface composition within these footprints is

Table 6.1: Details of the nested model domain configuration.

Domain	Model	Domain size in direction X	Y	Z	Horizontal grid spacing	Vertical grid spacing
W1	WRF	1527 km	1495 km	20705 m	15 km	58-358 m, 61 layers
W2	WRF	307 km	305 km	20705 m	3 km	58-358 m, 61 layers
W3	WRF	104 km	104 km	20705 m	1 km	58-358 m, 61 layers
P1/3	PALM	6144 m	6144 m	3072 m	16 m	16 m, 1.05 stretch after 640 m (max 32 m)
P2/4	PALM	1152 m	1152 m	576 m	2 m	2 m

derived following Jongen et al. (2024b) distinguishing four surface cover types: impervious surfaces, high vegetation, low vegetation, and open water. While the footprint is dynamic in this analysis, the Q_E per surface cover type is kept constant at the value of 10:00 local time on 29 June 2019. This reveals the effect of the footprint on observed ET without showing the influence of ET dynamics. The same Q_E per surface cover type is multiplied by the footprint fraction of the surface cover types for every timestep to yield the fictive ET observed with EC. In the end, the distribution of the EC Q_E is determined for the footprints of 29 June 2019, of every 10:00 of the year, and of the entire year.

6.2.2 Model configuration

To simulate 29 June 2019, a model configuration was designed to resemble the real urban area and forcing (Figure 6.1 and Table 6.1). The first three nesting levels (W1-3) use the mesoscale Weather Research and Forecasting (WRF) model version 4.1.4 (Powers et al., 2017) to downscale the ECMWF operational reanalysis data (Muñoz-Sabater et al., 2021) to a resolution suitable for LES. The smallest WRF domain provides atmospheric initial conditions and forcing representative for the modeled day. Below the WRF domains, two LES nesting levels exist using PALM 23.04rc (P1-4) (Maronga et al., 2020). P1 (ROTH) and P3 (TUCC) are the parent domains to the nested domains P2 (ROTH) and P4 (TUCC). We will shortly introduce both modelsand discuss the details of the configuration.

To drive the PALM model simulations, the mesoscale WRF model version 4.1.4 (Powers et al., 2017) was employed to generate the meteorological boundary conditions. The ECMWF operational reanalysis drives the largest WRF domain (W1) with a grid spacing of 15 km. The reanalysis is at a 0.25-degree spatial resolution and is provided at the start and every 6 h at the model boundaries. Within the parent domain, we have 2 nested domains (W2/W3) with a resolution of 3 and 1 km respectively. The domains contain 100x100, 101x101, and 103x103 grid cells. All domains use 61 vertical model levels, and a time step of 90 sec in the outer domain. The model runs for 30 hours starting on 28 June 2019 at 8:00 local time. The physical parameterizations that have been employed are

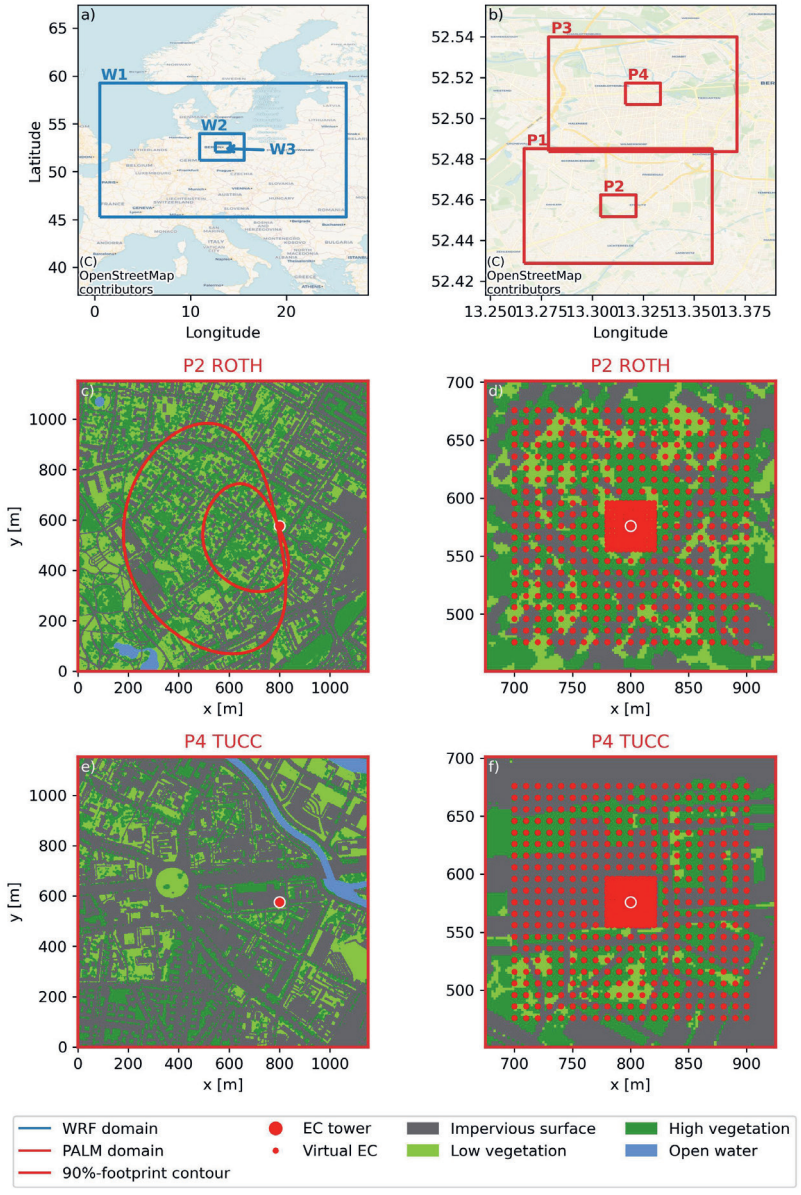


Figure 6.1: WRF and PALM model configuration with the a) the WRF domains, b) the PALM domains, c) the ROTH PALM child domain with example footprints, d) the ROTH PALM child domain zoomed around the EC tower, e) the TUCC PALM child domain, and f) the TUCC PALM child domain zoomed around the EC tower showing the 40-m and 200-m grids of virtual ECs. Half of the virtual EC towers in the smaller grid are omitted for clarity.

WSM3 for the microphysics, the Grell-Freitas ensemble scheme for the deep convection (only outer domain), RRTMG radiation scheme for long- and shortwave radiation, and the MYNN boundary-layer and surface-layers schemes, and the NOAH land surface scheme. Output was stored every 30 minutes.

For the LES domains, we use PALM 23.04rc including PALM-4U components making it suitable for the urban environment (Maronga et al., 2015; Maronga et al., 2020). PALM is designed to run on highly parallel computers like the cluster used for this study, Snellius, the Dutch national supercomputer. The PALM model is based on the non-hydrostatic, incompressible Boussinesq equations with prognostic equations for the three wind components, the potential temperature, total water content, the subgrid-scale kinetic energy, and a scalar. Subgrid scale turbulence is parameterized with a 1.5-order Deardorff (1980) turbulence closure model. Small subgrid structures (e.g. chimneys, balconies, edges) are represented with a roughness length of 0.05 m (Letzel et al., 2012). The Cartesian topography in PALM enables explicitly resolving solid obstacles such as buildings and orography. PALM includes an array of modules. In our configuration, the following schemes are active the radiation model, land surface model, urban surface model, plant canopy model, and offline and self-nesting. The offline nesting allows WRF to force PALM, while the self-nesting supports multiple PALM domains to be linked.

From the smallest WRF domain, W3, one PALM dynamic driver is created using the WRF4PALM Python packag for each site (Lin et al., 2021). The dynamic driver is only connected to the PALM parent domain, while the PALM child domain is one-way connected to the parent domain. The PALM parent domains (P1/3) have a horizontal resolution of 16 m and the child domains of 2 m. For each site, two PALM static drivers are created according to the methodology described by Heldens et al. (2020), as the parent and child PALM domain have their own static drivers. The PALM child domains are 1152x1152x576 m with a 2-m resolution and are extended towards the wind origin rather than centered around the EC system. Therefore, the EC systems are located at 800,576 m capturing more of the footprint and minimizing effects from the domain edge. The child domains are centered within the parent domains that are 6614x6614x3072 m with a 16-m resolution. In the z-direction, the grid cells are stretched by a factor of 1.05 after 640 m to a maximum of 32 m. The PALM simulations last from 8:00 until 18:00 local time with the first hour omitted from the analysis to account for the spin-up effects.

Within both PALM child domains, wind speed and specific humidity are output at high frequency in two grids around the EC systems, a larger 200-m and a smaller 40-m grid (Figure 6.1d/f). The first grid has a virtual EC every twenty meters in a square of two hundred meters totaling 441 virtual towers. The second grid has a tower every two meters in a square of forty meters also totaling 441 virtual towers. The output frequency coincides with the variable timestep in PALM, which is typically 0.20 s for ROTH and 0.19 s for

TUCC. Q_E can be calculated from this high-frequency LES output according to:

$$Q_E = \rho_a L_v \overline{w'q'} \tag{6.1}$$

where ρ_a is the air density, L_v the latent heat of vaporization ($2.5 \cdot 10^6$ J kg^{-1}). The primes indicate the deviation from the temporal mean calculated per 30 minutes in this study. This follows the same principle as EC observations. Like the virtual ECs from the LES, the EC observations are processed at 30-min resolution and are quality controlled according to the literature (Foken et al., 2004). Our results are less sensitive to the turbulence realization due to the spatial extent of the grid and the number of virtual ECs allowing us to focus on the footprints.

The grids are used to study the effect of the sensor placement on the Q_E observations. The inter-grid differences are leveraged to show how the measured Q_E varies with distance. Each virtual EC will have a slightly different footprint depending on its location. From a certain distance separating two virtual ECs, the variation in Q_E will no longer increase and the observations become independent. A variogram describes the variation with the semivariance against the distance between the observations. The range is the distance after which the semivariance no longer increases and two observations contain no information about each other. To find the range we fit a spherical variogram (Oliver, 2009) through the semivariances found in the virtual EC grids. The range is estimated for every half hour. When EC sensors are placed further apart than this distance, their observations would be independent.

6.3 Results

Only the variation in the EC footprint causes observed ET to vary 50 W m^{-2} within the day of the case study (Figure 6.2a). Surprisingly, the ET variation related to the footprint does not increase when all footprints from 10:00 local time from a year or even a full year are considered (Figure 6.2b/c). At both sites, the sensitivity follows a similar distribution regardless of how many footprints are considered. Therefore, our case study may capture most of the footprint influence on ET despite lasting only a single day

Before analyzing the PALM simulation, we compare the observations and model results from WRF and PALM at the ROTH 40-m EC system (Figure 6.3). Differences between PALM and the observations appear to be driven by the WRF forcing (Figure 6.3a/b). The air temperature in the PALM and WRF results is nearly identical and is around 3 °C lower than the observations during the analysis period (Figure 6.3a). The 40-m wind speed is higher than the observations by on average 1.3 m s^{-1} in WRF and 1.6 m s^{-1} in PALM (Figure 6.3b). The average wind speed in PALM is higher than in WRF as the wind speed does not drop in the afternoon in PALM. By staying high PALM follows a similar pattern as the observations with a positive bias. The tendency of the wind direction is the same in

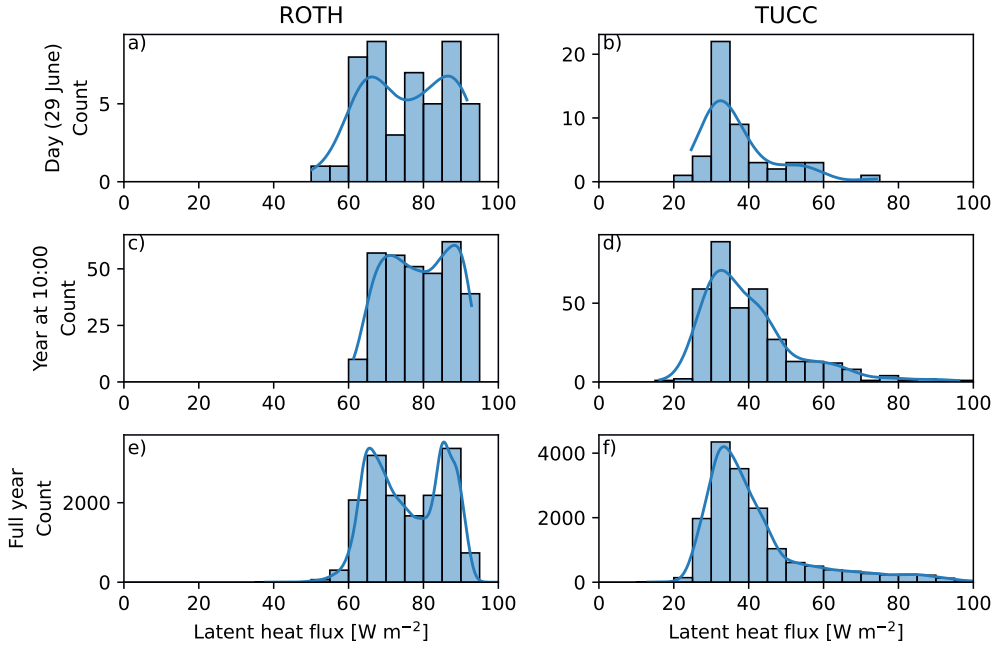


Figure 6.2: Sensitivity of ET to the changing footprint estimated by taking the varying surface cover fractions in the analytically estimated footprint (Kljun et al., 2015) while keeping the ET flux per surface cover constant. For this calculation, the ET fluxes per surface cover are used derived by (Jongen et al., 2024b) for 10:00 local time on 29 June 2019. The considered footprints include all footprints during a) 29 June 2019, b) 10:00 local time every day of 2019, and c) every half-hour in 2019.

the models as in the observations with different fluctuations around the tendency (Figure 6.3c). The results at TUCC are comparable to those at ROTH (Figure A.6.1).

The comparison of Q_E reveals the effect of the coarser resolution of W3 compared to P2 and P4 when the validation at ROTH (Figure 6.3d) is compared with TUCC (Figure A.6.1d). On the one hand, the WRF results show an almost ideal daily cycle at ROTH and a constant zero at TUCC, as WRF represents the EC footprint with one grid cell with one surface cover. On the other hand, Q_E is much more dynamic in the PALM results, as the changing footprint is captured by resolving the turbulent motions with grid cells of 2 m.

To assess the modeled turbulence characteristics, we concentrate on a 30-minute period and compare the high-frequency observations and model results (Figure 6.4). The instantaneous values should not be compared, as the model output is only one realization of the chaotic character of turbulence. The variation in vertical wind speed is similar in the observations ($\sigma_w = 0.67$ m s^{-1}) and PALM results (($\sigma_w = 0.99$ m s^{-1}), Figure 6.4a). The high-frequency

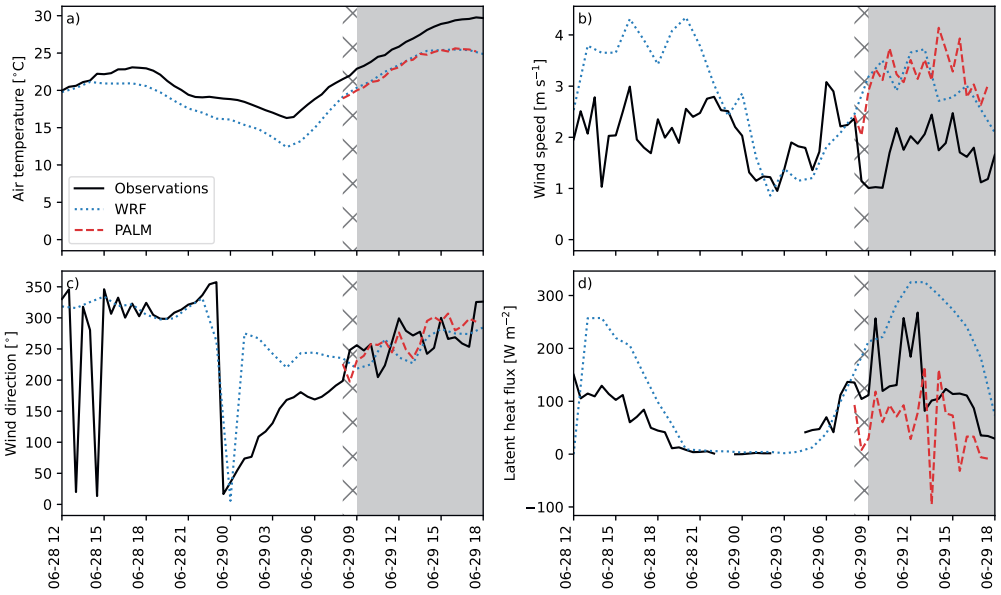


Figure 6.3: Comparison of the observations (solid black line) with the WRF (W3, dotted blue line) and PALM (P2, dashed red line) model results at the location of the 40-m ROTH EC system for a) air temperature, b) wind speed, c) wind direction, and d) latent heat flux. The shading indicates the analysis period, while the hatch indicates the PALM spinup.

signals depicted in Figures 6.4a and b look very similar, as the mean vertical wind speed is 0.21 m s^{-1} for the observations and 0.04 m s^{-1} for PALM. More fluctuations are visible in the observations given the higher temporal resolution of the observations (10 Hz) than of the PALM output (on average 5 Hz).

The specific humidity in PALM is 3.7 g kg^{-1} higher than in the observations (Figure 6.4c). While the observations show a large negative deviation after 10:10 (Figure 6.4d), the variation over the entire 30 minutes appears comparable (observations: $\sigma = 0.14$ g kg^{-1}, PALM: $\sigma = 0.19$ g kg^{-1}). Looking at the product of the deviations, we see lower and higher moisture exchange alternating in both observations and PALM resulting in a Q_E of 64 W m^{-2} in the observations and 60 W m^{-2} in PALM (Figure 6.4e).

The Q_E dynamics per surface cover look alike at both sites but differ from the dynamics derived from observations and conceptual models by Jongen et al. (2024b). PALM has higher Q_E for low vegetation than the derived values and lower Q_E for high vegetation. The PALM results are averaged for the entire modeled domain, while the observations are performed and conceptual models are given parameter values at the research garden at ROTH. The low vegetation at ROTH is often shaded, which may explain the lower Q_E.

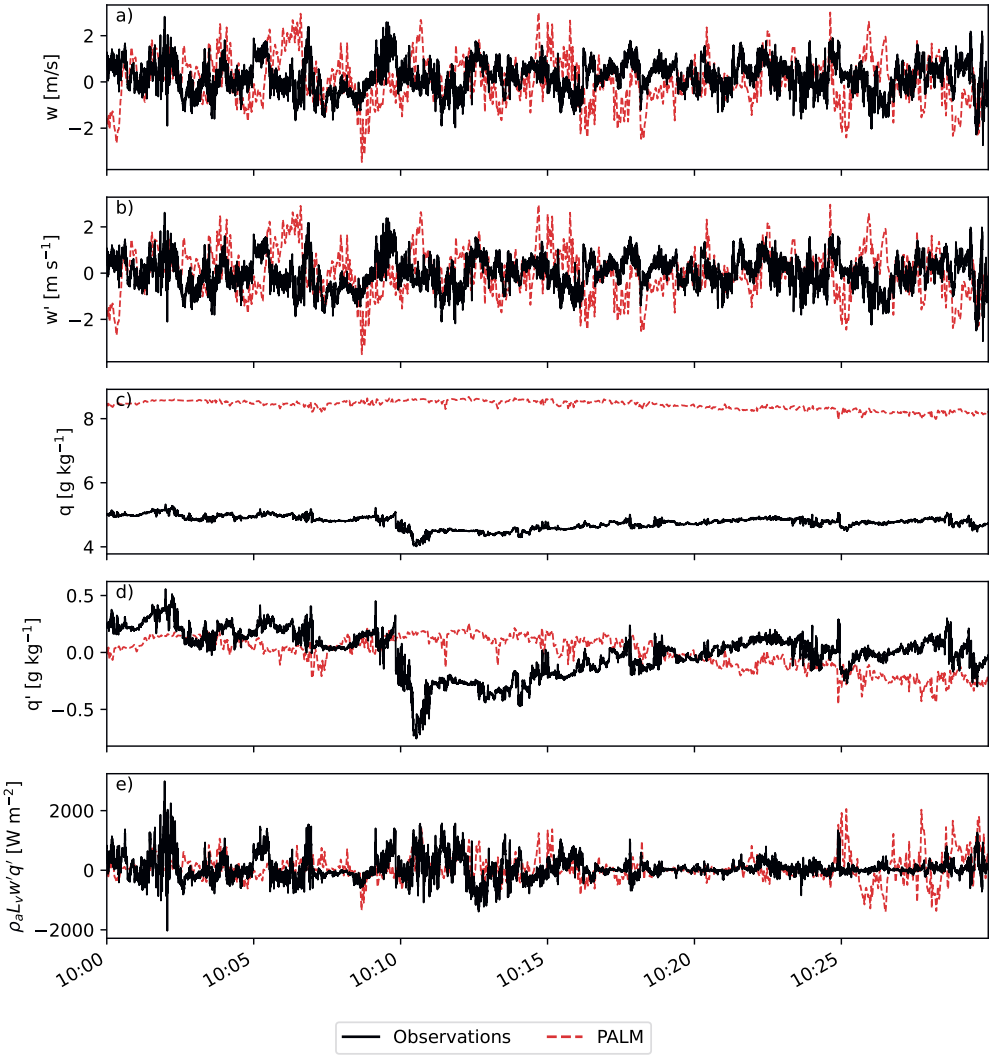


Figure 6.4: Comparison of the observations (solid black line) with the PALM model results (P2, dashed red line) at a high temporal resolution for a) vertical wind speed, b) vertical wind speed deviations from the mean vertical wind speed, c) specific humidity, d) specific humidity deviations from the mean specific humidity, and e) latent heat flux at the 40-m EC at ROTH. The observations are at 10 Hz and the PALM output is on average at 5 Hz. The high-frequency output from the EC system is not quality-controlled.

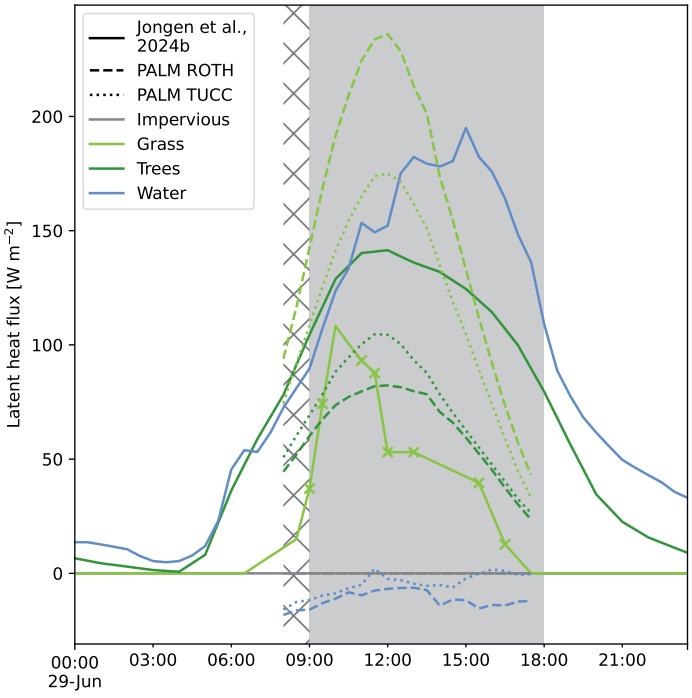


Figure 6.5: Time series of the surface latent heat flux per surface cover in the ROTH and TUCC PALM simulations (P2/4, dashed and dotted lines) with estimates derived from observations and conceptual models (solid lines, from Jongen et al. (2024b)). The PALM values are averaged over the entire simulated domain. Details on the observations are given in Jongen et al. (2024b). The crosses indicate high-quality EC observations (flag 0, Foken et al., 2004).

Both model results from both PALM domains and the observations indicate no Q_E from impervious surfaces, as the last precipitation was 9 days before the simulations. Open water shows very different dynamics with negative fluxes in the PALM simulations. The negative Q_E is likely the consequence of moisture condensation from warm air over colder open water (by default 283 K in PALM). Even though this differs substantially from the estimates from Jongen et al. (2024b), it hardly affects the results due to the minimal presence of open water at both sites especially in the footprint given the easterly wind direction.

Overall, Q_E is lower at the more impervious TUCC site than at the suburban ROTH site (Figure 6.6a). The observations do not differ substantially between the 40-m and 30-m EC systems at ROTH and the model results are also comparable. The modeled values deviate from the observations, which is at least partly due to the difference in Q_E values at the surface shown in Figure 6.5. The surface cover is modeled the same in the whole PALM domain suggesting the footprint causes the virtual ECs to show a large variation in

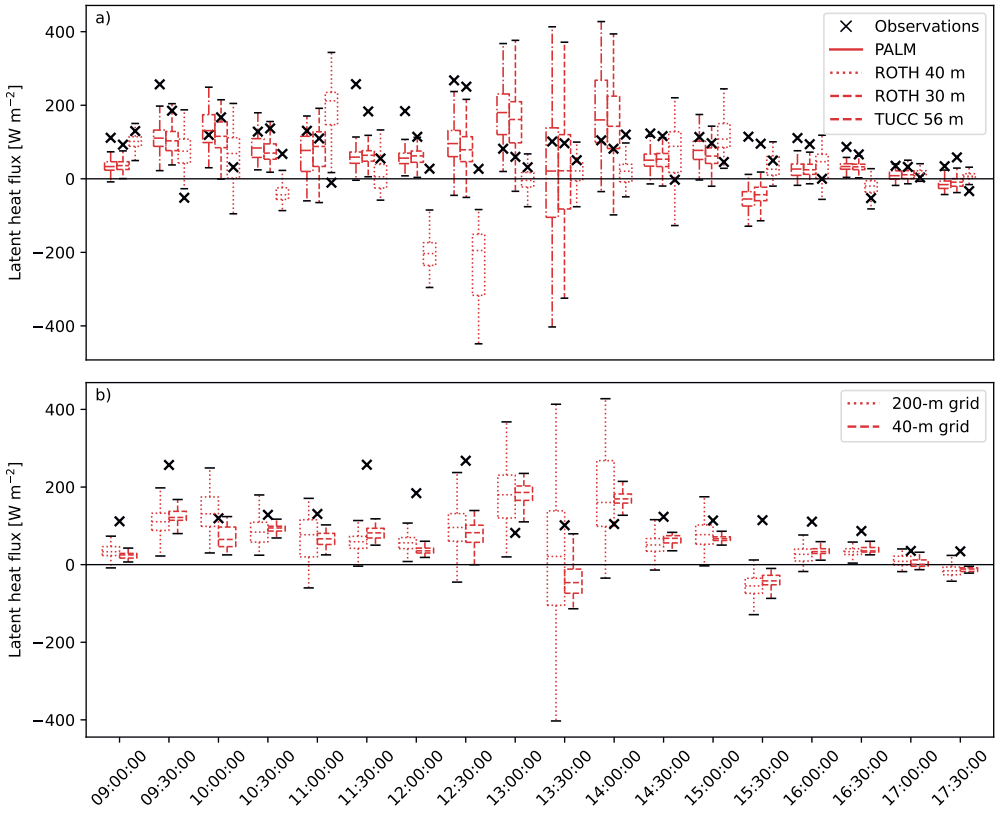


Figure 6.6: Time series of Q_E from the EC observations (crosses) and the virtual EC grids in PALM (boxplots) for a) the 200-m grids at all three EC systems (ROTH: 40 m and 30 m; TUCC: 56 m) and b) for the 200-m and 40-m grid at the 40-m ROTH EC during 29 June 2019. The boxplots indicate the distribution of Q_E within the virtual EC grids.

Q_E. Yet, PALM produces mostly realistic Q_E values. Two moments stand out as PALM produces unlikely Q_E values. At 12:00 and 12:30 local time, the modeled Q_E is unusually low at TUCC. At 13:30, the Q_E distribution is wider than at the other times at both the 40-m and 30-m levels at ROTH. A humidity budget analysis could not explain these phenomena.

Virtual ECs simulate a narrower distribution of Q_E values when situated closer together (Figure 6.6b). While both grids are centered around the EC system, the narrower distribution from the 40-m grid is not necessarily closer to the observations than the 200-m grid. From here onwards, the virtual ECs are compared with each other rather than with observations to focus on the effect of the footprint instead of the differences in surface Q_E.

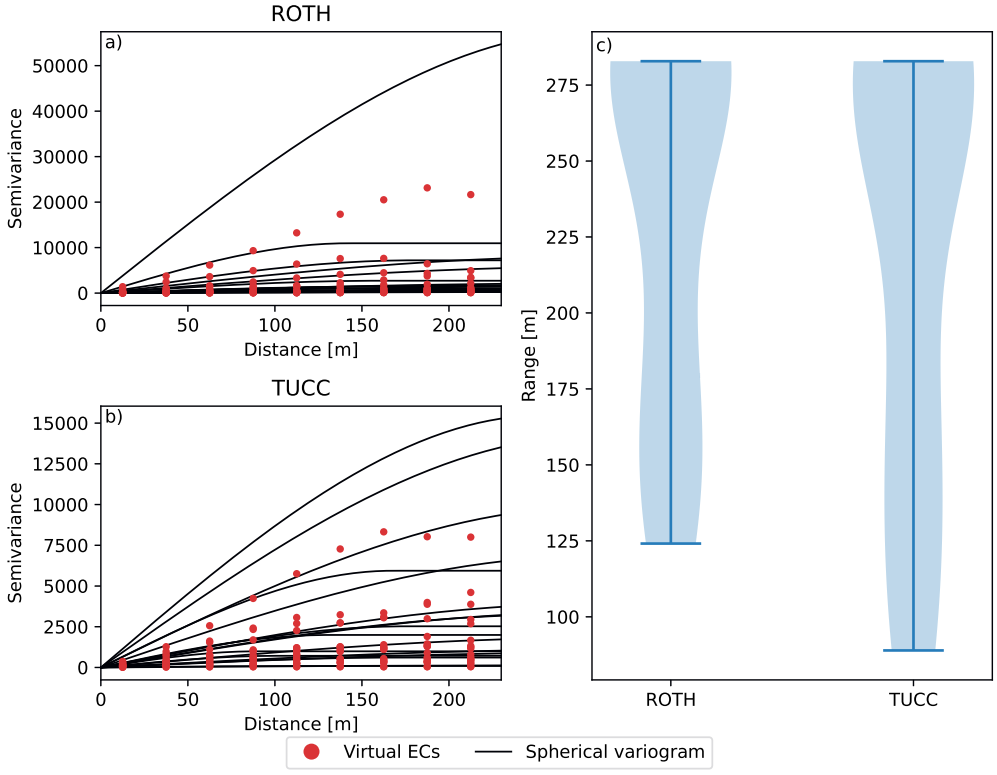


Figure 6.7: Spherical variograms of the Q_E from the 200-m virtual EC grids derived for every 30 minutes at a) ROTH and b) TUCC. The black lines are the spherical variograms fitter per 30-min period. The range of the spherical variograms is the distance beyond which the EC observations become independent. Panel c shows the range distribution at both sites where the width of the plot depicts the number of half hours with that range.

The virtual EC grid shows the variation in simulated Q_E increases until virtual ECs are between 125 and 283 m apart (Figure 6.7). The variograms vary over time and are likely dependent on the dominant eddy size that is being sampled. The range coincides with the maximum distance in the grid (283 m, diagonal in 200-m grid) for half the time (10 and 8 of the 18 timesteps for ROTH and TUCC respectively). The remainder of the time the range lays within the grid. Virtual ECs further apart contain no additional information about each other compared to the mean. The footprint of the virtual EC is different enough to become an independent observation.

6.4 Discussion

Here, we have modeled footprints explicitly with an analytical model and implicitly with LES 1) to assess the importance of the temporal and spatial footprint dynamics and 2) to assess the capacity of LES to simulate the turbulent transport of water vapor. The analytically modeled footprints show that observed Q_E is affected by the footprint variation found in a day as in a year. Therefore, we recommend that footprints always be considered when interpreting EC observations. With the implicit footprints, the LES is capable of producing realistic Q_E values but fails to closely mimic the observations. The mismatch starts at the surface boundary conditions and in the mesoscale model and propagates through the nested domains. By comparing virtual ECs in a grid, the range over which two EC systems would become independent is determined yielding values below 283 m half of the time. Below, we discuss the PALM results in more detail.

The PALM results show LES is able to simulate realistic Q_E time series, but it is unable to closely mimic observations. To a large extent, the lateral boundary conditions are responsible for the mismatch between PALM and observations, as the PALM simulations follow WRF closely. Moreover, the land surface model may not reflect the actual surface Q_E introducing errors regardless of the footprint. Idealized simulations could address part of these issues, but only when they are built heterogeneously and not with the typical repeated arrays. The repeated arrays would not allow for capturing the influence of the footprint, as the surface is homogenous and in that setting the EC method is insensitive to the footprint.

While the temporal footprint dynamics are relevant for interpreting and analyzing existing EC systems, the spatial variation in the footprints calls for rethinking sensor placements. Footprint descriptions are irrelevant for data interpretation when they merely use the average footprint climatology or even the surroundings within 1 km. Our results show that shifting an EC system 283 m creates independent observations for half the time, while the average surroundings would not change as much. Additionally, the similarity between the 40-m and 30-m EC systems at ROTH suggests standalone towers may not require to be as high as is currently assumed. LES may be suitable to test the required height and ideal location of an EC tower.

Our study shows that Q_E observations highly depend on the EC footprint. However, the quantification of the dependence is site- and even event-specific. The consistency of the temporal sensitivity between the sites seems to suggest that footprints in urban areas are as important over a day as over a year even during days with relatively constant conditions. The site-specificness lies mainly in spatial sensitivity and stems from the configuration of the urban fabric. This is apparent in the comparison between the ROTH and TUCC sites, where TUCC has lower ranges in the variogram than ROTH. Their sensitivities to the

temporal and spatial variation in the footprint stress the necessity of interpreting the Q_E observations considering the footprint.

In our LES results, the footprint is implicitly modeled by simulating the turbulent transport between the surface and the virtual EC sensors. Yet, this footprint is not explicitly present in the output. For this reason, we could not test whether the model domains captured most of the footprint, as predicted based on the analytical footprints. Moreover, the meteorological conditions differed between the observations used for the analytically estimated footprint and the PALM model results determining the implicit footprint.

The implicitness of the footprints prevents a comparison of analytical and LES-derived footprints. An explicit LES-derived footprint analysis could provide insight into the temporal dynamics. In the future, the method developed by Auvinen et al. (2017) may be further developed to allow this. To evade the high computational demand of this method, new algorithms would have to be developed circumventing the need for forward Lagrangian modeling. These methods exist but currently can only work on small grids, as all complete flow fields need to be saved. Yet, they show backward and forward Langragrian yields the same results (Cai and Leclerc, 2007). Efforts to reconstruct neighborhood scale fluxes like by Salmond et al. (2012) may also benefit from LES-derived footprints minimizing the uncertainty in the footprint.

6.5 Conclusion

This study highlights the critical role of considering footprint variability when analyzing eddy-covariance observations of the latent heat flux (Q_E) in urban environments. While analytical footprint models are computationally efficient, they do not capture the complexity of urban flow fields. For a case study of two sites in Berlin, we show that Large-Eddy Simulations (LES) can capture these complex urban footprints for a case study, but struggle to capture exact Q_E values from real scenarios and currently do not explicitly output the footprints. The results demonstrate that Q_E observations are sensitive to the dynamic footprints that change in response to the meteorological conditions. Surprisingly, Q_E observations are affected as much by the footprint variation in a day as in a year. Moreover, LES model results show that EC flux observations EC observations become decorrelated within 283 meters indicating how specific the observations are to the sensor location rather than the neighborhood. These findings emphasize the need for a careful approach when interpreting EC data in heterogeneous urban areas. Incorporating the time-dependent footprint in EC analyses may lead to a better understanding of Q_E. When Q_E is better understood, water requirements of neighborhoods can be estimated better including the effects of planned changes. Ultimately these estimates facilitate more efficient urban water resources management.

Acknowledgments

We would like to thank Matthias Suehring for sharing the static driver files and Fred Meier for providing the observations from the Urban Climate Observatory Berlin.

Chapter 7

Synthesis

7.1 Introduction

The preceding chapters aim to unravel the influence of the water balance on the urban energy balance and indirectly the urban climate. Now is the time to draw up the balance sheet. This thesis' aim is founded on the idea that the energy and water balances are irrefutably linked through a shared flux. While it is the same flux, it can be called Q_E, ET, or evaporation. The physical link between the balances is undeniable, but its effects prove to be disguised and difficult to indicate. To lift the veil, seven research questions have been formulated around three topics: urban water storage, the water balance in ULSMs, and the link between surface cover and ET.

In the preceding chapters, these topics are addressed by combining observations and models on scales between the patch and neighborhood levels (Figure 7.1). First, Chapter 2 assesses the water storage simultaneously quantifying the limited water storage capacity and the recession rate of ET after rainfall for 14 cities. In Chapter 3, the water balance is evaluated in 19 ULSMs from the Urban-PLUMBER project at 20 urban sites without observations. This evaluation demonstrates the potential for ULSM improvement, especially regarding water closure and surface runoff. Building on this evaluation, Chapter 4 focuses on the surface runoff in the same ULSMs showing that all three runoff mechanisms should be included: infiltration, saturation, and interception runoff. Switching from multiple cities to one, the following two chapters study the link between the surface and ET in Berlin. Where Chapter 5 unlocks the link between patch- and neighborhood-scale in both directions using EC footprints as the key, Chapter 6 explores the opportunities to refine our knowledge of the temporal and spatial variability in these footprints with large-eddy simulations.

This final chapter aims to bring these chapters together beginning with a discussion of their main findings, in which the research questions introduced in the first chapter are answered. This discussion is followed by a reflection on three broader themes overarching the preceding chapters. The first theme returning in all chapters is the *challenge of studying the urban environment*. I describe the factors contributing to this challenge and illustrate these with a case study. The second theme is *urban climate modeling* including but also beyond land surface modeling. Here, I reflect on the current state of urban modeling and how this could develop in the future. The third theme is the *connection between water and thermal comfort*. In all preceding chapters, this connection has been made through cooling by ET, but now the stage is given to other pathways that link water and thermal comfort. After these themes, I present how these findings and urban climate research, in general, could be employed towards application and what is needed to make fundamental knowledge applicable.

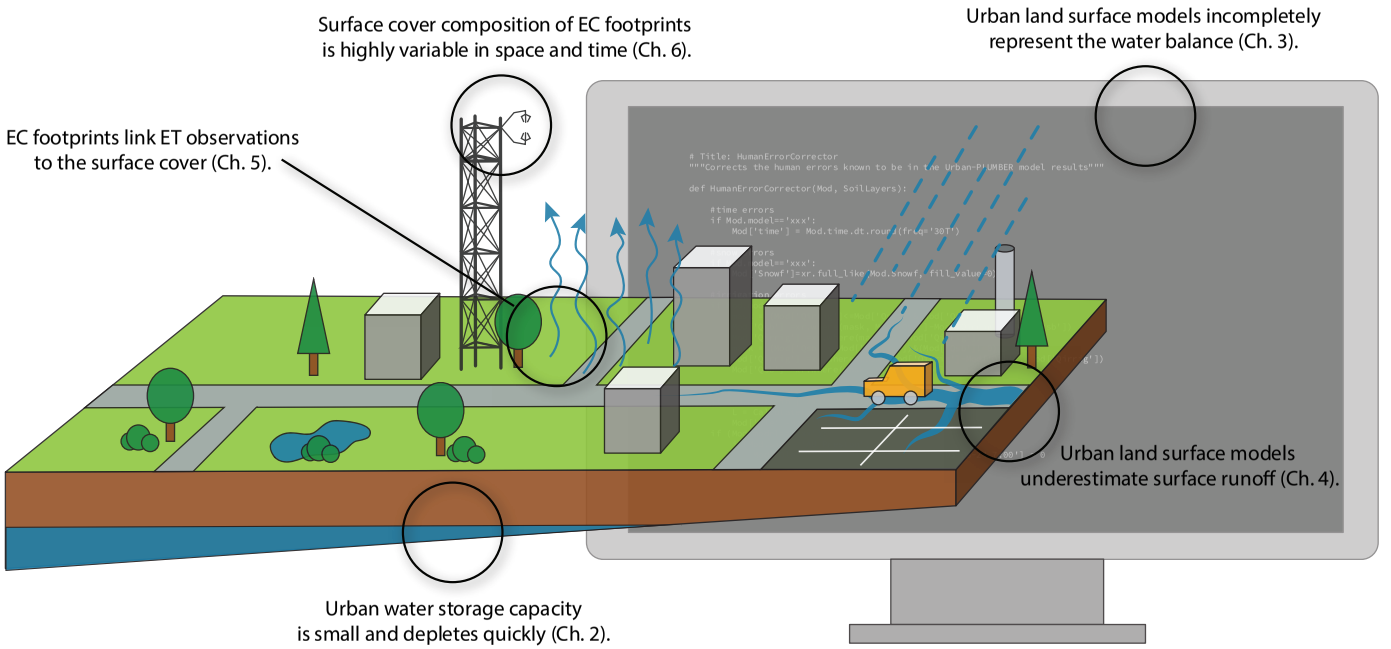


Figure 7.1: Graphical abstract summarizing the main findings of this thesis.

7.2 Main findings

7.2.1 Water limitation

The first main research question addresses the limited knowledge about urban water storage. The urban water balance shows precipitation or irrigation can instantaneously support ET. At times without precipitation or irrigation, water storage enables ET thereby ensuring ET's cooling function in dry periods. Despite the importance of water storage to endure these dry periods, its capacity had not been quantified in urban areas. The first logical question is thus: *What is the urban water storage capacity?* Only the capacity does not reveal how long ET can be sustained leading to the question *How quickly does the storage empty through evaporation?*

Chapter 2 answers these questions by analyzing the ET recession during dry periods shortly after rainfall at 14 urban sites worldwide. The water storage capacities available for ET were found to be between 1.3 and 28.4 mm being mostly lower than 13.4 mm. Seasonally, this water storage capacity changes with higher values during summer plausibly due to higher vegetation activity. The water storages empty quickly reaching 37% of the initial ET between 1.8 to 20.1 days, but in most cases within 10.4 days.

All 14 urban sites had less water storage capacity and shorter timescales of ET recession than natural areas. Yet, values vary substantially between sites with medians from 2.8 to 13.4 mm. No relation is found between the water storage capacity and the site characteristics related to both the urban form and the meteorological conditions. Chapter 2 describes two reasons for the missing relations. First, seasonal dependency may disguise the relations with site characteristics. Second, even though the gathered urban flux dataset is unique in the number of sites and observed years, urban areas vary so widely that 14 sites are insufficient to unravel the role of separate site characteristics. One additional reason arises from Chapter 5, but I will build some suspense and reveal this in Section 7.2.3 where the story revolves around the connection between the surface and the fluxes.

The low water storage capacity and quick ET decline showcase urban areas as water-limited environments. While intercity differences remain unexplained in Chapter 2, the results in Chapter 5 imply surface cover is relevant. The ET dynamics and related water storage are studied by attributing ET to four common surface cover types in urban areas. After rainfall, ET declines at a different pace for each surface cover type. A rapid decline is seen for impervious surfaces implying a low water storage capacity, which can explain why urban areas have lower water storage capacities than natural areas. Vegetation shows slower recession, where faster recession is seen for high than low vegetation. The rapid ET decline from impervious surfaces demonstrates water limitation is related to urbanization, as impervious surfaces are a signature trait of urban areas.

7.2.2 Water balance in urban land surface models

Urban areas expand and numerical weather and climate models reach finer grid spacing. Together these developments increase the need to include the interaction between the urban surface and the overlying atmosphere in modeling. ULSMs fulfill this need. In the meantime, their evaluation has traditionally focused on the energy balance, which contains the main target variables of these models. The directly linked water balance has not been evaluated, leaving the question: *How well does the water balance perform in current ULSMs?* Here, I will focus on ULSMs used standalone, i.e. offline, as in the previous chapters. The implications for online runs will be discussed in Section 7.3.2 from a broader (urban) climate modeling perspective.

Chapter 5 reveals the water balance has serious shortcomings in nearly all ULSMs. The water balance was found unclosed in over half of the examined model runs. The non-closure can be explained in multiple models by internal consistency between the modeled water storage and the water fluxes. Other models disregard snow, have numerically unstable thin soil layers, miss a water availability feedback between runoff and ET, or include infinite soil moisture sources or sinks. Apart from closure, models struggle to capture the magnitude and timing of ET. This is in line with previous studies, as ET is part of the energy balance and has been shown to be the most difficult energy flux in previous evaluations. These evaluations did not consider runoff, but this thesis shows ULSMs capture the magnitude and timing of runoff less well than those of ET.

While no ULSM passes all checks of a good water balance, some include a more complete and accurate water balance than others. As ULSMs are built with mainly the energy balance in mind, this provides the ideal opportunity to answer the question: *What are the effects of the water balance representation on the model performance for the surface energy balance?* Surprisingly, a better water balance could not be shown to improve the performance in the energy balance and even the shared latent heat flux. However, the absence of evidence is not evidence of absence. The start of this thesis introduces the three drivers of ET, of which water is only one. Models also (mis)represent the other two drivers, energy and transport. For example, airflow parameterizations struggle to capture the heterogeneity in building geometry (Li et al., 2020; Lu et al., 2024). When a poor airflow representation yields flawed estimations for the ET driver transport, a good water balance will not save the model's performance. This illustrates the challenge of isolating the effect of one model component.

While model evaluation may teach us how well models perform, a logical next step is to aim to increase this performance. Hence, I posed the next question in the introduction: *How can the water balance representation be improved?* While current ULSMs show clear deficiencies in the water balance, the identified non-closure problems can be fixed. For runoff, the evaluation yields no direct potential pathways for model improvements. Therefore, Chapter 4 zooms further into the runoff parameterizations to expose the causes

of the poor runoff performance. Three runoff generation processes are crucial to accurately simulate urban runoff: infiltration, saturation, and interception excess. While the first two are well known, the last is specific to urban areas and describes the excess of the small storage of impervious surfaces. These surfaces cannot saturate and allow very minimal infiltration. Their small storage capacity plays a role in the low water storage capacities found in Chapter 2 and is indicated by the quick ET recession in Chapter 5. Current ULSMs frequently miss one or more runoff generation processes leading to an underestimation of surface runoff. The inclusion of these processes in new model versions will likely lead to improved model results.

The water balance evaluation exposed shortcomings in ULSMs described above that may explain some results from Chapter 6. In this chapter, a land surface model provides the surface ET as boundary conditions to the LES. The modeled ET deviates from the ET derived from small-scale observations and conceptual models in Chapter 5. The land surface model in the LES is based on HTESSEL, a predecessor of CHTESSEL (Balsamo et al., 2009). The latter is part of the water balance model evaluation discussed above and thus the shortcomings found in the water balance evaluation may affect its performance. This land surface model is now developed independently from the original ULSM as part of the LES. Therefore, awareness of the water balance issues in ULSMs would be beneficial for the wider modeling community.

7.2.3 Surface cover, evapotranspiration, and eddy-covariance footprints

Understanding the effect of the surface on ET is critical for urban water management. In urban settings, the neighborhood ET is a composite flux from the heterogeneous surface. Where ULSMs simulate the surface-atmosphere interaction with a constant surface composition, the surface influencing the observations is dynamic for EC observations. Accounting for the changing surface compositions in the footprint is essential to interpret the ET observations and link them to the surface, which leads to the question: *How can the heterogeneous surface be linked to neighborhood ET?*

In Chapter 5, this link is made in two directions. One starts at the surface using observations and conceptual models for each surface cover type to reconstruct the neighborhood ET (bottom-up). The other begins with the observed neighborhood ET using a linear system of equations to decompose these observations into the contributions from different surface cover types (top-down). Both methods find the same relative differences between the surface cover fraction in the footprint and ET contribution. ET contribution is relatively low for impervious surfaces and relatively high for vegetation and open water compared to its footprint fraction. The patch-scale observations and models from the bottom-up method results also unveil the temporal dynamics of ET separated by surface cover type, where the contrasting drydowns show the role of water. Earlier I discussed the implications of these results for the water storage capacity.

For both methods, analytically modeled EC footprints are the key to connecting the patch to the neighborhood scale by providing the surface cover composition for each moment. The analytical footprint model reveals the surface cover composition is far from constant. A striking example is the impervious surface covering between 25 and 85% of the footprint at one of the Berlin sites. While the values are only valid for this specific spatial configuration, the site is not an extreme example of urban heterogeneity. Chapter 6 reiterates the temporal volatility of ET observations by showing they are as sensitive to footprint changes within a day as those within a year.

The high temporal variability of the footprint composition results from the spatial surface heterogeneity. This temporal variation is found with an EC system installed in one place. As the variation relates to the surface configuration, this gives rise to the last question raised in the introduction: *To what extent are EC observations sensitive to sensor placement in an urban setting?* The LES in Chapter 6 mimic EC observations successfully. Therefore, the LES can be used to substitute observations and a dense virtual EC network is installed that would be impossible in the real world. This network demonstrates that EC systems placed 280 m apart provide independent observations for half the time at the studied Berlin sites. An EC system can thus not reliably represent a certain neighborhood as summarized in local climate zones.

By now, I have not answered the question left unanswered earlier. Why could I not explain the variation in water storage capacities across the sites in Chapter 2? Two options were given before, but one last option remains. This option may have been clear after reading Chapters 5 and 6, but there is no escape after this section. As the changing footprint is not considered in the surface descriptions in Chapter 2, the footprint dynamics may explain why the storage differences cannot be explained. Unfortunately, estimating footprint-weighted surface descriptions is not always possible. Even the relatively simple analytical footprint models require meteorological variables that are not readily available everywhere like the boundary-layer and roughness height. Moreover, detailed surface cover classification maps are needed to translate the footprint into surface cover contributions.

All of the above supports the statement that considering the EC footprints is crucial for improving our ET understanding. The high temporal and spatial variability shows that the surface represented by EC observations is specific to the exact moment and location. Their variability makes it important to consider the footprints, especially combined with the surface cover types' contrasting ET dynamics.

7.3 Perspectives and future directions

7.3.1 The urban challenge

While the same physical laws apply to urban areas as to any other area, concepts and models cannot always be directly transferred. Hence, urban versions of land surface models exist to capture the unique urban features. These features set urban areas apart and force us to reconsider our understanding of the surface-atmosphere interaction. These features are the heterogeneity within the city and the presence of human activity (Figure 7.2). On top of that, the diversity among cities hampers knowledge transfer between cities. Together they create the urban challenge.

Heterogeneity

Urban heterogeneity sparks an interwoven network of interactions. All these interactions create variations and complicate performing representative observations. Consequently, urban measurements cannot possibly fulfill WMO standards (WMO, 2023), and no official observation stations are situated in cities. Related to this is the bias of (flux) observation availability towards natural systems (Jansen et al., 2023). As stressed previously in this thesis, footprints are essential to get the most out of the rare EC flux observations of the heterogeneous urban mosaic. Their dynamics and impact on the observations provide insight into the influence of urban heterogeneity.

While footprint dynamics reveal the effect of heterogeneity, their changes prevent the full water balance from being observed for the same area. Essentially, the footprint shifts the boundaries of the observed "catchment". For what catchment ET is observed, thus changes through time. At the same time, runoff has a different catchment altogether. This is not determined by the EC sensor placement and the meteorological conditions but rather by the topography and human interventions. These interventions may be canals, water pumps, or flow obstructions. Further below, the human contribution to the urban challenge will be discussed later more completely. In addition to these changing and non-matching catchments, the drinking water supply ignores natural water divides as water is imported from other catchments to fulfill the water demand (Deines et al., 2016). In conclusion, observing the urban water balance is an elusive undertaking in real urban areas.

Scale models may allow a more comprehensive observation set of the water balance. They have been applied extensively to study atmospheric aerodynamics and thermodynamics (Kanda, 2009; Lirola et al., 2017; Zhao et al., 2023). A recent study shows the tree effect on thermal comfort with significant heat stress reduction during the day (Demotes-Mainard et al., 2022). While the trees were irrigated, the water fluxes were not observed. However, the study shows the potential of scale models to observe the urban water balance.

Figure 7.2: A bird's eye view of two cities illustrates the urban challenge depicting the heterogeneity within and differences between urban areas. Moreover, every bit of the surface has been changed by humans. The upper panel shows New York City, the United States of America, and the lower Berat, Albania. Pictures taken by the author.

As the failure to adhere to WMO standards suggests, the perfect representative site does not exist in the city. Yet, the consequence does not have to be that no observations are performed. Rather than pursuing this fairy tale site, good metadata helps to interpret and benefit from data from available sites. Metadata protocols have been established to improve the metadata quality (Muller et al., 2013) and should be adopted by existing and new urban observation networks.

While standardized metadata helps to interpret the existing observations, it does not overcome the observations' limited representativity. Observations from more sites are required to capture the heterogeneity within cities. Logistics often limit the availability of potential sites. Mobile, remotely sensed, and crowd-sourced observations are three ways to circumvent the logistical limitations. Mobile observations are installed on a vehicle and may cover entire neighborhoods or cities. Different vehicles may be used such as cars (Belušić et al., 2014) or (cargo) bikes (Heusinkveld et al., 2010). Remotely sensed observations can cover larger areas using platforms such as drones (Feng et al., 2022) or satellites (Logan et al., 2020) for the sensors.

Crowd-sourcing can harvest observations from a variety of data sources. The data source most closely resembling standard meteorological observations is personal weather stations already installed by citizens. Personal weather stations can provide meteorological variables like air temperature (Meier et al., 2017), precipitation (De Vos et al., 2017), and wind (Droste et al., 2020). Other opportunistic data sources combine mobile and crowd-sourced observations and can help increase the observation density. Examples are the cellphones (Song et al., 2021) and cars (Marquès et al., 2022). Crowd-sourced observations are typically concentrated in cities because of the presence of people.

Human activity

While people enable crowd-sourcing, the same people also affect the urban climate. Two clear examples are the fluxes introduced in the energy and water balance. In the energy balance, its name reveals the culprit of the anthropogenic heat flux. The flux has many sources all releasing heat while using energy. These sources are related to among others building heating and transportation. Heating sources are more prevalent in winter. Transport sources are non-stationary and peak during rush hours. In short, the anthropogenic heat flux is spatially and temporally variable. Direct observations are challenging, but the flux can be estimated top-down from energy use at large scales (neighborhood and larger), bottom-up from estimating each source, and as the energy balance residual (Sailor et al., 2015). The first cannot capture spatial details and the last requires long time series to validate the assumption that heat storage can be neglected. The multitude of and diversity in sources contribute to the urban challenge.

In the water balance, humans introduce the anthropogenic moisture flux. While the flux is commonly called irrigation, moisture is also released during combustion and metabolism.

Except for irrigation, the sources coincide with those of the anthropogenic heat flux with the same temporal and spatial patterns. Yet, the moisture flux expressed in W m^{-2} is smaller than its sensible heat counterpart (Narumi et al., 2009; Demirel, 2012). It has also been studied less extensively. As such, the moisture flux is not part of global inventories of the anthropogenic heat flux (Jin et al., 2019). While this part of the anthropogenic moisture flux may be small, irrigation may contribute as much as precipitation depending on background climate and water availability (Grimmond and Oke, 1986). Urban irrigation occurs mainly at the microscale limiting insight into this flux to estimations (Grimmond and Oke, 1986; Mitchell et al., 2001; Zeisl et al., 2018; Kokkonen et al., 2018). Irrigation adds to the intricacies of urban climate due to a combination of uncertainty and a potentially significant contribution to the water balance.

These "new" fluxes do not adhere to the same rules as the other fluxes, because they are driven by human behavior instead of physics. As was mentioned earlier, human intervention allows water to be imported artificially and the same is true for energy. For example, water is imported as drinking water transported from another catchment and used for garden irrigation. For energy, an example is fossil fuel transported to the city and combusted there releasing heat into the atmosphere.

Next to the newly introduced fluxes, people change the surface affecting the existing fluxes. On the scale of kilometers, urbanization transforms vast areas drastically. Such surface cover change will change the drivers and dynamics of the turbulent heat fluxes (Jansen et al., 2023). On a much smaller scale, people continuously change their surroundings. Examples are a tree planted in a garden or solar panels installed on the roof. While these changes may not noticeably change local scale fluxes, they may alter the microclimate of their direct surroundings further adding to the complexity of the urban climate.

Diversity

Urban areas vary within one city (heterogeneity), but they also vary among each other in many ways. Lipson et al. (2022b) show urbanization leads to impervious cover from nearly none to nearly complete. In their dataset, cities are spread over different climates with varying mean air temperatures and precipitation. Yet, more observations are available at the Northern Hemisphere temperate or continental locations than for other regions and climates (Grimmond, 2006a; Lipson et al., 2022b). Impervious fraction, background temperature, and precipitation are just three factors that can differ between cities.

Many more city characteristics than the impervious fraction change how cities affect their climate. In their metadata protocol, (Muller et al., 2013) describe the minimum metadata needed to describe urban measuring networks properly. This list outlines more characteristics relevant to urban climate including but not limited to the geographic setting, the geometry of surrounding buildings, and vegetation characteristics. Additional differences are found in building materials with different albedo, heat capacity, and heat

conductance. Many building materials are used around the world and which one is common in a particular city is often subject to local availability (Varghese, 2015). Additionally, culture and history create a city's unique identity and affect its architecture (Graham, 2012). All these variations play into the mix of interactions in cities that are heterogeneous within themselves. Hence, equations, concepts, and rules of thumb cannot be transferred to cities globally without considering these variations.

To capture the full diversity of cities, the field of urban climate has to pursue sites in the undersampled spectrum of all characteristics. These studies should include observations considering the challenges and opportunities described above and models.

Facing the challenge

To illustrate the challenge, I will describe an attempt to close the water balance in the city center of Amsterdam, the Netherlands. Despite the high data availability in the Netherlands, no comprehensive dataset exists that contains all the fluxes in the water balance. So, each one has to be examined on its own, starting with the input fluxes: precipitation and irrigation. For precipitation, the high data availability proves true, as multiple rain gauges are operated in the vicinity of Amsterdam. The spatial variability can even be examined in detail with radar observations. However, observation standards mean no rain gauges are operated within the city boundaries of Amsterdam itself. Irrigation is not directly observed, but piped water input can give an indication. As far as this human water input is measured, water use data are owned by commercial parties. Moreover, sharing this data raises privacy concerns. As in many cases, data is only retrievable when aggregated. In the case of Amsterdam, monthly data is available at the municipality scale. When examining the output fluxes, more and more the urban challenge presents itself. Amsterdam is equipped with an eddy-covariance system, but its dynamic footprint in the heterogeneous city center creates a spatial mismatch with the other datasets. Runoff to and through the canals is not observed but may be estimated with model simulations. Again, this represents another spatial extent. Sewage is equivalent to a black hole swallowing water without a tally. Water is stored in the groundwater and the canals. Water levels are recorded for both. While the open water levels can be interpolated easily, the urban heterogeneity and especially manmade obstacles make the groundwater level an irregular and discontinuous plane.

As the patchwork of datasets makes clear, this attempt to close the water balance is faced with the heterogeneity and human activity part of the challenge. Our focus on one city has spared this attempt from the diversity among cities. Still, most acquired datasets are unique to Amsterdam or otherwise the Netherlands and the approach would have to be adapted before it can be applied in another city. Yet, the water balance puzzle can be put together opportunistically taking advantage of the available datasets. However, the urban challenge makes our puzzle miss pieces while others are faded. These missing

and faded pieces muddle the final balance of this endeavor. As a result, over 70% of the precipitation and irrigation neither leaves nor is stored in the city center according to the available observations of runoff, ET, and water storage (Van Hal, 2023).

7.3.2 Urban climate modeling

Computational resources continue to grow, paving the way for new possibilities in modeling. A possibility particularly interesting for urban areas is increasing the spatial resolution. At a higher resolution, models capture more of the heterogeneity and processes. However, the costs of these computational resources warrant they are deployed purposefully. Operational numerical weather prediction is moving towards hectometer scales. During the 2024 Olympics in Paris, for instance, multiple institutes ran numerical weather prediction models at 100-m resolution in a temporary operational test setting (WMO, 2024). While possible in test and research settings (e.g. Ronda et al., 2017; Belair et al., 2018; Lean et al., 2019), models will have to be further developed to reach the sub-kilometer/hectometer scale operationally (Lean et al., 2024). After summarizing the benefits and unanswered questions of this sub-kilometer scale, I put this in the context of LES and climate models. The last part of this section focuses on the (development of) ULSMs lying below all these models and general model evaluation insights.

Increasing model resolution benefits the simulations in multiple ways, of which three important ones are described here. First, the higher spatial resolution can better describe the surface heterogeneity. With this better description, the temperature could be separately predicted for neighborhoods. Much more of the variation and heterogeneity is captured. Second, the more detailed orography helps capture small peak features in wind and precipitation. Third, the higher resolution moves convection from parameterized to resolved eliminating the need for the parameterizations.

Despite these advantages, higher spatial resolution not only increases the computational costs. Conceptually questions remain unanswered. When more of the surface heterogeneity is captured, urban characteristics can no longer be assumed homogenous within a grid cell, and concepts like local climate zones (Stewart and Oke, 2012) fail to describe the high variation. To properly describe the variation, high-detail datasets are required that may not always be available and without these datasets no additional surface details can be provided to the model. Additionally, the high variation between grid cells will also lead to advection becoming more important. Moreover, the discussed spatial scales are in the so-called turbulent grey zone, where the dominant turbulent length scales are similar to the spatial resolution and eddies are partially resolved (Honnert et al., 2020). This violates the assumptions of numerical weather prediction models with courser grids that separate the resolved large-scale dynamics from the parameterized turbulence. However, the hectometer scale models are not large-eddy simulations either that resolve the larger eddies and parameterize only the small-scale turbulence.

Numerical weather prediction falls between the spatial detail of climate models and LES. Global climate simulations run for much longer timescales than weather models. While spatial resolution is increasing, sub-kilometer resolutions are not feasible at the current state-of-the-art (Mauritsen et al., 2022). While these simulations will resolve larger metropolitan areas, the heterogeneity within the city cannot be captured. On the other end of the spectrum, LES offers spatial resolutions that drop below a meter for domains of a few square kilometers. This potential cannot yet be harvested operationally. In LES, the boundary conditions are an important limitation to represent real scenarios and in extension thereof operational applications, as seen in Chapter 6.

The different applications highlight the trade-off between temporal and spatial details and temporal and spatial extent. For each application, the drawbacks and additional costs urge considering the inherent value of increasing resolution. Details may be relevant for one purpose like renovating a street but irrelevant for another like a weather forecast. A higher resolution through switching model types or increasing the resolution is not necessarily required, justifiable, or beneficial. Now the availability of computational resources is less often the limiting factor, we should look critically at the environmental and economic costs of running intensive simulations (Silva et al., 2024). What model is most suitable, depends on the goal of the simulation. Who would check the details of an operational LES forecast that predicts the temperature at every meter in a street? On the other hand, how can urban climate be incorporated into the layout of the same street when it is simulated in blocks of 100 m?

Hectometer scale operational forecasts are not here now. Yet, urban climate modeling is entering the operational realm in the form of ULSMs. This year, 2024, ECMWF will include a ULSM based on CHTESSEL in its operational setup (McNorton et al., 2021; McNorton and Balsamo, 2023). With their increasing use in an online and even operational setting, it is worthwhile to consider the implications of the shortcomings found in ULSM water balance representations in Chapter 3 and 4.

When linked to a spatially distributed meteorological model, the water balance closure issues uncovered in Chapter 3 will be passed on to other grid cells creating or removing water and may cause the model to drift. Furthermore, water storage inconsistencies will result in untrustworthy ET during longer dry periods. Last, the unreliable runoff simulations are more relevant in a coupled setting, as the runoff generated in one cell will be input for its neighbor impacting the model simulation beyond one cell. Therefore, ULSMs cannot afford to overlook the water balance any longer.

When summarizing the main findings earlier, I discussed possible improvements to the water balance representation in ULSMs. The increasing resolution of numerical weather prediction models makes ULSMs increasingly relevant as part of weather and climate modeling. How can these models be further developed? Here, I will present possibilities available to improve ULSMs beyond the water balance presented earlier.

The first possibility is a new evaluation strategy considering the time-dependent footprints of the EC observations. Now the models are provided with the surface properties in the average footprint or even the average of the surroundings within a certain radius. The EC observations, however, represent the surface cover from a dynamic footprint, which is relevant even within a day as shown in Chapter 5. Given the variation observed in that chapter, a lot of understanding is lost by evaluating a model prescribed with average surface characteristics with observations with a dynamic footprint composition.

These evaluations could be improved by accounting for the dynamic footprint of eddy-covariance. Many ULSMs follow the tiling approach with one tile representing a certain surface cover type. The tiles are combined based on the averaged surface properties rather than the surface properties of the footprint at that moment. As the tiles are often modeled separately, they can be combined based on the relevant footprint. This is similar to an observation operator used in for example land surface temperature evaluation with satellites in urban areas (Yang et al., 2020) or rainfall evaluation with radar (Ryzhkov et al., 2011). Observation operators ensure the modeled variable resembles the observed variable more closely.

A side note on footprint modeling is warranted by their importance for observation interpretation. Currently, no analytical footprint model exists dedicated to urban surfaces. Idealized LES could inform how the footprint changes due to the building characteristics and the urban plan. Very few studies have estimated footprints in idealized urban settings. In the end, the idealized simulations will have to be compared with realistic ones. Auvinen et al., 2017 expose the challenges with footprint modeling over realistic urban canopies. These challenges stem from the combination of highly variable flows and knowing the observed location rather than the source of ET. While these challenges limited them to modeling a single footprint, their analysis shows that the surface cover composition is significantly different between the analytical and numerical approach. The urban roughness is not properly represented in the current analytical models and turbulence cannot smoothen out the preferential flow paths in urban canyons.

Next to considering footprints during evaluation, urban heterogeneity may be captured more comprehensively in ULSMs. While edge effects occur at the interface of different surface covers, the urban mosaic is now mostly summarized by surface fractions regardless of spatial organization. The omnipresence of edges in a mosaic warrants caution with the assumption that their effect is negligible. To set the scene, 20% of the world's forests are within 100 m of their boundary (Haddad et al., 2015). In urbanized regions like the northeastern United States, 11% of the forest is within only 30 m of a road or building (Smith et al., 2018). Changes are visible in the soil conditions including moisture, temperature, and composition. These changes cohere with changes in growth and carbon uptake of vegetation (Reinmann and Hutyra, 2017). All these changes will affect the

urban climate for example higher soil moisture increasing ET and more vegetation growth increasing shading.

Currently, most edge effects are not represented in ULSMs. Notable exceptions are urban canyon models that have 2D interactions within the street canyon. This exception also demonstrates the benefits of including edge effects, as models with vegetation in the street canyon outperform those without (Lipson et al., 2024). Yet, these models represent non-urban surfaces separately from the urban canyon following a tile approach, while these share edges with the urban canyons.

While canyon models explicitly represent the edges within the canyon, examples exist where an edge effect of the water balance is represented more conceptually. The effect in question is the buffering of high runoff volumes from impervious surfaces by the pervious area (Hopkins et al., 2015). By no means all ULSMs include this buffering, but two models sketch how (in)completely this is considered. The first model has a fudge factor, which describes the fraction of water exceeding the storage capacity per surface cover tile that is redivided over all tiles (Meili et al., 2020). The second model also redistributes the runoff exceeding the storage capacity, but rather than one fudge factor a parameter is defined for each tile combination (Järvi et al., 2011). In both cases, the edge length and the spatial configuration is ignored while this will determine the extent of the interaction. So, improvements are possible even for included edge processes.

After sketching these pathways for ULSM development, I will finish my reflection on modeling by zooming out and considering the lessons learned that apply to model evaluation in general. These lessons are based on the model evaluation in Chapter 3 and 4. This evaluation was conducted with data gathered by but independently from the main Urban-PLUMBER project. This thesis has a novel focus on the water balance, which revealed new possibilities for model improvement. ULSMs have more model components that have received little or no attention in model evaluation such as the seasonal vegetation changes and human heat fluxes. The Urban-PLUMBER data set provides an opportunity to assess these components as well. Model evaluation projects should aim for such spin-offs that reuse their efforts and provide insights into otherwise unstudied model components.

Model components remain unevaluated when no observations are available. Chapter 3 demonstrates that the absence of observations is not inevitably the end of quantitative evaluation. Logical checks allow diagnosing model errors due to internal model inconsistencies, misconceptualization, and human error. These logical checks should be deliberately designed to confirm whether the model results adhere to physical laws like the conservation of mass and fall within the expected magnitude. This strategy may even be applied alongside observation-based evaluation, as this evaluation will be affected by the uncertainty inherent to observations (Aerts et al., 2023). When using observations, error metrics should be chosen thoughtfully. Error metrics capture specific behavior depending on the evaluated variable. An example is the normalized standard deviation for the turbulent heat

fluxes in Chapter 3 and Lipson et al. (2024). This metric shows how well the amplitude of the daily cycle is mimicked, as most variation is within the daily cycle. In the end, specific metrics pinpoint more precisely where and how models go wrong, whether these metrics are based on observations or not.

As models are built and run by people, human errors can be made throughout the modeling process. Additional mistakes may be made during post-processing and in communication with others. As model intercomparisons aim to understand model rather than modeler performance, these human errors cloud the sought-after relations between model and performance. Recently, model intercomparisons have begun to include a step, in which the results are subjected to simple checks such as the sign and magnitude of the modeled variables (Menard et al., 2020; Lipson et al., 2024). When models fail these checks, the modelers are offered the chance to correct their mistakes. In this process, observations are not shared and no calibration is possible. While this process will reduce human error in the modeling and post-processing, fully eliminating human error is unattainable. Inherently, a model comparison also compares the participating modelers (Clarke, 2008). A modeler rather than a model intercomparison study could shed light on the influence of modelers on the model results. Apart from modelers, models differ in many more ways, which hampers the attribution of errors to certain conceptual choices. To circumvent this maze of differences, one could build a single modeling framework with multiple options for concepts and implementations. In such a framework, individual choices can be tested and their effects isolated.

7.3.3 Water and thermal comfort

Throughout this thesis, water and thermal comfort are linked through the cooling potential of ET regulated by water availability and limitation. Yet, humidity also plays a role in human thermal comfort (Sobolewski et al., 2021). Intuitively we know that a hot and dry place is more manageable than a hot and humid place. While the urban heat island is arguably the most studied phenomenon in urban climate (Stewart, 2011; Kim and Brown, 2021), the urban moisture island has received less attention. For a first impression, I assessed the urban moisture island for observational networks in five cities: Amsterdam (The Netherlands), Ghent (Belgium), Novi Sad (Serbia), Phoenix (The United States of America), and Sydney (Australia). The first results suggest that perhaps the lack of consistency between the cities explains why not much has been written about this topic yet.

All networks show the classical urban heat island with the highest values at night and lower, sometimes negative, values during the day (Figure 7.3). Shifting the focus from air temperature to specific humidity, the similarity in the patterns disappears and every city has a distinct fingerprint. In Amsterdam, the urban area is more humid during the day and slightly less moist at night, while Novi Sad mirrors this pattern. Ghent is always

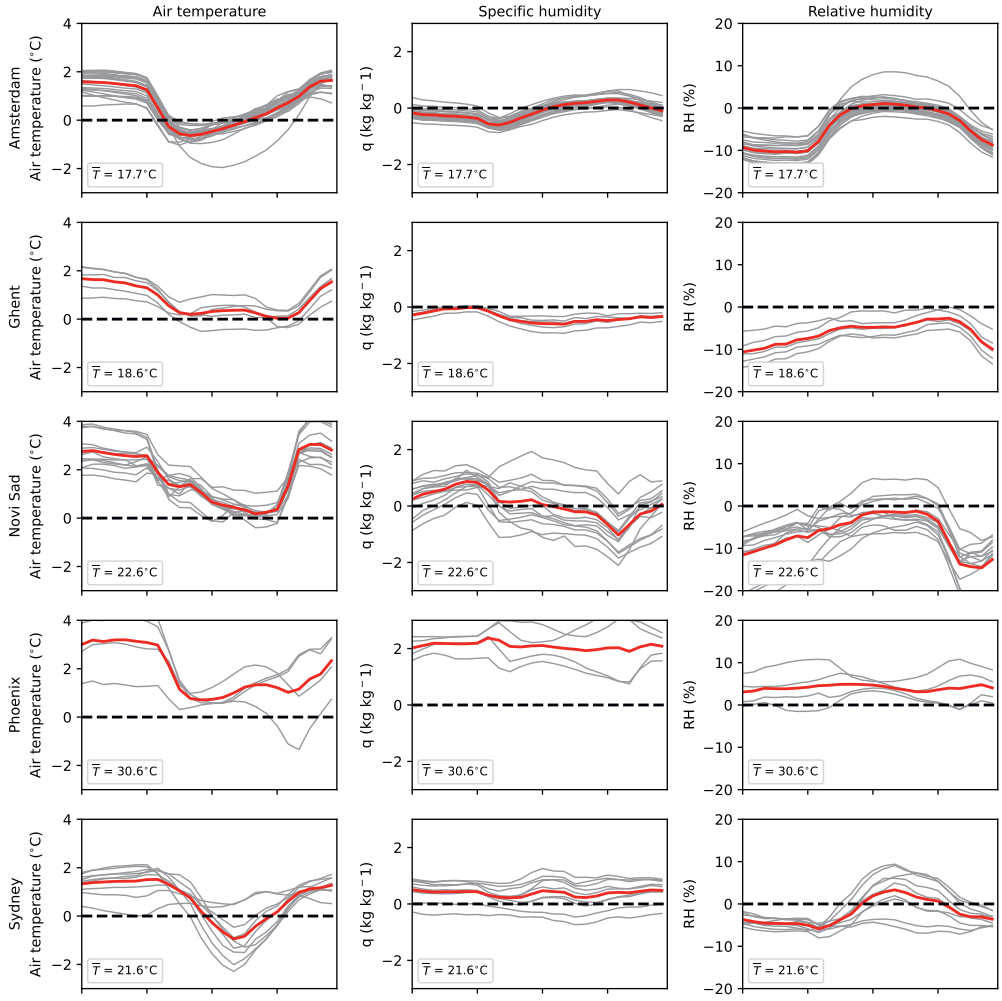


Figure 7.3: Average daily trend in summer of the difference between air temperature (first column), specific humidity (second column), and relative humidity (third column) at multiple urban sites and a rural reference site in Amsterdam (Ronda et al., 2017), Ghent (Caluwaerts et al., 2020), Novi Sad (Šećerov et al., 2019), Phoenix (AZMet, 2024), and Sydney (Ulpiani et al., 2022). Each grey line represents one observation location with the median in red. The temperature in the box is the mean temperature during the observed period. Summer is defined as JJA for the northern hemisphere and DJF for the southern hemisphere.

drier than its surroundings and Phoenix always wetter. In Sydney, the variation is higher between the urban stations than with the rural reference. Despite the differences in specific humidity, the relative humidity is more comparable amongst cities with larger differences at night and close to no difference during the day. Here, Phoenix is the odd one out

with constantly higher relative humidity. Some of these patterns may be easier to explain than others, but a comprehensive concept explaining urban moisture dynamics is not yet available.

During the day another factor is relevant for thermal comfort. Shading influences thermal comfort more strongly than air temperature and humidity within an urban canyon. Water may seem of little relevance for shade. However, when trees experience water stress during a drought, trees may start shedding their leaves (Tyree et al., 1993). The shed leaves reduce the shading for the remainder of the growing season. More extremely, prolonged water stress may induce tree mortality (Allen et al., 2015; McDowell et al., 2018; Brodribb et al., 2020). When a tree dies, its shading effects are lost not only for that growing season but also for the years after. A new tree will potentially require decades to regrow to the size of the original tree. Therefore, ET estimations are essential for urban water management to prevent water stress and watering trees should not lightly be put low on the priority list during a water shortage.

Humidity's role in thermal comfort and water stress leading to leaf shedding and tree mortality shows that the effect of the urban water balance on thermal comfort is not merely through ET. These other links cannot be ignored when choices are made regarding water management. Humidity-dependent thermal comfort may cancel (a part of) the cooling effects from ET, while increased tree mortality may deprive a neighborhood of proper shading for years or even decades. Urban water management is troubled by the complexity of the urban challenge. This complexity has no beginning or end. The circle is closed by humidity's negative feedback on ET.

7.4 Towards application

The title of this section starts with the word *towards*, because more steps are to be taken before application can be fully achieved for two reasons. The first is the number of disciplines relevant to the functioning of a city beyond the urban climate field. The second is the urban challenge discussed elaborately only in Section 7.3.1. Cities are built across the world. Their characteristics reflect different locations, climates, and cultures, creating a rich palette. Within these cities, even more hues and shades differentiate neighborhoods. These contrasting situations cause dissimilar responses to measures to mitigate heat stress. Acknowledging these challenges, I will sketch how urban climate research could be applied and become more applicable, and what hurdles currently prevent this. Here, I focus on how the urban climate field can start hurdling instead of how it could await developments in other fields to remove the hurdles.

To start, I will revisit the main findings of this thesis and how they could be applied. ET in cities was found to be water-limited due to low water storage capacities. How quickly ET declines after rainfall depends strongly on the surface cover. While these conclusions

are valuable lessons that may inform choices in the future, the findings concerning ULSMs are not as easily transferred to practice. They point out how ULSMs could be improved, but these models are not built to show the impact of changes made to the urban area. ULSMs are built with research and numerical weather forecasting in mind, yet ULSMs have the potential to predict the effect of changes to the urban environment. At this moment, they are not tailored for this purpose. Their parameters limit the characteristics, of which the impact can be determined, and their output is not translated to variables directly relevant to the people living in cities.

ULSMs are not a one-off when it comes to refraining from taking the next step towards application. First, not a single scale is sufficient to understand the impact of changes on urban climate. Scale-connecting studies will help to map intervention effects across scales. Chapter 5 and 6 are examples that show how the effect of surface changes on the neighborhood ET can be estimated. Second, mitigation measures and their effects remain specific to their circumstances. Intervention outcomes are ambiguous and hard to isolate. Whether a net positive or negative effect is reached depends on the exact urban setting. This holds for green and blue infrastructure, which are common strategies to cool cities. Trees can cool through ET and shading but may heat their surrounding by preventing ventilation (Meili et al., 2021). Water bodies cool through evaporation but absorb most incoming radiation and can store this energy due to a high heat capacity (Gunawardena et al., 2017). Third, two-way communication has to be facilitated between science and practice. On one hand, research findings should be made accessible without jargon. On the other hand, the right questions need to be formulated and answered to make urban climate research applicable. Variables helpful for understanding are not always the same as the ones that are relevant for people. While studying the surface temperature can benefit our fundamental understanding, one may question its relevance for a pedestrian compared to air temperature or even more comprehensive thermal comfort.

A fourth factor is that scientific disciplines tend to study the urban climate only from their perspective (González et al., 2021). This thesis puts the spotlight on the connection between the water and energy balance and in doing so it bridges the disciplines of hydrology and meteorology. Combining knowledge and concepts from both fields opened the door to new methods and synergies, such as ULSM evaluation using the curve number method. Bridging disciplines is thus not only possible but also helpful.

Yet, two disciplines are a step in a journey. Many other disciplines are connected to urban climate. Connections may present themselves in several ways. Urban climate may have implications beyond human thermal comfort, such as the required hardiness of trees (Lanza and Stone Jr, 2016) or freeze-thaw damage to buildings and monuments (Guilbert et al., 2019). These disciplines interact with urban climate. As explained earlier, water and thermal comfort are connected manifold. Urban greenery can help to cool urban areas (Norton et al., 2015) and may increase happiness at the same time (Syamili et al.,

2023). Also, social factors matter too as cooling is not distributed fairly across the urban demographic (Rocha et al., 2024). Within a city, all these fields have to be balanced. Therefore, their knowledge has to be connected and made accessible to ensure comfortable and well-functioning cities. This requires a dedicated effort to integrate urban climate insights with themselves and the related fields.

Change inevitably follows human settlement and this change continues during their presence. All these changes enable us to live close together in cities. Simultaneously, they affect the urban climate, in which water plays a key role. Whenever we change our (urban) environment, its climate should be considered to improve the quality of life.

Appendices

A.2 Additional table and figures Chapter 2

Table A.2.1: Same as Table 2.1, but with results from the analysis with ET corrected for the amount of available solar energy.

City	Drydowns	Days	ET_0 (mm d^{-1})	λ (day)	$t_{\frac{1}{2}}$ (day)	S_0 (mm)	R^2
Amsterdam	16	61	0.6 – 2.1 (1.5)	2.8 – 7.6 (5.2)	1.9 – 5.2 (3.6)	3.9 – 14.8 (6.6)	0.60
Arnhem	39	148	0.9 – 1.3 (1.1)	1.6 – 2.9 (2.2)	1.1 – 2.0 (1.6)	2.1 – 3.2 (2.6)	0.80
Basel (AESC)	109	445	0.9 – 1.2 (1.1)	4.1 – 5.2 (4.7)	2.8 – 3.6 (3.3)	3.9 – 5.4 (4.7)	0.75
Basel (KLIN)	150	623	1.2 – 1.4 (1.3)	5.5 – 7.2 (6.3)	3.8 – 5.0 (4.4)	6.4 – 9.3 (7.4)	0.66
Berlin (ROTH)	9	36	0.6 – 1.9 (0.8)	4.8 – 13.7 (11.9)	3.3 – 9.5 (8.2)	4.2 – 22.1 (11.5)	0.79
Berlin (TUCC)	30	122	0.4 – 0.9 (0.6)	2.4 – 4.0 (2.8)	1.7 – 2.8 (2.0)	1.2 – 3.1 (1.8)	0.68
Helsinki	41	177	1.7 – 2.0 (1.8)	3.4 – 7.8 (5.0)	2.4 – 5.0 (3.5)	6.6 – 11.9 (8.6)	0.80
Heraklion (HECKOR)	3	13	0.9 – 3.4 (2.9)	0.8 – 5.0 (1.7)	0.6 – 3.5 (1.2)	1.5 – 14.3 (2.9)	0.86
Lodz	55	249	1.3 – 1.8 (1.6)	3.2 – 4.8 (3.9)	2.2 – 3.3 (2.7)	4.2 – 7.6 (5.5)	0.70
Melbourne	2	9	0.7 – 1.8 (1.2)	1.6 – 10.2 (5.9)	1.1 – 7.1 (4.1)	1.1 – 17.9 (9.5)	0.65
Mexico City	9	52	0.8 – 1.5 (1.4)	4.8 – 14.6 (9.5)	3.3 – 10.1 (6.6)	5.6 – 19.1 (11.4)	0.60
Seoul	7	39	1.1 – 2.7 (1.7)	1.7 – 8.2 (4.3)	1.2 – 5.7 (3.0)	5.5 – 9.7 (8.9)	0.53
Singapore	8	43	1.3 – 1.6 (1.4)	6.2 – 17.7 (8.8)	4.3 – 12.3 (6.1)	9.3 – 24.6 (12.5)	0.76
Vancouver	61	282	1.3 – 1.7 (1.4)	4.9 – 7.8 (6.1)	3.4 – 5.4 (4.2)	6.7 – 10.0 (7.7)	0.60

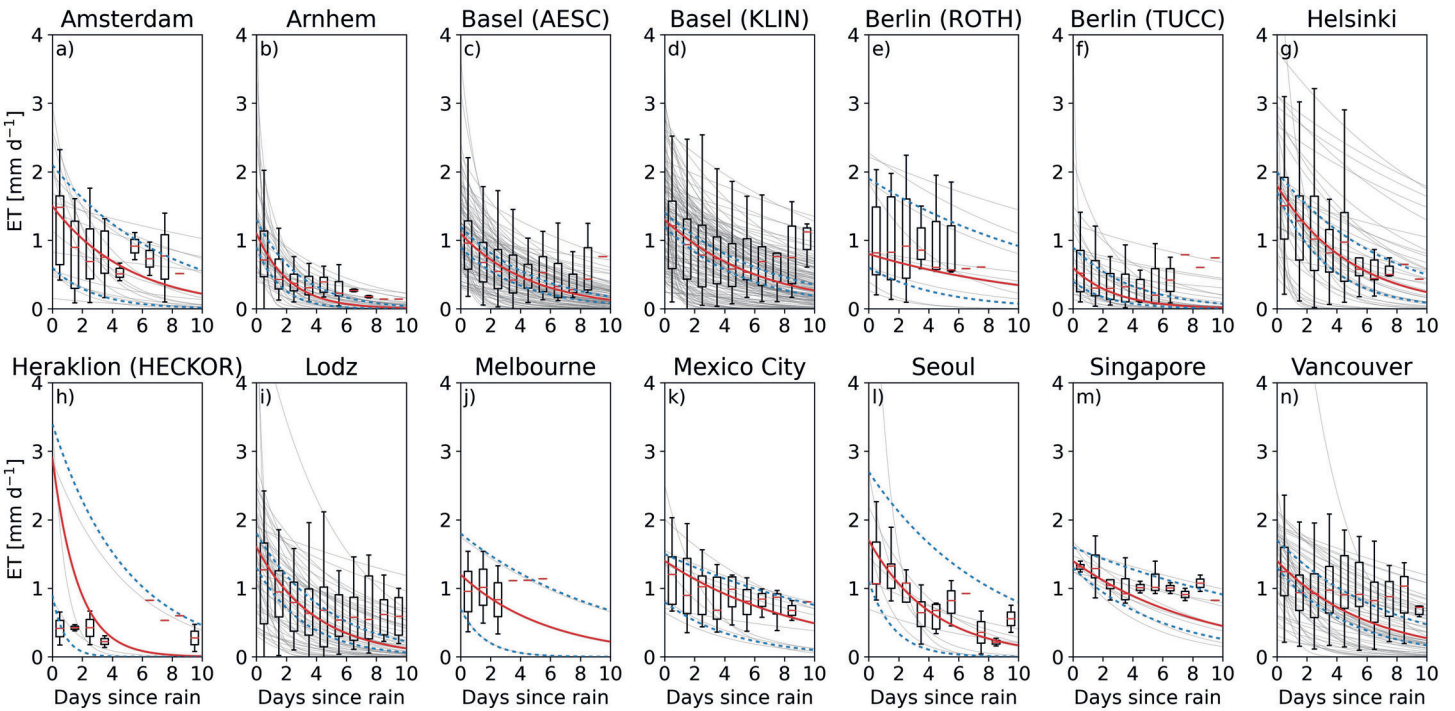


Figure A.2.1: Same as Figure 2.2, but with results from the analysis with ET corrected for the amount of available solar energy at the surface. This correction is performed by multiplying the evaporative fraction by the average available energy over the drydown.

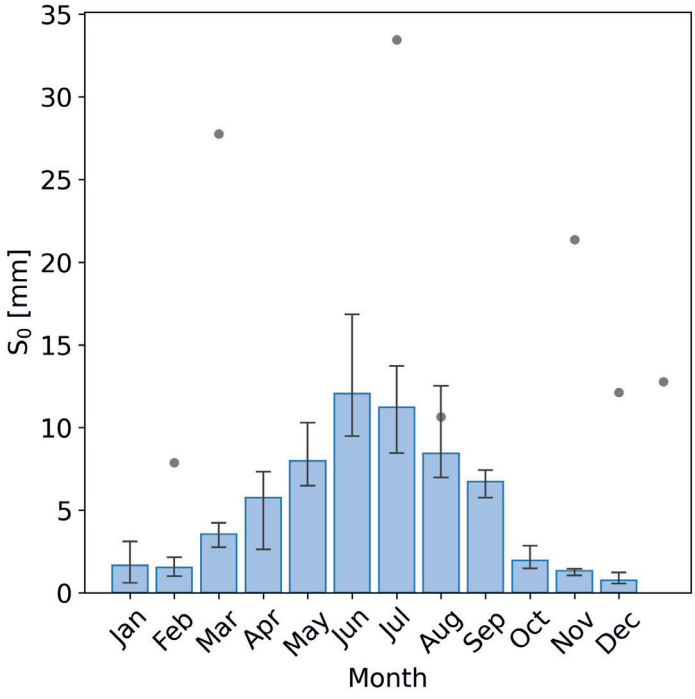


Figure A.2.2: Same as Figure 2.3, but with results from the analysis with ET corrected for the amount of available solar energy. This correction is performed by multiplying the evaporative fraction by the average available energy over the drydown.

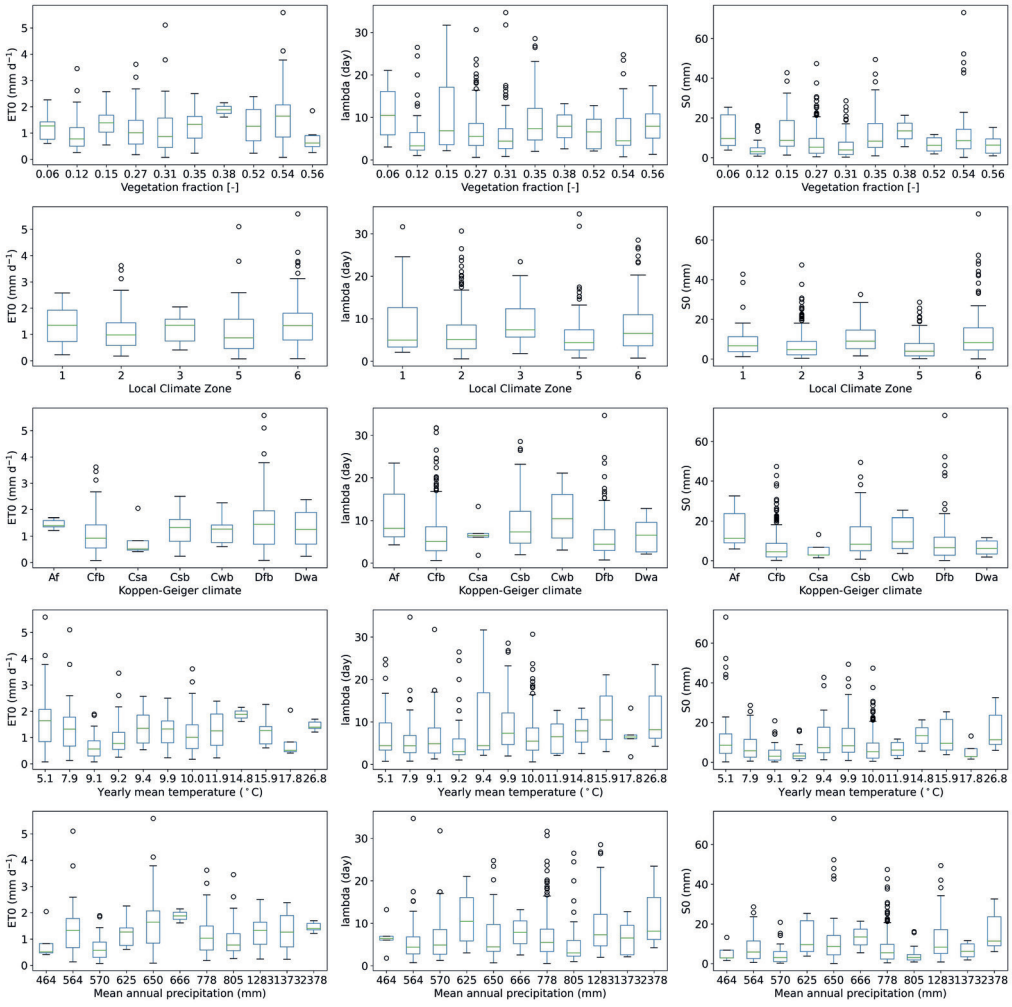


Figure A.2.3: Estimated model parameters (left column: ET_0, middle column: λ, right column: S_0) as function of climatological and urban form site characteristics from top to bottom the vegetation fraction, local climate zone, Köppen-Geiger climate, yearly mean temperature and mean annual precipitation.

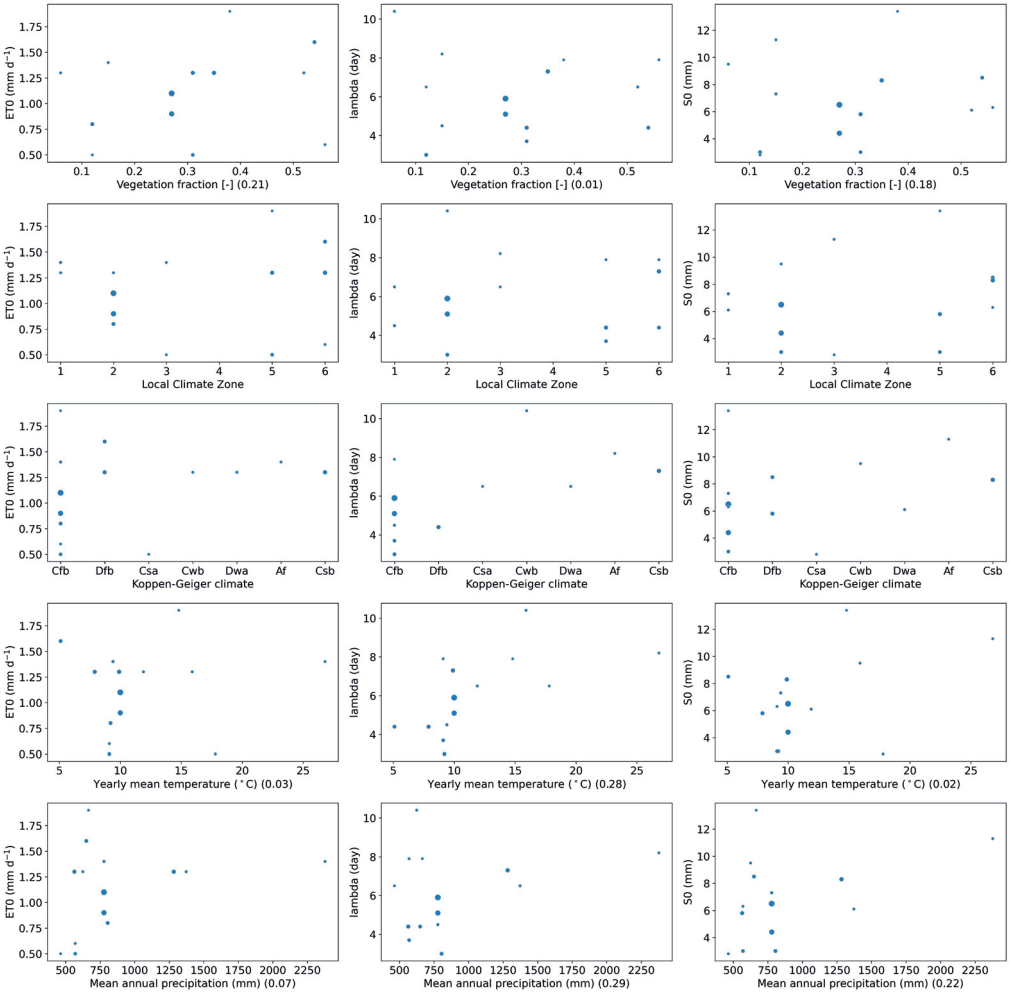


Figure A.2.4: Estimated model parameters as function of climatological and urban form site characteristics. The size of the dots indicates the number of drydowns. Between brackets the correlation coefficient is displayed based on a weighted linear regression (based on the number of drydowns per city) for the quantitative site characteristics.

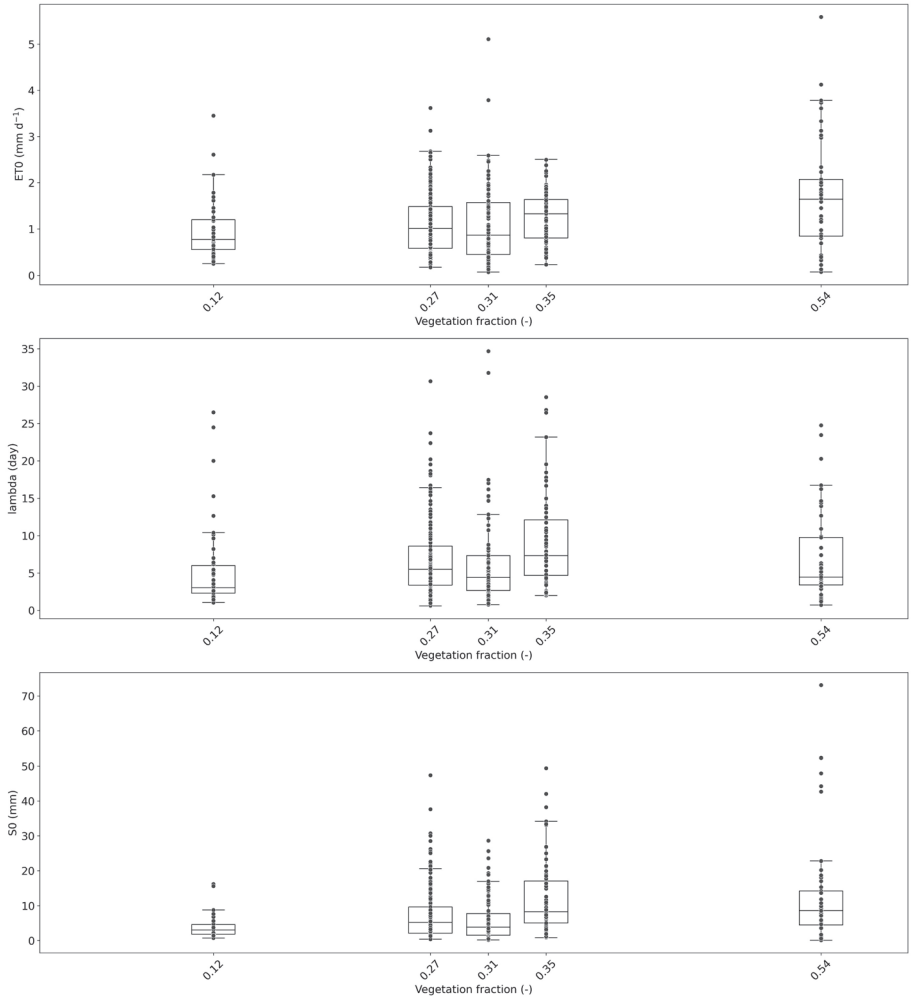


Figure A.2.5: Boxplots of estimated model parameters as function of vegetation fraction. Only locations with at least 20 drydowns are taken into account.

A.3 Additional figures Chapter 3

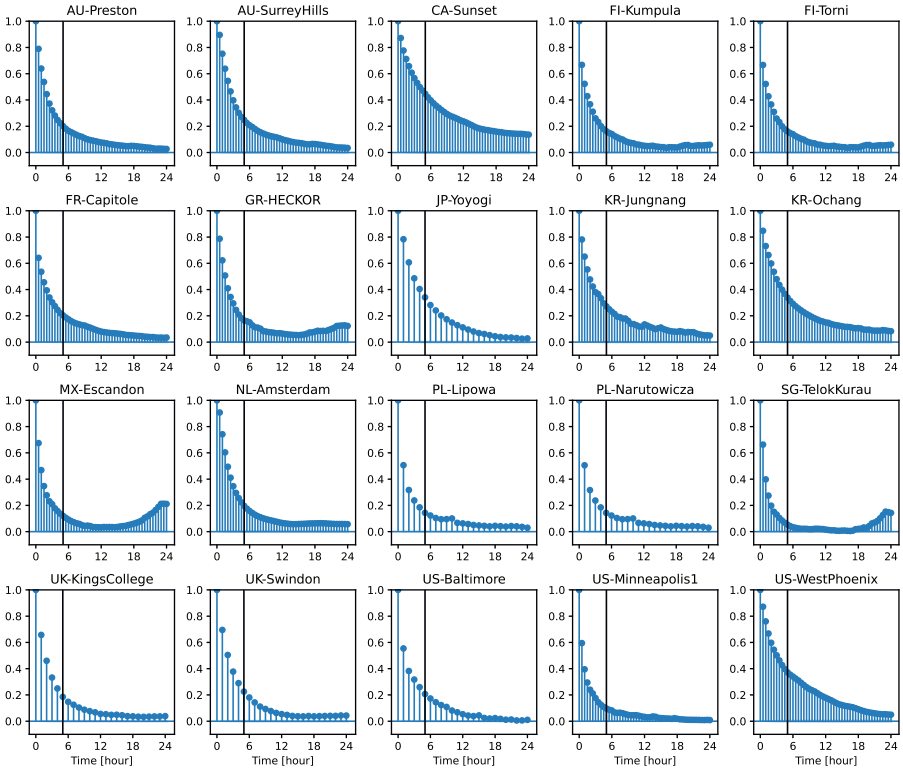


Figure A.3.1: Auto-correlation of (half-)hourly precipitation for all sites to determine separate events with the threshold of 5 hours (vertical line).

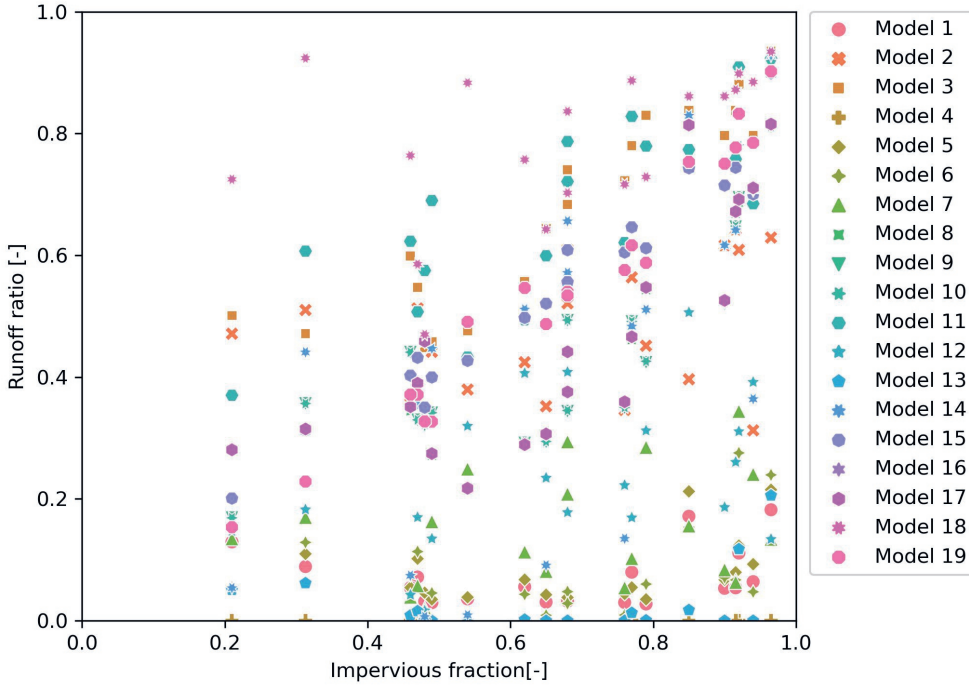


Figure A.3.2: Relation between the site impervious fraction and runoff ratio per model run (R/P) for each model (marker).

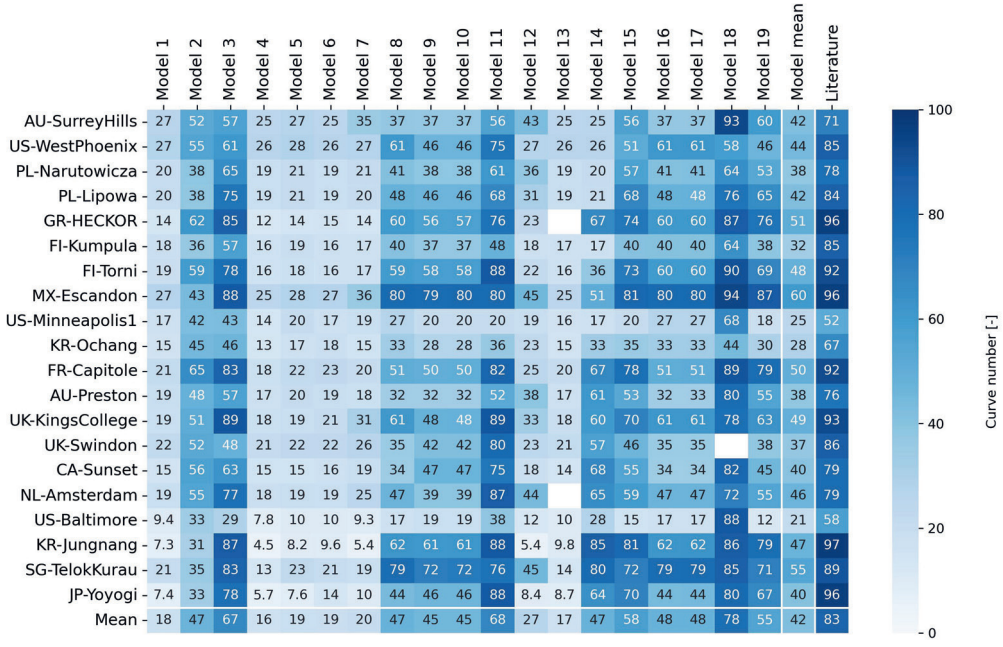


Figure A.3.3: Curve numbers estimated from modeled runoff for each model run and from curve number reference tables in Croshney et al. (1985), see methods section.

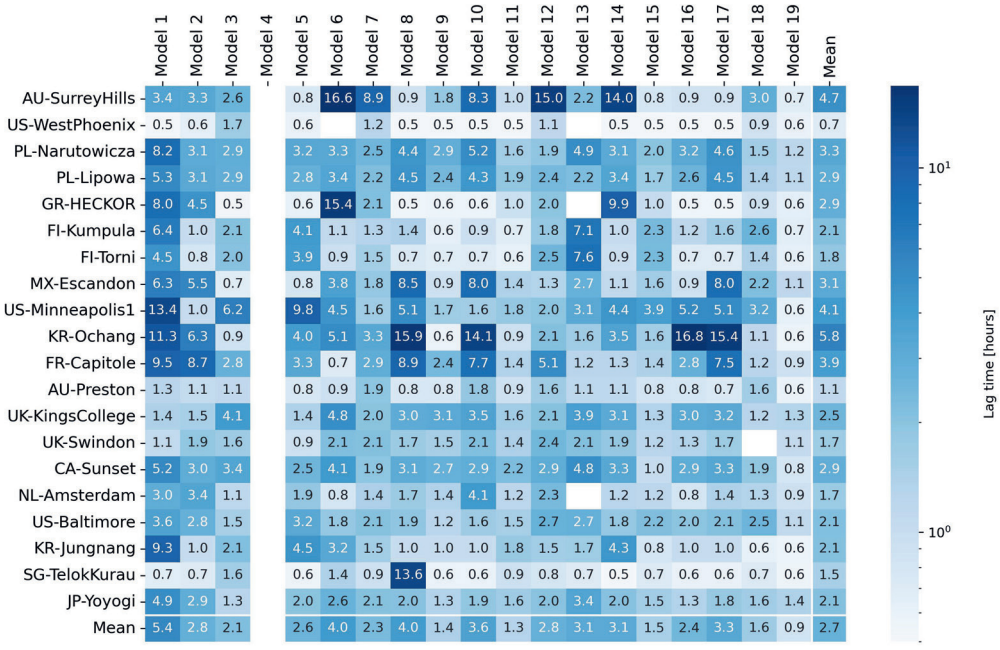


Figure A.3.4: Average lag times (hours) for all model runs at each site. Note that the color scale is logarithmic.

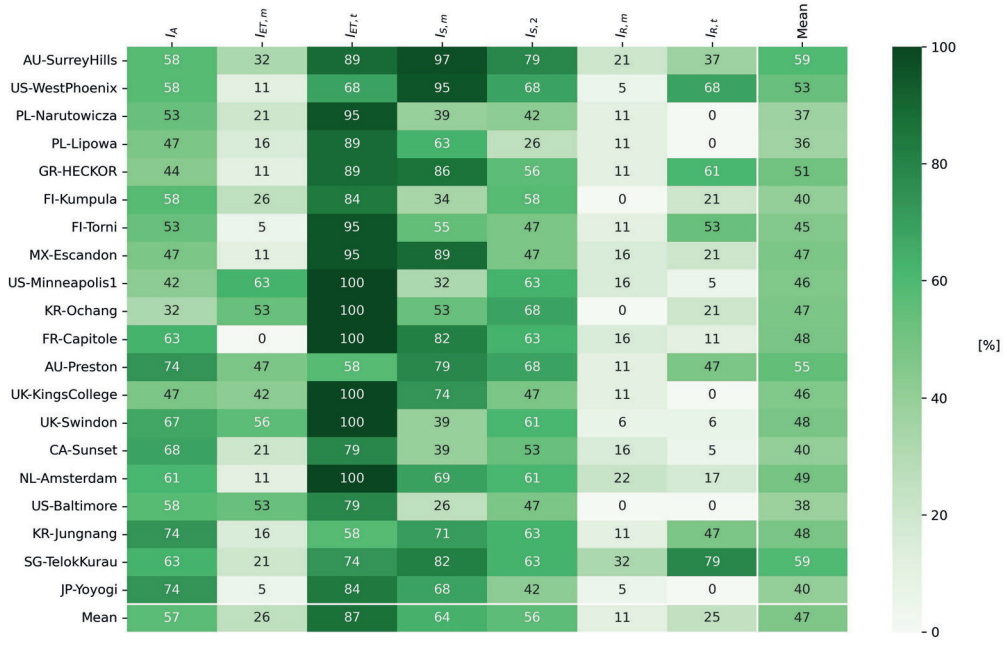


Figure A.3.5: As Figure 3.4, but are averaged per site.

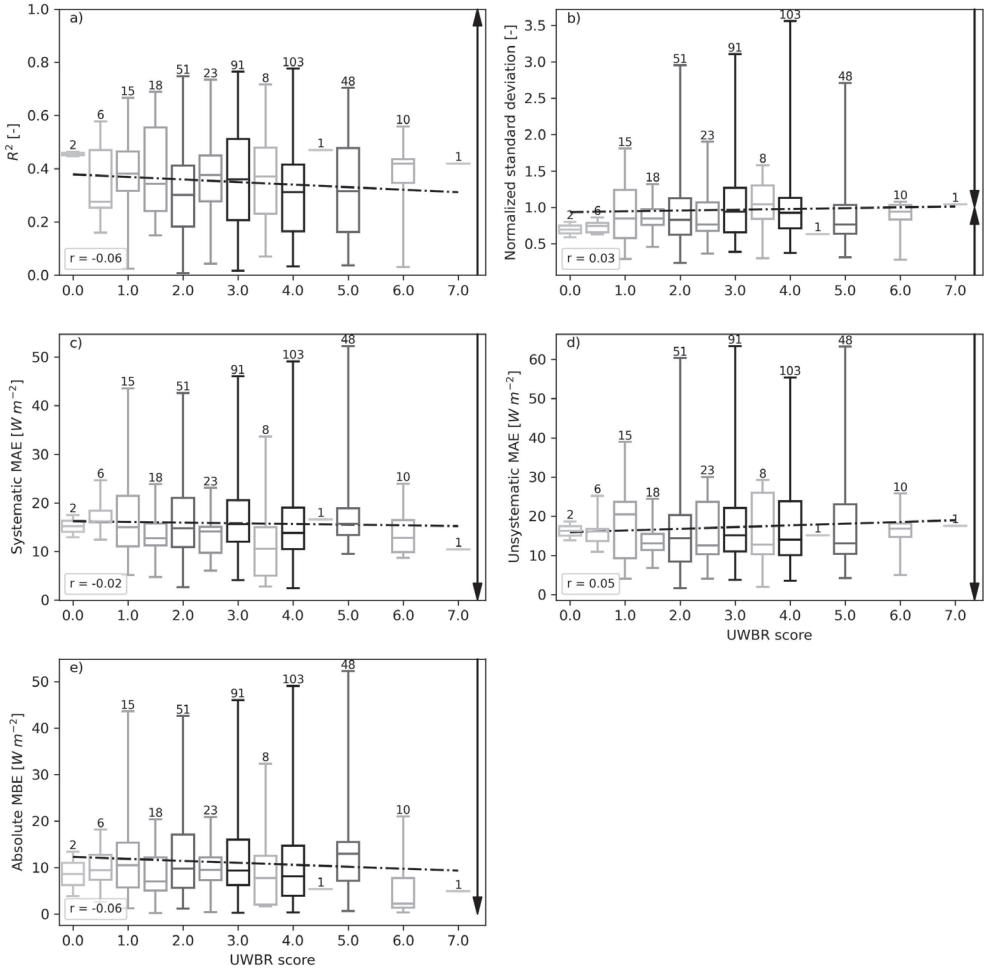


Figure A.3.6: Relation between the urban water balance representation (UWBR) score (0.5) classes and five error metrics for the (half-)hourly modeled Q_E: (a) R^2, (b) normalized standard deviation, (c) systematic, (d) unsystematic MAE, and (e) absolute MBE. The number of model runs per class is given above the box. Whiskers indicate the minimum and maximum values. A Wald (1943) test indicates a significant correlation (*). Arrows indicate the direction of better model performance.

A.5 Additional figure Chapter 5

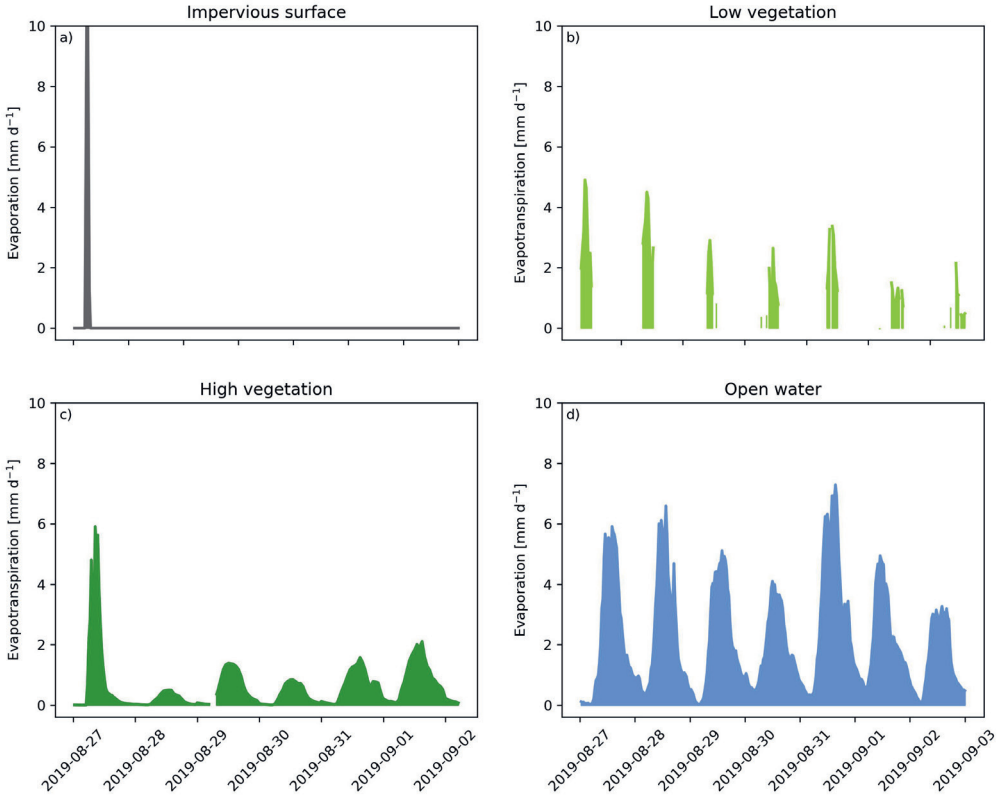


Figure A.5.1: As Figure 5.7, but for one of the last drydowns of the warm season. This drydown occurred between 27-08-2019 and 03-09-2019.

A.6 Additional figure Chapter 6

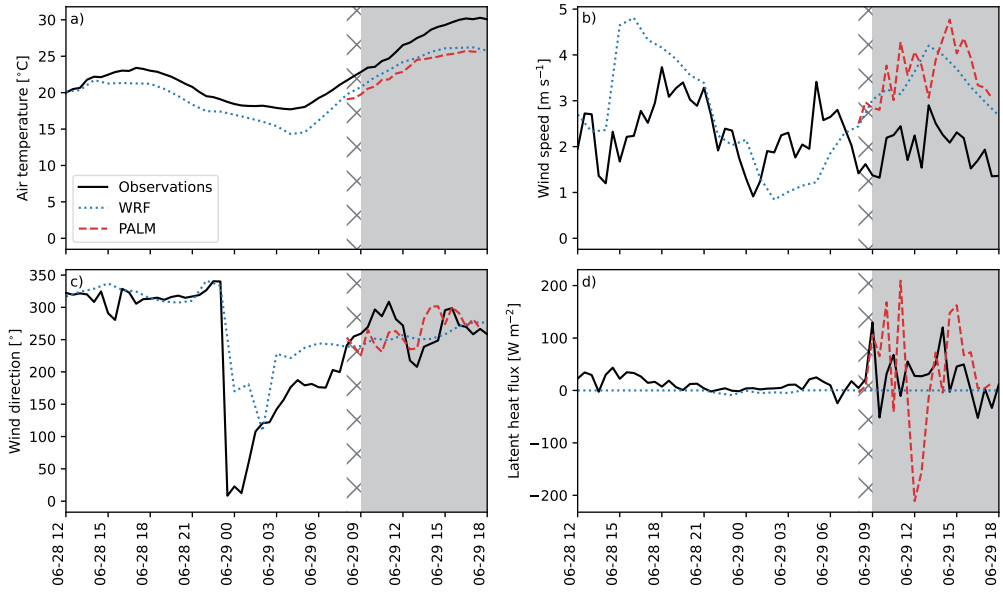


Figure A.6.1: As Figure 6.3, but at the 56-m EC system at TUCC using the P4 PALM domain.

References

Aerts, J. P., J. M. Hoch, G. Coxon, N. C. van de Giesen, and R. W. Hut (2023). "On the importance of observation uncertainty when evaluating and comparing models: A hydrological example". *EGUsphere* 2023, 1–26.

Allen, C. D., D. D. Breshears, and N. G. McDowell (2015). "On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene". *Ecosphere* 6 (8), 1–55.

Allen, L., F. Lindberg, and S. Grimmond (2011). "Global to city scale urban anthropogenic heat flux: Model and variability". *International Journal of Climatology* 31 (13), 1990.

Allen, R. G., M. Tasumi, and R. Trezza (2007). "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model". *Journal of Irrigation and Drainage Engineering* 133 (4), 380–394.

Amt für Statistik Berlin-Brandenburg (2019). *Das Statistische Informationssystem Berlin-Brandenburg*. Date accessed: 9 September 2021.

Anys, M. and M. Weiler (2024). "Rainfall interception by urban trees: Event characteristics and tree morphological traits". *Hydrological Processes* 38 (4), e15146.

Arnold Jr, C. L. and C. J. Gibbons (1996). "Impervious surface coverage: The emergence of a key environmental indicator". *Journal of the American planning Association* 62 (2), 243–258.

Asadian, Y. and M. Weiler (2009). "A new approach in measuring rainfall interception by urban trees in coastal British Columbia". *Water Quality Research Journal* 44 (1), 16–25.

Aston, A. (1979). "Rainfall interception by eight small trees". *Journal of Hydrology* 42 (3-4), 383–396.

Aubinet, M., T. Vesala, and D. Papale (2012). *Eddy covariance: A practical guide to measurement and data analysis*. Springer Science & Business Media.

Auvinen, M., L. Järvi, A. Hellsten, Ü. Rannik, and T. Vesala (2017). "Numerical framework for the computation of urban flux footprints employing large-eddy simulation and Lagrangian stochastic modeling". *Geoscientific Model Development* 10 (11), 4187–4205.

Avissar, R. (1992). "Conceptual aspects of a statistical-dynamical approach to represent landscape subgrid-scale heterogeneities in atmospheric models". *Journal of Geophysical Research: Atmospheres* 97 (D3), 2729–2742.

AZMet (2024). *AZMet - Arizona Meteorological Network*. Date accessed: 4 May 2024.

Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts (2009). "A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System". *Journal of Hydrometeorology* 10 (3), 623–643.

Baptista, M. D., S. J. Livesley, E. G. Parmehr, M. Neave, and M. Amati (2018). "Variation in leaf area density drives the rainfall storage capacity of individual urban tree species". *Hydrological Processes* 32 (25), 3729–3740.

Barker, F., R. Faggian, and A. J. Hamilton (2011). "A history of wastewater irrigation in Melbourne, Australia". *Journal of Water Sustainability* 1 (2), 31–50.

Baruch, E. M., K. A. Voss, J. R. Blaszczak, J. Delesantro, D. L. Urban, and E. S. Bernhardt (2018). "Not all pavements lead to streams: Variation in impervious surface connectivity affects urban stream ecosystems". *Freshwater Science* 37 (3), 673–684.

Bastiaanssen, W. G., M. Menenti, R. Feddes, and A. Holtslag (1998). "A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation". *Journal of Hydrology* 212, 198–212.

Belair, S., S. Leroyer, N. Seino, L. Spacek, V. Souvanlassy, and D. Paquin-Ricard (2018). "Role and impact of the urban environment in a numerical forecast of an intense summertime precipitation event over Tokyo". *Journal of the Meteorological Society of Japan. Ser. II*.

Belušić, D., D. H. Lenschow, and N. J. Tapper (2014). "Performance of a mobile car platform for mean wind and turbulence measurements". *Atmospheric Measurement Techniques* 7 (6), 1825–1837.

Berne, A., G. Delrieu, J.-D. Creutin, and C. Obled (2004). "Temporal and spatial resolution of rainfall measurements required for urban hydrology". *Journal of Hydrology* 299 (3-4), 166–179.

Berthier, E., H. Andrieu, and F. Rodriguez (1999). "The Rezé urban catchments database". *Water Resources Research* 35 (6), 1915–1919.

Best, M. J., G. Abramowitz, H. Johnson, A. Pitman, G. Balsamo, A. Boone, M. Cuntz, B. Decharme, P. Dirmeyer, J. Dong, et al. (2015). "The plumbing of land surface models: Benchmarking model performance". *Journal of Hydrometeorology* 16 (3), 1425–1442.

Best, M. J. and S. Grimmond (2016). "Modeling the partitioning of turbulent fluxes at urban sites with varying vegetation cover". *Journal of Hydrometeorology* 17 (10), 2537–2553.

Best, M. J., M. Pryor, D. Clark, G. Rooney, R. Essery, C. Ménard, J. Edwards, M. Hendry, A. Porson, N. Gedney, et al. (2011). "The Joint UK Land Environment Simulator (JULES), model description–Part 1: Energy and water fluxes". *Geoscientific Model Development* 4 (3), 677–699.

Beven, K., R. Lamb, P. Quinn, R. Romanowicz, J. Freer, et al. (1995). "Topmodel." *Computer Models of Watershed Hydrology.*, 627–668.

Bhaduri, B., J. Harbor, B. Engel, and M. Grove (2000). "Assessing watershed-scale, long-term hydrologic impacts of land-use change using a GIS-NPS model". *Environmental Management* 26, 643–658.

Bian, Z., Y. Gu, J. Zhao, Y. Pan, Y. Li, C. Zeng, and L. Wang (2019). "Simulation of evapotranspiration based on leaf area index, precipitation and pan evaporation: A case study of Poyang Lake watershed, China". *Ecohydrology & Hydrobiology* 19 (1), 83–92.

Bjorkegren, A., C. S. B. Grimmond, S. Kotthaus, and B. Malamud (2015). "CO_2 emission estimation in the urban environment: Measurement of the CO_2 storage term". *Atmospheric Environment* 122, 775–790.

Blyth, E. M., V. K. Arora, D. B. Clark, S. J. Dadson, M. G. De Kauwe, D. M. Lawrence, J. R. Melton, J. Pongratz, R. H. Turton, K. Yoshimura, et al. (2021). "Advances in land surface modelling". *Current Climate Change Reports* 7 (2), 45–71.

Boer-Euser, T. de, L. Bouaziz, J. De Niel, C. Brauer, B. Dewals, G. Drogue, F. Fenicia, B. Grelier, J. Nossent, F. Pereira, et al. (2016). "Looking beyond general metrics for model comparison–lessons from an international model intercomparison study". *Hydrology and Earth System Sciences Discussions* 2016, 1–29.

Boese, S., M. Jung, N. Carvalhais, A. J. Teuling, and M. Reichstein (2019). "Carbon–water flux coupling under progressive drought". *Biogeosciences* 16 (13), 2557–2572.

Bonifacio-Bautista, M., M. Ballinas, A. Jazcilevich, and V. L. Barradas (2022). "Estimation of anthropogenic heat release in Mexico City". *Urban Climate* 43, 101158.

Bonneau, J., T. D. Fletcher, J. F. Costelloe, P. J. Poelsma, R. B. James, and M. J. Burns (2018). "Where does infiltrated stormwater go? Interactions with vegetation and subsurface anthropogenic features". *Journal of Hydrology* 567, 121–132.

Botzen, W., M. Martinius, P. Bröde, M. Folkerts, P. Ignjacevic, F. Estrada, C. Harmsen, and H. Daanen (2020). "Economic valuation of climate change–induced mortality: Age-dependent cold and heat mortality in the Netherlands". *Climatic Change* 162 (2), 545–562.

Boussetta, S., G. Balsamo, G. Arduini, E. Dutra, J. McNorton, M. Choulga, A. Agustí-Panareda, A. Beljaars, N. Wedi, J. Munõz-Sabater, et al. (2021). "ECLand: The ECMWF land surface modelling system". *Atmosphere* 12 (6), 723.

Brakensiek, D. L. and W. J. Rawls (1983). "Green-Ampt infiltration model parameters for hydrologic classification of soils". In: *Advances in irrigation and drainage: Surviving external pressures*. ASCE, 226–233.

Bringfelt, B. (1986). *Test of a forest evapotranspiration model*. SMHI.

Brodribb, T. J., J. Powers, H. Cochard, and B. Choat (2020). "Hanging by a thread? Forests and drought". *Science* 368 (6488), 261–266.

Brutsaert, W. and J. L. Nieber (1977). "Regionalized drought flow hydrographs from a mature glaciated plateau". *Water Resources Research* 13 (3), 637–643.

Bryant, M. L., S. Bhat, and J. M. Jacobs (2005). "Measurements and modeling of throughfall variability for five forest communities in the Southeastern US". *Journal of Hydrology* 312 (1-4), 95–108.

Burba, G. and D. Anderson (2010). *A brief practical guide to eddy covariance flux measurements: Principles and workflow examples for scientific and industrial applications*. Li-Cor Biosciences.

Cai, X. and M. Y. Leclerc (2007). "Forward-in-time and backward-in-time dispersion in the convective boundary layer: The concentration footprint". *Boundary-Layer Meteorology* 123, 201–218.

Caluwaerts, S., R. Hamdi, S. Top, D. Lauwaet, J. Berckmans, D. Degrauwe, H. Dejonghe, K. De Ridder, R. De Troch, F. Duchêne, et al. (2020). "The urban climate of Ghent, Belgium: A case study combining a high-accuracy monitoring network with numerical simulations". *Urban Climate* 31, 100565.

Cao, W., Y. Zhou, B. Güneralp, X. Li, K. Zhao, and H. Zhang (2022). "Increasing global urban exposure to flooding: An analysis of long-term annual dynamics". *Science of The Total Environment* 817, 153012.

Carlyle-Moses, D. E., S. Livesley, M. D. Baptista, J. Thom, and C. Szota (2020). "Urban trees as green infrastructure for stormwater mitigation and use". *Forest-Water Interactions*, 397–432.

Chakravarti, I., R. Laha, and J. Roy (1967). *Handbook of methods of applied statistics. Volume I: Techniques of computation, descriptive methods, and statistical inference*. John Wiley & Sons, Incorporated.

Châtelet, É. (1740). *Institutions de Physique*. Paris.

Chen, H., J. J. Huang, H. Liang, W. Wang, H. Li, Y. Wei, A. Z. Jiang, and P. Zhang (2023). "Can evaporation from urban impervious surfaces be ignored?" *Journal of Hydrology* 616, 128582.

Chin, D. A. (2022). "Minimum rainfall for applying the curve-number method". *Journal of Irrigation and Drainage Engineering* 148 (8), 06022003.

Chow, W. T. (2017). *Eddy covariance data measured at the CAP LTER flux tower located in the West Phoenix, AZ neighborhood of Maryvale from 2011-12-16 through 2012-12-31*.

Chow, W. T., F. Salamanca, M. Georgescu, A. Mahalov, J. M. Milne, and B. L. Ruddell (2014a). "A multi-method and multi-scale approach for estimating city-wide anthropogenic heat fluxes". *Atmospheric Environment* 99, 64–76.

Chow, W. T., T. J. Volo, E. R. Vivoni, G. D. Jenerette, and B. L. Ruddell (2014b). "Seasonal dynamics of a suburban energy balance in Phoenix, Arizona". *International Journal of Climatology* 34 (15), 3863–3880.

Christen, A., N. Coops, B. Crawford, R. Kellett, K. Liss, I. Olchovski, T. Tooke, M. Van Der Laan, and J. Voogt (2011). "Validation of modeled carbon-dioxide emissions from an urban neighborhood with direct eddy-covariance measurements". *Atmospheric Environment* 45 (33), 6057–6069.

Christen, A. and R. Vogt (2004). "Energy and radiation balance of a central European city". *International Journal of Climatology* 24 (11), 1395–1421.

Clarke, R. T. (2008). "Issues of experimental design for comparing the performance of hydrologic models". *Water Resources Research* 44 (1).

Conte, M. and D. Contini (2019). "Size-resolved particle emission factors of vehicular traffic derived from urban eddy covariance measurements". *Environmental Pollution* 251, 830–838.

Coutts, A. M., J. Beringer, and N. J. Tapper (2007a). "Characteristics influencing the variability of urban CO_2 fluxes in Melbourne, Australia". *Atmospheric Environment* 41 (1), 51–62.

– (2007b). "Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia". *Journal of Applied Meteorology and Climatology* 46 (4), 477–493.

Crawford, B. and A. Christen (2015). "Spatial source attribution of measured urban eddy covariance CO_2 fluxes". *Theoretical and Applied Climatology* 119, 733–755.

Crawford, B., S. Grimmond, and A. Christen (2011). "Five years of carbon dioxide fluxes measurements in a highly vegetated suburban area". *Atmospheric Environment* 45 (4), 896–905.

Cronshey, R., R. Roberts, and N. Miller (1985). "Urban hydrology for small watersheds (TR-55 Rev.)" In: *Hydraulics and Hydrology in the Small Computer Age*. ASCE, 1268–1273.

Dai, Y., J. Boekee, B. Schilperoort, M.-C. ten Veldhuis, and B. J. Van de Wiel (2023). "Wind machines for frost damage mitigation: A quantitative 3D investigation based on observations". *Agricultural and Forest Meteorology* 338, 109522.

Dardanelli, J. L., J. Ritchie, M. Calmon, J. M. Andriani, and D. J. Collino (2004). "An empirical model for root water uptake". *Field Crops Research* 87 (1), 59–71.

De Bruin, H. (1979). "Neerslag, openwaterverdamping en potentieel neerslagoverschot in Nederland, frequentieverdelingen in het groeiseizoen". *WR 79-04*.

De Vos, L., H. Leijnse, A. Overeem, and R. Uijlenhoet (2017). "The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam". *Hydrology and Earth System Sciences* 21 (2), 765–777.

Deardorff, J. W. (1980). "Stratocumulus-capped mixed layers derived from a three-dimensional model". *Boundary-Layer Meteorology* 18, 495–527.

Deines, J. M., X. Liu, and J. Liu (2016). "Telecoupling in urban water systems: An examination of Beijing's imported water supply". *Water International* 41 (2), 251–270.

Deletic, A. B. and C. Maksimovic (1998). "Evaluation of water quality factors in storm runoff from paved areas". *Journal of Environmental Engineering* 124 (9), 869–879.

Demirel, Y. (2012). *Energy: Production, conversion, storage, conservation, and coupling*. Springer Science & Business Media.

Demotes-Mainard, S., S. Herpin, A. Boukouya, S. Mballo, B. Dubuc, L. Ledroit, C. Le Lebras, D. Lemesle, and P.-E. Bournet (2022). "Impacts of the urban environment on well-watered tree architectural development and tree climate services". In: *XXXI International Horticultural Congress (IHC2022): II International Symposium on Greener Cities: Improving Ecosystem Services 1374*, 189–196.

Demuzere, M., J. Kittner, A. Martilli, G. Mills, C. Moede, I. D. Stewart, J. van Vliet, and B. Bechtel (2022). "A global map of Local Climate Zones to support Earth system modelling and urban scale environmental science". *Earth System Science Data Discussions* 2022, 1–57.

Denich, C. and A. Bradford (2010). "Estimation of evapotranspiration from bioretention areas using weighing lysimeters". *Journal of Hydrologic Engineering* 15 (6), 522–530.

Droste, A. M., B. G. Heusinkveld, D. Fenner, and G.-J. Steeneveld (2020). "Assessing the potential and application of crowdsourced urban wind data". *Quarterly Journal of the Royal Meteorological Society* 146 (731), 2671–2688.

Duarte Rocha, A., S. Vulova, C. van der Tol, M. Förster, and B. Kleinschmit (2022). "Modelling hourly evapotranspiration in urban environments with SCOPE using open remote sensing and meteorological data". *Hydrology and Earth System Sciences* 26 (4), 1111–1129.

Dunne, T. and R. D. Black (1970). "Partial area contributions to storm runoff in a small New England watershed". *Water Resources Research* 6 (5), 1296–1311.

DWD (2021). *Vieljährige Mittelwerte*. Date accessed: 9 September 2021.

Ek, M., K. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. Tarpley (2003). "Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model". *Journal of Geophysical Research: Atmospheres* 108 (D22).

Ennos, R. (2010). "Urban cool". *Physics World* 23 (08), 22.

Euser, T., W. Luxemburg, C. Everson, M. Mengistu, A. Clulow, and W. Bastiaanssen (2014). "A new method to measure Bowen ratios using high-resolution vertical dry and wet bulb temperature profiles". *Hydrology and Earth System Sciences* 18 (6), 2021–2032.

Fan, X., H. Shao, M. Sakamoto, K. Kuga, L. Lan, D. P. Wyon, K. Ito, M. P. Bivolarova, C. Liao, and P. Wargocki (2022). "The effects of ventilation and temperature on sleep quality and next-day work performance: Pilot measurements in a climate chamber". *Building and Environment* 209, 108666.

Fank, J. (2011). "Lysimeters: A tool for measurements of soil fluxes". In: *Encyclopedia of Agrophysics*. Ed. by J. Gliński, J. Horabik, and J. Lipiec. Dordrecht: Springer Netherlands, 428–431.

Fedora, M. and R. Beschta (1989). "Storm runoff simulation using an antecedent precipitation index (API) model". *Journal of Hydrology* 112 (1-2), 121–133.

Feigenwinter, C., R. Vogt, and A. Christen (2012). "Eddy covariance measurements over urban areas". In: *Eddy covariance a practical guide to measurement and data analysis*. Ed. by M. Aubinet, T. Vesala, and D. Papale. Dordrecht: Springer Netherlands. Chap. 16, 377–397.

Feng, L., Y. Liu, Y. Zhou, and S. Yang (2022). "A UAV-derived thermal infrared remote sensing three-temperature model and estimation of various vegetation evapotranspiration in urban micro-environments". *Urban Forestry & Urban Greening* 69, 127495.

Fenner, D., A. Christen, N. Chrysoulakis, S. Grimmond, M. Van Hove, S. Kotthaus, F. Meier, W. Morrison, and M. Zeeman (2023). *Impact of a large isolated city on the mixed layer height during different weather conditions*. Tech. rep. Copernicus Meetings.

Fenner, D., F. Meier, D. Scherer, and A. Polze (2014). "Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001–2010". *Urban Climate* 10, 308–331.

Fisher, R. A. and C. D. Koven (2020). "Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems". *Journal of Advances in Modeling Earth Systems* 12 (4), e2018MS001453.

Fletcher, T. D., H. Andrieu, and P. Hamel (2013). "Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art". *Advances in Water Resources* 51, 261–279.

Foken, T., M. Göockede, M. Mauder, L. Mahrt, B. Amiro, and W. Munger (2004). "Post-field data quality control". In: *Handbook of micrometeorology: A guide for surface flux measurement and analysis*. Springer, 181–208.

Foken, T., R. Leuning, S. R. Oncley, M. Mauder, and M. Aubinet (2012). "Corrections and data quality control". In: *Eddy covariance*. Springer, 85–131.

Fortuniak, K. (2003). "A slab surface energy balance model (SUEB) and its application to the study on the role of roughness length in forming an urban heat island". *Acta Universitatis Wratislaviensis* 2542, 368–377.

Fortuniak, K., K. Kłysik, and M. Siedlecki (2006). "New measurements of the energy balance components in Łódź". In: *Preprints, sixth International Conference on Urban Climate, Göteborg, Sweden*, 12–16.

Fortuniak, K., W. Pawlak, and M. Siedlecki (2013). "Integral turbulence statistics over a central European city centre". *Boundary-Layer Meteorology* 146, 257–276.

Fouillet, A., G. Rey, F. Laurent, G. Pavillon, S. Bellec, C. Guihenneuc-Jouyaux, J. Clavel, E. Jougla, and D. Hémon (2006). "Excess mortality related to the August 2003 heat wave in France". *International Archives of Occupational and Environmental Health* 80, 16–24.

Franssen, H. H., R. Stöckli, I. Lehner, E. Rotenberg, and S. I. Seneviratne (2010). "Energy balance closure of eddy-covariance data: A multisite analysis for European FLUXNET stations". *Agricultural and Forest Meteorology* 150 (12), 1553–1567.

Gaines, J. M. (2016). "Water potential". *Nature* 531 (7594 S1), S54–S54.

Gallo, K., A. McNab, T. R. Karl, J. F. Brown, J. Hood, and J. Tarpley (1993). "The use of a vegetation index for assessment of the urban heat island effect". *Remote Sensing* 14 (11), 2223–2230.

Gao, K., J. Feng, and M. Santamouris (2024). "Are grand tree planting initiatives meeting expectations in mitigating urban overheating during heat waves?" *Sustainable Cities and Society* 113, 105671.

Gash, J., P. Rosier, and R. Ragab (2008). "A note on estimating urban roof runoff with a forest evaporation model". *Hydrological Processes* 22 (8), 1230–1233.

Gasparrini, A. and B. Armstrong (2011). "The impact of heat waves on mortality". *Epidemiology* 22 (1), 68–73.

Gasparrini, A., Y. Guo, F. Sera, A. M. Vicedo-Cabrera, V. Huber, S. Tong, M. d. S. Z. S. Coelho, P. H. N. Saldiva, E. Lavigne, P. M. Correa, et al. (2017). "Projections of temperature-related excess mortality under climate change scenarios". *The Lancet Planetary Health* 1 (9), e360–e367.

Gentine, P., D. Entekhabi, A. Chehbouni, G. Boulet, and B. Duchemin (2007). "Analysis of evaporative fraction diurnal behaviour". *Agricultural and Forest Meteorology* 143 (1-2), 13–29.

Gerrits, A. M. J. (2010). "The role of interception in the hydrological cycle". PhD thesis. TU Delft.

Gillefalk, M., D. Tetzlaff, C. Marx, A. Smith, F. Meier, R. Hinkelmann, and C. Soulsby (2022). "Estimates of water partitioning in complex urban landscapes with isotope-aided ecohydrological modelling". *Hydrological Processes* 36 (3), e14532.

Gillefalk, M., D. Tetzlaff, R. Hinkelmann, L.-M. Kuhlemann, A. Smith, F. Meier, M. P. Maneta, and C. Soulsby (2021). "Quantifying the effects of urban green space on water partitioning and ages using an isotope-based ecohydrological model". *Hydrology and Earth System Sciences* 25 (6), 3635–3652.

González, J. E., P. Ramamurthy, R. D. Bornstein, F. Chen, E. R. Bou-Zeid, M. Ghandehari, J. Luvall, C. Mitra, and D. Niyogi (2021). "Urban climate and resiliency: A synthesis report of state of the art and future research directions". *Urban Climate* 38, 100858.

Goret, M., V. Masson, R. Schoetter, and M.-P. Moine (2019). "Inclusion of CO_2 flux modelling in an urban canopy layer model and an evaluation over an old European city centre". *Atmospheric Environment: X* 3, 100042.

Graham, B. (2012). *Culture, Urbanism and Planning*. Ashgate Publishing, Ltd.

Graham, P., L. Maclean, D. Medina, A. Patwardhan, and G. Vasarhelyi (2004). "The role of water balance modelling in the transition to low impact development". *Water Quality Research Journal* 39 (4), 331–342.

Granger, R. and N. Hedstrom (2011). "Modelling hourly rates of evaporation from small lakes". *Hydrology and Earth System Sciences* 15 (1), 267–277.

Grimmond, C. S. B. and T. R. Oke (1986). "Urban water balance: 2. Results from a suburb of Vancouver, British Columbia". *Water Resources Research* 22 (10), 1404–1412.

– (1999). "Heat storage in urban areas: Local-scale observations and evaluation of a simple model". *Journal of Applied Meteorology and Climatology* 38 (7), 922–940.

Grimmond, S. (2006a). "Progress in measuring and observing the urban atmosphere". *Theoretical and Applied Climatology* 84 (1-3), 3–22.

Grimmond, S. (2006b). "Progress in measuring and observing the urban atmosphere". *Theoretical and Applied Climatology* 84 (1), 3–22.

Grimmond, S., M. J. Best, J. Barlow, A. Arnfield, J.-J. Baik, A. Baklanov, S. Belcher, M. Bruse, I. Calmet, F. Chen, et al. (2009). "Urban surface energy balance models: Model characteristics and methodology for a comparison study". In: *Meteorological and Air Quality Models for Urban Areas*. Springer, 97–123.

Grimmond, S., M. Blackett, M. J. Best, J.-J. Baik, S. Belcher, J. Beringer, S. Bohnenstengel, I. Calmet, F. Chen, A. Coutts, et al. (2011). "Initial results from Phase 2 of the international urban energy balance model comparison". *International Journal of Climatology* 31 (2), 244–272.

Grimmond, S., M. Blackett, M. J. Best, J. Barlow, J. Baik, S. Belcher, S. Bohnenstengel, I. Calmet, F. Chen, A. Dandou, et al. (2010). "The international urban energy balance models comparison

project: First results from phase 1". *Journal of Applied Meteorology and Climatology* 49 (6), 1268–1292.

Grimmond, S., V. Bouchet, L. T. Molina, A. Baklanov, J. Tan, K. H. Schlünzen, G. Mills, B. Golding, V. Masson, C. Ren, et al. (2020). "Integrated urban hydrometeorological, climate and environmental services: Concept, methodology and key messages". *Urban Climate* 33, 100623.

Grimmond, S. and T. R. Oke (1991). "An evapotranspiration-interception model for urban areas". *Water Resources Research* 27 (7), 1739–1755.

– (2002). "Turbulent heat fluxes in urban areas: Observations and a Local-scale Urban Meteorological Parameterization Scheme (LUMPS)". *Journal of Applied Meteorology and Climatology* 41 (7), 792–810.

Guilbert, D., S. Caluwaerts, K. Calle, N. Van Den Bossche, V. Cnudde, and T. De Kock (2019). "Impact of the urban heat island on freeze-thaw risk of natural stone in the built environment, a case study in Ghent, Belgium". *Science of The Total Environment* 677, 9–18.

Gunawardena, K. R., M. J. Wells, and T. Kershaw (2017). "Utilising green and bluespace to mitigate urban heat island intensity". *Science of the Total Environment* 584, 1040–1055.

Gurnell, A., M. Lee, and C. Souch (2007). "Urban rivers: Hydrology, geomorphology, ecology and opportunities for change". *Geography Compass* 1 (5), 1118–1137.

Haacke, N. (2022). "Scaling and connectivity assessment of critical source areas of diffuse pollution in urban catchments under a changing regime of high-intensity storms". PhD thesis. Technischen Universität Berlin.

Haddad, N. M., L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, et al. (2015). "Habitat fragmentation and its lasting impact on Earth's ecosystems". *Science Advances* 1 (2), e1500052.

Haggard, B., P. Moore Jr, and K. Brye (2005). "Effect of slope on runoff from a small variable-slope box". *Journal of Environmental Hydrology* 13 (2), 25–45.

Hamdi, R. and G. Schayes (2007). "Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities". *Atmospheric Chemistry and Physics* 7 (17), 4513–4526.

Hamdi, R., H. Kusaka, Q.-V. Doan, P. Cai, H. He, G. Luo, W. Kuang, S. Caluwaerts, F. Duchêne, B. Van Schaeybroek, et al. (2020). "The state-of-the-art of urban climate change modeling and observations". *Earth Systems and Environment*, 1–16.

Hamdi, R., P. Termonia, and P. Baguis (2011). "Effects of urbanization and climate change on surface runoff of the Brussels Capital Region: A case study using an urban soil–vegetation–atmosphere-transfer model". *International Journal of Climatology* 31 (13), 1959–1974.

Hammond, M. J., A. S. Chen, S. Djordjević, D. Butler, and O. Mark (2015). "Urban flood impact assessment: A state-of-the-art review". *Urban Water Journal* 12 (1), 14–29.

Harshan, S., M. Roth, E. Velasco, and M. Demuzere (2017). "Evaluation of an urban land surface scheme over a tropical suburban neighborhood". *Theoretical and Applied Climatology* 133 (3-4), 867–886.

Hartogensis, O., C. Watts, J. Rodriguez, and H. De Bruin (2003). "Derivation of an effective height for scintillometers: La Poza experiment in Northwest Mexico". *Journal of Hydrometeorology* 4 (5), 915–928.

Heaviside, C., H. Macintyre, and S. Vardoulakis (2017). "The urban heat island: Implications for health in a changing environment". *Current Environmental Health Reports* 4, 296–305.

Heaviside, C., S. Vardoulakis, and X.-M. Cai (2016). "Attribution of mortality to the urban heat island during heatwaves in the West Midlands, UK". *Environmental Health* 15 (1), 49–59.

Helbig, N. and H. Löwe (2012). "Shortwave radiation parameterization scheme for subgrid topography". *Journal of Geophysical Research: Atmospheres* 117 (D3).

Heldens, W., C. Burmeister, F. Kanani-Sühring, B. Maronga, D. Pavlik, M. Sühring, J. Zeidler, and T. Esch (2020). "Geospatial input data for the PALM model system 6.0: Model requirements, data sources and processing". *Geoscientific Model Development* 13 (11), 5833–5873.

Hellsten, A., S.-M. Luukkonen, G. Steinfeld, F. Kanani-Sühring, T. Markkanen, L. Järvi, J. Lento, T. Vesala, and S. Raasch (2015). "Footprint evaluation for flux and concentration measurements for an urban-like canopy with coupled Lagrangian stochastic and large-eddy simulation models". *Boundary-Layer Meteorology* 157 (2), 191–217.

Henderson-Sellers, A., K. McGuffie, and A. Pitman (1996). "The project for intercomparison of land-surface parametrization schemes (PILPS): 1992 to 1995". *Climate Dynamics* 12 (12), 849–859.

Hengl, T. (2018). *Sand content in % (Kg/Kg) at 6 Standard Depths (0, 10, 30, 60, 100 And 200 Cm) At 250 M Resolution (Version v0. 2) [Data set]*. DOI: `https://doi.org/10.5281/zenodo.2525662`.

Herrmann, D. L., L. A. Schifman, and W. D. Shuster (2018). "Widespread loss of intermediate soil horizons in urban landscapes". *Proceedings of the National Academy of Sciences* 115 (26), 6751–6755.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. (2020). "The ERA5 global reanalysis". *Quarterly Journal of the Royal Meteorological Society* 146 (730), 1999–2049.

Hertwig, D., S. Grimmond, M. A. Hendry, B. Saunders, Z. Wang, M. Jeoffrion, P. L. Vidale, P. C. McGuire, S. I. Bohnenstengel, H. C. Ward, et al. (2020). "Urban signals in high-resolution weather and climate simulations: Role of urban land-surface characterisation". *Theoretical and Applied Climatology* 142 (1), 701–728.

Herwitz, S. R. (1985). "Interception storage capacities of tropical rainforest canopy trees". *Journal of Hydrology* 77 (1-4), 237–252.

Heusinkveld, B. G., L. Van Hove, C. Jacobs, G. Steeneveld, J. Elbers, E. Moors, and A. Holtslag (2010). "Use of a mobile platform for assessing urban heat stress in Rotterdam". In: *Proceedings of the 7th Conference on Biometeorology*, 433–438.

Hirano, T., H. Sugawara, S. Murayama, and H. Kondo (2015). "Diurnal variation of CO_2 flux in an urban area of Tokyo". *Sola* 11, 100–103.

Hirschi, M., D. Michel, I. Lehner, and S. I. Seneviratne (2017). "A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration". *Hydrology and Earth System Sciences* 21 (3), 1809–1825.

Ho, J. Y., Y. Shi, K. K. Lau, E. Y. Ng, C. Ren, and W. B. Goggins (2023). "Urban heat island effect-related mortality under extreme heat and non-extreme heat scenarios: A 2010–2019 case study in Hong Kong". *Science of the Total Environment* 858, 159791.

Hong, S.-O., J. Kim, Y.-H. Byun, J. Hong, J.-W. Hong, K. Lee, Y.-S. Park, S.-S. Lee, and Y.-H. Kim (2023). "Intra-urban variations of the CO_2 fluxes at the surface-atmosphere interface in the Seoul Metropolitan Area". *Asia-Pacific Journal of Atmospheric Sciences*, 1–15.

Hong, J.-W., J. Hong, J. Chun, Y. H. Lee, L.-S. Chang, J.-B. Lee, K. Yi, Y.-S. Park, Y.-H. Byun, and S. Joo (2019). "Comparative assessment of net CO_2 exchange across an urbanization gradient in Korea based on eddy covariance measurements". *Carbon Balance and Management* 14 (1), 1–18.

Hong, J.-W., K. Lee, and J. Hong (2020a). *Observational data of Ochang and Jungnang in Korea, EAPL at Yonsei University [Data set].*

Hong, J.-W., S.-D. Lee, K. Lee, and J. Hong (2020b). "Seasonal variations in the surface energy and CO_2 flux over a high-rise, high-population, residential urban area in the East Asian monsoon region". *International Journal of Climatology.*

Honnert, R., G. A. Efstathiou, R. J. Beare, J. Ito, A. Lock, R. Neggers, R. S. Plant, H. H. Shin, L. Tomassini, and B. Zhou (2020). "The atmospheric boundary layer and the "gray zone" of turbulence: A critical review". *Journal of Geophysical Research: Atmospheres* 125 (13), e2019JD030317.

Hopkins, K. G., D. J. Bain, and E. M. Copeland (2014). "Reconstruction of a century of landscape modification and hydrologic change in a small urban watershed in Pittsburgh, PA". *Landscape Ecology* 29, 413–424.

Hopkins, K. G., N. B. Morse, D. J. Bain, N. D. Bettez, N. B. Grimm, J. L. Morse, M. M. Palta, W. D. Shuster, A. R. Bratt, and A. K. Suchy (2015). "Assessment of regional variation in streamflow responses to urbanization and the persistence of physiography". *Environmental Science & Technology* 49 (5), 2724–2732.

Horton, R. E. (1933). "The role of infiltration in the hydrologic cycle". *Eos, Transactions American Geophysical Union* 14 (1), 446–460.

Howard, L. (1833). *The climate of London deduced from meteorological observations made in the metropolis and various places around it.* Vol. 1. London: W. Phillips, George Yard, Lombard Street.

Hsieh, C.-I., M. Siqueira, G. Katul, and C.-R. Chu (2003). "Predicting scalar source-sink and flux distributions within a forest canopy using a 2-D Lagrangian stochastic dispersion model". *Boundary-Layer Meteorology* 109, 113–138.

Huang, J. Y., T. Black, R. Jassal, and L. L. Lavkulich (2017). "Modelling rainfall interception by urban trees". *Canadian Water Resources Journal/Revue Canadienne des Ressources Hydriques* 42 (4), 336–348.

Huang, J., Z. Tu, P. Du, J. Lin, and Q. Li (2010). "Uncertainties in stormwater runoff data collection from a small urban catchment, Southeast China". *Journal of Environmental Sciences* 22 (11), 1703–1709.

Huang, W. T. K., P. Masselot, E. Bou-Zeid, S. Fatichi, A. Paschalis, T. Sun, A. Gasparrini, and G. Manoli (2023). "Economic valuation of temperature-related mortality attributed to urban heat islands in European cities". *Nature Communications* 14 (1), 7438.

Hudson, N. W. (1985). *Soil conservation, second edition*. Cornell University Press.

Ichinose, T., K. Shimodozono, and K. Hanaki (1999). "Impact of anthropogenic heat on urban climate in Tokyo". *Atmospheric Environment* 33 (24-25), 3897–3909.

Ishidoya, S., H. Sugawara, Y. Terao, N. Kaneyasu, N. Aoki, K. Tsuboi, and H. Kondo (2020). "O_2: CO_2 exchange ratio for net turbulent flux observed in an urban area of Tokyo, Japan, and its application to an evaluation of anthropogenic CO_2 emissions". *Atmospheric Chemistry and Physics* 20 (9), 5293–5308.

Jacobs, C., J. Elbers, R. Brolsma, O. Hartogensis, E. Moors, M. T. R.-C. Márquez, and B. van Hove (2015). "Assessment of evaporative water loss from Dutch cities". *Building and Environment* 83, 27–38.

Jacobs, C., L. Klok, M. Bruse, J. Cortesão, S. Lenzholzer, and J. Kluck (2020). "Are urban water bodies really cooling?" *Urban Climate* 32, 100607.

Jacobson, C. R. (2011). "Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review". *Journal of Environmental Management* 92 (6), 1438–1448.

Jansen, F. A., H. J. Jongen, C. M. Jacobs, F. C. Bosveld, A. J. Buzacott, B. G. Heusinkveld, B. Kruijt, M. van der Molen, E. Moors, G.-J. Steeneveld, et al. (2023). "Land cover control on the drivers of evaporation and sensible heat fluxes: An observation-based synthesis for the Netherlands". *Water Resources Research* 59 (11), e2022WR034361.

Jansen, F. A. and A. J. Teuling (2020). "Evaporation from a large lowland reservoir–(dis) agreement between evaporation models from hourly to decadal timescales". *Hydrology and Earth System Sciences* 24 (3), 1055–1072.

Jansen, F. A., R. Uijlenhoet, C. M. Jacobs, and A. J. Teuling (2022). "Evaporation from a large lowland reservoir–observed dynamics and drivers during a warm summer". *Hydrology and Earth System Sciences* 26 (11), 2875–2898.

Järvi, L., S. Grimmond, and A. Christen (2011). "The surface urban energy and water balance scheme (SUEWS): Evaluation in Los Angeles and Vancouver". *Journal of Hydrology* 411 (3-4), 219–237.

Järvi, L., Ü. Rannik, T. V. Kokkonen, M. Kurppa, A. Karppinen, R. D. Kouznetsov, P. Rantala, T. Vesala, and C. R. Wood (2018). "Uncertainty of eddy covariance flux measurements over an urban area based on two towers". *Atmospheric Measurement Techniques* 11 (10), 5421–5438.

Jin, K., F. Wang, D. Chen, H. Liu, W. Ding, and S. Shi (2019). "A new global gridded anthropogenic heat flux dataset with high spatial resolution and long-term time series". *Scientific Data* 6 (1), 139.

Jin, L., S. Schubert, D. Fenner, F. Meier, and C. Schneider (2021). "Integration of a building energy model in an urban climate model and its application". *Boundary-Layer Meteorology* 178 (2), 249–281.

Jongen, H. J., M. Lipson, A. J. Teuling, S. Grimmond, J.-J. Baik, M. Best, M. Demuzere, K. Fortuniak, Y. Huang, M. G. De Kauwe, R. Li, J. McNorton, N. Meili, K. Oleson, S.-B. Park, T. Sun, A. Tsiringakis, M. Varentsov, C. Wang, Z.-H. Wang, and G. J. Steeneveld (2024a). "The Water Balance Representation in Urban-PLUMBER Land Surface Models". *Journal of Advances in Modeling Earth Systems* 16 (10), e2024MS004231.

Jongen, H. J., S. Vulova, F. Meier, G. J. Steeneveld, F. A. Jansen, D. Tetzlaff, B. Kleinschmit, N. Haacke, and A. J. Teuling (2024b). "Attributing urban evapotranspiration from eddy-covariance to surface cover: Bottom-up versus top-down". *Water Resources Research* 60 (9), e2024WR037508.

Jongen, H. J., G.-J. Steeneveld, J. Beringer, A. Christen, N. Chrysoulakis, K. Fortuniak, J. Hong, J.-W. Hong, C. M. Jacobs, L. Järvi, et al. (2022). "Urban water storage capacity inferred from observed evapotranspiration recession". *Geophysical Research Letters* 49 (3), e2021GL096069.

Jongen, H. J., A. J. Teuling, D. Li, E. Akinlabi, S. Gadde, and G.-J. Steeneveld (2024c). "Modeling urban evapotranspiration with large-eddy simulations: The spatial and temporal influence of eddy-covariance footprints". *In preparation*.

Jongen, H. J., A. J. Teuling, M. Lipson, S. Grimmond, M. Best, and G.-J. Steeneveld (2024d). "Do Urban-PLUMBER land surface models underestimate surface runoff?" *In preparation*.

Kahler, D. M. and W. Brutsaert (2006). "Complementary relationship between daily evaporation in the environment and pan evaporation". *Water Resources Research* 42 (5).

Kanda, M. (2009). "Review of Japanese urban models and a scale model experiment". *Meteorological and Air Quality Models for Urban Areas*, 39–46.

Kanda, M., A. Inagaki, M. O. Letzel, S. Raasch, and T. Watanabe (2004). "LES study of the energy imbalance problem with eddy covariance fluxes". *Boundary-Layer Meteorology* 110, 381–404.

Karl, T., C. Lamprecht, M. Graus, A. Cede, M. Tiefengraber, J. Vila-Guerau de Arellano, D. Gurarie, and D. Lenschow (2023). "High urban NO_x triggers a substantial chemical downward flux of ozone". *Science Advances* 9 (3), eadd2365.

Karsisto, P., C. Fortelius, M. Demuzere, C. S. B. Grimmond, K. Oleson, R. Kouznetsov, V. Masson, and L. Järvi (2016). "Seasonal surface urban energy balance and wintertime stability simulated using three land-surface models in the high-latitude city Helsinki". *Quarterly Journal of the Royal Meteorological Society* 142 (694), 401–417.

Kim, S. W. and R. D. Brown (2021). "Urban heat island (UHI) variations within a city boundary: A systematic literature review". *Renewable and Sustainable Energy Reviews* 148, 111256.

Kirchner, J. W. (2009). "Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward". *Water Resources Research* 45 (2).

Klaassen, W., F. Bosveld, and E. De Water (1998). "Water storage and evaporation as constituents of rainfall interception". *Journal of Hydrology* 212, 36–50.

Kljun, N., P. Calanca, M. Rotach, and H. P. Schmid (2015). "A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP)". *Geoscientific Model Development* 8 (11), 3695–3713.

Kljun, N., P. Calanca, M. Rotach, and H. Schmid (2004). "A simple parameterisation for flux footprint predictions". *Boundary-Layer Meteorology* 112, 503–523.

Kljun, N., M. Rotach, and H. Schmid (2002). "A three-dimensional backward Lagrangian footprint model for a wide range of boundary-layer stratifications". *Boundary-Layer Meteorology* 103, 205–226.

Kokkonen, T., S. Grimmond, A. Christen, T. Oke, and L. Järvi (2018). "Changes to the water balance over a century of urban development in two neighborhoods: Vancouver, Canada". *Water Resources Research* 54 (9), 6625–6642.

Koopmans, S., B. Heusinkveld, and G. Steeneveld (2020). "A standardized Physical Equivalent Temperature urban heat map at 1-m spatial resolution to facilitate climate stress tests in the Netherlands". *Building and Environment* 181, 106984.

Kormann, R. and F. X. Meixner (2001). "An analytical footprint model for non-neutral stratification". *Boundary-Layer Meteorology* 99, 207–224.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006). *World map of the Köppen-Geiger climate classification updated*.

Kotthaus, S. and S. Grimmond (2014a). "Energy exchange in a dense urban environment–Part I: Temporal variability of long-term observations in Central London". *Urban Climate* 10, 261–280.

– (2014b). "Energy exchange in a dense urban environment–Part II: Impact of spatial heterogeneity of the surface". *Urban Climate* 10, 281–307.

Kowalczyk, E., Y. Wang, R. Law, H. Davies, J. McGregor, and G. Abramowitz (2006). "The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model". *CSIRO Marine and Atmospheric Research Paper* 13, 42.

Kraemer, R. and N. Kabisch (2022). "Parks under stress: Air temperature regulation of urban green spaces under conditions of drought and summer heat". *Frontiers in Environmental Science* 10.

Krayenhoff, E. S. and J. A. Voogt (2007). "A microscale three-dimensional urban energy balance model for studying surface temperatures". *Boundary-Layer Meteorology* 123 (3), 433–461.

Kuhlemann, L.-M., D. Tetzlaff, A. Smith, B. Kleinschmit, and C. Soulsby (2020). *Soil Moisture data for grassland, shrub and trees at the Steglitz Urban Ecohydrological Observatory [Data set]*.

Kuhlemann, L.-M., D. Tetzlaff, A. Smith, B. Kleinschmit, and C. Soulsby (2021). "Using soil water isotopes to infer the influence of contrasting urban green space on ecohydrological partitioning". *Hydrology and Earth System Sciences* 25 (2), 927–943.

Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura (2001). "A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models". *Boundary-Layer Meteorology* 101 (3), 329–358.

Laaidi, K., A. Zeghnoun, B. Dousset, P. Bretin, S. Vandentorren, E. Giraudet, and P. Beaudeau (2012). "The impact of heat islands on mortality in Paris during the August 2003 heat wave". *Environmental Health Perspectives* 120 (2), 254–259.

Lanza, K. and B. Stone Jr (2016). "Climate adaptation in cities: What trees are suitable for urban heat management?" *Landscape and Urban Planning* 153, 74–82.

Lavoisier, A. L. (1789). *Traite Elementaire de chimie*. Paris.

Lean, H. W., J. F. Barlow, and C. H. Halios (2019). "The impact of spin-up and resolution on the representation of a clear convective boundary layer over London in order 100 m grid-length versions of the Met Office Unified Model". *Quarterly Journal of the Royal Meteorological Society* 145 (721), 1674–1689.

Lean, H. W., N. E. Theeuwes, M. Baldauf, J. Barkmeijer, G. Bessardon, L. Blunn, J. Bojarova, I. A. Boutle, P. A. Clark, M. Demuzere, et al. (2024). "The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models". *Quarterly Journal of the Royal Meteorological Society*.

Leclerc, M. Y., S. Shen, and B. Lamb (1997). "Observations and large-eddy simulation modeling of footprints in the lower convective boundary layer". *Journal of Geophysical Research: Atmospheres* 102 (D8), 9323–9334.

Lemonsu, A., V. Viguie, M. Daniel, and V. Masson (2015). "Vulnerability to heat waves: Impact of urban expansion scenarios on urban heat island and heat stress in Paris (France)". *Urban Climate* 14, 586–605.

Leopold, L. B. (1968). *Hydrology for urban land planning: A guidebook on the hydrologic effects of urban land use*. Vol. 554. US Geological Survey.

Letzel, M. O., C. Helmke, E. Ng, X. An, A. Lai, and S. Raasch (2012). "LES case study on pedestrian level ventilation in two neighbourhoods in Hong Kong". *Meteorologische Zeitschrift* 21 (6), 575–589.

Li, D., W. Liao, A. J. Rigden, X. Liu, D. Wang, S. Malyshev, and E. Shevliakova (2019a). "Urban heat island: Aerodynamics or imperviousness?" *Science Advances* 5 (4), eaau4299.

Li, Q., E. Bou-Zeid, S. Grimmond, S. Zilitinkevich, and G. Katul (2020). "Revisiting the relation between momentum and scalar roughness lengths of urban surfaces". *Quarterly Journal of the Royal Meteorological Society* 146 (732), 3144–3164.

Li, X., Y. Zhou, S. Yu, G. Jia, H. Li, and W. Li (2019b). "Urban heat island impacts on building energy consumption: A review of approaches and findings". *Energy* 174, 407–419.

Lietzke, B., R. Vogt, C. Feigenwinter, and E. Parlow (2015). "On the controlling factors for the variability of carbon dioxide flux in a heterogeneous urban environment". *International Journal of Climatology* 35 (13), 3921–3941.

Lin, D., B. Khan, M. Katurji, L. Bird, R. Faria, and L. E. Revell (2021). "WRF4PALM v1. 0: A mesoscale dynamical driver for the microscale PALM model system 6.0". *Geoscientific Model Development* 14 (5), 2503–2524.

Lipson, M. J., S. Grimmond, M. Best, G. Abramowitz, A. Coutts, N. Tapper, J.-J. Baik, M. Beyers, L. Blunn, S. Boussetta, E. Bou-Zeid, M. D. Kauwe, C. de Munck, M. Demuzere,

S. Fatichi, K. Fortuniak, B.-S. Han, M. Hendry, Y. Kikegawa, H. Kondo, D.-I. Lee, S.-H. Lee, A. Lemonsu, T. Machado, G. Manoli, A. Martilli, V. Masson, J. McNorton, N. Meili, D. Meyer, K. Nice, K. Oleson, S.-B. Park, M. Roth, R. Schoette, A. Simon, G.-J. Steeneveld, T. Sun, Y. Takane, M. Thatcher, A. Tsiringakis, M. Varentsov, C. Wang, and Z.-H. Wang (2023). "The Urban-PLUMBER model evaluation project: Phase 1 results". In: 11^{th} International Conference on Urban Climate. Sydney, Australia.

Lipson, M. J., S. Grimmond, M. Best, G. Abramowitz, A. Coutts, N. Tapper, J.-J. Baik, M. Beyers, L. Blunn, S. Boussetta, E. bou-Zeid, M. G. D. Kauwe, C. de Munck, M. Demuzere, S. Fatichi, K. Fortuniak, B.-S. Han, M. Hendry, Y. Kikegawa, H. Kondo, D.-I. Lee, S.-H. Lee, A. Lemonsu, T. Machado, G. Manoli, A. Martilli, V. Masson, J. McNorton, N. Meili, D. Meyer, K. A. Nice, K. W. Oleson, S.-B. Park, M. Roth, R. Schoetter, A. Simon, G.-J. Steeneveld, T. Sun, Y. Takane, M. Thatche, A. Tsiringakis, M. Varentsov, C. Wang, Z.-H. Wang, and A. J. Pitman (2024). "Evaluation of 30 urban land surface models in the Urban-PLUMBER project: Phase 1 results". *Quarterly Journal of the Royal Meteorological Society* 150 (758), 126–169.

Lipson, M. J., S. Grimmond, M. Best, W. T. Chow, A. Christen, N. Chrysoulakis, A. Coutts, B. Crawford, S. Earl, J. Evans, et al. (2022a). *Site data archive for "Harmonized gap-filled dataset from 20 urban flux tower sites" for the Urban-PLUMBER project (Version v1) [Data set].*

Lipson, M. J., S. Grimmond, M. J. Best, W. T. Chow, A. Christen, N. Chrysoulakis, A. Coutts, B. Crawford, S. Earl, J. Evans, et al. (2022b). "Harmonized gap-filled datasets from 20 urban flux tower sites". *Earth System Science Data Discussions*, 1–29.

Lipson, M. J., M. Thatcher, M. A. Hart, and A. Pitman (2018). "A building energy demand and urban land surface model". *Quarterly Journal of the Royal Meteorological Society* 144 (714), 1572–1590.

Lirola, J. M., E. Castañeda, B. Lauret, and M. Khayet (2017). "A review on experimental research using scale models for buildings: Application and methodologies". *Energy and Buildings* 142, 72–110.

Liu, C., J. Liu, W. Shao, J. Lu, and H. Gao (2022). "The considerable water evaporation induced by human perspiration and respiration in megacities: Quantifying method and case study in Beijing". *International Journal of Environmental Research and Public Health* 19 (14), 8638.

Liu, L. and M. B. Jensen (2018). "Green infrastructure for sustainable urban water management: Practices of five forerunner cities". *Cities* 74, 126–133.

Liu, X., X.-X. Li, S. Harshan, M. Roth, and E. Velasco (2017). "Evaluation of an urban canopy model in a tropical city: The role of tree evapotranspiration". *Environmental Research Letters* 12 (9), 094008.

Liu, Y., J. Xiao, W. Ju, K. Xu, Y. Zhou, and Y. Zhao (2016). "Recent trends in vegetation greenness in China significantly altered annual evapotranspiration and water yield". *Environmental Research Letters* 11 (9), 094010.

Logan, T., B. Zaitchik, S. Guikema, and A. Nisbet (2020). "Night and day: The influence and relative importance of urban characteristics on remotely sensed land surface temperature". *Remote Sensing of Environment* 247, 111861.

Lu, J., N. Nazarian, M. A. Hart, E. S. Krayenhoff, and A. Martilli (2024). "Representing the effects of building height variability on urban canopy flow". *Quarterly Journal of the Royal Meteorological Society* 150 (758), 46–67.

MacKay, M. D., G. Meyer, and J. R. Melton (2022). "On the discretization of Richards equation in Canadian land surface models". *Atmosphere-Ocean*, 1–11.

Manabe, S. (1969). "Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface". *Monthly Weather Review* 97 (11), 739–774.

Manoli, G., S. Fatichi, E. Bou-Zeid, and G. G. Katul (2020). "Seasonal hysteresis of surface urban heat islands". *Proceedings of the National Academy of Sciences* 117 (13), 7082–7089.

Maronga, B., S. Banzhaf, C. Burmeister, T. Esch, R. Forkel, D. Fröhlich, V. Fuka, K. F. Gehrke, J. Geletič, S. Giersch, et al. (2020). "Overview of the PALM model system 6.0". *Geoscientific Model Development* 13 (3), 1335–1372.

Maronga, B., M. Gryschka, R. Heinze, F. Hoffmann, F. Kanani-Sühring, M. Keck, K. Ketelsen, M. O. Letzel, M. Sühring, and S. Raasch (2015). "The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: Model formulation, recent developments, and future perspectives". *Geoscientific Model Development* 8 (8), 2515–2551.

Marquès, E., V. Masson, P. Naveau, O. Mestre, V. Dubreuil, and Y. Richard (2022). "Urban heat island estimation from crowdsensing thermometers embedded in personal cars". *Bulletin of the American Meteorological Society* 103 (4), E1098–E1113.

Martinez, G. S., C. Linares, A. Ayuso, V. Kendrovski, M. Boeckmann, and J. Diaz (2019). "Heat-health action plans in Europe: Challenges ahead and how to tackle them". *Environmental Research* 176, 108548.

Masoner, J. R. and D. I. Stannard (2010). "A comparison of methods for estimating open-water evaporation in small wetlands". *Wetlands* 30 (3), 513–524.

Masson, V. (2000). "A physically-based scheme for the urban energy budget in atmospheric models". *Boundary-Layer Meteorology* 94 (3), 357–397.

Masson, V., L. Gomes, G. Pigeon, C. Liousse, V. Pont, J.-P. Lagouarde, J. Voogt, J. Salmond, T. R. Oke, J. Hidalgo, et al. (2008). "The Canopy and Aerosol Particles Interactions in TOulouse Urban Layer (CAPITOUL) experiment". *Meteorology and Atmospheric Physics* 102, 135–157.

Mauder, M., M. Cuntz, C. Drüe, A. Graf, C. Rebmann, H. P. Schmid, M. Schmidt, and R. Steinbrecher (2013). "A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements". *Agricultural and Forest Meteorology* 169, 122–135.

Mauder, M. and T. Foken (2006). "Impact of post-field data processing on eddy covariance flux estimates and energy balance closure". *Meteorologische Zeitschrift* 15 (6), 597–610.

Mauder, M., T. Foken, M. Aubinet, and A. Ibrom (2021). "Eddy-covariance measurements". In: *Springer handbook of atmospheric measurements*. Springer, 1473–1504.

Mauder, M., T. Foken, and J. Cuxart (2020). "Surface-energy-balance closure over land: A review". *Boundary-Layer Meteorology* 177 (2), 395–426.

Mauder, M., C. Liebethal, M. Göckede, J.-P. Leps, F. Beyrich, and T. Foken (2006). "Processing and quality control of flux data during LITFASS-2003". *Boundary-Layer Meteorology* 121, 67–88.

Mauritsen, T., R. Redler, M. Esch, B. Stevens, C. Hohenegger, D. Klocke, R. Brokopf, H. Haak, L. Linardakis, N. Röber, et al. (2022). "Early development and tuning of a global coupled cloud resolving model, and its fast response to increasing CO_2". *Tellus Series A: Dynamic Meteorology and Oceanography* 74, 346–363.

McColl, K. A., W. Wang, B. Peng, R. Akbar, D. J. Short Gianotti, H. Lu, M. Pan, and D. Entekhabi (2017). "Global characterization of surface soil moisture drydowns". *Geophysical Research Letters* 44 (8), 3682–3690.

McDonnell, J. J. (2013). "Are all runoff processes the same?" *Hydrological Processes* 27 (26), 4103–4111.

McDowell, N., C. D. Allen, K. Anderson-Teixeira, P. Brando, R. Brienen, J. Chambers, B. Christoffersen, S. Davies, C. Doughty, A. Duque, et al. (2018). "Drivers and mechanisms of tree mortality in moist tropical forests". *New Phytologist* 219 (3), 851–869.

McGrane, S. J. (2016). "Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: A review". *Hydrological Sciences Journal* 61 (13), 2295–2311.

McMahon, T., B. Finlayson, and M. Peel (2016). "Historical developments of models for estimating evaporation using standard meteorological data". *Wiley Interdisciplinary Reviews: Water* 3 (6), 788–818.

McNorton, J. and G. Balsamo (2023). *An urban scheme for the IFS*. Date accessed: 9 October 2024.

McNorton, J., G. Arduini, N. Bousserez, A. Agustí-Panareda, G. Balsamo, S. Boussetta, M. Choulga, I. Hadade, and R. Hogan (2021). "An urban scheme for the ECMWF Integrated Forecasting System: Single-column and global offline application". *Journal of Advances in Modeling Earth Systems*, e2020MS002375.

Meier, F., D. Fenner, T. Grassmann, M. Otto, and D. Scherer (2017). "Crowdsourcing air temperature from citizen weather stations for urban climate research". *Urban Climate* 19, 170–191.

Meier, F. and D. Scherer (2012). "Spatial and temporal variability of urban tree canopy temperature during summer 2010 in Berlin, Germany". *Theoretical and Applied Climatology* 110, 373–384.

Meili, N., G. Manoli, P. Burlando, E. Bou-Zeid, W. T. Chow, A. M. Coutts, E. Daly, K. A. Nice, M. Roth, N. J. Tapper, et al. (2020). "An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (UT&C v1. 0)". *Geoscientific Model Development* 13 (1), 335–362.

Meili, N., G. Manoli, P. Burlando, J. Carmeliet, W. T. Chow, A. M. Coutts, M. Roth, E. Velasco, E. R. Vivoni, and S. Fatichi (2021). "Tree effects on urban microclimate: Diurnal, seasonal,

and climatic temperature differences explained by separating radiation, evapotranspiration, and roughness effects". *Urban Forestry & Urban Greening* 58, 126970.

Menard, C. B., R. Essery, G. Krinner, G. Arduini, P. Bartlett, A. Boone, C. Brutel-Vuilmet, E. Burke, M. Cuntz, Y. Dai, et al. (2020). "Scientific and human errors in a snow model intercomparison". *Bulletin of the American Meteorological Society*, 1–46.

Menzer, O. and J. P. McFadden (2017). "Statistical partitioning of a three-year time series of direct urban net CO_2 flux measurements into biogenic and anthropogenic components". *Atmospheric Environment* 170, 319–333.

Miao, S. and F. Chen (2014). "Enhanced modeling of latent heat flux from urban surfaces in the Noah/single-layer urban canopy coupled model". *Science China Earth Sciences* 57 (10), 2408–2416.

Mignot, E., X. Li, and B. Dewals (2019). "Experimental modelling of urban flooding: A review". *Journal of Hydrology* 568, 334–342.

Miralles, D. G., T. Holmes, R. De Jeu, J. Gash, A. Meesters, and A. Dolman (2011). "Global land-surface evaporation estimated from satellite-based observations". *Hydrology and Earth System Sciences* 15 (2), 453–469.

Mitchell, V. G., R. G. Mein, and T. A. McMahon (2001). "Modelling the urban water cycle". *Environmental Modelling & Software* 16 (7), 615–629.

Monteith, J. L. (1965). "Evaporation and environment". In: *Symposia of the Society for Experimental Biology*. Vol. 19. Cambridge University Press (CUP) Cambridge, 205–234.

Morin, E., Y. Enzel, U. Shamir, and R. Garti (2001). "The characteristic time scale for basin hydrological response using radar data". *Journal of Hydrology* 252 (1-4), 85–99.

Moriwaki, R., M. Kanda, H. Senoo, A. Hagishima, and T. Kinouchi (2008). "Anthropogenic water vapor emissions in Tokyo". *Water Resources Research* 44 (11).

Mu, Q., M. Zhao, and S. W. Running (2011). "Improvements to a MODIS global terrestrial evapotranspiration algorithm". *Remote Sensing of Environment* 115 (8), 1781–1800.

Mukherjee, S. and A. K. Mishra (2021). "Increase in compound drought and heatwaves in a warming world". *Geophysical Research Letters* 48 (1), e2020GL090617.

Muller, C. L., L. Chapman, S. Grimmond, D. T. Young, and X.-M. Cai (2013). "Toward a standardized metadata protocol for urban meteorological networks". *Bulletin of the American Meteorological Society* 94 (8), 1161–1185.

Muñoz-Sabater, J., E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, et al. (2021). "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications". *Earth System Science Data* 13 (9), 4349–4383.

Nachtergaele, F. (2001). "Soil taxonomy—a basic system of soil classification for making and interpreting soil surveys". *Geoderma* 99 (3-4), 336–337.

Narumi, D., A. Kondo, and Y. Shimoda (2009). "Effects of anthropogenic heat release upon the urban climate in a Japanese megacity". *Environmental Research* 109 (4), 421–431.

Nicolini, G., G. Antoniella, F. Carotenuto, A. Christen, P. Ciais, C. Feigenwinter, B. Gioli, S. Stagakis, E. Velasco, R. Vogt, et al. (2022). "Direct observations of CO_2 emission reductions due to COVID-19 lockdown across European urban districts". *Science of the Total Environment* 830, 154662.

Nimac, I., I. Herceg-Bulić, M. Žuvela-Aloise, and M. Žgela (2022). "Impact of North Atlantic Oscillation and drought conditions on summer urban heat load-a case study for Zagreb". *International Journal of Climatology* 42 (9), 4850–4867.

Nordbo, A., L. Järvi, S. Haapanala, J. Moilanen, and T. Vesala (2013). "Intra-city variation in urban morphology and turbulence structure in Helsinki, Finland". *Boundary-Layer Meteorology* 146, 469–496.

Nordbo, A., S. Launiainen, I. Mammarella, M. Leppäranta, J. Huotari, A. Ojala, and T. Vesala (2011). "Long-term energy flux measurements and energy balance over a small boreal lake using eddy covariance technique". *Journal of Geophysical Research: Atmospheres* 116 (D2).

Norton, B. A., A. M. Coutts, S. J. Livesley, R. J. Harris, A. M. Hunter, and N. S. Williams (2015). "Planning for cooler cities: A framework to prioritise green infrastructure to mitigate high temperatures in urban landscapes". *Landscape and Urban Planning* 134, 127–138.

NRCS (2004a). *Estimation of Direct Runoff from Storm Rainfall*. in Part 630 Hydrology National Engineering Handbook.

– (2004b). *Hydrologic Soil-Cover Complexes*. in Part 630 Hydrology National Engineering Handbook.

Oke, T. R. (1982). "The energetic basis of the urban heat island". *Quarterly Journal of the Royal Meteorological Society* 108 (455), 1–24.

Oke, T. R., G. Mills, A. Christen, and J. A. Voogt (2017). *Urban climates*. Cambridge university press.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, and S. Grimmond (2008). "An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities". *Journal of Applied Meteorology and Climatology* 47 (4), 1038–1060.

Oleson, K. W. and J. Feddema (2020). "Parameterization and surface data improvements and new capabilities for the Community Land Model Urban (CLMU)". *Journal of Advances in Modeling Earth Systems* 12 (2), e2018MS001586.

Oleson, K. W., A. Monaghan, O. Wilhelmi, M. Barlage, N. Brunsell, J. Feddema, L. Hu, and D. Steinhoff (2015). "Interactions between urbanization, heat stress, and climate change". *Climatic Change* 129, 525–541.

Oliver, M. A. (2009). "The variogram and kriging". In: *Handbook of applied spatial analysis: Software tools, methods and applications*. Springer, 319–352.

Ostos, E. J. and E. Luyando (1998). "Long-term association between pan evaporation and the urban heat island in Mexico City". *Atmósfera* 11 (1), 45–60.

Oswald, C. J., C. Kelleher, S. H. Ledford, K. G. Hopkins, A. Sytsma, D. Tetzlaff, L. Toran, and C. Voter (2023). "Integrating urban water fluxes and moving beyond impervious surface cover: A review". *Journal of Hydrology*, 129188.

Oudin, L., B. Salavati, C. Furusho-Percot, P. Ribstein, and M. Saadi (2018). "Hydrological impacts of urbanization at the catchment scale". *Journal of Hydrology* 559, 774–786.

Ouédraogo, A. A., E. Berthier, B. Durand, and M.-C. Gromaire (2022). "Determinants of evapotranspiration in urban rain gardens: A case study with lysimeters under temperate climate". *Hydrology* 9 (3), 42.

Pataki, D. E., H. R. McCarthy, E. Litvak, and S. Pincetl (2011). "Transpiration of urban forests in the Los Angeles metropolitan area". *Ecological Applications* 21 (3), 661–677.

Paul, M. and J. Meyer (2001). "Streams in the urban landscape". *Annual Review of Ecology and Systematics* 32, 333–365.

Paul-Limoges, E., S. Wolf, F. D. Schneider, M. Longo, P. Moorcroft, M. Gharun, and A. Damm (2020). "Partitioning evapotranspiration with concurrent eddy covariance measurements in a mixed forest". *Agricultural and Forest Meteorology* 280, 107786.

Pawlak, W., K. Fortuniak, and M. Siedlecki (2011). "Carbon dioxide flux in the centre of Łódź, Poland - analysis of a 2-year eddy covariance measurement data set". *International Journal of Climatology* 31 (2), 232–243.

Payero, J. O. and S. Irmak (2008). "Construction, installation, and performance of two repacked weighing lysimeters". *Irrigation Science* 26, 191–202.

Pearlmutter, D., E. Kruger, and P. Berliner (2009). "The role of evaporation in the energy balance of an open-air scaled urban surface". *International Journal of Climatology* 29 (6), 911.

Penman, H. (1956). "Evaporation: An introductory survey." *Netherlands Journal of Agricultural Science* 4 (1), 9–29.

Peters, E. B., R. V. Hiller, and J. P. McFadden (2011). "Seasonal contributions of vegetation types to suburban evapotranspiration". *Journal of Geophysical Research: Biogeosciences* 116 (G1).

Petrucci, R. H., F. G. Herring, and J. D. Madura (2010). *General chemistry: Principles and modern applications*. Pearson Prentice Hall.

Porson, A., P. A. Clark, I. Harman, M. J. Best, and S. Belcher (2010). "Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations". *Quarterly Journal of the Royal Meteorological Society* 136 (651), 1514–1529.

Powers, J. G., J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, D. J. Gochis, R. Ahmadov, S. E. Peckham, et al. (2017). "The Weather Research and Forecasting model: Overview, system efforts, and future directions". *Bulletin of the American Meteorological Society* 98 (8), 1717–1737.

Price, K. (2011). "Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review". *Progress in Physical Geography* 35 (4), 465–492.

Qin, H.-p., Z.-x. Li, and G. Fu (2013). "The effects of low impact development on urban flooding under different rainfall characteristics". *Journal of Environmental Management* 129, 577–585.

Qin, L., C. Yan, L. Yu, M. Chai, B. Wang, M. Hayat, Z. Shi, H. Gao, X. Jiang, B. Xiong, et al. (2022). "High-resolution spatio-temporal characteristics of urban evapotranspiration measured

by unmanned aerial vehicle and infrared remote sensing". *Building and Environment* 222, 109389.

Qiu, G. Y., C. Yan, and Y. Liu (2023). "Urban evapotranspiration and its effects on water budget and energy balance: Review and perspectives". *Earth-Science Reviews*, 104577.

Qiu, G.-y., H.-y. Li, Q.-t. Zhang, C. Wan, X.-j. Liang, and X.-z. Li (2013). "Effects of evapotranspiration on mitigation of urban temperature by vegetation and urban agriculture". *Journal of Integrative Agriculture* 12 (8), 1307–1315.

Ramamurthy, P. and E. Bou-Zeid (2014). "Contribution of impervious surfaces to urban evaporation". *Water Resources Research* 50 (4), 2889–2902.

Ramírez, B. H., L. A. Melsen, L. Ganzeveld, R. Leemans, and A. J. Teuling (2018). "Tropical Montane Cloud Forests in the Orinoco River basin: Inferring fog interception from through-fall dynamics". *Agricultural and Forest Meteorology* 260, 17–30.

Randrup, T. B., E. G. McPherson, and L. R. Costello (2001). "Tree root intrusion in sewer systems: A review of extent and costs". *Journal of Infrastructure Systems 7: 26-31* 7, 26–31.

Rannik, Ü., A. Sogachev, T. Foken, M. Göckede, N. Kljun, M. Y. Leclerc, and T. Vesala (2012). "Footprint analysis". In: *Eddy covariance: A practical guide to measurement and data analysis*. Ed. by M. Aubinet, T. Vesala, and D. Papale. Dordrecht: Springer Netherlands, 211–261.

Rawls, W. J., A. Shalaby, and R. H. McCuen (1981). "Evaluation of methods for determining urban runoff curve numbers". *Transactions of the ASAE* 24 (6), 1562–1566.

Reinmann, A. B. and L. R. Hutyra (2017). "Edge effects enhance carbon uptake and its vulnerability to climate change in temperate broadleaf forests". *Proceedings of the National Academy of Sciences* 114 (1), 107–112.

Robine, J.-M., S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J.-P. Michel, and F. R. Herrmann (2008). "Death toll exceeded 70,000 in Europe during the summer of 2003". *Comptes Rendus Biologies* 331 (2), 171–178.

Rocha, A. D., S. Vulova, M. Förster, B. Gioli, B. Matthews, C. Helfter, F. Meier, G.-J. Steeneveld, J. F. Barlow, L. Järvi, et al. (2024). "Unprivileged groups are less served by green cooling services in major European urban areas". *Nature Cities*, 1–12.

Rocha, A. D., S. Vulova, F. Meier, M. Förster, and B. Kleinschmit (2022). "Mapping evapotranspirative and radiative cooling services in an urban environment". *Sustainable Cities and Society* 85, 104051.

Ronda, R., G. Steeneveld, B. Heusinkveld, J. Attema, and A. Holtslag (2017). "Urban finescale forecasting reveals weather conditions with unprecedented detail". *Bulletin of the American Meteorological Society* 98 (12), 2675–2688.

Rondinelli, W. J., B. K. Hornbuckle, J. C. Patton, M. H. Cosh, V. A. Walker, B. D. Carr, and S. D. Logsdon (2015). "Different rates of soil drying after rainfall are observed by the SMOS satellite and the South Fork in situ soil moisture network". *Journal of Hydrometeorology* 16 (2), 889–903.

Rosekind, M. R., K. B. Gregory, M. M. Mallis, S. L. Brandt, B. Seal, and D. Lerner (2010). "The cost of poor sleep: Workplace productivity loss and associated costs". *Journal of Occupational and Environmental Medicine* 52 (1), 91–98.

Ross, S. L. and T. Oke (1988). "Tests of three urban energy balance models". *Boundary-Layer Meteorology* 44 (1), 73–96.

Roth, M., C. Jansson, and E. Velasco (2017). "Multi-year energy balance and carbon dioxide fluxes over a residential neighbourhood in a tropical city". *International Journal of Climatology* 37 (5), 2679–2698.

Rutter, A., A. Morton, and P. Robins (1975). "A predictive model of rainfall interception in forests. II. Generalization of the model and comparison with observations in some coniferous and hardwood stands". *Journal of Applied Ecology*, 367–380.

Ryu, Y.-H., J.-J. Baik, and S.-H. Lee (2011). "A new single-layer urban canopy model for use in mesoscale atmospheric models". *Journal of Applied Meteorology and Climatology* 50 (9), 1773–1794.

Ryzhkov, A., M. Pinsky, A. Pokrovsky, and A. Khain (2011). "Polarimetric radar observation operator for a cloud model with spectral microphysics". *Journal of Applied Meteorology and Climatology* 50 (4), 873–894.

Sadegh, M., A. AghaKouchak, A. Flores, I. Mallakpour, and M. R. Nikoo (2019). "A multi-model nonstationary rainfall-runoff modeling framework: Analysis and toolbox". *Water Resources Management* 33 (9), 3011–3024.

Sailor, D. J. (2011). "A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment". *International Journal of Climatology* 31 (2), 189–199.

Sailor, D. J., M. Georgescu, J. M. Milne, and M. A. Hart (2015). "Development of a national anthropogenic heating database with an extrapolation for international cities". *Atmospheric Environment* 118, 7–18.

Saleem, J. A. and G. D. Salvucci (2002). "Comparison of soil wetness indices for inducing functional similarity of hydrologic response across sites in Illinois". *Journal of Hydrometeorology* 3 (1), 80–91.

Salmond, J., M. Roth, T. Oke, A. Christen, and J. Voogt (2012). "Can surface-cover tiles be summed to give neighborhood fluxes in cities?" *Journal of Applied Meteorology and Climatology* 51 (1), 133–149.

Salvucci, G. D. (2001). "Estimating the moisture dependence of root zone water loss using conditionally averaged precipitation". *Water Resources Research* 37 (5), 1357–1365.

Santamouris, M., C. Cartalis, A. Synnefa, and D. Kolokotsa (2015). "On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings — A review". *Energy and Buildings* 98, 119–124.

Santamouris, M. (2014). "Cooling the cities - a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments". *Solar energy* 103, 682–703.

Sarangi, C., Y. Qian, J. Li, L. R. Leung, T. Chakraborty, and Y. Liu (2021). "Urbanization amplifies nighttime heat stress on warmer days over the US". *Geophysical Research Letters* 48 (24), e2021GL095678.

Saxton, K., W. Rawls, J. S. Romberger, and R. Papendick (1986). "Estimating generalized soil-water characteristics from texture". *Soil Science Society of America Journal* 50 (4), 1031–1036.

Scherer, D., F. Ament, S. Emeis, U. Fehrenbach, B. Leitl, K. Scherber, C. Schneider, and U. Vogt (2019). "Three-dimensional observation of atmospheric processes in cities". *Meteorologische Zeitschrift* 28 (2), 121–138.

Schifman, L. A., D. L. Herrmann, W. D. Shuster, A. Ossola, A. Garmestani, and M. E. Hopton (2017). "Situating green infrastructure in context: A framework for adaptive socio-hydrology in cities". *Water Resources Research* 53 (12), 10139–10154.

Schilperoort, B., M. Coenders-Gerrits, W. Luxemburg, C. Jiménez Rodríguez, C. Cisneros Vaca, and H. Savenije (2018). "Using distributed temperature sensing for Bowen ratio evaporation measurements". *Hydrology and Earth System Sciences* 22 (1), 819–830.

Schmid, H. and T. Oke (1990). "A model to estimate the source area contributing to turbulent exchange in the surface layer over patchy terrain". *Quarterly Journal of the Royal Meteorological Society* 116 (494), 965–988.

Schmutz, M., R. Vogt, C. Feigenwinter, and E. Parlow (2016). "Ten years of eddy covariance measurements in Basel, Switzerland: Seasonal and interannual variabilities of urban CO_2 mole fraction and flux". *Journal of Geophysical Research: Atmospheres* 121 (14), 8649–8667.

Schuepp, P., M. Leclerc, J. MacPherson, and R. Desjardins (1990). "Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation". *Boundary-Layer Meteorology* 50 (1-4), 355–373.

Schulz, J.-P. and G. Vogel (2020). "Improving the processes in the land surface scheme TERRA: Bare soil evaporation and skin temperature". *Atmosphere* 11 (5), 513.

Šećerov, I. B., S. M. Savić, D. D. Milošević, D. M. Arsenović, D. M. Dolinaj, and S. B. Popov (2019). "Progressing urban climate research using a high-density monitoring network system". *Environmental Monitoring and Assessment* 191, 1–19.

Senate Department for the Environment, Mobility, Consumer and Climate Protection Berlin (2014). *Straßenbefahrung 2014 - [WMS]*. Date accessed: 21 September 2023.

Senate Department for Urban Development, Building and Housing (2012). *Building and Vegetation Heights 2012*. Date accessed: 21 September 2023.

– (2013). *Biotope Types 2013*. Date accessed: 21 September 2023.

Shellito, P. J., E. E. Small, A. Colliander, R. Bindlish, M. H. Cosh, A. A. Berg, D. D. Bosch, T. G. Caldwell, D. C. Goodrich, H. McNairn, et al. (2016). "SMAP soil moisture drying more rapid than observed in situ following rainfall events". *Geophysical Research Letters* 43 (15), 8068–8075.

Shuster, W. D., J. Bonta, H. Thurston, E. Warnemuende, and D. Smith (2005). "Impacts of impervious surface on watershed hydrology: A review". *Urban Water Journal* 2 (4), 263–275.

Shuster, W. D., S. D. Dadio, C. E. Burkman, S. R. Earl, and S. J. Hall (2015). "Hydropedological assessments of parcel-level infiltration in an arid urban ecosystem". *Soil Science Society of America Journal* 79 (2), 398–406.

Silva, C., R. Vilaça, A. Pereira, and R. Bessa (2024). "A review on the decarbonization of high-performance computing centers". *Renewable and Sustainable Energy Reviews* 189, 114019.

Smith, I. A., L. R. Hutyra, A. B. Reinmann, J. K. Marrs, and J. R. Thompson (2018). "Piecing together the fragments: Elucidating edge effects on forest carbon dynamics". *Frontiers in Ecology and the Environment* 16 (4), 213–221.

Sobolewski, A., M. Młynarczyk, M. Konarska, and J. Bugajska (2021). "The influence of air humidity on human heat stress in a hot environment". *International Journal of Occupational Safety and Ergonomics* 27 (1), 226–236.

Solcerova, A., F. Van de Ven, and N. Van de Giesen (2019). "Nighttime cooling of an urban pond". *Frontiers in Earth Science* 7, 156.

Somarakis, G., S. Stagakis, N. Chrysoulakis, M. Mesimäki, and S. Lehvävirta (2019). "ThinkNature nature-based solutions handbook".

Song, K., X. Liu, and T. Gao (2021). "Potential application of using smartphone sensor for estimating air temperature: Experimental study". *IEEE Internet of Things Journal* 9 (16), 14300–14306.

Stagakis, S., N. Chrysoulakis, N. Spyridakis, C. Feigenwinter, and R. Vogt (2019). "Eddy Covariance measurements and source partitioning of CO_2 emissions in an urban environment: Application for Heraklion, Greece". *Atmospheric Environment* 201, 278–292.

Starke, P., P. Göbel, and W. Coldewey (2010). "Urban evaporation rates for water-permeable pavements". *Water Science and Technology* 62 (5), 1161–1169.

Steeneveld, G. J., S. Koopmans, B. G. Heusinkveld, and N. E. Theeuwes (2014). "Refreshing the role of open water surfaces on mitigating the maximum urban heat island effect". *Landscape and Urban Planning* 121, 92–96.

Steeneveld, G.-J., S. van der Horst, and B. Heusinkveld (2020). "Observing the surface radiation and energy balance, carbon dioxide and methane fluxes over the city centre of Amsterdam". In: *EGU General Assembly Conference Abstracts*, 1547.

Stewart, I. D. (2011). "A systematic review and scientific critique of methodology in modern urban heat island literature". *International Journal of Climatology* 31 (2), 200–217.

Stewart, I. D. and T. R. Oke (2012). "Local climate zones for urban temperature studies". *Bulletin of the American Meteorological Society* 93 (12), 1879–1900.

Stewart, R. D., A. S. Bhaskar, A. J. Parolari, D. L. Herrmann, J. Jian, L. A. Schifman, and W. D. Shuster (2019). "An analytical approach to ascertain saturation-excess versus infiltration-excess overland flow in urban and reference landscapes". *Hydrological Processes* 33 (26), 3349–3363.

Sumner, D. M. and J. M. Jacobs (2005). "Utility of Penman–Monteith, Priestley–Taylor, reference evapotranspiration, and pan evaporation methods to estimate pasture evapotranspiration". *Journal of Hydrology* 308 (1-4), 81–104.

Sun, R., Y. Wang, and L. Chen (2018). "A distributed model for quantifying temporal-spatial patterns of anthropogenic heat based on energy consumption". *Journal of Cleaner Production* 170, 601–609.

Sützl, B. S., G. G. Rooney, A. Finnenkoetter, S. I. Bohnenstengel, S. Grimmond, and M. van Reeuwijk (2021). "Distributed urban drag parametrization for sub-kilometre scale numerical weather prediction". *Quarterly Journal of the Royal Meteorological Society* 147 (741), 3940–3956.

Syamili, M., T. Takala, A. Korrensalo, and E.-S. Tuittila (2023). "Happiness in urban green spaces: A systematic literature review". *Urban Forestry & Urban Greening*, 128042.

Taha, H. (1997). "Urban climates and heat islands: Albedo, evapotranspiration, and anthropogenic heat". *Energy and Buildings* 25 (2), 99–103.

Templeton, N. P., E. R. Vivoni, Z.-H. Wang, and A. P. Schreiner-McGraw (2018). "Quantifying water and energy fluxes over different urban land covers in Phoenix, Arizona". *Journal of Geophysical Research: Atmospheres* 123 (4), 2111–2128.

Teuling, A. J., S. I. Seneviratne, R. Stöckli, M. Reichstein, E. Moors, P. Ciais, S. Luyssaert, B. Van Den Hurk, C. Ammann, C. Bernhofer, et al. (2010). "Contrasting response of European forest and grassland energy exchange to heatwaves". *Nature Geoscience* 3 (10), 722–727.

Teuling, A. J., S. I. Seneviratne, C. Williams, and P. Troch (2006). "Observed timescales of evapotranspiration response to soil moisture". *Geophysical Research Letters* 33 (23).

Tewari, M., F. Chen, H. Kusaka, and S. Miao (2007). "Coupled WRF/Unified Noah/urban-canopy modeling system". In: *NCAR WRF Documentation, NCAR, Boulder*. Vol. 122. Citeseer, 1–22.

Thatcher, M. and P. Hurley (2012). "Simulating Australian urban climate in a mesoscale atmospheric numerical model". *Boundary-Layer Meteorology* 142 (1), 149–175.

Theeuwes, N. E., A. Solcerova, and G. J. Steeneveld (2013). "Modeling the influence of open water surfaces on the summertime temperature and thermal comfort in the city". *Journal of Geophysical Research: Atmospheres* 118 (16), 8881–8896.

Theeuwes, N. E., G.-J. Steeneveld, R. J. Ronda, and A. A. Holtslag (2017). "A diagnostic equation for the daily maximum urban heat island effect for cities in Northwestern Europe". *International Journal of Climatology* 37 (1), 443–454.

Thommen, O. (2005). "Heat wave 2003 and mortality in Switzerland". *Swiss Medical Weekly* 135 (1314), 200–205.

Tingsanchali, T. (2012). "Urban flood disaster management". *Procedia Engineering* 32, 25–37.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl (2009). "Earth's global energy budget". *Bulletin of the American Meteorological Society* 90 (3), 311–324.

Troch, P. A., A. Berne, P. Bogaart, C. Harman, A. G. Hilberts, S. W. Lyon, C. Paniconi, V. R. Pauwels, D. E. Rupp, J. S. Selker, et al. (2013). "The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange". *Water Resources Research* 49 (9), 5099–5116.

Twine, T. E., W. Kustas, J. Norman, D. Cook, P. Houser, T. Meyers, J. Prueger, P. Starks, and M. Wesely (2000). "Correcting eddy-covariance flux underestimates over a grassland". *Agricultural and Forest Meteorology* 103 (3), 279–300.

Tyree, M. T., H. Cochard, P. Cruiziat, B. Sinclair, and T. Ameglio (1993). "Drought-induced leaf shedding in walnut: Evidence for vulnerability segmentation". *Plant, Cell & Environment* 16 (7), 879–882.

Ulpiani, G., M. A. Hart, G. Di Virgilio, A. M. Maharaj, M. J. Lipson, and J. Potgieter (2022). "A citizen centred urban network for weather and air quality in Australian schools". *Scientific Data* 9 (1), 129.

United Nations (2018). *World urbanization prospects, the 2018 revision*.

Valente, F., J. David, and J. Gash (1997). "Modelling interception loss for two sparse eucalypt and pine forests in Central Portugal using reformulated Rutter and Gash analytical models". *Journal of Hydrology* 190 (1-2), 141–162.

Valeo, C. and S. Moin (2000). "Variable source area modelling in urbanizing watersheds". *Journal of Hydrology* 228 (1-2), 68–81.

Van de Vijver, E., N. Delbecque, A. Verdoodt, and P. Seuntjens (2020). "Estimating the urban soil information gap using exhaustive land cover data: The example of Flanders, Belgium". *Geoderma* 372, 114371.

Van Dijke, A. J. H., R. Orth, A. J. Teuling, M. Herold, M. Schlerf, M. Migliavacca, M. Machwitz, T. C. van Hateren, X. Yu, and K. Mallick (2023). "Comparing forest and grassland drought responses inferred from eddy covariance and Earth observation". *Agricultural and Forest Meteorology* 341, 109635.

Van Emmerik, T., A. Rimmer, Y. Lechinsky, K. Wenker, S. Nussboim, and N. Van de Giesen (2013). "Measuring heat balance residual at lake surface using Distributed Temperature Sensing". *Limnology and Oceanography: Methods* 11 (2), 79–90.

Van Hal, N. (2023). "Quantifying the water balance of Amsterdam". MA thesis. Wageningen University and Research.

Varghese, P. (2015). *Building materials*. PHI Learning Pvt. Ltd.

Velasco, E., M. Roth, S. Tan, M. Quak, S. Nabarro, and L. Norford (2013). "The role of vegetation in the CO_2 flux from a tropical urban neighbourhood". *Atmospheric Chemistry and Physics*.

Velasco, E., R. Perrusquia, E. Jiménez, F. Hernández, P. Camacho, S. Rodríguez, A. Retama, and L. Molina (2014). "Sources and sinks of carbon dioxide in a neighborhood of Mexico City". *Atmospheric Environment* 97, 226–238.

Velasco, E., S. Pressley, R. Grivicke, E. Allwine, T. Coons, W. Foster, B. Jobson, H. Westberg, R. Ramos, F. Hernández, et al. (2009). "Eddy covariance flux measurements of pollutant gases in urban Mexico City". *Atmospheric Chemistry and Physics* 9 (19), 7325–7342.

Velasco, E., S. Pressley, R. Grivicke, E. Allwine, L. T. Molina, and B. Lamb (2011). "Energy balance in urban Mexico City: Observation and parameterization during the MILAGRO/MCMA-2006 field campaign". *Theoretical and Applied Climatology* 103, 501–517.

Velasco, E. and M. Roth (2010). "Cities as net sources of CO_2: Review of atmospheric CO_2 exchange in urban environments measured by eddy covariance technique". *Geography Compass* 4 (9), 1238–1259.

Vesala, T., L. Järvi, S. Launiainen, A. Sogachev, Ü. Rannik, I. Mammarella, E. S. Ivola, P. Keronen, J. Rinne, A. Riikonen, et al. (2008a). "Surface–atmosphere interactions over complex urban terrain in Helsinki, Finland". *Tellus B: Chemical and Physical Meteorology* 60 (2), 188–199.

Vesala, T., N. Kljun, Ü. Rannik, J. Rinne, A. Sogachev, T. Markkanen, K. Sabelfeld, T. Foken, and M. Y. Leclerc (2008b). "Flux and concentration footprint modelling: State of the art". *Environmental Pollution* 152 (3), 653–666.

Vulova, S., L.-M. Kuhlemann, D. Tetzlaff, C. Soulsby, and B. Kleinschmit (2019). "Assessment of evapotranspiration from urban vegetation across space and time: A case study in Berlin". In: *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 1–4.

Vulova, S., F. Meier, A. D. Rocha, J. Quanz, H. Nouri, and B. Kleinschmit (2021). "Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence". *Science of The Total Environment* 786, 147293.

Vulova, S., A. D. Rocha, F. Meier, H. Nouri, C. Schulz, C. Soulsby, D. Tetzlaff, and B. Kleinschmit (2023). "City-wide, high-resolution mapping of evapotranspiration to guide climate-resilient planning". *Remote Sensing of Environment* 287, 113487.

Wald, A. (1943). "Tests of statistical hypotheses concerning several parameters when the number of observations is large". *Transactions of the American Mathematical Society* 54 (3), 426–482.

Walsh, C. J., T. D. Fletcher, and A. R. Ladson (2005a). "Stream restoration in urban catchments through redesigning stormwater systems: Looking to the catchment to save the stream". *Journal of the North American Benthological Society* 24 (3), 690–705.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan (2005b). "The urban stream syndrome: Current knowledge and the search for a cure". *Journal of the North American Benthological Society* 24 (3), 706–723.

Wan, Z., K. Zhang, X. Xue, Z. Hong, Y. Hong, and J. J. Gourley (2015). "Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States". *Water Resources Research* 51 (8), 6485–6499.

Wang, C., Z.-H. Wang, and Y.-H. Ryu (2021). "A single-layer urban canopy model with transmissive radiation exchange between trees and street canyons". *Building and Environment* 191, 107593.

Wang, K., P. Wang, Z. Li, M. Cribb, and M. Sparrow (2007). "A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature". *Journal of Geophysical Research: Atmospheres* 112 (D15).

Wang, W. and J. Shu (2020). "Urban renewal can mitigate urban heat islands". *Geophysical Research Letters* 47 (6), e2019GL085948.

Wang, W. and K. J. Davis (2008). "A numerical study of the influence of a clearcut on eddy-covariance fluxes of CO_2 measured above a forest". *Agricultural and Forest Meteorology* 148 (10), 1488–1500.

Wang, Y. P., E. Kowalczyk, R. Leuning, G. Abramowitz, M. R. Raupach, B. Pak, E. van Gorsel, and A. Luhar (2011). "Diagnosing errors in a land surface model (CABLE) in the time and frequency domains". *Journal of Geophysical Research: Biogeosciences* 116 (G1).

Wang, Z.-H., E. Bou-Zeid, and J. A. Smith (2013). "A coupled energy transport and hydrological model for urban canopies evaluated using a wireless sensor network". *Quarterly Journal of the Royal Meteorological Society* 139 (675), 1643–1657.

Ward, H. and S. Grimmond (2017). "Assessing the impact of changes in surface cover, human behaviour and climate on energy partitioning across Greater London". *Landscape and Urban Planning* 165, 142–161.

Ward, H., Y. Tan, A. Gabey, S. Kotthaus, and S. Grimmond (2018). "Impact of temporal resolution of precipitation forcing data on modelled urban-atmosphere exchanges and surface conditions". *International Journal of Climatology* 38 (2), 649–662.

Ward, H. C. (2017). "Scintillometry in urban and complex environments: A review". *Measurement Science and Technology* 28 (6), 064005.

Ward, H. C., J. G. Evans, and S. Grimmond (2013). "Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in Swindon, UK". *Atmospheric Chemistry and Physics* 13 (9), 4645–4666.

Ward, H. C., S. Kotthaus, L. Järvi, and S. Grimmond (2016). "Surface Urban Energy and Water Balance Scheme (SUEWS): Development and evaluation at two UK sites". *Urban Climate* 18, 1–32.

Weedon, G. P., C. Prudhomme, S. Crooks, R. J. Ellis, S. S. Folwell, and M. J. Best (2015). "Evaluating the performance of hydrological models via cross-spectral analysis: Case study of the Thames Basin, United Kingdom". *Journal of Hydrometeorology* 16 (1), 214–231.

Weedon, G. P., E. L. Robinson, J. P. Bloomfield, S. Turner, E. J. Crane, and M. J. Best (2023). "Geological controls of discharge variability in the Thames Basin, UK from cross-spectral analyses: Observations versus modelling". *Journal of Hydrology* 625, 130104.

Wei, W. and J. Shu (2020). "Urban renewal can mitigate urban heat islands". *Geophysical Research Letters*.

Weng, Q., D. Lu, and J. Schubring (2004). "Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies". *Remote Sensing of Environment* 89 (4), 467–483.

Wenzel Jr, H. G. and M. L. Voorhees (1981). *Evaluation of the urban design storm concept*.

Wetzel, P. J. and J.-T. Chang (1987). "Concerning the relationship between evapotranspiration and soil moisture". *Journal of Climate and Applied Meteorology* 26 (1), 18–27.

Wilby, R. L. (2007). "A review of climate change impacts on the built environment". *Built Environment* 33 (1), 31–45.

Williams, C. A. and J. D. Albertson (2004). "Soil moisture controls on canopy-scale water and carbon fluxes in an African savanna". *Water Resources Research* 40 (9).

Willmott, C. J. (1982). "Some comments on the evaluation of model performance". *Bulletin of the American Meteorological Society* 63 (11), 1309–1313.

WMO (2019). *Guidance on Integrated Urban Hydrometeorological, Climate and Environmental Services - Volume I*. Geneva.

– (2023). *Guide to Instruments and Methods of Observation Volume I – Measurement of Meteorological Variables*. Tech. rep. WMO-No. 8. Geneva: World Meteorological Organization (WMO).

– (2024). *Paris Olympics will advance research into weather forecasting and urban meteorology*. Date accessed: 18 October 2024.

Wong, N. H., C. L. Tan, D. D. Kolokotsa, and H. Takebayashi (2021). "Greenery as a mitigation and adaptation strategy to urban heat". *Nature Reviews Earth & Environment* 2 (3), 166–181.

Wong, T. H. (2006). "Water sensitive urban design-the journey thus far". *Australasian Journal of Water Resources* 10 (3), 213–222.

Wouters, H., M. Demuzere, K. De Ridder, and N. P. van Lipzig (2015). "The impact of impervious water-storage parametrization on urban climate modelling". *Urban Climate* 11, 24–50.

Wu, K., K. J. Davis, N. L. Miles, S. J. Richardson, T. Lauvaux, D. P. Sarmiento, N. V. Balashov, K. Keller, J. Turnbull, K. R. Gurney, et al. (2022). "Source decomposition of eddy-covariance CO_2 flux measurements for evaluating a high-resolution urban CO_2 emissions inventory". *Environmental Research Letters* 17 (7), 074035.

Wu, X., S. Moustakas, N. Bezak, M. Radinja, M. B. Alivio, M. Mikoš, M. Dohnal, V. Bares, and P. Willems (2024). "Assessing the performance of blue-green solutions through a fine-scale water balance model for an urban area". *Science of The Total Environment*, 174750.

Xiao, Q. and E. G. McPherson (2011). "Rainfall interception of three trees in Oakland, California". *Urban Ecosystems* 14, 755–769.

Xu, C., M. Rahman, D. Haase, Y. Wu, M. Su, and S. Pauleit (2020). "Surface runoff in urban areas: The role of residential cover and urban growth form". *Journal of Cleaner Production* 262, 121421.

Yang, J., M. S. Wong, H. C. Ho, E. S. Krayenhoff, P. Chan, S. Abbas, and M. Menenti (2020). "A semi-empirical method for estimating complete surface temperature from radiometric surface temperature, a study in Hong Kong City". *Remote Sensing of Environment* 237, 111540.

Yang, Y., D. Long, and S. Shang (2013). "Remote estimation of terrestrial evapotranspiration without using meteorological data". *Geophysical Research Letters* 40 (12), 3026–3030.

Yang, Z.-L., R. Dickinson, A. Henderson-Sellers, and A. Pitman (1995). "Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1 (a)". *Journal of Geophysical Research: Atmospheres* 100 (D8), 16553–16578.

Yao, L., W. Wei, and L. Chen (2016). "How does imperviousness impact the urban rainfall-runoff process under various storm cases?" *Ecological Indicators* 60, 893–905.

Yu, M., H. Wu, J. Yin, X. Liang, and S. Miao (2022). "Improved delineation of urban hydrological processes in coupled regional climate models". *Water Resources Research*, e2022WR032695.

Zeeman, M., K. Benjamins, F. Briegel, M.-A. Drouin, G. Feigel, D. Fenner, S. Kotthaus, R. Hilland, F. Meier, W. Morrison, et al. (2023). *A modular data management approach for environmental observation campaigns in multiple cities*. Tech. rep. Copernicus Meetings.

Zeisl, P., M. Mair, U. Kastlunger, P. M. Bach, W. Rauch, R. Sitzenfrei, and M. Kleidorfer (2018). "Conceptual urban water balance model for water policy testing: An approach for large scale investigation". *Sustainability* 10 (3), 716.

Zhang, K., J. S. Kimball, and S. W. Running (2016). "A review of remote sensing based actual evapotranspiration estimation". *Wiley interdisciplinary reviews: Water* 3 (6), 834–853.

Zhang, W., M. Yang, S. Zhang, L. Yu, F. Zhao, D. Chen, S. Yang, H. Li, S. Zhang, R. Li, et al. (2023). "Applicability assessment of five evapotranspiration models based on lysimeter data from a bioretention system". *Ecological Engineering* 194, 107049.

Zhao, L., X. Lee, R. B. Smith, and K. Oleson (2014). "Strong contributions of local background climate to urban heat islands". *Nature* 511 (7508), 216.

Zhao, Y., L. W. Chew, Y. Fan, C. Gromke, J. Hang, Y. Yu, A. Ricci, Y. Zhang, Y. Xue, S. Fellini, et al. (2023). "Fluid tunnel research for challenges of urban climate". *Urban Climate* 51, 101659.

Zhou, Q. (2014). "A review of sustainable urban drainage systems considering the climate change and urbanization impacts". *Water* 6 (4), 976–992.

Zhou, Q., G. Leng, J. Su, and Y. Ren (2019). "Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaptation". *Science of the Total Environment* 658, 24–33.

Zimmermann, F., H. Köstler, C. Grabowski, M. Moeck, C. Saure, and K.-H. Kielhorn (2015). *Beschreibung der Biotoptypen*.

Acknowledgements

"Ah, we zijn weer hier!"

Welcome to the most read part of the thesis! While writing the synthesis, you zoom out overlooking the research you performed in four years. While writing the acknowledgements, you zoom out overlooking the people that helped you do that research. Writing a thesis is called a journey, trajectory, pathway, or rollercoaster, and each of those is true in its own way. At times it is a struggle and a times it is a delight. This is a small window into the world behind the written, rewritten, edited, and then polished words of the science. The view from this window reveals many a place.

One may see the same office in Lumen that also hosted my master's thesis. Looking a little further one finds the ferry over the Rhine. Yet, this process took me well beyond there. A little spotlight shines on Berlin, the city most prominent in this thesis. Overseas lays Boston where I tried my hardest to understand LES. The mists around the Appalachians hide where conceiving the title of this thesis was a welcome distraction from my 20-kg backpack. In the far distance, Sydney was the birthplace of the section on urban moisture in the synthesis. More importantly than these places are the people. Let me introduce you to the wonderful people who make these places what they are and make them impossible to grasp. My not-too-sincere apologies for the number of words that I need to do them justice. Those of you who know me may have seen this coming.

I want to start by thanking both my promotors. I truly appreciate how this endeavor has been shared by the three of us from the proposal writing until the last checks of this book. In all this time, we have met with everyone with the exceptions in the single digits. In light of this commitment and involvement in my research, I would not know who to address first. Therefore, I decided to follow the chronological order of meeting you.

Ryan, you invited me to be ambitious too. This started during our very first meeting in Forum to discuss my ideas for an MSc thesis. This thesis was the starting point of the proposal that became in turn this book. In Cordoba, I witnessed this drive to inspire when the bar was lowered for the next decade and you started with the words: "The world is burning..." More than one endeavor would not have started or come to fruition without your ambition and some chapters would have never been written. With a manuscript

approaching the finish line, you always made sure that every word contributed to the storyline and the final message. I hope this is visible in this thesis.

Gert-Jan, still during my MSc thesis, I nervously walked into your office. I felt like a hydrologist in disguise, but over time you taught me enough about meteorology that I at least feel like a half-decent meteorologist. In the same office we met, we spent quite a few hours digging through the codes of models uncovering the most surprising fixes in models. Your practical approach to science has undoubtedly saved me from many reviewer comments. Whenever I check my manuscripts now, I hear you say: "What would a reviewer say about this?"

Saying that many more contributed to the research in this thesis, would be an understatement. My appreciation is great for all the co-authors who enabled me to study cities all over the world with a world of models. Mathew, thank you for sharing the Urban-PLUMBER models without your tremendous effort to gather these models much of my work would not have been possible. I appreciated our discussions on the details of the models. Many odd results were explained in these talks. Stenka, I hope my gratitude is as warm as the welcome you gave me when I visited Berlin. Somehow you kept finding the time to brainstorm or clarify the Berlin observations. Sue, thank you for the critical perspective on the methods and the clarity of my drafts. Fred, your help was critical in establishing my collaborations in Berlin and uncovering every type of observation I could think of. For all the others, I benefitted greatly from all your critical reflections on the science and the storytelling, thank you!

A warm thank you to all the colleagues from both the hydrology and the meteorology groups. Combining two chair groups meant I missed more than I would have liked. You all organized too many fun activities, which is characteristic of both groups. For over four years, I have been in the big office or "the penthouse". I shared it with many people, for a short time, others for a longer one, some frequently, others just now and then. So, I am confident this is not a complete list. Thank you for the beehive vibe and always having someone to talk to, Awad, Bas, Derk, Germano, Hadeel, Haftu, Jingwei, João, Miranda, Nazwa, Niguse, Noortje, Paolo, Rose, Titus, and Yvette! Louise, we met well before we had the chance to sit in an office together. Thank you for our walks and runs that were an escape from the home office, as well as for the dinners and heartfelt conversations about any topic. Aris, Bert, Bianca, Dragan, Esther, and Lucas. Thank you for the shared passion and frustration of studying urban areas and the answers to many questions that made seemingly complex things easy.

When travel became possible again, this opened the door for the Tour the PhD. I want to thank the all that Joris, Rose, Sjoukje, Steven, and Tessa. The discussions during the Tour were the start of a much larger project. Tessa, when we started, I did not realize how much of a challenge coordinating this would be, but you showed my how easy, fun, and efficient it can be to work collaboratively. I also want to thank you, Femke, for having me

join on your last chapter. While not every discussion was directly relevant to the chapter, I learnt a lot from you.

While Wageningen has been my home base during my PhD, I was hosted by wonderful people during two trips. My longest time abroad was in Boston. Thank you, Dan, for teaching me not only the technical skills but also your conceptual way of thinking, discussing your own and my research as well as science in general, and letting me experience the academic culture in the US. A shoutout to the rest of the Urban Fluid Mechanics lab, I enjoyed your willingness to show me the way in Boston, BU, and PALM. Chaitu and Vijay, you have been great housemates, thank you for showing me more of Boston than the university. The other Dan, our snow hike was an experience to never forget. While you were not in Boston, Shrinidhi, I want to thank you as without you no run on Snellius would have ever finished. My stay in Australia would not have been possible without Negin, Melissa, Steven, and Kerry. Negin, Mellissa, and the Climate-Resilient City lab members, I am as happy as surprised that you found time to host us in the middle of organizing a conference.

Without staying sane there is no thesis, so here is to everyone that provided good times and peace of mind. To Nienke, the first call is always to you, so let this not be an exception. Not always close, but always there. Somehow we managed to finish our theses nearly simultaneously. To Lynn, your stories never fail to entertain. The door will always be open. While you also worked at MAQ, Kim, you seem more in place here together with Christoph. Our Friday nights were the best way to kick of the weekend, even though we sometimes needed a reminder that the science talk was enough. Every now and then it feels like a little family whether that is a sweatpants Friday night or a Cretan canyon walk.

To Vogelvrij, I hope we may stay as free despite seeing each other less than at the start of this PhD. Let's keep up our getaways! I am sorry I had to miss the most recent one. To Wim and Karina, within a few steps on the floor, dancing was all I could think about. A wonderful way to clear the mind. The competition team and the late-night Wednesday club made sure the time of the floor was worthwhile too. Also, to my neighbors in Andelst, I am still getting used to waving to people in a car, but you have made me feel welcome from the get-go creating a home for our home. To my (extended) family, Christmas has become and remains a highlight of the year. To my parents and brothers, who are always there when needed. I enjoy how everyone makes the time to share in all our moments.

Judith, where should I begin? You could have been in nearly every paragraph of my obnoxiously long acknowledgements, but I promised you your own moment. From start to finish, let's revisit it all. For the start, we go back to the very beginning before I even considered a PhD. You put me on the track of the WIMEK talent program and doing a PhD. You have been my office companion at home and later even in the penthouse. Many non-working pieces of code were fixed by your incredible rubber duck imitations. Endless

chatter about ideas passed by and your eye for detail has saved me too often. We have even written papers and given a seminar together. Through this journey we followed each other where we could, which led us to a myriad of destinations. Staying put is by no means one of our talents. Yet, we always had something to switch off work for a moment. Whether that is creating a new home, hiking our own hike, or planning a wedding. Who knows where our adventures will lead us next? I cannot wait to keep *lekker langzaam lanterfanten* with you.

Harro Jongen

Authorship statement

The general research direction of this thesis has been developed by me as part of the WIMEK graduate program. The proposal was developed together with my promoters. Based on the proposal, I conducted the presented research formulating the specific research questions. I wrote the introduction (Chapter 1) and the synthesis (Chapter 7) with minor suggestions from my promoters. Chapters 2–6 are the result of collaborations, for which the contribution is outlined below by chapter.

HJ = Harro Jongen (WUR)

AC = Andreas Christen (UF)
BK = Birgit Kleinschmidt (TUB)
CW = Chenghao Wang (UoO)
DT = Doerthe Tetzlaff (IGB/HUB)
EV = Erik Velasco (indep.)
FM = Fred Meier (TUB)
JBe = Jason Beringer (UWA)
JM = Joe McNorton (ECMWF)
GS = Gert-Jan Steeneveld (WUR)
KO = Keith Oleson (NSF NCAR)
MB = Martin Best (MetOffice)
MK = Martin de Kauwe (UoBr)
MR = Matthias Roth (UoSi)
NC = Nektarios Chrysoulakis (FORTH)
NM = Naika Meili (NUS/ETH)
RL = Ruidong Li (UCL/TU)
RV = Roland Vogt (UoBa)
SGr = Sue Grimmond (UoR)
SV = Stenka Vulova (TUB)
WP = Włodzimierz Pawlak (UoL)
ZW = Zhihua Wang (ASU)

AT = Aristofanis Tsiringakis (ECMWF)
CJ = Cor Jacobs (WEnR)
DL = Dan Li (BU)
EA = Emmanual Akinlabi (BU)
FJ = Femke Jansen (WUR)
JBa = Jong-Jin Baik (SNU)
JH = Jinkyu Hong (YU)
JW = Je-Woo Hong (KEI)
KF = Krzysztof Fortuniak (UoL)
LJ = Leena Järvi (UoH)
MD = Matthias Demuzere (B-Kode)
ML = Mathew Lipson (UNSW)
MV = Mikhail Varentsov (LMSU)
NH = Nasrin Haacke (TUB)
NT = Natalie Theeuwes (KNMI)
RT = Ryan Teuling (WUR)
SGa = Srinidhi Gadde (UoT)
SP = Sueng-Bu Park (UoSe)
TS = Ting Sun (UCL)
YH = Yuqi Huang (UoO)

Chapter 2

Conceptualization:	HJ, RT
Data collection:	AC, CJ, EV, GS, JBe, JH, JW, KF, LJ, MR, NC, NT, RV, WP
Data analysis and interpretation:	HJ in consultation with GS, RT
Writing of manuscript:	HJ
Revision of manuscript and approval for publication:	AC, CJ, EV, GS, JBe, JH, JW, KF, LJ, MR, NC, NT, RT, RV, WP

Chapter 3

Conceptualization:	GS, HJ, ML, RT, SGr
Modeling:	AT, CW, GS, JBa, JM, MB, MD, MK, ML, MV, KF, KO, NM, RL, SP, TS, YH, ZW
Data analysis and interpretation:	HJ in consultation with GS, RT
Writing of manuscript:	HJ
Revision of manuscript and approval for publication:	AT, CW, GS, JBa, JM, MB, MD, MK, ML, MV, KF, KO, NM, RL, RT, SGr, SP, TS, YH, ZW

Chapter 4

Conceptualization:	HJ
Data collection:	MB, ML, SGr
Data analysis and interpretation:	HJ in consultation with GS, RT
Writing of manuscript:	HJ
Revision of manuscript and approval for publication:	GS, MB, ML, RT, SGr

Chapter 5

Conceptualization:	HJ
Data collection:	MB, ML, SGe
Data analysis and interpretation:	HJ in consultation with GS, RT
Writing of manuscript:	HJ
Revision of manuscript and approval for publication:	GS, MB, ML, RT, SGe

Chapter 6

Conceptualization:	HJ
Modeling:	HJ in consultation with DL, EA, SGa
Data analysis and interpretation:	HJ in consultation with GS, RT
Writing of manuscript:	HJ
Revision of manuscript and approval for publication:	DL, EA, GS, RT, SGa

List of publications

Peer-reviewed Journal Publications

Arheimer, B., C. Cudennec, A. Castellarin, S. Grimaldi, K. V. Heal, C. Lupton, A. Sarkar, F. Tian, J.-M. K. Onema, S. Archfield, G. Blöschl, P. L. B. Chaffe, B. F. Croke, M. Dembélé, C. Leong, A. Mijic, G. M. Mosquera, B. Nlend, A. O. Olusola, M. J. Polo, M. Sandells, J. Sheffield, T. C. van Hateren, M. Shafiei, S. Adla, A. Agarwal, C. Aguilar, J. C. Andersson, C. Andraos, A. Andreu, F. Avanzi, R. R. Bart, A. Bartosova, O. Batelaan, J. C. Bennett, M. Bertola, N. Bezak, J. Boekee, T. Bogaard, M. J. Booij, P. Brigode, W. Buytaert, K. Bziava, G. Castelli, C. V. Castro, N. C. Ceperley, S. K. R. Chidepudi, F. H. S. Chiew, K. P. Chun, A. G. Dagnew, B. Dekongmen, M. del Jesus, A. Dezetter, J. A. do Nascimento Batista, R. C. Doble, N. Dogulu, J. P. Eekhout, A. Elçi, M. Elenius, D. C. Finger, A. Fiori, S. Fischer, K. Förster, D. Ganora, E. G. Ellouze, M. Ghoreishi, N. Harvey, M. Hrachowitz, M. Jampani, F. Jaramillo, **H. J. Jongen**, K. Y. Kareem, U. T. Khan, S. Khatami, D. G. Kingston, G. Koren, S. Krause, H. Kreibich, J. Lerat, J. Liu, M. M. de Brito, G. Mahé, H. Makurira, P. Mazzoglio, M. Merheb, A. Mishra, H. Mohammad, A. Montanari, N. Mujere, E. Nabavi, A. Nkwasa, M. E. O. Alegria, C. Orieschnig, V. Ovcharuk, S. S. Palmate, S. Pande, S. Pandey, G. Papacharalampous, I. Pechlivanidis, G. Penny, R. Pimentel, D. A. Post, C. Prieto, S. Razavi, S. Salazar-Galán, A. S. Namboothiri, P. P. Santos, H. Savenije, N. J. Shanono, A. Sharma, M. Sivapalan, Z. Smagulov, J. Szolgay, J. Teng, A. J. Teuling, C. Teutschbein, H. Tyralis, A. van Griensven, A. J. van Schalkwyk, M. van Tiel, A. Viglione, E. Volpi, T. Wagener, L. Wang-Erlandsson, M. Wens, and J. Xia (2024). "The IAHS Science for Solutions decade, with Hydrology Engaging Local People IN a Global world (HELPING)". *Hydrological Sciences Journal* 69 (11), 1417–1435.

Boekee, J., S. J. van der Linden, M.-C. ten Veldhuis, I. E. Verouden, P. J. Nollen, Y. Dai, **H. J. Jongen**, and B. J. van de Wiel (2024). "Rethinking the Roughness Height: An Improved Description of Temperature Profiles over Short Vegetation". *Boundary-Layer Meteorology* 190 (7), 31.

Hateren*, T. C. van, **H. J. Jongen***, H. Al-Zawaidah, J. G. Beemster, J. Boekee, L. Bogerd, S. Gao, C. Kannen, I. van Meerveld, S. I. de Lange, F. Linke, R. B. Pinto, J. O. Remmers, J. Ruijsch, S. R. Rusli, R. C. van de Vijsel, J. P. Aerts, S. M. Agoungbome, M. Anys, S. Blanco Ramírez, T. van Emmerik, L. Gallitelli, G. Chiquito Gesualdo, W. Gonzalez Otero, S. Hanus, Z. He, S. Hoffmeister, R. O. Imhoff, T. Kerlin, S. M. Meshram, J. Meyer, A. Meyer Oliveira, A. C. Müller, R. Nijzink, M. Scheller, L. Schreyers, D. Sehgal, P. F. Tasseron, A. J. Teuling,

M. Trevisson, K. Waldschläger, B. Walraven, C. Wannasin, J. Wienhöfer, M. J. Zander, S. Zhang, J. Zhou, J. Y. Zomer, and B. W. Zwartendijk (2023). "Where should hydrology go? An early-career perspective on the next IAHS Scientific Decade: 2023–2032". *Hydrological Sciences Journal* 68 (4), 529–541.

Jansen, F. A., **H. J. Jongen**, C. M. Jacobs, F. C. Bosveld, A. J. Buzacott, B. G. Heusinkveld, B. Kruijt, M. van der Molen, E. Moors, G.-J. Steeneveld, C. van der Tol, Y. van der Velde, B. Voortman, R. Uijlenhoet, and A. J. Teuling (2023). "Land cover control on the drivers of evaporation and sensible heat fluxes: An observation-based synthesis for the Netherlands". *Water Resources Research* 59 (11), e2022WR034361.

Jongen, **H. J.**, M. Lipson, A. J. Teuling, S. Grimmond, J.-J. Baik, M. Best, M. Demuzere, K. Fortuniak, Y. Huang, M. G. De Kauwe, R. Li, J. McNorton, N. Meili, K. Oleson, S.-B. Park, T. Sun, A. Tsiringakis, M. Varentsov, C. Wang, Z.-H. Wang, and G. J. Steeneveld (2024a). "The Water Balance Representation in Urban-PLUMBER Land Surface Models". *Journal of Advances in Modeling Earth Systems* 16 (10), e2024MS004231.

Jongen, **H. J.**, S. Vulova, F. Meier, G. J. Steeneveld, F. A. Jansen, D. Tetzlaff, B. Kleinschmit, N. Haacke, and A. J. Teuling (2024b). "Attributing Urban Evapotranspiration From Eddy-Covariance to Surface Cover: Bottom-Up Versus Top-Down". *Water Resources Research* 60 (9), e2024WR037508.

Jongen, **H. J.**, G.-J. Steeneveld, J. Beringer, A. Christen, N. Chrysoulakis, K. Fortuniak, J. Hong, J.-W. Hong, C. M. Jacobs, L. Järvi, F. Meier, W. Pawlak, M. Roth, N. E. Theeuwes, E. Velasco, R. Vogt, and A. J. Teuling (2022b). "Urban water storage capacity inferred from observed evapotranspiration recession". *Geophysical Research Letters* 49 (3), e2021GL096069.

* Contributed equally to this work.

Other Scientific Publications

Jongen, **H. J.** (2024). "The Urban-PLUMBER land surface model evaluation project: the water balance representation". In: *IAUC newsletter*. Vol. 90.

Jongen, **H. J.**, M. Lipson, A. J. Teuling, S. Grimmond, and G.-J. Steeneveld (2022a). "Crucial consistency of the water balance in urban land surface models". In: *EGU General Assembly Conference Abstracts*, EGU22–7046.

Jongen, **H. J.**, A. J. Teuling, and G.-J. Steeneveld (2022c). "Urban water storage capacity inferred from observed evapotranspiration recession". In: *IAUC newsletter*. Vol. 83.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Harro Joseph Jongen

born on 22 April 1997 in Arnhem, the Netherlands

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Wageningen, 3 April 2025

SENSE coordinator PhD education

Dr Ir Peter Vermeulen

The SENSE Director

Dr Jampel Dell'Angelo

The SENSE Research School declares that **Harro Joseph Jongen** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 52.4 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2021)
- Research in context activity: Writing of community opinion paper 'Where should hydrology go? An early-career perspective on the next IAHS Scientific Decade: 2023-2032' (2023)

Other PhD and Advanced MSc Courses

- Introduction to machine learning. PE&RC and WIMEK (2021)
- Effective HPC for Climate and Weather, Centre of Excellence in Simulation of Weather and Climate in Europe, Online/Germany (2021)
- Scientific Integrity, Wageningen Graduate Schools (2021)
- Communication with the Media and the General Public, Wageningen Graduate Schools (2021)
- Effective and efficient communication in academia and beyond, Wageningen Graduate Schools (2022)
- Scientific writing, Wageningen Graduate Schools (2022)
- CHESS Summer School on Land-Atmosphere Interaction Processes and Convection, University of Bergen (2022)
- Start to teach, Wageningen University (2023)
- Teaching and supervising thesis students, Wageningen University (2023)

Management and Didactic Skills Training

- Chairman of WIMEK PhD council (2021-2024)
- Organization 'Tour the PhD' visiting research institutes abroad (2022)
- Organization writing week HWM (2023)
- Supervising MSc student with thesis entitled 'Quantifying the Water Balance of Amsterdam' (2024)
- Assisting practicals in the BSc courses 'Water 1' (2023-2024) & 'Introduction Soil Water Atmosphere' (2023)
- Assisting practicals in the MSc courses 'Urban hydrometeorology' (2021-2023) & 'Catchment and Climate Hydrology' (2023)

Selection of Oral Presentations

- *The water balance in a city: sponge or desert?* BBOS Autumn Symposium 2021, 20-22 October 2021, Soesterberg, The Netherlands
- *Crucial consistency of the water balance in urban land surface models*. EGU 2022, 23-27 May 2022, Vienna, Austria
- *The Urban-PLUMBER land surface model evaluation project: the water balance representation* & *Reconstructing urban evapotranspiration observed with eddy-covariance with landcover contributions in the footprint*. ICUC 2023, 28 August – 1 September 2023, Sydney, Australia

This research received funding from the WIMEK graduate school.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Harro J. Jongen & Judith Jongen-Boekee