Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decomposition of Wastewater Sources in England

Barnaby Dobson

bdobson@tcd.ie

Trinity College Dublin

Robin Maes-Prior

Imperial College London

Christopher Jackson

British Geological Survey

Article

Keywords:

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10040792/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at npj Clean Water on September 4th, 2026.
See the published version at https://doi.org/10.1038/s41545-026-00630-2.

Decomposition of Wastewater Sources in England

Barnaby Dobson[1,2*], R. Maes-Prior[2] and C. Jackson[3]

[1]Trinity College Dublin, Dublin, Ireland.
[2]Imperial College London, London, UK.
[3]British Geological Survey, Keyworth, UK.

*Corresponding author(s). E-mail(s): bdobson@tcd.ie;

Abstract

The flow arriving at a wastewater treatment plant mixes municipal sewage, rainfall, soil moisture and groundwater, yet is routinely interpreted as if its composition were fixed. Despite the uncertainty this introduces to any case where the signal is used, from combined sewer overflow management to wastewater-based epidemiology, source-resolved composition is relatively underexplored. Here we introduce WWCompose, an empirical-conceptual model that recovers the dynamic contributions of rainfall, soil moisture and groundwater to daily influent, and apply it to approximately $\mathbf{2,900}$ treatment plants across England using open national data. Predicted flows match observations with high skill (mean Kling–Gupta efficiency $\mathbf{0.68}$; $\mathbf{80\%}$ of plants above $\mathbf{0.6}$), with fitted parameters organising around hydrogeology, permeability and urbanisation. Non-municipal water is widespread: municipal flow averages $\mathbf{59\%}$ nationally, and at individual works may vary between $\mathbf{100\%}$ in dry weather to $< \mathbf{5\%}$ during the wet conditions that often dominate periods of interest.

1 Introduction

Influent arriving to a Wastewater Treatment Plant (WWTP) constitutes water from multiple sources and pathways. It is composed of household and trade wastewater, direct storm water runoff and drainage, rainfall-derived inflow and infiltration [1], and groundwater [2]. The composition of influent matters because the same observed flow can create different interpretations for flooding [3], combined sewer overflow spills [4], diffuse urban pollution [5], receiving-water exposure [6], wastewater-based epidemiology [7], and the design or operation of treatment technology [8]. For example, the consequences of a sewer spill depend strongly on whether the discharged flow is primarily rainwater or municipal effluent [6], while interpretation of SARS-CoV-2 wastewater samples depends on the number of people contributing to the influent [9]. Both cited examples implicitly assumed a static influent composition, introducing significant uncertainty given its known dynamic variability [10]. Rainfall-derived inflow and infiltration, and groundwater sources are established as major influent contributors [11][12], observed to compose up to 70% of influent at some WWTPs [13][11]. However, they are difficult to estimate owing to multiple data, scale, and conceptual limitations [14]. The mechanism for groundwater ingress is that the water table

rises above some pipes in a sewer network and enters through cracks or gaps [15], this is categorised as phreatic zone sewer infiltration [14]. Although this process is conceptually straightforward, variability in pipe depths and groundwater depth make quantifying phreatic zone sewer infiltration incredibly challenging from a data perspective [16]. Rainfall-derived inflow and infiltration is referred to as vadose zone sewer infiltration [14], which captures a diverse set of potential processes in the near-surface such as preferential flow paths [17] and connectivity with hydrological features such as rivers [18][19]. Although characterising each mode completely requires comprehensive data and modelling efforts [16], dynamic variability in phreatic zone infiltration can generally be attributed to variability in groundwater level [20] and variability in the vadose zone to soil moisture [21].

A combination of physical modelling [22], statistical data analysis and machine learning [23][24], and conceptually based empirical methods [25] have all been demonstrated to achieve good predictive performance of WWTP influent in literature. These studies are typically focussed on few WWTPs further highlighting the data limitations in sewer infiltration. The only studies that investigate infiltration on more than 10 WWTPs [26][27][28], do not make predictions about WWTP influent and so cannot be validated.

In this study, we identify that the open data environment in England currently supports a nationwide assessment and validation of WWTP influent composition in 2,900 locations. We build on existing conceptually based empirical methods [25] to test the hypothesis that daily influent flow contains recoverable signatures from dynamic data on rainwater, soil moisture, and groundwater level. The outputs of such an approach facilitate quantifying dynamic wastewater composition with applications ranging from sewer spills to wastewater epidemiology.

2 Results

2.1 National model performance

We validate our approach by comparing predicted influent with observed daily flow at WWTPs across England, finding high model performance in many WWTPs Fig. 1, with an average $KGE = 0.68$ and 80% of works having a $KGE > 0.6$. The spread over cross-validation repetitions is narrow relative to the national performance gradient, indicating that the fitted decompositions are not dominated by arbitrary split-to-split variability. The few, $< 3\%$, locations below $KGE < 0.4$ are interpretable failures rather than random outliers: visual inspection commonly shows operational signatures in the effluent record; primarily abrupt changes in mean flow or variance. The main medium-performance band $KGE \approx 0.4 - 0.6$ constitutes 13% of WWTPs and is concentrated in the east of England, consisting of smaller plants with more variable operational regimes and drier catchments. Drier climates and more permeable geology are typically harder to fit, with the driest 5% of catchments having an average $KGE = 0.64$ and higher permeability catchments with $KGE = 0.57$, see Fig. 1 and section 2.4 for further details.

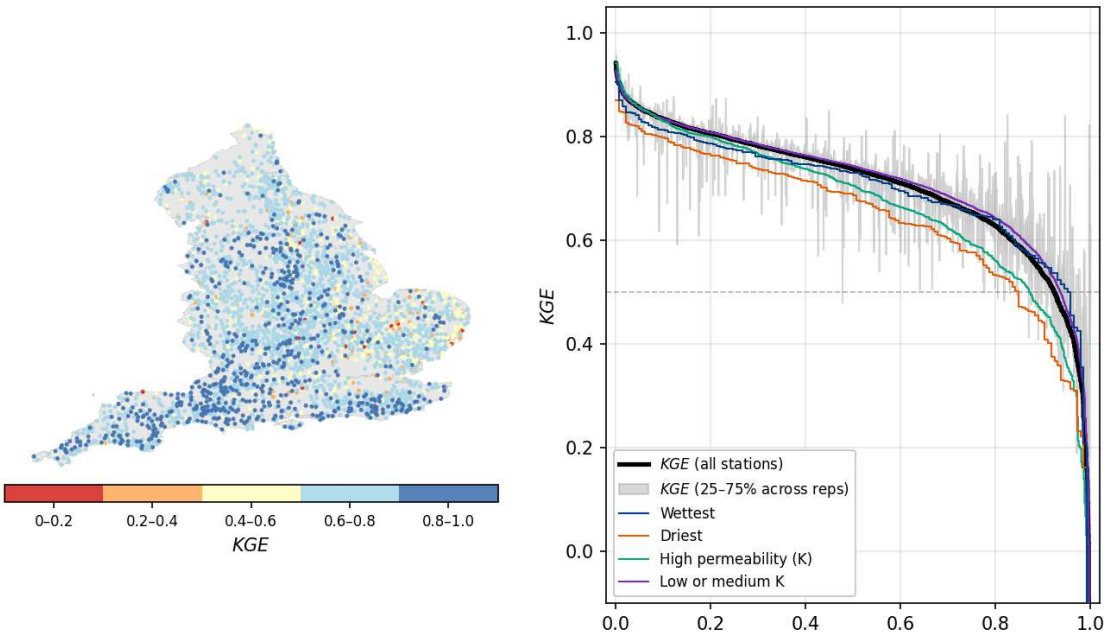


Fig. 1 National model performance across England WWTPs. Left: full-record Kling–Gupta efficiency (KGE) for each plant. Right: exceedance probability (x-axis) for a given KGE value (y-axis), different coloured lines show different groups, while the grey band shows variability over cross-validation repetitions.

2.2 Spatial structure in fitted parameters

The fitted parameters exhibit distinct spatial organisation (Fig. 2) varying from parameter to parameter. For example, the constant scaling parameter for groundwater contributions C_{GW} broadly traces major aquifers, the constant scaling parameter for municipal contributions C_{Mun} increases around urban areas, and the exponent applied to precipitation contributions E_P increases in the wettest regions. Although some common spatial structure exists across fitted parameters, it is weak: the top four components of a principal component analysis explain only 40% of the variance across the 17 parameters, and the top eight explain 65%. This does not support the existence of distinct wastewater catchment archetypes. The upper precipitation gate threshold $T_{P,\mathrm{gate}}^{\mathrm{hi}}$, field capacity θ_i, saturated water content θ_s, and wetting-front suction head ψ are omitted from Fig. 2 because they show limited variability or no distinctive national spatial pattern.

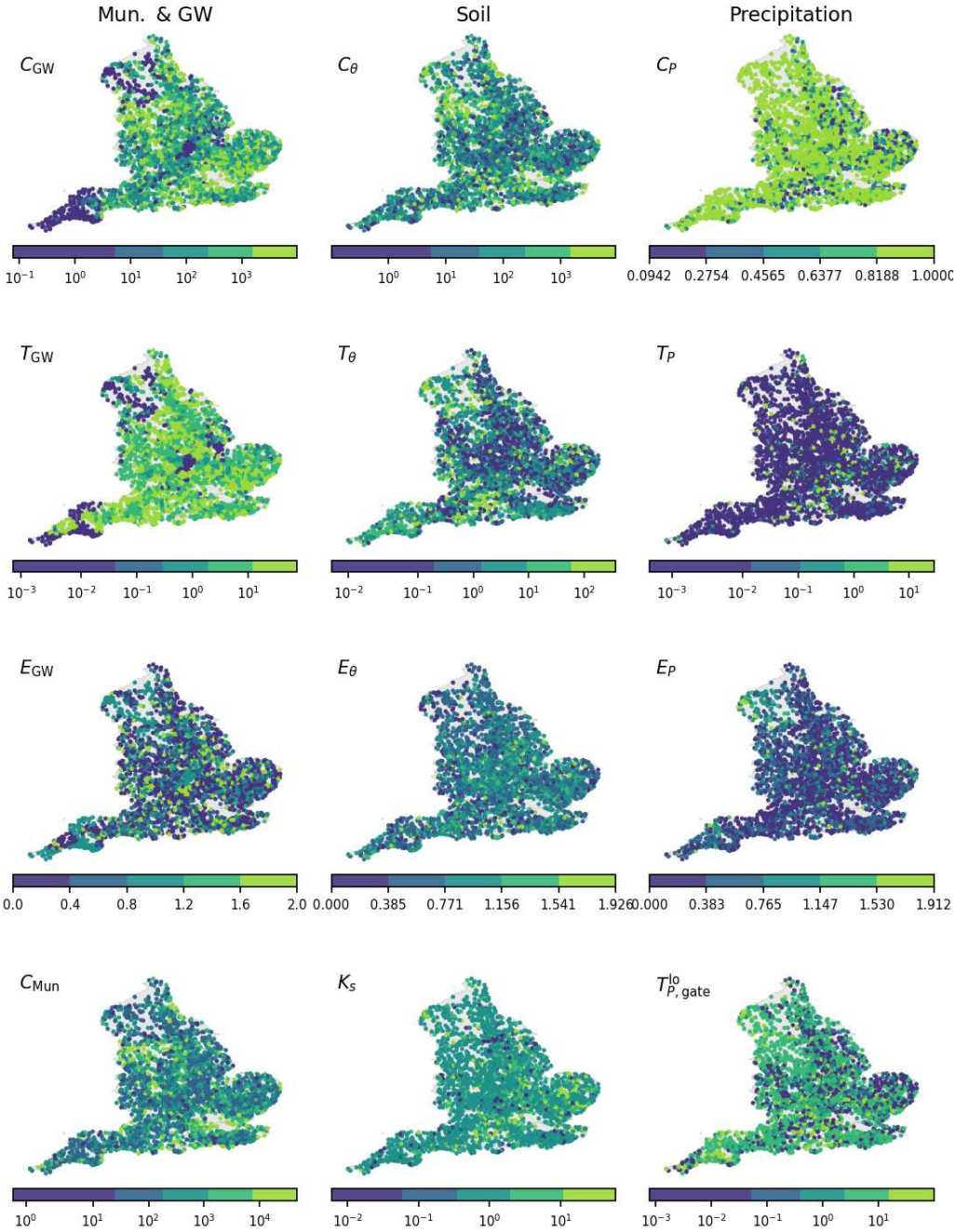


Fig. 2 Spatial structure in fitted model parameters. Per-WWTP average parameter values. GW stands for groundwater, θ for soil moisture, P for precipitation, Mun for municipal. C refers to a linear factor applied to that model component, T a threshold above which a contribution must rise to contribute to influent, and E an exponent applied to contributions above T. K_s refers to saturated conductivity. Further details are provided in Section 5.

2.3 National composition

The fitted models derive the composition fractions for a given WWTP on a given day. These daily compositions are averaged over the entire data period, $2010 - 2024$, and shown in Fig. 3. At a national scale, municipal is the largest contribution (59%), followed by soil moisture (21%), groundwater (13%), and precipitation (7%). Despite its prevalence, only 0.1% of catchments are $> 95\%$ municipal, highlighting the rarity of a textbook fully separate sewer system. Sensible spatial groupings are evident – we see higher

precipitation composition in the west, lower soil moisture composition upon chalk, and higher groundwater composition above major aquifers – but compositions within these categories are still highly variable.

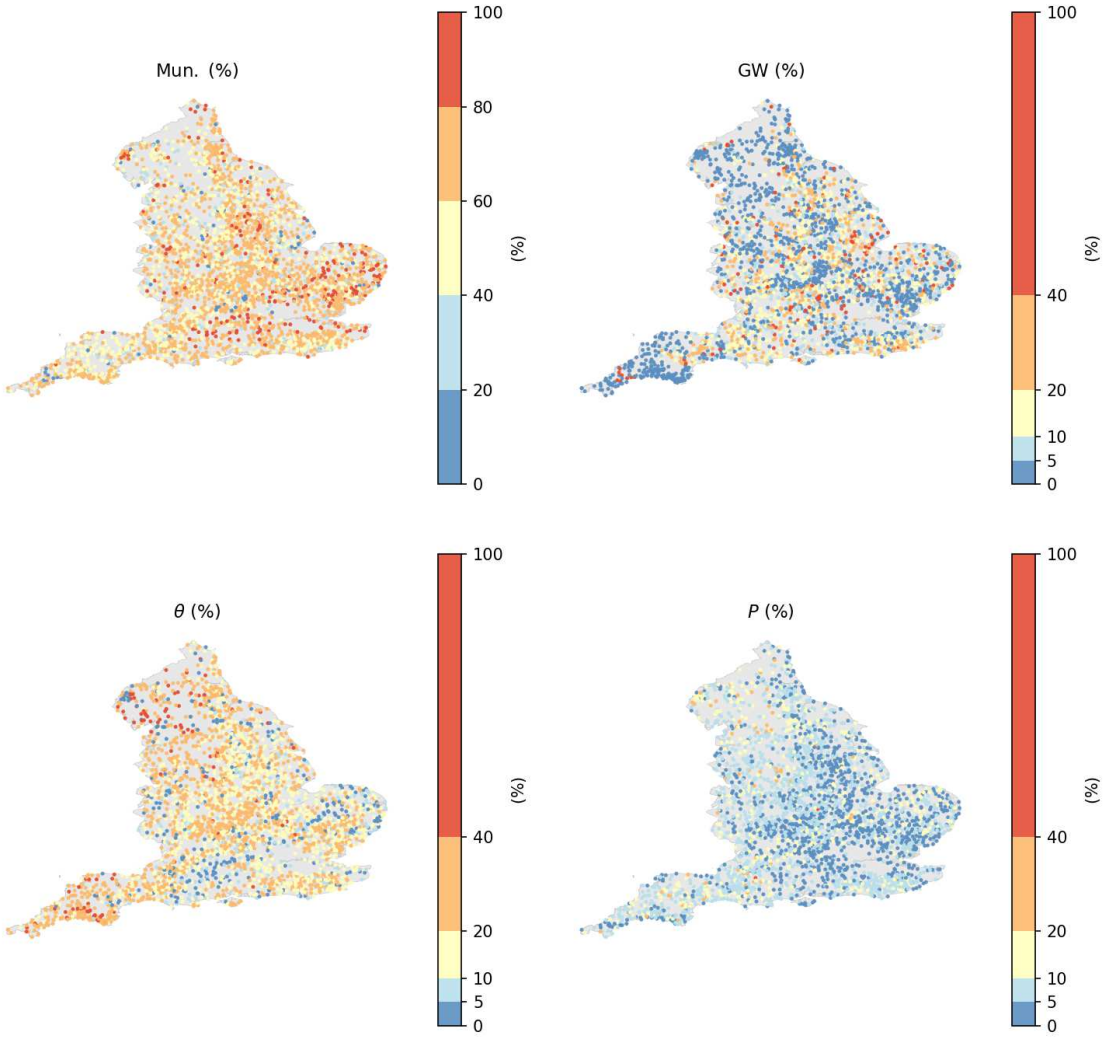


Fig. 3 National wastewater composition. Per-plant average composition. $Mun.$ stands for municipal, GW for groundwater, θ for soil moisture, and P for precipitation.

2.4 Parameter and composition variability

Figures 1 and 3 indicate the importance of aridity and permeability towards model behaviour. We investigate these categories in greater detail with Fig. 4, showing the Kolmogorov-Smirnov test statistic, signed by direction of Spearman rank, between parameters, compositions, and performance values in these different categories. The results support the interpretation of model behaviour, below.

Greater permeability provides more opportunity for surface water dissipation, reducing soil moisture and rainfall composition, $\theta(\%)$ and $P(\%)$. Reductions in these sources correspondingly increases the

composition proportion of groundwater and municipal, $GW(\%)$ and $Mun(\%)$. Municipal composition increases by more than groundwater composition in these cases because many WWTPs do not have pipes at the depth of the groundwater table, but nearly all have significant municipal components. We see that some parameters contributing to soil moisture (the constant scaling parameter for soil moisture contributions C_θ, the exponent applied to soil moisture contributions E_θ, saturated hydraulic conductivity K_s) do increase, but also that the soil moisture contribution threshold T_θ increases, suggesting that while overall soil moisture contributions decrease (decrease in $\theta(\%)$), contributions occur in a more peaked manner.

WWTPs in drier areas show higher municipal composition and reduced composition from all other sources. Reductions in T_θ, the lower precipitation gate threshold $T^{\mathrm{lo}}_{P,\mathrm{gate}}$, and increases in $T^{\mathrm{hi}}_{P,\mathrm{gate}}$, E_θ support more variable but weaker (decreased C_θ and the constant scaling parameter for precipitation contributions C_P) contributions from direct and indirect soil moisture. In contrast, reductions in the exponent applied to groundwater contributions E_{GW} and E_P and increases in the groundwater contribution threshold T_{GW} and the precipitation contribution threshold T_P imply decreased variability in contributions from groundwater and precipitation. Wetter WWTPs generally reverse trends shown in drier ones, supporting higher precipitation and direct soil moisture contributions, with reduced municipal, groundwater and indirect soil moisture contributions.

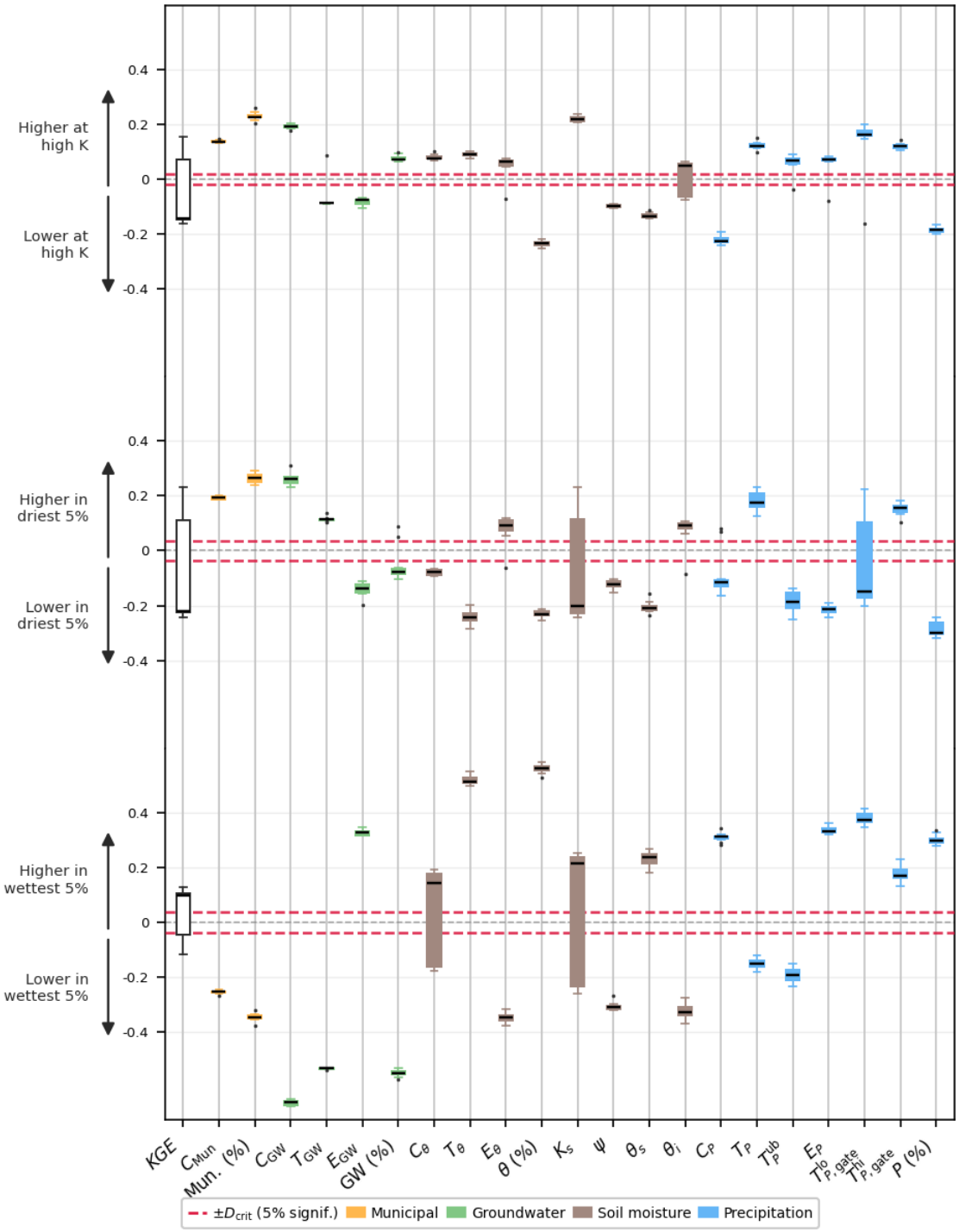


Fig. 4 Effect of permeability classification on fitted composition and parameters. Boxplots show signed Kolmogorov–Smirnov statistic for different parameters, compositions, and performance. The top row performs the test between high vs low/medium permeability, middle row for driest, and bottom row for wettest WWTPs.

2.5 Case studies

Three exemplar sites were selected to highlight distinct behaviours and shortcomings in the approach, shown in Fig. 5.

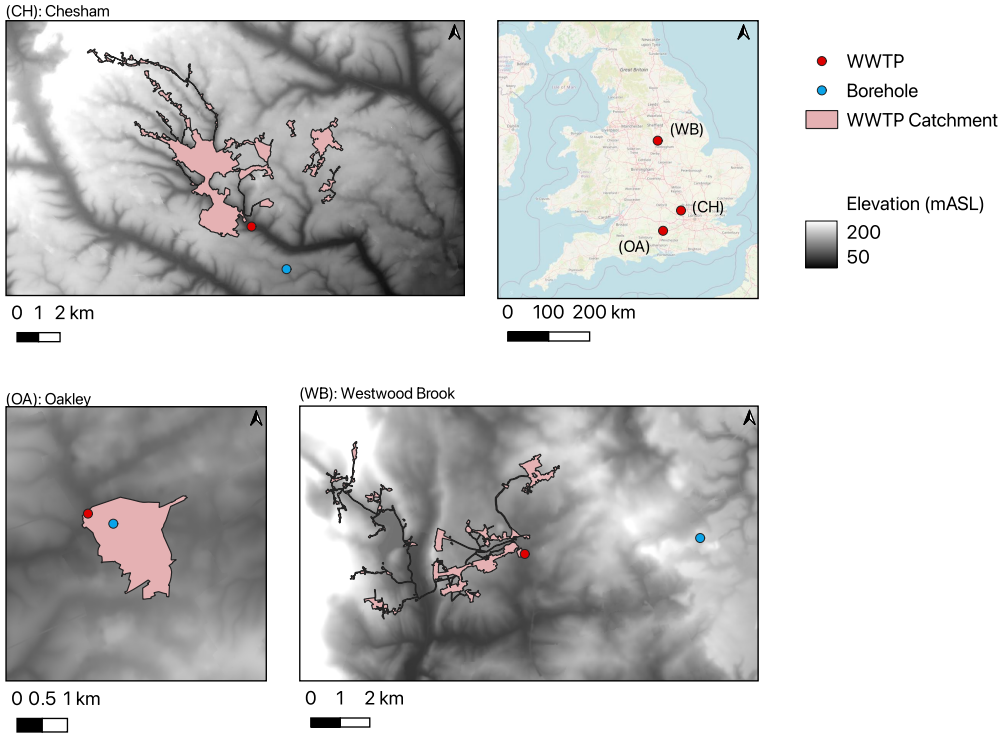


Fig. 5 Map of exemplar case study locations. Selected WWTPs span contrasting regions. Elevation data is from Terrain Tiles, accessed on 2026-06-01 from https://registry.opendata.aws/terrain-tiles.

The fitted soil moisture, borehole groundwater level, dynamic composition and predicted versus simulated flow at the WWTP are shown in Fig. 6

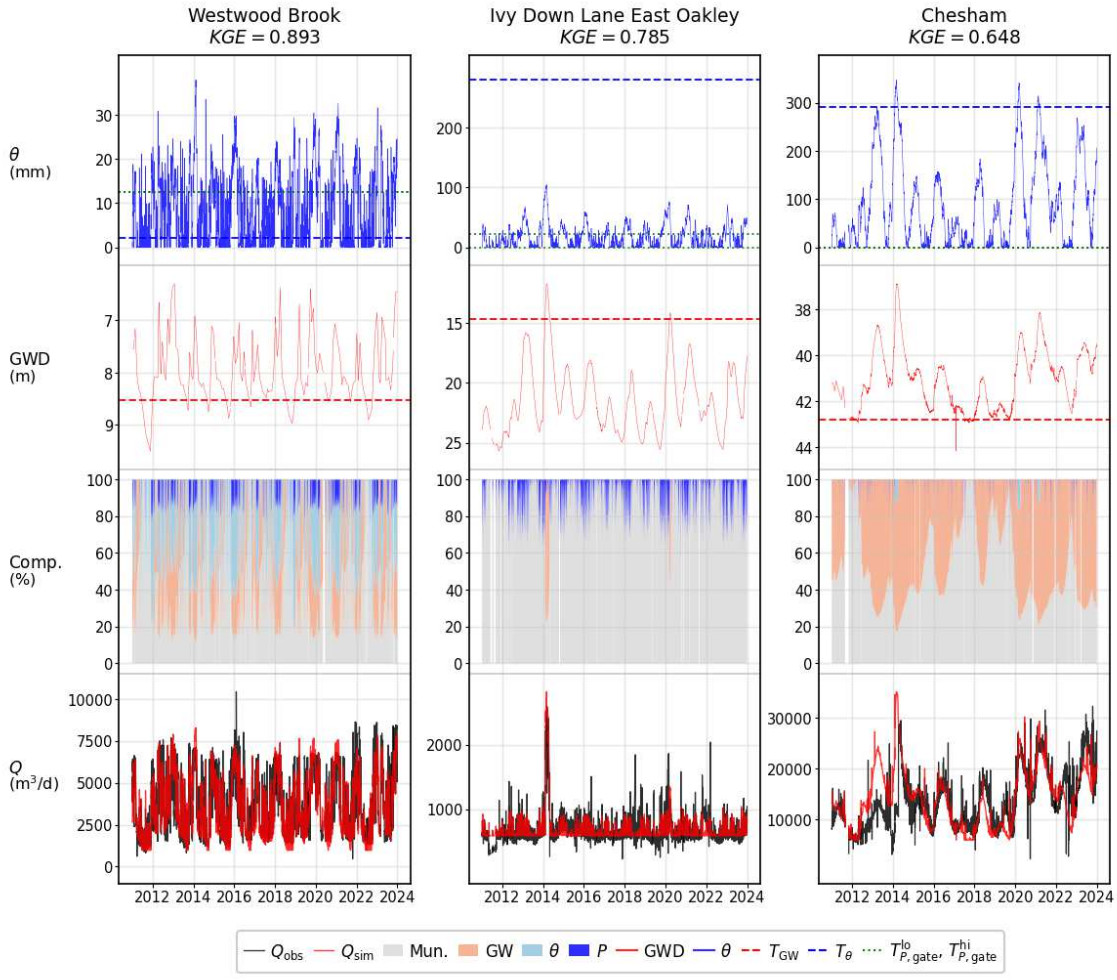


Fig. 6 Selected site time series and inferred wastewater composition. Top row shows fitted soil moisture (θ), second row shows borehole observations formulated as groundwater depth below surface datum (GWD), third row shows dynamic composition of influent, bottom row shows predicted and observed WWTP flows.

Westwood Brook WWTP is a high precipitation site upon mudstone, sandstone and siltstone bedrock with no superficial deposits, classified as a low permeability site. It has a balanced composition across sources, its model performance is in the top 1.5% of stations. Its municipal contribution varies between $20 - 100\%$, with some periods exhibiting an equal split between municipal, rainfall, phreatic, and vadose zone infiltration. It exemplifies a WWTP where composition analysis is of critical importance. Despite sewer overflows in this catchment spilling 88 times in 2025 for a total of 650 hours [29], assessment by the private water company that operates it does not mention infiltration.

Oakley WWTP is a high preciptitation site upon clay, silt, sand and gravel superficial deposits above a chalk bedrock, classified as a low permeability site. Oakley demonstrates ephemeral phreatic zone infiltration particularly well, with groundwater overtopping into the sewer network in 2014. Any other variability is generally precipitation dominated and highlights the difficulty in achieving the highest, $> 0.75KGE$, performance values in the wetter catchments, Fig. 1, which we attribute to the dominant influence of heterogeneous sewer infrastructure during rainfall events.

Chesham WWTP is a low precipitation site upon chalk bedrock with no superficial deposits, classified as a high permeability site. It is also a site of special attention to the private water company that operates the plant owing to its widespread phreatic zone infiltration confirmed by field samples. The WWTP flow variability tracks groundwater level in contrast to soil moisture, which instead has peaks occurring 3-6 months earlier than groundwater level peaks. The significant depths recorded at the assigned borehole $(38 - 42m)$ are deeper than any pipe would be set, indicating that the infiltration is likely occurring in a localised area, presumably adjacent to the rivers in the catchment, further highlighting the data limitations for more detailed assessments.

3 Discussion and Outlook

Our central result is the high flow-prediction accuracy of WWCompose across a never before demonstrated number of WWTPs. Such high performance indicates that source-resolved wastewater composition is possible in many locations. The performance distribution in Fig. 1 shows that dominant hydrological controls can be recovered from effluent records even without explicit sewer-network topology. The composition maps, Fig. 3, reveal widespread non-municipal flow arriving at WWTPs. When disaggregated over time, for example, as shown in Westwood Brook, Fig. 6, the importance of wastewater decomposition becomes even more apparent, with non-municipal sources dominating during wetter periods. The spatial parameter patterns in Fig. 2 and detailed parameteric assessment in Fig. 4 further indicate that the fitted quantities organise around hydrogeology, permeability, urbanisation, and soil in broadly intuitive relationships. Although WWTP archetypes were not apparent, the clearest catchment characteristics to delineate behavioural differences were aridity and permeability, supporting results from [30].

We identify a variety of use cases for WWCompose, discussed below.

Understanding the impact of combined sewer overflow spills is a key focus area of successive UK governments at present, who have committed £60 billion to their remediation [31]. Existing literature has already indicated that 80% of total monitored spill duration is likely to be driven by soil moisture or groundwater sources [4]. Wastewater decomposition facilitates prioritising spills based on their proportion that is municipal, and targeting investment to the sites with most concentrated sewage spills. Furthermore, if a site has been selected for remediation, decomposition can direct appropriate solutions. For example, Nature Based Solutions (NBS) are a commonly proposed sewer spill intervention [32], and they may reduce spills that occur during high precipitation composition periods, however, they may worsen spills that occur during high soil moisture composition periods and will either not impact or potentially worsen, depending on their spatial extent, spills that occur during high groundwater composition periods.

A variety of research fields benefit from more granular detail in the influent to a WWTP. Provision of municipal composition is directly usable for wastewater epidemiologists [9], and, if paired with sample timings, may reduce the large uncertainty in interpretation attributable to the unknown human-waste fraction [33]. In contrast, pairing composition with influent chemical samples may facilitate better quantifying diffuse urban pollution, which also has a significant uncertainty owing to the variability in composition of a given sample [34], and has already been demonstrated for extreme precipitation events in urban drainage systems [35]. More widely, dynamic simulation of wastewater influent is required for analysis and design of WWTP technology [36]. Phenomenological models, such as [37] have demonstrated the value of splitting influent into different processes to generate such simulations. WWCompose profiles could be further developed to link to pollutant sources to predict water quality profiles that achieve similar dynamic water quality which could be used for WWTW performance analysis.

Wastewater networks are the key driver of urban hydrology and are integral even at catchment scale hydrology [38], interacting with fluvial and pluvial flooding. Identifying composition enables better quantification of wastewater processes in both hydrological and hydrodynamic pipe models. The aggregated conceptual representation may be broadly compatible with lumped hydrology models and could potentially be integrated into large scale hydrology initiatives such as CAMELS-GB [39]. For more detailed urban drainage models, results are partially transferrable. Soil moisture representation in the Stormwater Management Model (SWMM) [40], a popular and open-source model for urban drainage modelling, includes a Green-Ampt infiltration model and power-law relationship for soil-to-pipe inflows as in WWCompose. Although this behaviour is defined in SWMM as a groundwater representation, based on the analysis presented in this work, it would more accurately be described to capture soil moisture-driven infiltration. The prevalence of groundwater contributions in some locations highlights the importance of enabling large scale aquifer groundwater table overtopping in SWMM-like models, which currently requires external packages to implement [41]. As with sewer spills, the integrated picture provided by

WWCompose may be used to guide flood interventions, helping to understand the interplay between precipitation-driven, soil moisture-driven, or groundwater-driven flooding.

4 Conclusions

National-scale wastewater composition can be inferred from open data with accuracy typically $>$ 0.6 KGE across English WWTPs, the largest such assessment demonstrated to date. The results show that non-municipal water is widespread even where municipal flow dominates long-term averages, and that permeability, groundwater, soil wetness, precipitation, and operational controls each leave identifiable signatures in effluent records. Aridity and permeability emerge as the clearest controls on composition, whereas distinct catchment archetypes do not. By turning a routinely monitored but rarely decomposed signal into a source-resolved, time-varying quantity, WWCompose provides a transferable basis for prioritising sewer overflow interventions, sharpening wastewater-based epidemiology, and representing wastewater within hydrological and urban drainage models. More broadly, it demonstrates that the dynamic provenance of wastewater is recoverable wherever comparable open data exist.

5 Methods

Our approach, called Wastewater Compose (WWCompose), extends existing empirical-conceptual studies such as the 'Triangle Method' and 'Moving Minimum' methods proposed by [27] and 'Quasi-linear Model' approach proposed by [25], which further provided dynamic estimates of dry-weather flow components. These studies recognise the difficulty in acquiring sufficient information to fully characterise a sewer network and instead create a conceptual model, splitting wastewater into constituent components, with few parameters that can be fitted to existing data. WWCompose, described in full in 5.1, assumes that the dynamic components of influent (rainfall, soil moisture, and groundwater) must exceed a specified threshold after which they contribute to influent with a linear and exponential component. We use climate data to force the model and, where possible, constrain these parameters with geospatial data, as described in 5.2. WWCompose is fit using best practice split-sample cross-validation and a sequential fitting process with final re-optimization, described in 5.3.

5.1 Model structure

Simulated effluent flow is decomposed into municipal, groundwater, soil-wetness, and effective-precipitation terms:

$$Q_{\mathrm{sim}}(t) = C_{\mathrm{Mun}} + Q_{\mathrm{GW}}(t) + Q_{\theta}(t) + Q_{P}(t). \tag{1}$$

Besides C_{Mun}, which is assumed constant over the study period, each term is calculated by the distance a variable is above a threshold, however, the precise formulation differs for each term, as follows. When bounds are specified below, such as $C_{Mun} \in [0, 1e6]$, these are the bounds within which the optimizer fits parameters, described in further detail in 5.3. Although population and municipal water use may vary seasonally, we consider this simplification acceptable given the larger climate-driven variability of the other sources.

5.1.1 Groundwater

Groundwater depth, $d(t)$ units (m), at the assigned borehole reduces groundwater contributions, and increases when the water table is shallow:

$$Q_{\mathrm{GW}}(t) = C_{\mathrm{GW}} \max\big(T_{\mathrm{GW}} - d(t), 0\big)^{E_{\mathrm{GW}}}, \tag{2}$$

with fitted coefficient $C_{\mathrm{GW}} \in [0, 1e4]$, threshold $T_{\mathrm{GW}} \in [0, 100]$, and exponent $E_{\mathrm{GW}} \in [0, 2]$.

5.1.2 Soil moisture

A scalar soil-store depth $\theta(t)$ (mm) is forced by rainfall $P(t)$ and reduced by $\mathrm{ET}_0(t)$, using a daily Green–Ampt-style recurrence [42]. We use an explicit form to improve computational efficiency during fitting:

$$\theta^*(t) = \max\big(\theta(t-1) - \mathrm{ET}_0(t), 0\big), \tag{3}$$

$$f(t) = \begin{cases} K_s(\psi\,\Delta\theta + \varepsilon)/\varepsilon, & \theta^*(t) < \varepsilon, \\ K_s(1 + \psi\,\Delta\theta/\theta^*(t)), & \theta^*(t) \geq \varepsilon, \end{cases} \tag{4}$$

$$\theta(t) = \theta^*(t) + \min\big(P(t), f(t)\big). \tag{5}$$

Here $\Delta\theta = \theta_s - \theta_i$ with θ_s and θ_i dimensionless saturated water content and field capacity respectively; P, ET_0, θ, and θ^* use depth units (mm). K_s is saturated hydraulic conductivity, ψ is wetting-front suction head, $f(t)$ is infiltration capacity, and ε is set to $1e-6$ to avoid division by 0. In the lower branch, ψ multiplies only the moisture deficit $\Delta\theta$, not cumulative storage θ^*. When $\theta^*(t) < \varepsilon$, the upper branch is a numerical guard; infiltration is then unconstrained by capacity ($f \gg P(t)$), so all rainfall infiltrates and $\theta(t) = \theta^*(t) + P(t)$. We fit K_s, ψ, θ_s, and θ_i within the soil probability bounds within WWTP catchment, calculated using SoilGrids (§5.2).

The soil-wetness contribution to flow is thus,

$$Q_\theta(t) = C_\theta\ \max\big(\theta(t) - T_\theta, 0\big)^{E_\theta}, \tag{6}$$

with fitted $C_\theta \in [0, 1e4]$, $T_\theta \in [0, 500]$, and $E_\theta \in [0, 2]$.

5.1.3 Precipitation

In addition to the generic value over threshold term, for example, eq. 6, effective precipitation also uses capped rainfall excess intended to represent the capacity of the urban drainage system,

$$e_P(t) = \min\left(\max(P(t) - T_P, 0), T_P^{\mathrm{ub}}\right) \tag{7}$$

Furthermore, unlike the other components which cannot be meaningfully scaled, we multiply precipitation by the impervious area, A_{imp}, calculated using [43][44],

$$\widetilde{Q}_P(t) = C_P\, A_{\mathrm{imp}}\, e_P(t)^{E_P} \tag{8}$$

To capture the influence of semi-pervious surfaces reaching saturation and then contributing to precipitation-driven runoff, we include an interaction with antecedent soil moisture, $\theta_{\mathrm{lag}}(t)$:

$$g_P(t) = \min\left(\max\left(\frac{\theta_{\mathrm{lag}}(t) - T^{\mathrm{lo}}_{P,\mathrm{gate}}}{T^{\mathrm{hi}}_{P,\mathrm{gate}} - T^{\mathrm{lo}}_{P,\mathrm{gate}}}, 0\right), 1\right) \tag{9}$$

The precipitation term is

$$Q_P(t) = g_P(t)\,\widetilde{Q}_P(t). \tag{10}$$

Fitted parameters are $C_P \in [0,1]$, $T_P \in [0,100]$, $T_P^{\mathrm{ub}} \in [0,100]$, $E_P \in [0,2]$, and the gate limits $T^{\mathrm{lo}}_{P,\mathrm{gate}} \in [0,400]$, $T^{\mathrm{hi}}_{P,\mathrm{gate}} \in [0,600]$.

5.2 Data sources and alignment

We use daily effluent flow records for English WWTPs from 2010–2024, [30], to fit WWCompose, requiring the assumption that, at a daily timescale, influent and effluent will be comparable. Diurnal variability and stormwater storage will introduce errors here, but we consider these acceptable given that influent is less widely monitored than effluent. The analysis only uses effluent observations certified as good quality in the raw data.

HadUK gives catchment-mean daily rainfall [45], which drives both the soil store and the excess-rainfall term (§5.1). ET_0 comes from HadUK via FAO-56 Penman–Monteith [46].

SoilGrids (sand, silt, clay, bulk density) at a fixed depth supplies the values at uncertainty quantiles (Q0.05, Q0.95) [47]. Rosetta maps each to van Genuchten parameters [48]; these yield Green–Ampt K_s, ψ from K_s, θ_s, and θ_i (here at $-33\,\mathrm{kPa}$) [42]. The quantiles defines optimisation bounds for K_s, ψ, θ_s, and θ_i.

Borehole groundwater depths are retrieved from the English Environment Agency's hydrology data API, and assigned to WWTPs. If the associated borehole data series is absent or constant, no groundwater term is present in Eq. 1.

Permeability class K, used to interpret results, is assigned from superficial deposits where present and bedrock where absent, using expert knowledge informed by lithological descriptions of the British Geological Survey's $1:50,000$ scale superficial and bedrock geology maps.

5.3 Calibration and validation

Calibration fits daily simulated totals to observed effluent over the aligned window at each site within the study period 2010–2024 (§5.2). Parameters are estimated by maximising Kling–Gupta efficiency (KGE) on raw daily flows:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \tag{11}$$

where r is correlation, α is the variability ratio, and β is the mean ratio between simulated and observed flows.

Each aligned record is partitioned into ten contiguous time splits. For each repetition, one split is withheld for validation and parameters are trained on the remaining splits. The held-out segment moves across repetitions so that out-of-sample skill is summarised over different hydro-meteorological periods.

WWCompose contains 17 parameters and the objective is non-convex, so we use sequential optimization with joint refinement. Parameters are fit in reverse order of assumed variability of the component that they refer to (i.e., municipal, then groundwater, then soil moisture, then precipitation). A joint refinement then optimises all fitted parameters simultaneously. Trivially tied solutions (i.e., a $< 1e-6KGE$ difference) are forced to favour the more physically explainable process; WWCompose prefers municipal over groundwater and rainfall over soil-moisture because conventional knowledge suggests that these are the more direct routes for influent to enter the sewer network.

Every fitting stage uses differential evolution, a population-based global search, followed by local polishing of the best candidate (bound-constrained quasi-Newton refinement of the same objective). We use the implementation from Python SciPy [49]. The results shown in this paper use a population size of 250 with maximum iterations of $3,000$, which testing revealed to provide converged objectives. These settings for WWCompose can be run for an individual WWTP on a MacBook Air M2 2022 in under 30 minutes, while the full set of experiments were performed using High-Performance Computing, see Acknowledgements.

References

[1] Staufer, P., Scheidegger, A. & Rieckermann, J. Assessing the performance of sewer rehabilitation on the reduction of infiltration and inflow. *Water Research* **46**, 5185–5196 (2012). URL https://doi.org/10.1016/j.watres.2012.07.001.

[2] Becker, B., Reichel, F., Bachmann, D. & Schinke, R. High groundwater levels: Processes, consequences, and management. *WIREs Water* **9** (2022). URL https://doi.org/10.1002/wat2.1605.

[3] Thapa, J. B., Jung, J. K. & Yovichin, R. D. A qualitative approach to determine the areas of highest inflow and infiltration in underground infrastructure for urban area. *Advances in Civil Engineering* **2019** (2019). URL https://doi.org/10.1155/2019/2620459.

[4] Lipp, A. G. & Dobson, B. Infiltration drives a heavy–tailed distribution of combined sewer overflow spill durations. *Environmental Research Communications* **8**, 041017 (2026).

[5] Thorndahl, S., Schaarup-Jensen, K. & Rasmussen, M. R. On hydraulic and pollution effects of converting combined sewer catchments to separate sewer catchments. *Urban Water Journal* **12**, 120–130 (2013). URL https://doi.org/10.1080/1573062x.2013.831915.

[6] Giakoumis, T. & Voulvoulis, N. The pollution load of combined sewer overflows and risks to england's waterbodies: relating event duration monitoring data to discharge consents from wastewater treatment works. *Environmental Science: Water Research & Technology* (2026).

[7] Carneiro, R. B. *et al.* A critical review of wastewater-based epidemiology as a tool to evaluate the unintentional human exposure to potentially harmful chemicals. *Analytical and bioanalytical chemistry* **417**, 495–511 (2025).

[8] Dirckx, G., Fenu, A., Wambecq, T., Kroll, S. & Weemaes, M. Dilution of sewage: Is it, after all, really worth the bother? *Journal of Hydrology* **571**, 437–447 (2019). URL https://doi.org/10.1016/j.jhydrol.2019.01.065.

[9] Guerrero-Latorre, L. *et al.* The catalan surveillance network of sars-cov-2 in sewage: design, implementation, and performance. *Scientific reports* **12**, 16704 (2022).

[10] Pangle, L. A., Diem, J. E., Milligan, R., Adams, E. & Murray, A. Contextualizing inflow and infiltration within the streamflow regime of urban watersheds. *Water Resources Research* **58** (2022). URL https://doi.org/10.1029/2021wr030406.

[11] Karpf, C. & Krebs, P. Modelling of groundwater infiltration into sewer systems. *Urban Water Journal* **10**, 221–229 (2013). URL https://doi.org/10.1080/1573062x.2012.724077.

[12] Li, B., Chen, J. & Long, M. Quantifying cell binding kinetics mediated by surface-bound blood type b antigen to immobilized antibodies. *Chinese Science Bulletin* **53**, 3634–3641 (2008). URL https://doi.org/10.1007/s11434-008-0517-9.

[13] Ye, L., Qian, Y., Zhu, D. Z. & Huang, B. Inflow and infiltration assessment of a prototype sanitary sewer network in a coastal city in china. *Water Science and Technology* **88**, 2940–2954 (2023). URL https://doi.org/10.2166/wst.2023.386.

[14] Zeydalinejad, N., Javadi, A. A. & Webber, J. L. Global perspectives on groundwater infiltration to sewer networks: A threat to urban sustainability. *Water Research* **262**, 122098 (2024). URL https://doi.org/10.1016/j.watres.2024.122098.

[15] Su, X., Liu, T., Beheshti, M. & Prigiobbe, V. Relationship between infiltration, sewer rehabilitation, and groundwater flooding in coastal urban areas. *Environmental Science and Pollution*

Research **27**, 14288–14298 (2019). URL https://doi.org/10.1007/s11356-019-06513-z.

[16] Zeydalinejad, N., Javadi, A. A., Jacob, M. & Webber, J. L. Evaluating groundwater infiltration risk in sewer networks under climate change. *Stochastic Environmental Research and Risk Assessment* **40**, 41 (2026).

[17] Su, X., Belvedere, P., Tosco, T. & Prigiobbe, V. Studying the effect of sea level rise on nuisance flooding due to groundwater in a coastal urban area with aging infrastructure. *Urban Climate* **43**, 101164 (2022). URL https://doi.org/10.1016/j.uclim.2022.101164.

[18] Wittenberg, H. & Aksoy, H. Groundwater intrusion into leaky sewer systems. *Water Science and Technology* **62**, 92–98 (2010). URL https://doi.org/10.2166/wst.2010.287.

[19] Martínez, C., Vojinovic, Z., Price, R. & Sanchez, A. Modelling infiltration process, overland flow and sewer system interactions for urban flood mitigation. *Water* **13**, 2028 (2021). URL https://doi.org/10.3390/w13152028.

[20] Zeydalinejad, N. An overview of the methods for evaluating the resilience of groundwater systems. *MethodsX* **10**, 102134 (2023). URL https://doi.org/10.1016/j.mex.2023.102134.

[21] Zhang, M. *et al.* Quantifying rainfall-derived inflow and infiltration in sanitary sewer systems based on conductivity monitoring. *Journal of Hydrology* **558**, 174–183 (2018). URL https://doi.org/10.1016/j.jhydrol.2018.01.002.

[22] Dunning, C., Feinstein, D., Hunt, R. & Krohelski, J. Simulation of ground-water flow, surface-water flow, and a deep sewer tunnel system in the menomonee valley, milwaukee, wisconsin (2004). URL https://doi.org/10.3133/sir20045031.

[23] Zeydalinejad, N. Artificial neural networks vis-à-vis modflow in the simulation of groundwater: a review. *Modeling Earth Systems and Environment* **8**, 2911–2932 (2022). URL https://doi.org/10.1007/s40808-022-01365-y.

[24] Aßmann, E. *et al.* Augmentation of wastewater-based epidemiology with machine learning to support global health surveillance. *Nature Water* **3**, 753–763 (2025).

[25] Karpf, C. & Krebs, P. Quantification of groundwater infiltration and surface water inflows in urban sewer networks based on a multiple model approach. *Water Research* **45**, 3129–3136 (2011). URL https://doi.org/10.1016/j.watres.2011.03.022.

[26] Dirckx, G., Bixio, D., Thoeye, C., De Gueldre, G. & Van De Steene, B. Dilution of sewage in flanders mapped with mathematical and tracer methods. *Urban Water Journal* **6**, 81–92 (2009). URL https://doi.org/10.1080/15730620802541615.

[27] Weiß, G., Brombach, H. & Haller, B. Infiltration and inflow in combined sewer systems: long-term analysis. *Water Science and Technology* **45**, 11–19 (2002). URL https://doi.org/10.2166/wst.2002.0112.

[28] Jenssen Sola, K., Bjerkholt, J. T., Lindholm, O. G. & Ratnaweera, H. Infiltration and inflow (i/i) to wastewater systems in norway, sweden, denmark, and finland. *Water* **10**, 1696 (2018).

[29] Lipp, A. G., Dawe, J. & Balaji, S. www. sewagemap. co. uk and poopy: Open-source tools for understanding and communicating the environmental impacts of combined sewer overflows in real-time. *EarthArXiv* (2025).

[30] Maes-Prior, R., Dobson, B. & Mijic, A. Drivers of wastewater dynamics: A statistical analysis of england's large wastewater treatment works. *Water Science and Technology* 295 (2026).

[31] Department for Environment, Food and Rural Affairs. Storm overflows discharge reduction plan. Tech. Rep., HM Government, London (2023). URL https://www.gov.uk/government/publications/storm-overflows-discharge-reduction-plan. Expanded plan, updated 25 September 2023.

[32] Moghanlo, S. J. & Raimondi, A. Impacts of blue-green infrastructures on combined sewer overflows. *Nature-Based Solutions* **7**, 100208 (2025).

[33] van der Drift, A.-M. R. *et al.* Wastewater surveillance studies on pathogens and their use in public health decision-making: a scoping review. *Science of the Total Environment* **993**, 179982 (2025).

[34] Chen, Y., Yang, Z., Dong, J., Hong, N. & Tan, Q. Understanding phosphorus fractions and influential factors on urban road deposited sediments. *Science of the Total Environment* **921**, 170624 (2024).

[35] Hu, J., Ouyang, W., Hou, C., Ji, K. & Yang, Z. Effects of urban stormwater pollution on watershed diffuse loads under extreme precipitation conditions. *Journal of Hydrology* **654**, 132802 (2025).

[36] Martin, C. & Vanrolleghem, P. A. Analysing, completing, and generating influent data for wwtp modelling: A critical review. *Environmental Modelling & Software* **60**, 188–201 (2014).

[37] Gernaey, K. V., Flores-Alsina, X., Rosen, C., Benedetti, L. & Jeppsson, U. Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. *Environmental Modelling & Software* **26**, 1255–1267 (2011).

[38] Coxon, G. *et al.* Wastewater discharges and urban land cover dominate urban hydrology signals across england and wales. *Environmental Research Letters* **19**, 084016 (2024).

[39] Coxon, G. *et al.* CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in great britain. *Earth System Science Data* **12**, 2459–2483 (2020). URL https://doi.org/10.5194/essd-12-2459-2020.

[40] Gironás, J., Roesner, L. A., Rossman, L. A. & Davis, J. A new applications manual for the storm water management model (swmm). *Environmental Modelling and Software* **25**, 813–814 (2010). URL https://doi.org/10.1016/j.envsoft.2009.11.009.

[41] Zheng, Y., Chen, S., Qin, H. & Jiao, J. J. Modeling the spatial and seasonal variations of groundwater head in an urbanized area under low impact development. *Water* **10**, 803 (2018).

[42] Green, W. H. & Ampt, G. A. Studies on soil physics: Part i. the flow of air and water through soils. *The Journal of Agricultural Science* **4**, 1–24 (1911). URL https://doi.org/10.1017/S0021859600001441.

[43] Dobson, B., Jovanovic, T., Alonso-Álvarez, D. & Chegini, T. Swmmanywhere: A workflow for generation and sensitivity analysis of synthetic urban drainage models, anywhere. *Environmental Modelling & Software* **186**, 106358 (2025).

[44] Dobson, B., Alonso-Álvarez, D. & Chegini, T. Swmmanywhere: Synthesise urban drainage network models anywhere in the world. *Journal of Open Source Software* **10**, 7729 (2025).

[45] Hollis, D., McCarthy, M., Kendon, M., Legg, T. & Simpson, I. HadUK-Grid—a new UK dataset of gridded climate observations. *Geoscience Data Journal* **6**, 151–159 (2019). URL https://doi.org/10.1002/gdj3.78.

[46] Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. *Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56* (Food and Agriculture Organization of the United Nations, Rome, 1998). FAO Irrigation and Drainage Paper No. 56.

[47] Poggio, L. *et al.* SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* **7**, 217–240 (2021). URL https://doi.org/10.5194/soil-7-217-2021.

[48] Schaap, M. G., Leij, F. J. & van Genuchten, M. T. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* **251**, 163–176 (2001). URL https://doi.org/10.1016/S0022-1694(01)00466-8.

[49] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).

Acknowledgements. Co-author Jackson was supported by the British Geological Survey via NERC national capability funding. Co-author Maes-Prior was supported by the Engineering and Physical Sciences Research Council and Mott MacDonald EP/X524773/1. We acknowledge computational resources and support provided by the Imperial College Research Computing Service (http://doi.org/10.14469/hpc/2232).