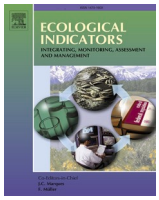
Ecological Indicators 176 (2025) 113650

Contents lists available at ScienceDirect

Ecological Indicators

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



Original Articles

Green and blue infrastructure but not anthropogenic drivers affect invertebrate communities in urban ecosystems

Joeri Morpurgo [a,*], Roy P. Remme [a], Quinten D. Mudde [b], Emilie A. Didaskalou [a], Kornelia J. Serwatowska [a], Krijn B. Trimbos [a], Mingming Hu [a], Peter M. van Bodegom [a]

[a] Institute of Environmental Sciences CML, Leiden University, Einsteinweg 2, 2333 CC Leiden, the Netherlands
[b] Royal Netherlands Institute of Sea Research, Estuarine and Delta Systems (NIOZ), 4400 AC Yerseke, the Netherlands

ARTICLE INFO

Keywords:
Biodiversity
Green infrastructure
Remote sensing
Species distribution model
Urban ecology
Urban planning

ABSTRACT

Rapid urbanization alters habitat quality and connectivity, influencing species dispersal and ultimately shaping community assembly. Increasingly, urban environments and their vegetation are shown to be important for the conservation of biodiversity. However, urban species distributions remain poorly understood. By integrating DNA-based sampling with species distribution models (SDMs), we aim to quantitatively assess urban community assembly while using a novel green infrastructure classification framework intended to assess both urban biodiversity and ecosystem services.

We sampled invertebrate distributions in The Hague, the Netherlands, using two complementary DNA-based methods: traditional bulk trapping (n = 205) and environmental DNA (eDNA) sampling (n = 207). Species were identified using DNA sequencing and Operational Taxonomic Units, with presence-absence data used to develop SDMs driven by vegetation and anthropogenic indicators.

The SDMs generally outperformed random models (59.5 %) and performed strong during calibration (90.4 %, AUC > 0.70). They highlighted that vegetation density, structure, and proximity to water are the primary drivers of invertebrate distributions, while direct anthropogenic pressures play a minimal role. At the same time, very few models (1.3 %) performed well during validation.

This finding indicates challenges in predicting species distributions and suggest that dispersal is not a limiting factor in The Hague's urban environment, as current green infrastructures appear sufficient to support many species. The model overfitting and low validation performance also indicate the need to refine biodiversity indicators for urban environments, as typically used vegetation indicators do not predict species distributions well. The absence of dispersal limitations that suggest that the urban environment acts as one large *meta*-community, indicates that ensuring sufficient green infrastructure in the urban environment should be the first priority to enhance biodiversity in the urban environment.

1. Introduction

Cities have long been considered ecological deserts. With urbanization rapidly increasing in many parts of the world, these ecological deserts increasingly replace the original habitat and local species are lost where cities are built (Güneralp and Seto, 2013). This loss in biodiversity is exacerbated by the tendency to build cities in biodiversity hotspots. Recent models suggest that a third of the land-dwelling vertebrates are affected by urbanization (Simkin et al., 2022). Within cities, new habitat is built, often with properties very different from those of natural habitat. A strategy to mitigate biodiversity loss is the implementation of urban green infrastructure. (i.e. a network of natural area managed to deliver ecosystem services and protect biodiversity in cities (Tzoulas et al., 2007; European Commission, 2014)). However, compared to the natural areas, urban green infrastructure exhibits uniquely high levels of anthropogenic and environmental stressors. The anthropogenic stressors include human disturbance, introduction of exotic species, nutrient loading, and routine green maintenance among others (Peters et al.,

* Corresponding author.
E-mail addresses: j.morpurgo@cml.leidenuniv.nl (J. Morpurgo), r.p.remme@cml.leidenuniv.nl (R.P. Remme), quinten.mudde@nioz.nl (Q.D. Mudde), e.a.didaskalou@cml.leidenuniv.nl (E.A. Didaskalou), k.serwatowska@cml.leidenuniv.nl (K.J. Serwatowska), trimbos@cml.leidenuniv.nl (K.B. Trimbos), hu@cml.leidenuniv.nl (M. Hu), p.m.van.bodegom@cml.leidenuniv.nl (P.M. van Bodegom).

https://doi.org/10.1016/j.ecolind.2025.113650
Received 17 February 2025; Received in revised form 24 April 2025; Accepted 21 May 2025
Available online 4 June 2025

2015; Orr et al., 2022; Vilmi et al., 2017). Simultaneously, the cities provide harsh environments due to their more extreme climate, comprising of the urban heat island, droughts and flash floods (Masson et al., 2020). Urban dwelling species will necessarily have to deal with these abundant stressors (McCloy et al., 2024).

Despite these prevailing stressors, cities are also increasingly being recognized as refuges and potential hotspots for biodiversity, known to harbour even rare, endangered and red-listed species (Boakes et al., 2023). Simultaneously, evidence is starting to mount that core ecological drivers, such as increasing habitat availability, increasing connectivity among green patches and reducing stressors, are readily enhancing urban biodiversity (Beninde et al., 2015; LaPointe et al., 2015). Several reviews and *meta*-analyses show that both the green infrastructure and man-made built infrastructure shape urban biodiversity. In a *meta*-analysis of 87 studies that investigated the effect of green infrastructure on urban biodiversity, species richness was best predicted by habitat richness, vegetation structure and density of green infrastructure (Beninde et al., 2015). In parallel, a review on the spatial connectivity of urban green infrastructure at city scale does show the importance of ecological connectivity for understanding species distributions (LaPointe et al., 2015). Subsequently, the design of built infrastructure and anthropogenic conditions, such as urban form, policy and the socioeconomic status of neighbourhood communities also seem to be shaping urban biodiversity (Kuras et al., 2020; Schell et al., 2020). For example, small urban parks within denser urban environments have less diverse communities due to a reduction in the number of endemic species and species with low tolerance to urban environments (Amaya-Espinel et al., 2019; Fischer et al., 2015). Likewise, neighbourhoods with higher socioeconomic status tend to have higher plant and animal diversity (Leong et al., 2018). These studies indicate that urban biodiversity is moulded by cities, their green and built infrastructures, layout and use.

Currently, knowledge on important urban biodiversity drivers is growing, yet empirical evidence on the drivers of individual distributions of species and their assembly into communities is largely lacking (Pickett et al., 2011; Andrade et al., 2021). To support urban biodiversity, better understanding urban species environmental requirements regarding the design of the quantity, spatial lay-out and quality of urban green and built infrastructure is key. These spatial features and their design are critical for a plethora of ecological processes such as dispersal and functioning of trophic networks. Considering green and built features as additional environmental filters next to I) regional climate, II) spatial and socio-economic urban development, III) species interactions and IV) their functional traits, may help us understand how urban species assemble in to communities (Aronson et al., 2016; Zhao et al., 2023). If traditional ecological theories developed in natural habitats apply, conventional actions to preserve target species and their associated biodiversity should prove effective in cities. Yet, if species and their assembly into communities differ in urban environments, different approaches for biodiversity support are needed in cities. By understanding the critical features of urban community assembly, we can further elucidate the underlying mechanisms of urban biodiversity, and in turn develop guidelines that support, protect, and conserve urban target species and biodiversity more effectively (IUCN et al., 2022), ultimately finetuning the critical design factors of urban green infrastructure to support urban biodiversity.

Invertebrates provide crucial ecosystem services and are essential for ecosystem health through pollination, nutrient cycling, soil bioturbation and they serve as crucial links in trophic networks (Buchholz and Egerer, 2020; Belovsky and Slade, 2000; Drager et al., 2016; Eggleton, 2020). For example, flower-visiting insects are indispensable for pollination crucial for sustaining plant and invertebrate populations throughout the urban and *peri*-urban environment (Eggleton, 2020). The trophic cascade from plants to invertebrates to soil nutrients is highly important for plant diversity, soil health, and overall ecosystem health (Belovsky and Slade, 2000). Despite their evident importance, invertebrate numbers in protected areas continue to dwindle at alarming rates (Hallmann et al., 2017). This decline is known to threaten the stability of many critical ecosystem functions (Wagner et al., 2021). Here, urban environments may also serve as genetic reservoirs as some animal species have been recorded to disperse from rural to urban environments and vice versa (Brashear et al., 2015; Varner et al., 2014). Yet, substantial evidence of cities acting as genetic reservoirs for invertebrates is lacking. Evidence does suggest that urban environments can enhance geneflow for some species, but the geneflow to rural areas remains unknown (Collins et al., 2024). Simultaneously, urban environments are known to harbour rare and endangered species different from local communities, yet their drivers are poorly known (Andrade et al., 2021; Boakes et al., 2023). Considering the above, there is an urgent need to study the drivers of invertebrate species fundamental niche and their resulting distributions so that we can design cities and their layout with green infrastructure that meaningfully supports invertebrate communities.

Understanding the drivers of urban invertebrate species can be achieved by Species Distribution Modelling (SDM). These models provide estimates of the importance of environmental features by modelling species presence and absence against the environmental variables (Dutta et al., 2022; Naimi and Araújo, 2016). To date, such an approach is rare at the city scale, but has been shown to be successful for plants, leafhoppers and grasshoppers in identifying their drivers (Kattwinkel et al., 2009). In parallel, recent developments in DNA-sequencing techniques, specifically environmental DNA (eDNA) and bulk metabarcoding, now allow for rapid sampling of whole communities (Cristescu and Herbert, 2018, Gleason et al., 2021). Integrating the novel DNA-sequencing with SDMs allows us to rapidly assess the entire community assembly at the landscape scale.

Through the integration of DNA-based data and SDMs, we aim to understand urban species, their distribution, and their drivers, with unprecedented taxonomic coverage and spatial coverage in a city. Through this integrative framework, we seek to identify the primary ecological and landscape-level drivers that shape invertebrate assemblages in the urban environment. The framework enables us to investigate the effects of urban green and blue infrastructure on species and their distribution and draw inferences on ecological connectivity and biodiversity hotspots. This knowledge can be translated into guidelines to make it easier for urban planners and municipalities to develop strategies that effectively support and enhance urban biodiversity through target species (Kirk et al., 2021; Pierce et al., 2024).

2. Methods

We investigated urban invertebrate species distributions in a three-step process. First, we selected relevant sampling sites for invertebrate trapping and soil collection through conditional Latin Hypercube Sampling (cLHS). Second, we extracted, amplified, and sequenced the DNA from both bulk invertebrate samples and soil samples. Sequenced samples were processed with the DADA2 bioinformatics pipeline, clustered into Operational Taxanomic Units (OTUs) and BLASTed against the BOLD reference database to assign taxonomy. Finally, we modelled the identified species and their distributions with predictors thought to be ecologically important, at multiple spatial resolution. The resulting SDMs were used to infer ecological hotspots at the species level and to assess the most important drivers of their distributions.

2.1. Sampling and location

Invertebrates and soil were sampled once per location using the DNA from bulk invertebrate and soil eDNA in a broad range of green infrastructure types in The Hague, The Netherlands (52∘04′57.0″N 4∘17′40.3″E). The sampling included a mix of sparsely to densely vegetated sites, and green spaces with short and tall vegetation, including parks, residential green, natural areas and buffer strips

(Morpurgo et al., 2023). We excluded the Leidschenveen district in our sampling design, as it was both isolated from the most of The Hague and also logistically infeasible to get there given our sampling equipment and design. We sampled 207 sites on non-rainy days from the 15th of June until the 4th of August 2022 (Fig. 1). Sampling consisted of collecting soil and setting traps on Monday or Tuesday, and retrieving traps two days later on Wednesday and Thursday.

Sample site selection aimed to cover the full range of indicators expected to be of influence on invertebrate diversity. To ensure the full coverage of the relevant indicators, we used conditional Latin Hypercube Sampling (cLHS). This method is recommended in cases where sampling is logistically limited, as it allows for maximum coverage in feature space with the least amount of samples. During the sampling campaign, new random sites were drawn on a weekly basis, based on cLHS, while accounting for already sampled spots and excluding potential sampling locations that were private through visual assessment with Google Maps. This dynamic sampling approach also aims to prevent spatial and temporal clustering of the sample locations. The indicators used in the cLHS were largely based on a review covering green infrastructure and its relation to several biodiversity taxa (Morpurgo et al., 2023). Our included indicators cover the basic needs for invertebrates, such as food, shelter and water. In addition, we added data on buildings and land uses to fully represent the urban landscape (Table S1).

At every site, we sampled with both pitfall and pan traps (n = 205, Fig. 1) and soil eDNA (n = 207, Fig. 1). By combining traditional trapping and eDNA, we were able to both compare methods and capture most invertebrate biodiversity through off-setting any potential trapping biases (Montgomery et al., 2021). The combination of pan and pitfall traps allows for capturing both species that have high dispersal capacity (i.e. flying insects) and species with low dispersal capacity (i.e. walking or crawling insects), which are likely to face different barriers in cities. For the soil eDNA sample, we dug up approximately 500 ml of topsoil that was mixed in a plastic bag. From this mixture, a sample of 50 ml soil was stored in a falcon tube and stored in a freezer at −20 °C. For the invertebrate bulk sampling, one pitfall and three colours of pan traps (white, yellow, blue) were set with 0.1 % perfume-free and colourless soap per location (Everingham, 2023). The pitfall trap was transparent and measured 80 mm length by 115 mm diameter, which is considered optimal given the relatively small size of our targeted invertebrate species (Brown and Matthews, 2016). Yellow and blue pan traps were 40 mm deep and 105 mm in diameter. The white pan traps were 55 mm deep and 110 mm in diameter. The selected colours are known to attract different pollinator species, and thus we combined them to off-set any species-colour related biases (Vrdoljak and Samways, 2012). After 48 h, trapped invertebrates were filtered through coffee filters, and scraped into 50 ml falcon tubes containing 96 % ethanol and stored in a freezer at −20 °C.

2.2. DNA extractions

To extract eDNA from soil, we added 15 ml of saturated phosphate buffer (Na_2HPO_4; 0.12 m) to 15 g of soil and homogenized the samples by vigorously shaking, followed by stirring at 600 rpm for 10 min for further homogenization (Taberlet et al., 2012). After shaking, 2 ml of the soil-buffer mixture was transferred in a 2.5 ml tube and samples were further extracted using the Qiagen DNeasy PowerSoil Pro Kit following the manufacturer's protocol starting from step 3. The resulting eDNA extracts were stored at −20 ℃ until further use.

We used an adjusted protocol from Beentjes et al. (2019) to extract DNA from the invertebrate bulk samples. First, we homogenized the bulk using the IKA Ultra-Turrax Tube Drive with 15 steel beads of 5 mm diameter per 50 ml tube for 5 min at max speed. For each sample, we transferred 2 ml of the mixture into a 2.5 ml tubes twice, with each tube serving as a subsample. We centrifuged the subsamples at 16,000 g for 3 min to separate the homogenized invertebrate bulk from the ethanol. After, we removed the supernatant ethanol and dried the samples at 50 ℃ until visually dry. The dried subsamples were combined and digested using 20 µL proteinase K and 180 µL ATL buffer for 24 h at 56 ℃ and 600 rpm. The digested samples were centrifuged at 10000g for 3 min to separate the non-digestible material from the eluted DNA. If eluted DNA did not separate, the sample was centrifuged at 16,000 g for 3 min.

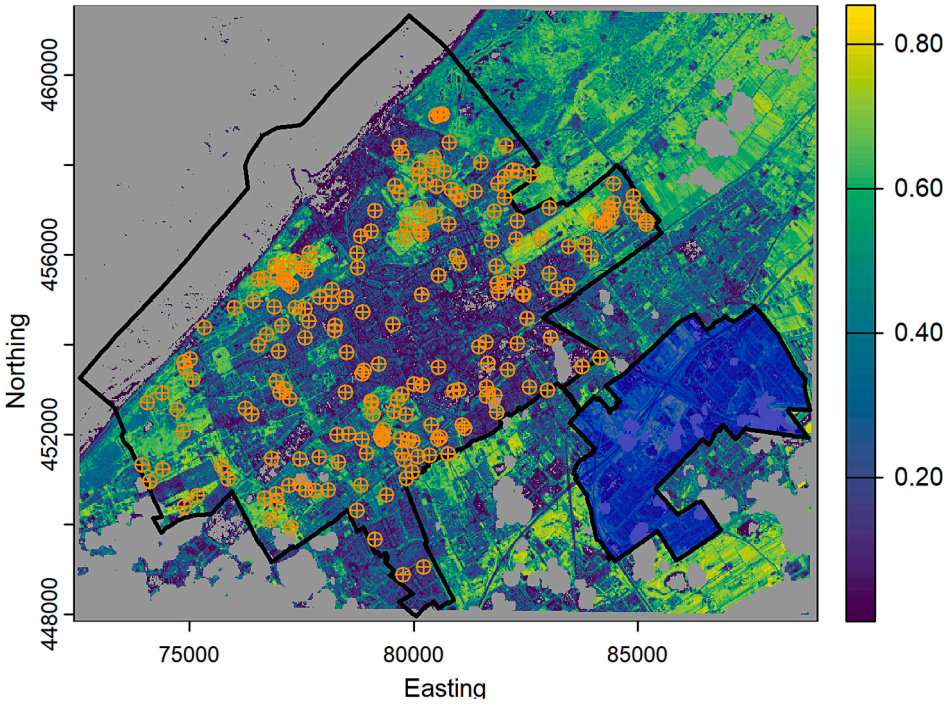


Fig. 1. Sampling locations against a background of scaled NDVI in The Hague, The Netherlands. The orange crossed circles indicate a sample being taken there. The blue polygon indicates the excluded zone (Leidschenveen). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Finally, the recovered supernatant containing the DNA was transferred into a new tube and we followed the manufacture's protocol from step 3 onwards. Resulting DNA extracts were stored at −20 ℃ until further use.

2.3. PCR amplification and DNA sequencing

For both soil and invertebrate samples, we targeted the mitochondrial gene Cytochrome *c* oxidase I (COI), but used two different primer sets that previous research suggested to be optimal for assessing general invertebrate communities for soil eDNA and bulk DNA extracts (Leray et al., 2013; Elbrecht et al., 2019). For the soil eDNA extracts, we used the primer mlCOIintF and dgHCO2198 resulting in an amplicon of 313 base pairs (Table S2; Leray et al., 2013). For invertebrate bulk DNA extracts, we used the forward primer BF1 and the reverse primer BR2 resulting in an amplicon of 316 base pairs (Table S2; Elbrecht et al., 2019). Both primer sets included Illumina tails, to attach indexing sequences required for DNA sequence demultiplexing.

For the first PCR round, samples were amplified in triplicate in 25 μL reactions containing 2 μL of DNA extract, 8 μL of DNase/RNase free water, 12.5 μL Taqman, Environmental Master Mix 2.0 and 2.5 μL of each primer (10uM). PCR was performed in a Bio-Rad C1000 touch Thermal Cycler, using the following program: initial denaturation at 98 ℃ for 10 min, 30 cycles of denaturation at 95 ℃ for 15 s, annealing at 45 ℃ for 30 s, and extension at 72 ℃ for 40 s, followed by a final elongation at 72 ℃ for 5 min. Samples were visually checked on 2 % precast agarose E-gels™ with SYBR™ Safe (Invitrogen) to ensure successful amplification before proceeding. All samples' triplicates were combined into one library and cleaned with NucleoMag NGS Clean-up and Size Select beads (MACHEREY-NAGEL) at a ratio of 0.9:1 (beads/sample). For the second PCR round, libraries were dual indexed in 20 μl reactions containing 2 μl of library (1st PCR product), 6 μl of DNase/Rnase free water, 10 μl Taqman, Environmental Master Mix 2.0 and 1.5 μl of each IDT10 index (10 μM). The PCR program used was: initial denaturation at 98 ℃ for 10 min, 8 cycles of denaturation at 95 ℃ for 15 s, annealing at 55 ℃ for 30 s, and extension at 72 ℃ for 40 s, followed by a final elongation at 72 ℃ for 5 min.

As sequencing is a random process, equimolar pooling was chosen to equalize the chances of the DNA being sequenced (Muller et al., 2019). We equimolarly pooled the library based on DNA concentrations measured with an 4200 TapeStation from Agilent using the D1000 ScreenTapes. In particular, this type of pooling dilutes samples with abundant PCR products to allow samples with little PCR product to have an equal sequencing chance. The soil and invertebrate bulk pooled libraries were then cleaned with NucleoMag NGS Clean-up and Size Select beads (MACHEREY-NAGEL) at a ratio of 0.9:1 (beads/sample). Finally, the two pooled libraries (soil & invertebrate bulk) were equimolarly pooled and sequenced by the Illumina NovaSeq 6000 SP system on a 250PE flow cell.

2.4. Bioinformatics

To identify the sequenced COI DNA, we used a software combination of DADA2, CutAdapt, VSEARCH, BLAST and the public BOLD reference database. First, we started with DADA2 to filter reads with a higher-than-expected error of 1, excluding low-quality reads (Callahan et al., 2016; Elbrecht et al., 2019). The primers from the filtered reads were removed using Cutadapt, and reads shorter than 300 bp were omitted as they are substantially shorter than the target amplification area (Martin, 2011). After, we used the Big Data Paired-end DADA2 workflow to generate chimera-free amplicon sequence variants, but decreasing the maxEE argument to a stricter 1. Using VSEARCH, the chimera-free amplicon sequence variants were clustered into OTUs using the 97 % similarity cut-off, typically representing a species (Elbrecht et al., 2018; Rognes et al., 2016). These OTUs were identified using BLAST with the BOLD COI public database (BOLDSYSTEMS, 2023; Altschul et al., 1990). The BLAST was set to give a maximum of 10 matches, which also had to align with the expected location of the amplicon within the COI marker. For every OTU, we filtered to a single taxonomic annotation per OTU based on sequentially the highest bitscore, lowest e-value and highest percentage similarity (>70 %).

For the SDMs, we also omitted any OTU that was identified as a non-invertebrate species. To assess the impact of dispersal capacity on the SDM performance, we grouped OTUs in three groups based on estimated dispersal capacity by taxonomic groups. The dispersal groups were invertebrates that primarily fly (flying; *Diptera* and *Hymenoptera*), facultative flying invertebrates that fly and crawl (facultative flying; *Coleoptera, Dermaptera, Hemiptera* and *Orthoptera*) and ground-dwelling invertebrates (ground; *Annelida, Arachnida, Collembola, Diplopoda, Isopoda* and *Symphyla*).

2.5. Species distribution models

We used SDMs based on the presence or absence of the identified species from either the eDNA or bulk metaBarcoding datasets to infer invertebrate species distributions. For robust SDMs, we chose a minimum of 15 presences (out of the 200 + samples) for an OTU to be modelled. The SDMs were modelled across eight spatial resolutions (10*10 m, 20*20 m, 30*30 m, 40*40 m, 50*50 m, 100*100 m, 200*200 m and 500*500 m, to capture maximum foraging and dispersal distances of most invertebrates (Auger et al., 2024)) and two sets of drivers (CUGIC and ECO, Table 1). The first set of drivers focused on land use variables as identified in CUGIC (Consolidated Urban Green Infrastructure Classification, Morpurgo et al. 2023):

$$Species = Vegetationdensity + Vegetationstructure + LULC + \epsilon$$

The second set of drivers, hereafter called ECO, includes more predictors that are considered ecologically important to invertebrates:

$$Species = Vegetationdensity + Vegetationstructure + LULC + LAI + Waterdistance + Addressdensity + Noise + \epsilon.$$

where in both formulas *Species* denotes the presence or absence of an identified invertebrate species. The term *Vegetationdensity* contains numerical information between 0——1 on the density of foliage in a given cell, using scaled NDVI (Carlson and Ripley, 1997). The term *Vegetationstructure* is a numerical indicator between 1 and 2, where 1 describes all vegetation being below 1 m high, and 2 describes all vegetation that is above 1 m high. Upon aggregation to a larger spatial scale, this term describes the proportion of low vs. high vegetation in a cell. The land use and land cover is described categorically in the *LULC* term, including I) natural, II) park, III) residential and IV) buffer strip. For the ECO model, we also included numerical information on leaf area index with the term *LAI*, and distance to water with the term *WaterDistance* to include data on species' needs (Pfeifer et al., 2012; Crocker et al., 2017). *Addressdensity* and *Noise* are indicators that reflect the level of anthropogenic influence that poses potential barriers to species (Den Dulk et al., 1992; Van Der Waals, 2002). The final ∈ term encompasses the remaining error in both the CUGIC and ECO models.

The SDMs used 70 % of the data to calibrate the model and used the remaining 30 % of the data to validate the accuracy of the model. All SDMs were run in the R package SDM, using several model structures (GAM, MDA, FDA, MARS, CART, MAXENT) and replicated 10 times to increase the robustness of the analysis (Naimi and Araújo, 2016). This resulted in 960 models trained and validated per OTU, including 2 sets of drivers, 6 model structures, 8 spatial resolutions and 10 replications. Accuracy was evaluated using traditional AUC values thresholds against the mean of the 480 validation AUCs per OTU (Gorunescu, 2011).

2.6. Drivers of SDM performance

To understand the drivers impacting the SDMs, we assessed their

Table 1
Description of the variables used in the Species Distribution Models and their method of aggregation. The variables LAI, Water distance, Address density and noise only apply to the ECO model. Aggregation methods describe either a mean (Bilinear), a sum method or the most frequent value (modal). Data statistics are identical for both the soil and bulk dataset, unless stated differently.

Variable at 10*10 m	Description	Aggregation method	20*20 m	30*30 m	40*40 m	50*50 m	100*100 m	200*200 m	500*500 m
Species (Reponse variable)	A binary variable representing the presence (1) or absence (0) of an OTU identified to a species at > 97 % identical DNA-match with the BOLD public data package (BOLDSYSTEMS, 2023). The soil eDNA contained 47 OTUs, and the invertebrate bulk contained 163 OTUs.	No aggregation	No aggregation	No aggregation	No aggregation	No aggregation	No aggregation	No aggregation	No aggregation
Vegetation density	A numerical variable that uses the average normalized difference vegetation index (NDVI), retrieved from Sentinel-2 on June — September 2022, and scaled accordingly to both the NDVI values found in our sample sites (NVDI: Min = 0.05; Max = 0.90). The range of scaled vegetation densities was 0.01—–0.75, with an average value of 0.36 and a SD of 0.20.	Bilinear	Min = 0.03 Avg = 0.37 SD = 0.20 Max = 0.77	Min = 0.01 Avg = 0.36 SD = 0.20 Max = 0.77	Min = 0.03 Avg = 0.37 SD = 0.19 Max = 0.76	Min = 0.03 Avg = 0.36 SD = 0.19 Max = 0.75	Min = 0.03 Avg = 0.35 SD = 0.17 Max = 0.75	Min = 0.08 Avg = 0.36 SD = 0.16 Max = 0.74	Min = 0.09 Avg = 0.35 SD = 0.14 Max = 0.63
Vegetation height	A binary variable that uses Vegetation density in combination with AHN4 LiDAR data to classify vegetation below 1 m (1) or vegetation with over 10 % above 1 m (2) (AHN, 2024). Upon aggregation, cells were averaged to create an indicator on vegetation structure, between 0 and 1. The minimum (indicating all vegetation < 1 m) and maximum values (all vegetation above > 1 m) were 1 and 2, respectively. The average value was 1.4 and SD 0.49.	Bilinear	Min = 1 Avg = 1.45 SD = 0.41 Max = 2	Min = 1 Avg = 1.45 SD = 0.37 Max = 2	Min = 1 Avg = 1.46 SD = 0.34 Max = 2	Min = 1 Avg = 1.44 SD = 0.31 Max = 2	Min = 1 Avg = 1.45 SD = 0.27 Max = 2	Min = 1 Avg = 1.45 SD = 0.25 Max = 1.99	Min = 1 Avg = 1.45 SD = 0.20 Max = 1.83
LULC (land-use/ landcover	A categorical variable with four classes that describe the respective classes in the CUGIC classification (Morpurgo et al., 2023). The classes are Park (SOIL = 51; BULK = 50), Natural (Nat, SOIL = 27; BULK = 26), Residential (Resi, SOIL = 101; BULK = 100) and Buffer zone (Buff, SOIL = 28; BULK = 29).	Modal	Park = 50 Nat = 28 Resi = 100 Buff = 29	Park = 51 Nat = 28 Resi = 100 Buff = 28	Park = 49 Nat = 27 Resi = 102 Buff = 29	Park = 51 Nat = 27 Resi = 103 Buff = 26	Park = 48 Nat = 25 Resi = 107 Buff = 27	Park = 42 Nat = 27 Resi = 112 Buff = 26	Park = 46 Nat = 34 Resi = 104 Buff = 23
LAI (Leaf Area Index)	A numerical variable describing the Leaf Area Index as a supplementary vegetation indicator retrieved from SNAP with Sentinel-2 imagery (European Space Agency (ESA), 2023). The range of this value is 0.02—–2.95 with a mean of 0.51 and a SD of 0.37.	Bilinear	Min = 0.07 Avg = 0.54 SD = 0.37 Max = 2.85	Min = 0.02 Avg = 0.52 SD = 0.34 Max = 2.15	Min = 0.10 Avg = 0.55 SD = 0.32 Max = 2.32	Min = 0.08 Avg = 0.53 SD = 0.30 Max = 2.25	Min = 0.07 Avg = 0.55 SD = 0.30 Max = 1.87	Min = 0.11 Avg = 0.55 SD = 0.27 Max = 1.67	Min = 0.15 Avg = 0.52 SD = 0.24 Max = 1.66
Water distance	A numerical variable describing the distance to the nearest fresh-water (Kadaster, 2022b). The range of the values are 0—–782.6. The SOIL mean was 147.7 with a SD of 157.8. The BULK mean was 148.7 with a SD of 157.9	Bilinear	Min = 1 Avg = 147 SD = 157 Max = 781	Min = 0 Avg = 147 SD = 158 Max = 774	Min = 0 Avg = 146 SD = 158 Max = 783	Min = 2 Avg = 146 SD = 155 Max = 768	Min = 10 Avg = 146 SD = 156 Max = 773	Min = 10 Avg = 143 SD = 151 Max = 728	Min = 13 Avg = 140 SD = 143 Max = 834
Address density	A numerical variable describing the number of addresses found within the cell (Kadaster, 2022a). The range of these values is 0—–8. The SOIL mean was 0.44 and SD was 1.34. For	Sum	Min = 0 Avg = 2 SD = 3 Max = 19	Min = 0 Avg = 3 SD = 5 Max = 24	Min = 0 Avg = 6 SD = 9 Max = 79	Min = 0 Avg = 9 SD = 12 Max = 55	Min = 0 Avg = 36 SD = 42 Max = 222	Min = 0 Avg = 142 SD = 136 Max = 558	Min = 0 Avg= SD= Max=

(*continued on next page*)

Table 1 (*continued*)

Variable at 10*10 m	Description	Aggregation method	20*20 m	30*30 m	40*40 m	50*50 m	100*100 m	200*200 m	500*500 m
	the BULK, the mean was 0.41 and the SD was 1.28.								
Noise	A numerical variable describing the modeled amount of noise (Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2020). The range of these values was 39—–71, with a mean of 53.7 and a standard deviation of 7.0.	Bilinear	Min = 40 Avg = 54 SD = 7 Max = 73	Min = 39 Avg = 54 SD = 7 Max = 72	Min = 39 Avg = 54 SD = 7 Max = 74	Min = 39 Avg = 54 SD = 7 Max = 72	Min = 40 Avg = 54 SD = 7 Max = 71	Min = 40 Avg = 54 SD = 6 Max = 71	Min = 39 Avg = 55 SD = 5 Max = 73

performance against random models while varying potentially impacting settings. For every SDM, the validation AUC was compared to 100 randomly generated models with identical settings (i.e., a random model) to assess if the SDM with empirical data performed significantly better than the random models (Raes and ter Steege, 2007). The AUCs of all models that performed better than random were used in a linear mixed model to assess if spatial resolution, dispersal group (flying, facultative flying, ground), spatial aggregation or the set of drivers (CUGIC or ECO) were important drivers of AUC.

For dispersal, we added an interaction with model type and spatial resolution as the drivers or spatial scale may have different effects on species with low dispersal capacity compared to species with high dispersal capacity (Wunderlich et al., 2022). Additionally, this model accounts for the effect of the calibration AUC and controls for species average differences through the random effects. The resulting model was analysed with an Tukey post-hoc test to assess difference between dispersal groups and model types. This leads to the following model:

$$vAUC = driverset + dispersal + resolution + dispersal \times driverset + dispersal \times resolution + \zeta species + cAUC + \epsilon.$$

where *vAUC* describes the mean validation AUC from the SDMs for a particular species, and is predicted by the rest of the terms. The $\zeta species$ term describes the random effects for every species, cancelling out average differences in AUC between species. The term *cAUC* accounts for the AUC retrieved during the calibration phase.

2.7. Variable importance of SDM performance

To assess the variable importance, we extracted the delta AUC from every SDM. The delta AUC indicates how much the model is affected by the exclusion of a variable (Naimi and Araújo, 2016). Next, we used a Tukey post-hoc test to test for differences in variable importance within the CUGIC and ECO driver sets, and within dispersal groups. Finally, the most important variable was calculated by the frequency of a variable selected to be the most impactful across driver sets and dispersal groups.

3. Results

3.1. Comparing sampling methods

In our samples, we detected 1108 OTUs in the soil eDNA and 330 OTUs in the invertebrate bulk in at > 15 sites, making them suitable for the SDMs. Among those, we assigned 81 soil OTUs and 200 bulk OTUs to an invertebrate species using a 70 % similarity threshold (Table 2). Most annotated bulk OTUs, were matched to the taxonomic level of species (n = 135) or order (n = 51).

For the soil eDNA annotated OTUs, most matched to the taxonomic level of order (n = 42), family (n = 16), or species (n = 14). *Diptera*, *Isopoda* and *Coleoptera* were the most dominant taxa in the bulk invertebrate method, while *Collembola*, *Coleoptera* and *Hymenoptera* appeared proportionally more frequently in soil eDNA. Nonetheless, invertebrate bulk retrieved more OTUs for every taxa, except for *Hymenoptera*, *Annelida* and *Symphyla*. Running the analysis on a combined dataset resulted in 257 OTUs, indicating that only 24 OTUs overlapped between both datasets. Especially the taxa of *Coleoptera* (+15, 63 %), *Collembola* (+12, 55 %) and *Hymenoptera* (+8, 50 %) showed a substantial increase in coverage by combining the methods.

Table 2
The 10 most frequently found and identified OTUs of soil eDNA, invertebrate bulk metabarcoding and the combined dataset. Taxa described are OTUs identified at > 70 % match using BLAST with the BOLD database. Remaining OTUs are summed into the "Other" group.

Taxa	Soil eDNA (n = 81)	Invertebrate bulk (n = 200)	Combined (n = 257)
Diptera	9	71	74
Coleoptera	17	24	39
Collembola	19	21	33
Isopoda	1	28	27
Hymenoptera	16	12	24
Arachnida	5	12	17
Hemiptera	0	15	15
Annelida	5	1	6
Diplopoda	1	2	2
Symphyla	2	0	2
Other	6	14	18

3.2. SDM predictive performance and maps

Calibrating SDMs (n = 4496) resulted in a high proportion of AUC scores, indicating good model performance across driver sets (n = 4051, 90.4 %; AUC > 0.70). However, during model validation the model performance dropped very strongly (with only n = 57, 1.3 %, having AUC > 0.70). Even so, most models performed better than random models (n = 2664, 59.5 %, p < 0.05). The SDM performance seems independent of the spatial resolution of the predictors, as there was neither visual clustering of models with the identical spatial resolution nor a statistical significant effect (Table S2). These results indicate that for both sets of drivers, the predictors likely overfit the species niches, resulting in a poor predictive model (AUC < 0.70). This particular problem is more apparent in the ECO model, where the additional predictors seemed to increase the calibration AUCs values higher, while the validation AUCs values remained largely unaffected (Fig. 2).

3.3. Global drivers of SDM performance

The linear mixed model assessing the validation AUCs of significant SDMs (n = 2664, 59.3 %), shows that dispersal strategy and driver set (CUGIC-model) drive the variation in validation AUC (Table 3). In contrast, the relationship between spatial resolution and AUC fluctuated between a positive and negative effect, and was rarely statistically significant.

Comparing the driver sets shows that the ECO-type has statistically significantly lower AUC values compared to the CUGIC-type. This

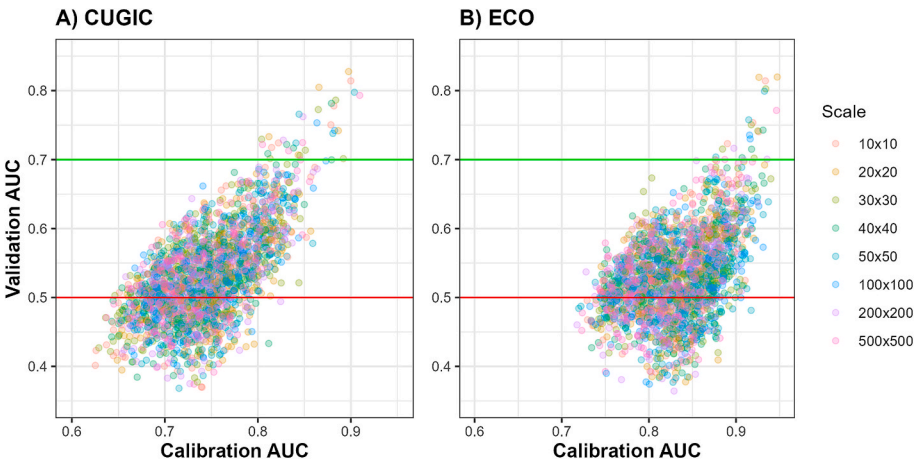


Fig. 2. Scatterplot of the AUC values during calibration and validation phase of the species distribution modelling, across eight different spatial resolutions. Panel A) shows the resulting AUCs from the SDMs with the CUGIC predictor set and panel B) shows the results with the ECO predictor set. The red line shows the performance of 0.50 AUC, which indicates that the SDM has no classification power. The green line shows the performance of 0.70 AUC, where the SDM has a fair or better classification performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Linear mixed model and the Tukey post-hoc results of AUC by dispersal strategy and driver set. The results for spatial resolution have been omitted due to its inconsistent effect (for full table see S2). Significant terms are highlighted with a bold font ($p < 0.05$).

Predictor	*Value*	*std.error*	*t-value*	*p-value*
ECO-model v. CUGIC-model	−0.05	0.004	−17.56	**<0.001**
Calibration AUC	0.62	0.025	24.35	**<0.001**
Posthoc − CUGIC				
Flying v. facultative flying	−0.007	0.005	−1.346	0.37
Flying v. ground	−0.013	0.004	−3.024	**0.007**
Ground v. facultative flying	−0.006	0.005	−1.314	0.39
Posthoc − ECO:				
Flying v. facultative flying	−0.0003	0.005	−0.078	0.99
Flying v. ground	−0.007	0.004	−1.705	0.20
Ground v. facultative flying	−0.007	0.005	−1.436	0.32

implies that the additional variables do not include extra relevant information for the SDMs (value = −0.05, $p < 0.001$; Fig. 3). However, this effect interacted with the dispersal strategies. The post-hoc analysis shows that for the CUGIC models, flying invertebrates had a consistently lower AUC than ground-dwelling invertebrates (estimate = −0.013, p = 0.007). In contrast, for the ECO-model, none of the dispersal groups performed significantly better than the other. These results suggest that the lower statistical power of the ECO models caused a loss in statistical power to distinguish effects of differences in dispersal capacity.

3.4. Variable importance of the SDMs

The ANOVA and post-hoc analysis on variable importance of significant SDMs (n = 2664, 59.3 %), shows that vegetation density was the most important variable for predicting species presence (Fig. 4). In the CUGIC models, vegetation density was the most important variable for ground and facultative flying invertebrates, while flying invertebrates SDMs were equally dependent on LULC. A similar pattern was visible in the ECO models, where vegetation density is the most important variable for ground and facultative flying invertebrates. However, the distance to water was more important to flying invertebrates, followed by vegetation density.

In contrast, the least important variables for predicting species presence were both noise and address density. The noise variable scored consistently as the least important variable for all the invertebrate SDMs. The address density scored lowest for the ground-dwelling and facultative flying invertebrates, but performeds slightly better in the case of flying invertebrates. These results on most and least important variables indicate e that ecological indicators are likely important for predicting invertebrate distribution, although different in strength per taxa, while indicators on anthropogenic pressures are unlikely to be important for assessing invertebrate distribution.

3.5. Heatmaps of species distributions

The validation threshold of AUC-score > 0.70 resulted in the inclusion of 14 different species containing 57 SDMs explaining their distribution well. The best performing SDMs per species were most frequently aggregated to 20*20 m resolution (n = 5) and used the CUGIC predictor set (n = 11). The SDMs that explained species distribution well mostly came from the invertebrate bulk metabarcoding data (n = 9). These results indicate that indicators on urban green infrastructure at a local resolution were the best at modelling the observed urban species.

The resulting distribution heatmaps show niches scattered across The Hague (Fig. 5). For example, *Paristoma notabilis* and *Philoscia moscrum*strongly gravitated towards forested regions. Their most likely habitat shows strong overlap in their distribution, indicating that they are similarly affected by changes to the environment. In contrast, the *Sarcoptiformes* species do not overlap with the other two species, and their distribution is more present in either the urban center or the city outskirts. The *Trixoscelis obscurella* species visually also has similar hotspots closer to the city center compared to the ground-dwelling invertebrates.

Interestingly, the maps for the *Hymenoptera, Entobryomorpha* and *Mesaphorura* species all show coarser grids, indicating their best model includes data from larger spatial scales, even though their dispersal styles are different. These findings reflect that urban species occupy different habitats across the city but are also able to sharing identical hotspots. These results indicate that species-specific SDMs may provide better overview of several urban invertebrate species their distribution compared to broad biodiversity patterns.

4. Discussion

The presented results provide a first glimpse at the potential for highly detailed coverage of invertebrate species and their distributions in urban contexts. The integration of DNA-based data and SDMs allows us to identify important drivers for urban biodiversity. Our results show that green and blue infrastructures are the most important predictors for

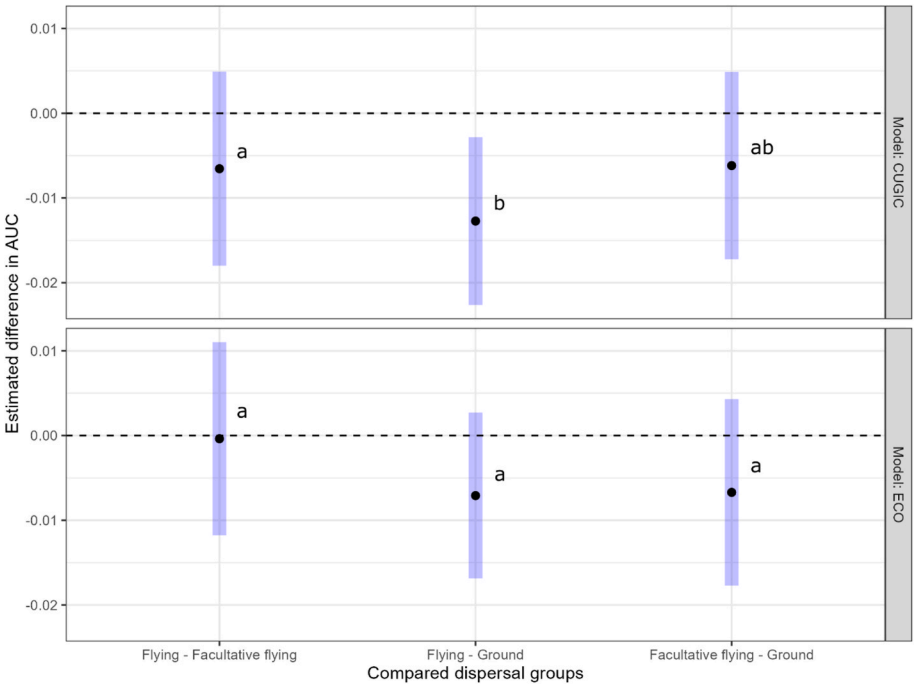


Fig. 3. Estimated differences in SDM performance (AUC) across predictor sets (CUGIC/ECO), comparing groups with different dispersal capacities. The horizontal dotted line in both panels indicates no difference between the compared groups. The black dot shows the estimated difference between the groups, and the blue box shows the 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the distribution of urban invertebrate species. In contrast, indicators on urbanisation such as address density and noise are consistently the least important. These results indicate that within the urban environment invertebrate diversity seems to be driven largely by the presence of habitat (green and blue infrastructure), instead of the quality of the habitat(s) in relation to the prevalence of anthropogenic stressors (noise and address density). Our findings show that combining DNA-based data and SDMs facilitate both assessing the importance of spatial features and their role in assessing community assembly of urban biodiversity.

4.1. *Ecological implications*

Utilizing DNA-based techniques to infer species presence presents a novel avenue for rapidly assessing biodiversity with extensive coverage of species. Our results show that incorporating both eDNA and bulk metabarcoding enhances the coverage, showcasing their complementary nature and the high potential for uncovering vast quantities of species previously undetectable or at reduce labour (Kirse et al., 2021). These techniques become even more powerful in the near future, with reference databases increasing in coverage and techniques to retrieve sequenced DNA becoming less biased (Cristescu and Herbert, 2018).

The SDMs based on the eDNA data showed that urban species distributions dependent on green and blue infrastructure, while anthropogenic pressures showed minimal impact on species distributions. Several studies have shown similar effects, suggesting that available habitat impacts species distribution more than anthropogenic influences. For example, in Singapore, the presence of urban vegetation and parks were the main drivers of butterfly species presence (Jain et al., 2021). For weaver birds' movements in Cape Town, increasing urbanization appears inconsequential while the presence of ecologically crucial wetlands plays a decisive role in the bird's distribution (Calder et al., 2015). This is in line with *meta*-analyses on urban biodiversity showing that habitat – quantified as vegetation density, abundance and structure – is the most important driver of urban biodiversity (McKinney, 2008; Beninde et al., 2015). These results on green and blue infrastructure all suggest that sufficient terrestrial and aquatic habitat is the main limiting factor to enhancing urban biodiversity, and that creating habitat through urban green and blue spaces is the key to increasing urban biodiversity.

Contrary to our expectations, we did not find that anthropogenic pressures such as address density and noise are important for the distribution of urban species. Anthropogenic pressures have been shown to be useful for estimating species distributions when including rural territories, and with studies even suggesting that anthropogenic pressures are a pivotal driver of biodiversity (McKinney, 2008; Jain et al., 2021; Beninde et al., 2023). For example, a study examining 1200 terrestrial species over the complete rural to urban gradient in the greater Los Angeles, showed pronounced association between species distributions and urbanization (Beninde et al., 2023). Rural to urban species richness changes have been shown to occur mostly due to turnover from locally unique species, to globe-trotting species (Mckinney, 2008). This suggests that the species found in cities are (pre-)adapted to anthropogenic pressures and can therefore survive well in stressful environments (i.e. urban exploiter; Kark et al., 2007). This theory would explain the lack of a meaningful contribution of anthropogenic pressures to our models, as (pre-)adapted species would not be meaningfully affected by anthropogenic pressures. Indicators on anthropogenic pressures would be most informative in models that cover areas with little to no human impacts, as we expect that species most sensitive to anthropogenic pressures are present there (e.g. Glisson et al., 2017).

Using the commonly used indicators for urban green infrastructure research, the majority of SDMs performed better than random, yet surprisingly few reached the traditionally held standards of AUCs for assessing SDM performance. The predictors that we expected to be important for urban biodiversity, based on previous research (Beninde et al., 2015; Morpurgo et al., 2023)—such as address density, noise, and distance to water— caused overfitting of the models rather than enhancing their performance.

There are two possible explanations for why our models failed to meet the performance threshold (AUC $>$ 0.70). Our first explanation

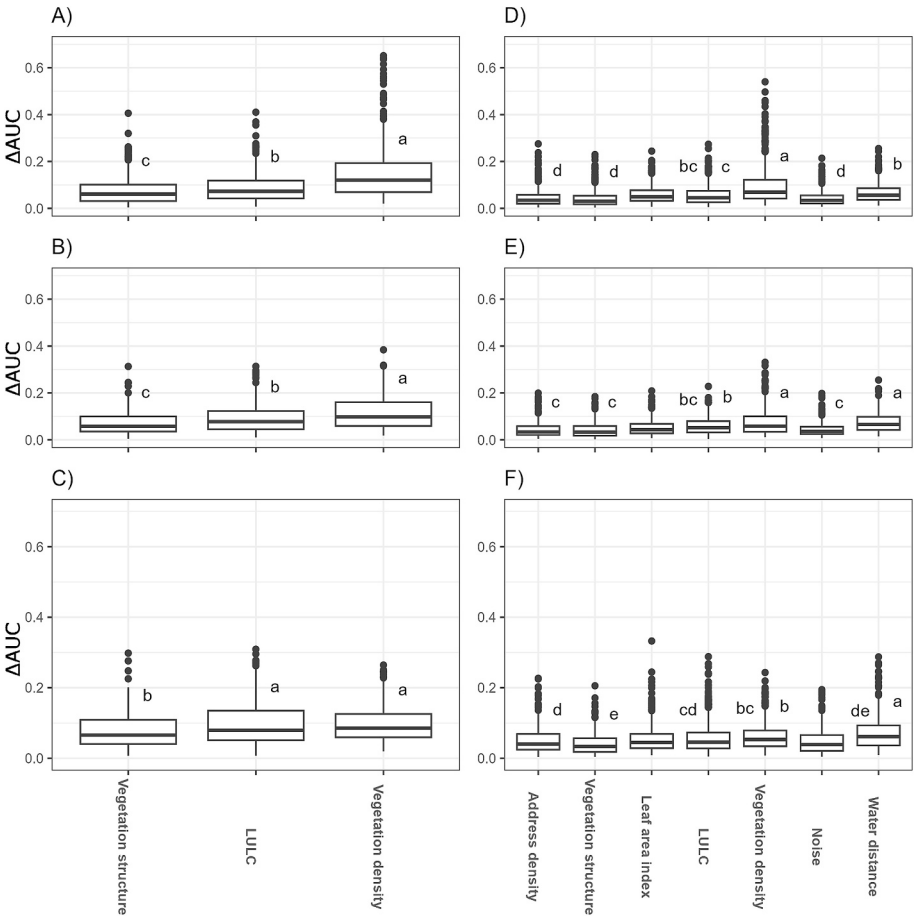


Fig. 4. Variable importance scores by model and dispersal group. The left panels, A) — C), show the results for the CUGIC SDMs and right panels, D) — F), the ECO SDMs. The first row, A) and D), show the results for the ground invertebrates. The second row, B) and E), show the results for the facultative flying invertebrates. The third row, C) and F), show the results for the flying invertebrates. The AUC variable importance displays the effect of exclusion of the variable in the model on the performance of the SDM. Letters next to the boxes are corresponding to a post-hoc analysis, grouping variables that are not significantly different from each other ($p > 0.05$). The alphabetical order indicates which group is the most important for the SDM. LULC = land-use/landcover.

revolves around a potential mismatch in the selection of predictors, which are meant to be informative for explaining biodiversity metrics but may lack species-specific niche information. To model species niches accurately, it is important to consider the life-history of the organism, a coinciding relevant spatial resolution(Aguirre-Gutiérrez et al., 2013; Wunderlich et al., 2022), and environmental drivers given the organism (Araújo and Guisan, 2006). In our study, there was no difference in performance across eight different spatial resolutions, supporting the argument that species-specific drivers of niches might have been missing. Here, we suggest some predictors that may improve urban SDMs performances. Recently developed micro-climatic data could be particularly vital for urban SDMs, with natural systems already showing temperature variations of up to 10 °C (Suggitt et al., 2011). Several studies have already shown that micro-climate data at a resolution of 25 m^2 varyingly enhances model performance for many species (Lembrechts et al., 2019). We hypothesize that local climate data is likely more important in cities, as urban micro-climates are more extreme due to the urban heat island effect, flooding and droughts. Also, indicators specific to pollinators such as floral availability or nesting substrate, and indicators specific to ground-dwelling invertebrates such as soil quality, pollution levels, and organic matter content, have been suggested to enhance predictions of their distributions (Santorufo et al., 2014; Silva et al., 2014). Our results, alongside existing literature, are underscoring the necessity of exploring other predictors that may be more informative for the distribution of urban species.

A more radical second explanation for the low performance of most SDMs could be that our species are (pre-)adapted to the urban environment, are free to disperse between patches and act as one *meta*-community. Urban environments are known to be biotically homogenised and populated mostly by generalist species that easily adapt to novel food sources and habitats, while also having high dispersal capacity (Kark et al., 2007). Given that we accounted for most of the important requirements for urban invertebrate species, and given that the observed species share important ecological traits such as, resources, habitat, mutualists, neutral theory may provide an elegant solution to the poorly performing SDMs (Leibold and McPeek, 2006). The congregation of species with similar traits in the urban environment makes a strong case for it being an ecosystem with neutral dynamics. Given neutral theory, the assembly and distribution of urban invertebrate species would be primarily driven by birth, death, and immigration instead of the environment constraints (Bell, 2000). Neutral theory would explain our results as it would suggest that habitat availability, i.e. the amount of green space, would acts as a limiting factor rather than quality characteristics (such as noise, address density) of urban green infrastructure.

4.2. *Implications for urban planning*

Green and blue infrastructure were identified as the most important drivers for urban invertebrate species distributions, indicating that the protection and enhancement of these infrastructures is imperative. This is in line with other work from biodiversity studies (Beninde et al., 2015;

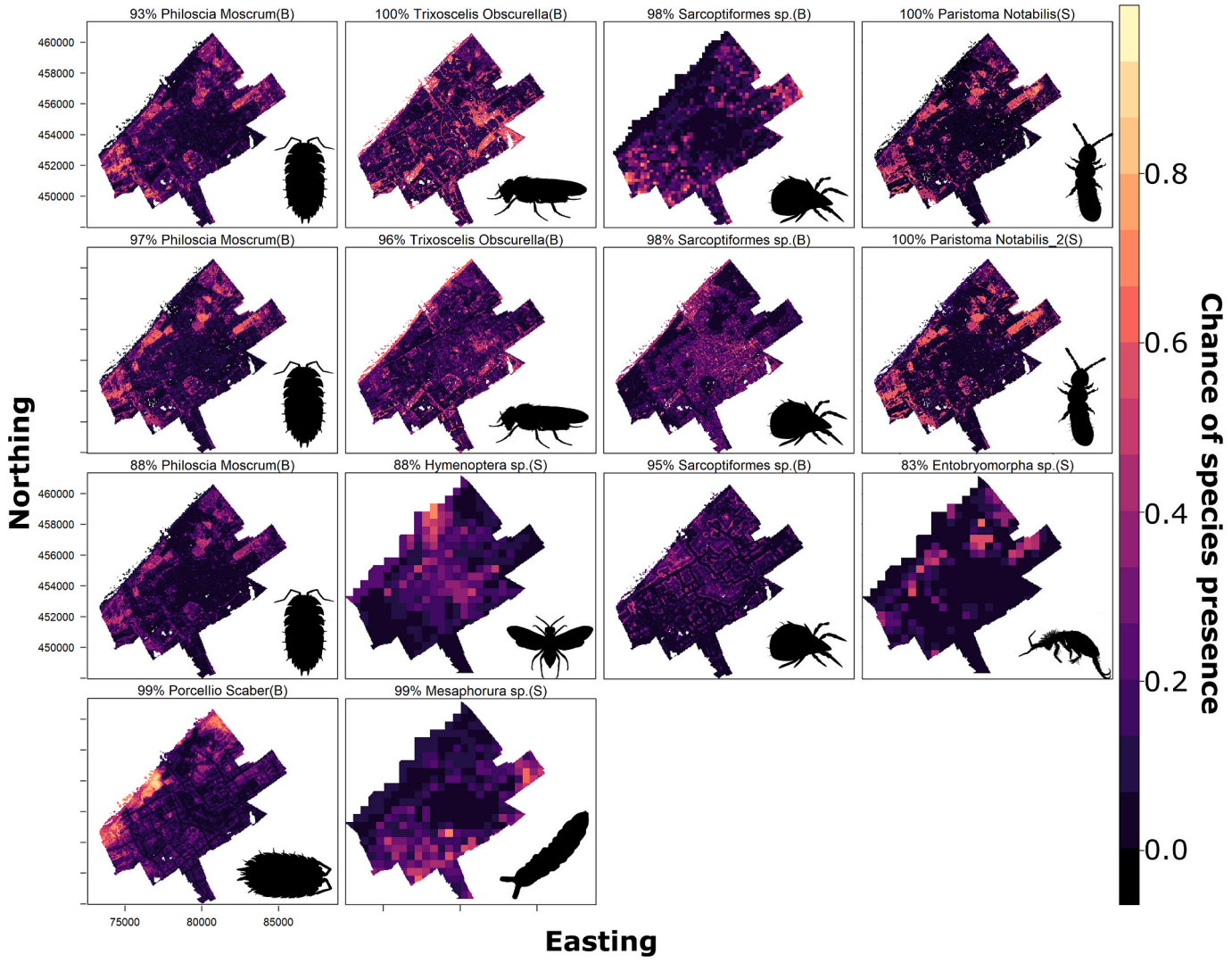


Fig. 5. Predicted distributions of the 14 species where the SDMs had an AUC > 0.70. The species names indicate the best match between the modelled OTU and the BOLD reference data, where the percentages indicate the overlap between the OTU and the reference data. An (S) indicates the SDM is based on data from the soil eDNA dataset and a (B) indicates the SDM is based on the data from the invertebrate bulk. These heatmaps reflect the best performing SDM, by AUC value, and therefore have different spatial resolutions.

McKinney, 2008). Urban green and blue infrastructure also provide other benefits, including for climate adaptation and human health, where increasing quantities of urban nature provide more benefits (Bratman et al., 2019; Remme et al., 2021; Morpurgo et al., 2023). Moreover, anthropogenic pressures do not meaningfully affect urban invertebrates distributions. This indicates that the incorporation of urban green infrastructures through green roofs, street greenery, and rain gardens likely have a strong positive effect on urban biodiversity. Our results therefore strengthens the plea for the prioritisation of planning and preservation of green spaces, parks, gardens, and other potential habitats within cities to support and enhance urban invertebrate biodiversity alongside human-centred benefits.

5. Conclusion

Enhancing invertebrate biodiversity is important for safeguarding biodiversity targets and sustaining ecosystem functions essential to humans. For urban invertebrate biodiversity, our analysis shows that environmental predictors on habitat availability, food and water are the most important drivers for urban invertebrate species distributions. Within the urban environment, species distributions are less impacted by anthropogenic pressures than by habitat availability. We argue that DNA-based SDMs allows for the assessment of urban invertebrate species, their distribution and the drivers of urban community assembly. Further research should prioritize investigating understanding the mechanisms of urban invertebrate distribution as much of it is left unexplained. We urge policymakers and urban planners to invest in more green and blue infrastructure as they were the most important for invertebrate species' presence.

CRediT authorship contribution statement

Joeri Morpurgo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Roy P. Remme:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Quinten Mudde:** Writing – review & editing, Methodology, Investigation. **Emilie A. Didaskalou:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Data curation. **Kornelia J. Serwatowska:** Writing – review & editing, Validation, Supervision, Software, Methodology, Investigation. **Krijn B.**

Trimbos: Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Mingming Hu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Peter M. Van Bodegom:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work reports financial support was provided by Dutch Research Council through 'Merian Fund Cooperation China-The Netherlands (CAS) Green Cities 2019'. No. 482.19.704. This work reports financial support was provided by CML Impact fund of Leiden University.

Acknowledgements

We would like to thank the following people for their contribution during fieldwork, lab work, or in consultation: Jenny Lüdtke, Anne Wilts, Klaas Land, Sofie Rasmussen, Rianne van Duinen, Iris de Wolf, Laura Zantis, Stephanie Cap, Mike Slootweg, Ilmer Benda, Mike Lakerveld, Puck Bramer, Meilin Remijn, Sandhya S Jhingoer, Niels Raes, Yali Si, Kat Stewart, Jan Macher, and Kevin Beentjes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2025.113650.

Data availability

The data that support the findings of this study are openly available on Dryad at: https://doi.org/10.5061/dryad.0vt4b8h8j

References

Andrade, R., Franklin, J., Larson, K.L., Swan, C.M., Lerman, S.B., Bateman, H.L., Warren, P.S., York, A., 2021. Predicting the assembly of novel communities in urban ecosystems. Landsc. Ecol. 36 (1), 1–15. https://doi.org/10.1007/s10980-020-01142-101142-1.

AHN, 2024. Actueel Hoogte Bestand Viewer. Actueel Hoogte Bestand 4. https://www.ahn.nl/ahn-viewer.

Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M., Biesmeijer, J.C., 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria – dutch hoverflies as a case study. PLoS One 8 (5), e63708. https://doi.org/10.1371/journal.pone.0063708.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amaya-Espinel, J.D., Hostetler, M., Henríquez, C., Bonacic, C., 2019. The influence of building density on Neotropical bird communities found in small urban parks. Landsc. Urban Plan. 190, 103578. https://doi.org/10.1016/j.landurbplan.2019.05.009.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33 (10), 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Aronson, M.F.J., Nilon, C.H., Lepczyk, C.A., Parker, T.S., Warren, P.S., Cilliers, S.S., Goddard, M.A., Hahs, A.K., Herzog, C., Katti, M., la Sorte, F.A., Williams, N.S.G., Zipperer, W., 2016. Hierarchical filters determine community assembly of urban species pools. Ecology 97 (11), 2952–2963. https://doi.org/10.1002/ecy.1535.

Auger, G., Pottier, J., Mathieu, J., Jabot, F., 2024. Space use of invertebrates in terrestrial habitats: Phylogenetic, functional and environmental drivers of interspecific variations. Glob. Ecol. Biogeogr. 33 (4). https://doi.org/10.1111/geb.13811.

Bell, G., 2000. The distribution of abundance in neutral communities. Am. Nat. 155 (5), 606–617. https://doi.org/10.1086/303345.

Beentjes, K.K., Speksnijder, A.G.C.L., Schilthuizen, M., Hoogeveen, M., Pastoor, R., van der Hoorn, B.B., 2019. Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. PLoS One 14 (12), e0226527. https://doi.org/10.1371/journal.pone.0226527.

Belovsky, G.E., Slade, J.B., 2000. Insect herbivory accelerates nutrient cycling and increases plant production. Proc. Natl. Acad. Sci. USA 97. https://doi.org/10.1073/pnas.250483797.

Beninde, J., Veith, M., Hochkirch, A., 2015. Biodiversity in cities needs space: A meta-analysis of factors determining intra-urban biodiversity variation. In Ecol. Lett. 18 (6), 581–592. https://doi.org/10.1111/ele.12427.

Beninde, J., Delaney, T.W., Gonzalez, G., Shaffer, H.B., 2023. Harnessing iNaturalist to quantify hotspots of urban biodiversity: the Los Angeles case study. Front. Ecol. Evol. 11. https://doi.org/10.3389/fevo.2023.983371.

Boakes, Z., Stafford, R., Bramer, I., Cvitanović, M., Hardouin, E.A., 2023. The importance of urban areas in supporting vulnerable and endangered mammals. Urban Ecosystems. https://doi.org/10.1007/s11252-023-01492-z.

BOLDSYSTEMS, 2023, October 20. BOLDSYSTEMS Data Packages. Data Packages Downloadable DNA Barcoding Datasets. https://www.boldsystems.org/index.php/datapackages.

Brashear, W.A., Ammerman, L.K., Dowler, R.C., 2015. Short-distance dispersal and lack of genetic structure in an urban striped skunk population. J. Mammal. 96 (1), 72–80. https://doi.org/10.1093/jmammal/gyu004.

Bratman, G.N., Anderson, C.B., Berman, M.G., Cochran, B., de Vries, S., Flanders, J., Folke, C., Frumkin, H., Gross, J.J., Hartig, T., Kahn, P.H., Kuo, M., Lawler, J.J., Levin, P.S., Lindahl, T., Meyer-Lindenberg, A., Mitchell, R., Ouyang, Z., Roe, J., Daily, G.C., 2019. Nature and mental health: An ecosystem service perspective. Sci. Adv. 5 (7). https://doi.org/10.1126/sciadv.aax0903.

Brown, G.R., Matthews, I.M., 2016. A review of extensive variation in the design of pitfall traps and a proposal for a standard pitfall trap design for monitoring ground-active arthropod biodiversity. Ecol. Evol. 6 (12), 3953–3964. https://doi.org/10.1002/ece3.2176.

Buchholz, S., Egerer, M.H., 2020. Functional ecology of wild bees in cities: towards a better understanding of trait-urbanization relationships. Biodivers. Conserv. 29 (9–10), 2779–2801. https://doi.org/10.1007/s10531-020-02003-8.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13 (7), 581–583. https://doi.org/10.1038/nmeth.3869.

Calder, J.-L., Cumming, G.S., Maciejewski, K., Oschadleus, H.D., 2015. Urban land use does not limit weaver bird movements between wetlands in Cape Town, South Africa. Biol. Conserv. 187, 230–239. https://doi.org/10.1016/j.biocon.2015.04.021.

Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens. Environ. 62 (3), 241–252. https://doi.org/10.1016/S0034-4257(97)00104-1.

Collins, C.M., (Tilly), Audusseau, H., Hassall, C., Keyghobadi, N., Sinu, P.A., Saunders, M. E., 2024. Insect ecology and conservation in urban areas: An overview of knowledge and needs. Insect Conserv. Diversity 17 (2), 169–181. https://doi.org/10.1111/icad.12733.

Cristescu, M.E., Hebert, P.D.N., 2018. Uses and misuses of environmental DNA in biodiversity science and conservation. Annu. Rev. Ecol. Evol. Syst. 49 (1), 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306.

Crocker, W., Maute, K., Webb, C., French, K., 2017. Mosquito assemblages associated with urban water bodies; implications for pest and public health threats. Landsc. Urban Plan. 162, 115–125. https://doi.org/10.1016/j.landurbplan.2017.02.006.

Den Dulk, C. J., van de Stadt, H., Vliegen, J. M., 1992. A new measure for degree of urbanization: the address density of the surrounding area. Maandstatistiek van de Bevolking (Hague, Netherlands : 1982), 40(7), 14–27.

Drager, K.I., Hirmas, D.R., Hasiotis, S.T., Bents, T.C., 2016. Effects of ant (*Formica subsericea*) nest development on physical and hydrological properties in a coarse-textured soil. Soil Sci. 181 (3–4), 166–174. https://doi.org/10.1097/SS.0000000000000145.

Dutta, T., Sharma, S., Meyer, N.F.v., Larroque, J., Balkenhol, N., 2022. An overview of computational tools for preparing, constructing and using resistance surfaces in connectivity research. Landsc. Ecol. 37 (9), 2195–2224. https://doi.org/10.1007/s10980-022-01469-x.

Eggleton, P., 2020. Annual Review of Environment and Resources The State of the World's Insects. https://doi.org/10.1146/annurev-environ-012420.

Elbrecht, V., Vamos, E.E., Steinke, D., Leese, F., 2018. Estimating intraspecific genetic diversity from community DNA metabarcoding data. PeerJ 6, e4644.

Elbrecht, V., Braukmann, T.W.A., Ivanova, N.v., Prosser, S.W.J., Hajibabaei, M., Wright, M., Zakharov, E.v., Hebert, P.D.N., Steinke, D., 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7, e7745.

European Commission. (2014). Building a Green for Europe Environment. In European Union. https://doi.org/10.2779/54125.

European Space Agency (ESA), 2023. SNAP.

Everingham, S., 2023. Detailed protocols: Pan-trapping add on (European sites). https://www.bug-net.org/detailed-protocols-pan-trapping-add-on-european-sites/.

Fischer, J.D., Schneider, S.C., Ahlers, A.A., Miller, J.R., 2015. Categorizing wildlife responses to urbanization and conservation implications of terminology. Conserv. Biol. 29 (4), 1246–1248. https://doi.org/10.1111/cobi.12451.

Gleason, J.E., Elbrecht, V., Braukmann, T.W.A., Hanner, R.H., Cottenie, K., 2021. Assessment of stream macroinvertebrate communities with eDNA is not congruent with tissue-based metabarcoding. Mol. Ecol. 30 (13), 3239–3251. https://doi.org/10.1111/mec.15597.

Glisson, W.J., Conway, C.J., Nadeau, C.P., Borgmann, K.L., 2017. Habitat models to predict wetland bird occupancy influenced by scale, anthropogenic disturbance, and imperfect detection. Ecosphere 8 (6). https://doi.org/10.1002/ecs2.1837.

Gorunescu, F., 2011. Classification Performance Evaluation (pp. 319–330). https://doi.org/10.1007/978-3-642-19721-5_6.

Güneralp, B., Seto, K.C., 2013. Futures of global urban expansion: Uncertainties and implications for biodiversity conservation. Environ. Res. Lett. 8 (1). https://doi.org/10.1088/1748-9326/8/1/014025.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D., de Kroon, H., 2017.

More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One 12 (10), e0185809. https://doi.org/10.1371/journal.pone.0185809.
IUCN, NATURA Global to Local Thematic Working Group, & Urban Biodiversity Hub. (2022, April 14). The Urban Butterfly Effect. The Urban Butterfly Effect. http://www.ubhub.org/pdfs/tube2022.pdf.
Jain, A., Zeng, Y., Webb, E.L., 2021. Critical dependence of butterflies on a non-native host plant in the urban tropics. Front. Ecol. Evol. 9. https://doi.org/10.3389/fevo.2021.655012.
Kadaster, 2022a. BAG. In Basisregistratie Adressen en Gebouwen v2.
Kadaster, 2022b. Waterdeel. Basisregistratie Grootschalige Topografie. https://denhaag.dataplatform.nl/#/data/6e9f56f7-5fb0-4c43-a2a4-819bac5ea143.
Kark, S., Iwaniuk, A., Schalimtzek, A., Banker, E., 2007. Living in the city: can anyone become an 'urban exploiter'? J. Biogeogr. 34 (4), 638–651. https://doi.org/10.1111/j.1365-2699.2006.01638.x.
Kattwinkel, M., Strauss, B., Biedermann, R., Kleyer, M., 2009. Modelling multi-species response to landscape dynamics: mosaic cycles support urban biodiversity. Landsc. Ecol. 24 (7), 929–941. https://doi.org/10.1007/s10980-009-9371-7.
Kirk, H., Garrard, G.E., Croeser, T., Backstrom, A., Berthon, K., Furlong, C., Hurley, J., Thomas, F., Webb, A., Bekessy, S.A., 2021. Building biodiversity into the urban fabric: A case study in applying Biodiversity Sensitive Urban Design (BSUD). Urban For. Urban Green. 62, 127176. https://doi.org/10.1016/j.ufug.2021.127176.
Kirse, A., Bourlat, S.J., Langen, K., Fonseca, V.G., 2021. Unearthing the potential of soil eDNA metabarcoding—towards best practice advice for invertebrate biodiversity assessment. Front. Ecol. Evol. 9. https://doi.org/10.3389/fevo.2021.630560.
Kuras, E.R., Warren, P.S., Zinda, J.A., Aronson, M.F.J., Cilliers, S., Goddard, M.A., Nilon, C.H., Winkler, R., 2020. Urban socioeconomic inequality and biodiversity often converge, but not always: A global meta-analysis. Landsc. Urban Plan. 198, 103799. https://doi.org/10.1016/j.landurbplan.2020.103799.
LaPoint, S., Balkenhol, N., Hale, J., Sadler, J., van der Ree, R., 2015. Ecological connectivity research in urban areas. Funct. Ecol. 29 (7), 868–878. https://doi.org/10.1111/1365-2435.12489.
Leibold, Mathew. A., & McPeek, Mark. A. (2006). Coexistence of the niche and neutral perspectives in community ecology. Ecology, 87(6), 1399–1410.
Lembrechts, J.J., Nijs, I., Lenoir, J., 2019. Incorporating microclimate into species distribution models. Ecography 42 (7), 1267–1279. https://doi.org/10.1111/ecog.03947.
Leong, M., Dunn, R.R., Trautwein, M.D., 2018. Biodiversity and socioeconomics in the city: a review of the luxury effect. Biol. Lett. 14 (5), 20180082. https://doi.org/10.1098/rsbl.2018.0082.
Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front. Zool. 10 (1), 34. https://doi.org/10.1186/1742-9994-10-34.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17 (1), 10. https://doi.org/10.14806/ej.17.1.200.
Masson, V., Lemonsu, A., Hidalgo, J., Voogt, J., 2020. Urban climates and climate change. Annu. Rev. Env. Resour. 45 (1), 411–444. https://doi.org/10.1146/annurev-environ-012320-083623.
McCloy, M.W.D., Andringa, R.K., Maness, T.J., Smith, J.A., Grace, J.K., 2024. Promoting urban ecological resilience through the lens of avian biodiversity. Front. Ecol. Evol. 12. https://doi.org/10.3389/fevo.2024.1302002.
McKinney, M.L., 2008. Effects of urbanization on species richness: A review of plants and animals. Urban Ecosystems 11 (2), 161–176. https://doi.org/10.1007/s11252-007-0045-4.
Montgomery, G.A., Belitz, M.W., Guralnick, R.P., Tingley, M.W., 2021. Standards and best practices for monitoring and benchmarking insects. Front. Ecol. Evol. 8. https://doi.org/10.3389/fevo.2020.579193.
Morpurgo, J., Remme, R.P., van Bodegom, P.M., 2023. CUGIC: The consolidated urban green infrastructure classification for assessing ecosystem services and biodiversity. Landsc. Urban Plan. 234. https://doi.org/10.1016/j.landurbplan.2023.104726.
Muller, B.H., Mollon, P., Santiago-Allexant, E., Javerliat, F., Kaneko, G., 2019. In-depth comparison of library pooling strategies for multiplexing bacterial species in NGS. Diagn. Microbiol. Infect. Dis. 95 (1), 28–33. https://doi.org/10.1016/j.diagmicrobio.2019.04.014.
Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39 (4), 368–375. https://doi.org/10.1111/ecog.01881.
Orr, J.A., Rillig, M.C., Jackson, M.C., 2022. Similarity of anthropogenic stressors is multifaceted and scale dependent. Natural Sciences 2 (1). https://doi.org/10.1002/ntls.20210076.
Peters, H., O'Leary, B.C., Hawkins, J.P., Roberts, C.M., 2015. Identifying species at extinction risk using global models of anthropogenic impact. Glob. Chang. Biol. 21 (2), 618–628. https://doi.org/10.1111/gcb.12749.
Pfeifer, M., Disney, M., Quaife, T., Marchant, R., 2012. Terrestrial ecosystems from space: a review of earth observation products for macroecology applications. Glob. Ecol. Biogeogr. 21 (6), 603–624. https://doi.org/10.1111/j.1466-8238.2011.00712.x.
Pickett, S.T.A., Cadenasso, M.L., Grove, J.M., Boone, C.G., Groffman, P.M., Irwin, E., Kaushal, S.S., Marshall, V., McGrath, B.P., Nilon, C.H., Pouyat, R.V., Szlavecz, K., Troy, A., Warren, P., 2011. Urban ecological systems: Scientific foundations and a decade of progress. J. Environ. Manage. 92 (3), 331–362. https://doi.org/10.1016/j.jenvman.2010.08.022.
Pierce, J.R., Costadone, L., Mannetti, L., Morpurgo, J., Green, C.E., Halder, M.D., Guijosa, P.A.L., Bogan, A.L., Galt, R., Hughes, J., 2024. Urban Nature Indexes tool offers comprehensive and flexible approach to monitoring urban ecological performance. Npj Urban Sustain. 4 (1), 22. https://doi.org/10.1038/s42949-024-00143-2.
Raes, N., ter Steege, H., 2007. A null-model for significance testing of presence-only species distribution models. Ecography 30 (5), 727–736. https://doi.org/10.1111/j.2007.0906-7590.05041.x.
Remme, R. P., Frumkin, H., Guerry, A. D., King, A. C., Mandle, L., Sarabu, C., Bratman, G. N., Giles-Corti, B., Hamel, P., Han, B., Hicks, J. L., James, P., Lawler, J. J., Lindahl, T., Liu, H., Lu, Y., Oosterbroek, B., Paudel, B., Sallis, J. F., … Daily, G. C. (2021). An ecosystem service perspective on urban nature, physical activity, and health. Proceedings of the National Academy of Sciences, 118(22). https://doi.org/10.1073/pnas.2018472118.
Rijksinstituut voor Volksgezondheid en Milieu (RIVM). (2020). Atlas Leefomgeving. Geluid in Nederland. https://www.atlasleefomgeving.nl/kaarten?config=3ef897de-127f-471a-959b-93b7597de188&activateOnStart=layercollection&gm-x=155000.00000000003&gm-y=456478.71437812824&gm-z=2.894059846274022&gm-b=1544180834512,true,1;1544715737496,true,0.8.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Santorufo, L., Cortet, J., Arena, C., Goudon, R., Rakoto, A., Morel, J.-L., Maisto, G., 2014. An assessment of the influence of the urban environment on collembolan communities in soils using taxonomy- and trait-based approaches. Appl. Soil Ecol. 78, 48–56. https://doi.org/10.1016/j.apsoil.2014.02.008.
Schell, C.J., Dyson, K., Fuentes, T.L., des Roches, S., Harris, N.C., Miller, D.S., Woelfle-Erskine, C.A., Lambert, M.R., 2020. The ecological and evolutionary consequences of systemic racism in urban environments. Science 369 (6510). https://doi.org/10.1126/science.aay4497.
Silva, D.P., Gonzalez, V.H., Melo, G.A.R., Lucia, M., Alvarez, L.J., de Marco, P., 2014. Seeking the flowers for the bees: Integrating biotic interactions into niche models to assess the distribution of the exotic bee species Lithurgus huberi in South America. Ecol. Model. 273, 200–209. https://doi.org/10.1016/j.ecolmodel.2013.11.016.
Simkin, R.D., Seto, K.C., McDonald, R.I., Jetz, W., 2022. Biodiversity impacts and conservation implications of urban land expansion projected to 2050. In: Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.2117297119.
Suggitt, A.J., Gillingham, P.K., Hill, J.K., Huntley, B., Kunin, W.E., Roy, D.B., Thomas, C. D., 2011. Habitat microclimates drive fine-scale variation in extreme temperatures. Oikos 120 (1), 1–8. https://doi.org/10.1111/j.1600-0706.2010.18270.x.
Taberlet, P., Prud'homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J., Melodelima, C., Pompanon, F., Coissac, E., 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Mol. Ecol. 21 (8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x.
Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., James, P., 2007. Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. Landsc. Urban Plan. 81 (3), 167–178. https://doi.org/10.1016/j.landurbplan.2007.02.001.
van der Waals, J., 2002. The compact city and the environment: a review. Tijdchr. Econ. Soc. Geogr. 91 (2), 111–121. https://doi.org/10.1111/1467-9663.00099.
Varner, D.M., Hepp, G.R., Bielefeld, R.R., 2014. Movements and seasonal use of habitats by rural and urban female mottled ducks in southeast Florida. J. Wildl. Manag. 78 (5), 840–847. https://doi.org/10.1002/jwmg.734.
Vilmi, A., Alahuhta, J., Hjort, J., Kärnä, O.-M., Leinonen, K., Rocha, M.P., Tolonen, K.E., Tolonen, K.T., Heino, J., 2017. Geography of global change and species richness in the North. Environ. Rev. 25 (2), 184–192. https://doi.org/10.1139/er-2016-0085.
Vrdoljak, S.M., Samways, M.J., 2012. Optimising coloured pan traps to survey flower visiting insects. J. Insect Conserv. 16 (3), 345–354. https://doi.org/10.1007/s10841-011-9420-9.
Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R., Stopak, D., 2021. Insect decline in the Anthropocene: Death by a thousand cuts. Proc. Natl. Acad. Sci. USA 118 (2). https://doi.org/10.1073/PNAS.2023989118.
Wunderlich, R.F., Mukhtar, H., Lin, Y.-P., 2022. Comprehensively evaluating the performance of species distribution models across clades and resolutions: choosing the right tool for the job. Landsc. Ecol. 37 (8), 2045–2063. https://doi.org/10.1007/s10980-022-01465-1.
Zhao, Z., Borzée, A., Li, J., Chen, S., Shi, H., Zhang, Y., 2023. Urban bird community assembly mechanisms and driving factors in university campuses in Nanjing, China. Animals 13 (4), 673. https://doi.org/10.3390/ani13040673.