Dear Associate editor,

We thank you very much for your support and for the comments. Accordingly, we made the requested modifications (cf. the answer below included in your text). We also came back to the referee's answers, added up our responses by specifying the new line numbers of the revised parts in the revised manuscript (later below). Our answers are indicated by red marking.

In addition to all the comments, we improved some parts:

- **The figures were adjusted to color blindness** (as already mentioned in the previous answers to the referees): the colors were kept as before, but some variations in the pin shapes and strips in the box background.

cf. figures 2, 3, 4, 5, S1, S2, S3, S4, S5, S6, S7, S8 and S9 (-> all the figures, except Fig. 1)

- **The seed densities were corrected** (as already mentioned in the previous answers to the referees).

Lines 112-115: "sowing of 14 pollinator-friendly forbs at a density of 0.7 g m−2. […] four grass species at a density up to 0.7 g m−2. […] the hemiparasitic forb Rhinanthus minor […] was sown at a 6 g m−2 density."
changed to lines 134-139: "the sowing of 14 pollinator-friendly forbs at a density of 0.06 g m^{-2}. […] four grass species at a density up to 0.6 g m^{-2}. […] the hemiparasitic forb *Rhinanthus minor* […] was sown at a 5 g m^{-2} density."

- **The units in the flux equation were corrected.**

Lines 213-215: "µ" was missing and now changed to "µg GHG m^{-2} s^{-1}".

- **The soil sampling protocol was clarified the protocol** (as already mentioned in the previous answers to the referees).

Also in line with further comments made by the reviewers for clarification, some clarifications in the 2.3.2 section in the overall but the major changes were made from lines 242-247: "At KMP sites, samples for C and N were collected with an auger (d = 2.3 cm) down to a 30 cm depth. Altogether, six individual samples were collected systematically at the lawn and the meadow. At VKI, we used an auger (d = 1.7 cm) to collect eight samples and then pooled them together. The fresh samples were sieved (mesh size = 2 mm) and the fresh 245 weights of the smaller and larger soil fractions were weighed. The soil was dried at 105 ℃ for 24 h. Total soil C and N contents were determined from dried and milled samples of soil smaller than 2 mm with an elemental CN analyzer (LECO, Michigan, USA)."
changed to lines 288-293: "Other soil samples were collected to determine C and N content. At KMP sites, samples for C and N were collected with a soil auger down to 30 cm depth. Altogether, six individual samples were collected at the lawn and the meadow. At VKI, we used a soil auger to collect eight samples and then pooled them together. The irrigated lawn samples were first sieved with a 2-mm mesh sieve and then dried at 105 ℃ for 24 h. At the other intensive sites, the samples were first dried at 105 ℃ for 48 h, and then sieved with a 2-mm mesh sieve. Total soil C and N contents were determined from the dried and milled samples of soil with grain size smaller than 2 mm with an elemental CN analyzer (LECO, Michigan, USA). "

- **Some coordinates were clarified.**

Line 273: "(60°'12'1".1" N, 24°'57'4".7" E)"
changed to line 324: "(60°12'14.0" N, 24°57'38.9" E)".

We remain at your disposal for any further details and discussion.
Kind regards,

Justine Trémeau on behalf of all authors

Associate editor's review – 11.10.2023

Dear Authors,

Thank you for your detailed responses to the constructive comments and suggestions offered by the two referees.

Both reviewers recognized the significance and novelty of your approach and findings and agreed that the manuscript should be conditionally accepted. However, the changes the reviewers suggest fall into the category of 'major revisions'. Therefore, I recommend 'major revisions' before considering the revised manuscript for publication in Biogeosciences. For your revision, please also take into consideration the following points:

- The poor performance of JSBACH: As the second reviewer pointed out, it would be helpful if you explain why the model performed poorly. Your long response to the comment was not clear as to how you would discuss the large gaps between the observed and modeled data points. Please provide an in-depth discussion of critical factors that need to be considered for improving model performance.
We improved the materials and method section by adding the motivation, an additional analysis with adjusted soil carbon in the model and also some metrics (R^2 and RMSE) to compare with the measurements.
In the results section, a description of the additional analysis is provided and the metrics are now available in table 2. Some additional information can be found in the supplementary materials such as the dynamics graphs of 2021 and 2022 with the adjusted soil carbon simulation (Fig. S6 and S7) and NEE, which result from the latter simulation.
Finally, the discussion was improved with a deeper discussion on the JSBACH performance.

Materials and method:
Lines 318-321: "To derive annual estimates of NEE, TER and GPP, the intensive sites were simulated using the JSBACH (Reick et al., 2013), which is the land component of the Max Planck Institute Earth system model (MPI-ESM) (Giorgetta et al., 2013). JSBACH is a process-based model and calculates the dynamic carbon cycle and key driving factors including seasonal dynamics in leaf area, momentary CO2 fluxes, evapotranspiration, soil moisture, litter production and soil carbon dynamics."

Lines 346-353: "The model was used to derive the annual average GPP, TER and NEE for the period 2005–2022. The simulations were set up to represent habitats where the soil organic matter accumulates over time from the litter of standing vegetation. This was achieved by running a long spin-up period (thousands of years). As is often the case in urban areas the soils at the intensive sites had not been accumulated from the litter of the current vegetation, and therefore the soil carbon pools in the model are not equal to the ones present at the sites. Due to this, the simulations may not reproduce the observed TER, but instead represent a more general situation for these habitats. However, we also performed additional simulations where the soil carbon pools were adjusted to meet the observed TER values in 2021 and 2022. The agreement between the simulated and observed carbon fluxes was evaluated by R2 and RMSE."

Results:
Lines 449-451: "In order to represent the average fluxes of the different habitats, the model was primarily initialized to reach a steady state with each vegetation type and therefore, the simulated soil carbon pool represents the result of long-term carbon input of the site itself."

Lines 454-476: "The maximum soil moisture in the model is limited by the field capacity (Fig. S2). At the same time, the modelled soil temperature between 10 and 15 cm depth agreed with the measured soil temperature at 10 cm depth (Fig. S3). Modelled LAI varied in accordance with the observations especially at the mesic meadow and in the second year at the dry meadow (Fig. S4). The dry meadow was established very recently and was found to be sparse and inhomogeneous with respect to the vegetation, and therefore the drought response was challenging to capture in the model simulation particularly in 2021. The drought response for the non-irrigated lawn agreed with the observations in 2021, while in 2022 the 460 recovery was too strong in the simulation. At the irrigated lawn, the simulated LAI decreased later in 2022 than the observed LAI (Fig. S4).
Simulated TER followed the dynamics in the observed seasonal cycle and, for example, showed a similar response to the 2022 drought but generally underestimated the overall level of the non-irrigated lawn and the dry meadow. The drought response was stronger in the mesic and dry meadow (Fig. 3e-h, Fig. S5e-h). For the irrigated lawn the benefits of irrigation may have been overestimated, seen as a weaker drought response in the model than in the observations. The irrigation used in the model was estimated as an average over the whole area, while there probably was less irrigation where the measurement equipment was installed. The highest R2 values for TER were observed at the irrigated lawn and mesic meadow and the lowest at the dry meadow (Table 2). Adjusting the carbon pools to fit closer to observed TER values in 2021 and 2022 (Fig. S6-7) mainly decreased RMSE but did not notably improve R2 values at the different vegetation types in 2021 and 2022 (Table 2).
[…] Simulated daily mean GPP followed the observations closely for non-irrigated lawn and mesic meadow (R2 between 0.41 and 0.85, Table 2), while the dry meadow was difficult to simulate especially in 2021 (Table 2, Fig. S5)."

Lines 485-487: "According to the simulations that were adjusted to meet the TER observations, annual NEE were 328, 2663, 1476 and 2140 g CO2 m−2 yr−1 in 2021 and −195, 1496, 571 and 1205 g CO2 m−2 yr−1 in 2022, at the irrigated lawn, the non-irrigated lawn, the mesic meadow and the dry meadow respectively (Table S4)."

Table 2. R^2 and root mean square error (RMSE, mg CO_2 m^{-2} s^{-1}) calculated for total ecosystem respiration (TER) and gross primary production (GPP) in the standard simulation where the soil carbon was stabilized based on the standing vegetation, and for the adjusted simulations (adj soil c) where the soil carbon pool was set so that simulated TER met the observations in 2021 and 2022.

		TER				GPP			
		standard		adj soil c		standard		adj soil c	
Year	**Site**	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
2021	irrigated lawn	0.86	0.07	0.84	0.07	0.17	0.13	0.15	0.11
	non-irrigated lawn	0.61	0.24	0.47	0.18	0.41	0.10	0.42	0.10
	mesic meadow	0.80	0.23	0.79	0.19	0.85	0.12	0.87	0.10
	dry meadow	0.15	0.20	0.04	0.19	0.02	0.13	0.01	0.13
2022	irrigated lawn	0.62	0.09	0.68	0.08	0.38	0.13	0.39	0.11
	non-irrigated lawn	0.61	0.13	0.53	0.10	0.65	0.08	0.66	0.09
	mesic meadow	0.71	0.12	0.67	0.11	0.85	0.12	0.86	0.10
	dry meadow	0.23	0.20	0.26	0.16	0.49	0.11	0.49	0.11

+ additional tables and figures in the Supplementary materials: Fig. S6 and S7 and Tables S4 and S5

Discussion:
Lines 598-612: "Although the highest ecosystem activities and emissions take place in the warm summer months, to fully validate the annual balance, it would be useful to have some additional wintertime observations. Nevertheless, we can be quite confident in the simulated photosynthetic production, since

GPP principally occurs during the snow-free seasons. Furthermore, soil respiration is highly related to soil carbon quality and quantity (Davidson and Janssens, 2006). Here, it seems evident that the studied soil was not stabilized but, rather, that growing media with unknown properties was either brought to the site or some earlier vegetation type built the soil carbon storage. This is most evident in the young sites (less than 15 years old) and most managed sites i.e., the lawns and the dry meadow. It highlights unpredictable features of urban soils which are characterized by high anthropogenic disturbances, unknown origins and changes in land use. Thus, the quality and quantity of organic matter cannot be connected to a linear history even at a city scale (Pouyat et al., 2006; Setälä et al., 2016; Ivashchenko et al., 2019; Sushko et al., 2019; Cambou et al., 2021). Here, we chose to estimate the carbon balance over an extended time in stabilized conditions and all our sites were estimated to be small sinks of carbon varying between 63–336 g CO2 m−2 yr −1. However, adjusting the model parameters to reproduce values closer to the observed TER values turned most of the sites to sources of atmospheric CO2 during the measurement years 2021 and 2022. Even if those values represent the current situation at the study sites, those represent also an unstabilized state. Whereas the extended runs stand for long-term carbon balances of these vegetation types."

- Line 22: "momentary"? Please use a term that can indicate the duration or interval of the monitoring.
The interval of monitoring was added to the sentence.

Lines 23-24: "The process-based ecosystem model JSBACH was utilized together with the momentary observations collected approximately every second week on CO2 exchange to quantify the annual carbon (C) balance of these sites."

- Lines 25-25 (& 613): Please reformulate the sentences to highlight your key findings on the inter-system differences in CO2, CH4, and N2O fluxes. It is also not clear whether the term 'carbon sink' here refers to the sink for both CO2 and CH4.
The sentence was reformulated as followed:
Lines 26-29: "We found that without considering the impact of management and unstabilized soil, lawns are clear sinks of carbon (NEE = −336 ± 187 and −157 ± 139 g CO2 m−2 yr−1) compared with meadows (NEE = −151 ± 198 and −63 ± 134 g CO2 m−2 yr−1), and the conversion from a lawn to a meadow did not affect the fluxes of CH4 and N2O."

- Line 156: Please describe how you ensured that CO2 flux measurements were conducted after the plants had been adjusted physiologically to the change from light to dark conditions (e.g., by putting intervals between the measurements).
A short explanation was added to the revised manuscript.

Lines 191-192: "There was a pause of at least one minute in between the single light intensities to ventilate the chamber and to allow the plants some time to acclimatize to the changing light level."

- Line 161 CO2, H2O, N2O and CH4 analyzer (DX4015, Gasmet Technologies Oy, Vantaa, Finland): Is this a CRDS-based analyzer or a system of multi-analyzers based on different analytical principles? Please provide essential methodological information including the principles, calibration methods, and QC measures (particularly regarding the very low concentrations of CH4 and N2O mentioned in Line 200).
CO2, H2O, N2O and CH4 analyzer (DX4015, Gasmet Technologies Oy, Vantaa, Finland) is a FTIR-based (Fourier Transform Infrared) analyzer. Some details were added concerning the methods. The same already-described quality control process was applied to the Gasmet data, which does not change the results.

Line 203: "This Fourier Transform Infrared-based analyzer was zero calibrated at the beginning of each measurement day."

Lines 245-248: “Regarding the satellite sites, the filtering process described above was also applied to the Gasmet dataset resulting in discarding 23 measurements of TER out of 384 recordings over the three years of measurements, but no CH_4, nor N_2O measurements (384 recordings of each in total).”

- Tables and figures (and associated text): Please pay attention to putting a space between the mass unit and gas name (e.g., mg CO2).
This was corrected over the manuscript.

I would like to ask you to make all the changes easily identifiable in a marked-up manuscript based on your point-by-point responses to the reviewer comments. If possible, please add up your responses to the original reviewer comments and specify the line numbers of the revised parts in your final responses accompanying the revised manuscript.

Sincerely,

Ji-Hyung Park
Associate Editor, Biogeosciences”

First reviewer’s comments – 18.08.2023

Dear reviewer,
We would like to thank you for your time, effort and support! We have revised the manuscript according to all comments which has resulted, for example, in the removal of species richness analysis and the clarification of the plant functional type analysis. Please find below our point-by-point responses in red color.
Best regards,
Justine Trémeau on behalf of all authors

General comments
This paper contributes some valuable case study data to an important area of research and management. The authors combination of GHG, drought and PFTs is interesting and fruitful approach. I particularly enjoyed the drought dataset and discussion, which I though was especially authoritative and novel. The figures look really good. The paper is well referenced, though add Marshall et al. 2023 Ecological Solutions and Evidence. The conclusions are clear and useful. In general the authors are careful not to overstate their findings and are clear about which results are robust and which are more circumstantial, except when it comes to plant biodiversity. Only non-grass species were identified in grassland systems, which is definitely a weakness. Thus I am uncomfortable with the species richness conclusions. I suggest to drop species richness entirely from the paper and refer only to plant functional types (which are well studied). Or, you could drop the Poaceae sp. taxon from the analyses and make the species richness models explicitly *forb* richness, rather than species richness. And in the title, perhaps ‘biodiversity’ would be more accurate as ‘plant functional types’. I am also missing a brief justification for why intensive study sites were combined with satellite sites. Why not study more sites intensively?
Thank you very much for the kind appreciation! We do understand your finding about the species richness, we dropped out this analysis from the manuscript and adjusted the related section accordingly. The title was changed according to your comment and a short statistical analysis of the different plant functional type cover between lawns and meadows was added (lines 389-390 (methods) and lines 536-541 (results)).
We agree that it would be optimal to study a high number of sites in detail. However, observing CO2 exchange in a temporally representative manner to determine annual carbon balance is laborious and thus expensive. Therefore, we had two sets of sites: 1) intensive sites with good temporal coverage of CO2 exchange for the carbon balance and 2) satellite sites with better spatial coverage to answer the other

research questions. However, as the methods partly overlap, we combined the datasets in the research question 4), now related to plant functional types. We further clarified this at the end of the revised Introduction and in the Materials and Methods.

The title was changed to: "Lawns and meadows in urban green space - A comparison from perspectives of greenhouse gas, drought resilience and plant functional types" and the introduction focuses less on diversity.

Lines 94-96: "The empirical data set included intensive sites with high temporal coverage of CO2 exchange to estimate the carbon balance and satellite sites with high spatial coverage to study the transformation process."

Line 106-108: "In order to study the CO2 exchange with high temporal coverage and to compare the carbon balances of urban grasslands, two pairs of urban lawns and urban meadows, situated 4 kilometers apart, were selected for the intensive measurement sites (Fig. 1)."

Lines 130-131: "To study the transformation process from an urban lawn into an urban meadow, six other locations belonging to local student housing associations were selected for high spatial coverage GHG measurements in the urban green space."

Specific comments and technical corrections

Summary

Line 20 – make clear that this is two sites not four

Contrary to what was mentioned in the previous answer, we did not emphasize it. That is indeed two locations, but four different sites and we chose in the end to reformulate the sentences without mentioning the sites as pairs. We wanted to make the abstract simpler and found that mentioning the pairs was some non-crucial information here (while it is more than crucial why/how they were paired in the main text, cf. below).

Lines 21-26: "Four of these sites i.e., an irrigated lawn, an old mesic meadow, a non-irrigated lawn, and a young dry meadow, were more intensively studied in 2021-2022. […] On the remaining sites, we studied the initial dynamics of conversion from lawns to meadows by transforming part of lawns to meadows in late 2020 and conducting measurements from 2020 to 2022."

Introduction

Line 47 – what is species cortege?

We changed the term to "vegetation community".

Lines 81-83: "However, the relationship between species richness, productivity and resistance is not straightforward but influenced by many factors, such as management, vegetation community and environmental conditions (Vogel et al., 2012; De Keersmaecker et al., 2016; Jung et al., 2020)."

Line 58 – mention some typical species (in the UK often Lolium perenne for a hardwearing sports turf, or Agrostis stolonifera/Festuca rubra for a fine-leaved display lawn)

The most abundant species (according to Tonteri and Haila, 1990) were listed in the introduction.

Lines 53-57: "These lawns typically contain certain selected species of grasses, such as in Southern Finland Agrostis capillaris, Alopecurus pratensis, Dactylis glomerata, Festuca ovina, F. pratensis, F. rubra and Poa annua (Tonteri and Haila, 1990), yet they can also harbor other species of forbs and grasses that are spontaneously established, thus giving lawns the ability to behave like semi-natural grasslands (Thompson et al., 2004; Fischer et al., 2013)."

Line 65 – Marshall et al 2023 also highly relevant: https://besjournals-onlinelibrary-wiley-com.ezp.lib.cam.ac.uk/doi/full/10.1002/2688-8319.12243
True, it is an interesting study. We added the reference in the revised manuscript (line 62).

Methods

line 90 – mention the criteria used to match lawns and meadows e.g. distance, …?
We matched the plots with distance. This is now clarified this in the revised manuscript.

Line 154: "In this study, the lawn and the meadow within a pair were paired considering the short distance between each other."

line 93 - in the UK the first two species are non-native invasive garden escapes. Is this the case in Finland? If so, mention that.
For sure, *Lupinus polyphyllus* is indeed a non-native invasive garden escapes; however, the others are neither alien nor invasive, they are archaeophytes (i.e. introduced in ancient times by humans). In the revised MS, we included a note after *Lupinus* that it is a non-native invasive species.

Line 112-113: "Currently, Aegopodium podagraria, Lupinus polyphyllus (non-native invasive), Dactylis glomerata, Anthriscus sylvestris, Elymus repens, Lamium album and Urtica dioica are the dominant species."

Line 96 – mowing robot! Cool
Yes!

Line 120 – mention species. Weak if the paper includes a plant biodiversity perspective.
We mentioned broadly the species found at the satellite sites, and it is now in line with the revised manuscript, which does not include plant species biodiversity perspective anymore.

Lines 133-134: "All of the six lawns were predominantly covered by grasses (mostly Poa pratensis, Festuca rubra and Lolium perene)."

Methods – I am missing a (brief) comment on why intensive and extensive study approaches were chosen. why 2 x intensive and 8 x extensive, and how sites were allocated to the two study approaches?
Thank you for this comment which made us improve the description! As stated above, it is not feasibly to study labour-intensive aspects with high temporal and spatial coverage and therefore, we combined two approaches to have them both. As shortly explained above, originally, the intensive site study was designed for one project and the satellite site study designed for another project. The first one focuses on carbon sequestration, so the GPP was included in the study, and different types of grasslands, where one of the meadows is old, were studied. The second project focuses on conversion from a lawn to a meadow, so all meadows were transformed at the same time following the same procedure end of 2020 (so young meadows) and there, only TER (no GPP), in addition to CH4 and N2O fluxes were studied. Moreover, for the case of the extensive sites, their locations were chosen from those made available by the housing associations and their number was larger to fulfill the public interest the transformation initiative

produced. This is why we make the distinction between the different sites. The usage of these two sets of sites is now motivated and it improved the description throughout the M&M section.

Lines 94-96: "The empirical data set included intensive sites with high temporal coverage of CO2 exchange to estimate the carbon balance and satellite sites with high spatial coverage to study the transformation process."

Line 106-108: "In order to study the CO2 exchange with high temporal coverage and to compare the carbon balances of urban grasslands, two pairs of urban lawns and urban meadows, situated 4 kilometers apart, were selected for the intensive measurement sites (Fig. 1)."

Lines 130-131: "To study the transformation process from an urban lawn into an urban meadow, six other locations belonging to local student housing associations were selected for high spatial coverage GHG measurements in the urban green space."

Section 2.2 – this is not my area of expertise so I can't comment.

Line 235 – why this difference in approach? A mistake/for convenience? Does it matter for the results?
The sampling actually took place during different years as KMP lawn was studied already during a previous project and the other intensive sites were included later. Unfortunately, the timing of the sampling was missing earlier but we included the sampling dates in the revised manuscript. Truly, spatial coverage in KMP lawn was wider due to a higher number of sampling points (cf. lines 163-166) but the number is still sufficient and thus the difference should not hinder the comparison of the sites. We also want to stress that this data describes the sites as background and is not used to answer any research questions.

Lines 163-163: "The soil temperature, soil moisture, soil samples and plant inventories were collected close to the GHG measurement quadrats except for the irrigated lawn where we used the soil samples collected from the same plot by Ahongshangbam et al. (2023)"

Lines 277-289: "To determine the overall soil characteristics, we collected altogether 1 L of soil up to 25 cm depth at each of the sites. At irrigated lawn, the pooled sample comprised 16–18 individual samples collected in 2020 around the site with a thin auger (d = 2.3 cm). At the rest of the remaining intensive sites, the samples were collected in 2021 with a slightly larger auger (d = 5.0 cm) to collect 4 individual samples that were then pooled together."

Line 263 – combining all grass species to one species is a weakness of the paper, especially as the study sites are grass-dominated grassland systems, and biodiversity is mentioned in the title. It contrasts with the thorough flux, soil and drought methods. Suggest to reframe the paper from 'biodiversity' (for which you don't have data as grass species were not identified except in broad non-taxonomically defined groups) and towards 'plant functional types', for which you have good data.
As described earlier, the species richness perspective was removed from the reviewed manuscript, and the revised manuscript now focuses only on the plant functional types as suggested by the reviewer. We also added a new table (Table 4) to show the cover estimates of different functional types and a short comparison of plant functional type's distributions between lawns and meadows.

Line 331- corrections for multiple tests?
We have now added to the analysis a Kenward-Roger correction (Kenward et al., 1997), because of the small size of the dataset.

Lines 399-400: "P-values were adjusted by a Kenward-Roger approach (Kenward and Roger, 1997) from R lmerTest package (Kuznetsova et al., 2017)"

Results

Figure 2 looks good
Thank you!

It would be easier to follow the results if you used the type of lawn/meadow (following fig 2) instead of the abbreviated site names. i.e. line 373, KMP meadow -> mesic meadow, VKI meadow -> dry meadow.
We changed the names in the results section as suggested and mentioned it in the method.

Lines 108-111: "In Kumpula (KMP) neighborhood, an old, mesic and mesotrophic urban meadow, which could be associated to the "Mesic perennial anthropogenic herbaceous vegetation (V39)" in the EUNIS classification (European Environment Agency, 2023) and hereinafter referred to as "mesic meadow", was paired with a highly managed urban lawn, hereinafter referred to as "irrigated lawn"."

Lines 118-119: "In Viikki (VKI) neighborhood, the urban lawn and urban meadow – hereinafter referred to as "non-irrigated lawn" and "dry meadow" – are situated inside a public park 60 meters apart from each other."

Discussion

Line 480 – or even net positive
We added the aspect of net positive here.

Lines 575-576: "Carbon neutrality, or even a net positive scenario, is nowadays one of the major goals for cities and states in mitigating climate change (IPCC, 2022)."

Line 482 – "a very fertile and a very poor meadow" . What is meant here? Poor in what? Earlier you call the sites 'mesic meadow' i.e. middling, neither especially fertile nor nutrient poor, and 'dry meadow' i.e. lacking moisture . nutrient availability is related to soil moisture, but they are distinct concepts
One is a mesic, mesotrophic meadow and the other one is a dry sandy (->xeric) nutrient-poor meadow. This is now clarified in the manuscript in the discussion, but also in the materials and method.

Lines 108-111: "In Kumpula (KMP) neighborhood, an old, mesic and mesotrophic urban meadow, which could be associated to the "Mesic perennial anthropogenic herbaceous vegetation (V39)" in the EUNIS classification (European Environment Agency, 2023) and hereinafter referred to as "mesic meadow", was paired with a highly managed urban lawn, hereinafter referred to as "irrigated lawn". The meadow is an old agricultural field on which farming practices were abandoned about 40 years ago. "

Lines 121-123: "In 2020, a section of the park's lawns was transformed into a dry and nutrient-poor urban meadow by replacing the topsoil with a layer of recycled sand and sowing seeds (Riikonen and Karilas, 2021)."

Lines 576-578: "Our aim was to understand how CO2 sequestration differs between lawns and meadows by studying contrasting vegetation types, a mesic mesotrophic meadow and a dry nutrient-poor meadow together with irrigated and non-irrigated lawns, to determine the full range of their GHG exchange."

Line 500 good point
Thank you!

Line 511 – yes, I think time since establishment would be important.

Yes, we do agree. It has been clarified that in next studies, it would be valuable to compare meadows with varying fertility but comparable age.

Lines 630-631: "For these reasons, it would be useful in the future to compare homogeneous meadows with varying fertility – from mesic mesotrophic to dry nutrient-poor, but comparable age."

Line 515 – more quadrats would have helped to represent better the variation present at the site?
Yes, probably, but again we cannot say anything about the inhomogeneity, which would probably be the same with more quadrats.

Line 576 and line 585 – no link between plant diversity and C and N cyles in your study could be because you didn't identify grass species?
It is possible, but the section about "plant richness" is now removed from the revised manuscript and the discussion has been updated accordingly.

Conclusions

Line 614 – Meadows do increase biodiversity -> meadows have higher species richness and PFTs
It is changed as suggested in the revised manuscript.

Lines 329-731: "However, while meadows are known to have higher species richness and diversity of plant functional types, we found that meadows do not increase carbon sequestration compared with lawns on an annual scale, and even that NEE is usually more negative in lawns than in meadows."

Line 618 – new sentence .Nonetheless…
It is changed as suggested (line 733).

Line 620 – I am uncomfortable with this conclusion given your approach to plant identification. You could drop species richness entirely from the paper and use only PFTs. Or, you could drop the Poaceae sp. Taxon from the analyses and make the species richness models explicitly forb richness, rather than species richness. And then line 620 should read "Regarding the plant diversity, the FORB richness does not seem to affect significantly 620 the C and N cycles. Some PFTs are correlated to specific C and N status".
We understand your point of view and we chose to discard the richness section. We corrected the discussions and conclusion accordingly.

Lines 736-739: "Regarding the vegetation, forbs proportion was found to be higher in meadows and some plant functional type proportions, including grasses, legumes, forbs, horsetails and mosses, are correlated to specific C and N status. Nevertheless, as this analysis was done with an only one-year dataset, it would be necessary to investigate with more urban grasslands, on the link between plants and C and N cycle."

Second reviewer's comments – 18.09.2023

Dear reviewer,
We would like to thank you for your time, effort and support! We have revised the manuscript according to all comments which has resulted, for example, in unified terminology, in the removal of species richness analysis, as well as in the clarification and further discussion about the JSBACH modelling. Please find below our point-by-point responses in red color.
Best regards,
Justine Trémeau on behalf of all authors

Trémeau et al. provide an interesting study about urban green places, their biodiversity, and their ecosystem functions. They have a novel study setting and new results, and the language in the manuscript is mostly clear. I personally would have split this manuscript in two (e.g., one focusing on flux magnitudes and modeling, and another one on the key ecological and soil drivers) to be able to analyze all the core topics in enough detail. For example, JSBACH seems to perform quite badly (especially for the non-irrigated lawn and dry meadow) and it would have been useful to explore this in detail. However, as this manuscript is now close to its final stage, I do not think such a split makes sense, and have provided my major (or moderate) and minor feedback considering the current structure below.

Major issues:

1. **Biodiversity data**

The terminology related to the biodiversity has to be edited. I think you study species richness and the cover of certain plant functional types, and would use these terms throughout the manuscript. It would also be good to report the number of species that you observed and discuss those in relation to earlier studies.
According to the comments from both reviewers, the species richness analysis is now removed from the manuscript, and where it is expected, the term "diversity" has been changed to "plant functional types". Therefore, we did not include the number of species either. Instead, we added a new table (Table 4) to show the cover estimates of different functional types and a short comparison of plant functional type's distributions between lawns and meadows.

The title was changed to: "Lawns and meadows in urban green space - A comparison from perspectives of greenhouse gas, drought resilience and plant functional types".

Lines 550-554: "Table 4. Cover proportions (%) of the different plant functional types: grasses (Poaceae), legumes (Fabaceae), forbs (other families of flowering vascular plants, which do not belong to one of the listed categories), trees, sedges (Carex), horsetails (Equisetum), and mosses (Bryophyta), inventoried on June 27th and 21st July 2022. The "total" column is the sum of all the plant functional type cover, a fully covered quadrat with only one layer of vegetation should have a 100% cover, >100% indicates layered vegetation (short and tall grassland plants) and <100% indicates the presence of bare soil."

Line 18: Here and throughout – I don't think you are studying plant functional diversity. I generally consider plant functional diversity as diversity metrics derived from trait measurements (see e.g. https://www.frontiersin.org/articles/10.3389/fpls.2022.993051/full). You are considering species richness metrics here. Be careful with the wording throughout the manuscript.
As mentioned above, we removed the species richness analysis and focused on plant functional types. We have carefully gone through the manuscript to be consistent with the terminology.

2. **Cumulative fluxes are highly uncertain due to the poor performance of JSBACH I am not sure if the utilization of JSBACH here makes a lot of sense after seeing the very poor performance of the model. Understanding why the models perform so badly would be important. Also, why was JSBACH chosen as a first step? Is it the best model to predict fluxes in northern grasslands?**

JSBACH is the land model of the earth system model ICON and is thus widely used especially in Europe. There is not a dedicated model for boreal, urban grasslands as those differ greatly from agricultural ecosystems, especially by management and cultivated species and therefore, JSBACH can be assumed to describe these studied habitats as well as any other process-based ecosystem model.

In some cases, the observed emissions (TER) were high compared to the input of carbon from the standing vegetation. It seems evident that the studied soil was not stabilized during the measurements. Such mismatch is a result of urban construction where growing media with unknown properties was brought to the site, and it is not a result of long-term carbon input by the standing vegetation. This is most evident in the young sites (less than 15 years old) i.e., the non-irrigated lawn and the dry meadow, highlighting the variable and unpredictable characteristics of the urban soils that are commonly brought on the site from somewhere else or the land use has changed and therefore, the quality and quantity of organic matter is not connected to the history of the site.

Here, the simulations included a long spin-up period, i.e. in the simulations the soil carbon pools are a result of the carbon input from the vegetation over a long time. The simulations therefore represent carbon fluxes typical for these habitats over longer time periods. We decided to choose this approach as we wanted to be more general and show the longer-term effects of these vegetation types than just two cases where the ecosystems have not stabilized yet. Thus, instead of just fitting the model to the observations and giving the carbon balance of these four sites with unconnected soil underneath, we wanted to give a more general overview of the different habitats in the long term. Therefore, the model was originally initialized to reach a steady state with each vegetation type and the simulated soil carbon pool represents the result of long-term carbon input of the habitat itself. We realized now that we had not sufficiently motivated the use of the model and the approach in the earlier version of the manuscript. In the revised manuscript, we have now clarified why we have used the approach.

It is also possible to better fit the model fluxes to the observations by adjusting the soil carbon pools, as shown in Figures S6 and S7. This, however, does not necessarily result in fluxes that would represent the average fluxes of these habitats over extended periods. Some outliers can still be seen in Figure S6 and S7, mostly data points with large ranges. Furthermore, the soil respiration in the model likely reacts slowly to short hot periods, as the flux is driven by a two-week running mean of the temperature. In the revised manuscript, we added annual balances also by this approach i.e., forcing the model simulation to the observations. In addition, the new figures (S6 and S7) were added in the supplementary materials and discussion about the effect of this decision in the Discussion section.

For details/changes in the manuscript cf. answer to the editor (second section of this document)

Do you come to the same conclusions about the general sink strength when you compare the pure measurements?

The measurements were only conducted every two weeks and there are uncertainties according to the day and the time of the day. Thus, it is not possible to give annual fluxes like with JSBACH. However, it seems evident from the data that the annual balance would have been higher based on the TER observations that were more often higher than the model simulation. In the revised manuscript, as mentioned earlier, we now also provide annual estimates that are estimated with a model version that is forced closer to the observations as already described earlier. R2 and RMSE values are also indicated (Table 2 above).

420: Perhaps add some kind of a model performance metric (R2, RMSE) here to demonstrate that dry meadow estimates, for example, are highly uncertain.

We added some metrics (R2 and RMSE) for discussing the model performance regarding the annual cycle and the changes in the carbon fluxes during a dry period.

Cf. Table 2 (above)

503: This comparison seems quite uncertain here after seeing the major issues with the model. I would put more emphasis on comparing the measured growing season fluxes (and SOC, SON, species richness, and plant communities in general) with earlier studies.
As described earlier, we modified the discussion to include the purpose of the modelling exercise and the uncertainty considering the annual balances of especially TER and NEE of these sites. We also added more values to compare with older studies.

Lines 593-595: ". Decina et al. (2016) found that during the growing season, TER was about 0.198 ± 0.006 mg CO2 m−2 s−1 at urban lawns in Boston, USA, which is lower than the values measured at our sites."

Lines 613-621: "Many have reported urban grasslands to be C sources at least in certain conditions (Allaire et al., 2008; Hiller et al., 2011; Bezyk et al., 2018), but some have also reported sinks of about 20–180 g CO2 m−2 yr−1 (Thienelt and Anderson, 2021). However, most of the studies focus on the respiration rate, omitting GPP (Kaye et al., 2005; Decina et al., 2016; Lerman and Contosta, 2019; Sushko et al., 2019; Upadhyay et al., 2021). Jasek et al. (2020) found that the annual mean of TER for a mix of different types of urban grassland is 424 ± 43 g C m−2 yr−1 in Krakow, Poland, and Kaye et al. (2005) measured that irrigated lawns in Northern Colorado, USA, emits around 2777 ± 273 g C m−2 yr−1, where our annual TER values vary between 1181 ± 48 g C m−2 yr−1 at the irrigated lawn and 481 ± 37 g C m−2 yr−1 at the dry meadow. In the end, we found that lawns were620 stronger sinks than the meadows which is in line with Peoplau et al. (2016) who found that soil C was higher in lawns than in meadows, which was mainly attributed to the clippings left at the site and fertilization."

3. **Methods section needs to be clearer**

I suggest the authors carefully revise the methods section and make sure that it is chronological. Right now the sections are sometimes a bit hard to follow because they can start with a description of sample collection, then go to laboratory analysis, then talk about field description and study design again. I would always start with study design, then move on to field sampling, and then laboratory analysis.
Thank you for the feedback! We have carefully reviewed and edited the manuscript and re-organized the materials and method section in a better way. A "sampling design" section (section 2.1.4, lines 153-164) was added after the site description, in which we describe the design (the transects and four quadrats along the transects), the flux measurements, soil sampling and the statistical sections were also slightly reordered following your order suggestion.

135: Somewhere here, specify how many collars you had and how many times they were visited.
The number of collars has been clarified in the new "2.1.4 sampling design" section and a sentence was added to specify how many times each site, at least one collar at a site, was visited (including also the number of visits, when we were able to measure the light response). A timeline table was also added to the Supplementary materials (Table S2, see below).

Lines 155-156: "GHG flux measurements were conducted at each site along a transect. All the data was collected at four fixed quadrats of approximately 1 m2 distributed along a single transect."

Lines 175-177: "Over the two years of measurements, KMP lawn was visited 23 times, including 15 days of LR, KMP meadow 23 times, including 16 days of LR, VKI lawn and meadow 20 times, including 15 days of LR, and at each site, most of the time the four collars were measured at a site (Table S2)."

Lines 196-198: “In total, the sites were visited thirteen times: three times before transformation, including two times with CH4 and N2O measurements and one time with only CO2, and ten times after transformation, including six times with CH4 and N2O measurements and four times with only CO2 (Table S2).”

133: Maybe it is mentioned here somewhere, but if you were taking dark measurements at a collar, how many replicates did you take?
At each site, we measured four different plots/collars only once. This was clarified in the methods.

Lines 186-192: “The duration of each chamber closure was at least 2 minutes. To determine the light response of NEE, we repeated the measurement five times at each collar under different PAR intensities – 100%, 50%, 25%, 12.5% and 0% of the prevailing sunlight level – created by shading the chamber with netted or opaque fabrics. We first measured without covering the chamber and then covered it with 1–3 layers of mesh cloth. Eventually, we measured the last recording with an opaque cover.”

Lines 198-199: “A single measurement was performed at each of the four selected 1-m2 quadrats at each site (Table S2).”

138: How many measurements did you take to derive the light response? Did you take measurements at several light intensities? (I now see that this is mentioned on line 155 – perhaps make sure the text is in a logical location always; on line 155 also describe the PAR values for the three different light levels)
By default, five measurements were collected at each collar at different light intensities from full ambient sun light (transparent chamber) to no light at all (opaque cover), with three intermediate measurements done at 50%, 25% and 12.5% of full ambient sun light. If some of these needed to be discarded due to poor quality and possible technical errors, at least three measurements were needed to derive the light response curve. A sentence was added to clarify the minimum of measurements needed for the LR curve, and the different light intensities are now explicitly written.

Lines 186-188: “To determine the light response of NEE, we repeated the measurement five times at each collar under different PAR intensities – 100%, 50%, 25%, 12.5% and 0% of the prevailing sunlight level – created by shading the chamber with netted or opaque fabrics.”

Line 258: “In order to fit the LR curve for one collar, a minimum of three measurements done at different light intensities was required.”

162: I am a bit confused now, earlier it was said that it was with a dark chamber. I also find it a bit challenging to follow which gases were measured and why there were differences. It might help to show these various measurement details with a timeline or a table format.
We use two different terminologies to describe the same chamber. We chose “opaque” as a consensus. We appreciate the table idea; we added one to the supplementary materials.

Lines 194-195: “At the satellite sites, GHG measurements were conducted with an opaque chamber monthly, between May–September 2021–2022.”

+ Table S2

Table S2. Measurement campaign timeline.

Total number of TER measurements (total number of GPP measurements, number of quality-controlled GPP measurements) measured with the LI-COR device at the intensive sites. All TER measurements passed the quality control.
Total number of TER, CH_4 and N_2O measurements (−number of TER measurements removed in the quality control, if any) measured with the Gasmet device at the satellite sites. All CH_4 and N_2O measurements passed the quality control.
Total number of TER measurements measured with the Vaisala device at the satellite sites. All measurements passed the quality control.

		Intensive sites				Satellite sites											
		KMP		VKI		JMT3		JMT7		KO4		PK3		SK1		SMT12	
	Week	lawn	meadow	lawn	meadow	lawn	meadow	lawn	meadow	lawn	meadow	lawn	meadow	lawn	meadow	lawn	meadow
2020	29					4	4	4	4	4	4	4	4	4	4	4	4
	34					4	4	4	4	4	4 (−2)	4	4 (−1)	4	4	4	4
	41					4	4	4	4	4	4	4	4	4	4	4	4
	Total					12	12	12	12	12	12 (−2)	12	12 (−1)	12	12	12	12
2021	20					4	4	4	4	4	4	4	4	4	4	4	4
	23	3 (3, 3)	3 (3, 3)														
	24			4 (4, 4)	4 (4, 3)					4	4	4	4				
	25	4 (4, 4)	4 (4, 4)			4	4	4	4					4	4	4	4
	26			4 (4, 3)	4 (4, 2)												
	27	4 (4, 4)	3 (3.3)														
	28			4 (4, 4)	4 (4, 2)	4	4	4	4	4	4	4	4	4	4	4	4
	29	4 (4, 2)	4 (4, 4)														
	30			4 (4, 4)	4 (4, 4)												
	31		4 (4, 3)														
	32			4 (4, 2)	4 (4, 4)												
	33	4 (2, 2)	3 (1, 2)			4	4	4	4	4	4	4	4	4	4	4	4
	34			4 (4, 4)	3 (2, 2)												
	35	2 (1, 1)															
	36			4 (0, 0)	4 (0, 0)												
	37	4 (0, 0)	4 (0, 0)														
	38					4	4	4	4 (−1)	4	4	4	4	4	4	4 (−2)	4
	40	4 (3, 2)	3 (3, 3)														
	41			4 (0, 0)	4 (0, 0)												
	44	4 (4, 4)	4 (0, 0)	4 (0, 0)	4 (0, 0)												
	48	4 (0, 0)	4 (0, 0)														
	Total	37 (27,36 14)	(27,36 17)	(24,36 20)	(24,35 17)	(22, 20	20	20	20 (−1)	20	20	20	20	20	20	20 (−2)	20
2022	19	4 (0, 0)	4 (1, 1)														
	20			4 (4, 3)	4 (4, 1)	4	4	4	4	4	4 (−2)	4	4 (−1)	4	4	4	4 (−1)
	21	4 (4, 4)	4 (4, 2)														
	22			4 (4, 4)	4 (4, 4)												
	23	2 (1, 1)	4 (2, 2)														
	24			4 (4, 4)	4 (4, 4)	4	4	4	4	4	4	4	4	4	4	4	4
	25	4 (4, 4)	4 (3, 3)														
	26			4 (4, 4)	4 (3, 3)												
	27	4 (3, 3)	4 (4, 4)														
	28			4 (4, 4)	4 (4, 4)	4 (−1)	4 (−1)	4	4 (−2)	4	4	4	4 (−1)	4 (−1)	4 (−1)	4	4 (−1)
	29	4 (0, 0)	4 (0, 0)														
	30			4 (4, 4)	4 (4, 4)												
	31	4 (3, 3)	3 (1, 1)														
	32			4 (4, 3)	4 (4, 4)												
	33	5 (4, 4)	4 (2, 2)			4	4	4	4	4	4	4	4	4	4	4	4
	34			4 (4, 4)	4 (1, 1)												
	35	4 (0, 0)	4 (4, 1)														
	37	4 (3, 3)	4 (0, 0)														
	38					4	4	4 (−1)	4 (−1)					4	4 (−3)	4	4
	39			4 (4, 0)	4 (4, 0)					4	4	4	4				
	40	4 (0, 0)	4 (0, 0)														
	42			4 (0, 0)	4 (0, 0)												
	43	4 (2, 1)	4 (1, 1)														
	45	4 (0, 0)	4 (0, 0)	4 (0, 0)	4 (0, 0)												
	Total	51 (24,51 15)	(24,51 16)	(22,44 30)	(36,44 25)	(32,	20 (−1)	20 (−1)	20 (−1)	20 (−3)	20	20 (−2)	20	20 (−2)	20 (−1)	20 (−4)	20	20 (−2)
Total years	over 3	88 (51,87 29)	(51,87 33)	(46,80 50)	(60,79 42)	(54,	52 (−1)	52 (−1)	52 (−1)	52 (−4)	52	52 (−4)	52	52 (−3)	52 (−1)	52 (−4)	52 (−2)	52 (−2)

167: Should the number and distribution of plots be presented already earlier with the study designs? Perhaps with a description of how many times they were visited in total?
As already mentioned earlier, we clarified the description by adding a section right after the site description, called "2.1.4 Sampling design" section, in which we described the number and distribution of the plots at each site. We also added two sentences to describe how many times the sites were visited in total (TER), and the number of times when we recorded LR.

Lines 155-156: "GHG flux measurements were conducted at each site along a transect. All the data was collected at four fixed quadrats of approximately 1 m2 distributed along a single transect."

Lines 175-177: "Over the two years of measurements, KMP lawn was visited 23 times, including 15 days of LR, KMP meadow 23 times, including 16 days of LR, VKI lawn and meadow 20 times, including 15 days of LR, and at each site, most of the time the four collars were measured at a site (Table S2)."

Lines 196-198: "In total, the sites were visited thirteen times: three times before transformation, including two times with CH4 and N2O measurements and one time with only CO2, and ten times after transformation, including six times with CH4 and N2O measurements and four times with only CO2 (Table S2)."

205: Earlier it was said that ca. 10 % of PAR variability was accepted?
The 10% of PAR variability was used on the field, and then the 150 µmol m−2 s−1 threshold was used by the Python script. Thus, there were two control steps at two different phases. However, after double checking, it appears that the first filtering is more restrictive than the second and no measurements were deleted from the second filtering. We chose to remove the mention of this second phase of filtering.

230: Did I understand correctly that no soil temperature or moisture data were collected during the measurements? This seems problematic to me since these are two key drivers of GHG fluxes. After seeing Fig. 2 I think you measured these but I am not sure
As stated in the materials and method section, soil temperature and soil moisture were all measured at the same time as the fluxes were recorded. To make it clearer, the description was moved from the flux measurement protocol subsection to soil measurement section (section 2.3).

Line 163: "The soil temperature, soil moisture, soil samples and plant inventories were collected close to the GHG measurement quadrats […]."

Lines 269-274: "2.3.1 Soil temperature and soil moisture
Soil temperature and soil moisture were always measured together with the chamber measurements within the 1-m2 270 quadrats at both the intensive and the satellite sites. One replicate of soil temperature was measured within the quadrats at a 10-cm depth with a handheld soil thermometer (HH376, Omega Engineering Inc., Connecticut, USA), and 4–5 replicates of soil moisture at 5 cm depth were measured with a handheld setup (ML3 ThetaProbe and HH2 Moisture Meter, Delta-T Devices Ltd., Cambridge, UK)."

248: I would state the number of samples, study designs, and collection time early in the paragraph instead of late in the paragraph.
We added a section "2.1.4 Sampling design" (lines 153-164) much earlier in the materials and method, and added a sentence at the beginning of the paragraphs to clarify that it is the start of the satellite site's paragraph (after the paragraph describing the method for the intensive sites).

Line 276: "At the intensive sites, overall soil characteristics, soil density and C and N contents were analyzed."

Line 296: “At the satellite sites, overall soil characteristics, soil C and N contents and soil particle distribution were also determined.”

273: Why not the SMEAR station that was used earlier?
The data from the Finnish Meteorological Institute (FMI) is produced for the operational weather forecast and likely has a more rigorous quality control. The stations are situated on the same campus area and show very similar data. The FMI observations were gapfilled using the SMEAR III data.

276: How were the LAI data used in the model? Did it have to be gapfilled?
The LAI was not gapfilled for the simulations. The model does not use the observed LAI, except for the maximum value of LAI. The JSBACH has its own phenology model that calculates the growth and shedding rates based on environmental conditions (mainly temperature and soil moisture). We clarified it.

Lines 330-333: “The maximum LAI for each site (Table S3) was set based on Sentinel-2 data (Nevalainen et al., 2022; Nevalainen, 2022), whereas the model simulated the seasonal LAI dynamics driven by temperature and precipitation. In addition, in the simulation of the mesic meadow the shedding of the grass was activated 65 days after the growth started, to better simulate the observed LAI.”

283: Do I read this correctly – the main property that differentiates the lawns and meadows in the model is the satellite-derived LAI (because the meteorological data were the same) and the flux measurements?
The model doesn’t use the satellite derived LAI as input. Only the maximum LAI, used by the phenology model, was estimated from the observations. This is now clarified in the revised Materials and Methods section. The model calculates the growth and shedding rates from the environmental conditions. The model also calculates the soil moisture, which depends on the precipitation, temperature, radiation, vegetation and soil properties. The fluxes are not used by the model either, all fluxes are calculated by the model. It is true that the LAI has a big impact on the fluxes etc., however, as the model derives the LAI several factors will affect the simulated fluxes.

Lines 330-333: “The maximum LAI for each site (Table S3) was set based on Sentinel-2 data (Nevalainen et al., 2022; Nevalainen, 2022), whereas the model simulated the seasonal LAI dynamics driven by temperature and precipitation. In addition, in the simulation of the mesic meadow the shedding of the grass was activated 65 days after the growth started, to better simulate the observed LAI.”

Lines 335-337: “However, the volumetric field capacity and wilting point for each site were adjusted based on the soil moisture measurements (Table S3). The root depths of the lawns were assumed to be shallower than of the meadows (Table S3).”

295: Can you explain how the model was using the GHG flux measurements? Were the parameters somehow altered based on that information?
Only the GPP was used to adjust the model parameters. The soil carbon pools, that the model uses for calculating the heterotrophic soil respiration, were accumulated by running a long spin-up period (thousands of years). The simulated total ecosystem respiration will therefore not always reproduce observed fluxes. For example in cases where the vegetation has been changed recently, which may often be the case in urban areas. In the revised manuscript, we clarify how the simulations were performed.

Lines 346-352: “The model was used to derive the annual average GPP, TER and NEE for the period 2005–2022. The simulations were set up to represent habitats where the soil organic matter accumulates over time from the litter of standing vegetation. This was achieved by running a long spin-up period (thousands of years). As is often the case in urban areas the soils at the intensive sites had not been accumulated from the litter of the current vegetation, and therefore the soil carbon pools in the model are 350 not equal to the ones present at the sites. Due to this, the simulations may not reproduce the observed TER, but instead

represent a more general situation for these habitats. However, we also performed additional simulations where the soil carbon pools were adjusted to meet the observed TER values in 2021 and 2022."

295: Did you simulate for the peak winters too? Fig. 3 only shows visualizations that lack winter months. How did the model perform during the winter?
The model was run for full years, also including the winter months. We don't have measurements for the winter months and therefore no direct comparison was made. The only flux that is active during the winter is the heterotrophic soil respiration, which decreases slowly when the soil temperature decreases, and reaches a minimum in late winter. We added some additional information.

Lines 493-494 (in the caption of Fig. 3): "The model simulated the whole year but for clarity, Jan-March and Dec are not visible as there were no measurements during those months."

Lines 483-484: "Additionally, it was found that 13–16 % of TER occurred between November and March."

Lines 596-597: "The model was also run for winter months with the heterotrophic soil respiration decreasing at the same time as the soil temperature until reaching a minimum in late winter."

301: reference to FAO
It was added to the reviewed manuscript (cf. the "added references" section at the very end of this document).

Lines 557-559: "The Penman-Monteith equation (Monteith, 1965) was used to calculate the Potential Evapotranspiration (PET) with the FAO-56 method (Allen et al., 1998)."

310: components - resistance and recovery - we s…
It is corrected in the reviewed manuscript.

Lines 367-369: "As resilience is the result of both components – resistance and recovery – we studied in this paper only the resistance, which was defined as "the magnitude of disturbance that a system can absorb before shifting from one state to another" (Capdevila et al., 2021)."

329: SOC and SON not defined yet. Also, how do you know that you do not have inorganic C? Did you test for carbonates?
Thank you, we defined SOC and SON in the revised manuscript. We assumed that the share of inorganic C and N is very low and added a sentence about that in the revised manuscript.

Lines 293-294: "Total soil C and N contents were assumed to be only organic components without carbonates i.e., soil organic carbon (SOC) and soil organic nitrogen (SON)."

Lines 301-303: "Total soil C and N contents were assumed to be only organic components without carbonates, and SOC and SON were determined from freeze-dried and milled samples with an elemental CN analyzer (LECO, Michigan, USA)."

335: What is plant functional category? Not defined yet. This should be described in detail in the methods section.
Now, it is described in the plant inventory section.

Lines 310-312: "In each of the quadrats (see Section 2.1.4), the cover of each plant species was estimated and classified in one of the following plant functional type categories: forb, grass, horsetail, legume, moss,

sedge or tree, where forb includes all other families of flowering vascular plants, which do not belong to one of the already listed categories."

4. **Some potential issues with the statistics need to be addressed**

333: What is location? The coordinates? The name of the study design (intensive, spatial)?
Thank you for noting this! "Location" factor describes the site of the lawn/meadow pairs. We changed the formulation to make it clearer in the revised manuscript.

Lines 391-392: "Then, the effects of vegetation were assessed by applying linear mixed-effects models (LMMs), using the R lme4 package (Bates et al., 2015), where a factor describing the locations of the lawn/meadow pairs (Fig. 1) was included as a random effect."

335: Did you check the correlation of the predictor variables? It is generally not recommended to put variables that correlate >0.7 with each other to a multivariate linear model. Predictor collinearity can for example reverse the slopes and cause many other issues in the model.
It had not been done so far, but now it is corrected: correlation factor between Sedges and Horsetails was 0.73, so Sedges factor was removed from the analysis since sedges covered only 0.5% of one of our sites vs. horsetails were found in two sites and covered 30% of one of them. The correlation table now figures in the supplementary materials.

Lines 394-396: "If predictive variables i.e., plant functional types correlated over 0.70, one of the correlated variables was discarded from the LMM according to the lowest cover proportion and the lowest occurrence in the dataset."

Lines 543-545: "Next, we explored possible correlations between plant functional types and variables related to C and N cycles. First, we discarded sedges from the analysis since it was correlated with horsetails (> 0.7, Table S7) and only present at one site in a low proportion (Table 3)."

Table 4. Cover proportions (%) of the different plant functional types: grasses (*Poaceae*), legumes (*Fabaceae*), forbs (other families of flowering vascular plants, which do not belong to one of the listed categories), trees, sedges (*Carex*), horsetails (*Equisetum*), and mosses (*Bryophyta*), inventoried on June 27th and 21st July 2022. The "total" column is the sum of all the plant functional type cover, a fully covered quadrat with only one layer of vegetation should have a 100% cover, >100% indicates layered vegetation (short and tall grassland plants) and <100% indicates the presence of bare soil.

Experiment	Location	Management	Grasses	Legumes	Forbs	Trees	Sedges	Horsetails	Mosses	Total
Intensive	KMP	lawn	70.5	6.0	21.5	0.0	0.0	0.0	0.0	98.0
		meadow	30.0	16.5	70.3	0.0	0.0	0.0	0.0	116.8
	VKI	lawn	92.3	0.3	7.5	0.0	0.0	0.0	0.0	100.0
		meadow	6.3	38.3	23.3	0.0	0.0	0.0	0.0	67.8
Satellite	JMT-3	lawn	78.8	6.6	10.1	0.0	0.0	0.0	0.0	95.4
		meadow	6.5	9.1	42.7	0.4	0.0	0.0	0.0	58.8
	JMT-7	lawn	21.3	20.6	37.5	0.0	0.0	0.0	0.0	79.4
		meadow	31.9	13.0	68.3	0.0	0.0	0.0	0.0	113.1
	KO-4	lawn	78.8	3.9	15.1	0.0	0.0	0.3	0.0	98.0
		meadow	10.0	0.4	53.0	0.0	0.0	0.0	0.0	63.4
	PK-3	lawn	39.8	0.8	10.2	0.2	0.0	0.0	38.6	89.5
		meadow	23.8	7.0	104.8	0.0	0.5	30.0	0.0	166.0
	SK-1	lawn	65.0	9.7	24.3	0.0	0.0	0.0	0.0	99.0
		meadow	81.3	2.0	52.8	0.0	0.0	0.0	0.0	136.0
	SMT-12	lawn	66.3	2.3	27.8	1.3	0.0	0.0	0.0	97.7
		meadow	15.8	1.8	106.0	0.3	0.0	0.0	0.0	123.8

Table S6. Correlation values between cover proportion of the plant functional types. The plant functional types are legumes (*Fabaceae*), grasses (*Poaceae*), forbs (other families of flowering vascular plants, which do not belong to one of the listed categories), trees, sedges (*Carex*), horsetails (*Equisetum*), and mosses (*Bryophyta*).

	Grasses	Legumes	Forbs	Trees	Sedges	Horsetails	Mosses
Grasses	1.00						
Legumes	−0.42	1.00					
Forbs	−0.55	0.21	1.00				
Trees	−0.24	−0.28	0.11	1.00			
Sedges	−0.14	0.08	0.36	−0.15	1.00		
Horsetails	0.08	0.01	0.11	−0.21	0.73	1.00	
Mosses	0.03	−0.31	−0.31	0.33	−0.07	−0.10	1.00

339: What do you mean by testing any regression?
The original wording was wrong, by "running" would have been better. Nevertheless, it was completely removed from the sentence for better clarification.

Lines 396-397: "Next, the normal distribution of each regression's residuals was visually checked with Quantile-Quantile (Q-Q) plots."

440: Did you control for soil moisture, soil temperature and air temperature variations between the two groups here? Did you do the measurements at the control and experiment during the same day and same conditions? How robust is this comparison?
Soil temperature and soil moisture were measured at the same time as the fluxes were measured. There were no statistical differences between the two treatments (Fig. S7). Each individual lawn/meadow pair was always measured consecutively during the same day to achieve as similar conditions as possible. It is clarified in the revised manuscript.

Lines 383-386: "Considering the transformation from lawns into meadows at the satellite sites, Shapiro-Wilk tests were used to evaluate the normal distribution of each variable (i.e., CO2 fluxes, CH4 fluxes, NO2 fluxes, soil moisture and soil temperature) by year and treatment (meadow vs. lawn). Then, differences in given variables between lawns and meadows were assessed either by t-tests or Mann-Whitney U tests depending on the normality of the data."

Lines 528-530: "The measurements at the control lawn and transformed meadow of each site were conducted consecutively during the same day, and therefore the environmental conditions were considered to be as similar as possible. Indeed, soil moisture and temperature did not differ between the treatments during the measurements (Fig. S9)."

450: But the R^2 is still high for some variables? Why is that? Also, is it possible that the variables would be non-linearly related but you did not test that? I would consider exploring the non-linearity here too. Also, how much variability did you have for species richness metrics (and how does that compare with earlier studies)? If there is not much variability in the first place, it cannot really explain the fluxes either.
This richness section has been deleted from the manuscript as mentioned above, because the grass identification was missing in the study.

Minor

19: Probably good to write these as full words when first mentioned (carbon dioxide,…). Same with GHG, C, and N.
They are now mentioned in full word.

In the abstract:
Lines 20-21: "In eight lawns and eight meadows in the Helsinki Metropolitan Area, Finland, carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) fluxes were measured using manual chambers and plant functional types were recorded."

Lines 23-24: "The process-based ecosystem model JSBACH was utilized together with the momentary observations collected approximately every second week on CO_2 exchange to quantify the annual carbon (C) balance of these sites."

In the main text:
Lines 43-45: "In the urban context, green spaces not only sequester atmospheric carbon (C) but provide vital ecosystem services, such as cooling, recreation, purification of air and water, and risk reduction for flooding (Niemelä et al., 2010; Belmeziti et al., 2018; 45 Lampinen et al., 2021; Shen et al., 2023)."

Lines 91-92: "In addition, we were interested to see if we could detect any connections between the different plant functional types and carbon and nitrogen (N) pools."

Lines 168-171: "At the intensive sites, the carbon dioxide (CO_2) exchange measurements were conducted more frequently than at the satellite sites, and they also included the light response (LR) of net ecosystem exchange (NEE). At the satellite sites, the measurements also included methane (CH_4) and nitrous oxide (N_2O) exchanges, but measurements were conducted with opaque chambers, thus disabling photosynthesis."

34: Slightly confusing to refer to "these intensive study sites" and "six other pairs" here when they have not been introduced in the abstract
We made it clearer; we don't mention the pairs anymore in the abstracts.

Lines 20-26: "In eight lawns and eight meadows in the Helsinki Metropolitan Area, Finland, carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) fluxes were measured using manual chambers and plant functional types were recorded. Four of these sites i.e., an irrigated lawn, an old mesic meadow, a non-irrigated lawn, and a young dry meadow, were more intensively studied in 2021-2022. […] On the remaining sites, we studied the initial dynamics of conversion from lawns to meadows by transforming part of lawns to meadows in late 2020 and conducting measurements from 2020 to 2022."

25: Can you give some numbers for the sink strength (e.g. just for the growing season)?
It is now added to the revised abstract.

Lines 26-28: "We found that without considering the impact of management and unstabilized soil, lawns are clear sinks of carbon (NEE = −336 ± 187 and −157 ± 139 g CO_2 m−2 yr−1) compared with meadows (NEE = −151 ± 198 and −63 ± 134 g CO_2 m−2 yr−1) […]."

25: slightly confusing way to use the word "whereas" here. It usually is tightly linked to the previous sentence (e.g., lawns are clear sinks, whereas meadows are sources) – now these two sentences do not seem to be linked at all
This is now corrected; the organization of the sentences was changed.

Lines 26-30: "We found that without considering the impact of management and unstabilized soil, lawns are clear sinks of carbon (NEE = −336 ± 187 and −157 ± 139 g CO_2 m^{-2} yr^{-1}) compared with meadows (NEE = −151 ± 198 and −63 ± 134 g CO_2 m^{-2} yr^{-1}), and the conversion from a lawn to a meadow did not affect the fluxes of CH_4 and N_2O. Moreover, the mesic meadow was more resistant to drought events than the non-irrigated lawn."

29: I would write a stronger concluding sentence
Thank you, it makes sense. We changed the concluding sentence as follows:

Lines 30-32: "Lastly, the proportion of flowering plants (forbs, including legumes) was found to be higher in meadows than in lawns. Even though social and economic aspects also steer urban development, these results can guide planning when considering biodiversity and carbon-smartness."

31: You need more references throughout the paragraph. I would also consider including the urban perspective here all ready
We added more references throughout the paragraph and included the urban perspective already in the second paragraph of the introduction (also considering the fact that the introduction was adjusted to focus less on the plant diversity aspect).

Lines 41-42 and 44: "(Niemelä et al., 2010; […])"
Line 45: "([…]; Shen et al., 2023)"
Line 47: "(Haaland and van den Bosch, 2015)"
Lines 50-51: "(Hedblom et al., 2017; Ignatieva et al., 2020)"
Line 52: "(Ignatieva et al., 2015, 2017; Zobec et al., 2020; Paudel and States, 2023)"
Line 55: "(Tonteri and Haila, 1990)"
Line 62: "([…]; Marshall et al., 2023)"
Line 64: "(Smith et al., 2015; Unterweger et al., 2017; Norton et al., 2019; Ignatieva et al., 2020)"
Line 67: "(Selhorst and Lal, 2013; Kong et al., 2014; Wang et al., 2022)"
Line 68: "(Kaye et al., 2005; Pouyat et al., 2006)"
Line 70: "(Kaye et al., 2005; Upadhyay et al., 2021)"

For more details see the "added references" section at the very end of this document.

55-58: References needed
We slightly reshaped the paragraph and added key references here as well.

Lines 50-57: "Within urban landscapes, lawns constitute one of the most common features of green space (Hedblom et al., 2017; Ignatieva et al., 2020), and are usually subjected to frequent and intense management to fulfil social, aesthetic and recreational purposes (Ignatieva et al., 2015, 2017; Zobec et al., 2020; Paudel and States, 2023). Although many types of lawns exist, most of the urban green spaces worldwide are dominated by turfgrass lawns. These lawns typically contain certain selected species of grasses, such as in Southern Finland Agrostis capillaris, Alopecurus pratensis, Dactylis glomerata, Festuca ovina, F. pratensis, F. rubra and Poa annua (Tonteri and Haila, 1990), yet they can also harbor other species of forbs and grasses that are spontaneously established, thus giving lawns the ability to behave like semi-natural grasslands (Thompson et al., 2004; Fischer et al., 2013)."

64: i.e. meadows? Make sure that meadows are clearly defined too
It is now defined clearly.

Line 60: "An alternative approach to creating more environmentally friendly and biodiverse urban grasslands consists of substituting or modifying part of the lawnscape from the short, monocultural,

homogeneous setting of grass species into meadows, a flowering setting with extensive management and possible active incorporation of forbs (Southon et al., 2017; Lane et al., 2019; Norton et al., 2019; Bretzel et al., 2020; Marshall et al., 2023)."

70: Nice and clear research questions!
Thank you!

77: Are the recently created meadows not urban because that word is not used here but it is used in the other vegetation types?
No, they are all urban, but we just omitted how crucial it could be to have it. We clarified the nomenclature in the introduction and added a sentence at the beginning of the materials and method section.

Lines 91-93: "In addition, we were interested to see if we could detect any connections between the different plant functional types and carbon and nitrogen (N) pools. To answer these questions, we measured GHG fluxes in urban lawns, in recently created urban meadows and in an older urban meadow in the Helsinki Metropolitan Area in Finland."

Line 104: "The data collection took place on intensive and satellite sites which all are considered urban."

125: It would be more interesting to show the distribution of urban green spaces here instead of roads.
We agree and we also thought to show urban green spaces first, but there is no urban green spaces layer available for the whole Helsinki Metropolitan Area, but just for some cities. Therefore, we don't have enough material to improve the map further. The main idea of the map is to show how the sites are spread around the sites and not for example the green corridors which are not in the focus of this manuscript.

130: Add uncertainty estimates or standard deviations? Also, what is the type of vegetation column, it was not described in the text and seems to be a mixture of genus, moisture and mixture definitions?
We added standard deviation for the values, that were not just from a single measurement. The vegetation type column is here to give an idea of the vegetation found at each site. We changed it to 'Vegetation's characteristics' for a better understanding.

Table 1. Site characteristics at the intensive and satellite sites. The sown mixture indicates seeds of pollinator-friendly forbs, grass species and *Rhinanthus minor* which were sown in late 2020 and veg. refers to vegetation. (± standard deviation) calculated for at least four replicates, at each VKI site there was only one replicate for each variable. Due to Finnish classification, approximations were made for sand (0.06–2.0 mm), silt (0.002–0.06 mm) and clay (<0.002 mm).

Experiment	Location	Management	Soil texture	Vegetation's characteristic	Irrigation	Sand (%)	Silt (%)	Clay (%)	pH	SOC (%)	SON (%)	CN ratio
Intensive	KMP	irrigated lawn	sandy loam	Grasses (Poa)	Yes	69.5	22.1	8.4	5.6	3.8 (± 1.2)	0.25 (± 0.06)	14.8 (± 2.8)
		meadow	sandy loam	Mesic veg.	No	59.1	31.8	9.1	6.5	5.6 (± 1.8)	0.46 (± 0.16)	12.3 (± 0.7)
	VKI	lawn	sandy loam	Grasses (Fescue)	No	72.2	25.6	2.2	6.3	2.8	0.19	14.5
		meadow	sandy loam	Xerophilic veg.	No	72.4	17.1	10.5	6.1	2.5	0.17	14.9
Satellite	JMT-3	lawn	silt loam	Grasses	No	5.2	68.9	26.0	5.3 (± 0.1)	5.0 (± 0.4)	0.46 (± 0.04)	10.8 (± 0.1)
		meadow	loamy sand	Sown mixture	No	73.7	23.5	2.8	6.5 (± 0.2)	4.6 (± 0.7)	0.18 (± 0.03)	25.7 (± 0.2)
	JMT-7	lawn	loamy sand	Grasses	No	82.7	14.9	2.4	6.5 (± 0.1)	4.0 (± 0.5)	0.26 (± 0.04)	15.0 (± 0.5)
		meadow	loamy sand	Sown mixture	No	81.4	16.0	2.6	6.5 (± 0.2)	3.2 (± 0.3)	0.21 (± 0.02)	15.3 (± 0.6)
	KO-4	lawn	sandy loam	Grasses	No	66.2	29.4	4.4	6.4 (± 0.3)	4.4 (± 1.2)	0.29 (± 0.08)	15.4 (± 0.6)
		meadow	sandy loam	Sown mixture	No	63.3	32.7	4.0	6.1 (± 0.4)	4.0 (± 0.7)	0.19 (± 0.04)	20.9 (± 1.7)
	PK-3	lawn	sandy loam	Grasses	No	68.8	27.6	3.6	6.2 (± 0.1)	4.0 (± 0.7)	0.19 (± 0.03)	21.1 (± 0.4)
		meadow	sandy loam	Sown mixture	No	68.4	27.1	4.5	6.1 (± 0.1)	5.4 (± 0.3)	0.32 (± 0.04)	17.2 (± 1.8)
	SK-1	lawn	sandy loam	Grasses	No	57.3	33.9	8.8	5.6 (± 0.1)	4.5 (± 0.3)	0.37 (± 0.03)	12.3 (± 0.5)
		meadow	sandy loam	Sown mixture	No	54.1	36.5	9.4	5.4 (± 0.0)	4.2 (± 0.1)	0.35 (± 0.01)	12.0 (± 0.3)
	SMT-12	lawn	loam	Grasses	No	47.7	42.7	9.7	5.8 (± 0.1)	3.6 (± 0.3)	0.29 (±0.05)	12.8 (± 1.3)
		meadow	loam	Sown mixture	No	51.8	39.6	8.6	6.1 (± 0.2)	4.4 (± 0.7)	0.31 (± 0.05)	14.2 (± 0.4)

365: The uncertainties in the plot are standard deviations across years? So should the standard deviations be the same for the two years for the same time period?
The wording in the caption was apparently confusing, but the standard deviation was calculated for each day of measurements with the four replicates measured at a site. It is now rephrased in the caption.

Lines 421-426: "Figure 2. (a) Weekly precipitation (Prec.) and mean air temperature (Tav) recorded by FMI meteorology station, (b) manually measured soil moisture at a 5 cm depth, (c) manually measured soil temperature at a 10 cm depth, (d) leaf area index from satellite Sentinel-2 and (e) manually measured green cover at the intensive study sites with standard deviations calculated for each site at a daily scale. The irrigated lawn is KMP lawn, the non-irrigated lawn is VKI lawn, the mesic meadow is KMP meadow, and the dry meadow is VKI meadow. Red rectangles indicate the drought periods according to the SPEI and light grey rectangles represent the summer season. In panel (a), Tmax and Tmin represent the weekly instantaneous maximum and minimum mean temperatures."

479: Can you give a clearer and stronger concluding sentence here?
We arranged a stronger concluding sentence as recommended, as follows:

Line 571-573: "Furthermore, as expected, the forbs proportion was found to be higher in meadows than in lawns and seems to be negatively associated with N2O fluxes and C/N ratio, and positively to SOC."

485: Are you talking about annual fluxes here now?
It is the case, and it has been specified in the revised version.

Lines 580-583: "At an annual level, irrigation increased the GPP of a lawn by over 40% and the sink (NEE) by more than 100%. It has already been found that water input improves carbon uptakes (Thienelt and Anderson, 2021) and demonstrated by Zirkle et al. (2011) that irrigated lawns store up to 10 g C m−2 yr−1 more carbon in soil than non-irrigated ones. Also, our analysis on annual NEE indicated that an irrigated lawn is a stronger sink, approximately 50 g C m−2 yr−1 higher than a non-irrigated lawn."

555: not just water use efficiency, but also light and nutrient use efficiency?
It is now added to the revised manuscript.

Lines 570-572: "Such a high diversity of plants and niches, as we can find in meadows, leads to a better nutrient, light and water-use efficiency (De Boeck et al., 2006; Walde et al., 2021), and helps to face water shortage."

613: Is this statement true? On some years the average CH4 and N2O fluxes were positive.
This is true, in some years the fluxes were positive, but when comparing lawns to meadows they are similar, so the positive fluxes are not due to the transformation it-self. To make it clearer, this part was slightly rephrased to make.

Lines 727-728: "The transformations from lawns to meadows did not imply any additional negative climate effects in terms of ecosystem respiration, CH4 and N2O fluxes."

617: What do you mean by larger scale?
It was there to mention a higher number of sites, which has is now clarified.

Lines 733-734: "Nonetheless, it would be necessary to check this result with a larger number of sites and investigate the overall resilience with longer time series dataset, to help stakeholders and city planners to make better decisions to optimize land-use."

626: What about the analysis codes, will you share them too?
Gasmet and LI-COR codes will be shared in the new Code availability section at the end of the manuscript.

Lines 745-747: "Code availability
The scripts used to calculate LI-COR and Gasmet fluxes can be found at: https://github.com/hvekuri/Chamber-codes/ (Vekuri, 2023)."

Added references:

Ahongshangbam, J., Kulmala, L., Soininen, J., Frühauf, Y., Karvinen, E., Salmon, Y., Lintunen, A., Karvonen, A., and Järvi, L.: Sap flow and leaf gas exchange response to a drought and heatwave in urban green spaces in a Nordic city, Biogeosciences, 20, 4455–4475, https://doi.org/10.5194/bg-20-4455-2023, 2023.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements, edited by: Food and Agriculture Organization of the United Nations, Food and Agriculture Organization of the United Nations, Rome, 300 pp., 1998.

Bezyk, Y., Dorodnikov, M., Górka, M., Sówka, I., and Sawiński, T.: Temperature and soil moisture control CO2 flux and CH4 oxidation in urban ecosystems, Geochemistry, 125989, https://doi.org/10.1016/j.chemer.2023.125989, 2023.

Cambou, A., Saby, N. P. A., Hunault, G., Nold, F., Cannavo, P., Schwartz, C., and Vidal-Beaudet, L.: Impact of city historical management on soil organic carbon stocks in Paris (France), J Soils Sediments, 21, 1038–1052, https://doi.org/10.1007/s11368-020-02869-9, 2021.

Costa, K. H. and Groffman, P. M.: Factors Regulating Net Methane Flux by Soils in Urban Forests And Grasslands, Soil Science Society of America Journal, 77, 850–855, https://doi.org/10.2136/sssaj2012.0268n, 2013.

Davidson, E. A. and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change, Nature, 440, 165–173, https://doi.org/10.1038/nature04514, 2006.

Decina, S. M., Hutyra, L. R., Gately, C. K., Getson, J. M., Reinmann, A. B., Short Gianotti, A. G., and Templer, P. H.: Soil respiration contributes substantially to urban carbon fluxes in the greater Boston area, Environmental Pollution, 212, 433–439, https://doi.org/10.1016/j.envpol.2016.01.012, 2016.

European Environment Agency. EUNIS habitat types hierarchical view - revised groups: https://eunis.eea.europa.eu/habitats-code-browser-revised.jsp?expand=#level_23466, last access: 26 September 2023.

Finnish Meteorological Institute: SmartMet Server, https://github.com/fmidev/smartmet-plugin-timeseries, last access: 30 March 2023a.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, 2013.

Groffman, P. M. and Pouyat, R. V.: Methane Uptake in Urban Forests and Lawns, Environ. Sci. Technol., 43, 5229–5235, https://doi.org/10.1021/es803720h, 2009.

Haaland, C. and van den Bosch, C. K.: Challenges and strategies for urban green-space planning in cities undergoing densification: A review, Urban Forestry & Urban Greening, 14, 760–771, https://doi.org/10.1016/j.ufug.2015.07.009, 2015.

Hedblom, M., Lindberg, F., Vogel, E., Wissman, J., and Ahrné, K.: Estimating urban lawn cover in space and time: Case studies in three Swedish cities, Urban Ecosyst, 20, 1109–1119, https://doi.org/10.1007/s11252-017-0658-1, 2017.

Ignatieva, M., Ahrné, K., Wissman, J., Eriksson, T., Tidåker, P., Hedblom, M., Kätterer, T., Marstorp, H., Berg, P., Eriksson, T., and Bengtsson, J.: Lawn as a cultural and ecological phenomenon: A conceptual framework for transdisciplinary research, Urban Forestry & Urban Greening, 14, 383–387, https://doi.org/10.1016/j.ufug.2015.04.003, 2015.

Ignatieva, M., Eriksson, F., Eriksson, T., Berg, P., and Hedblom, M.: The lawn as a social and cultural phenomenon in Sweden, Urban Forestry & Urban Greening, 21, 213–223, https://doi.org/10.1016/j.ufug.2016.12.006, 2017.

Ignatieva, M., Haase, D., Dushkova, D., and Haase, A.: Lawns in Cities: From a Globalised Urban Green Space Phenomenon to Sustainable Nature-Based Solutions, Land, 9, 73, https://doi.org/10.3390/land9030073, 2020.

Ivashchenko, K., Ananyeva, N., Vasenev, V., Sushko, S., Seleznyova, A., and Kudeyarov, V.: Microbial C-availability and organic matter decomposition in urban soils of megapolis depend on functional zoning, Soil & Environment, 38, 31–41, https://doi.org/10.25252/SE/19/61524, 2019.

Jasek-Kamińska, A., Zimnoch, M., Wachniew, P., and Różański, K.: Urban CO2 Budget: Spatial and Seasonal Variability of CO2 Emissions in Krakow, Poland, Atmosphere, 11, 629, https://doi.org/10.3390/atmos11060629, 2020.

Kaye, J. P., McCulley, R. L., and Burke, I. C.: Carbon fluxes, nitrogen cycling, and soil microbial communities in adjacent urban, native and agricultural ecosystems, Global Change Biology, 11, 575–587, https://doi.org/10.1111/j.1365-2486.2005.00921.x, 2005.

Kenward, M. G. and Roger, J. H.: Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, Biometrics, 53, 983–997, https://doi.org/10.2307/2533558, 1997.

Kong, L., Shi, Z., and Chu, L. M.: Carbon emission and sequestration of urban turfgrass systems in Hong Kong, Science of The Total Environment, 473–474, 132–138, https://doi.org/10.1016/j.scitotenv.2013.12.012, 2014.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B.: lmerTest Package: Tests in Linear Mixed Effects Models | Journal of Statistical Software, Journal of Statistical Software, 82, https://doi.org/10.18637/jss.v082.i13, 2017.

Lindén, L., Riikonen, A., Setälä, H., and Yli-Pelkonen, V.: Quantifying carbon stocks in urban parks under cold climate conditions, Urban Forestry & Urban Greening, 49, 126633, https://doi.org/10.1016/j.ufug.2020.126633, 2020.

Marshall, C. A. M., Wilkinson, M. T., Hadfield, P. M., Rogers, S. M., Shanklin, J. D., Eversham, B. C., Healey, R., Kranse, O. P., Preston, C. D., Coghill, S. J., McGonigle, K. L., Moggridge, G. D., Pilbeam, P. G., Marza, A. C., Szigecsan, D., Mitchell, J., Hicks, M. A., Wallis, S. M., Xu, Z., Toccaceli, F., McLennan, C. M., and Eves-van den Akker, S.: Urban wildflower meadow planting for biodiversity, climate and society: An evaluation at King's College, Cambridge, Ecological Solutions and Evidence, 4, e12243, https://doi.org/10.1002/2688-8319.12243, 2023.

Metzger, M. J., Shkaruba, A. D., Jongman, R. H. G., and Brunce, R. H. G.: Description of the European Environmental Zones and Strata, Wageningen, Alterra, 2012.

Monteith, J. L.: Evaporation and environment, Symposia of the Society for Experimental Biology, 19, 205–234, 1965.

Niemelä, J., Saarela, S.-R., Söderman, T., Kopperoinen, L., Yli-Pelkonen, V., Väre, S., and Kotze, D. J.: Using the ecosystem services approach for better planning and conservation of urban green spaces: a Finland case study, Biodivers Conserv, 19, 3225–3243, https://doi.org/10.1007/s10531-010-9888-8, 2010.

Paudel, S. and States, S. L.: Urban green spaces and sustainability: Exploring the ecosystem services and disservices of grassy lawns versus floral meadows, Urban Forestry & Urban Greening, 84, 127932, https://doi.org/10.1016/j.ufug.2023.127932, 2023.

Pouyat, R. V., Yesilonis, I. D., and Nowak, D. J.: Carbon Storage by Urban Soils in the United States, Journal of Environmental Quality, 35, 1566–1575, https://doi.org/10.2134/jeq2005.0215, 2006.

Selhorst, A. and Lal, R.: Net Carbon Sequestration Potential and Emissions in Home Lawn Turfgrasses of the United States, Environmental Management, 51, 198–208, https://doi.org/10.1007/s00267-012-9967-6, 2013.

Setälä, H. M., Francini, G., Allen, J. A., Hui, N., Jumpponen, A., and Kotze, D. J.: Vegetation Type and Age Drive Changes in Soil Properties, Nitrogen, and Carbon Sequestration in Urban Parks under Cold Climate, Frontiers in Ecology and Evolution, 4, 2016.

Shen, J., Peng, Z., and Wang, Y.: From GI, UGI to UAGI: Ecosystem service types and indicators of green infrastructure in response to ecological risks and human needs in global metropolitan areas, Cities, 134, 104176, https://doi.org/10.1016/j.cities.2022.104176, 2023.

Smith, L. S., Broyles, M. E. J., Larzleer, H. K., and Fellowes, M. D. E.: Adding ecological value to the urban lawnscape. Insect abundance and diversity in grass-free lawns, Biodivers Conserv, 24, 47–62, https://doi.org/10.1007/s10531-014-0788-1, 2015.

Sushko, S., Ananyeva, N., Ivashchenko, K., Vasenev, V., and Kudeyarov, V.: Soil CO2 emission, microbial biomass, and microbial respiration of woody and grassy areas in Moscow (Russia), J Soils Sediments, 19, 3217–3225, https://doi.org/10.1007/s11368-018-2151-8, 2019.

Tonteri, T. and Haila, Y.: Plants in a boreal city: Ecological characteristics of vegetation in Helsinki and its surroundings, southern Finland, Annales Botanici Fennici, 27, 337–352, 1990.

Unterweger, P. A., Schrode, N., and Betz, O.: Urban Nature: Perception and Acceptance of Alternative Green Space Management and the Change of Awareness after Provision of Environmental Information. A Chance for Biodiversity Protection, Urban Science, 1, 24, https://doi.org/10.3390/urbansci1030024, 2017.

Upadhyay, S., Singh, R., Verma, P., and Raghubanshi, A. S.: Spatio-temporal variability in soil CO2 efflux and regulatory physicochemical parameters from the tropical urban natural and anthropogenic land use classes, Journal of Environmental Management, 295, 113141, https://doi.org/10.1016/j.jenvman.2021.113141, 2021.

Vekuri, H.: Chamber-codes, https://github.com/hvekuri/Chamber-codes, 2023.

Walde, M., Allan, E., Cappelli, S. L., Didion-Gency, M., Gessler, A., Lehmann, M. M., Pichon, N. A., and Grossiord, C.: Both diversity and functional composition affect productivity and water use efficiency in experimental temperate grasslands, Journal of Ecology, 109, 3877–3891, https://doi.org/10.1111/1365-2745.13765, 2021.

Wang, R., Mattox, C. M., Phillips, C. L., and Kowalewski, A. R.: Carbon Sequestration in Turfgrass–Soil Systems, Plants, 11, 2478, https://doi.org/10.3390/plants11192478, 2022.

Zobec, M., Betz, O., and Unterweger, P. A.: Perception of Urban Green Areas Associated with Sociodemographic Affiliation, Structural Elements, and Acceptance Stripes, Urban Science, 4, 9, https://doi.org/10.3390/urbansci4010009, 2020.

Discarded reference:

Hundertmark, W. J., Lee, M., Smith, I. A., Bang, A. H. Y., Chen, V., Gately, C. K., Templer, P. H., and Hutyra, L. R.: Influence of landscape management practices on urban greenhouse gas budgets, Carbon Balance and Management, 16, 1, https://doi.org/10.1186/s13021-020-00160-5, 2021.