

Peer Review File

Key determinants of individuals' greenhouse gas emissions in Germany interplay and are mostly domain-specific



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality

17th May 23

Dear Mr Brandenstein,

Your manuscript titled "Predicting individual greenhouse gas emissions in everyday life: A machine learning approach" has now been seen by 2 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to

- address these criticisms (that is, either to incorporate the suggestions or provide a compelling argument why the point made by the reviewer is not valid, or relevant to the editorial threshold as outlined below)

AND

- meet our editorial thresholds as outlined below,

then we would be happy to look at a substantially revised manuscript.

In the following, we list our main editorial concerns.

****Editorial threshold 1: Provide comprehensive and novel insights into the internal and external drivers of individual greenhouse gas emissions in different domains of life, and discuss the significance and novelty of your results compared to the literature

****Editorial threshold 2: Demonstrate that using a holistic approach that takes into account the complex relationships between different drivers can help predict individual greenhouse gas emissions and target mitigation policies

****Editorial threshold 3: Ensure to fully describe and justify your methods, including choice of model predictors and limitation of your approach, and discuss potential changes in the effects of the drivers over time

If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter) and any completed checklist:

[Link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Niheer Dasandi, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-8708-837X

Aliénor Lavergne, PhD
Associate Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements (Download the link to your computer as a PDF.)

For your information, you can find some guidance regarding format requirements summarized on the following checklist:(<https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf>) and formatting guide (<https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf>).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This paper uses survey data and machine learning method along with basic data analysis to identify the factors that contribute to emissions. I have made detailed comments on the attached pdf. Due to insufficient novelty and ineffective presentation, I suggest rejecting this paper.

My detailed recommendations are in the attached, but I will briefly summarize the main points.

The literature review is not enough. Although three issues are clear, each point is not developed well. I would recommend that this is expanded a bit to describe the evidence a little more precisely. I don't agree some statements. For example, "ML models have not yet been applied to identify factors influencing climate-relevant behavior", to my knowledge there are some related papers as follows:

1. Shi, X., Wang, K., Cheong, T. S., & Zhang, H. (2020). Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. *Energy Economics*, 92, 104942.
2. Wang, A., Lam, J. C., Song, S., Li, V. O., & Guo, P. (2020). Can smart energy information interventions help householders save electricity? a svr machine learning approach. *Environmental Science & Policy*, 112, 381-393.
3. Exploring household emission patterns and driving factors in Japan using machine learning methods

The language and presentation is poor. The language used in the article is lacking in quality and the overall presentation is subpar. The Introduction should provide background information and clearly state the research question and any gaps in current knowledge. As it stands, the presentation is not sufficiently clear. Too many breaks and turns of phrase reduce the quality of this paper.

Furthermore, the structure of the article requires improvement. As a study investigating factors that affect a particular outcome, it is essential to clearly define and specify the considered factors in the methodology section. This crucial aspect is lacking in the current manuscript, resulting in a lack of clarity in the presentation of the research. If it were up to me, I would explore the impact of various factors from different domains, investigate the correlations between internal and external factors, and discuss the differences in the effects of these factors over time to analyze time effects. The current structure of the article lacks coherence and appears disorganized.

Reviewer #2 (Remarks to the Author):

This is an interesting article taking a new approach to calculating the GHG emissions of individuals using machine learning techniques. The authors focus especially on mobility and diet – two categories for which their statistical model predicted emissions with the highest confidence. Further, the authors chose to analyze factors that influence levels of individual GHG emissions by considering person-related (internal) and external factors.

Although the machine learning tools for predicting GHG emissions are powerful and with the application of this method authors do contribute to carbon footprint studies, the results for mobility are not very novel and the analyses and discussion lack depth. However, the analysis of diet in relation to people's values is an interesting addition to previous studies and it connects social psychology with carbon footprinting analysis in an interesting way.

To strengthen the article here are some more detailed comments:

\ In the Intro, I missed references to the newest IPCC report and especially the fifth chapter (Demand chapter). There is a whole section on sociocultural lifestyle changes. Another important reference point

is “1.5 -Degrees Lifestyles: Toward a fair consumption space for all” report. Please also check the article connecting affluence and environmental impacts: Scientists’ warning on affluence | Nature Communications

\ The mobility discussion – The predictors for mobility are not that novel. I do not know what exactly you checked when analyzing predictors of GHG mobility but it might be beneficial for your article and analysis to drill more in the results presented in the article. For example, retired with lower income individuals have lower GHG mobility emissions – how does it differ between car and air transport? Does location matter for their mobility (rural/urban), and how about their physical health? Do those with poorer health fly more? Please consider providing in the supplementary the results of the statistical analyses where you control for this type of variables while analyzing the predictors of GHG emissions. Maybe a table with the list of all variables considered in your regressions and the results of the models would clarify and summarize in a better way your analysis than the number of figures you have currently.

\ I think your discussion of results linked to mobility would benefit from reading “The political economy of car dependence: A systems of provision approach” by Giulio Mattioli et al (The political economy of car dependence: A systems of provision approach - ScienceDirect). You write about the attitudes and emotions around mobility choices but there is a lack of discussion of systematic lock-ins (which would fall under your external factors). Is it about liking to sit on the bus or is it about not being able to afford to live closer to your workplace, or not being able to afford collective transport, or having poor availability of collective transport? Considering the concept of needs escalation in relation to car and air transport might also enrich your discussion: please check Lina’s Brand Correa article: Full article: Understanding (and tackling) need satisfier escalation (tandfonline.com)

\ Considering car usage for work commute a complementary analysis of time use was done by Wiedenhofer. In his article, he writes that German households in dense urban areas with little car use tend to have more leisure travel by plane, which negates some of their emission reductions stemming from lower use of car. Please consider this reference for your discussion. Check it here: Household time use, carbon footprints, and urban form: a review of the potential contributions of everyday living to the 1.5 °C climate target - ScienceDirect

\ Some of the categories/words used are not clear and need more explanation. For example, in line 243 you write about the internal factor “perceived behavioral control”– from reading the main text of the article it is not clear to me what it means. I’m aware that you do not have that much place but maybe a table giving a short definition of internal and external factors you use, would be something that could be useful to the reader? I also recommend re-reading the article to search for terms that might be a bit difficult to understand. For example, in line 538 you write “private space emissions” – what kind of direct emissions exactly are these?

\ Statal analyses: the average explained variance of the GHG categories is still quite low (max 30%). It is important that GHG emissions are accurate for the majority of your sample, rather than the minority.

\ The drivers and barriers of impactful behaviour – the mobility: did you check for multicollinearity (for example calculating variance inflation factors) – I would expect car ownership to be closely related to mobility GHG emissions. In your mobility statistical analysis did you control for income or place of living? I didn’t see this analysis in the supplementary. From what I see you run a number of models but it would be great to see some results of them in the supplementary.

\ Discussion: In the section about diet, you provide some examples of policy intervention, which is great. I would like to see in addition a bit more in-depth reasoning for not advising to have sustainable food options in supermarkets. I understand that your analysis shows that this type of intervention are not helpful, but did you consider other factors (than what you used in your analysis) at play? I strongly encourage you to read Fuchs’s article, especially part four and maybe add a sentence about the role of power lens on consumption and absolute reduction in addition to how to change people’s habits, and if we embark on challenging the meat industry what is that we are against Power: the missing element in sustainable consumption and absolute reductions research and action -

ScienceDirect

\ I got confused by your sentence in lines 530-531 – how CCS or hydrogen are helping us meet the goal of 1,5 C degrees and emission reductions by 50% by 2030? Isn't it a narrative of climate delay to rely on future technologies that are not fully developed yet for emission reductions, while we already have available technologies that are feasible and sufficient for large decarbonization? I recommend IPCC report on Renewable Energy Sources and Climate Change Mitigation; and the Discourses of climate delay article for a reference to how techno-optimism can be a climate delay strategy: Discourses of climate delay | Global Sustainability | Cambridge Core.

Response Letter - Predicting individual greenhouse gas emissions in everyday life: A machine learning approach

Manuscript number COMMSENV-23-0367-T

Dear reviewers,

we sincerely thank you for your thoughtful comments on our manuscript and the possibility to resubmit a revised version for publication in Nature Communications Earth & Environment.

In the following we describe how we aimed to address all comments. We did our best to resolve all raised concerns and to integrate all the requested information into our manuscript while sticking to the journal guidelines. Due to the latter, we still report some of the requested information in the supplementary materials only but made sure that they can be easily found by adding cross references to the main text. For your convenience, the most relevant changes to the main text (including revised Figures and Tables) and supplementary materials are included in this response letter.

Yours sincerely,

The authors

Comments Reviewer #1

We addressed all major comments and concerns raised by reviewer #1 in the main text of our article. We resolved specific comments (in the pdf version of our initial submission) directly in the manuscript (e.g., unclear sentences) but explain and quote all major changes directly below.

Comment #1 (p. 3): What does "impactful" mean here? I think a lot of literature about driving factors would capture some significant factors. If you think previous studies just focus on less important behavior, how to claim your selected behavior is strongly related to individuals' emissions? How do you explain "systematically"? More described information needed.

Author response: Thank you for making us aware of this unfortunate formulation. We added a short explanation of what we mean by "impactful behavior".

"Beyond studying factors in isolation and neglecting their associations, previous study designs often did not capture impactful behavior, that is, behavior high in GHG emissions (cf. Kennedy et al., 2014; Moser & Kleinhüchelkotten, 2018). In detail, studies often investigated specific types of sustainable behavior that can be assessed rather easily but may only exert relatively low GHG emissions, such as self-reported recycling or water-saving behaviors. Notably, the determining factors of such low GHG behaviors can differ drastically from those influencing high GHG behaviors (Gifford et al., 2011; Huddart Kennedy et al., 2015), which may have led researchers and policymakers alike to focus on less important factors for climate change mitigation. On the other hand, studies that did measure high GHG behaviors (such as mobility and consumption patterns or energy use) often only measured parts of the respective behaviors (e.g., traveled distance by car but not accounting for car type or fuel consumption) which may also affect the obtained relative importance of factors. As a result, measuring sustainable behavior and related GHG emissions comprehensively and focusing on behaviors with high GHG impacts seems necessary to devise solutions effectively battling climate change (Gifford et al., 2011; Huddart Kennedy et al., 2015; Stern, 2011)." (p.4)

Comment #4 (p. 5): I recommend to show in a better structure, now the words are too small to read. And what is the innovation of this figure? You did the meta analysis from literature review? It is not informative enough to me.

Author response: We agree that the figure was too difficult to read in its previous form and lacked a more thorough description. We thus replaced the figure with a table which presents all relevant factors analyzed in the literature review. The factors illustrated in this table are also the ones we used in our models to predict domain-specific footprint. To adhere to the journals styling guidelines, we inserted the table in the Methods section (Table 1, p. 27-34) and deliver further information about the single predictors (e.g., item sources, scale reliability and previous findings) in the supplementary materials, section 2.

Comment #11 (p. 7): I don't think so. There are many related papers like
1. Shi, X., Wang, K., Cheong, T. S., & Zhang, H. (2020). Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. *Energy Economics*, 92, 104942.
2. Wang, A., Lam, J. C., Song, S., Li, V. O., & Guo, P. (2020). Can smart energy information interventions help householders save electricity? a svr machine learning approach. *Environmental Science & Policy*, 112, 381-393.

3. Exploring household emission patterns and driving factors in Japan using machine learning methods

Author response: We thank Reviewer 1 for bringing these papers to our attention. We carefully reviewed them and discussed their approach and the differences to our study in the revised Introduction

and Methods sections of our article. We further used the specific ML models used in the papers you mentioned to compare their prediction performance to our Random Forest model (see response to your next comment).

Introduction:

“Right now, only a few studies exist which at least satisfy some of the criteria outlined above to a certain degree. For instance, studies have analyzed the influence of multiple factors on mobility behavior (e.g., Hunecke et al., 2007), accounted for complex interplays of factors using ML models to identify high emission household (e.g., Chen et al., 2022), or measured behavior in GHG emissions more comprehensively (e.g., Kennedy et al., 2014). However, no previous work has yet fully met all these study requirements in a single analysis which would be necessary to comprehensively investigate drivers and barriers of sustainable behavior and derive effective demand-side solutions.” (p. 4f)

Methods:

“Despite their increasing popularity in environmental and social sciences, ML models have not yet been widely applied to identify factors influencing climate-relevant behavior. Although a few recent studies deployed ML models to analyze influencing factors of emissions, the used models either were not able to account for complex interactions between factors (Shi et al., 2020), focused on specific intervention effects on GHG emissions (Wang et al., 2020) or were used to identify overall household emission clusters (Chen et al., 2022). In our study, however, we aim to predict individual GHG emissions (in all relevant life domains) through a multitude of internal and external factors and account for their complex interplay. Doing so, we applied different ML models. First, we chose the popular Random Forest ML model (*RF*) due to its ability to approximate any input-output mapping function, ability to cope with small to medium-sized dataset and – on average – higher prediction accuracies on survey data compared to other ML or traditional models (Breiman, 2001; Kern et al., 2019; Liaw & Wiener, 2002). We also directly compared the RFs performance to two other popular ML models, a linear LASSO regression (*LASSO*) and non-linear support vector machines (*SVM*), which delivered promising results in related work on predicting GHG emissions mentioned earlier (Shi et al., 2020; Wang et al., 2020). We compared the ML models’ performance to a traditional linear ordinary least squares linear regression model (*LM*) which represents the current practice of predicting climate-relevant behavior. For detailed information about the used ML models and full model results, see supplementary section 4.1 and 5.1.” (p. 36f)

Comment #12 (p. 7): Why random forest? More evidence

Author response: In the revised manuscript, we now deliver more evidence for why we chose Random Forest and why we mostly report findings from this model. In short, previous research has shown that Random Forest are among the best performing models on tabular data, if the relationships between predictors and the outcome are complex and manifold (e.g., Kern et al., 2019). To make this choice more transparent, we also tested two alternative ML models (Lasso Regression and SVM) which are commonly used in the literature. We included the figure comparing the performance of all models to this response letter (Figure 1). The performance of the two other ML models on our dataset is subpar to the RF on average, supporting our choice of model. We added the overall results and argumentation for using the RF to the Methods and Results section of our manuscript as follows:

“First, we chose the popular Random Forest ML model (*RF*) due to its ability to approximate any input-output mapping function, ability to cope with small to medium-sized dataset and – on average – higher prediction accuracies on survey data compared to other ML or traditional models (Breiman, 2001; Kern et al., 2019; Liaw & Wiener, 2002). We also directly compared the RFs performance to two other popular ML models, a linear LASSO regression (*LASSO*) and non-linear support vector machines (*SVM*), which delivered promising results in related work on predicting GHG emissions mentioned earlier (Shi et al., 2020; Wang et al., 2020). We compared the ML models’ performance to a traditional

linear ordinary least squares linear regression model (LM) which represents the current practice of predicting climate-relevant behavior.” (p.36f)

“The results show that all ML models outperformed the standard LM. In detail, the RF exhibited the highest prediction performance of all models across domains (on average), followed by the SVM, LASSO and LM (see Figure 1 and supplementary section 5.1 for full model description and discussion).” (p.7)

“Since the RF performed the best among all models and prediction of GHG emissions worked best in domains mobility and diet, we focused on analyzing factor importance of the RF in these domains.” (p.10)

Supplementary section 5.1:

“We compared the performance of the Random Forest Regressor (RF), LASSO Regression (LASSO), Support Vector Machine (SVM) and the Linear Regression (LM). All ML models have also been applied to predict GHG emissions in previous studies and yielded good results (Shi et al., 2020; Wang et al., 2020). For more information about the individual models, see section 4.

The results are depicted in Figure 1 and Table 1. The model comparison across life domains indicates the RF to perform the best on average. Explained variance (R^2) was the highest and the overall prediction error (MAE) the lowest (except in domain consumption) across all domains in the RF. The SVM performed the second best and matched the prediction error performance of the RF in some domains (mobility, heating) and even surpassed it in one domain (consumption). However, the RF still outperformed the SVM in explained variance across domains and performed much better in the diet domain. Also, the SVM test set performance did not match the expected performance in the CV in some domains (consumption, electricity), indicating that the SVM may have overfitted the data. Resultingly, we chose the RF as our final ML model to report results on in the main article and to further investigate factor relationships. Further, the regularized LASSO alleviated some of the problems of the LM. The LASSO consistently showed lower prediction variance throughout the CV as well as slightly higher explained variance and lower prediction error on average compared to the LM (except mobility domain). This indicates that some of the factors did not contribute much to the prediction accuracy and their coefficients were shrunk towards zero in the LASSO compared to the LM. However, compared to more complex models like RF and SVM, the linear LASSO still performed much worse. This result further indicates that sustainable behavior is influenced by complex higher-level interactions of internal and external factors, which the SVM and RF can accommodate for, whereas the LASSO, by default, cannot.” (p.53)

Comment #20 (p. 14): I try to understand this long paragraph, but I am confused by many digressions and punctuations. There are factors related to mobility being expressed here, and internal and external factors may have some correlation. However, I did not observe clear arguments or sufficient evidence. The entire article shows an unclear structure and a confused expression, which is not academic and scientific enough. I recommend to restructure it and compare more ML models based on the prediction performance. I think the survey data is valuable, to find a more innovative question and improve the academic language.

Author response: We restructured the corresponding paragraph and – in line with your suggestion – ran model comparisons between multiple ML models (see previous comments). However, the main focus of the article is not to work out differences in ML model performances but to show that using ML models in general is an essential point when analyzing driving factors of individual GHG emissions due to theoretical considerations and our empirical results. This has not been addressed in previous studies and – resultingly – the relative importance of factors driving sustainable behavior is still unclear which,

in turn, can lead to ineffective mitigation strategies. We did our best to make this research aim clear in our revised Introduction and throughout the manuscript. As requested and described above, we compared different candidate models and the Random Forest outperformed all other models. Thus, the main findings of the Random Forest model seem central to our research aim.

Comment #21 (p. 17): The factors should be introduced in advance. They are not the same as in the figure1.

Author response: We thank Reviewer 1 for pointing us to these inconsistencies. In the revised manuscript, we now aligned the factor names throughout the Methods and Results sections.

As Figure 1 was difficult to read, we got rid of it and now report all the relevant information about the factors in Table 1. This table (which we included in the Methods section), contains a summarized version of all factors/predictors identified in the literature review and used in our models. We included this table at the end of this response letter for your convenience. However, since a full description of the factors (including previous study results, scales and wordings typically reported in the Methods section of an article) would be too extensive, we still cross-reference readers to the supplementary materials for more details.

Results

“For our survey, we first identified the relevant factors of individual sustainable behavior in a comprehensive literature review and selected factors which are grounded in established theories on sustainable behavior as well as those which were identified as relevant in meta analyses and large scale studies. In general, the identified drivers and barriers of sustainable behavior can be categorized into internal (i.e., person-related) and external (i.e., situation-related) factors (Kollmuss & Agyeman, 2002). Internal factors mainly include psychological factors such as individual beliefs, attitudes, values, and intentions. External factors subsume political, social, economic, and cultural conditions people find themselves in. A detailed description of the literature review and all factors included in our analysis can be found in the Methods section and supplementary sections 1 and 2.” (p. 6)

Methods

“In general, the identified drivers and barriers of sustainable behavior can be categorized into internal (i.e., person-related) and external (i.e., situation-related) factors (Kollmuss & Agyeman, 2002). Internal factors consisted of constructs derived from the Theory of Planned Behavior, Value-Belief Norm Theory, and habit formation. These include attitudes, (personal) norms and values, perceived behavioral control, behavioral intentions, climate change awareness and personal responsibility beyond others. We further considered environmental knowledge, emotions towards sustainable behaviors, political attitudes and voting intention as well as media consumption and demographics as internal factors. External factors subsume political, social, economic, and cultural conditions people find themselves in. They were highly domain-specific and included items like accessibility to information and feedback about energy and heating behavior, situational possibilities of eating sustainably or infrastructural mobility circumstances (e.g., access to public transportation).” (p. 26)

Summarized feedback of Reviewer #1

First comment: This paper uses survey data and machine learning method along with basic data analysis to identify the factors that contribute to emissions. I have made detailed comments on the attached pdf. Due to insufficient novelty and ineffective presentation, I suggest rejecting this paper.

My detailed recommendations are in the attached, but I will briefly summarize the main points.

The literature review is not enough. Although three issues are clear, each point is not developed well. I would recommend that this is expanded a bit to describe the evidence a little more precisely. I don't agree some statements. For example, "ML models have not yet been applied to identify factors influencing climate-relevant behavior", to my knowledge there are some related papers as follows:

1. Shi, X., Wang, K., Cheong, T. S., & Zhang, H. (2020). Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. *Energy Economics*, 92, 104942.
2. Wang, A., Lam, J. C., Song, S., Li, V. O., & Guo, P. (2020). Can smart energy information interventions help householders save electricity? a svr machine learning approach. *Environmental Science & Policy*, 112, 381-393.
3. Exploring household emission patterns and driving factors in Japan using machine learning methods

Author response: Thanks for this detailed suggestion. We elaborated more on our three main points in the introduction of the manuscript and worked out differences between the studies you provided and ours. Please find the changes on pages 2-4.

Second comment: The language and presentation is poor. The language used in the article is lacking in quality and the overall presentation is subpar. The Introduction should provide background information and clearly state the research question and any gaps in current knowledge. As it stands, the presentation is not sufficiently clear. Too many breaks and turns of phrase reduce the quality of this paper.

Author response: We carefully proofread the article, corrected typos and smoothed the language. Further, we revised the Introduction section to clearly state the scientific value of our work (see p. 2 – 5).

Third comment: Furthermore, the structure of the article requires improvement. As a study investigating factors that affect a particular outcome, it is essential to clearly define and specify the considered factors in the methodology section. This crucial aspect is lacking in the current manuscript, resulting in a lack of clarity in the presentation of the research. If it were up to me, I would explore the impact of various factors from different domains, investigate the correlations between internal and external factors, and discuss the differences in the effects of these factors over time to analyze time effects. The current structure of the article lacks coherence and appears disorganized.

Author response: In line with your suggestion and as described in our response to your previous comments, we went into more detail about the investigated factors in the Methods section, included a tabular overview of all factors used in the models (Table 1), including their description and exemplar items. We also cross-referenced readers for further details about measurement specifics, original source and codebook to the supplementary materials (section 2). Since we don't have time series data at the moment, we unfortunately cannot observe effects over time. However, this is an important point and we addressed it in more detail in our Discussion section. Further, we included more details about factor associations/correlations in the manuscript as follows:

"We first performed a systematic literature review to identify the relevant factors of individual sustainable behavior. We selected those factors which are grounded in established theories on sustainable behavior as well as those which were relevant in meta analyses, large scale studies and

climate change mitigation reports. Since the final list of identified factors was extensive, we provide a summarized overview in the following (see Table 1). A more in-depth description of the literature review, identified factors and previous research results can be found in the supplementary sections 1 and 2.

In general, the identified drivers and barriers of sustainable behavior can be categorized into internal (i.e., person-related) and external (i.e., situation-related) factors (Kollmuss & Agyeman, 2002). Internal factors consisted of constructs derived from the Theory of Planned Behavior, Value-Belief Norm Theory, and habit formation. These include attitudes, (personal) norms and values, perceived behavioral control, behavioral intentions, climate change awareness and personal responsibility beyond others. We further considered environmental knowledge, emotions towards sustainable behaviors, political attitudes and voting intention as well as media consumption and demographics as internal factors. External factors subsume political, social, economic, and cultural conditions people find themselves in. They were highly domain-specific and included items like accessibility to information and feedback about energy and heating behavior, situational possibilities of eating sustainably or infrastructural mobility circumstances (e.g., access to public transportation).” (p.26)

“Notably, average footprints and the observed drivers and barriers of sustainable behavior might not be the same everywhere across the globe (e.g., Noll et al., 2022) and thus the relative impact of regulatory measures might also vary between countries as well as over time (Ivanova et al., 2016; Ivanova et al., 2017; Kennedy et al., 2014; Markandya et al., 2009). Therefore, future research should extend the proposed design to non-Western countries and also observe changes in behavior and their drivers and barriers over time, another knowledge gap identified in the recent IPCC report (Creutzig et al., 2022). Being able to flexibly adjust regulative measures in different transition phases and contexts based on knowledge about the respective factors and their interaction could speed up climate change mitigation drastically (Creutzig et al., 2022).” (p. 24)

“Due to our overall study aim of identifying the most important factors driving and hindering sustainable behavior, we did not follow up on all potential interactions or relationships between predictors implicitly captured by our ML models. However, since there are likely to be more important interactions that could lead to even more tailored regulative strategies, future studies could use our data set and test for specific interaction effects.” (p.24)

Comments Reviewer #2

Comment 1: In the Intro, I missed references to the newest IPCC report and especially the fifth chapter (Demand chapter). There is a whole section on sociocultural lifestyle changes. Another important reference point is “1.5 -Degrees Lifestyles: Toward a fair consumption space for all” report. Please also check the article connecting affluence and environmental impacts: Scientists’ warning on affluence.

Author response: We thank Reviewer 2 for bringing the IPCC report chapter and the additional articles to our attention. We revised our Introduction as well as our Discussion and discussed our findings in light of the respective articles. We also reframed strategies to battle climate change in terms of supply-side vs. demand-side solutions.

“Mitigation strategies are often categorized as *supply-side* and *demand-side* solutions. While the more traditional supply-side solutions focus on technological advancements to, for instance, decarbonize supply chains and improve energy efficiency of appliances, demand-side solutions directly address individual behavior and needs, such as promoting reduced car usage or energy consumption (Creutzig et al., 2018; Wiedmann et al., 2020). Supply-side solutions are traditionally considered as crucial for reaching global climate goals; but there is also a growing acknowledgment of the importance of demand-side solutions in tackling climate change from various angles. This shift mirrors the understanding that relying solely on potential technological advancements on the supply-side is unlikely to effectively accomplish global climate and sustainability objectives in due time (Wiedmann et al., 2020).” (p.2)

“To date, studies have recognized a wide range of individual and situational factors that are relevant to individual behavior in various life domains such as the role of social norms, values and structural barriers for electricity consumption or mobility behavior (Clayton et al., 2015; Creutzig et al., 2022). While identifying singular factors can advance our general knowledge about existing drivers of and barriers to sustainable behavior, the intricacies of their relationships and their absolute and relative strength of influence on sustainable behavior is still widely unknown. This issue has also been raised in the most recent IPCC report: Although a lot of evidence linking singular factors to individual sustainable behavior exists, there still is a high need for investigating and understanding the mutual interactions of factors and their relative importance for individual behavioral change (Creutzig et al., 2022, pp. 570-571).” (p. 2f)

“Employed and wealthier people exhibited much higher mobility related GHG emissions than people who are retired and/or have lower incomes (supplementary section 5.2.1, Figure 1). This finding is consistent with previous studies investigating impactful behavior and the environmental impact of affluent citizens (Huddart Kennedy et al., 2015; Hunecke et al., 2007; Kennedy et al., 2014; Wiedenhofer et al., 2018; Wiedmann et al., 2020).” (p.10)

Comment 2: The mobility discussion – The predictors for mobility are not that novel. I do not know what exactly you checked when analyzing predictors of GHG mobility but it might be beneficial for your article and analysis to drill more in the results presented in the article. For example, retired with lower income individuals have lower GHG mobility emissions – how does it differ between car and air transport? Does location matter for their mobility (rural/urban), and how about their physical health? Do those with poorer health fly more? Please consider providing in the supplementary the results of the statistical analyses where you control for this type of variables while analyzing the predictors of GHG emissions. Maybe a table with the list of all variables considered in your regressions and the results of the models would clarify and summarize in a better way your analysis than the number of figures you have currently.

Author response: We agree that testing specific hypotheses about predictor/outcome associations (as the ones you brought up) are important to advance in our understanding of what drives and hinders sustainable behavior. This approach has often been used in previous studies on sustainable behavior, as

argued in our introduction. However, the ML modeling approach which is at the core of our present work is not intended for such typical hypothesis testing. Nevertheless, the ML approach has some advantages we make use of in the present study. It can deal with the multitude of potential relationships and interactions between internal and external factors with sustainable behavior autonomously. Additionally, this approach can inform us about the singular relative importance of factors and their interactive importance which we consider important information when it comes to decision making regarding regulative strategies. Of course, the data set we generated also allows to test for more specific relationships between the factors we included and we plan to test them in future work.

To address your points in the manuscript, we rewrote parts of the Introduction and Methods section and discussed the importance of analyzing singular relationships between variables as the ones you suggested. We also really liked your suggestion of including a table with all predictors used in our models so that readers can see which factors we “controlled for”. You can find it at the end of this response letter.

“Due to our overall study aim of identifying the most important factors driving and hindering sustainable behavior, we did not follow up on all potential interactions or relationships between predictors implicitly captured by our ML models. However, since there are likely to be more important interactions that could lead to even more tailored regulative strategies, future studies could use our data set and test for specific interaction effects.” (p.24)

Comment 3: I think your discussion of results linked to mobility would benefit from reading “The political economy of car dependence: A systems of provision approach” by Giulio Mattioli et al (The political economy of car dependence: A systems of provision approach - ScienceDirect). You write about the attitudes and emotions around mobility choices but there is a lack of discussion of systematic lock-ins (which would fall under your external factors). Is it about liking to sit on the bus or is it about not being able to afford to live closer to your workplace, or not being able to afford collective transport, or having poor availability of collective transport? Considering the concept of needs escalation in relation to car and air transport might also enrich your discussion: please check Lina’s Brand Correa article: Full article: Understanding (and tackling) need satisfier escalation (tandfonline.com)

Author response: We agree that these are very important considerations. Our results suggest that it is a combination of both, that is, having poor availability and/or insufficient ease of use as well as holding positive attitudes and emotions among other psychological variables. However, Reviewer 2 is right that we did not extensively discuss the role of systematic lock-ins previously. In the revised manuscript we now go into more detail and also included references to the mentioned articles as follows:

“These findings are in line with previous studies on external factors of mobility behavior (Abou-Zeid et al., 2012; De Witte et al., 2008) and also point towards so-called “lock-in effects”: The situational and infrastructural circumstances of individuals’ living sectors may lock them in to certain behaviors (i.e., car usage) to fulfill their needs, like going to work or grocery shopping, which makes direct behavioral interventions less effective (Akenji et al., 2021; Brand-Correa et al., 2020; Mattioli et al., 2020). These circumstances can only be improved by investing in reliable, fast, and ample public transportation infrastructure and changes in city designs and land use (Mattioli et al., 2020; Wiedenhofer et al., 2018). Unlocking travel mode choices and satisfying individual needs through targeting mobility circumstances is likely to lead to changes in mobility behavior and may also reduce the strong effect of income on mobility GHG emissions found in our analysis and in previous studies (Creutzig et al., 2022; Wiedenhofer et al., 2018). However, when implementing such measures, policy makers must be aware of “rebound effects” (i.e., the saved time and money is spent on other consumables or increased leisure air travel) that may negate some of the GHG savings (Wiedenhofer et al., 2018). An alternative approach to reduce the particularly strong effect of income on mobility related GHG emissions may also include more general socio-ecological transformations, such as reducing contracted working times and discussing ways of economic de-growth while still maintaining (or even improving) well-being.

However, the wider implications of such measures have been debated (Wiedenhofer et al., 2018).” (p. 21)

Comment 4: Considering car usage for work commute a complementary analysis of time use was done by Wiedenhofer. In his article, he writes that German households in dense urban areas with little car use tend to have more leisure travel by plane, which negates some of their emission reductions stemming from lower use of car. Please consider this reference for your discussion. Check it here: Household time use, carbon footprints, and urban form: a review of the potential contributions of everyday living to the 1.5 °C climate target – ScienceDirect

Author response: We thank reviewer 2 for bringing this work to our attention. We discussed emissions reduction strategies in the mobility domain based on our findings and included potential backfire effects mentioned in the article (see response to your previous comment).

Comment 5: Some of the categories/words used are not clear and need more explanation. For example, in line 243 you write about the internal factor “perceived behavioral control”– from reading the main text of the article it is not clear to me what it means. I’m aware that you do not have that much place but maybe a table giving a short definition of internal and external factors you use, would be something that could be useful to the reader? I also recommend re-reading the article to search for terms that might be a bit difficult to understand. For example, in line 538 you write “private space emissions” – what kind of direct emissions exactly are these?

Author response: We agree that some of the constructs/factors we introduced were not sufficiently explained in the first submission. Consistent with your suggestion, we included a short description of each factor in our models in Table 1 included in the Methods section of our manuscript. We also included this table at the end of the response letter.

We also went through the entire manuscript and added explanations of terms that might not be familiar to readers. In the specific case you mentioned, we rephrased the sentence as follows:

“Further, whereas the present study investigated individual behavior, emissions directly caused by individuals in their daily life only constitute part of a population’s overall GHG emissions, around 18% of total GHG emissions in Germany (Umweltbundesamt, 2022a), which seems representative of many Western civilizations (European Environment Agency, 2019; Ivanova et al., 2017; United States Environmental Protection Agency, 2020). Therefore, mitigation efforts must be extended to infrastructural, agricultural, and industrial sectors which emit the remaining part of total GHGs.” (p.24f)

Comment 6: Statical analyses: the average explained variance of the GHG categories is still quite low (max 30%). It is important that GHG emissions are accurate for the majority of your sample, rather than the minority.

Author response: In general, the explained variances in our models are comparable or even surpass results reported in previous studies (domain mobility and diet). Notably, we only focus on explained variance on out-of-sample data which is – by design – almost always lower than the explained variance on the dataset the models are fitted on due to the variance-bias tradeoff. However, performance on out-of-sample data presents a more realistic estimate of how we could expect models to perform in the population (see p. 37). The variance still unexplained by our models may be attributable to multiple sources such as measurement errors, potentially missing factors and accuracy of footprint estimation.

We discussed these points and the overall performance of our models in our discussion and the supplementary materials in more detail:

“Notably, the absolute prediction performance (i.e., explained variance) of our models in the specific domains is difficult to compare with existing literature due to differences in methodologies. In our approach, we only focus on prediction performance on out-of-sample data (vs. in-sample) and measured behavior in GHG emissions (vs. scales). However, our prediction results still come close or even surpass results reported in previous work which come closest to our approach (e.g., Gatersleben et al., 2002; Hunecke et al., 2007). Variance still unexplained by our models may be attributable to multiple sources such as measurement errors, missing factors in the models and the accuracy of footprint estimations.” (supplementary section 5.1, p. 54)

Comment 7: The drivers and barriers of impactful behaviour – the mobility: did you check for multicollinearity (for example calculating variance inflation factors) – I would expect car ownership to be closely related to mobility GHG emissions. In your mobility statistical analysis did you control for income or place of living? I didn’t see this analysis in the supplementary. From what I see you run a number of models but it would be great to see some results of them in the supplementary.

Author response: In general, multicollinearity is not a problem in ensemble methods like Random Forest due to their inherent structure. As we described in the supplementary materials (section 4.1), Random Forest are comprised of single decision trees, each getting a random subsample of predictors and a bootstrap sample of cases in the dataset. As a result, single trees are decorrelated. Although this procedure does not entirely solve the problem of highly correlated predictors, the overall ordering (i.e., relative importance) of variables is not affected by high correlations between predictors, which is the most important criteria for our study purpose. If you want to know more about the influence of multicollinearity in RFs and an experiment to support our claim, please check:

Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot. Variable selection using Random Forests. *Pattern Recognition Letters*, Elsevier, 2010, 31 (14), pp.2225-2236.

We included a short note about this in the supplementary, where we describe the RFs working principles:

“For every decision tree in the RF, a random subsample of predictors as well as bootstrap subsample of samples is drawn. Consequently, the single trees are decorrelated and the overall RF is regularized, yielding models with less variance. This procedure also makes RFs more robust to highly correlated predictors in the model so that the overall ordering (i.e., relative importance) of predictors is not affected, if multicollinearity in the data exists (Genuer et al., 2010).” (section 4.1., p. 46f)

Regarding your other points: Yes, we did include income and place of living in our models. As some information got easily lost in the first submission, we now display all predictors used in our models in the new Table 1 (Methods and end of this response letter). We also looked at the relationships between these variables and behavior in more detail, see supplementary section 5.2.1, Figures 1 and 7. In general, the results we report in the supplementary are more in-depth analyses of our RF model (i.e., dependence plots of the calculated SHAP values which represent the contribution of each predictor to the overall model prediction). We did this to dive into the relationships of some interesting singular factors with behavior. However, as mentioned in the main article and in our answer to your second comment, investigating all potential singular relationships and/or interactions would be too extensive and is beyond the scope of our article.

Comment 8: Discussion: In the section about diet, you provide some examples of policy intervention, which is great. I would like to see in addition a bit more in-depth reasoning for not advising to have sustainable food options in supermarkets. I understand that your analysis shows that this type of intervention are not helpful, but did you consider other factors (than what you used in your analysis) at

play? I strongly encourage you to read Fuchs's article, especially part four and maybe add a sentence about the role of power lens on consumption and absolute reduction in addition to how to change people's habits, and if we embark on challenging the meat industry what is that we are against Power: the missing element in sustainable consumption and absolute reductions research and action – ScienceDirect

Author response: Thank you for the suggestion of factors to discuss regarding the diet domain. We would not advise directly against having sustainable food options in supermarkets but rather argue that increased availability is unlikely to help people shift their dietary patterns and therefore reduce diet-related GHG emissions, as informed by our results. All the factors we considered in the diet domain are listed in Table 1 included in the revised Methods section. You can also find it at the end of the response letter for your convenience. The article you mentioned is a great addition to our discussion of driving factors and potential mitigation strategies. We added a sentence about how to approach reduced meat consumption on a societal scale alongside our suggestions of potential mitigation strategies:

“In contrast to previous suggestions, factors like availability of sustainable food options in supermarkets, their identifiability or individual income do not seem to contribute much to dietary choices and thus may not be promising targets for encouraging GHG-emission-friendly diets. Crucially, however, this does not imply that targeting availability and prices of conventional food options (meat products in particular) would also be inefficient in reducing GHG emissions. Quite the contrary, implementing policies that shift financial and structural power from conventional food lobbies and the meat industry to the organic sector may substantially complement the effectiveness of campaigns promoting dietary shifts (Fuchs et al., 2016).” (p. 22f)

Comment 9: I got confused by your sentence in lines 530-531 – how CCS or hydrogen are helping us meet the goal of 1,5 C degrees and emission reductions by 50% by 2030? Isn't it a narrative of climate delay to rely on future technologies that are not fully developed yet for emission reductions, while we already have available technologies that are feasible and sufficient for large decarbonization? I recommend IPCC report on Renewable Energy Sources and Climate Change Mitigation; and the Discourses of climate delay article for a reference to how techno-optimism can be a climate delay strategy: Discourses of climate delay | Global Sustainability | Cambridge Core.

Author response: We agree if we would follow suit with mitigation strategies we already know to help us reach our GHG targets, CCS and hydrogen would not be required to the extent we gave it credit for. We are certainly not experts in these technological solutions and the topic is much more complex then presented here. To abide to your suggestion, we deleted the corresponding paragraph and focused more on supply-side and demand-side solutions to dwelling emissions which are discussed in current literature.

“However, [supply-side solutions like] large-scale decarbonization of the electrical grid as well as building and adaptation of more energy efficient homes may not be enough to meet global climate goals. Within the demand-side framework, researchers have advocated for implementing policies that more directly target resource-intensive living standards. As mentioned previously, the size of dwellings and their type (single family vs. multifamily housing) significantly contribute to shelter-related GHGs. Thus, promoting broader societal changes by directly addressing city planning and individual lifestyles, for instance by supporting compact city designs and multifamily housings as well as incentivizing individuals to reduce their living space while still maintaining personal well-being, might also be necessary (Akenji et al., 2021; Creutzig et al., 2022; Wiedmann et al., 2020).“ (p. 23)

Table 1. Overview of previously investigated factors (predictors) in the models to predict sustainable behavior.

INTERNAL FACTORS		EXTERNAL FACTORS	
Name of predictor (factor type)	Description [number and type of items, example item]	Name of predictor (factor type) [domain]	Description [number and type of items, example item]
Behavioral Beliefs (psychological), TPB	Attitudes towards sustainable behavior [mean score of 2 items, numerical "From my point of view, saving electricity is unimportant"]	Evaluation of state financial subsidies for (energetic) renovation (political) [<i>Shelter: Heating</i>]	Satisfaction, transparency, accessibility of financial subsidies for energetical renovation state [mean score of 3 items, numerical "In 2019 the financial subsidies of the state for an (energetic) renovation and installation of more efficient heating sources were insufficient"]
Personal Norms (psychological), TPB	Opinions and habits of peers towards sustainable behavior [mean score of 2 items, numerical "Most of the people who are close to me keep their consumption behavior and spending (e.g. on clothing, everyday objects) as low as possible."]	Evaluation of financial subsidies of city / municipality for renovation (political) [<i>Shelter: Heating</i>]	Satisfaction, transparency, accessibility of financial subsidies for energetical renovation municipality [mean score of 3 items, numerical "In 2019 the financial subsidies of my city / municipality for an (energetic) renovation and installation of more efficient heating sources were insufficient"]
Perceived Control (psychological), TPB	Perceived behavioral control over sustainable behavior [mean score of 2 items, numerical "For me, reducing my consumer behavior and my consumption expenditure (e.g. on clothing, everyday objects) is hard"]	Evaluation of information about (energetic) renovation (political) [<i>Shelter: Heating</i>]	Satisfaction, transparency, accessibility of information regarding energetical renovation [mean of 3 items, numerical "Information for an (energetic) renovation and installation of more efficient heating sources are satisfactory"]
Behavioral Intention (psychological), TPB	Future intention to behave sustainably [1 item, numerical, "When you think about your future behavior (regardless of current pandemic-related limitations), how much do you agree with the	Evaluation of information about sustainable heating behavior (political) [<i>Shelter: Heating</i>]	Accessibility and transparency of ecological heating behaviour Information in general [mean score of 2 items, numerical "Information on environmentally friendly heating behavior at home

	following statements? - I intend going without a car and using alternative means of transport (bicycle, public transport, etc.)."] General human values, Schwartz scales <i>Subscales:</i> Self-Direction, Power, Universalism, Achievement, Security, Stimulation, Conformity/Tradition, Hedonism, Benevolence [all subscales: mean value of 3 items, numerical]		(e.g. setting the water temperature, ventilation behavior, thermostat regulation) are:"] Information about electricity savings behavior municipality [1 item, categorical In 2019, my municipality / city offered free advice on energy efficiency, energetic renovation and / or energy saving.]
Values (psychological), VBN		Energy advice municipality (political) [<i>Shelter:</i> <i>Electricity</i>]	
NEP Growth/Technology (psychological), VBN	New ecological paradigm: Growth, Harmony and Technology [sum score of 3 items, numerical "In order to survive, people have to live in harmony with nature."]	Employer support of alternative mobility options (situational/infrastructural) [<i>Mobility</i>]	Financial and situational support of using alternative means of transportation at work [1 item, categorical "In 2019, my employer promoted the use of alternative mobility options (e.g. the possibility of a work bike, more vacation days without a flight, job tickets)."]
Awareness of Consequences (psychological), VBN	Awareness of consequences of climate change [4 items, numerical "The consequences of climate change are being dramatized."]	Homeoffice (situational/infrastructural) [<i>Shelter, Mobility</i>]	Possibility and use of Homeoffice [2 items, binary coded "How many days a week did you work from home on average in 2019 (remote work / home office)?"
Ascription of Responsibility (psychological), VBN	Acknowledgement of responsibility towards the environment/climate [sum score of 2 items, numerical "Climate change is mainly caused by humans."]	Tenant/owner of house/flat (situational/infrastructural) [<i>Shelter:</i> <i>Electricity</i>]	Ownership structure of housing [1 item, categorical "Were you tenant or owner of your apartment / house in 2019?"
Perceived Consumer Effectiveness (psychological), VBN	Perceived Effectiveness of one's own behavior for environment/climate change [mean score of 2 items, numerical "I can't do much to protect the environment on my own."]	Involved in care of relatives (situational/infrastructural) [<i>Mobility</i>]	[1 item, binary coded "At the end of 2019, did you have relatives in need of care outside of your household that you looked after?"

Personal Normative Beliefs (psychological), VBN	Personal normative beliefs on climate obligations [sum score of 2 items, numerical "I feel obliged to do everything in my power to stop climate change."]	Possibility to renovate apartment energetically (situational/infrastructural) [<i>Shelter: Heating</i>]	[1 item, categorical "If I wanted, I could ensure that my (rented) apartment / house would be energetically renovated and / or a more efficient heating source would be installed."]
Habits (psychological)	Habit of behaving sustainably [1 item, numerical "Saving electricity is something I do automatically."]	Flight status (situational/infrastructural) [<i>Mobility</i>]	[1 item, categorical "Did you fly at least once in the past three years (2018 to 2021)?"]
Emotions (psychological)	Emotions towards sustainable behavior (happiness and satisfaction): [mean score of 2 items, numerical "How do you think you would feel if you were to eat sustainably in the near future?"]	Car possession (situational/infrastructural) [<i>Mobility</i>]	Car possession [1 item, categorical "Do you own a car?"]
Knowledge (psychological)	Knowledge about determinants/processes of climate change[sum of correct answers to 9 items, multiple choice format"Compared to the corresponding amount of vegetables in terms of calories, beef is... harmful for the environment"]	Satisfaction alternative mobility options (situational/infrastructural)[<i>Mobility</i>]	Satisfaction with infrastructure using different mobility types (bike, public transport, walking)[mean score 3 items, numerical "How satisfied were you in 2019 with the connection, travel time and frequency (for public transport) as well as the transport infrastructure of the following mobility options in your residential area for typical destinations such as your work, supermarket or leisure"]
Political efficacy (psychological)	Internal and external political efficacy mean scores of subscales: <i>Internal</i> : [4 items, numerical] <i>External</i> : [4 items, numerical] ["I can understand and assess important political questions very well.", "The politicians strive to	Availability of special mobility options (situational/infrastructural) [<i>Mobility</i>]	Availability of car sharing, cycle paths, park & ride infrastructure, public transport discounts and other special mobility options [sum score of 5 items, categorical "Were there the following mobility options in your area in 2019? - Public transport offers (e.g. for professional groups, family cards, etc.)"]

	maintain close contact with the population."]		
Political trust (psychological)	Trust in different official institutions [mean score of 9 items, numerical "Please indicate below to what extent you trust the respective institutions or groups of people. - German federal government"]	Satisfaction special mobility (situational/infrastructural) [<i>Mobility</i>]	Satisfaction with special mobility options, such as car sharing, cycle paths, park & ride infrastructure, public transport discounts and other special mobility options [mean score of 5 items, numerical "How satisfied were you with the following mobility option (s) near you in 2019? - infrastructure for "park and ride""]
Big Five (psychological)	Big Five personality factors sum scores of Subscales: <i>Conscientiousness</i> [3 items, numerical] <i>Openness to Experience</i> [3 items, numerical] <i>Extraversion</i> [3 items, numerical] <i>Agreeableness</i> [3 items, numerical] <i>Neuroticism</i> [3 items, numerical] ["I am someone who..."	Development of e-Mobility Infrastructure (situational/infrastructural) [<i>Mobility</i>]	Development of charging infrastructure and conditions for e-Mobility [mean score of 2 items, numerical "How good were the infrastructural conditions in 2019 at your typical everyday destinations for electric cars (e.g. free parking spaces, extra lanes / parking spaces for electric cars)?"
Voting Intention: Party (psychological)	Voting intention in the next German federal election [1 item, categorical "If there was a federal election next Sunday, which party would you vote for with your ZWEITSTIMME?"	Distance to nearest public transport (situational/infrastructural) [<i>Mobility</i>]	Distance to nearest bus and train station [2 items, integer number (Distance in km)]
Ideology (psychological)	Self-Placement in one-dimensional political Spectrum [1 item, numerical "Many people use the terms "left" and "right" to denote different political attitudes. We have a yardstick here that runs from left to right. When	Added travel time alternative mobility (situational/infrastructural) [<i>Mobility</i>]	Change in time spent commuting to when using alternative transportation [3 items, recoded to one item "How much more/less time did you have to spend on average if you had used alternative means of transport (e.g. public transport, bicycle, etc.)

	you think of your own political views, where would you place those views on this scale? Please tick the appropriate box."]		instead of a car to get to work?""]
Political participation (psychological)	Political participation regarding environmental issues <i>Sustainability as Topic of Conversation:</i> [1 item, categorical] "There are various ways in which one can try to improve something in Germany or to prevent something from getting worse. Have you done any of this over the past 12 months? - Discussed the topic of "sustainability" with friends / family / other people"] <i>Member of an Environmental Organisation:</i> [1 item, categorical] "Some people are members of different groups or associations. Please indicate whether you are a member of the following groups yourself and how you are involved. - environmental organization"]	Care of relatives (situational/infrastructure) [Mobility]	Regular care work for relatives outside of own house [1 item, binary] "At the end of 2019, did you have relatives in need of care outside of your household that you looked after?"
Conspiracy belief: Climate Change (psychological)	Conspirational thinking about human-cause climate change [1 item, numerical (4 point scale)] "Man-made climate change is an excuse to patronize citizens or to tax them."]	Eating out regularly (situational/infrastructure) [Diet]	Regular visits to canteens/restaurants [1 item, binary] "Did you eat out regularly (at least once a week) in 2019 (e.g. canteen / restaurant at work)?""]

Media consumption (psychological)	Number of regularly consumed (non-) mainstream media formats: Frequency of Use, Media for News Consumption, Mainstream Media Consumption, Non-Mainstream Media Consumption [3 items each, sum score "Please indicate how often you have used the medium mentioned on average over the past 12 months. - TV", "Please indicate how often you have used the medium mentioned on average over the past 12 months. - News programs on TV (also online via the media library)", "Which sources of information do you mainly use to find out about politics in the media? (Multiple choices possible)"]	Availability sustainable groceries (situational/infrastructurel) [Diet]	Availability of seasonal, organic or regional groceries [1 item, numerical "In order to be able to buy organic, regional or seasonal products, I had to go a long way and undergo disturbing circumstances in 2019."]
Citizen Norms: Sustainability (psychological)	Environmental citizenship norms [1 item, numerical (7 point scale)] "There are different opinions about what constitutes a "good citizen". How important would you rate the following behaviors to be a good citizen? - Behave sustainably and environmentally friendly in everyday life"]	Availability sustainable food in restaurant (situational/infrastructurel) [Diet]	Price and options of vegetarian and vegan dishes in frequently visited restaurants/canteens [sum score of 5 items, numerical] "The restaurants / canteens I ate in regularly offer a wide range of vegetarian dishes."]
Citizen Norms: Political Consumerism (psychological) [Consumption, Diet]	Citizenship norms [1 item, numerical (7 point scale) "There are different opinions about what constitutes a "good citizen". How important would you rate the following behaviors to be a good citizen? - Select products according to political, ethical or ecological	Possibilities of sustainable consumption (situational/infrastructurel) [Consumption]	Possibility to rent and buy second hand [mean score of 2 items, categorical "In 2019 there were opportunities to borrow utensils (e.g. lawn mowers, drills, etc.) that I rarely need near my place of residence."]

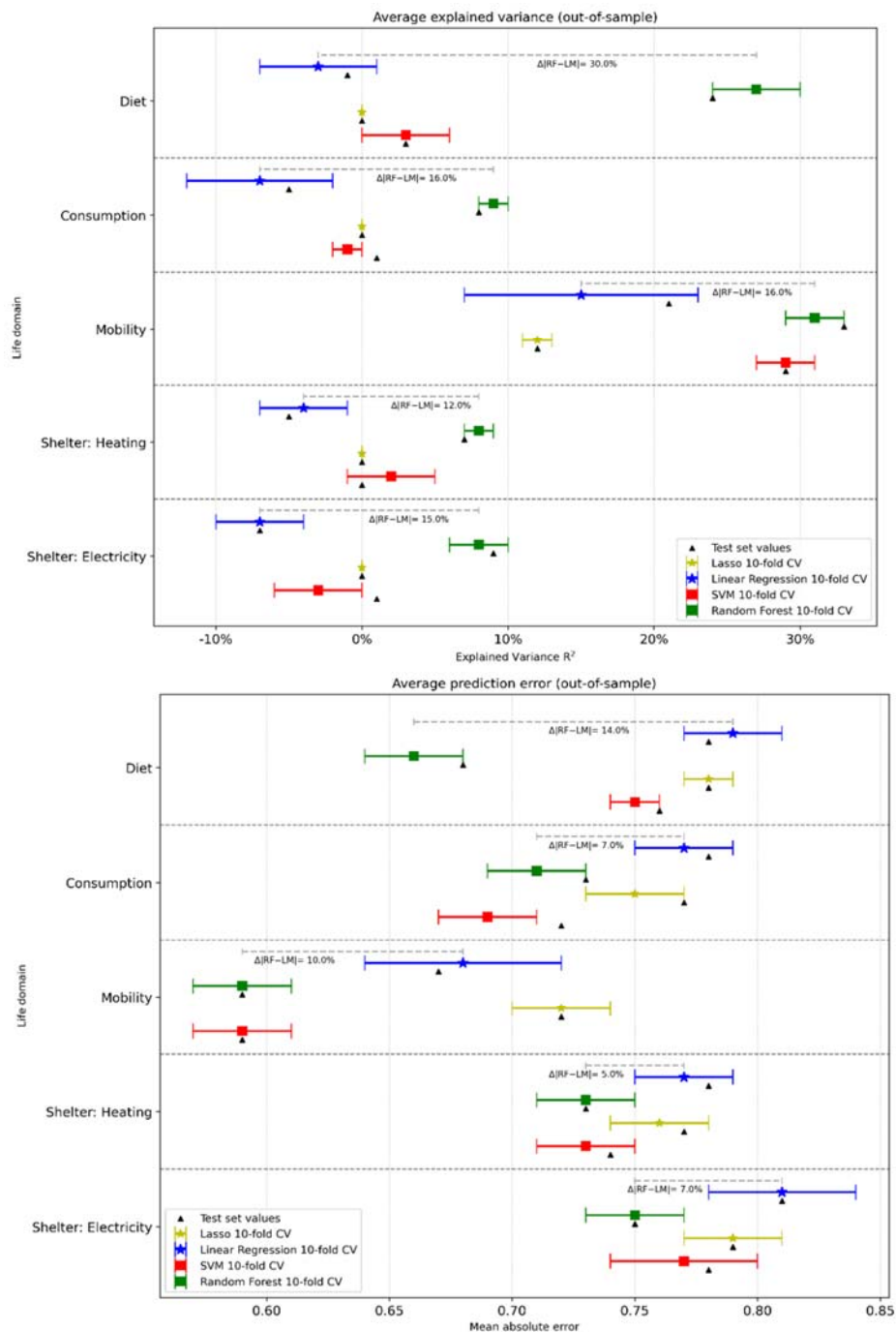
aspects, even if they are more expensive"]

Gender (demographical)	Self-identified gender of respondent [1 item, categorical (male, female, other)]	Knowledge electricity (technological) [<i>Shelter: Electricity</i>]	<i>Knowledge of annual Electricity Consumption</i> [1 item, binary coded "I know my annual electricity consumption from 2019."] <i>Knowledge of Electricity Consumption of Household Devices</i> [1 item, numerical "In 2019 I knew exactly which devices in my household were using the most electricity."]
Federal home state (demographical)	1 item, categorical	Traceability electricity (technological) [<i>Shelter: Electricity</i>]	<i>Power Consumption in own Household</i> [1 item, numerical "In 2019 I was able to see and understand my current power consumption exactly at all times. Information: For instance, using a smart meter can make electricity consumption traceable."]
Income (demographical)	Monthly income in € 1 item, numerical	Traceability heating (technological) [<i>Shelter: Heating</i>]	<i>Heating consumption traceability in own Household</i> [1 item, numerical "In 2019 I was able to see and understand my current heating consumption exactly at all times."]
Professional status (demographical)	Job status [1 item, categorical looking for a job, apprenticeship, student, employee, civil servant, self-employed, retired]	Traceability consumption (technological) [<i>Consumption</i>]	<i>Traceability of sustainability of new Products</i> [1 item, numerical "When I buy new everyday objects (e.g. furniture, electronics or clothing), I can see exactly how sustainably they have been produced."]

Highest educational qualification (demographical)	1 item, categorical	Traceability diet (technological) [Diet]	Identifiability of sustainable Food [2 items, numerical "The labeling of ecological, regional and organic food is easy to understand."]
Living area (demographical)	Area of residency (urban/rural) 1 item, categorical		
PLZ (demographical)	First three ZIP Code digits 1 item, numerical (integer number)		
Age (demographical)	1 item numerical (integer number)		

Note. Predictor names are shown as used in the models. Parentheses behind predictor names indicate respective subcategory of internal/external factor as shown in Figure 1, brackets indicate the respective domain the factor was used in to predict footprints. Predictors for internal factors were used in all five domains to predict footprints and phrased accordingly. TPB: Theory of Planned Behavior, VBN: Value Belief Norm Theory. Further information about the used items, previous results on predictor associations with sustainable behavior, comprehensive item phrasings and scale reliability can be found in supplementary section 2.

Figure 1. Prediction performance of individual GHG emissions in all life areas (cross-validation and independent test set) using different ML models and LM.



Note. Explained variance (upper plot) and mean absolute prediction error (lower plot) of the Random Forest (RF), Linear Regression (LM), Lasso Regression (LASSO) and Support vector machine (SVM) models. Colored error bars indicate performance in the 10-fold cross-validation (CV) on the training set, spanning a 1SD interval. Star and square indicators represent average performance across folds. Upward pointing triangles below the error bars indicate prediction performance on the independent test set. Dashed lines indicate absolute performance difference between the LM and RF. Values are rounded to two decimal places. Negative R^2 values indicate a worse model fit than always predicting the mean GHG emission ($R^2 = 0$) (see supplementary section 4.3).

9th Oct 23

Dear Mr Brandenstein,

Your manuscript titled "Predicting individual greenhouse gas emissions in everyday life: A machine learning approach" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

Please note that it may still be possible for your paper to be published before the end of 2023, but in order to do this we will need you to address these points as quickly as possible so that we can move forward with your paper.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a [CC BY license](http://creativecommons.org/licenses/by/4.0) (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/article-processing-charges>

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be

asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

[Link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Alienor Lavergne, PhD
Associate Editor
Communications Earth & Environment

Niheer Dasandi, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-8708-837X

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The article provides a detailed response to the questions I had previously raised. The structure and expression have also improved significantly compared to before. In this revision, there is a comprehensive review of the research direction in terms of method determination and the analysis of emission impact factors. The questionnaire data source for the research is also presented in the SI, and the accuracy of the model results is explained.

However, the figures in the manuscript are quite blurry, and the quality of the images could be improved.

The section on "Drivers and barriers of impactful sustainable behavior in domains shelter and consumption" is as important as the second section, and some of the results' figures could be included in the main text rather than being placed entirely in the SI. However, this also highlights a concern: the main content focuses on the discussion of factors in four domains, but the depth of the analysis seems somewhat lacking.

The current research aims to identify influencing factors, and I still perceive the level of innovation in this paper as moderate. As the author mentions on Page 4, "However, no previous work has yet fully met all these study requirements in a single analysis which would be necessary to comprehensively investigate drivers and barriers of sustainable behavior and derive effective demand-side solutions." In my view, the research approach in this paper and the literature references are quite similar. "All these study requirements" likely refer to analyzing as many driving factors as possible. On Page 20, "We found many important factors in our models for different life domains to resemble those identified in previous studies, but we also found the importance of quite a few factors to deviate from previous

literature." Due to methodological and data differences, it's expected that the results obtained would differ somewhat, and I don't consider this a significant research discovery.

The title of the paper is "Predicting individual greenhouse gas emissions in everyday life." I initially thought that after identifying the factors, the paper would go on to predict daily carbon emissions based on the identified factors and provide different scenario results. This would have been an interesting study. Of course, offering policy recommendations after identifying influencing factors, as done in the paper, is a valuable analysis, but it differs from what I expected in terms of "predicting emissions."

Overall, this revision has seen significant improvement. The overall flow is good, and the literature review is thorough. However, in terms of research depth and innovation, I still find it somewhat lacking.

Reviewer #2 (Remarks to the Author):

Thank you for your detailed answer to all of the raised comments and for making the necessary edits. I do not have more comments. Good luck!

Response Letter - Predicting individual greenhouse gas emissions in everyday life: A machine learning approach

Manuscript number COMMSENV-23-0367-T

Dear reviewers,

we sincerely thank you again for your thoughtful comments on our revised manuscript and the possibility to resubmit a revised version for publication in Nature Communications Earth & Environment. In the following we describe how we aimed to address all final concerns and comments.

Yours sincerely,

The authors

Reviewer #1 (Remarks to the Author):

Comment #1: The article provides a detailed response to the questions I had previously raised. The structure and expression have also improved significantly compared to before. In this revision, there is a comprehensive review of the research direction in terms of method determination and the analysis of emission impact factors. The questionnaire data source for the research is also presented in the SI, and the accuracy of the model results is explained. However, the figures in the manuscript are quite blurry, and the quality of the images could be improved.

Author response: Thanks for your feedback, we appreciate it! We redid all figures to improve the resolution as well as to comply with the standards of the journal and editorial requests.

Comment #2: The section on "Drivers and barriers of impactful sustainable behavior in domains shelter and consumption" is as important as the second section, and some of the results' figures could be included in the main text rather than being placed entirely in the SI. However, this also highlights a concern: the main content focuses on the discussion of factors in four domains, but the depth of the analysis seems somewhat lacking.

Author response: This is a very good point! In the latest revision, we transferred the respective predictor importance figures for domains consumption and shelter (Figure 4 and 5) into the main manuscript and also added a short discussion of the results to expand on these factors:

"An explanation for the lower prediction accuracy in our models may be that shelter related emissions are under lower personal control than emissions in other domains 18, 61. This is indicated by the relatively high factor importance of perceived control in our models (Figure 4) (...)" (p.14)

"Although in our models we found factors which were relevant for consumer behavior in previous studies (e.g., income⁵⁹ and habits⁶⁷) to also be the most important predictors of consumption GHG emissions (Figure 5), the overall model prediction performance remains relatively low (Figure 1)." (p.16)

Comment #3: "The title of the paper is "Predicting individual greenhouse gas emissions in everyday life." I initially thought that after identifying the factors, the paper would go on to predict daily carbon emissions based on the identified factors and provide different scenario results. This would have been an interesting study. Of course, offering policy recommendations after identifying influencing factors, as done in the paper, is a valuable analysis, but it differs from what I expected in terms of "predicting emissions."

Author response: We see that the previous title could have led to incorrect assumptions about our study. Thus, we suggested a different title in the Editorial requests which better reflects our main findings and study approach: **Key determinants of individuals' greenhouse gas emissions in Germany interplay and are mostly domain-specific**

Reviewer #2 (Remarks to the Author):

Thank you for your detailed answer to all of the raised comments and for making the necessary edits. I do not have more comments. Good luck!

Author response: Thanks a lot!