

Sodium constraints on megaherbivore communities in Africa

Corresponding Author: Dr Andrew Abraham

Version 0:

Decision Letter:

10th June 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Dr Abraham,

Your manuscript entitled "Sodium constraints on megaherbivore communities in Africa" has now been seen by 4 reviewers, whose comments are attached. The reviewers have raised a number of concerns which will need to be addressed before we can offer publication in Nature Ecology & Evolution. We will therefore need to see your responses to the criticisms raised and to some editorial concerns, along with a revised manuscript, before we can reach a final decision regarding publication.

In particular, several of the reviewers would like to see greater clarity and transparency in the methods reporting, and additional evaluation of the validity and predictive accuracy of the models.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file [OPTIONAL: in Microsoft Word format].

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a "Response to reviewers" document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at <http://www.nature.com/natecolevol/info/final-submission>. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

* Extended Data Figures - please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

[redacted]

Reviewer expertise:

Reviewer #1: Sodium, herbivory, grasslands

Reviewer #2: Ecosystem modelling, biogeochemistry

Reviewer #3: Herbivores, biogeochemistry

Reviewer #4: Spatial mapping, machine learning

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors examine the macroecology of sodium and its consequences for the distributions of African herbivores. I have reviewed prior submission versions of this manuscript and am satisfied with the large amount of revisions conducted by the authors that have greatly improved this study.

I have only a few minor comments:

Line 67: Would be better to cite the original studies, rather than the review paper here.

Line 79: Would be good to add one sentence about the predicted mechanism linking Na limitation with body size

Line 123: restate what values and relationships you refer to here

Line 128: Can this be quantified or perhaps a cite a figure/table here?

Line 190: delete this "and"

Line 225: We found a very similar result with insect herbivores in Welts et al. 2020 GEB, <https://doi.org/10.1111/geb.13119>, see Figure 5.

Line 288: This caveats paragraph could use a more general topic sentence

Reviewer #2 (Remarks to the Author):

In "Sodium constraints on megaherbivore communities in Africa", the authors set out to investigate how the spatial availability of Sodium (Na) influences large herbivore distribution across Sub-Saharan Africa. Na is a critical nutrient for herbivores but crucially it is not for plants, creating a stoichiometric mismatch that may influence the ecology of both trophic groups. The authors use empirical and remote sensing alongside a suite of machine learning algorithm to show that Na plant content (i) varies by up to 3 orders of magnitude across sub-Saharan, (ii) drives herbivore Na uptake, and (iii) explains large herbivore distribution patterns at continental scales.

Thank you for the opportunity to review this manuscript. I enjoyed reading through your work and found it a timely, exciting new contribution to the study of animal spatial and nutritional ecology. I was impressed not only by the scope of your work and the wealth of empirical data you assembled to answer your research question, but also by the rigour and attention to detail you put into your study design, analyses, and interpretation. This is a very well-designed, well-executed study. The discussion raises some interesting points and opens up a number of exciting new research opportunities, while seamlessly weaving your new results with current knowledge on the topic. Bravo!

As much as I liked the study, I do have some comments mostly focusing on clarifying certain aspects of your work to improve your manuscript's clarity and accessibility by a broad audience. I also have several comments on your code, as I encountered some challenges when I tried to replicate your results. I hope that these comments will help strengthen the manuscript overall and make even more citeable and far-reaching. I can certainly see myself citing it extensively in the future.

Main comments

As a first general comment, throughout the main text, I noticed that there is quite a bit of inconsistency in the use of the "large herbivore" label. For example, At l. 85, large herbivores are defined as all herbivores with body mass > 2kg. Later, at l. 215, we're introduced to 'meso-herbivores', whose body size ranges 50-500 kg. Further on, at l. 245, large herbivores are described as all herbivores < 500 kg. Please, revise this definition so it is both available to readers early in the text and consistent throughout the manuscript. This lack of clarity in defining the study's focal group of animals makes it hard to follow the authors' discussion of their results and casts some doubt on the study's conclusions.

Second, I encountered an issue when trying to replicate the analyses with the code and data provided. When I tried to fit the GAMM to the full dataset (all herbivores), I obtained a very different Table S7A where the tensor product interaction term between Diet_NA_selective and Log10_body_mass did not appear at all. Conversely, the GAMM output for the mega-herbivore only analysis in my run of the code was exactly the same as Table S7B. I am no expert in GAMMs, but this apparent inconsistency struck me as an odd and potentially serious issue as it directly affects one of the main results of the analysis---namely, that Na constraints are not uniform across body sizes (l. 201-210)---which is the basis for the separate analyses on mega-herbivores and a large part of the manuscript's discussion.

So, I ask the authors: is it to be expected that the results table for a GAMM would change among different runs of the model, to the point where one term is completely absent from it? Has it happened to them before? It could be that this issue is due to problems with my R setup. Indeed, running the code provided was not at all a smooth process (see below), so I cannot exclude that this different result depends on coding rather than statistical issues. Nevertheless, at a minimum I think the authors should review their code to ensure reproducibility and consistency in the results.

The next issue I found was with the data collection and management methods used. The description of data collection protocols and of the data itself is, in my opinion, lacking. I realize space is limited in Nature journals, but I identified at least four instances in which key information on data collection and handling was reported in the SI but really belonged in the main text.

1. The SI is the only place where readers are informed that, in some circumstances, the authors had to simulate data to include studies that only reported pooled foliar Na data in their dataset (see l. 71-80 of S1.1 in the Supplementary Information). This is an important piece of information that needs to be at least mentioned in the main text for readers to correctly interpret and evaluate the results. As well, even after reading the SI, it is unclear (i) what percentage of the final dataset is made up of these simulated data (we are only told that 67% of the final dataset is raw data; does this mean 33% is simulated data?), and (ii) if the maps of Na concentration are sensitive to the inclusion of these simulated data. I would encourage the authors to move this text to the manuscript's methods and include a brief discussion of whether excluding these data has any effect on the results.

2. The authors note that they collate data on foliar Na content from several field campaigns (l. 340-341) over the period 1994-2021 (l. 62 in Text S1, Supplementary Information). However, I could not find anywhere in the main text or in the supplementary materials a discussion of (i) whether the field protocols of these different field campaigns were comparable across years and field sites, (ii) what actions the authors took to ensure comparability if this was not the case. Additional details on these two items would both strengthen the manuscript and improve reproducibility of the study.

3. Similarly, the authors briefly mention that they conducted a field campaign to integrate the limited fecal Na data available in the literature referring the reader to Table S3 (l. 394-396). However, Table S3 and the related Text S1.2 offer little additional detail on this data collection effort. In fact, only in Text S1.2 does the reader find out the length of the field campaign (2017-2021) and that it was further integrated with data from Kenya collected in 1998-2000. Again, this sparks questions about what protocols were used, how comparable the data from 1998 were to those from 2021, and which are the 28 herbivore species (l. 159) included in the final dataset for fecal Na (could these be identified somehow in Table S4?). These issues should be addressed before the manuscript is published.

4. In general, the different dataset assembled for this study often cover very different time spans. For example, the foliar Na data span 1994-2021, whereas NPP was calculated for 2001-2010, and Na aerosol deposition was modeled only for 2006. Likewise, herbivore fecal data covers 1998-2000 in Kenya only and 2017-2021 across 10 other sites in southern Africa. Of course, relying on data from disparate sources is often unavoidable when asking questions at very large spatial scales. However, I would like to see the main text's discussion address this additional source of variation and uncertainty more explicitly---particularly as the authors conclude their manuscript with a strong emphasis on how quickly things can change in the context of global change.

Minor comments

l. 83-86: this sentence requires references to support the assertion that the variables listed are predictors of large herbivore abundance

l. 101: as you show in figure S2, a lot of your data points for foliar Na concentration are concentrated (no pun intended) within the studied Serengeti-Mara ecosystem. I was wondering if you explored whether having so many data points in one ecosystem has any effects on your overall conclusions re foliar Na distribution maps?

* it also appears that the Serengeti-Mara data are quite variable, at least as much as the rest of the data scattered across the continent. Is the Median foliar Na legend the same for the inset too? Or should the inset have a separate legend

* relatedly, I noticed all of the data points in the Serengeti-Mara inset are squares, i.e., C4 plants. Are there no data for C3 plants in this ecosystem? Or should panel (B) have a similar inset?

l. 147: as I mentioned above, can you provide details on the identity of these 28 species and whether they are included in the larger sample of 65 species included in your GAMMs? I expect they are, but it would be nice to have confirmation in the text.

l. 148: could you include a map of the sites from Table S3? It would be useful to compare it to the maps in Figure S2 to get an idea of how closely the herbivore fecal Na sampling locations map onto the distribution of plant foliar Na sampling sites.

l. 187-189: this list of variables used in your GAMMs is useful but felt incomplete when, later on, you list different variables by the amount of variance explained. Table S5 is quite large and by all means belongs in the SI, but maybe you could streamline your in-text lists of variables between here and l. 205-208 so readers do not have to wonder, e.g., where does human development index come from all of a sudden.

l. 223: I noticed you only report results from the unselective scenario for the mega-herbivore group. Is it because you implicitly assume these herbivores are not picky eaters due to their high food requirements caused by their large body size? Alternative, is this related to your later sentence discussing how you removed the degree of omnivory in diet from the mega-herbivore analyses (l. 482)? If that is so, it needs to be explained more explicitly as currently it is unclear why there are no "unselective" scenario results for the mega-herbivores

l. 242: this sentence, particularly from "with 55%" onwards, is unclear to me. Could you rephrase it?

l. 394: "scant data", it would be useful to have a quantitative measure (e.g., %) here rather than a qualitative one---for both literature-sourced and field-collected data.

l. 415: typo, "seclctive" -> "selective"

l. 419: Please indicate which figures the last sentence ("All results are plotted on log10 scales.") refers to

Figures

Figure 3: in panels (a) and (b) what do the colors of the bars mean? Are they related to the colors in the partial dependence plot (panel (c))? If so, what do the yellow bars mean?

Figure 4: is the legend on a log10 scale here as well?

Supplementary information comments

Given the amount of information relevant and necessary to understand the main text that the supplementary information (SI) contains, I offer a few comments on this document as well. I hope the authors will find these comments useful to improve both the main text and the SI.

l. 60-61: do you provide the references for these 16 published studies anywhere?

l. 71-80: I discussed this portion extensively in my comments on the main text above, but I want to note here that it is also not clear where else in the manuscript or SI is the 'total number of observations' reported. Please, add a cross-reference here.

l. 120-121: where these three overestimated sites included in subsequent analyses?

l. 159: could you list these 28 herbivore species here or in a table? Or highlight them in Table S4?

l. 171: it is unclear to me what the authors mean by 'diet' here: is it % omnivory in diet, as mentioned later (l. 173), or is it diet stoichiometry? If it is % omnivory, I would recommend to use this term consistently throughout instead of switching back and forth with diet.

l. 276: I do not doubt that you found substantial correlation between some of the anthropogenic factors, but I would rather see this result reported in a figure or table.

l. 333: typo, "predcited" -> "predicted"

l. 355-382: this section reads very much like l. 288-305 in the main text and I am not sure why it is repeated here. I would encourage the authors to instead use this space to discuss some of the other limitations of their study (see my other comments above for some ideas).

Table S2: reading through the manuscript, I really wanted this table in the main text rather than in the SI. Given how often you refer to your "ensemble of machine learning models", having a quick reference to them would be very useful.

Tables

Table S3: some entries (e.g., Gorongosa, Palabora) are missing data on latitude, longitude, or both

Table S4: two potential improvements here; (i) mark the species that were sampled for fecal Na data (l. 159) and (ii) mark whether a species is a 'large', 'meso', or 'mega' herbivore

Figures

Figure S1: Three comments on this figure

1. the axes should state which the base of the log used: natural, 2, 10, ...?
2. it is unclear if this figure is purely conceptual or whether data underlie the relationships shown. Also, why are ranges for the b exponent reported for "dietary Na intake" and "Na req (adult maintenance)" but not for "Na req (growing, lactating, etc)"? Maybe Duvall et al. (2023) explain this, but it should appear in the caption of this figure too.
3. the two smaller panels for Medium-Na and High-Na on the right should at a minimum state if the "Dietary Na intake" line (the more visible one) is the same as the Low-Na panel, but ideally also offer more information on the other lines shown.

Figure S4a: typo in panel 3, "Neaural" -> Neural, and in panel 5, "iteraratively" -> "iteratively"

Figure S10, l. 560: what do the numbers in parenthesis refer to? Are they meant to be panel labels? If so, the figure is not showing them. Or are they referring to the rows of panels? If so, they do not match them either.

Reviewer #2 (Remarks on code availability):

I would to commend the authors for provided all code and data needed to reproduce their analyses in a well structured, organize format. The folder provided is easy to navigate, intuitively organized, and has an informative README file. The code itself is well written, cleanly organized, and appropriately commented for easy of reading and running it.

Unfortunately, the code also cannot be run as provded. This is due to some of the packages ('rgdal', 'bgm') being no longer available on the CRAN nor actively supported. This means, for instance, that lines reading shapefiles and rasters, as well as the map-generating lines, cannot run and readers need to find an alternative to 'rgdal' to re-run the code. This is a significant hindrance when trying to reproduce the authors' results and needs be addressed before publication.

In the case of 'rgdal', I suggest the authors switch to either the 'sf' or 'terra' packages to run their spatial analyses, as these packages are actively supported and available on the CRAN, thus ensuring reproducibility of the analyses in the future. In my own attempts at runnig the code I found that substituting the functions 'rgdal::readOGR' with 'sf::read_sf' and 'rgdal::spTransform' with 'sf::st_transform' seems to fix most problems caused by using 'rgdal'. As well, changing 'ggplot2::geom_map' with 'ggplot2::geom_sf' (and changind relevant arguments in it) allows for generating the maps from the manuscript.

For package bgm, the authors should consider whether to move to the newest version available on the CRAN to maintain reproducibility.

Another issue I found was that some functions are called either before the relevant package is loaded or from a package that is not loaded at all, so running the code proceeds in halting way as one tries to figure out which package does the function belong to. This happens with function 'grid.arrange' from package 'gridExtra' (not loaded in the script) and function 'inSide' from package 'mgcv', called at l. 348-387 but the package is loaded at l. 655. Please, review the code and make sure all packages are correctly loaded at the beginning of the script to avoid these issues.

l. 318-332: Plot map of sodium concentration for C3 plants throws an error with `scale_fill_gradientn()` supplying discrete values to a continuous scale (same at l. 335-345)---could not find a solution, please double check this

l. 434 and l. 454: these comments are not describing what the following lines of code do. Each comment states the code below compared median to 95th percentile for C3 (l. 434) and C4 (l.454) plants but, in reality, the code is comparing median C4 to median C3 (l. 434) and C4's 95th percentile to C3's 95th percentile (l. 454). Please, correct these comments

The 'for' loop at l. 1118-1138 did not run for me, and instead threw an error: "Columns 'x' and 'y' don't exist." The issue seems related to object 'den', which does not contain any variable named 'x' or 'y', but rather 'x2' and 'y2'. Are these the intended variables to be used here?

Relatedly, the second 'if' statemente at l. 1130 does not work and instead throws an error, "Error in .local(x, i, j = j, ..., value) : y is already present as a coordinate name!". Please, check this too.

The same problem affects the mega-herbivore-only analysis of parameter importance (l. 1195-1225). Please, check it as well.

l. 1346: the file path given to argument 'dsn' is incorrect

Finally, as mentioned above, the outcome of the all-herbivore GAMM was different in my run of the code than what is reported in SI Table S7A. Unfortunately, I could not figure out whether this was due to the coding issues listed above or some other problem with the analyses. Please, double check this too.

Reviewer #3 (Remarks to the Author):

This is a fascinating and well-written paper that I'm excited to see published. Generally, the nutrient requirements of animals beyond nitrogen and phosphorus – and the ecological consequences of these nutrient requirements – are understudied at scales larger than individual animal physiology, and are overdue for greater consideration. Research on sodium intake by large herbivores in particular often centers around behavior at natural salt licks or anthropogenic mineral supplementation, and far less attention has been paid to the impacts of foliar sodium on herbivore ecology and biogeography. The maps of foliar Na concentrations alone are extremely useful contributions; overall, this manuscript fills an important gap and should have significant impact moving forward. I do have some comments for the authors, including one substantive concern, that I hope will improve the manuscript and clarify any methodological choices and limitations to conclusions prior to dissemination.

My biggest question relates to the GAMMs testing whether foliar Na (and its interaction with body size) measurably improves the performance of models of herbivore density. I like this approach overall, and as a broader technical point, I appreciate the care taken to detect and avoid overfitting and predictor variable autocorrelation in each of the individual models. However, the outputs of some models (i.e. predicted foliar Na) were predictors for the other models (i.e. estimated dietary Na as a predictor of herbivore density), and there is the potential for undetectable correlation or “double-counting” among predictor variables for these models. For example, distance to coasts could implicitly be incorporated in both models (in the herbivore density model, via anthropogenic pressures, which are frequently concentrated near coasts along with centers of human population, and in the foliar Na model, via sea salt aerosol deposition), and primary productivity is also a factor in both models. Thus, certain predictors are, at least indirectly, included more than once in the herbivore density GAMMs. Granted, both of these variables may have opposite directional effects on herbivore density and foliar Na, and thus perhaps these particular examples are not of immense concern. However, it is necessary to demonstrate that this potential for “statistics on statistics” is not an issue; as a scientist with admitted “small-data” tendencies, I get nervous about drawing too many conclusions from models built from modeled predictor variables, particularly when some variables may inform multiple levels of estimates (aridity vs. distance to water, exchangeable Na & K and CEC vs. soil fertility include other examples of potentially related predictors across models). To assuage this concern, I would recommend some discussion and assessment of correlations between predictor variables across foliar Na and density models in the Supplement. Perhaps the Supplement could also include a more basic LMM (similar to the one relating fecal Na to dietary Na; Fig S7) regressing the mean Hempson et al. herbivore density observations against dietary Na for sites with empirical foliar Na measurements, if sufficient data exist. The R^2 would likely be low, but it could be useful to visualize how estimated dietary Na alone relates to field observations of herbivore density.

Was any consideration given to the year or season of Na sampling in the analysis? 1994-2021 (the period of direct foliar Na sampling by the authors) is a relatively long period of time, and presumably the data sourced from published literature span an even wider range of time. I agree that spatial differences in foliar Na resulting from climatic, edaphic, and biogeographic factors across sampling sites probably have a greater effect than single site year-over-year variation. However, foliar sodium levels are not static and can vary strongly across points in a growing season and across years as environmental conditions fluctuate; the timing of sampling across studies may influence how representative values are of the foliar sodium available to herbivores more generally in each habitat. [Edit: I see this is addressed in Text S3; however, I think it is of sufficient importance to mention at least briefly in the Discussion. Is there any reason to believe that foliar Na tends to be measured during certain points in the growing season? Could Na limitation shift seasonally across space or plant type?]. I tried to investigate this myself with the data provided and found something is funky with the Sach et al. 2020 foliar Na dataset – month and year both have strange, impossible values. However, when these data are removed it does seem that many more observations were collected in June and August compared to other months.

Regarding Figures 2a and S7 – could these be presented in non-log space? Studies (in this journal no less!) have shown that log-transformed data are extremely difficult for people to interpret, including trained ecologists (Menge et al. 2018). Particularly for the selective scenario (Figure 2a), the R^2 is high and the relationship compelling; however, it appears (on the log scale at least) that the relationship could be non-linear and in fact there is a threshold of foliar Na (somewhere around 1000 mg/kg) at which fecal Na concentrations saturate. This is roughly the same threshold at which dietary Na concentrations shift from having a positive to negative impact on megaherbivore density as per Figure 3d – might some greater consideration of this threshold be fruitful?

Line 201 – Based on table S6, should this not be dietary Na from the *selective* scenario?

Lines 235-237 and similar points in Discussion and Figure 4 – This is an extremely neat observation. Any insights from the Na model as to what factors are driving lower foliar Na in these areas? The Discussion later mentions soil infertility and other previous hypotheses explaining the lack of megaherbivores in these parts of western Africa, but does not detail why foliar Na could be lower here (greater distance from the Eastern coast? However, this does not lead to reduced Na at similar distances in southern Africa). This seems of particular importance given that elephant, giraffe, and black rhino pre-colonial ranges encompass these areas. Has something changed? Is Na availability here lower than it was historically? Generally, given the amount of space dedicated to this discussion of “missing megaherbivores”, it could be helpful to incorporate some representation of current megaherbivore densities by species into Figure 4. Are elephants and giraffes less dense in these areas with high potential for Na deficiency? Is there any reason to believe that has changed since colonization, or do intrinsic local factors render such areas consistently ill-suited to high densities of megaherbivores?

Lines 254-255 – I'd also reference Figure 3c here.

The Discussion does a great job of elegantly explaining the results and implications while giving plenty of room for caveats and the limitations of these findings.

- Julia D. Monk

References:

Menge, D. N. L. et al. Logarithmic scales in ecological data presentation may cause misinterpretation. *Nat Ecol Evol* 2, 1393–1402 (2018).

Reviewer #3 (Remarks on code availability):

With a few tweaks to account for different R versions and package retirements/updates (see below), I was able to run the code and reproduce the results and figures with little trouble.

For publication, it would be helpful if the authors could update to use the *sf* package since *rgdal* has now been retired. I understand analyses were conducted on a prior version of R, but for future users who may want to replicate the analysis or learn from the methods applied, *sf* spatial functions have greater longevity (it didn't take me that long to replace these functions when reviewing the code).

For the Dataset column in multiple datasets, it could be helpful to put everyone's full name, in case others looking to collaborate on similar projects want to be able to contact the relevant scientists (I know that Roo is Andrew Abraham, but not everyone does!).

Reviewer #4 (Remarks to the Author):

The text is clearly written, the figures are beautiful, the manuscript is pleasing to read.

I've been asked to provide comments on the spatial mapping component of the study.

Broad scale mapping of ecological variables using ML technics has become a common task nowadays, although there is no standard & widely-accepted procedure to perform the assessment of broad-scale mapping model's predictive performance – which is problematic. Among the numerous studies that have performed such task in the past years, I find the present study to be particularly transparent, rigorous and comprehensive in the way they performed and presented their model validation. I have no major (negative) criticism on the spatial mapping component of the manuscript, and will just address two comment to the authors.

First, as a naïve reader who does not know much about foliar Na, I am fairly clueless after reading the main text of your manuscript and the supplementary data on whether or not your mapping model(s) are effective at mapping foliar Na. In the main text (section “Patterns of foliar sodium across sub-Saharan Africa” – L. 99), there is no comment about mapping model predictive performance. You simply state that you mapped it. In the Methods either. One needs to go to the Supplementary – in particular to the beautiful figure S4b – to see some RMSEs. But again, as a naïve reader I'm not sure how to interpret RMSE values – not only because there is no unit (add one), but also because it is not directly evident to me if ex. 0.6 is small or large. I'm a bit skeptical about it : are we in the same unit as in fig. S2B & C ? It would be surprising, because it would mean the mean prediction error (ex. RMSE = 0.5) is about ~ 5 % of the lowest observed value (~ 10 Mg. Kg⁻¹)? This seems fishy. A suggestion is to (i) add a second Y axis to the scatterplots in Fig. S4b with RMSE values in % (which would be easier to interpret), and/or (ii) add scatterplots of associated R² - so that readers have some clue on the share of variance explained by models. If presented RMSEs are indeed of the order of 5%, then I definitely suggest adding R² values.

Second, in my opinion, your model predictive power assessment would have been complete if you've had compared your models to a purely spatial model i.e. a model based on observations geographical coordinates alone. It would be very easy to add the trend line of a purely spatial model to Fig. S4b (together with all other models). This would provide some context / a baseline to help the interpretation of your models RMSE values.

Minor comments

Table S2 : row of “RF”, mtry ranges from 1 to 289 ? Isn't there only 11 predictors? Why 289?

*****END*****

Version 1:

Decision Letter:

8th October 2025

Dear Dr. Abraham,

Thank you for submitting your revised manuscript "Sodium constraints on megaherbivore communities in Africa" (NATECOLEVOL-25041294A). It has now been seen again by the original reviewers and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Ecology & Evolution, pending minor revisions to satisfy the reviewers' final requests and to comply with our editorial and formatting guidelines.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTeX)-- we can not proceed with PDFs at this stage.

If you have not done so already, please ensure that you also email us a completed copy of the Reporting summary :

Reporting summary: <https://www.nature.com/documents/nr-reporting-summary.pdf>

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Ecology & Evolution. Please do not hesitate to contact me if you have any questions.

[redacted]

Reviewer #2 (Remarks to the Author):

Thank you for the opportunity to review a revised submission of this manuscript. I reviewed the original submission and I enjoyed reading through this revised version. Overall, I found the manuscript much improved. The authors did an excellent job of revising the manuscript and incorporating the comments of all reviewers. Well done! I have no further major comments on this work.

Matteo Rizzuto

Reviewer #2 (Remarks on code availability):

Code remarks

I commended the authors on the extensive rework of their code to update the R packages used, expand on several of the analyses reported, and improve the organization of their script. I was able to reproduce the results reported in both the manuscript and the SI. Below, I report a few minor concerns and issues I had with the code, which I am believe can be easily fixed.

* when producing maps of sodium concentration for C3 and C4 plants (l. 339-348 and l. 353-362), the `fill` argument of `geom_tile` needs to be set to `ensemble$pred` otherwise the code fails (at least on my machine)

* line 757 is supposed to prepare data to produce a correlation map across variables in the model, but the three variables called in the `which()` function are not part of the `den` dataframe (i.e., "age", "BP", "cholesterol"). They are not in either version of the `Herbivore_density_data` dataset

* l. 988 and 989, the `geom_text` call to add labels does not work and throws an error:

Error in `annotate()`:

! Problem while setting up geom aesthetics.

! Error occurred in the 4th layer.

Caused by error in `list_sizes()`:

! `x$label` must be a vector, not an expression vector.

this is apparently a relatively recent issue (<https://github.com/tidyverse/ggplot2/issues/6622>), so no fault to the authors but just thought I'd flag it in case by the time the paper is published a workaround has been found or the authors want to change how the R^2 and RMSE information is presented in these graphs (this affects all other `geom_text` calls in the cross validation plots)

* when fitting model `megaherbivore_selective_Na_gamm` (l. 1561-1586), there is an extra call for `s(x, y, bs = "so", xt =`

list(bnd = bound),sp=2) +" in the bam() function call, which leads the model fitting to failure. Once removed, the results match Table S7B

* when drawing partial dependence plots, the function call for `draw()` generates the following error:

Error in `map()`:

! In index: 14.

Caused by error in `filter()`:

! In argument: `is\_inside`.

Caused by error:

!`.1` must be of size 10000 or 1, not size 0.

Backtrace:

1. gratia::draw(all_herbivore_selective_Na_bodymass_gamm)

23. dplyr::dplyr_internal_error(...)

Unfortunately, I could not find a solution for this error but I thought I would flag it in case the authors have better luck.

Reviewer #3 (Remarks to the Author):

I thank the authors for their incredibly thorough and careful consideration of all reviewer comments, and consider the manuscript much improved. I hope to see it published soon, and have just a few final (minor) suggestions to improve the strength of the conclusions and the clarity for readers.

1. I am not fully satisfied by the discussion of the issue of differences in sampling timescales for foliar Na and fecal Na. Figure R3 shows the histograms of year of collection for foliar and fecal samples, and indicates that most collection of foliar samples happened after 2000, which aligns with other covariates. However, a significant proportion (roughly 1/3?) of data were still collected prior to 2000; furthermore, most fecal samples were collected in 2018 and afterwards. Thus, while the peaks of these two histograms are similar, their distributions are very different; thus, in analysis, it remains likely that many, if not most, fecal Na concentrations are being regressed against foliar Na data collected some 10-50 years prior, or against simulated values based on those data. More concretely, it would be helpful if the supplement included a plot showing the relationship between year of sampling and foliar Na concentrations for the entire dataset. This would reveal whether there are regional trends in foliar Na concentrations over time (sensu Kaspari and Welte 2024, as the authors discuss), or if year of sampling bears no relationship to foliar Na (which would reduce the need for concern); this would also allow readers to see for themselves, as the response to reviewers indicates, that this variation is much smaller than the spatial variation in foliar Na presented in the manuscript. Similarly, was there any relationship between the temporal and spatial clustering of these leaf data (i.e. were samples collected in similar parts of Africa in the same years, even if not at the same site)? Seeing these considerations in the supplement would increase the reader's confidence in the relevance of linking foliar Na and fecal Na data that were collected across different time periods and using that relationship to infer the impacts of foliar Na availability on dietary Na deficiencies and species distributions.

2. I appreciate the careful consideration of "double-counting" in response to my previous review. The correlation matrix in Figure S3a should indicate which variables are included in the foliar Na predictions and which are used to model herbivore density.

Line 295 – typo; nutrient vs. nutritent

- Julia Monk

Reviewer #4 (Remarks to the Author):

I read with attention the rebuttal letter of A. J. Abraham et al. and the modifications made to their manuscript. The authors fully accounted for my comments, by notably adding (1) an additional baseline model (ie. a purely spatial model) and (2) an additional metric (ie. the R^2) to facilitate the assessment of the predictive power of their foliar Na concentration prediction models. I am satisfied with the validation approach of the mapping models, which I think is robust and transparent.

Open Access This Peer Review 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 cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review 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 <https://creativecommons.org/licenses/by/4.0/>

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors examine the macroecology of sodium and its consequences for the distributions of African herbivores. I have reviewed prior submission versions of this manuscript and am satisfied with the large amount of revisions conducted by the authors that have greatly improved this study.

Thank you for reviewing our paper again and providing a positive assessment – it is much appreciated.

I have only a few minor comments:

Line 67: Would be better to cite the original studies, rather than the review paper here.

We now cite the original studies.

Line 79: Would be good to add one sentence about the predicted mechanism linking Na limitation with body size

We now state: *“Recent studies^{5,22} have suggested that due to discrepancies between allometric scaling relationships of Na requirement and dietary Na intake, large-bodied herbivores are predicted to be particularly vulnerable to Na limitation (figure S1).”* (L70-72).

Line 123: restate what values and relationships you refer to here

We have altered this sentence to read: *“Aridity was among the most important predictors of both median and 95th percentile values, but its relationship with foliar Na was variable and nonlinear, suggesting that its effects are context dependent (figure S6).”* (L118-120).

Line 128: Can this be quantified or perhaps a cite a figure/table here?

We now point readers to Figure 1, which shows the spatial pattern of plant Na concentrations we refer to.

Line 190: delete this “and”

Done.

Line 225: We found a very similar result with insect herbivores in Welts et al. 2020 GEB, <https://doi.org/10.1111/geb.13119>, see Figure 5.

It is nice to see the consistency between our results. We have now included the reference of Welts et al. (2020) and highlight this to readers by stating *“... aligning with results assessing drivers of insect abundance in North America¹.”* (L221-222).

Line 288: This caveats paragraph could use a more general topic sentence

We now introduce the caveats paragraph by opening with: *“Our methodological approach did, however, incur a number of limitations.”* (L287).

Reviewer #2 (Remarks to the Author):

In "Sodium constraints on megaherbivore communities in Africa", the authors set out to investigate how the spatial availability of Sodium (Na) influences large herbivore distribution across Sub-Saharan Africa. Na is a critical nutrient for herbivores but crucially it is not for plants, creating a stoichiometric mismatch that may influence the ecology of both trophic groups. The authors use empirical and remote sensing alongside a suite of machine learning algorithm to show that Na plant content (i) varies by up to 3 orders of magnitude across

sub-Saharan, (ii) drives herbivore Na uptake, and (ii) explains large herbivore distribution patterns at continental scales.

Thank you for the opportunity to review this manuscript. I enjoyed reading through your work and found it a timely, exciting new contribution to the study of animal spatial and nutritional ecology. I was impressed not only by the scope of your work and the wealth of empirical data you assembled to answer your research question, but also by the rigour and attention to detail you put into your study design, analyses, and interpretation. This is a very well-designed, well-executed study. The discussion raises some interesting points and opens up a number of exciting new research opportunities, while seamlessly weaving your new results with current knowledge on the topic. Bravo!

Thank you for your thorough review and enthusiasm for our paper. We have now incorporated all your comments and suggestions into our updated manuscript.

As much as I liked the study, I do have some comments mostly focusing on clarifying certain aspects of your work to improve your manuscript's clarity and accessibility by a broad audience. I also have several comments on your code, as I encountered some challenges when I tried to replicate your results. I hope that these comments will help strengthen the manuscript overall and make even more citeable and far-reaching. I can certainly see myself citing it extensively in the future.

Main comments

As a first general comment, throughout the main text, I noticed that there is quite a bit of inconsistency in the use of the "large herbivore" label. For example, At l. 85, large herbivores are defined as all herbivores with body mass > 2kg. Later, at l. 215, we're introduced to 'meso-herbivores', whose body size ranges 50-500 kg. Further on, at l. 245, large herbivores are described as all herbivores < 500 kg. Please, revise this definition so it is both available to readers early in the text and consistent throughout the manuscript. This lack of clarity in defining the study's focal group of animals makes it hard to follow the authors' discussion of their results and casts some doubt on the study's conclusions.

We have tightened our language on defining "large-herbivores" throughout the manuscript. Specifically, we define large-herbivores in the introduction as animals >2kg (L74); this corresponds with the herbivore body size range we analyse in our paper (see Table S4). To avoid confusion in our results, we have now removed the term "meso-herbivore" and instead refer to "herbivores within the body mass range: 50-500kg" at L209. We do, however, still define megaherbivores as animals > 1000kg (L212) as this is a commonly-used and well-defined term; we ensure clarity by using this term to describe which animals appear impacted or not by sodium concentrations (e.g., in our abstract L29 and results L216-224).

Second, I encountered an issue when trying to replicate the analyses with the code and data provided. When I tried to fit the GAMM to the full dataset (all herbivores), I obtained a very different Table S7A where the tensor product interaction term between Diet_NA_selective and Log10_body_mass did not appear at all. Conversely, the GAMM output for the mega-herbivore only analysis in my run of the code was exactly the same as Table S7B. I am no expert in GAMMs, but this apparent inconsistency struck me as an odd and potentially serious issue as it directly affects one of the main results of the analysis---namely, that Na constraints are not uniform across body sizes (l. 201-210)---which is the basis for the separate analyses on mega-herbivores and a large part of the manuscript's discussion.

So, I ask the authors: is it to be expected that the results table for a GAMM would change among different runs of the model, to the point where one term is completely absent from it? Has it happened to them before? It could be that this issue is due to problems with my R setup. Indeed, running the code provided was not at all a smooth process (see below), so I cannot exclude that this different result depends on coding rather than statistical issues. Nevertheless, at a minimum I think the authors should review their code to ensure reproducibility and consistency in the results.

Thank you for carefully checking our code. For clarification, the GAMM results should be the same each time the model is run.

In our original analysis, we generated different GAMM models with and without diet_Na (both individually and as a tensor product with body mass) by commenting dietary_Na in and out of the model. The uploaded R script in our Figshare account had dietary_Na commented out and therefore produced different results to Table S7A, which shows our best GAMM which includes dietary Na interacting with body mass. Indeed, the model you ran here was showing the results of our base GAMM (first row of Table S7A). We had left the megaherbivore-only model with dietary Na included, hence why it appeared exactly the same as Table S7B.

We recognise that this was poor practice and apologise for the confusion caused. We have now re-written this section of code to explicitly define each model separately and describe which model corresponds with each row in Table S7A and Table S7B in our R script notes. When running the code now, you should see the exact same output as Table S7A under at Lines 1366-1615 in the R script.

The next issue I found was with the data collection and management methods used. The description of data collection protocols and of the data itself is, in my opinion, lacking. I realize space is limited in Nature journals, but I identified at least four instances in which key information on data collection and handling was reported in the SI but really belonged in the main text.

We have now updated the methods in the main text and supplement to include the requested information. Please see our rewritten study caveats section in the Discussion (L287-310), Methods (L342-363; L403-439) and Supplementary Information (L56-110; L394-448).

1. The SI is the only place where readers are informed that, in some circumstances, the authors had to simulate data to include studies that only reported pooled foliar Na data in their dataset (see l. 71-80 of S1.1 in the Supplementary Information). This is an important piece of information that needs to be at least mentioned in the main text for readers to correctly interpret and evaluate the results. As well, even after reading the SI, it is unclear (i) what percentage of the final dataset is made up of these simulated data (we are only told that 67% of the final dataset is raw data; does this mean 33% is simulated data?), and (ii) if the maps of Na concentration are sensitive to the inclusion of these simulated data. I would encourage the authors to move this text to the manuscript's methods and include a brief discussion of whether excluding these data has any effect on the results.

Thank you for this helpful request. We now explicitly outline the number of observations in the final dataset that are from raw data (i.e., 1 sample included as 1 observation; $n=2,768$) and those that were simulated from pooled mean values ($n=1,490$) in the methodology of our main text (L342-363). We have also ensured that we better identify which data are raw vs simulated in our datasheet included in our Figshare repository.

As we are interested in calculating the median/95th percentile for each location for our plant Na concentration database, it was important that we did not underestimate the contribution of averaged values collated from the literature, as these sites likely correspond more closely to our median estimates. We could not simply weight calculations by the sample size as we were calculating the median not mean and were also interested in calculating the 95th percentile, which is based on the spread of data. In this case, it was necessary to simulate some data (33% of the final dataset) to generate a range of plausible values from the mean when the raw data was missing. At sites where we had a large sample size ($n>20$), the data was almost always normally distributed, with a standard deviation approximately equal to the mean; where necessary, we therefore generated pseudo-data using a normal distribution around the mean-value obtained from literature sources.

We appreciate that this is not ideal and would have much preferred to include the raw data in all instances. However, unfortunately, this was not possible. We note that simulated data only contribute to 24 locations out of a total of 289 geographical locations in our final database. Therefore, as we build our statistical models using only one median or 95th percentile value from each grid cell, the effect on model performance for predicting plant Na concentrations is relatively small.

Nevertheless, to ensure that our simulation procedure did not influence the outcome of our plant Na models, we have rerun the models using only the raw data only and compared model outputs to our results using the full dataset. See below for comparison of the (i) raw only and (ii) full dataset outputs for the median scenario (we also checked the 95th percentile scenario which had correspondingly very similar outcomes):

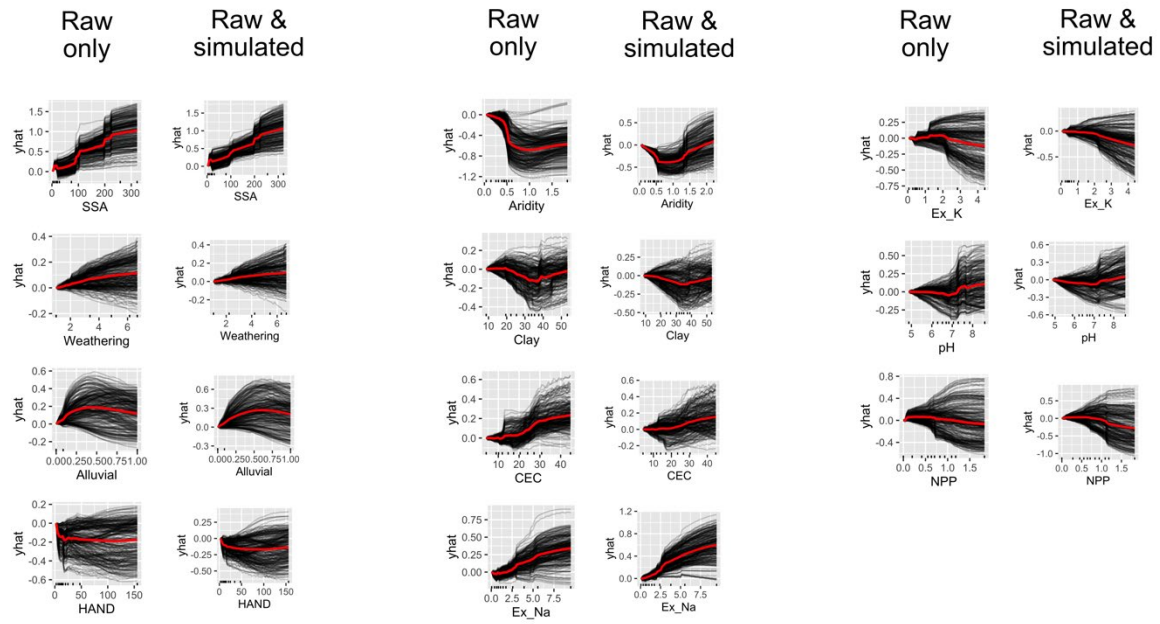


Figure R1. Comparison of model partial dependence plots of the ensemble of machine learning models used to predict plant Na concentration using (i) raw data only (n=2768; 265 locations), and (ii) raw and simulated data together (n=4,258; 289 locations). Note that relationships are very similar between model comparisons.

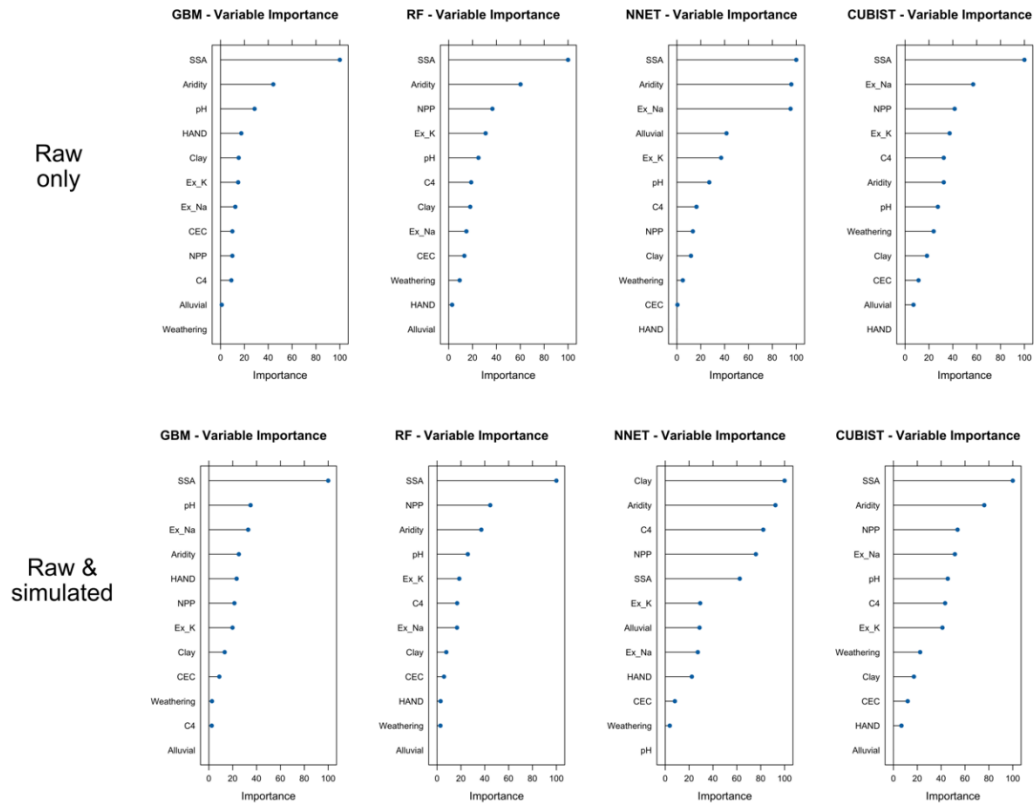


Figure R2. Comparison of model variable importance plots of the individual machine learning models used to predict plant Na concentration using (i) raw data only (n=2768; 265 locations), and (ii) raw and simulated data together (n=4,258; 289 locations). Note that variable importance rankings are similar between models-dataset pairs.

As the model outputs are similar in both cases, the use of simulated data does not alter our predictions of foliar Na across sub-Saharan Africa, but inclusion of simulated data did extend the geographical distribution of our observations giving more confidence to our conclusions as model predictions are better when nearby training values are included (see Figure S4).

2. The authors note that they collate data on foliar Na content from several field campaigns (l. 340-341) over the period 1994-2021 (l. 62 in Text S1, Supplementary Information). However, I could not find anywhere in the main text or in the supplementary materials a discussion of (i) whether the field protocols of these different field campaigns were comparable across years and field sites, (ii) what actions the authors took to ensure comparability if this was not the case. Additional details on these two items would both strengthen the manuscript and improve reproducibility of the study.

This is an important point. Our final dataset (including data from both published and unpublished datasets) comes from ~35 data sources and spans the period 1963-2018. Inevitably, this incurs challenges of comparison between field protocols and laboratory techniques; a common challenge for any study collating rarely collected data at continental-scales.

We now clarify in our SI that: *“Inevitably, there is some slight discrepancies between plant collection protocols across these field campaigns (e.g., targeting of specific plant species or leaf age). We therefore only included observations if foliar samples were healthy leaves that did not show obvious signs of herbivory or disease damage. Measurements were analysed for Na using an Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES), Inductively Coupled Plasma Mass Spectrometer (ICP-MS) or Horiba Compact Natrium Na Ion Meter LAQUAtwin readers in various labs in Africa, Europe and the USA. As with other studies collating data from multiple field campaigns and laboratories over long time periods (e.g., Kattge et al., 2021), there may have been small discrepancies between the laboratories and methodologies used to estimate foliar Na concentration. However, we are confident that such discrepancies do not impede our ability to analyse gradients across the continent that span >4 orders of magnitude (Figure S2).” (L64-74).*

3. Similarly, the authors briefly mention that they conducted a field campaign to integrate the limited fecal Na data available in the literature referring the reader to Table S3 (l. 394-396). However, Table S3 and the related Text S1.2 offer little additional detail on this data collection effort. In fact, only in Text S1.2 does the reader find out the length of the field campaign (2017-2021) and that it was further integrated with data from Kenya collected in 1998-2000. Again, this sparks questions about what protocols were used, how comparable the data from 1998 were to those from 2021, and which are the 28 herbivore species (l. 159) included in the final dataset for fecal Na (could these be identified somehow in Table S4?). These issues should be addressed before the manuscript is published.

These details were an oversight on our part. We now include more details related to the specific field campaign undertaken in both the methodology of the main text and text S1.

In the main text we now state: *“There is a paucity of large-herbivore fecal samples analysed for Na in published literature sources covering sub-Saharan Africa. Consequently, we combined the scant data (n=319; 6 species) found using the search terms [“sodium”, “Africa”, “faeces”, “feces”] in Google Scholar and Web of Science (June 2022) with unpublished data from previous field studies (Laikipia 1998-2000; Hluhluwe – Imfolozi Game Reserve 2018 and Gorongosa National Park 2021-2022) and our own specific field campaign (n=1037; see table S3). All unpublished data represented samples of fresh dung from one individual animal and were analysed using either Inductively coupled plasma optical emission spectrometer (ICP-OES) or Inductively coupled plasma mass spectrometer (ICP-MS). In total, we collated 1,356 geolocated fecal Na concentration values from 28 herbivore species across 20 wildlife reserves spanning ~4,000 km from equatorial Kenya to the Eastern Cape of South Africa.” (L409-419).*

We have also updated our SI document to include data on the timing, number of samples and specific protocols using for the inclusion of previously collected unpublished data: *“Additional unpublished data (n=146) had already been collected in Laikipia (Kenya) in 1998-2000, Hluhluwe – Imfolozi Game Reserve (South Africa) in 2018 and Gorongosa (Mozambique) in 2021-2022. All samples were collected as fresh single faecal samples from one individual. Specific locations of all fecal collections in our final database are outlined in table S3. South African*

samples specifically collected for this project were collected fresh and frozen within 8 hours at -20°C. Samples were transported to the Endocrine Research Laboratory, University of Pretoria, for lyophilisation at -54°C and ~0.96 mbar for 5-7 days until completely dry. Dry fecal samples were subsequently pulverised using a pestle and mortar to homogenise the sample. At the University of Pretoria Soil Sciences Laboratory, 0.25-0.30g of dried fecal sample was digested in 10ml of Suprapur Nitric acid (65%) and analysed for Na concentration using a SPECRO GENESIS Inductively coupled plasma optical emission spectrometer (ICP-OES). Samples from Hluhluwe – Imfolozi Game Reserve (South Africa), Gorongosa (Mozambique) and Laikipia (Kenya) were analysed using an Inductively coupled plasma mass spectrometer (ICP-MS). These original data were added to 319 observations from six peer-reviewed publications (table S3) for a final dataset of 1,356 observations from 28 large herbivore species. The large herbivore species from which we obtained faecal Na observations are outlined in Table S4.” (L165-180).

4. In general, the different dataset assembled for this study often cover very different time spans. For example, the foliar Na data span 1994-2021, whereas NPP was calculated for 2001-2010, and Na aerosol deposition was modeled only for 2006. Likewise, herbivore fecal data covers 1998-2000 in Kenya only and 2017-2021 across 10 other sites in southern Africa. Of course, relying on data from disparate sources is often unavoidable when asking questions at very large spatial scales. However, I would like to see the main text's discussion address this additional source of variation and uncertainty more explicitly---particularly as the authors conclude their manuscript with a strong emphasis on how quickly things can change in the context of global change.

We appreciate that foliar and dung Na signals may change through time and in places we have mismatch between the exact periods where data was collected, and covariates determined. However, we have attempted to align all covariates with the closest period that most observations were collected. We note, for example, that while our dataset is collected over a long-period of time (1963-2021); >80% of the plant and faecal observations were collected after 2000, coincident with the time periods that our covariates are calculated for.

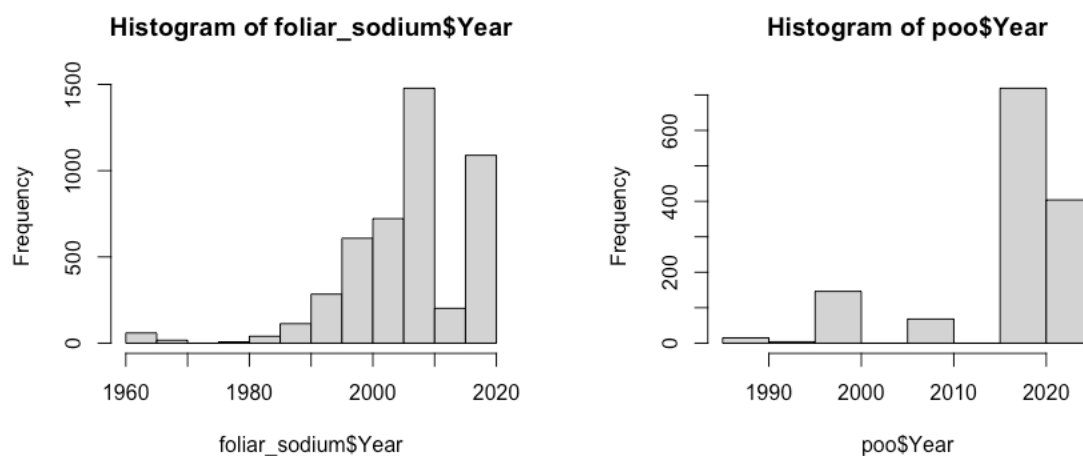


Figure R3. Histograms showing when the plant (left) and faecal data were collected. Note that most plant data was collected during the period 2000-2010, which aligns with the period over which most of our covariates were calculated.

While we would like to assess how Na availability for large-herbivores changes through time, we are not confident that our dataset, which (i) spans strong spatial gradients and (ii) does not have repeat measurements in the same place is suitable for robustly making such conclusions. In coming years, we hope to bolster our dataset after which we may potentially be able to look at how plant and faecal Na have changed in recent decades.

We now take space in our discussion to reflect on the role that both intra- and interannual variations in plant and faecal Na concentrations may play in determining our results, and how future studies can attempt to encapsulate this temporal aspect (L293-296).

Minor comments

I. 83-86: this sentence requires references to support the assertion that the variables listed are predictors of large herbivore abundance

We have included references to Olff et al. (2002); Santini et al. (2022) and Pranzini et al. (2024).

I. 101: as you show in figure S2, a lot of your data points for foliar Na concentration are concentrated (no pun intended) within the studied Serengeti-Mara ecosystem. I was wondering if you explored whether having so many data points in one ecosystem has any effects on your overall conclusions re foliar Na distribution maps?

* it also appears that the Serengeti-Mara data are quite variable, at least as much as the rest of the data scattered across the continent. Is the Median foliar Na legend the same for the inset too? Or should the inset have a separate legend

* relatedly, I noticed all of the data points in the Serengeti-Mara inset are squares, i.e., C4 plants. Are there no data for C3 plants in this ecosystem? Or should panel (B) have a similar inset?

Thank you for this important point. We have indeed incorporated this potential issue of spatial autocorrelation into our machine learning framework. 1) We weight each grid cell by the total number of observations collated in that grid cell; in this respect the Serengeti area was a distributed field campaign and each 0.1-degree grid cell only has 1-2 samples, thus lowering the impact on overall model prediction. 2) When assessing model performance, we specifically undertake a buffered leave one out cross validation to evaluate how model performance changes with distance from the training point to consider how well the model can predict in regions away from data clusters, such as the Serengeti. Further, we then use k-means clustering and calculate the R2 and RMSE as an average from different regions to prevent the high-number of samples from the Serengeti from giving overconfidence in model performance. For an overview of the pipeline we follow please see Figure S4a.

Foliar Na concentration in the Serengeti-Mara system is quite variable due a number of underlying geological and climate gradients, most notably the difference in volcanic soils in the SE and alluvial soils in the west. This creates the observed patterns of foliar Na as correctly identified by the Reviewer. We refer the Reviewer to L132-136 where we highlight that: *“Our work provides a detailed multifactorial analysis of spatial patterns in foliar Na at the continental scale, revealing major variation that cannot be explained by simple proxies such as distance from coastline^{9,10,31} (figure S4). While sea-salt deposition is indeed a primary source of Na, we find that local and regional geology (e.g., watercourses, Na-rich geological substrates³²) and climate also play important roles.”*

All the data from the Serengeti-Mara system are grass samples as this was the protocol for that field campaign. It would be nice to have comparable samples from trees; but those data are unfortunately missing.

I. 147: as I mentioned above, can you provide details on the identity of these 28 species and whether they are included in the larger sample of 65 species included in your GAMMs? I expect they are, but it would be nice to have confirmation in the text.

We now outline which 28 species had faecal samples collected in Table S4.

I. 148: could you include a map of the sites from Table S3? It would be useful to compare it to the maps in Figure S2 to get an idea of how closely the herbivore fecal Na sampling locations map onto the distribution of plant foliar Na sampling sites.

This is a nice suggestion. We have now integrated a map of both faecal collections (brown points) and plant collections (blue points) into Figure S7.

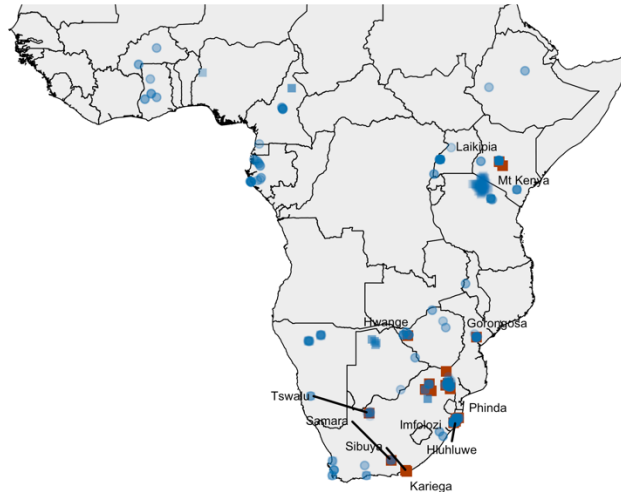


Figure R4. Map of field data for plant collections (blue points) and faecal points (brown points).

I. 187-189: this list of variables used in your GAMMs is useful but felt incomplete when, later on, you list different variables by the amount of variance explained. Table S5 is quite large and by all means belongs in the SI, but maybe you could streamline your in-text lists of variables between here and I. 205-208 so readers do not have to wonder, e.g., where does human development index come from all of a sudden.

We have integrated this helpful suggestion. As we are very limited by text space in this manuscript and cannot afford redundancy, we have removed the first list here, where we had originally tried to collect our variables into meaningful groups, and instead explicitly refer to the individual variables and how their mechanism influences herbivore density in the results paragraph (L199-204).

I. 223: I noticed you only report results from the unselective scenario for the mega-herbivore group. Is it because you implicitly assume these herbivores are not picky eaters due to their high food requirements caused by their large body size? Alternative, is this related to your later sentence discussing how you removed the degree of omnivory in diet from the mega-herbivore analyses (I. 482)? If that is so, it needs to be explained more explicitly as currently it is unclear why there are no "unselective" scenario results for the mega-herbivores

As we ran multiple model comparisons (see Table S6); however, we only report the results of the best model. For the GAMM model using all herbivore densities, the best model included dietary Na from the selective scenario as an interactive effect with body mass. For the mega-herbivore only model, the best model used dietary Na from the unselective scenario, possibly reflecting the less selective diet habits of megaherbivores (Clauss et al., 2013). We now stress this point more clearly in our revised text (L218-219).

I. 242: this sentence, particularly from "with 55%" onwards, is unclear to me. Could you rephrase it?

We have rewritten this to state: *"In general, we found that Na availability is low for wild herbivores compared to values considered optimal for domestic livestock production^{5,43} ($>500 \text{ mg kg}^{-1}$), with 55-90% of the land area exhibiting foliar concentrations $<500 \text{ mg kg}^{-1}$, depending on functional group and feeding scenario (table 1)."* (L237-241).

I. 394: "scant data", it would be useful to have a quantitative measure (e.g., %) here rather than a qualitative one---for both literature-sourced and field-collected data.

We have added details on the number of observations collected from the literature ($n=319$) and collected in our specific field campaign ($n=1037$). (L4011-415).

I. 415: typo, "seclctive" -> "selective"

Done.

I. 419: Please indicate which figures the last sentence ("All results are plotted on log10 scales.") refers to

This was confusing and unnecessary information for the methods. As we clearly outline that the axes are on log10 scale in the figure legend, we have removed this sentence.

Figures

Figure 3: in panels (a) and (b) what do the colors of the bars mean? Are they related to the colors in the partial dependence plot (panel (c))? If so, what do the yellow bars mean?

We apologise. The figure legend denoting what the colours of the bars mean (environmental, anthropogenic, functional trait, latent spatial factors, latent phylogenetic factors) was missing from our original version of this figure. We have corrected this to include the legend.

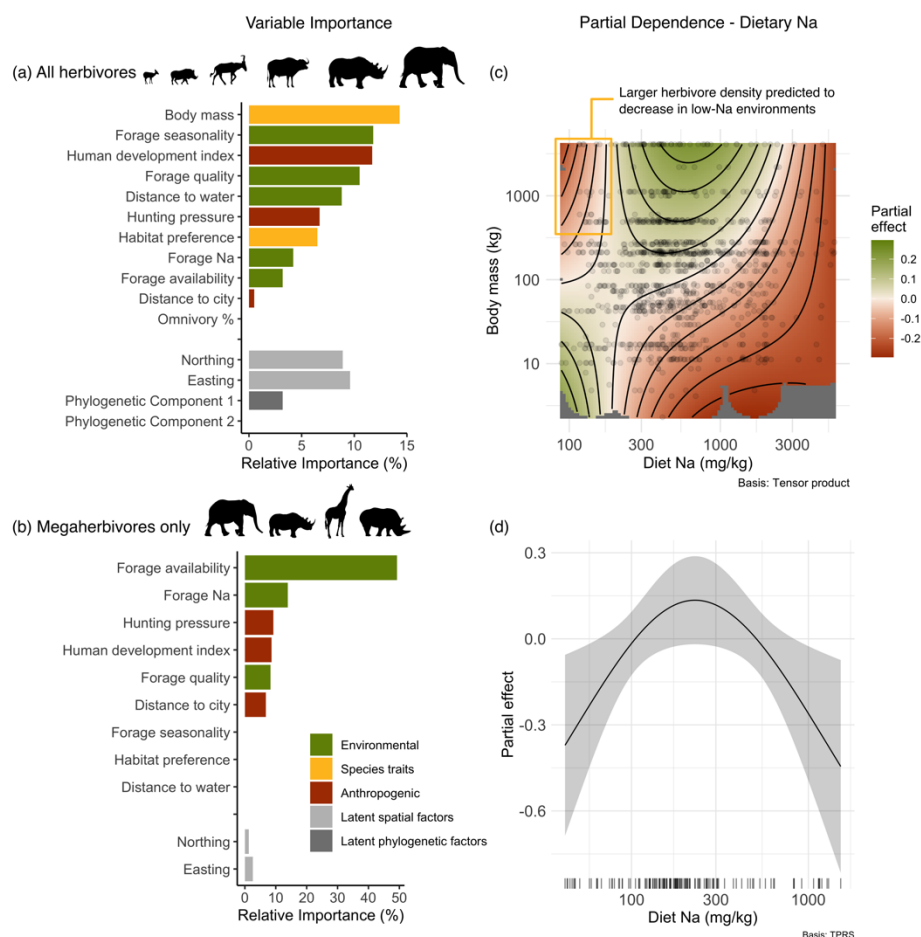


Figure R5. Updated Main text Fig 3 with legend now included.

Figure 4: is the legend on a log10 scale here as well?

The legend for figure 4 is on log10 scale for the dietary Na availability, and also for the newly included megaherbivore density points (see below). We now explicitly refer readers to this in our figure legend (L744-746).

Supplementary information comments

Given the amount of information relevant and necessary to understand the main text that the supplementary information (SI) contains, I offer a few comments on this document as well. I hope the authors will find these comments useful to improve both the main text and the SI.

I. 60-61: do you provide the references for these 16 published studies anywhere?

We now point readers to our plant Na database which includes the details of where, when and by who all data used in our study were collected.

I. 71-80: I discussed this portion extensively in my comments on the main text above, but I want to note here that it is also not clear where else in the manuscript or SI is the 'total number of observations' reported. Please, add a cross-reference here.

We have now substantially rewritten this section to ensure that readers understand where the data were collated from and what % of data were kept in the final dataset, and the proportion of this that was simulated from mean values.

I. 120-121: where these three overestimated sites included in subsequent analyses?

These sites were used to validate the modelled results of sea salt aerosol. We extracted the modelled value of sodium aerosol deposition for each plant sample. We have now modified our text to ensure that readers understand that the empirical data from these sites was not used directly to infer sea salt aerosol deposition, but simply as a validation tool (SI L136-138).

I. 159: could you list these 28 herbivore species here or in a table? Or highlight them in Table S4?

We now highlight which species we collected faecal samples from in Table S4.

I. 171: it is unclear to me what the authors mean by 'diet' here: is it % omnivory in diet, as mentioned later (I. 173), or is it diet stoichiometry? If it is % omnivory, I would recommend to use this term consistently throughout instead of switching back and forth with diet.

We did refer to % omnivory in this case and have added it in to ensure readers understand our intention.

I. 276: I do not doubt that you found substantial correlation between some of the anthropogenic factors, but I would rather see this result reported in a figure or table.

We now include a correlation matrix between all variables used to (i) model foliar Na and (ii) herbivore density – see updated figure S3. We have also included additional discussion about the possible role of double counting (see comment by Reviewer 3), but note that: *“In general, however, our variables used for modelling foliar Na concentration had low correlation coefficients (<0.3); those that had high correlation coefficients (e.g., NDVI, NPP, Aridity) were either explicitly included in our model to predict herbivore density or had negative relationships with foliar Na concentrations. Importantly, this provided an opportunity to tease apart the effects of Na and total forage availability in shaping herbivore communities at the continental scale as the limiting mechanism (i.e., not enough forage or Na) were opposed to each other.”* (SI L444-449).

I. 333: typo, "predcited" -> "predicted"

Done.

I. 355-382: this section reads very much like I. 288-305 in the main text and I am not sure why it is repeated here. I would encourage the authors to instead use this space to discuss some of the other limitations of their study (see my other comments above for some ideas).

We have now substantially revised this section and define our study limitations within: spatial data availability; temporal data availability; species-level data availability; missing covariates and modelling procedures. Please see SI L395-449 for details on each of these sections.

Table S2: reading through the manuscript, I really wanted this table in the main text rather than in the SI. Given how often you refer to your "ensemble of machine learning models", having a quick reference to them would be very useful.

Unfortunately, we are very limited by space in this manuscript, although we appreciate that it would have been useful to include this table – among others – in the main text. We have left Table S2 in the supplement.

Tables

Table S3: some entries (e.g., Gorongosa, Palabora) are missing data on latitude, longitude, or both

We have now added latitude and longitude values for all entries.

Table S4: two potential improvements here; (i) mark the species that were sampled for fecal Na data (l. 159) and (ii) mark whether a species is a 'large', 'meso', or 'mega' herbivore

We now highlight which species we collected faecal samples from in Table S4. As we have dropped the language of “meso-herbivores” we do not identify size-classifications.

Figures

Figure S1: Three comments on this figure

1. the axes should state which the base of the log used: natural, 2, 10, ...?
2. it is unclear if this figure is purely conceptual or whether data underlie the relationships shown. Also, why are ranges for the b exponent reported for "dietary Na intake" and "Na req (adult maintenance)" but not for "Na req (growing, lactating, etc)"? Maybe Duvall et al. (2023) explain this, but it should appear in the caption of this figure too.
3. the two smaller panels for Medium-Na and High-Na on the right should at a minimum state if the "Dietary Na intake" line (the more visible one) is the same as the Low-Na panel, but ideally also offer more information on the other lines shown.

This figure is purely conceptual, however the body mass exponents that we include are derived from Duvall et al. 2023, which was generated on a log10 axis. We have now updated this figure to show log10 axes. To help readers understand this we start our figure caption by saying: “A conceptual figure of theoretical drivers...”. Panels for medium-Na and high-Na are also conceptual and are simply to show that as Na availability increases, different sized animals will come out of Na deficit at different times. We now highlight this in our figure legend.

The exponent for adults at requirement and pregnant/lactating females is the same; lactating females just require twice as much Na – hence the slope of the lines is equal. We have now repositioned the b exponent label in between the lines and have added a note in our figure legend.

Figure S4a: typo in panel 3, "Neaual" -> Neural, and in panel 5, "iteraratively" -> "iteratively"

Done.

Figure S10, l. 560: what do the numbers in parenthesis refer to? Are they meant to be panel labels? If so, the figure is not showing them. Or are they referring to the rows of panels? If so, they do not match them either.

We apologise for this typo. We have now rewritten this figure caption, highlighting that the top row refers to model validation exercises using all available herbivore density and the bottom row uses megaherbivore species only.

Reviewer #2 (Remarks on code availability):

I would to commend the authors for provided all code and data needed to reproduce their analyses in a well structured, organize format. The folder provided is easy to navigate, intuitively organized, and has an informative README file. The code itself is well written, cleanly organized, and appropriately commented for easy of reading and running it.

Thank you for carefully working through our code and identifying a number of bugs. We have now rewritten our analysis code, updated R packages and checked that it runs smoothly on a number of different computers. The updated code can be found at: <https://figshare.com/s/166a498424860b2fbd8a>

Unfortunately, the code also cannot be run as provided. This is due to some of the packages ('rgdal', 'bgm') being no longer available on the CRAN nor actively supported. This means, for instance, that lines reading shapefiles and rasters, as well as the map-generating lines, cannot run and readers need to find an alternative to 'rgdal' to re-run the code. This is a significant hindrance when trying to reproduce the authors' results and needs be addressed before publication.

We have updated our code to use the recommended R packages throughout.

In the case of 'rgdal', I suggest the authors switch to either the 'sf' or 'terra' packages to run their spatial analyses, as these packages are actively supported and available on the CRAN, thus ensuring reproducibility of the analyses in the future. In my own attempts at running the code I found that substituting the functions 'rgdal::readOGR' with 'sf::read_sf' and 'rgdal::spTransform' with 'sf::st_transform' seems to fix most problems caused by using 'rgdal'. As well, changing 'ggplot2::geom_map' with 'ggplot2::geom_sf' (and changing relevant arguments in it) allows for generating the maps from the manuscript.

We have switched to using the 'sf' package throughout and made associated changes.

For package bgm, the authors should consider whether to move to the newest version available on the CRAN to maintain reproducibility.

We do not have a package "bgm" and therefore assume this a typo for "gbm". We have updated our version of "gbm"

Another issue I found was that some functions are called either before the relevant package is loaded or from a package that is not loaded at all, so running the code proceeds in halting way as one tries to figure out which package does the function belong to. This happens with function 'grid.arrange' from package 'gridExtra' (not loaded in the script) and function 'inSide' from package 'mgcv', called at l. 348-387 but the package is loaded at l. 655. Please, review the code and make sure all packages are correctly loaded at the beginning of the script to avoid these issues.

Thanks for catching this. We now load all packages at the beginning of the script to avoid this error.

l. 318-332: Plot map of sodium concentration for C3 plants throws an error with `scale_fill_gradientn()` supplying discrete values to a continuous scale (same at l. 335-345)---could not find a solution, please double check this

This section of code should run smoothly now.

l. 434 and l. 454: these comments are not describing what the following lines of code do. Each comment states the code below compared median to 95th percentile for C3 (l. 434) and C4 (l.454) plants but, in reality, the code is comparing median C4 to median C3 (l. 434) and C4's 95th percentile to C3's 95th percentile (l. 454). Please, correct these comments

We apologise and have updated the notes in this section of the script.

The 'for' loop at l. 1118-1138 did not run for me, and instead threw an error: "Columns 'x' and 'y' don't exist." The issue seems related to object 'den', which does not contain any variable named 'x' or 'y', but rather 'x2' and 'y2'. Are these the intended variables to be used here?

We have now corrected these variable names and the code should run smoothly.

Relatedly, the second 'if' statement at l. 1130 does not work and instead throws an error, "Error in .local(x, i, j = j, ..., value) : y is already present as a coordinate name!". Please, check this too.

The same problem affects the mega-herbivore-only analysis of parameter importance (l. 1195-1225). Please, check it as well.

l. 1346: the file path given to argument 'dsn' is incorrect

We have updated all file paths to the correct source.

Finally, as mentioned above, the outcome of the all-herbivore GAMM was different in my run of the code than what is reported in SI Table S7A. Unfortunately, I could not figure out whether this was due to the coding issues listed above or some other problem with the analyses. Please, double check this too.

This was not due to your coding error, but poor practice on our part. We again apologise; the new version of the code should be much easier to implement and show the correct output.

Reviewer #3 (Remarks to the Author):

This is a fascinating and well-written paper that I'm excited to see published. Generally, the nutrient requirements of animals beyond nitrogen and phosphorus – and the ecological consequences of these nutrient requirements – are understudied at scales larger than individual animal physiology, and are overdue for greater consideration. Research on sodium intake by large herbivores in particular often centers around behavior at natural salt licks or anthropogenic mineral supplementation, and far less attention has been paid to the impacts of foliar sodium on herbivore ecology and biogeography. The maps of foliar Na concentrations alone are extremely useful contributions; overall, this manuscript fills an important gap and should have significant impact moving forward. I do have some comments for the authors, including one substantive concern, that I hope will improve the manuscript and clarify any methodological choices and limitations to conclusions prior to dissemination.

Thank you for your excitement towards our paper and useful comments that have helped improve the clarity and structure of our manuscript.

My biggest question relates to the GAMMs testing whether foliar Na (and its interaction with body size) measurably improves the performance of models of herbivore density. I like this approach overall, and as a broader technical point, I appreciate the care taken to detect and avoid overfitting and predictor variable autocorrelation in each of the individual models. However, the outputs of some models (i.e. predicted foliar Na) were predictors for the other models (i.e. estimated dietary Na as a predictor of herbivore density), and there is the potential for undetectable correlation or “double-counting” among predictor variables for these models. For example, distance to coasts could implicitly be incorporated in both models (in the herbivore density model, via anthropogenic pressures, which are frequently concentrated near coasts along with centers of human population, and in the foliar Na model, via sea salt aerosol deposition), and primary productivity is also a factor in both models. Thus, certain predictors are, at least indirectly, included more than once in the herbivore density GAMMs. Granted, both of these variables may have opposite directional effects on herbivore density and foliar Na, and thus perhaps these particular examples are not of immense concern. However, it is necessary to demonstrate that this potential for “statistics on statistics” is not an issue; as a scientist with admitted “small-data” tendencies, I get nervous about drawing too many conclusions from models built from modeled predictor variables, particularly when some variables may inform multiple levels of estimates (aridity vs. distance to water, exchangeable Na & K and CEC vs. soil fertility include other examples of potentially related predictors across models). To assuage this concern, I would recommend some discussion

and assessment of correlations between predictor variables across foliar Na and density models in the Supplement. Perhaps the Supplement could also include a more basic LMM (similar to the one relating fecal Na to dietary Na; Fig S7) regressing the mean Hempson et al. herbivore density observations against dietary Na for sites with empirical foliar Na measurements, if sufficient data exist. The R² would likely be low, but it could be useful to visualize how estimated dietary Na alone relates to field observations of herbivore density.

This is a fair point. We now included a correlation matrix across all predictors used to both model foliar Na concentration and herbivore density (see updated Figure S3 and below). There is strong correlation between some variables, although most have low correlation (<0.3). Importantly, those that had high correlation coefficients (e.g., NDVI, NPP, Aridity) that could be involved in double counting were either explicitly included in our model to predict herbivore density or had negative relationships with foliar Na concentrations. This provided an opportunity to tease apart the effects of Na and total forage availability in shaping herbivore communities at the continental scale as the limiting mechanism (i.e., not enough forage or Na) were opposed to each other.

Soil pH does have a positive relationship with foliar Na, although the effects of pH on herbivore density are mediated by its influence on plant dynamics; factors we already include directly in our model using NDVI (total forage availability) and soil fertility (forage quality).

Human development index also had a positive relationship with foliar Na; a trend that may influence outcomes as it is likely that low human development index may have larger negative impacts on large-herbivore communities (see Text S2 for detailed explanation of considered anthropogenic factors and expected mechanisms). However, HDI was included directly in our model predicting large-herbivore density and thus was not vulnerable to issues of double counting.

Notably, no single variable had a strong relationship with herbivore abundance, reflecting the multiple factors that shape large-herbivore density across the continent.

Following Reviewer 3's advice, we examined if there was sufficient overlap between empirically collected data for plant Na concentration and herbivore density data. In total, there were only 54 observations for herbivore density where empirical foliar Na samples had been collected (although we note that these were not necessarily well matched in time – see R2 and R3's concerns re: inter-annual variability). However, this subset of data only represented 5 geographical locations which makes analysis very vulnerable to location-specific idiosyncrasies. Following our previous protocol, we tried generating a generalised additive mixed model with this dataset. However, these models were poor; the fixed effects (i.e., dietary Na availability) only explained 5% and 7% of the variation in herbivore density for the median and 95th percentile scenarios respectively, but with dissimilar relationships between models. Given the small sample size, we do not feel that this is robust analysis and are wary of drawing any conclusions. We have decided not to include it in our supplementary material. Future assessments that link mineral concentrations and herbivore densities directly in the same reserves will help improve rigour of such analyses. We now make this point in SI L450-452.

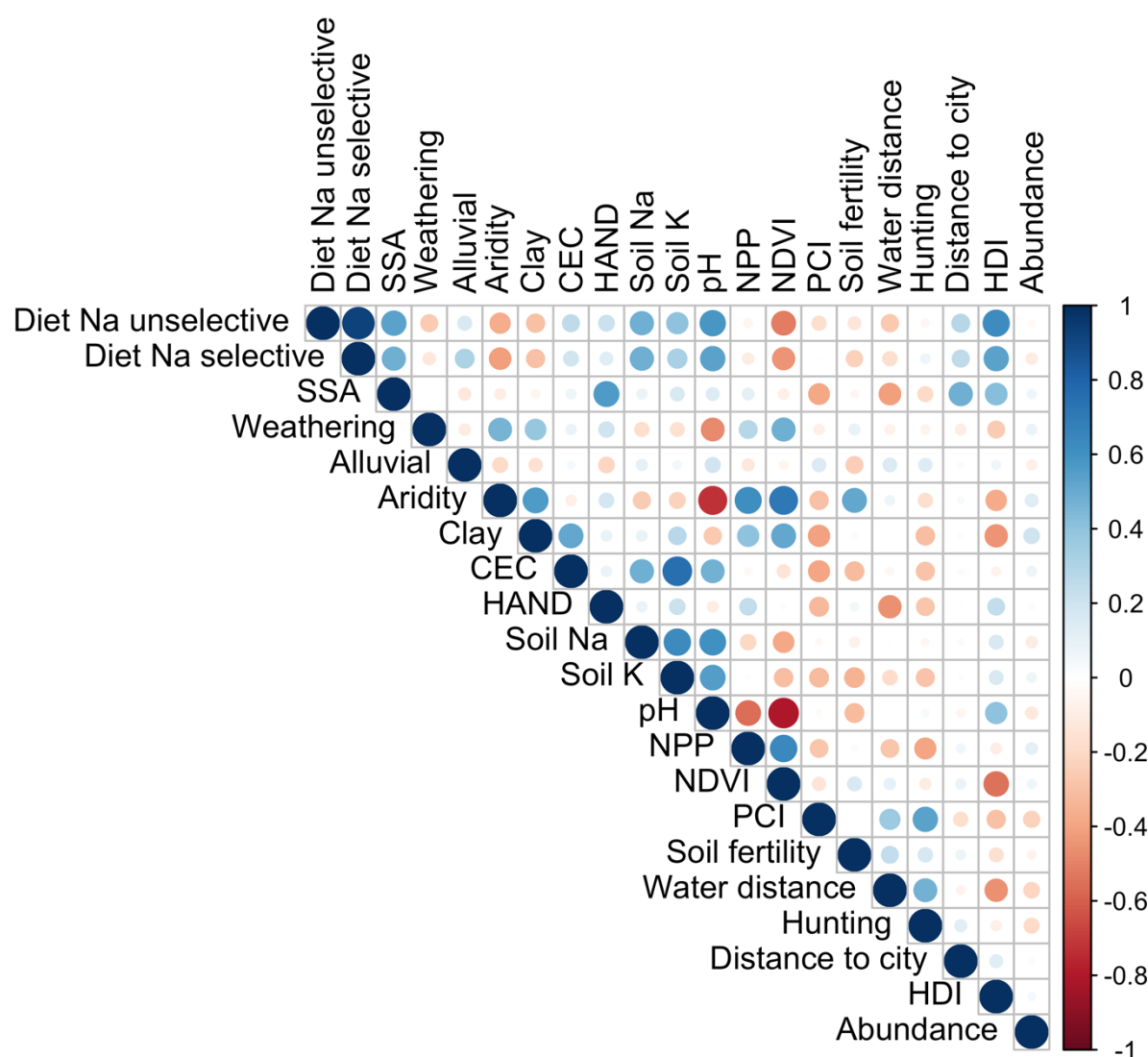


Figure R6. Correlation matrix for all variables used in predicting foliar Na concentration and large-herbivore density.

Was any consideration given to the year or season of Na sampling in the analysis? 1994-2021 (the period of direct foliar Na sampling by the authors) is a relatively long period of time, and presumably the data sourced from published literature span an even wider range of time. I agree that spatial differences in foliar Na resulting from climatic, edaphic, and biogeographic factors across sampling sites probably have a greater effect than single site year-over-year variation. However, foliar sodium levels are not static and can vary strongly across points in a growing season and across years as environmental conditions fluctuate; the timing of sampling across studies may influence how representative values are of the foliar sodium available to herbivores more generally in each habitat. [Edit: I see this is addressed in Text S3; however, I think it is of sufficient importance to mention at least briefly in the Discussion. Is there any reason to believe that foliar Na tends to be measured during certain points in the growing season? Could Na limitation shift seasonally across space or plant type?]. I tried to investigate this myself with the data provided and found something is funky with the Sach et al. 2020 foliar Na dataset – month and year both have strange, impossible values. However, when these data are removed it does seem that many more observations were collected in June and August compared to other months.

This comment falls in line with similar suggestions made by Reviewer 2 (see above).

Seasonal and annual differences can influence foliar Na concentration and have been quantified at a number of site-specific studies (e.g., Curtis et al. 1970). Na tends to accumulate passively throughout the growing season and decadal-scale nutrition dilution due to climate change is a potentially an important factor for lowering foliar Na concentrations (Kaspari and Welti, 2024). However, such intra- and inter-annual changes are typically much smaller than the four orders of magnitude differences we observe across sub-Saharan Africa. In this first iteration of mapping Na concentrations across the continent, we therefore decided to focus on the major spatial patterns; however, we acknowledge the role of temporal variation and now emphasise this point in the discussion (L294-296) and supplementary information (SI L414-422).

In time, we hope that our dataset can be bolstered to try and examine such temporal nuances. Thanks for catching the issue with the dates for the Sach et al. (2019) data. We have now corrected our dataset to when these plant samples were collected: Sept 2017.

Regarding Figures 2a and S7 – could these be presented in non-log space? Studies (in this journal no less!) have shown that log-transformed data are extremely difficult for people to interpret, including trained ecologists (Menge et al. 2018). Particularly for the selective scenario (Figure 2a), the R^2 is high and the relationship compelling; however, it appears (on the log scale at least) that the relationship could be non-linear and in fact there is a threshold of foliar Na (somewhere around 1000 mg/kg) at which fecal Na concentrations saturate. This is roughly the same threshold at which dietary Na concentrations shift from having a positive to negative impact on megaherbivore density as per Figure 3d – might some greater consideration of this threshold be fruitful?

Thank you for this keen observation. We now fit generalised additive mixed-effects models to allow for potential non-linear fits, with dietary Na as a fixed effect and genus as a random effect. This does indeed improve model performance and has a threshold point (correctly identified at ~1000-3000 mg/kg in the selective dietary scenario).

We have however, kept our figure in the main manuscript with log10 axis as our main focus is on low values and the extreme upper values in our dataset (e.g., >5000 mg/kg) obscure the ability to see patterns at low dietary Na intake when on linear scales. Instead, we now highlight the importance of interpreting these plots in log10 space explicitly in the figure legend (L718).

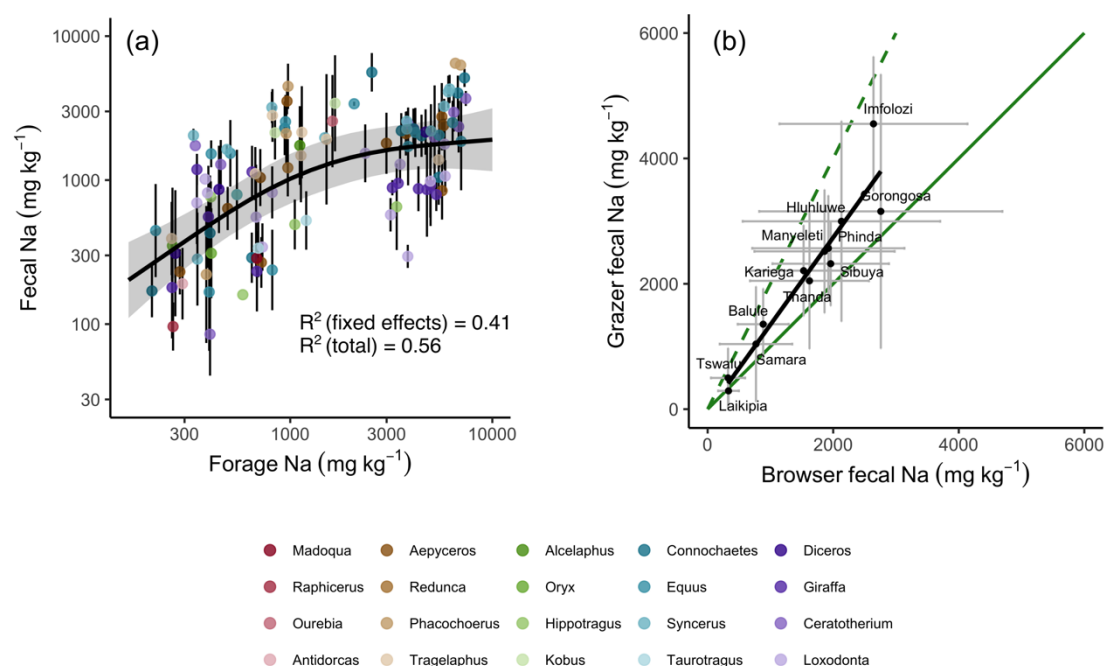


Figure R7. Updated version of Figure 2.

Line 201 – Based on table S6, should this not be dietary Na from the *selective* scenario?

Thanks for this catch. We have corrected this to “selective”.

Lines 235-237 and similar points in Discussion and Figure 4 – This is an extremely neat observation. Any insights from the Na model as to what factors are driving lower foliar Na in these areas? The Discussion later mentions soil infertility and other previous hypotheses explaining the lack of megaherbivores in these parts of western Africa, but does not detail why foliar Na could be lower here (greater distance from the Eastern coast? However, this does not lead to reduced Na at similar distances in southern Africa). This seems of particular importance given that elephant, giraffe, and black rhino pre-colonial ranges encompass these areas. Has something changed? Is Na availability here lower than it was historically? Generally, given the amount of space dedicated to this discussion of “missing megaherbivores”, it could be helpful to incorporate some representation of current megaherbivore densities by species into Figure 4. Are elephants and giraffes less dense in these areas with high potential for Na deficiency? Is there any reason to believe that has changed since colonization, or do intrinsic local factors render such areas consistently ill-suited to high densities of megaherbivores?

The low foliar Na concentrations in parts of Central and West Africa are indicative of low Na inputs and high leaching rates. The prevailing wind direction in West Africa does not blow sea salt aerosol Na as far into this part of the continent, while there is an absence of geologic inputs of Na through rock weathering (see Figure S3). We now emphasize this point in discussion (L256-257).

The idea of adding the empirical megaherbivore census counts into Fig 4 is a good one; we now include these which highlight that megaherbivore densities are generally low in the low-Na environments and that fewer surveys have been done in these regions (presumably due to challenges of keeping viable populations megaherbivores in these landscapes).

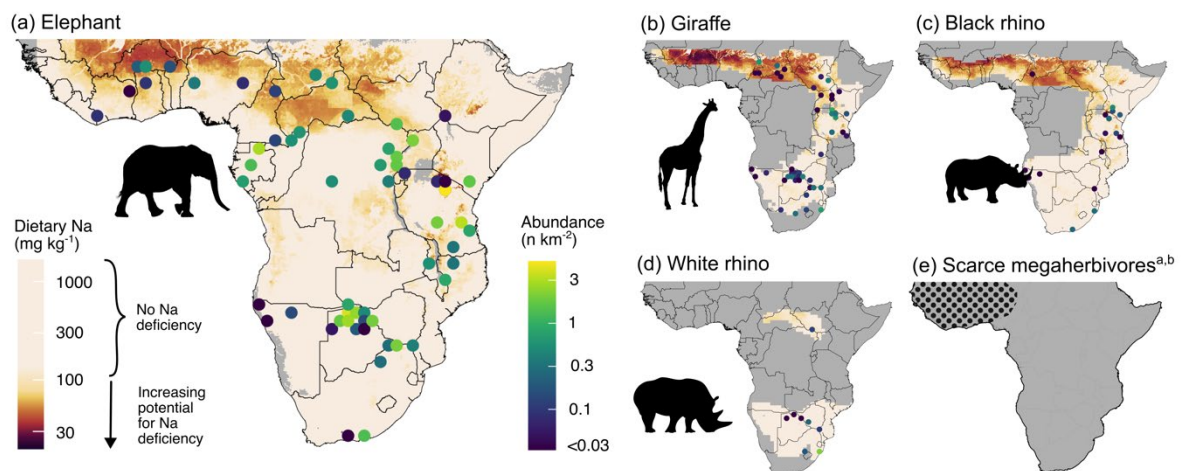


Figure R8. Updated main text Fig 4 showing megaherbivore abundance. We now state: “*Note: both the dietary Na concentrations (mg kg^{-1}) and megaherbivore density (n km^{-2}) legends are plotted on a $\log_{10}$ scale to highlight region where low plant Na constrains megaherbivore abundance.*” (L744-746).

Lines 254-255 – I’d also reference Figure 3c here.

Done.

The Discussion does a great job of elegantly explaining the results and implications while giving plenty of room for caveats and the limitations of these findings.

Thanks for your insightful comments; they have greatly helped improve our manuscript.

- Julia D. Monk

References:

Menge, D. N. L. et al. Logarithmic scales in ecological data presentation may cause misinterpretation. *Nat Ecol Evol* 2, 1393–1402 (2018).

Reviewer #3 (Remarks on code availability):

With a few tweaks to account for different R versions and package retirements/updates (see below), I was able to run the code and reproduce the results and figures with little trouble.

We're glad to hear that our code was accessible. We have now updated and refined our code further based on helpful comments both here and by Reviewer 2.

For publication, it would be helpful if the authors could update to use the sf package since rgdal has now been retired. I understand analyses were conducted on a prior version of R, but for future users who may want to replicate the analysis or learn from the methods applied, sf spatial functions have greater longevity (it didn't take me that long to replace these functions when reviewing the code).

We have now updated our analysis code and checked this on various computers to ensure compatibility.

For the Dataset column in multiple datasets, it could be helpful to put everyone's full name, in case others looking to collaborate on similar projects want to be able to contact the relevant scientists (I know that Roo is Andrew Abraham, but not everyone does!).

Good catch. We have now corrected this to include all data contributors' full names.

Reviewer #4 (Remarks to the Author):

The text is clearly written, the figures are beautiful, the manuscript is pleasing to read.

Thank you for your positive assessment of our paper.

I've been asked to provide comments on the spatial mapping component of the study.

Broad scale mapping of ecological variables using ML technics has become a common task nowadays, although there is no standard & widely-accepted procedure to perform the assessment of broad-scale mapping model's predictive performance – which is problematic. Among the numerous studies that have performed such task in the past years, I find the present study to be particularly transparent, rigorous and comprehensive in the way they performed and presented their model validation. I have no major (negative) criticism on the spatial mapping component of the manuscript, and will just address two comment to the authors.

Thank you for taking the time to check our spatial mapping. We are happy to hear that our methods were clear and comprehensive. Following your insightful suggestions for improvements, we now emphasise the quality of our foliar Na model predictions in the manuscript and include a purely spatial model for baseline comparison. Further, we now also include R^2 as an additional model performance metric to compare across models, which highlights that our ensemble model is consistently the best performing model for predicting foliar Na concentrations.

First, as a naïve reader who does not know much about foliar Na, I am fairly clueless after reading the main text of your manuscript and the supplementary data on whether or not your mapping model(s) are effective at mapping foliar Na. In the main text (section “Patterns of foliar sodium across sub-Saharan Africa” – L. 99), there is no comment about mapping model predictive performance. You simply state that you mapped it. In the Methods either. One needs to go to the Supplementary – in particular to the beautiful figure S4b – to see some RMSEs.

We now highlight that: “As the data was right-skewed, we trained models on $\log_{10}$ foliar Na concentration. The ensemble model was consistently the best performing model with robust prediction accuracy ($R^2 > 0.35$; $RMSE < 0.4 \log_{10}$ units); although model performance decreased with distance from training points across all models tested (Figure S4).” (L100-103).

But again, as a naïve reader I’m not sure how to interpret RMSE values – not only because there is no unit (add one), but also because it is not directly evident to me if ex. 0.6 is small or large. I’m a bit skeptical about it : are we in the same unit as in fig. S2B & C ? It would be surprising, because it would mean the mean prediction error (ex. $RMSE = 0.5$) is about $\sim 5\%$ of the lowest observed value ($\sim 10 \text{ Mg. Kg}^{-1}$)? This seems fishy. A suggestion is to (i) add a second Y axis to the scatterplots in Fig. S4b with RMSE values in % (which would be easier to interpret), and/or (ii) add scatterplots of associated R^2 - so that readers have some clue on the share of variance explained by models. If presented RMSEs are indeed of the order of 5%, then I definitely suggest adding R^2 values.

As our foliar Na concentration data was right-skewed, we modelled the $\log_{10}$ concentration. We have updated our manuscript and figure of the methodological pipeline we used (Fig S4a) to better reflect this.

Following Reviewer 4’s helpful suggestion, we have now added RMSE values (i.e., $\log_{10}$ units) into our plots in Figure S4b (see below). We now also show R^2 values of model performance to highlight predictive capacity; in the cases of both the median and 95th percentile scenario, the ensemble model is the best predicting model while all models showing decrease performance with increasing buffer distance (see Ploton et al., 2020 for similar effects).

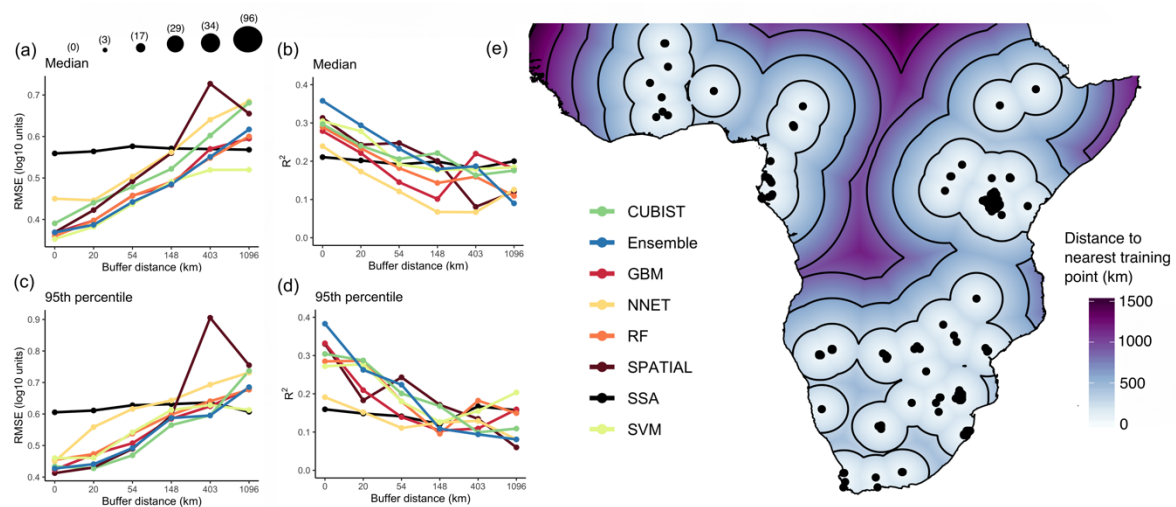


Figure R9. Updated Figure S4b.

Second, in my opinion, your model predictive power assessment would have been complete if you’ve had compared your models to a purely spatial model i.e. a model based on observations geographical coordinates alone. It would be very easy to add the trend line of a purely spatial model to Fig. S4b (together with all other models). This would provide some context / a baseline to help the interpretation of your models RMSE values.

We have additionally generated a purely spatial model and included this in our updated Figure S4b. We fitted a kernel smooth to our foliar Na concentration points based on their geographical coordinates using the ‘spatstat’ package. We set sigma (the smoothing parameter) to 0.5 degrees resolution ($\sim 55\text{km}$) to account for the continental-scale of our dataset and 0.1-degree groups of our foliar Na concentration data.

The purely spatial model was similar in performance to some of the individual machine-learning models but was consistently worse than the ensemble model. We now comment on these results extensively in a new section in the Supplementary Results (SI L312-326).

Minor comments

Table S2 : row of “RF”, mtry ranges from 1 to 289 ? Isn’t there only 11 predictors? Why 289?

Thanks for catching this; we have updated this to 11 to reflect the 11 predictors used in our random forest model.

Cited references:

Curtis, C. R., Lauver, T. L., & Francis, B. A. (1977). Foliar sodium and chloride in trees: seasonal variations. *Environmental Pollution (1970)*, 14(1), 69-80.

Duvall, E. S., Griffiths, B. M., Clauss, M., & Abraham, A. J. (2023). Allometry of sodium requirements and mineral lick use among herbivorous mammals. *Oikos*, 2023(9), e10058.

Kaspari, M., & Welte, E. A. (2024). Nutrient dilution and the future of herbivore populations. *Trends in Ecology & Evolution*, 39(9), 809-820.

Olf, H., Ritchie, M. E., & Prins, H. H. (2002). Global environmental controls of diversity in large herbivores. *Nature*, 415(6874), 901-904.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., ... & Péliissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature communications*, 11(1), 4540.

Pranzini, N., Maiorano, L., Cosentino, F., Thuiller, W., & Santini, L. (2024). The role of species interactions in shaping the geographic pattern of ungulate abundance across African savannah. *Scientific Reports*, 14(1), 19647.

Sach, F., Dierenfeld, E. S., Langley-Evans, S. C., Watts, M. J., & Yon, L. (2019). African savanna elephants (*Loxodonta africana*) as an example of a herbivore making movement choices based on nutritional needs. *PeerJ*, 7, e6260.

Santini, L., Benítez-López, A., Dormann, C. F., & Huijbregts, M. A. (2022). Population density estimates for terrestrial mammal species. *Global Ecology and Biogeography*, 31(5), 978-994.

Welte, E. A., Kuczynski, L., Marske, K. A., Sanders, N. J., de Beurs, K. M., & Kaspari, M. (2020). Salty, mild, and low plant biomass grasslands increase top-heaviness of invertebrate trophic pyramids. *Global Ecology and Biogeography*, 29(9), 1474-1485.

Reviewer #2 (Remarks to the Author):

Thank you for the opportunity to review a revised submission of this manuscript. I reviewed the original submission and I enjoyed reading through this revised version. Overall, I found the manuscript much improved. The authors did an excellent job of revising the manuscript and incorporating the comments of all reviewers. Well done! I have no further major comments on this work.

Matteo Rizzuto

Thank you for the careful review of our paper. We especially appreciate the coding checks, which we have modified accordingly and uploaded a revised R script to our Figshare repository.

Reviewer #2 (Remarks on code availability):

Code remarks

We checked all code using latest version of R studio to ensure that all code works with latest packages and systems. The final code and data is now published at our Figshare repository.

I commened the authors on the extensive rework of their code to update the R packages used, expand on several of the analyses reported, and improve the organization of their script. I was able to reproduce the results reported in both the manuscript and the SI. Below, I report a few minor concerns and issues I had with the code, which I am believe can be easily fixed.

* when producing maps of sodium concentration for C3 and C4 plants (l. 339-348 and l. 353-362), the `fill` argument of `geom\_tile` needs to be set to `ensemble\$pred` otherwise the code fails (at least on my machine)

This appears to be machine specific as this code works on a range of computers checked with the code as is – no change.

* line 757 is supposed to prepare data to produce a correlation map across variables in the model, but the three variables called in the `which()` function are not part of the `den` dataframe (i.e., "age", "BP", "cholesterol"). They are not in either version of the Herbivore_density_data dataset

This was an accidental error and these variables are indeed not included in our model. This line of code has been deleted.

* l. 988 and 989, the `geom\_text` call to add labels does not work and throws an error:

Error in `annotate()`:

! Problem while setting up geom aesthetics.

! Error occurred in the 4th layer.

Caused by error in `list\_sizes()`:

! `x\$label` must be a vector, not an expression vector.

this is apparently a relatively recent issue (<https://github.com/tidyverse/ggplot2/issues/6622>), so no fault to the authors but just thought I'd flag it in case by the time the paper is published a workaround has been found or the authors want to change how the R² and RMSE information is presented in these graphs (this affects all other `geom\_text` calls in the cross validation plots)

Thanks for raising this. We have amended the `geom\_label` call accordingly.

* when fitting model megaherbivore_selective_Na_gamm (l 1561-1586), there is an extra call for "s(x, y, bs = "so", xt = list(bnd = bound),sp=2) +" in the bam() function call, which leads the model fitting to failure. Once removed, the results match Table S7B

Thanks for this catch – the repeated line has been deleted.

* when drawing partial dependence plots, the function call for ``draw()`` generates the following error:

Error in ``map()``:

i In index: 14.

Caused by error in ``filter()``:

i In argument: ``is_inside``.

Caused by error:

! ``..1`` must be of size 10000 or 1, not size 0.

Backtrace:

1. `gratia::draw(all_herbivore_selective_Na_bodymass_gamm)`

23. `dplyr:::dplyr_internal_error(...)`

Unfortunately, I could not find a solution for this error but I thought I would flag it in case the authors have better luck.

We did not experience this error when checking our code on multiple laptops.

Reviewer #3 (Remarks to the Author):

I thank the authors for their incredibly thorough and careful consideration of all reviewer comments, and consider the manuscript much improved. I hope to see it published soon, and have just a few final (minor) suggestions to improve the strength of the conclusions and the clarity for readers.

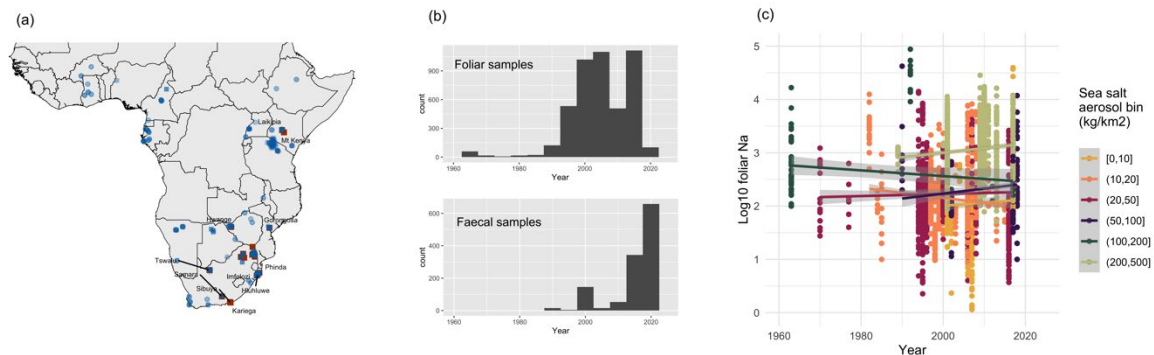
Thank you for your support towards our paper – it's greatly appreciated.

1. I am not fully satisfied by the discussion of the issue of differences in sampling timescales for foliar Na and fecal Na. Figure R3 shows the histograms of year of collection for foliar and fecal samples, and indicates that most collection of foliar samples happened after 2000, which aligns with other covariates. However, a significant proportion (roughly 1/3?) of data were still collected prior to 2000; furthermore, most fecal samples were collected in 2018 and afterwards. Thus, while the peaks of these two histograms are similar, their distributions are very different; thus, in analysis, it remains likely that many, if not most, fecal Na concentrations are being regressed against foliar Na data collected some 10-50 years prior, or against simulated values based on those data. More concretely, it would be helpful if the supplement included a plot showing the relationship between year of sampling and foliar Na concentrations for the entire dataset. This would reveal whether there are regional trends in foliar Na concentrations over time (sensu Kaspari and Welte 2024, as the authors discuss), or if year of sampling bears no relationship to foliar Na (which would reduce the need for concern); this would also allow readers to see for themselves, as the response to reviewers indicates, that this variation is much smaller than the spatial variation in foliar Na presented in the manuscript. Similarly, was there any relationship between the temporal and spatial clustering of these leaf data (i.e. were samples collected in similar parts of Africa in the same years, even if not at the same site)? Seeing these considerations in the supplement would increase the reader's confidence in the relevance of linking foliar Na and fecal Na data that were collected across different time periods and using that relationship to infer the impacts of foliar Na availability on dietary Na deficiencies and species distributions.

This is an important point, which we have followed up on. In our updated Supplementary Information, we now present the histograms of foliar Na samples and faecal Na samples together to highlight temporal incongruities between sampling efforts. As the Reviewer highlights, due to the potential issue of climate induced nutrient dilution, this may lead to a gradual decrease in plant Na concentrations over time.

Following the Reviewers suggestion, we have plotted foliar Na concentration through time for different regions of Africa based on sea salt deposition (our most important predictor of foliar Na). This plot reveals that the majority of the variance is indeed explained by spatial and phylogenetic factors. However, we note that: while time does not appear to be the primary determinant of foliar Na (i.e., there is not a consistent pattern of decreasing foliar Na within each region) and thus any temporal component is not likely to significantly impact our assessment of foliar Na as a driver of faecal Na at the continental scale, increasing the size of the current dataset and more rigorous analysis may elucidate temporal patterns such as seasonal, episodic extreme events (such as drought or cyclones) and CO2 induced nutrient dilution.

We add these points to our Supplementary Discussion under *Temporal data availability*. Please see below for our updated Fig. S7.



2. I appreciate the careful consideration of “double-counting” in response to my previous review. The correlation matrix in Figure S3a should indicate which variables are included in the foliar Na predictions and which are used to model herbivore density.

We have now included which variables are used in which model into a revised Figure S3a.

Line 295 – typo; nutrient vs. nutritent

Good catch!

- Julia Monk

Reviewer #4 (Remarks to the Author):

I read with attention the rebuttal letter of A. J. Abraham et al. and the modifications made to their manuscript. The authors fully accounted for my comments, by notably adding (1) an additional baseline model (ie. a purely spatial model) and (2) an additional metric (ie. the R^2) to facilitate the assessment of the predictive power of their foliar Na concentration prediction models. I am satisfied with the validation approach of the mapping models, which I think is robust and transparent.

Thank you for your insightful suggestions; we are happy to hear that our methods were robust and transparent.