**Subject editor**

I read the manuscript with great interest. Quantifying and mapping sampling bias is indeed a critical step for all studies that use species occurrences records. The sources of geographic biases that the authors try to estimate their impacts (airports, roads, rivers) are well known and discussed both formally and informally in the literature.  
[REPLY]: Some kind response

However, as far as species distribution models go, the true bias is not in the geographic distribution of occurrences, but on the environmental space. The pretty maps built by species distribution modelers tend to hide an ugly fact: species distribution models are not spatially explicit. So, I think the first important topic that the authors have missed is that estimating geographic sampling bias is not a good diagnosis of potential statistical biases in predictions of species distribution models, as a biased geographic sampling may, or may not, be biased in environmental space. There is a substantial literature and software tools on environmental biases in species distributions, and I think those should be accounted for in the manuscript.

[REPLY}:

* Add a paragraph on existing tools with environmental bias, including references to the suggested literature
* Change the text a little to go away from the species distribution moelling, an stress that our tool, provides an intuitive estimate of the biasing effect, which is comparable among bias types and datasets.

For example, suppose an island with a high mountain in the center and a highway around the island (say, Tenerife). There may very well be many species at lower altitudes that do not occur at higher altitudes, because of the environmental gradient caused by the central mountain. In that case, the "geographic bias" may truly reflect the distribution of the species, as roads and airports would also be at lower altitude.

[REPLY} Yes, good point, but:

* We estimate the effect on record number not species number
* Yes, one should preferably use sampbias within one biome

Quantifying sampling biases is, of course, of great importance. However, the software tool and the software note leave an important question untouched. If I have a geographically biased dataset, then what should I do? Of course, the immediate answer is to organize a field expedition to the area that is undersampled. However, what do authors think modelers should do? Sub-sample areas that are close to roads and airports?

[REPLY]: Yes, of course, that is THE question. I do not know. I think it is expected to much to solve this, but we should have suggestions. Any ideas? In general, I think we should move the text away from the modelling and more sell it as a way to quantify bias, and compare it across datasets.

**Reviewer: 1**

The paper presents a software to quantify the effects of different sources of accessibility bias in species occurrence data. The method estimates sampling rate for cells considering the distance of the bias factor by using Bayesian inference. Although the software implements only one method for bias quantification, I understand that the implementation of such method is a first important step to the quantification of bias accessibility in occurrence samples. Below I describe some general issues that should be taken into account in order to increase its potential.

1- When calculating the decay in sampling rate for multiple factors, the final effect is a cumulative function of each bias. However, as you stated in the text, some biasing factors can be correlated, i.e., corresponding to similar accessibility bias. Thus, considering them as cumulative could exacerbate the real effect of the sampling bias in a cell. If my rationale is true, would be useful to have additional options to combine the effects or to consider the correlation among them or at least provide an analysis of such correlation.

[REPLY]: OK, Daniele will it be possible to include this somehow? E.g. test for correlation, or set a flag to use the different factors cumulatively?

2- The definition of the study area as the geographic extent of the dataset could hampering the use of the package, as most studies have a predefined geographic area for interest. Thus, the usability of the package would be increased if the study area can be defined by the user.

[REPLY]: Yes so far it was possible to limit the extent to a convex hull around all empirical records, but I will include an option to limit them to a user provided polygon.

3- In Fig.2, what means negative values of sampling rate in a cell? Also, the legend name would be more informative as “estimated sample rate”.

[REPLY] Daniele, can you add an explanation to the figure caption

4- Finally, I understand that the implementation of such method is a first step to the quantification of bias accessibility in occurrence samples, but it would be great that in future versions the package implements other methods of bias quantification such as those of Ruete, 2015 and Fernández & Sakamura 2015.

[REPLY]: I am not sure here. This sounds like a lot of work, I am unsure about the gain. The first one is implemented in MatLab, and the secondone partly in R, but in a very weird way, using some other kind of App. I can contact the authors and ask them if they were interested in implementing their stuff, but otherwise I am not sure this is worth it. What do you think?

**Reviewer: 2**

Comments to the Author

The manuscript addresses a highly relevant topic, and presents a very important new tool for the analysis of biodiversity, biogeography and macroecology.

[REPLY] Thank you

Only a few points should be explained in the text for the article to be suitable for publication. Only one point should be improved in the methodology to make it more precise (grid cell). These points will be addressed in the sequence:

1 - (page 7) is the grid generated by sampbias of equal area per grid? Or do you use degrees as a measure? I suggest that the grid be made with cells of the same area This problem can be more serious at larger scales. I think this is a crucial change to the method. If this is already done (cells of equal area), make this clear in the text.

[REPLY]: Currently it is a lat/lon grid, since the spatial extent is anyways limited. To implement this, we can switch to an option of a user-provided raster.

2 - Explain in the text the reasons for the priors' choices. In addition, make it clear whether the user can change the parameters and the prior calculation method.

[Reply] Daniele.

3 - Is there a minimum limit of samples (species points) for the analysis to be performed? What should happen if a database with few samples is used in a very large study area? I think this can be shown in the text.

[REPLY] I suggest we do a simulation with series, with and without bias and an increasing number of records, to see, what the effect is, pick an empirical threshold on this and then sen a warning if there are less than this threshold of records.

4 - (page 11) make it clear what each parameter of the script represents in the equation presented in the text

[REPLY] I’ll clarify this in the text.