General comments

Zhao and co-authors present very interesting world-wide statistical analysis of soil physical and chemical characteristics variability in respect to climatic drivers. They create "biome" plots and maps of soil properties. The selected statistical approach is very innovative, which makes the manuscript sound and worth publishing. I am very skeptical about the direct causality. This issue is, however, relatively well covered in the discussion section of the manuscript. Nevertheless, there is one issue that is not covered in the discussion section at all and that is very important according to my opinion. Authors completely ignore soil orders and associated soil horizons. They report data for top 30 cm of soil. These 30cm can contain either single horizon or several very different horizons with very different physical and chemical properties.

My major concern surrounds the results presentation. Authors need to provide more information about the various statistical analyses they used to make Figures 4 – 6 and they also need to clarify various threshold they defined. The manuscript often contains either very vague or very strong statements unsupported by the results (see specific comments). For this type of presented results I think it is especially important to present uncertainty in quantitative terms. Any potential user of the extrapolated maps information/database should be aware of the limitations.

Specific comments:

Page 1, Lines 16 – 17: What is the "critical MAP for the transition from alkaline to acidic soil"? This is some model parameter?

Page 1, Line 18: two dots

Page 1, Lines 18 – 19: I do not understand the meaning of the last sentence of the abstract. Can authors clarify its meaning?

Page 1, Lines 21 – 23: This is very vague statement that deserves more clarity.

Page 1, Lines 25 – 26: Again, the statement is very vague. I would suggest more specific statement.

Page 1, Lines 29 – 30: What doesn't mean "soil stewardship for societal well-being"

Page 2, Lines 25 – 26: Please check the superscripts.

Page 2, Line 30: Third and very important soil-forming factor is the bedrock. I think authors should mention it right away in the introduction, not only in discussion section.

Page 3, Line 20: Please specify the type of pedologic data that GSD contains. Way these data wasn't used in analysis?

Page 4, Line 4: Authors excluded 10% of all observations (those above and below 95% and 5% quantile respectively). Is there a specific reason for that? In the case of this dataset, it is very difficult to identify outliers. First 30 cm of soil can include one or several different soil horizons with very different physical and chemical properties. Values identified as outliers might be very likely correct and reflect the difference between different soil horizons of different soil orders. Quick look on Fig. 4a suggest to me that all peatlands were very likely removed from the dataset. For that reason I would strongly suggest to keep all data in the dataset.

Page 4, Line 18: Please check the superscript.

Page 4, Lines 24 - 25: Can authors clarify the uncertainty estimation of C to N ratio? The uncertainty of the ratio composed of two variables, each with its uncertainty, should be calculated differently than the uncertainty of a single variable.

Page 4, 2.4. Statistical analysis: Based on presented results, this section requires more detailed information. Most importantly, authors should clarify the statistics reported in figures 4 – 6. See also the specific comments bellow.

Page 5, Line 15: Can authors also report here calculated total amounts of SOC and STN (Tab. 1)? It would be very interesting to compare this estimate with previous estimates in respect to chosen statistical approach in the discussion section. The estimates reported in Tab. 1 doesn't seem to me very different from previous estimates. Does it mean that the approach selected by authors is not so different in terms of the outputs?

Page 5, Line 25: "soillayer"

Page 5, Line 26: The "saturation curve" is mentioned here for the first time. Why saturation curve, what does it mean and how it was calculated/estimated? All that should be thoroughly explained in the statistical analysis section. Also explain the term "saturation threshold".

Page 5, Line 29: Authors should also explain the difference between "cold" and "warm" climates. How was the temperature threshold defined? How the arbitrary selected threshold affects the results?

Page 6, Line 11 - 12: Again, this is very strong statement. Given the issues surrounding the true causality in the found relationships discussed in this section I would suggest to make a less strong statement.

Page 6, Line 30 - 31: The transition definition and calculation is not explained in statistical analysis section nor reported in results section. It is impossible to review these results without detailed explanation.

Page 7, Line 10: Please check the subscript.

Page 7, Line 12 - 20: C to N ratio does not represent very good proxy for substrate quality so the discussion is very speculative at this point.

Page 7, Line 21 - 24: This is very speculative. Isn't the correlation simple given by the fact that plant derived organic material always contain C and N no matter what the limitation is?

Page 8, 4.4 Uncertainties in mapping soil properties at the global scale: Can authors quantify the uncertainty? According to Tab. S4, statistical model explains sometimes less than 20% of variability. In addition, 10% of all data were removed. I believe that the uncertainty is very important to state unambiguously so any potential user of the database and maps knows the limitations.

Page 8, Line 28: "Ourregion-specific"

Page 8 and 9, Conclusions: Again, very strong statements unsupported by the analysis.

Tables:

I would suggest to show only R^2 in table S4. It would improve the table alignment and reading.

Figures:

Fig. 4: I have concerns regarding MAP and MAT thresholds. Authors should clarify their definition and use. Specifically, arctic have typically low MAP but because of low MAT, their soils are very often water saturated.

Fig. 5: Please explain the part c of the figure in statistical analysis section.

Fig. 7: Results presented in this figure also require more information. How it was calculated? Was the increase of explained variability by a specific explanatory variable compared to null or some standard model?

Fig. 8: I did not find any reference to this figure in the main text.