The review of the PhD-thesis

MGR. MARTIN MACEK
PATTERNS AND PROCESSES IN SPATIAL DISTRIBUTION OF PLANT SPECIES ACROSS SCALES

The thesis consists of six chapters, of which the first chapter is the general overview and summary of the thesis, and the remaining five are individual papers – three of them already published in respected journals, one submitted and one ms. (which is in the form that generally permits immediate submission). Of the five papers, Martin Macek is the first author of three (two of them already published), and second author of remaining two. His contribution to all the papers is substantial (and clearly stated in the text). The set of the five papers is rather heterogeneous, nevertheless, the author has done good job in the first chapter connecting the various papers and justifying their interconnection (in this way, the author has demonstrated his ability to be a PI of large projects in future, and his ability to write appropriate reports for the granting agencies).

Striking common feature of all the papers is the use of sophisticated methods of data analysis, and their (in my view appropriate) application for testing of biologically relevant hypotheses. With this respect, the use of GIS for predicting the climate and it variability on small spatial scale and the use of the climatic variables to understand the variability and plant communities (Chapter 3) are typical examples. The three papers analyzing the altitudinal gradient data in the Himalaya form a coherent set, and I like the sophisticated tests of various hypotheses, and I highly appreciate the insightful use of various null models. I would also stress that the papers deals with topics that are timely and of interest to many ecologists, from very practical questions related to contentions about the natural regeneration in the bark-beetle affected forests, to rather theoretical aspects of the species distribution along altitudinal gradients, which is also a hotly debated topic in recent literature.

In summary, I am convinced that the author demonstrated his ability of scientific work and particularly, that his analytical skills will be highly valued in any ecological team engaged in use of complicated data sets for testing ecological hypotheses. Consequently, I am convinced that awarding of the PhD is fully warranted.

In České Budějovice, July 2nd, 2020 Jan Lepš

The papers present interesting topics and approaches, some particularly worth of discussion. I have first some questions for which I want to have the candidate's explanation during the defense, and finally, at the end, I list few minor comments of rather technical nature, that I hope might be of some use to the author.

Questions to be answered

Chapter 2

It seems to me that the elevation is the most important factor determining the success of natural regeneration. Mortality is apparently independent of the elevation, but if we measure success of regeneration by comparison of pre-outbreak density and density of seedlings, the regeneration is not so successful above 1300 m a.s.l. What is the cause? Might it be the higher cover of *Callamagrostis villosa* there? (On the other hand, the patchy distribution of spruce individuals close to the treeline might be quite nice and corresponding to natural conditions, so we perhaps do not need to be worries so much about lower success of regeneration there.)

Chapter 3

I know that everything is about correlations, no experiment is feasible. You explain the improvement of the fit based on fine scale temperature data (as compared to WorldClim) by response to temperature. I would have another explanation. The fine scale temperature is better correlated with topographic characteristics (particularly SWI, TPI). I am inclined to believe that the effect on plant composition is not the direct effect of temperature, but simply temperature is correlated with SWI, but SWI is also reflecting soil water (or other soil characteristics, and similarly for TPI) – consequently, what is in the statistical model effect of temperature on a small scale, is in reality direct effect of soil water (or other soil characteristics). Would it be possible (for the set where both temperatures and topographic characteristics are available) do a variation partitioning between temperature and topography?

General statistical question inspired by Fig. S4: the role of outliers. I generally hate excluding points, because they are outliers, and consider it to be a dangerous procedure, particularly, if we want to use tests. However, some outliers can change the results considerably, particularly for small data sets. The low correlation of Tmin5 between 2017 and 2018 is caused to a large extent by a single point. Moreover, this 2017 value is the only observation during the whole 4 year period below 5 (with generally very few observations below 6 – see also the outlier in the 2017 histogram). I would be inclined to check the (mal)functioning of the datalogger or similar possibilities. Or is there some other explanation?

Chapter 4

The problem here is that the elevation is NOT an environmental gradient, but just surrogate for real environmental characteristics, that are correlated with elevation (but everybody, including myself, uses elevation as environmental gradient). I would consider the fact that the elevation is good predictor of temperature, but rather poor predictor of moisture (due to the patchiness of moisture) as the most probable, most parsimonious a sufficient explanation for the asymmetry. Do we really need the other explanations? At this point, I am rather surprised that there is only one third skewed, I would expect more. And I do not accept your explanation that it is because of penalization in AICc (without it, the more complicated model would be always better). According to my opinion, the symmetric curve is highly improbable, and the rejection of the null hypothesis of symmetry might be a matter of number of observations. I thought about other possibility. Would it be possible to predict the skewness on the basis

of different R^2 for the prediction of temperature and moisture? This might be a new null hypothesis against which you can test your data, or even have the symmetry and this value of skewness as two possibilities and then use some Bayesian approaches to decide between them. Could you comment on this possibility?

Chapter 5

In chapter 4, you argue that Ladakh is optimal for the study of species elevational distribution because of sparse vegetation – because of this, the species distribution along altitudinal gradient is determined mainly by climatic factors. On the contrary, in chapter 5, you demonstrate that the lower elevation limit is (partially) caused by competition – so which of the views is correct (I would subscribe to the second one). And I also like your discussion with John Lawton regarding details – without them, the bold first-order phenomena might be pretty misleading.

Chapter 6

If I see it correctly, then MDE reflects purely neutral (non-biological processes), whereas MPA includes some biological processes (what else would prevent the species optima to be evenly distributed along the gradient studied). If you agree with this, I would recommend stressing it more in the paper. Then, I see the basic question whether there is a maximum of the richness within the domain, or whether the function is monotonically decreasing (naturally, there must be zero above the limit for vascular plants. The first thing that I would do (with the dataset you have available) would be to test for such a maximum in the dependence of species richness in the 100m x 100m plots as a function of elevation. If there is a maximum within the domain, then the decreasing part of the curve toward lower elevations would enable some biological interpretation (physiological, species pool). Is there any problem with such a simple analysis?

Minor and specific comments

Although part of the papers has been already published, in the thesis, the layout of all the chapters is unified (and not simply copied from the final form of the papers), which looks nice. However, more attention could have been paid to keep all the subscripts and superscripts properly typeset.

p. 72 - the commonly used abbreviation for principal coordinate analysis is PCoA, whereas PCA is classical name of principal component analysis (mixing these two might cause confusion).

p. 78 Table 2 I know that the software provides the significance for intercept, but the null hypothesis that the intercept is zero is pointless.

p. 117 I hate the description of statistical analysis in the form “we used a xxx R package (citation)”, even though it is more and more common in published papers. I would appreciate just a single sentence what you have used, using this package (probably) the generalized linear models, with (I guess) binomial canonical distribution (because you have just presence absence data, and I guess that eHOF is also appropriate for quantitative data, thus it must enable other distributions).

p. 164 –be careful with the statistical terminology. The distribution function is by definition non-decreasing – what is decreasing is the probability density function (probably, also in Fig. 5 it would be for the red line better to use the probability density function, not just probability function, which is the term that I do not know).