

Report on Sensitivity analysis of CMI algorithm

Mariona Roig{e}

Contents

1	The Match Climates algorithm for pest risk analysis (PRA)	1
2	Sensitivity Analysis of the CM algorithm	2
2.1	Climate similarity projection	2
2.2	Case study using the species <i>Mentha aquatica</i>	2
2.2.1	Number of AWAY points	2
2.3	Number of HOME locations that obtained each CMI values . . .	5
2.3.1	The CMI = 0.7 threshold value	5
2.3.2	Species specificity of the 0.7 threshold	7
	References	9

[TITLE]

1 The Match Climates algorithm for pest risk analysis (PRA)

The climatic similarity between two locations can be used to inform a key question in pest risk analysis (PRA). PRA aims to assess the risk that a specific species poses to an area of interest (PRA area). One way to infer whether the species of interest could become established in the PRA area is by comparing the climate of abroad locations where the species is present to the climate of the region of interest.

The match climates algorithm (Sutherst and Maywald (1985)) allows the user to compare the meteorological data from different places. For a given pair of locations, it answers questions such as: Are the extreme minimum and maximum temperatures similar? Do the two locations have similar amounts of rainfall and similar seasonal rainfall patterns? The numerical output of the algorithm is an index that ranges between 0 (poorly matched) and 1 (perfectly matched).

Climate matching has been widely applied in PRA to estimate the potential geographic distributions of weeds, insects and mites. However, there is a lack of studies that address its predictive validity. A first attempt at evaluating the

predictions of CMI for its use in PRA is found in Phillips et al. (2018), which evaluated whereas the predictions of climatic similarity for a list of 35 invasive ants in New Zealand were consistent with its establishment status. Phillips et al. (2018) concluded that using the results of CM seemed to be informative to predict ants invasions in New Zealand. However, few caveats were identified and recommended for further study.

2 Sensitivity Analysis of the CM algorithm

2.1 Climate similarity projection

A specific form of using CMI is to construct projections for a PRA area where the species of concern is not present yet. A projection is a model-derived estimate of the climatic suitability of an area of interest.

The steps to build a projection for a PRA area (HOME) are: 1. Obtain the presence points where the species is known to be present abroad 2. Divide HOME into a grid of cells with their respective climate data 3. Compare the climate of the AWAY points with the climate of the HOME locations using the CM algorithm 4. Obtain for each of the HOME climate cells, a CMI value (from 0 to 1) that represents the **maximum** climatic similarity obtained between the HOME point when compared to all the AWAY points.

2.2 Case study using the species *Mentha aquatica*

Here we perform a sensitivity analysis of the match climates algorithm to the number of overseas points used in the construction of the HOME projection.

To do so, we chose the species *Mentha aquatica* because of its large number of overseas presence points (189,835). Then we built 1000 HOME projections, starting with a sample size of 100 AWAY points and increasing it by 100 each iteration until reached a sample size of 100,000. The analysis took over three weeks time to run.

2.2.1 Number of AWAY points

More AWAY points allows the algorithm to compare the climate data of each HOME point to more potential AWAY matches thus increasing the likelihood of obtaining a better match (higher CMI). Therefore, we expect that a larger number of AWAY points will increase the average CMI obtained in the HOME projection.

The results of *Mentha aquatica* case study show how the number of AWAY points and the CMI have a positive relationship (Figure 1).

- (1) Scatterplot of number of away points and mean CMI



Figure 2: Relationship between number of away points and mean CMI

At the same time the variability of CMI values (measured by their sd) seems to decrease with higher sample sizes (number of AWAY points used in calculating the HOME projection) (Figure 2).

(2) Scatterplot of number of away points and variability of CMIs

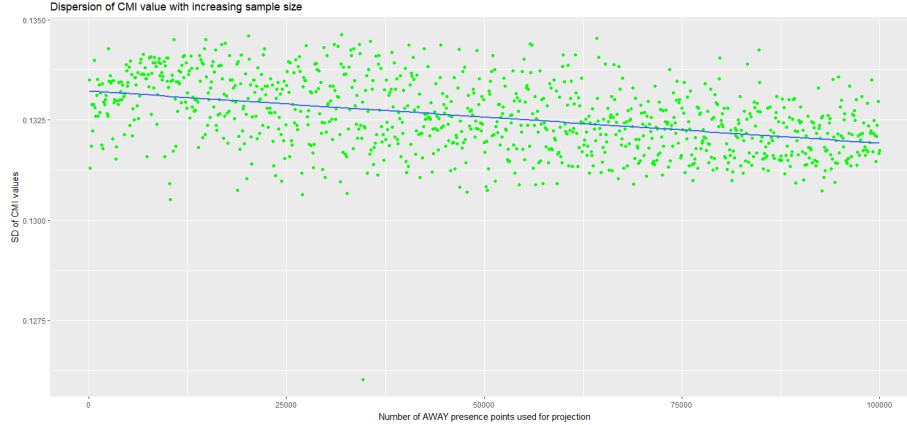


Figure 3: Decrease of CMI variability with higher sample sizes

An interesting question is whether there is a minimum amount of AWAY points that ensures a stable projection. Figure 1 seems to show a plateau after 50,000 AWAY points. This result suggests that a very large number of AWAY points are required to build 'good projections'.

A consequence of using over 50,000 AWAY presence points is that it increases the likelihood of including the Influential AWAY points in the sample. Influential

AWAY points are those that become the best match for a big number of HOME locations.

The importance of the **influential** AWAY points

The influential AWAY points are those overseas points that have matched the most with HOME locations when constructing the projection.

Figure 3 shows that the greater number of AWAY points used, more AWAY points the projection contains. Consequently, projections built using few AWAY points will be very sensitive to the quality of their most Influential AWAY points.

- (3) Relationship between the number of best matched points and sample size

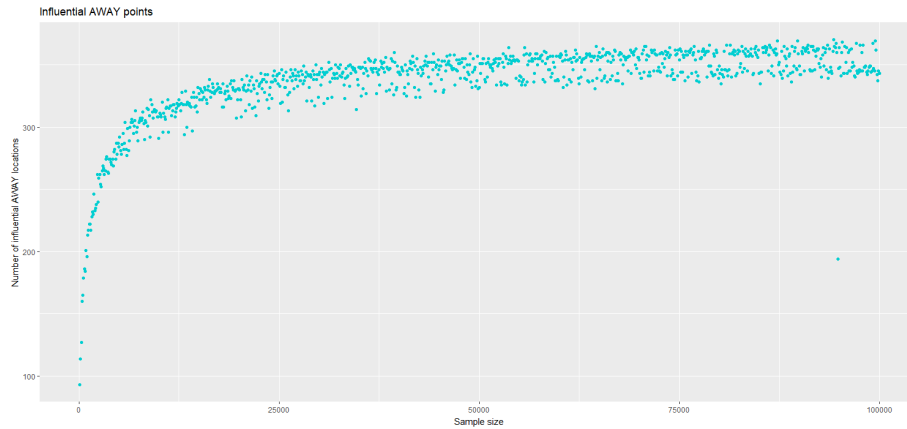


Figure 4: Scatterplot

2.3 Number of HOME locations that obtained each CMI values

2.3.1 The CMI = 0.7 threshold value

Because CMI ranges between 0 and 1, the 0.7 CMI threshold is generally interpreted as indicating two climates are sufficiently similar for a species to persist in both. However that threshold is rather arbitrary and potentially species specific.

Each facet plot below shows the number of New Zealand locations that obtained each CMI value for a given sample size. The red vertical line shows where the 0.7 CMI value lies and how many locations are over it.

- (7) Boxplot ??? Mariona: I don't think I understand this box plot, or how it relates to the facet plots.

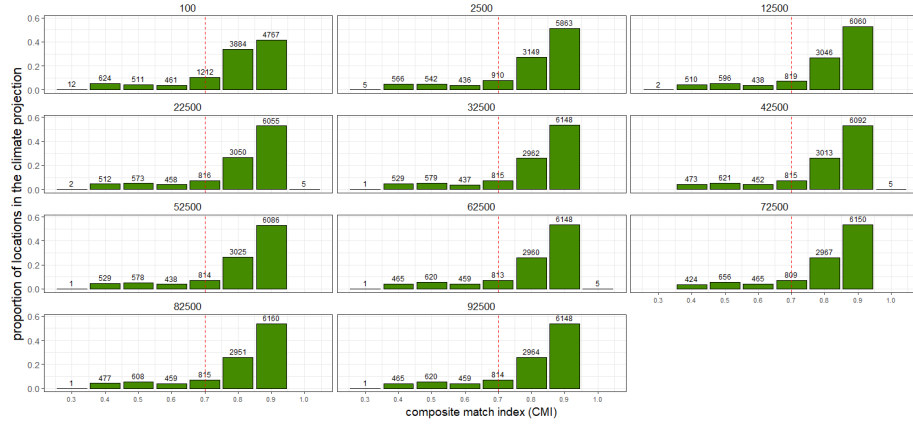


Figure 5: Facet plots for 11 sample sizes from 100 to 100,000

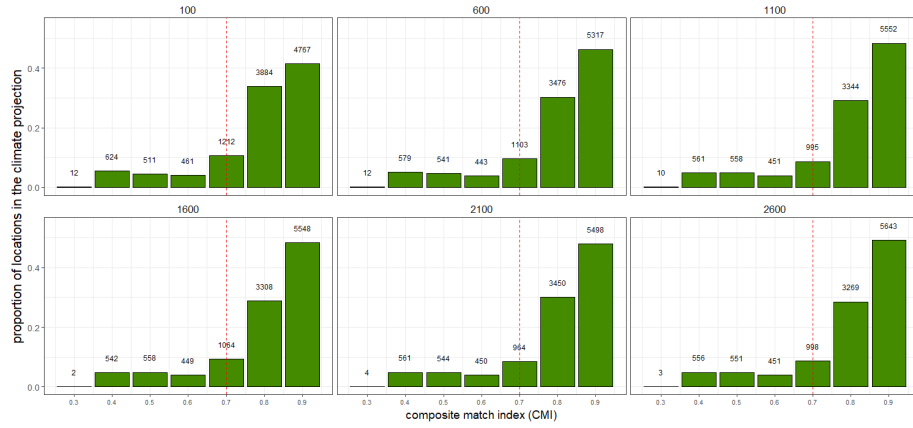


Figure 6: 'Zoom' to sample sizes from 100 to 2500

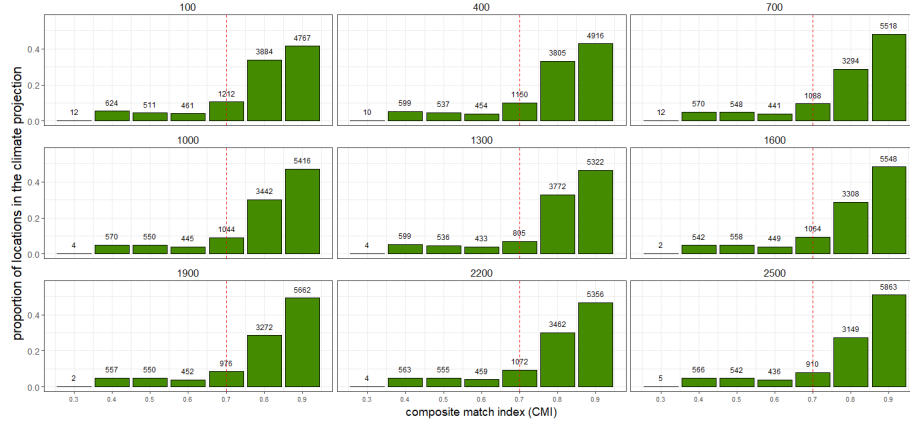


Figure 7: More detailed ‘Zoom’ to sample sizes from 100-2500

2.3.1.1 Proportion of HOME to AWAY locations

(Extracted from Phillips *et al.* 2018) The number of HOME locations with climates favourable for establishment will generally increase with the proportion of CMIs over 0.7. However, the total number of overseas climate cells from which the proportion for each species was calculated must also be considered when interpreting such data. For example, *Nylanderia vaga* (Forel), *Strumigenys xenos* Brown, *Pheidole rugulosa* Forel and *P. proxima* Mayr had CMIs from only three or fewer non-New Zealand locations. This suggests their distributions were under-represented by the data and, had CMIs for more locations been available, some would likely have been less than 0.7, particularly for tropical or subtropical species such as *N. vaga*. Thus, the proportions of CMIs $C \geq 0.7$ could be too high for these species.

2.3.2 Species specificity of the 0.7 threshold

A potential way to evaluate whether the 0.7 threshold is suitable for the species of concern is to calculate the climatic variability of the species known abroad range.

A species with all its occurrence points climatically similar to each other will probably have a narrow climate niche. In that case, a 0.7 threshold value might be too low. On the other hand, a species with a very broad climatic range could indicate the species is ongoing a process of niche expansion, for which then a 0.7 threshold could be too high. This will also be related to the number of AWAY points used to calculate the projection (more points, more probability of a wider climatic range).

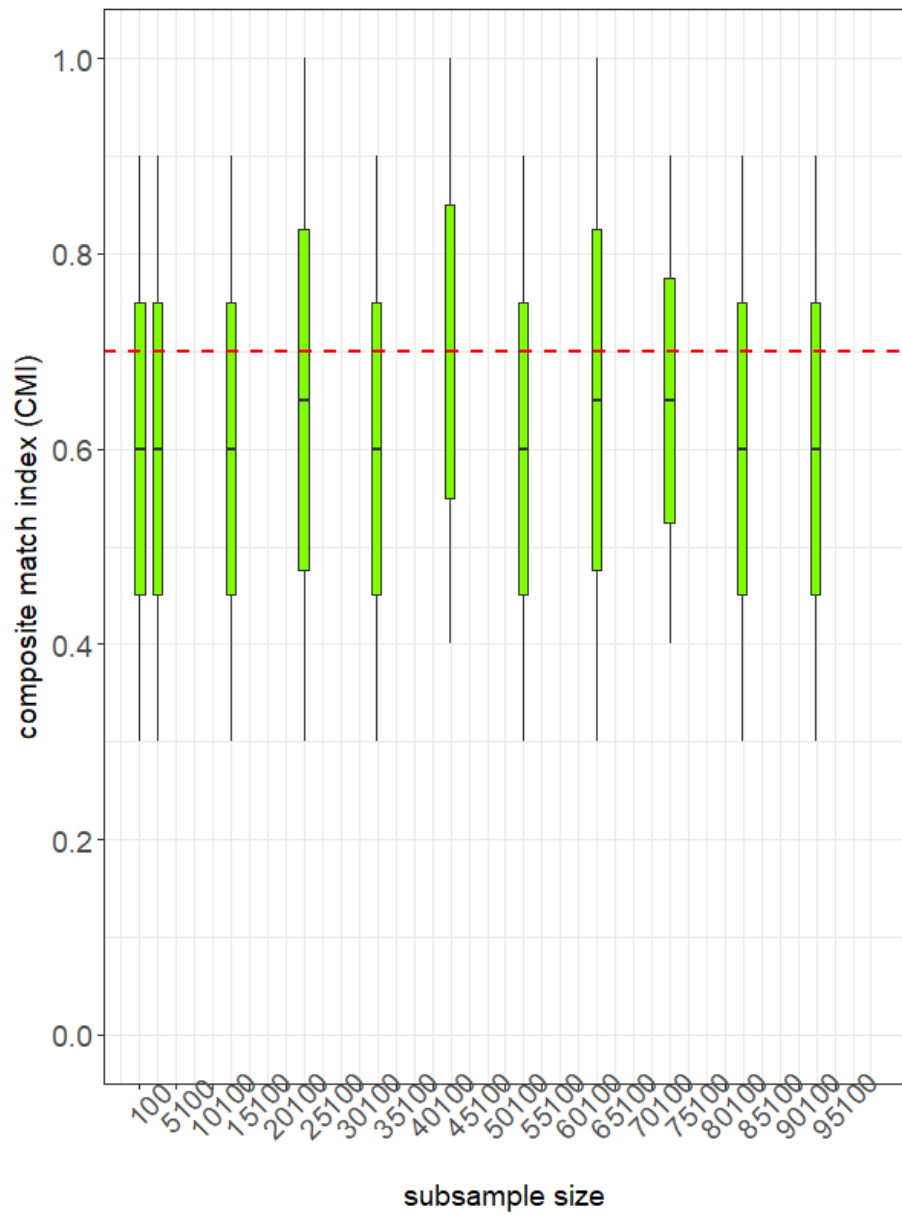


Figure 8: Box plot of the proportions of CMIs over 0.7

References

- Phillips, Craig B., John M. Kean, Cor J. Vink, and Jocelyn A. Berry. 2018. “Utility of the CLIMEX ‘match climates regional’ algorithm for pest risk analysis: an evaluation with non-native ants in New Zealand.” *Biological Invasions* 20 (3). Springer International Publishing: 777–91. <https://doi.org/10.1007/s10530-017-1574-2>.
- Sutherst, R. W., and G. F. Maywald. 1985. “A computerised system for matching climates in ecology.” *Agriculture, Ecosystems and Environment* 13 (3-4): 281–99. [https://doi.org/10.1016/0167-8809\(85\)90016-7](https://doi.org/10.1016/0167-8809(85)90016-7).