Report on Sensitivity analysis of CMI algorithm

Mariona Roig'{e}

Contents

[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 maixmum 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 Sensitiviy 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

Hypotheses: More AWAY points mean more one to one comparisons HOME vs AWAY are gonna be carried, increasing the chance of finding a higher CMI value as a best match for each HOME location.

1. Do more away points increase average CMI in the HOME projection?

2. Do more AWAY points decrease the variability of the CMI values obtained in the HOME projection?

3. Is there a minimum amount of AWAY points that ensures a stable projection?
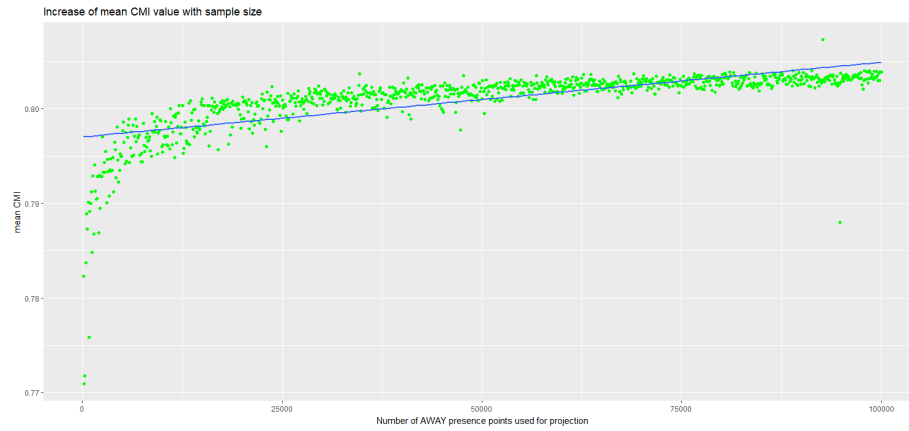
START: We want to know the climatic similarity of NZ to the non-NZ range of a species. E.g. *Mentha Aquatica*

mentha_aquatica_non_nz_cms.RData
Mentha Aquatica overseas presence points and CMs
R object name: pq_cms
Type: data frame. 169035 obs. of 10 variables

WorldWithoutNZDataByWeek24Apr2018.RData
World Climate data
R object name: A
Type: data frame. 3062160 obs. of 10 variables

01_get_nz_projections.r
i) Import M.A overseas presence points and world climate data
ii) Take a random sample of size 'm' of M.A overseas presence points
iii) Filter the world climate data to the locations where the species occurs using the climate location IDs in Mentha_aquatica_non_nz_cms.RData

Mentha_aquatica_PpClIData.csv
Climate data for the non-nz locations where M.A occurs.
The climate locations are 0.5° × 0.5° grids.
R object name: op_sub
Type: data frame. 'm' obs. of 10 variables

01_get_nz_projections.r
iv) Call bat file RunAllNzProjection.bat
v) snooze and wait while RunAllNzProjection.bat runs

ClimateMatch_V3.exe
i) Run CMI between the 'm' MA presence points and all the 11474 location points within NZ climate data

NZDataByWeek28Mar2018.csv
New Zealand climate data
Type: data frame. 684493 obs. of 13 variables, when met. 11478 locations

RunAllNzProjection.bat
i) Call the application ClimateMatch_V3.exe
ii) Import NZ climate data NzDataByWeek28Mar2018.csv
iii) Tell ClimateMatch_V3.exe to run in 7 cores and name its outputs

01_get_nz_projections.r
vi) Finish the snooze time and call JoinOutFiles.bat

out1.csv   out2.csv   out3.csv
out4.csv   out5.csv   out6.csv
out7.csv

JoinOutFiles.bat
i) Imports all the out* csv files
ii) join them in one arly object
iii) outputs them as Mentha_aquatica_all_nz_proj_i.csv where i is the loop number in case we are trying different values of m

Mentha_aquatica_all_nz_proj_i.csv

01_get_nz_projections.r
vii) Read the output of JoinOutFiles.bat
viii) Tidy it so that it contains only the info needed as HOME projection
ix) Save it as *.csv file

Mentha_aquatica_all_nz_proj_i.csv
Best CMI between all NZ locations (596493 points) and the 'n' random overseas MA-pp selected when subsampling. Also called 'HOME projection'
R object name: hm_proj
Type: 11471 obs. of 6 variables
i refers to the loop number if we are evaluating different values of 'm'

02_combine_all_projections_in_one_rdata_file.r
i) Import 'i' number of Mentha_aquatica_all_nz_proj_i.csv files
ii) Tidy and combine all i files into one dataframe
iii) Save as an R object

02_nz_proj_cms_DATE.RData
R object that contains a dataframe summarizing all the home projections obtained over i iterations.
R object name: proj_comp
Type: 11469438 obs. of 11 variables

fGetColorsToPlotsAClimaticNcrProjection.r
R function

fStatIndicesOfCharfdString.r
R function

fGgplotNzClimaticNcrProjection.r
R function

03_make_maps_of_NZ_projections.r
i) Import 2_nz_proj_cms_DATE.RData
ii) Use function fGgplotNzClimaticNcrProjection.r to plot maps for each i

04_summarize_cms.r
i) Import 02_nz_proj_cms_DATE.r
ii) Descriptive statistics
iii) Calculate facet plot CMI proportions per species
iv) Make boxplot of overseas CMI per species
v) Calculate and plot influential away points

NZ_outline_and_admins_sho.RData
R object name: nz_borders
Type: Shapefile with the outline and NZ administrative borders

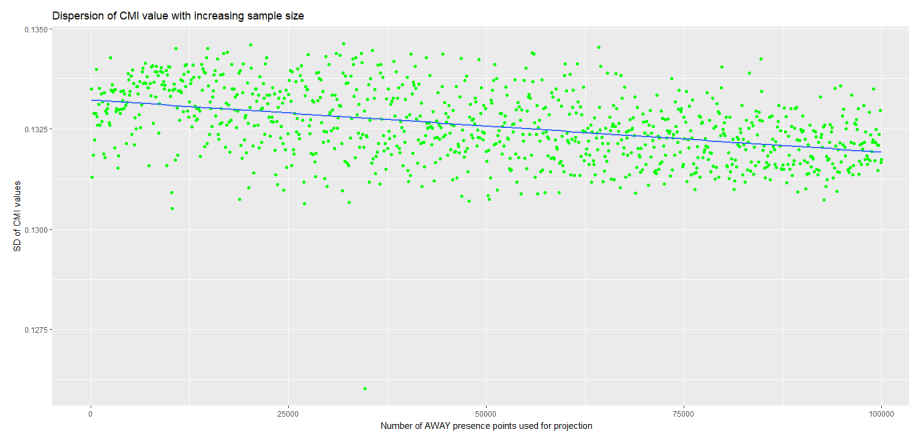INTERPRET PLOTS

Figure 2: Regression line

Figure 3: Regression line

2.2.2 Number of HOME locations that obtained each CMI values

2.2.2.1 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.
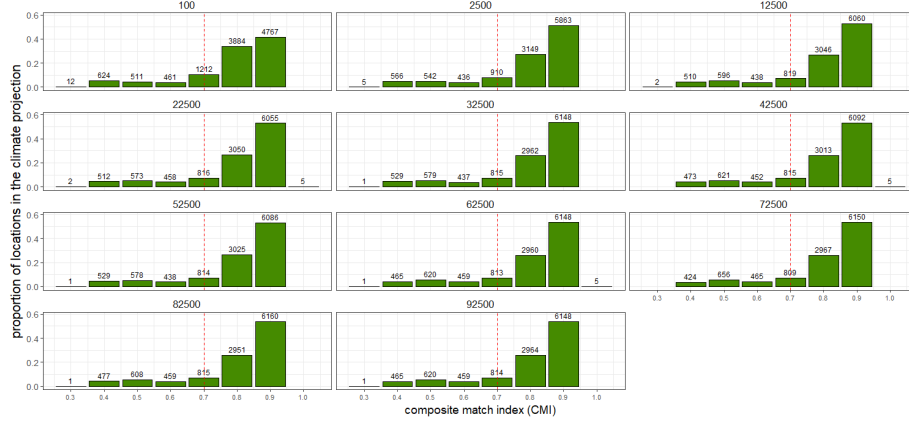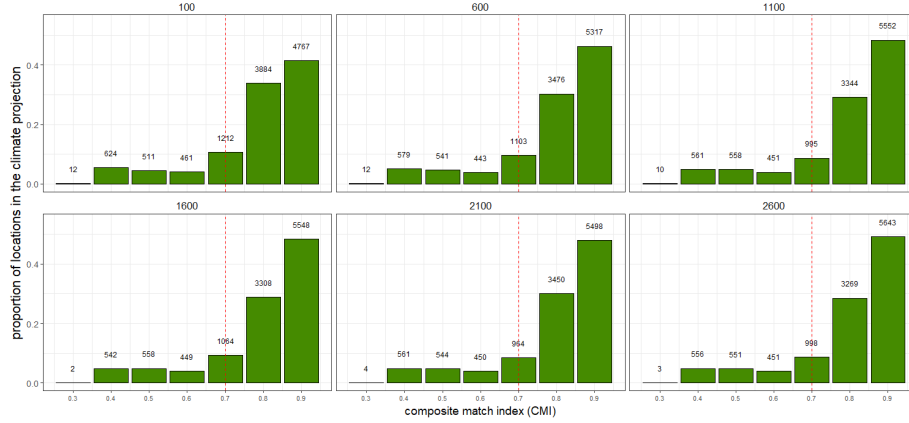

Figure 4: Facet plot

Figure 5: Facet plot

2.3 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.

5
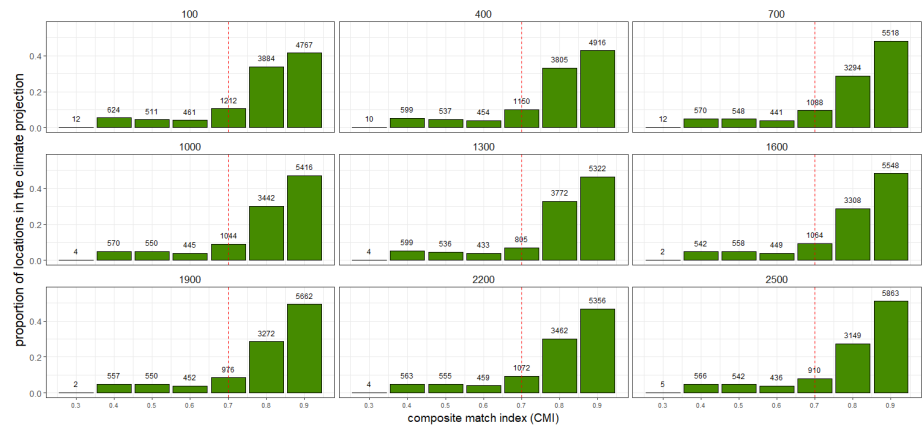
Figure 6: Facet plot zoom
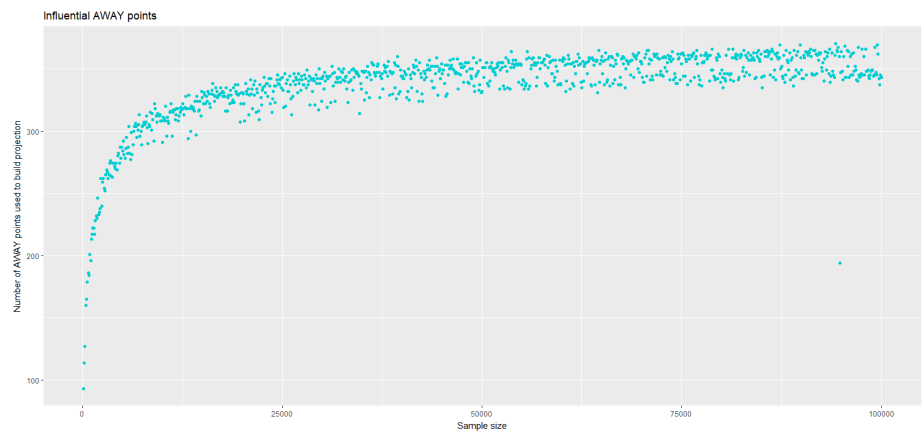
Figure 7: Scatterplot

6

2.3.1 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 (Fig. 3). 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 (Fig. 3). 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 0.7 could be too high for these species (Fig. 3).

2.3.2 Species specificity

Imagine a species with all its occurrence points very similar to each other. You could assume that species has a very narrow climate range, therefore maybe a 0.7 threshold value is too low and we would be overconsidering the threat it poses to the HOME location. On the contrary, imagine a species that has a very broad climatic range (which could be indicative of niche expansion) and its best match with HOME location is 0.6. Wouldn't it be wise to consider the species for further PRA since it has potential to keep expanding its climatic niche? One solution could be to correct the CMI threshold by some variability measure of the species climatic niche?

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).

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.