Multi-model analysis for projecting the global distribution of *Halyomorpha halys*

Senait D. Senay

International Science & Technology Practice & Policy (InSTePP)

University of Minnesota

Senait D. Senay[1, 4&6], Craig B. Phillips[3, 4], Ursula Torres[2], John Kean[3, 4], Robert C. Venette[5] and Susan P. Worner[2, 4]

[1] International Science & Technology Practice & Policy (InSTePP), Dept. of Applied Economics, University of Minnesota, 1994 Buford Ave., 248E Ruttan Hall, St. Paul, MN 55108, USA
[2]BioProtection Research Centre, PO Box 85084, Lincoln University, Lincoln 7674, New Zealand
[3]AgResearch Limited, Private Bag 4749, Christchurch 8140, New Zealand
[4]Better Border Biosecurity
[5]Northern Research Station, Forest Service, United States Department of Agriculture, 1992 Folwell Ave., St. Paul, MN 55108
[6]ssenay@umn.edu

Bio-Protection
Bioprotection science for New Zealand

agresearch
āta mātai, mātai whetū

B3 Science Solutions for BETTER BORDER BIOSECURITY
www.b3nz.org

University of Minnesota
Driven to Discover℠
Minnesota Invasive Terrestrial Plants and Pests Center

College of Food, Agricultural and Natural Resource Sciences
University of Minnesota

INSTEPP
INTERNATIONAL SCIENCE & TECHNOLOGY PRACTICE & POLICY

Department of APPLIED ECONOMICS
University of Minnesota

Standardized multi-factor modelling: Why do we need it?

- *There are numerous sources of uncertainty whenever prediction is involved.*

- *In a correlative SDM setting, quality of presence data, model type, model parametrization and model evaluation are known to be major sources of uncertainty.*

- *Similar to climate studies, using various models rather than a single model allows us to better estimate best and worst case scenarios, confidence interval and consistency between models.*

- *An ensemble modelling framework can be useful in weeding out weak models. [as there are no universally best models]*

Background – *Halyomorpha halys*

Halyomorpha halys

- *Native to China, N. & S. Korea and Japan*
- *Established in the US (1990's)*
- *Established in Europe (2000's)*
- *Established in Canada (2014)*
- *Multi-host pest { more than 100 host species recorded}*
- *Currently expanding its invaded range*

Picture courtesy: https://blog.doc.govt.nz/2014/11/29/brown-marmorated-stink-bug/

General setting

- *Geographic extent: Global and NZ*
- *Occurrence dataset: 3,419*
- *Environmental covariates: 35 bioclim variables from CLIMOND + Elevation layer*
  - ✓ *Spatial resolution = 10" Global resolution & 30' NZ extent (interpolated from CLIMOND data)*

Research design & Methods

- *Two types of presence datasets = Initial and Refined*
- *Two types of modelling approaches = Presence-only (MaxEnt) & Presence-Absence (7 model ensemble MMF)*
- *Standard modelling procedures*
  - ✓ *Niche analysis results used to inform modelling decisions*
  - ✓ *Background and pseudo-absence datasets were methodologically generated*
  - ✓ *Centralized variable selection carried out using a random forest classifier*
  - ✓ *Multiple-model evaluation methods used.*

…Methods - Modelling

Presence-only (MaxEnt)

- *Software = MaxEnt v.3.3.3k*
- *Features (functions) = LQPH*
- *Background points = All locations (37k+ and 40k+ for initial and refined scenarios respectively)*
- *A moderate regularization multiplier= (β=2)* *(Radosavljevic and Anderson, 2014)*
- *10-fold cross-validation*

Presence-absence (MMF)

- *A 7 model Framework*
  - *QDA, LOG, NB, CART, RF, ANN, SVM*
- *Built-in customized model specific parametrization*
- *Top 3 models selected based on Kappa score and are combined to give weighted average, best and worst scenarios*
- *10-fold cross-validation*

$$P_{EA} = \frac{\sum(\kappa_{m_i} \times P_{m_i})}{\sum \kappa_{m_i}} \text{----------} (1)$$

Where: κ *is the Kappa score of the* $_{i^{th}}$ *high performing model* (m)*, and* P *is the predicted probability value for a given cell by the* i^{th} *model* (m).

$$P_{EB} = \min([P_{m_i}]) \text{----------} (2)$$

$$P_{EW} = \max([P_{m_i}]) \text{----------} (3)$$

Where: $[P_{m_i}]$ *is a vector of predicted probabilities for a given cell* $(P_{m_1}, \cdots, P_{m_n})$ *by high performing models* $m_1 \cdots m_n$.

Result - Presence datasets

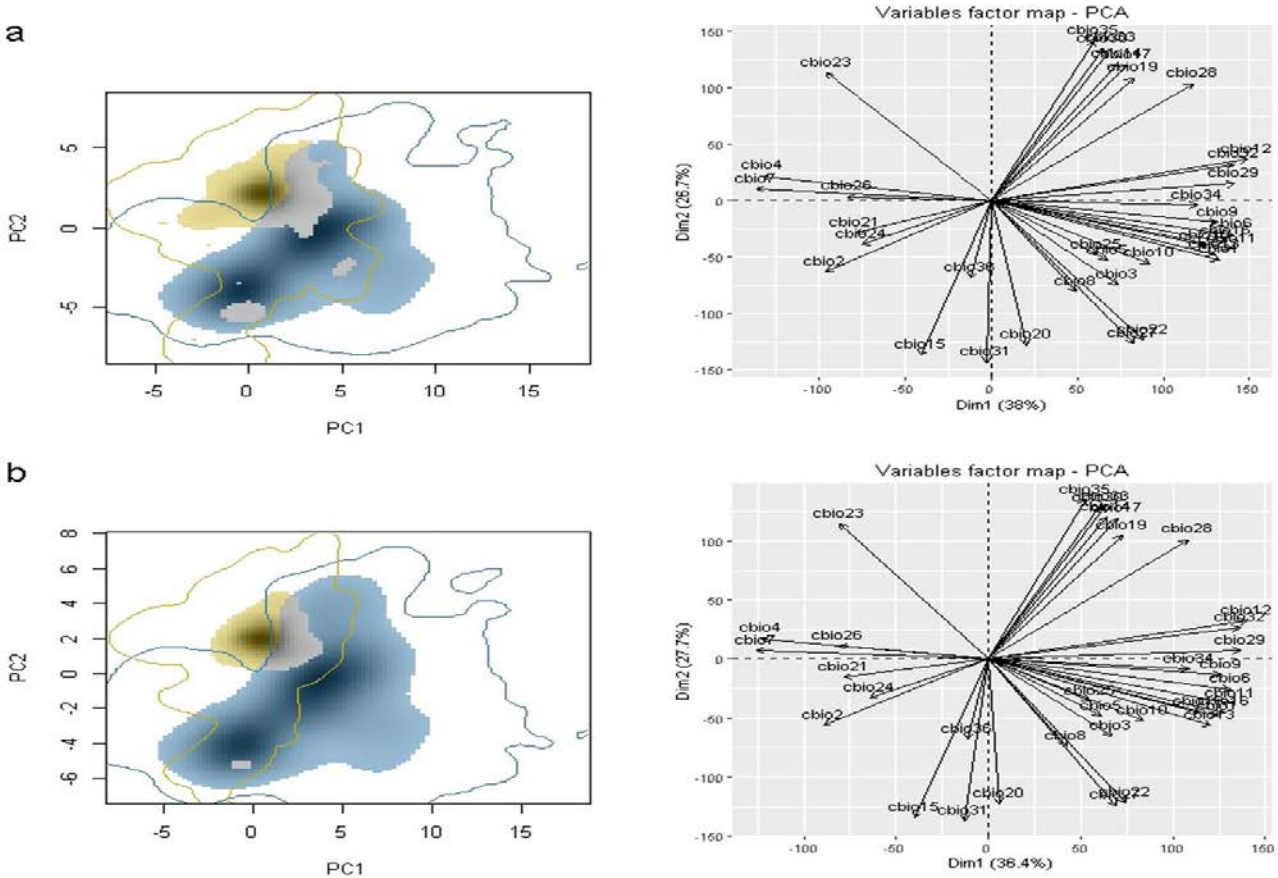
Initial dataset (hh.set1)
Total = 3,189
After duplicates were removed = 492
Unique to this dataset = 22
Removed locations include: Alaska, Florida and Arizona.

Refined dataset (hh.set2)
Total = 3,384
After duplicates were removed = 547
Unique to this dataset = 77
Added locations include: More locations in Japan

Result - Niche analysis

Stability of the inferred climatic niche of Halyomorpha halys

- *PCA performed on the 36 predictors with respect to the H. halys distributional data.*
- *Eco-regions were used to delimit the invaded and native ranges.*
- *A kernel density function using a 100*100 cell was applied to obtain the smoothed density of occurrences in each cell in the PCA environmental space (based of 1st by 2nd component matrix)*

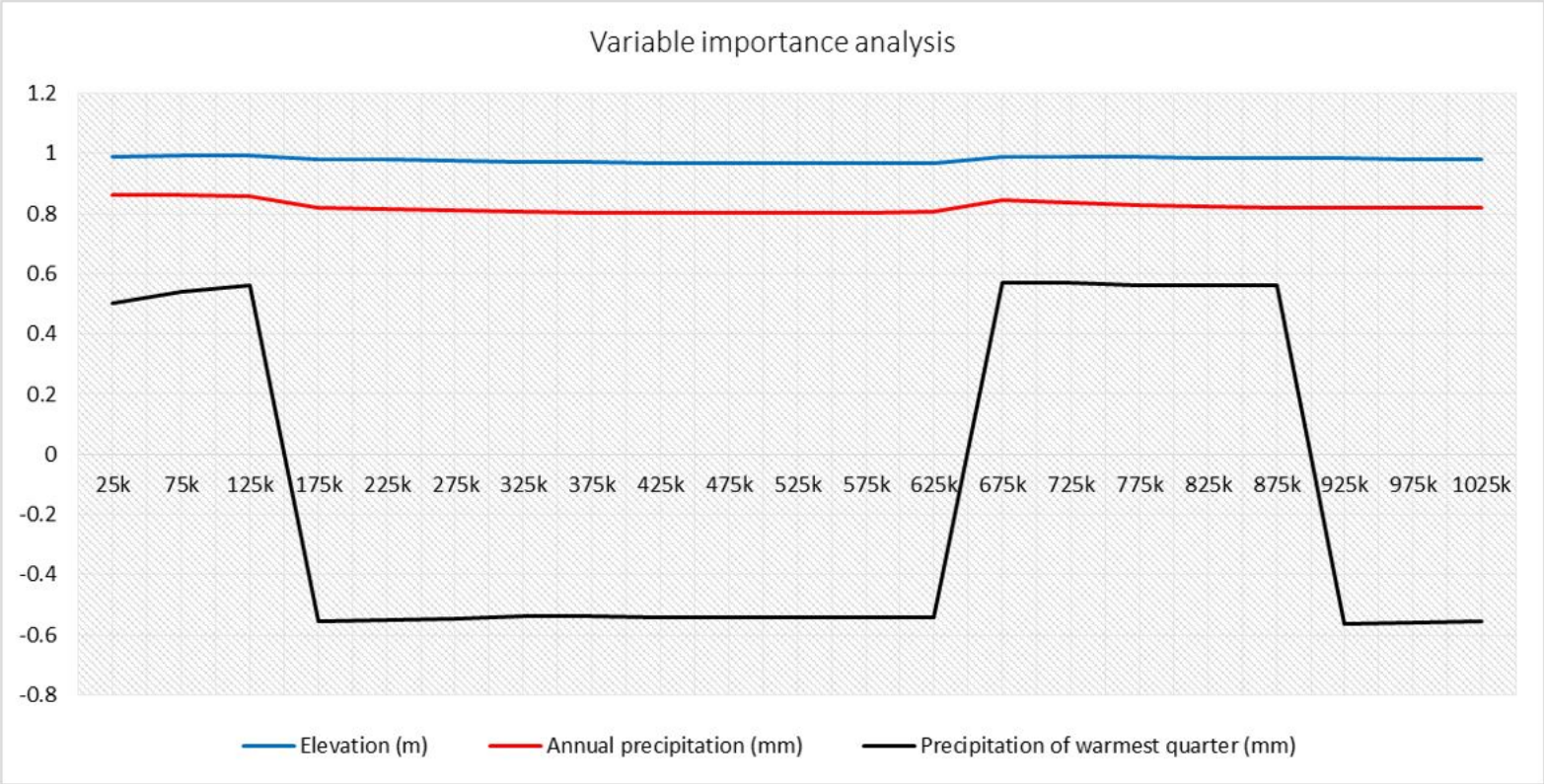



```
Stability = 0.63
Unfilling = 0.89
Expansion = 0.37
```

Result - Background data

Background data

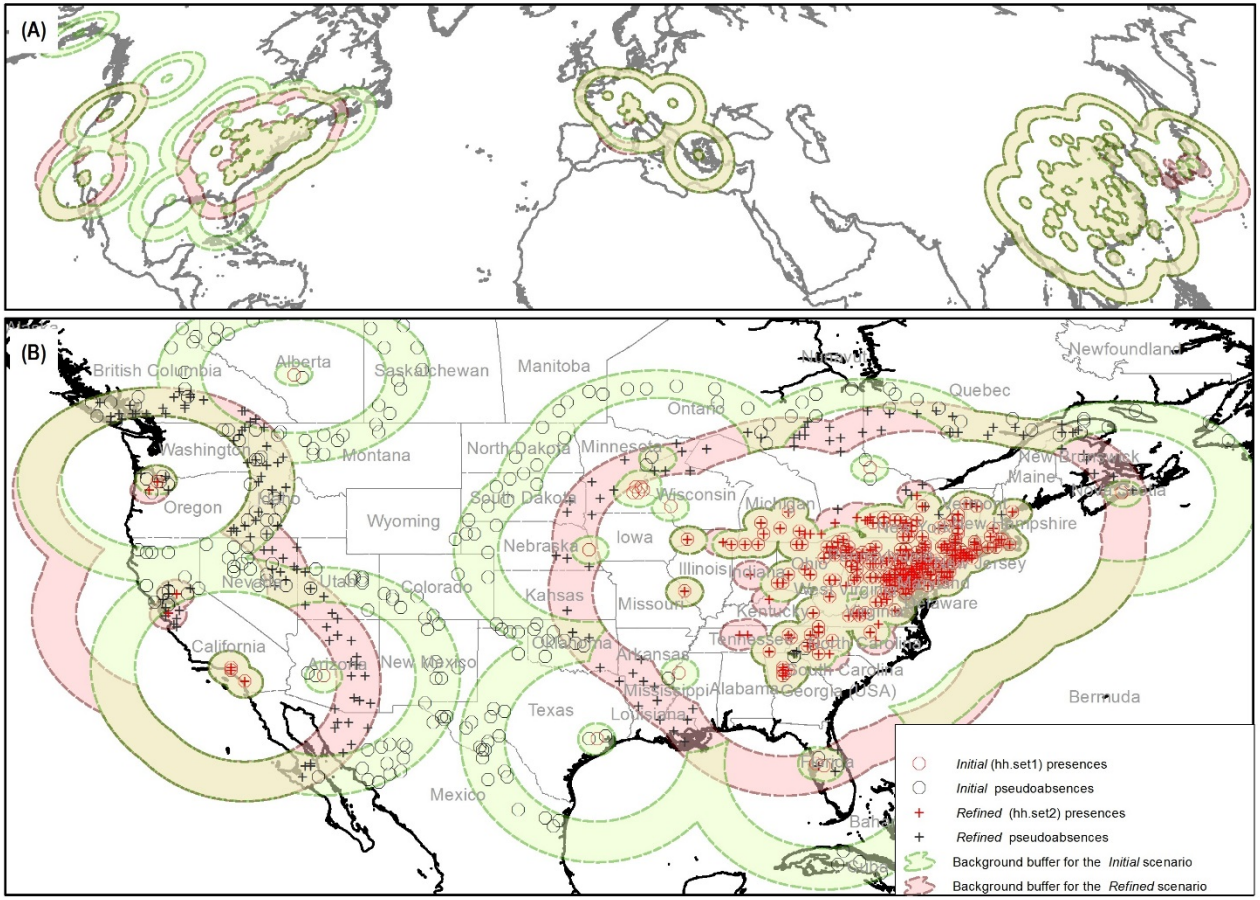
- *Variable importance analysis used to delimit optimum geographic distance to buffer background data*
- *Same distances identified both for the initial and refined scenario*



Result - Pseudo-absence locations

Pseudo-absence points

- *One-Class Support Vector Machines (OCSVM) used to environmentally profile the background data into different classes based on how dissimilar they are to the presence locations*
- *K-means clustering used to select centroids of clusters generated from the most dissimilar OCSVM class.*



Result - Variable selection

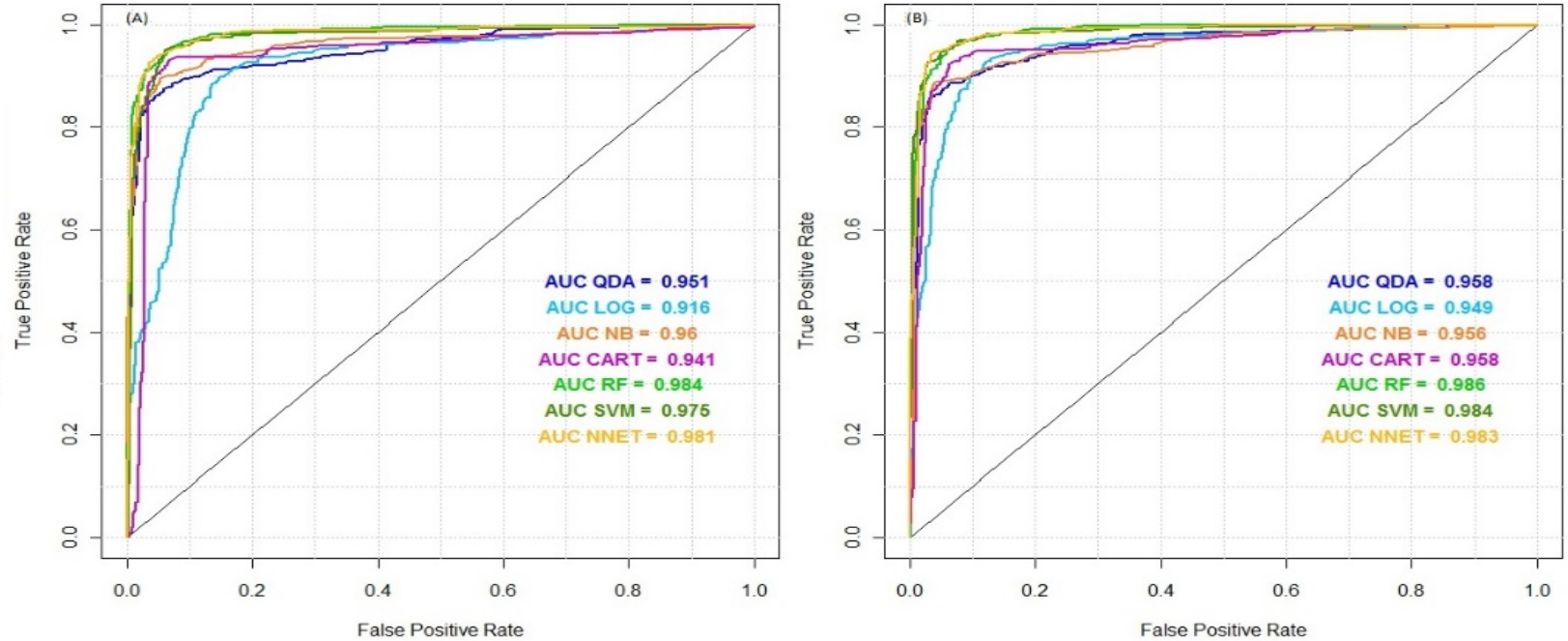
Variables selected using a Random forest classifier

Table 2. Variables chosen for modelling in declining order of importance. (A) Based on the Initial presence dataset, (B) Based on the refined presence dataset

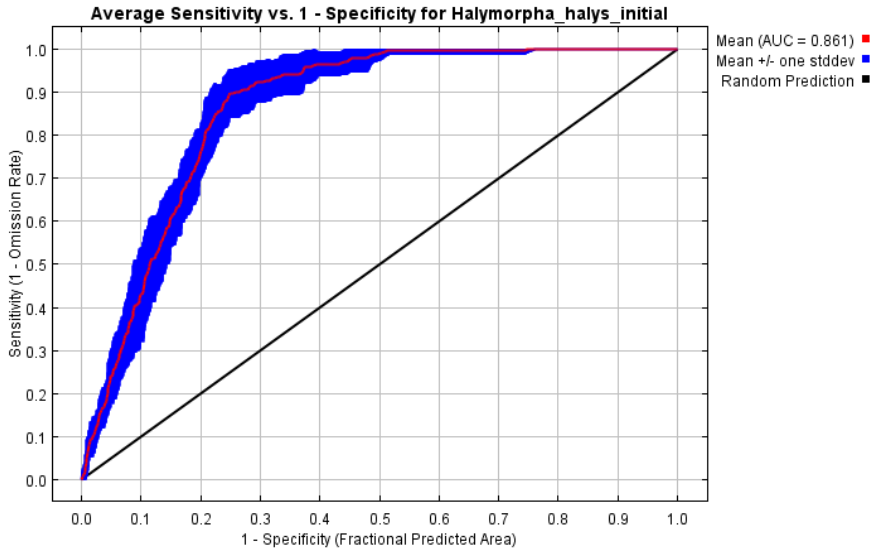
(A) Selected variables for the *Liberal scenario*	Df	Deviance	AIC
-	-	717.67	735.67
Annual mean radiation (W m-2) [var 20]	1	728.64	744.64
Mean temperature of warmest quarter (°C) [var 10]	1	729.53	745.53
Annual mean temperature (°C) [var 01]	1	739.55	755.55
Max temperature of warmest week (°C) [var 05]	1	743.12	759.12
Mean moisture index of warmest quarter [var 34]	1	746.93	762.93
Precipitation of warmest quarter (mm) [var 18]	1	755.74	771.74
Highest weekly radiation (W m-2) [var 21]	1	761.79	777.79
Isothermality (Bio02 ÷ Bio07) [var 03]	1	785.46	801.46
(B) Selected variables for the *Conservative scenario*	Df	Deviance	AIC
-	-	611.64	637.64
Annual mean temperature (°C) [var 01]	1	611.65	635.65
Radiation of warmest quarter (W m-2) [var 26]	1	611.7	635.7
Max temperature of warmest week (°C) [var 05]	1	612.08	636.08
Annual mean radiation (W m-2) [var 20]	1	613.42	637.42
Precipitation of warmest quarter (mm) [var 18]	1	613.66	637.66
Isothermality (Bio02 ÷ Bio07) [var 03]	1	615.05	639.05
Mean moisture index of warmest quarter [var 34]	1	615.11	639.11
Highest weekly radiation (W m-2) [var 21]	1	619.34	643.34
Mean temperature of wettest quarter (°C) [var 08]	1	625.51	649.51
Mean temperature of coldest quarter (°C) [var 11]	1	720.22	744.22
Temperature seasonality (C of V) [var 04]	1	803.22	827.22

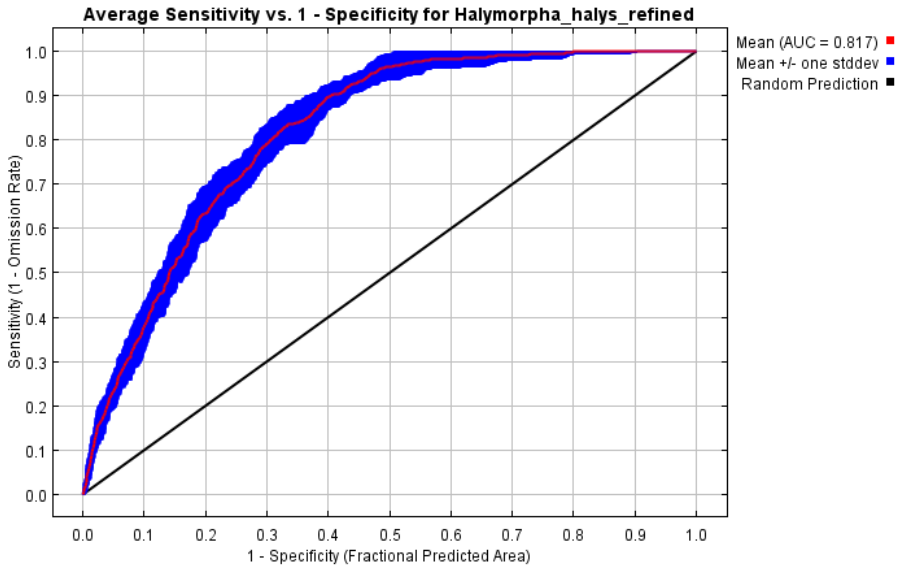
Evaluation – confusion matrix based

MMF



MaxEnt



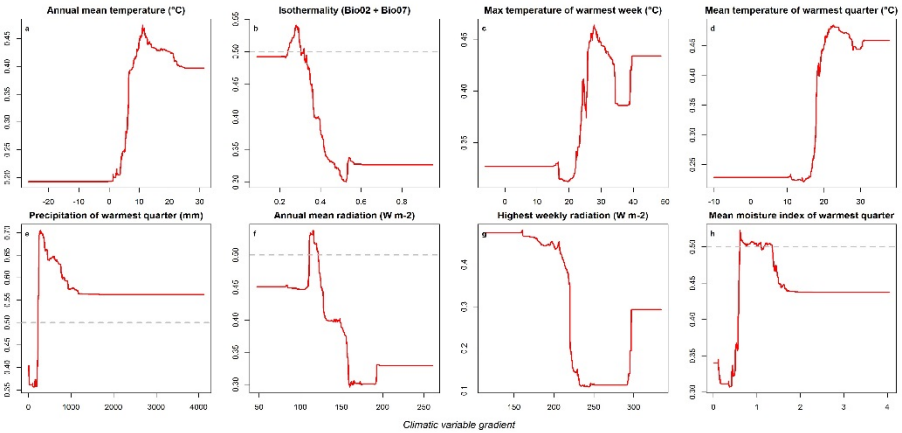


Evaluation – Response curves (initial)

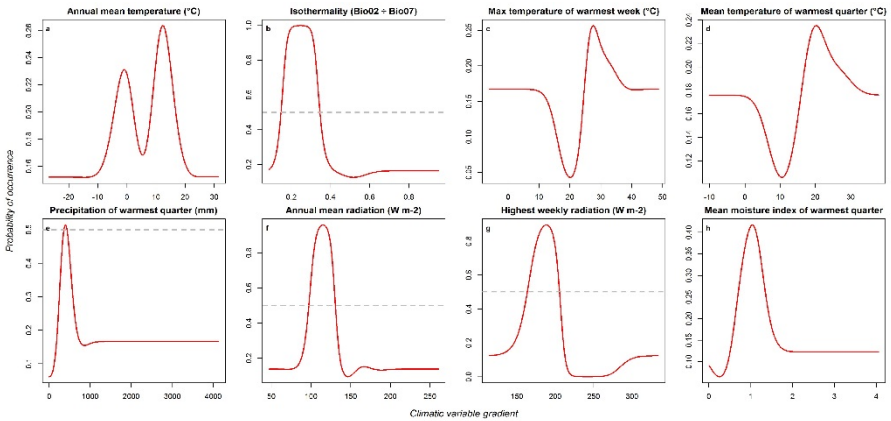
Annual mean radiation (W m-2) [var 20]
Mean temperature of warmest quarter (°C) [var 10]
Annual mean temperature (°C) [var 01]
Max temperature of warmest week (°C) [var 05]

Mean moisture index of warmest quarter [var 34]
Precipitation of warmest quarter (mm) [var 18]
Highest weekly radiation (W m-2) [var 21]
Isothermality (Bio02 ÷ Bio07) [var 03]

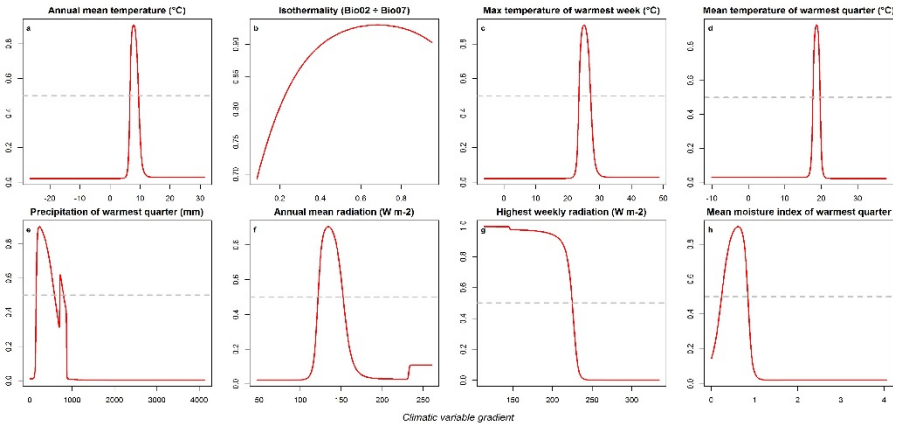
RF



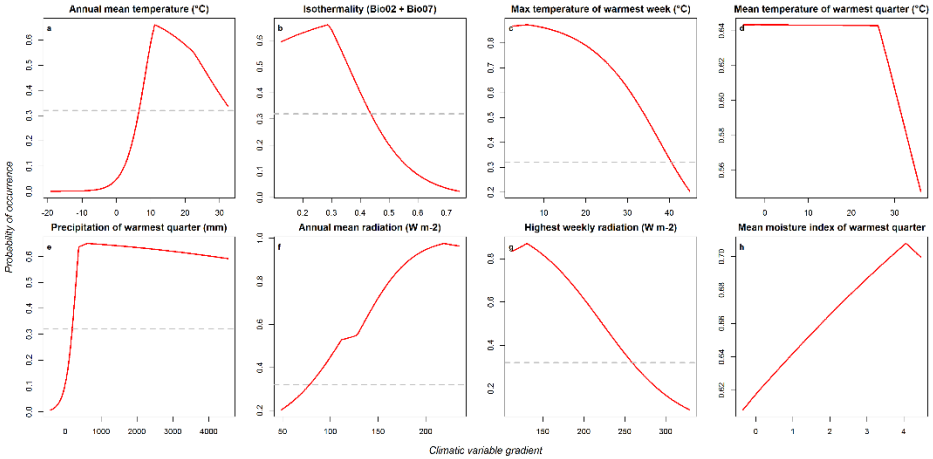
SVM



NNET

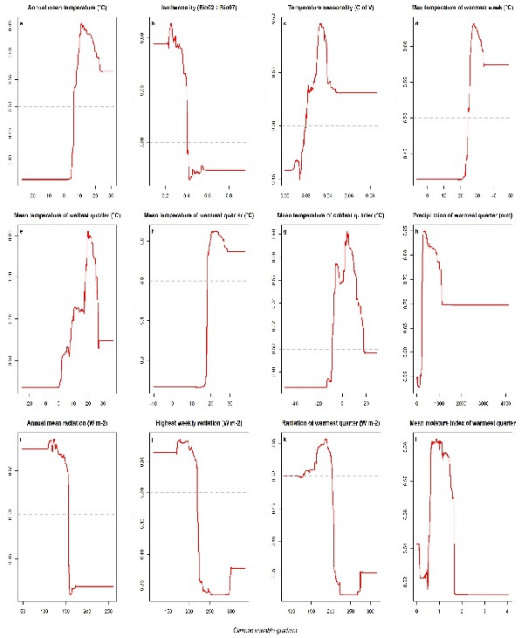


MaxEnt

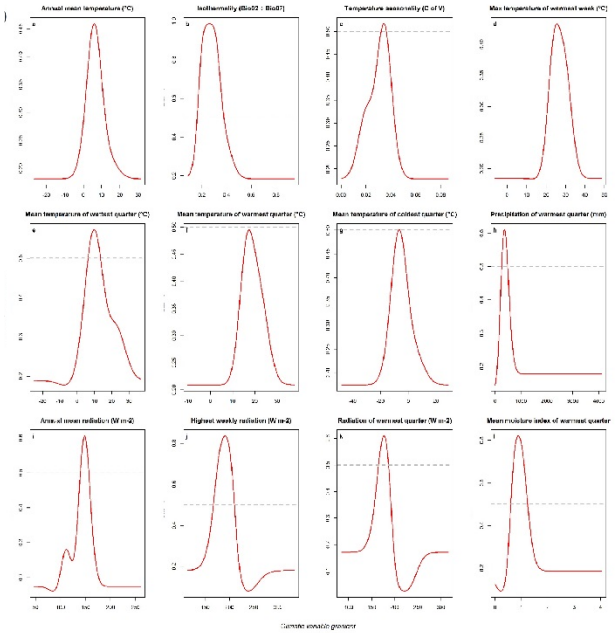


Evaluation – Response curves (refined)

RF

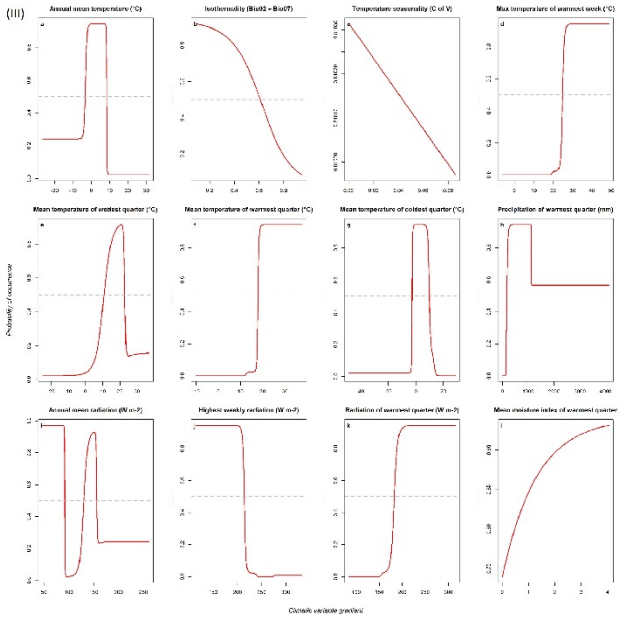


SVM

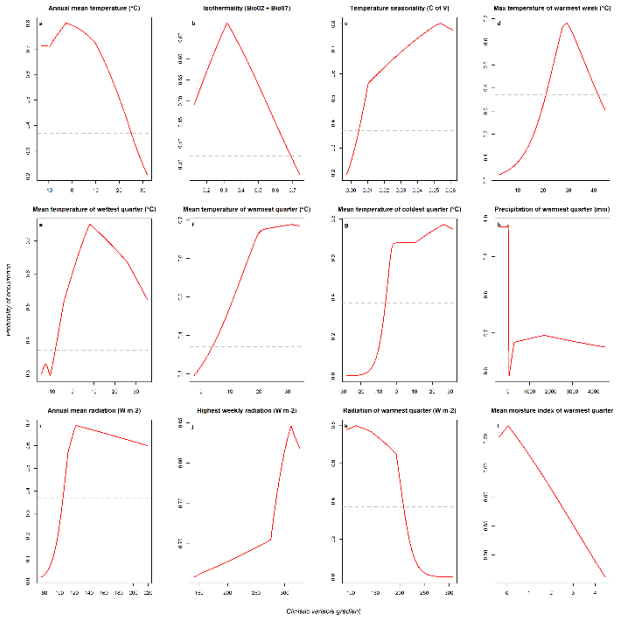


Annual mean temperature (°C) [var 01]	1
Radiation of warmest quarter (W m-2) [var 26]	2
Max temperature of warmest week (°C) [var 05]	3
Annual mean radiation (W m-2) [var 20]	4
Precipitation of warmest quarter (mm) [var 18]	5
Isothermality (Bio02 ÷ Bio07) [var 03]	6
Mean moisture index of warmest quarter [var 34]	7
Highest weekly radiation (W m-2) [var 21]	8
Mean temperature of wettest quarter (°C) [var 08]	9
Mean temperature of coldest quarter (°C) [var 11]	10
Temperature seasonality (C of V) [var 04]	11
Mean temperature of warmest quarter (°C) [var 10]	12

NNET

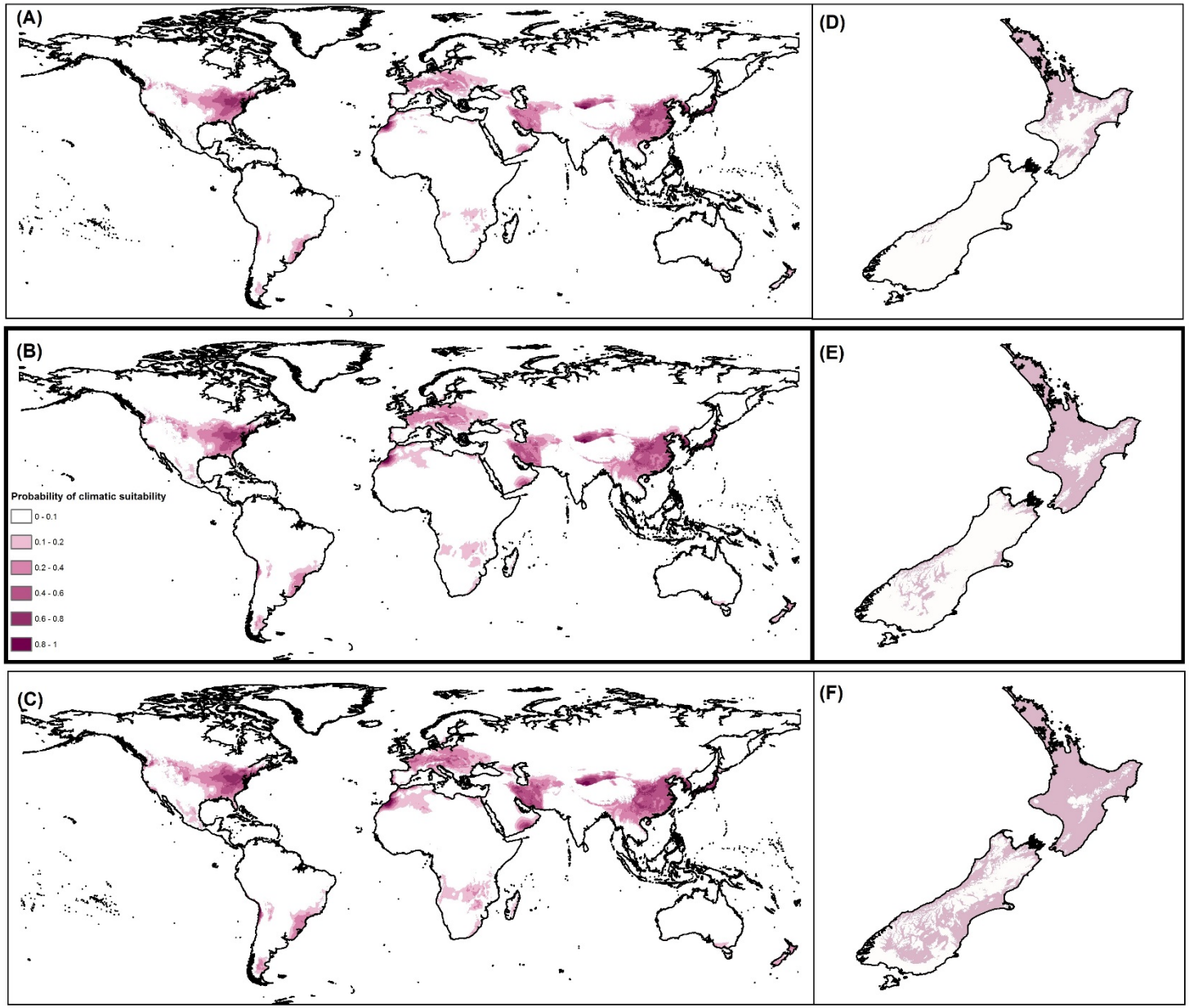


MaxEnt



Annual mean temperature (°C) [var 01]	1
Radiation of warmest quarter (W m-2) [var 26]	2
Max temperature of warmest week (°C) [var 05]	3
Annual mean radiation (W m-2) [var 20]	4
Precipitation of warmest quarter (mm) [var 18]	5
Isothermality (Bio02 ÷ Bio07) [var 03]	6
Mean moisture index of warmest quarter [var 34]	7
Highest weekly radiation (W m-2) [var 21]	8
Mean temperature of wettest quarter (°C) [var 08]	9
Mean temperature of coldest quarter (°C) [var 11]	10
Temperature seasonality (C of V) [var 04]	11
Mean temperature of warmest quarter (°C) [var 10]	12

Presence-only *H. halys* potential distribution projections

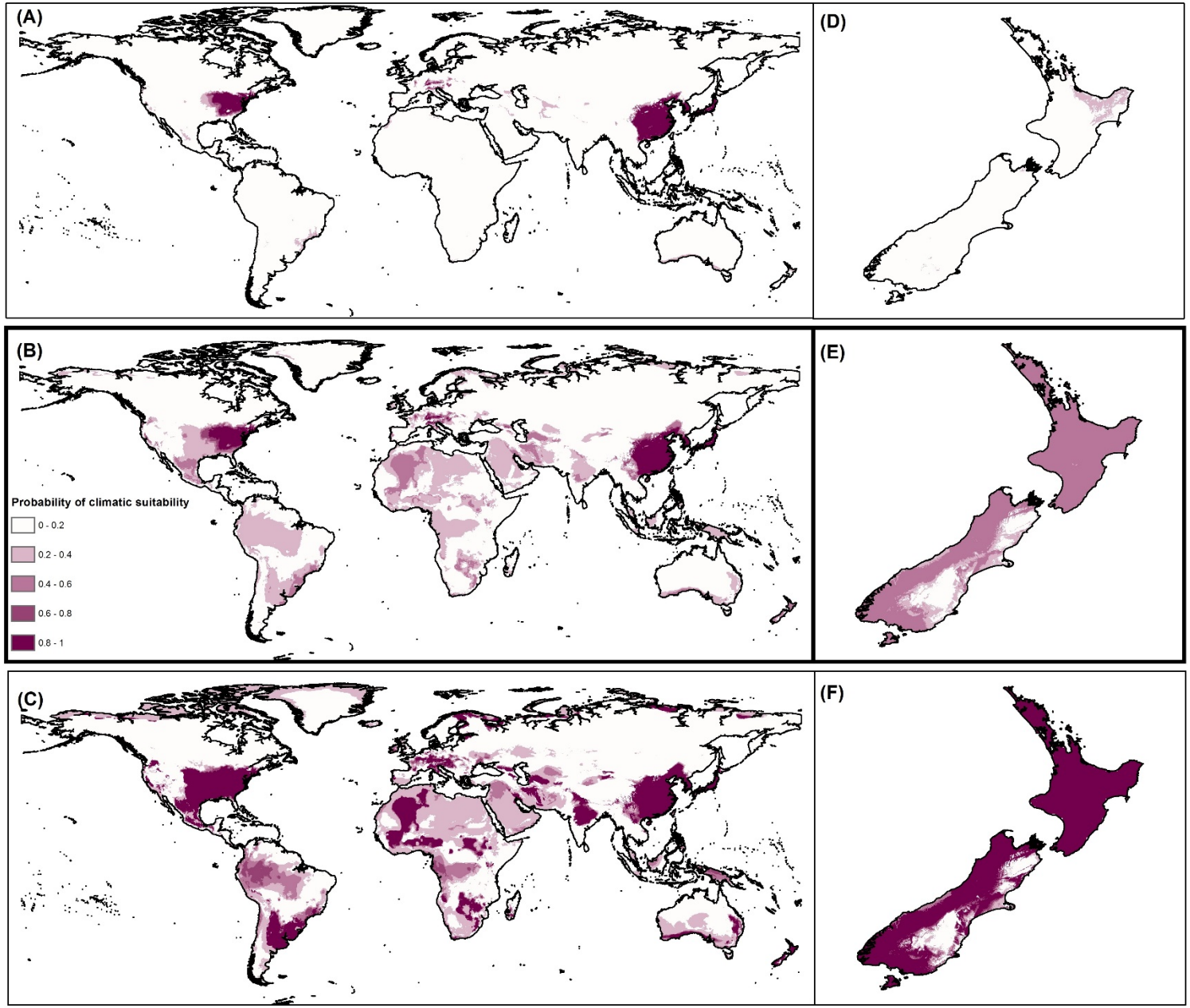


Best scenario

Average

Worst scenario

Presence-absence *H. halys* potential distribution projections

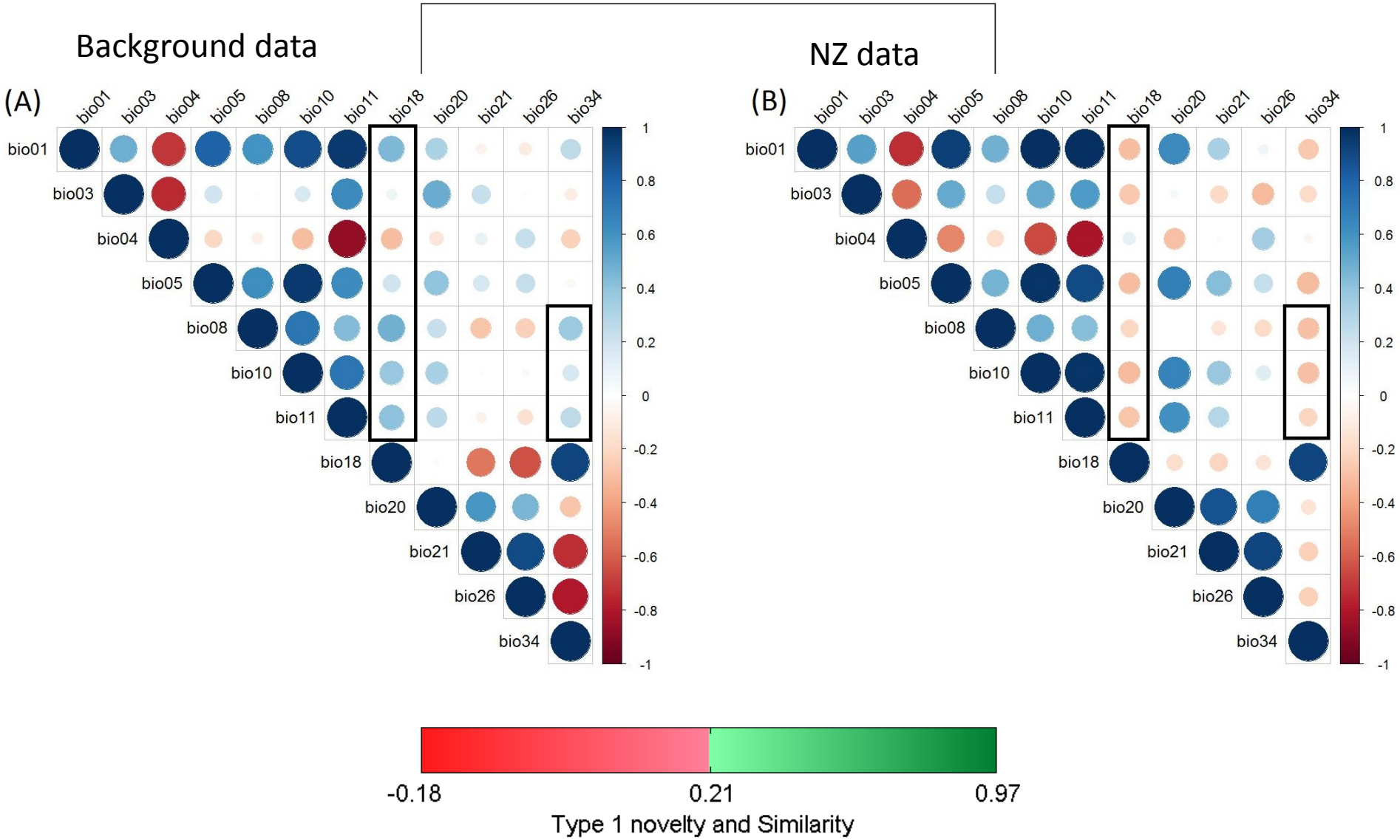


Best scenario

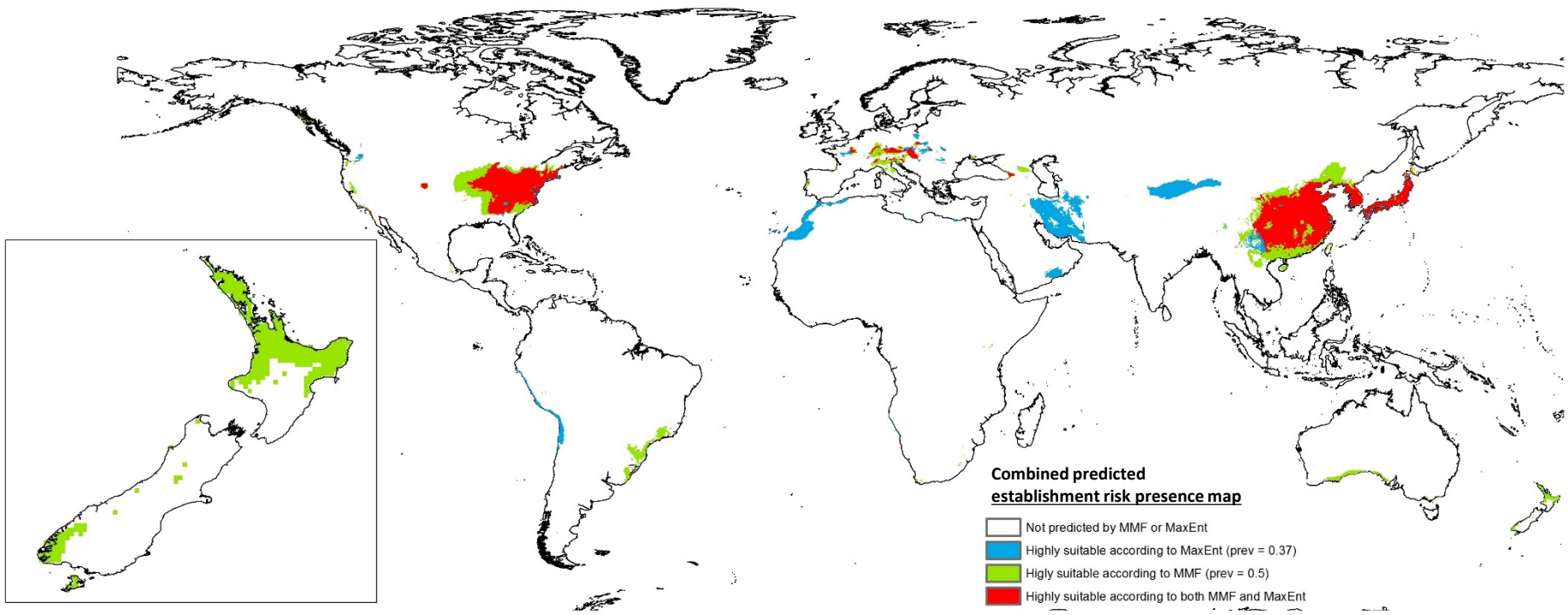
Weighted average

Worst scenario

H. halys distribution projections for New Zealand



Ensemble map – *H. halys* establishment risk (I/0)



P/A prev. = 0.5
P-O prev. = 0.37

Conclusion

- *Standardized modelling frameworks allow grounded comparisons between different model results.*

- *Using multiple models in potential distribution predictions could reveal probable but different spatial distribution trajectories*

- *While using ensemble models it is important to weed out weaker and low-performing models so as not to include unnecessary variation in predictions.*

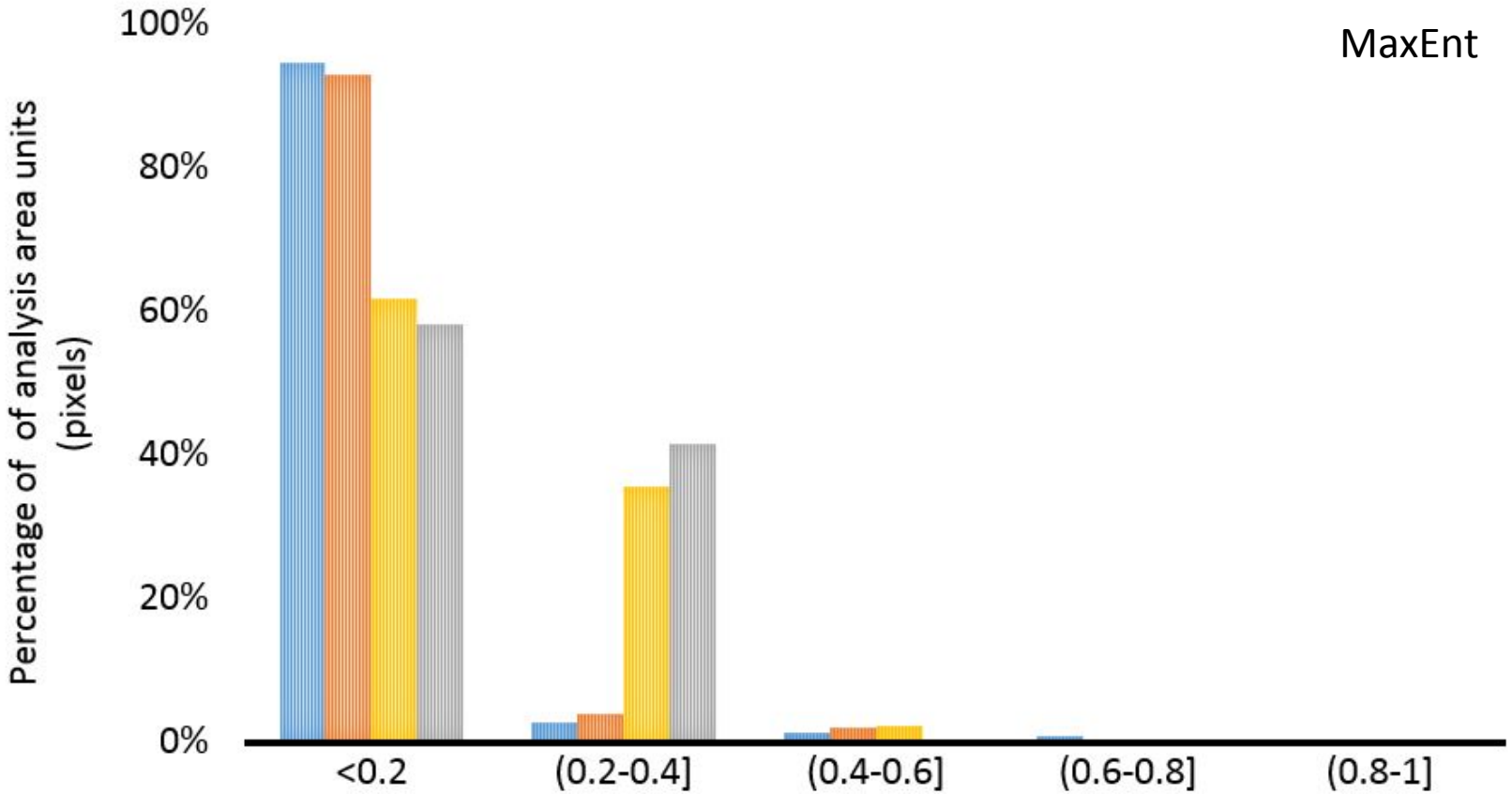
- *H. halys is expanding its invasion, however from the fragmented climatically suitable sites predicted in Europe it appears that its expansion in North America might be faster than in Europe. The fact that there are suitable environments that are still not occupied by H.halys all over its invaded range increases the likelihood of its expansion.*
- *Follow up studies on H. halys dispersal and effect of control mechanisms on range expansion are highly recommended in order to estimate the potential for eradication or containment.*

- *H. halys proved to be a difficult pest to predict due to:*
  1. *Its association with urban/human environments*
  2. *also possibly because it has variant acclimation potential sin different environments.*

Thank you!

www.instepp.umn.edu

Evaluation-predicted area



	<0.2	(0.2-0.4]	(0.4-0.6]	(0.6-0.8]	(0.8-1]
Global initial	94.77%	2.89%	1.42%	0.92%	0.00%
Global refined	93.12%	4.09%	2.15%	0.57%	0.07%
New Zealand initial	61.85%	35.75%	2.39%	0.00%	0.00%
New Zealand refined	58.30%	41.70%	0.00%	0.00%	0.00%