Revisiting the use of Self-Organizing Maps (SOM) to predict to risk of invasion of pre-emergent pest species

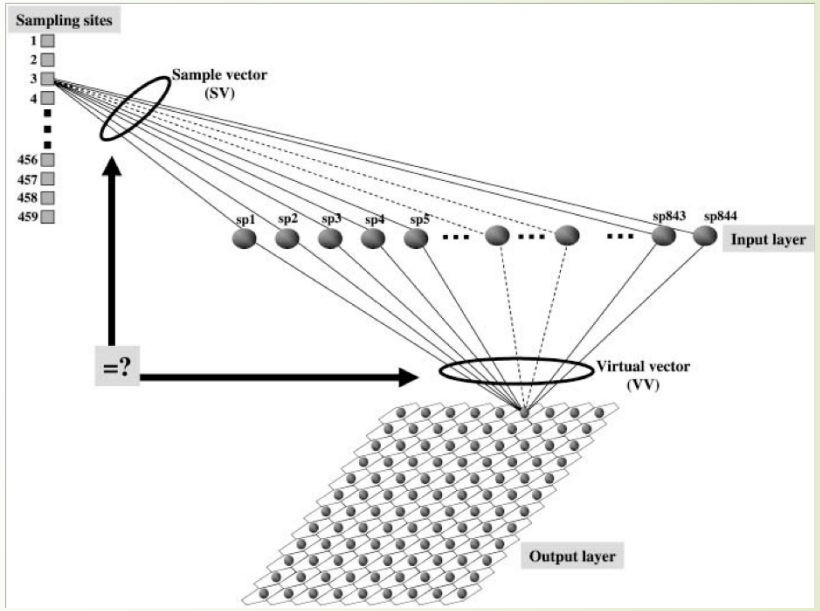
Karl Suiter; Durham, North Carolina

Godshen Pallipparambil; Center for Integrated Pest Management, NC State University

2019 Annual Meeting of the International Pest Risk Research Group

Pest Prioritization Using Self Organizing Maps (SOM)

Susan Worner



Worner S, Gevrey M, Eschen R, Kenis M, Paini D, Singh S, Watts M, Suiter K (2013) Prioritizing the risk of plant pests by clustering methods; self-organising maps, k-means and hierarchical clustering. NeoBiota 18: 83-102. https://doi.org/10.3897/neobiota.18.4042

SOM Background

- Developed in 1982 by Teuvo Kohonen
- Neural network algorithm using unsupervised competitive learning
- Primarily used for organization and visualization of complex data
- 2-dimentional pictorial representation of complex data
- Uses SOM Toolbox, version 2.0 (http://www.cis.hut.fi/somtoolbox)
- Analysis and visualization performed using Matlab

SOM Invasive Species Data

- CABI Crop Protection Compendium
- 238 geopolitical regions
- 103793 distribution records
- Data clean-up. Removed records labeled ….
  - No information available; Unconfirmed record; No Information Listed; Last reported Introduced, not established; Introduced, estab. uncertain; Eradicated; Absent, unreliable record; Absent, no pest record; Absent, confirmed by survey; Absent, formerly present; Absent, intercepted only; Absent, invalid record; Absent, never occurred; Absent, reported not confirmed
- 98100 clean distribution records
- 5636 distinct pest species
- **1,341,368 data points**

CABI Invasive Species Data Detail

Classification	Number	%
Arachnida	153	2.71%
Bacteria	226	4.01%
Chordata	26	0.46%
Chromista	104	1.85%
Diplopoda	1	0.02%
Colembolla	5	0.09%
Fungi	987	17.51%
Insecta	3025	53.67%
Mollusca	23	0.41%
Nematoda	139	2.47%
Plantae	603	10.70%
Protista	11	0.20%
Virus	333	5.91%
Total	**5636**	

Classification	Number	%
Acarida	4	2.61%
Acarina	1	0.65%
Araneae	2	1.31%
Astigmata	7	4.58%
Mesostigmata	4	2.61%
Prostigmata	127	83.01%
Sarcoptiformes	2	1.31%
Trombidiformes	6	3.92%
Total	**153**	

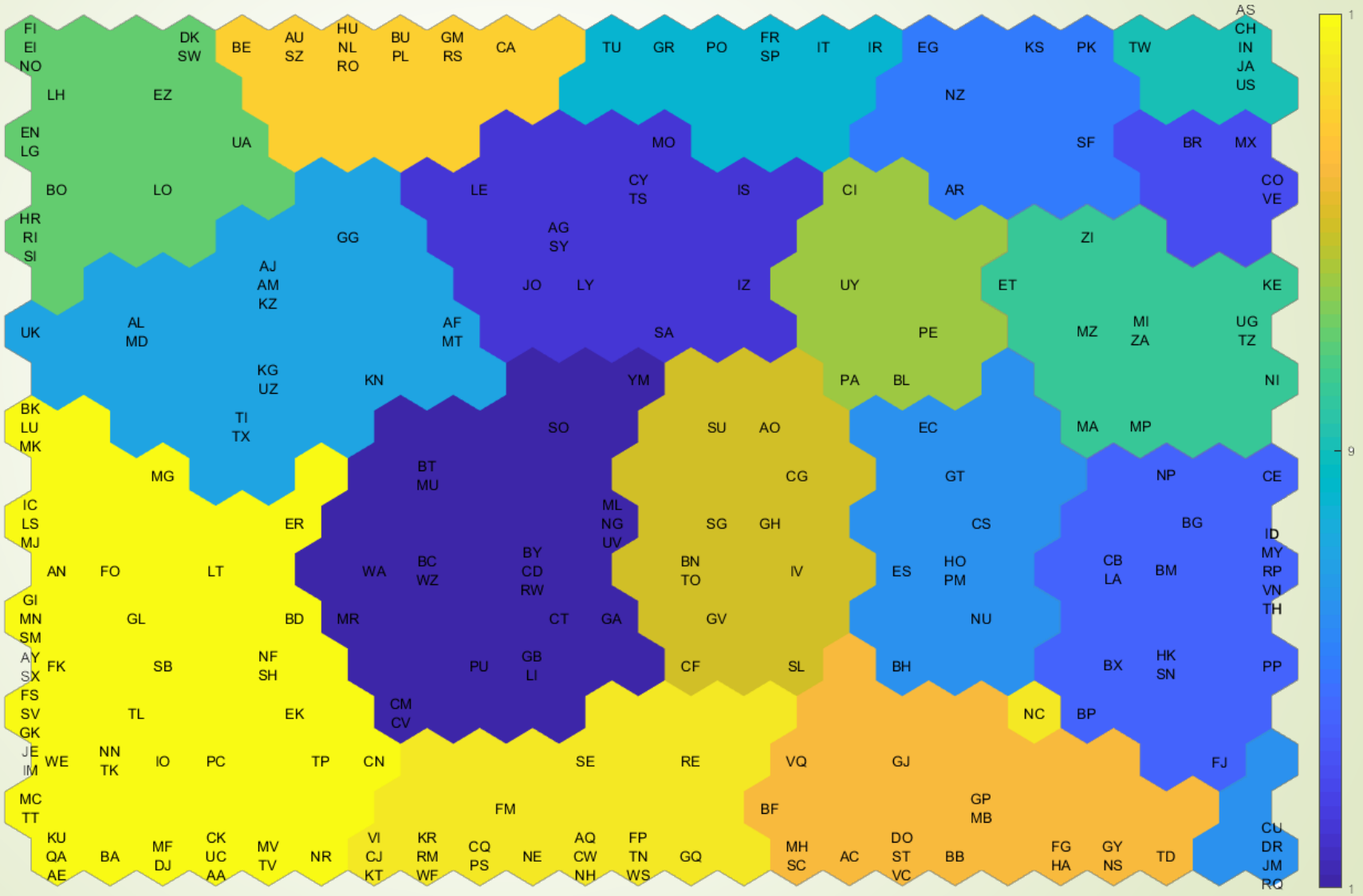
Classification	Number	%
Blattodea	51	1.69%
Coleoptera	805	26.61%
Dermaptera	2	0.07%
Diptera	259	8.56%
Hemiptera	824	27.24%
Hymenoptera	125	4.13%
Lepidoptera	778	25.72%
Orthoptera	78	2.58%
Phasmatodea	1	0.03%
Psocoptera	7	0.23%
Spirobolida	1	0.03%
Thysanoptera	94	3.11%
Total	**3025**	

SOM Data Matrix

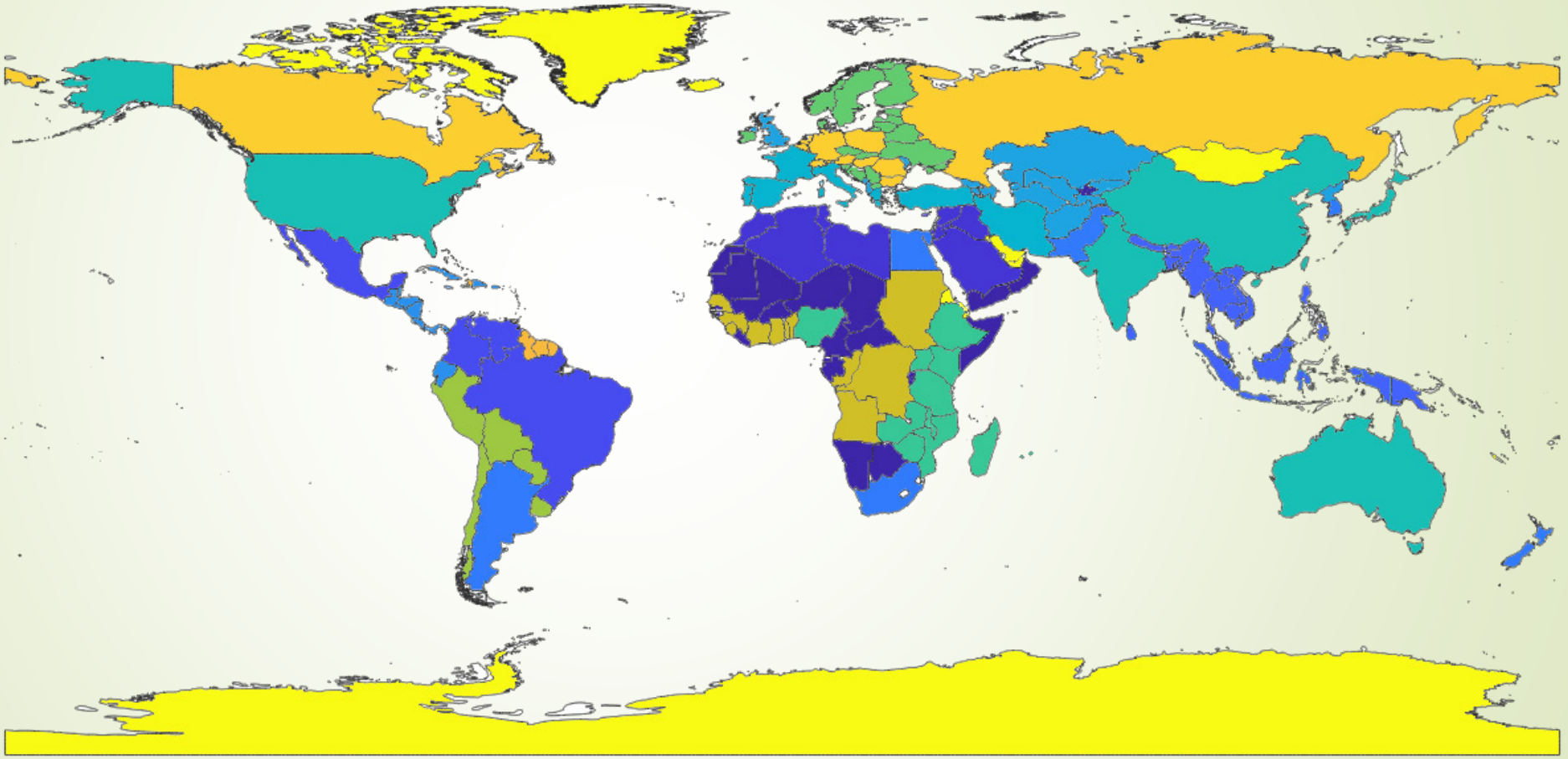
Xylosandrus ater	Xylosandrus compactus	Xylosandrus crassiusculus	Xylosandrus discolor	Xylosandrus germanus	Xylosandrus morigerus	Xylosandrus mutilatus	Xylotrechus altaicus	Xylotrechus namanganensis	Xylotrechus quadripes	Xylotrupes gideon	Yponomeuta evonymella	Yponomeuta malinellus	Yponomeuta padellus	Yponomeuta plumbellus	Zabrotes subfasciatus	Zabrus tenebrioides	Zaprionus indianus	Zaprionus tuberculatus	Zeuzera coffeae	Zeuzera multistrigata	Zeuzera pyrina	Ziziphus mauritiana	Ziziphus spina-christi	Zonitoides arboreus	Zonocerus elegans	Zonocerus variegatus	Zophiuma butawengi	Zophopetes cerymica	Zophopetes dysmephila	Zulia charon	Zulia colombiana	Zygaena filipendulae	Zygina nivea	Zygogramma bicolorata	Zygotylenchus guevarai	
5601	5602	5603	5604	5605	5606	5607	5608	5609	5610	5611	5612	5613	5614	5615	5616	5617	5618	5619	5620	5621	5622	5623	5624	5625	5626	5627	5628	5629	5630	5631	5632	5633	5634	5635	5636	FIPS_Code
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	AF
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AL
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AY
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	AG
0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AQ
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AN
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	AO
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AC
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AJ
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	AR
0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AS
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	AU
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	BF
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BA
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	BG
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AM
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	BB
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BE
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BD
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BT
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BL

0 = Absent 1 = Present

SOM Visualization



SOM Visualization



Detailed SOM Visualization

AU - Austria
BE – Belgium
BU – Bulgaria
CA – Canada
GM – Germany
HU – Hungary
NL – Netherlands
PL – Poland
RO – Romania
RS – Russian Federation
SZ – Switzerland

Poland Pest Risk Rankings

- Analyses allow us to visualize how geo-political regions are associated based on their known pest assemblages
- Use of region Best Matching Unit (BMU) neuron weights (i.e., Euclidean Distance) can be used to assign risk rankings to pests not known to occur in a region
- Objective prioritization and ranking of pests

Poland Pest Risk Rankings

Risk	Pres/Abs	Species	Classification
0.8169900000	0	*Tospovirus*: Tomato spotted wilt virus	Virus
0.8169200000	0	*Cynodon dactylon*	Plantae
0.8166000000	0	*Medicago polymorpha*	Plantae
0.8148800000	0	*Cryphonectria parasitica*	Fungi
0.8146000000	0	*Amaranthus albus*	Plantae
0.8142800000	0	*Amaranthus blitum*	Plantae
0.8142200000	0	*Cuscuta campestris*	Plantae
0.8128600000	0	*Diaspidiotus perniciosus*	Insecta
0.8128600000	0	*Plasmopara halstedii*	Chromista
0.7921700000	0	*Tuta absoluta*	Insecta
0.7894500000	0	*Cydalima perspectalis*	Insecta
0.7878100000	0	*Corynespora cassiicola*	Fungi
0.7874800000	0	*Xanthomonas campestris* pv. *campestris*	Bacteria
0.7626600000	0	*Caulimovirus*: Cauliflower mosaic virus	Virus
0.7056900000	0	*Capsicum annuum*	Plantae
0.7036500000	0	*Nepovirus*: Grapevine fanleaf virus	Virus
0.7033100000	0	*Microdochium panattonianum*	Fungi
0.7032500000	0	*Urocystis cepulae*	Fungi
0.7031300000	0	*Amaranthus hybridus*	Plantae
0.7019500000	0	*Grapholita molesta*	Insecta
0.7018900000	0	*Eleusine indica*	Plantae
0.6809200000	0	*Heliotropium europaeum*	Plantae
0.6808500000	0	*Xanthomonas arboricola* pv. *pruni*	Bacteria

Poland Pest Risk Rankings

Hemiptera

Risk	Pres/Abs	Species
0.81286000	0	*Diaspidiotus perniciosus*
0.65395000	0	*Icerya purchasi*
0.65237000	0	*Pseudaulacaspis pentagona*
0.59262000	0	*Pseudococcus viburni*
0.59085000	0	*Ceresa alta*
0.59013000	0	*Metcalfa pruinosa*
0.56781000	0	*Saissetia oleae*
0.51841000	***0***	***Halyomorpha halys***
0.49423000	0	*Dreyfusia nordmannianae*
0.47066000	0	*Phloeomyzus passerinii*

Lepidoptera

Risk	Pres/Abs	Species
0.79217000	***0***	***Tuta absoluta***
0.78945000	0	*Cydalima perspectalis*
0.70195000	0	*Grapholita molesta*
0.67895000	0	*Scrobipalpa ocellatella*
0.67888000	0	*Etiella zinckenella*
0.67647000	0	*Mythimna loreyi*
0.59229000	0	*Trichoplusia ni*
0.56820000	0	*Saturnia pyri*
0.56786000	0	*Sitotroga cerealella*
0.56753000	0	*Corcyra cephalonica*

Poland Pest Risk Rankings

Fungi

Risk	Pres/Abs	Species
0.81488000	0	*Cryphonectria parasitica*
0.78781000	0	*Corynespora cassiicola*
0.70331000	0	*Microdochium panattonianum*
0.70325000	0	*Urocystis cepulae*
0.67915000	0	*Rosellinia necatrix*
0.67900000	0	*Apiognomonia erythrostoma*
0.67855000	0	*Pyrenophora chaetomioides*
0.67850000	0	*Macrophomina phaseolina*
0.65237000	0	*Botryosphaeria obtusa*
0.59234000	0	*Phaeocryptopus gaeumannii*

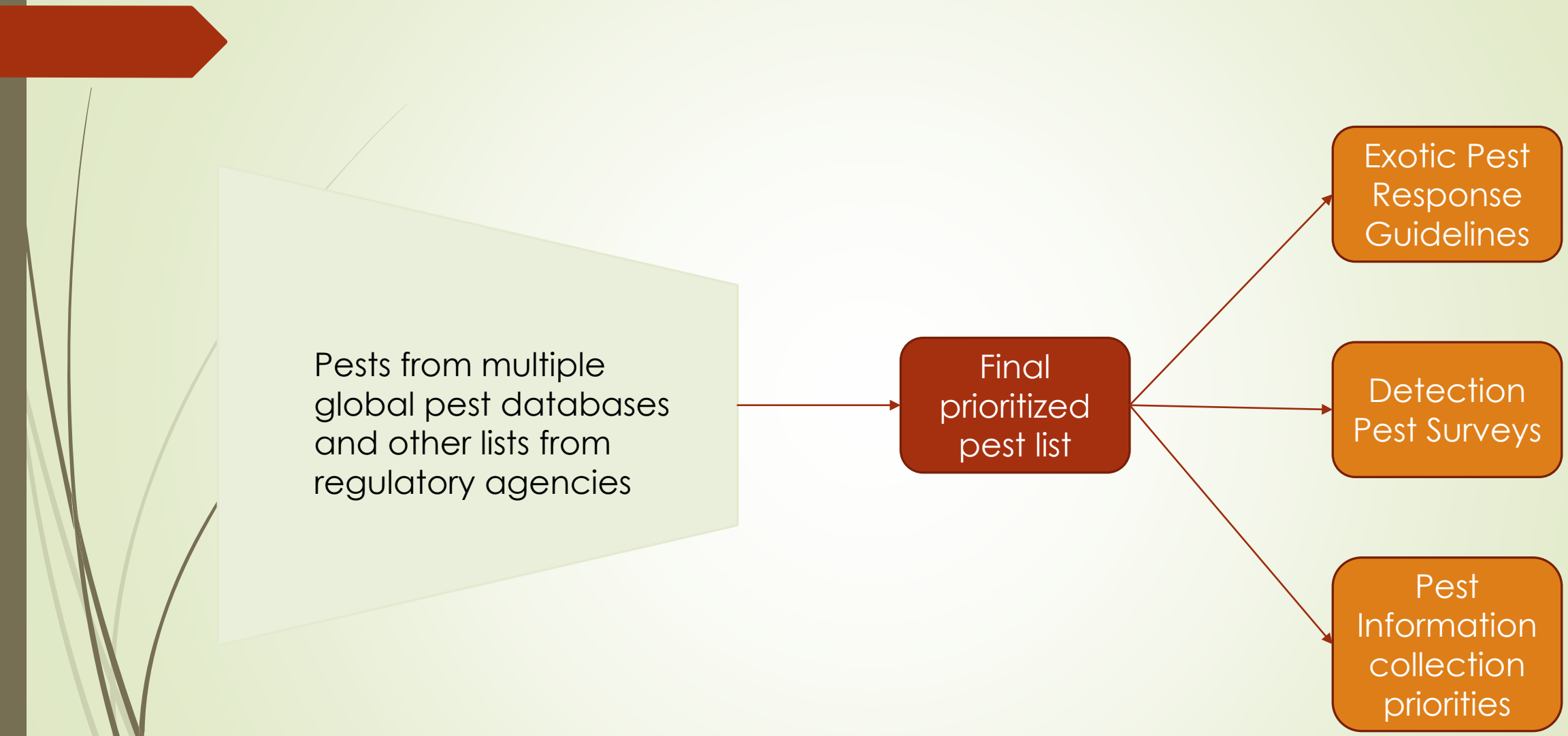
Viruses

Risk	Pres/Abs	Species
0.81699000	0	*Alfamovirus*: Alfalfa mosaic virus
0.76266000	0	*Allexivirus*: Garlic virus B
0.70365000	0	*Allexivirus*: Garlic virus C
0.54270000	0	*Allexivirus*: Shallot virus X
0.45852000	0	*Alphanecrovirus*: Olive latent virus 1
0.45644000	0	*Alphanecrovirus*: Olive mild mosaic virus
0.43180000	0	*Ampelovirus*: Little cherry virus 2
0.38656000	0	*Ampelovirus*: Plum bark necrosis stem pitting-associated virus
0.38498000	0	*Anulavirus*: Pelargonium zonate spot virus
0.35772000	0	*Apscaviroid*: Apple dimple fruit viroid

European BMSB Risk Rankings

Country	Ranking Number (out of 824 Hemipterans)	Risk Ranking
Croatia	2	0.46002000
Turkey	6	0.37027000
Portugal	7	0.59404000
Poland	8	0.51841000
Netherlands	11	0.59265000
Ukraine	19	0.24040000
Belarus	26	0.29510000
Bosnia and Herzagovena	28	0.05645000
Macedonia	31	0.05645000
United Kingdom	32	0.16955000
Czech Republic	34	0.08857600
Latvia	49	0.10311000
Denmark	50	0.08610100
Moldova	51	0.03498800
Albania	54	0.03498800
Estonia	66	0.10311000
Ireland	81	0.00156060
Lithuania	81	0.01981200

SOM analysis results can be integrated in the pest regulatory continuum


Pests from multiple global pest databases and other lists from regulatory agencies
Final prioritized pest list
Exotic Pest Response Guidelines
Detection Pest Surveys
Pest Information collection priorities

Thank you