Big Data-Model Integration as a Multi-scale Approach to Predicting the Spread of Vector-Borne Diseases: an End-to-End Vision and Operational Framework

A USDA-ARS Grand Challenge Project led by Drs. Debra Peters and Luis Rodriguez

- Goal: to develop a strategy and operational framework for complex ecological problems requiring large amounts of diverse data and scientific expertise.

- Approach: based on spatio-temporal modeling of cross-scale interactions coupled with human-guided machine learning.

- Utility of approach: illustrated using Vesicular Stomatitis (VS), an infectious disease of livestock that leads to economic losses, quarantines, and restrictions in national and international trade.

Vesicular stomatitis as a model disease

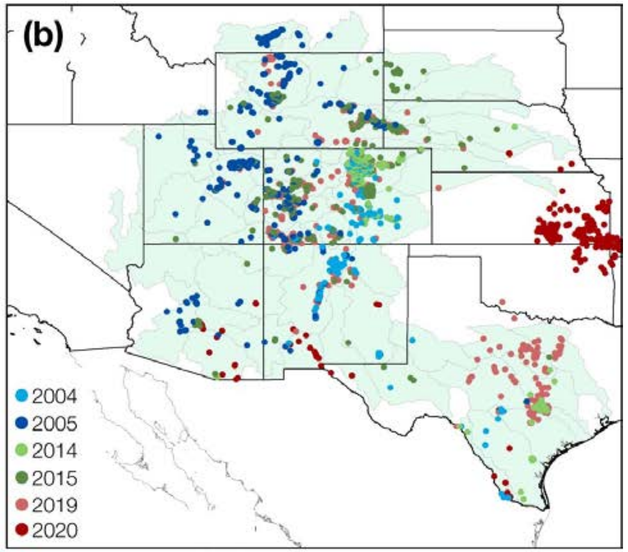
- VS virus is an arthropod-borne RNA virus (Rhabdovirus)
- Two serotypes: Indiana (VSIV) and New Jersey (VSNJV)
- Multiple hosts (e.g., cattle, horses, pigs, humans, wildlife) and insect vectors (blackflies, sand flies, biting midges)
- Clinical signs in cattle and pigs resemble foot-and-mouth disease (eradicated in US in 1926) making rapid diagnosis important
- Reportable disease to USDA Animal and Plant Health Inspection Service (APHIS)
- Sporadic outbreaks in the US (ca. 6 to 10 yr intervals) caused by strains originating in endemic areas of southern Mexico
- VS is the most commonly reported vesicular disease of livestock in the Americas

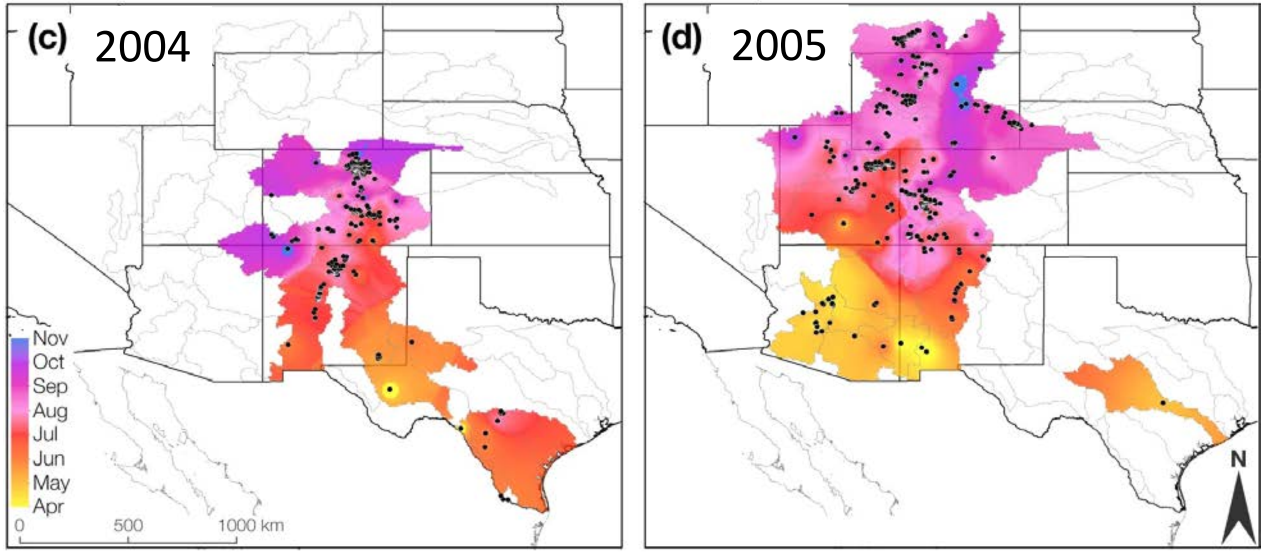
Distribution of Vesicular Stomatitis in the US

- In the US, two outbreak cycles occurred in 2004-2006 and 2012-2015 in an area >1.1 M km^2 of the western US (data from USDA APHIS)
- Most recent outbreak (2019-2020) extended range into eastern KS and southern MO
- Initial question: what explains spatial variability in VS occurrence?

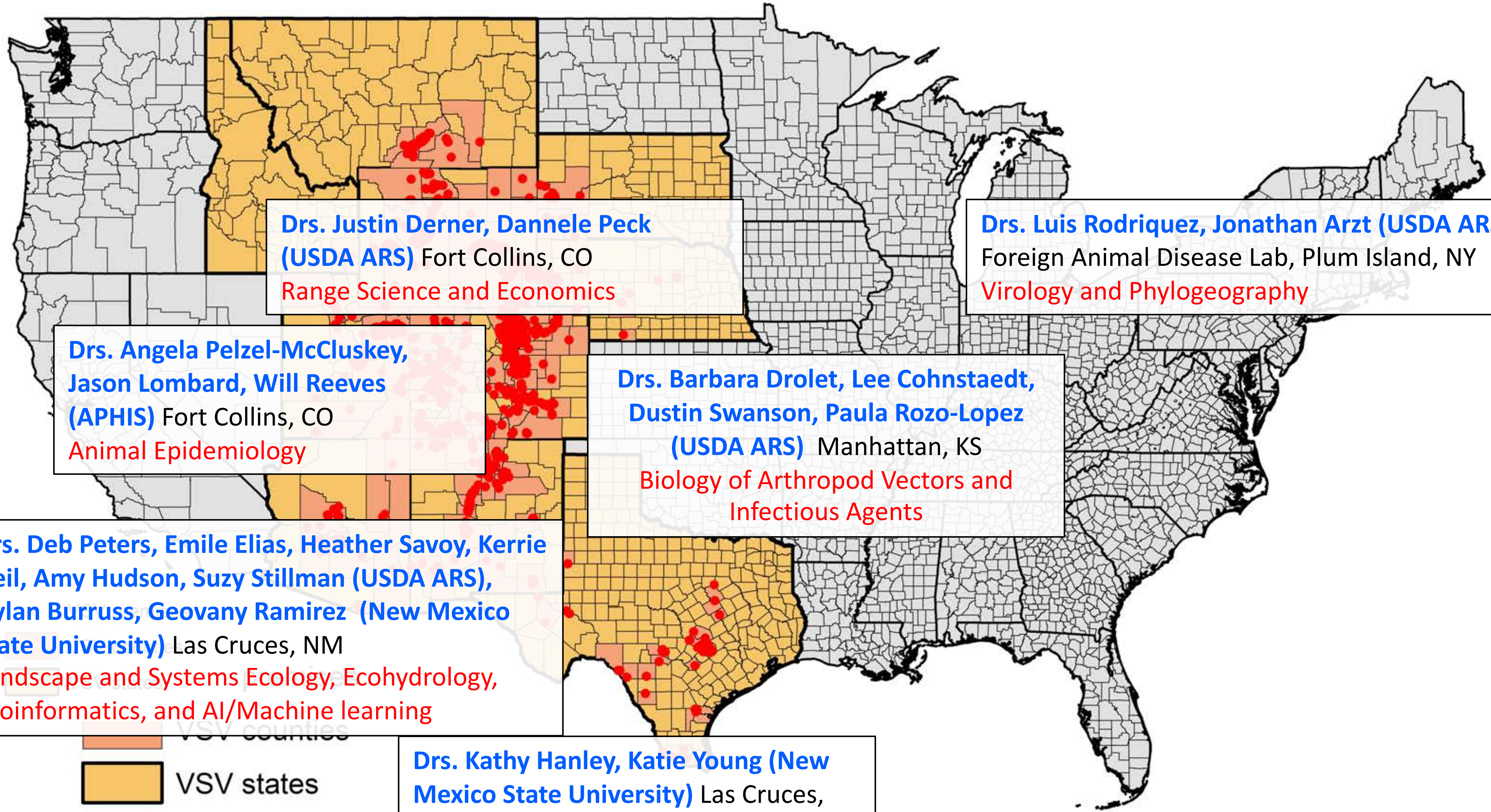
(a)

Outbreak Year	States	Counties	Premises Infected
2004	CO, NM, TX	43	294
2005	AZ, CO, ID, MT, NE, NM, TX, UT, WY	71	445
2006	WY	3	13
2009	NM, TX	3	5
2012	CO, NM, TX	12	36
2014	AZ, CO, NE, TX	32	435
2015	AZ, CO, NE, NM, SD, TX, UT, WY	79	823
2019	CO, KS, NE, NM, OK, TX, UT, WY	114	1142
2020	AR, AZ, KS, MO, NE, NM, OK, TX	61	288





Step 1 – Create a Trans-disciplinary Team



Drs. Justin Derner, Dannele Peck (USDA ARS) Fort Collins, CO
Range Science and Economics

Drs. Luis Rodriquez, Jonathan Arzt (USDA ARS)
Foreign Animal Disease Lab, Plum Island, NY
Virology and Phylogeography

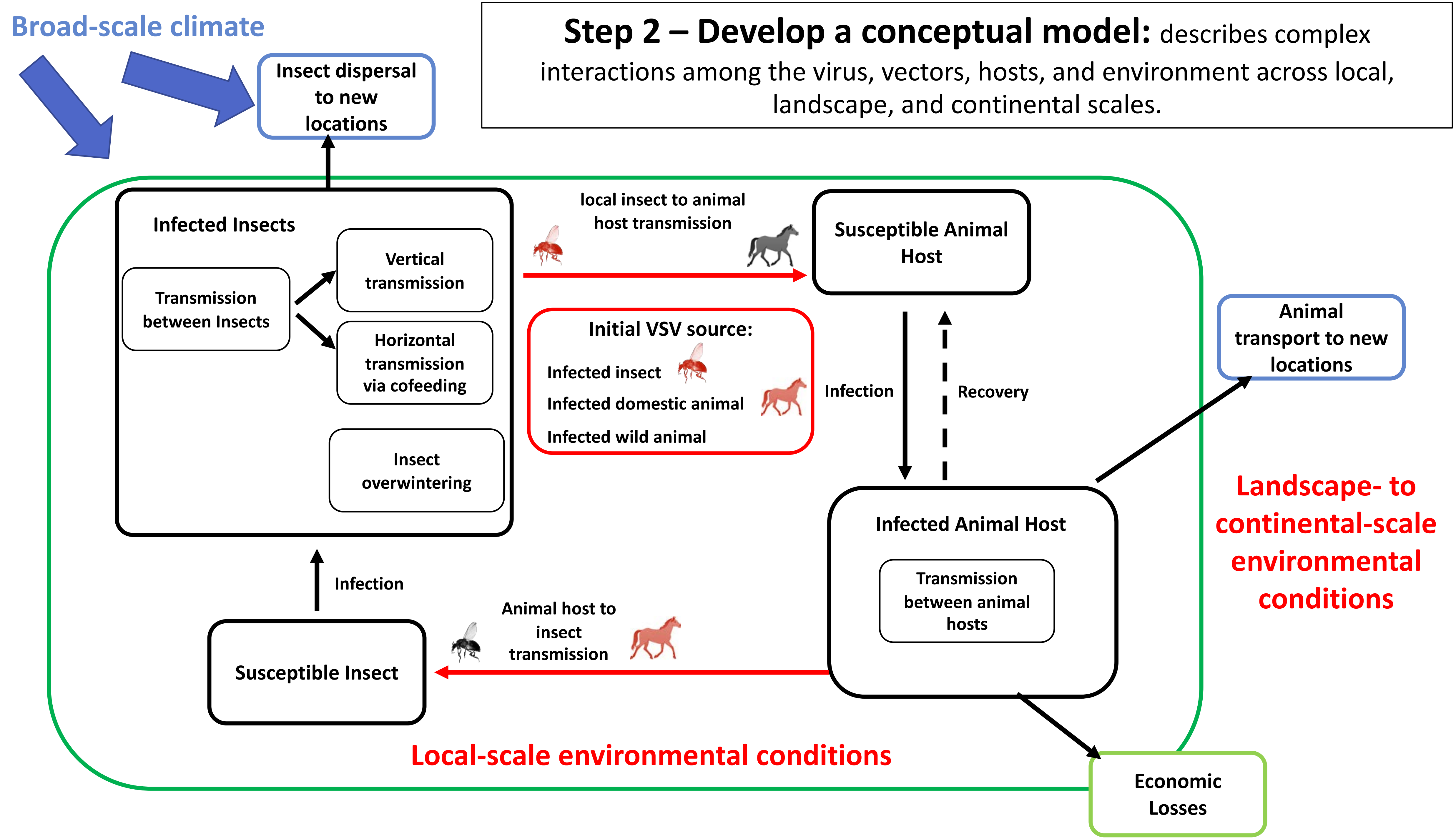
Drs. Angela Pelzel-McCluskey, Jason Lombard, Will Reeves (APHIS) Fort Collins, CO
Animal Epidemiology

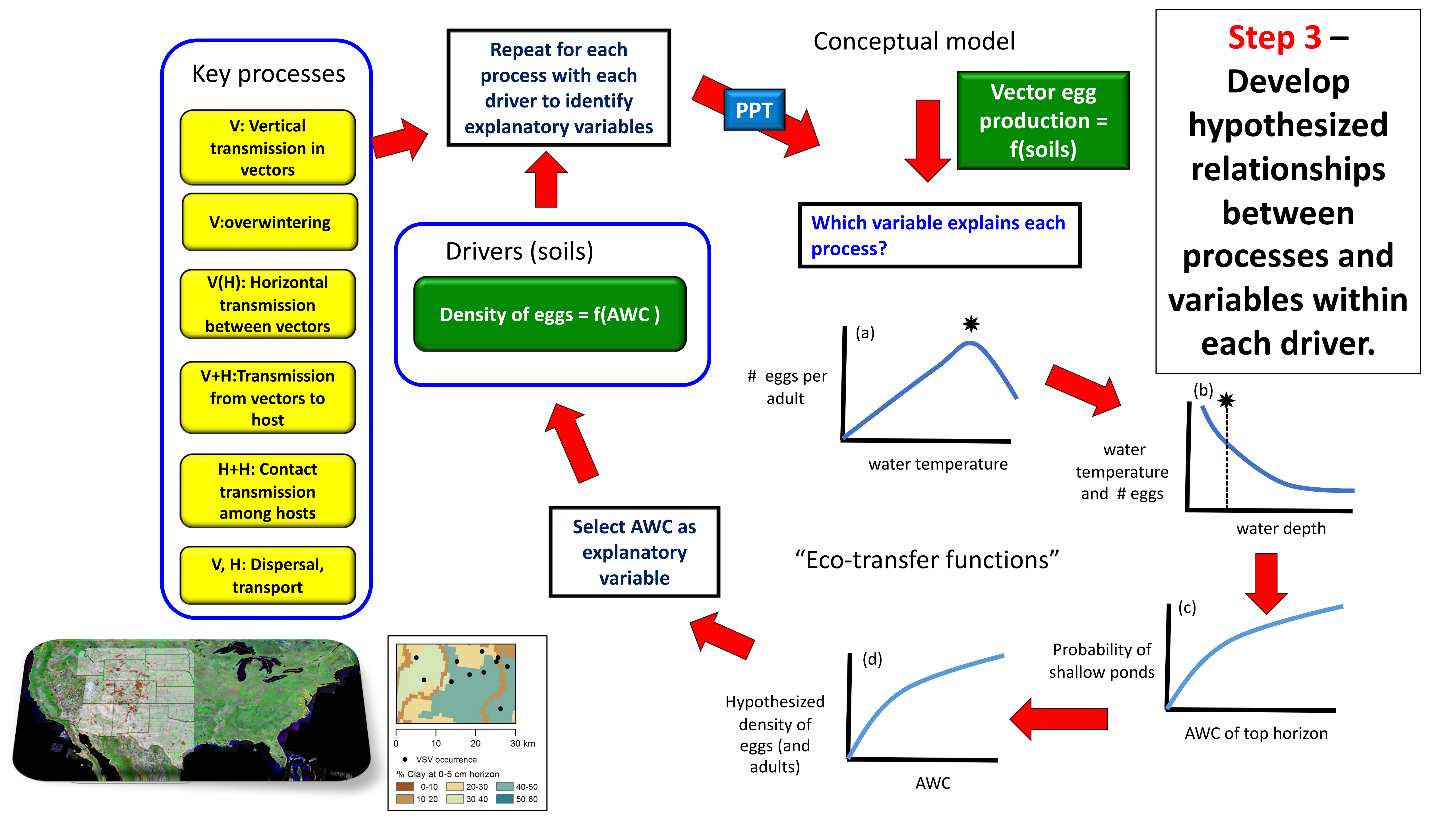
Drs. Barbara Drolet, Lee Cohnstaedt, Dustin Swanson, Paula Rozo-Lopez (USDA ARS) Manhattan, KS
Biology of Arthropod Vectors and Infectious Agents

Drs. Deb Peters, Emile Elias, Heather Savoy, Kerrie Geil, Amy Hudson, Suzy Stillman (USDA ARS), Dylan Burruss, Geovany Ramirez (New Mexico State University) Las Cruces, NM
Landscape and Systems Ecology, Ecohydrology, Ecoinformatics, and AI/Machine learning

Drs. Kathy Hanley, Katie Young (New Mexico State University) Las Cruces, NM Disease ecology, Epidemiology

Step 2 – Develop a conceptual model: describes complex interactions among the virus, vectors, hosts, and environment across local, landscape, and continental scales.




Key processes
V: Vertical transmission in vectors
V:overwintering
V(H): Horizontal transmission between vectors
V+H:Transmission from vectors to host
H+H: Contact transmission among hosts
V, H: Dispersal, transport
Repeat for each process with each driver to identify explanatory variables
PPT
Conceptual model
Vector egg production = f(soils)
Step 3 – Develop hypothesized relationships between processes and variables within each driver.
Which variable explains each process?
Drivers (soils)
Density of eggs = f(AWC)
(a)
eggs per adult
water temperature
(b)
water temperature and # eggs
water depth
Select AWC as explanatory variable
"Eco-transfer functions"
(c)
Probability of shallow ponds
AWC of top horizon
(d)
Hypothesized density of eggs (and adults)
AWC
0
10
20
30 km
VSV occurrence
% Clay at 0-5 cm horizon
0-10
10-20
20-30
30-40
40-50
50-60

Step 4 – Identify the datasets associated with the variables in the eco-transfer functions (e.g., 484 for VS)

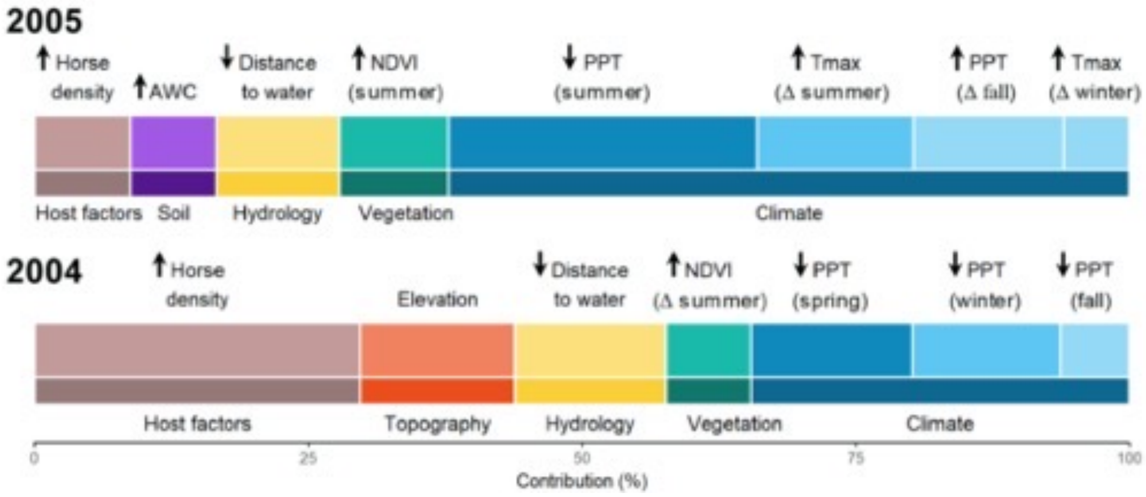
Step 5 – Standardize and harmonize the data in time and space (e.g., VS occurrence data at 4 km x 4 km in continuous grid across western US)

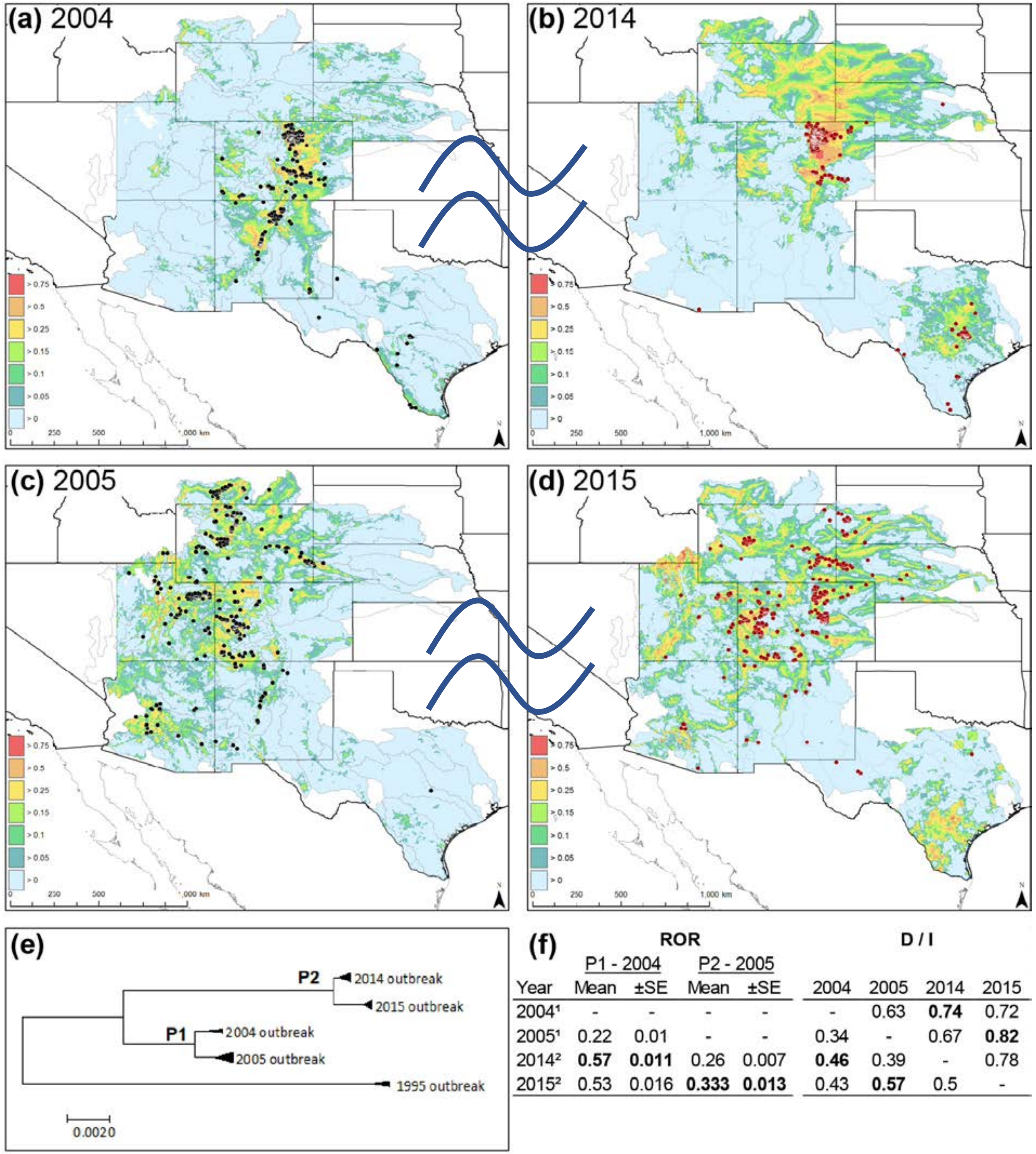
Peters et al. (2020). Ecosphere

Response variables	Source of data	Temporal resolution	Spatial resolution	Variables	State variable
VSV case occurrence	AM Pelzel-McCluskey (USDA-APHIS-VS databases)	Daily data (2003-2016)	point	NA	host
VSV lineage	LL Rodriguez	Daily data (2003-2015)	point	NA	virus
Host factors	**Source of data**	**Temporal resolution**	**Spatial resolution**	**Variables**	**Process+**
Animals[2]	https://quickstats.nass.usda.gov/	2002, 2007, 2012 data	county	Horse (density)	Dispersal (D, C, H)
Animal premises[2]	https://quickstats.nass.usda.gov/	2002, 2007, 2012 data	county	Farm & ranch (density)	Dispersal (D, C, H)
Environmental drivers	**Source of data**	**Temporal resolution**	**Spatial resolution**	**Environmental Variables**	**Process+**
Pedology	http://www.soilinfo.psu.edu/index.cgi?soil_data&conus&data_cov&fract [NRCS]	Static maps [STATSGO]	900 m	Soil properties: % clay, AWC[1]	Biting midge (V)
Hydrology*	https://www.sciencebase.gov/catalog/item/51360134e4b03b8ec4025bfa [USGS]	Static maps	30 m	Location of water bodies	Black fly (V)
	https://waterdata.usgs.gov/nwis/sw [USGS]	Daily data (2003-2016)	30 m	Stream flow	Black fly (V)
	http://giovanni.gsfc.nasa.gov/giovanni/ [NASA]	Monthly data (2003-2016)	12 km	Runoff (cm) Soil moisture (%)	Black fly (V) Biting midge (V)
Topography	http://www2.jpl.nasa.gov/srtm/ [NASA]	Static DEM	900 m	Elevation (m)	OW, V
Climatology*	http://www.prism.oregonstate.edu/normals/ [OSU]	Daily, monthly (2003-2016); long-term average data (1981-2010)	4 km	Minimum, maximum temperature (° C); precipitation (cm)	OW, V V, H
Drought*	http://climate.colostate.edu/~drought [NOAA]	Monthly data (2002-2015)	12 km	Evaporative Demand Drought Index (EDDI)	V, H
Land surface properties*	https://lpdaac.usgs.gov/node/78 [NASA]	Monthly imagery; MODIS (2003-2016)	5.6 km	Vegetation greenness (NDVI)	V, H

+ predominant process(es) expected to be important: D: dispersal; C: contact transmission; H: horizontal transmission; V: vertical transmission; OW: overwintering (other processes are either less important or there is insufficient data on importance)

Steps 6 and 7 – Conduct analyses and interpret results (e.g., machine learning using MaxEnt for distribution models)





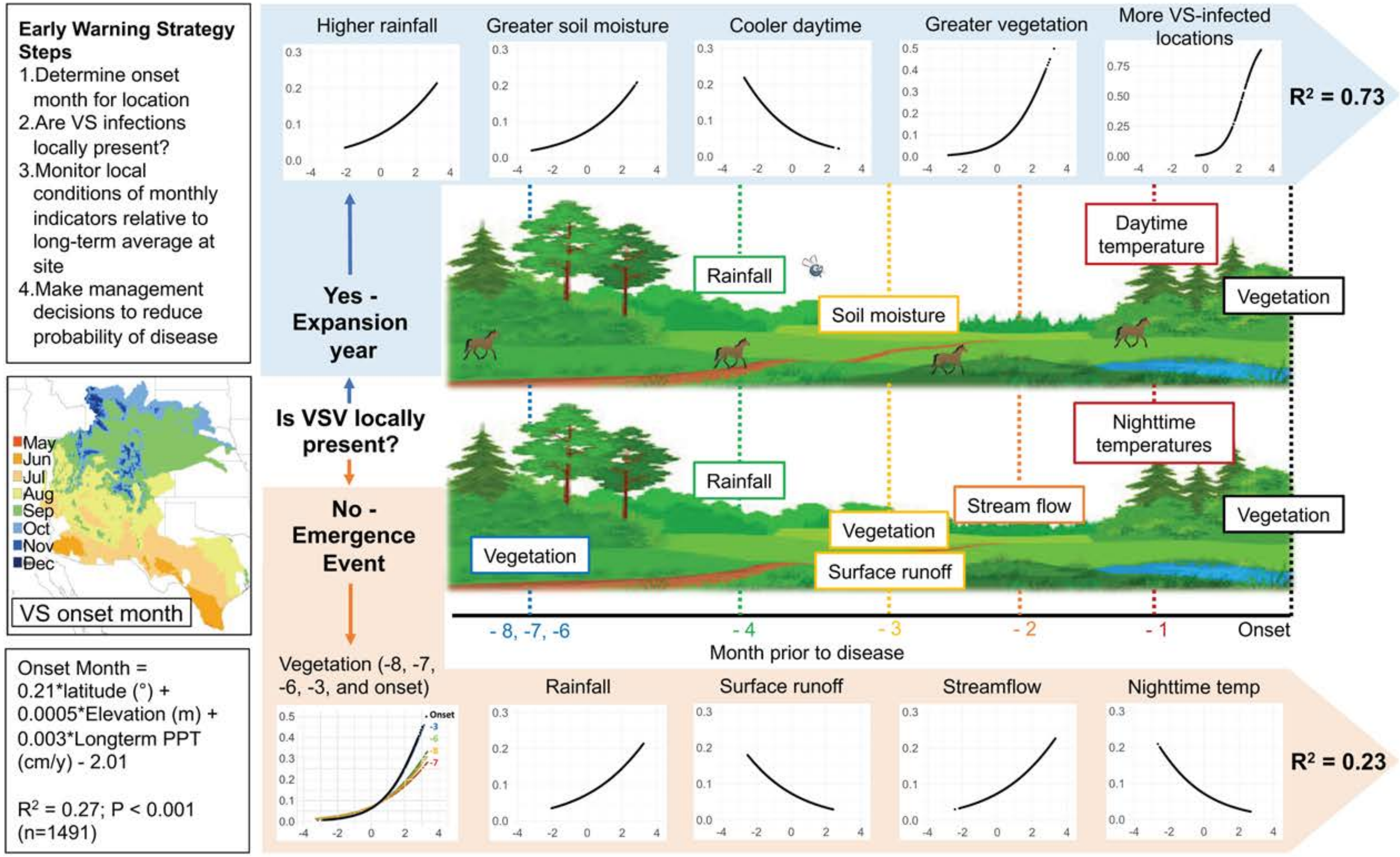
	ROR				D / I			
	P1 - 2004		P2 - 2005					
Year	Mean	±SE	Mean	±SE	2004	2005	2014	2015
2004[1]	-	-	-	-	-	0.63	**0.74**	0.72
2005[1]	0.22	0.01	-	-	0.34	-	0.67	**0.82**
2014[2]	**0.57**	**0.011**	0.26	0.007	**0.46**	0.39	-	0.78
2015[2]	0.53	0.016	**0.333**	**0.013**	0.43	**0.57**	0.5	-

Step 8 – Conduct experiments to test new hypotheses

H: Black flies main vector in incursion years

H: Biting midges important vector in expansion years

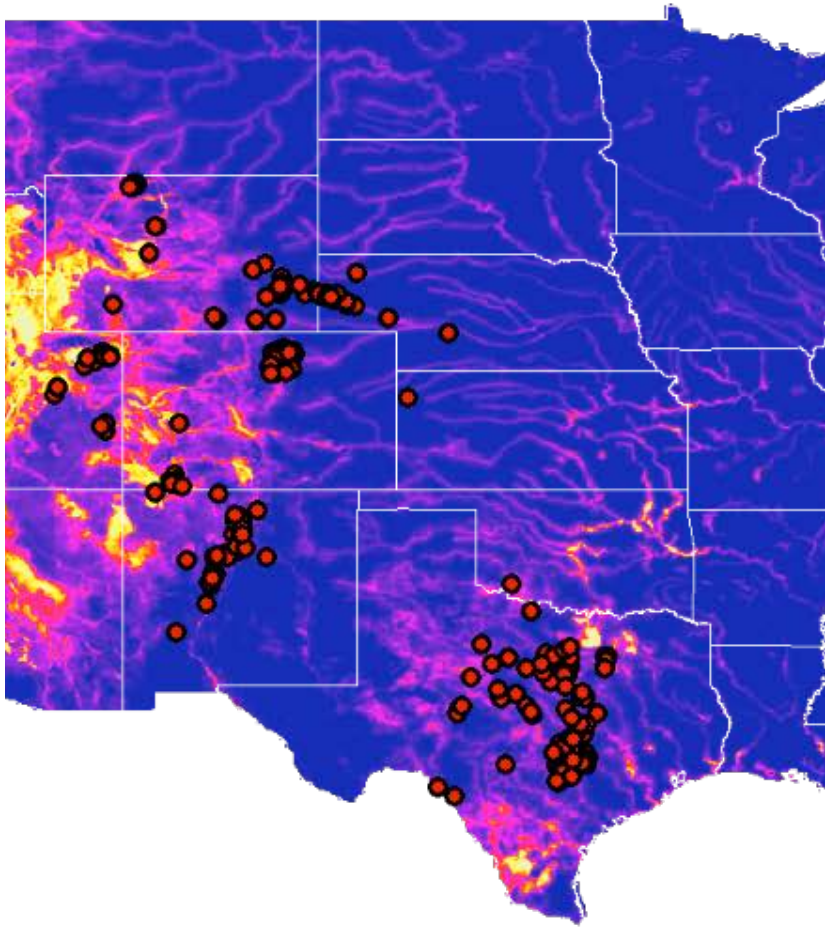
Step 9 – Develop early warning strategies based on variables and processes related to patterns

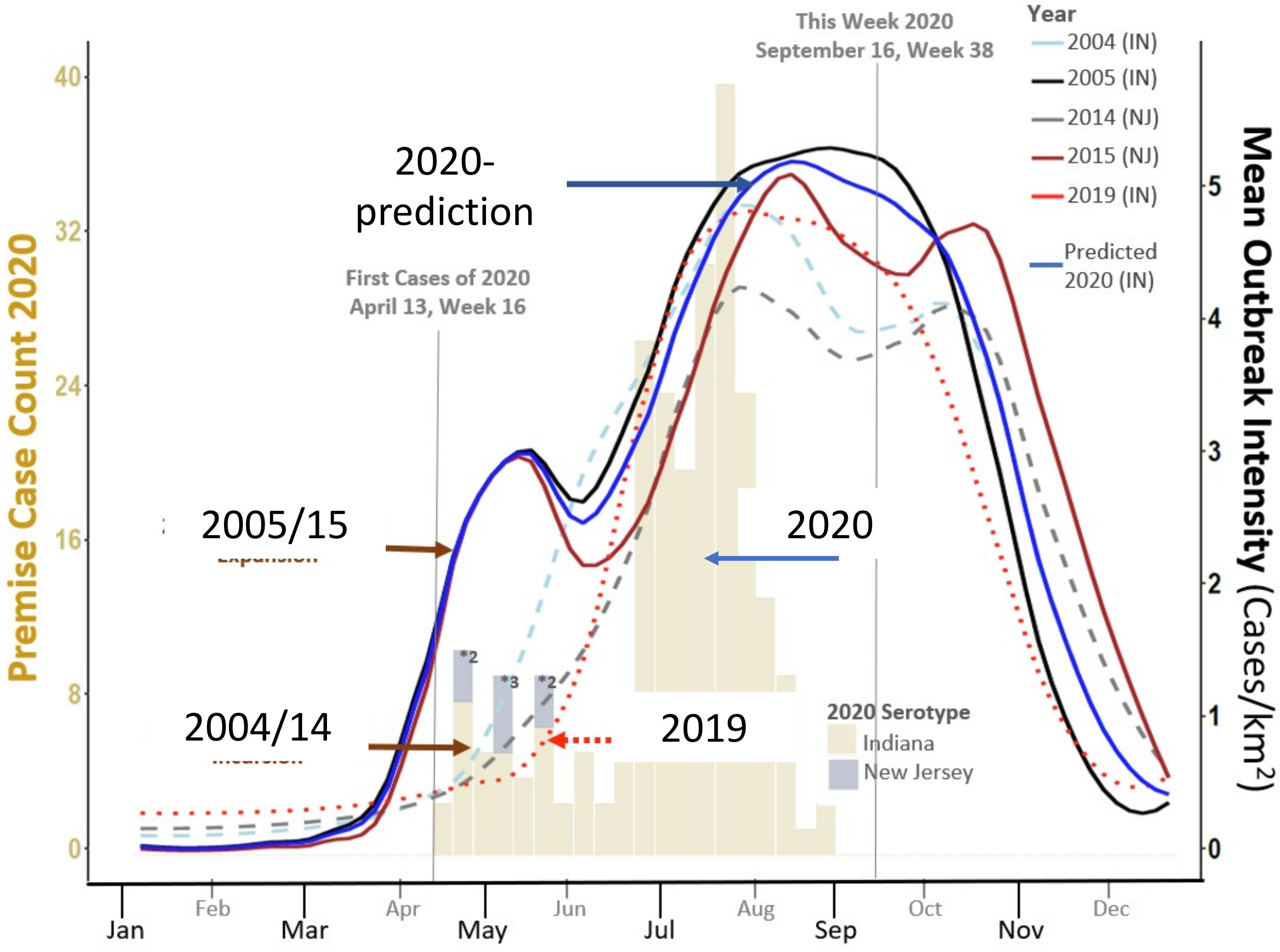


Step 10 – Apply approach to additional vector-borne diseases (e.g., West Nile)

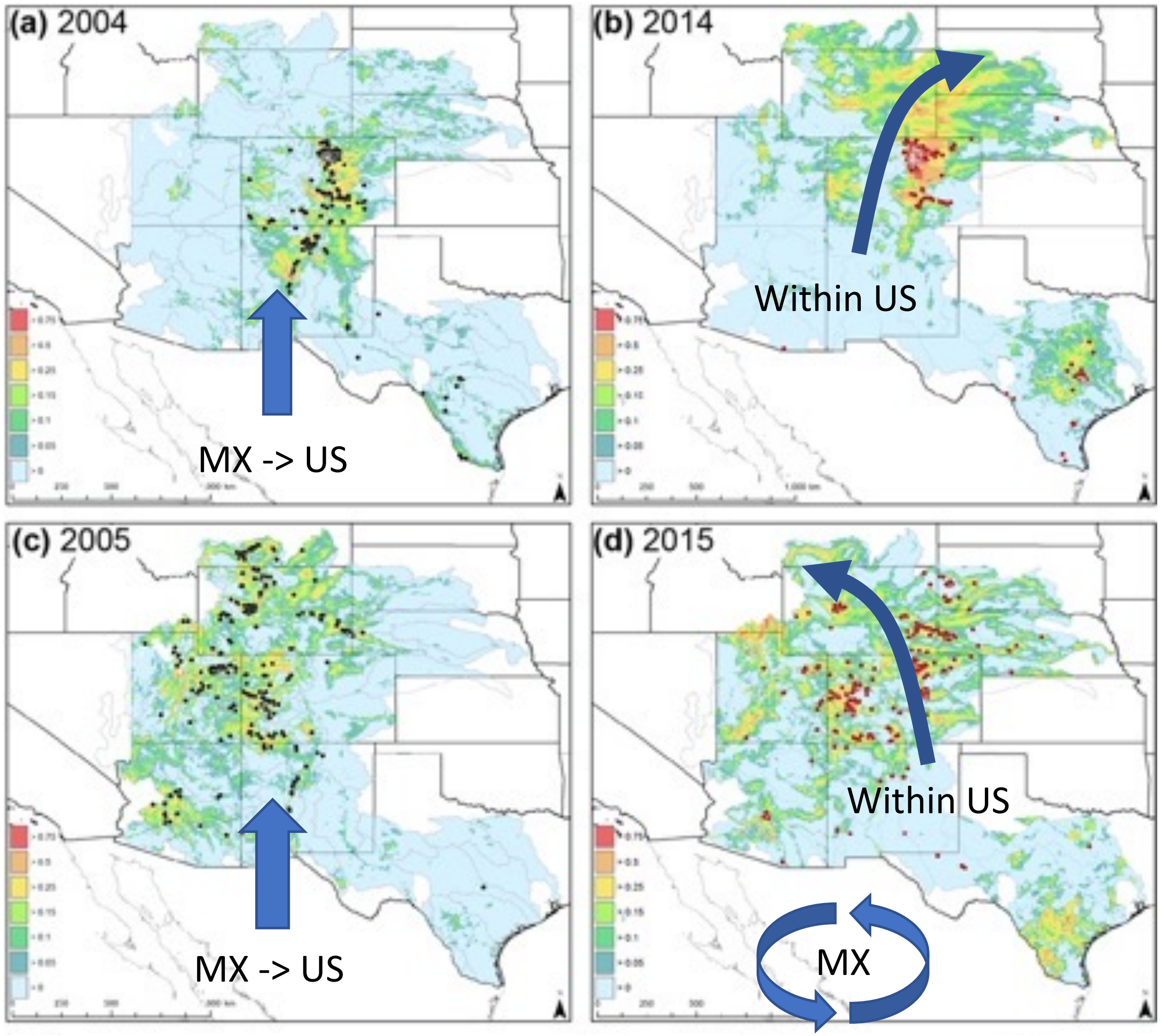
Step 11 – Predict future dynamics and spread of disease based on past outbreaks (VS outbreak in 2019-20 IN serotype; NJ 2004-2015)

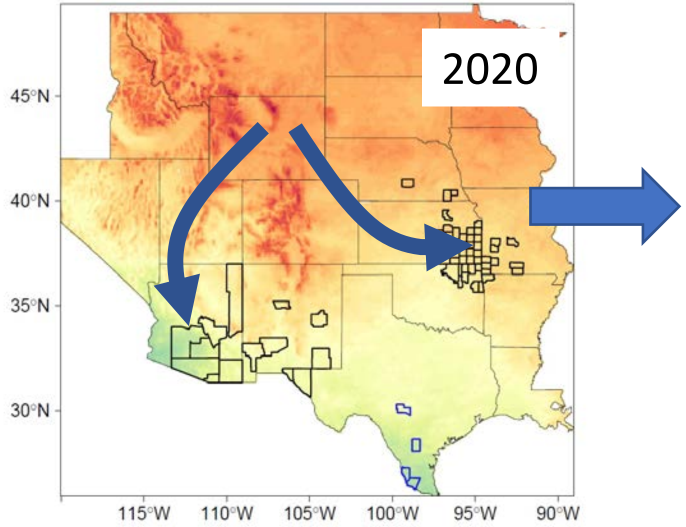
2019 VS-IN (points) using 2005 VS-NJ MaxEnt habitat model





Additional VS dynamics under study (MaxEnt for spatial patterns; deep neural networks; multi-scale numerical models; spatio-temporal models)





Future outbreaks

Changing climate

Insights for End to End Predictions

- Strategically identify a trans-disciplinary team of scientists and technical experts including an expert who can integrate disciplinary interests and system dynamics
- Recognize importance of the iterative process needed to build a meaningful data cube (data + metadata + scientific expertise + technical expertise -> feedback to data)
- Focus on pattern + process relationships to guide data/variable selection and analyses
- Regular team meetings and addition of scientists with new skillsets are needed to maintain a successful trans-disciplinary project through time
- Be adaptive when the next outbreak happens (natural systems are inherently variable)

Challenges

- Limited data availability (e.g., insect vectors response to environment under field conditions)
- Ecological response differences between serotypes (IN, NJ) unknown
- Multiple computational tools needed for different parts of problem require skilled personnel

Key Papers

Peters, D.P.C., Burruss, N.D., Rodriguez, L.L., McVey, D.S., Elias, E.H., Pelzel-McCluskey, A.M., Derner, J.D., Schrader, T.S., Yao, J., Pauszek, S.J., Lombard, J., Archer, S.R., Bestelmeyer, B.T., Browning, D.M., Brungard, C.W., Hatfield, J.L., Hanan, N.P., Herrick, J.E., Okin, G.S., Sala, O.E., Savoy, H.M. and Vivoni, E.R. 2018. An integrated view of complex landscapes: a big data-model integration approach to trans-disciplinary science. BioScience 68: 653–669. https://doi.org/10.1093/biosci/biy069

Elias, E., McVey, D.S., Peters, D., Derner, J.D., Pelzel-McCluskey, A., Schrader, T.S. and Rodriguez, L. 2019. Contributions of Hydrology to Vesicular Stomatitis Virus Emergence in the Western USA. Ecosystems 22: 416–433. https://doi.org/10.1007/s10021-018-0278-5

Peters, D.P.C., McVey, D.S., Elias, E.H., Pelzel-McCluskey, A.M., Derner, J.D., Burruss, N.D., Schrader, T.S., Yao, J., Pauszek, S.J., Lombard, J., and Rodriguez, L.L.. 2020. Big data-model integration and AI for vector-borne disease prediction. Ecosphere 11: e03157. https://doi.org/10.1002/ecs2.3157

VS Story Map: https://arcg.is/08fub5