Citizen Science Helps Predict Risk of Emerging Infectious Disease

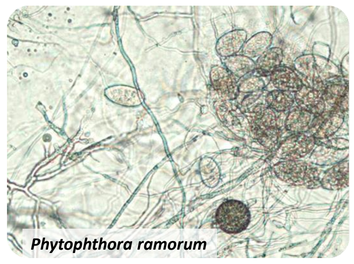
Email: rkmeente@ncsu.edu

Ross K. Meentemeyer[1], John B. Vogler[1], and Matteo Garbelotto[2]

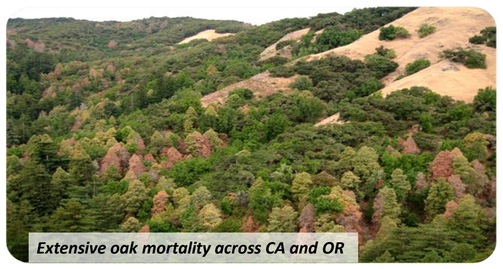
http://gis.ncsu.edu

[1]Center for Earth Observation, Department of Forestry & Environmental Resources, North Carolina State University; [2]Forest Pathology & Mycology Laboratory, University of California, Berkeley

I. INTRODUCTION


Phytophthora ramorum


Symptomatic Leaves


Extensive oak mortality across CA and OR

Citizen science holds great potential for **advancing spatial prediction of biological invasions** by providing inexpensive **location-based, time series data** of unprecedented quantity and distribution. In 2008, we developed a citizen science program to detect the spread of the emerging forest disease **Sudden Oak Death (SOD)** across the metropolitan region of the San Francisco Bay Area in California, including under-sampled habitat within urban areas and along the wildland-urban interface. Each year, our "**SOD Blitz**" program used **crowdsourcing methods** to encourage citizens to collect leaf tissue symptomatic of this disease and submit it to our lab for molecular diagnosis. **Hundreds of volunteers** contributed **5,861 georeferenced samples** between 2008 and 2012 with 3,600 (61.4%) samples testing positive. Geospatial and statistical analyses of sample distributions and disease-environment interactions through time, and assessment of questionnaires linked to submitted samples helped us answer questions about the:

Sudden Oak Death
- Causal agent: *Phytophthora ramorum*
- *P. ramorum* introduced in mid-1990's
- Origin of *P. ramorum* is not known
- Disease outbreaks in coastal forests of CA and OR
- Killed millions of oak (*Quercus* sp.) & tanoak (*Lithocarpus densiflora*) trees

- ❑ **Spatial scale of disease clustering over time**
- ❑ **Environmental and societal drivers of disease risk**
- ❑ **Locations at risk of disease spread in subsequent years**
- ❑ **Disease detection success of citizen science participants**

II. The SOD BLITZes – CITIZEN SCIENCE at WORK

1. Hold community meetings each Spring

- ❑ Train participants to identify SOD symptoms on hosts
- ❑ Explain data collection process (sampling, forms, GPS)
- ❑ Define collection areas and distribute materials

2. Collect symptomatic leaf tissue

- ❑ Collection by individual participants on weekends
- ❑ Bag samples, tag trees, record data, locate w/ GPS
- ❑ Submit samples and forms at central location

3. Test samples for P. ramorum

- ❑ Testing at UC Berkeley diagnostic laboratory
- ❑ Polymerase chain reaction (PCR) analysis of DNA
- ❑ PCR eliminates false negatives & false positives

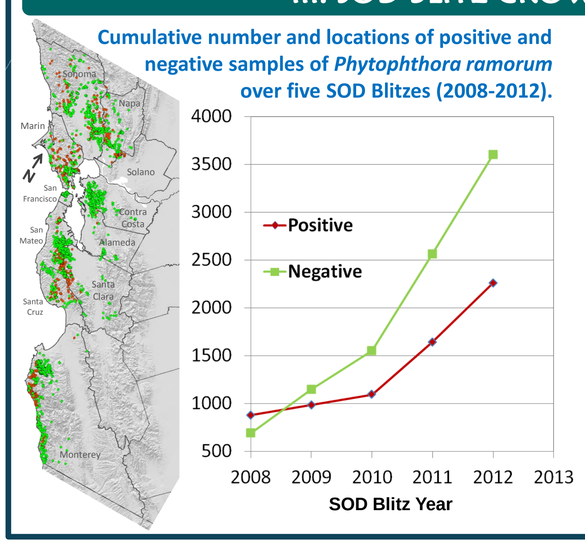
4. Map disease distribution

- ❑ Produce & share maps of presence AND absence data
- ❑ Discuss management options with community

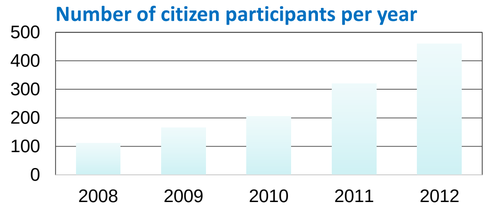
-> Learn more at sodblitz.org & sodmap.org

III. SOD BLITZ CROWDSOURCED DATA

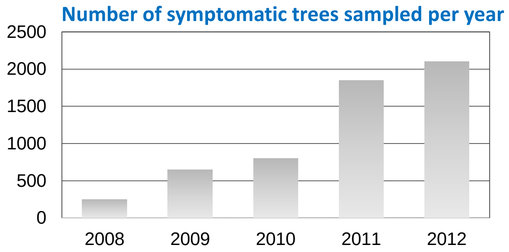
Cumulative number and locations of positive and negative samples of *Phytophthora ramorum* over five SOD Blitzes (2008-2012).



Number of citizen participants per year



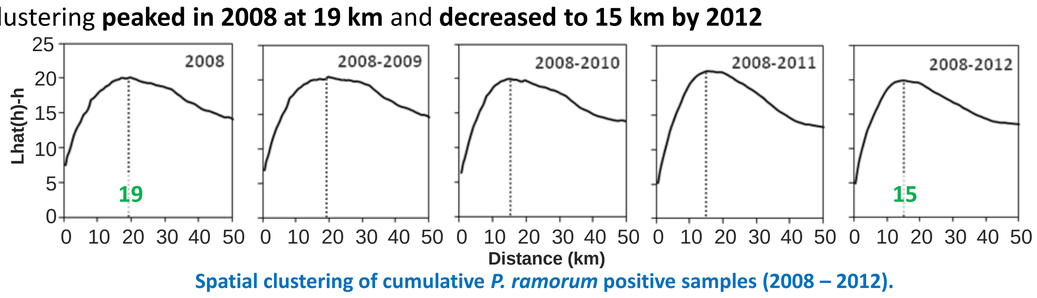
Number of symptomatic trees sampled per year



IV. KNOWLEDGE AND DISCOVERY

1. What did the SOD Blitz tell us about the spatial scale that disease clusters over time?

- ➢ Spatial cluster analysis (*Ripley's K with Lhat-h transformation*) of pathogen occurrence revealed:
  - ❑ Disease **clustering decreased through time** as the disease spread further
  - ❑ Clustering **peaked in 2008 at 19 km** and **decreased to 15 km by 2012**



Spatial clustering of cumulative *P. ramorum* positive samples (2008 – 2012).

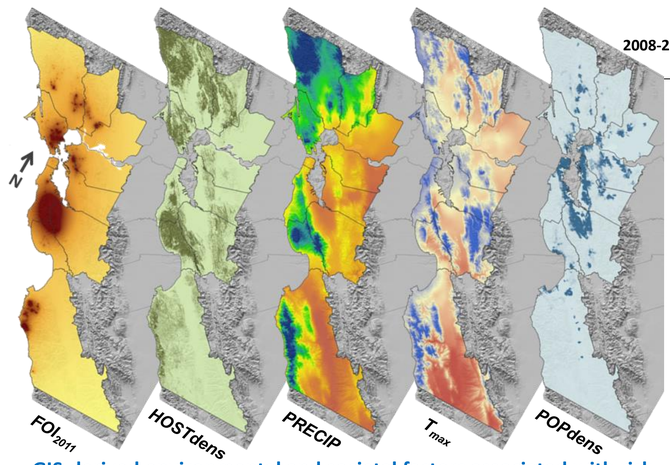
2. Did the SOD Blitz improve our understanding of the drivers of disease risk over time?

- ➢ *GLM logistic regression models* of disease-environment interactions, calibrated and analyzed incrementally following each year of the SOD Blitz, showed that:
  - ❑ Through 2010, **disease risk increased** each year at locations **with higher levels** of **force of infection (*FOI*), host density (*HOSTdens*),** and **precipitation (*PRECIP*)**
  - ❑ In 2011, additionally, we found a **negative association** with **max. temperature (T_{max})**
  - ❑ In 2012, **human population density** (***POPdens***) emerged as a fifth factor with infection risk **lower in more populated urban locations**

Generalized linear logistic regression models of *P. ramorum* infection probability.

Model	Sample size	Equation	Predictive accuracy
2008	879(+) 689(-)	6.19882 + 0.16551*FOI_{2007} + 0.01188*$HOST_{dens}$ + 0.00737**PRECIP*	0.61
2008-2009	983(+) 1145(-)	5.1664 + 0.05235*FOI_{2008} + 0.01488*$HOST_{dens}$ + 0.01671**PRECIP*	0.67
2008-2010	1093(+) 1551(-)	4.56107 + 0.03154*FOI_{2009} + 0.01830*$HOST_{dens}$ + 0.01491**PRECIP*	0.72
2008-2011	1640(+) 2562(-)	0.06359 + 0.02408*FOI_{2010} + 0.01728*$HOST_{dens}$ + 0.01741**PRECIP* - 0.25731*T_{max}	0.71
2008-2012	2261(+) 3600(-)	0.03347 + 0.01418*FOI_{2011} + 0.01897*$HOST_{dens}$ + 0.01242**PRECIP* - 0.21431*T_{max} - 0.00019*POP_{dens}	N/A*

Validation



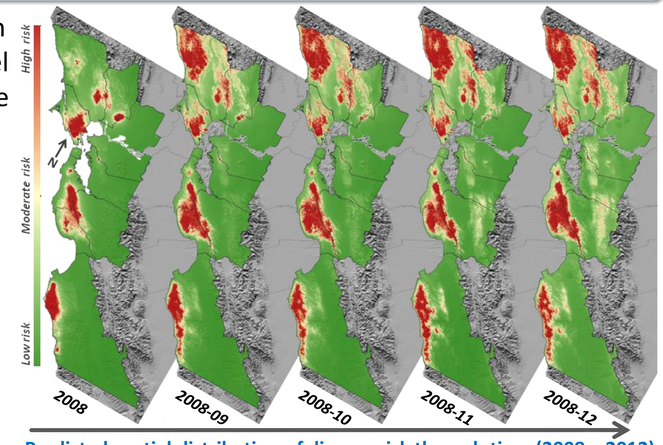
GIS-derived environmental and societal factors associated with risk of *P. ramorum* infection.

- ***Predictive accuracy*** refers to model accuracy when applied to independent presence-absence data from a subsequent year of SOD Blitz sampling
- FOI_{20XX} refers to force of infection based on distribution of SOD-infected locations in year indicated
- ***HOSTdens*** is the density of foliar host species
- ***PRECIP*** is the mean monthly precipitation over 30 years (1981-2010) during Dec-May wet season
- T_{max} is the mean monthly maximum temperature over 30 years (1081-2010) and during Dec-May wet season
- ***POPdens*** is the number of persons per habitable square kilometer based on 2010 Census data
- All variables in each model are significant at $P < 0.001$.
- All models are significant at $P < 0.0001$.

* Validation of 2008-2012 model is possible with release of SOD Blitz 2013 results.

3. Can we use the SOD Blitz to predict locations at risk of disease?

- ➢ Each year, in the GIS, we applied the equation of the corresponding logistic regression model to maps of the significant variables to produce *spatially-explicit predictions of infection probability* through time. We found:
  - ❑ **Greatest risk** emerged in **coastal forests** of **Monterey**, **Santa Cruz**, **San Mateo**, **Marin**, and northern **Sonoma** Counties
  - ❑ Models predicted **increases over time in amount of land area facing high risk**
  - ❑ **Ability to predict risk of disease spread** – based on previous years of sampling – **improved 10%** (61% - 71%) during SOD Blitz



Predicted spatial distribution of disease risk through time (2008 – 2012).

4. Did education and experience of citizen science participants affect probability of disease detection?

- ➢ *A/B split testing* of proportional leaf tissue sample outcomes (positive & negative for *P. ramorum*) between laypersons and professionals revealed:
  - ❑ In **2011**, **laypersons** had **higher detection success**
  - ❑ In **2012**, both **groups performed equally** well

		Successes	Total	Success Rate	p-value	Improvement
2011	Professional	14	74	11% – 29% (19%)	—	—
	Layperson	74	225	27% – 39% (33%)	0.022	9.2% – 115% (74%)
2012	Professional	106	375	24% – 33% (28%)	—	—
	Layperson	298	1,045	26% – 31% (29%)	0.93	-18% – 19% (0.88%)

Split tests of disease detection success rates among participant groups.

V. CONCLUSIONS -> CITIZEN SCIENCE IS WORKING for SOD!

1) Our ability to predict risk of disease spread is improving with each year of the SOD Blitz, confirming the importance of continued annual Blitzes. **2)** Citizens with little or no prior knowledge of plant pathology and SOD are just as likely to contribute positive samples as self-identified professionals, confirming the effectiveness of SOD Blitz informational sessions and training efforts. **3)** Citizen science is producing valuable, time series data for understanding locations at risk of disease spread and improving our understanding of this emerging infectious disease.

Research funded by: NSF, USFS, Gordon and Betty Moore Foundation