Citizen science helps predict risk of emerging infectious disease

Ross K Meentemeyer[1,2*], Monica A Dorning[2], John B Vogler[2], Douglas Schmidt[3], and Matteo Garbelotto[3,4]

Engaging citizen scientists is becoming an increasingly popular technique for collecting large amounts of ecological data while also creating an avenue for outreach and public support for research. Here we describe a unique study, in which citizen scientists played a key role in the spatial prediction of an emerging infectious disease. The yearly citizen-science program called "Sudden Oak Death (SOD) Blitz" engages and educates volunteers in detecting the causal pathogen during peak windows of seasonal disease expression. We used these data – many of which were collected from under-sampled urban ecosystems – to develop predictive maps of disease risk and to inform stakeholders on where they should prioritize management efforts. We found that continuing the SOD Blitz program over 6 consecutive years improved our understanding of disease dynamics and increased the accuracy of our predictive models. We also found that self-identified non-professionals were just as capable of detecting the disease as were professionals. Our results indicate that using long-term citizen-science data to predict the risk of emerging infectious plant diseases in urban ecosystems holds substantial promise.

Front Ecol Environ 2015; 13(4): 189–194, doi:10.1890/140299

Mitigating threats to biodiversity and ecosystem function from unexpected outbreaks of emerging infectious disease hinges on scientists' ability to detect and predict disease spread across broad spatial extents in a timely manner (Crowl *et al.* 2008). The economic cost of collecting sufficient data to develop empirical models of non-human diseases, such as those affecting plants and wildlife in ecological communities, is often prohibitive but may be offset by involving volunteer citizen scientists in the data collection process. Citizen-science programs offer promising new approaches for increasing the extent and frequency of sampling efforts (Dickinson *et al.* 2012). For example, the world is beginning to see potential for accelerated responses to natural disasters – through rapid compilation of volunteered geographic information at the forefront of an event (Goodchild and Glennon 2010). Currently, engaging citizen volunteers in collecting timely georeferenced data on the spread of emerging pathogens is an under-explored opportunity that could be used to help predict disease risk while simultaneously educating the public about disease control and prevention and involving stakeholders in the scientific and planning processes (Dickinson *et al.* 2012).

Citizen scientists are increasingly being called upon to survey the abundance and distribution of organisms (see review in Dickinson *et al.* 2010), but with few notable applications focused on pests and pathogens (eg ZomBee Watch, House Finch Disease Survey). When responding to the threat of emerging diseases, the goal is often not just to monitor spread but to predict locations where future outbreaks are imminent. However, predictive models are only as good as the available data and, as such, concerns about increased observer error and sampling bias in citizen-science observations are still common (Crall *et al.* 2011; Kremen *et al.* 2011). Fortunately, assessment of data quality and adequate training for volunteers can reduce observer error (Gardiner *et al.* 2012), and targeted sampling strategies can limit sampling bias (Dickinson *et al.* 2010). Targeted approaches are especially useful for analyzing seasonal events and directing data collection to coincide with peak windows for observation (eg timing of disease expression, flight of migratory birds). Citizen-science programs that limit observer error and sampling bias through targeted sampling techniques and well-designed volunteer training sessions can also produce extensive datasets that are critically needed for predicting disease spread with empirical models.

Using the Sudden Oak Death (SOD) Blitz program (www.sodblitz.org) as a case study, we focus on two questions that shed light on the value of citizen science augmented with geographic information and crowdsourcing (collecting data by soliciting contributions from the public) for responding to emerging infectious diseases: (1) does the SOD Blitz improve our understanding of pathogen habitat and our ability to predict disease risk?, and (2) did the educational background and professional experience of our citizen-science participants affect the probability of disease detection? The presence of SOD is a major public concern in coastal California and Oregon due to the widespread mortality of millions of socially and ecologically important trees. SOD is caused by the generalist and invasive plant pathogen *Phytophthora ramorum*,

[1]*Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC* [*]*(rkmeente@ncsu.edu);* [2]*Center for Geospatial Analytics, North Carolina State University, Raleigh, NC;* [3]*Forest Pathology and Mycology Laboratory, University of California–Berkeley, Berkeley, CA;* [4]*Department of Environmental Science, Policy, and Management, University of California–Berkeley, Berkeley, CA*

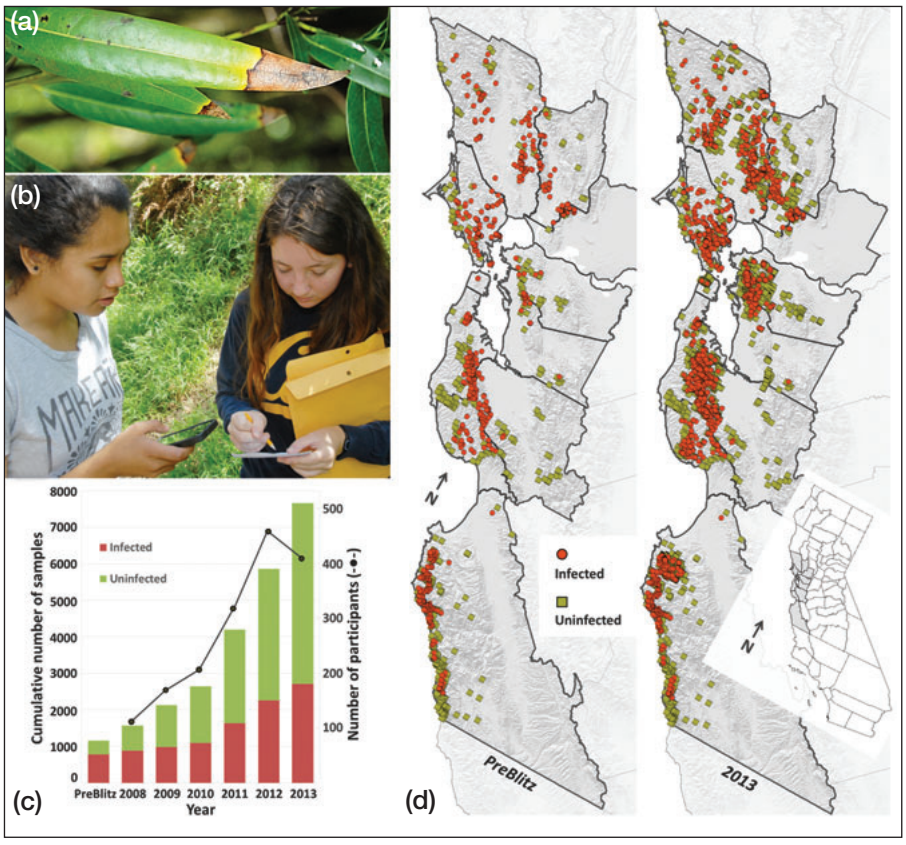

Figure 1. *(a) Symptomatic leaves on the reservoir host, California bay laurel (*Umbellularia californica*), Santa Cruz County, CA. (b) Young citizen scientists locating and collecting symptomatic leaves with the aid of the SODmap mobile application, Alameda County, CA. (c) Cumulative infected (red) and uninfected (green) foliar samples collected PreBlitz and by citizen-science participants over the 6-year Blitz sampling period between 2008 and 2013. (d) Locations of PreBlitz and cumulative SOD Blitz samples by 2013 within the 11-county study extent. (a) and (b) Courtesy of M Garbelotto Lab, UC Berkeley.*

which affects numerous plant species and is killing millions of oak (*Quercus* spp) and tanoak (*Notholithocarpus densiflorus*) trees in California and Oregon (Rizzo and Garbelotto 2003; Cobb *et al.* 2013). In the early 2000s, the first SOD monitoring networks were established on public lands, with very few observations collected in metropolitan areas or at the wildland–urban interface. This gap in our observation network has hindered our ability to predict disease spread and effectively prioritize future detection and management efforts, particularly in areas where residents are being affected by loss of trees. The multiyear SOD Blitz citizen-science program expands monitoring and detection in under-sampled urban areas and helps private landowners protect threatened trees.

Methods

The SOD Blitz program

The SOD Blitz program was initiated in 2008 and continues today with the goal of tracking disease spread and prioritizing treatment of vulnerable trees at high-risk locations in California. Crowdsourcing of SOD Blitz volunteers allows us to use targeted sampling techniques to collect data during the narrow window of peak disease expression (ie display of symptomatic leaves) and in areas previously underrepresented in our sampling efforts, including urban areas and private properties located throughout the central California coastal region. Each year, we advertise Blitz events via press releases, televised news stories, newspaper and radio announcements, email campaigns, and two websites (www.suddenoakdeath.org and www.matteolab.org). We also tap into networks of community groups focused on conservation, green spaces, herbaria, nature preserves, amateur mycology, and botany. SOD Blitzes are integrated into these groups' official monthly activities, which helps ensure high participation levels. Participants originate from a wide spectrum of communities, from high-school students and concerned homeowners to docents, arborists, firefighters, and K–12 teachers.

All of our program's citizen scientists participate in workshops to gain a general education on the history and impacts of the disease, and for on-the-ground training to detect disease symptoms and collect symptomatic plant tissue (Figure 1a) for laboratory diagnosis. During the "Blitzes", trained participants use a symptom detection guide and a mobile mapping tool (Garbelotto *et al.* 2014) to identify and map symptomatic trees (Figure 1b). Storage packets for symptomatic tissue allow each participant to sample leaves from up to 16 trees during a given annual Blitz. Submitted leaf tissue samples undergo species-specific molecular assays involving polymerase chain reaction (PCR) analysis (Hayden *et al.* 2006) to determine presence or absence of the pathogen. Maps of disease distribution and risk predictions (described below) are available online (www.sodmap.org), enabling participants to visualize their efforts and understand the threats posed by SOD.

Predicting disease risk

For each year from 2008 through 2013, we combined disease presence and absence data (leaf sample locations linked to PCR results) collected by participants with other research observations collected prior to the SOD Blitz program (PreBlitz samples 2000–2007; see description of species data in Václavík *et al.* 2012). We analyzed the rela-

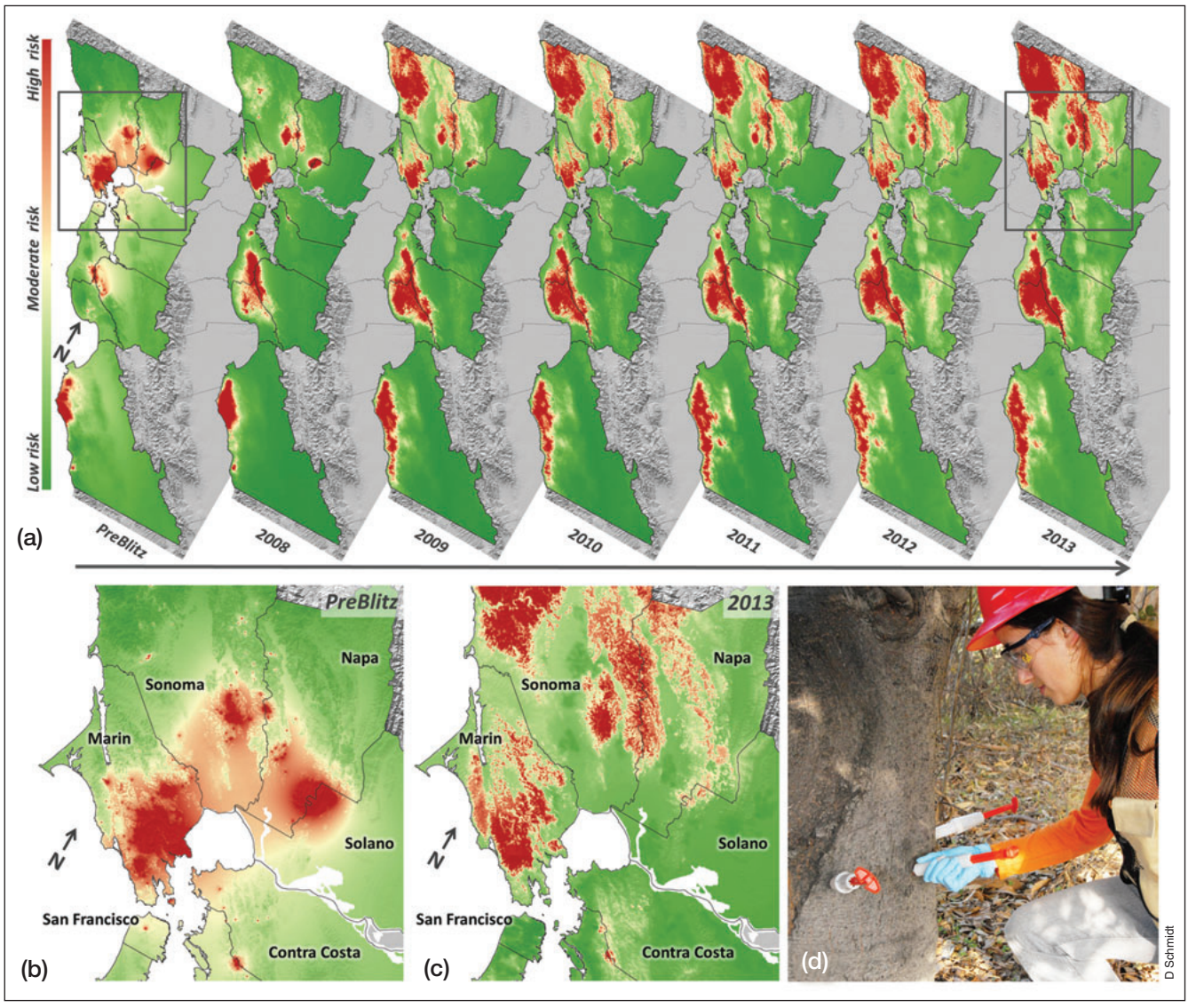

Figure 2. *(a) Predicted spatial distribution of disease risk through time (PreBlitz – 2013). (b) Detail view of predicted disease risk (PreBlitz inset) in counties north of San Francisco. (c) Detail view of predicted disease risk (2013 inset) based on cumulative SOD data (PreBlitz – 2013). (d) Application of preventative treatment (phosphonate compound injections) to a valuable host tree, coast live oak (*Quercus agrifolia*), located in a high-risk area, Marin County, CA. (d) Courtesy of M Garbelotto Lab, UC Berkeley.*

tive importance of environmental and societal drivers of the probability of infection by evaluating all possible generalized linear model (GLM) regression equations from combinations of hypothesized site suitability factors (WebTable 1) and selecting the best model based on Akaike's Information Criterion (AIC). After each annual SOD Blitz, we used the locations of the positive samples to calculate a negative exponential dispersal kernel (Meentemeyer *et al.* 2012) that estimated the "force of invasion" expected across the landscape during the subsequent year of sampling to understand changes in inoculum load through time and space (WebTable 1; WebFigure 1a). We developed seven models (PreBlitz and 2008–2013), with each model generated by incremental inclusion of one additional year of observations and knowledge of the force of invasion at any given site. Within the geographic information system (GIS), we applied the equation of the best model for each year to produce spatially explicit maps of infection probability through time for the entire study extent (Figure 2a). Locations were considered "high risk" if infection likelihood exceeded the optimal threshold probability according to the receiver operating characteristic (ROC; Manel *et al.* 2001). We validated each model by comparing infection risk probability maps to locations of positive and negative samples collected in the following year of the SOD Blitz; overall predictive accuracy (WebTable 2) is based on correct predictions of a subsequent year's positive and negative samples in high- and low-risk areas, respectively.

Evaluating SOD Blitz volunteers

Beginning with the 2011 SOD Blitz, we administered questionnaires to assess participant backgrounds, and

prior knowledge of and experiences with SOD and the SOD Blitzes. We linked each participant questionnaire to a submitted and lab-verified foliar sample. We then used the Z test for difference of proportions to determine whether the chance of successful detection of *P ramorum* infection differs significantly between self-identified professionals (eg participants in research, extension, plant pathology, forestry, or environmental fields) and self-identified non-professionals with little or no prior knowledge of SOD biology. We also analyzed changes within each group over time.

Results and discussion

Our SOD Blitz program increased the number of foliar samples (Figure 1a) in the central California SOD detection network by more than 560% and increased the mean density of samples by 475% (from 0.04 km^{-2} to 0.23 km^{-2}). More than 1600 citizens (Figure 1b) participated in the program between 2008 and 2013, generating 6504 georeferenced foliar samples of potential *P ramorum* infection (1929 confirmed infected; 4575 confirmed uninfected) (Figure 1, c and d). PreBlitz samples collected between 2000 and 2007 from multiple sources indicated 786 infected and 373 uninfected sites (total: 1159) (Figure 1d). Explicit observations of disease absence are critical to predicting species distributions (Václavík and Meentemeyer 2009); thus, the substantial increase in the number of confirmed uninfected sites is particularly valuable for predicting disease risk. The number of volunteers and foliar samples collected each year have generally increased since the start of the program (Figure 1c), with many volunteers (10–20% in a given year) returning for subsequent Blitz events after completing their first year of sampling.

Question 1: does the SOD Blitz improve our understanding of pathogen habitat and our ability to predict disease risk?

In models for all years, the risk of infection was greater at locations with (1) increasing force of invasion (WebFigure 1a) during the previous year, (2) greater host plant density (WebFigure 1b), and (3) higher average monthly precipitation during the wet season (WebFigure 1c; WebTable 2). As the program progressed and more data were collected, new variables emerged as predictors of infection risk. In the 2011 model, mean maximum temperature during the wet season (WebFigure 1d) was added as a significant predictor, having a negative influence on infection risk. In both the 2012 and 2013 models – based on yet more citizen-science data – human population density (WebFigure 1e) emerged as a fifth significant factor; risk of infection was lower in more populated urban locations. While these results are generally consistent with previous research on habitat conditions conducive to disease spread (eg Meentemeyer *et al.* 2008), our analysis of citizen-science data collected in 2012 and 2013 revealed the possibility of a different association between human population density and risk of infection (WebTable 2).

Our ability to predict risk of infection – based on overall predictive accuracy – improved each year as the density and distribution of data collection increased over the course of the SOD Blitz program (WebTable 2; Figure 1d). We predicted infection risk using PreBlitz data with an accuracy of 65%. By 2012, incorporating 5 years of SOD Blitz data, our overall predictive accuracy increased to 78% (WebTable 2). Spatially explicit predictions of infection probability through time (Figure 2, a–c) show that the greatest risk of infection occurs in coastal forests of Monterey, Santa Cruz, San Mateo, Marin, and northern Sonoma counties. With each year of additional data collection, our SOD Blitz models predicted increases in the amount of land area facing high risk of infection; the area at high risk increased from 662 km^2 according to PreBlitz model to 3907 km^2 based on the 2013 model (WebTable 2). The most notable increases in risk occurred in northern Sonoma County, in Santa Cruz County between the San Jose Valley and Pacific Ocean, and in the southern portion of the Big Sur in Monterey County (Figure 2a). Our ability to identify areas of high risk informs citizens where they may need to apply control efforts to spare economically, culturally, or aesthetically important trees. These risk maps are also being used to help prioritize areas for disease management actions (Figure 2d). Note that our predictions of disease risk represent the probability of pathogen occurrence within 100-m × 100-m grid cell areas, not infection prevalence of the plant host population.

The continuation of the SOD Blitz program each year is making it possible for us to regularly update maps of infection risk and improve our understanding of the pathogen's preferred habitat. Because SOD is an emerging forest disease that is in disequilibrium with its environment, it is important to regularly collect additional data as the disease colonizes new locations and spreads into its full potential niche. This is vital to updating our understanding of the pathogen's habitat and to improving the accuracy of our predictive models, given that models generated at earlier stages of disease invasion are known to underpredict disease risk (Václavík and Meentemeyer 2012).

Question 2: did the educational background and professional experience of our citizen scientists affect probability of disease detection?

By linking lab-verified leaf tissue samples with participant questionnaires from the 2011, 2012, and 2013 Blitzes, we could empirically determine whether or not the chance of successful detection of *P ramorum* infection differed between and among self-identified professionals and self-identified non-professionals. Z tests of foliar sample outcomes between groups revealed that in 2011 non-professionals contributed a significantly greater propor-

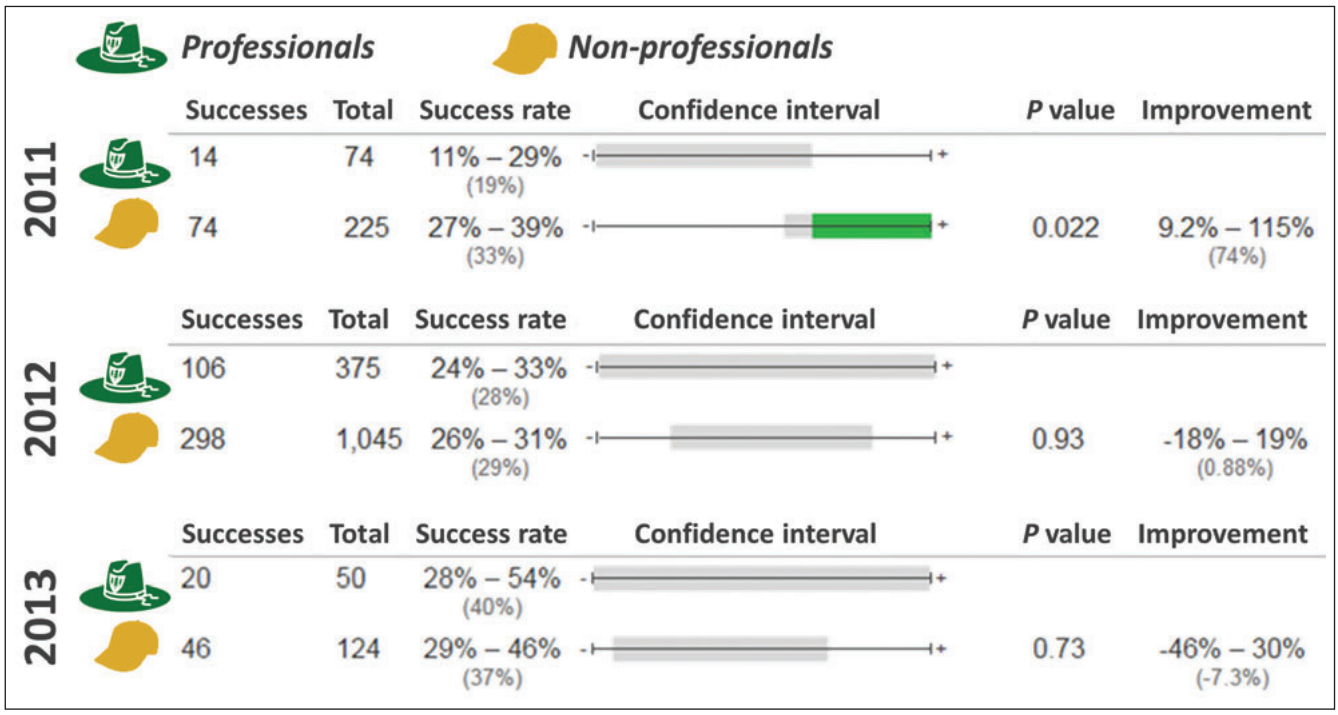

Figure 3. *Z tests comparing disease detection success rates between self-identified professionals (baseline group) and self-identified non-professionals for the 2011, 2012, and 2013 SOD Blitzes. "Successes" indicate the number of leaf tissue samples that tested positive for the pathogen out of the "Total" number of samples submitted. The "Success rate", "Confidence interval" (shaded horizontal bars), and observed value (in parentheses) estimate how precisely we know the actual success rate based on the observed data. The P value ranges from 0 to 1. Lower P values indicate greater confidence that the chances of success differ significantly (P < 0.05) between the groups. "Improvement" estimates how much better (green portion of confidence interval) or worse (% increase or decrease in parentheses) the non-professionals performed relative to the professionals.*

tion of positive samples (33% tested positive for infection) as compared with professionals (19%) (Figure 3). In 2012 and 2013, we observed no significant difference in the proportion of positive samples contributed by the two groups (Figure 3). Between 2012 and 2013, we found that the rate of disease detection increased for both groups, which suggests that the effectiveness of our education and training efforts is increasing. This result is also consistent with previous reports (see Dickinson *et al.* 2010) that have shown how retention of citizen volunteers (in our case 10–20%) can improve the effectiveness of citizen-science programs over time. Furthermore, amateur citizen scientists were no less successful at identifying and collecting positive leaf samples than volunteers with professional backgrounds in science. These findings support claims that data collected by citizen scientists can be reliable (Cohn 2008), particularly when guided by effective education and technical training programs (Gardiner *et al.* 2012).

Conclusions

Continuation of the SOD Blitz program over 6 consecutive years increased the density and distribution of observations in our disease monitoring network, enhanced our understanding of disease spread, and improved our ability to predict locations where trees are at high risk of infection. We found that our trained amateur citizen scientists were just as likely to correctly identify symptomatic vegetation as were scientific professionals. Using a carefully designed sampling strategy with trained volunteers, we increased the quantity and quality of disease distribution data and developed more accurate predictive models. Online dissemination of program results empowers citizens by providing immediate access to the data they collected, as well as to the tools used to track the spread of SOD across the landscape and in their own backyards. Regularly updated risk maps can also be used by citizen scientists and other stakeholders to engage in community-based management practices that help protect trees in high-risk areas. The accessibility of the program's results via the internet has attracted considerable media attention for the SOD Blitz, with more than 40 news articles highlighting the program since 2008. This media coverage has generated additional support for the SOD Blitz program and has further disseminated the results by publishing revised maps of SOD distribution based on those downloadable from the program's website (www.sodmap.org).

In summary, citizen-science efforts have enhanced pathogen risk management and response. The emergence of pests and pathogens is an inherently spatial process; understanding and modeling their progress across a region requires well-distributed, geographically explicit

data. Getting informed citizens on the ground quickly and using the appropriate tools could improve initial response efforts, increasing the potential for eradication or slowing disease spread, and limiting impacts to ecosystems and society. Involving citizens also provides important outreach and educational opportunities that inform the public about the threat of emerging infectious diseases and the actions they can take to help limit their spread. We call on all scientists to engage, educate, and empower citizens using bottom-up approaches in response to emerging pests and pathogens.

Acknowledgements

This research was supported by grants from the US National Science Foundation (NSF; DEB-EF-0622677 and EF-0622770) as part of the joint NSF–National Institutes of Health Ecology and Evolution of Infectious Diseases program. We also acknowledge financial support from US Forest Service's State and Private Forestry Organization and the Gordon and Betty Moore Foundation. Our thanks to the following: A Flint (US Geological Survey) for downscaling PRISM climate data; W Dillon (North Carolina State University [NCSU]) for climate data processing and averaging; W Feng (University of North Carolina [UNC] Charlotte), J Koch (NCSU), and W Dillon for developing and coding a tool for automated calculation of force of invasion; A Hjarding (UNC Charlotte) and Z Arcaro (NCSU) for data entry of SOD Blitz survey questionnaires; and S Mascheretti, D Barbosa, and T Mohr (University of California–Berkeley) for processing foliar samples.

References

Cobb RC, Rizzo DM, Hayden J, *et al.* 2013. Biodiversity conservation in the face of dramatic forest disease: an integrated conservation strategy for tanoak (*Notholithocarpus densiflorus*) threatened by sudden oak death. *Madrono* **60**: 151–64.

Cohn J. 2008. Citizen science: can volunteers do real research? *BioScience* **58**: 192–97.

Crall AW, Newman GJ, Stohlgren TJ, *et al.* 2011. Assessing citizen science data quality: an invasive species case study. *Conserv Lett* **4**: 433–42.

Crowl TA, Crist TO, Parmenter RR, *et al.* 2008. The spread of invasive species and infectious disease as drivers of ecosystem change. *Front Ecol Environ* **6**: 238–46.

Dickinson JL, Zuckerberg B, and Bonter DN. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annu Rev Ecol Evol S* **41**: 149–72.

Dickinson JL, Shirk J, Bonter D, *et al.* 2012. The current state of citizen science as a tool for ecological research and public engagement. *Front Ecol Environ* **10**: 291–97.

Garbelotto M, Maddison ER, and Schmidt D. 2014. SODmap and SODmap Mobile: two tools to monitor the spread of Sudden Oak Death. *Forest Phytophthoras* **4**; doi:10.5399/osu/fp.4.1.3560.

Gardiner MM, Allee LL, Brown PMJ, *et al.* 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Front Ecol Environ* **10**: 471–76.

Goodchild MF and Glennon JA. 2010. Crowdsourcing geographic information for disaster response: a research frontier. *Int J Digital Earth* **3**: 231–41.

Hayden K, Ivors K, Wilkinson C, and Garbelotto M. 2006. TaqMan chemistry for *Phytophthora ramorum* detection and quantification, with a comparison of diagnostic methods. *Phytopathology* **96**: 846–54.

Kremen C, Ullman KS, and Thorp RW. 2011. Evaluating the quality of citizen-scientist data on pollinator communities. *Conserv Biol* **25**: 607–17.

Manel S, Williams HC, and Ormerod SJ. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. *J Appl Ecol* **38**: 921–31.

Meentemeyer RK, Anacker BL, Mark W, and Rizzo DM. 2008. Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecol Appl* **18**: 377–90.

Meentemeyer RK, Haas SE, and Václavík T. 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu Rev Phytopathol* **50**: 379–402.

Rizzo DM and Garbelotto M. 2003. Sudden oak death: endangering California and Oregon forest ecosystems. *Front Ecol Environ* **1**: 197–204.

Václavík T and Meentemeyer RK. 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecol Model* **220**: 3248–58.

Václavík T and Meentemeyer RK. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Divers Distrib* **18**: 73–83.

Václavík T, Kupfer J, and Meentemeyer RK. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *J Biogeogr* **39**: 42–55.

