# The perils of naïve use of open-source data: A comment on "Spatiotemporal distribution of sudden oak death in the US and Europe"

Susan J. Frankel [a,*] , Matteo Garbelotto [b], Chris Jones [c] , Niklaus J. Grünwald [d] , Robert C. Venette [e]

[a] *USDA Forest Service, Pacific Southwest Research Station, 1323 Club Dr, Vallejo, CA 94592 USA*
[b] *Department of Environmental Science, Policy and Management, University of California, 54 Mulford Hall, Berkeley, CA 94720, USA*
[c] *Center for Geospatial Analytics, College of Natural Resources, Jordan Hall 5120, Campus Box 7106, North Carolina State University, Raleigh, NC 27695 USA*
[d] *Horticultural Crops Disease and Pest Management Research Unit, United States Department of Agriculture (USDA) Agricultural Research Service, Corvallis, OR 97330, USA*
[e] *USDA Forest Service, Northern Research Station, 1561 Lindig Ave. St., St. Paul, MN USA 55108 USA*

## ARTICLE INFO

## ABSTRACT

Kang et al. (2024) present a spatiotemporal analysis of *Phytophthora ramorum* outbreaks from 2005 to 2021 in the United States and Europe. However, the analysis and conclusions are flawed because of a lack of understanding of the pathosystems analyzed which led the authors to select improper methods for their analysis. The open-source data analyzed does not include sampling over all seasons of the year. Sampling is primarily conducted in the spring which makes the data unbalanced and inappropriate for examination of seasonality without transformation. Differences in characteristics, and significant driving factors (e.g., relative humidity) between the locations where infection clusters occur, irrigated nurseries with complex sources of inoculum and modified environments versus natural forests subject to only ambient environmental conditions, were not considered when analyzing relationships between moisture conditions and pathogen spread. Additional occurrence records exist for *P. ramorum* in the United States and the United Kingdom, but they were not included in the analysis. Clear descriptive language and proper study design are required to understand how environmental conditions influence *P. ramorum* establishment and spread so they can inform forest management and regulations to protect the resources at risk. An understanding of the temporal and spatial dynamics of Sudden Oak Death, Sudden Larch Death, Ramorum Blight and other diseases caused *by P. ramorum* is critical to serve as the basis for management strategies to limit losses and pathogen spread. The use of publicly available data presents specific challenges that need to be considered in spatiotemporal analyses to obtain meaningful results.

## 1. Short communication

In a recent paper, Kang et al. (2024) set out to conduct a "17-year long-term spatiotemporal analysis of SOD [sudden oak death] outbreaks from 2005 to 2021 in the US and Europe" to characterize patterns of spread and to infer effects of climatological variation on disease dynamics. The authors rely on an analysis of publicly available data provided by Garbelotto et al. (2020) and by the European Union Risk Analysis for *Phytophthora ramorum* (RAPRA) program. However, rather than elucidating conditions for plant pathogen spread, the paper serves as a cautionary example of the perils of using publicly available data

without fully understanding the limitations of the data and the details of the pathosystem for which the data were collected. For a list of technical errors see Supplemental File, Technical Review.

*Understanding data collection and data limitations.* The paper addresses an important topic, distribution patterns and aggregation characteristics for "Sudden Oak Death" outbreaks over multiple years. However, the analysis makes incorrect assumptions about the data collection and about the environmental conditions in nurseries vs. wildlands. As a result, the study provides misleading conclusions and may lead to confusion. Notably for the SODmap project, samples are collected only in the spring because that is the only time of the year when the pathogen

can be readily isolated (Hayden et al., 2004; Vettraino et al., 2009). Given the known association between rainfall and increasing disease incidence (Garbelotto et al., 2021; Kozanitas et al., 2022) sampling is always done in the rainy spring months and not across different seasons, hence, the seasonality results reported in Kang et al. (2024), Fig. 4 are not valid and an artifact of the data collection.

Large datasets, such as the SODmap, that generate point data in a haphazard, unsystematic way cannot be used for spatiotemporal analyses without necessary geostatistical transformations. For an unsystematic data set, one can make inferences on trends of infection over time and across sites, for example, using the approach described in Lione et al. (2017). That approach identifies geographic quadrants across California at a scale such that the intrinsic errors (e.g., GIS or operator errors) in the data will not significantly affect the outcome. One can then compare different quadrants for infection rate and compare the same quadrant for infection rate over time. The analysis has to be done across units not simply across the landscape as in Kang et al. (2024). Another possible approach is to use Monte Carlo simulations using the entire datasets to calculate changes in a metric and then see whether a specific hypothesis (e.g. infection rates increase with rainfall) is statistically more likely to occur than simply by chance, given the points in the dataset, as in Osmundson et al. (2022).

In the Kang et al. (2024) study, sampling intensity is an issue for negatives (but not so for positives), so, computationally one needs to determine the sampling threshold needed to identify a positive. One of the parameters is how large of an area is needed to define the geographic unit, to maximize discriminatory power given the a) intensity of sampling in the database and b) technical or operator error. Once this has been done, then one can determine how many quadrants are positive and when they turn positive or negative.

In addition, there are also several issues that may emerge when employing citizen science datasets to make direct inferences on the distribution of a disease. First, negative results may have limited value in areas that were poorly sampled. Host abundance and density, as well as the expected number of positives in any given year, affect the minimum number of samples necessary for a putative negative to be safely regarded as a negative. Second, apparent spatial and temporal "clusters" may emerge that may simply reflect how and when samples were collected and not reflect disease dynamics. For example, we might expect to have more reports of *P. ramorum* in and near population centers simply because there are more people there to file reports. Similarly, we might expect more samples reported near major trails and none in steep, inaccessible areas. Population density, land ownership, and land use have been shown to influence *P. ramorum* infection in Sonoma County, California (Cushman and Meentemeyer, 2008). These factors are not considered in Kang et al. (2024) and thus bias the results substantially.

*Understanding data limitations.* There are large gaps in the data used in this study which is missing both years of information and large geographic areas. The data from SODmap are an incomplete representation of the areas studied. For example, the authors seem to have assumed that all horticultural nursery *P. ramorum* detections in the US are listed in the SODmap data set (Garbelotto, 2024; Meentemeyer et al., 2015) but they are not. The hundreds of detections of *P. ramorum* in nurseries in America are only maintained by the USDA Animal and Plant Health Inspection Service as part of the US federal quarantine program and the data is not publicly available. As a result, the study is missing detections for at least 25 US states from 2005 to 2021 (USDA APHIS, 2023).

*Understanding the pathosystems.* This paper has dozens of technical misstatements, often of the most fundamental type, that reflect a poor understanding of the pathosystems. The pathogen is stated to be a bacterium, but it is an oomycete (Garbelotto and Hayden 2012; Grünwald et al., 2009); the primary host tree species, *Notholithocarpus densiflorus* is labelled with an old name (*Lithocarpus*) that was revised >15 years ago. The authors fail to distinguish between "Sudden Oak Death" a

disease on some species of oaks (*Quercus*) and tanoak (*Notholithocarpus densiflorus*) (LeBoldus et al., 2022) and the pathogen that causes it, *Phytophthora ramorum* (Werres et al., 2001). Contrary to the paper's title, the disease, Sudden Oak Death, does not occur in Europe, but the pathogen, *Phytophthora ramorum*, is present there causing Sudden Larch Death and Ramorum Blight and is subject to regulatory control (EPPO, 2024).

There are >130 known hosts of *P. ramorum* (USDA APHIS, 2022) including ferns, shrubs, and trees each with different epidemiology and distinct implications for spread. Some species support pathogen sporulation and others do not (Garbelotto et al., 2021). These differences which drive localized pathogen survival and spread are not considered in the Kang et al. (2024) analyses.

The analysis combines nursery detections and forest records to explore disease development and speculates on what drives infection and spread. It examines whether detections were correlated with those at its neighboring locations. However, the combination of this data fails to account for fundamental differences in disease development and conditions in nurseries, gardens, and forests. Most nurseries and gardens grow plants in containers or in the ground under irrigation (Green et al., 2021). The irrigation confounds the precipitation records; it alters the actual amount of water the plants are exposed to. In this study, the frequency and distribution pattern of the moisture in irrigated nurseries is not accounted for. *Phytophthora ramorum* infection is driven by moisture patterns (DiLeo et al., 2014; Tooley et al., 2009) so irrigation confounds the presumed relationship between the precipitation record and disease development. Thus, rainfall explains less of the data variance in nurseries than in wildlands (Pastalka et al., 2017). Relative humidity is critical for infection and must be included for nursery analysis (Garbelotto et al., 2021; Tooley et al., 2009; Pastalka et al., 2017). Due to irrigation *P. ramorum* in nurseries can be sampled during the whole growing season, but not in forests where rain on the Pacific Coast is typically restricted to late fall, winter and early spring.

Nurseries may be located in non-forested areas so the pathogen has no contiguous hosts to spread to. In contrast, forest environments with host tree species may provide opportunities for pathogen spread. *Phytophthora ramorum* spreads via windblown rain in forest environments (Davidson et al., 2005, 2011; Rosenthal et al., 2021). For nurseries, pathogen movement out of facilities occurs due to different mechanisms including inadvertent sales of infected plants, changes in placement of infected container nursery plants including movement to other nurseries, or movement of the pathogen via irrigation run-off or transfer of infested soil (Green et al., 2021). Shipments of plants sold have been shown to move the pathogen on a regional and continental scale (Goss et al. 2009), whereas the pathogen spreads locally in forests. These fundamental, significant differences that influence pathogen spread are not accounted for in the Kang et al. (2024) analysis. Also, the scale and intensity of sampling is very different between the data sets from nurseries vs. wildlands so they cannot simply be combined for analysis.

## 2. Closing remarks

Scientific writing requires detailed, correct description of the specific problem, data and the methods used for analysis. The publication of numerous technical errors misinforms readers and provides inaccurate conclusions. Science, management and regulatory programs rely on consistent standards for science communication and require use of clear, descriptive language to reduce ambiguity, and potential misinterpretations.

The proper application of analytical techniques requires a thorough understanding of the environments and types of problems being assessed. The lack of understanding of the limitations of this publicly available data and failure to consider epidemiologically important differences between infections in nurseries and forests led to conclusions that are not biologically based. The authors did not contact the data steward which might have prevented these mistakes. The increasing use

of publicly available data and autonomous bots makes the likelihood of these types of errors more common, but communication and collaboration can prevent these types of errors. It could be beneficial for journal editors to invite the lead scientist that produced the original data to serve as a referee of the paper that is re-using their data. Data providers might appreciate the importance of accepting the review invitation. To prevent uninformed reuse, data could be shared with some degree of precision, enough for instance for webpage display, but authors could be required to ask the authors for the "real data" for re-analysis. Other best practices include crediting data sources by citation and providing evidence of personal communication with them to make sure data is interpreted correctly. Invitations for data creators to serve as co-authors in any new analysis may also prevent improper analyses.

## CRediT authorship contribution statement

**Susan J. Frankel:** Writing – review & editing, Writing – original draft, Conceptualization. **Matteo Garbelotto:** Writing – review & editing, Writing – original draft. **Chris Jones:** Writing – review & editing, Writing – original draft. **Niklaus J. Grünwald:** Writing – review & editing, Writing – original draft. **Robert C. Venette:** Writing – review & editing, Writing – original draft, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2025.110553.

## Data availability

No data was used for the research described in the article.

## References

Cushman, J.H., Meentemeyer, R.K., 2008. Multi-scale patterns of human activity and the incidence of an exotic forest pathogen. Journal of Ecology 96 (4), 766–776.

Davidson, J.M., Patterson, H.A., Wickland, A.C., Fichtner, E.J., Rizzo, D.M., 2011. Forest type influences transmission of *phytophthora ramorum* in California oak woodlands. Phytopathology 101 (4), 492–501.

Davidson, J.M., Wickland, A.C., Patterson, H.A., Falk, K.R., Rizzo, D.M., 2005. Transmission of *phytophthora ramorum* in mixed-evergreen forest in California. Phytopathology 95 (5), 587–596.

DiLeo, M.V., Bostock, R.M., Rizzo, D.M., 2014. Microclimate impacts survival and prevalence of *phytophthora ramorum* in *Umbellularia californica*, a key reservoir host of sudden oak death in Northern California forests. PLoS One 9, e98195. https://doi.org/10.1371/journal.pone.0098195.

EPPO Global Database, 2024. Phytophthora Ramorum. https://gd.eppo.int/taxon/PHYTRA/reporting [Accessed November 20, 2024].

Garbelotto, M., 2024. SODmap Project. UC Berkeley Forest Pathology and Mycology Lab. https://nature.berkeley.edu/matteolab/?page_id=755 (Accessed November 25, 2024.).

Garbelotto, M., Hayden, K.J., 2012. Sudden oak death: interactions of the exotic oomycete *phytophthora ramorum* with naïve North American hosts. Eukaryotic Cell 11 (11), 1313–1323.

Garbelotto, M., Popenuck, T., Hall, B., Schweigkofler, W., Dovana, F., Goldstein de Salazar, R., Schmidt, D., Sims, L., 2020. Citizen science uncovers *phytophthora ramorum* as a threat to several rare or endangered California manzanita species. Plant Dis. 104, 3173–3182.

Garbelotto, M., Schmidt, D., Popenuck, T., 2021. Pathogenicity and infectivity of *Phytophthora ramorum* vary depending on host species, infected plant part, inoculum potential, pathogen genotype, and temperature. Plant Pathol. 70 (2), 287–304.

Goss, E.M., Larsen, M., Chastagner, G.A., Givens, D.R., Grünwald, N.J., 2009. Population genetic analysis infers migration pathways of *Phytophthora ramorum* in US nurseries. PLoS Pathog. 5 (9), e1000583. https://doi.org/10.1371/journal.ppat.1000583.

Green, S., Cooke, D.E., Dunn, M., Barwell, L., Purse, B., Chapman, D.S., Valatin, G., Schlenzig, A., Barbrook, J., Pettitt, T., Price, C., 2021. PHYTO-THREATS: addressing threats to UK forests and woodlands from Phytophthora; identifying risks of spread in trade and methods for mitigation. Forests 12 (12), 1617.

Grünwald, N.J., Goss, E.M., Ivors, K., Garbelotto, M., Martin, F.N., Prospero, S., Hansen, E., Bonants, P.J.M., Hamelin, R.C., Chastagner, G., Werres, S., Rizzo, D.M., Abad, G., Beales, P., Bilodeau, G.J., Blomquist, C.L., Brasier, C., Brière, S.C., Chandelier, A., Davidson, J.M., Denman, S., Elliott, M., Frankel, S.J., Goheen, E.M., de Gruyter, J., Heungens, K., James, D., Kanaskie, A., McWilliams, M.G., Man in 't Veld, W.A., Moralejo, E., Osterbauer, N.K., Palm, M.E., Parke, J.L., Perez Sierra, A.M., Shamoun, S.F., Shishkoff, N., Tooley, P.W., Vettraino, A.M., Webber, J., Widmer, T.L., 2009. Standardizing the nomenclature for clonal lineages of the sudden oak death pathogen. *Phytophthora ramorum*. Phytopathology 99 (7), 792–795.

Hayden, K.J., Rizzo, D., Tse, J., Garbelotto, M., 2004. Detection and quantification of *phytophthora ramorum* from California forests using a real-time polymerase chain reaction assay. Phytopathology 94, 1075–1083.

Kang, X., Chen, W., Li, S., Wang, T., Yu, L., Singh, R.P., 2024. Spatiotemporal distribution of sudden oak death in the US and Europe. Agric For Meteorol 346, 109891.

Kozanitas, M., Metz, M., Osmundson, T., Serrano, M., Garbelotto, M., 2022. The epidemiology of Sudden Oak death disease caused by *Phytophthora ramorum* in a mixed bay laurel-oak woodland provides important clues for disease management. Pathogens 11 (2), 250.

LeBoldus, J.M., Navarro, S.M., Kline, N., Ritokova, G., Grünwald, N.J., 2022. Repeated emergence of sudden oak death in Oregon: chronology, impact, and management. Plant Dis. 106 (12), 3013–3021.

Lione, G., Gonthier, P., Garbelotto, M., 2017. Environmental factors driving the recovery of bay laurels from *Phytophthora ramorum* infections: an application of numerical ecology to citizen science forests 8(8), 293.

Meentemeyer, R.K., Dorning, M.A., Vogler, J.B., Schmidt, D., Garbelotto, M., 2015. Citizen science helps predict risk of emerging infectious disease. Front Ecol Environ 13 (4), 189–194.

Osmundson, T., Bergemann, S., Rasmussen, R., Garbelotto, M., 2022. Using point data to assess biogeographical signal, endemicity and factors associated with macrofungal diversity in the data-poor Pacific oceanic island bioregion. J. Biogeogr. 49, 891–903. https://doi.org/10.1111/jbi.14354.

Pastalka, T., Rooney-Latham, S., Kosta, K., Suslow, K., Huffman, V., Ghosh, S., Schweigkofler, W., 2017. Monitoring using a sentinel plant system reveals very limited aerial spread of *phytophthora ramorum* from infected ornamental plants in a quarantine research nursery. Plant Health Prog. 18 (1), 9–16.

Rosenthal, L.M., Fajardo, S.N., Rizzo, D.M., 2021. Sporulation potential of *phytophthora ramorum* differs among common California plant species in the Big Sur region. Plant Dis. 105 (8), 2209–2216.

Tooley, P.W., Browning, M., Kyde, K.L., Berner, D., 2009. Effect of temperature and moisture period on infection of Rhododendron 'Cunningham's White' by *phytophthora ramorum*. Phytopathology 99 (9), 1045–1052.

US Department of Agriculture. Animal and Plant Health Inspection Service, 2022. APHIS List of Regulated Hosts and Plants Proven Or Associated with *Phytophthora ramorum*. https://www.aphis.usda.gov/sites/default/files/usdaprlist.pdf [Accessed November 20, 2024].

US Department of Agriculture. Animal and Plant Health Inspection Service, 2023. Risk of *Phytophthora ramorum* to the United States. Version 2. Raleigh, NCOctober 5, 2023.

Vettraino, A., Sukno, S., Vannini, A., Garbelotto, M., 2009. Diagnostic sensitivity of different methods used by two laboratories for the detection of *phytophthora ramorum* on multiple natural hosts. Plant Pathol. 59 (2), 289–300. https://doi.org/10.1111/j.1365-3059.2009.02209.x.

Werres, S., Marwitz, R., Man In't Veld, W.A., De Cock, A.W.A.M., Bonants, P.J.M., De Weerdt, M., Themann, K., Ilieva, E., Baayen, R.P., 2001. *Phytophthora ramorum* sp. nov., a new pathogen on *Rhododendron* and *Viburnum*. Mycol. Res. 105, 1155–1165. https://doi.org/10.1016/S0953-7562(08)61986-3.