

Predicting Movement of Nursery Hosts Using a Linear Network Model¹

Steve McKelvey,² Frank Koch,³ and Bill Smith⁴

Abstract

There is widespread concern among scientists and land managers that *Phytophthora ramorum* may be accidentally introduced into oak-dominated eastern U.S. forests through the transfer of the pathogen from infected nursery plants to susceptible understory forest species (for example, *Rhododendron* spp.) at the forest-urban interface. Inspection programs can be made more efficient by identifying locations throughout the U.S. that are most likely to receive infected nursery stock. We develop a spatial network model framework utilizing potential interstate nursery stock movements on a bipartite network, adopting a Bayesian approach to model probabilities of transmission of *P. ramorum* from entry to destination nodes within the network. As the goal of this paper is to present a general model framework, no specific risk analysis is presented. In the face of future discoveries of *P. ramorum* infections in the eastern U.S., instances of models from this framework can be used to help identify sites with a high risk of infection based on observed infection patterns.

Key words: Bayesian analysis, *Phytophthora ramorum*, nursery shipments, transportation network.

Introduction

Although *Phytophthora ramorum* has only been detected in natural forest landscapes in California and Oregon, there is some concern that the pathogen may be accidentally introduced into oak-dominated eastern U.S. forests. A noteworthy potential pathway for such introduction is transfer of the pathogen from infected nursery plants to susceptible understory forest species (for example, *Rhododendron* spp.) at the forest-urban interface. Genetic evidence suggests the commercial plant trade has played a major role in spreading the pathogen (Ivors and others 2004, Ivors and others 2006); indeed, there appears to be continued exchange of *P. ramorum* genetic material between Europe and North America (Rizzo and others 2005). Of more immediate concern, confirmed sporulating hosts include a number of genera, particularly from the Ericaceae family, that are both widely distributed in natural forests of the U.S. and commonly sold as ornamental plants by nurseries (Englander and Tooley 2003, Tooley and others 2004, Rizzo and others 2002, Rizzo and others 2005). Despite current federal restrictions (see APHIS 2007), the interstate shipment of west coast nursery stock remains a serious risk, as is evident from the detection of infected plants in nurseries in Indiana, Maine, and other states in 2006. The United

¹ A version of this paper was presented at the Sudden Oak Death Third Science Symposium, March 5–9, 2007, Santa Rosa, California.

² Department of Mathematics, Statistics and Computer Science, Saint Olaf College, Northfield, MN USA 55057.

³ Department of Forestry and Environmental Resources, North Carolina State University, Research Triangle Park, NC USA 27709.

⁴ Southern Research Station, USDA Forest Service, Research Triangle Park, NC USA 27709.

States Department of Agriculture-Animal and Plant Health Inspection Service (USDA-APHIS) and the USDA-Forest Service (USDA-FS) have ongoing national *P. ramorum* survey programs; these programs can be made more efficient by identifying locations throughout the U.S. that are most likely to receive infected nursery stock.

Towards this end, we have been developing a model framework based on a bipartite network incorporating interstate nursery stock movement. The network depicts nursery stock shipment from sources on the west coast to destinations further east where *P. ramorum* has not been detected in the natural forest landscape. We constructed the network using geographic information system (GIS) data layers and relational databases from the U.S. Department of Transportation and other federal agencies. The foundation of our network is the Freight Analysis Framework (FAF₂) Commodity Origin-Destination database (U.S. Department of Transportation 2006). The Commodity Origin-Destination database is a relational database describing the movement of commodities, in both tonnage and monetary value, between statistical regions. The statistical regions in the database consist of major metro areas, the remaining areas of states, and entire states. These regions are treated as nodes in the model network.

In our model framework, probabilities of transmission of *P. ramorum* from source to destination nodes within the network are based on nursery stock flow volumes along the links connecting each origin/destination pair. See equation (2). The 43 data categories in the FAF₂ are not specific enough to break out nursery stock as a specific category of commodity. The nursery stock tonnages are approximated by estimating what proportion of all commodities consists of nursery stock and applying this proportion to each link in our network (US Departments of Commerce and Transportation 2005).

Other approaches to identifying possible paths of *P. ramorum* infection are certainly possible. The model framework in this paper is quite general and requires only the identification of nodes that are sources of infection, nodes that are at risk of becoming newly infected and some measure of the likelihood that *P. ramorum* will be spread from each currently infected node to the currently uninfected nodes. The network we develop using FAF₂ information provides a solid test network for our model and is a good representation of the kind of network and corresponding infection likelihood information that a user of this model would need to construct.

As new infections are observed a Bayesian update of conditional probabilities allows us to make comparative statements about the likelihoods of various paths of infection, identifying likely proximate sources of the infection. The same analysis can identify new sites that are at high risk of infestation through importation of *P. ramorum* from these proximate sources even if *P. ramorum* has not yet been detected. This allows managers to selectively target resources in their efforts to mitigate the spread of *P. ramorum*.

The need for such a tracking mechanism is plain after considering the two maps shown in figures 1 and 2. These figures have the FAF₂ statistical regions as a backdrop. The lines and points in figures 1 and 2 are derived from USDA-APHIS trace forward data (APHIS, unpublished data).

Figure 1 shows the links to regions where *P. ramorum* host plants, not necessarily infected by *P. ramorum*, were shipped by commercial nurseries from California and Oregon during the years 2003 through 2006. These flows are derived from USDOT data (U.S. Department of Transportation 2006). The dots on the figure show the actual location of nurseries that received host plants from California and Oregon. The network of actual shipments resolved to localities rather than regions would be far more complex.

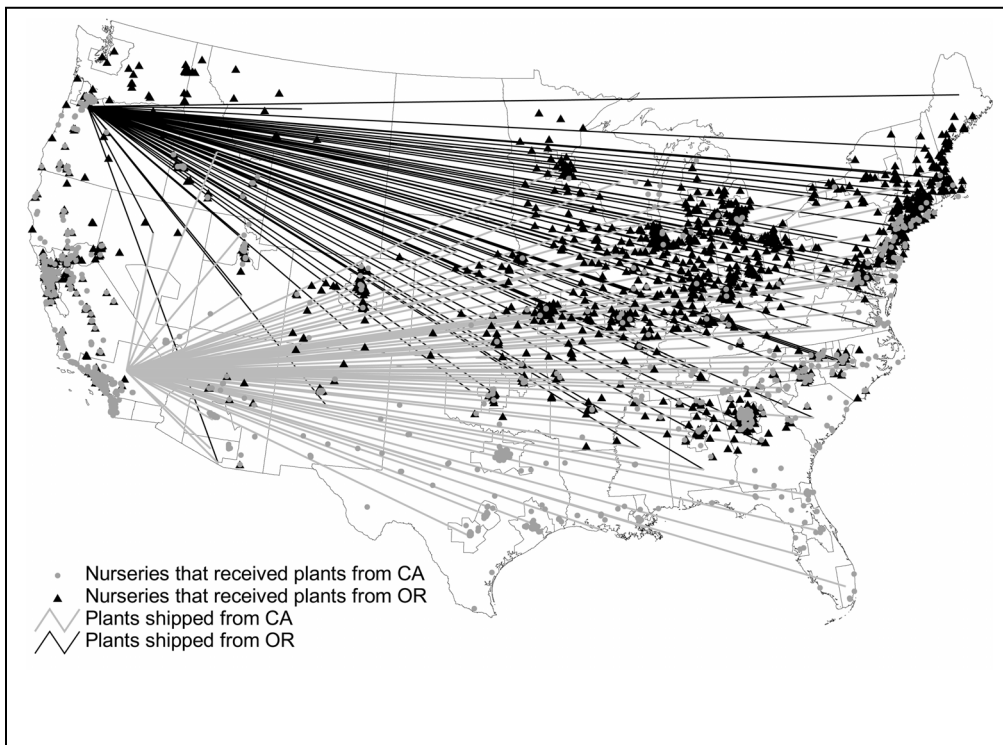


Figure 1—U.S. *P. ramorum* host plant shipping pattern.

Figure 2 shows the dynamic nature of the flow of nursery stock plants that were found to be positive for *P. ramorum*. In this figure we clearly see how the source of infected plants has shifted northward. During the earlier years of 2003 and 2004 southern California was the most common source of detected infected host plants in nursery stock. In 2005 and 2006 northern Oregon became a significant source of infected nursery stock.

The models we propose will help inspectors quickly identify future shifts in the source of *P. ramorum* infected nursery stock, increasing the likelihood of detecting and mitigating the spread of the pathogen past the urban/forest interface.

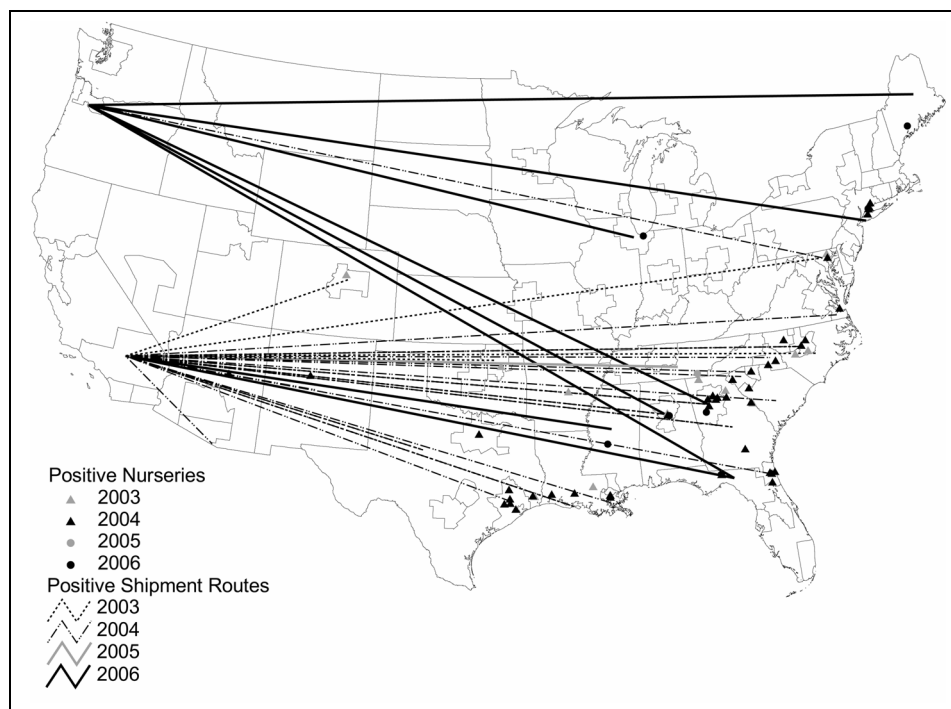


Figure 2—U.S. confirmed *P. ramorum* infected host plant shipments.

Model Outline

The scenario studied here involves a short time period in which new infections by *P. ramorum* are detected in previously pristine sites. We assume that all the newly detected infections result from nursery stock shipped from a single previously infected source. The goal of the model is to assign a probability to each source indicating the likelihood that a source is the sole origin of the infection. Once these conditional probabilities are determined, probabilities of undetected infection can be assigned to all other susceptible sites, allowing managers to focus subsequent inspection efforts on sites most at risk of *P. ramorum* infection.

To this end, a directed bipartite network is defined with source nodes corresponding to known domestic sources of *P. ramorum* infected nursery stock and destination nodes corresponding to sites that receive nursery stock from infected sources. Nursery stock flows along the links of this network, each of which leads from a source node to a destination (or site) node. The flow rate along each link equals the amount of nursery stock flow, by weight, carried from the source to the destination.

A standard Bayesian probabilistic model is imposed on this network to update probabilities of each source node as being the origin of recently found new infections. A priori probabilities of being the solitary origin of *P. ramorum* infection can be assigned to each source based on many factors, including historical observations, rate of importation from possible foreign sources of *P. ramorum* and other relevant factors. Which factors are most relevant will become clearer as the biology and ecology of *P. ramorum* becomes better understood.

Probabilities of transmission are also assigned to each link based on nursery stock flow rates along the link. With these probabilities determined, a simple application of Bayes' Rule can serve to update the probabilities that a given source site is the unique origin of the newly uncovered infection pattern.

Model Notation

Here we begin the formal description of our model. Let

$$S = \{s_1, s_2, \dots, s_i, \dots, s_m\}$$

be the set of nodes corresponding to potential sources of *P. ramorum* infection. These sites will be areas of known *P. ramorum* infection that export significant amounts of relevant nursery stock. Similarly, let

$$D = \{d_1, d_2, \dots, d_j, \dots, d_n\}$$

be the set of nodes corresponding to destinations of nursery stock shipments. At the beginning of the model's time period these destinations are assumed to be free from *P. ramorum* infection.

To reflect the results of the destination inspections carried out during the time frame of the model, define D_I to be the set of destination nodes found to be infected, define D_C to be the set of destination nodes found to be "clean," in other words, uninfected, and let D_U be the set of destination nodes for which infection status is unknown. The union of these three sets is D , the set of all destination nodes.

Let E be the event that the infection pattern determined by the sets D_I , D_C and D_U has developed during the model time period.

Furthermore, let F_i be the event that source s_i is the sole origin of the observed infection pattern. The assumption that exactly one source caused the observed infection pattern implies that the events F_i are disjoint and together cover the universe of outcomes.

The flow rates of nursery stock play an important role in our model. To this end, let f_{ij} be the flow of stock from source s_i to destination d_j .

Lastly, we will need to compute probabilities that a flow rate from a specific source to a specific destination results in the initial infection of that destination during the model time period. This computation will be based upon a parameter p defined to be the probability that a single unit of flow will result in the initial infection of the flow's destination.

Ranking Sources as Origins of Infection

The first goal of our analysis is to determine, for each source s_i , the probability that s_i is the origin of the observed infection pattern. In terms of our notation what we seek

is the conditional probability $P(F_i|E)$. By a direct application of Bayes' Rule we see this probability can be computed as

$$P(F_i | E) = \frac{P(E | F_i)P(F_i)}{\sum_{k: s_k \in S} P(E | F_k)P(F_k)} \quad (1)$$

for every i such that $s_i \in S$.

To use (1) we need values for $P(E|F_i)$ and $P(F_i)$ for every source s_i . The probabilities $P(F_i)$ are the a priori probabilities of each source being the origin of infection. These probabilities are assigned by users of the model based on historical observations, rates of importation from possible foreign sources of *P. ramorum* and other relevant factors. It should also be noted that the $P(F_i)$ terms in (1) can be replaced by nonnegative relative likelihoods, quantities that need not sum to precisely one, without affecting the validity of the expression.

The probabilities $P(E|F_i)$ of achieving the observed infection pattern given s_i is the origin of the infection are calculated from nursery stock flow rates and an assumption of independent destination infections. Recall that p denotes the probability of a single unit of flow infecting a previously uninfected destination. If f denotes the flow from some infected source to some destination, the probability of the infection of the destination is given by

$$P(\text{new infection given flow } f) = 1 - (1 - p)^f. \quad (2)$$

Assuming that infection of different destinations are independent events, we can now derive the conditional probability $P(E|F_i)$ for each source s_i . Recalling that f_{ij} denotes the flow of nursery stock from source s_i to destination d_j during the model time period, we see

$$P(E | F_i) = \left[\prod_{j: d_j \in D_I} (1 - (1 - p)^{f_{ij}}) \right] \left[\prod_{j: d_j \in D_C} (1 - p)^{f_{ij}} \right]. \quad (3)$$

The first term on the right hand side of (3) represents the probability of all the infections that were realized. The second term gives the probability of all the non-infection events. The destinations for which no information is available, the elements of D_U , play no role in this computation.

At this point we are able to compute all the quantities required by (1) to compute the conditional probabilities for each source. Doing so allows us to rank the sources according to the probability that they were the origin of the new infection pattern. The updated probabilities also allow us to identify destinations from the D_U set that are at high risk of having also been infected during the model time period.

Identifying High Risk Destinations

Assigning probabilities of infection to each destination in D_U can be achieved by the common technique of partitioning the universe of possibilities into a small number of disjoint events. In our context the relevant partition comprises the events $F_i|E$ for each source s_i . This gives us

$$\begin{aligned}
 P(\text{site } d_j \text{ is newly infected} | E) &= \sum_{i: s_i \in S} P(\text{site } d_j \text{ is infected by source } s_i | E \cap F_i) P(F_i | E) \\
 &= \sum_{i: s_i \in S} (1 - (1 - p)^{f_{ij}}) P(F_i | E)
 \end{aligned} \tag{4}$$

where the probabilities $P(F_i | E)$ can be computed according to (1).

Application to Large Networks

To be of practical use this model must be applicable to moderately large networks. Using custom software, written by McKelvey in the Java programming language, the model above has been applied to the nationwide network of nursery stock flows described in the Introduction. This network consists of nine regional sources in California, Oregon and Washington and 101 destination regions throughout the country (see figure 3). On this network the software performed all the calculations seemingly instantaneously.

The number of operations required to perform the computations in this model grows as the number of nodes and links in the network. If m is the number of sources in the scenario and n is the number of destinations, the time and memory necessary to perform all the calculations is bounded from above by a number proportional to $m+n+mn$. Thus, for large networks, a doubling in the size of the network increases the upper bounds on the time and memory needed for the computation by roughly a factor of four. This quadratic growth indicates that scaling issues are unlikely to pose serious impediments to applying the model to large scale problems.



Figure 3—U.S. national stock flows.

Future Research

Future research will be directed toward loosening the independence assumptions and the more problematic assumption that the entire newly observed infection pattern has precisely one source as its origin. While these assumptions lead to a model that is useful in relating nursery stock flow patterns and recent infection activity to the risk of future infection events, the current methodology fails to exploit patterns in the commodity flow or the possibility of several concurrent infection origins. In addition, the possibility of infection introduced at transshipment points is not addressed currently. All of these issues will be addressed in future work.

Literature Cited

Englander, L.; Tooley, P. 2003. Plant hosts in the nursery industry - PLANTS MOVE! How might the movement of plants in the nursery industry contribute to the spread of *Phytophthora ramorum* to new areas? Sudden Oak Death Online Symposium. www.apsnet.org/online/SOD (website of The American Phytopathological Society). doi:10.1094/SOD-2003-LE.

Ivors, K.L.; Hayden, K.; Bonants, P.J.M.; Rizzo, D.M.; Garbelotto, M. 2004. AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycological Research* 108: 378–392.

Ivors, K.L.; Garbelotto, M.; Vries, I.D.E.; Ruyter-Spira, C.; Hekkert, B.T.E.; Rosenzweig, N.; Bonants, P. 2006. Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Molecular Ecology*. 15: 1493–1505.

Rizzo, D.M.; Garbelotto, M.; Davidson, J.M.; Slaughter, G.W.; Koike, S.T. 2002. *Phytophthora ramorum* and Sudden Oak Death in California: I. Host Relationships. Pages 733-740 In: Proceedings of the fifth symposium oak woodlands: oaks in California's changing landscape. 2001. Gen. Tech. Rep. PSW-GTR-184. R.B. Standiford, D. McCreary, K.L. Purcell, eds. Southwest Research Station, Forest, Service, U.S. Department of Agriculture, Albany, CA: 733–740.

Rizzo, D.M.; Garbelotto, M.; Hansen, E.M. 2005. *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology*. 43: 13.1–13.27.

Tooley, P.W.; Kyde, K.L.; Englander, L. 2004. Susceptibility of selected ericaceous ornamental host species to *Phytophthora ramorum*. *Plant Disease*. 88: 993–999.

USDA APHIS. 2007. 7 CFR Part 301. *Phytophthora ramorum*; Quarantine and Regulations. *Federal Register* 72(38): 8585-8604.

U.S. Department of Commerce, Census Bureau (USCB) and U.S. Department of Transportation, Bureau of Transportation Statistics (USDOT-BTS). 2005. Commodity Flow Survey 2002 [CD-ROM]. C1-E02-ECFS-00-US1. Washington, DC: Economic Census 2002.

U.S. Department of Transportation, Federal Highway Administration, Office of Freight Management and Operations (USDOT-FHA). 2006. FAF₂ Commodity Origin-Destination Database [data tables]. http://ops.fhwa.dot.gov/freight/freight_analysis/faf/fafstate2state.htm.