Predictive Modeling and Categorizing Likelihoods of Quarantine Pest Introduction of Imported Propagative Commodities from Different Countries

ByeongJoon Kim,[1,2] Seung Cheon Hong,[1,*] Daniel Egger,[3] Catherine S. Katsar,[4] and Robert L. Griffin[4]

The present study investigates U.S. Department of Agriculture inspection records in the Agricultural Quarantine Activity System database to estimate the probability of quarantine pests on propagative plant materials imported from various countries of origin and to develop a methodology ranking the risk of country–commodity combinations based on quarantine pest interceptions. Data collected from October 2014 to January 2016 were used for developing predictive models and validation study. A generalized linear model with Bayesian inference and a generalized linear mixed effects model were used to compare the interception rates of quarantine pests on different country–commodity combinations. Prediction ability of generalized linear mixed effects models was greater than that of generalized linear models. The estimated pest interception probability and confidence interval for each country–commodity combination was categorized into one of four compliance levels: "High," "Medium," "Low," and "Poor/Unacceptable," Using K-means clustering analysis. This study presents risk-based categorization for each country–commodity combination based on the probability of quarantine pest interceptions and the uncertainty in that assessment.

KEY WORDS: Action rate; categorization; quarantine pests; risk-based inspection; uncertainty

1. INTRODUCTION

International trade plays a key role in the introduction of species to new environments (Essl, Winter, & Pysek, 2012; Levine & D'Antonio, 2003; Meyerson & Mooney, 2007; Pimentel, Lach, Zuniga, & Morrison, 2000; Pysek et al., 2010; Seebens et al., 2015; Tatem, Hay, & Rogers, 2006). About 50,000 exotic species have been introduced into the United States, and the number is increasing (Pimentel et al., 2000; Pimentel, Zuniga, & Morrison, 2005). Biological invasions cause economically and ecologically significant damage to agricultural and natural resources by decreasing native species diversity and abundance in invaded locations (Pimentel et al., 2000; Pysek et al., 2012; Simberloff et al., 2013; Vila et al., 2011). Agricultural quarantine inspections (AQI) include activities that help reduce pest threat to U.S. agriculture by providing information that can be used for risk-based decision making. AQI data are collected from U.S. ports of entry and incorporated into a database called Agricultural Quarantine Activity System (AQAS). McCullough, Work, Cavey, Liebhold, and Marshall (2006) found that arthropods account for over 75% of pest

[1]Center for Integrated Pest Management, North Carolina State University, Raleigh, NC, USA.
[2]Personal Line Auto Insurance: Research and Development, Progressive Corporation, Mayfield, OH, USA.
[3]Center for Quantitative Modeling, Pratt School of Engineering, Duke University, Durham, NC, USA.
[4]USDA–APHIS–CPHST, Raleigh, NC, USA.
*Address correspondence to Seung Cheon (Steve) Hong, USDA-APHIS-PPQ, 1730 Varsity Drive, Suite 400, Raleigh, NC 27606, USA; schong227@gmail.com.

interceptions when monitoring activities check for arthropods, plant pathogens, weeds, and mollusks. They also reported on a large historical interception record of 725,000 pests collected between 1984 and 2000 and noted that the most interceptions were associated with travelers' baggage (62%), followed by cargo (30%) and propagative material (7%).

Management of quarantine species is applicable to a variety of invasion processes (Blackburn et al., 2011; Lodge et al., 2006). In attempting to understand the significance of the initial invasion stage for effective management, Puth and Post (2005) alluded to the importance of inspection program designs at ports of entry. One of the first barriers to the introduction of quarantine species is the inspection program. In this article, the "quarantine species" is defined as "a plant pest that is not known to occur or only has limited distribution in the United States"(Convention, 2007). Interception rates of quarantine species at ports of entry vary depending on the type of cargo and mode of transportation. Ornamental plant material such as cut flowers and plants for planting were found to be major vectors for the introduction of exotic species in Europe between 1995 and 2004 (Kenis, Rabitsch, Auger-Rozenberg, & Roques, 2007; Liebhold, Brockerhoff, Garrett, Parke, & Britton, 2012). The study found that more than 40% of interceptions were associated with cut flowers and propagative plants.

Several studies have investigated different pathways and their historical interception rates of pests in detection and monitoring programs (Areal et al., 2008; McCullough et al., 2006; Robinson, Burgman, & Cannon, 2011; Surkov, Lansink, van Kooten, & van der Werf, 2008). A pathway for quarantine inspection is defined as any collection of inspection items from a population and can vary depending on commodity, transportation route, or trading countries/partners (Robinson et al., 2011). One study investigated potential inspection resource allocation on high- and low-risk pathways using a simulation of quarantine inspection in Australia (Robinson et al., 2011). They showed that reducing inspection frequency for lower-risk pathways was an effective strategy for ensuring the risk remained below an established threshold while maintaining a full inspection frequency for high-risk pathways. Similarly, Govindaraju, Bebbington, and Wrathall (2010) showed the feasibility of a partial inspection and skip sampling program in New Zealand for imported food materials based on historical inspection records.

Interception rates are likely to vary depending on both country of origin and specific pest (Eschen, Roques, & Santini, 2015; Kenis et al., 2007). The interception and establishment of quarantine pests showed significant interactions with the type of imported woody plants in a study by Eschen et al. (2015). Another study showed that the likelihood of pest detection varied mainly with the genus of cut flowers (Areal et al., 2008), implying that risk management based on plant pest detection efforts may need to target those plant genera with the highest probability of carrying quarantine pests.

For the categorization of pathways with different risk or compliance levels, a specified value of risk is used as a cutoff. If the interception rate for a pathway is greater than the cutoff value, it is categorized as high risk, while if the probability of quarantine pest interception for a pathway is smaller than the threshold value, it is considered low risk. An appropriate threshold value depends on the resources available for inspection considered with the number and type of pests, their probability of establishment, and the impact. The impact is the resulting damage of agricultural and environmental resources due to the introduction of quarantine species. The assessment of expected impact is a complex process including many disciplines such as population biology and economics (Sakai et al., 2001). The threshold value can also be affected by available inspection resources. For example, inspection resources can be expressed as amount of time spent in inspections, number of shipments that one inspector can examine per day, or other relevant units. Due to the lack of data about the time available for inspection compared to other tasks, it is difficult to estimate inspection resources. If resources were unlimited, the threshold value could be entirely determined by risk. Since inspection resources are limited, it may be determined by iteration as a way of choosing how to allocate those limited resources (Robinson et al., 2011).

Quarantine pest interception records can be used to calculate interception rates statistically using a binomial probability distribution. It is important to account for the uncertainty of interception rates when risk categorization of pathways is considered; however, only a few studies have considered the uncertainty associated with the interception rate for risk ranking (Robinson et al., 2011). The U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) established the Propagative Monitoring and Release Program (PMRP) to expedite the movement of high-volume

imports of plant material with low risk for the introduction of quarantine pests into the United States. The program used interception records, host suitability, and other parameters for risk categorization. The risk categorization of the PMRP was based on a combination of the qualitative (e.g., host suitability and expert opinion) and quantitative (e.g., number of actions and shipments) approaches. The applicability of predictive statistical models such as regression or machine learning, however, was limited due to data inconsistency. We therefore proposed statistical models for categorizing risk based on quarantine pest interception rates and their uncertainty in country–commodity combinations in order to help build a new framework for quarantine pest detection and management programs associated with the inspection of plants for planting. In our study, we investigated combinations of propagative plant genera and their countries of origin as potential pathways of introduction for quarantine pests. We analyzed agriculture inspection and quarantine activity records archived at the USDA–APHIS–Plant Protection and Quarantine (PPQ) AQAS from October 2014 to September 2015 to determine interception rates of quarantine pests by country of origin and commodity and to develop a potential categorization protocol for implementation in a risk-based inspection program. The objectives of this study are (1) to estimate interception probability parameters of genus-origin combinations for carrying quarantine pest species and (2) to develop a methodology that takes account of uncertainty for ranking risk of genus-origin import combinations for their potential of carrying quarantine species.

2. METHODOLOGY

2.1. AQAS Database

The USDA–APHIS–PPQ AQAS records quarantine activities performed by the Department of Homeland Security, Customs and Border Protection, and APHIS PPQ at U.S. ports of entry. The AQAS data include multiple variables such as quantities, types, and countries of origins for propagative plants, fruits, vegetables, cut flowers, lumber, and other products. We chose to analyze inspection records of propagative plant materials for our study. The volume of imports for regulated plant products for consumption (e.g., fruits, vegetables, cut flowers) is far greater than for propagative plant material but

Table I. Overview of FY 2015 Inspection Data Sets

	Inspection Data
Data collection period	October 2014–January 2016
Total inspections	128,653
Total training records	90,057
Total test records	38,596
Countries of commodity origin	93
Commodity genera	1,729
Country–commodity combinations	1,527
Interceptions	565
Plant quantity	1,415,645,938
Plant inspection stations	14

we chose to analyze inspection records for propagative material because the importation of plants for planting is generally considered to be a higher risk than imports of other regulated plant products and a history of pest interception data was readily available for analysis in AQAS. Inspection records of propagative materials collected from October 1, 2014 to January 31, 2016 in calendar year were obtained from AQAS. The data were split up into two parts of 70% and 30% for training and test data sets, which were used for developing predictive models and validation study, respectively. The AQAS data are generated as follows. After inspection is completed, an inspector assigns a disposition, which indicates the action taken on a given commodity presented for entry into or through the United States. If a quarantine pest is detected, the shipment is assigned a quarantine-action-related disposition. The assigned disposition for inspection was recorded by country–commodity combination and converted to a binary response variable for data analysis. If an inspection event was assigned a quarantine-action-related disposition code for a pest, then it was considered a positive interception and assigned a value of 1; otherwise, it was assigned a value of 0.

2.2. Data Exploration

Inspection records of importing propagative materials collected from October 2014 to January 2016 were used for fitting statistical models and validation study. There were a total 128,653 of inspection records during the period (Table I). Data were randomly split up into two groups with a 7:3 ratio for training and test data sets, respectively. The training data set was used to fit statistical models, while the test data set was used to evaluate the predictive

performance of the fitted models. The training and test data sets had a total of 90,057 and 38,596 inspection records of shipments (i.e., country–commodity combinations) with 1,729 different genera of propagative plant materials from 93 countries, respectively. There were a total of 5,497 country–commodity genus combinations inspected during the period. Among them, 1,527 country–commodity combinations with at least greater than 10 shipments were selected for data analysis. A total of 565 shipments were detected with actionable pests from 14 plant inspection stations (PIS).

2.3. Variables for Statistical Analysis

The dependent and independent variables used for statistical analysis are summarized in Table II. The original phytosanitary activity data were aggregated by country–commodity combination. Plant quantity and number of inspections conducted during overtime shift work were used as variables for each country–commodity combination. Other categorical variables (i.e., PIS, Month, and Pathway) were converted to numbers for each country–commodity combination. For example, if a country–commodity combination was imported through three PISs (e.g., Miami, FL, John F. Kennedy International Airport, NY, and Los Angeles, CA), then the PIS value for that country–commodity would be 3. Thus, numeric variables are used to represent diversity within categorical variables during the import period. Country–commodity combinations with 10 or more shipments in the training data were used for further statistical testing. This provided 1,398 country–commodity combinations from the training data set to use in testing model fit (Table I).

2.4. Statistical Analysis

The relationship of selected inspection-related variables to the likelihood of carrying a quarantine invasive pest species was determined with a generalized linear model (GLM) with Bayesian approach and a generalized linear mixed effects model (GLMM) using R version 3.2.5. The GLM and GLMM were used to analyze interception data and selected variables (Table II) (R Development Core Team, 2015). Starting with the full models (GLM1 and GLMM1), the best models were developed by eliminating variables until all variables in the model were significant using a likelihood ratio test and backward elimination (Table III). The four

selected models (two full and two reduced models) are summarized in Table III. We then compared the predictive performance and compliance of these models. Prior to that, logistic regression analysis with the maximum likelihood method was initially conducted to account for the pattern of quarantine pest interception using the glm function in R. The initial statistical modeling attempt did not fit the data due to issues of convergence validity, collinearity, and/or perfect separation. Multicollinearity was tested by computing variance inflation factors (Davis, Hyde, Bangdiwala, & Nelson, 1986) and the result showed strong multicollinearities, especially among commodity genera and countries (data not presented). This may be because there are many plant commodity genera and countries (Table I) and interception is highly dependent on these variables. We also found that some commodity genera and countries are multicollinear to each other. Consequently, it is likely to induce the instable coefficient estimation (e.g., greater standard errors) in logistic regression analysis (Zorn, 2005). In our study, we used Bayesian logistic regression analysis as an alternative method to resolve these issues and to obtain stable coefficient estimates (Gelman, Jakulin, Pittau, & Su, 2008). A weakly informative prior distribution, the Student *t*-distribution with one degree of freedom (Cauchy), was used as a prior distribution in a Bayesian logistic regression analysis. The advantage of this prior distribution is the ability to achieve stable estimation when there are issues in logistics regression, as described above (Gelman et al., 2008).

The conditional posterior distribution for β follows multivariate normal distribution with a mean of β and a standard deviation σ^2, where the marginal posterior distribution for σ^2 is a scaled inverse-χ^2. The coefficients β_i and the unknown scale σ_i have the following distribution (Gelman et al., 2008):

$$\beta_i \sim N\left(\mu_i, s^2{}_i\right) \quad \text{and} \quad \{\sigma_i^2 \sim \text{Inv-}\chi^2\left(v_i, s_i^2\right),$$

where coefficients β_i follow *t*-prior distribution with centers μ_i, and scales s_i with v_i are degrees of freedom.

The posterior distribution was used to generate coefficients based on a Monte Carlo approximation with 1,000 iterations. The Bayesian approach with a Student-*t* prior distribution provided more robust inferences in logistic regression analysis than other prior distributions (Gelman et al., 2008; Lange, Roderick, & Jeremy, 1989). Bayesian logistic regression was conducted with the bayesglm function in the "arm" package in R. Akaike information

Variable	Variable Type and Description
Action	Response. Total number of quarantine pests for country–commodity combination; used as positive response for interceptions
Shipment	Response. Total number of shipments for country–commodity combination; used as response variable for total trials for each combination
Country	Categorical. Country of commodity origin
Commodity	Categorical. Genus of propagative plants
Plant quantity	Numerical. Total number of imported propagative plants
Overtime	Numerical. Total number of inspections conducted during nonregular time
PIS	Numerical. Number of unique plant inspection stations for each country–commodity pair (range: 1–14)
Month	Numerical. Number of months in which inspections occurred for each country–commodity pair (range: 1–12, January–December)
Pathway	Numerical. Number of unique pathways for each country–commodity pair (range: 1–7; e.g., air baggage, air cargo, land border baggage, land border cargo, mail/express mail, and maritime cargo)

Models[a]	Variables[b] Used
GLM1	Country of origin + Commodity + log(Plant quantity) + log(Overtime) + PIS + Month + Pathway
GLM2	Country of origin + Commodity + log(Plant quantity) + log(Overtime) + Month + Pathway
GLMM1	log(Plant quantity) + log(Overtime) + PIS + Month + Pathway + Country of origin/Commodity[c]
GLMM2	log(Plant quantity) + log(Overtime) + Month + Pathway + Country of origin/Commodity[c]

[a]Model estimation methods for GLM and GLMM: For GLM, the bayesglm function using arm package was used; for GLMM, the glmer function using lme4 package in R was used.

[b]Country of origin and commodity are categorical variables. Plant quantity is the total number of plants for country–commodity combinations imported during FY 2015. Overtime is the number of country–commodity combinations that came through at overtime hour. PIS (plant inspection station) represents the number of different stations; month, the number of months; and pathway, the number of different pathways for each country–commodity combination. For GLMMs, logarithms of plant quantity and overtime variables were used.

[c]Random effects: Commodity nested within country of origin.

criterion (AIC) values revealed that the Bayesian approach improved the model fit over that obtained from the maximum likelihood method. AIC values for the Bayesian approach were 1,200s, while the values of standard logistics regression were 18,000 and 14,000 for GLM1 and GLM2, respectively. As a result, the convergence issue was resolved (Gelman et al., 2008). For these reasons, further analysis was conducted using results from the GLMs with the Bayesian approach.

The GLMM1 and 2 were fitted with the glmer function in the "lme4" package. In GLMM, the fixed effects were plant quantity, overtime, PIS, month, and pathway. The random effect was commodity genus nested within country, which accounts for the varying probabilities of carrying quarantine pest species among different country–commodity combinations and for overdispersion (Table III).

2.5. Estimation of Probability and Prediction Interval for Country–Commodity Combinations Carrying Quarantine Species

For both GLMs and GLMMs, prediction intervals were obtained using simulated coefficients based on selected models, as shown in Table III. Each model was simulated 1,000 times with a Monte Carlo approximation to generate coefficients, followed by multivariate normal distribution with means and standard deviation. The lower 2.5th and upper 97.5th percentiles of simulated coefficients were chosen and implemented into each model to calculate the probability of carrying quarantine-significant pest species for each country–commodity combination. Three levels (median, minimum, and maximum) of probabilities for country–commodity combinations were used for clustering analysis later.

2.6. Model Evaluation and Comparison

The models were evaluated using the randomly selected 30% data (i.e., test data set) from the whole data set. In the validation test, predicted probabilities of carrying quarantine pests for country–commodity combinations were compared with observed interception rates using quantitative error index statistics (Legates & McCabe, 1999; Moriasi et al., 2007; Willmott, 1981; Willmott et al., 1985; Willmott, Robeson, & Matsuura, 2012). Moriasi et al. (2007) recommended a combination of multiple quantitative statistics, including dimensionless techniques and error index statistics, for model evaluation. We compared the performance of four different models using modified index of agreement (md), as well as six error index statistics: Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), mean squared error (MSE), root MSE (RMSE), percent bias (PBIAS), and RMSE-observations standard deviation ratio (RSR) (Legates & McCabe, 1999; Moriasi et al., 2007; Nash & Sutcliffe, 1970). The md measures the degree of model prediction error and its value varies between 0 and 1, indicating no agreement and a perfect match, respectively. NSE ranges from negative infinite to 1 and indicates how well simulated and observed action rates fit the 1:1 line. The closer to 1 the more accurate the model is. A value of MAE, MSE, or RMSE provides a deviation of the model prediction error. A smaller value indicates better model performance. PBIAS is a measure of the overall tendency of simulated values. The lower the value is the greater the accuracy. Positive and negative values indicate over- and underestimation biases, respectively. RSR varies from 0 to positive value. A lower value of RSR indicates lower RMSE or residual variation, implying better simulation performance. Equations are the following:

$$md = 1 - \frac{\sum_{i=1}^{n} |Y_i^{obs} - Y_i^{sim}|^j}{\sum_{i=1}^{n} (|Y_i^{sim} - Y^{mean}| + |Y_i^{obs} - Y^{mean}|)^j},$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Y_i^{sim} - Y_i^{obs})^2}{\sum_{i=1}^{n} (Y_i^{obs} - Y^{mean})^2},$$

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |Y_i^{obs} - Y_i^{sim}|,$$

$$MSE = \frac{1}{N} \sum_{i=1}^{n} (Y_i^{obs} - Y_i^{sim})^2,$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (Y_i^{obs} - Y_i^{sim})^2},$$

$$PBIAS = 100 \frac{\sum_{i=1}^{n} (Y_i^{sim} - Y_i^{obs})}{\sum_{i=1}^{n} (Y_i^{obs})},$$

$$RSR = \frac{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y_i^{sim})^2}}{\sqrt{\sum_{i=1}^{n} (Y_i^{obs} - Y^{mean})^2}}.$$

2.7. Categorization of Country–Commodity Combinations by Their Uncertainty

Country–commodity combinations were categorized into High, Medium, Low, and Poor/Unacceptable compliance groups based on the simulated probabilities described above. Categorization involved two steps: (1) splitting country–commodity combination groups with high or low variance of estimated probabilities of carrying quarantine pests and (2) categorizing combinations in each group into compliance levels using predetermined thresholds. To identify country–commodity pairs with highly reliable (i.e., low variance) probabilities of carrying quarantine pests, the minimum, maximum, and their absolute difference were used. Because the estimated probabilities provide 95% confidence intervals for likelihood of carrying quarantine-significant pest species for each country–commodity combination, the differences between minimum and maximum values represent a relative magnitude of uncertainty for the estimated probability. If the difference is smaller, the median probability has a narrower range, suggesting relatively high accuracy of the median probability. The absolute differences were used in combination with minimum and maximum probabilities to cluster 1,398 country–commodity combinations into "High" and "Low" variation groups. The purpose of this clustering was to obtain a set of country–commodity combinations with relatively narrow simulated probability distributions in order to decrease the uncertainty of estimated probabilities of carrying quarantine-significant pest species. The "Low" variation group represents country–commodity combinations with relatively high accuracy for estimation of the probability of carrying quarantine pests, whereas the "High" variation group contains those with a wider range of likelihoods. To cluster country–commodity combinations into "High" and "Low" variation groups, K-means clustering (K-means function in R) was conducted

Table IV. Compliance-Level Thresholds for Country–Commodity Risk Rating Based on Their Estimated Probabilities and Confidence Intervals

Uncertainty Group	Compliance Categorization	Predicted Probability (p) of Carrying Quarantine Pests
Low variation	High	$p \leq 0.003$
	Medium	$0.003 < p \leq 0.03$
	Low	$0.03 < p \leq 0.10$
	Poor/Unacceptable	$p > 0.10$
High variation	Medium	$p \leq 0.03$
	Low	$0.03 < p \leq 0.10$
	Poor/Unacceptable	$p > 0.10$

(Hartigan & Wong, 1979). The number of clusters was determined by the minimum cluster number above which total variance explained by clustering is greater than or equal to 95%. The "High" and "Low" variation groups were determined by investigating distributions of minimum, maximum, and their difference based on selected clusters. The cluster group with the smallest distribution of probabilities was considered a "Low" variation group and the rest were considered "High" variation groups.

2.8. Categorization of Country–Commodity Combinations Based on Compliance Using Predetermined Thresholds

The country–commodity combinations in "Low" or "High" variation groups were further categorized into their compliances based on predetermined thresholds (Table IV). For the "Low" variation group, compliance groups were divided into four groups based on the probability (P) of carrying quarantine pests and threshold ranges (personal communication with APHIS): "High": $p \leq 0.003$, "Medium": $0.003 < p \leq 0.03$, "Low": $0.03 < p \leq 0.10$, and "Poor": $p > 0.10$. For the "High" variation group, country–commodity combinations were categorized as follows: "Medium": $p \leq 0.03$, "Low": $0.03 < p \leq 0.10$, and "Poor": $p > 0.10$. No country–commodity combinations in the "High" variation group were categorized as "High" compliance because of the high uncertainty of the estimated probability of carrying quarantine pests.

3. RESULTS

3.1. Exploratory Data Analysis

During FY 2015, about 50% of shipments into the United States came from Costa Rica, Guatemala, and Mexico (Supporting Information Table SI). About 50% of shipments with detections of quarantine-significant pest species also came from these countries. More than 50% of the propagative plant material by volume was imported from Guatemala, Costa Rica, and El Salvador, in order of volume. Approximately 780 million individual propagative plants were imported from these three countries in about 44,000 shipments.

Petunia was the most frequently imported plant genus by number of shipments (2,200 shipments in FY 2015), followed by *Calibrachoa*, *Euphorbia*, *Verbena*, and *Salvia* (Supporting Information Table SII). By volume, the greatest amount of *Calibrachoa* was imported, followed by *Petunia*, *Pelargonium*, and *Impatiens* (some genera are not shown in the tables given in Supporting Information). The top five plant genera for carrying quarantine pests were *Tillandsia*, *Dracaena*, *Codiaeum*, *Hedera*, and *Salvia.*

By country–commodity combination, *Dracaena* from Costa Rica was the most frequently imported (Supporting Information Table SIII). A total of 881 out of 99,584 shipments imported during FY 2015 were *Dracaena* from Costa Rica. The second most frequent combination was *Petunia* (Israel), followed by *Phalaenopsis* (Taiwan), *Tillandsia* (Guatemala), and *Calibrachoa* (Israel). *Tillandsia* (Guatemala) had the greatest number of pest interceptions, followed by *Dracaena* (Costa Rica), *Codiaeum* (Costa Rica), *Hedera* (Guatemala), and *Schefflera* (Costa Rica). By volume of plants, *Hedera* (Guatemala), *Pelargonium* (Mexico), *Calibrachoa* (El Salvador), *Impatiens* (Guatemala), and *Euphorbia* (El Salvador) were the top five imported.

3.2. Estimation of Probability and Prediction Interval for Country–Commodity Combinations Carrying Exotic Species

The estimated probabilities of 1,398 country–commodity combinations carrying quarantine-significant pest species were plotted with 95% confidence intervals (Fig. 1). The x-axes of all figures are ordered by observed quarantine action disposition rates of country–commodity combinations, which allows for comparison of the pattern of predicted
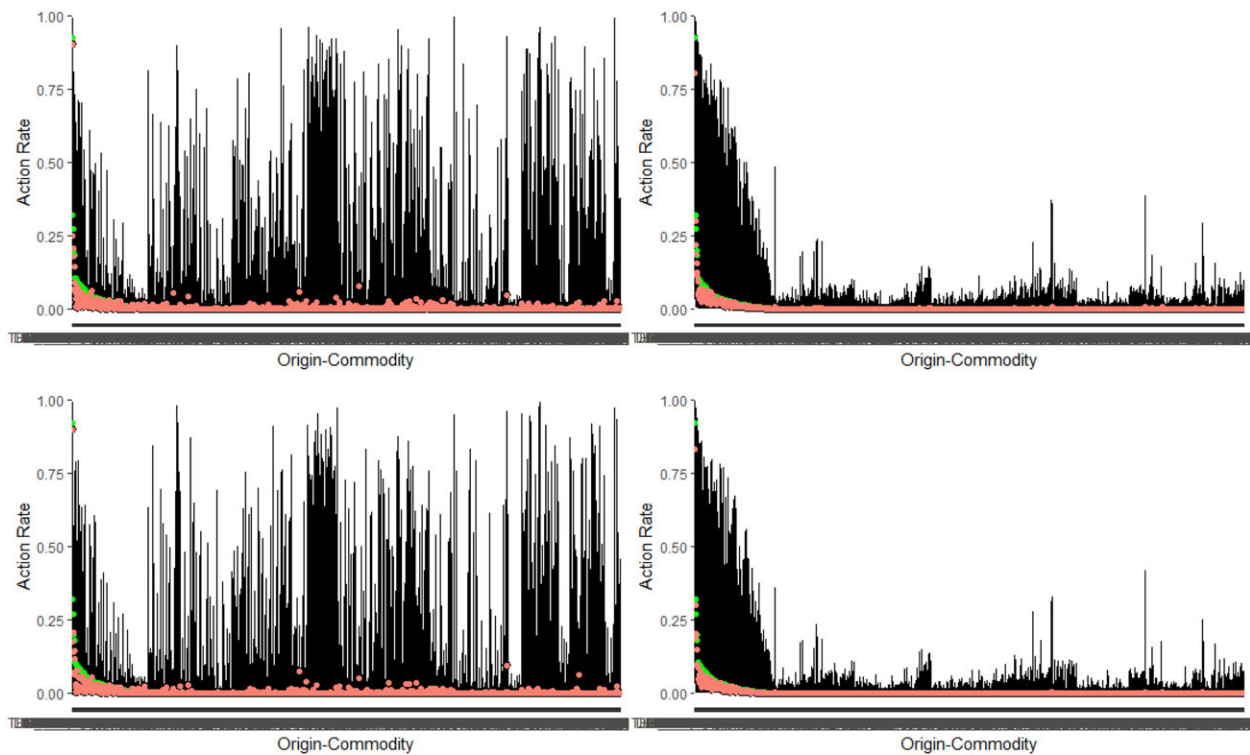
Fig. 1. Quarantine action disposition rates of observation (green dots) and simulation (salmon dots) with credible intervals (vertical bars) of country–commodity combinations from different statistical models: GLM1 (top left), GLM2 (bottom left), GLMM1 (top right), GLMM2 (bottom right).

action rates among the four figures. The result shows that predicted interception rates from the four models are highly correlated with each other and also with observed action rates (Fig. 1; Supporting Information Table SVII). The range of correlations is from 0.92 to 0.99 (Supporting Information Table SVII). The correlation is slightly greater between models with the same estimation methods (i.e., GLM1 vs. GLM2 or GLMM1 vs. GLMM2 compared to GLMs vs. GLMMs). Although predicted action rates are highly correlated to each other regardless of models, confidence intervals are dramatically different between GLMs and GLMMs (Fig. 1). In GLMMs, random effects of country–commodity combinations have five times greater variance than residuals, whereas zero variance for country is observed in either model (Supporting Information Table SIV). This result suggests that interception rates of quarantine pests on commodity genera vary depending on their countries of origin. Coefficients with 95% confidence intervals for GLMMs and posterior distributions of coefficients for selected

predictor variables in GLMs are shown in Supporting Information Fig. S1.

3.3. Model Evaluation and Comparison

The md indicates that predicted probabilities from all four models have relatively good agreement with observed interception rates in the training and the test data sets (Table V). The overall range of the md is 0.67 to 0.77. When compared with the observed quarantine pest interception rate in the test data set, however, the agreements dramatically dropped to between 0.50 and 0.52. In the training data set, the NSE values indicate that predicted probability is fairly matched, with observed value into a 1:1 line, while NSE values for the test data set are below zero, indicating that the observed mean interception rate is a better predictor. The MAE, MSE, and RMSE suggest that overall errors between predicted and observed values are very small in both the training and the test data sets. PBIAS for training data shows that the predicted probabilities of GLMMs were

Table V. Summary of Goodness-of-Fit Tests Between Predicted Probability and Observed Rate in Training and Test Data Sets

Predicted Probability Versus Observed Value with Training Data				
	GLM1	GLM2	GLMM1	GLMM2
md	0.67	0.67	0.77	0.77
NSE	0.24	0.12	0.47	0.47
MAE	0.00	0.00	0.00	0.00
MSE	0.00	0.00	0.00	0.00
RMSE	0.01	0.01	0.01	0.01
PBIAS (%)	4.49	7.56	40.97	41.19
RSR	0.81	0.87	0.68	0.68

Predicted Probability Versus Observed Value with Test Data				
	GLM1	GLM2	GLMM1	GLMM2
md	0.50	0.50	0.52	0.52
NSE	−2.19	−2.27	−2.90	−2.86
MAE	0.01	0.01	0.01	0.01
MSE	0.00	0.00	0.00	0.00
RMSE	0.03	0.03	0.03	0.03
PBIAS (%)	−9.68	−3.94	30.26	30.09
RSR	1.67	1.68	1.83	1.82

md: modified index of agreement; NSE: Nash–Sutcliffe efficiency; MAE: mean absolute error; MSE: mean squared error; RMSE: root MSE; PBIAS: percent bias; RSR: ratio of RMSE to the standard deviation of the observation.

overestimated (i.e., positive values) compared to Bayesian GLMs. Among the four models, PBIAS values for GLMs are closer to zero than those in GLMMs for the training data. RSR values, ratios of RMSE to the standard deviation of observation, indicate that predicted values from GLMMs are closer to observed values in the training data set while simulated probabilities from GLMs were better than those from GLMMs in the test data set.

3.4. Categorization of Country–Commodity Combinations

The K-means cluster analysis shows that five clusters of country–commodity combinations account for at least 95% variance of estimated quarantine pest interception probabilities and intervals, regardless of model used (Supporting Information Table SV). Among the five clusters of country–commodity combinations, those in the "rank 1" group had action disposition rates with the shortest intervals (Supporting Information Fig. S2), implying lower uncertainty than the other clusters. Thus, country–commodity combinations in the rank 1 group were considered to be a "Low" variation

group in terms of quarantine action disposition rate estimates, while the rest of the combinations were grouped as "High" variation because combinations in those groups have relatively wider intervals (Fig. 2). This suggests that estimated probabilities in the "Low" variation/uncertainty group are more accurate than those in the "High" variation/uncertainty group. When GLMM was used, a greater number of combinations were categorized as "Low" variance than when GLM was used (Table VI). With GLMs, 662 and 697 country–commodity combinations were grouped as "Low" variance, whereas 1,002 and 933 combinations belonged to "Low" when using GLMMs.

Using the predetermined thresholds (Table IV) of probabilities of carrying quarantine-significant pest species, country–commodity combinations were also categorized into four compliance levels: "High," "Medium," "Low," and "Poor/Unacceptable" (Table VII, Fig. 3). Country–commodity combinations at the "High" compliance level were only possible in the "Low" uncertainty group (Table VII). In total, 467 and 480 country–commodity combinations were categorized as "High" compliance for GLM1 and GLM2, respectively, while 991 and 925 combinations were classified into the "High" compliance group for GLMM1 and GLMM2. Country–commodity combinations at a "High" compliance level are regarded as having a lower probability of carrying quarantine-significant pest species, which suggests that the trading partners for these combinations (i.e., the countries in the country–commodity pairs) have cleaner commodities than those with lower compliance levels. For the purpose of enhancing cost-effective inspection, the sampling intensity for country–commodity combinations in the "High" compliance group may be reduced below the current sampling intensity. The numbers of "High" compliance country–commodity combinations from GLMMs were approximately twice as high as those from GLMs, while the numbers of "Medium" or "Low" compliance combinations were greater in GLMs than in GLMMs. The number of combinations in the "Medium" or "Low" compliance groups were 926, 912, 400, and 468 for GLM1, GLM2, GLMM1, and GLMM2, respectively. The numbers of country–commodity combinations in "Poor/Unacceptable" compliance groups were between five and seven, regardless of the model used (Table VII). A greater number of total country–commodity combinations were observed from GLMs than from GLMMs.
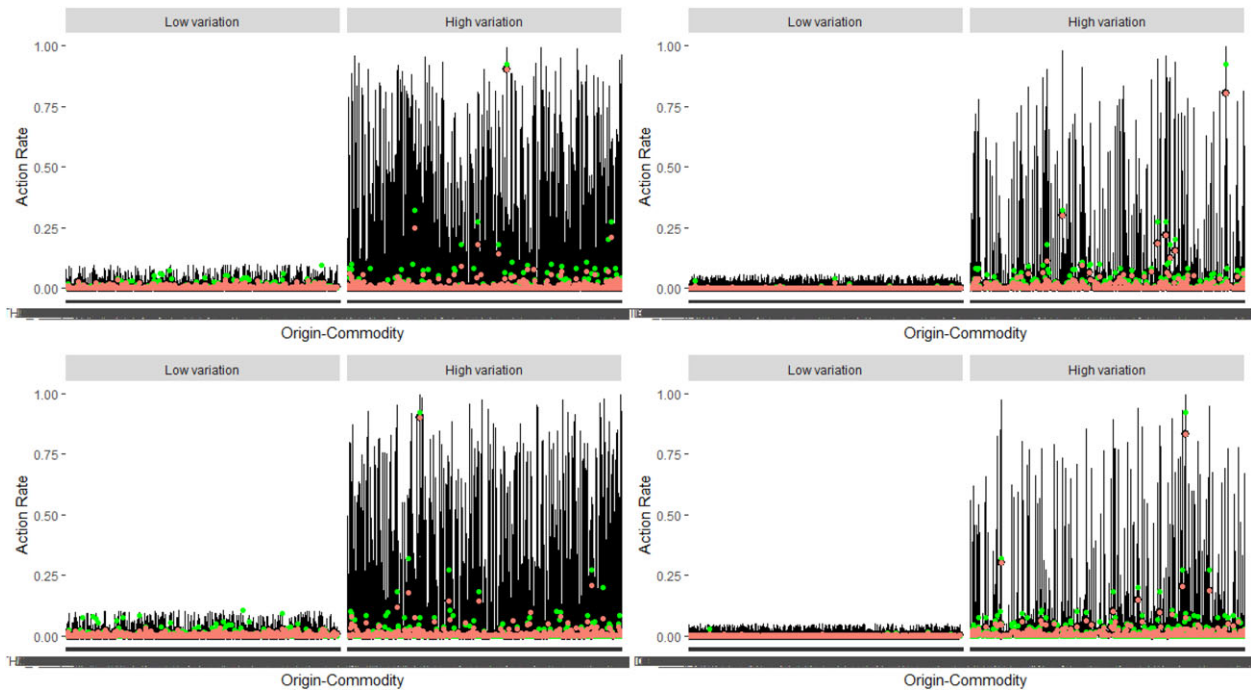
Fig. 2. Observed (green dots) and predicted (salmon dots) quarantine action disposition rates of country–commodity combinations by Low and High variation of four statistical models (clockwise from top left: GLM1, GLMM1, GLMM2, GLM2).

Table VI. Frequency of Country–Commodity Combinations in Clustering and Risk Rating Groups Resulting from Different Statistical Model Fittings

Cluster[a]	GLM1	GLM 2	GLMM1	GLMM2	Variance or Uncertainty Level	Further Rating
1	662	697	1,002	933	Low	Used[b]
2	269	248	235	292	High	Not used
3	178	179	65	68	High	Not used
4	163	151	54	58	High	Not used
5	126	123	42	47	High	Not used
Total OCs ≥10 shipments	1,398	1,398	1,398	1,398		
Not used	3,603	3,603	3,603	3,603		
Total	5,001	5,001	5,001	5,001		

[a]Clusters in order of increasing interval variance.
[b]Only country-commodities with low uncertainty for action rates were used for further rating.

4. DISCUSSION

In this study, we categorized country–commodity combinations into different compliance levels based on simulated interception rates of quarantine species and predetermined thresholds. We also compared the categorization results among models. To do this, we used two steps. First, country–commodity combinations were separated into small and large variance groups based on the confidence intervals of the estimated probabilities of carrying quarantine pest species. Second, each group was further partitioned into compliance levels (High, Medium, Low, and Poor/Unacceptable) using thresholds (Table VII).

We found that the use of statistically estimated probabilities of interception and their confidence intervals is a feasible and promising approach for rating compliance of country–commodity combinations. The predictive performance of models with the test data set showed that GLMMs have slightly better predictive power than GLMs (Table V) based on md. Validation with an entirely new data set not used for modeling is essential for measuring predictive performance. Since the ultimate goal of

Table VII. Compliance Levels of Country–Commodity Combinations with Uncertainty Levels from Different Statistical Models

| Variance or Uncertainty Level	Compliance Level	GLM1	GLM2	GLMM1	GLMM2
Low variance					
	High	467 (33.4%)	480 (34.3%)	991 (70.9%)	925 (66.2%)
	Medium	193 (13.8%)	215 (15.4%)	11 (0.8%)	8 (0.6%)
	Low	2 (0.1%)	2 (0.1%)	0 (0%)	0 (0%)
	Poor/Unacceptable	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Total	662 (47.4%)	697 (49.9%)	1,002 (71.7%)	933 (66.7%)
High variance	Medium	701 (50.1%)	667 (47.7%)	360 (25.8%)	426 (30.5%)
	Low	30 (2.1%)	28 (2%)	29 (2.1%)	34 (2.4%)
	Poor/Unacceptable	5 (0.4%)	6 (0.4%)	7 (0.5%)	5 (0.4%)
	Total	736 (52.6%)	701 (50.1%)	396 (28.3%)	465 (33.3%)
Grand total		1,398			
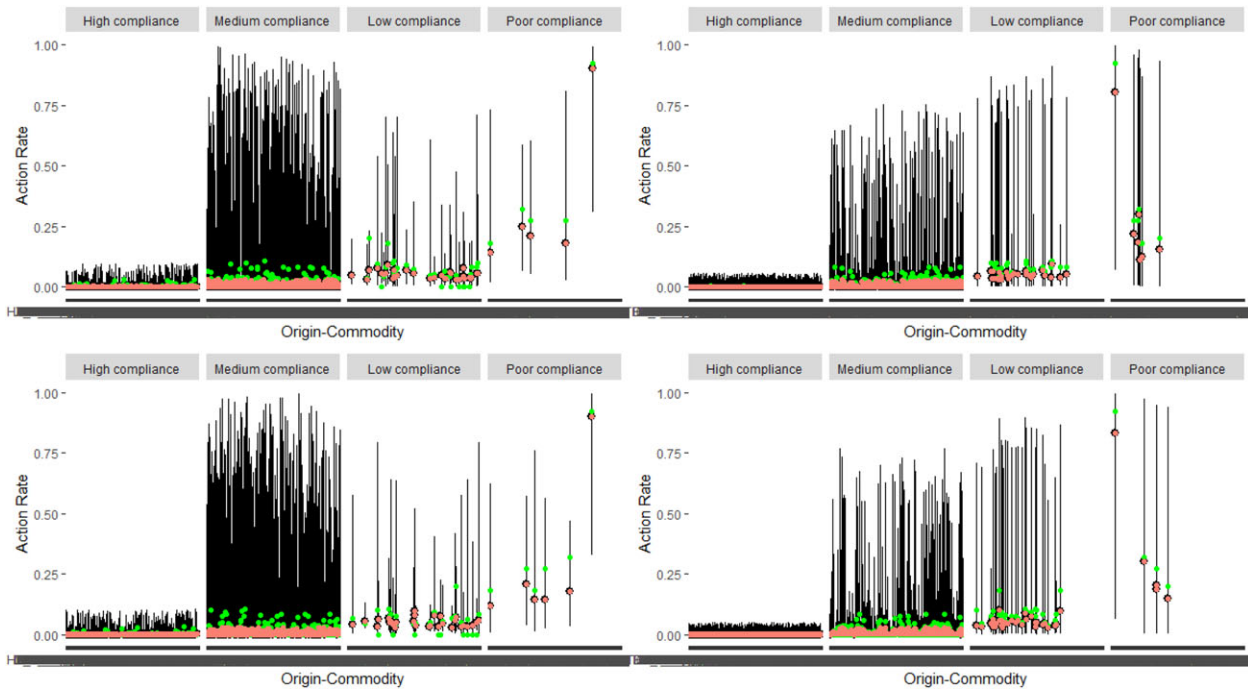
Fig. 3. Quarantine action disposition rates of simulation (salmon dots) and observation (green dots) by grouping with compliance levels (within Low variance) and High variance based on FY 2015 values: GLM1 (top left), GLM2 (top right), GLMM1 (bottom left), and GLMM2 (bottom right).

the study is to predict interception probability for future imported country–commodity combinations, the predictive performance with a new data set (i.e., test data) provides a more realistic assessment of the best model choice(s) in our study (Shmueli, 2010). The predictive performance of all models was much better with the training data set than with the test data set. Indices for goodness-of-fit tests showed that the observed interception rates corresponded relatively well to the models with the training data, while the indices dropped dramatically with test data (e.g.,

modified indices of agreement in Table V), implying overfitting of the training data, especially in GLMMs. This result indicates that models used in our study might exhibit patterns specific to the training data set rather than true properties of the unknown function. In other words, the test data set does not demonstrate the pattern found in the training data set (James, Witten, Hastie, & Tibshirani, 2013). It may imply that more variables should be explored to account for variability. The split based on inspection time was necessary because the goal of the modeling

is to predict interception probabilities of newly imported commodities in future. Additional validation studies with training and test data sets selected in various ways have shown that results can differ depending on split methods (Supporting Information Table SVI). When the test data set had a similar amount of data to the training set, the predictive performance with the test data had more predictive power. Regardless of data split method, the predictive power was always greater in GLMMs than in GLMs. The different results among various validation studies may be due to both data split method (random vs. chronological) and amount of data (training vs. test) (Jamieson et al., 2013; Shmueli, 2010). In our study, the predictive power tends to be better when training and test data sets have similar amounts of data and are split by random selection (Validation study 2 in Supporting Information Table SVI). In this article, the result from the initial validation study was presented since it was our first choice and it may better represent the reality of the inspection situation in predicting compliance levels of individual imported consignments.

Calculated compliance levels for country–commodity pairs were also influenced by the type of statistical models used (Table VII). In general, more country–commodity combinations were categorized into the "High" compliance group when using GLMMs than when using GLMs. A minimum of 33% and a maximum of 71% of commodities were classified as being in the "High" compliance group. This result indicates that a country–commodity combination can be categorized into one compliance level (e.g., "High" compliance) in one statistical model and into another compliance level (e.g., "Medium") in another model. Although an in-depth discussion of statistical algorithms is beyond the scope of this study, inconsistency of grouping among models may be the result of different algorithms of parameter estimation and variable selection. Since the choice of model may significantly affect the compliance level of country–commodity combinations for quarantine pests, the choice(s) of model for applying results to the real world should be made cautiously, with careful consideration of operational aspects such as the operation system, the availability of inspection resources, and the policies in place. For example, if available inspection resources are limited, then models with relatively greater numbers of "High" compliance levels (e.g., GLMM1 or GLMM2) can be used for developing an inspection program so that more inspection effort can be focused on those country–commodity pairs with high-risk levels (i.e., "Medium" and/or "Low" compliance). Depending on the availability and capacity of inspection resources, categorization results from multiple models can be selected to efficiently accommodate inspection limitations. Instead of using the result from one model, the use of combined categorization results from four models can be used to prioritize the allocation of inspection resources (Shmueli, 2010).

The separation of country–commodity combinations into "Low" and "High" variability groups may be a useful step before compliance categorization with thresholds because our method only allowed combinations with "Low" variation to obtain the "High" compliance country–commodity categorization (Fig. 3). Confidence intervals in the "High" compliance group were smaller than those in other compliance groups. This means that the estimated quarantine pest probabilities in the "High" compliance group have relatively narrower interval ranges than other clustered groups, implying that the estimates are more accurate than those in other groups. In the study, the "High" compliance group (i.e., probability of carrying quarantine pest species ≤ 0.003) was selected only from the "Low" variation commodity group (Figs. 2 and 3). The purpose of this limitation was to obtain country–commodity combinations with a relatively high degree of certainty that they have a low probability of carrying quarantine-significant pest species. This is very important for the operation aspects of inspection program implementation. When allocating inspection effort based on compliance levels, reduced inspection resources will be assigned to the "High" compliance group, so the chance of passing quarantine-significant pest species may be higher unless the categorization is accurate. Although there were differences depending on which model was used, the country–commodity pairs in the upper limit of intervals in the "Low" variation group had at most about a 10% probability of carrying exotic species. GLMs have greater upper limits than GLMMs in our results (data not shown). Like the categorization of compliance level, the classification results of the high and low variation groups varied depending on the model estimation method and variable selection (Table VII; Fig. 2). Thus, careful use of statistical modeling and variables is required. Although the inconsistency among models may be inevitable, more data collection and careful choice of models and variables may reduce it to some extent.

The K-means function in R has several algorithms as options (the default is "Hartigan–Wong"), which vary in optimization methods and data types and Euclidean distance is used to calculate sum of squared deviations from data points and a centroid (Forgy, 1965; Hartigan & Wong, 1979; Lloyd, 1982; MacQueen, 1967; Morissette & Chartier, 2013; Slonim, Aharoni, & Crammer, 2013). Since results from different algorithms were almost identical in our study (data not shown), we present the clustering for variation from the default "Hartigan–Wong" algorithm (Fig. 2). Although the effect of clustering algorithms on grouping variations was negligible in this study, confirmation is still recommended for future studies.

Selection of threshold values for compliance levels may be one of the most important and complicated steps in designing a compliance categorization process. In this study, thresholds were determined by a USDA-APHIS expert panel review of cluster analysis results in order to reflect the realities of the inspection system. To support decision making in the selection of optimal thresholds by APHIS expert panel review, the simulation study of cluster analysis generated the best split of country–commodity based on interception rate. Cluster cutoff values were used as thresholds for placing the combinations into compliance categories. Thus, the compliance levels actually indicate probabilities of carrying quarantine pest species that are statistically different among compliance groups in our study. We considered the input of the expert panel to modify the threshold values so that they were more closely aligned with actual inspection systems. It may be more reasonable to consider operational and policy aspects in addition to analytical calculations when determining compliance-level thresholds. For example, limited inspection resources can be allocated differently for different country–commodity pairs depending on compliance levels. Such resource limitations should be reflected in setting threshold values. Further study is required to determine optimum threshold values for categorization of compliance level.

Categorization of imported commodities based on their probabilities of carrying exotic species is important for improving inspection programs. Such categorizations can be applied to the allocation of inspection resources; for example, by differentiating sampling efforts depending on categorical compliance levels (Robinson et al., 2011). Our study shows how interception data for propagative plant materials can be used in statistical models to generate estimated probabilities of quarantine pest presence,

along with confidence intervals, all of which can be used to categorize country–commodity combinations into compliance levels. Although more research on the operational aspects of categorization is needed, this approach for categorizing commodities based on estimated probability and the corresponding uncertainty is a realistic method for improving inspection programs by allocating resources based on the likelihood of intercepting quarantine-significant species.

5. CONCLUSION

Our study aims at developing and comparing statistical models and categorizing compliance levels of country–commodity combinations based on predicted interception rates. GLM and GLMM were used to estimate interception rates of country–commodity combinations and their uncertainty (i.e., 95% confidence ranges of predicted interception rates). Model validation with a new test data set was used to assess predictive performance of statistical modeling. In this particular study, GLMMs slightly outperformed GLMs when compared on predictive performance, although their predictabilities with the new "future" test data set were lower than those with the training data. This implies that there may be overfitting in the statistical modeling or that the test data set was generically different from the training data set. Additional studies with more data are necessary to confirm the trend. Predicted interception rates and their confidence intervals were highly influenced by the statistical models used. Thus, careful consideration is required to apply the information to the development of inspection programs with different monitoring intensities based on compliance levels of country–commodity combinations. Nonetheless, the results from our study with empirical inspection data will help to provide a guideline for assessing risk of invasive quarantine species introductions and for developing inspection programs for country–commodity combinations and other possible pest introduction pathways.

possible, in part, by a Cooperative Agreement from the U.S. Department of Agriculture's Animal and Plant Health Inspection Service (APHIS). It may not necessarily express APHIS' views.

REFERENCES

Areal, F. J., Touza, J., MacLeod, A., Dehnen-Schmutz, K., Perrings, C., Palmieri, M. G., & Spence, N. J. (2008). Integrating drivers influencing the detection of plant pests carried in the international cut flower trade. *Journal of Environmental Management*, *89*(4), 300–307. https://doi.org/10.1016/j.jenvman.2007.06.017

Blackburn, T. M., Pysek, P., Bacher, S., Carlton, J. T., Duncan, R. P., Jarosik, V., … Richardson, D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*, *26*(7), 333–339. https://doi.org/10.1016/j.tree.2011.03.023

Convention, I. P. P. (2007). *International standards for phytosanitary measures no. 5: Glossary of phytosanitary terms*. Rome, Italy: Food and Agriculture Organization.

Davis, C. E., Hyde, J. E., Bangdiwala, S. I., & Nelson, J. J. (1986). An example of dependencies among variables in a conditional logistic regression. In S. H. Moolgavkar & R. L. Prentice (Eds.), *Modern statistical methods in chronic disease epidemiology* (pp. 140–147). Hoboken, NJ: Wiley.

Eschen, R., Roques, A., & Santini, A. (2015). Taxonomic dissimilarity in patterns of interception and establishment of alien arthropods, nematodes and pathogens affecting woody plants in Europe. *Diversity and Distributions*, *21*(1), 36–45. https://doi.org/10.1111/ddi.12267

Essl, F., Winter, M., & Pysek, P. (2012). Trade threat could be even more dire. *Nature*, *487*(7405), 39.

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*, 768–769.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*, 1360–1383. https://doi.org/10.1214/08-AOAS191

Govindaraju, K., Bebbington, M., & Wrathall, T. (2010). Statistical evaluation of the New Zealand Food Safety Authority sampling protocol for imported food. *Risk Analysis*, *30*(5), 817–826. https://doi.org/10.1111/j.1539-6924.2009.01346.x

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Berlin, Germany: Springer.

Jamieson, L., DeSilva, H., Worner, S., Rogers, D., Hill, M., Walker, J., & Zydenbos, S. (2013). A review of methods for assessing and managing market access and biosecurity risks using systems approaches. *New Zealand Plant Protection*, *66*, 1–9.

Kenis, M., Rabitsch, W., Auger-Rozenberg, M.-A., & Roques, A. (2007). How can alien species inventories and interception data help us prevent insect invasions? *Bulletin of Entomological Research*, *97*(5), 489–502. https://doi.org/10.1017/S0007485307005184

Lange, K. L., Roderick, J. A. L., & Jeremy, M. G. T. (1989). Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association*, *84*(408), 881–896. https://doi.org/10.2307/2290063

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic

model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Levine, J. M., & D'Antonio, C. M. (2003). Forecasting biological invasions with increasing international trade [Predicción de las Invasiones Biológicas con el Incremento del Comercio Internacional]. *Conservation Biology*, *17*(1), 322–326. https://doi.org/10.1046/j.1523-1739.2003.02038.x

Liebhold, A. M., Brockerhoff, E. G., Garrett, L. J., Parke, J. L., & Britton, K. O. (2012). Live plant imports: The major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment*, *10*(3), 135–143. https://doi.org/10.1890/110198

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Lodge, D. M., Williams, S., MacIsaac, H. J., Hayes, K. R., Leung, B., Reichard, S., … McMichael, A. (2006). Biological invasions: Recommendations for US policy and management. *Ecological Applications*, *16*(6), 2035–2054. https://doi.org/10.1890/1051-0761(2006)016[2035:bIrfup]2.0.co;2

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA*, 281–297.

McCullough, D. G., Work, T. T., Cavey, J. F., Liebhold, A. M., & Marshall, D. (2006). Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. *Biological Invasions*, *8*(4), 611–630. https://doi.org/10.1007/s10530-005-1798-4

Meyerson, L. A., & Mooney, H. A. (2007). Invasive alien species in an era of globalization. *Frontiers in Ecology and the Environment*, *5*(4), 199–208. https://doi.org/10.1890/1540-9295(2007)5[199:iAsiae]2.0.co;2

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900.

Morissette, L., & Chartier, S. (2013). The K-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, *9*(1), 15–24.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I: A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

Pimentel, D., Lach, L., Zuniga, R., & Morrison, D. (2000). Environmental and economic costs of nonindigenous species in the United States. *BioScience*, *50*(1), 53–65.

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, *52*(3), 273–288. https://doi.org/10.1016/j.ecolecon.2004.10.002

Puth, L. M., & Post, D. M. (2005). Studying invasion: Have we missed the boat? *Ecology Letters*, *8*(7), 715–721. https://doi.org/10.1111/j.1461-0248.2005.00774.x

Pysek, P., Jarosik, V., Hulme, P. E., Kuhn, I., Wild, J., Arianoutsou, M., … Winter, M. (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12157–12162. https://doi.org/10.1073/pnas.1002314107

Pysek, P., Jarosik, V., Hulme, P. E., Pergl, J., Hejda, M., Schaffner, U., & Vila, M. (2012). A global assessment of invasive plant impacts on resident species, communities and ecosystems: The interaction of impact measures, invading species' traits and environment. *Global Change Biology*, *18*(5), 1725–1737. https://doi.org/10.1111/j.1365-2486.2011.02636.x

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robinson, A., Burgman, M. A., & Cannon, R. (2011). Allocating surveillance resources to reduce ecological invasions: Maximizing detections and information about the threat. *Ecological Applications*, 21(4), 1410–1417. https://doi.org/10.1890/10-0195.1

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., … Ellstrand, N. C. (2001). The population biology of invasive species. *Annual Review of Ecology and Systematics*, 32(1), 305–332.

Seebens, H., Essl, F., Dawson, W., Fuentes, N., Moser, D., Pergl, J., … Blasius, B. (2015). Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology*, 21(11), 4128–4140. https://doi.org/10.1111/gcb.13021

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. https://doi.org/10.1214/10-STS330

Simberloff, D., Martin, J. L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., … Vila, M. (2013). Impacts of biological invasions: What's what and the way forward. *Trends in Ecology & Evolution*, 28(1), 58–66. https://doi.org/10.1016/j.tree.2012.07.013

Slonim, N., Aharoni, E., & Crammer, K. (2013). Hartigan's K-means versus Lloyd's K-means: Is it time for a change? *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China*.

Surkov, I. V., Lansink, A., van Kooten, O., & van der Werf, W. (2008). A model of optimal import phytosanitary inspection under capacity constraint. *Agricultural Economics*, 38(3), 363–373. https://doi.org/10.1111/j.1574-0862.2008.00306.x

Tatem, A. J., Hay, S. I., & Rogers, D. J. (2006). Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), 6242–6247. https://doi.org/10.1073/pnas.0508391103

Vila, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarosik, V., Maron, J. L., … Pysek, P. (2011). Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, 14(7), 702–708. https://doi.org/10.1111/j.1461-0248.2011.01628.x

Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., … Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research-Oceans*, 90(NC5), 8995–9005. https://doi.org/10.1029/JC090iC05p08995

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32(13), 2088–2094. https://doi.org/10.1002/joc.2419

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis*, 13(2), 157–170. https://doi.org/10.1093/pan/mpi009

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table SI. Summary of Top 10 Countries Based on Shipment

Table SII. Summary of Top 10 Commodities

Table SIII. Top 10 Country–Commodity Combinations

Table SIV. Variance of Random Effects from GLMMs

Table SV. Total Sum of Squares Explained by Each Group

Table SVI. Modified Index of Agreement Between Observed and Predicted Rates with Train and Test Data Sets Selected in Various Ways

Table SVII. Correlation Coefficients of Observed and Predicted Action Rates from Statistical Models

Fig. SI. Coefficients with 95% credible intervals for GLMM and posterior distributions of selected coefficients in GLM.

Fig. S2. Clustering results (clockwise from top left: GLM1, GLMM1, GLMM2, GLM2).