Appendix: Integrating public health surveillance and environmental data to model presence of Histoplasma in the United States

1 Additional Details

1.1 Histoplasmosis Case Definition (pre-2016)

The following table contains the histoplasmosis case definitions prior to 2016 for the states included in our analysis.

State	Pre-2016 definition			
	Clinical description	**Confirmed**	**Probable**	**Suspect**
Alabama	A case of acute pulmonary histoplasmosis is defined as an influenza-like illness with two or more of the following symptoms: fever/ chills, cough, chest pain, and/or myalgia/arthralgia. Histoplasmosis infection may also be asymptomatic or present with non-influenza-like symptoms including abdominal pain, mediastinal lymphadenopathy, hepatosplenomegaly, and skin lesions.	Lab Criteria: • Isolation of H. capsulatum from culture of sputum, urine, tissue, etc. • Identification of organisms in tissues by histopathology. • A fourfold rise between acute and convalescent CF titers. Case Classification: • A case diagnosed as acute pulmonary histoplasmosis AND meets at least 1 of the confirmatory lab criteria. • A case with clinically compatible symptoms (not diagnosed as acute pulmonary histoplasmosis) AND meets at least 1 of the confirmatory lab criteria. • Detection of H. capsulatum antigen in urine or serum. Results of ≥2 EIA units or ≥0.6ng/mL are considered positive. • Detection of H and M band by immunodiffusion.	Lab Criteria: A single CF titer ≥1:32 and M band detection. Case Classification: • A case diagnosed as acute pulmonary histoplasmosis AND meets presumptive (probable) lab criteria. • A case with clinically compatible symptoms (not diagnosed as acute pulmonary histoplasmosis) AND meets presumptive lab criteria. • A case with clinically compatible symptoms, no confirmatory lab evidence, AND is epi-linked to suspected source during an outbreak.	Case Classification: • An asymptomatic case that meets confirmatory (confirmed) or presumptive (probable) lab criteria.
Arkansas	N/A	Cases were considered to be confirmed if the patient had a measured fever ≥101°F (≥38.3°C), and either cough, chest pain, shortness of breath, or abnormal CXR, and at least one positive culture, antigen, or serologic test for Histoplasma.	Cases were considered probable if the patient has symptoms consistent with histoplasmosis (self-reported fever and either cough, chest pain, or shortness of breath) and at least one positive culture, antigen, or serologic test for Histoplasma. Because subclinical illness and illness for which no histoplasmosis tests were performed were observed in this outbreak, when a confirmed case was identified, the definition of a probable case was broadened to include any person exposed to the site or event who also had clinical features of fever ≥101°F (≥38.3°C) and at least one of cough, chest pain, shortness of breath, or abnormal CXR within 3 weeks of exposure, even in the absence of laboratory testing for Histoplasma.	N/A
Delaware	N/A	N/A	N/A	N/A
Illinois	A case of acute respiratory histoplasmosis is defined as an influenza-like illness with two or more of the following symptoms: fever, cough, shortness of breath or chest pain.	Acute Respiratory Histoplasmosis: A patient that is clinically compatible, and culture positive for H. capsulatum or has a four-fold rise in titers collected 2-4 weeks apart (numbers 1 or 2 in laboratory criteria).	Acute Respiratory Histoplasmosis: A case that is clinically compatible, not culture confirmed, does not have a four-fold rise in titer, but positive by any of the other laboratory methods listed above (numbers 3-8 in	N/A

Table 1: Histoplasmosis case definitions prior to 2016

State	Pre-2016 definition			
	Clinical description	**Confirmed**	**Probable**	**Suspect**
	Acute histoplasmosis may occur in the absence of respiratory symptoms. The symptoms of acute disseminated histoplasmosis include fever and weight loss.	Acute Disseminated Histoplasmosis: A case that has fever and weight loss with or without respiratory symptoms, and a positive culture from an extrapulmonary site.	laboratory criteria). A case is also considered probable with clinically compatible symptoms, no laboratory confirmation, and an epi link to a suspected source during an outbreak. Acute Disseminated Histoplasmosis: A case that has fever and weight loss with or without respiratory symptoms, and evidence of H. capsulatum by either histopathology staining or DNA probe of a specimen from an extrapulmonary site.	
Indiana	Infections may take one of five forms: • Asymptomatic - individual has no clinical signs or symptoms, but as immunological evidence of infection. • Acute disseminated - is an illness of short duration that involves other organs in addition to the lungs. It is marked by cough, exhaustion and enlargement of the liver and spleen. • Acute benign respiratory - characterized by weakness, fever, chest pains, and cough. Symptom severity is dependent on magnitude of exposure to fungal conidia. • Chronic disseminated - a prolonged illness involving organs other then the lungs. It may include fever, anemia, hepatitis, pneumonia, endocarditis, meningitis, and ulcers of the mouth, tongue, nose, and larynx. • Chronic pulmonary - resembles tuberculosis.	Lab Tests Sufficient for Case Confirmation: • Culture - a positive culture for Histoplasmosis capsulatum. • Histological - any pathological finding indicating an infection with histoplasma. • Complement-fixing antibodies - presence of antibodies to yeast (Y) or mycelial (M) antigens in dilutions greater than 1:16 in patients with a compatible clinical presentation and no other explanation for his/her illness. • Immunodiffusion testing - The presence of H band antibodies is indicative of recent infection (within 6 months). The presence of H band antibodies (with or without M band) with a compatible clinical presentation and no other explanation for his/her illness. • Serum or urine antigen - A positive test with a compatible clinical presentation.	N/A	N/A
Kentucky	A systemic fungal infection of varying severity caused by Histoplasma capsulatum. Infection may be asymptomatic or take one of four clinical forms: • Acute benign respiratory - mild respiratory illness with general malaise, fever, chills, headache, myalgia, chest pains, nonproductive cough and scattered small calcifications of the lung. • Acute disseminated - debilitating fever, GI symptoms, bone marrow suppression, lymphadenopathy. Most frequent in children and	• Isolation of H. capsulatum from culture of bone marrow, sputum, or lesions, OR • Histological demonstration of intracellular yeast cells from bone marrow or tissue biopsy, OR • Detection of H. capsulatum polysaccharide antigen in urine or serum, OR • Rise in CF titers to either histoplasmin or yeast-phase antigen.	N/A	N/A

Table 1: Histoplasmosis case definitions prior to 2016 (cont.)

State	Pre-2016 definition			
	Clinical description	**Confirmed**	**Probable**	**Suspect**
	immunosuppressed; fatal if not treated. • Chronic pulmonary - clinically and radiologically resembles chronic pulmonary tuberculosis with cavitations, usually in middle-aged and elderly persons with underlying emphysema • Chronic disseminated - low-grade fever, weight loss, weakness, liver and spleen enlargement, mucosal ulcers, subacute course with slow progression; fatal if not treated.			
Michigan	A case of acute histoplasmosis is defined as an influenza-like illness with two or more of the following symptoms: fever/chills, cough, chest pain, weakness, or myalgia/arthralgia.	In order for a patient to be considered to have an acute case of histoplasmosis, they must have a clinically compatible illness coupled with laboratory evidence of infection. Lab Criteria: • A four-fold rise in compliment fixation titer between serum specimens collected 2-4 weeks apart • Identification of the organism in tissues by histopathology • Isolation of the organism from cultures	Lab Criteria: • Complement fixation titer to the yeast-phase antigen ≥1:32 or • H band detected by Immunodiffusion testing or • Detection of antigen in body fluids including urine, serum, cerebral spinal fluid, and broncho-alveolar lavage fluid	N/A
Minnesota	N/A	N/A	N/A	N/A
Mississippi	N/A	N/A	N/A	N/A
Nebraska	N/A	Lab Criteria: • Four-fold rise in complement fixation titer between acute and convalescent sera, OR • Culture isolation and confirmation of Histoplasma capsulatum from a clinical specimen, OR • Identification of Histoplasma capsulatum in tissue using an organism-specific PCR assay	Lab Criteria: • Complement fixation titer to the yeast-phase antigen ≥1:32, OR • H band detected by immunodiffusion testing, OR • Detection of antigen in a body fluid including urine, serum, csf, and bronchoalveolar lavage fluid, OR • Detection of characteristic yeast forms in tissue that are morphologically consistent with histoplasmosis, OR • Detection of characteristic intracellular yeast forms in a phagocyte in a peripheral blood smear	N/A
Pennsylvania	N/A	A case of histoplasmosis was defined as 1) a laboratory-confirmed H. capsulatum infection or 2) self-reported fever and two additional symptoms (i.e., headache, cough, chest pain, or difficulty breathing). Laboratory-confirmation was defined as either a urine or	N/A	N/A

Table 1: Histoplasmosis case definitions prior to 2016 (cont.)

State	Pre-2016 definition			
	Clinical description	**Confirmed**	**Probable**	**Suspect**
		serum Histoplasma antigen enzyme immunoassay (EIA) test result of >0.6 ng/mL.		
Wisconsin	A systemic fungal infection of varying severity caused by Histoplasma capsulatum. Infection may be asymptomatic or take one of four clinical forms: • Acute benign respiratory - mild respiratory illness with general malaise, fever, chills, headache, myalgia, chest pains, nonproductive cough and scattered small calcifications of the lung. • Acute disseminated - debilitating fever, GI symptoms, bone marrow suppression, lymphadenopathy. Most frequent in children and immunosuppressed; fatal if not treated. • Chronic pulmonary - clinically and radiologically resembles chronic pulmonary tuberculosis with cavitations, usually in middle-aged and elderly persons with underlying emphysema. • Chronic disseminated - low-grade fever, weight loss, weakness, liver and spleen enlargement, mucosal ulcers, subacute course with slow progression; fatal if not treated.	Clinically compatible illness with laboratory confirmation. Lab Criteria: • Isolation of H. capsulatum from culture of bone marrow, sputum, or lesions, OR • Histologic demonstration of intracellular yeast cells from bone marrow or tissue biopsy, OR • Detection of H. capsulatum polysaccharide antigen in urine or serum, OR • Rise in CF titers to either histoplasmin or yeast-phase antigen. • Positive serology test for anti-H band antibody	N/A	N/A

Table 1: Histoplasmosis case definitions prior to 2016 (cont.)

1.2 Model Specification

The proposed model assumes for county i, the probability of the presence of *H. capsulatum* is given by

$$\psi_i = \Phi\left(\mathbf{X}_i\boldsymbol{\alpha} + \eta_i\right),$$

where $\mathbf{X}_i$ is a vector of environmental covariates, $\boldsymbol{\alpha}$ a vector of regression coefficients, and η_i a spatial random effect. It is well known that the spatial random effect can be confounded with the environmental variables [6]. To alleviate this confounding and reduce the dimension of the spatial random effect, we assume the spatial random effect depends on Moran's I basis functions [5, 3]. That is,

$$\psi_i = \Phi\left(\mathbf{X}_i'\boldsymbol{\alpha} + \mathbf{K}_i\boldsymbol{\gamma}\right), \tag{1}$$

where $\mathbf{K}$ is an $n \times q$ matrix of Moran's I basis functions, and $\boldsymbol{\gamma}$ is a q-dimensional random effect.

Following Hughes and Haran (2013) [4], we assume $\boldsymbol{\gamma}$ is multivariate normal with mean zero and precision matrix $\mathbf{K}'(\mathbf{D} - \mathbf{A})\mathbf{K}/\sigma^2$, where $\mathbf{A}$ is an adjacency matrix whose (i, j) element is 1 if county i and county j share a border and is 0 otherwise, and $\mathbf{D}$ is a diagonal matrix whose (i, i)th entry is the total number of neighbors of county i. This form for the random effect $\boldsymbol{\gamma}$ along with the inclusion of the Moran's I basis functions yields conditional posterior distributions for $\boldsymbol{\alpha}$ that have the same variance as under the non-spatial ordinary linear model and effectively removes confounding between the fixed and random effects, permitting interpretation of $\boldsymbol{\alpha}$ as in the ordinary linear model.

Improper uniform priors on the real line were used for all regression coefficients, and inverse gamma prior distributions with parameters 0.5, 0.5 were assumed for the variance parameters σ^2 and τ^2. The temporal autocorrelation parameter ρ was assumed to follow a uniform distribution over $(0, 1)$. The posterior distribution was simulated using a Markov chain Monte Carlo (MCMC) algorithm coded in Matlab. For computational efficiency, we employ the data augmentation approach of Albert and Chibb (1993) [1]. That is, we assume latent, Gaussian versions of the detection and occupancy processes, given by $\tilde{Y}_{it}$ and $\tilde{Z}_i$, and then define $Y_i = \mathrm{I}\left(\tilde{Y}_{it} > 0\right)$ and $Z_i = \mathrm{I}\left(\tilde{Z}_i > 0\right)$, where $\mathrm{I}(\cdot)$ denotes the indicator function. The data augmentation scheme permits Gibbs sampling updates in the MCMC algorithm. Convergence was assessed by visual inspection of trace plots.

Our model is also used to simulate from the posterior predictive distribution of ψ_i for counties in states outside of our study region. Doing so requires the availability of covariate information at these additional counties. In addition, the Moran's I basis functions must be computed as in [2] by initially included all locations of interest (those in the study area in addition to the counties to be predicted) and then subsetting the rows based on whether or not the corresponding county is in the study area. For county j not in the study area, draws are generated from the posterior predictive distribution of ψ_j by computing $\psi_j = \Phi\left(\mathbf{X}_j'\boldsymbol{\alpha} + \mathbf{K}_j\boldsymbol{\gamma}\right)$ for each $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ simulated by the MCMC algorithm.

2 Results

2.1 Reversible Jump MCMC

Reversible Jump MCMC was used to assist with variable selection and select the final subsets of covariates to be used for the occupancy and detection processes [7]. For the occupancy process, we initially considered the proportion of the county with land-cover of cultivated crops, the log of farm nitrogen, the log of non-farm nitrogen, the log of the mean elevation of the county, the latitude and longitude of the county centroid, the proportion of the county that is covered with water, and the proportion of the county that is undeveloped. For the detection process, we initially considered the proportion of the county with land-cover of cultivated crops, the average maximum temperature for the county in that month, soil moisture, the log of the population density, the log of the percent of the population that is non-white, the log of the percent of the population that is Latino, the proportion of the population that work in construction, the proportion that work in agriculture, and the proportion of the population with private insurance. We removed variables that had a relatively low estimated probability of inclusion following the RJMCMC algorithm. This led to the removal of undeveloped in the occupancy process and average maximum temperature, percent Latino, and percent working in construction from the detection process. The remaining variables were included in the final analysis.

2.2 Additional Results

Below are plots of the estimated detection probabilities $p_{it} = P(Y_{it} = 1 | Z_i = 1)$ for each county and each month in the study period. We see a lot of spatial heterogeneity in detection with strong temporal auto-correlation.

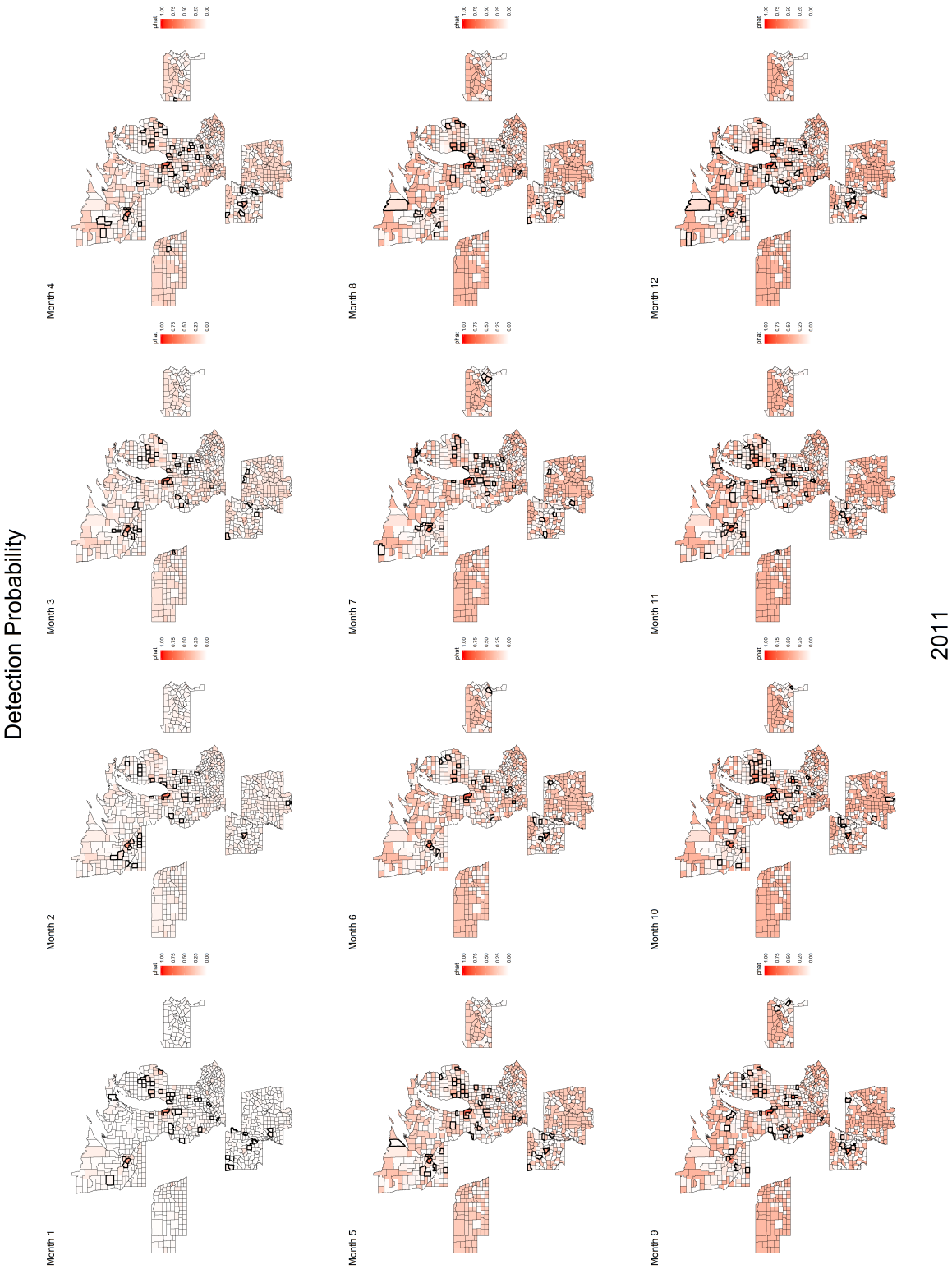

Figure 1: Estimated detection posterior probabilities for each county p_{it} for every month in 2011. The black outlined boxes are areas counties where at least one case of histoplasmosis was reported.

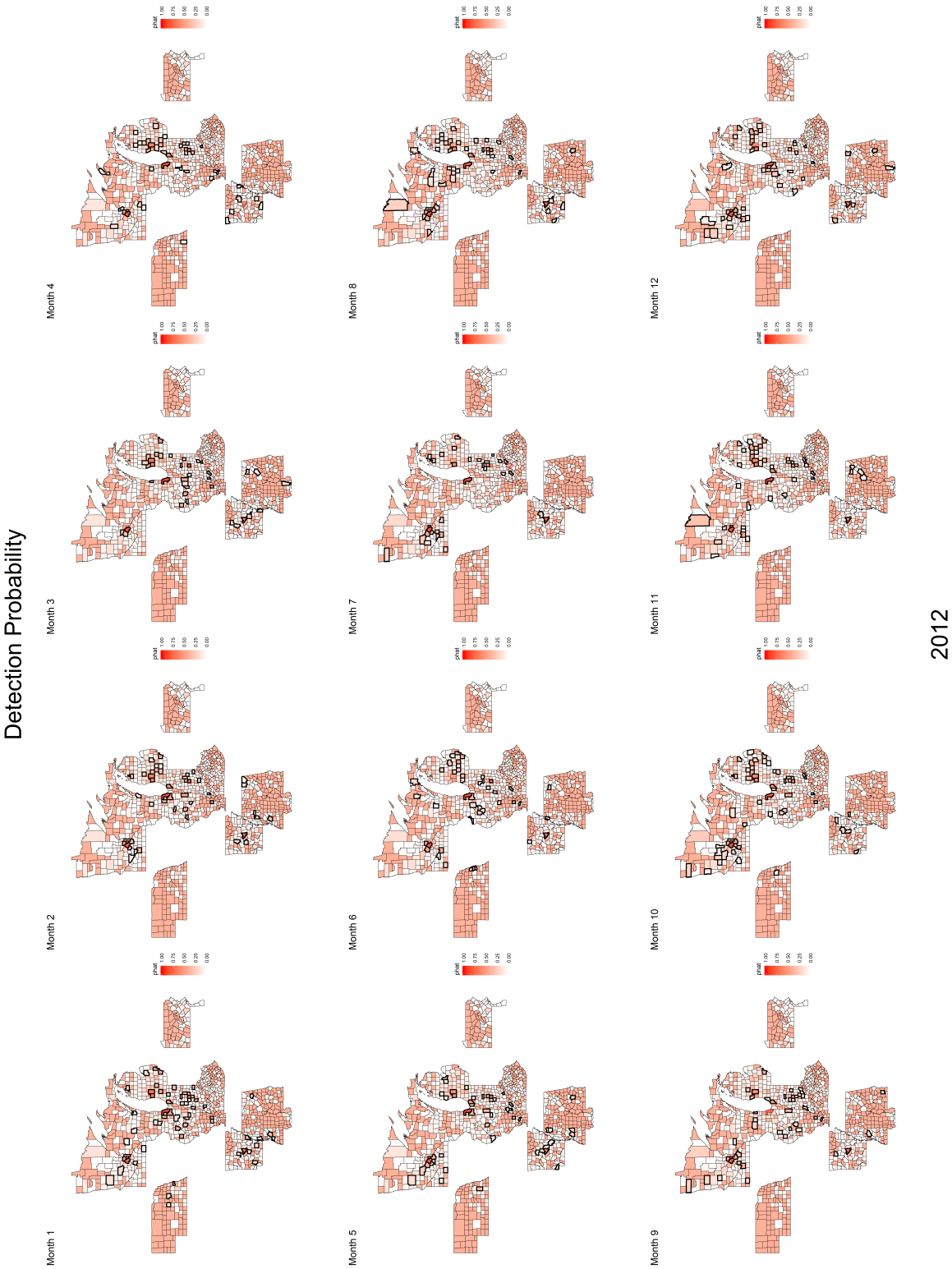

Figure 2: Estimated detection posterior probabilities for each county p_{it} for every month in 2012. The black outlined boxes are areas counties where at least one case of histoplasmosis was reported.

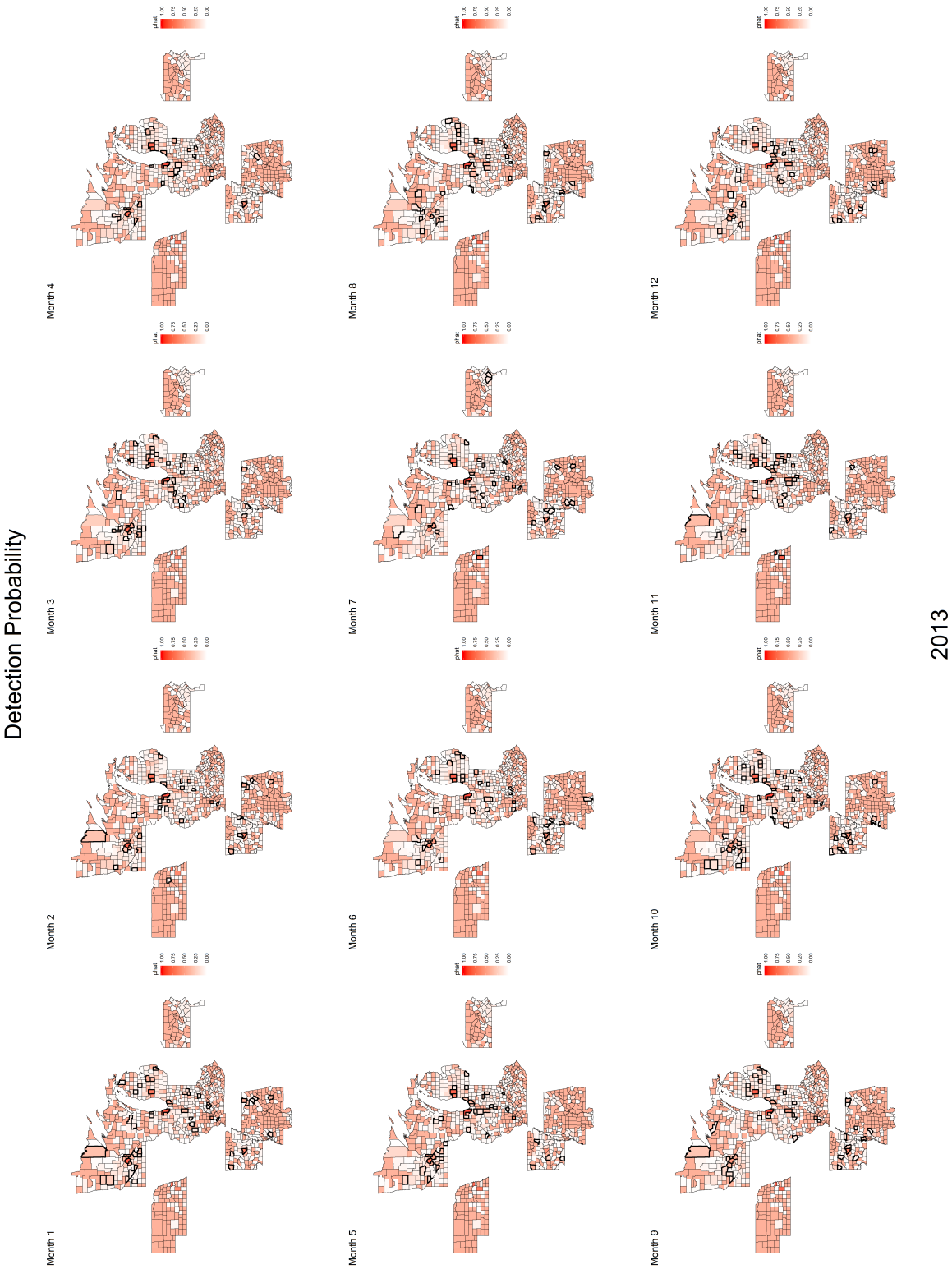

Figure 3: Estimated detection posterior probabilities for each county p_{it} for every month in 2013. The black outlined boxes are areas counties where at least one case of histoplasmosis was reported.

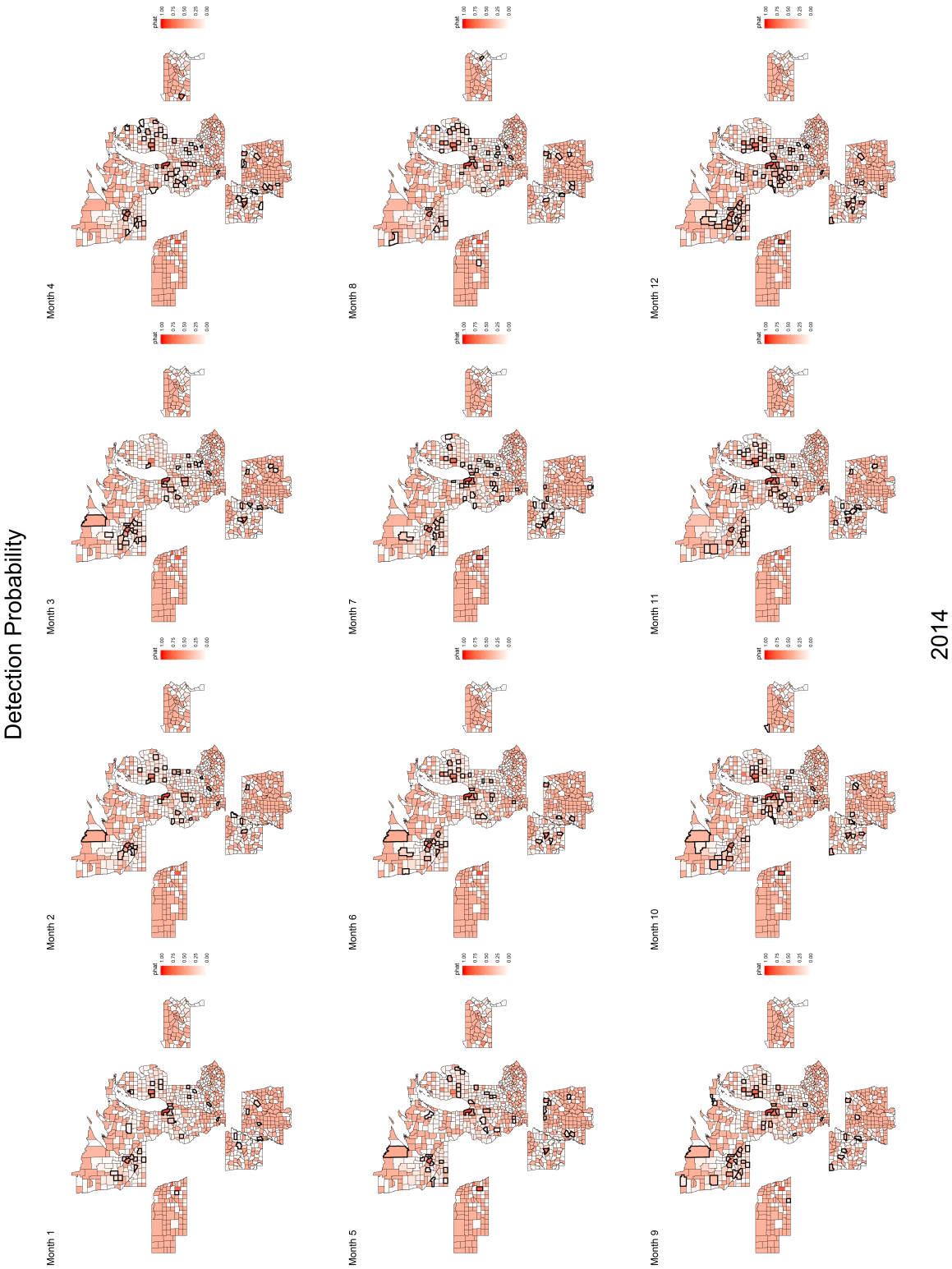

Figure 4: Estimated detection posterior probabilities for each county p_{it} for every month in 2014. The black outlined boxes are areas counties where at least one case of histoplasmosis was reported.

References

[1] J.H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 88(422):669–679, 1993.

[2] Jonathan R. Bradley, Scott H. Holan, and Christopher K. Wikle. Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. The Annals of Applied Statistics, 9(4):1761–1791, 2015.

[3] S.A. Hepler and R. Erhardt. A spatiotemporal model for multivariate occupancy data. Environmetrics, 32(2), 2021.

[4] John Hughes and Murali Haran. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(1):139–159, 2013.

[5] Devin S Johnson, Paul B Conn, Mevin B Hooten, Justina C Ray, and Bruce A Pond. Spatial occupancy models for large data sets. Ecology, 94(4):801–808, 2013.

[6] Brian J. Reich, James S. Hodges, and Vesna Zadnik. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics, 62(4):1197–1206, 2006.

[7] Christian Robert and George Casella. Monte Carlo statistical methods. Springer Science & Business Media, 2013.