PSTAT236 Project 1

Seonga Cho, Imani Russell, Sam Sambado

ABSTRACT

Global amphibian populations have declined at alarming rates during the past three decades. This rapid decline is attributed to land use change, the illegal pet trade, and the emergence of the fungal pathogen, *Batrachochytrium dendrobatidis* (Bd). However, Bd impacts frog species and individuals within the same population in interestingly different ways. This variation in the prevalence and intensity of Bd infection in amphibians has led researchers to explore the impact of environmental factors such as seasonal patterns of temperature and rainfall as well as the physical reservoirs (i.e. ponds) amphibians complete their life cycles in. We have analyzed several frog populations from the San Francisco Bay Area across various environmental characteristics to predict Bd infection loads at individual sites. We worked through a tutorial to apply univariate and co-kriging methods to better understand how environmental factors can be used to predict Bd loads across the San Francisco Bay Area.

INTRODUCTION

Outbreaks of emerging infectious diseases are becoming more problematic as climate changes, often driving declines in wildlife communities at unprecedented rates. One of the most lethal examples is the fungus, *Batrachochytrium dendrobatidis* (Bd), which causes chytridiomycosis and is responsible for the decline and extinction of hundreds of amphibian species worldwide (1). Due to its far-reaching effects, chytridiomycosis has been described as the "biggest loss of biodiversity attributable to a pathogen" (2). While some frog populations are driven to extinction by infection, other populations of the same species may coexist with infections, showing no signs of disease-driven decline (3). One example is the Mountain yellow-legged frog complex (MYLF), consisting of 2 closely-related species distributed throughout the Sierra Nevada. These species are generally considered highly susceptible to Bd - with several populations experiencing epizootic disease dynamics (disease-driven declines), while other populations experience enzootic dynamics, resisting decline even with Bd infections.

Bd was first described in the late 1990s (4). In amphibians, Bd infects keratinized tissues (the mouthparts of tadpoles and the skin of subadults and adults). When the zoospore encounters amphibian skin, it encysts and creates a sporangium in which zoospores are produced asexually. After 2-7 days, the sporangium releases zoospores to the outer surface of the skin (5), which can either reinfect the same host or enter the surrounding environment and infect new hosts.

In order for a pathogen to persist in populations, the pathogen (e.g. Bd), the host (e.g. amphibian), and the environment (e.g. ponds) need to be synchronized in a particular way for there to be disease transmission. As a non-directly transmitted pathogen, Bd zoospores can live in the environment without a host. This aspect of Bd, and of non-directly transmitted pathogens, are one of the many mysteries surrounding the success of these pathogens. Currently, it is still unknown how long Bd zoospores can persist in the environment without a host, and where in the environment they are persisting. This provides an unique opportunity to expand disease investigation beyond the host and to explore the role of environmental reservoirs such as ponds which are subjugated to dynamical, seasonal influxes of rainfall and temperature differences in Northern California.

Our report focuses on various environmental factors that may predict host infection loads at sampling sites. This serves as a summary of our experience following the Co-Kriging in R technical note (by Rossiter, 2018) and our learning experience applying this approach to real, (often messy) ecological data.

THE DATA SET

The data used was collected by Tatum Katz, a PhD student in the Ecology, Evolution and Marine Biology Department. Sampling occurred at the same pond sites in January, March, July, and August of 2020 when COVID restrictions allowed researchers to go into the field. Data collected includes 546 individual amphibians at 22 pond sites across 4 regional parks (Blue Oak Ranch Reserve, Pleasanton Ridge Regional Park, Garin/Dry Creek Pioneer Regional Park in Alameda County, and Briones Regional Park in Contra Costa County).The variables of interest are listed below:

- infection load (qpcr_qty) ← **target variable**: Infection load is measured for each individual amphibian via skin swab.
- sample site (code): the pond where measurements are taken
- (property): the entity to which the site belongs to
- water temperature (water_temperature_c): temperature of the pond
- (ph): pH of the pond
- conductivity (conductivity_uS): conductivity of the pond
- dissolved solids (total_dissolved_solids_ppm): amount of dissolved solids in the pond
- salinity (salinity_ppt): salinity of the pond

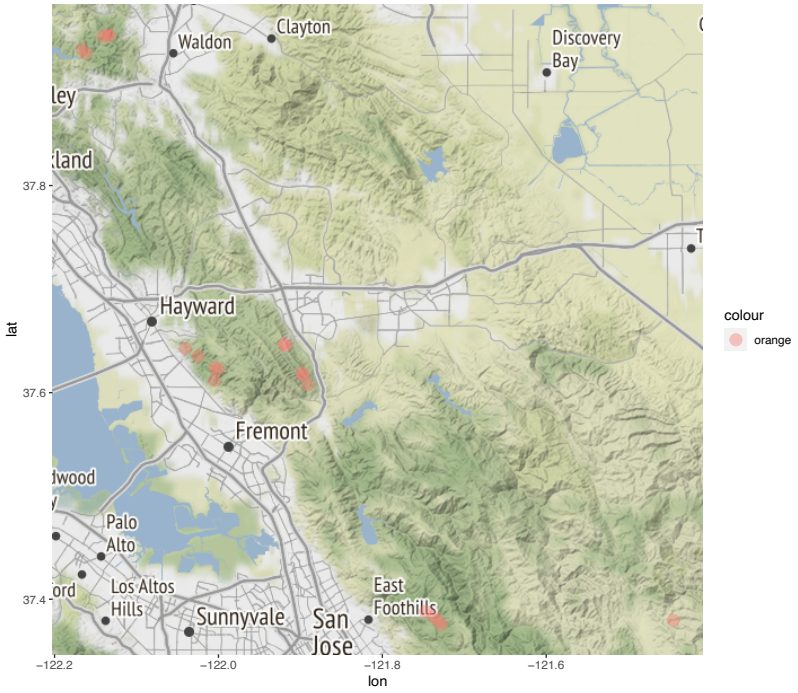


Figure 1: Map of sample sites in the San Francisco Bay area.

METHODS

We had originally planned to analyze amphibian infection and community assemblage across the UC Natural Reserve System (UC NRS) which spans thousands of miles across California's latitudinal gradient that Imani has gathered for her PhD work. However, after starting to work with the data, we found that the sampling area was too large and discontinuous to create an interpolation grid. Instead, we decided to use the other dataset. These data were collected at several pond sites across the San Francisco Bay area, which allowed us to create an interpolation grid.

Each amphibian was given a unique ID, identified to species, measured snout to vent (svl), weighed, sexed, aged and swabbed for pathogen detection. The pathogen detection swabs were brought back to the laboratory and tested for Bd with a quantitative polymerase chain reaction (qPCR). After qPCR, samples were labeled as

	Winter (N=127)	Spring (N=82)	Summer (N=148)	Early Fall (N=187)	Total (N=544)	p value
zoospore						0.002
N-Miss	64	4	2	111	181	
Mean (SD)	11.725 (73.623)	1837.330 (7614.183)	33.153 (143.276)	237.859 (1007.121)	459.967 (3616.453)	
Range	0.000 - 582.273	0.000 - 47266.320	0.000 - 1034.752	0.000 - 8188.000	0.000 - 47266.320	

(a)

	Blue Oak (N=179)	Briones (N=80)	Garin/Dry Creek (N=98)	Pleasanton Ridge (N=187)	Total (N=544)	p value
zoospore						0.117
N-Miss	35	23	19	104	181	
Mean (SD)	592.101 (4043.041)	1345.189 (6431.732)	16.906 (112.640)	44.507 (172.156)	459.967 (3616.453)	
Range	0.000 - 47266.320	0.000 - 47266.320	0.000 - 998.928	0.000 - 1034.752	0.000 - 47266.320	
amphibian_svl						< 0.001
Mean (SD)	34.765 (22.892)	66.000 (23.073)	29.990 (26.813)	30.102 (30.424)	36.895 (29.078)	
Range	1.000 - 106.000	16.000 - 97.000	1.000 - 107.000	1.000 - 105.000	1.000 - 107.000	
amphibian_weight						< 0.001
Mean (SD)	39.547 (30.167)	46.550 (19.748)	28.898 (27.021)	25.396 (31.283)	33.794 (29.735)	
Range	1.000 - 95.000	8.000 - 84.000	1.000 - 100.000	1.000 - 99.000	1.000 - 100.000	

(b)

Figure 2: Summary table of (a) zoospore counts by season, and (b) zoospores and amphibian traits by property.

positive/negative for Bd and positive samples were quantified based on infection load (the amount of zoospores). However, not all samples have been processed in the lab yet so some individuals have NAs for infection load. In addition to the amphibian data, environmental data was collected at each monthly visit. The environmental data of the pond water that we wanted to analyze for this project includes: pH, temperature, conductivity (uS), and salinity (ppt).

Our target variable is amphibian infection load (quantified by the qPCR) and our co-variables of interest are amphibian weight and svl as well as pond water pH, temperature, conductivity, and salinity.

We chose these co-variables because we know that amphibians complete their life cycle in ponds and ponds serve as reservoirs for fungal pathogens, like Bd. However ponds are not static environments and can change in quantity (amount of water available in pond) and quality (pH, salinity, turbidity, etc.) throughout the year based on seasonal rainfall and temperature. Seasonal rainfall and temperature can also vary from site to site within regional parks due to topographic features and pond usage of other vertebrate species (cows, birds, humans, etc.). We believe our co-variables are part of important processes to predict infection loads of Bd in amphibians based on previous work (6).

Data cleanup

We received 4 separate data files corresponding to each season. For ease of analysis, we merged the data files into 1 master data file. The winter data file was missing infection loads for individual animals, which were found in Imani's dataset, so we matched the correct infection loads with the individual amphibians.

Before merging the data files together, we calculated the average infection loads for each site during each season, and appended those quantities to the environmental data. At this point, we were able to combine the 4 data files into one master file. Average infection loads quantified by qPCR were multiplied by 80 to account for dilution of the DNA during laboratory analyses and represent standard infection load units (zoospore quivalents, or ZE) based on previous literature estimates.

Follow-along with technical note: Co-Kriging in R (Rossiter, 2018

Disease data is often highly right-skewed (with a large proportion of 0's). Because of this, we took a commonly used approach of separating the data into positive and negative infection loads, and analyzing just the positive infection load data. The other option of this approach is to then also analyze both positive and negative infections together but as binary outcomes (infected vs not infected). For this analysis, we used only the positive infection load data approach, not the binary approach, which limited the amount of individuals we could analyze to 44 individuals.

Interpolation grid

At this point, we made created an interpolation grid in ArcGIS 10.7 and QGIS 3.6 to begin the spatial analysis. We created 100m*100m grids in the research area. After creating grids, the research area was covered by 414,090 grids. There are two reasons for defining the grid size as 100m. 100m is a relatively narrow and homogeneous research unit, when considering the frog's migration pattern. The second reason of defining 100m for the spatial unit is for minimizing the calculation load. Because of the huge research area, the calculation burden is immerse for this case. So, This research applied 100m grid for the analysis. After applying the auto generated best semivariogram model, The univariate kriging result was shown like Fig. 12. Due to the lack of observation points, it is hard to check the significant results in kriging prediction. Therefore, using more variables would improve the prediction performance. So, this research will apply multiple co-variables by co-kriging approach.

Target variable

We visualized the distribution of the target variable data (infection load) both untransformed and log(10) transformed (Figure 8). The untransformed data is highly right-skewed, and we see the log(10) transformed data is more normally distributed. Therefore, we will work with the log(10) transformed data for our analysis.

We also looked at the proportion of samples that are higher than various levels/thresholds. The only meaningful threshold above 1 ZE (which denotes a positive

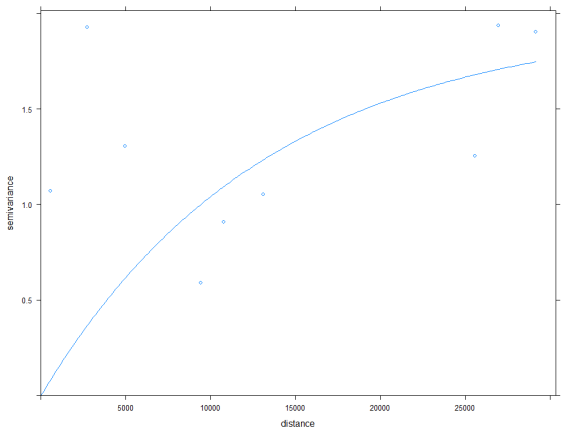


Figure 3: Exponential Semivariogram Model

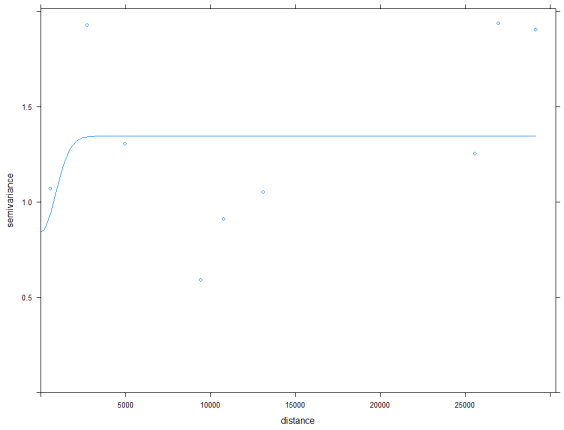


Figure 4: Gaussian Semivariogram Model

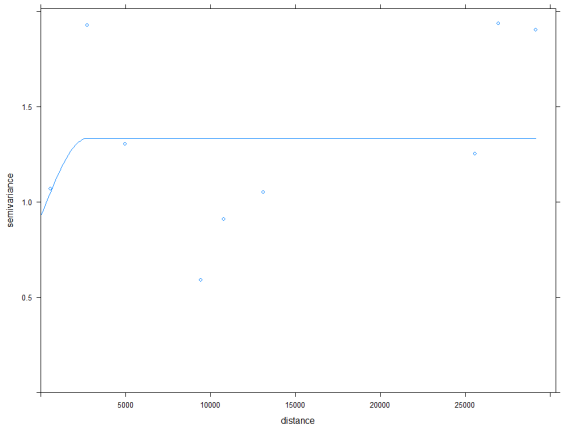


Figure 5: Spherical Semivariogram Model

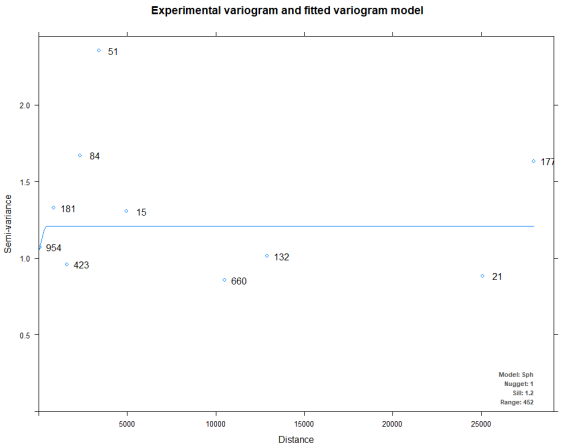


Figure 6: Auto generated best Semivariogram Model

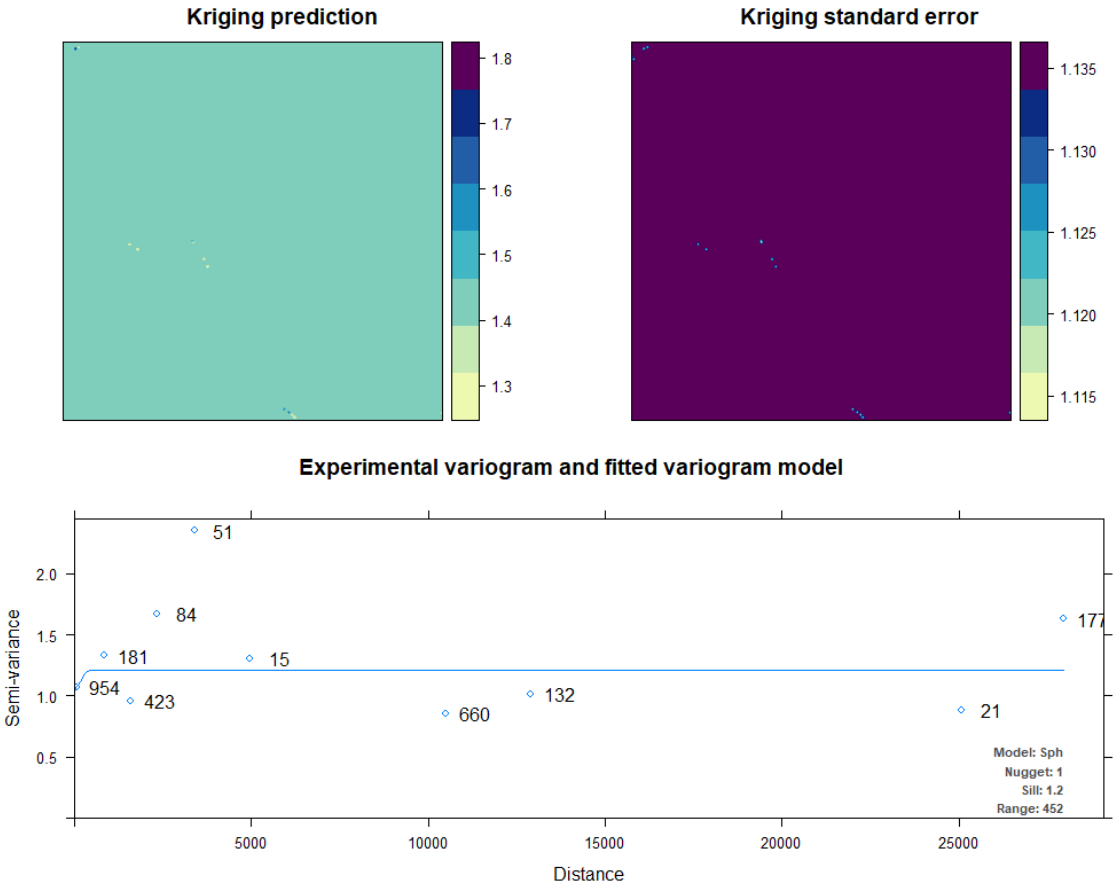


Figure 7: Auto generated best Semivariogram Model

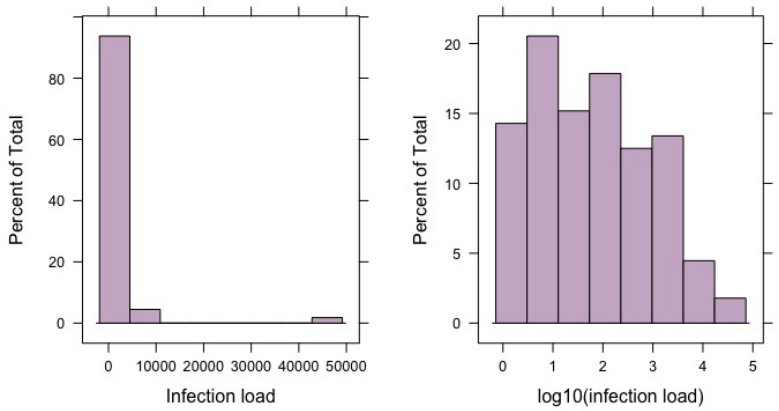


Figure 8: Untransformed vs log(10)-transformed infection loads.

infection) is 10,000, which has been identified as the "lethal" threshold of infection for host, generally (3). By doing this, we can see that 42% of our samples are above 100 ZE, 18.8% of samples are above 1,000 ZE, 1.8% of samples are above 10,000 ZE. No hosts were sampled with infection loads above 100,000 ZE.

Co-variables

We also chose covariables based on theory - what we hypothesized could affect infection loads across populations (this list can be found in the **DATA SET** section). We also examined the distributions of the covariables and their log(10) transformations for normality (Figure 9 and 10).

We found that the distribution of almost all variables (SVL, mass, pH, water conductivity, total dissolved solids, and salinity) became more normal after log(10) transformation. However, water temperature distribution only slightly changed, and pH remained left-skewed (even after also square root and cube root transformations). Because of this, we decided to proceed with the analysis using the log(10) transformations of all variables.

We plotted the autogenerated best variogram to estimate the fit (Figure 6) This showed that there is not really any spatial correlation within the infection load data. We plotted variograms for both untransformed and log(10)-transformed infection loads (using the spherical, exponential, and gaussian models, and see that none reveal spatial correlation.

Ordinary kriging: Interpolation of infection loads

We start off our analysis with univariate kriging. Because we can use kriging to predict target variable values where only co-variables were measured, Rossiter creates an

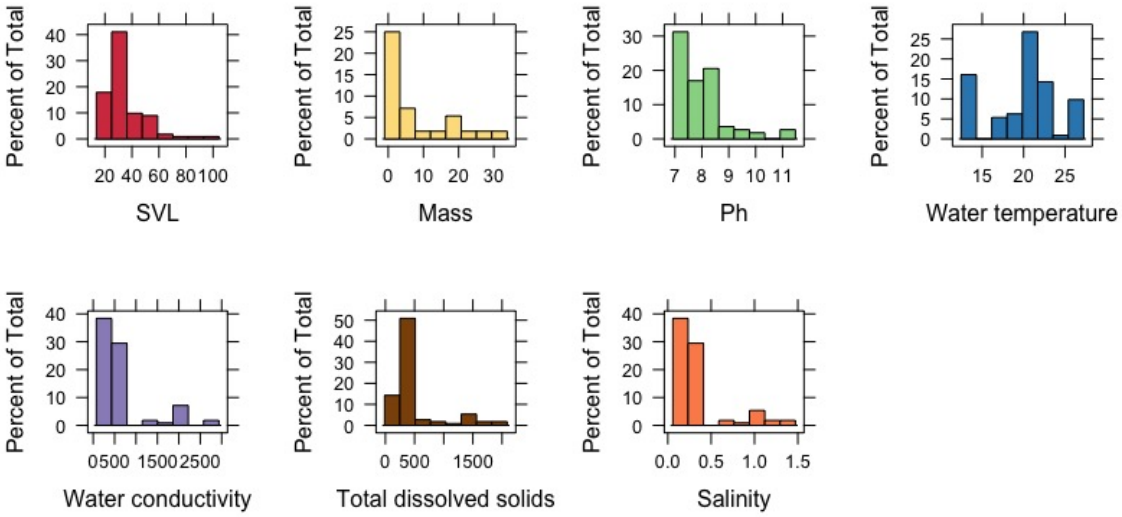


Figure 9: Untransformed covariable values.

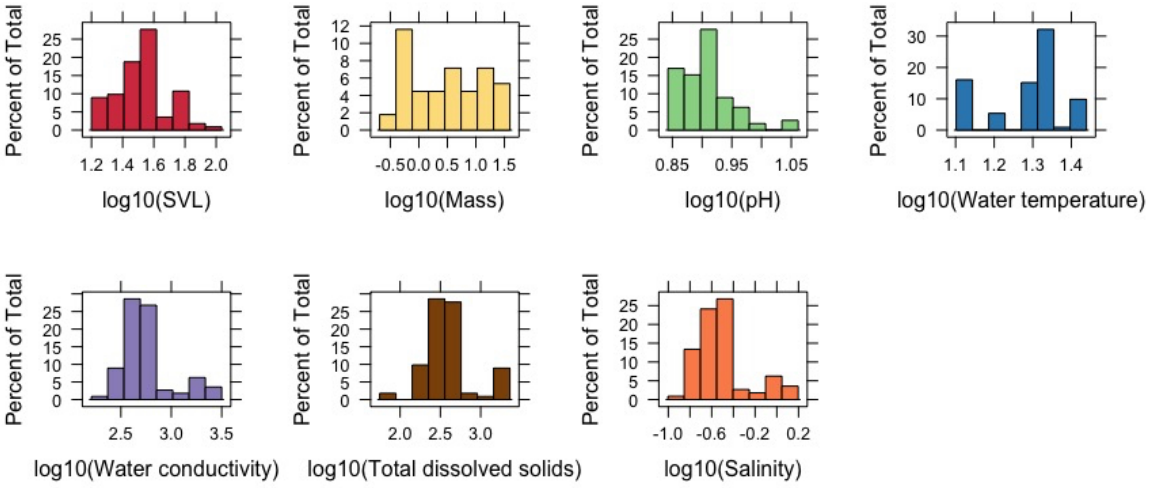


Figure 10: log(10)-transformed covariable values.

exercise to simulate under-sampling of our target variable - infection load. However, because we had large amounts of missing data for all variables (not just the target) and a small sample size, we decided to forego this exercise.

Co-kriging: Modelling a bivariate co-regionalisation

To begin our co-kriging analysis, we used co-regionalization to model the spatial structure of (a) svl, and (b) weight, (b) pH, (c) water temperature, (d) conductivity, (e) total dissolved, and (f) salinity and its covariance with infection load (by using

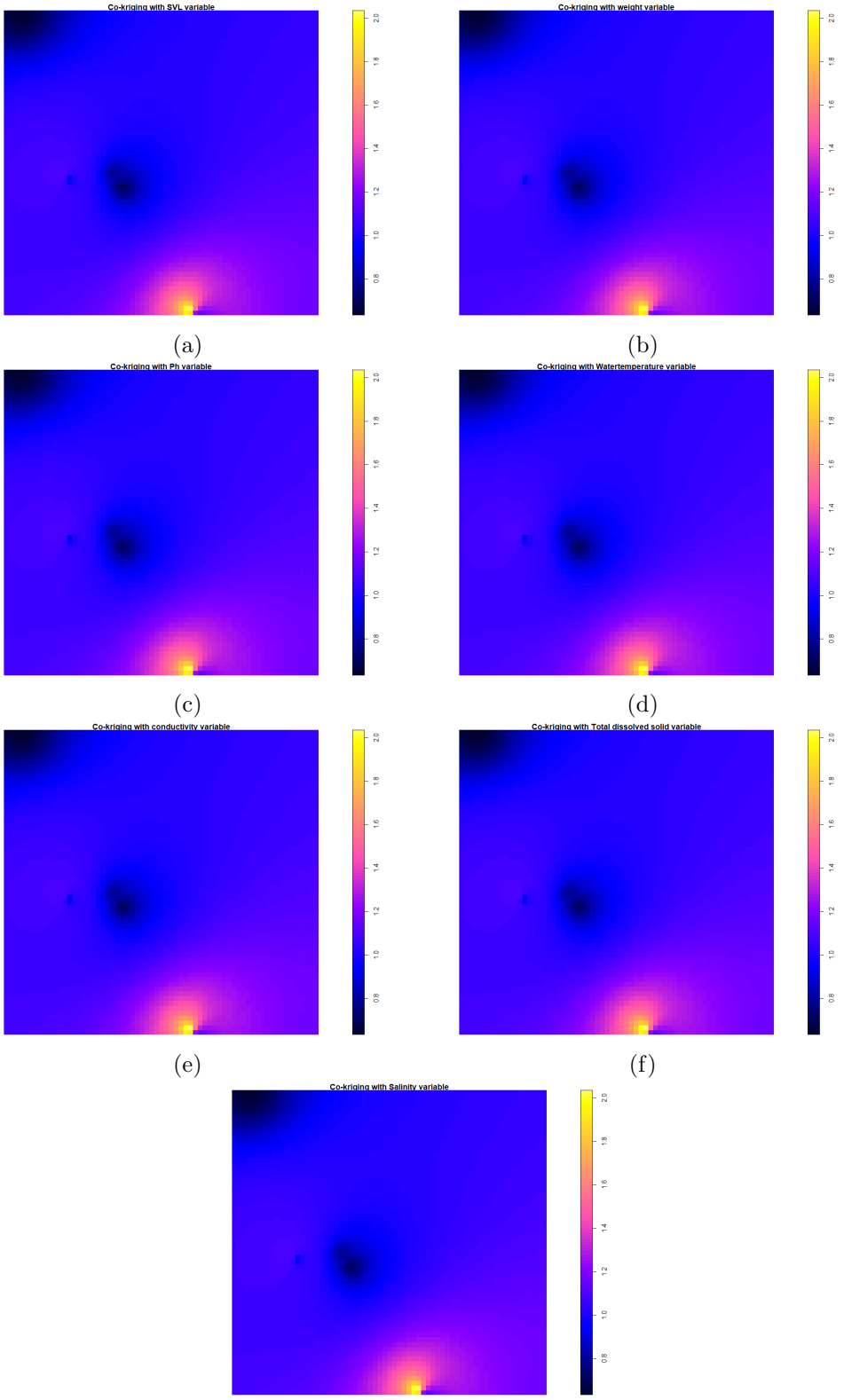


Figure 11: Co-variable maps (a) svl, (b) weight, (c) pH, (d) water temperature, (e) conductivity, and (f) salinity.

this model to interpolate).

Modelling the co-variable: Simple and Universal kriging

To begin our co-kriging analysis, we used co-regionalization to model the spatial structure of (a) svl, and (b) weight, (b) pH, (c) water temperature, (d) conductivity, (e) total dissolved solids, and (f) salinity and their covariance with infection load (by using this model to interpolate). We began by creating plots of infection loads vs the different co-variables (Figure 16).

We can see that there is a negative relationship between infection load and water temperature, pH, salinity, svl, and weight. We find a positive relationship between infection load and total dissolved solids.

Since our relationship was strong enough for co-kriging, we modelled the co-regionalization of the variables by fitting models to direct and cross-variograms at the same time.

We started by modelling 7 variables together ((a) svl, and (b) weight, (b) pH, (c) water temperature, (d) conductivity, (e) total dissolved, and (f) salinity). Can also try just 2 as Rossiter did.

Rossiter mentions in his text that for multivariate kriging, you need to work with objects of class gstat, so we built these objects from the variograms we created previously. After creating the gstat object, we were able to create the variograms. In variograms, we applied 452m for range, nugget value 1, and sill value 1.2 for spherical model. We found this value by automated optimized in R. For comparison, different parameters were applied for other experimental variograms.

We then fit a linear model of co-regionalization to the gstat object by adding variogram models to fit direct and cross variograms for all variables (Figure 18).

Comparison of co-kriging and ordinary kriging

After using the two methods of kriging to predict infection loads across our sample sites, we compare the efficacy of the two methods.

We first compared the differences in predictions as created maps to display these difference. Due to the concentrated observations, all of the kriging maps show the high prediction value on the bottom side. And other regions show similar prediction values among the research area.

DISCUSSION AND CONCLUSIONS

Infectious diseases are part of dynamic processes that fluctuate with environmental inputs (e.g. temperature, rainfall) and host traits (e.g. species, life stage). However,

teasing apart the most important influences on disease transmission can be difficult and often limited in scope.

Often disease mitigation efforts will focus on total pathogen prevalence by region (e.g. property) (Figure 15) without considering the variation that can be attributed within a small geographical space (e.g. site code). Within our dataset we found variation across seasons (Figure 2a)(Figure 12a) and individual ponds (Figure 13), which we hypothesize is attributed to the quality of the environmental reservoir (e.g. pond).

When we run basic statistic models, such as generalised linear models (log(zoospore + 1) $\sim$ the different co-variables), we can see that variables such as amphibian svl (Figure 16e), and pond variables such as temperature (Figure 16a), salinity (Figure 16c), pH (Figure 16b), have negative impacts while dissolved solids (Figure 16d) has a positive impact on the log count of zoospores.

Due to the inherent messy nature of ecological data that can be incomplete and often skewed, we took the multivariate approach, kriging, to predict missing target variables (i.e. infection loads) where co-variables (i.e. amphibian traits and pond quality) were measured more consistently. Through working through co-kriging methods, we discovered many of the difficulties of working with ecological disease data, which is nearly always extremely non-normal. This project allowed those of us working with disease data to better understand the requirements and first steps for kriging analyses, and and to pursue more appropriate approaches for working with non-Gaussian data in spatial analyses.

AUTHOR CONTRIBUTIONS

Seonga Cho created and coded interpolation grid and kriging analyses. Imani Russell cleaned data, coded initial attempt at analyses, and wrote report. Sam Sambado cleaned data, coded summary statistics, and wrote report.

REFERENCES

[1] Skerratt, L. F. *et al.* Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth* **4**, 125–134 (2007).

[2] Scheele, B. C. *et al.* Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science* **363**, 1459–1463 (2019).

[3] Briggs, C. J., Knapp, R. A. & Vredenburg, V. T. Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. *Proceedings of the National Academy of Sciences* **107**, 9695–9700 (2010).

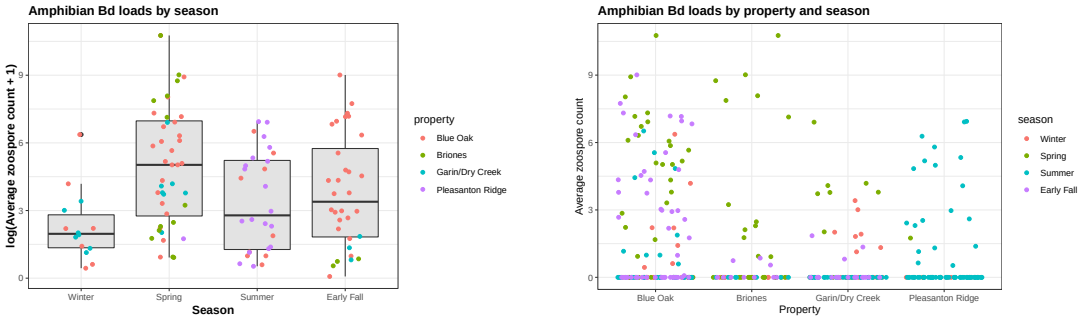
[4] Longcore, J. E., Pessier, A. P. & Nichols, D. K. Batrachochytrium dendrobatidis

gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia* **91**, 219–227 (1999).

[5] Rosenblum, E. B., Voyles, J., Poorten, T. J. & Stajich, J. E. The deadly chytrid fungus: a story of an emerging pathogen. *PLoS Pathog* **6**, e1000550 (2010).

[6] Wilber, M. Q., Knapp, R. A., Toothman, M. & Briggs, C. J. Resistance, tolerance and environmental transmission dynamics determine host extinction risk in a load-dependent amphibian disease. *Ecology Letters* **20**, 1169–1181 (2017).

SUPPLEMENTAL FIGURES



(a) (b)

Figure 12: Amphibian Bd loads (log transformed zoospores) by property and season.

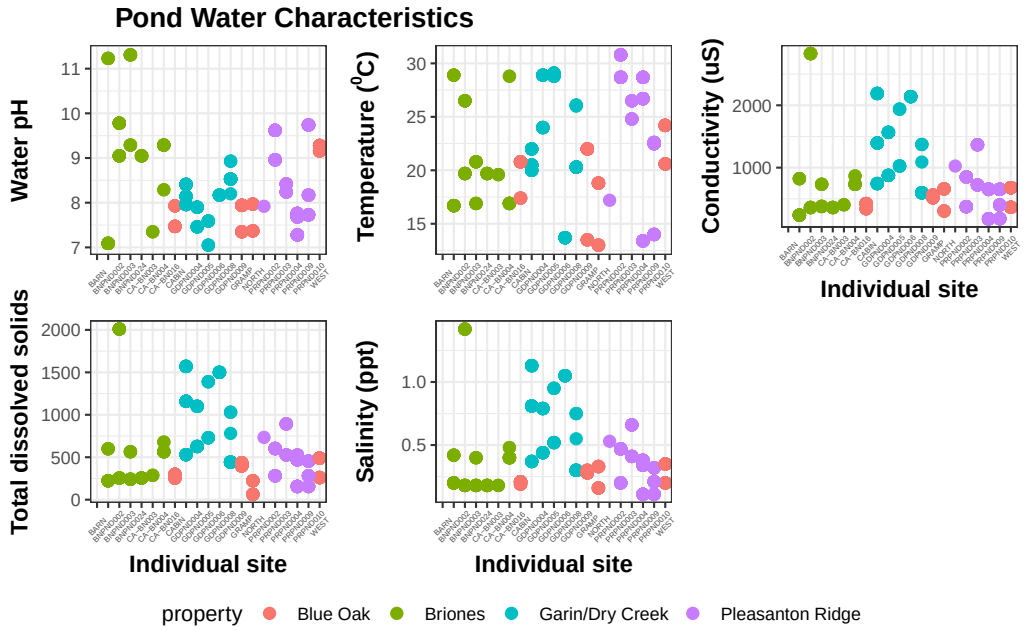


Figure 13: Environmental characteristics of individual ponds by property type.

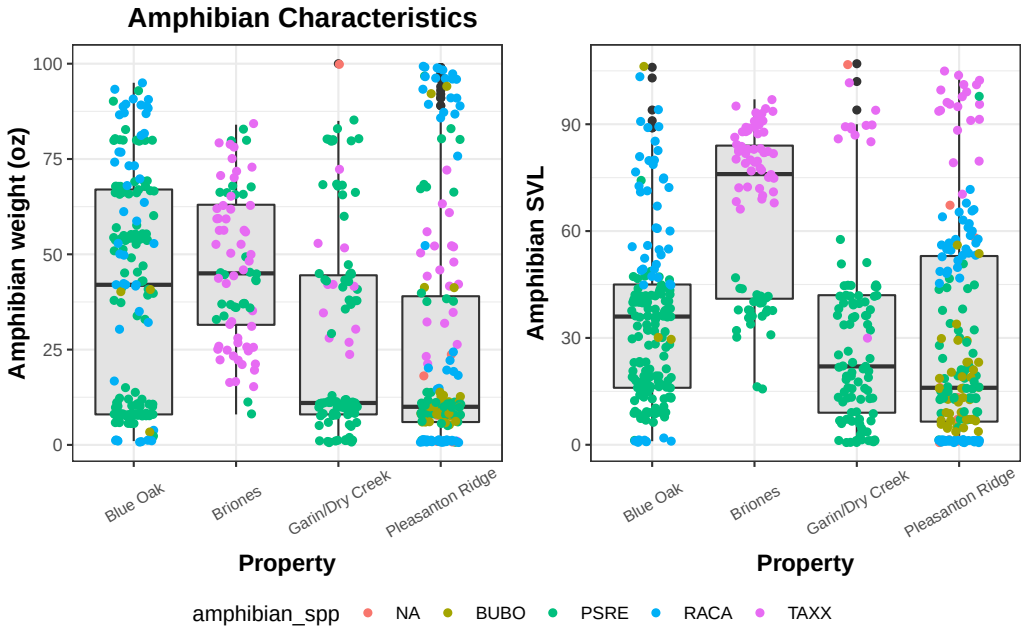


Figure 14: Amphibian characteristics (weight and SVL) by property type.

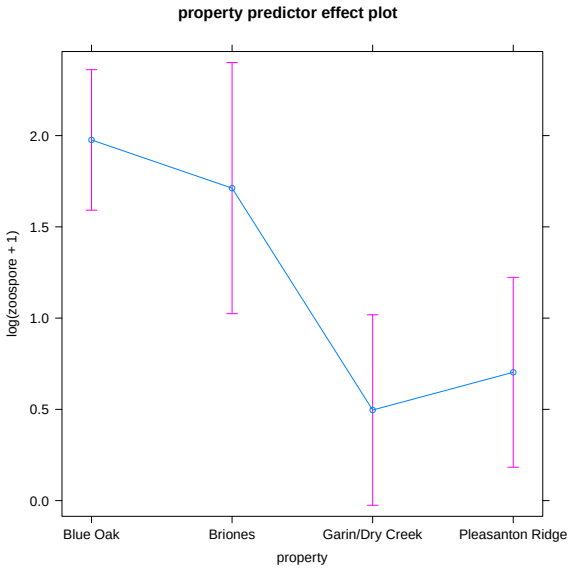


Figure 15: Property as a predictor effect for zoospore count.

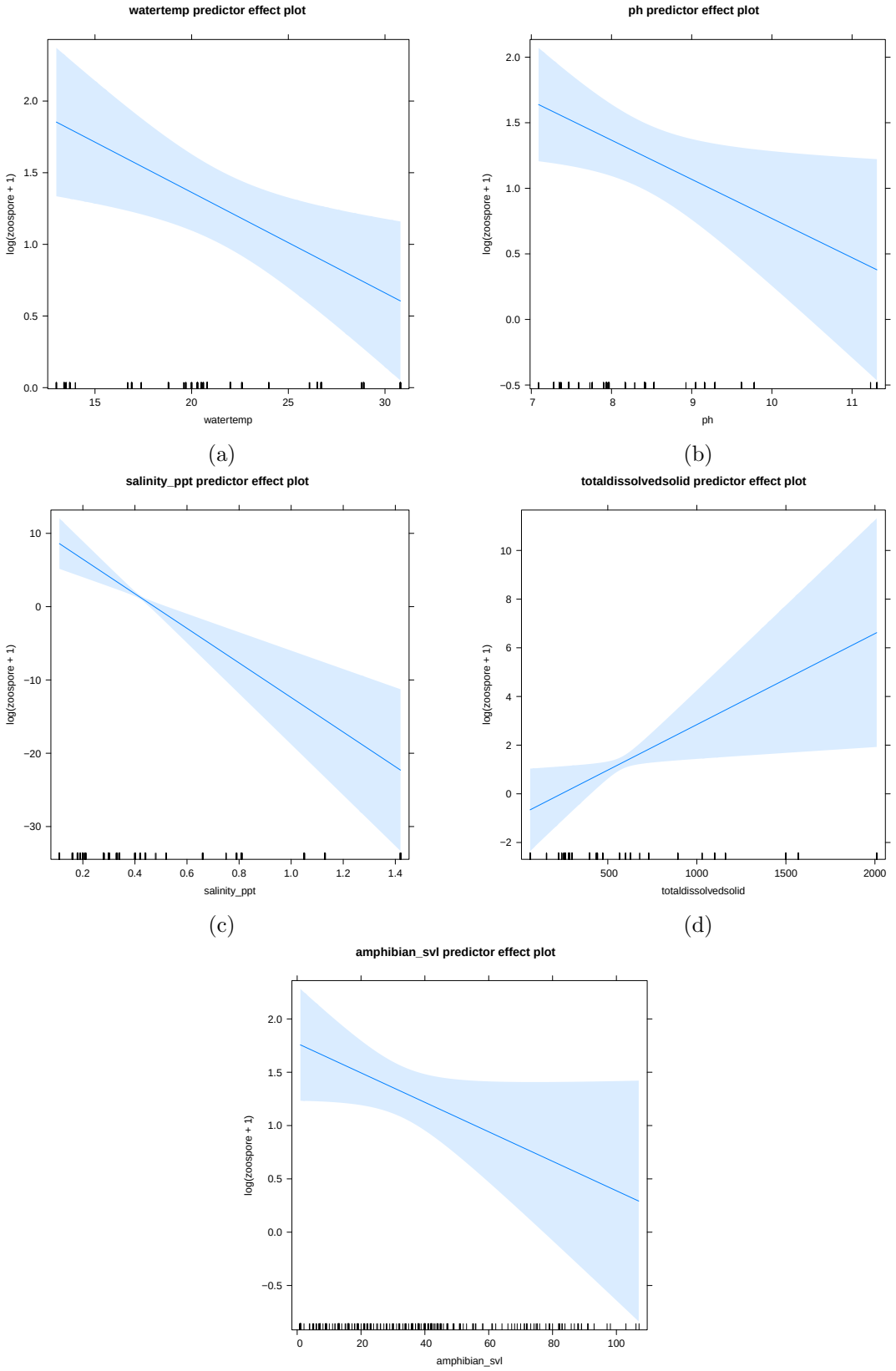
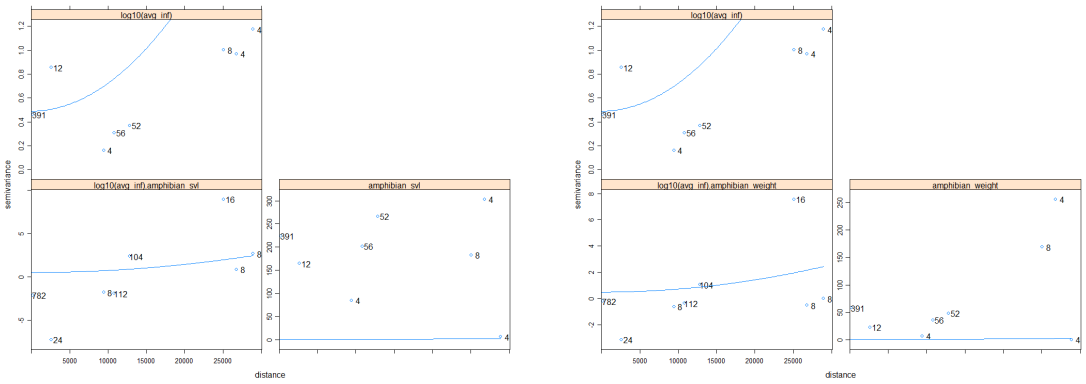
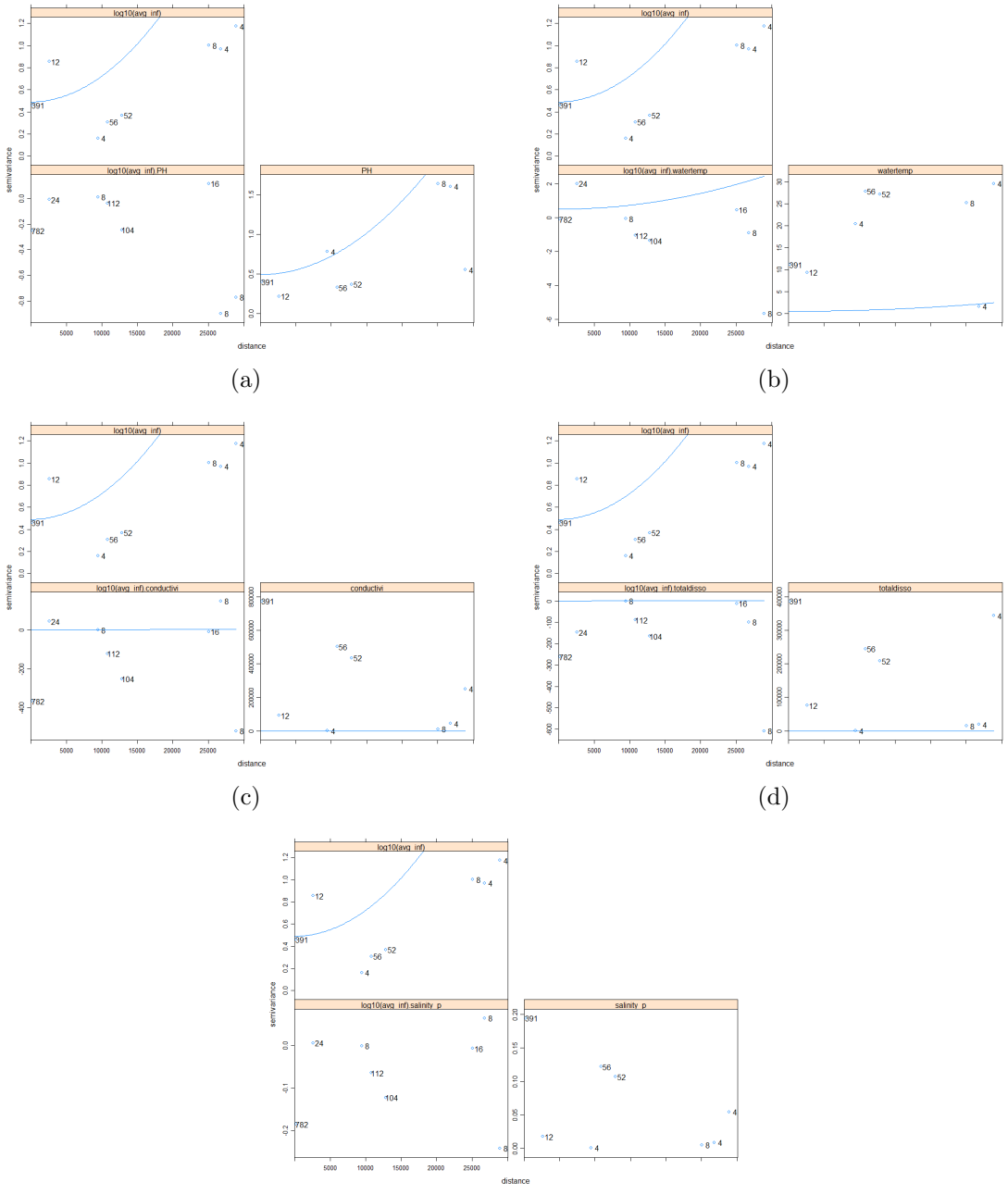


Figure 16: Water and amphibian characteristics as predictors for zoospore counts in amphibians.



(a) (b)

Figure 17: Amphibian variable semivariograms (a) svl, and (b) weight.



(a) (b) (c) (d) (e)

Figure 18: Pond water variable semivariograms (a) pH, (b) water temperature, (c) conductivity, (d) total dissolved, and (e) salinity.