BAYESIAN DATA MINING TECHNIQUES IN PUBLIC HEALTH AND BIOMEDICAL APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Seonghye Jeon

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2012

BAYESIAN DATA MINING TECHNIQUES IN PUBLIC HEALTH AND BIOMEDICAL APPLICATIONS

Approved by:

Professor Brani Vidakovic, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Youngja Park
Department of Medicine
Emory University School of Medicine

Professor Xiaoming Huo
School of Industrial Systems and Engineering
Georgia Institute of Technology

Professor Julie Swann
School of Industrial Systems Engineering
Georgia Institute of Technology

Professor Yajun Mei
School of Industrial Systems and Engineering
Georgia Institute of Technology

Date Approved: 26 March 2012

To my family, in Korea and in Atlanta.

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor Dr. Brani Vidakovic for his guidance and support in countless ways. Without his enlightening instruction, great kindness and patience, unwavering enthusiasm and trust, I could not have completed my thesis. I also want to extend my gratitude to my thesis committee members, Dr. Yajun Mei, Dr. Youngja Park, Dr. Xiaoming Huo and Dr. Julie Swann for all their careful reviewing and insightful comments. Their detailed reviews and constructive comments greatly improved this thesis.

Many thanks to my CDC-ORISE fellowship supervisors, Laurie Barker and Susan Griffin. They have provided me a wonderful environment and full support to pursue my research. I could not have imagined having better supervisors. Thank you also to my fellow graduate students, especially Hyojung Kang and Chia-Jung Chang. You all have made my life in Atlanta so much more meaningful.

Finally, and most importantly, I would like to thank my family. I wouldn't have made it this far without their constant prayer, support, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis is focused on Bayesian data mining techniques for *non-traditional data* in the health care and biomedical fields, and comprises three main topics.

Meta-analysis of clinical trials with incomplete data set

This methodology is motivated by meta-analysis that compares the effectiveness of glass-ionomer (GI) to resin-based (RB) sealants in preventing caries in permanent molars of children. While GI and RB sealants are two of the most commonly used dental sealant materials, there is strong debate regarding their comparative effectiveness. Although GI to RB sealant meta-analyses have been performed in the past, this research is the first to find difference by including studies that were excluded previously solely because they failed to report data in a paired manner. The conclusion is based on 12 clinical studies performed on 2,012 children, that had incomplete information and inconsistent conclusions.

To retrieve full paired table information from studies that only reported marginal totals, we propose a Bayesian approach based on Metropolis random walk and a hierarchical Bayesian model. The procedure relies on Markov chain Monte Carlo simulation, recovers a full table and reports its accuracy in terms of credible sets for the cell counts. The performance of the model is remarkable; from 17 matched-pair study examples, four of the estimated tables coincided exactly with the original tables and nine tables deviated only (+1,-1) in terms of cell counts.

Analysis of spatial repeated measures with excess zero and no covariate information

Water fluoridation, the addition of fluoride to drinking water to prevent tooth decay, is listed as one of ten greatest public health achievements of the 20th century by the Centers for Disease Control and Prevention. However, fluoride is present naturally in water in concentrations well above recommended levels, which may have several long-term adverse effects. Therefore, describing the availability of natural fluoride by county can inform public health efforts in prevention of tooth decay and dental fluorosis.

This work is to estimate United States county level natural fluoride occurrence in the ground water. To accommodate the limitations of the data, we propose a two-step hierarchical Bayesian approach. It models spatial repeated measures with excess zeroes and it is robust even when covariate information is not available. The approach starts with: (i) an autologistic model that estimates the probability of zero observations, and proceeds to (ii) an excess-zero model that describes the underlying physical phenomena of the zero-inflated data. The proposed hierarchical Bayesian structure produces improved discretized estimates and reflects the continuous behavior of the natural fluoride occurrence better. By employing the hyper prior structure, state level estimates can be obtained as well. Furthermore, the model can answer pragmatic questions such as "the probability that a specific county has naturally occurring fluoride above EPA recommended level".

Image classification based on the overall image regularity in wavelet domain

Breast cancer is one of the most common forms of cancer among women in the United States; an estimated 1 in 8 women born today will be diagnosed during her lifetime. Since the causes of breast cancer are not yet fully understood, early detection

is still the best strategy for improving prognoses. Mammography is currently the most effective method for detecting breast cancer early; however, radiological interpretation of mammogram images is a challenging task. The appearance of even normal tissue is highly variable and complex, and signs of early disease are often indistinct.

We propose a diagnostic based on the properties of overall image backgrounds; this procedure currently is an unused diagnostic modality in mammograms. Generally speaking, normal/healthy images tend to be more irregular than cancer images. The overall regularity of the image is assessed through wavelet analysis, which is then summarized by a few measures. These estimators are evaluated based on their ability to classify digitized mammogram images from a clinical database, for which the true disease status is known by biopsy. This research presents a two-fold approach: (i) generalization of the covariance wavelet spectra to the complex domain and (ii) the estimation of Hurst parameter and phase information as discriminatory descriptors. The most accurate classification rates from this work achieve 86%; these rates vary slightly with the choice of wavelet basis, levels used and size of training set.

CHAPTER I

INTRODUCTION

The emerging research issues in evidence-based healthcare decision-making and explosion of comparative effectiveness research (CER) are evident proof of the effort to thoroughly incorporate the rich data currently available within the system. However, in most cases, the data involves some challenging issues such as the required information is not fully reported due to the lack of statistical knowledge (incomplete or missing data issue), data itself exhibits non-traditional patterns (change of support, zero inflation) or it is highly correlated and irregular. Therefore a traditional statistical approach is not directly applicable.

The flexibility of Bayesian data mining techniques lends its strength to handle the challenging issues in the biomedical and health care domains. The well-known advantages of the Bayesian approach include no requirement of closed form analytical solution, the elicitation of prior beliefs, good finite sample performance and the ability to incorporate unknown variability. Also, the Bayesian approach integrates different data sources from observational data to controlled experiments, and answers pragmatic questions like that are usually of immediate interest of health practitioners.

This thesis comprises three distinct topics on Bayesian data mining techniques for *non-traditional data* in the health care and biomedical fields. In the following sections, we provide the background information on the statistical approaches used in this thesis.

1.1 Meta-analysis

Meta-analysis involves combining the results of several studies that address related hypotheses. The goal of the meta-analysis is obtaining a reasonable summary measure so that one can draw a conclusion from the independent and inconsistent individual studies. It is often an important component of a systematic review procedure. One of the most popular usages of meta-analysis can be found in combining several clinical trials of a medical treatment, in an effort to obtain a better understanding of how well the treatment works.

The procedure begins with formulating the problem, which implies identifying the right question. Then a set of appropriate studies is identified through literature review and systematic search and included into the meta-analysis based on selection criteria. Once a set of studies is obtained, then it requires statistical approaches to answer two questions: the selection of a measure to compare the treatment and control group, and a statistical method to combine the summary measure from individual studies.

1.1.1 Selection of a summary measure

The meta-analyst usually has little or no control over the choice of summary measure, because most of the decision is determined by what was employed in the primary studies. However, different summary measures are reported across the primary studies and therefore it becomes the job of the analyst to create a summary statistic that is comparable across all the studies. In some situations this task involves interpolating missing information. Here we present three classes of study outcomes: discrete data, continuous data and a miscellaneous set of outcome measures. For the following sections, we assume the simplest study design, a parallel study that compares control and treatment group.

Discrete data: risk differences, relative risks and odds ratios

Three widely used measures for discrete data are the difference between two probabilities (risk difference), the ratio of two probabilities (relative risk) and the ratio of the odds for the treatment to the odds for the control group (odds ratio). Risk difference is easy to interpret; it is defined within [0 1] boundary, and it is approximately normally distributed for modest sample sizes. On the other hand, relative risk and odds ratios are typically analyzed on the logarithmic scale and transformed back to the original scale. Each of the three measures and its variance is summarized in Table 1. In the table, 'c' denotes control group and 't' denotes treatment group; n_c and n_t are the total number of samples in the control and treatment groups, respectively; a, b, c and d denote the number of observations in each of the cells defined by the treatment and outcome table.

Table 1: Summary measures and their variance for discrete data meta-analysis.

	Risk difference	log(Relative risk)	log(Odds ratio)
Parameter	$p_t - p_c$	$\log(p_t/p_c)$	$\log\left(\frac{p_t/(1-p_t)}{p_c/(1-p_c)}\right)$
Variance	$\frac{p_t q_t}{n_t} + \frac{p_c q_c}{n_c}$	$\frac{q_t}{n_t p_t} + \frac{q_c}{n_c p_c}$	$\frac{1}{n_a} + \frac{1}{n_b} + \frac{1}{n_c} + \frac{1}{n_d}$

Continuous data: means and effect sizes

When the primary studies have continuous outcome measures, they usually report means, $\bar{x}$ in each treatment arm. Then the analyst may use the mean difference as a summary measure. Let $\bar{x}_c$ and $\bar{x}_t$ be the study means for control group and treatment group, respectively. Then the difference in means is

$$y = \bar{x}_t - \bar{x}_c$$

with its variance s^2 equal to

$$s^2 = {s_p}^2 \left(\frac{1}{n_c} + \frac{1}{n_t}\right),$$

where

$$s_p{}^2 = \frac{(n_c - 1){s_c}^2 + (n_t - 1){s_t}^2}{n_t + n_c - 2}$$

and ${s_c}^2$, ${s_t}^2$ is equal to the control and treatment group sample variance, respectively.

If there is no direct common measure across all the studies, one can transform the study-specific summary to a standardized scale-free statistic. For example the difference of means divided by the variability of the measures (standardized mean difference) can be used. Let us assume that each outcome follows a normal distribution:

$$Y_c \sim \mathcal{N}(\mu_c, \sigma^2)$$

$$Y_t \sim \mathcal{N}(\mu_t, \sigma^2)$$

Then the standardized mean difference is defined as

$$\delta = \frac{\mu_t - \mu_c}{\sigma}$$

with its variance

$$\left(\frac{1}{n_c} + \frac{1}{n_t}\right) + \frac{\delta^2}{2(n_c + n_t)}$$

Since the results from the primary studies are transformed into unitless measures, they can be pooled into the meta-analysis even when the 'event' is measured in different ways across the studies.

Other measures

When the test statistics are reported as summary measures across the studies, it is possible to recover the estimated effect size if the required information is also reported. For instance, in case the z-statistic is reported, the standardized mean difference can be calculated as follows:

$$\delta = z\sqrt{\frac{1}{n_c} + \frac{1}{n_t}}$$

1.1.2 Assess the variations between studies

Once a measure is given as an effect estimate, meta-analysis methods are variations on a weighted average of them from the different studies. Therefore, variation across studies (heterogeneity) have to be evaluated first. Let us assume that we have "k" independent studies. Then a statistical test for homogeneity of study means is equivalent to testing

$$H_0 : \theta_i = \theta_j = \theta$$

for all $i, j = 1, 2, \cdots, k$ against

$$H_1 : \theta_i \neq \theta_j$$

for at least one pair (i, j). Under the null hypothesis, the k study-specific summary statistics have a common mean θ. Therefore, under H_0 the test statistics Q_W follows chi-square distribution with degrees of freedom $k - 1$,

$$Q_W = \sum_{i}^{k} W_i \left(Y_i - \hat{\theta}_{MLE} \right) \sim \chi^2_{k-1}$$

where $\hat{\theta}_{MLE} = \sum W_i Y_i / \sum W_i$ and $W_i = 1/s_i^2$.

If H_0 is rejected, then we may conclude that the study means arose from two or more distinct populations and proceed by either identifying covariates that could stratify studies into the homogeneous subset or using a random-effects model. If we fail to reject H_0, then it is assumed that the between-study variation is small and the analyst may use θ_{MLE} as an estimate for the common mean, θ. This is called the 'chi-squared test' and is the most well known statistical test to evaluate the heterogeneity. However, it has a low power when studies are few in number or have small sample

size. Also the test has high power to detect a small amount of heterogeneity when there are many studies in a meta-analysis.

Some argue that statistical heterogeneity is inevitable and it will always exist whether or not we detect it using a statistical test [1]. Thus, the result of the heterogeneity test should be irrelevant to the choice of analysis. Methods have been developed for quantifying inconsistency across studies to assess its impact on the meta-analysis. A useful statistic for quantifying inconsistency is I^2:

$$I^2 = \frac{Q - df}{Q} \times 100\%$$

where Q is the chi-squared statistic and df is its degrees of freedom [1, 2]. This measure describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error.

1.1.3 Combine summary measures

There are at least three sources of variation to consider before combining summary measures across studies: sampling error, study-level characteristics and inter-study variation. The fixed-effects model assumes each study is measuring the same underlying parameter and that there is no inter-study variation. On the other hand, the random-effects model assumes that each study is associated with a different but related parameter.

Two common analysis approaches in the meta-analysis are fixed-effects and random-effects model. The difference between the two models is in the variance estimation of the summary measure. The fixed-effects model uses the inverse-variance as a weight for each study. The random-effects model makes an adjustment to the study weights according to the extent of variation (heterogeneity). The two models will give identical results when there is little or no heterogeneity among the studies. However, when the variability between studies is significant, confidence intervals for the summary

statistics will be wider if the random-effects method is used rather than a fixed-effects method, and the corresponding statistical claims will be more conservative.

Fixed-effects model

When the within-study variation s_i^2 is assumed to be known, the log-likelihood of θ is proportional to $\sum_i \frac{(Y_i-\theta)^2}{s_i^2}$. Then the maximum likelihood estimator (MLE) is obtained as follows:

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}$$

where $W_i = 1/s_i^2$.

Random-effects model

The random-effects model assumes that the study summary measure Y_i is a draw from a normal distribution:

$$Y_i \sim \mathcal{N}(\theta_i, {s_i}^2).$$

And each study-specific mean θ_i is assumed to be sampled from a prior distribution:

$$\theta_i \sim \mathcal{N}(\theta, \tau^2).$$

Note that θ is the average treatment effect across the studies and τ^2 is the inter-study variation. After averaging over the study-specific effects, the posterior distribution of each Y_i is again Normal with mean θ and variance ${s_i}^2 + \tau^2$.

If τ^2 is known, then the MLE of θ is given by:

$$\hat{\theta}(\tau)_{MLE} = \frac{\sum_i W_i(\tau) Y_i}{\sum_i W_i(\tau)}$$

with $W_i(\tau) = 1/\left(s_i^2 + \tau^2\right)$.

However, in most cases τ^2 is unknown. The most common methods of estimating τ^2 are the method of moments (MOM) and the Bayesian approach. The MOM estimator is obtained from the chi-squared homogeneity test. By equating Q_W with its expected value, DerSimonian and Laird [3] proposed a non-iterative MOM estimator of τ^2,

$$\hat{\tau}^2_{DL} = \max\left\{0, \quad \frac{Q_W - (k-1)}{\sum W_i - \frac{\sum W_i^2}{\sum W_i}}\right\}.$$

Then the DerSimonian-Laird estimator becomes,

$$\hat{\theta}_{DL} = \frac{\sum_i w_i(\hat{\tau}_{DL})Y_i}{\hat{\tau}_{DL}}$$

where $w_i(\hat{\tau}_{DL}) = 1/\left(s_i^2 + \hat{\tau}^2_{DL}\right)$.

A Bayesian approach incorporates the uncertainty in the estimates of hyperparmeters θ and τ^2. Let $\theta \sim \mathcal{N}(0, a^2)$ and $\tau^{-2} \sim \mathcal{G}amma(c, d)$. Then the joint posterior distribution of the unknown parameters of interest is

$$\prod_i p\left(\theta_i | y_i, s_i^2\right) p\left(\theta_i | \theta, \tau^2\right) p(\theta) p(\tau^2).$$

By taking the integral over the nuisance parameters, the posterior distribution of the summary statistics can be obtained as follows:

$$\hat{\theta}_B = \mathbb{E}\left(\theta | y, \tau^2\right) = \int_\theta \theta \int_{\theta_i, \tau^2} \left\{p(V) d\theta_i d\tau^2\right\} d\theta$$

Typically the integral is analytically intractable, in which case approximation methods such as Monte Carlo approximations to the posterior may be employed.

Frequentist and Bayesian random-effects models differ in how they estimate unknown parameters of interest. For example, in estimating the between-study variation τ^2, the frequentist model assumes that there is one true value of τ^2 which can be estimated from the observed data. On the other hand, Bayesian model treats τ^2 as a

realization of a random variable with an underlying distribution. The unknown mean and the variance of the distribution can be estimated from the data, while this procedure incorporates more sources of variability than the frequentist approach. This is a clear advantage of the Bayesian approach especially when there are heterogeneities between studies are significant.

Another difference is that the Bayesian approach starts with a prior belief, and updates it using the available data. We can estimate both the point estimate and its uncertainty from this updated distribution. For example, in conducting the meta-analysis to compare treatment and control group, we can assume that there is no difference in the effectiveness *a priori*, by employing an appropriate prior. Consequently, Bayesian model tends to have a wider confidence interval. If not only the frequentist model but also the Bayesian meta-analysis favors one over the other, it is a strong evidence that one is superior to the other.

1.1.4 Sensitivity Analysis

Once the studies are collected and analyzed, the meta-analyst needs to assess the appropriateness of the assumptions that have been made. Sensitivity analysis, a systematic approach of investigating how sensitive the results are depending on the method of analysis or changes in the data, should be considered. First of all, an exploratory analysis of the primary data should be performed to understand important features of the data. In case when we have enough number of studies, a box plot of the study effects can indicate spread (skewness, multi-modal, etc) or tails (presence of outliers).

The meta-analyst is recommended to use both a fixed-effects model and a random-effects model and compare the results from each model. Sensitivity to the distributional assumptions can be assessed by assuming different distributions for the study

effects and comparing subsequent inferences. Moreover, within a model, one should determine how sensitive the pooled estimate is to any one study or a group of studies.

1.2 Spatial data analysis

Modeling of spatial data reflects the empirical regularity that neighboring areas or observations tend to be similar, and that similarity typically diminished as distance increases. Geostatistical data includes geochemical readings, species distribution, or disease events in relation to a pollution source, and a continuous spatial framework is more relevant allowing interpolation between observed point readings. On the other hand, in discrete spatial data analysis, the data is typically aggregated with observations consisting of counts or of regional indicators.

Let $\mathbf{s} \in \mathcal{R}^d$ be a location in d-dimensional space, and $\mathbf{X}(\mathbf{s})$ be an observed value at $\mathbf{s}$. Then the full dataset can be modeled as a multivariate random process

$$\{\mathbf{X}(\mathbf{s}) : \mathbf{s} \in D\}$$

where $\mathbf{s}$ varies over the $D \subset \mathcal{R}^d$. The spatial observations can be categorized into three cases [4]:

1. *Geostatistical data*

$\mathbf{X}(\mathbf{s})$ is a random vector at location $\mathbf{s}$ that varies *continuously* over D. In this case D is a fixed subset of $\mathcal{R}^d$ which includes a d-dimensional rectangle of positive volume. It is also called *point source data*, where the locations are fixed points in space that may occur anywhere in D.

2. *Lattice data*

As in the geostatistical data, D is a fixed subset of regular or irregular shape. However

it only contains countably many sites in $\mathcal{R}^d$, where each site has its neighborhood information. It is referred to as *regional summary data models* where the source data are counts or averages over geographical regions.

3. *Point pattern data*

The index set of D gives the locations of random events (spatial point pattern).

1.2.1 Geostatistical data modeling

Let $X(s_i)$ be the spatial process associated with a particular variable (for example, chemical concentration) observed at location $\{s_i : s_i \in D \subset \mathcal{R}^d\}$. The simplest model is to assume $X(s_i)$ to be a second order stationary process:

$$\mathbb{E}\left[\mathbf{X}(\mathbf{s})\right] = \mu$$

$$Cov\left[X(s_i), X(s_j)\right] = C(s_i - s_j) < \infty$$

where $C(\cdot)$ is the *covariogram*, which is analogous to the autocovariance functions in time series analysis. Then the *variogram* is defined as

$$Var\left[X(s_i) - X(s_j)\right] = 2\gamma(s_i - s_j)$$

and $\gamma(s_i - s_j)$ is called *semivariogram*. If the variogram can be defined as a function of the Euclidean distance between s_i and s_j, then the spatial process is *isotropic*; in this case $2\gamma(s_i - s_j) = 2\gamma(d_{ij})$ and $C(s_i - s_j) = C(d_{ij})$. The *sill* is defined as $\lim_{d\to\infty} 2\gamma(d)$ and the *nugget* as $\lim_{d\to 0} 2\gamma(d)$. The nugget can have non-zero value, due to the microscale variability or measurement error.

For a second-order stationary Gaussian process, the model can be written as

$$X(s_i) = \mu + W(s_i) + \epsilon(s_i) \tag{1.2.1}$$

where μ is the global mean, $W(s)$ is spatial correlation and $\epsilon(s)$ captures random errors such as measurement error. We can use parametric covariance function for an isotropic spatial model,

$$Cov\left(W(s_i), W(s_j)\right) = \sigma^2 \rho(d_{ij}; \phi).$$

Most common choice of ρ is

$$\rho(d_{ij}; \phi) = exp(-\phi d_{ij}), \phi > 0$$

so that the covariance decreases with respect to the distance at exponential rate ϕ. The random error is usually assumed to be $\epsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau^2)$. Then the variogram can be derived as follows:

$$2\gamma(\tau^2, \sigma^2, \phi) = 2\left[\tau^2 + \sigma^2\left(1 - \rho(d_{ij}; \phi)\right)\right] \tag{1.2.2}$$

where τ^2 is the nugget and $\tau^2 + \sigma^2$ is the sill.

In many practical applications, it is of interest to predict the unobserved value $X(t)$ at some target location t. Basic spatial interpolation methods include inverse distance weighting, which attenuates the variable with decreasing proximity from the observed location. Kriging is another model, which is based on the minimum mean squared error approach. Assume that we seek predictions at target locations $(t_1, \cdots, t_m)$, given response data at source locations $(s_1, \cdots, s_n)$. From (1.2.1) and (1.2.2), the joint distribution of $\mathbf{X}_1 = (X(s_1), \cdots, X(s_n))'$ and $\mathbf{X}_2 = (X(t_1), \cdots X(t_m))'$ is

$$\left[(\mathbf{X_1}, \mathbf{X_2})' \,|\mu, \tau^2, \sigma^2, \phi\right] \sim \mathcal{MVN}_{m+n}\left(\mu \mathbf{1}_{m+n}, \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}\right) \tag{1.2.3}$$

where $\sum_{ij} = Cov(X_i, X_j), i, j = 1, 2$. For example, the ij^{th} element of $\sum_{11}$ is $\sigma^2 exp(-\phi d_{ij})$. Then the conditional distribution of the target values $\mathbf{X_2}$, given source

values $\mathbf{X_1}$ and the model parameters is

$$\left[\mathbf{X_2}|\mathbf{X_1}, \mu, \tau^2, \sigma^2, \phi\right] \sim \mathcal{MVN}_m\left(\mu_{2.1}, \sum_{2.1}\right)$$

with

$$\mu_{2.1} = \mu\mathbf{1}_m + \sum_{21}\sum_{11}^{-1}\left(\mathbf{X}_1 - \mu\mathbf{1}_n\right)$$

$$\sum_{2.1} = \sum_{22} - \sum_{21}\sum_{11}^{-1}\sum_{12}.$$

If the spatial correlation bewteen $X(s_i)$ and $X(s_j)$ depends not only on the Euclidean distance but on its direction as well, the process $\mathbf{X}(\mathbf{s})$ is *anisotropy*. A common model for anisotropy process is *geometric range anisotropy*, where we assume a positive definite matrix B with variogram

$$2\gamma(\mathbf{h}_{ij}) = 2\gamma\sqrt{\mathbf{h}_{ij}' B \mathbf{h}_{ij}}.$$

Note that when $B = \phi^2 I$, the term $\sqrt{\mathbf{h}_{ij}' B \mathbf{h}_{ij}} = \phi d_{ij}$ which is equal to an isotropy process. Therefore, the anisotropy process is a generalization of the isotropy process with its variogram equal to

$$2\gamma(\tau^2, \sigma^2, B) = 2\left[\tau^2 + \sigma^2\left(1 - \rho\sqrt{\mathbf{h}_{ij}' B \mathbf{h}_{ij}}\right)\right] \tag{1.2.4}$$

and τ^2 is nugget and $\tau^2 + \sigma^2$ is sill.

1.2.2 Lattice data modeling

Within a region, suppose that counts or rates are observed for subgroups of the population; an observation is denoted by y_{ij} where $i = 1, \cdots, I$ indexes the regions and $j = 1, \cdots, J$ indexes the subgroups (or repeated measure in each region). The main effect for region, say η_i, offers many possibilities. For example, we might have a set of independent variables collected for each region. Then they can be incorporated

into the model through a specified parametric function (typically linear). In addition to or in the absence of covariate information, we can include regional random effects ϕ_i to capture spatially correlated heterogeneity. One of the most common way to model the spatial effect is a conditional autoregressive (CAR) model. In CAR model, we assume that the conditional density of $\phi_i|\phi_{j\neq i}$ is proportional to

$$exp\left[-\frac{\lambda}{2}\left(k_i\phi_i - \textstyle\sum_{j\neq i} w_{ij}\phi_j\right)^2\right]$$

where $w_{ij} \geq 0$ is a weight reflecting the influence of ϕ_j on the expectation of ϕ_i, and $k_i > 0$ is a "sample size" associated with region i.

Two special types of modeling have been discussed in the literature. First model requires a matrix of inter-region distances d_{ij}. Then the weight w_{ij} is equal to $g(d_{ij})$ for a suitable decreasing function $g(\cdot)$. For example, $g(\cdot)$ can be based on the estimated variogram of the observations. Since $w_{ij} = w_{ji}$, the symmetry of B requires k_i to be constant. The second approach defines a set δ_i of neighbors of region i, or as regions within a prescribed distance of region i. Let n_i be the number of neighbors of region i and let $w_{ij} = 1/n_i$ if $j \in \delta_i$, and 0 otherwise. Note that B is symmetric if $k_i = n_i$. Then we obtain

$$\phi_i|\phi_{j\neq i} \sim \mathcal{N}\left(\bar{\phi}_i, \frac{1}{\lambda n_i}\right)$$

where $\bar{\phi}_i = n_i^{-1}\sum_{j\in\delta_i}\phi_j$. Since the sum of all rows and the sum of all columns are zero, B is singular, and hence the joint density is improper. This second model can be combined with the first one, by setting $w_{ij} = g(d_{ij})$ for $j \in \delta_i$. The improper CAR model can be resolved by adding sum-to-zero constraint $\sum_{i=1}^{I}\phi_i = 0$.

It is important to incorporate both spatially correlated heterogeneity (CH) and uncorrelated heterogeneity (UH) into the model. It is because unobserved effects within a study could take on a variety of forms. Convolution models with CAR CH random effects have been shown to be robust under simulation to a wide range of

underlying true risk models [5].

1.2.3 Residual Spatial Autocorrelation

After fitting the model, it needs to be addressed whether any residual spatial structure has been left within the data. Since a good model fit should leave residuals with little or no spatial correlation, a test for spatial correlation in the residuals from a model fit can be a useful guide. Various statistics could be used to assess the autocorrelation, but the most common measure is Moran's I which is defined as a ratio of quadratic forms,

$$I = \mathbf{e}'\mathbf{W}\mathbf{e}/\mathbf{e}'\mathbf{e}$$

where $\mathbf{e} = \{e_1, \cdots, e_m\}$ and $e_i = (y_i - \hat{y}_i)/\sqrt{var(\hat{y}_i)}$. The W is the 0/1 $m \times m$ adjacency matrix for the regions with elements w_{ij}. For a Poisson count data model, this residual can be defined as $e_i = (y_i - \hat{y}_i)/\sqrt{\hat{y}_i}$. Positive (negative) values indicate positive (negative) spatial autocorrelation in the residuals. Values range from -1 (indicating perfect dispersion) to $+1$ (perfect correlation). A zero value indicates a random spatial pattern. Moran's I values can be transformed to Z-scores, in which values greater than 1.96 or smaller than -1.96 indicate statistically significant spatial autocorrelation with level 0.05.

1.3 Wavelet analysis

A wide range of complex structures in nature is characterized by irregular behavior. Examples of such irregular signals in both time and scale are abundant in medicine, physics, economics, and geosciences. Although irregular, such signals can be well modeled by multifractal processes. Multifractal behavior can be quantified by statistical similarity of patterns at many different scales. The regularity index describes

the strength of the similarity. However, the scaling is usually complex and can be inhomogeneous in both time and scale. Multifractal formalism has been developed to quantify the irregular scaling pattern.

1.3.1 Self-similarity and Scaling

Self-similar processes are widely used to model various real-world phenomena including engineering, physics, medicine, biology, engineering, art, economics, astronomy and chemistry. Multiscale methods (wavelets, wavelet-like decompositions, general time/frequency representations) provide tools and environments to evaluate and classify high dimensional data that exhibit scaling laws and long range dependence.

A process $Y(t)$ is self-similar with self-similarity index $H > 0$ (H-ss) if

$$Y(at) \stackrel{\mathrm{d}}{=} a^H Y(a).$$

Here $\stackrel{\mathrm{d}}{=}$ denotes equality in all finite-dimensional distributions. An H-ss process with stationary increments exhibits long range dependence (LRD) when $H > 1/2$. If the distributional properties of a process are intrinsically invariant to changes of a scale, the process has *scaling property*. The scaling behavior of a signal is closely related to singularity of wavelet coefficients. The singularity (Hölder exponent) and the self-similarity (Hurst exponent) are obtained through multi-scale analysis of wavelet transforms.

To detect the phenomena of singularity and self-similarity using wavelets, let us consider an L_1-normalized orthogonal wavelet basis comprised of $\psi_{j,k}(t) = 2^j\psi(2^j t - k)$. We assume that the $\psi(t)$ has $\mathcal{R}$ vanishing moments:

$$\int t^r \psi(t)dt = 0, r = 0, \ldots, \mathcal{R} - 1$$

The coefficients of discrete wavelet transform of a process Y are defined by

$$d_{j,k} = \int_{-\infty}^{\infty} Y(t)\psi_{j,k}(t)dt$$

which carries information on the local difference of the process near to the position k on a dyadic scale j.

If $Y(t)$ is of Hölder exponent H, then

$$\lim_{k2^{-j}\to t} |d_{j,k}| = \mathcal{O}(2^{-jH}) \tag{1.3.1}$$

for any wavelet with $\mathcal{R} > H$ [6, 7]. Therefore, the decay of the local differences of a process is related to the singularity of the signal, provided that the decomposing wavelet is more regular than the process.

The Equation (1.3.1) also serves as a basis for wavelet based estimation of H:

$$\log_2 \mathbb{E}|d_{j,k}|^q = -jqH + C_q$$

where C_q is a constant depending on q, wavelet function ψ, and the magnitude of the signal.

We assume that the wavelet coefficients are uncorrelated, and hence independent, as has been approximately the case in various contexts. Then the sum of squared wavelet coefficients represents expected level-wise energies. Abry et al. [8, 9] defined the *wavelet spectra* of a process with stationary increments as follows:

$$e_2(j) = \mathbb{E}(d_{j,\cdot})^2$$

where $d_{j,\cdot}$ stands for an arbitrary wavelet coefficient from the level j. The plot of $(j, \log_2 e(j))$ is referred to as *theoretical logscale diagram*.

Let $\hat{e}(j)$ be an estimator of $e(j)$. Then, the plot of $(j, \log_2 \hat{e}(j))$ is called *logscale diagram* or *scalogram*. Since the wavelet transform is linear, the wavelet spectra gives complete second-order description of the random process. As in the case of Fourier counterparts, energy is large at dominant scales.

Let, for some $q \in \mathbb{R}$

$$e_q(j) = \mathbb{E}|d_{j,\cdot}|^q.$$

The plot of $(j, \log_2 e(j))$ is q^{th} order theoretical logscale diagram. One way to estimate $e_q(j)$ is to use the empirical moments of the wavelet coefficients at dyadic scale j,

$$\hat{e}_q(j) = \frac{1}{n_j} \sum_k |d_{j,k}|^q$$

where n_j is the number of $d_{j,k}$ available at dyadic scale j. The plot $(j, \log_2 \hat{e}_q(j))$ is called q^{th} order logscale diagram (q-*LD*) or q^{th} scalogram. Straight lines in q-*LDs* imply empirical evidence for monotone scaling.

1.3.2 Fractionality

Brownian motion and its families are widely used to model the complex structure of the real-world data. Brownian motion $B(t)$ is a random process continuous in time t and is defined as follows:

(1) $B(0) = 0$

(2) $B(t)$ is Gaussian random variable with zero mean and variance t

(3) For any choice n and $0 \leq t_1 \leq \cdots \leq t_n$, the increments $\big[B(t_2) - B(t_1)\big], \ldots,$ $\big[B(t_n) - B(t_{n-1})\big]$ are independent and stationary.

The covariance of Brownian motion is then

$$\mathbb{E}\big[B(t)B(s)\big] = \frac{1}{2}\big(t + s - |t - s|\big) = \min[t, s].$$

The difference process

$$Y(n) = B_H(n+1) - B_H(n)$$

is called fractional Gaussian noise (fGn). The covariance function of fGn is

$$\gamma(h) = \frac{E|X(1)|^2}{2}[(h+1)^{2H} - h^{2H} + (h-1)^{2H}].$$

Fractional Brownian motion (fBm) is a generalization of Brownian motion. A zero mean Gaussian process with stationary self-similar increments is called fractional

Brownian motion with Hurst exponent H (fBm_H) if

$$B_H(t) \sim \mathcal{N}(0, \sigma^2|t|^{2H})$$

and

$$B_H(t+\tau) - B_H(t) \overset{\text{d}}{=} B_H(\tau) - B_H(0) \overset{\text{d}}{=} \tau^H B_H(1).$$

As a zero mean Gaussian process, $B_H(t)$ could be alternatively defined via its covariance structures:

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{\sigma^2}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right]$$

Fractional Brownian motion has been commonly used to characterize a wide range of structures in natural phenomena that exhibit self-similarity and long-range dependence. The fractal dimension captures the regularity of the signal, and it is measured by the Hurst exponent H. Mandelbrot and Van Ness [10] extended fBm by adopting time-varying Hurst exponent. This stochastic process is called multifractional Brownian motion (mBm) and is used to model both a long-range dependence and path regularity varying with time.

The multifractional Brownian motion $\big(W(t)\big)_{t\geq 0}$ with Hurst function $H(t)$ and its scaling factor C has covariance function:

$$\mathbb{E}[W(t)W(s)] = \frac{C}{2} g(H(t), H(s))\{|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)}\}$$

for $H \in \mathbb{C}^{\eta}([0,1])$, $s, t \in [0,1]$, and

$$g(H(t), H(s)) = \frac{\sqrt{K(2H(t))K(2H(s))}}{K(H(t)+H(s))}.$$

The fractional Brownian motion is a non-stationary process whose sample paths exhibit a homogeneous degree of smoothness or regularity. And for many applications, this regularity may be too homogeneous. In particular, one may want the sample paths of a process that exhibits differing degrees of regularity as a function of time. In the following section, we introduce processes with fractal characteristics that exhibit such degree of complexity, which are often referred to as *multifractals*.

1.3.3 Multifractality

Multifractal processes exhibit patterns of locally varying scaling behavior similar to that encountered in real data sets. Since multifractal processes are non-stationary, standard approaches in time series analysis such as Fourier transform are not appropriate because the Fourier transform is not localized in time. It was first introduced in the context of turbulence and applied in many other contexts such as DNA patterns research, earthquake distribution analysis, signal processing and internet data traffic modeling.

In order to study the varying local properties of multifractal processes, tools able to localize information both in time and frequency are appropriate. Given that wavelets are local in both frequency/scale (via dilations) and in time (via translations), the wavelet defined multiscale analysis is more suitable in this setting.

The distribution of the regularity indices that vary in time and scale is called the multifractal spectrum (MFS). MFS of a process is a summary of its scaling and singularity properties. Let us consider the local singularity strength of wavelet coefficients as follows:

$$\alpha(t) = \lim_{k2^{-j} \to t} -\frac{1}{j} \log_2 |d_{j,k}|.$$

Small values of $\alpha(t)$ reflect the more irregular behavior at time t. Any inhomogeneous process has a collection of local singularity strength measures, and their distribution $f(\alpha)$ forms the MFS.

Therefore, the multifractal spectrum (MFS) summarizes variable degrees of scaling in signals. In the case of fractals, scaling refers to the propagation of energy when the signals or images are inspected as various resolutions. The dynamics of the scaling can be used as discriminatory descriptors, thus, multifractality provides an additional window through which to look at the data and makes inference not possible with standard statistical approaches.

Rather than operating with MFS as a density function, we summarize it by a small number of meaningful descriptors. Theoretically, the MFS of fBm which represents monofractal process, consists of three geometric parts: the vertical line, the maximum point, and the right slope. However, it is rare to obtain such a perfect spectrum in practice, even for a well-simulated fractional Brownian motion.

Although there are estimation errors, the MFS can be approximately summarized by three canonical descriptors (multifractal descriptors) without loss of the discriminant information. The proposed summaries are (1) the spectral mode (Hurst exponent, H), (2) left slope (LS) or left tangent (LT) and (3) width spread (broadness, B) or right slope (RS) or right tangent (RT).

The broadness (B) can be defined as follows [11]. Suppose that α_1 and α_2 are two roots which satisfy the equation $f(\alpha) + C = 0$ and $\alpha_1 < \alpha_2$. The broadness (B) of MFS is then $B = \alpha_2 - \alpha_1$. Since $\alpha(q)$ is discrete, it is intractable to find the exact root of the equation $f(\alpha) + C = 0$. To get around this, we first find the two closest points $(\alpha_i^l, f(\alpha_i^l))$ and $(\alpha_i^u, f(\alpha_i^u))$ for each i such that $f(\alpha_i^l) < -C$ and $f(\alpha_i^u) > -C$ for $i = 1, 2$ and then obtain the two solutions α_1, α_2 by interpolation.

The slopes LS and RS and tangents LT and RT can be obtained using the

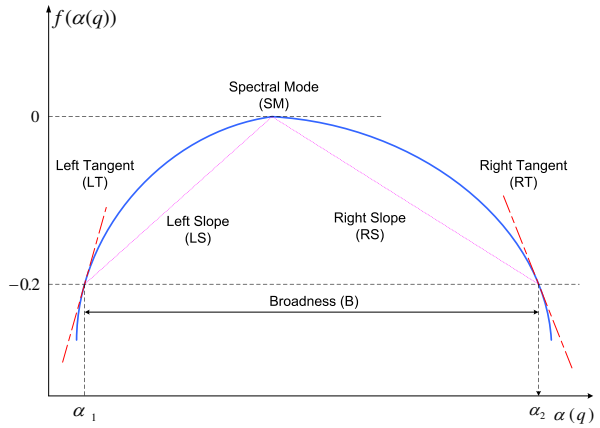


Figure 1: Geometric descriptors of MFS

interpolation technique, as computed by

$$LS = \frac{C}{H - \alpha_1}\ , \qquad RS = \frac{-C}{\alpha_2 - H}$$

$$LT = \frac{f(\alpha_1^u) - f(\alpha_1^l)}{\alpha_1^u - \alpha_1^l}\ , \qquad RT = \frac{f(\alpha_2^u) - f(\alpha_2^l)}{\alpha_2^u - \alpha_2^l}$$

1.3.4 Wavelets

Wavelet analysis decomposes signals and images into component waves of varying durations, called wavelets. Wavelets are unconditional bases that are local and decompose signals and images into a hierarchy of detailed features. They can be used for a wide variety of fundamental data processing tasks, such as compression, removing noise, dimension reduction or classification.

Multiresolution analysis (MRA) shows how discrete signals are synthesized by beginning with a very low resolution signal and successively adding details to create higher resolution versions, ending with a complete synthesis of the signal at the finest resolution. The signal f is expressed as a sum of a lower resolution (or averaged signal,

A) added with a signal made up of fluctuations (or details, **D**). These fluctuations provide the added details necessary to produce the full resolution signal f. In general, if the number of signal values ($N = 2^J$) is divisible K times by 2, then a K-level MRA,

$$f = \mathbf{A}^{J-K} + \mathbf{D}^{J-K} + \cdots + \mathbf{D}^{J-2} + \mathbf{D}^{J-1}$$

can be performed on the signal f.

Wavelet decomposition of a function $f(t)$ can be written as follows;

$$f(t) = \sum_k c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k d_{j,k}\psi_{j,k}(t)$$

with

$$c_{j_0,k} = \int f(t)\phi_{j_0,k}(t)dt,\ d_{j,k} = \int f(t)\psi_{j,k}(t)dt$$

where $\sum_k c_{j_0,k}\phi_{j_0,k}(t)$ and $\sum_k d_{j,k}\psi_{j,k}(t)$ represent $\mathbf{A}^{j_0}$ and $\mathbf{D}^j$, respectively. The index j_0 indicates the coarsest scale or lowest resolution of analysis, and smaller j correspond to higher resolutions of the analysis.

A mother wavelet $\psi(t)$, is an oscillatory function which decays rapidly in time. The shifted versions of a low-pass scaling function $\phi(t)$ is called father wavelet. The wavelet at resolution level j and location k is given by

$$\phi_{j_0,k}(t) = 2^{j_0/2}\phi(2^{j_0}t - k),\ \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$$

The set $\{\phi_{j_0,k}, \psi_{j,k}; j \geq j_0, k \in \mathbb{Z}\}$ forms an orthonormal basis of $L_2(\mathbb{R})$, the space of square integrable functions on $\mathbb{R}$.

The wavelet transform are readily generalized to multidimensional case. The 2-D wavelet basis functions are constructed via translations and dilations of a tensor product of univariate wavelets and scaling functions:

$$\phi(t_1, t_2) = \phi(t_1)\phi(t_2)$$

$$\psi^h(t_1, t_2) = \phi(t_1)\psi(t_2)$$

$$\psi^v(t_1, t_2) = \psi(t_1)\phi(t_2)$$

$$\psi^d(t_1, t_2) = \psi(t_1)\psi(t_2)$$

The symbols h, v, d stand for horizontal, vertical and diagonal directions, respectively. For technical reasons we consider L_1-normalization of wavelets atom instead of standard L_2-normalization:

$$\phi_{j_0,k}(t) = 2^{j_0}\phi(2^{j_0}t_1 - k_1, 2^{j_0}t_2 - k_2)$$

$$\psi^i_{j,k}(t) = 2^j\psi^i(2^j t_1 - k_1, 2^j t_2 - k_2)$$

for $i = h, v, d$ and where $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$ and $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$. All the wavelets that we consider in this paper are compactly supported, i.e., they are zero outside some compact interval. For more technical details and overview on wavelets, see Vidakovic [12].

1.3.5 Complex wavelets

Although the discrete real-valued wavelet transform (DWT) is a powerful tool for analyzing the complex structures of the data, it has three disadvantages that undermine its usage in many applications. First, it is not shift invariant because small input signal shifts generate changes in all DWT coefficients. Second, DWT analysis lacks the phase information that accurately describes non-stationary signal behavior. Third, the two-dimensional (2-D) separable DWT suffers from poor directionality because DWT coefficients reveal only three spatial edge orientations, diagonal, horizontal and vertical [13].

The discrete complex wavelet transform (DCWT) has been advocated to alleviate the limitations of DWT [14, 15]. To illustrate the construction of complex wavelet filter, let us consider the Daubechies wavelets. The Daubechies wavelets are compactly

supported orthogonal wavelets with a pre-assigned degree of smoothness, which is specified by vanishing moments $\mathcal{R}$. Increasing values of $\mathcal{R}$ correspond to smoother wavelets. For a given value of $\mathcal{R}$ there are $2\mathcal{R}-1$ possible solutions to the equations that define the Daubechies wavelets, but not all are distinct. For example, when $\mathcal{R}=3$ there are four solutions. However, only two are distinct: two solutions give the real extremal phase wavelet; the other two are a conjugate pair, hence giving equivalent complex-valued wavelets.

Apart from the Haar wavelet, complex wavelets with an odd number of vanishing moments are only compactly supported wavelets which are symmetric [16]. As well as these symmetric solutions, asymmetric complex-valued wavelets occur when there are four or more vanishing moments. Another advantage of the complex wavelets is that they produce the phase information as well as the spectrum modulus. It is known that phase and spectrum are collaborating in a nontrivial way to describe the data: phases encode most of the coherent (in space and scale) structure of the image and the spectrum mostly encode the strength of local information that could be corrupted with noise [15]. Due to this advantages, complex wavelet has been used in various areas including motion estimation [17], texture image modeling [18], image denoising [19] and NMR spectra classification [20].

CHAPTER II

ECOLOGICAL-TYPE INFERENCE IN MATCHED-PAIR STUDIES WITH FIXED MARGINAL TOTALS

In this chapter we propose a Bayesian approach for matched-pair studies that only reported marginal totals to retrieve full paired table. Since traditional ecological inference that assumes parallel tables are not directly applicable, we propose an algorithm based on Metropolis random walk and a hierarchical Bayes model. The procedure relies on MCMC simulation, recovers a full table and reports its accuracy in terms of credible sets for the cell counts. The performance of the model is remarkable; from 17 matched-pair study examples, four of the estimated tables exactly coincided with the original tables and nine tables deviated only (+1,-1) in terms of cell counts. This methodology is motivated by meta-analysis of split-mouth designs assessing the effectiveness of glass-ionomer (GI) to resin-based (RB) sealants in preventing caries in juvenile permanent molars.

2.1 Introduction

Meta-analyses synthesize quantitative data, typically from randomized controlled trials (RCT), to compare the effectiveness of different interventions to treat disease. In a typical RCT, patients are randomly assigned to different treatment groups. Another type of RCT commonly used in dental, ophthalmology, and pharmacology trials is a matched pair design where interventions are applied to the same patient. For example, in randomized split-mouth trials comparing the effectiveness of tooth specific interventions to prevent decay, one tooth in a subject is randomly selected to receive

treatment A while the contralateral tooth in the same subject receives treatment B. Another example is cross-over trials testing the efficacy of drugs. In this design, a patient is randomly administered treatment A or B in the first time period and then administered the remaining treatment in the second time period. The link between the split-mouth design and cross-over trials is apparent – the tooth location in the split-mouth design is analogous to time in the cross-over design.

Matched pair design has statistical advantages. Because the control and test groups are subject to the same environment, this design controls for many confounding factors. Thus differences in outcomes between test and control groups are likely attributable to the treatment. Moreover, since control and test groups receive both interventions, matched pair studies usually require no more than half the number of the subjects to produce the same precision as parallel group studies [21]. The matched pair design for comparing control and treatment groups can be represented as Table 2.

Table 2: Matched-pair design table

		Control		
		Event	Non-event	Total
Treatment	Event	y_{11}	y_{12}	$y_{1\cdot}$
	Non-event	y_{21}	y_{22}	$y_{2\cdot}$
Total		$y_{\cdot 1}$	$y_{\cdot 2}$	n

The sample size n relates to the number of paired observations and cell counts y_{ij} represent the number of pairs for which one of the four combinations of events/non-events was observed. The results in Table 2 can be summarized as a traditional 2×2 contingency table, ignoring the pairing:

In the following discussion, we refer to Table 2 as *paired* and to Table 3 as *parallel*. When only Table 3 is reported, the information on individual counts y_{ij} is missing and the paired table has the form as in (2.1.1),

Table 3: Contingency table summarizing Table 2 as if two parallel groups were considered

	Event	Non-event	Total
Treatment	$y_{1\cdot}$	$y_{2\cdot}$	n
Control	$y_{\cdot 1}$	$y_{\cdot 2}$	n

$$\begin{array}{cc|c} & & y_{1\cdot} \\ & & y_{2\cdot} \\ \hline y_{\cdot 1} & y_{\cdot 2} & n \end{array} \tag{2.1.1}$$

While the majority of matched pair studies are well conducted, reporting of results differs. Some studies reported the full information as in Table 2, others reported only the total number of events for each group, $y_{1\cdot}$ and $y_{\cdot 1}$. A major challenge in meta-analysis of matched-pair studies is proper incorporation of those studies that only provide the marginal totals. Meta-analyses of the effectiveness of dental sealant materials have either excluded those studies [22] or treated them as if they were parallel group studies [23]. However, assuming that they are parallel group studies is the same as assuming that the cell counts $y_{1\cdot}$ and $y_{\cdot 1}$ in Table 3 are independent. Since $y_{1\cdot}$ and $y_{\cdot 1}$ have a positive correlation, ignoring pairing inflates the variance of the risk measures, increases the type II error [24, 25], and may lead to inaccurate statistical conclusions. Also, meta-analyses may unnecessarily limit their statistical power by excluding studies whose aggregate reports still contain considerable information about the problem.

Since only marginal totals are available, as in (2.1.1), this problem can be referred to as *ecological inference*. Typically, ecological inference deals with several groups (geographical areas) in which the number of individuals and number of cases are known. The goal is to infer the number of cases in terms of an explanatory variable,

such as exposure. The body of research in ecological inference includes numerous papers, monographs and edited volumes. A monograph by King [26] and edited volume by King, Tanner and Rosen [27] provide comprehensive overviews of the current state of research in ecological inference and a wealth of references.

Solutions to the ecological problem are based on a single binomial model with the probability as a weighted average, convolution of two binomial distributions or its normal approximation, conditional distribution of binomials given their sum, method of bounds, ecological regression, and nonlinear neighborhood method, to list just a few approaches. Bayesian solutions are discussed by Good [28], Fienberg and Holland [29], Diaconis and Sturmfels [30], Evans, Gilula, and Guttman [31], Wakefield [32], Jackson, Best and Richardson [33], Dobra [34], Dobra, Tebaldi and West [35], Salway and Wakefield [36], Liu et al. [37], among others.

This chapter focuses on a Bayesian approach to ecological-type inference in the matched pair studies because the traditional methods of ecological inference that assume parallel tables are not directly applicable. The proposed solution builds on work of Diaconis and Sturmfels [30], Dobra [34], and Dobra, Tebaldi and West [35], whose Markov-basis simulation approach leads to a powerful Bayesian computational methodology. The proposed procedure is based on a hierarchical Bayes model which incorporates prior information about the within-pair association. This information can be combined via the prior/hyper-prior structure, or as a separate Bayesian analysis that is conducted on studies with full information.

The proposed algorithm recovers the full table and reports its accuracy in terms of credible sets for the cell counts. This methodology will support inclusion of more well conducted studies into meta-analyses, increasing their statistical power even though the table entries may not be known with certainty.

2.2 Methods

In this section we describe a Bayesian approach for estimating individual pair counts. It is based on a Metropolis sampling scheme that is built on work of Dobra, Tebaldi, and West [35]. The algorithm takes the Bayesian estimator of the pivotal cell count y_{11} and uses it to estimate other cell counts from the relative marginal totals. The model can easily incorporate any prior information available.

2.2.1 Preliminaries

Assume that the table $\boldsymbol{y}$ as in (2.1.1) is partially observed, that is, all the marginal totals $y_{1\cdot}$, $y_{2\cdot}$, $y_{\cdot 1}$ and $y_{\cdot 2}$ are known, while the individual entries y_{11}, y_{12}, y_{21}, and y_{22} are unknown. If no restrictions on the cell counts are given, the likelihood is the product of independent Poisson distributions with cell intensities $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$; if the total count n is given, the likelihood becomes multinomial with cell probabilities $\boldsymbol{p}\,(\boldsymbol{\lambda})$, while when all marginals are given, the likelihood is a conditional multinomial distribution. For 2×2 tables, the conditional multinomial probability involves only the *pivotal cell*, the count y_{11},

$$\boldsymbol{P}\,(\boldsymbol{y}) = \frac{\frac{n!}{y_{11}!(y_{1\cdot}-y_{11})!(y_{\cdot 1}-y_{11})!(n-y_{1\cdot}-y_{\cdot 1}+y_{11})!}\exp\{y_{11}\theta\}}{\sum_z \frac{n!}{z!(y_{1\cdot}-z)!(y_{\cdot 1}-z)!(n-y_{1\cdot}-y_{\cdot 1}+z)!}\exp\{z\theta\}} \tag{2.2.1}$$

with y_{11} and z satisfying $\max\{0, y_{1\cdot} + y_{\cdot 1} - n\} \leq y_{11}, z \leq \min\{y_{1\cdot}, y_{\cdot 1}\}$, and

$$\theta = \log p_{11} - \log p_{12} - \log p_{21} + \log p_{22}, \tag{2.2.2}$$

where p_{ij} are the probabilities associated with the cell (i, j). Note that $\exp\{\theta\}$ measures the association in 2×2 tables [38], and it stands for the product of probabilities of concordant pairs divided by the product of probabilities of discordant pairs. For parallel tables, $\exp\{\theta\}$ is simply the odds ratio. We will see later that distributions

on the space of tables with fixed marginals can be introduced by selecting sampling probabilities, and that prior information can guide the distribution choice.

2.2.2 Metropolis random walk on a space of paired tables with fixed marginal counts

Denote by $\mathcal{Y}$ the set of all tables with nonnegative integer entries consistent with the observed marginal counts as in (2.1.1). Diaconis and Sturmfels [30] propose a general algorithm that samples from the space $\mathcal{Y}$ according to a specified target distribution by forming a Markov chain among the tables. The chain moves to the next state if an element of the Markov basis is added to the table that represents the current state. In terms of matrices, the subsequent state in the chain is simply the sum of two matrices (tables): the current state and an element from the basis. It can be shown that the Markov basis $\mathcal{M}$ for 2×2 tables consists of two elements

$$\mathcal{M} = \left\{ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \right\}$$

and that any two tables $\boldsymbol{y}$, $\boldsymbol{y}'$ from $\mathcal{Y}$ can be connected by applying finite number of steps from $\mathcal{M}$,

$$\boldsymbol{y}' = \boldsymbol{y} + \sum_{i=1}^{k} \boldsymbol{f}_i,$$

where $\boldsymbol{f}_i$ is selected from $\mathcal{M}$, and k is a finite number. It is easy to see that adding an element from $\mathcal{M}$ preserves the marginal totals. Also, if the selected move $\boldsymbol{f}$ makes any of the cell entries negative, the move is not accepted and the chain stays in the current state until the acceptable move is selected. Dobra, Tebaldi and West [35] expand the methodology of Diaconis and Sturmfels [30] to tables of arbitrary size and dimension, and with various subsets of marginal constraints.

To generate a table with a specific distribution on $\mathcal{Y}$, one forms the Markov chain with a properly designed Metropolis acceptance step. Let π be the target

distribution function over $\mathcal{Y}$ and $\boldsymbol{\phi}$ be the value of the parameter. Given the current state $(\boldsymbol{y}\ , \boldsymbol{\phi}\ ^*)$, a proposal table $\boldsymbol{y}\ '$ is generated by adding an element of the Markov basis f_i to the current table $\boldsymbol{y}$. The move from $\boldsymbol{y}$ to $\boldsymbol{y}\ '$ is accepted with probability

$$\rho = \min\left\{1, \frac{\pi(\boldsymbol{y}\ '|\boldsymbol{\phi}\ ^*)}{\pi(\boldsymbol{y}\ |\boldsymbol{\phi}\ ^*)}\right\}. \tag{2.2.3}$$

This expression is quite simple since we operate with a symmetric proposal distribution and thus the acceptance probability in (2.2.3) depends only on the target probability.

At any state of the chain, the table entries are known and the likelihood on $\mathcal{Y}$ is multinomial

$$\prod_{i,j} \frac{n!}{y_{11}!\ y_{12}!\ y_{21}!\ y_{22}!}\ p_{ij}^{y_{ij}}. \tag{2.2.4}$$

The prior on lexicographically ordered cell probabilities is Dirichlet,

$$\boldsymbol{p}\ = \{p_{ij}\} \sim \mathcal{D}irichlet(\boldsymbol{q}\ , s), \tag{2.2.5}$$

where the hyper-parameter $\boldsymbol{q}$ is vector of probabilities attracting tables from the Markov chain, while the hyper-parameter s measures the strength of this attraction. Parameter s is often called "prior sample size" and in applications it should be smaller than the observed size of table, n.

The hyper-parameter $\boldsymbol{q}$ incorporates any prior information available. If there are a number of studies that are reported in a paired manner, those studies can be used to calibrate $\boldsymbol{q}$. One way is performing a meta-analysis with a prior placed on the parameter $\boldsymbol{q}$. Taking the Bayesian estimator (for example, mean or median) of the posterior distribution of $\boldsymbol{q}$, we can use it as a hyper parameter of the Dirichlet prior in (2.2.5).

For the likelihood in (2.2.4) the Metropolis acceptance probability is

$$\rho = \min\left\{1, \frac{y_{11}!\ y_{12}!\ y_{21}!\ y_{22}!}{y'_{11}!\ y'_{12}!\ y'_{21}!\ y'_{22}!}\ \exp\{\pm\theta^*\}\right\},$$

where y_{ij} and y'_{ij} are entries of tables $\boldsymbol{y}$ and $\boldsymbol{y}'$. The table $\boldsymbol{y}$ represents the current state of the chain while $\boldsymbol{y}'$ is the proposal. The parameter θ^* is as in (2.2.2), with probabilities generated by the Dirichlet prior. The sign of θ^* is determined by the selected element f_i from the Markov basis in the current step.

The proposal with the probability ρ is equivalent, with respect to the acceptance criterion (2.2.3), to the suggestion by Dobra, Tebaldi and West [35] who define the probability measure on $\mathcal{Y}$ as a product of independent Poisson distributions with the goal to incorporate the log-linear model via Poisson intensities.

2.3 Examples

In this section we first demonstrate how our methodology can be used to retrieve the full table from a randomized split mouth study examining the effectiveness of dental sealants. We next compare the variance obtained with our methodology to that obtained under the assumption that control and treatment groups are parallel. Finally in the last section, we assess the performance by comparing the full table obtained from our methodology to the actual full table for studies included in two meta-analyses: one comparing the effectiveness of different sealant materials in randomized split-mouth trials [22], and the other comparing two types of insulin in randomized cross-over trials [39].

In this chapter, risk ratio (RR) is adopted as a measure to compare the performance of the control and the treatment groups. The risk ratio (RR) of the matched pair studies requires marginal totals and can be estimated from Table 2 as

$$RR = \frac{y_{11} + y_{12}}{y_{11} + y_{21}} = \frac{y_{1\cdot}}{y_{\cdot 1}} \tag{2.3.1}$$

which is the same as in the parallel studies. However, care needs to be taken in the estimation of the standard error. The correct standard error of the log risk ratio for

paired design is

$$SE(\log(RR))_{paired} = \sqrt{\frac{1}{y_{1\cdot}} + \frac{1}{y_{\cdot 1}} - \frac{2y_{11}}{y_{1\cdot}y_{\cdot 1}}}, \tag{2.3.2}$$

as opposed to

$$SE(\log(RR))_{parallel} = \sqrt{\frac{1}{y_{1\cdot}} + \frac{1}{y_{\cdot 1}} - \frac{2}{n}}, \tag{2.3.3}$$

for the parallel layout. The standard error in (2.3.2) is lower than the error in (2.3.3) if $y_{11}y_{22} > y_{12}y_{21}$. Details on the derivation of paired and parallel variances can be found in the Appendix. Next, we provide four examples that illustrate the proposed methodology.

Example 1. We start with the fully observed paired table reported by Forss [40]

$$\boldsymbol{t} = \begin{array}{|cc||c|} \hline 8 & 8 & 16 \\ 15 & 66 & 81 \\ \hline\hline 23 & 74 & 97 \\ \hline \end{array} \tag{2.3.4}$$

and pretend for a moment that only marginal totals (16, 81) and (23, 74) are available.

Next, we set a Metropolis random walk over $\mathcal{Y}$ with the acceptance probability given as

$$\rho = \min\left\{1, \left(\frac{p_{11}p_{22}}{p_{12}p_{21}}\right)^{\pm 1} \prod_{i,j} \frac{y_{ij}!}{y'_{ij}!}\right\}, \tag{2.3.5}$$

where, as before, y_{ij} and y'_{ij} are entries of tables $\boldsymbol{y}$ and $\boldsymbol{y}\,'$. The parameter $\boldsymbol{p}$ is given Dirichlet prior with the probability vector $\boldsymbol{q}$ as (0.05, 0.1, 0.1, 0.75) and the prior sample size s is half of the observed size of the table.

The chain can start at any element of $\mathcal{Y}$, say table $\begin{array}{|cc|} \hline 0 & 16 \\ 23 & 58 \\ \hline \end{array}$. In Figure 2, the upper panel shows the MCMC evolution of the pivot cell y_{11}. Given the marginal counts, y_{11} is sufficient to recover the full table. The bottom panel in Figure 2 gives

the frequencies of the chain's visits to states of y_{11} from which posterior measures can be estimated. The posterior mean of y_{11} is 7.61, while the posterior mode and median are 8 which coincide with the original cell count. The 95% credible set for y_{11}, based on 10,000 simulations, is [4, 11].

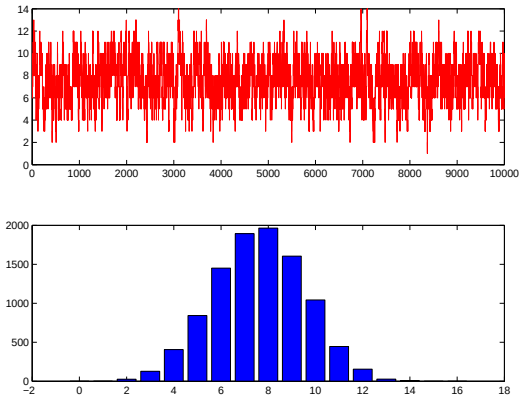


Figure 2: Posterior samples and distribution of the pivotal cell y_{11}

Example 2. If the Markov chain walk is generated over the space of tables with the same marginals as Forss [40], then it is instructive to look at the behavior of log-risk-ratio variances. When conditional multinomial sampling over space $\mathcal{Y}$ is applied, 98% of tables visited by the chain have variance of the effect size smaller than the variance obtained from the marginals only, as in (3.3).

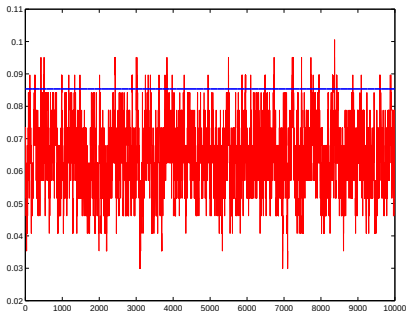


Figure 3: Variance of the log-risk-ratio for the tables from $\mathcal{Y}$ for 10,000 samples. The log-risk-ratio variance for the parallel table (solid line) is provided for comparison.

In the following examples, we start from a list of completely reported paired tables, and report a single table as parallel. Using the Metropolis random walk model proposed in this chapter with relatively non-informative prior, we recover the paired table. Performance of the method is easily assessed by comparing the recovered entries with those from the original tables.

Example 3. Randomized split mouth (RSM) studies

A recent Cochrane systematic review on the effectiveness of dental sealants included meta-analyses comparing resin-based (RB) sealants to both glass ionomer (GI) sealants and to no sealant. The meta-analysis comparing RB and GI sealant materials found that there were an insufficient number of studies to estimate a summary effect measure [22]. This review, however, omitted studies of randomized split-mouth design that did not report full paired tables. We apply our methodology to the 3 studies included in this meta-analysis and compare the resulting estimated paired tables to the actual paired tables. For the Cochrane Review meta-analysis of RB versus no sealant, we apply our methodology to the 9 included studies. The Cochrane Collaboration conducted meta-analyses on studies of RSM design for each year of follow-up, and the example is based on that structure. The Metropolis random walk simulation is implemented by MATLAB with the mean of Dirichlet prior as $s \times \boldsymbol{q}$ where $s = n/2$ and $\boldsymbol{q} = (q_{11}, q_{12}, q_{21}, q_{22})$. The simulation is iterated 100,000 times after a burn-in of 10,000 iterations.

Table 4: GI vs. RB sealants (3 years follow-up), $\boldsymbol{q} = (0.06\ 0.10\ 0.10\ 0.74)$, $s = n/2$

	Original Data				Metropolis RW			
Study	y_{11}	y_{12}	y_{21}	y_{22}	y_{11}	y_{12}	y_{21}	y_{22}
Poulsen 2001	7	37	6	156	7	37	6	156
Kervanto-Seppälä 2008	2	25	5	625	1	26	6	624
Arrow 1995	3	3	28	378	2	4	29	377

Table 5: RB vs. No sealant (1 year follow-up), $\boldsymbol{q} = (0.02\ 0.06\ 0.06\ 0.86)$, $s = n/2$

	Original Data				Metropolis RW			
Study	y_{11}	y_{12}	y_{21}	y_{22}	y_{11}	y_{12}	y_{21}	y_{22}
Bojanini 1976	2	6	79	188	5	3	76	191
Charbeneau 1979	11	5	82	104	12	4	81	105
Sheykholeslam 1978	3	2	49	132	3	2	49	132

Table 6: RB vs. No sealant (2 years follow-up), $\boldsymbol{q} = (0.04\ 0.08\ 0.08\ 0.80)$, $s = n/2$

	Original Data				Metropolis RW			
Study	y_{11}	y_{12}	y_{21}	y_{22}	y_{11}	y_{12}	y_{21}	y_{22}
Brooks 1979	22	3	64	144	18	7	68	140
Charbeneau 1979	29	4	100	53	30	3	99	54
Sheykholeslam 1978	10	1	79	85	9	2	80	84

Table 7: RB vs. No sealant (3 years follow-up), $\boldsymbol{q} = (0.06\ 0.10\ 0.10\ 0.74)$, $s = n/2$

	Original Data				Metropolis RW			
Study	y_{11}	y_{12}	y_{21}	y_{22}	y_{11}	y_{12}	y_{21}	y_{22}
Brooks 1979	23	4	63	111	20	7	66	108
Charbeneau 1979	47	5	96	45	47	5	96	45
Hunter 1988	35	9	163	302	32	12	166	299

Example 4. Cross-over trials

Cross-over trial designs are widely used in many branches of medicine, such as clinical pharmacology, pediatrics, cancer, or schizophrenia [21]. In contrast to a parallel group trial, each individual in a cross-over trial receives two or more treatments but in a random order. A meta-analysis conducted in 1997 comparing insulin lispro versus regular insulin included 8 large randomized clinical trials [39]. Five of the trials were cross-over trials and 3 were of parallel group design. The analysis treated the 5 cross-over trials as if they were of parallel group design when they pooled the 8 studies. We apply our methodology to the 5 cross-over trials and compare the resulting estimated paired tables to the actual paired tables.

Table 8: Insulin Lispro vs. Regular Insulin, $\boldsymbol{q} = (0.01\ 0.03\ 0.03\ 0.93)$, $s = n/2$

	Original Data				Metropolis RW			
Study	y_{11}	y_{12}	y_{21}	y_{22}	y_{11}	y_{12}	y_{21}	y_{22}
Anderson 1997b	7	17	26	927	6	18	27	926
Vignati 1997	3	8	9	353	2	9	10	352
Rowe 1996	1	1	4	82	1	1	4	82
Holleman 1997	2	5	12	174	3	4	11	175
Holcombe 1997	4	9	14	431	3	10	15	430

2.4 Application to meta-analysis of dental sealants effectiveness

The methodology described in this chapter is applied to the meta-analysis of split-mouth designs assessing the effectiveness of glass-ionomer (GI) to resin-based (RB) sealants in preventing caries in juvenile permanent molars. Although GI to RB sealant meta-analyses have been performed in the past, the evidence for the effectiveness of GI sealants relative to RB sealants was inconclusive. In this section, we reexamine the effectiveness of GI sealants to RB sealants using data from both split mouth studies included in the Cochrane Review [22] and split mouth studies that were excluded solely because they failed to report data in a paired manner.

2.4.1 Data Description

Two commonly used dental sealant materials are resin-based (RB) and glass ionomer (GI). Resin-based sealants are recommended for use in clinical and school settings because of strong evidence of effectiveness [41, 42]. A Cochrane Review of the effectiveness of different sealant materials in preventing caries in permanent molars (8 studies, 6 of split mouth design and 2 of parallel group design) found study results so divergent they did not conduct a meta-analysis. And two other reviews [43, 44] found no statistical difference between GI and RB sealants in their meta-analyses.

It should be noted that after the Cochrane Review stratified findings from the 6 split-mouth studies by year since sealant placement there were very few studies from which to estimate a summary measure. For example, there were no studies 1 year after placement and only 1 study with usable data 2 years after placement. Some well designed split mouth studies [45, 46, 47, 48, 49, 50] were not included in the original review because they failed to report a paired variance or full paired table [43]. Failure to appropriately report statistical findings for paired data is not unique to sealant studies and methods exist for estimating the paired variance from such trials if the dependency of the data can be estimated [21, 51].

To estimate the paired variance for the split mouth studies excluded from the Cochrane Review, we use the Metropolis random walk and a hierarchical Bayes model described in this chapter to retrieve the full paired table from the marginal totals. For studies where there were no caries in the control group (RB), the denominator of the risk ratio (RR) was 0 and thus the RR was undefined. While the Cochrane Review deemed the RR inestimable in such studies, we added 0.5 to each cell count as recommended by Agresti [52] to estimate the RR.

2.4.2 Meta-analysis

The RR and its 95% confidence interval are calculated from the adjusted data of each study. The variation across studies (heterogeneity) was evaluated with both the chi-squared test statistics Q and the quantity I^2. Two random effects models, DerSimonian-Laird (DSL) which is one of the frequentist methods and full Bayesian, are used to incorporate the heterogeneity between studies into the meta-analysis. Frequentist (DSL) and Bayesian models differ in how they estimate between-study variation (Δ^2) and in how their results are to be interpreted.

The Bayesian meta-analysis model used in this chapter differs from the traditional

Bayesian model; we place a prior distribution on the relative cell probability instead of the summary measure itself [53]. From Table 10, the multinomial distribution can be used to model outcomes of RSM studies

$$\boldsymbol{y} \sim \mathcal{M}ultinomial(\boldsymbol{n}, \boldsymbol{p}) \tag{2.4.1}$$

where $\boldsymbol{n} = y_{11} + y_{12} + y_{21} + y_{22}$ is the study size and $\boldsymbol{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ is the vector of corresponding cell probabilities.

Table 9: Randomized split mouth design

		RB		
		Caries	No Caries	Total
GI	Caries	y_{11}	y_{12}	$y_{1\cdot}$
	No Caries	y_{21}	y_{22}	$y_{2\cdot}$
Total		$y_{\cdot 1}$	$y_{\cdot 2}$	n

As a conjugate prior for p, we use Dirichlet prior

$$\boldsymbol{p} \sim \mathcal{D}irichlet(\boldsymbol{\alpha}) \tag{2.4.2}$$

where $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$. Here, $\boldsymbol{\alpha}$ represents the hyper-parameter of the Dirichlet prior. We utilize non-informative gamma priors on the hyper-parameter $\boldsymbol{\alpha}$ and take each component to be smaller than the average study size of the corresponding year, m; α_{11} to be smaller than $m/10$, α_{12} and α_{21} to be smaller than $m/5$, and α_{22} to be smaller than $m/2$. Non-informative prior on $\boldsymbol{\alpha}$ implies that GI and RB sealants do not differ in their effectiveness a priori. The posterior mean of $\boldsymbol{\alpha}$ is used as a measure of the Bayesian meta-analysis. The pooled risk ratio is then defined as $\frac{\alpha_{11}+\alpha_{12}}{\alpha_{11}+\alpha_{21}}$. Although the conjugate prior is adopted, the posterior distribution is not analytically tractable. We use Markov chain Monte Carlo (MCMC) method implemented by OpenBUGS v 3.0.2. The simulation starts at an initial value of $\boldsymbol{\alpha}$ and is iterated 100,000 times after a burn-in of 10,000 iterations. As a result our Bayesian model does not require the addition of 0.5 to cell counts of 0 to estimate RR.

2.4.3 Result

Heterogeneity was present in years 2 and 3 (Table 10). Adding studies that did not report findings for paired data did not change the direction of the pooled RR (it was always greater than 1) but did result in the pooled RR becoming statistically significant at year 2 (Table 10). At year 2 the DSL and Bayesian summary RR were respectively, 2.237 (95%CI: 1.141-4.387) and 2.149 (95%CI: 1.124-3.957). The DSL and Bayesian models yielded very similar RRs. Both found that at 2 years RB was superior to GI and no difference in the other years. Although 2 other meta-analyses examined the effectiveness of GI to RB sealants, this meta-analysis is the first to find a difference. Because we have no evidence that trials were sufficiently powered and because we did find a difference at year 2, we would argue that in settings where it is possible to keep surfaces dry, RB should be recommended over GI sealant. The results are summarized in Table 10, Figure 4 and 5.

2.5 Discussion

We demonstrated a Bayesian approach to the ecological-type inference in matched-pair studies when only fixed marginal totals are available. The methodology is based on the Metropolis random walk and hierarchical Bayes model and it incorporates any prior information available. The performance of the model is remarkable; from Examples 3 and 4, four of the estimated tables exactly coincided with the original tables. Even with a fairly non-informative prior, nine of the estimated tables deviated only ± 1 from the original tables in terms of the pivotal cell size y_{11}.

Developing a methodology to retrieve paired data from matched paired design studies that do not report paired tables is an important issue in meta-analyses. A study of methodological issues in cross-over trials [21] reported that presenting data as paired tables was uncommon in the published work. This same study also searched

Table 10: Adjusted data and the result of meta-analyses

Study	Both sound (+)	GI+ RB-	GI- RB+	Both carious (-)	Risk Ratio (95% CI)	Het. Chi^2 (p-value)	I^2
1 YEAR							
Ganesh 2006	100.5	0.5	0.5	0.5	1.000 (0.141, 7.099)		
Mills 1993	59.5	0.5	0.5	0.5	1.000 (0.141, 7.099)		
POOLED (DSL)					1.000 (0.250, 3.999)	0.000 (1.000)	0%
POOLED (Bayesian)					1.113 (0.343, 2.674)		
Karlzen-Reuterv 1995	73.5	1.5	0.5	0.5	0.500 (0.070, 3.550)		
Rock 1996	151	0	6	1	7.000 (1.140, 42.971)		
Raadal 1996	133.5	0.5	3.5	0.5	4.000 (0.563, 28.397)		
Sipahier 1995	79	1	2	4	1.200 (0.646, 2.230)		
POOLED (DSL)					1.434 (0.781, 2.634)	5.642 (0.223)	11%
POOLED (Bayesian)					1.593 (0.640, 3.456)		
2 YEARS							
Ganesh 2006	100.5	0.5	0.5	0.5	1.000 (0.141, 7.099)		
Mills 1993	59.5	0.5	0.5	0.5	1.000 (0.141, 7.099)		
Poulsen 2001	191	2	9	1	3.333 (1.017, 10.922)		
POOLED (DSL)					2.003 (0.813, 4.934)	1.673 (0.433)	0%
POOLED (Bayesian)					1.362 (0.404, 3.642)		
Forss 1994	141	2	2	5	1.000 (0.571, 1.751)		
Karlzen-Reuterv 1995	72	1	0	1	0.500 (0.125, 1.999)		
Raadal 1996	131.5	0.5	0.5	0.5	6.000 (0.845, 42.596)		
Rock 1996	116	0	14	2	8.000 (2.188, 29.249)		
Williams 1996	273	1	16	5	3.500 (1.704, 7.190)		
POOLED (DSL)					2.237 (1.141, 4.387)	20.696 (0.008)	61%
POOLED (Bayesian)					2.149 (1.124 3.957)		
3 YEARS							
Arrow P 1995	378	28	3	3	0.194 (0.087, 0.431)		
Kervanto-Seppälä 2008	625	5	25	2	3.857 (1.767, 8.422)		
Poulsen 2001	156	6	37	7	3.385 (1.978, 5.793)		
POOLED (DSL)					1.379 (0.233, 8.170)	38.800 (0.000)	95%
POOLED (Bayesian)					1.300 (0.482 2.896)		
Karlzen-Reuterv 1995	70	3	1	0	0.333 (0.035, 3.205)		
Raadal 1996	122.5	0.5	10.5	0.5	11.000 (1.549, 78.093)		
Rock 1996	105	1	21	3	6.000 (2.348, 15.333)		
POOLED (DSL)					2.004 (0.590, 6.808)	49.540 (0.000)	90%
POOLED (Bayesian)					1.834 (0.824, 3.662)		
$\geq$ 4 YEARS							
Forss 1994	66	8	15	8	1.438 (0.881, 2.346)		
Williams 1996	189	11	17	5	1.375(0.791, 2.390)		
POOLED (DSL)					1.410 (0.977, 2.034)	0.014 (0.906)	0%
POOLED (Bayesian)					1.456 (0.590, 3.093)		

the Cochrane Controlled Trials Registry and found that 8% of trials included the term "cross over" in the title or abstract. Another search of the Cochrane Database of Systematic Reviews found that 184 reviews (18%) contained a free text term referring to cross-over trials. Further examination of these reviews found that over 50% of

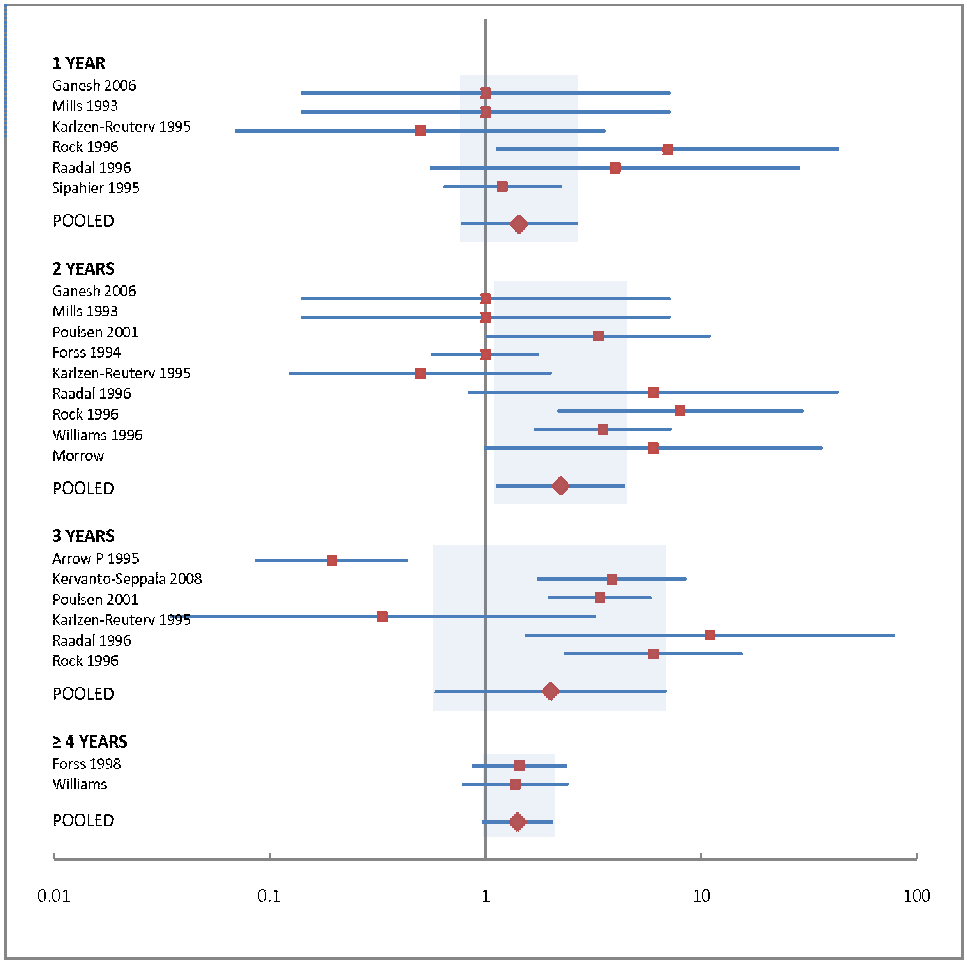


Figure 4: DerSimonian-Laird Meta-analysis

reviews either treated cross-over trials as if they were parallel groups or only used data from the first time period assuming that in doing so it would better approximate a parallel group. If the Cochrane systematic review meta-analysis of resin-based versus glass-ionomer sealants had included RSM trials that did not report data as paired tables, the number of studies would have increased from 2 to 6 in year 1, from 3 to 9 in year 2, and from 3 to 6 in year 3.

For the Metropolis algorithm, the prior information is incorporated via the mean of the Dirichlet prior. Two modifications of the algorithm are possible: (i) one may consider placing the prior on the parameter θ directly; a reasonable support for such

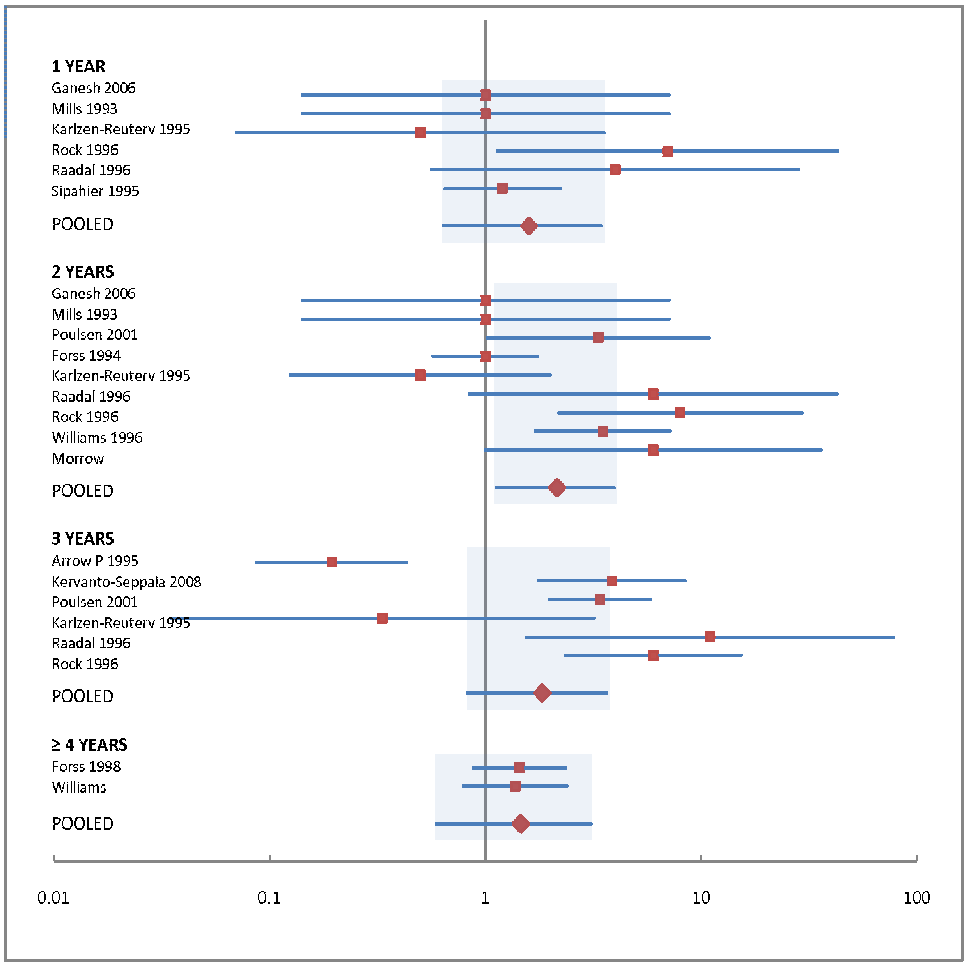


Figure 5: Bayesian Meta-analysis

priors is the set of non-negative θ's since the pairing is well justified when the association is positive, and (ii) it is possible to introduce an additional level of hierarchy in the model to incorporate prior information by a hyper-prior placed on the mean of the Dirichlet prior. The Metropolis random walk from Section 2 is expanded to paired tables of size $k \times k$ in a straightforward manner. The corresponding parallel table in that case is of size $2 \times k$, the elements from the Markov basis are of size $k \times k$ with four non-zero elements consisting of 1 and -1 arranged in such a way to preserve marginal sums when added. The Metropolis acceptance criterion is readily generalized as well.

CHAPTER III

BAYESIAN AUTOLOGISTIC EXCESS-ZERO MODEL WITH APPLICATION TO SPATIAL REPEATED MEASURES ZERO-INFLATED DATA

3.1 Introduction

It is common to observe excessive number of zeros in count data analysis. For example, in estimating US county level natural fluoride concentration in the ground water, significant number of counties have no natural fluoride occurrence and therefore generate excessive zeros. More examples can be found in ecology[82, 55], health [56, 57], manufacturing [58] and psychology [59, 60]. Especially, many environmental, epidemiology, disease mapping applications yield zero inflated data in a spatial context [61, 62, 63].

Several statistical methods to model count data with a large number of zeros have been proposed. The use of discrete mixture models is the most common approach. For example, hurdle models [64, 65] consist of a point mass at zero and a truncated count model for the nonzero part. Zero-inflated models [58] are a mixture of a discrete distribution that explains excessive zeros and a regular untruncated count model. In other words, a zero-inflated model separate zeros into 'structural' zeros (event does not occur in nature) and 'chance' zeros (event does occur but observed to be zero by statistical randomness). Although the distinction between 'structural' and 'chance' zeros seems reasonable in theory, how to distinguish zero observations is rather a practical issue.

Once a model is selected, associated parameters can be estimated from independent variables. Classical (frequentist) solutions to excess-zero models are based on

maximum likelihood estimate (MLE) and the likelihood ratio (LR) test. However, in the reality, models could have multiple modes [67]. Also, the large sample approximation theory which serves as the basis for classical inference for non-normal data often requires the use of nonstandard asymptotic theory. As an alternative, several authors have proposed a Bayesian approach; it does not require closed form analytical solutions, explicitly elicits prior beliefs, shows good finite sample performance and has the ability to incorporate unknown variability. More on Bayesian solutions to excess zero count models are discussed by Angers and Biswas [67], Ghosh et al. [68], Neelon et al. [69].

Parameter estimation usually involves a set of independent variables. For instance, the probability of structural zero π_0 and the Poisson parameter λ in ZIP model can be estimated using logit and log-log link. Lambert [58] suggested a zero-inflated(τ) model:

$$\log(\lambda) = X\beta, \quad \text{logit}(\pi_0) = \tau(X\beta)$$

where X is the independent variable and β is parameters of interest. However, when there is no explanatory variable available, the estimates become questionable. In a spatial context, it is possible that there is no information other than the neighborhood structure.

In this paper, we suggest a Bayesian approach to model spatial repeated measure count data with excessive zero when there is no covariate information available. The model comprises of two parts; we first estimate the probability of zero observation (both structural and chance zeros) using an autologistic model and use this probability as a basis to model the underlying spatial phenomenon of the excess zero data. To demonstrate the proposed method and evaluate the performance, we use the natural fluoride concentration data in the ground water of the counties in Georgia.

3.2 Background

In this section, we briefly describe the autologistic model, hurdle and zero-inflated model which serve as the basis of our approach. The autologistic model is commonly used to illustrate spatial binary data on the lattice. Hurdle and zero-inflated models are the most popular approaches in modeling count data with excessive zeros.

3.2.1 Autologistic Model

Besag [70] first suggested modeling spatially-distributed variables via a conditioning on neighborhoods. Instead of considering the full conditional distributions within a likelihood function, a likelihood simply based on a product of the local conditioning within a neighborhood was proposed. The most common and widely used example is with binary outcome $\{Z_i\}$. Let θ_i be the parameter associated with Z_i and δ_i be the neighborhood of the ith small area. Then the model becomes,

$$P(Z_i|\{Z_j\}_{j\in\delta_i}) = \theta_i^{Z_i}(1-\theta_i)^{1-Z_i} \tag{3.2.1}$$

and

$$\theta_i = \frac{\exp\{\alpha_i + \lambda S_{\delta_i}\}}{1+\exp\{\alpha_i + \lambda S_{\delta_i}\}}, \quad S_{\delta_i} = \sum_{j\in\delta_i} Z_i\,. \tag{3.2.2}$$

Since Z_i is a binary variable, $\sum_{j\in\delta_i} Z_i$ is the total number of 'observations of interest' in the neighborhood. For the lattice data, δ_i is usually defined as a set of lattice that shares a borderline with the ith small area. Figure 6 shows the example of Fulton county (Georgia) and its neighborhood counties.

The likelihood is then

$$L(Z_i|\theta_i) = \prod_i \frac{[\exp\{\alpha_i + \lambda S_{\delta_i}\}]^{Z_i}}{1+\exp\{\alpha_i + \lambda S_{\delta_i}\}} \tag{3.2.3}$$

In estimating the probability of zero observations, we treat Z_i as follows: $Z_i = 1$ if we observe zero from the ith small area and $Z_i = 0$ if we observe non-zero value. Then

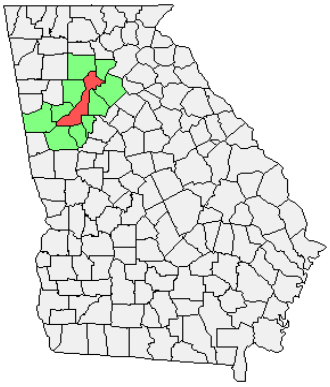


Figure 6: GA Fulton county (red) and its neighborhood counties δ_i (green)

θ_i represents the probability that the observation at ith small area is zero, which can be estimated from (3.2.2) and (3.2.3).

3.2.2 Hurdle Model

In the hurdle model, we assume that non-zero observations (counts of one, two or more) occur from crossing a threshold or hurdle [64]. The probability of crossing this hurdle involves a binary sampling model, while the sampling of non-zero counts involves a truncated count model such as truncated Poisson or truncated binomial (sampling confined to values y above zero).

Let f_1 and f_2 be probability densities for binary sampling and non-zero counts sampling, respectively. Then the probability of the two stages is given by

$$f(0) = f_1(0) \tag{3.2.4a}$$

$$f(k) = \frac{1 - f_1(0)}{1 - f_2(0)} f_2(k), \quad k = 1, 2, ... \tag{3.2.4b}$$

In fact, $\frac{1}{1-f_2(0)} f_2(k)$ is a truncated distribution of f_2. If $f_1(0) > f_2(0)$, it is zero inflated (excessive zeros) whereas $f_1(0) < f_2(0)$ implies zero deflation.

For example, one can choose f_1 to be a Bernoulli distribution with $f_1(0) = \pi_H$ and f_2 be a Poisson distribution with mean λ. Then the likelihood of the Poisson-hurdle

model for the count observation Y has the following form:

$$P(Y=0) = \pi_H, \;\; 0 \leq \pi_H \leq 1 \tag{3.2.5a}$$

$$P(Y=k) = \frac{1-\pi_H}{1-e^{-\lambda}} \frac{\lambda^k e^{-\lambda}}{k!}, \;\; \lambda > 0, \; k = 1, 2, ... \tag{3.2.5b}$$

Let Z be the latent random variable sampled from f_2. The mean and variance of the hurdle model is then

$$\mathbb{E}(Y) = \frac{1-f_1(0)}{1-f_2(0)} \mathbb{E}(Z) \tag{3.2.6}$$

$$\mathbf{Var}(Y) = \frac{f_1(0)-f_2(0)}{1-f_1(0)} [\mathbb{E}(Y)]^2 + \gamma \mathbb{E}(Y) \tag{3.2.7}$$

where $\gamma = \frac{\mathbf{Var}(Z)}{\mathbb{E}(Z)}$ denotes the coefficient of dispersion of the latent random variable Z. A detailed derivation is provided in the Appendix.

3.2.3 Zero-inflated Model

Under zero-inflated densities for count data, zero counts may result from two processes: they may be either true zeroes or result from a stochastic mechanism. Another terminology is structural vs. random zeroes [82]. For example, in estimating the natural fluoride level, a structural zero corresponds to no natural fluoride occurrence whereas random zero implies that the fluoride does occur in nature but observed to be zero due to the randomness. The random mechanism could be described by a Poisson or negative binomial density. Then the zero-inflated model can be written as a two part mixture model consisting of a degenerate distribution at zero with an untruncated distribution.

Let π_Z be the probability of structural zero, then the likelihood for the zero-inflated model can be written as follow:

$$f(0) = \pi_Z + (1-\pi_Z) f_D(0), \;\; 0 \leq \pi_Z \leq 1 \tag{3.2.8a}$$

$$f(k) = (1 - \pi_Z) f_D(k), \;\; k = 1, 2, ... \tag{3.2.8b}$$

where f_D is the underlying distribution which describes the random mechanism.

For example, the zero-inflated Poisson model has the following form:

$$P(Y = 0) = \pi_Z + (1 - \pi_Z) e^{-\lambda}, \;\; 0 \leq \pi \leq 1 \tag{3.2.9}$$

$$P(Y = k) = (1 - \pi_Z) \frac{\lambda^k e^{-\lambda}}{k!}, \;\; \lambda > 0, \, k = 1, 2, ... \tag{3.2.10}$$

If the probability of structural zero, π_Z, is greater than zero, then the zeros are inflated. Unlike hurdle model, zero-inflated model does not allow zero deflation. In this sense, we can consider zero-inflated model as a special case of hurdle model.

The mean and variance of the zero inflated variable Y are given by,

$$\mathbb{E}(Y) = (1 - \pi_Z) \mathbb{E}(Y_D) \tag{3.2.11}$$

$$\mathbf{Var}(Y) = \frac{\pi_Z}{1 - \pi_Z} [\mathbb{E}(Y)]^2 + \gamma \mathbb{E}(Y). \tag{3.2.12}$$

Again, $\gamma = \frac{\mathbf{Var}(Y_D)}{\mathbb{E}(Y_D)}$ is the coefficient of dispersion of the latent random variable Y_D. If $\gamma \geq 1$, i.e., the latent variable Y_D is not under-dispersed, then the distribution of Y has a over-dispersed distribution. On the other hand, if $\gamma < 1$ and Y_D is under-dispersed, then Y has a under-dispersed distribution if and only if $\mathbb{E}(Y_D) < \frac{1-\gamma}{\pi_Z}$ [68].

3.2.4 Example of Excess-zero Data Analysis

For the purpose of illustration, we take an excess-zero data example from Bohning et al [71]. This paper demonstrates count data modeling from a dental epidemiological study in Belo Horizonte that evaluated the effectiveness of school programs in reducing caries. Six interventions were randomized to six schools, and all children of a given school received the same treatment. 797 school children were examined both before and after the trial, and their dental status was evaluated by the number of decayed,

Table 11: Intervention method for Belo Horizonte study

School 1	Oral health education
School 2	All four methods together
School 3	Control group
School 4	Enrichment of the school diet with rice bran
School 5	School mouthwash with 0.2% sodium fluoride (NaF) solution
School 6	Oral hygiene

Table 12: Data description for Belo Horizonte study

Response variable	DMFT index at the end of the study
Explanatory variables	DMFT index at the beginning of the study Gender of the school child (0-female; 1-male) Ethnic group (1-dark; 2-white; 3-black) School index

missing and filled teeth (DMFT index). The detailed information for the intervention method and reported data are summarized in Table 11 and Table 12.

However, excessive zeros in the response variable violate the variance-mean relationship of the Poisson distribution. Figure 7 represents the histogram of the response variable. Its homogeneous Poisson fit (red line) clearly shows that there is a zero-inflation, which implies that a large portion of students had zero number of decayed, missing or filled teeth at the end of the study.

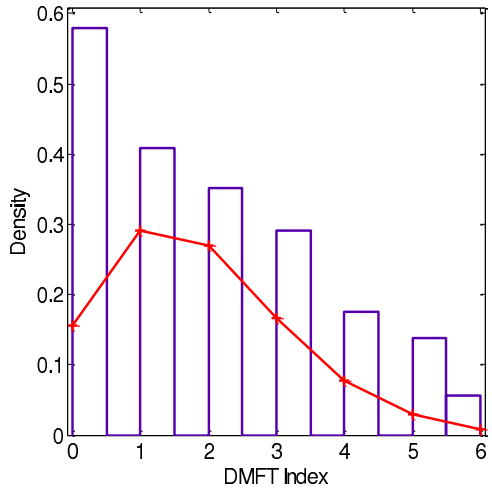


Figure 7: Histogram of the DMFT index at the end of the Belo Horizonte study

Now let us analyze this data using two excess-zero models described in Sections 2.2 and 2.3. Under the hurdle model, the probability of crossing the hurdle $(1 - \pi_H)$ is the probability that an individual would have at least one decayed missing or filled tooth at the end of the trial. Only after a child crosses this hurdle, (s)he would have positive number of DMFT index which then will be analyzed using (3.2.2). On the other hand, the zero-inflated model assumes that the children with zero DMFT index can be categorized into two groups: those who have zero risk of caries/cavities and those who have low risk and they happen to have zero DMFT index at the end of study due to the randomness.

3.3 Model Description

This section describes the autologistic excess-zero model and its implementation in a full Bayesian scheme. The model comprises two parts: (1) estimate the probability of zeros using the autologistic model and (2) use the estimated probability as a basis of a hurdle or zero-inflated model. In the following analysis we assume that no other covariate is available except for the neighborhood information. The model can easily be extended by including covariates as linear predictors of the corresponding link function.

We first introduce a latent binary random variable Z_i, which is equal to one if the ith observation has value zero. We assume

$$Z_i \sim \mathcal{B}ernoulli(\theta_i) \tag{3.3.1}$$

where θ_i is the probability of zero observation. To estimate θ_i, we adopt the convolution model that incorporates spatially correlated randomness and random errors [72]. Assuming a logit link function,

$$\text{logit}(\theta_i) = \alpha + u_i + v_i \tag{3.3.2}$$

where $u_i \sim \mathcal{N}(0, \sigma_u^2)$ is the uncorrelated heterogeneity and v_i is the spatially correlated heterogeneity. If we assume $v_i | v_{j\{\forall j \neq i\}} \sim \mathcal{N}(\sum_{j \in \delta_i} \frac{1}{N_i} v_j, \frac{\sigma^2}{N_i})$ with N_i be the number of neighborhood for ith small area, then it becomes the conditional autoregressive (CAR) model. Once the posterior distribution of θ is obtained, we take the average or median of the posterior distribution as an estimated probability of zero observation, $\hat{\theta}_i$.

For the hurdle model, $\hat{\theta}_i$ can be set equal to π_{H_i} in (3.2.2). In case of the zero-inflated model, the probability of zero observations is the sum of the structural zero and random zero probabilities:

$$\hat{\theta}_i = \widehat{\pi}_{Z_i} + (1 - \widehat{\pi}_{Z_i}) f_D(0) \tag{3.3.3}$$

Therefore, the probability of structural zeros given $\hat{\theta}_i$ is

$$\widehat{\pi}_{Z_i} = \frac{\hat{\theta}_i - f_D(0)}{1 - f_D(0)} \tag{3.3.4}$$

When the underlying distribution f_D is Poisson, the estimate becomes

$$\widehat{\pi}_{Z_i} = \frac{\hat{\theta}_i - e^{-\lambda}}{1 - e^{-\lambda}}. \tag{3.3.5}$$

Once we obtain $\hat{\theta}_i$, next part of the analysis is straightforward. More details and actual implementation of the model with repeated measures will be described in the following section.

3.4 Analysis of the Natural Fluoride Concentration Data

Water fluoridation, the addition of fluoride to drinking water to prevent tooth decay, is listed as one of the ten greatest public health achievements of the 20th century by the Centers for Disease Control and Prevention (CDC). However, fluoride can occur naturally in water in concentrations well above recommended levels, which can

have several long-term adverse effects including dental fluorosis, skeletal fluorosis, and weakened bones.

Although describing the availability of natural fluoride by county can inform public health efforts in prevention of tooth decay and dental fluorosis, the effort has been limited to few states. The county-level estimates, however, enable policy practitioners to assess the county-level risk of high natural fluoride occurrence and hence can improve the water fluoridation policy. It is also valuable information to residents so they can be aware of how much natural fluoride is present in their county. Furthermore, this map can be a useful asset for future research; one of the nation's most comprehensive public health databases, National Health and Nutrition Examination Survey (NHANES) provides survey and examination results by county. Therefore, persistent research questions in the field such as the possible correlation between high natural fluoride occurrence vs. dental fluorosis can benefit from this county-level natural fluoride occurrence map.

3.4.1 Data Description

Natural fluoride level data is extracted from the Water Fluoridation Reporting System (WFRS) on September of 2009. WFRS contains 56,702 public water system (PWS) information across the Unites States and the information includes state and county of each PWS' state and county, ID and name, number of population served, water system type and its respective fluoride level. Among 10 different types of water systems, this analysis is limited to the ground water systems, which provides the estimate of the natural fluoride concentration.

Although the fluoride concentration exhibits a continuous occurrence in nature, the data is reported county level and therefore imposes a change of support issue [73]. Moreover, the fluoride level is reported only up to second decimal places. Therefore

the analysis would be invalid if we treat the reported data as continuous. To accommodate these limitations, we treat the reported data as a discrete lattice data instead of continuous geostatistical data by the following data adjustment procedure: each observation is multiplied by 10 and rounded up which results in a set of discrete integer data. Then each observation is assumed to be a repeated measure for the corresponding county (lattice).

Among 3,219 counties and county equivalents, 1,043 of them have no public ground water system data. For counties with no reported measurement, we take observations from neighborhood counties based on the definition in Figure 6. To reduce the effect from outliers, we discard neighborhood observations that are outside of $\pm 3\sigma_{\delta_i}$ with σ_{δ_i} being the standard deviation of neighborhood observations for county i.

Even after the data adjustment, there are other issues that need to be addressed. First of all, the data have excessive zeros (5,035 out of 20,632 data points), i.e., 24.4% of the observations are zero. Despite all the constraints, there is no covariate information from which we can estimate parameters other than the county neighborhood structure.

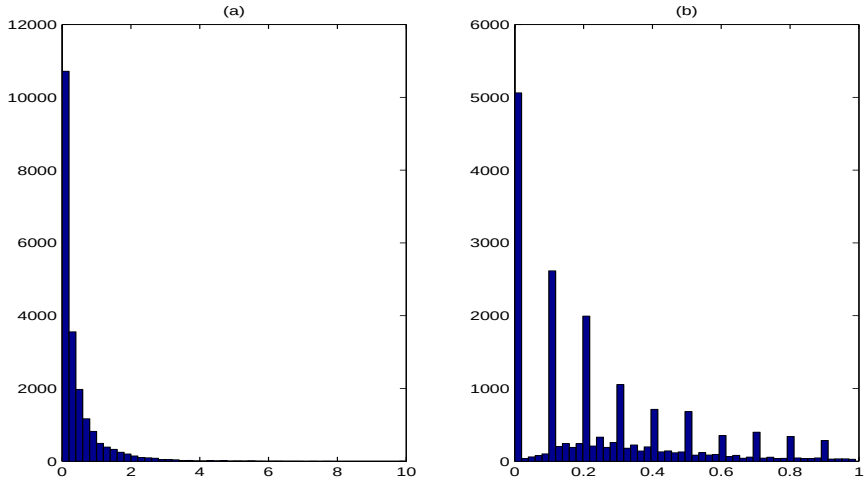


Figure 8: (a) Histogram of the US natural fluoride concentration data. Note the excessive zero density. (b) Histogram of the range [0 1]. See the discrete pattern of the reported data.

3.4.2 Implementation under fully Bayesian scheme

We apply the Bayesian autologistic excess-zero model described in this paper to the Georgia natural fluoride concentration data. It includes 159 counties with 1,051 observations. Some counties have multiple observations (up to 71 data points) while 28 counties have no public ground water system. This analysis can easily be extended to a larger data set that includes multiple states.

3.4.2.1 Estimating the probability of zeros using autologistic model

First we estimate the probability θ_i that county i has zero natural fluoride concentration given single or multiple continuous-type observations. This measurement includes both 'no natural fluoride occurrence (structural zero)' and 'zero observations due to the randomness of the natural fluoride distribution (random zero)'. For example, Decatur county has 10 public ground water systems with reported level 0, 0, 0, 0, 0, 0, 0, 0.24, 0.45 and 0.46. Although the average is 0.115 which is above zero, seven zero measurements implies that θ_i is not simply zero but could be fairly high. Also, we can borrow information from its neighboring counties to accommodate the continuous occurrence of the natural fluoride.

To model the repeated measure for each county, we introduce a latent binary variable Z_{ij} which is equal to 1 if the jth observation in county i is zero. We can also extend this model by employing different threshold value other than zero, to estimate the probability of having natural fluoride occurrence above the threshold. As described in (3.3.1) we assume that each Z_{ij} is generated from Bernoulli distribution with an underlying probability θ_{ij}. This probability is estimated via logit link function that maps (0, 1) interval into the real number domain:

$$\text{logit}(\theta_{ij}) = \beta_{0i} + \beta_{1i}X_{ij} + v_i + u_{ij} \tag{3.4.1}$$

where X_{ij} is the observed value, v_i is spatial effect of county i and u_{ij} is random

error of jth observations in county i. If (3.4.1) and (3.3.2) are compared, then it is apparent that the linear combination of actual observation $\beta_{0i}+\beta_{1i}X_{ij}$ is used instead of constant α. We assume fairly non-informative prior on β_{0i} and β_{1i},

$$\beta_{0i} \sim \mathcal{N}ormal(\mu_{\beta_0}, \sigma^2_{\beta_0})$$

with hyper prios

$$\mu_{\beta_0} \sim \mathcal{N}ormal(0, 100)$$

$$\sigma_{\beta_0} \sim \mathcal{U}niform(0, 100).$$

The same (prior - hyper prior) system is used for β_{1i}. Note that the hyper priors regulate the state level estimate.

Let the posterior estimate be $\hat{\beta_{0i}}$, $\hat{\beta_{1i}}$ and $\hat{v_i}$. Then the posterior probability that we would observe zero observation in county i can be represented as follows:

$$\hat{\theta_i} = \hat{\beta_{0i}} + \hat{\beta_{1i}}\bar{X}_i + \hat{v_i} \tag{3.4.2}$$

where $\bar{X}_i$ is the average of the measurements in county i.

3.4.2.2 Modeling the count data using the hurdle or zero-inflated model

Once the probability of zero-observations is obtained, the count data modeling is straightforward. For the illustration purpose, we choose general Poisson, hurdle-Poisson and zero inflated-Poisson models together with the analogous of negative binomial models. Poisson model is easy to interpret but the assumption of equal mean and variance are inappropriate for over-dispersed data. On the other hand, the negative binomial model is more flexible, although it can over-fit the data and the parameters lack a straightforward interpretation compare to the Poisson case. Therefore, the underlying distribution needs to be carefully chosen based on the characteristic of the data set.

One of the major issues in identifying mixture models using parametric densities is the possibility of facing flat likelihoods, and this issue is especially critical in cases where no independent variable is available. Wasserman[74] states that improper priors yield improper posteriors in mixture models. Moreover, even proper priors may lead to poor posterior solutions, if the prior is too vague and the sample size is small. Therefore, to specify appropriate priors that are objective and also effective in estimation, Congdon[75] suggests the use of mildly informative proper priors based on subject matter knowledge. Thus, instead of vague non-informative prior, we employ semi-informative prior that captures the variation and range of the data set.

Poisson model has one parameter; rate λ_{ij}, which is equal to the tenfold mean occurrence of natural fluoride. Assuming log link,

$$\log(\lambda_{ij}) = \eta_{ij}$$

which maps a positive number into a real number domain, we employ a normal prior on η_{ij},

$$\eta_{ij} \sim \mathcal{N}(\mu_{\eta_i}, \sigma^2_{\eta_i}) \tag{3.4.3}$$

with $\mu_{\eta_i} \sim \mathcal{N}(0, 100)I(-5, 5)$ and $\sigma^2_{\eta_i} \sim \mathcal{G}amma(1, 1)I(0.01, 10)$. The bound index $I(a, b)$ is to constrain each parameter to fall between a and b. Instead of allowing parameters float freely, we constrain them to ensure the convergence of posterior distribution.

The negative binomial model has two parameters, the positive real number r_{ij} and the probability p_{ij}. Note that the tenfold mean occurrence of natural fluoride is equal to $\frac{p_{ij}r_{ij}}{1-p_{ij}}$. We assume a fairly non-informative prior on each parameter,

$$r_{ij} \sim \mathcal{G}amma(\alpha_{r_i}, \kappa_{r_i})$$

$$p_{ij} \sim \mathcal{B}eta(\alpha_{p_i}, \kappa_{p_i})$$

but restrict the range of hyper-priors by imploying Gamma hyper priors

$$\mathcal{G}amma(1,1)I(0.01,10)$$

on α_{r_i}, κ_{r_i}, α_{p_i} and κ_{p_i}. This hierarchical structure allows the county level natural fluoride concentration (prior) to be guided by a state level distribution (hyper-prior).

3.4.3 Model Checking

For complex Bayesian models, exact calculation and inference of posterior may not be feasible. As an alternative, a Markov chain Monte Carlo (MCMC) algorithm draws samples iteratively from the full conditional distributions of the model parameters. Then the posterior quantities of interest can be approximated from the sample.

MCMC algorithm is implemented by the software package WinBUGS [76]. For each model, we run two initially dispersed Markov chain for 60,000 iterations. After discarding the first 10,000 samples as a burn-in to ensure the convergence, we only retain every 50th draw to reduce autocorrelation. Therefore the summary statistics for the parameter values are based on thinned samples of 2,000 iterations. To check the convergence of the MCMC method, deviance from each chain is monitored. MCMC diagnostics, such as trace plots and Brooks-Gelman-Rubin scale reduction statistics [77] indicates that we attain rapid convergence and efficient mixing of the chains.

3.4.4 Model Comparison

Standard model selection criterion such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) comprises of model's relative fit and penalty for the model complexity. Since more complex models usually provide better fit, the penalty term offset gains in model fit due solely to added complexity.

Model complexity is assessed by the number of model parameters. For fixed effect

models, it is easy to determine while for random effect models it is less clear. The deviance information criterion (DIC) is proposed to estimate the number of effective parameters in a Bayesian hierarchical model that includes both fixed and random effects [78]. Moreover, DIC is easily calculated from the samples generated by a MCMC simulation unlike AIC and BIC that require calculating the likelihood and its maximum over the parameter.

Note that the deviance is defined as

$$D(\theta) = -2\log[L(y|\theta)] + C$$

where $L(y|\theta)$ is the likelihood function of the data y with unknown parameters θ, and C is a constant. The posterior mean of the deviance, $\bar{D}(\theta) = E_\theta[D(\theta)]$, measures how well the model fits the data. The smaller $\bar{D}(\theta)$ is, the better the fit. Define $\hat{D}(\theta) = D(E[\theta|y])$, which is the value of $D(\theta)$ evaluated at the posterior mean of the parameter θ. Then the effective number of parameters of the model, i.e., complexity penalty, is computed as $p_D = \bar{D}(\theta) - \hat{D}(\theta)$. The DIC is defined as

$$DIC = \bar{D}(\theta) + p_D.$$

If models differ in DIC by more than three, the one with the smaller DIC is usually considered to have the better fit [78].

We compare six models based on DIC and the result is summarized in Table 13. Poisson models have lower DIC than negative binomial models, which implies that Poisson underlying distribution is more adequate for this data.

3.4.5 Results and Discussion

One of the advantages with the proposed model is that it can answer pragmatic questions like "what is the probability (risk) that a certain county would have natural fluoride occurrence above the EPA recommended level". On January 7th of 2011, the

Table 13: Model Comparison

Model	$\bar{D}$	$\hat{D}$	p_D	DIC
Poisson	230.445	180.086	49.638	280.083
Hurdle-Poisson	221.927	178.760	43.167	265.094
ZIP	218.929	178.712	40.216	259.144
NB	208.710	37.587	171.123	379.833
Hurdle-NB	754.589	537.726	216.862	971.451
ZINB	196.714	28.585	168.129	364.844

U.S. Department of Health and Human Services (HHS) and the U.S. Environmental Protection Agency (EPA) announced a guideline on fluoride in drinking water. The new recommendation is 0.7 milligrams of fluoride per liter of water that replaces the current recommended range of 0.7 to 1.2 milligrams. Therefore it will be of interest to the policy practitioners and the residents to have the map of the estimated risk with the new recommendation level. In our analysis, threshold value is set to be zero to obtain $\hat{\theta}_i$ in (3.4.2).

Since zero inflated Poisson (ZIP) model has the lowest DIC from Table 13, we estimate Georgia county level natural fluoride concentration based on ZIP model (Figure 9). Figure 10 shows the simple average measurement for each county.

Comparing Figure 9 and Figure 10, hierarchical Bayesian structure produces improved discretized estimates. For example, Houston county (salient dark region in the middle from Figure 10) has distinctively high natural fluoride concentration when the estimate are obtained from simple average. It is because simple average does not fully borrow strength from the neighborhood since it only takes the average of repeated measures from the county of interest itself. However, when the autologistic-ZIP model is applied, we obtain smoother estimate for Houston county that includes mid-high level natural fluoride occurrence in neighborhood counties. This is more in lined with the continuous occurrence of the natural fluoride.

Also, by employing the hyper prior structure, the state level estimates can be

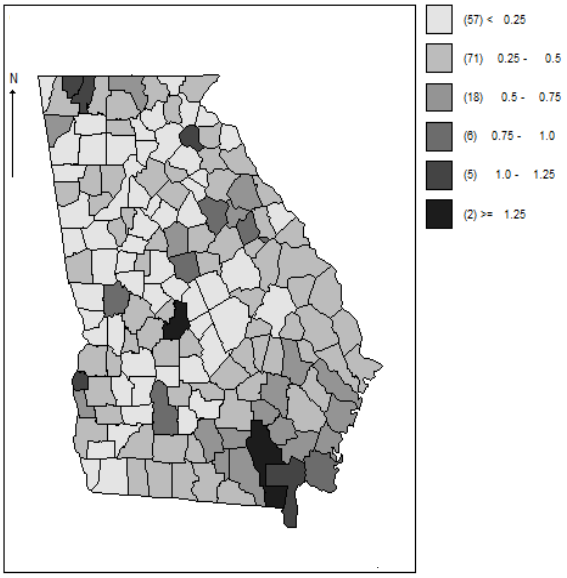


Figure 9: Autologistic-ZIP

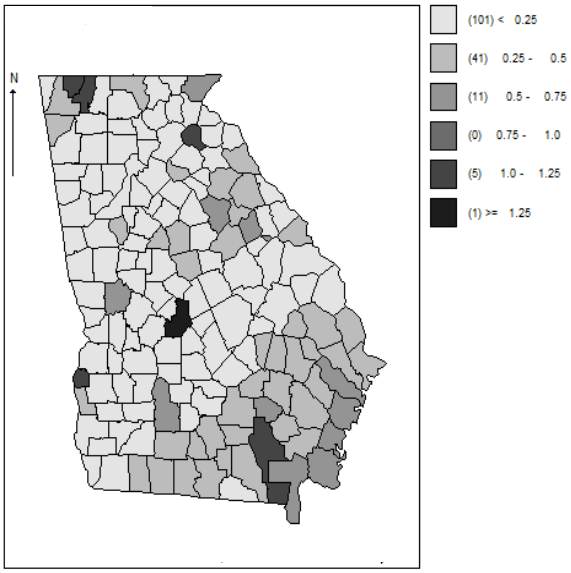


Figure 10: Simple Average

obtained as well. While posterior estimates of β_{0i} and β_{1i} from (3.4.1) and μ_{η_i} from (3.4.3) inform us the county-level information, μ_{β_0} and μ_{β_1} from (3.4.1) provide the state level information.

3.5 Conclusion

We have described a Bayesian approach to fit the spatial repeated measure count data with excessive zeros which is robust without any available independent variables. The model first estimates the probability of zero observations (both structural and chance zeros) from the autologistic model. The estimated probability is then used as a basis of hurdle and zero-inflated model.

For the Georgia county level natural fluoride concentration data, zero-inflated models are more adequate than the hurdle models based on the DIC measure. This result coincides with the geographical intuition since some counties have no natural fluoride occurrence whereas some other counties have very low occurrence with the random chance of zero observation. Also, models with Poisson underlying distribution have constantly lower DIC than negative binomial models.

The model suggested in this paper produce improves county level estimates that reflect the continuous behavior of the natural fluoride occurrence better, compare to the simple average measures. By employing the hyper-prior structure, state level estimates can be obtained as well. The methods discussed here can be extended to semi-continuous two-part models, which assume a continuous (e.g., log-normal or Gamma) rather than Poisson or negative binomial distributions for the non-zero values.

This work is motivated to improve the water fluoridation policy. Although describing the availability of natural fluoride by county can inform public health efforts in prevention of tooth decay and dental fluorosis, the effort has been limited to few states. The county-level estimates will not only be a valuable information to policy practitioners and residents, but also for future research. It is well known that the natural fluoride occurrence is closely related to the geological feature, contact time in the aquifer, climate and chemical composition of groundwater. And it is also assumed

that agricultural and industrial activities can cause high natural fluoride concentration [79]. In this sense, the correlation between man-made factors and high risk of natural fluoride occurrence can be assessed using this county-level map.

CHAPTER IV

DATA REGULARITY INDICES IN COMPLEX WAVELET DOMAIN

Breast cancer is the second leading cause of death in women in the United States. Mammography is currently the most effective method for detecting breast cancer early; however, radiological interpretation of mammogram images is a challenging task. On the other hand, many medical images demonstrate a certain degree of self-similarity over a range of scales which can help us in their description and classification.

In this work, we generalize the scaling-mixing wavelet spectra to the complex wavelet domain. In this domain, we estimate Hurst parameter and phase and use them as discriminatory descriptors to classify mammographic images to benign and malignant. The proposed methodology is tested on a set of images from the University of South Florida Digital Database for Screening Mammography (DDSM).

4.1 Introduction

The National Cancer Institute estimates that 1 in 8 women born today will be diagnosed with breast cancer during her lifetime [80]. Breast cancer is one of the most common forms of cancer among women in the United States, second only to non-melanoma skin cancer. A national objective has been set by the U.S Department of Health and Human Services to reduce the female breast cancer death rate from 22.9 per 100,000 females in 2007 down to 20.6 by the year 2020 – a 10% improvement [81]. One of the most important tools toward that goal is advanced precision of screening

technologies. Early detection is the best method for improving prognosis and also leads to less invasive options for both specific diagnosis and treatment.

Mammography is currently the most efficient and prevalent method for detecting a breast cancer early, before it is substantial enough to feel or cause symptoms. However, the radiological interpretation of mammogram images is a difficult task since the appearance of even normal tissue is highly variable and complex, and signs of early disease are often small or indistinct. Reading a mammogram image is a skill that physicians develop over time, and confidently stating whether findings are cancerous or not is often quite difficult. Suspicious findings are commonly clarified by follow-up images, ultrasound, or MRI. On the other hand, it has been estimated that $10 - 30\%$ of cancers which could have been detected are missed [82]. Thus, improving both the specificity and the sensitivity of mammographic diagnoses is an important goal in improving prognoses while also reducing the number of unnecessary procedures or surgical operations.

In high frequency and irregular data collected in real-life settings (both naturally occurring and human-made), a commonly occurring phenomenon is that of regular scaling. Examples of this have been found in a variety of systems and processes including economics (stock market, exchange rate fluctuations), telecommunications (internet data), physics (hydrology, turbulence), geosciences (wind and rainfall patterns), and several applications in biology and medicine (DNA sequences, heart rate variability, auditory nerve-spike trains). The irregular behaviors of these complex structures are difficult or impossible to quantify by standard modeling techniques; but when observations are inspected at different scales, there is in fact a regular relationship between the behavior at each scale. This phenomenon has been demonstrated in many medical images, leading to the diagnostic use of tools capable of quantifying statistical similarity of data patterns at various scales.

The standard measure of regular scaling is the Hurst exponent. This measure can

also be connected to measures of long memory and fractality in signals and images and is viewed as an informative summary. Many techniques for estimating the Hurst exponent exist, and assessing the accuracy of these estimations can be complicated. Wavelet transforms are powerful tool in estimating the Hurst exponent and modeling statistical similarity at different scales. For example, Nicolis et al. [83] proposed a method based on the wavelet spectra for extracting the self-similarity measures in an isotropic and anisotropic spaces. Ramírez-Cobo et al. [84] demonstrated a wavelet-based spectra method for estimating Hurst exponent in time-varying two-dimensional rainfall maps.

For an efficient representation of the image or signal, the wavelet basis is desirable to be orthogonal, symmetric and to have compact support. [85]. An orthogonal basis has a variety of theoretical and practical advantages: it leads to more efficient algorithms, and establishing properties of a representation is often easier with orthogonal bases. Symmetry guarantees an orientation-free representation of features, preventing distortion when the data is approximated in its basis representation. Moreover, the computational cost of performing wavelet transforms depends heavily on the support size of a basis. Apart from the Haar wavelet, complex wavelets with an odd number of vanishing moments are only compactly supported wavelets which are symmetric [16]. Due to this advantages, complex wavelet has been used in various areas including motion estimation [17], texture image modeling [18], image denoising [19] and NMR spectra classification [20]. However, there is no literature that addresses the complex wavelet-based spectra and its self-similarity measures.

The novelty of this chapter is to use the scale-mixing wavelet spectra based on *complex wavelet transforms* for estimating the Hurst exponent. We then focus on the estimated Hurst exponent and show its ability to differentiate cancerous from normal tissue visible in the backgrounds of mammogram images, and compare this performance with its counterpart obtained from real-valued wavelet transform. Moreover

complex wavelet transform produces an additional measure, phase information. We also demonstrate the classification power of the phase information and use it as an additional modality in the discriminatory analysis.

A further novelty of our work is the use of the information contained in the background tissue of images. Most of the references found in literature dealing with breast cancer detection methods are based on microcalcifications [86, 87, 106, 89]. Only recently the information contained in the background is taken into consideration [83, 90]. This classifying measure based on background tissue would be a new tool to be used in combination with existing clinical diagnostic tools, thus improving the power of non-invasive diagnostic techniques.

4.2 The Dataset

The collection of digitized mammograms we analyzed was obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM)

`http://marathon.csee.usf.edu/Mammography/Database.html`.

The DDSM is described in detail in Heath et al [91]. Images from this database containing suspicious areas are accompanied by pixel-level "ground truth" information relating locations of suspicious regions to what was assessed and verified through biopsy. We selected 105 normal (benign) cases from volumes normal-01, and 98 cancer cases from volumes cancer-01 and cancer-02. Each case contains four mammograms (two for each breast: the craniocaudal (CC) and mediolateral oblique (MLO) projections) from a screening exam. We considered only the CC projections, using the right breast image for all normal cases, and the cancerous breast (right or left) image for cancer cases. A sub-image of size 1024×1024 was taken from each case for analysis. An example of a sub-image is provided in Fig. 11.

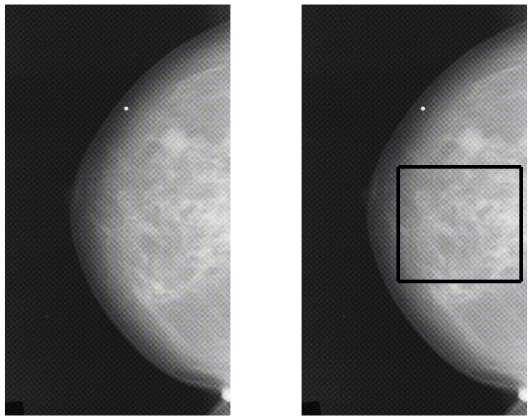


Figure 11: *Left panel*: right CC mammogram corresponding to a malignant case. *Right panel*: sub-image of size 1024×1024 considered for the analysis.

4.3 Discrete complex wavelets

This section discusses discrete complex wavelet transforms. Unlike the popular method used in this context [92], the proposed method is orthogonal and minimal.

4.3.1 Complex wavelet basis

The construction of the complex wavelet basis associated with multiresolution analysis follows the usual approach proposed by Mallat [93] and Daubechies [94]. Details on the construction and properties of complex wavelets can be found in [16, 14, 95, 96]. In analogy to real case, the wavelet function $\psi(x)$ for the complex wavelet is given by

$$\psi(x) = \frac{1}{\sqrt{2}} \sum_{k} (-1)^k 2h^*_{1-k} \phi(2x - k) \tag{4.3.1}$$

where ϕ is the scaling function, h is the low pass filter and the * indicates the complex conjugate. The representation of wavelets in 2-D can be done through the tensor product of univariate scaling functions and wavelets as follows:

$$\begin{aligned} \phi(x,y) &= \phi(x)\cdot\phi(y) \\ \psi_h(x,y) &= \phi(x)\cdot\psi(y) \\ \psi_v(x,y) &= \psi(x)\cdot\phi(y) \\ \psi_d(x,y) &= \psi(x)\cdot\psi(y) \end{aligned} \tag{4.3.2}$$

where symbols h, v, d in (4.3.2) stand for horizontal, vertical and diagonal directions, respectively. The atoms capture image features in the corresponding directions.

4.3.2 The complex scale-mixing 2-D wavelet transform

The discrete complex wavelet transform (DCWT) can be considered as a complex-valued extension of the standard discrete wavelet transform (DWT). It uses complex-valued filtering (analytic filter) for transforming the real/complex signals. Complex wavelet coefficients can be computed by Mallat's algorithm [93]

$$c_{j-1,l} = \sum_k h^*_{k-2l} c_{j,k} \tag{4.3.3}$$

and

$$d_{j-1,l} = \sum_k g^*_{k-2l} c_{j,k} \tag{4.3.4}$$

where h is as in (4.3.1) and g is the quadrature mirror filter. The * denotes the complex conjugate.

Conversely, the reconstruction is given by

$$c_{j,k} = \sum_l c_{j-1,l} h_{k-2l} + \sum_l d_{j-1,l} g_{k-2l}. \tag{4.3.5}$$

Moreover, the real and imaginary coefficients are used to compute the modulus and phase information. The wavelet coefficients can be written as

$$d_{j,k} = Re(d_{j,k}) + i \cdot Im(d_{j,k})$$

with magnitude

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + Im(d_{j,k})^2}$$

and phase

$$\angle d_{j,k} = \arctan\left(\frac{Im(d_{j,k})}{Re(d_{j,k})}\right)$$

when $|Re(d_{j,k})| > 0$.

There are many versions of the 2-D wavelet transforms which lead to different tessellations, or tilings [84]. Let us define the complex wavelet atoms as follows

$$\phi_{(j_1,j_2),\mathbf{k}}(\mathbf{x}) = 2^{(j_1+j_2)/2}\ \phi(2^{j_1}x - k_1, 2^{j_2}y - k_2) \tag{4.3.6}$$

$$\psi_{\delta,(j_1,j_2),\mathbf{k}}(\mathbf{x}) = 2^{(j_1+j_2)/2}\ \psi_\delta(2^{j_1}x - k_1, 2^{j_2}y - k_2), \tag{4.3.7}$$

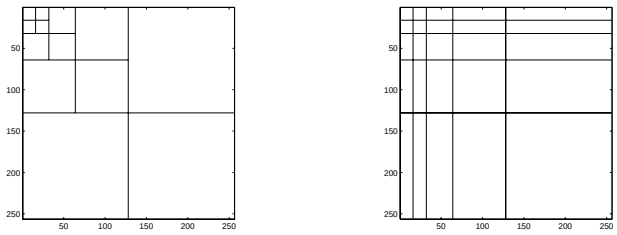
where δ is one of directions h, v, or d, and $(j_1, j_2) \in \mathbb{Z}^2$. Then, any function $f \in \mathcal{L}_2(\mathbb{R}^2)$ can be represented as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\mathbf{k}} c_{(J_0,J_0),\mathbf{k}}\ \phi_{(J_0,J_0),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j>J_0}\sum_{\mathbf{k}} d_{(J_0,j),\mathbf{k}}\ \psi_{h,(J_0,j),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j>J_0}\sum_{\mathbf{k}} d_{(j,J_0),\mathbf{k}}\ \psi_{v,(j,J_0),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j_1,j_2>J_0}\sum_{\mathbf{k}} d_{(j_1,j_2),\mathbf{k}}\ \psi_{d,(j_1,j_2),\mathbf{k}}(\mathbf{x}), \end{aligned}$$

and a 2-D wavelet transform, which we call the *scale-mixing wavelet transform* is obtained. The scale-mixing detail coefficients are defined as

$$\begin{aligned} d_{(J_0,j),\mathbf{k}} &= 2^{(J_0+j)/2}\int f(\mathbf{x})\ \psi_h^*(2^{J_0}x - k_1, 2^j y - k_2)\ dx\ dy, \\ d_{(j,J_0),\mathbf{k}} &= 2^{(j+J_0)/2}\int f(\mathbf{x})\ \psi_v^*(2^j x - k_1, 2^{J_0}y - k_2)\ dx\ dy, \\ d_{(j_1,j_2),\mathbf{k}} &= 2^{(j_1+j_2)/2}\int f(\mathbf{x})\ \psi_d^*(2^{j_1}x - k_1, 2^{j_2}y - k_2)\ dx\ dy, \end{aligned} \tag{4.3.8}$$

where ψ^* is a complex conjugate of ψ. Note that (j_1, j_2) in (4.3.6) and (4.3.7) can be indexed as $(j_1, j_1 + s)$ with $s \in \mathbb{Z}$.



(a) (b)

Figure 12: Tessellations for 2-D wavelet transforms. (a) Traditional 2-D transform of depth 4; (b) Scale-mixing wavelet transform of depth 4.

Similar to the traditional one- and two-dimensional cases, the complex scale-mixing detail coefficients are linked to the original image (2-D signal) through a matrix equation. Suppose that a $2^n \times 2^n$ image (matrix) A is to be transformed into the wavelet domain. The complex wavelet matrix W is first composed by the complex scaling and wavelet filter coefficients h_k and g_k as in Vidakovic [12]. Note that the wavelet filter is given by $g_k = (-1)^k h^*_{1+N-k}$ and N is a shift parameter which affects the location of the wavelet. Then the rows of A are transformed by a one-dimensional transform given by the wavelet matrix W, resulting in WA'. The same is repeated on the rows of WA'. The result is

$$B = W(WA')' = WAW', \tag{4.3.9}$$

the scale-mixing wavelet transform of matrix A, which will be the basis for defining the scale-mixing spectra. It represents a finite-dimensional implementation of (4.3.8) for signal $f(\mathbf{x})$ sampled in a form of matrix A..

The tessellation induced by transform in (4.3.9) is shown in Figure 12 (b). A more general transform can be obtained as an iterative repetition of the transform in (4.3.9) with depth k, applied only on the "smooth part" of the previous iterative step.

The scale-mixing 2-D transform is operationally appealing. Constructing an appropriate W is computationally fast and, since W is orthogonal, the inverse transform

is straightforward:

$$A = W'BW.$$

By inspecting the tessellation in Figure 12, several hierarchies of detail spaces can be identified. The diagonal hierarchy interfaces coefficients with the same component scales and coincides with the diagonal hierarchy in the traditional 2-D spectra. One level above and below the diagonal hierarchy are hierarchies of detail spaces that interface the scales that differ by 1. For the hierarchy above the diagonal, the scales along x-direction are interfaced by the next coarser scale along y-direction. For the hierarchy below the diagonal, the roles of x and y are interchanged.

The orthogonality of W implies

$$\text{trace}(AA') = \text{trace}(BB')$$

for $B = WAW'$, implying the total energy in the image A

$$E = \text{trace}(AA')$$

is preserved.

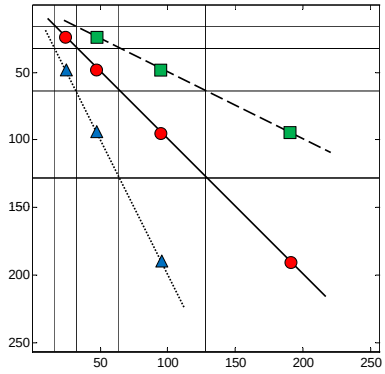


Figure 13: (a) Three detail-space hierarchies generating the scale-mixing 2-D transform, where (j_1, j_2) is indexed as $(j, j+s)$, $s \in \mathbb{Z}$. Circles correspond to $s = 0$, triangles to $s = 1$, and squares to $s = -1$. The scales (j_0, j), $j_0 = 7$ (squares), and (j, j_0), $j_0 = 6$ (triangles) are shown in the figure.

4.3.3 The complex scale-mixing wavelet spectra

The scale-mixing spectra is defined in terms of the complex scale-mixing coefficients in (4.3.8)

$$S(j) = \log_2 \mathbb{E}\left(|d_{(j,j+s),\mathbf{k}}|^2\right), \tag{4.3.10}$$

where $j, s \in \mathbb{Z}$ are fixed. Note that $s = 0$ in (4.3.8) corresponds to the diagonal 2-D spectra.

To calibrate the scale-mixing spectra, consider now a 2-D fractional Brownian motion, $B_H(\mathbf{u})$. For such a process, the scale-mixing detail coefficients are given by

$$d_{(j,j+s);\mathbf{k}} = 2^{j+\frac{s}{2}} \int B_H(\mathbf{u}) \psi^* \left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) d\mathbf{u}. \tag{4.3.11}$$

where ψ^* denotes the complex conjugate of ψ^d, the wavelet atom in the diagonal direction defined in (4.3.7). These coefficients are random variables with zero mean and variance [97], which leads to

$$\begin{aligned}\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{2j+s} \int & \psi\left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) \\ & \times \psi^*\left(2^j v_1 - k_1, 2^{j+s} v_2 - k_2\right) \mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] d\mathbf{u}\, d\mathbf{v}.\end{aligned} \tag{4.3.12}$$

As in Veitch and Abry [98], it is assumed here that the coefficients within and across the scales are uncorrelated.

From (4.3.12), it can be shown that

$$\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{-j(2H+2)}\, V_{\psi,s}(H), \tag{4.3.13}$$

where $V_{\psi,s}(H)$ is an expression depending on ψ, H and s, but not on the scale j,

$$V_{\psi,s}(H) = -\frac{\sigma_H^2}{2} \int\int \psi(\mathbf{p}+\mathbf{q})\, \psi^*(\mathbf{q})\, |\mathbf{p}_s|^{2H} 2^{-s}\, d\mathbf{p}\, d\mathbf{q}. \tag{4.3.14}$$

A proof of (4.3.13) is provided in the Appendix. By taking logarithms in (4.3.13)

$$\log_2 \mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = -(2H+2)j + \log_2 V_{\psi,s}(H) \tag{4.3.15}$$

for $j \in \mathbb{Z}$, the Hurst exponent can be estimated from the slope of the linear equation (4.3.15). Finally, the empirical counterpart of (4.3.15) is a regression defined on pairs

$$\left(j, \log_2 \frac{1}{n} \sum_{j,j+s} \left| d_{(j,j+s),\mathbf{k}} \right|^2 \right), \quad j, s \in \mathbb{Z}. \tag{4.3.16}$$

The slope of the regression would estimate the Hurst exponent, i.e., $H = -(slope + 2)/2$. Instead of the sample mean in (4.3.16), different location measures could be used, such as the median.

Although (4.3.16) is based on the ordinary least squares (OLS) regression, the variance of wavelet coefficients is not constant. Therefore we can improve the estimator by using more robust approach that incorporates heteroscedasticity. [98] used weighted linear regression to improve the estimator. This method weights each level by the inverse of the variance of that level. Hamilton et al. [90] proposed estimation methods that are based on a weighted average of all pair-wise slopes s_{ij} between levels i and j. Given a weight w_{ij}, the estimator of the overall slope in (4.3.16) is then $\sum_{i,j} w_{ij} s_{ij} / \sum_{i,j} w_{ij}$. Different types of weights are proposed, from which we obtain more robust estimation methods. In this work we adopted robust estimators proposed in the literature along with the OLS regression based estimator to perform more comprehensive comparison.

4.3.4 The complex phase information

It is known that phase and spectrum are collaborating in a nontrivial way to describe the data. While phases encode most of the coherent (in space and scale) structure of the image, the spectrum mostly encode the strength of local information that could be corrupted with noise [15]. For this reason phase information has been used in edge detection and in the reconstruction of images. A classical illustration is given in Oppenheim and Lim [99] where the image reconstruction is more driven by the phase of the Fourier transform rather than by the magnitude.

Recently, many research studies have focused on using phase information from the complex wavelet transforms [100, 101, 102, 103]. In the wavelet domain the phase of a coefficient near an isolated feature varies linearly with its distance from the feature. Despite of the numerous literatures focused on the usage of phase in detecting edges and the feature orientations, the discriminatory power of phase in the complex wavelet domain has not yet been studied and is unknown.

Although the phase of coefficients at each level does not have any scaling property as the wavelet-based spectra, the summary statistics of the phases turn out to be discriminatory. In the following section, we demonstrate how phase information can be used as a classification modality.

4.4 Mammogram Classification

In this section we illustrate how the complex wavelet-based spectra and the phase information can be used to classify digitized mammograms. We demonstrate that the spectra slope and phase, as descriptors of digitized images, have good discriminatory power. It is straightforward to implement the described analysis in various scientific areas in which 2-D data are instrumental, such as geoscience or industrial applications.

For every sub-image of size 1024×1024, we performed discrete real-valued wavelet transform (DWT) and discrete complex wavelet transform (DCWT) using Daubechies 6 tap filter. After each transform, we estimated the slope of wavelet spectra using traditional ordinary least squares regression (OLS) along with four robust estimation methods. The robust estimation approaches include Abry-Veitch weighted regression (AV), modified level enhanced OLS (MEOLS), harmonic average weighted slopes (HA) and modified HA (MHA). For more details on these robust estimators, we refer the reader to [98, 90]. Note that the wavelet spectra slope is used as a predictor instead of the Hurst exponent. It is because the estimated Hurst exponents H are

empirical, and the slowly decaying spectra (with slope > -2) could cause H to be negative.

For each classification method, we randomly selected 67% of the data as a training set to fit the classifier and used the remaining 33% of the data to test performance. The random selection of training and testing sets was repeated 10,000 times, so the reported prediction errors are averaged over 10,000 runs. Performance of each model was compared in terms of sensitivity, specificity, and overall correct classification rate.

The most parsimonious classification approach would be the logistic regression involving only the wavelet spectra slope as a predictor. The result is summarized in Table 14. The robust estimation methods show superior performance over OLS estimator, with rate ranging from 0.58 to 0.86. Although the performance of DWT and DCWT is comparable, the overall sensitivity and correct classification rates are higher with complex wavelet transform.

Table 14: Logistic classification based on the wavelet spectra slope; five different estimation methods were compared, each with real-valued (DWT) and complex (DCWT) wavelet transform.

	Method	Sensitivity	Specificity	Correct Classification
OLS	DWT	0.2472	0.7304	0.4744
	DCWT	0.2912	0.7135	0.4892
AV	DWT	0.7396	0.7901	0.7634
	DCWT	0.7635	0.8064	0.7838
MEOLS	DWT	0.5795	0.6768	0.6248
	DCWT	0.6175	0.7039	0.6585
HA	DWT	0.6168	0.7175	0.6651
	DCWT	0.6573	0.7361	0.6958
MHA	DWT	0.8599	0.8298	0.8436
	DCWT	0.8627	0.8489	0.8545

One of the interesting findings is that the phase contains information to classify normal and malignant images. Since the features and directions of background tissue is best preserved in the level of finest detail, we focus on the phase information of

finest detail. Figure 14 shows the estimated density of the phase average and variance at the finest level. While the average of malignant and normal cases are similar, the variance is quite different; the phase from normal images have higher variance, which implies more irregularity. This finding is consistent with universal paradigm in medical signal and image processing, that increased regularity of signals and images is often associated with pathologies. In this case, we hypothesize control mammograms have no clusters of consistent features and edges in the detailed wavelet space.

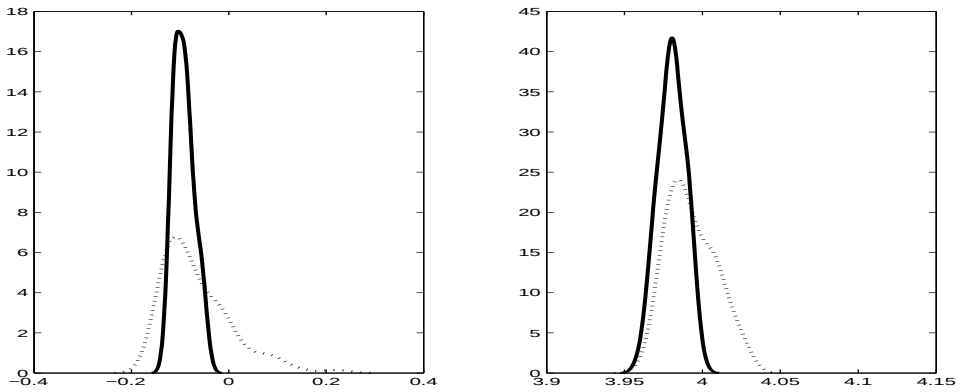


Figure 14: *Left panel*: Estimated density of phase average at the finest level. *Right panel*: Estimated density of phase variance at the finest level. Both are obtained from 105 normal cases and 98 cancer cases. The solid line corresponds to malignant cases and the dotted line to normal cases.

To validate the discriminatory power of the phase information, we assessed the logistic models by using each of the three summary statistics (average, variance, Q3-Q1) of the phases at the finest level. As Table 15 shows, the variance as a single predictor best classifies malignant and normal images with correct classification rate being 67%. Other measures such as kurtosis, median, coefficient of variation and mean absolute deviation retain classification power in a range of 50-60%.

We then conducted classification analysis based on the complex wavelet spectra slope and the phase variance. Figure 15 shows a scatter plot of cases by complex spectra slope versus phase variance, illustrating the differentiation between benign

Table 15: Logistic classification based on summary statistics of phase information at the finest level.

	Method	Sensitivity	Specificity	Correct Classification
Phase	Average	0.6018	0.5275	0.5533
	Variance	0.7004	0.6384	0.6654
	Skewness	0.5934	0.5279	0.5487
	Q1	0.6218	0.5121	0.5546
	Q3	0.6201	0.5080	0.5526

and malignant cases.

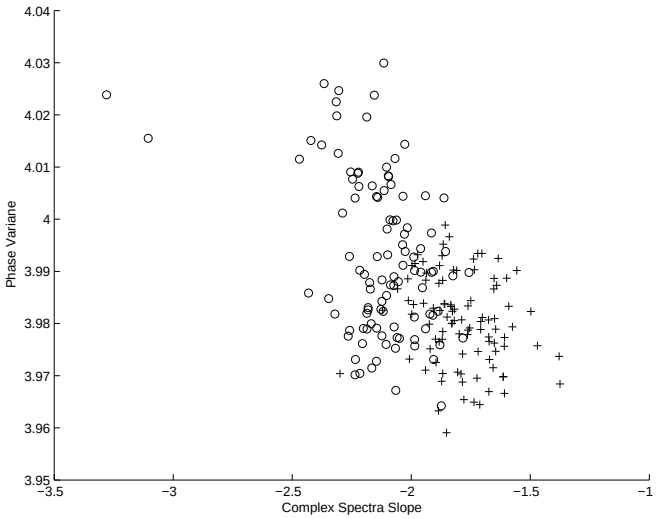


Figure 15: Scatter plot of Complex spectra slope (obtained by MHA estimation method) versus Phase variance. The symbols denote: *circles* for normal mammographies, *crosses* for malignant mammographies.

We combined complex wavelet spectra obtained from five different estimation methods with the phase variance. For each of the five pairs, we performed logistic, linear, quadratic and support vector machine (SVM) classification. By comparing results in Table 16 with Tables 14 and 15, we conclude that the overall performance improved significantly regardless of the pair combination. The improvement is especially notable for OLS; from Table 14, the OLS estimator itself had almost no discriminatory power (correct classification rate below 0.5). However by combining

it with the phase variance, the rates have increased up to 67%. Several wavelet bases and level combinations have been compared but the results and the conclusion remain the same.

Table 16: Linear, Quadratic and SVM classification based on the pair of complex wavelet spectra and the phase variance.

	Method	Sensitivity	Specificity	Correct Classification
(OLS; phase variance)	Logistic	0.6900	0.6338	0.6578
	Linear	0.7376	0.6067	0.6701
	Quadratic	0.8091	0.5365	0.6687
	SVM	0.4568	0.9012	0.6723
(AV; phase variance)	Logistic	0.7851	0.8009	0.7919
	Linear	0.8171	0.7744	0.7951
	Quadratic	0.8390	0.7522	0.7943
	SVM	0.7650	0.8127	0.7882
(EOLS; phase variance)	Logistic	0.6857	0.6390	0.6581
	Linear	0.7363	0.6069	0.6696
	Quadratic	0.8076	0.5264	0.6627
	SVM	0.4422	0.9087	0.6684
(MEOLS; phase variance)	Logistic	0.7095	0.7265	0.7155
	Linear	0.7629	0.6548	0.7072
	Quadratic	0.8145	0.5630	0.6849
	SVM	0.5472	0.8291	0.6839
(HA; phase variance)	Logistic	0.7395	0.7589	0.7471
	Linear	0.7805	0.7003	0.7392
	Quadratic	0.8221	0.6155	0.7157
	SVM	0.6372	0.7900	0.7113
(MHA; phase variance)	Logistic	0.8561	0.8574	0.8559
	Linear	0.8814	0.8320	0.8559
	Quadratic	0.8827	0.8297	0.8554
	SVM	0.8243	0.8826	0.8525

4.5 Conclusions

In this chapter we propose a complex scale-mixing 2-D wavelet transform in the context of assessing regularities of 2-D objects. The proposed transform is implemented

by matrix operations, and it guarantees orthogonality, low computational cost, directional insights, interplay between scales, and a straightforward inverse transform. We then explore the spectra and self-similarity measures based on the proposed complex wavelet transform. Their discriminatory power is demonstrated in the context of mammogram image classification. The procedure is based on background tissues of images rather than mammogram features such as microcalcification and tumor masses, which is an unused diagnostic modality in the field. The estimated Hurst exponent and phase information turn out to be discriminatory summaries in mammogram image classification. Although phase information has been used mainly for edge detection and image reconstructions in the literature, we identify that the summary measures of phase contribute to the correct classification of cancer and normal images.

To obtain the estimates of Hurst exponent, we use four robust estimation methods (AV, MEOLS, HA, MHA) along with the ordinary least squares estimator. From the logistic classification model, we found that robust Hurst exponent estimates and the phase variance at the finest level have enough power to differentiate between benign and malignant cases.

It is well known that for the real wavelets there is no symmetric and compactly supported scaling function defining an orthogonal MRA. Complex wavelets assure symmetry, compact support and orthogonality of decomposing atoms which is desirable in image representations. This advantage is proved through the higher sensitivity and correct classification rate across different representation scenarios.

Another benefit of using the proposed model is that we can combine the Hurst exponent estimates with the phase variance to identify whether the images contained evidence of malignancy. By combining these two measures, we obtain significantly improved correct classification rate. Note that the most accurate classification rates achieve 86% with (MHA; phase variance) pair. However, MHA and MEOLS employ

weights that heavily emphasize the fine detail levels, which is in accordance with the empirical observation that finer levels in real wavelet decompositions are critical for correct mammogram image classification [90]. Therefore complex wavelets provide a more robust tool in the sense that combination of an under-performing summary for real wavelets becomes discriminatory when combined with phase information.

CHAPTER V

CONCLUSION

This thesis deals with analyzing non-traditional data in health-care and biomedical domains. First of all, the second chapter focuses on the meta-analysis of clinical trials with incomplete data set. This requires a new type of ecological-type inference in matched-pair studies when only fixed marginal totals are available. The proposed methodology is based on the Metropolis random walk and hierarchical Bayesian model and it incorporates any prior information available. The performance of the proposed model is verified through 17 matched-pair study examples. This work is motivated by meta-analysis that compares the effectiveness of two types of dental sealant materials. By using this model, we are able to incorporate more clinical studies (that were previously excluded solely because they failed to report data in a paired manner) and find the difference for the first time among similar meta-analyses.

Developing a methodology to retrieve paired data from matched paired design studies that do not report paired tables is an important issue in meta-analyses. A study of methodological issues in cross-over trials reported that presenting data as paired tables was uncommon in the published work. Another search of the Cochrane Database of Systematic Reviews found that over 50% of reviews either treated cross-over trials as if they were parallel groups or only used data from the first time period assuming that in doing so it would better approximate a parallel group.

For the Metropolis algorithm, the prior information is incorporated via the mean of the Dirichlet prior. Two modifications of the algorithm are possible: (i) one may consider placing the prior on the parameter θ directly, and (ii) it is possible to introduce an additional level of hierarchy in the model to incorporate prior information by

a hyper-prior placed on the mean of the Dirichlet prior. The Metropolis random walk algorithm is expanded to paired tables of size $k \times k$ in a straightforward manner.

This chapter demonstrates that the flexibility of Bayesian approach lends its strength to handle the challenging issues in evidence-based healthcare decision-making. I plan to explore more general Bayesian approaches in this field. The study ranges from asking the right question (what works for whom, for which medical problems, and under what circumstances) to obtaining and incorporating existing information (pivotal trials, systematic reviews, and rigorously conducted observational studies) to designing the study of interest (usually randomized clinical trials: RCT) to performing adaptive RCT study (that allows interventions to be added and less effective ones dropped without restarting the study) to drawing statistically valid conclusion. Compared with the traditional frequentist methods, the Bayesian approach can reduce the sample size, time, and cost required to obtain decision-relevant information.

The third chapter deals with the spatial repeated measures with zero-inflation and no covariate information. This model comprises of two steps: 1) autologistic model that estimates the probability of zero observations and 2) excess-zero model that describes the underlying geographical phenomenon of the zero-inflated data. This methodology is motivated and illustrated with the example of estimating U.S. county level natural fluoride concentration in the ground water. The proposed hierarchical Bayesian structure produces improved discretized estimates and reflects the continuous behavior of the natural fluoride occurrence better. By employing the hyper-prior structure, state level estimates can be obtained as well.

The methods discussed here can be extended to semi-continuous two-part models, which assume a continuous (e.g., log-normal or Gamma) rather than Poisson or negative binomial distributions for the non-zero values. This model can also be extended to a recursive system; after the excess-zero modeling (ii), the probability

of zero observation can be updated in the autologistic model (i), and this feedback loop continues until it reaches an equilibrium stage. This model can be applied to any spatial count data with excess zero, and be extended as covariate information becomes available.

This work is motivated to improve the water fluoridation policy, the addition of fluoride to drinking water to prevent tooth decay. However, fluoride can occur naturally in water in concentrations well above recommended levels which imposes several long-term adverse effects such as dental fluorosis, skeletal fluorosis, and weakened bones. Although describing the availability of natural fluoride by county can inform public health efforts in prevention of tooth decay and dental fluorosis, the effort has been limited to a few states. The county-level estimates will not only be a valuable information to policy makers and residents, but also for future research; one of the nation's most comprehensive public health database, National Health and Nutrition Examination Survey (NHANES) provides survey and examination results by county. Therefore, persistent research questions in the field such as possible correlation between high natural fluoride occurrence vs. dental fluorosis can benefit from this county-level natural fluoride occurrence map.

Finally, the last chapter covers image classification based on the overall image regularity in complex wavelet domain. We first propose a complex scale-mixing 2-D wavelet transform; the proposed transform guarantees orthogonality, low computational cost, directional insights, interplay between scales, and a straightforward inverse transform. We then explore the spectra and self-similarity measures based on the proposed complex wavelet transform. Their discriminatory power is demonstrated in the context of mammogram image classification. The procedure is based on background tissues of images rather than mammogram features such as microcalcification and tumor masses, which is an unused diagnostic modality in the field.

The estimated Hurst exponent and phase information turn out to be discriminatory summaries in mammogram image classification. Although phase information has been used mainly for edge detection and image reconstructions in the literature, we identify that the summary measures of phase contribute to the correct classification of cancer and normal images. It is well known that for the real wavelets there is no symmetric and compactly supported scaling function defining an orthogonal MRA. Complex wavelets, on the other hand, assure symmetry, compact support and orthogonality of decomposing atoms which is desirable in image representations. This advantage is proved through the higher sensitivity and correct classification rate across different representation scenarios.

Another benefit of using the proposed model is that we can combine the Hurst exponent estimates with the phase variance to identify whether the images contained evidence of malignancy. By combining these two measures, we obtain significantly improved correct classification rate. Complex wavelets provide more robust tool in the sense that combination of an under-performing summary for real wavelets becomes discriminatory when combined with phase information.

Mammography devices that provide three-dimensional images of the breast for cancer screening and diagnostics was approved by the U.S. Food and Drug Administration in 2011. As three dimensional images become available, 3D wavelet transform can be applied to assess the overall regularity. Thus, promising new indicators may be added to current screening techniques that will improve the specificity and sensitivity of breast cancer diagnoses. The implication of this research is improved prognoses along with a reduction in the number of additional procedures for specific diagnosis.

CHAPTER VI

APPENDIX

6.1 Derivation of paired and parallel variance

Let us assume that x is a random variable and $f(x)$ is an arbitrary function that is infinitely differentiable in a neighborhood of Ex. Then the 1^{st} order Taylor expansion around Ex is

$$f(x) \approx f(Ex) + f'(Ex)(x - Ex)$$

Therefore, the variance and the covariance of $f(x)$ can be approximated as follows;

$$Var\big(f(x)\big) \approx \big[f'(Ex)\big]^2 Var(x) \tag{6.1.1}$$

$$Cov\big(f(x), f(y)\big) \approx f'(Ex)f'(Ey)Cov(x, y) \tag{6.1.2}$$

Now, let us consider the variance of the log risk ratio (RR). For p_t the probability of event for the treatment group and p_c that of the control group, the log RR is

$$\log(RR) = \log \frac{p_t}{p_c} \tag{6.1.3}$$

and the variance is

$$Var\big[\log(RR)\big] = Var\big[\log(p_t)\big] + Var\big[\log(p_c)\big] - 2Cov\big[\log(p_t), \log(p_c)\big] \tag{6.1.4}$$

For matched-pair studies in Table 2, p_t and p_c can be approximated as $\frac{y_{11}+y_{12}}{n}$ and $\frac{y_{11}+y_{21}}{n}$ respectively. If we assume that each cell count $(y_{11}, y_{12}, y_{21}, y_{22})$ is generated from multinomial distribution with underlying probability $(p_{11}, p_{12}, p_{21}, p_{22})$, then

$$Var(y_{11}) = np_{11}(1 - p_{11}) \tag{6.1.5}$$

and

$$Cov(y_{11}, y_{12}) = -np_{11}p_{12} \tag{6.1.6}$$

By applying (6.1.1) and (6.1.2) to (6.1.4), we obtain

$$Var\big[\log(RR)\big] \approx \frac{Var(p_t)}{(Ep_t)^2} + \frac{Var(p_c)}{(Ep_c)^2} - 2\frac{Cov(p_t, p_c)}{Ep_t Ep_c} \tag{6.1.7}$$

Also, from (6.1.5) and (6.1.6)

$$Ep_t \approx \frac{y_{11} + y_{12}}{n},\ Ep_c \approx \frac{y_{11} + y_{21}}{n},$$

$$Var(p_t) \approx \frac{(y_{11} + y_{12})(y_{21} + y_{22})}{n^3},\ Var(p_c) \approx \frac{(y_{11} + y_{21})(y_{12} + y_{22})}{n^3}$$

and

$$Cov(p_t, p_c) \approx \frac{y_{11}y_{22} - y_{12}y_{21}}{n^3}$$

If we assume that p_t and p_c are independent (which is not correct for the matched-pair studies), then the variance of the log risk ratio is approximately

$$Var\big[\log(RR)\big] = \frac{1}{y_{11} + y_{12}} + \frac{1}{y_{11} + y_{21}} - \frac{2}{n} \tag{6.1.8}$$

which is equal to the square of (2.3.3). However, if we consider the positive correlation between p_t and p_c, then the variance is

$$Var\big[\log(RR)\big] = \frac{1}{y_{11} + y_{12}} + \frac{1}{y_{11} + y_{21}} - \frac{2y_{11}}{(y_{11} + y_{12})(y_{11} + y_{21})} \tag{6.1.9}$$

which is equal to the square of (2.3.2).

6.2 Derivation of excess zero model mean and variance

6.2.1 Hurdle Model

From Section 3.2.2, we obtain

$$\mathbb{E}(Y) = \sum_{y\geq 0} y \cdot P(Y = y) = \frac{1 - f_1(0)}{1 - f_2(0)} \sum_{y\geq 1} y f_2(y) = \frac{1 - f_1(0)}{1 - f_2(0)} \mathbb{E}(Z)$$

where $\mathbb{E}(Z)$ is the expected value of confined samples of y greater than zero. In the same manner,

$$\mathbb{E}(Y^2) = \frac{1 - f_1(0)}{1 - f_2(0)} \mathbb{E}(Z^2)$$

and hence

$$\begin{aligned}
\mathbf{Var}(Y) &= \frac{1 - f_1(0)}{1 - f_2(0)} \mathbb{E}(Z^2) - \big[\frac{1 - f_1(0)}{1 - f_2(0)} \mathbb{E}(Z)\big]^2 \\
&= \frac{1 - f_1(0)}{1 - f_2(0)} \big(1 - \frac{1 - f_1(0)}{1 - f_2(0)}\big) [\mathbb{E}(Z)]^2 + \frac{1 - f_1(0)}{1 - f_2(0)} \mathbf{Var}(Z) \\
&= \frac{f_1(0) - f_2(0)}{1 - f_1(0)} [\mathbb{E}(Y)]^2 + \gamma \mathbb{E}(Y).
\end{aligned}$$

6.2.2 Zero-inflated Model

From Section 3.2.3, we obtain

$$\mathbb{E}(Y) = \sum_{y\geq 0} y \cdot P(Y = y) = (1 - \pi_Z) \sum_{y\geq 1} y f_D(y) = (1 - \pi_Z) \mathbb{E}(Y_D).$$

Since

$$\mathbb{E}(Y^2) = (1 - \pi_Z) \mathbb{E}(Y_D^2),$$

the variance can be written as follows:

$$
\begin{aligned}
\mathbf{Var}(Y) &= (1-\pi_Z)\mathbb{E}(Y_D^2) - (1-\pi_Z)^2[\mathbb{E}(Y_D)]^2 \\
&= \pi_Z(1-\pi_Z)[\mathbb{E}(Y_D)]^2 + (1-\pi_Z)\mathbf{Var}(Y_D) \\
&= \frac{\pi_Z}{1-\pi_Z}[\mathbb{E}(Y)]^2 + \gamma\mathbb{E}(Y)
\end{aligned}
$$

6.3 Derivation of (4.3.13)

The scale-mixing detail coefficients of a 2-D fBm (4.3.11) are

$$d_{(j,j+s);\mathbf{k}} = 2^{j+\frac{s}{2}} \int B_H(\mathbf{u})\psi\left(2^j u_1 - k_1, 2^{j+s}u_2 - k_2\right) d\mathbf{u}.$$

These coefficients are random variables with zero mean and variance [97]

$$\begin{aligned}\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{2j+s} \int \psi\left(2^j u_1 - k_1, 2^{j+s}u_2 - k_2\right) \\ \times \psi^*\left(2^j v_1 - k_1, 2^{j+s}v_2 - k_2\right) \mathbb{E}\left[B_H(\mathbf{u})B_H(\mathbf{v})\right] d\mathbf{u}\; d\mathbf{v}. \qquad (6.3.1)\end{aligned}$$

Since

$$\mathbb{E}\left[B_H(\mathbf{u})B_H(\mathbf{v})\right] = \frac{\sigma_H^2}{2}\left(|\mathbf{u}|^{2H} + |\mathbf{v}|^{2H} - |\mathbf{u}-\mathbf{v}|^{2H}\right),$$

and

$$\int \psi\left(2^j u_1 - k_1, 2^{j+s}u_2 - k_2\right) d\mathbf{u} = \int \psi\left(2^j v_1 - k_1, 2^{j+s}v_2 - k_2\right) d\mathbf{v} = 0,$$

it can be easily seen that (6.3.1) becomes

$$\begin{aligned}\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = -\frac{\sigma_H^2}{2}\, 2^{2j+s} \int\int \psi\left(2^j u_1 - k_1, 2^{j+s}u_2 - k_2\right) \\ \times \psi^*\left(2^j v_1 - k_1, 2^{j+s}v_2 - k_2\right) |\mathbf{u}-\mathbf{v}|^{2H} d\mathbf{u}\; d\mathbf{v}.\end{aligned}$$

Next, define substitutions

$$\mathbf{p} \;=\; (p_1, p_2) = \left(2^j(u_1 - v_1), 2^{j+s}(u_2 - v_2)\right),$$

Then, if $\mathbf{p}_s \equiv (p_1, 2^{-s}p_2)$,

$$\begin{aligned}\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] &= -\frac{\sigma_H^2}{2}\, 2^{2j+s} \int\int \psi(\mathbf{p}+\mathbf{q})\; \psi^*(\mathbf{q})\; 2^{-2jH}\; |\mathbf{p}_s|^{2H}\; 2^{-4j-2s}\; d\mathbf{p}\; d\mathbf{q} \\ &= -\frac{\sigma_H^2}{2}\, 2^{-j(2H+2)} \int\int \psi(\mathbf{p}+\mathbf{q})\; \psi^*(\mathbf{q})\; |\mathbf{p}_s|^{2H} 2^{-s}\; d\mathbf{p}\; d\mathbf{q} \\ &= 2^{-j(2H+2)}\; V_{\psi,s}(H),\end{aligned}$$

where

$$V_{\psi,s}(H) = -\frac{\sigma_H^2}{2} \int\int \psi(\mathbf{p}+\mathbf{q})\; \psi^*(\mathbf{q})\; |\mathbf{p}_s|^{2H} 2^{-s}\; d\mathbf{p}\; d\mathbf{q}, \qquad (6.3.2)$$

is an integral depending on ψ, H, and s, but not on the scale j.

Bibliography

[1] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *British Medical Journal*. 327:557-560, 2003.

[2] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 21:1539-1558, 2002.

[3] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 7:188-188, 1986.

[4] Cressie N. Statistics for Spatial Data. *John Wiley & Sons, Inc.* 1993.

[5] Lawson AB, Biggeri A, Boehning D, Lesaffre E, Viel J-F, Clark A, Schlattmann P, Divino F. Disease mapping models: An empirical evaluation. *Statistics in Medicine*. 19:2217-2242, 2000.

[6] Jaffard S. Local behavior of Riemann‘s function. *Contemporary Mathematics* 189:287, 1995.

[7] Gonçalves P, Riedi R, Baraniuk R. A simple statistical analysis of wavelet-based multifractal spectrum estimation. *The Proceedings 32nd Asilomar Conference on Signals, Systems and Computers* 1:287-291, Citeseer, 1998.

[8] Abry P, Flandrin P, Taqqu M, Veitch D. Wavelets for the analysis, estimation and synthesis of scaling data. *Self-Similar Network Traffic and Performance Evaluation* 39-88, 2000.

[9] Abry P, Flandrin P, Taqqu M, Veitch D. Self-similarity and long-range dependence through the wavelet lens. *Long-range dependence: theory and applications* 527-56, 2002.

[10] Mandelbrot B, Van Ness J. Fractional Brownian motions, fractional noises and applications. *SIAM review* 10:422-37, 1986.

[11] Shi B, Moloney K, Leonard KV, Jacko J, Sainfort F. Multifractal discrimination model (mdm) of high-frequency pupil diameter measurements for human-computer interaction. *Quantitative Medical Data Analysis Using Mathematical Tools and Statistical Techniques* 16:333-50, 2007.

[12] Vidakovic B. Statistical Modeling by Wavelets. *John Wiley & Sons*, 1999.

[13] Fernandes FCA, van Spaendonck RLC, Burrus CS. Multidimensional, mapping-based complex wavelet transforms. *IEEE Transactions on Image Processing* 14:110-24, 2005.

[14] Lina JM, Mayrand M. Complex Daubechies wavelets. *Applied and Computational Harmonic Analysis* 2:219-29, 1995.

[15] Clonda D, Lina JM, Goulard B. Complex Daubechies wavelets: properties and statistical image modeling. *Signal Processing* 84:1-23, 2004.

[16] Lawton W. Applications of complex valued wavelet transforms to subband decomposition. *IEEE Transactions on Signal Processing* 41:3566-8, 1993.

[17] Magarey J, Kingsbury N. Motion estimation using a complex-valued wavelet transform. *IEEE Transactions on Signal Processing* 46:1069-84, 1998.

[18] Portilla J, Simoncelli EP. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision* 40:49-71, 2000.

[19] Achim A, Kuruoglu EE. Image denoising using bivariate α-stable distributions in complex wavelet domain. *IEEE Signal Processing Letters* 12:17-20, 2005.

[20] Kim SB, Wang Z, Oraintara S, Temiyasathit C, Wongsawat Y. Feature selection and classification of high-resolution NMR spectra in the complex wavelet transform domain. *Chemometrics and Intelligent Laboratory Systems* 90:161-8, 2008.

[21] Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology.* 31:140-149, 2002.

[22] Ahovuo-Saloranta A, Hiiri A, Nordblad A, Makela M, Worthington HV. Pit and fissure sealants for preventing decay in the permanent teeth of children and adolescents.*Cochrane Database of Systematic Reviews 2008.* Issue 4. Art. No.: CD001830. DOI:10.1002/14651858.CD001830.pub3.

[23] Muller-Bolla M, Lupi-Pegurier L, Tardieu C, Velly AM, Antomarchi C. Retention of resin-based pit and fissure sealants: A systematic review.*Community Dentistry and Oral Epidemiology.* 34(5):321-36, 2006.

[24] Vaeth M, Poulsen S. Comments on a commentary: statistical evaluation of split mouth caries trials.*Community Dentistry and Oral Epidemiology.* 26:80-83, 1998.

[25] Wacholder S, Weinberg CR. Paired versus two-sample design for a clinical trial of treatments with dichotomous outcome: power considerations. *Biometrics.* 38:801-812, 1982.

[26] King G. A solution to the ecological inference problem. Princeton University Press: Princeton, NJ 1997.

[27] King G, Tanner MA, Rosen O. Ecological inference: new methodological strategies. Cambridge University Press: Cambridge, UK 2004.

[28] Good IJ. A Bayesian significance test for multinomial distribution (with discussion). *Journal of the Royal Statistical Society. Series B: Statistical methodology.* 29:399-431, 1967.

[29] Fienberg SE, Holland PW. On the choice of flattening constants for estimating multinominal probabilities.*Journal of Multivariate Analysis*. 2:127-134, 1972.

[30] Diaconis P, Sturmfels B. Algebraic algorithms for sampling from conditional distributions.*The Annals of Statistics*. 26(1):363-397, 1998.

[31] Evans M, Gilula Z, Guttman I. Latent class analysis of two-way contingency tables by Bayesian methods.*Biometrika*. 76(3):557-563, 1989.

[32] Wakefield J. Ecological inference for 2×2 tables. *Journal of the Royal Statistical Society. Series A: Statistics in Society*. 167(3):385-445, 2004.

[33] Jackson C, Best N, Richardson S. Improving ecological inference using individual-level data.*Statistics in Medicine*. 25:2136-2159, 2006.

[34] Dobra A. Markov bases for decomposable graphical models. *Bernoulli*. 9(6):1093-1108, 2003.

[35] Dobra A, Tebaldi C, West M. Data augmentation in multi-way contingency tables with fixed marginal totals. *Journal of Statistical Planning and Inference*. 136:355-372, 2006.

[36] Salway R, Wakefield J. Gamma generalized linear models for pharmacokinetic data. *Biometrics*. 64(2):620-626, 2008.

[37] Liu J, Gustafson P, Cherry N, Burstyn I. Bayesian analysis of a matched case-control study with expert prior information on both the misclassification of exposure and the exposure-disease association. *Statistics in Medicine*. 28(27):3411-3423, 2009.

[38] Edwards AWF. The measure of association in a 2×2 table. *Journal of the Royal Statistical Society. Series A: General*.126(1):109-114, 1963.

[39] Brunelle RL, Llewelyn J, Anderson JH Jr, Gale EAM, Koivisto VA. Meta-analysis of the effect of insulin lispro on severe hypoglycemia in patients with Type I diabetes. *Diabetes Care*. 21:1726-1731, 1998.

[40] Forss H, Halme E. Retention of a glass ionomer cement and a resin-based fissure sealant and effect on carious outcome after 7 years. *Community Dentistry and Oral Epidemiology*. 26(1):21-25. 1998.

[41] Beauchamp BJ, Caufield PW, Crall JJ, Donly K, Feigal R, Gooch B, Ismail A, Kohn W, Siegal M, Simonsen R. Evidence-based clinical recommendations for the use of pit-and-fissure sealants: A report of the American Dental Association Council on Scientific Affairs. *Journal of the American Dental Association* 139:257-68, 2008.

[42] Gooch BF, Griffin SO, Kolavic G, Kohn WG, Rozier RG, Siegal M, Fontana M, Brunson D, Carter N, Curtis DK, Donly KJ, Haering H, Hill LF, Hinson HP, Kumar J, Lampiria L, Mallatt M, Meyer DM, Miller WR, Sanzi-Schaedel SM, Simonsen R, Truman BI, Zero DT. Preventing dental caries through school-based sealant programs: Updated recommendations and reviews of evidence. *Journal of the American Dental Association* 140:1356-65, 2009.

[43] Beiruti N, Frencken JE, van't Hof MA, van Palenstein Helderman WH. Caries preventive effect of resin-based and glass ionomer sealants over time: a systematic review. *Community Dentistry and Oral Epidemiology* 34:403-9, 2006.

[44] Yengopal V, Mickenaustch S, Bezerra AC, Leal SC. Caries-preventive effect of glass ionomer and resin-based fissure sealants on permanent teeth: a meta analysis. *Journal or Oral Science* 51:373-82, 2009.

[45] Arrow P, Riordan PJ. Retention and caries preventive effects of a GIC and a resin-based fissure sealant. *Community Dentistry and Oral Epidemiology* 23:282-5, 1995.

[46] Forss H, Halme E. Retention of a glass ionomer cement and a resin-based fissure sealant and effect on carious outcome after 7 years. *Community Dentistry and Oral Epidemiology* 26:21-5, 1998.

[47] Kervanto-Seppälä S, Lavonius E, Pietila I, Pitkaniemi J, Meurman JH, Kerosuo E. Comparing the caries-preventive effect of two fissure sealing modalities in public health care: a single application of glass ionomer and a routine resin-based sealant programme. A randomized split-mouth clinical trial. *International Journal of Paediatric Dentistry* 18:56-61, 2008.

[48] Poulsen S, Beiruti N, Sadat N. A comparison of retention and the effect on caries of fissure sealing with a glass-ionomer and a resin-based sealant. *Community Dentistry and Oral Epidemiology* 29:298-301, 2001.

[49] Ganesch M, Tandon S. Clinical evaluation of FUJI VII sealant material. *Journal of Clinical Pediatric Dentistry* 31:52-7, 2006.

[50] Mills RW, Ball IA. A clinical trial to evaluate the retention of a silver cermet-ionomer cement used as a fissure sealant. *Operative Dentistry* 18:148-54, 1993.

[51] Barker LK, Griffin SO, Jeon S, Gray SK, Vidakovic B. Ecological-type inference in matched pair studies with fixed marginal totals. *Statistics in Medicine* 30(5):541-548, 2011.

[52] Agresti A. An Introduction to Categorical Data Analysis. *John Wiley & Sons, Inc* New York, 1996.

[53] Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine* 21:1601-23, 2002.

[54] Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, Low-Choy SJ, Tyre AJ, Possingham HP. Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. *Ecology Letters* 8:1235-46, 2005.

[55] Wenger SJ, Freeman MC. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology* 89:2953-9, 2008.

[56] Simons JS, Neal DJ, Gaher RM. Risk for marijuana-related problems among college students: an application of zero-inflated negative binomial regression. *American Journal of Drug and Alcohol Abuse* 32:41-53, 2006.

[57] Lee AH, Wang K, Scott JA, Yau KW, McLachlan GJ. Multi-level zero-inflated Poisson regression modeling of correlated count data with excess zeros. *Statistical Methods in Medical Research* 15:47-61, 2006.

[58] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1-14, 1992.

[59] Cheung YB. Growth and cognitive function of Indonesian children: zero-inflated proportion models. *Statistics in Medicine* 25:3011-22, 2006.

[60] Atkins DC, Gallop RJ. Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. *Journal of Family Psychology* 21:726-35, 2007.

[61] Agarwal DK, Gelfand AE, Citron-Pousty SC. Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* 9:341-55, 2002.

[62] Fahrmeir L, Echavarria LO. Structured additive regression for overdispersed and zero-inflated count data. *Applied Stochastic Models in Business and Industry* 22:351-69, 2006.

[63] Gschlößl S, Czado C. Modeling count data with overdispersion and spatial effects. *Statistical Papers* 48:531-52, 2008.

[64] Mullahy J. Specification and testing of some modified count data models. *Journal of Econometrics* 33:341-65, 1986.

[65] Heilbron DC. Zero-altered and other regression models for count data with added zeros. *Biometric Journal* 36:531-47, 1994.

[66] Ridout M, Demetrio CGB, Hinde J. Models for count data with many zeros. *International Biometric Conference* Cape Town, 1998.

[67] Angers JF, Biswas A. A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis* 42:37-46, 2003.

[68] Ghosh SK, Mukhopadhyay P, Lu JC. Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference* 136:1360-75, 2006.

[69] Neelon BH, O'Malley AJ, Normand ST. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modeling* 10(4):421-439, 2010.

[70] Besag JE. Nearest-neighbor systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B* 34:75-83, 1972.

[71] Bohning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated poisson model and the decayedand missing and filled teeth index in dental epidemiology. *Journal of Royal Statistical Society. Series A* 162:195-209, 1999.

[72] Calder CA, Cressie N. Some topics in convolution-based spatial modeling. *The Proceedings of the 56th Session of the International Statistics Institute* Lisborn, Portugal, 2007.

[73] Gotway CA, Young LJ. Combining incompatible spatial data. *Journal of the American Statistical Association* 97:632-48, 2002.

[74] Wasserman L. Bayesian model selection and model averaging. *Journal of Mathematical Psychology* 44:92-107, 2000.

[75] Congdon P. Applied Bayesian modeling. *John Wiley and Sons*, 2003.

[76] Spiegelhalter D, Thomas A, Best N, Gilks W. Bayesian inference using Gibbs sampling manual. *MRC Biostatistics Unit, Institute of Public Health, Cambridge* 2003.

[77] Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434-55, 1998.

[78] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64:583-639, 2002.

[79] Brunt R, Vasak L, Griffioen J. Fluoride in groundwater: probability of occurrence of excessive concentration on global scale. *International groundwater resources assessment centre* 2004.

[80] Altekruse, S.F. Kosary, C.L. Krapcho, M. Neyman, N. Aminou, R. Waldron, W. SEER Cancer Statistics Review: 1975-2007. *National Cancer Institute*, Bethesda, MD. 2010.

[81] U.S. Department of Health and Human Services. Healthy People 2020. http://healthypeople.gov/HP2020. Washington, DC. 2010.

[82] Martin, J., Moskowitz, M., Milbrath, J. Breast cancer missed by mammography. *American Journal of Roentgenology* 37:142-162, 1979.

[83] Nicolis, O., Ramírez-Cobo, P., Vidakovic, B. 2-D wavelet-based spectra with applications. *Computational Statistics and Data Analysis* 55: 738-751, 2011.

[84] Ramírez-Cobo, P., Lee, K. S., Molini, A., Porporato, A., Katul, G., Vidakovic, B. A wavelet-based spectral method for extracting self-similarity measures in time-varying two-dimensional rainfall maps. *Journal of Time Series Analysis* 32:351-363, 2011.

[85] Gao, R. Yan, R. Wavelets: Theory and Applications for Manufacturing. *Springer* 2010

[86] Wang, T., Karayiannis, N. Detection of microcalcifications in digital mammograms using wavelets. *IEEE Transactions on medical imaging* 17:498-509, 1998.

[87] Netsch, T. Peitgen, H. Scale-space signatures for the detection of clustered microcalcifications in digital mammograms. *IEEE Transactions on medical imaging* 18:774-786, 1999.

[88] Kestener P, Lina JM, Saint-Jean P, Arneodo A. Waelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Analysis and Stereology* 20:169-74, 2001.

[89] El-Naqa, I., Yang, Y., Wernick, M., Galatsanos, N., Nishikawa, R. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on medical imaging* 21:1552-1563, 2002.

[90] Hamilton, E. K., Jeon, S., Ramírez-Cobo, P., Lee, K. S., Vidakovic, B. Diagnostic classification of digital mammograms by wavelet-based spectral tools: A comparative study. *The Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine* 384-389, 2011.

[91] Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P. The digital database for screening mammography. *The Proceedings of the 5th International Workshop on Digital Mammography* Toronto, Canada, 2000.

[92] Selesnick IW, Baraniuk RG, Kingsbury NG. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine* 2005.

[93] Mallat, S. G. A Wavelet Tour of Signal Processing. *Academic Press* San Diego, 1998.

[94] Daubechies, I. Ten Lectures on Wavelets. *Society for Industrial and Applied Mathematics*. Philadelphia, PA. 1992.

[95] Strang, G., Nguyen, T. Wavelets and Filter Banks. *Wellesley-Cambridge Press* Wellesley, MA. 1996.

[96] Zhang, X.-P., Desai, M., Peng, Y. Orthogonal complex filter banks and wavelets: Some properties and design. *IEEE Transactions on Signal Processing* 47:1039-1048, 1999.

[97] Heneghan, C., Lown, S., Teich, M. Two dimensional fractional Brownian motion: Wavelet analysis and synthesis. *Image analysis and interpretation, proceedings of the IEEE Southwest Symposium* 213-217, 1996.

[98] Veitch, D. Abry, P. A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory* 45:878-897, 1999.

[99] Oppenheim, A. Lim, J. The importance of phase in signals. The Proceedings of the IEEE 69:529-541, 1981.

[100] Anderson, R., Kingsbury, N., Fauqueur, J. Determining multiscale image feature angles from complex wavelet phases. *International conference on Image processing* Toronto, Canada, 2005.

[101] Hua, G. Orchard, M. T. Image reconstruction from the phase or magnitude of its complex wavelet transform. *The Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* 3261-3264, 2008.

[102] Miller, M. Kingsbury, N. Image modeling using interscale phase properties of complex wavelet coefficients. *IEEE Transactions on Image Processing* 17:1491-1499, 2008.

[103] Rakvongthai, Y. Oraintara, S. On phase statistics of complex wavelet coefficients at edges. *European Conference on Signal Processing* 2-6, 2008.

[104] Barber S, Nason GP. Real nonparametric regression using complex wavelets. *Journal of Royal Statistical Society Series B* 66:927-39, 2004.

[105] Priebe CE, Solka JL, Lorey RA, Rogers GW, Poston WL, Kallergi M, Quian W, Clarke LP, Clark RA. The application of fractal analysis to mammographic tissue classification. *Cancer Letters* 77:183-9, 1994.

[106] Kestener P, Lina JM, Saint-Jean P, Arneodo A. Waelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image Analysis and Stereology* 20:169-74, 2001.

[107] Bocchi L, Coppini G, Nori J, Valli G. Detection and clustered microcalcifications in mammograms using fractals models and neural networks. *Medical Engineering & Physics* 26:303-12, 2004.

[108] Chen CC, Daponte JS, Fox MD. Fractal features analysis and classification in medical imaging. *IEEE Transactions on Medical Imaging* 8:133-42, 1989.

[109] Kuklinski WS. Utilization of fractal image models in medical image processing. *Fractals* 2:363-9, 1994.

VITA

Seonghye Jeon is a Ph.D. candidate in Industrial & Systems Engineering (ISyE) at Georgia Institute of Technology with a specialization in statistics. She joined the Ph.D. program in 2007 and was taken under the advisement of Dr. Brani Vidakovic a year later. Her research interests include high dimensional data classification, data mining, clinical trials meta-analysis, spatial interpolation and wavelets. She has been working with the Centers for Disease Control and Prevention (CDC) since September 2009 as an ORISE research fellow. She earned both her bachelors and masters degree in ISyE from Korea Advanced Institute of Science and Technology (KAIST), in 2004 and 2007.