Alastair Patrick Thomas

Spatial Inference of Coseismic Slip on the Cascadia Subduction Zone

Any undertitle is written here

Overall comments:

Define terms, especially ones statisticians won't understand, since a statistician will grade your thesis (I select a grader, right?).

Could add in section on data processing, + subsections on SPDE method, Single vs. Multi-event models + likelihood. Careful with your descriptions ported over from the literature, change them and make them your own. Convert plots to pdfs so they are less pixellated, and move scales to side so plots take less space.

Change notation so $E[y|\ldots] = \mu$ not u, & $\mu = G\,T\exp\{\alpha I + Ax\}$

α not μ.

Overall, nice work, but it can definitely still be improved.

Master's thesis in Mathematics
Supervisor: John Paige
June 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematics

NTNU

ABSTRACT

Write an abstract/summary of your thesis, and state your main findings here.

A summary should be included in both English and any second language, if this is applicable, regardless if the thesis is written in English or in your preferred language. These should be on separate pages, the English version first.

PREFACE

Write the preface of your thesis here.

You may include acknowledgements and thanks as part of your preface on this page, or you may add it as a new chapter after the preface.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

The following is an alphabetical list of abbreviations used throughout this thesis:

- **CSZ** Cascadia Subduction Zone
- **GADM** Global Administrative Areas

CHAPTER

ONE

INTRODUCTION

Earthquakes are a natural occurence which can affect everybody, but especially certain zones around the globes. These zones are usually centred around the boundary of tectonic plates [**empty citation**]. One doesn't have to look too far back in history to see the devastation that earthquakes can wreck, the Turkey 2023 [**empty citation**] caused a death doll of over 65,000 with 1.5 million left homeless. This was the result of just an earthquake, however when an earthquake triggers a tsunami the damage can be even worse. The Japan earthquake and tsunami of 2011.

- Need to define all the terms in here

We don't currently understand the slip distribution. How much damage an earthquake could cause. Talk about work that has already been done and why what I'm doing can help

Literature Review

- What. Subduction zone, plate under each other. - Where - History - Map of plates

Slip distribution -> tsunami -> damage limitation.

1.1 Cascadia Subduction Zone

Coseismic subsidence is simply a measure, usually in meters, of how far the earth's surface sunk because of an earthquake. It is also possible to observe coseismic uplift, when the earths surface rises, as well as horizontal displacement

1.2 Project Description

1.3 Thesis Summary

CHAPTER

TWO

DATA

define in Ch1

"which" requires comma beforehand, "that" doesn't

In this chapter the data sources which are used when modelling the slip distribution along the CSZ are presented and explained. There are two key data sets ~~which~~ that have been used. The first is a compilation of coseismic displacement estimates ~~which are based on comparisons between coastal marsh displacements and a dislocation model predictions data set~~ [1]. The second data set ~~is that which~~ gives the geometry of the CSZ and is the Cascadia subset of Slab2 [2]. Other data sets related to the geological activity of the CSZ do exist, for example turbidite records [3] [4], however for now the work will focus on the aforementioned data sets.

define w/ "subsidence" + "uplift" + "coseismic"

define briefly

not [3,4] or [3-4]?

2.1 Coseismic Subsidence

add more citatations, since you say it is "well documented"

It is well documented that there have been many megathrust events along the CSZ throughout history, [3]. Of these the events, the ones from the past ∼ 6500yrs have left several measurable geological markers along the North American coast which are able to be studied. Leonard et al. [1] ~~have~~ compiled a data collection of coseismic subsidence estimates. The data set is summarised in Figure 4.3.1.

2.1.1?

Coseismic subsidence can be estimated/occurs due to

~~These subsidence estimates are made~~ from the repeated burying of peat-mud couplets[1] in marsh sediments along the Cascadia coastline during megathrust events. During said events, coseismic subsidence drops the marsh down into lower intertidal areas, thus becoming buried soils. In the time between seismic events, gradual uplift, as well as sediment build up from tidal activity bring the soils back up to a higher intertidal zone where the development of organic-rich soil is facilitated. Each of these intertidal zones, as classified by elevation, has its own unique collection of organic material which mainly depends on how tolerant the organisms are to tidal exposure. The amount of coseismic subsidence, is estimated by the elevation difference between each buried soil layer and the sediment lying directly over it. It has been assumed that the sediment overlaying the buried soils gives the paleoelevation prior to a megathrust event. It has also been assumed that when the earth subsided during a megathrust event, it was covered within a

this is directly from Leonard et al. Change the wording to make it your own!!

define or remove

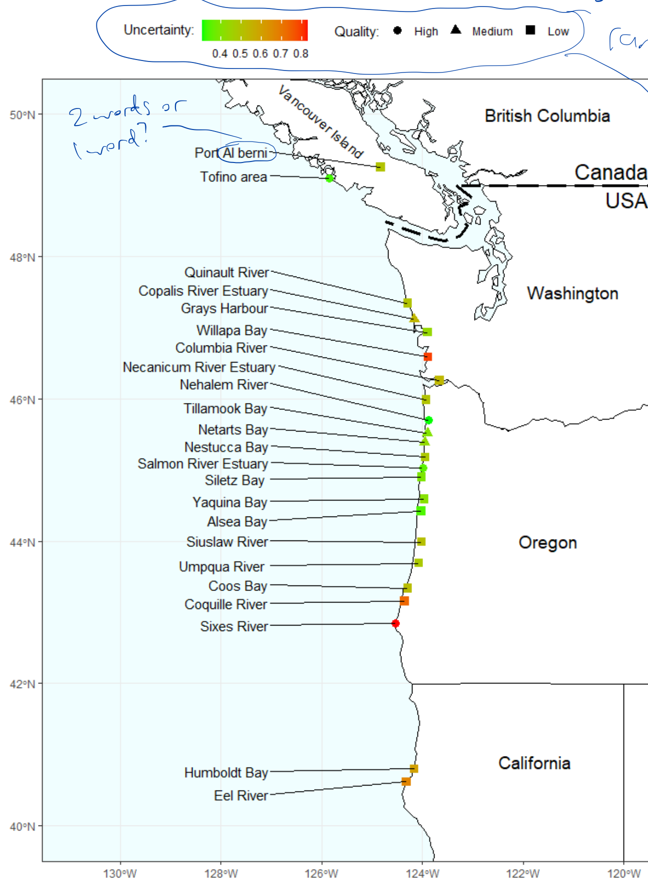
[1] Peat-mud couplets are alternating layers of peat and mud deposits.

nice!

define

few weeks by tidal deposits and thus the effect of uplift in this short period was insignificant.

discuss dataset details here (e.g. # of sites, # earthquakes, range of # of observations per site).



2 words or 1 word?

nice figure!

Figure 2.1.1: Illustration of the ~~locations~~ sites ~~at which~~ where subsidence ~~data has~~ estimates have been collected along the CSZ. Each location is coloured by the mean of the uncertainty in the recorded subsidence's at that location and the shape of each location site represents the mode of the quality of samples there.

subsidence estimates

define "error" here. MOE? SD? 1.96·SD?

The error in the measurements comes from the estimation of the paleoelevation, which is done by evaluating the organic content as well as macrofossil and microfossil content in each layer. Most of the data comes from measurements based on organic material and macro fossil analysis which usually results in an error of $\pm 0.5 - 0.8m$. However some sites have microfossil samples available, which combined with transfer function analysis [5] can reduce the error to $\pm 0.2 - 0.3m$. Leonard et al. [1] have therefore categorised the ~~data points~~ estimates into three distinct quality groups:

define

don't italicize. Use reg. text m or \mathrm{m}

1. High Quality - Estimates based on microfossil analysis.

2. Medium Quality - Estimates based on organic content with good macrofossil analysis.

3. Low Quality - Estimates based mainly on organic content.

don't these also have to do with dating macrofossils on 1 side vs 2 sides of sediment layers?

This is 1700 event sites/subsidence estimates, the most of any EQ in the dataset.
& has

displacement/subsidence estimate

In the data set, each ~~data point~~ has been assigned an "event", labelled as either $T1, T2, ..., T12$ or any of these with an a or b appended to the end. The events without an a or b represent full, or nearly full margin ruptures[2]. The events with an a or b at the end represent earthquakes which affected only some of the margin. To label each subsidence record with a specific coseismic event, Leanord et al used published radiocarbon dates for soils to calculate an age distribution using OxCal [6] as well as correlating the subsidence data with turbitide records[3] [4]. For more details please see Figure 4 from [1].

briefly describe

displacement estimates sites

This data set compromises of 523 ~~data points~~ spread over 23 ~~locations~~, 12 full margin ruptures and 9 partial margin ruptures. Figure 4.3.1 shows that these sites run the length of the Cascadia coastline, starting on Vancouver Island in the north at Port Alberni, running all the way down to Northern California in the south at Eel River. It is also evident from Figure 4.3.1 that the spatial locations of ~~data points~~ the sites needs to be considered carefully. Firstly the data is collected almost exclusively along the coastline, however it can be seen from Figure 2.2.1 that the CSZ extends with longitude ? throughout the region. Furthermore the Northern and Southern extremes of the CSZ could be under-represented in the data set as there ~~is~~ are only 2 ~~locations~~ sites in Northern California, and similarly there ~~is~~ are only 2 ~~locations~~ sites on Vancouver Island.

We will only use the full ruptures.

Port Allberni

2.2 Cascadia Subduction Zone Geometry

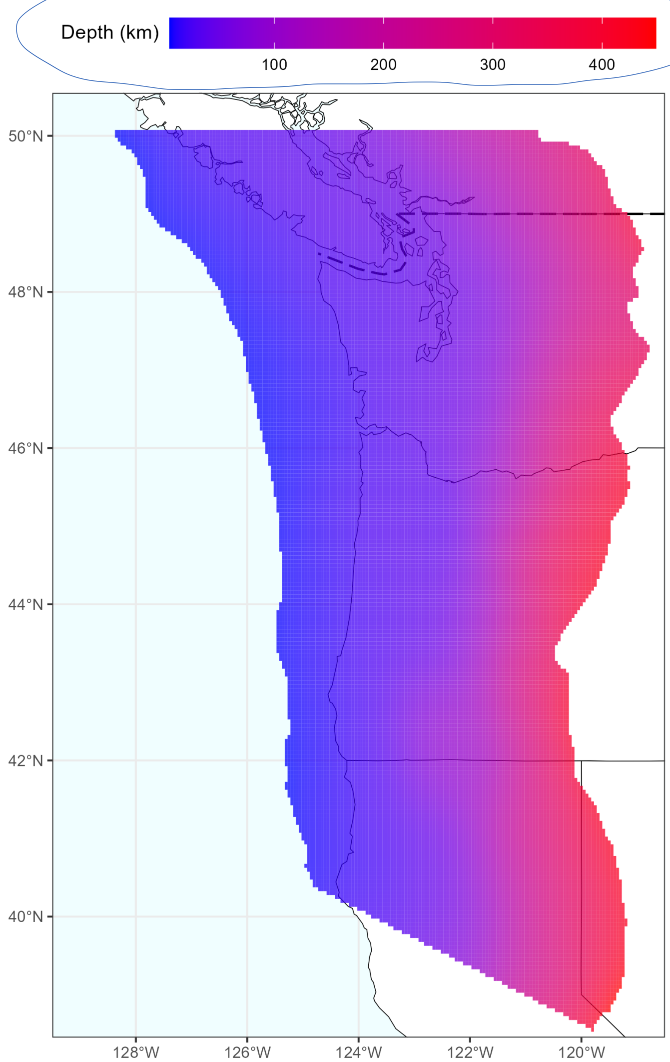
In order to model the slip distribution of the CSZ during a megathrust event, it is essential to obtain an accurate geometry of the CSZ. Most important to the modelling presented in this thesis is to obtain the depths across the entirety of the subducting plate. This is because the underlying slip distribution will be calculated across a finite triangular mesh which is created to encapsulate the 3D fault geometry of the CSZ. Furthermore, the Okada model [7], which relies on the fault geometry will be used to calculate subsidence values at the surface, thus allowing for inference about the slip distribution from the subsidence data set presented in Section 2.1. An ~~good~~ accurate and high resolution geometry is therefore critical.

In 2018, Hayes et al. published an updated geometrical model of all the subduction zones in the world [2] which has been named "Slab2". This presents the detailed geometry of more than 24 million km^2 of subducted zones, of which the CSZ is included. They have included data from active-source seismic data interpretations, receiver functions, local and regional seismicity catalogs (from both regional networks and relocation studies), and seismic tomography models and as such claim that the Slab2 model gives a better geometry than other available subduction zone geometries, especially at shallower depths. This is particularly important in the modelling done here, as the inference will be on a fault geometry that is limited to the shallower depths of the CSZ. As specifically mentioned to model the CSZ, tomographic data was used to extend the geometry to greater depths than had been previously modelled.

unclear. Depends on what you mean (and what they/Slab2 means) by "shallow".

[2] A full margin rupture is a seismic event (earthquake) which effects the entirety of the tectonic plate boundary.

The specific Slab2 data set used to create a 3D triangular mesh (as described in Section 4.2) for the modelling in this thesis can be found online [8]. From this the depths and across a given coorindate grid data are used. The Slab2 data for the CSZ can be seen in Figure 2.2.1. This roughly shows that the CSZ is shallowest on the western side and increases in depth as one travels eastwards over the CSZ.



shallower

deeper

Figure 2.2.1: Illustration of the point depths for the CSZ as given by the Slab2 [2] data set.

2.3 ~~Data Plotting~~ Country and State Boundaries

As this thesis is concerned with the spatial analysis of the CSZ, it is only natural that the majority of the data, methods and results will be best presented in terms of their spatial coordinates, overlayed on top of a base map of the CSZ. To create a basemap, data was taken from the GADM database [9], which gives the state and province boundaries of America and Canada respecitively. This was combined with data about the US-Canada border [10] and plotted using the ggplot2 [11] package in R.

2.4 Data Processing

- "Uncertainties" is subsidence data
- Triangulation of fault geometry
- Removing non full margin EQs
- (Removing or adding sites
  if they were bad / if they are new since Leonard et al.)
- Rakes are assumed to be 90 degrees, i.e. downdip. (or maybe this should be in the next Ch)

?

CHAPTER

THREE

THEORY

In this chapter the theoretical building blocks of the modelling done for the CSZ are introduced. Firstly a surface deformation model, the Okada model [7], which is used to take the slip distribution to the surface subsidence data is described. Then a spatial model, which aims to capture the underlying distribution of tectonic movements in the CSZ during megathurst events, is obtained. Finally the negative log likelihood of the model is derived, thus enabling inference about model parameters from the ~~subsidence~~ data presented in ~~Section~~ 2.~~1~~.

Chapter 2

3.1 Okada Model

megathrust events

A key step in the modelling of ~~the~~ CSZ is to link the coseismic slip to the coseismic subsidence. This step enables the spatial model of the underlying slip distribution to be fitted to the subsidence data set (Section 2.1) and is achieved through the application of the Okada model [7]. based on the fault geometry (Section 2.2).

The Okada ~~This~~ model takes the ~~an~~ underlying fault geometry, which is described here as a triangular mesh of subfaults, whereby each subfault is defined by its geometric parameters, most notable y the locations and depths of the corners of the subfaults. It is assumed that the fault is embedded in an elastic[1], homogeneous[2] and isotropic[3] half-space[4].

Given the average coseismic slip amount and direction over each subfault, ~~A slip distribution is then assumed over the fault, with each individual subfault having a given value of slip.~~ The Okada model then uses Green's functions to describe the response of the half-space to point sources of deformation (coseismic slips). This will give the contribution of each individual subfault to the vertical and horizontal movement at a grid of points on the earths surface. Finally, the Okada model is linear, so the total movement at each point on the surface grid can be calculated by summing up the contributions of each subfault.

nice

[1] An elastic space is one which returns to it's original state once external forces are removed.

[2] In a homogeneous space, all the properties of the material are assumed to be uniform throughout.

[3] Isotropic means that the response to any external forces will be the same in any direction.

[4] A half-space is one in which only a certain section of the space is considered for analysis.

Spatial Model for Coseismic Slips: Single Event

3.2 ~~Slip Distribution Model~~

add subsection 3.2.1 Model Formulation

The spatial model presented here describes just one mega a thrust event. For example the ~a 9.0 magnitude earthquake of 1700 could be modelled here. As introduced in Section 2.1, a data set of subsidence values, with associated standard deviations is available. From this data set, the data points from any singular earthquake are selected, let this be N subsidence values, and subsequently N standard deviation values. Let the vector $\boldsymbol{y} \in \mathbb{R}^N$ be the vector of subsidence ~~data points~~, and $\boldsymbol{\sigma} \in \mathbb{R}^N$ be the vector of standard deviations. sub/SD estimates

Subsidence can be ~~calculated~~ predicted when the coseismic slip is known via the Okada model [7]. The slip values are given as an average across each of the K triangular subfaults, ~~created~~ discussed in Section 4.2. ~~Therefore~~ Let $\boldsymbol{s} \in \mathbb{R}^K$ be the random vector of these average ~~slip~~ values. To take the subsidence ~~values~~ predictions calculated by the Okada model to the observed subsidence values, an additive error, $\boldsymbol{\epsilon} \in \mathbb{R}^N$, based on the observed standard deviation data can be added. Thus it is possible to write the equation for observed subsidence as:

$$\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\epsilon} = G\boldsymbol{s} + \boldsymbol{\epsilon} \tag{3.1}$$

whereby each term $\boldsymbol{\epsilon}$ is assumed to be independently normally distributed with a mean of 0 and variance given by the corresponding value in $\boldsymbol{\sigma}$ squared. That is:

$$\epsilon_i|\sigma_i^2 \sim \mathrm{N}\left(0, \sigma_i^2\right). \tag{3.2}$$

The Okada matrix $G \in \mathbb{R}^{N \times K}$ is the matrix calculated from the Okada model with a ~~uniform~~ unit slip over the entire fault, giving deformation at the locations of the N data points. The predicted subsidence given the set of average slips is then Gs.)

The slip vector can be thought of as being the transformation of an underlying spatial distribution X. Given that $X = \boldsymbol{x}$ the slips will be given by:

Define X in words, make bold

$$\boldsymbol{s}|\lambda, \cancel{\mu}\,\alpha, \boldsymbol{x} = T\exp\left(\cancel{\mu}\,\alpha\boldsymbol{1} + A\boldsymbol{x}\right). \tag{3.3}$$

define A

$T \in \mathbb{R}^{K \times K}$ is the diagonal taper matrix which is added to gradually decreases the spatial slip values near down-dip edge of the fault (the deeper, eastern edge) to provide a more realistic representation of how the the slip diminishes towards the edge of the fault zone. The elements of T are calculated as: Say $T = T(\lambda)$

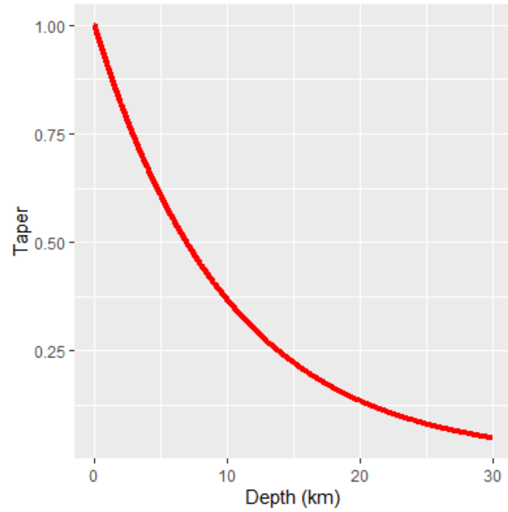
support w/ citations

$$T_{i,j} = \begin{cases} t(d_i|\lambda) & i = j \\ 0 & i \neq j \end{cases} \tag{3.4}$$

where d_i is the depth of the centroid of the i^{th} subfault in km and t is a taper function. → taking values in (0, 1]
The taper function used is

$$t(d|\lambda) = \exp\left(-\lambda d\right), \tag{3.5}$$

where λ is a decay rate parameter ~~which needs determining~~ to be estimated. Setting $\lambda = \frac{1}{10}$ gives the taper shown in Figure 3.2.1. ~~It is seen that~~ this is 1 ~~at the shallowest part of the CSZ~~ at depth 0 and tapers to nearly 0 by 30 km.



Make a subsection for constructing SPDE mesh.

Figure 3.2.1: An example of the taper function with the free parameter fixed as $\lambda = \frac{1}{10}$.

Say earlier.

make more precise & expand on this:
- X is the coefficients not the spatial field
- What is the spatial field?
- what are the basis functions?
- $u(s) = \sum_{k=1}^{K} X_k \phi_k(s)$
- What exactly has a Matérn covariance?

In (3.3) ~~μ~~ α represents the median log untapered slip and $A \in \mathbb{R}^{K\times K}$ is a projection matrix. In order to model the spatial field, an SPDE representation will be used [12]. This requires a triangular grid over which polygonal basis functions are defined. This grid ~~will be~~ is made such that the ~~corners~~ basis function centers align to the ~~centers~~ centroids of the triangular subfaults. Thus, to project from SPDE function coefficients to slip values, each row in A will be 0 except in that column which corresponds to the correct subfault centroid, whereby the value will be 1.

bold

Finally it is assumed that the underlying spatial distribution X, which is a spatially correlated vector of polygonal basis function coefficients, is Gaussian with $\mathbf{0}$ mean and a Matérn covariance function:

define mathematically

$$X = \boldsymbol{x}|\kappa,\tau \sim \text{MVN}\left(\mathbf{0}, \frac{1}{\tau}Q_{spde}^{-1}(\kappa)\right). \tag{3.6}$$

The Matérn covaraince usually also relies on ν, which is the order of Bessel function used in the Matérn function. The system is solvable for $\nu \in \{1, 2\}$ [12]. $\nu = 1$ is chosen here. Since the SPDE mathod is used the Matérn function is not actually required, however κ, an inverse range parameter, is still required to calculate the approximation. The marginal precision of the spatial field is given by τ.

~~3.3~~ 3.2.2 Negative Log Likelihood

The negative log likelihood is required to fit the model parameters to the data given. Here the observations $\boldsymbol{y}$ ~~and $\boldsymbol{\sigma}$~~ are given, the latent variable $\boldsymbol{x}$ gives the underlying distribution, and parameters $\boldsymbol{\theta} = \{\kappa, \tau, \alpha, \lambda\}$ define the model. The negative log likelihood is given as $-\ell = -log(\boldsymbol{L})$, where $\boldsymbol{L}$ is the likelihood. ~~The likelihood required can also be thought of as being the posterior distribution $p(\boldsymbol{x}, \boldsymbol{\theta}|\mathbf{y})$. In a Bayesian setting, which follows later in this thesis, this posterior is given as:~~

don't bold L

Leave Bayesian interpretation for later or go all in on Bayesian interpretation & leave out freq. interpretation.

freq. interp: joint density:

$$\boxed{p(y,x|\theta) = p(y|x,\theta)\,p(x|\theta)}$$

Bayes: post. dens = $p(\theta, x|y) \propto p(y|x,\theta)p(x|\theta)p(\theta)$

$$p(\boldsymbol{x}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{3.7}$$

In this frequentist setting $\boldsymbol{L}$ is the same, except the prior distribution in theta, $p(\boldsymbol{\theta})$ is disregarded. Therefore the likelihood is given as:

$$\boldsymbol{L} = p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}). \tag{3.8}$$

There are two distinct parts here: $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ and $p(\boldsymbol{x}|\boldsymbol{\theta})$. To find $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, one must look towards (3.1). It is seen that each element of $\boldsymbol{y}$ is a linear combination of a constant and a normally distributed variable with mean 0 and variance σ_i^2, $i=1,\dots,N$. Each element of $\boldsymbol{y}$ is therefore independently normally distributed, with the mean vector given by:

assumed to be conditionally

I think using μ & α in this way is more in line w/ GLMs, easier to remember/understand

$E[y|\dots]$ $\mu(\dots)$ α

$$\mathbb{E}[\boldsymbol{y}] = \boldsymbol{u} = G_{\boldsymbol{y}} T \exp\left(\mu\mathbf{1} + A\boldsymbol{x}\right), \tag{3.9}$$

and the standard deviations given by $\boldsymbol{\sigma} = \{\sigma_1, ..., \sigma_N\}$. Since each observation is independent, the combined likelihood, $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, is the product of N normal probabilities:

$$\mu(\dots) := \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}$$

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \left[\frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y_i - u_i}{\sigma_i} \right)^2 \right) \right]. \tag{3.10}$$

μ_i

be consistent, use boldsymbol

The second part $p(\boldsymbol{x}|\boldsymbol{\theta})$, is the density function of the spatial field X, which was assumed to be multivariate normally distributed, and thus has the pdf: cite (3.6)

$$p(\mathbf{x}|\boldsymbol{\theta}) = (2\pi)^{\frac{K}{2}} \left| \frac{1}{\tau} Q_{spde}^{-1}(\kappa) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau} \mathbf{x}^T Q_{spde}(\kappa) \mathbf{x} \right). \tag{3.11}$$

~~In reality this density is calculated through the SPDE approximation and not directly.~~ ← no, this density is itself the approximation, and it is calculated directly

Finally the negative log likelihood can therefore be written as:

This part is unnecessary since we have (3.7)

$$-\ell = -log(\boldsymbol{L}) = -log(p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})). \tag{3.12}$$

where the explicit formula is written out and simplified in Appendix B.

3.3 Spatial Model for Coseismic Slips: Multiple Event

3.3.1 Model Formulation

3.3.2 Neg. Log. Lik.

CHAPTER

FOUR

METHODS

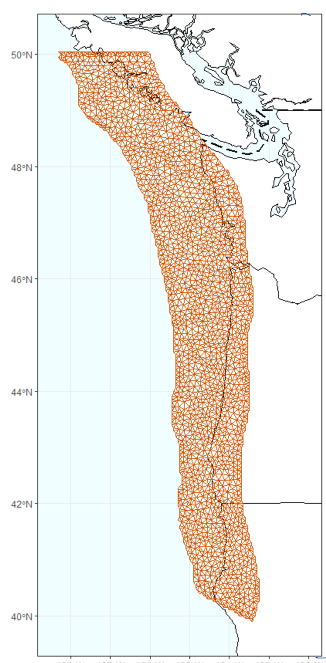
4.1 Okada Model Implementation

Basically say that the R code was created based on the geoclaw package and modified for our useage.

yes, definitely mention + cite geoclaw/ clawpack

4.2 Subfault Geometry

How the subfault mesh was created.



save as pdfs

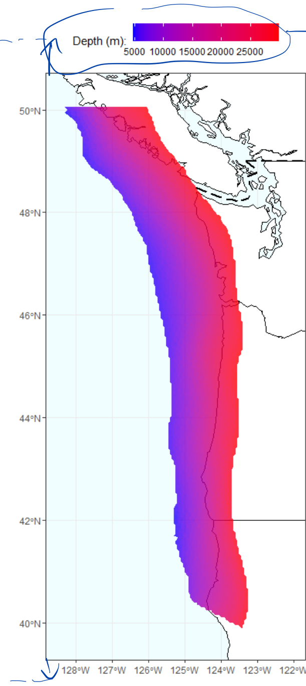


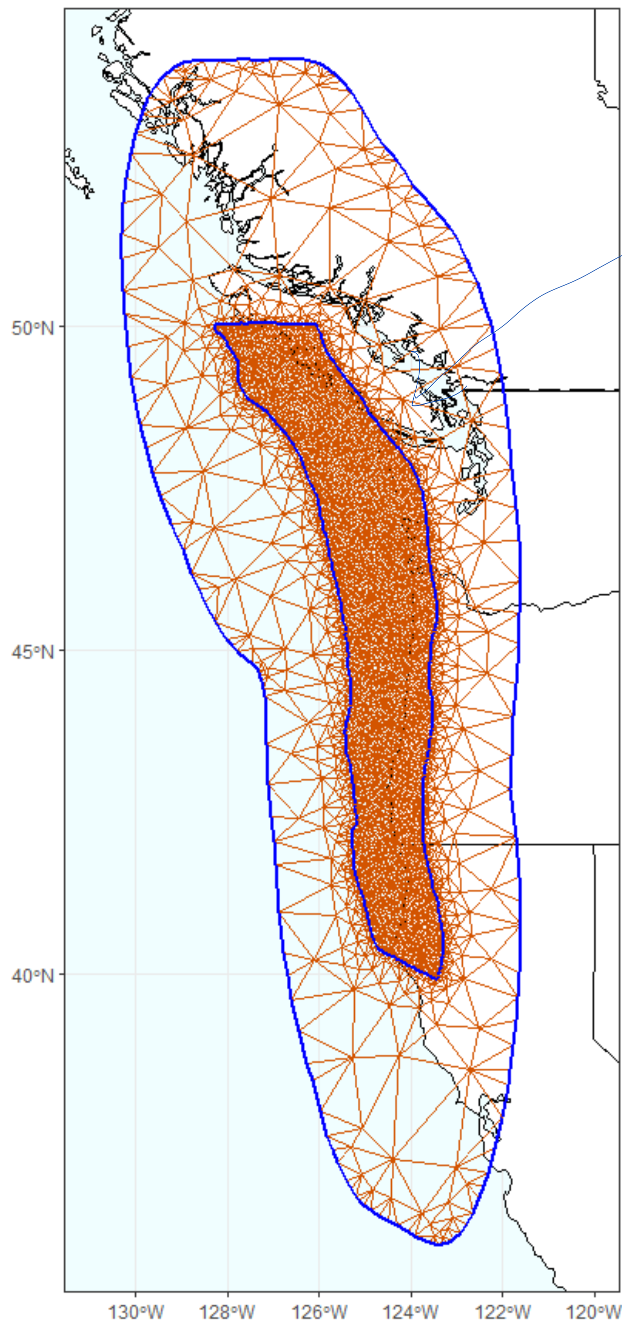
Figure 4.2.1: An illustration of the sub-fault mesh which is used in the Okada model.

Figure 4.2.2: An illustration of the depths at the centers of each sub-fault which is used in the Okada model.

use 1 Figure + 2 subplots or 2 subfigures labelled (a) + (b) if you want to cite them separately

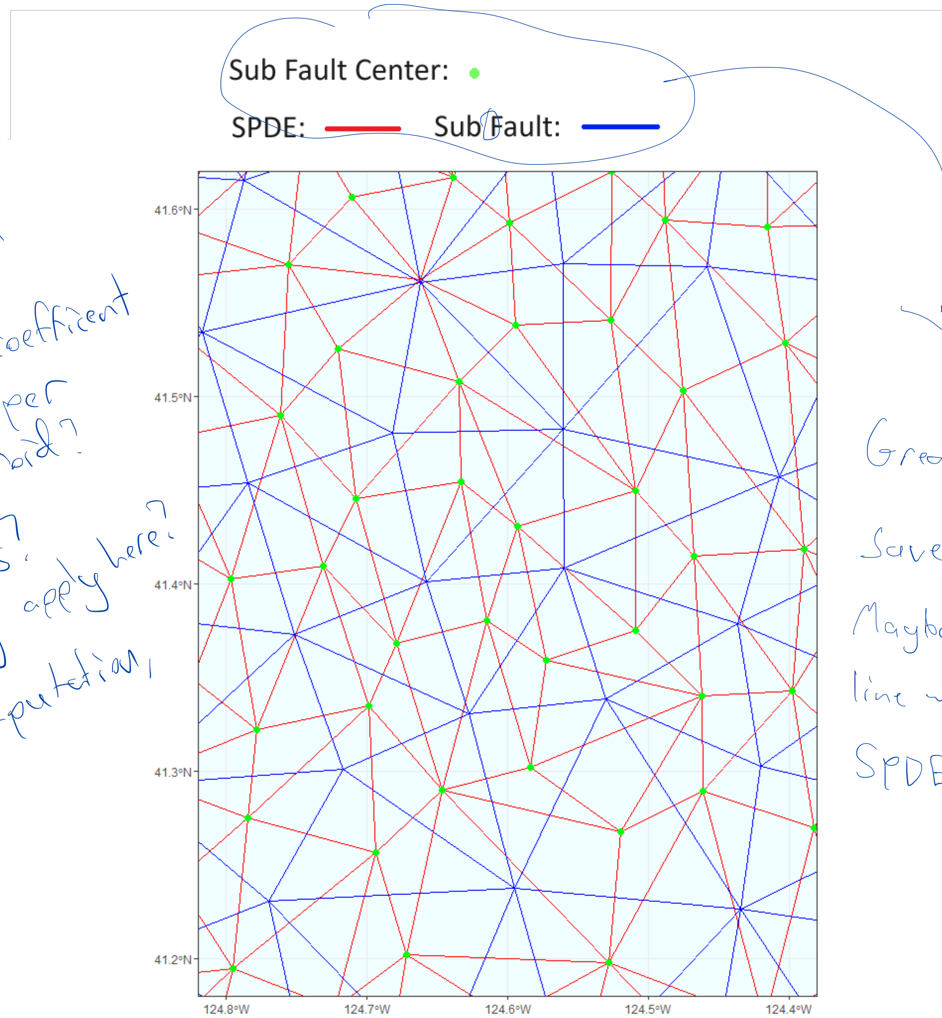
4.3 SPDE

Talk about how the SPDE approximation is used here. Explain the mesh.



save as pdf not png or whatever this format is it is pixellated

Figure 4.3.1: The triangular mesh which the underlying spatial field $\boldsymbol{x}$ is derived over. It has been created so that all the centroids of the subfault's **??** are aligned with corners in this mesh.



What are the advantages of only 1 basis coefficient being needed per subfault centroid? Disadvantages? ↳ do they apply here? Sparsity, computation, etc.

Great figure! Save as pdf. Maybe vary line width btwn SPDE + Subfault?

Figure 4.3.2: The SPDE mesh is shown in red and the sub faults are shown in blue. The ~~centers~~ Centroids of the sub faults are shown with green dots. This illustrates how the two meshes are overlayed, and why just one value from the SPDE approximation can be used to calculate the slip on each sub fault.

$$\begin{pmatrix} \hat{\boldsymbol{y}}^{(1)} \\ \vdots \\ \hat{\boldsymbol{y}}^{(M)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{G}^{(1)} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}^{(2)} & & \vdots \\ \vdots & & \ddots & \\ \boldsymbol{0} & \dots & & \boldsymbol{G}^{(M)} \end{pmatrix} \begin{pmatrix} \boldsymbol{S}^{(1)} \\ \vdots \\ \boldsymbol{S}^{(M)} \end{pmatrix}$$

4.4 Negative Log Likelihood in TMB

I will write this section once I have coded the model in TMB and I understand it more.

It should explain (at least vaguely) what TMB is, and why it is useful, particularly how it uses the negative log likelihood previously derived, and that it uses automatic differentiation to aid in the optimization process. You should discuss how it evaluates f(y|theta) by first optimizing f(y, x| theta) over x, and then using a Laplace approximation to approximate f(y|theta) = int f(y, x | theta) dx. Again, this does not have to be too specific about the inner-workings of TMB.

CHAPTER

FIVE

RESULTS

CHAPTER

SIX

DISCUSSION

Discuss your results here.

6.1 Future work

Include a section about what should or could be done in future research, or explain any recommended next steps based on the results you got. This should be the last section in the discussion.

CHAPTER

SEVEN

CONCLUSIONS

Give a concise summary of your research and finding here, and include a short summary of any future work as well.

REFERENCES

[1] Lucinda J Leonard et al. "Rupture area and displacement of past Cascadia great earthquakes from coastal coseismic subsidence". In: *Bulletin* 122.11-12 (2010), pp. 2079–2096.

[2] Gavin P Hayes et al. "Slab2, a comprehensive subduction zone geometry model". In: *Science* 362.6410 (2018), pp. 58–61.

[3] Chris Goldfinger et al. "Holocene earthquake records from the Cascadia subduction zone and northern San Andreas fault based on precise dating of offshore turbidites". In: *Annual Review of Earth and Planetary Sciences* 31.1 (2003), pp. 555–577.

[4] Chris Goldfinger et al. "Late Holocene rupture of the northern San Andreas fault and possible stress linkage to the Cascadia subduction zone". In: *Bulletin of the Seismological Society of America* 98.2 (2008), pp. 861–889.

[5] AD Hawkes et al. "Late Holocene coseismic deformation at the Cascadia subduction zone, Oregon, USA". In: *AGU Fall Meeting Abstracts*. Vol. 2008. 2008, T53B–1937.

[6] Christopher Bronk Ramsey. *OxCal (Version 4.0)*. Software. `https://c14.arch.ox.ac.uk/oxcal/OxCal.html`. 2001.

[7] Yoshimitsu Okada. "Surface deformation due to shear and tensile faults in a half-space". In: *Bulletin of the seismological society of America* 75.4 (1985), pp. 1135–1154.

[8] U.S. Geological Survey. *ScienceBase Catalog*. Accessed: [02/10/23]. Accessed 2023. URL: `https://www.sciencebase.gov/catalog/item/5aa312cde4b0b1c392ea3ef5`.

[9] GADM. *Global Administrative Areas*. Accessed: [02/10/23]. Accessed 2023. URL: `https://gadm.org/`.

[10] Homeland Infrastructure Foundation-Level Data (HIFLD). *Canada and US Border Dataset*. Accessed: [02/10/23]. Accessed 2023. URL: `https://hifld-geoplatform.opendata.arcgis.com/datasets/canada-and-us-border/explore`.

[11] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: `https://ggplot2.tidyverse.org`.

[12] Finn Lindgren, Håvard Rue, and Johan Lindström. “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73.4 (2011), pp. 423–498.

APPENDICES

APPENDIX

A

GITHUB REPOSITORY

All code and ~~latex~~ LaTeX-files used in this document are included in the Github repository linked below. Further explanations are given in the readme-file.

Github repository link

- `https://github.com/alastair-thomas/Cascadia`

APPENDIX

B

NEGATIVE LOG LIKELIHOOD

Is this Appendix necessary? Maybe if you move parts of nll calculations here so nothing is repeated.

An explicit formula for the negative log likelihood is derived here. The two defining equations (3.10 and 3.11) of the nll are:

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{N}\left[\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_i-u_i}{\sigma_i}\right)^2\right)\right].$$

$$p(\mathbf{x}|\boldsymbol{\theta}) = (2\pi)^{\frac{K}{2}}\left|\frac{1}{\tau}Q_{spde}^{-1}(\kappa)\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\tau}\mathbf{x}^T Q_{spde}(\kappa)\mathbf{x}\right).$$

The nll is defined as:

$$\begin{aligned}\mathcal{L} &= -log(p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})) \\ &= -log\prod_{i=1}^{N}\left[\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_i-u_i}{\sigma_i}\right)^2\right)\right] \\ &\quad\times(2\pi)^{\frac{K}{2}}\left|\frac{1}{\tau}Q_{spde}^{-1}(\kappa)\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2\tau}\mathbf{x}^T Q_{spde}(\kappa)\mathbf{x}\right).\end{aligned}$$

This can be simplified as:

$$\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{N} \left(log \left[\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y_i - u_i}{\sigma_i} \right)^2 \right) \right]^{-1} \right) \\
&\quad + log \left[(2\pi)^{\frac{K}{2}} \left| \frac{1}{\tau} Q_{spde}^{-1}(\kappa) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau} \mathbf{x}^T Q_{spde}(\kappa) \mathbf{x} \right) \right]^{-1} \\
&= \sum_{i=1}^{N} log \left[\sigma_i \sqrt{2\pi} \right] + \frac{1}{2} \left(\frac{y_i - u_i}{\sigma_i} \right)^2 \\
&\quad + log \left[(2\pi)^{\frac{K}{2}} \left| \frac{1}{\tau} Q_{spde}^{-1}(\kappa) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau} \mathbf{x}^T Q_{spde}(\kappa) \mathbf{x} \right) \right]^{-1} \\
&= N log(\sqrt{2\pi}) + \sum_{i=1}^{N} log(\sigma_i) + \frac{1}{2} \sum_{i=1}^{N} \left(\frac{y_i - u_i}{\sigma_i} \right)^2 \\
&\quad + log \left[(2\pi)^{\frac{K}{2}} \left| \frac{1}{\tau} Q_{spde}^{-1}(\kappa) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau} \mathbf{x}^T Q_{spde}(\kappa) \mathbf{x} \right) \right]^{-1}.
\end{aligned}$$

The term for $p(\boldsymbol{x}|\boldsymbol{\theta})$ is left un-simplified because this term is calculated via the SPDE approximation for computational efficiency.

HOLDER

- $\mathbf{y} \in \mathbb{R}^N$ - The subsidence data.
- $\lambda \in \mathbb{R}$ - longitude.
- $\phi \in \mathbb{R}$ - latitude.
- $d \in \mathbb{R}$ - depth (m).
- $\mathbf{z} \in \mathbb{R}^{N\times 2}$ - The location of the subsidence data. (λ, ϕ).
- $F \in \mathbb{R}^{M\times 3\times 3}$. The collection of triangular subfaults. One subfault $F_m = [(\lambda_1, \phi_1, d_1), (\lambda_2, \phi_2, d_2), (\lambda_3, \phi_3, d_3)]$.
- $\mathbf{f} \in \mathbb{R}^{M\times 3}$ - The centers of all the subfaults. This is where the slips will be calculated. Each subfault centroid is $f_m = (\lambda, \phi, d)$: the mean of the corners.
- $\mathbf{d} \in \mathbb{R}^3$ - The depth of the centroid of all the subfaults.
- $S \in \mathbb{R}^M$ - The random vector of slips. Each slip will be at the centroid of a triangular subfault.
- $\mathbf{s} \in \mathbb{R}^M$ - The average slip over each triangular subfault. Calculated at the centroid of the subfaults.
- $G_z \in \mathbb{R}^{N\times M}$ - The Okada matrix which takes the slips vector to a subsidence vector. Depends on the locations of the subsidence data $\mathbf{z}$. Calculated from the Okada model [7].
- $\epsilon \in \mathbb{R}^N$ - The errors in the subsidence data measurements. Where ϵ_i is one component.
- μ - Median untaper slip.
- $\mathbf{x} \in \mathbb{R}^K$ - Basis function coefficients, modelled via the spde approach.
- $T \in \mathbb{R}^{M\times M}$ - The taper matrix. $T_{i,i} = t(d_i|\lambda)$.
- t - The taper function.

- A - Projection matrix from the SPDE coefficients to the centroids of the fault geometry. The basis functions are chosen so that their corners correspond to the center of each subfault. Therefore the rows of A be 0 everywhere except in that column whereby the polygonal basis function relates to that subfault. Here it will be 1.
- $\frac{1}{\tau}$ - Marginal precision of the spatial field.
- Q_{spde} - Precision matrix for the polygonal basis function coefficients.
- σ_i^2 - Variance parameter for each measurement. Given.
- θ_s - The parameters in the spatial random field that we need to do inference on.