

# A spatial statistical model for full-margin Cascadia Subduction Zone earthquakes

J. Paige · P. Guttorp · D. A. Schmidt

Received: date / Accepted: date

**Abstract** There is risk of a magnitude 9 earthquake on the Cascadia Subduction Zone, yet few likelihood-based spatial models have been developed for the coseismic slip of such earthquakes. Current models either use a handful of predetermined earthquakes to represent the full range of those possible, or do not use readily available GPS-based fault locking and paleoseismological subsidence data. This study presents a set of stochastic spatial statistical models of spatial coseismic slip distributions that accounts for these available observations. The presented models can simulate future earthquakes and also exploit known properties of stochastic processes in order to infer the statistical distributions of coseismic slip and subsidence of historical earthquakes such as the 1700 event. The models consistently show microfossil-based subsidence estimate standard errors should be inflated by 1.25 times while the macrofossil-based estimate standard errors should be inflated by 1.75 times, and 95% confidence intervals for earthquake magnitudes for the different models are contained in the interval (8.52, 9.37). However, the varying assumptions of the models lead to different spatial slip distributions for earthquakes when little subsidence data is available, such as would be the case for future earthquakes without subsidence observations. This is especially true in the southern

---

We would like to thank the NSF GRFP (grant number DGE-1256082), and its DHS and EAR divisions (grants DMS-1106862, DMS-1401793, and EAR-1331412). Paige did the main research for the paper, and wrote the first draft. Guttorp and Schmidt contributed equally to the editing of the manuscript, and are listed alphabetically.

J. Paige  
Department of Statistics, University of Washington, Seattle, WA  
E-mail: paigejo@uw.edu

P. Guttorp  
Department of Statistics, University of Washington, Seattle, WA  
SAMBA Department, Norwegian Computing Center, Oslo, Norway.

D. A. Schmidt  
Department of Earth and Space Sciences, University of Washington, Seattle, WA

---

portion of the fault, where the GPS-based locking rate estimates tend to disagree with the historical subsidence estimates.

**Keywords** Cascadia subduction zone · earthquake modeling · paleoseismology · spatial statistics

## 1 Introduction

Seismic hazard studies strive to estimate what the next large earthquake will look like on a fault system in order to better understand the likely impacts to regional communities and to aid in mitigation planning. An estimation of the hazard is typically accomplished by using the crustal strain to constrain the spatial distribution and magnitude of locking on the fault as in McCaffrey et al. [14], or by using the earthquake history inferred from paleoseismic data as in Wang et al. [31]. However, observations of contemporary strain or records of past coseismic events are not necessarily a direct proxy for a future earthquake given the variability of rupture scenarios from one event to the next on the same fault Stein et al. [29]. An alternative approach utilizes a logic tree, where various rupture scenarios are assigned and weighted, to produce a probabilistic estimate of likely outcomes [33]. However, these trees are often limited to a small number of predetermined scenarios, can be ad-hoc, and likely do not capture the full range of possible events.

A more robust approach to estimate the seismic hazard is to develop a statistical model of fault rupture that incorporates the actual variability of past rupture events. In this work, we present an approach for constructing a fully probabilistic coseismic slip model using likelihood-based methods. A likelihood-based model in this context is one that involves a likelihood function giving the relative probabilities of observing the data given the model parameters such as in [3, 19]. Likelihood-based models are widely preferred among statisticians for their many useful properties, including providing theo-

retical frameworks for estimating parameter and observation uncertainty, and accounting for correlations in the data [1, Ch. 6-10]. We apply this technique to the Cascadia Subduction Zone (CSZ), which has one of the most detailed rupture histories of any fault system given the rich paleoseismic record [10]. This setting is ideal for likelihood-based models, since although the historical record is long, it has large uncertainty and is highly correlated in space. We use two disparate, but complimentary data sources to constrain the statistical model of fault slip and produce estimated uncertainties: (1) paleoseismic data from along the coastline that record the magnitude of subsidence for 21 previous coseismic events, and (2) the contemporary fault locking rate distribution on the megathrust that is derived from GPS data. This statistical slip model can be used predict the slip distribution of a future megathrust event, or conditioned on data to estimate the slip distribution of past coseismic events.

One class of spatial statistical models that is fully likelihood-based is known as stochastic process models, and the most common stochastic processes used to model spatial data are known as Gaussian processes. Using Gaussian processes or other stochastic process models for spatial data inherently involves spatial correlations: nearby data are often more similar than far away data. This is modeled using a correlation function, often denoted in the isotropic (direction-independent) case by  $\rho(d)$  for some distance  $d$ . Let  $Z(\mathbf{x}) \in \mathbb{R}$  be the value of a Gaussian process  $Z$  at location  $\mathbf{x} \in \mathbb{R}^2$ . Since  $Z$  is a Gaussian process, if  $Z$  has correlation function  $\rho(\cdot)$ , mean  $\mu$ , and variance  $\sigma^2$ , then the joint distribution of  $Z$  at multiple locations,  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , is the following multivariate normal distribution:

$$\begin{pmatrix} Z(\mathbf{x}_1) \\ Z(\mathbf{x}_2) \\ \vdots \\ Z(\mathbf{x}_n) \end{pmatrix} \sim \text{MVN}(\boldsymbol{\mu}, \sigma^2 \mathbf{C})$$

where  $\mathbf{C}$  is a correlation matrix with elements  $C_{ij} = \rho(\|\mathbf{x}_i - \mathbf{x}_j\|)$ . The Gaussian process parameters can be estimated with maximum (log) likelihood or by maximizing the posterior probability of the model parameters in a Bayesian setting. The log likelihood of an  $n$ -dimensional multivariate Gaussian random variable takes the following form for a set of observations,  $\mathbf{Z}$ , if the observations' mean vector and covariance matrix are  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{C}$  respectively:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{Z}) = -n \log(\pi) - (\mathbf{Z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}) - \log(|\boldsymbol{\Sigma}|) \quad (1)$$

where  $|\cdot|$  denotes the matrix determinant.

Under most models,  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  have relatively simple forms depending on a small set of model parameters over which the log likelihood or posterior probability is maximized. For instance, if we take  $\rho(d)$  to be a Matern correlation function, which is perhaps the most widely used family of spatial correlation functions due to its flexibility, then it takes the following form:

$$\rho(d; \nu, \phi) = \frac{\pi^{1/2} \phi^{2\nu}}{2^{\nu-1} \Gamma(\nu + 1/2)} (|d|/\phi)^\nu \mathcal{K}_\nu(|d|/\phi)$$

based on the formulation from page 31 of Stein [28], where  $\phi$  is the spatial scale parameter,  $\nu$  is a smoothness parameter, and  $\mathcal{K}_\nu(\cdot)$  is a modified Bessel function of the second kind or order  $\nu$ . Since the  $\nu$  parameter is poorly identified [5, Ch. 1.3], we will fix  $\nu = 3/2$ , in which case the correlation function takes the following simplified form:

$$\rho(d; \nu = \frac{3}{2}, \phi) = \left(1 + \frac{d}{\phi}\right) \exp -\frac{d}{\phi}.$$

This is faster to compute and has a closed form for its gradient with respect to  $\phi$ , which allows for more robust model fitting.

The observations corresponding to **Z** above will be the paleoseismic subsidence and fault locking rate estimates from Leonard et al. [10] and Evans et al. [4] respectively. Paleoseismic subsidence estimates are based on macro- and microfossil evidence in the soil of coastal marshes and estuaries. Not everywhere all locations along the coast are suitable for collecting such data—interseismic uplift of sandy soils can complicate the preservation of paleoseismic evidence [10]. Determining uncertainties of subsidence estimates can be particularly hard, since it can be challenging or impossible to separate coseismic and post-seismic deformation [30, 31]. Qualities of estimates may be highly variable, depending on all these factors. Hence, uncertainties are likely underestimated. In order to account for this, additional parameters are included in the proposed models in order to estimate by what factor, on average, uncertainties are over- or underestimated. The fitted values of these parameters may be of independent scientific interest.

To constrain the contemporary fault locking rate on the Cascadia megathrust, we use the spatial locking rate data product estimated from [24], who constrained their model using surface deformation captured by GPS data from the last three decades combined from several sources: McClusky et al. [15], Shen et al. [27], Hammond and Thatcher [7], Williams et al. [32], McCaffrey et al. [13], Loveless and Meade [12], and Herring et al. [8]. The data are summarized in Evans et al. [4]. Since the dataset is representative of the integrated locking rates over the last three decades, it is not necessarily proportional to the full amount of locking accumulated over the CSZ megathrust since the 1700 earthquake [30]. Hence, estimates of the locking rates and their uncertainties may have inaccuracies, and should not be considered more similar to a future Cascadia earthquake than any past subduction event.

It is therefore important to model the observations from each dataset with care, and make minimal assumptions about the strength of each observation.

All we require of the locking rate data is that the spatial correlation scale and overall variability of the locking rate estimates are similar (up to a constant of proportionality) to the coseismic slip of an earthquake independent of past historical earthquakes or the coming one. Similarly, we only assume the subsidence estimates are correct on average. Misstated uncertainties are accounted for using methods discussed in more detail in Sec. 2.2.

An outline of this paper is as follows: Sec. 2 discusses the specifics of the datasets used in this analysis along with the assumed CSZ fault geometry. It also presents the proposed spatial statistical models for CSZ slip distributions, how to fit them, and how to use them to estimate the spatial slip distributions of future and historical earthquakes. Sec. 3 shows the model parameters fit to the data, and gives summary statistics about the resulting earthquakes with a focus on two of the proposed models along with cross-validation results. Sec. 4 interprets the results from Sec. 3 within a broader scientific context. Lastly, Appendix A gives some mathematical proofs, and Appendix B gives the log likelihood gradients with respect to the parameters for easy model fitting. Appendix C shows some simulations of the 1700 event, and Appendix D shows summaries of the predictions for the 1700 event for all models.

## 2 Methods

There are three distinct datasets used in this analysis: paleoseismic subsidence estimates from 21 studies along the coast compiled in Leonard et al. [10], a GPS-based locking rate data product from Pollitz and Evans [24], and the discretized fault geometry from Pollitz et al. [25]. These datasets are used together to fit a fully likelihood-based spatial statistical model of coseismic slip that can account for both spatial and historical variation of CSZ full-margin fault ruptures.

## 2.1 Data

Paleoseismic data provide a record of a past coseismic event as in Leonard et al. [10]. In a subduction zone setting, this can include the deposition of sediment along the coast associated with the lowering of the ground relative to mean sea level and/or inundation of sediment-laden seawater with a tsunami, among other possible geological evidence. The paleoseismic data from Leonard et al. [10] contains subsidence estimates, uncertainties, a ‘quality’ indicator, latitude and longitude coordinates, an associated site, and an associated earthquake event, such as the 1700 event, for example. The quality indicator is an integer ranging from 1 to 3 depending on whether the estimate is based on whether the estimate is based on micro- or macro-fossil analyses and the type of macro-fossil analysis, where micro-fossil analysis are considered to be higher quality. Some estimates do not have associated uncertainties, and others do not have associated earthquakes, instead containing a subsidence estimate for an unknown past earthquake. The data are first culled so that any estimates without a specified earthquake event, subsidence, or uncertainty are removed. The resulting dataset contains 523 subsidence estimates ranging in time from the most recent 1700 event to earthquakes occurring nearly 7,000 years ago. While 196 of these are from the 1700 event, the majority are from older earthquakes. The data is shown in Fig. 1. All data and code for this analysis is available at Paige’s Github at <https://github.com/paigejo/M9>.

Fault locking rates on the Cascadia megathrust provide a proxy for where strain is currently accumulating. The locking rate data product used in this analysis is shown in Fig. 2 and is published in Pollitz and Evans [24]. GPS observations from the Western US over the last three decades is used to infer how tectonic plates shift and strain builds up along the CSZ megathrust. Locking rates with associated uncertainties are provided in mm/yr and with

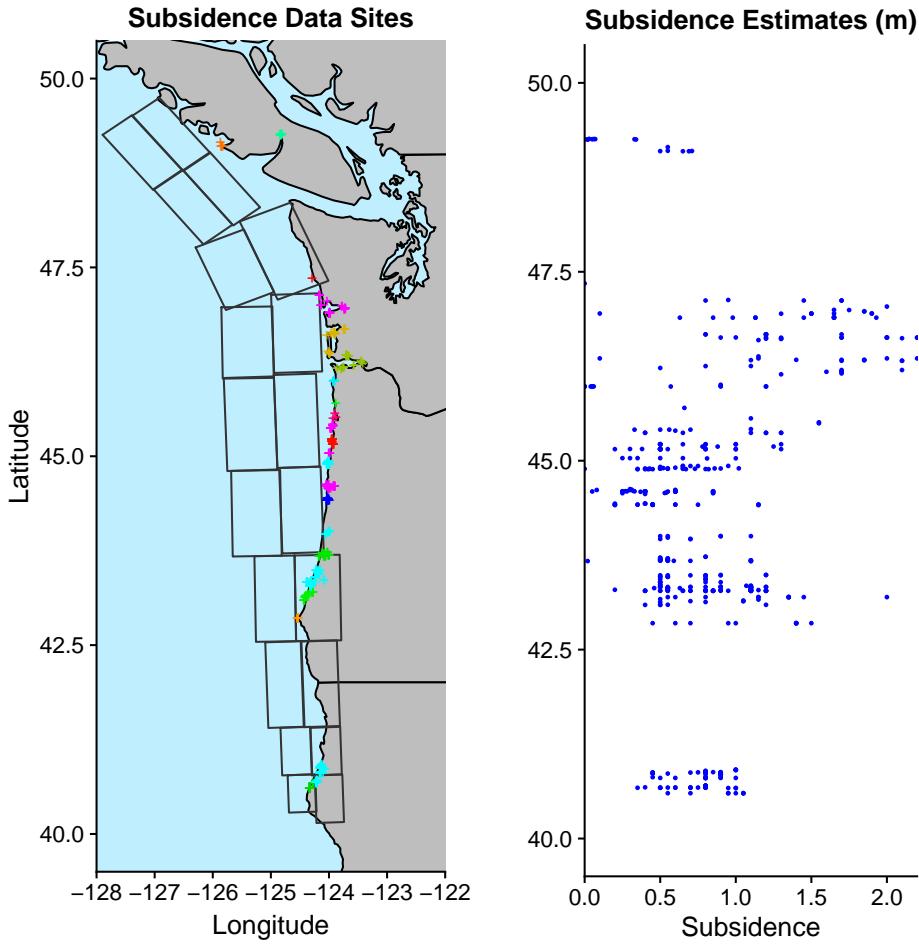


Fig. 1: Subsidence estimate study sites with the fault geometry and color coded by site (left), and the subsidence estimates themselves plotted on the same latitudinal scale (right). Subsidence estimates are from Leonard et al. [10] while the fault geometry is from Pollitz et al. [25].

latitude and longitude coordinates. As noted by Pollitz and Evans [24], the formal errors from their model are roughly one third the size of their derived empirical errors so we accordingly scale their formal error estimates by a factor of 3. The estimated locking rate uncertainties tend to increase with distance from onshore GPS observations. We use this locking rate spatial field to help estimate the covariance structure of possible CSZ earthquakes—the spatial

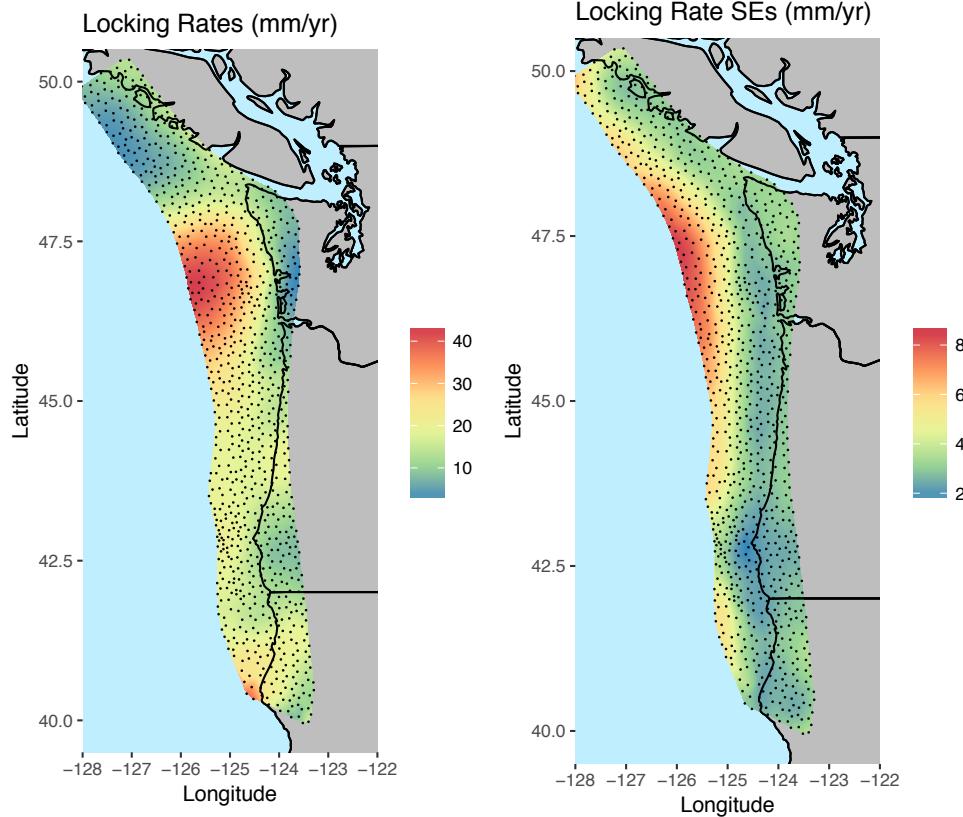


Fig. 2: Cascadia subduction zone fault locking rate estimates (left) and their standard errors (right) in mm/yr based on Evans et al. [4] when limited to 30km depths or shallower.

length scales on which coseismic slip varies—and how quickly slip decreases, on average, towards the deeper portion of the fault.

Different fault geometries for the CSZ megathrust lead to different surface deformations and therefore to different subsidence predictions. We assume the CSZ megathrust fault geometry from Pollitz et al. [25] that is depicted in Fig. 1 and is based on depth contours from McCrory et al. [16]. The CSZ megathrust is broken up into rectangular subfaults with uppermost edges running parallel to the strike direction, each subfault having depth as well as dip and strike angle information. These are then subdivided into smaller, rectangular subfaults

for finer scale resolution. The depth of the fault geometry at its deepest point is approximately 21km below sea level.

## 2.2 A spatial statistical model for full-margin earthquake slip

As in many previous studies (e.g. Leonard et al. [10], LeVeque et al. [11], Melgar et al. [18]) we assume a taper function such that average slip on the fault tapers to zero at its down-dip limit, which we assume is at depth 25km to allow flexibility in the taper while not being too much deeper than the fault geometry, avoiding drastic slip discontinuities as the edges of the fault geometry. This down-dip slip limit is similar to the one weighted most highly in Petersen et al. [23, p. 98-99], which used an estimated contour of 1cm/year locking rate for its down-dip slip limit. While some studies such as Evans et al. [4] and Wang et al. [31] assume two scenarios—one with the up-dip fault margin fully locked and one with the margin only partially locked—we assume the former for simplicity, and note that alternative scenarios should be considered in the future. We assume a normalized double exponential taper taking the form:

$$t(d; \lambda) = 1 - \frac{1 - \exp(-(d/d^*)^2 \lambda^2)}{1 - e^{-\lambda^2}} \quad (2)$$

so that the taper function smoothly varies from 1 at a depth of zero to 0 at the fixed fault down-dip slip limit  $d^*$ . Here  $d$  is the depth of the subfault centroid. The parameter  $\lambda$  is a rate that determines how quickly the slip tapers to zero. The taper function is shown for different values of  $\lambda$  in Figure 3. As  $\lambda \rightarrow 0$ , a straightforward application of l'Hôpital's rule shows  $\lim_{\lambda \rightarrow 0} t(d; \lambda) = 1 - (d/d^*)^2$ , which can be used to improve numerical accuracy for small values of  $\lambda$ .

In order to account for the latitudinal variability in the subsidence observations,  $\lambda$  is chosen to be a nonparametric function of latitude based on a five

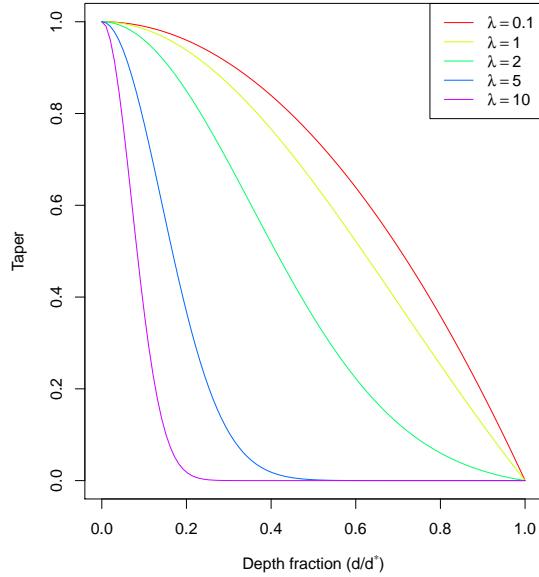


Fig. 3: Normalized double exponential taper functions from Eq. 2 with different rate parameters. The horizontal axis is the depth of the fault at a point divided by the down-dip slip limit,  $d^*$ .

knot B-spline basis expansion:

$$\boldsymbol{\lambda} = \boldsymbol{\Xi} \boldsymbol{\beta}$$

Here,  $\boldsymbol{\lambda}$  is the vector of taper rate parameters associated with the  $r$  grid cells in the fault geometry, and  $\boldsymbol{\Xi}$  is the B-spline basis matrix evaluated at the associated latitude values of those grid cells. Note that  $\boldsymbol{\beta}$  is a  $p$ -vector of coefficients for the B-splines forming the columns of  $\boldsymbol{\Xi}$  with  $p = 5$  in this analysis. Although we explored other basis functions and numbers of basis elements, we chose this basis due to its simplicity and stability over the latitudinal range of the data. Since there are large gaps in the subsidence data, using too many basis elements resulted in nearly unidentifiable models and unstable results.

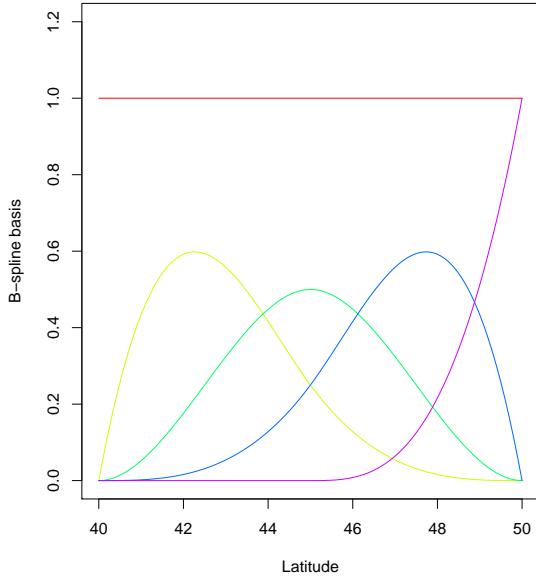


Fig. 4: A five knot B-spline basis is shown as a function of latitude, each color representing a different basis function.

The basis functions chosen are shown in Figure 4. Note that the chosen basis functions are symmetric in latitude, since the function space spanned by these elements is the same as the space spanned by replacing the constant basis function with a reflection of the magenta basis element around 45° latitude.

We now model coeseismic slip as a fully stochastic spatial process. Let  $\mathbf{S}$  be the vector of fault slip at the grid cells of the CSZ fault geometry. Then, letting  $\mathbf{T} = \text{diag}(\mathbf{t}(\mathbf{d}; \boldsymbol{\lambda}))$  be a diagonal taper matrix of taper function evaluations at each fault grid cell, we have:

$$\mathbf{S} = \mathbf{T}\mathbf{Z}$$

where  $\mathbf{Z}$  is taken to be a vector of values of a stochastic process occurring on the fault geometry. More specifically,  $\mathbf{Z}$  represents untapered coseismic slip, and could be modeled as a Gaussian process, a Gaussian process truncated at

zero to be nonnegative (a positive Gaussian process), a lognormal process, or some other stochastic process with the desired distributional model. We chose Gaussian and positive Gaussian process models for simplicity and also because preliminary results suggested they performed significantly better than lognormal process models. Moreover, we assume  $\mathbf{Z}$  has mean  $\mu_Z$ , variance  $\sigma_Z^2$ , and Matern correlation function  $\rho(\cdot; \phi, 3/2)$  depending on distance scale parameter  $\phi$ . Note that although  $\mu_Z$  and  $\sigma_Z$  are constant through space, the taper function induces nonstationarities in the marginal process mean and variance. Furthermore, historical subsidence data can be used to estimate more flexible stochastic process distributions for past earthquakes as will be discussed in Sec. 2.3. It should also be noted that  $\mathbf{S}$  as modeled here is much smoother than coseismic slips that might actually be observed, instead representing coseismic slip averaged (and therefore smoothed) over the fault geometry grid cells.

The model for the data is based on  $\mathbf{X}$ , the vector of GPS-based locking rate estimates, and  $\mathbf{Y}_i$ , the vector of subsidence observations corresponding to historical earthquake,  $i$ , and is as follows:

$$\mathbf{X} = \gamma \mathbf{S} + \boldsymbol{\xi} \quad (3)$$

$$\mathbf{Y}_i = \mathbf{G}_i \mathbf{S} + \tilde{\boldsymbol{\epsilon}}_i \quad (4)$$

for  $i \in \{1, 2, \dots, q\}$  with  $q$  being the number of earthquakes in the subsidence data. We therefore model the locking rate estimates as being proportional to some possible earthquake realization with proportionality constant  $\gamma$ , plus an error vector,  $\boldsymbol{\xi}$ . For simplicity, we assume that  $\boldsymbol{\xi}$  is independent Gaussian error using the uncertainty estimates from the locking rate data product inflated by a factor of three as discussed in Sec. 2.1. The subsidence resulting from fault slip is estimated using the Okada formulation [21] for a dislocation in an elastic

half-space with Poisson ratio 0.25. Given a fault geometry and the subsidence observation locations, the Okada model used to determine subsidence is a linear function of the slip over the discretized fault geometry, and is represented with the matrix  $\mathbf{G}_i$ . Lastly,  $\tilde{\epsilon}_i$  is an additive independent Gaussian noise vector where

$$\tilde{\epsilon}_{ij} = \psi_{ij}\epsilon_{ij},$$

$\epsilon_{ij}$  is the estimated measurement noise of subsidence observation  $j$  in earthquake  $i$ , and  $\psi_{ij}$  is an inflation factor for the  $j$ th subsidence observation in earthquake  $i$ . We assume  $\epsilon_{ij}$  is independent of  $\epsilon_{kl}$  for  $i \neq k$ . In light of concerns that the true uncertainties in the subsidence estimates may be different from those estimated, we felt that an inflation factor was a necessary addition to the model. For simplicity, two different inflations are estimated: one for macrofossil paleoseismic estimates and one for microfossil paleoseismic estimates. Letting

$$Q_{ij} = \begin{cases} 1, & Y_{ij} \text{ is a microfossil estimate,} \\ 0, & Y_{ij} \text{ is a macrofossil estimate,} \end{cases}$$

we have:

$$\psi_{ij} = Q_{ij}\psi_h + (1 - Q_{ij})\psi_l.$$

Hence,  $\psi_{ij}$  is either  $\psi_h$  or  $\psi_l$  depending on whether the subsidence estimate is a “high quality” microfossil estimate or a “low” to “middle” quality macrofossil estimate according to Leonard et al. [10]. Since the methodologies are completely different for micro and macrofossil estimates, there is no reason to assume the same uncertainty inflation factor for both.

Note that although assuming current locking rates are nearly proportional to the next earthquake’s slip is highly problematic [30], here we make a far weaker assumption: that current rates of locking are nearly proportional to

some earthquake in the past or future (with constant of proportionality  $\gamma$ ), or else they only provide information about the correlation scales and amount of variability in coseismic slip. In our exploratory analysis, historical earthquake magnitude data from Goldfinger et al. [6] suggested that the magnitudes of consecutive earthquakes were indeed not significantly related, lending some credibility to the assumption that different earthquakes' slip and subsidence distributions are independent conditional on the model parameters.

We consider three different taper models. One type of model that we will call the *combined* model, is as given above. However, it is also possible to slightly adjust the model by modifying the taper function for the locking rates to be based on spline coefficients  $\tilde{\beta} = \beta - \beta'$ , introducing  $p$  extra parameters in  $\beta'$ . This allows for the locking rate estimates to have a different taper from the subsidence data. The two taper functions can then be interpreted as a taper function based on locking rates and a taper function based on subsidence. These models will be referred to respectively as *locking* and *subsidence* taper models, and together as *separated* taper models.

Using the same taper for each dataset can in a sense average the taper functions that fit each separate dataset in a natural way with weights determined by uncertainties in the observations. However, modifying the taper function to allow it to be different for the locking rate dataset versus the subsidence dataset allows one to use the locking rate dataset to provide information about the correlation length of  $\mathbf{Z}$  while using the subsidence dataset to provide information about everything else. Hence, it allows predictions with the taper estimated either via the locking rate or the subsidence dataset, depending on which the user believes to more accurately represent the true underlying taper.

We fit both combined and separated taper models over a regular grid of values for  $\psi_l$  and  $\psi_h$ , ranging each from 0.5 to 2 by increments of 0.25, while for any fixed value of the uncertainty inflation parameters the model is estimated

via maximum likelihood and gradient descent using the BFGS method as implemented in the `constrOptim` function in R [20]. Gradients are derived in Appendix B.

### 2.3 Inferring historical coseismic slip from subsidence

After fitting the marginal model parameters, the conditional/predictive distribution for any historical coseismic slip (conditional on the subsidence observations associated with that earthquake) can be calculated relatively simply, especially in the Gaussian case. In the Gaussian case, computing the distribution is sometimes called ‘Kriging’ [5, 9] and is performed by considering the multivariate normal joint distribution of the data and the stochastic process. Assume  $\Sigma_D$  is the variance/covariance matrix of the data (in this case the data is the subsidence estimates for an earthquake),  $\Sigma_Z$  is the covariance matrix of the stochastic process (coseismic slip whose distribution is fit via maximum likelihood), and  $\Sigma_{ZD}$  is the cross-covariance matrix. Further assume the  $\mu_D$  and  $\mu_Z$  are the mean vectors of the data and the stochastic process respectively, and that  $\mathbf{Y}_i$  are the subsidence estimates for the earthquake of interest. Then the predictive distribution for the stochastic process of slip given the subsidence data for an earthquake is just multivariate normal with mean and covariance matrix given by the conditional formula for multivariate normal distributions:

$$\begin{aligned}\mu_{Z|D} &= \mu_Z - \Sigma_{ZD}\Sigma_D^{-1}(\mu_D - \mathbf{Y}_i) \\ \Sigma_{Z|D} &= \Sigma_Z - \Sigma_{ZD}\Sigma_D^{-1}\Sigma_{ZD}^T\end{aligned}$$

We can apply this same formula whether considering  $S$  to denote the stochastic process of coseismic slip over a continuous slip field, a set of areally averaged slip over a discrete geometry, or a joint distribution over the two. In the latter

case, a simulation from the distribution is a continuous slip field measured at a set of locations as well as its areally averaged counterpart.

It is important to note that, for ill-posed problems where the observations do not provide enough information about the model, the matrix  $\Sigma_D$  is singular or near-singular. Since paleoseismic subsidence is only measured along the coast and represent an integration of coseismic slip across the fault, they are unable to fully discern different types of earthquakes without certain modeling constraints. This is one reason we assume a down-sip slip limit of  $d^* = 25\text{km}$ , although it is certainly possible to use more advanced constraints.

Note that it is possible to produce negative slip under a Gaussian model. Although in our predictions for the 1700 earthquake the probability of negative slip is quite low, it is not insignificant for earthquakes with little or no subsidence data present such as earthquakes in the future or far in the past. It is, however, possible to throw out simulations with any negative slip, which corresponds to using a truncated Gaussian process. In fact, a simple proof given in Appendix A shows that the predictive distribution for a truncated Gaussian process is simply the predictive distribution for a Gaussian process, truncated (although the proof applies to any truncated distribution). Hence, drawing from the closed form Gaussian predictive distribution and then throwing out simulations that contain any negative value is equivalent to drawing from the truncated Gaussian process predictive distribution. However, it is recommended that those using this model take care when throwing out simulations with negative values or setting negative slip to zero, especially when there is a significant probability of negative slip. Throwing simulations or setting negative slip to zero can raise the mean of the coseismic slip enough to bias the predictions. Instead, if one is drawing a realization  $\mathbf{Z}$ , one can draw a realization from  $(\mathbf{Z} - \delta) | \{\mathbf{Z} - \delta > \mathbf{0}\}$  choosing  $\delta$  such that the expectation is approximately equal to the untruncated expectation:

$E[(\mathbf{Z} - \delta) | \{\mathbf{Z} - \delta > \mathbf{0}\}] \approx E[\mathbf{Z}]$ , where  $\{\mathbf{Z} - \delta > \mathbf{0}\}$  is the event that all elements of  $\mathbf{Z} - \delta$  are positive. Of course, this only accounts for biasing, and throwing out simulations with negative slip or setting negative slip to zero will also affect the process variance to a certain extent.

### 3 Application

#### 3.1 Marginal fit

	<i>Combined Taper</i>		<i>Separate Taper</i>		
	Estimates	SEs	Estimates	SEs	Units
$\mu_Z$	16.81	1.42	16.39	0.43	m
$\sigma_Z$	13.63	2.212	8.33	0.21	m
$\gamma$	1.96		1.92		$10^{-3}/\text{yr}$
$\beta_1$	-1.168	0.004	6.436	0.088	
$\beta_2$	-0.238	0.001	-11.734	0.237	
$\beta_3$	3.871	0.019	0.274	0.006	
$\beta_4$	2.477	0.007	-6.651	0.173	
$\beta_5$	1.005	0.006	-1.385	0.027	
$\beta'_1$			6.177	0.098	
$\beta'_2$			-14.225	0.274	
$\beta'_3$			6.552	0.212	
$\beta'_4$			-10.723	0.262	
$\beta'_5$			0.046	0.002	
$\phi$	153.0	8.7	129.8	2.3	km
$\psi_l$	1.75		1.75		
$\psi_h$	1.25		1.25		
Log Likelihood	-714.04		-619.02		

Table 1: Model parameter MLEs and their standard errors for the combined and separated taper models.

Fitting the combined model resulted in the parameter MLEs and standard errors given in Tab. 1, where standard errors are calculated using the hessian of the log-likelihood function. Table 1 shows that, for the combined and separate taper models,  $\mu_Z$  is 16.814 and 16.386 meters respectively. Similarly, the estimates for  $\sigma_Z$  are 13.633 and 8.328 respectively, which, along with a highly

significant  $\chi^2$  test, indicates that allowing separate tapers seems to explain a large amount of variation in the data. It is therefore likely that the processes influencing subsidence estimates and fault locking rate estimates are not exactly the same. However, fitting a model using both data sources is necessary, since locking rate estimates provide no information about the average magnitude of each earthquake, and the subsidence data provides provide minimal constraint on the slip spatial correlation structure and the down-dip slip profile given that the data are aligned north-south along the coastline.

For both combined and separated taper models, the optimal inflations for the macrofossil subsidence estimates ( $\psi_h$ ) are 1.25 while the optimal inflations for the microfossil estimates ( $\psi_l$ ) is 1.75. This consistency suggests this inflation is to some extent robust to model choice. Although estimate quality of course varies from study to study, it is nevertheless relevant that our models show the more recent microfossil estimates of subsidence seem to be better able to estimate their associated uncertainty. Note that this does not mean microfossil estimates are better estimates, just that their associated uncertainties better reflect the true underlying uncertainty in the data. In fact, after inflating uncertainties appropriately, the square root of the mean variance of the subsidence data is 0.45 cm for the microfossil estimates compared to 0.47 cm for the macrofossil estimates. This suggests that even if microfossil subsidence estimates often better estimate their uncertainty, estimate quality tends to be nearly identical for the micro- and macrofossil subsidence data used in this study.

The model fit yields a taper rate parameter ( $\lambda$ ) and a taper function varying through latitude. The taper rate for the combined and subsidence taper models increases in the far south, forcing coseismic slip to move farther offshore to prevent southern uplift. The paleoseismic evidence used in this study invariably suggested subsidence occurred in the south rather than uplift, yet fitting this

model using a taper parameter that is constant in latitude would force the model to produce uplift in the south running counter to the data, which is part of the reason the taper function varying by latitude was chosen. However, the taper rate is near zero across from Washington's Olympic Peninsula, allowing slip to occur closer to shore as shown in Figs. 5 and 6.

The normalized residuals of both datasets and of each model are displayed in Fig. 7. The subsidence residuals are biased for the combined and locking rate taper models, especially for the latter, where in the south predicted subsidence is on average approximately 2 standard errors below the observations leading to predicted uplifts. The subsidence taper model is unbiased, although some heteroscedasticity is not fully accounted for in the model. The locking rate residuals are smooth due to spatial correlations in the data through latitude. None of the normalized locking rate residuals look especially surprising; the combined and locking rate taper models are slightly less biased than the subsidence taper model, especially in the south since that is where the locking and subsidence datasets seem to disagree most. However, aside from that the normalized locking rate residuals of the subsidence taper model show no large systematic biases or poorly estimated variances.

A summary of the marginal model predictions is displayed in Figs. 5 and 6, which depicts mean coseismic slip for each fault element along with the distribution of possible earthquake magnitudes and the predicted versus observed subsidence for the Gaussian and positive Gaussian with adjusted mean models. The unadjusted positive Gaussian predictions are not shown, since the adjusted mean model marginal distributions tend to have improved predictions with the possible exception of the subsidence taper model, in which case the predictions are very similar. The subsidence predictions for the locking and combined taper model are clearly biased in all cases except for the positive normal, as previously noted in Fig. 7. The locking and combined ta-

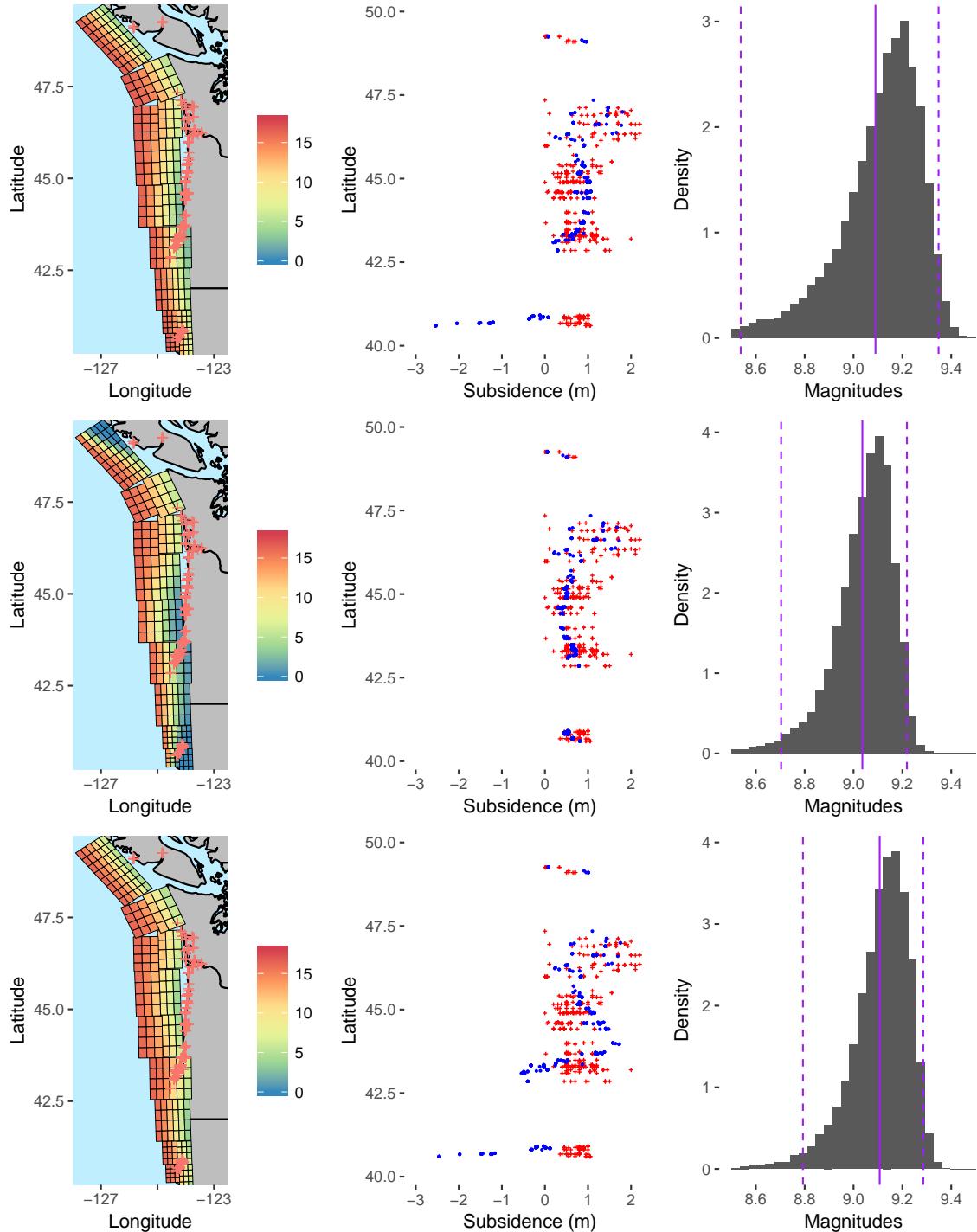


Fig. 5: Mean slip (left column), subsidence (middle column) with subsidence estimate (red '+') and mean predictions (blue points) along with magnitude distributions (right column) with mean (solid line) and 95% prediction bands (dashed lines). Results for the combined model are shown (top row), along with those for the subsidence taper model (middle row) and the locking rate taper model (bottom row) under Gaussian distributional assumptions.

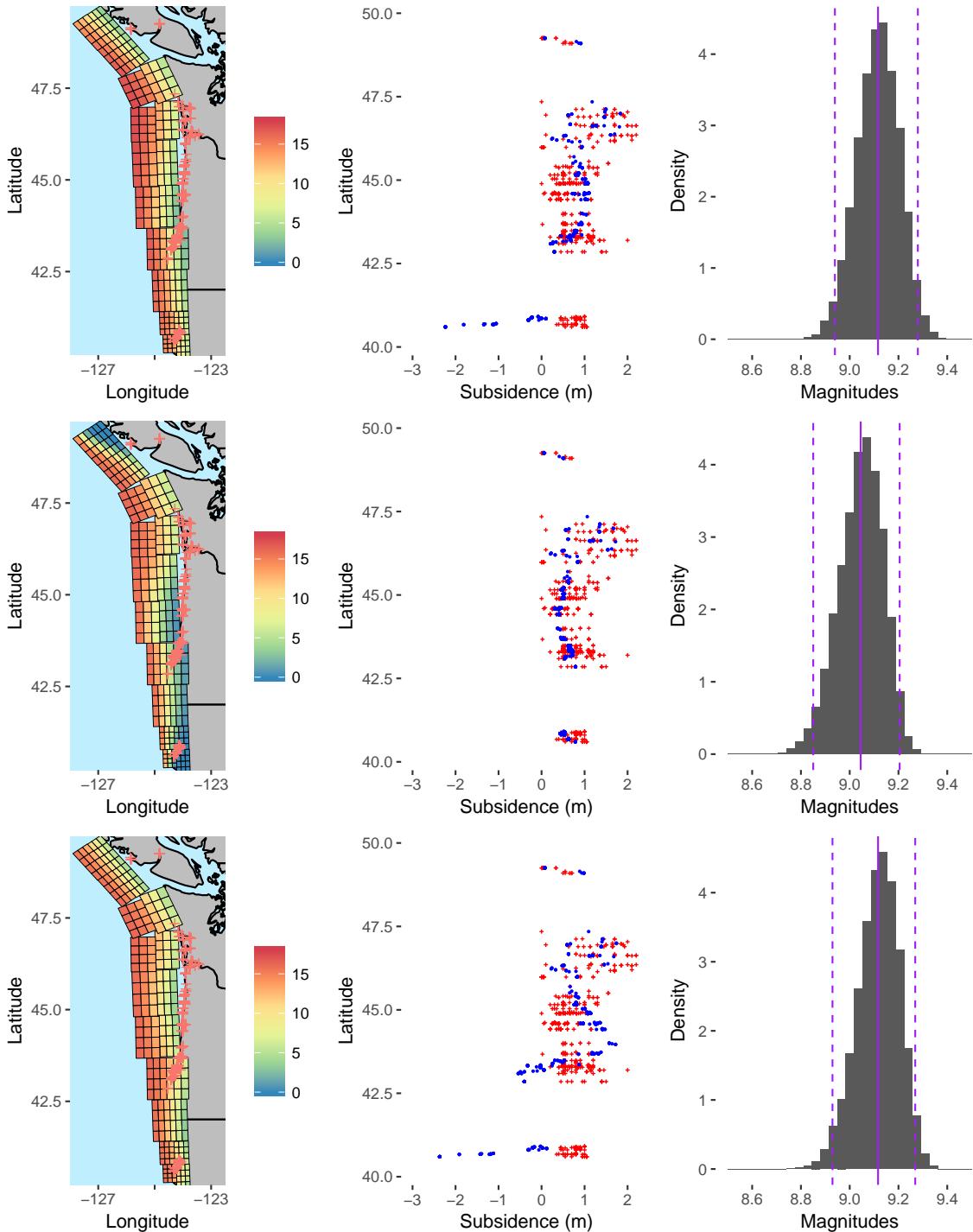


Fig. 6: Mean slip (left column), subsidence (middle column) with subsidence estimate (red '+') and mean predictions (blue points) along with magnitude distributions (right column) with mean (solid line) and 95% prediction bands (dashed lines). Results for the combined model are shown (top row), along with those for the subsidence taper model (middle row) and the locking rate taper model (bottom row) under positive Gaussian distributional assumptions with an adjusted mean.

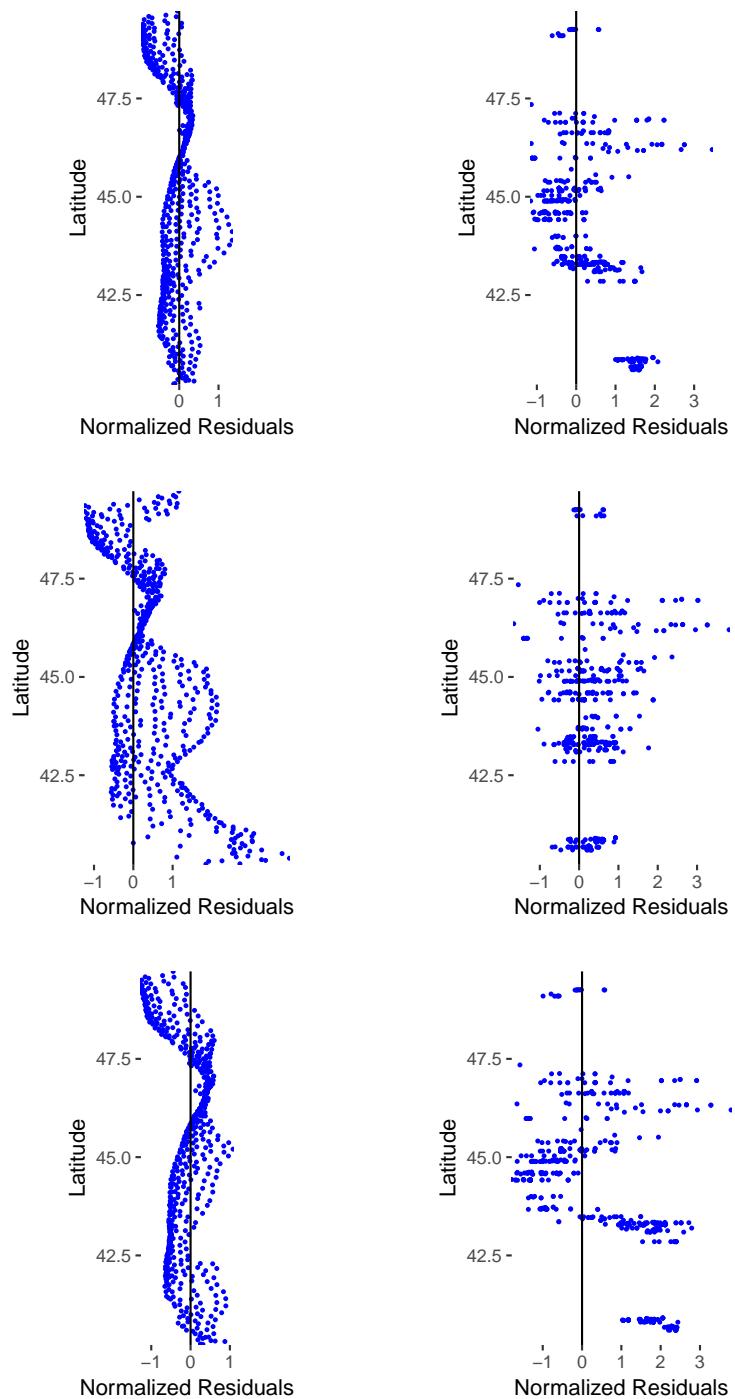


Fig. 7: Marginal residuals for the locking rate (left) and predictions for the subsidence (right) estimates. Results for the combined model are shown on top while those for the separated taper model are shown for the locking rate data (middle) and the subsidence data (bottom). Residuals are shown as a functions of latitude.

per models even predict uplift in the south when paleoseismic data exclusively indicates subsidence, and they predict slip will occur much deeper on the fault and further inland compared to the subsidence taper model. Both the combined and the locking taper models predicted uplift in the south, whereas the subsidence taper model predicts subsidence, and is best able to reproduce the mid-latitudinal variability in the subsidence estimates. Clarke Jr and Carver [2] suggests local faulting in the southern CSZ area, which could convolute either the locking or the subsidence datasets, causing the taper to be difficult to estimate in the south.

Figs. 5 and 6 indicate earthquake magnitudes between 8.6 and 9.4 with highly probability, with central estimates and 95% confidence bands given in Tab. 2. The mean-adjusted positive normal models have magnitude distribution concentrated almost entirely above 8.8 and below 9.3, whereas the normal distribution models have more spread in the possible earthquake magnitudes.

Taper	Distribution	Mean	95% CI
<i>Combined</i>	<i>Gaussian</i>	9.09	(8.52, 9.35)
	<i>Pos. Gaussian</i>	9.21	(9.03, 9.37)
	<i>Adj. Pos. Gaussian</i>	9.12	(8.94, 9.28)
<i>Subsidence</i>	<i>Gaussian</i>	9.00	(8.68, 9.22)
	<i>Pos. Gaussian</i>	9.07	(8.88, 9.22)
	<i>Adj. Pos. Gaussian</i>	9.05	(8.85, 9.21)
<i>Locking</i>	<i>Gaussian</i>	9.08	(8.77, 9.29)
	<i>Pos. Gaussian</i>	9.15	(8.96, 9.29)
	<i>Adj. Pos. Gaussian</i>	9.12	(8.93, 9.27)

Table 2: Mean magnitude and 95% confidence intervals for combined, subsidence, and locking taper models under normal, positive normal, and mean-adjusted positive normal distributional assumptions.

Predicted earthquake magnitudes are very similar for the three models, with mean earthquake magnitudes predicted as between 9.0 and 9.1 for all the models. The combined model has the highest uncertainty due to the larger

estimate for  $\sigma_Z$ . The consistent predicted magnitudes are encouraging, since no data or fixed parameters directly affect the predicted magnitudes, they are rather inferred indirectly from the subsidence data. However, subsidence residuals indicate that using the locking taper model may be problematic. We therefore recommend the combined or the subsidence taper models for future earthquake predictions.

A summary of the marginal coseismic slip distributions of the combined and subsidence taper models is shown in Fig. 8. More specifically, the 20th and 80th pointwise percentiles are shown along with the mean of the coseismic slip distributions. The combined models are more uncertain, as evidenced by the larger spread in the coseismic slip. However, the uncertainty is larger farther from shore due to less tapering (a taper function value near 1). Note that although each grid cell has a 20% chance of having slip larger than its 80th percentile, the probability that all grid cells have slip larger than all their 80th percentiles is much smaller. It is therefore not true that there is a 20% chance of observing an earthquake with at least as much slip as the pointwise 80th percentile slip displayed in Fig. 8.

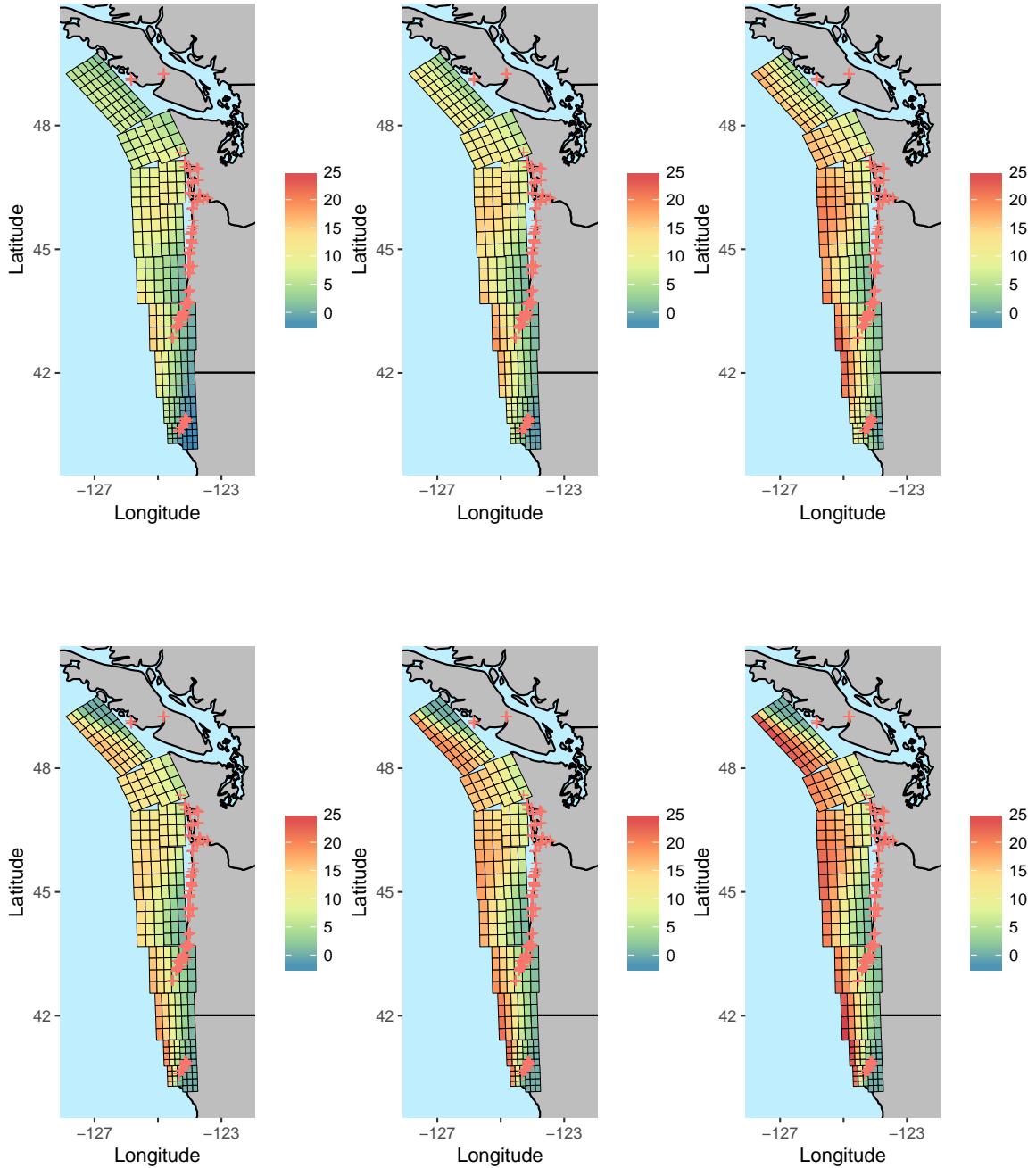


Fig. 8: 20th percentile (left column), mean (middle column) and 80th percentile (right column) coseismic slip. Results for the combined model are shown (top row), along with those for the subsidence taper model (bottom row). Subsidence estimate locations are shown using the red ‘+’ symbol.

### 3.2 Inferring coseismic slip of the 1700 event

Little is known about the slip distribution for the 1700 earthquake in Cascadia. It remains an open question as to where the peak slip was located along the margin, whether rupture propagated all the way to the trench, and how far down dip it extended. In order to answer these questions, the marginal models fit above are conditioned on subsidence data for the 1700 (or T1) event using the methodologies described in Sec. 2.3. Since predictions look somewhat similar after conditioning the marginal slip distribution on subsidence observations from 1700, only predictions for the combined taper model are shown. However, it is important to note that although subsidence predictions for all the models look remarkably similar, there are some slight differences in the 1700 inferred slip distributions among the three taper models, which are summarized in Appendix D.

Table 3 shows that magnitudes for the different taper and distributional models are similar, although the subsidence taper model tends to predict slightly larger magnitudes than the other models for the 1700 event. Figures in Appendix D shows that the subsidence model also tends to predict slightly larger coseismic slip than the other two models, although Tab. 2 shows that the marginal subsidence model tends to yield smaller magnitudes than the other taper models. Of course, the unadjusted positive Gaussian model predicts higher magnitudes than the Gaussian model since it is truncated at zero, forcing larger slip.

### 3.3 Validation

In order to validate the model, we use site-by-site cross-validation based on the subsidence data. In each cross-validation iteration, data from one site is left out while data from other sites is used to attempt to predict the ob-

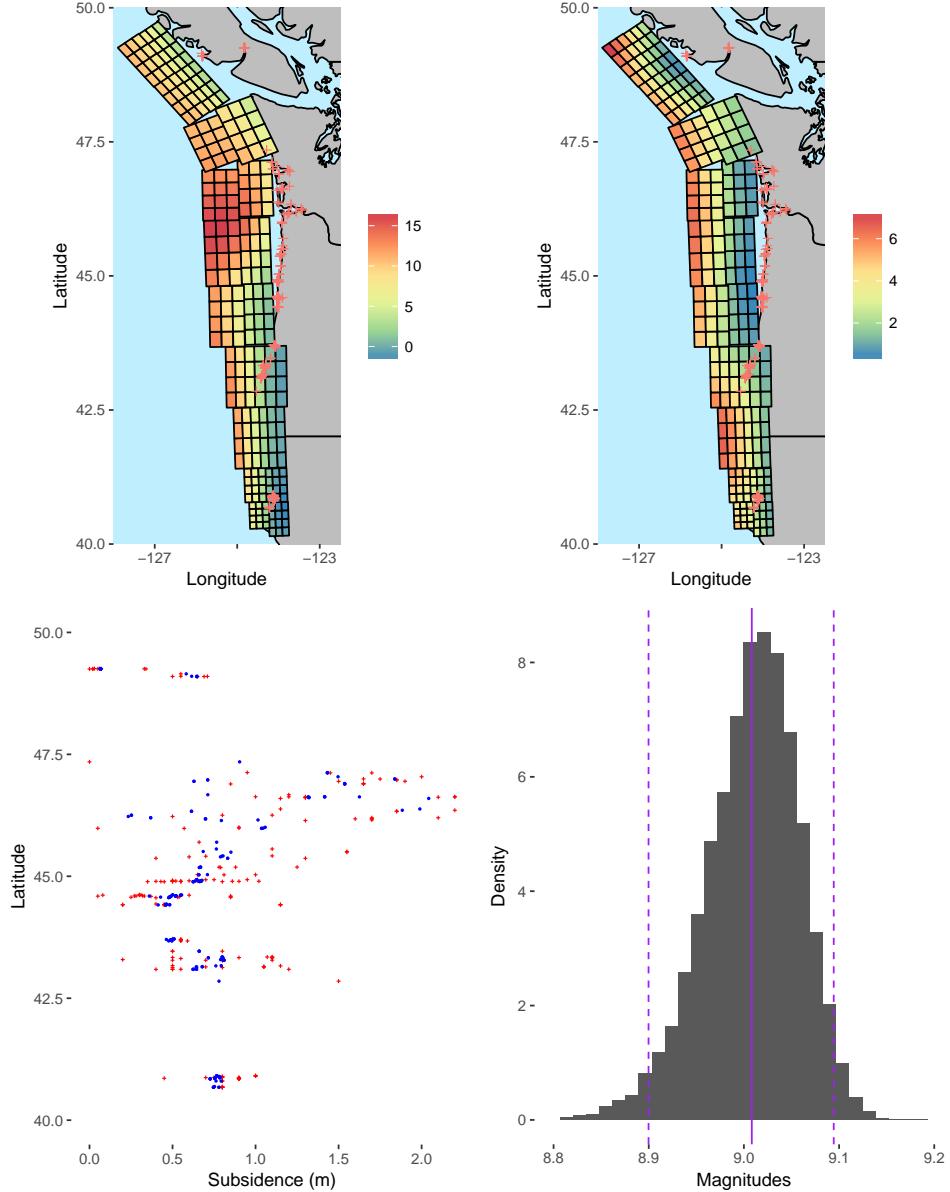


Fig. 9: Inference on the 1700 (T1) Cascadia subduction event for the combined taper model. Mean coseismic slip (top left) along with standard error (top right) are shown with subsidence predictions (blue dots, bottom left) and magnitudes (bottom right) with a 95% confidence band (dashed lines) and mean (solid line). Subsidence observations associated with the 1700 event are plotted as red ‘+’ symbols.

Taper	Distribution	Mean	95% CI
<i>Combined</i>	<i>Gaussian</i>	9.04	(8.95, 9.11)
	<i>Pos. Gaussian</i>	9.08	(9.02, 9.13)
<i>Subsidence</i>	<i>Gaussian</i>	9.11	(9.06, 9.15)
	<i>Pos. Gaussian</i>	9.11	(9.06, 9.15)
<i>Locking</i>	<i>Gaussian</i>	9.02	(8.94, 9.08)
	<i>Pos. Gaussian</i>	9.06	(9.01, 9.11)

Table 3: Inferred magnitude means and 95% confidence intervals for the 1700 subduction event under the three taper models discussed with Gaussian and positive Gaussian distributional assumptions.

servations at each left out site. Table 4 shows each subduction event in the cross-validation procedure and the number of sites and observations associated with it in our dataset. The focus here is specifically on validating with respect to the subsidence data since it is especially difficult to predict due to being spatially sparse and also noisy. Moreover, because we are leaving data out on a site-by-site basis, the prediction problem is more difficult than simply leaving out random observations—sites that are far from other sites will be more difficult to predict. We only use events T1 through T7 since these are major subduction events having at least 20 observations.

Event	T1	T2	T3	T4	T5	T6	T7
Number of Sites	21	9	10	16	16	9	11
Number of Observations	196	24	25	42	51	30	27

Table 4: Number of subsidence data sites and observations for CSZ subduction events T1 (the 1700 event) through T7.

Cross-validation mean square errors (MSEs) for normal and positive normal models are shown for the marginal (non-historical) distribution in Tab. 5, while MSEs for the predictions distributions are given in Tab. 6. We find that, unsurprisingly, the subsidence-based taper model is best able to predict the subsidence observations for both the marginal and predictive distributions with an average MSE of 0.278, followed by the combined model and the lock-

ing taper model in that order. However, the difference is much smaller for the predictive distributions than for the marginal distributions. Hence, the choice of taper model matters little when it comes to inference regarding historical earthquakes with a reasonable amount of subsidence data. The choice of modeling assumptions matters much more when it comes to predicting future earthquakes or historical earthquakes with few or no subsidence observations ( $n < 20$ ).

Bias correction by adjusting the positive normal distribution mean does not seem to improve the predictive distribution MSE substantially for events with a large number of observations ( $n > 20$ ), most likely because, adjusted or not, the mean of the predictive distribution is mostly determined by the data used to predict the subsidence at each site. On the other hand, the best class of models for the marginal distribution predictions appears to be the mean-adjusted positive normal models. Hence, for future earthquake predictions and for historical earthquakes with very few subsidence observations, mean-adjusted positive normal models will likely perform the best.

	Comb	Sub	Lock	CombPN	SubPN	LockPN	CombPNAdj	SubPNAdj	LockPNAdj
Avg	1.223	1.223	1.178	1.384	0.278	1.229	0.865	0.288	1.110
T1	1.014	1.014	0.985	1.100	0.275	1.019	0.707	0.288	0.924
T2	1.051	1.051	1.111	1.344	0.173	1.209	0.785	0.160	1.057
T3	2.667	2.667	2.325	3.344	0.334	2.491	1.928	0.335	2.191
T4	1.582	1.582	1.491	1.771	0.352	1.542	1.121	0.364	1.406
T5	1.065	1.065	1.082	1.156	0.325	1.115	0.778	0.334	1.029
T6	1.609	1.609	1.492	1.908	0.193	1.570	1.157	0.198	1.411
T7	0.583	0.583	0.729	0.545	0.186	0.734	0.329	0.194	0.683

Table 5: Marginal distribution site-by-site cross-validation MSEs for subduction events T1-T7 and their average for normal, positive normal, and adjusted mean positive normal models and for each taper model (combined, subsidence, and locking tapers).

	Comb	Sub	Lock	CombPN	SubPN	LockPN	CombPNAdj	SubPNAdj	LockPNAdj
Avg	0.368	0.368	0.375	0.369	0.362	0.380	0.368	0.363	0.380
T1	0.324	0.324	0.333	0.323	0.345	0.329	0.324	0.346	0.329
T2	0.208	0.208	0.203	0.278	0.112	0.261	0.238	0.110	0.249
T3	0.358	0.358	0.368	0.348	0.309	0.367	0.356	0.309	0.370
T4	0.540	0.540	0.539	0.510	0.499	0.527	0.502	0.496	0.524
T5	0.523	0.523	0.513	0.513	0.507	0.527	0.515	0.508	0.529
T6	0.358	0.358	0.381	0.395	0.332	0.438	0.401	0.334	0.440
T7	0.280	0.280	0.301	0.276	0.286	0.304	0.277	0.288	0.304

Table 6: Predictive distribution site-by-site cross-validation MSEs for subduction events T1-T7 and their average for normal, positive normal, and adjusted mean positive normal models and for each taper model (combined, subsidence, and locking tapers).

#### 4 Conclusions

We have presented a spatial statistical model for predicting coseismic slip on the Cascadia subduction zone conditioned on paleoseismic subsidence and contemporary locking rate data. This model can be used to predict past or future coseismic slip distributions with fully propagated uncertainties that account for spatial correlations and various data types. Spatial statistical models for coseismic slip have a number of benefits over methods used today and in the past. Since they are fully likelihood-based, they cleanly account for correlations in the data in a way that model fitting using only mean-squared error does not. Unlike in logic tree approaches, spatial statistical models can fit a fully continuous distribution of slip rather than only modeling a few discrete possible outcomes. They can be used to infer historical coseismic slip using simple formulas and to verify the estimated uncertainties in subsidence data for macro and microfossil based methods. Unlike models using a single taper function depending on depth, the proposed models can capture latitudinal variation in where coseismic slip is likely to occur.

Although all three taper models are valid predictive models, they have different assumptions about the behavior of the slip tapering. If the locking

rates and the subsidence is equally representative of the sets of observable coseismic slip and subsidence, then the combined taper is theoretically the correct taper to use since it weights each dataset based on the theoretically correct weights in the likelihood function. However, if one dataset is more representative of the true latent processes than the other, whether it is due to systematic biases in the data or a faulty assumption about the relationship of the data to the true underlying process, it may be more reasonable to use the subsidence or locking taper models. The difference between the taper models seems to make the largest difference in the southern portion of the CSZ, where Clarke Jr and Carver [2] suggests local faulting might make inference more difficult.

The assumptions of the different models matter less for inferring historical earthquake slip conditioned with  $> 20$  subsidence observations, but can make a difference when attempting to infer historic earthquake slip with few observations or no observations such as when making future earthquake slip predictions. The distributional assumption (normal, positive normal, and mean-adjusted positive normal) also matter more when there are fewer subsidence observations, although they understandably seem to make little difference when the bulk of the slip distribution is positive, such as for the subsidence taper model.

Although all distributional models (normal, positive normal, and mean-adjusted positive normal) and taper models (combined, subsidence, and locking) produce subsidence that is very similar to the paleoseismic estimates associated with the 1700 event, there are some minor differences in the slip and therefore the magnitude distributions. This is because the inverse problem of trying to infer slip distributions from subsidence estimates is relatively ill-posed—multiple different earthquakes can produce the same subsidence. This is why using both GPS-based locking rate data and subsidence estimates to

fit the marginal slip distribution before trying to infer the distribution of slip for any given earthquake is so important. It is also why the specific models and assumptions used to generate 1700 slip distributions matter. One possible direction of future work would be to fit multiple models to the data involving different assumptions and performing Bayesian model averaging techniques to cleanly weight multiple possible models in a fully probabilistic way.

Assuming coseismic slip is multivariate normal or positive normal may be reasonable for inference about historical earthquakes with a fair amount of subsidence data, but it is worth exploring alternative distributional forms in other settings. Although in our exploratory analysis we ruled out the multivariate lognormal distribution for its relatively poor predictions of past subsidence, Gaussian log-Gaussian distributions (cf. [22] and [5, p.150]) allow for a way to account for varying degrees of skewness in the data, and nonparametric spatial kernel stick-breaking priors (cf. [26], [5, p. 160]) could also be useful if marginal slip is considerably non-Gaussian (after accounting for the slip mean model).

Note that in this study we only attempt to model full-fault ruptures. If that assumption were loosened, the resulting distribution of slip over the CSZ megathrust will certainly be nonstationary. To account for this nonstationarity, one possibility would be to allow for the marginal mean and variance parameters, respectively  $\mu_Z$  and  $\sigma_Z^2$ , to vary through space. One relatively simple way would be to use B-spline basis expansions through latitude to represent the parameters. Mixture modeling approaches could also better model the earthquakes resulting from the different parts of the fault that might rupture.

It is also important to consider more accurate fault geometries. The Slab 1.0 geometry [17] is a triangularization of the fault that allows for a more flexible and smooth fault structure without gaps. In addition to being more

accurate, the geometry goes much deeper (100 km depth rather than 21 km), and therefore could allow for more accurate estimation of the down-dip slip limit and the slip taper structure at deeper parts of the megathrust. Since the current fault geometry only goes 21 km deep, there could be some unmodeled coseismic slip occurring deeper than the maximum depth of the fault geometry used in this analysis.

Of course, for any of these model changes, it is important to make sure the resulting model is simple enough to be identifiable. The model fitting and inference for this particular data is complicated by the fact that inferring slip based on subsidence is an ill-posed inverse problem, so it is important to be careful when making especially complex models with too many parameters. In addition, lack of data near the Olympic peninsula and the Oregon-California border makes determining the taper depths near those regions difficult. To an extent this can be mitigated by using penalized likelihoods or using priors in a fully Bayesian analysis, although it is also important to gather more paleoseismic data in these regions for better predictions nearby.

**Acknowledgements** We would like to thank the NSF GRFP for funding the first author (grant number DGE-1256082), and its DMS division for funding the first and second authors via grant DMS-1401793 and the second author via grant DMS-1106862. We would also like to thank the M9 project at the University of Washington for organizing a multi-disciplinary group studying the impacts of a magnitude 9 Cascadia subduction zone event, funded via NSF grant EAR-1331412. In addition, Professor Brian Atwater at the University of Washington and Dr. Carrie Garrison-Laney were very helpful when giving advice about subsidence estimates. Professor Randall Leveque at the University of Washington and Don-sub Rim who recently became a Professor at Columbia University provided code for the Okada model via the GeoClaw python package as well as the fault geometry used in this analysis, and Professor David Schmidt of the University of Washington had excellent advice regarding the interpretation of the GPS-based locking rate data and the Okada model.

## References

1. Casella G, Berger RL (2002) Statistical inference, vol 2. Duxbury Pacific Grove, CA
2. Clarke Jr SH, Carver GA (1992) Late holocene tectonics and paleoseismicity, southern Cascadia subduction zone. *Science* 255(5041):188
3. Duputel Z, Jiang J, Jolivet R, Simons M, Rivera L, Ampuero JP, Riel B, Owen SE, Moore AW, Samsonov SV, et al. (2015) The Iquique earthquake sequence of April 2014: Bayesian modeling accounting for prediction uncertainty. *Geophys Res Lett* 42(19):7949–7957
4. Evans EL, Loveless JP, Meade BJ (2015) Total variation regularization of geodetically and geologically constrained block models for the Western United States. *Geophys J Int* 202(2):713–727
5. Gelfand AE, Diggle P, Guttorp P, Fuentes M (2010) Handbook of spatial statistics. CRC press
6. Goldfinger C, Nelson CH, Morey AE, Johnson JE, Patton JR, Karabanov E, Gutierrez-Pastor J, Eriksson AT, Gracia E, Dunhill G, et al. (2012) Turbidite event history: Methods and implications for Holocene paleoseismicity of the Cascadia subduction zone. US Geol Survey Professional Paper 1661:170
7. Hammond WC, Thatcher W (2005) Northwest basin and range tectonic deformation observed with the global positioning system, 1999–2003. *J Geophys Res-Sol Ea* 110(B10)
8. Herring TA, Melbourne TI, Murray MH, Floyd MA, Szeliga WM, King RW, Phillips DA, Puskas CM, Santillan M, Wang L (2016) Plate Boundary Observatory and related networks: GPS data analysis methods and geodetic products. *Rev Geophys*

9. Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *J South Afr Inst Min Metall* 52(6):119–139
10. Leonard LJ, Currie CA, Mazzotti S, Hyndman RD (2010) Rupture area and displacement of past Cascadia great earthquakes from coastal coseismic subsidence. *Geol Soc Am Bull* 122(11-12):2079–2096
11. LeVeque RJ, Waagan K, González FI, Rim D, Lin G (2016) Generating random earthquake events for probabilistic tsunami hazard assessment. *Pure Appl Geophys* 173(12):3671–3692
12. Loveless JP, Meade BJ (2011) Stress modulation on the San Andreas fault by interseismic fault system interactions. *Geology* 39(11):1035–1038
13. McCaffrey R, Qamar AI, King RW, Wells R, Khazaradze G, Williams CA, Stevens CW, Vollick JJ, Zwick PC (2007) Fault locking, block rotation and crustal deformation in the Pacific Northwest. *Geophys J Int* 169(3):1315–1340
14. McCaffrey R, King RW, Payne SJ, Lancaster M (2013) Active tectonics of northwestern US inferred from GPS-derived surface velocities. *J Geophys Res-Sol Ea* 118(2):709–723
15. McClusky S, Bjornstad S, Hager BH, King R, Meade B, Miller M, Monastero F, Souter B (2001) Present day kinematics of the eastern California shear zone from a geodetically constrained block model. *Geophys Res Lett* 28(17):3369–3372
16. McCrory PA, Blair JL, Oppenheimer DH, Walter SR (2004) Depth to the Juan de Fuca slab beneath the Cascadia subduction margin: A 3-D model for sorting earthquakes. US Department of the Interior, US Geological Survey
17. McCrory PA, Blair JL, Waldhauser F, Oppenheimer DH (2012) Juan de Fuca slab geometry and its relation to Wadati-Benioff zone seismicity. *J*

Geophys Res-Sol Ea 117(B9)

18. Melgar D, LeVeque RJ, Dreger DS, Allen RM (2016) Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia Subduction Zone. *J Geophys Res-Sol Ea* 121(9):6658–6674
19. Minson S, Simons M, Beck J (2013) Bayesian inversion for finite fault earthquake source models I—Theory and algorithm. *Geophys J Int* 194(3):1701–1726
20. Nash JC (1990) Compact numerical methods for computers: linear algebra and function minimisation. CRC press
21. Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bull Seismol Soc Am* 75(4):1135–1154
22. Palacios MB, Steel MFJ (2006) Non-Gaussian Bayesian geostatistical modeling. *J Am Stat Assoc* 101(474):604–618
23. Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field E, Chen R, Rukstales KS, Luco N, Wheeler RL, Williams RA, Olsen AH (2014) Documentation for the 2014 update of the United States national seismic hazard maps. Open-File Report 2014-1091, U. S. Geological Survey, Reston, VA, DOI 10.3133/ofr20141091, URL <http://pubs.er.usgs.gov/publication/ofr20141091>
24. Pollitz F, Evans E (2017) Implications of the earthquake cycle for inferring fault locking on the Cascadia megathrust. *Geophys J Int* 209(1):167–185
25. Pollitz F, McCrory P, Wilson D, Svanc J, Puskas C, Smith RB (2010) Viscoelastic-cycle model of interseismic deformation in the northwestern United States. *Geophys J Int* 181(2):665–696
26. Reich BJ, Fuentes M (2007) A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Ann Appl Stat* pp 249–264

27. Shen Z, Agnew D, King R, Dong D, Herring T, Wang M, Johnson H, Anderson G, Nikolaidis R, van Domselaar M, et al. (2003) The SCEC crustal motion map, version 3.0. South Calif Earthquake Cent, Los Angeles (Available at <http://epicenter.usc.edu/cmm3>)
28. Stein ML (2012) Interpolation of spatial data: some theory for kriging. Springer Science & Business Media
29. Stein S, Geller RJ, Liu M (2012) Why earthquake hazard maps often fail and what to do about it. *Tectonophysics* 562:1–25
30. Wang K, Hu Y, He J (2012) Deformation cycles of subduction earthquakes in a viscoelastic Earth. *Nat* 484(7394):327
31. Wang PL, Engelhart SE, Wang K, Hawkes AD, Horton BP, Nelson AR, Witter RC (2013) Heterogeneous rupture in the great Cascadia earthquake of 1700 inferred from coastal subsidence estimates. *J Geophys Res-Sol Ea* 118(5):2460–2473
32. Williams TB, Kelsey HM, Freymueller JT (2006) GPS-derived strain in northwestern California: Termination of the San Andreas fault system and convergence of the Sierra Nevada–Great Valley block contribute to southern Cascadia forearc contraction. *Tectonophysics* 413(3):171–184
33. Witter RC, Zhang YJ, Wang K, Priest GR, Goldfinger C, Stimely L, English JT, Ferro PA (2013) Simulated tsunami inundation for a range of Cascadia megathrust earthquake scenarios at Bandon, Oregon, USA. *Geosphere* 9(6):1783–1803

## Appendix A: Proofs

The following lemma shows that, in order to draw from the conditional distribution of the positive normal distribution, it is sufficient to draw from the

normal distribution until the drawn random vector is entirely positive (letting  $A = \mathbb{R}_+^p$  be the set of multivariate positive real vectors).

**Lemma 1** *Assume  $\mathbf{Z}$  is a random vector in  $\mathbb{R}^p$  following some density,  $f_{\mathbf{Z}}(\mathbf{z})$ , and  $\tilde{\mathbf{Z}}$  is a random vector following the same density but truncated to be in  $P_{\mathbf{Z}}$ -measurable set,  $A$ , so that:*

$$f_{\tilde{\mathbf{Z}}}(\mathbf{z}) = \frac{1}{P_{\mathbf{Z}}(A)} f_{\mathbf{Z}}(\mathbf{z}).$$

*Then if  $\mathbf{Y} \sim F_{\tilde{\mathbf{Z}}}$ , we have:*

$$f_{\tilde{\mathbf{Z}}|\mathbf{Y}}(\mathbf{z}|\mathbf{Y}) \propto f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{Y}) \cdot \mathbb{1}\{\mathbf{z} \in A\}.$$

*Proof* Bayes rule implies:

$$\begin{aligned} f_{\tilde{\mathbf{Z}}|\mathbf{Y}}(\mathbf{z}|\mathbf{Y}) &= \frac{f_{\mathbf{Y}|\tilde{\mathbf{Z}}}(\mathbf{Y}|\tilde{\mathbf{Z}}) f_{\tilde{\mathbf{Z}}}(\mathbf{z})}{f_{\tilde{\mathbf{Z}}}(\mathbf{Y})} \cdot \frac{\mathbb{1}\{\mathbf{z} \in A\}}{P_{\mathbf{Z}}(A)} \\ &= \frac{f_{\mathbf{Y}|\tilde{\mathbf{Z}}}(\mathbf{Y}|\tilde{\mathbf{Z}})}{f_{\mathbf{Z}}(\mathbf{Y})} \cdot \mathbb{1}\{\mathbf{z} \in A\} \\ &\propto f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{Y}) \cdot \mathbb{1}\{\mathbf{z} \in A\}, \end{aligned}$$

completing the proof.  $\square$

## Appendix B: Gradients

Computing the subsidence likelihood gradient:

First we let  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T \ \mu \ \sigma \ \phi)^T$  be the vector of parameters with  $\boldsymbol{\beta}$  being the taper parameters and  $\mu$  and  $\sigma^2$  be the moments of  $\mathbf{Z}$ , the untapered

earthquake. The log-likelihood is given by:

$$\begin{aligned} l(\boldsymbol{\theta}) &\propto -\frac{1}{2} [(\mathbf{Y} - \boldsymbol{\mu}_y(\boldsymbol{\theta}))^T \Sigma^{-1}(\boldsymbol{\theta})(\mathbf{Y} - \boldsymbol{\mu}_y(\boldsymbol{\theta}))] - \frac{1}{2} \log(|\Sigma|) \\ &= -\frac{1}{2} [\mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \mathbf{Y} - 2\mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) + \boldsymbol{\mu}_y(\boldsymbol{\theta})^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta})] - \frac{1}{2} \log(|\Sigma|) \end{aligned}$$

The gradient is therefore:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} l(\boldsymbol{\theta}) &= -\frac{1}{2} \frac{\partial}{\partial \theta_i} [\mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \mathbf{Y} - 2\mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) + \boldsymbol{\mu}_y(\boldsymbol{\theta})^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta})] - \frac{1}{2} \frac{\partial}{\partial \theta_i} \log(|\Sigma|) \\ \frac{\partial}{\partial \theta_i} \mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \mathbf{Y} &= -\mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \Sigma^{-1}(\boldsymbol{\theta}) \mathbf{Y} \\ -2 \frac{\partial}{\partial \theta_i} \mathbf{Y}^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) &= -2\mathbf{Y}^T \left( \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_y(\boldsymbol{\theta})}{\partial \theta_i} - \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) \right) \\ \frac{\partial}{\partial \theta_i} \boldsymbol{\mu}_y(\boldsymbol{\theta})^T \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) &= \boldsymbol{\mu}_y(\boldsymbol{\theta})^T \left( \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_y(\boldsymbol{\theta})}{\partial \theta_i} - \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) \right) + \frac{\partial \boldsymbol{\mu}_y(\boldsymbol{\theta})^T}{\partial \theta_i} \Sigma^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_y(\boldsymbol{\theta}) \\ -\frac{1}{2} \frac{\partial}{\partial \theta_i} \log(|\Sigma(\boldsymbol{\theta})|) &= -\frac{1}{2} \text{tr} \left( \Sigma^{-1}(\boldsymbol{\theta}) \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \right) \end{aligned}$$

Gradient of the covariance of  $\mathbf{Y}$ :

At this point, we can compute the gradient of more specific terms. Let's start with the more difficult term, the covariance. Since  $\Sigma_{ik} = COV(Y_i, Y_k)$ , we find for  $i \neq j$ :

$$\begin{aligned} COV(Y_i, Y_j) &= COV((\mathbf{GT})_{i:} \mathbf{Z} + \boldsymbol{\epsilon}_i, (\mathbf{GT})_{j:} \mathbf{Z} + \boldsymbol{\epsilon}_j) \\ &= COV((\mathbf{GT})_{i:} \mathbf{Z}, (\mathbf{GT})_{j:} \mathbf{Z}) \\ &= (\mathbf{GT})_{i:} Var(\mathbf{Z}) (\mathbf{GT})_{j:}^T \end{aligned}$$

For  $i = j$  we must also add constant observation noise, but since this is constant it doesn't factor into the gradient. First note that

$$\begin{aligned}\frac{\partial}{\partial \mu} Var(\mathbf{Y}) &= (\mathbf{GT}) \left( \frac{\partial}{\partial \mu} Var(\mathbf{Z}) \right) (\mathbf{GT})^T \\ \frac{\partial}{\partial \phi} Var(\mathbf{Y}_i) &= \frac{\partial}{\partial \phi} \mathbf{G}_i \mathbf{T} Var(\mathbf{Z}) \mathbf{T} \mathbf{G}_i^T \\ &= \mathbf{G}_i \mathbf{T} \left( \frac{\partial}{\partial \phi} Var(\mathbf{Z}) \right) \mathbf{T} \mathbf{G}_i^T\end{aligned}$$

Now we compute the gradient with respect to  $\beta_i$ :

$$\begin{aligned}\frac{\partial}{\partial \beta_i} COV(Y_j, Y_k) &= \frac{\partial}{\partial \beta_i} (\mathbf{GT})_{j:} Var(\mathbf{Z}) (\mathbf{GT})_{k:} \\ &= \frac{\partial}{\partial \beta_i} \sum_{r,s} (\mathbf{GT})_{jr} COV(Z_r, Z_s) (\mathbf{GT})_{ks} \\ &= \frac{\partial}{\partial \beta_i} \sum_{r,s} G_{jr} t_r COV(Z_r, Z_s) G_{ks} t_s \\ &= \sum_{r,s} G_{jr} COV(Z_r, Z_s) G_{ks} \frac{\partial}{\partial \beta_i} (t_r t_s) \\ &= \sum_{r,s} G_{jr} COV(Z_r, Z_s) G_{ks} \left( \frac{\partial}{\partial \beta_i} t_r \cdot t_s + t_r \cdot \frac{\partial}{\partial \beta_i} t_s \right) \\ &= \sum_{k,s} G_{jr} G_{ks} \left[ COV(Z_r, Z_s) \left( \frac{\partial}{\partial \beta_i} t_r \cdot t_s + t_r \cdot \frac{\partial}{\partial \beta_i} t_s \right) \right] \\ &= G_{j:} \left[ Var(\mathbf{Z}) \cdot \left( \frac{\partial}{\partial \beta_i} \mathbf{t} \cdot \mathbf{t}^T + \mathbf{t} \cdot \frac{\partial}{\partial \beta_i} \mathbf{t}^T \right) \right] G_{k:}^T \\ \frac{\partial}{\partial \beta_i} Var(\mathbf{Y}) &= G \left[ Var(\mathbf{Z}) \cdot \left( \frac{\partial}{\partial \beta_i} \mathbf{t} \cdot \mathbf{t}^T + \mathbf{t} \cdot \frac{\partial}{\partial \beta_i} \mathbf{t}^T \right) \right] G^T\end{aligned}$$

Where  $\odot$  denotes the Hadamard (elementwise) product. The normalized taper function is given by:

$$t(d_j | \lambda_j, d^*) = 1 - \frac{1 - \exp\left(-\left(\frac{d_j}{d^*} \lambda_j\right)^2\right)}{1 - \exp(\lambda_j^2)}$$

where  $d_j$  is the depth for the  $j$ th fault grid cell,  $\lambda_j$  is the rate parameter for the grid cell with  $\boldsymbol{\lambda} = \Xi \boldsymbol{\beta}$  for fixed spline basis matrix  $\Xi$ , and  $d^*$  is the down-dip

slip limit for the fault. The derivative is then:

$$\frac{\partial}{\partial \beta_i} t_j = 2\lambda_j \Xi_{ji} \left[ \frac{(\frac{d_j}{d^*})^2 (1 - e^{-\lambda_j^2}) e^{-(\frac{d_j}{d^*} \lambda_j)^2} + (1 - e^{-(\frac{d_j}{d^*} \lambda_j)^2}) e^{-\lambda_j^2}}{(1 - e^{-\lambda_j^2})^2} \right]$$

We can therefore compute the derivative of the covariance between  $Y_i$  and  $Y_j$  without too much difficulty.

Gradient of the expectation of  $\mathbf{Y}$ :

The expectation of  $\mathbf{Y}$  is given by:

$$E[\mathbf{Y}] = \mathbf{G}\mathbf{T}E[\mathbf{Z}]$$

Hence, the gradients have the form:

$$\begin{aligned} \frac{\partial}{\partial \mu} E[\mathbf{Y}] &= \mathbf{G}\mathbf{T} \frac{\partial}{\partial \mu} E[\mathbf{Z}] \\ &= \mathbf{G}\mathbf{t} \\ \frac{\partial}{\partial \sigma_Z} E[\mathbf{Y}] &= \mathbf{G}\mathbf{T} \frac{\partial}{\partial \sigma_Z} E[\mathbf{Z}] = 0 \\ \frac{\partial}{\partial \beta_i} E[\mathbf{Y}] &= \frac{\partial}{\partial \beta_i} \mathbf{G}\mathbf{T}E[\mathbf{Z}] \\ &= \frac{\partial}{\partial \beta_i} \mathbf{G} \operatorname{diag}(E[\mathbf{Z}])\mathbf{t} \\ &= \mathbf{G} \operatorname{diag}(E[\mathbf{Z}]) \frac{\partial}{\partial \beta_i} \mathbf{t} \end{aligned}$$

Since  $\frac{\partial}{\partial \beta_i} t_j$  has already been derived, we can compute the gradient of the expectation of  $\mathbf{Y}$  and also the gradient of the log-likelihood.

Matern covariance gradient:

For Matern correlation function  $\rho(d; \phi, \nu)$  with smoothness parameter assumed to be  $\nu = 3/2$ , the covariance function takes the simplified form (under the

parameterization used on page 31 of Michael Stein's *Interpolation of Spatial Data: Some Theory for Kriging*):

$$\sigma_Z^2 \cdot \rho(d; \phi, 3/2) = \sigma_Z^2 \left(1 + \frac{d}{\phi}\right) \exp\left\{-\frac{d}{\phi}\right\}$$

Hence,

$$\begin{aligned} \frac{\partial}{\partial \sigma_Z} \sigma_Z^2 \cdot \rho(d; \phi, 3/2) &= 2\sigma_Z \cdot \rho(d; \phi, 3/2) \\ \frac{\partial}{\partial \phi} \sigma_Z^2 \cdot \rho(d; \phi, 3/2) &= \sigma_Z^2 \cdot \frac{\partial}{\partial \phi} \rho(d; \phi, 3/2) \\ \frac{\partial}{\partial \phi} \rho(d; \phi, 3/2) &= \frac{\partial}{\partial \phi} \left(1 + \frac{d}{\phi}\right) \exp\left\{-\frac{d}{\phi}\right\} \\ &= -\frac{d}{\phi^2} \exp\left\{-\frac{d}{\phi}\right\} + \left(1 + \frac{d}{\phi}\right) \left(-\frac{d}{\phi}\right) \left(\frac{d}{\phi^2}\right) \exp\left\{-\frac{d}{\phi}\right\} \\ &= -\frac{d}{\phi^2} \exp\left\{-\frac{d}{\phi}\right\} - \frac{d^2}{\phi^3} \rho(d) \\ &= \frac{-\frac{d}{\phi^2}}{1 + \frac{d}{\phi}} \rho(d) - \frac{d^2}{\phi^3} \rho(d) \\ &= -d \left(\frac{1}{\phi^2 + \phi d} + \frac{d}{\phi^3}\right) \rho(d) \end{aligned}$$

Now let  $\mathbf{D}$  be the distance matrix for the observations we compute the covariance of, and let  $\sigma_Z^2 \mathbf{C}$  be the covariance matrix of the latent process of interest where  $\mathbf{C}$  is the correlation matrix. Then,

$$\begin{aligned} \frac{\partial}{\partial \phi} \sigma_Z^2 \mathbf{C} &= \sigma_Z^2 \frac{\partial}{\partial \phi} \mathbf{C} \\ &= -\sigma_Z^2 \left( \frac{\mathbf{D} \odot \mathbf{D}}{\phi^3} + \frac{\mathbf{D}}{\phi^2 + \phi \mathbf{D}} \right) \odot \mathbf{C} \end{aligned}$$

where division is taken elementwise rather than inverting any of the above matrices.

Locking data likelihood gradient (normal model):

Under the normal model we have:

$$\mathbf{X} = \gamma \mathbf{T}\mathbf{Z} + \boldsymbol{\xi} \sim MVN \left( \gamma \mathbf{T}E[\mathbf{Z}], \gamma^2 \mathbf{T}Var(\mathbf{Z})\mathbf{T} + Var(\boldsymbol{\xi}) \right)$$

The log likelihood is therefore

$$\begin{aligned} l(\boldsymbol{\theta}) &\propto -\frac{1}{2} [(\mathbf{X} - \boldsymbol{\mu}_x(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta})(\mathbf{X} - \boldsymbol{\mu}_x(\boldsymbol{\theta}))] - \frac{1}{2} \log(|\boldsymbol{\Sigma}_x|) \\ &= -\frac{1}{2} [\mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \mathbf{X} - 2\mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) + \boldsymbol{\mu}_x(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta})] - \frac{1}{2} \log(|\boldsymbol{\Sigma}_x|) \end{aligned}$$

for  $\boldsymbol{\Sigma}_x = \gamma^2 \mathbf{T}Var(\mathbf{Z})\mathbf{T} + Var(\boldsymbol{\xi})$ . Note that  $\gamma$  is estimated conditional on the other parameters using the formula:

$$\begin{aligned} \hat{\gamma} &= \exp \left\{ \sum_i c_i \log(X_i) - \log(\mu_Z) - \log(t_i) \right\} \\ &= \frac{\exp \{ \sum_i c_i \log(X_i) \}}{\mu_Z} \exp \left\{ - \sum_i c_i \log(t_i) \right\} \end{aligned}$$

For weights  $c_i \propto Var(\xi_i)^{-1}$  summing to 1. Hence, we need to account for the gradient of  $\hat{\gamma}$  with respect to the other parameters in the optimization. Luckily, the gradient only depends on  $\mu_Z$  and  $\boldsymbol{\beta}$ , with:

$$\begin{aligned} \frac{\partial}{\partial \mu_Z} \hat{\gamma} &= -\hat{\gamma}/\mu_Z \\ \frac{\partial}{\partial \beta_j} \hat{\gamma} &= \frac{\exp \{ \sum_i c_i \log(X_i) \}}{\mu_Z} \exp \left\{ - \sum_i c_i \log(t_i) \right\} \cdot \sum_i -\frac{c_i}{t_i} \frac{\partial t_i}{\partial \beta_j} \\ &= \hat{\gamma} (\mathbf{c} \odot \mathbf{t})^T \frac{\partial \mathbf{t}}{\partial \beta_j} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma} &= \hat{\gamma} (\mathbf{c} \odot \mathbf{t})^T \frac{\partial \mathbf{t}}{\partial \boldsymbol{\beta}} \end{aligned}$$

The normal log likelihood gradient is then:

$$\begin{aligned}
\frac{\partial}{\partial \theta_i} l(\boldsymbol{\theta}) &= -\frac{1}{2} \frac{\partial}{\partial \theta_i} [\mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \mathbf{X} - 2 \mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) + \boldsymbol{\mu}_x(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta})] - \frac{1}{2} \frac{\partial}{\partial \theta_i} \log(|\boldsymbol{\Sigma}_x|) \\
\frac{\partial}{\partial \theta_i} \mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \mathbf{X} &= -\mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}_x(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \mathbf{X} \\
-2 \frac{\partial}{\partial \theta_i} \mathbf{X}^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) &= -2 \mathbf{X}^T \left( \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_x(\boldsymbol{\theta})}{\partial \theta_i} - \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}_x(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) \right) \\
\frac{\partial}{\partial \theta_i} \boldsymbol{\mu}_x(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) &= \boldsymbol{\mu}_x(\boldsymbol{\theta})^T \left( \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_x(\boldsymbol{\theta})}{\partial \theta_i} - \boldsymbol{\Sigma}_x^{-1} \frac{\partial \boldsymbol{\Sigma}_x(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) \right) + \frac{\partial \boldsymbol{\mu}_x(\boldsymbol{\theta})^T}{\partial \theta_i} \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \boldsymbol{\mu}_x(\boldsymbol{\theta}) \\
-\frac{1}{2} \frac{\partial}{\partial \theta_i} \log(|\boldsymbol{\Sigma}_x(\boldsymbol{\theta})|) &= -\frac{1}{2} \text{tr} \left( \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\Sigma}_x(\boldsymbol{\theta})}{\partial \theta_i} \right)
\end{aligned}$$

The gradient of these terms depends only on the gradient of  $\boldsymbol{\mu}_x(\boldsymbol{\theta})$  and of  $\boldsymbol{\Sigma}_x(\boldsymbol{\theta})$ , which are:

$$\begin{aligned}
\frac{\partial}{\partial \mu_Z} \boldsymbol{\Sigma}_x &= 2\hat{\gamma} \frac{\partial \hat{\gamma}}{\partial \mu_Z} \mathbf{T} \text{Var}(\mathbf{Z}) \mathbf{T} \\
\frac{\partial}{\partial \sigma_Z} \boldsymbol{\Sigma}_x &= \hat{\gamma}^2 \mathbf{T} \frac{\partial \boldsymbol{\Sigma}_Z}{\partial \sigma_Z} \mathbf{T} \\
\frac{\partial}{\partial \beta_i} \boldsymbol{\Sigma}_x &= \hat{\gamma}^2 \left[ \frac{\partial \mathbf{T}}{\partial \beta_i} \text{Var}(\mathbf{Z}) \mathbf{T} + \mathbf{T} \text{Var}(\mathbf{Z}) \frac{\partial \mathbf{T}}{\partial \beta_i} \right] \\
\frac{\partial}{\partial \phi} \boldsymbol{\Sigma}_x &= \hat{\gamma}^2 \mathbf{T} \frac{\partial \boldsymbol{\Sigma}_Z}{\partial \phi} \mathbf{T} \\
\frac{\partial}{\partial \mu_Z} \boldsymbol{\mu}_x &= \hat{\gamma} \mathbf{T} \frac{\partial}{\partial \mu_Z} E[\mathbf{Z}] + \left( \frac{\partial}{\partial \mu_Z} \hat{\gamma} \right) \mathbf{T} E[\mathbf{Z}] = \hat{\gamma} \mathbf{T} \mathbf{1} - (\hat{\gamma}/\mu_Z) \mathbf{t} \mu_Z = 0 \\
\frac{\partial}{\partial \sigma_Z} \boldsymbol{\mu}_x &= \hat{\gamma} \mathbf{T} \frac{\partial}{\partial \sigma_Z} E[\mathbf{Z}] = 0 \\
\frac{\partial}{\partial \beta} \boldsymbol{\mu}_x &= \hat{\gamma} \cdot \text{diag}(E[\mathbf{Z}]) \frac{\partial \mathbf{t}}{\partial \beta} + \mathbf{T} E[\mathbf{Z}] \left( \frac{\partial}{\partial \beta} \hat{\gamma} \right)^T \\
\frac{\partial}{\partial \phi} \boldsymbol{\mu}_x &= \hat{\gamma} \mathbf{T} \frac{\partial}{\partial \phi} E[\mathbf{Z}] = 0
\end{aligned}$$

## Appendix C: Example Simulations

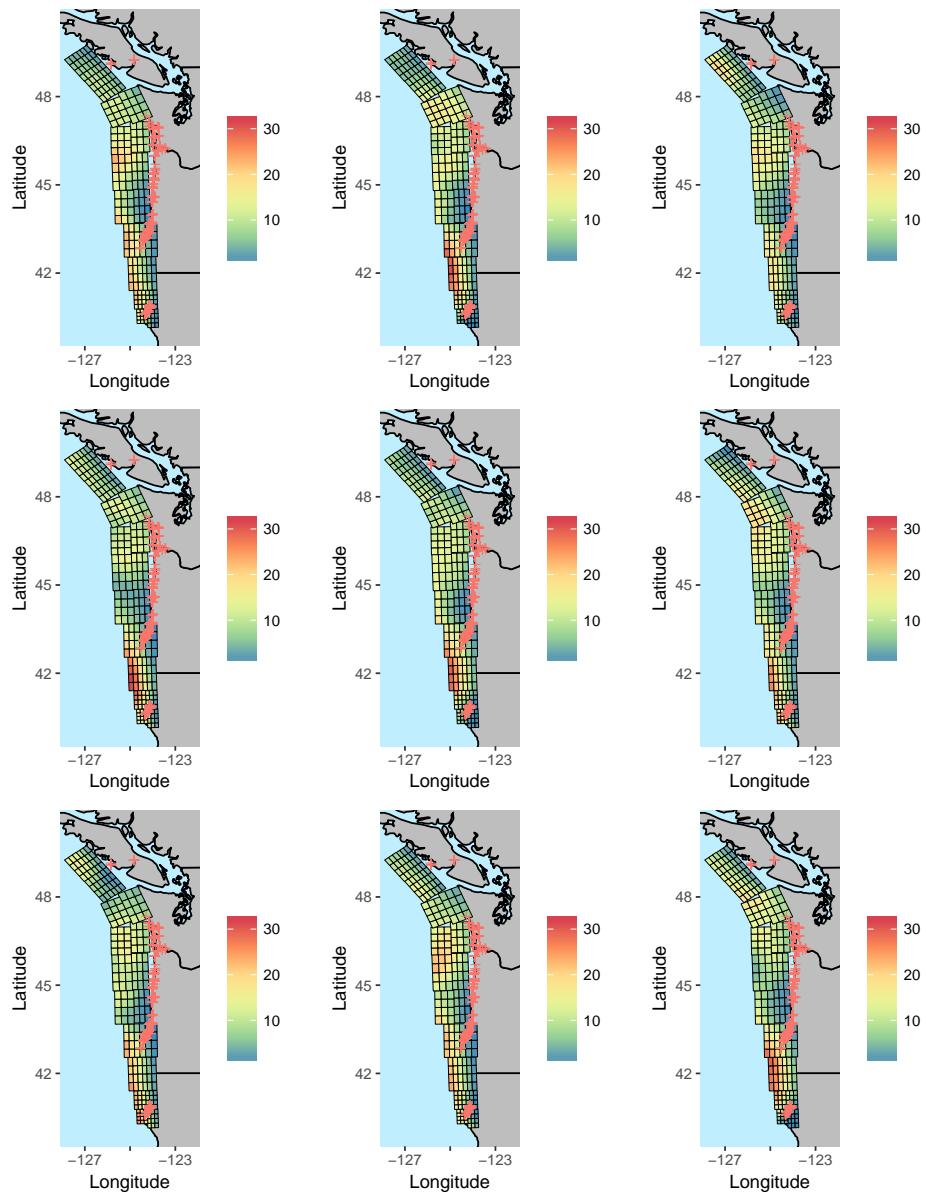


Fig. 10: Simulated 1700 (T1) Cascadia subduction event coseismic slip for the combined taper Gaussian model.

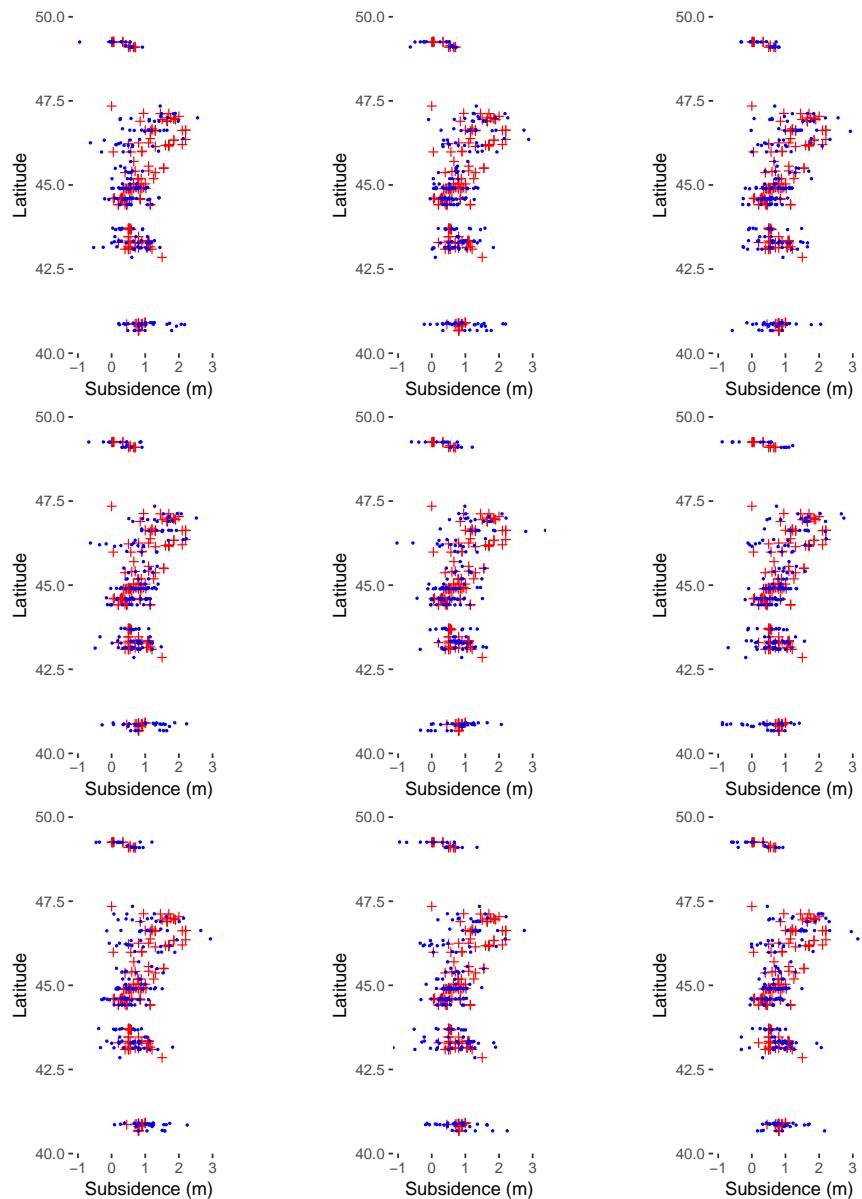


Fig. 11: Simulated 1700 (T1) Cascadia subduction event subsidence (blue dots) with observed subsidence estimates (red '+') for the combined taper Gaussian model.

**Appendix D: 1700 predictions for all models**

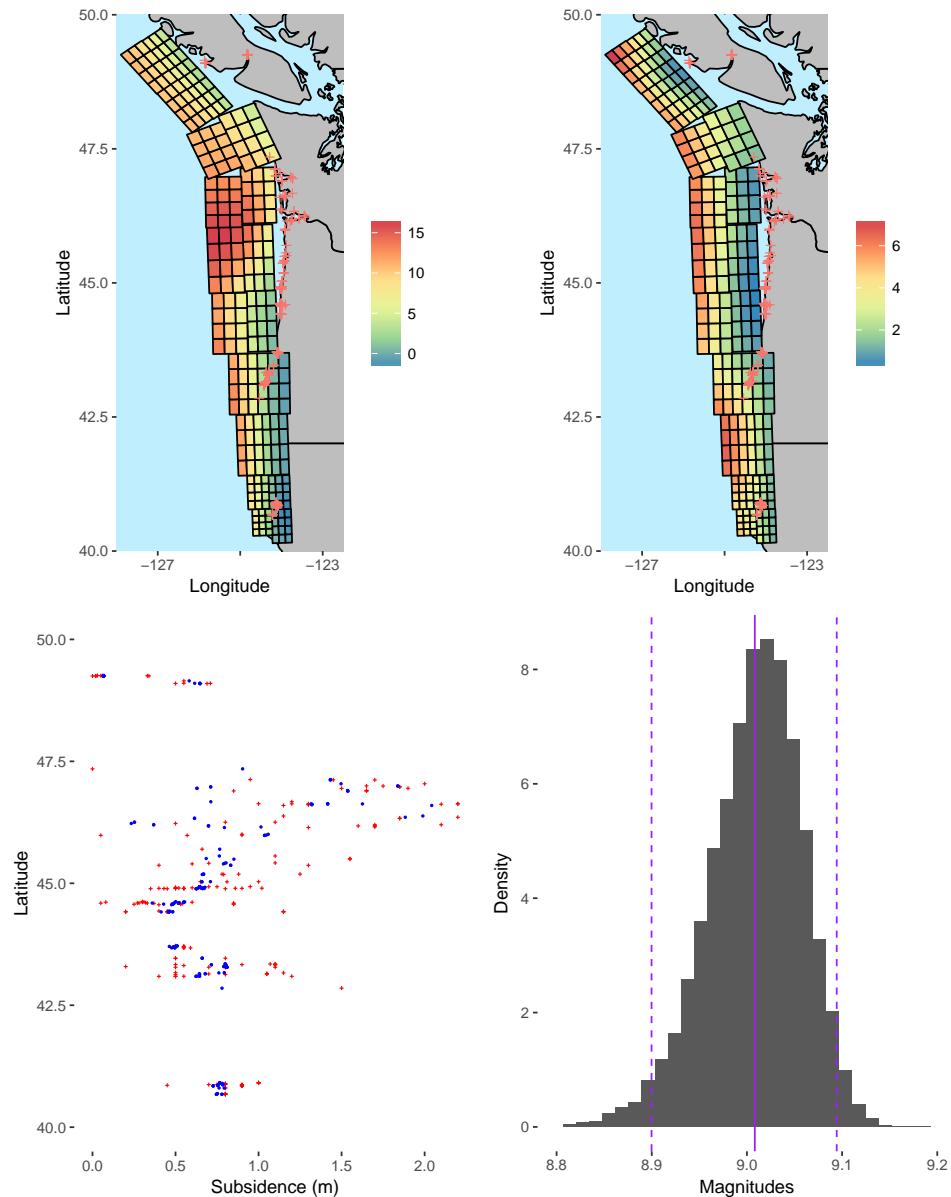


Fig. 12: Inference on the 1700 (T1) Cascadia subduction event for the combined taper Gaussian model. Mean coseismic slip (top left) along with standard error (top right) are shown with subsidence estimates and 95% prediction bands (blue bands, bottom left) and magnitudes (bottom right) with a 95% confidence band (dashed lines) and mean (solid line). Subsidence estimates associated with the 1700 event are plotted as red ‘+’ symbols.

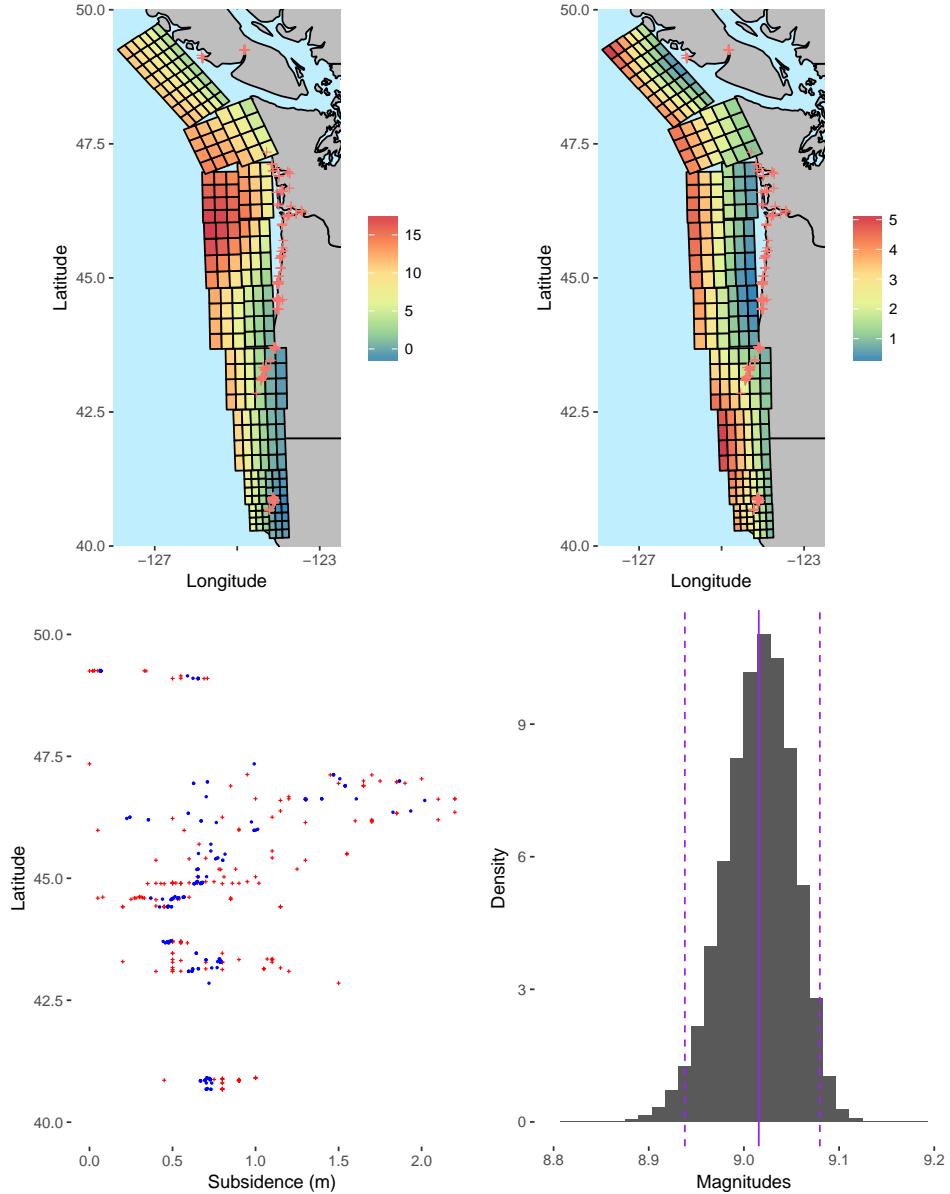


Fig. 13: Inference on the 1700 (T1) Cascadia subduction event for the subsidence taper Gaussian model. Mean coseismic slip (top left) along with standard error (top right) are shown with subsidence estimates and 95% prediction bands (blue bands, bottom left) and magnitudes (bottom right) with a 95% confidence band (dashed lines) and mean (solid line). Subsidence estimates associated with the 1700 event are plotted as red ‘+’ symbols.

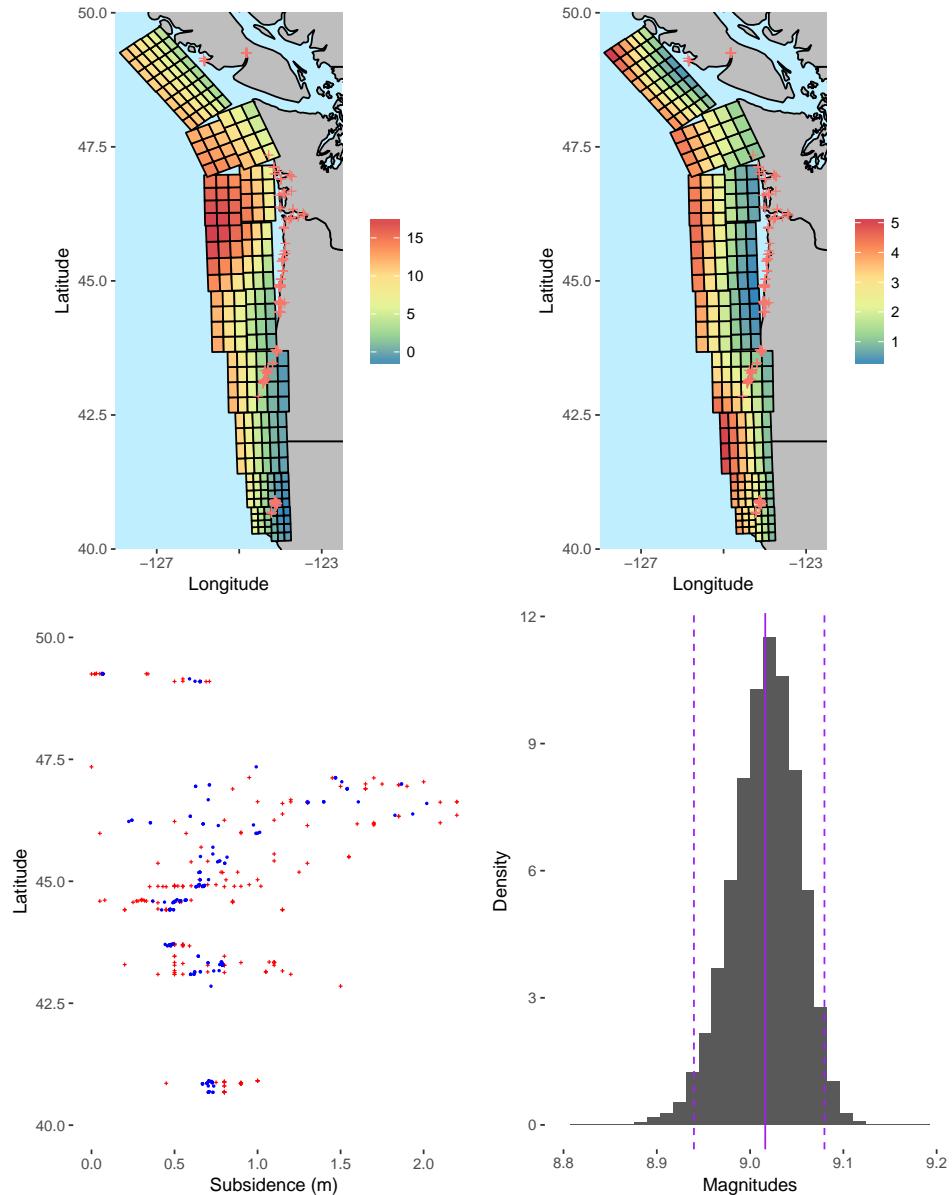


Fig. 14: Inference on the 1700 (T1) Cascadia subduction event for the locking taper Gaussian model. Mean coseismic slip (top left) along with standard error (top right) are shown with subsidence estimates and 95% prediction bands (blue bands, bottom left) and magnitudes (bottom right) with a 95% confidence band (dashed lines) and mean (solid line). Subsidence estimates associated with the 1700 event are plotted as red '+' symbols.