Investigations Into Non-Stationary Correlation Models for Site Terms of Ground-Motion Models due to Basins

Nicolas Kuehn*[1]

[1] *University of California, Los Angeles*

Abstract

Spatial correlation models of ground motions are typically isotropic ad stationary, meaning the correlation of observations at different stations depends only on the absolute inter-station distance. This is a simplifying assumption. In this work, spatial correlation models of ground-motion model site terms are developed. Non-stationary effects due to basins are introduced into the models, first as a separation due to basin index, and then also dependent on a predictor variable, which n this case is the distance to the basin boundary. Different models are compared via the generalization error estimated by cross-validation. Unsuprisingly, spatially correlated models perform better than a non-spatial base models. Models where the spatial correlation is separated inside and outside of a basin see a small improvement over a stationary models. Non-stationary models that depend on the distance to the basin boundary do not generally perform better. In this case, results depend slightly on the cross-validation scheme and the functional dependence on the predictor variable. The models oulined in this work provide a conceptual and computational framework to incorporate non-stationarity into spatial correlation models for ground motions.

1 Introduction

Ground-motion models (GMMs) provide the conditional distribution of a ground-motion parameter of interest (such as peak ground acceleration (PGA)) given earthquake and site related parameters such as magnitude, distance, or local geology. Empirical GMMs are typically estimated via regression, applying a mixed effects approach to account for systematic event and site terms (Al-Atik et al., 2010). Spatial correlation models of ground motion are typically estimated from residuals of empirical GMMs (e.g. Jayaram and Baker, 2009; Kuehn and Abrahamson, 2020; Loth and Baker, 2013).

Spatial correlation of systematic site terms has received less attention than correlation models for within-event residuals. The varying coefficient model (VCM) (Bussas et al., 2017;

*kuehn@ucla.edu

Gelfand et al., 2003) of Landwehr et al. (2016) contains a term that corresponds to a spatially correlated site term. Rahpeyma et al. (2018) modeled the site terms of their model for a small array in Iceland as correlated, and Chao et al. (2020) showed improved predictions of site terms in Taiwan when spatial correlations are taken into account. Recently, Chen et al. (2021) investigated site terms in New Zealand in terms of possible non-stationarity of their correlations. Even if spatial correlations of site terms are not explicitly considered, they are implicit in correlation models for within-event residuals.

Most published correlation models for ground motions are stationary and isotropic, meaning that the correlation depends only on (absolute) distance between stations, not on direction or locality. However, it has been recognized that this assumption is flawed, and that there are non-stationarities in ground-motion spatial correlations (Chen and Baker, 2019; Chen et al., 2021; Infantino et al., 2021; Kuehn and Abrahamson, 2020). Typically, studies of spatial non-stationarity of ground motion correlations estimate correlations between multiple sites, and associate differences in the correlation coefficients between stations to geological or rupture properties, which may explain these differences (Chen and Baker, 2019; Chen et al., 2021; Infantino et al., 2021). This is important, as it identifies the sources of non-stationarity, but it does not lead to a readily available forward model which can be used to sample large maps of spatially correlated ground motions.

Here, we investigate the possibility of modeling spatial correlations of site terms with a non-stationary correlation function. The main motivation is to provide a model (and modeling approach) for the development of nonergodic GMMs, but the same approach can be used to model spatial correlations of ground motions. There is overlap between nonergodic GMMs and spatial correlation models (Anderson and Uchiyama, 2011; Kuehn and Abrahamson, 2020; Walling and Abrahamson, 2012).

The spatial correlation model we use has a narrow focus and is very simple, but it is extendable to account for more complex cases. The model accounts for non-stationarities in site terms inside and outside of basins. Chen et al. (2021) found that there can be differences in the correlation between stations in and outside of basins in New Zealand. We incorporate this effect into the spatial correlation model. In addition, we account for possible different correlation length scales close to the basin boundary and the inside of a basin. The model is estimated for Southern California stations, using the NGA West-2 data set (Ancheta et al., 2014). We use the subset that was selected by Abrahamson et al. (2014) to develop their GMM.

2 Spatial Correlations of Site Terms

The general setup of a GMM is as follows

$$Y = f(\vec{x}) + \delta B + \delta S2S + \delta WS \tag{1}$$

where Y is the target variable (e.g. logarithmic peak ground acceleration (PGA), or the logarithm of pseudo-spectral acceleration (PSA) at different oscillator periods), $f(\vec{x})$ is a function relating the predictor variables $\vec{\xi}$ (such as magnitude, distance, time-averaged shear wave velocity in the upper 30m V_{S30}, and others), δB and $\delta S2S$ are systematic source and site terms, and δWS is the remaining within-event/within-site residual. Typically, the event and site terms are assumed to be independent and distributed according to a normal distribution with mean zero and standard deviation τ (for event terms) and ϕ_{S2S} (for site terms) (Al-Atik et al., 2010).

Assuming that site terms are independent means there is no information about the size of the site term for a new, previously unobserved station. However, as shown by Chao et al. (2020), one can improve the prediction of site terms by accounting for their spatial correlation. Thus, the site terms are not assumed to be independent anymore, but are distributed according to a multivariate normal distribution where the covariance matrix is determined by the spatial correlation model. An alternative, equivalent formulation is that the site terms are a function of location, where the prior distribution for the function is a Gaussian process (GP) (also called a Gaussian field) (Rasmussen and Williams, 2006). This leads to the following hierarchical model formulation

$$\begin{aligned} Y &\sim \mathcal{N}(\mu, \phi_{SS}) \\ \mu &= f(\xi) + \delta B + \delta S2S_0 + f_{stat}(\vec{t}_s) \\ \delta B &\sim N(0, \tau) \\ \delta S2S_0 &\sim N(0, \phi_{S2S,0}) \\ f_{stat} &\sim GP(0, k(\vec{t}_s), \vec{t}_s') \end{aligned} \tag{2}$$

where δB and $\delta S2S_0$ are non-spatially varying event and site terms, τ and $\phi_{S2S,0}$ are their respective standard deviations. We added a subscript 0 to the site terms to indicate that they are different from the site terms in Equation (1); $\delta S2S_0$ is the remaining systematic site term with the spatially correlated part removed. f_{stat} is the systematic, spatially correlated site term, where the spatial correlation is determined by the covariance function $k(\vec{t}_s, \vec{t}_s')$, with $\vec{t}_s$ denoting the station coordinate. While Equation (2) shows a spatial correlation model for site terms, the formulation works equivalently for event terms as well as within-event/within-site residuals (Kuehn and Abrahamson, 2020). The model of Equation (2) can be understood as a VCM with one spatially varying coefficient (a spatially varying constant dependent on station locations).

In spatial correlation models for ground motions, the exponential correlation function is almost universally used, which is defined as

$$k(\vec{t}, \vec{t}') = \omega^2 \exp\left[-\frac{|\vec{t}_s - \vec{t}_s'|}{\ell}\right] \tag{3}$$

where ω is the standard deviation of the GP, and ℓ is a spatial length scale. The exponential covariance function is a special case of the Matérn covariance, which is parameterized as

$$k_{mat}(\vec{t}_s, \vec{t}_s') = \omega^2 \frac{2^{(1-\nu)}}{\Gamma(\nu)} (\kappa|\vec{t}_s - \vec{t}_s'|)^\nu K_\nu(\kappa|\vec{t}_s - \vec{t}_s'|) \tag{4}$$

where Γ is the gamma function, K_ν is the modified Bessel function of the second kind, κ is a scale parameter and ν is a smoothness parameter. For $\nu = 0.5$, the Matérn covariance function becomes the exponential covariance function (Rasmussen and Williams, 2006). The Matérn function with $\nu = 1.5$ has occasionally been used in modeling spatial correlations of ground motions (Kuehn and Abrahamson, 2020; Ming et al., 2019; Tamhidi et al., 2020). We can define the spatial range $\ell = \sqrt{8\nu}/\kappa$, which corresponds to the distance where the correlation is 0.14 (Krainski et al., 2019), and has a similar meaning as the length scale in the exponential covariance function.

In this work, we use the Matérn covariance function with $\nu = 1$. This allows us to estimate the model parameters using Bayesian inference, via the *integrated nested Laplace approximation* (INLA) (Rue et al., 2009). INLA provides fast and efficient inference in spatial models, based on the *stochastic partial differential equation* (SPDE) approach (Bakka et al., 2018; Lindgren and Rue, 2015; Lindgren et al., 2011). Kuehn (2021b) compared nonergodic GMMs (modeled as VCMs) based on the exponential covariance function, as well as the Matérn covariance function with $\nu = 1$, and found that they give very similar results.

The Matérn covariance function as shown in Equation (4) is *isotropic*, meaning that the covariance between stations depends only on the absolute distance $|\vec{t}_s - \vec{t}|'_s$ (if the covariance depends only on the distance, it is called *stationary*). We use the isotropic and stationary model as the spatial base model, and compare three different non-stationary models to it.

A simple way to make the model non-stationary over all observations is to take into account whether the stations are in a basin or not:

$$k(\vec{t}_s, \vec{t}'_s) = \begin{cases} k_{mat}(\vec{t}_s, \vec{t}'_s) & \text{station in same basin} \\ k_{mat}(\vec{t}_s, \vec{t}'_s) & \text{station outside of basin} \\ 0 & \text{else} \end{cases} \tag{5}$$

In this model, the site terms within each basin, as well as the stations outside each basin, are modeled with an isotropic covariance function as before, but are assumed to be uncorrelated with stations outside the basin or in other basins.

The model of Equation (5) assumes that there is an abrupt change in the spatial correlation between site terms at basin boundaries. However, inside a basin the spatial correlation is stationary/isotropic, meaning the spatial correlation of two stations at the basin boundary and well inside a basin is the same if their distances are the same. In a second non-stationary model, we relax his assumption, by making the range ℓ of the covariance function dependent on the distance to the basin boundary. Such an approach was used by Kuehn and Abrahamson (2020), who modeled the spatial correlation of within-event residuals with a non-stationary covariance function where the length-scale is dependent of source-to-site distance, based on a covariance function proposed by (Paciorek and Schervish, 2006). Ingebrigtsen et al. (2014) showed how to incorporate predictor variables into a correlation structure in the SPDE approach. Here, we use the approach of Ingebrigtsen et al. (2014) and model the range of the spatial correlation as dependent on the distance to the basin boundary, i.e.

$$\ell = \exp\left[\theta_1 + \theta_2 \min(d_B, d_{ref})\right] \tag{6}$$

where d_B is the distance to the basin boundary, and d_{ref} is a reference distance introduced so that the range ℓ does not increase to infinity with increasing distance d_B.

Finally, one can combine the two non-stationary models, which leads to a model where stations in different basins are uncorrelated, and the correlation inside each basin, as well as outside of basins, depends on the distance to the basin.

3 Model and Implementation

We investigate the spatial correlation of the site terms by fitting a model like Equation (2) to residuals of the model of Abrahamson et al. (2014), hereafter called ASK14. Hence, our

target variable are the total residuals of ASK14, which we partition into event terms δB, site terms $\delta S2S$, and spatially correlated site terms f_{stat}. The data are described later in Section 4. While one could first partition the residuals into event terms and site terms, and then estimate a spatial correlation model for the site terms, this approach faces the problem that the site terms are associated with considerable uncertainty, in particular for stations that do not record many events. Estimating all terms of the model simultaneously takes this uncertainty in the random effects into account.

We compare 5 different models, which are distinguished based on the correlation structure for the site terms:

1. A "standard" model, with no spatial correlation between site terms, called *non-spatial*.
2. A model where the site terms are modeled with a stationary Maérn correlation function, called *stationary*.
3. A model where the site terms are modeled with a stationary Maérn correlation function, and only stations inside the same basin are correlated (cf. Equation (5)), called *stationary basin*.
4. A model where the range of the Matérn covariance function depends on distance to basin edge (cf. Equation (3)), and all stations are correlated, called *non-stationary*.
5. A model where only stations within the same basin are correlated, with the non-stationary correlation function dependent on distance to basin edge, called *non-stationary basin*.

Note that for simplicity, same basin means same basin index, i.e. outside of basin is treated as a group, just like a basin.

All models are implemented via the `R` package `R-INLA`[1] (Bivand et al., 2015; Lindgren and Rue, 2015; Rue et al., 2017), which allows one to perform approximate Bayesian inference in latent Gaussian models, and is The mathematical background of INLA is laid out in Rue et al. (2009). Gómez-Rubio (2020) or Krainski et al. (2019) provide introductions (with code examples) into the computations with `R-INLA`[2].

Spatial modeling with INLA is based on the SPDE approach, which allows for fast and efficient computation. Here, the spatial Gaussian field is approximated by basis functions which are evaluated on a triangular mesh (Lindgren and Rue, 2015; Lindgren et al., 2011). Thus, the model for the spatial terms is actually

$$\vec{f}_{mesh} = N(0, \mathbf{P}^{-1})$$

$$\vec{f}_{stat} = \mathbf{A}\ \vec{f}_{mesh} \tag{7}$$

where $\vec{f}_{mesh}$ is the spatial field evaluated at the nodes of the triangular mesh, $\vec{f}_{stat}$ is the spatial field a the observed station locations, and $\mathbf{A}$ is a projector matrix of size $N_{mesh} \times N_{stat}$ connecting the mesh nodes to the observation (station) locations. $\mathbf{P}^{-1}$ is the covariance matrix between the mesh nodes. When using a Matérn covariance function with $\nu = 1$, the precision

[1] https://www.r-inla.org/

[2] The books are available at https://becarioprecario.bitbucket.io/inla-gitbook/index.html and https://becarioprecario.bitbucket.io/spde-gitbook/index.html

matrix becomes sparse, which ca be exploited numerically. Kuehn (2021b) and Kuehn (2021a) showed how one can efficiently fit a nonergodic GMM with INLA. Here, we follow the The implementation of Krainski et al. (2019) for the non-stationary spatial models.

The hyperparameters to be estimated are the standard deviations ϕ_{SS}, τ, and $\phi_{S2S,0}$, and the standard deviation ω_{stat} and range ℓ_{stat} of the spatial field. In `R-INLA`, the standard deviations are internally represented as logarithmic precisions (precision is one divided by the variance). Fo ϕ_{SS}, we use a log-Gamma distribution as the prior for the corresponding log precision, with shape parameter 2 and rate parameter 0.5. Thus, the prior mean of the precision is 4.

The prior distributions for the other hyperparameters are based on a penalized complexity (PC) prior Fuglstad et al. (2019); Simpson et al. (2017). The PC prior encodes a notion of parsimony, treating the model with random effects as a more complex extension of a simpler base model. The PC prior shrinks the parameters towards the simpler base model, with zero marginal variance and infinite range. For details, see (Franco-Villoria et al., 2019; Fuglstad et al., 2019; Simpson et al., 2017). The PC prior is specified by setting the tail probability that the standard deviation of the spatial field is larger that a certain value, and its range smaller than a certain value. Here, we use the following values for the PC prior of the spatial field of the site terms:

$$Pr(\omega_{stat} > 0.5) = 0.01 \tag{8}$$

$$Pr(\ell < 100) = 0.99 \tag{9}$$

For the standard deviations τ and ϕ_{S2S}, we use the following values:

$$Pr(\tau > 0.5) = 0.01 \tag{10}$$

$$Pr(\phi_{S2S,0} > 0.5) = 0.01 \tag{11}$$

For the non-stationary models, we need to specify prior distributions for θ_1 and θ_2, which describe the dependence of the range on the distance to the basin edge. The priors for these parameters are normal distributions with mean zero and precision 1.

4 Data

We estimate the spatial correlation models using a subset of the NGA West 2 data set (Ancheta et al., 2014). We use the Southern California data of the subset that was used for the development of the ASK14 model. Here, we select only records from California with a latitude South of 34.5 degrees. In total, there are 6235 records from 149 events, recorded at 782 stations. Table 1 lists the number of records and stations in each basin in the study area.

In the NGA West 2 data set, there is a small number of stations which have the same coordinates, but different values of V_{S30}. Similarly, for a small number of cases there are different station IDs for the same coordinate. We combine stations with the same coordinate by assigning them the same ID, and the geometric mean of the V_{S30}-values.

For the analysis, the station coordinates are transformed into UTM coordinates. The mesh, needed for the SPDE approximation to he Matérn covariance function, is created to to contain the station UTM coordinates, with a maximum edge length of 4km, and an outer boundary of 30km. Figure 1 shows the mesh, together with the station coodinates and the basin boundaries.

Table 1: Number of stations in each basin.

Basin index	Basin name	Number of records	Number of stations
1	LA basin	727	172
2	Ventura Basin	24	9
3	San Fernando basin	145	47
4	San Gabriel Basin	182	28
5	Chino Basin	254	26
6	San Bernardino Basin	186	31
7	Coachella Valley Basin	141	14
8	Imperial Valley Basin	401	68
9	outside basin	4175	387

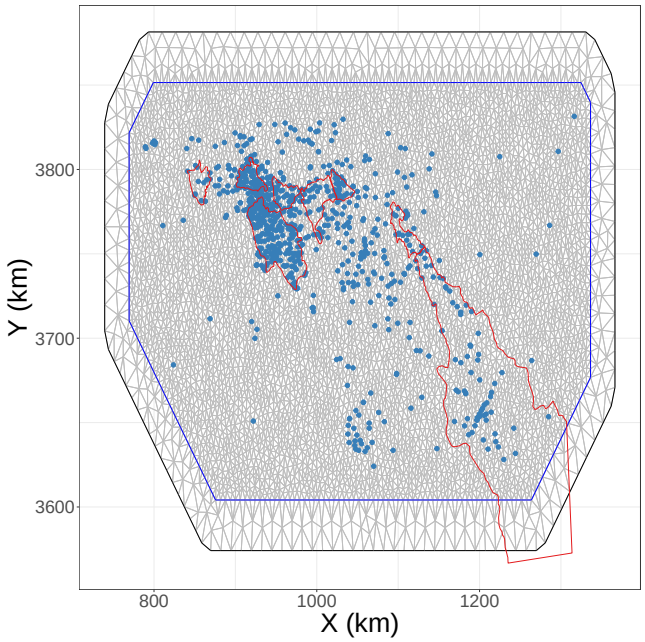

Figure 1: Plot of mesh, together with station locations and basin outlines.

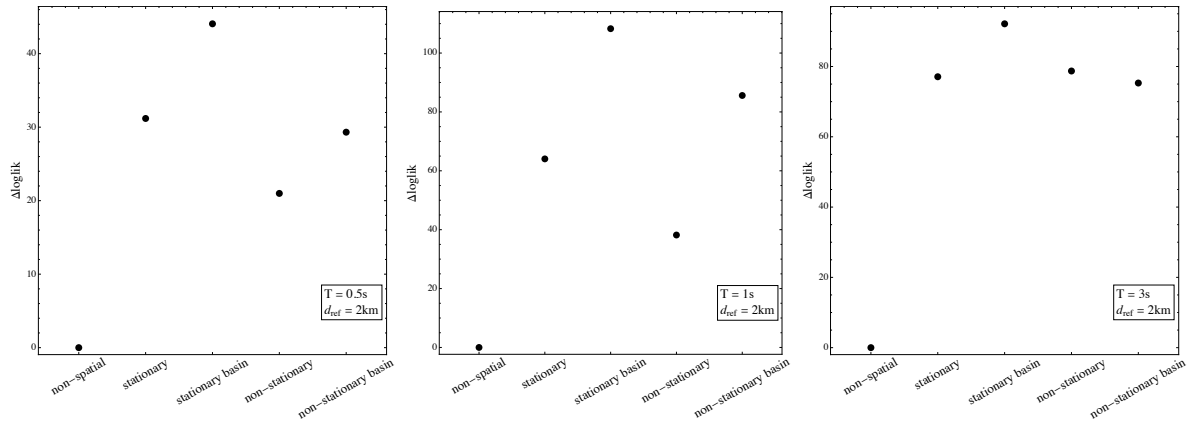

Figure 2: Differences in total log-likelihood of the different models, estimated by cross-validation, to the non-spatial model. Note the different scale for the different periods.

5 Results

5.1 Model Comparison

The different models described in the previous section are compared based on their predictive error, estimated by cross-validation. Since the non-stationary models affect the stations close to a basin boundary, we perform the cross-validation only over stations with small distances to a basin boundary d_B. We select all stations with $d_B < 3$km, and partition these into 10 groups. We loop over each group, leaving out all records from the current set of stations as test data, and estimate the different models on the remaining data. We then calculate predictions for the test data (both predictive mean and uncertainty), and compute the likelihood and residual of each test data point. The predictive uncertainty includes uncertainty in all estimated parameters (hyperparameters and random effects).

We perform the cross-validation procedure for three periods, $T = 0.5, 1., 3.$s, and calculate the total log-likelihood as the sum of the the log-likelihood values of each test data point, as well as the root mean squared error (RMSE) over all test data points. Figure 2 shows the difference in log-likelihood to the non-spatial model (larger value indicates a better model), and Figure 3 shows the RMSE values (smaller value indicates a better model). In general, one can see that the spatial models perform better than the non-spatial reference model, which is in line with results of e.g. Chao et al. (2020). The improvements are larger at longer periods. Assuming independence between stations inside a basin ad those outside leads to a small increase in likelihood and decrease of RMSE, indicating a better model performance. The non-stationary spatial models in general perform slightly worse their stationary counterparts.

For the results shown in Figures 2 and 3, a value of $d_{ref} = 2$km is used, which is the distance beyond which the spatial range becomes constant (cf. Equation (6)). We tested different values for d_{ref} (frpm 1km to 4km), but the results are not very sensitive to it. Results are show in the Appendix. The cross-validation results are a bit sensitive to the cut-off distance used to select the test stations: When only stations closer to the basin boundary are used, the non-stationary models tend to perform better compared to the stationary ones. Results for different cut-off values are also shown in the Appendix.

For computational reasons, the cross-validation procedure is only done for three periods

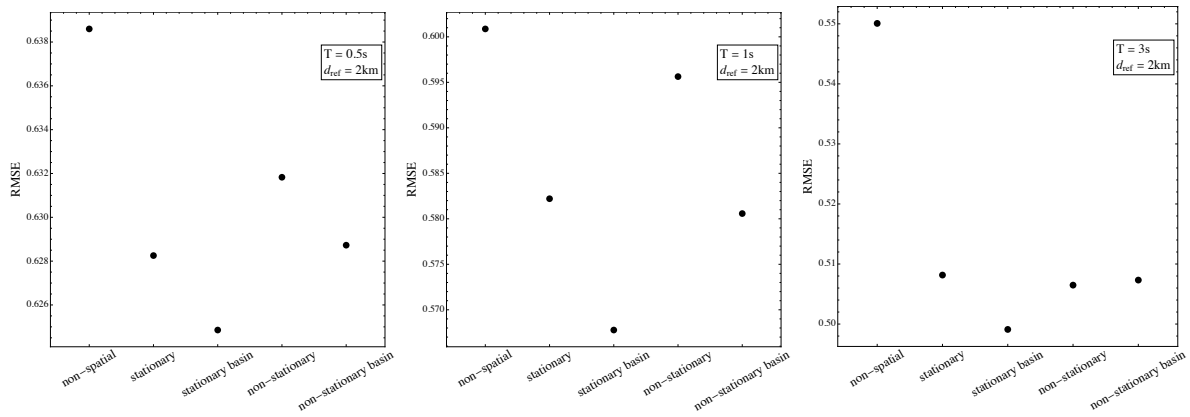

Figure 3: RMSE for the different models, based on cross-validation. Note the different scale for the different models.

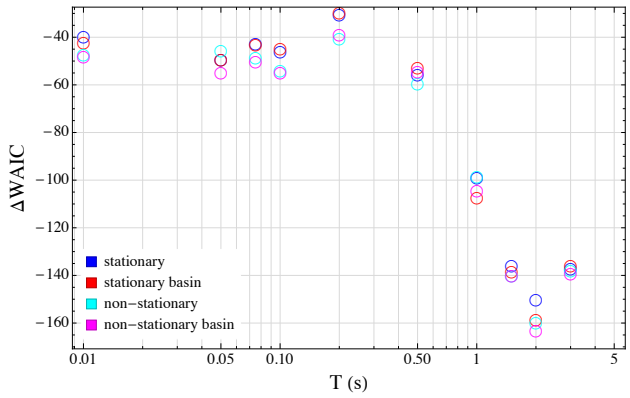

Figure 4: Differences in WAIC values to the non-spatial model. WAIC valus are based on all observations.

$T = 0.5$s, $T = 1$.s, and $T = 3$.s. Figure 4 shows a model comparison for 10 periods from $T = 0.01$s to $T = 33s$, based on the widely applicable information criterion (WAIC) (Vehtari et al., 2017; Watanabe, 2010). Displayed are the differences in WAIC to the non-spatial base model, with lower values indicating a better model. By contrast to the results based on cross-validation, the results shown in Figure 4 are based on all observations, not just the ones from stations close to a basin boundary. For Figure 4, a value of $d_{ref} = 2$km is used.

Overall, the results from WAIC are similar to the ones based on cross-validation, with the spatial models showing a clear improvement in model performance over the non-spatial base model. The difference in WAIC is larger at long periods. The different spatial models exhibit similar WAIC values. In general, the models that separate the spatial correlation of site terms inside and outside of a basin perform slightly better. Regarding the inclusion of a dependence of the spatial range on distance to the basin boundary, results are mixed, with an improvement at some, but not all periods.

5.2 Parameter Values

We now estimate the parameters of the different models for the 10 periods from = 0.01s to $T = 3$.s, using all data. The distance where the spatial range becomes constant for the non-stationary models is $d_{ref} = 2$km The values of the standard deviations ϕ_{SS}, τ, ϕ_{S2S}, and ω_{stat}

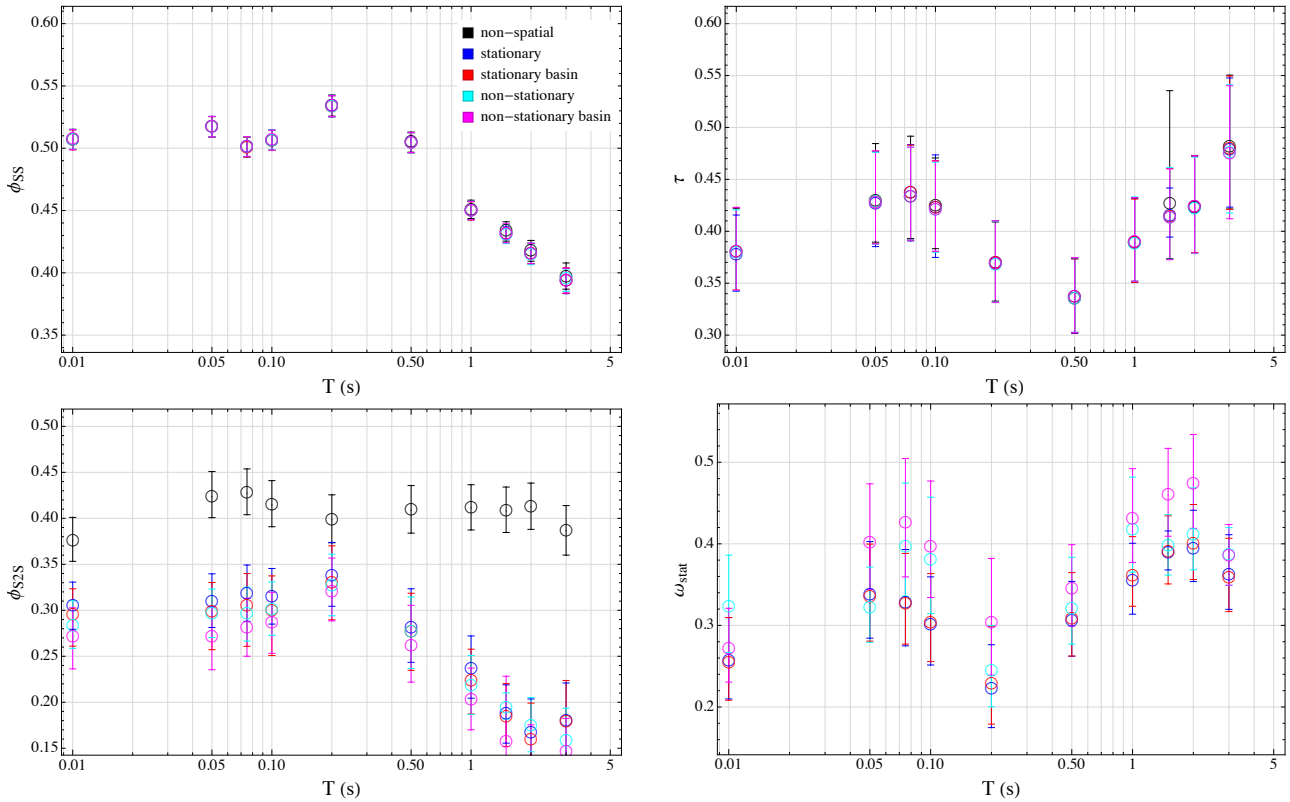

Figure 5: Estimated values of standard deviations (median of posterior distribution) for the different models. The error bars denote the 5% and 95% quantiles of the posterior distribution.

are shown in Figure 5. Both ϕ_{SS} and τ are very similar across the different models, which is not surprising since they describe the within-event/within-station variability as well as the variability of event terms, which is unaffected by the modeling of site terms. By contrast, the stationary models show much smaller values of the between-station variability ϕ_{S2S} compared to the stationary model; this reduction is offset by the variability associated with the spatially correlated terms, ω_{stat}. The reduction of ϕ_{S2S} due o the inclusion of spatial correlation of site terms is larger long periods, which corresponds to the larger increase in model performance seen in Figures 2 and 3. Within the spatial models, the estimated values of ϕ_{S2S} and ω_{stat} are similar.

Figure 6 shows the estimated values of the spatial range ℓ_{stat}. There s some variation between the models, but overall they show consistent behavior. The non-stationary models, where the spatial range depends on the distance to the basin, reach similar values to the stationary models for $d_b > d_{ref}$, i.e. where the spatial range flattens off.

6 Discussion

We have investigated the possible non-stationarity of spatial correlations of systematic site terms, from a perspective of directly modeling a non-stationary correlation function. Investigations into non-stationary spatial correlation of ground motions have become more common recently (e.g. Chen and Baker, 2019; Chen et al., 2021; Infantino et al., 2021). Incorporating he non-stationarity into the correlation function allows one to directly sample ground-motion values from a forward model.

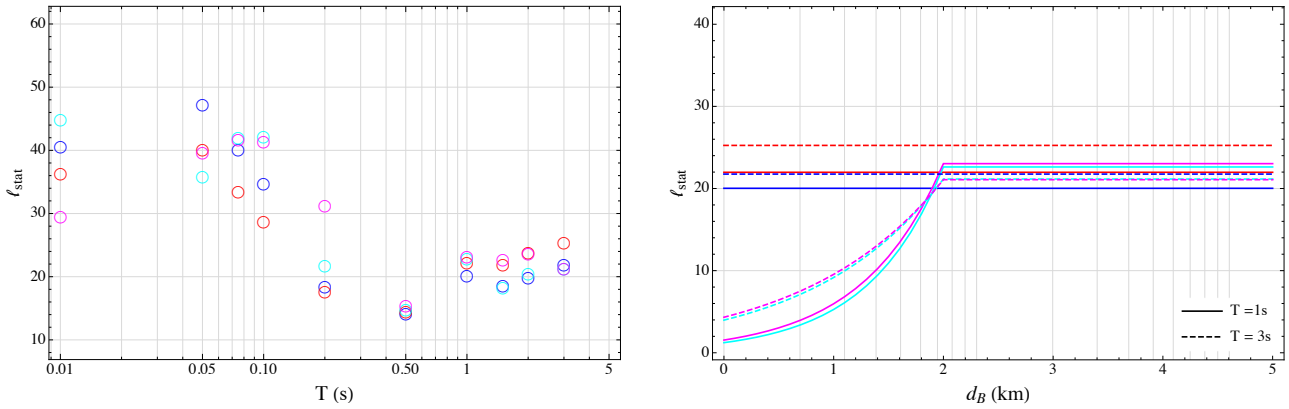

Figure 6: Left: Estimated values of the spatial rang ℓ_{stat} for the different spatial models. For the no-stationary models, the plateau value ($d_B > d_{ref}$ is plotted). Right: Dependence of spatial range on distance to basin boundary for two periods. The color-scale is the same as in Figure 5.

Our focus in this work is on non-stationarity in site terms due to basin effects. The models employed in this work are very simple, but we find that separating stations inside and outside of a basin in the correlation function leads to better model performance, both based on cross-validation and and WAIC. While these models are non-stationary, they can be considered "piece-wise stationary", as the spatial correlation model inside each basin is stationary. We also tested non-stationary models where the spatial range (or length scale) depends on the distance to the basin boundary. The assumption for this model would be that maybe the local geology at the basin boundary is more complex, and thus site terms are less correlated. Results with respect to these models are inconclusive, as they do not consistently perform better than the (piece-wise) stationary models. The ranking of the models is different for different periods, which test stations are used, as well as the distance d_{ref} where the spatial range becomes constant (cf. Equation (6)). We think it is best to not take this as evidence for or against the non-stationary models. The data is quite limited, and the models employed are still quite simple.

Part of the motivation for the current work is also to provide a conceptual and computational framework to model the non-stationarity of ground motions. For applications in e.g. risk analysis (Miller and Baker, 2015), one needs to be able to generate maps of spatially correlated ground-motion values, so it is beneficial to have a model of spatial non-stationarty that can directly be applied in a forward fashion. While the models used employed here are quite simple, they can be extended to include more predictor variables, to account for more complex differences in geology, or other effects such as source processes (Infantino et al., 2021; Schiappapietra and Smerzini, 2021). Other extensions one could think of is to relax the assumption that the spatial range is the same for all basins.

Code and Data Availability

R-INLA code to fit the models described in this work and data are available at https://github.com/nikuehn/NonStationary_SiteTerms.

Acknowledgements

We would like to thank Chukwuebuka Nweke for providing the basin boundaries, and Yousef Bozorgnia, Christine Goulet, Grigorios Lavrentiadis and other members of the nonergodic working group at UC Berkeley/UCLA for discussions. All models are implemented in the `R` computing environment (R Core Team, 2021), using the `R-INLA` package (Bivand et al., 2015; Lindgren and Rue, 2015; Rue et al., 2017). The basin distances were calculated with the computer program Mathematica. Plots were made with Mathematica, except for the mesh plot, which used the `R`-packages `ggplot2` (Wickham, 2016) and `inlabru` (Bachl et al., 2019). Other `R`-packages used are `sp` (Bivand et al., 2013) and `rgdal` (Bivand et al., 2021).

Partial support of Pacific Gas & Electric Company and California Department of Transportation are gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the sponsors.

References

Abrahamson, N. A., Silva, W. J., and Kamai, R. (2014). "Summary of the ASK14 Ground Motion Relation for Active Crustal Regions." *Earthquake Spectra*, 30(3), 1025–1055.

Al-Atik, L., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N. (2010). "The Variability of Ground-Motion Prediction Models and Its Components." *Seismological Research Letters*, 81(5), 794–801.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. (2014). "NGA-West2 Database." *Earthquake Spectra*, 30(3), 989–1005.

Anderson, J. G. and Uchiyama, Y. (2011). "A Methodology to Improve Ground-Motion Prediction Equations by Including Path Corrections." *Bulletin of the Seismological Society of America*, 101(4), 1822–1846.

Bachl, F. E., Lindgren, F., Borchers, D. L., and Illian, J. B. (2019). "inlabru: an R package for Bayesian spatial modelling from ecological survey data." *Methods in Ecology and Evolution*, 10(6), 760–766.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). "Spatial modeling with R-INLA: A review." *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), e1443.

Bivand, R., Keitt, T., and Rowlingson, B. (2021). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*, <https://CRAN.R-project.org/package=rgdal>. R package version 1.5-23.

Bivand, R. S., Gómez-Rubio, V., and Rue, H. (2015). "Spatial Data Analysis with R - INLA with Some Extensions." *Journal of Statistical Software*, 63(20), 1–31.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY, <https://asdar-book.org/>.

Bussas, M., Sawade, C., Kühn, N., Scheffer, T., and Landwehr, N. (2017). "Varying-coefficient models for geospatial transfer learning." *Machine Learning*, 106(9-10), 1419–1440.

Chao, S.-H., Lin, C.-M., Kuo, C.-H., Huang, J.-Y., Wen, K.-L., and Chen, Y.-H. (2020). "Implementing horizontal-to-vertical Fourier spectral ratios and spatial correlation in a ground-motion prediction equation to predict site effects." *Earthquake Spectra*, 875529302095244.

Chen, Y. and Baker, J. W. (2019). "Spatial Correlations in CyberShake Physics-Based Ground-Motion Simulations." *Bulletin of the Seismological Society of America*, 109(6), 2447–2458.

Chen, Y., Bradley, B. A., and Baker, J. W. (2021). "Nonstationary spatial correlation in New Zealand strong ground-motion data." *Earthquake Engineering & Structural Dynamics*, 50(13), 3421–3440.

Franco-Villoria, M., Ventrucci, M., and Rue, H. (2019). "A unified view on Bayesian varying coefficient models." *Electronic Journal of Statistics*, 13(2), 5334–5359.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2019). "Constructing Priors that Penalize the Complexity of Gaussian Random Fields." *Journal of the American Statistical Association*, 114(525), 445–452.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003). "Spatial Modeling With Spatially Varying Coefficient Processes." *Journal of the American Statistical Association*, 98(462), 387–396.

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman and Hall/CRC, Boca Raton, FL.

Infantino, M., Smerzini, C., and Lin, J. (2021). "Spatial correlation of broadband ground motions from physics-based numerical simulations." *Earthquake Engineering and Structural Dynamics*, 50(10), 2575–2594.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). "Spatial models with explanatory variables in the dependence structure." *Spatial Statistics*, 8(C), 20–38.

Jayaram, N. and Baker, J. W. (2009). "Correlation model for spatially distributed ground-motion intensities." *Earthquake Engineering & Structural Dynamics*, 38(15), 1687–1708.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F. K., and Rue, H. (2019). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC, Boca-Raton, FL.

Kuehn, N. (2021a). "A Comparison of Nonergodic Ground-Motion Models based on Geographically Weighted Regression and the Integrated Nested Laplace Approximation." *Engrxiv*.

Kuehn, N. (2021b). "Comparison of Bayesian Varying Coefficient Models for the Development of Nonergodic Ground-Motion Models." *Engrxiv*, 1–25.

Kuehn, N. M. and Abrahamson, N. A. (2020). "Spatial correlations of ground motion for nonergodic seismic hazard analysis." *Earthquake Engineering & Structural Dynamics*, 49(1), 4–23.

Landwehr, N., Kuehn, N. M., Scheffer, T., and Abrahamson, N. (2016). "A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients." *Bulletin of the Seismological Society of America*, 106(6), 2574–2583.

Lindgren, F. and Rue, H. (2015). "Bayesian Spatial Modelling with R - INLA." *Journal of Statistical Software*, 63(19), 1–25.

Lindgren, F., Rue, H., and Lindström, J. (2011). "An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach." *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(4), 423–498.

Loth, C. and Baker, J. W. (2013). "A spatial cross-correlation model of spectral accelerations at multiple periods." *Earthquake Engineering & Structural Dynamics*, 42(3), 397–417.

Miller, M. and Baker, J. (2015). "Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization." *Earthquake Engineering & Structural Dynamics*, 44(7), 1139–1156.

Ming, D., Huang, C., Peters, G. W., and Galasso, C. (2019). "An Advanced Estimation Algorithm for Ground-Motion Models with Spatial Correlation." *Bulletin of the Seismological Society of America*, 109(2), 541–566.

Paciorek, C. J. and Schervish, M. J. (2006). "Spatial modelling using a new class of nonstationary covariance functions." *Environmetrics*, 17(5), 483–506.

R Core Team (2021). "R: A Language and Environment for Statistical Computing, <https://www.r-project.org/>.

Rahpeyma, S., Halldorsson, B., Hrafnkelsson, B., and Jónsson, S. (2018). "Bayesian hierarchical model for variations in earthquake peak ground acceleration within small-aperture arrays." *Environmetrics*, 29(3), 1–19.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, <http://www.gaussianprocess.org/gpml/>.

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). "Bayesian Computing with INLA: A Review." *Annual Review of Statistics and Its Application*, 4(1), 395–421.

Schiappapietra, E. and Smerzini, C. (2021). "Spatial correlation of broadband earthquake ground motion in Norcia (Central Italy) from physics-based simulations." *Bulletin of Earthquake Engineering*, 19(12), 4693–4717.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science*, 32(1), 1–28.

Tamhidi, A., Kuehn, N. M., Ghahari, S. F., Taciroglu, E. T., and Bozorgnia, Y. (2020). "Conditioned Simulation of Ground Motion Time Series using Gaussian Process Regression." *Engrxiv*, 1–15.

Vehtari, A., Gelman, A., and Gabry, J. (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and Computing*, 27(5), 1413–1432.

Walling, M. and Abrahamson, N. A. (2012). "Non-Ergodic Probabilistic Seismic Hazard Analyses." *15th World Conference on Earthquake Engineering (15WCEE).*

Watanabe, S. (2010). "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research*, 11, 3571–3594.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, <https://ggplot2.tidyverse.org>.

A Model Comparison based on Cross-Validation with Different Parameter Settings

In the appedix, we show some model comparisons using different values of the reference distance d_{ref} (Figures A.1 and A.2), as well as different subsets of test stations (Figues A.3 and A.4).

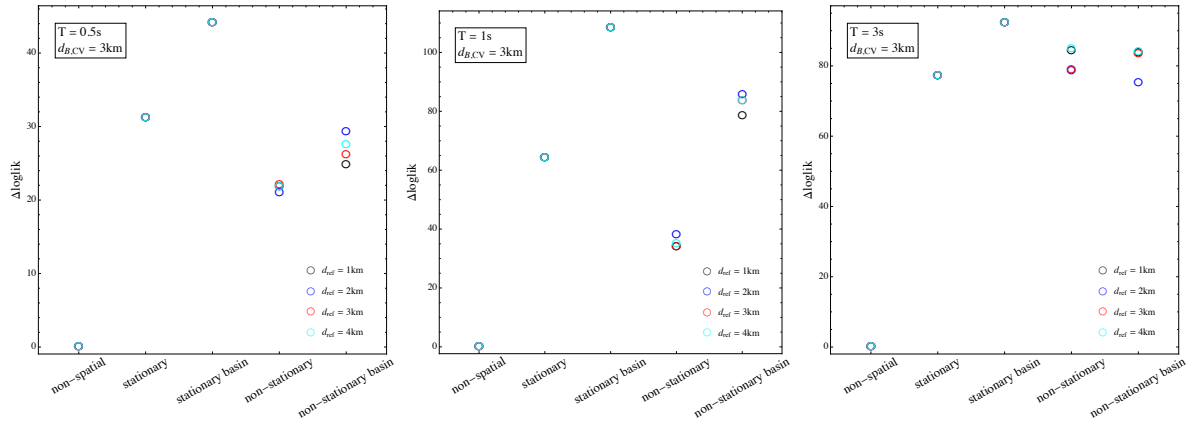

Figure A.1: Differences in total log-likelihood of the different models, estimated by cross-validation, to the non-spatial model, for different values of d_{ref}. Note the different scale for the different periods.

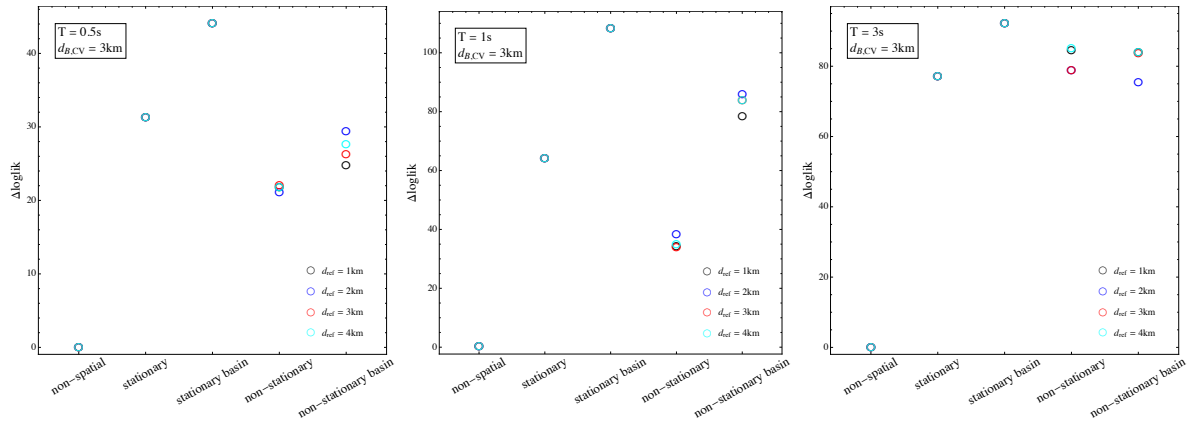

Figure A.2: RMSE values of the different models, estimated by cross-validation, for different values of d_{ref}. Note the different scale for the different periods.

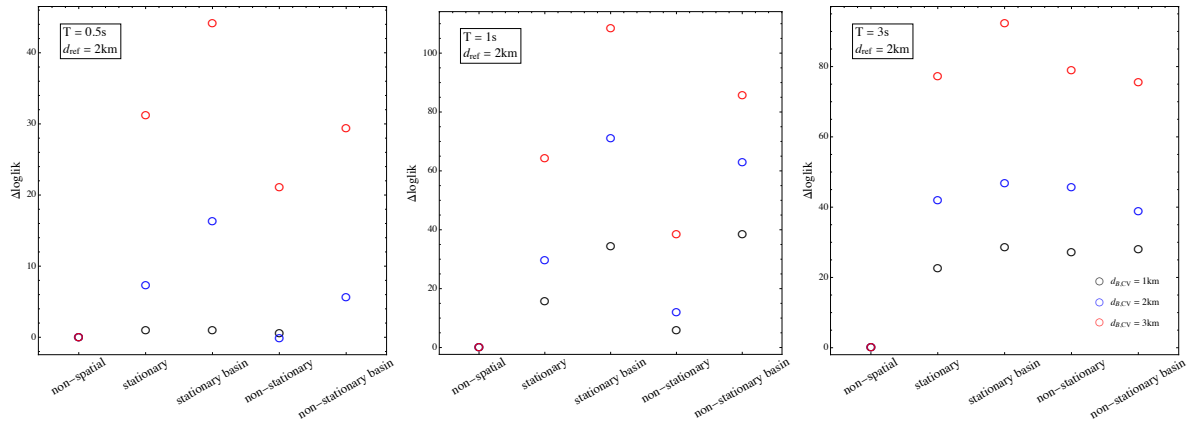

Figure A.3: Differences in total log-likelihood of the different models, estimated by cross-validation, to the non-spatial model, for different values of the cutoff distance used to select the test stations. Note the different scale for the different periods.

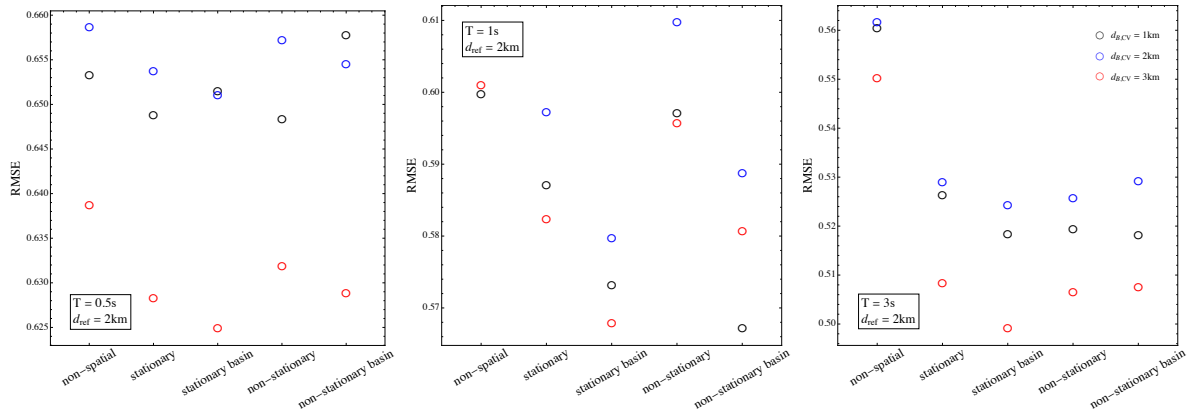

Figure A.4: RMSE valuesof the different models, estimated by cross-validation, for different values of the cutoff distance used to select the test stations. Note the different scale for the different periods.