

TNO PUBLIEK

Princetonlaan 6
3584 CB Utrecht
P.O. Box 80015
3508 TA Utrecht
The Netherlandswww.tno.nlT +31 88 866 42 56
F +31 88 866 44 75**TNO report****TNO 2021 R11742 | update 2022****Status of the TNO Model Chain Groningen per
October 1, 2021 and recommendations for the
public Seismic Hazard and Risk Analysis 2022**

Date	28 April 2022
Author(s)	
Copy no	
No. of copies	
Number of pages	55 (incl. appendices)
Number of appendices	4
Sponsor	Ministry of Economic Affairs and Climate Policy
Project name	Publieke SDRA Groningen
Project number	060.47420/01.04

All rights reserved.

No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

In case this report was drafted on instructions, the rights and obligations of contracting parties are subject to either the General Terms and Conditions for commissions to TNO, or the relevant agreement concluded between the contracting parties. Submitting the report for inspection to parties who have a direct interest is permitted.

© 2022 TNO

TNO PUBLIEK

Samenvatting

Status TNO Modelketen Groningen

De TNO Modelketen Groningen is ingericht om de publieke Seismische Dreigings- en Risico Analyse (SDRA) Groningen uit te voeren. De TNO Modelketen Groningen is onderverdeeld in drie hoofdcomponenten: het Seismisch Bronmodel (SSM), het Grondbewegingsmodel (GMM), en het Kwetsbaarheids- en Gevolgmodel (FCM). De technische status van alle beschikbare modelcomponenten per 1 oktober 2021 omvat de modelversies gebruikt voor de HRA 2019, HRA 2020, pSDRA 2021, en daarnaast alternatieve versies van het SSM en het FCM die door TNO geadviseerd zijn te gebruiken in de pSDRA 2021. Alle geïmplementeerde model versies beschreven in dit rapport kunnen gecombineerd worden voor gebruik in de publieke Seismische Dreigings- en Risico Analyse Groningen.

Aanbevolen modellen in de publieke SDRA Groningen 2022

Op basis van de beschikbare informatie per 1 oktober 2021 beschouwt TNO onderstaande (sub)modelversies als het meest geschikt voor gebruik in de publieke SDRA Groningen 2022. TNO is van mening dat deze geadviseerde modelversies de beschikbare wetenschappelijke kennis en inzichten het beste vertegenwoordigen.

- TNO adviseert gebruik te maken van de meest recente aardbevingscatalogus om het Seismisch Bronmodel (SSM) te kalibreren voor de publieke SDRA Groningen 2022. De door TNO geïmplementeerde SSM kalibratiemodule is de enige volledig transparante applicatie om deze taak uit te voeren.
- TNO adviseert het gebruik van SSM versie TNO-2020 in de publieke SDRA Groningen 2022. Deze modelversie wordt gekenmerkt door het gebruik van een distributie van stress covariante velden en een magnitudeverdeling die begrensd wordt door een Mmax verdeling.
- TNO adviseert het gebruik van GMM versie NAM-V6 in de publieke SDRA Groningen 2022. In dit model zijn de period-to-period correlaties geïmplementeerd voor zowel de grondbewegingen op het referentieniveau als de amplificatie functies van het site response model.
- TNO adviseert het gebruik van FCM versie TNO-2020 in de publieke SDRA Groningen 2022. Dit model representeert de best beschikbare kennis van de kwetsbaarheid van de Groningse gebouwenpopulatie en is in lijn met uitvoering van de 'Typologie-gebaseerde beoordeling van de veiligheid'.

Summary

Status TNO Model Chain Groningen

The TNO Model Chain Groningen is equipped to execute the public Seismic Hazard and Risk Analysis Groningen. The TNO Model Chain Groningen is subdivided into three main model components: Seismic Source Model (SSM), Ground Motion Model (GMM), and the Fragility and Consequence Model (FCM).

The technical status of all available model component per October 1, 2021 includes the model versions used for HRA 2019, HRA 2020, pSHRA 2021, as well as alternative versions of SSM and FCM recommended by TNO for use in pSHRA 2021. All implemented model versions mentioned in this report can be combined in the public Seismic Hazard and Risk Analysis Groningen.

Recommended models in the public SHRA Groningen 2022

Based on the available information as of October 1, 2021, TNO recommends the following versions of (sub) models as most suitable for use in the public SHRA Groningen 2022. In our opinion, these recommended model versions reflect the best available scientific knowledge to perform the next seismic hazard and risk analysis for the Groningen gas field.

- TNO recommend to include the most recent earthquake observations to perform the source model calibration for the public SHRA Groningen 2022. The TNO provided SSM calibration module is the only fully transparent application to perform this task.
- TNO recommends the use of SSM version TNO-2020 in the public SHRA Groningen 2022. This approach uses a full posterior distribution of stress covariate fields and uses a magnitude distribution that is truncated by a Mmax distribution.
- TNO recommends the use of GMM version NAM-V6 in the public SHRA Groningen 2022. In this model, the period-to-period correlations are implemented not only for the ground motions at reference level, but also for the amplification/attenuation functions of the site response model.
- TNO recommends the use of FCM version TNO-2020 in the public SHRA Groningen 2022. This model represents the best current knowledge of the fragility of the Groningen building stock and is in line with current practice within the framework of the 'typology-based safety assessment'.

Contents

	Samenvatting	2
	Summary	3
1	Introduction.....	5
2	Technical Status per October 1, 2021.....	6
2.1	Technical status SSM.....	6
2.2	Technical status GMM.....	9
2.3	Technical status FCM.....	10
2.4	Technical status EDB	11
3	Recommendations for the public Seismic Hazard and Risk Analysis (SHRA) Groningen 2022.....	12
3.1	Recommended SSM.....	12
3.2	Recommended GMM.....	13
3.3	Recommended FCM.....	14
4	References	15
5	Signature	17
	Appendices	
	A Seismic Source Model calibration	
	B On Bayesian inference of the exponentially tapered magnitude distribution	
	C Period-to-period correlation model for public SHRA Groningen	
	D Model description FCM version TNO-2020	
	E Additional material	

1 Introduction

As from 2021 the Probabilistic Seismic Hazard and Risk Analysis (SHRA) of the Groningen gas field, required for state approval of the Operational Strategy for the yearly gas production, is executed in the public domain. To fulfill this task TNO has independently rebuild and implemented the NAM HRA models in the so called TNO Model Chain Groningen, which has been successful in reproducing similar output using similar input (TNO, 2019; 2020c). This has resulted in the first analysis by TNO for gas year 2021-2022 reported as “public SHRA Groningen 2021” or in Dutch “publieke SDRA Groningen 2021” (TNO, 2021).

This report describes the status of the TNO Model Chain per October 1, 2021 (chapter 2), as well as TNO’s recommendations for model components to be used in the public SHRA Groningen 2022 (chapter 3). The scope is twofold: (i) supply an inventory of the available model components that form the basis of the TNO Model Chain; and (ii) give an overview of the TNO recommended model versions to run the public Seismic Hazard and Risk Analysis (SHRA) Groningen 2022 for the “Vaststellingsbesluit gasjaar 2022-2023”.

2 Technical Status per October 1, 2021

The TNO Model Chain Groningen is subdivided into three main model components: Seismic Source Model (SSM), Ground Motion Model (GMM), and the Fragility and Consequence Model (FCM).

The TNO Model Chain Groningen requires input that must be provided by external parties:

- Seismic Source model input:
 - Catalogue of induced earthquakes (KNMI)
 - Static: reservoir thickness, compressibility, fault data (NAM)
 - Dynamic: past and future reservoir pore pressure corresponding to the required production scenario (NAM)
- Extraction of the Exposure Database (EZK)

In the following these models are described in more detail. All TNO implemented model versions mentioned in this report can be combined in the public Seismic Hazard and Risk Analysis Groningen. Implemented model versions are in this report represented in a tabular format with reference to the model documentation, implementation report and model usage in past HRA/pSHRA.

2.1 Technical status SSM

In the TNO Model Chain Groningen the model versions listed below have been implemented, based on the scientific documentation and mathematical, numerical and/or algorithmic representation herein.

2.1.1 NAM-V5

The SSM version NAM-V5 was used for the HRA 2019. An overview of this model and the different sub models are listed in Table 1. Note that a number of sub models are continued to use in the HRA 2020 and pSHRA 2021, although part of a different SSM main version. The model calibration for this version is considered as external input.

Table 1 Overview of SSM version NAM-V5

SSM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V5	Bourne and Oates, 2018; Bourne and Oates, 2019	TNO, 2020a	x		
Sub models					
NAM-Model calibration provided as input		not part of HRA	x	x	x
Coulomb stress predictor for activity rate		TNO, 2020a	x		
Activity rate		TNO, 2020a	x	x	x
ETAS		TNO, 2020a	x	x	x
FMD: inverse power law b-value & Mmax distribution		TNO, 2020a	x		

2.1.2 NAM-V6

The SSM version NAM-V6 was used for the HRA 2020 and pSHRA 2021. An overview of this model and the different sub models are listed in Table 2. Note that a number of sub models were previously used in HRA 2019. The model calibration for this version is considered as external input.

Table 2 Overview of SSM version NAM-V6

SSM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V6	Bourne et al., 2019	TNO, 2020c		x	x
Sub models					
NAM-Model calibration provided as input		not part of HRA	x	x	x
Coulomb stress predictor for activity rate Coulomb stress predictor for to magnitude distribution		TNO, 2020c		x	x
Activity rate		TNO, 2020a	x	x	x
ETAS		TNO, 2020a	x	x	x
FMD: hyperbolic tangent b-value & Mmax distribution		TNO, 2020c		x	x
FMD: single b-value & exponential taper & Mmax distribution		TNO, 2020c		x	x

The NAM-V6 model is a further development, and builds upon NAM-V5. This version uses two distinct Coulomb stress fields, conditioned to respectively activity rate and magnitude distribution.

The functional form of the b-value of Magnitude Distribution (MD) changed from inverse power law to hyperbolic tangent. This MD is, like in NAM-V5, truncated by a Mmax distribution. This Mmax distribution is the result of expert elicitation (NAM, 2016).

A second MD was introduced in NAM-V6: a single b-value model truncated by an exponential taper. In contrast to the Mmax truncation, this taper model is, calibrated to the Groningen earthquake catalogue. To date there is no scientific consensus about the reliability of the taper calibration based on the Groningen earthquake catalogue, and therefore the predictive capability of this model (TNO, 2020b; Appendix B).

2.1.3 TNO-2020

SSM version TNO-2020 was proposed by TNO (2020d) for usage in the public SHRA 2021. An overview of this model and the different sub models are listed in Table 3. The TNO-2020 implementation is a selection of sub models that were previously used in HRA 2019 & 2020 and pSHRA 2021 and alternative implementations of sub models based on the original documentation (Bourne et al., 2019). These implementations are extensively documented (TNO, 2020a; 2020c).

Table 3 Overview of SSM version TNO-2020

SSM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
TNO-2020	Bourne et al., 2019, Appendix A & B	TNO, 2020c			
Sub models					
TNO-Model calibration		TNO, 2020a			

Coulomb stress distribution predictor for activity rate	TNO, 2020a			
Activity rate	TNO, 2020a	x	x	x
ETAS	TNO, 2020a	x	x	x
MD: hyperbolic tangent b-value & Mmax distribution	TNO, 2020c		x	x

The main update with reference to version NAM-V6 is the removal of the tapered MD model. This is not the introduction of a new seismological model, but rather the rejection of a model that got introduced in the context of the Groningen hazard and risk calculations by NAM in HRA 2020. The TNO-2020 version undoes the introduction of the tapered MD model and falls back on the previously used NAM-V5 model version. An in-depth justification for this is provided in Appendix B. The hyperbolic tangent model does not suffer from the same problem and is considered adequate for the purpose.

A final update compared to NAM-V6 is the use of a single Coulomb stress field distribution, instead of two maximum likelihood stress field realizations for activity rate and magnitude distribution respectively. This adaptation was already implemented in the first public version of the TNO Model Chain (TNO, 2020a) and enables the incorporation of uncertainty with respect to calibration of the Coulomb stress.

In addition to this, the model calibration is explicitly included as part of this model version. Appendix A describes the TNO calibration approach that was already implemented in the first public version of the TNO Model Chain (TNO, 2020a).

2.1.4 All available components

Table 4 lists all available SSM model components that can be used for a pSHRA. Apart from the three predefined model versions described one can decide to deviate from these options and compile an alternative model version.

Table 4 Overview of all available SSM sub models & version

SSM: all sub models	TNO software implementation	HRA		pSHRA
		2019	2020	2021
NAM-Model calibration provided as input	not part of HRA	x	x	x
TNO-Model calibration	TNO, 2020a			
Single Coulomb stress conditioned to activity rate	TNO, 2020a	x		
Coulomb stress predictor for activity rate Coulomb stress predictor for magnitude distribution	TNO, 2020c		x	x
Coulomb stress distribution	TNO, 2020a			
Activity rate	TNO, 2020a	x	x	x
ETAS	TNO, 2020a	x	x	x
MD: constant b-value & Mmax distribution	TNO, 2020a			
MD: inverse power law b-value & Mmax distribution	TNO, 2020a	x		
MD: hyperbolic tangent b-value & Mmax distribution	TNO, 2020c		x	x
MD: single b-value & exponential taper & Mmax distribution	TNO, 2020c		x	x

2.2 Technical status GMM

The TNO Model Chain Groningen has implemented the model versions listed below based on the scientific documentation and mathematical, numerical and/or algorithmic representation herein.

2.2.1 NAM-V5

The GMM version NAM-V5 was used for the HRA 2019.

Table 5 Overview of GMM version NAM-V5

GMM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V5	Bommer et al., 2017	TNO, 2020a	x		

2.2.2 NAM-V6

The GMM version NAM-V6 was used for the HRA 2020. The update from GMM NAM-V5 to NAM-V6 consists of updated input tables only. The model structure and logic is unchanged. In this model, like the NAM-V5, the period-to-period correlations are implemented not only for the ground motions at reference level, but also for the amplification/attenuation functions of the site response model.

Table 6 Overview of GMM version NAM-V6

GMM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V6	Bommer et al., 2019	TNO, 2020c		x	

Within the KEM research program, the quality of the model was assessed (KEM, 2020). Observed was that ground motions at short hypocentral distances were underestimated, caused by the damping model.

2.2.3 NAM-V6-2021

The GMM NAM-V6-2021 was used for the pSHRA 2021. This version has an alternative implementation of the correlation structure of the period-to-period residuals for the site response model. This model assumes no period-to-period correlation for the amplification/attenuation functions of the site response model. The latter choice is subject to scientific discussion (Appendix C).

Table 7 Overview of GMM version NAM-V6-2021

GMM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V6-2021	Bommer et al., 2019	TNO, 2020c			x

2.2.4 NAM-V7

During completion of this report the final version of the NAM-V7 model was published (13 October 2021). The information in this report is based on the provided documentation of the NAM-V7 model (Bommer et al., 2021) that was at that time in the final stage of completion and review. On the basis of the draft documentation it can be concluded that this model is substantially different from NAM-V6. Changes

are not limited to updated input tables, but include changes of functional forms, model logic and structure. Moreover, the underlying model set up is more complex.

On the basis of the draft documentation TNO meanwhile has started model implementation. However, the NAM-V7 model is not yet implemented in the TNO Model Chain Groningen. Therefore the model is not tested and differences and impact on hazard and risk compared to NAM-V6 have not yet been quantified. At this stage this model version is not available for use in the pSHRA Groningen.

2.3 Technical status FCM

The TNO Model Chain Groningen implemented the model versions listed below based on the scientific documentation and mathematical, numerical and/or algorithmic representation herein.

2.3.1 NAM-V6

The FCM version NAM-V6 was used for the HRA 2019.

Table 8 Overview of FCM version NAM-V6

FCM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V6	Crowley et al., 2019	TNO, 2020a	x		

2.3.2 NAM-V7

The FCM version NAM-V7 was used for the HRA 2020 and pSHRA 2021. The update from FCM NAM-V6 to NAM-V7 consists of updated parameter files. The model structure and logic is otherwise unchanged. The parameter files of only the most vulnerable unreinforced masonry classes were updated in NAM-V7. Updating not all unreinforced masonry vulnerability classes introduced a number of undesired inconsistencies (TNO, 2020b).

Table 9 Overview of FCM version NAM-V7

FCM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
NAM-V7	Crowley and Pinho, 2020	TNO, 2020a		x	x

2.3.3 TNO-2020

FCM version TNO-2020 was proposed by TNO (2020d) for usage in the public SHRA 2021. The update from FCM NAM-V7 to TNO-2020 consists of updated parameter files. The model structure and logic is otherwise unchanged. This model version is described in more detail in Appendix D.

Compared to NAM-V7 this model updates fragility and consequence model parameters and model uncertainties for the unreinforced masonry vulnerability classes. This model is based on extensive review and model validation executed within the framework of *typology based safety assessment for Groningen earthquakes* summoned by the Ministry of Economic Affairs and Climate Policy, and independently reviewed by the Advisory Board Safety Groningen (ACVG).

Table 10 Overview of FCM version TNO-2020

FCM version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
TNO-2020	Appendix D	TNO, 2020a			

2.4 Technical status EDB

The Exposure Database (EDB) is an extract of the building stock database and contains information specific for Hazard and Risk Modelling. This extract is provided by EZK. The TNO Model Chain implemented the EDB V7.1 (Arup, 2021a). The next version of the EDB (V8) is to be delivered in February 2022 (Arup, 2021b) and aims to update the *wierde* database field, which indicates whether the building is located on a *wierde*.

3 Recommendations for the public Seismic Hazard and Risk Analysis (SHRA) Groningen 2022

Based on the available information as of October 1, 2021, TNO recommends the following versions of (sub) models as most suitable for use in the public SHRA Groningen 2022 (Table 11). In our opinion, these recommended model versions reflect the best available scientific knowledge to perform the next seismic hazard and risk analysis for the Groningen gas field. The impact of the recommended model versions on the hazard and risk analysis is described in the pSHRA Groningen 2021 (TNO, 2021).

Table 11 Overview of recommended model versions public SHRA Groningen 2022

model version	documentation	TNO software implementation	HRA		pSHRA
			2019	2020	2021
SSM					
TNO-2020	Bourne et al., 2019, Appendix A & B	TNO, 2020c			
SSM sub models					
TNO-Model calibration		TNO, 2020a			
Coulomb stress distribution predictor for activity rate		TNO, 2020a			
Activity rate		TNO, 2020a	x	x	x
ETAS		TNO, 2020a	x	x	x
MD: hyperbolic tangent b-value & Mmax distribution		TNO, 2020c		x	x
GMM					
NAM-V6	Bommer et al., 2019, Appendix C	TNO, 2020c		x	
FCM					
TNO-2020	Appendix D	TNO, 2020a			

3.1 Recommended SSM

TNO recommends the use of version TNO-2020 in the public SHRA Groningen 2022, as described in paragraph 2.1. This approach uses a full posterior distribution of stress covariate fields and uses a Magnitude distribution that is truncated by a Mmax distribution and therefore more in line with the NPR NEN webtool. The impact of this model version on the hazard and risk analysis is described in the pSHRA Groningen 2021 (TNO,2021)

3.1.1 Calibration

TNO recommends to include the most recent earthquake observations to perform the public SHRA Groningen 2022. In contrast to previous years, the source model calibration should be performed by TNO, within the same Quality Assurance system that applies to the entire TNO Model Chain Groningen.

This approach maximizes the share of work done in the public domain, increases traceability and reproducibility, and reduces the dependency of the SHRA result on external inputs. Appendix A describes the TNO calibration approach.

TNO emphasizes that the externally delivered calibration for use in pSHRA 2021 could not be reproduced based on current NAM documentation and available NAM software code. The impact on risk is substantial (TNO, 2021).

3.1.2 *Coulomb stress field*

Regarding the Coulomb stress field, TNO recommends using a posterior distribution of conditioning parameters obtained from Bayesian inference from the observations. The general advantage of using posterior distributions rather than point estimates is that uncertainties/variabilities are accommodated and the result is more robust to variations in the input data.

If for some reason one does not wish to incorporate the Coulomb stress conditioning uncertainties/variabilities, TNO recommends to use a single Coulomb stress covariate conditioned to the activity rate (cf. NAM-V5).

3.1.3 *Magnitude model*

TNO (2020c) has demonstrated that a tapered Magnitude model cannot be calibrated reliably on an earthquake catalogue of the size that is available for Groningen and relies heavily on prior information that is currently inadequately specified and justified. For this reason the predictive capability of this model is considered very poor (TNO, 2020b).

In fact, the taper location and its hypothesized stress-dependence cannot be resolved from the observations. Either the model parameters or their prior distributions should be treated as epistemic uncertainty, in a similar fashion to M_{max} . This is described in more detail in Appendix B.

TNO strongly recommends that the stress-dependent exponential taper model, which receives 80% weight in the logic tree used in HRA 2020 and pSHRA 2021, should not be used for the calculation of hazard and risk. This is not the proposal of a new seismological model, but rather the rejection of a model that got introduced in the context of the Groningen hazard and risk calculations by NAM in HRA 2020. We propose to undo this introduction and to fall back on the previously used model. This is substantiated by the proposal to use only the hyperbolic tangent Magnitude Model present in NAM-V6.

3.2 **Recommended GMM**

TNO recommends the use of GMM version NAM-V6 in the public SHRA Groningen 2022. This version follows the original NAM implementation that was used for the HRA 2020. In this model, the period-to-period correlations are implemented not only for the ground motions at reference level, but also for the amplification/attenuation functions of the site response model. Justification for this implementation is described in Appendix C. This recommendation is strengthened by the observation that in the newly developed GMM NAM-V7 the period-to-period correlations are fully correlated for the site response model.

For the public SHRA Groningen model development TNO recommends to use the *Stage Gate* approach to carefully guard the process of making available newly developed models. With the finalisation including review of the NAM-V7 model the *Gate* towards the implementation *Stage* has been passed. After a successful implementation *Stage*, a testing and validation *Stage* will follow to judge if the implementation adequately represents the model. After a successful testing *Stage*

the impact of the model to the current Hazard and Risk methodology can be assessed. After this last *Stage* one can decide to include the model in the next public SHRA Groningen.

Although TNO started NAM-V7 implementation, the timeline for passing all *Gates* is uncertain and at this moment TNO cannot guarantee finalisation in order for EZK to make a timely decision whether to include this model version in the public SHRA Groningen 2022.

3.3 Recommended FCM

TNO recommends the use of FCM version TNO-2020 in the public SHRA Groningen 2022. This model represents the best current knowledge of the vulnerability of the Groningen building stock and is in line with current practice within the framework of the 'typology-based safety assessment'.

4 References

- Arup (2021a). EDB V7.1 Cover Note. File reference 229746_031.0_NOT2101, 1 February 2021.
- Arup (2021b). Exposure Database V8.0: Scope Proposal_Rev.0.02. File reference 229746_031.0_MEM4000, 19 April 2021.
- Bommer, J., B. Edwards, P. Kruiver, A. Rodriguez-Marek, P. Stafford, B. Dost, M. Ntinalexis, E. Ruigrok and J. Spetzler (2017). V5 Ground-Motion Model (GMM) for the Groningen Field, October 2017.
- Bommer, J., B. Edwards, P. Kruiver, A. Rodriguez-Marek, P. Stafford, B. Dost, M. Ntinalexis, E. Ruigrok and J. Spetzler (2019). V6 Ground-Motion Model (GMM) for Induced Seismicity in the Groningen Field With Assurance Letter, December 2019.
- Bommer, J., B. Edwards, P. Kruiver, A. Rodriguez-Marek, P. Stafford, M. Ntinalexis, E. Ruigrok and B. Dost (2021). V7 Ground-Motion Model for Induced Seismicity in the Groningen Gas Field (Draft for review & implementation only), 2 July 2021.
- Bourne, S. J. & Oates, S. J. (2018). Note for File: Finite rupture simulation for ground motion modelling and probabilistic seismic hazard and risk analysis for the Groningen gas field. Draft Internal NAM report.
- Bourne, S.J., Oates S.J., Scheefhals, R., Castellanos Nash, P., Mar-Or, A., Storck, T., Omid, P., and Van Elk, J. (2019). A Monte Carlo method for probabilistic seismic hazard and risk analysis of induced seismicity in the Groningen gas field, Shell, Alten Technical Software and NAM, May 2019.
- Bourne, S. J. & Oates, S. J. (2019). Evolution of induced earthquake magnitude distributions with increasing stress in the Groningen gas field, Restricted Draft, NAM, November 2019.
- Crowley, H., Pinho, R. and Cavalieri, F. (2019). Report on the v6 Fragility and Consequence Models for the Groningen Field, March 2019.
- Crowley, H. & Pinho, R. (2020). Report on the Fragility and Consequence Models for the Groningen Field (version 7), March 2020.
- KEM (2020). Evaluation, validation and improvement of the Site Amplification component of the Groningen Risk Model (KEM-02), F. Besseling, A. Bougioukos, J. de Greef, J. Pruiksmá, A. Tsouvalas, KEM 02 Research Question 2 report, Ministry of Economic Affairs and Climate Policy, 1 April 2020.
- NAM, (2016). Report on Mmax Expert Workshop. 8-10 March 2016. <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/cef44262-323a-4a34-afa8-24a5afa521d5>

TNO (2019). Comparative analysis of the NAM and TNO implementation in the Groningen Seismic Hazard and Risk Assessment, TNO 2019 R11997.

TNO (2020a). Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen. TNO 2020 R11052.

TNO (2020b). Advies vaststellingsbesluit Groningen gasveld 2020/2021, AGE 20-10.043, 11 mei 2020.

TNO (2020c). TNO Model Chain Groningen: Update and quick scan comparison of 2020 HRA model, TNO 2020 R11659, 6 november 2020.

TNO (2020d). Status of the TNO Model Chain Groningen per October 1, 2020 and recommendations for the public Seismic Hazard and Risk Analysis 2021, TNO 2020 R11464, 1 October 2020.

TNO (2021). Publieke Seismische Dreigings- en Risicoanalyse Groningen gasveld 2021, TNO 2021 R10441, 24 maart 2021.

5 Signature

Utrecht, 15-10-2021

TNO

J.A.J. Zegwaard
Head of department

A Seismic Source Model calibration

This appendix describes two aspects of the calibrations of the SSM:

- 1 Calibration as intrinsic part of the public SHRA
- 2 Technical description of the calibration procedures

At the request of EZK this documentation is submitted for external review. The review response is attached to this report.

A.1 Calibration as intrinsic part of the public SHRA

TNO (2020a) argued that the calibration of the Seismic Source Model (SSM) – a crucial step in the public Seismic Hazard and Risk Analysis (SHRA) – should be performed in the public domain. Here, we support this claim with four arguments:

- 1 **Public domain.** Source model calibration has previously not been embedded in the scope of a public SHRA (EZK, 2021). Source model calibration by NAM is unacceptable in the context of a public SHRA. Source model calibration by TNO maximizes the amount of work done in the public domain and reduces the dependency of the SHRA result on external inputs.
- 2 **Traceability.** Source model calibration by NAM is still largely a “black box”. Based on the documentation provided by NAM (2015a, 2015b) and several “question/reply” contacts with NAM, TNO has not been able to reproduce or replicate the calibration results of NAM. Source model calibration by TNO increases traceability and reproducibility.
- 3 **Quality control.** Source model calibration should be performed within the same quality control framework that applies to the entire TNO Model Chain Groningen. The calibration process should be reviewed and approved by an external authority. To our knowledge, the NAM calibration implementation and procedure has not been reviewed by any external party. The implementations of the activity rate model and the magnitude model in the TNO Model Chain Groningen (TNO, 2020b) have been evaluated by KEM (2020). For quality control, the numerical code of the entire TNO Model Chain Groningen – including calibration of the SSM – has been reviewed by Tessella (2020). The QAQC framework in which the TNO Model Chain is developed and maintained is ISO 9001 compliant and has been independently audited.
- 4 **Transparency.** Source model calibration has a significant effect on the risk outcome. The sensitivity study of TNO (2021, Figure 7.2) showed that “SSM calibration by TNO” resulted in a significantly different risk estimate (quantified as the Exposure Database average of mean LPR) than “SSM calibration by NAM”. The lack of transparency is undesirable in the context of safety perception by society.

A.2 Technical description of the calibration procedures

Both the theory and the step-by-step implementation of calibration of SSM are reported extensively in the report by TNO (2020b). These are repeated here for convenience. In addition, for the SSM-NAM-V6 stress-dependent exponential taper model, the calibration procedure is updated.

A.2.1 SSM-NAM-V5 calibration theory

The theory behind the Bayesian framework which is employed for the calibration of the source model follows below (section 2.2.1.4 page 22-23 in TNO report R11052 (TNO, 2020b):

The success of the seismicity rate model depends to a large degree on the various parameter settings. The parameters can be calibrated on the observed seismic data using a hindcast based on the historic production scenario. The hindcasted seismicity rate, with aftershock rate distributions conditional on the actual events is shown in equation (31). Extended with the magnitude distribution (30), with PDF f_m this gives:

$$\lambda_{xM}^{\text{obs}}(x, t, m, \gamma, \theta, \psi, \zeta, \ell_s, \mathcal{S}) = \lambda_x^{\text{obs}}(x, t, \gamma, \theta, \zeta, \mathcal{S}) f_m(m|c(x, t), \psi, \ell_s). \quad (1)$$

The total expected number of events over the observation periods is found by a temporal and spatial integral:

$$\Lambda^{\text{obs}}(\gamma, \theta, \zeta, \mathcal{S}) = \iint_{x,t} \lambda_x^{\text{obs}}(x, t, \gamma, \theta, \zeta, \mathcal{S}) dt dx. \quad (2)$$

The combination of (48) and (49) gives a probability distribution in space, time and magnitude for all events:

$$f_{xTM}(x, t, m|\gamma, \theta, \psi, \zeta, \ell_s, \mathcal{S}) = \frac{\lambda_x^{\text{obs}}(x, t, \gamma, \theta, \zeta, \mathcal{S})}{\Lambda^{\text{obs}}(\gamma, \theta, \zeta, \mathcal{S})} f_m(m|c(x, t), \psi, \ell_s). \quad (3)$$

In the context of parameter estimation the probability distribution is a likelihood function that can be applied to all observed earthquakes. Because of the normalization, the likelihood function above is not sensitive to the event count. However, a complementary likelihood expression for the number of observed earthquakes is found in (20):

$$p_N(n^{\text{obs}}|\Lambda^{\text{obs}}(\gamma, \theta, \zeta, \mathcal{S})). \quad (4)$$

The total likelihood is a product of (51) and n^{obs} evaluations of (47), one for every earthquake in the catalogue.

According to the Bayesian approach to parameter estimation, the posterior probability distribution for the parameters is obtained by a multiplication of the prior probability distribution and the likelihood.

A.2.2 SSM-NAM-V5 implementation of calibration

The implementation of this theory is described here (section 3.1.3 page 30-33 in TNO report R11052 (TNO, 2020b):

During the training phase, a Bayesian framework is applied to assign a likelihood score to each set of model parameters. Since during training, the activity rate model and the magnitude model are independent of each other (see also Section 2.2.1.2), but both models rely on the DSM, two independent posterior likelihood defined: first,

$LL_{AR}(\gamma, \theta, \zeta)$: the log-likelihood function depending on a combination of DSM covariate conditioning parameters (γ), main-shock activity rate parameters (θ) and ETAS clustering model parameters (ζ), and second $LL_M(\gamma, \psi)$, the log-likelihood function depending on a combination of DSM covariate conditioning parameters and magnitude parameters (ψ).

For any activity rate model (a model describing the number of events per unit time, independent of magnitude), the log-likelihood is given by:

$$LL_{AR}(\gamma, \theta, \zeta) = - \int_t \int_S \lambda_X(x, t) dSdt + \sum_{i=1}^n \log(\lambda_X(x_i, t_i))$$

where $\lambda_X(x, t)$ is the spatio-temporal event rate density (units: number of events per unit time per unit area, e.g. $m^{-2}year^{-1}$), n is the number of observed events in the time period under consideration and $\lambda_X(x_i, t_i)$ is the event rate density at the time-space location of an observed event.

For training we use the observation-conditioned total seismicity rate density λ_X^{obs} of Equation (31). The ETAS model functions $f_T(t)$ and $f_R(r)$ are the probability density function for temporal and spatial triggering defined as:

$$f_T(t) = \frac{p-1}{c} \left(\frac{t}{c} + 1 \right)^{-p},$$

$$f_R(r) = \frac{q-1}{\pi d} \left(\frac{r^2}{d} + 1 \right)^{-q},$$

where c, p respectively are the characteristic time and temporal power-law exponent parameters, defining the speed at which the aftershock rate decays over time. Also, d, q respectively are the characteristic area and spatial power-law exponent parameters of the ETAS model, defining the speed at which the aftershock rate decays spatially.

For any b-value model (a model describing the slope of the Gutenberg-Richter frequency magnitude distribution), the log-likelihood is given by:

$$LL_M(\gamma, \psi) = \sum_{i=1}^n \log[b(x_i, t_i) \log(10)] - \sum_{i=1}^n b(x_i, t_i) \log(10) (m_i - m_0), \quad (10)$$

where $b(x_i, t_i)$ is the b-value at the space-time location of the i^{th} event and m_i is the magnitude of the i^{th} event.

Box 1 outlines the numerical procedures followed in the implementation of the SSM V5 DSM. Box 2 outlines the training of the activity rate model and the b-value model.

BOX 1 IMPLEMENTATION OF SSM V5 DYNAMIC SUBSURFACE MODEL

DSM: Calculating smoothed incremental Coulomb stress change from input files (pore pressure change, reservoir thickness, reservoir compressibility, fault geometry) and model parameters $\gamma = \{r_{max}, \sigma\}$.

1. Obtain the topographic gradient $\Gamma(\mathbf{x})$ and fault density $\rho(\mathbf{x})$ on the base grid. To do so, only the points in the fault data which have a throw/thickness ratio $r \leq r_{max}$ are considered.
 - a. These points are assigned to the nearest base grid point, weighted by the fault area $A = l_{repr} t_{avg}$ to obtain the fault density grid $\rho(\mathbf{x})$.
 - b. These same points are assigned to the nearest base grid point, weighted by $\text{offset} \times l_{repr} t_{avg}$ to obtain the grid $\Gamma\rho(\mathbf{x})$.
 - c. The topographic gradient is obtained by $\Gamma(\mathbf{x}) = \frac{\Gamma\rho(\mathbf{x})}{\rho(\mathbf{x})}$.
2. Obtain the elastic modulus grid $H(\mathbf{x})$ by:

$$H(\mathbf{x}) = (H_s^{-1} + C_m(\mathbf{x}))^{-1},$$
 where $H_s = 10^{-5.3}$ and $C_m(\mathbf{x})$ is the reservoir compressibility grid.
3. Calculate the scalar value $\gamma = \frac{1-2\nu}{2-2\nu}$, where $\nu = 0.2$ is the Poisson ratio. Note that this results in a scalar factor on the incremental Coulomb stress change and therefore does not impact the seismicity forecast after model training. It is included for completeness only.
4. Calculate the vertical strain grid $\epsilon_{zz}(\mathbf{x}, t) = dP(\mathbf{x}, t)C_m(\mathbf{x})$, where $dP(\mathbf{x}, t)$ is the pore pressure change grid.
5. Calculate the (spatio-temporal) incremental Coulomb stress change $\Delta C(\mathbf{x}, t)$:

$$\Delta C(\mathbf{x}, t) = \gamma H(\mathbf{x})\epsilon_{zz}(\mathbf{x}, t)\Gamma(\mathbf{x}).$$
6. Set any negative and NaN values in $\Delta C(\mathbf{x}, t)$ to zero.
7. Obtain the smoothed incremental Coulomb stress change and smoothed fault density by applying a Gaussian kernel with characteristic length scale σ to the spatial (\mathbf{x}) dimensions of the $\Delta C(\mathbf{x}, t)$ grid and the $\rho(\mathbf{x})$ grid. This is implemented using [scipy.ndimage.gaussian filter](#) with $\text{sigma} = \frac{\sigma}{dx}$, where dx is the grid spacing of the $\Delta C(\mathbf{x}, t)$ and $\rho(\mathbf{x})$ grid and $\text{mode} = \text{"constant"}$.

BOX 2 IMPLEMENTATION OF SSM V5 MODEL TRAINING

SRM activity rate V5 training: Obtain log-likelihood for parameter vectors $\gamma = \{r_{max}, \sigma\}$, and $\theta = \{\theta_0, \theta_1\}$, and $\zeta = \{K, a\}$, and covariate ΔC

1. Obtain the smoothed incremental Coulomb stress change $\Delta C(\mathbf{x}, t)$ as described in Box 1, using parameters $\{r_{max}, \sigma\}$.
2. Obtain $\Delta C(\mathbf{x}, t_{start})$, $\Delta C(\mathbf{x}, t_{end})$, $\Delta C(\mathbf{x}_i, t_i)$, $\dot{\Delta C}(\mathbf{x}_i, t_i)$ and $\rho(\mathbf{x}_i)$ through spatial nearest neighbor interpolation and cubic spline temporal interpolation ([scipy.interpolate.CubicSpline](#), bc = "natural"). $\dot{\Delta C}(\mathbf{x}_i, t_i)$ is the time-derivative of $\Delta C(\mathbf{x}_i, t_i)$ and is obtained by using the [scipy.interpolate.CubicSpline](#) functionality nu = 1.
3. Obtain $\int_t \int_S \lambda(\mathbf{x}, t) dS dt$ numerically: $A = \Delta S \sum_x e^{\theta_0} \rho(\mathbf{x}) (e^{\theta_1 \Delta C(\mathbf{x}, t_{end})} - e^{\theta_1 \Delta C(\mathbf{x}, t_{start})})$, with ΔS the surface area of a grid cell.
4. Obtain $K \sum_{i=1}^n e^{a(M_i - M_0)}$ numerically: $B = K \sum_i [e^{a(M_i - M_0)}]$.
5. Obtain $\lambda(\mathbf{x}_i, t_i)$ numerically: $C_i = \rho(\mathbf{x}_i) \theta_1 \dot{\Delta C}(\mathbf{x}_i, t_i) e^{\theta_0 + \theta_1 \Delta C(\mathbf{x}_i, t_i)}$.
6. Obtain $\sum_{j=1}^{i-1} K e^{a(M_j - M_{min})} \left(\frac{p-1}{c} \left(\frac{t_i - t_j}{c} + 1\right)^{-p}\right) \left(\frac{q-1}{\pi d} \left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{d} + 1\right)^{-q}\right)$ numerically:

$$D_i = \frac{p-1}{c} \frac{q-1}{\pi d} K \sum_j \left[\left(\frac{1 + \delta t_{ij}}{c}\right)^{-p} \left(\frac{1 + \delta r_{ij}}{d}\right)^{-q} e^{a(M_{ij})} \right],$$

where $p = 1.35$, $q = 3.16$, $d = 4 \times 10^6 \text{ m}^2$, $c = 0.3 \text{ days}$. δt_{ij} , δr_{ij} , and M_{ij} are lower triangle matrices of inter-event time, inter-event distance and normalized event magnitude ($M - M_0$). E.g.:

$$\delta t_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ t_{12} & \ddots & 0 & 0 \\ \vdots & t_{ij} & \ddots & 0 \\ t_{1n} & t_{2n} & \dots & 0 \end{pmatrix}.$$

7. Obtain the log-likelihood numerically: $LL_{AR}(\gamma, \theta, \zeta) = -A - B + \sum_i \log(C_i + D_i)$.

SRM b-value model V5 training: Obtain log-likelihood for parameter vectors $\gamma = \{r_{max}, \sigma\}$ and $\psi = \{\beta_0, C_0, n\}$.

1. Obtain the smoothed incremental Coulomb stress change $\Delta C(\mathbf{x}, t)$ as described in Box 1, using parameters $\{r_{max}, \sigma\}$.
2. Obtain $\Delta C(\mathbf{x}_i, t_i)$ through spatial nearest neighbor interpolation and cubic spline temporal interpolation ([scipy.interpolate.CubicSpline](#), bc = "natural").
3. Obtain $b(\mathbf{x}_i, t_i)$ numerically: $b_i = \beta_0 \left(\frac{\Delta C(\mathbf{x}_i, t_i)}{C_0}\right)^{-n}$.
4. Obtain $b_i^* = \log(10) b_i$.
5. Obtain the log-likelihood numerically:

$$LL_M(\gamma, \psi) = \sum_i \log[b_i^*] - \sum_i [b_i^* \times (m_i - m_0)].$$

N.B.: In order to calculate the log-likelihood of a set of model parameters, the forward model does not have to be evaluated. The only information about the forward model that is required is:

- The number of events in the forward model (i.e. $\int_t \int_S \lambda_X(\mathbf{x}, t) dS dt$).
- The smoothed incremental Coulomb stress rate at the space-time location of the observed events.
- The time-derivative of the smoothed incremental Coulomb stress rate at the space-time location of the observed events.

Note that in the calibration procedure described above, the Coulomb Stress parameters are also allowed to be estimated during the calibration procedure (i.e. they are free parameters, rather than 'optimal' values).

In the V5 implementation of TNO, H_s is fixed at $10^{-5.3}$ while $\{\sigma, r_{\max}, \theta_0, \theta_1, K, a\}$ are calibrated simultaneously for the activity rate model, and $\{\sigma, r_{\max}, \beta_0, C_1, n\}$ are calibrated simultaneously for the V5 magnitude model. These calibrations can be combined to obtain the full $\{\sigma, r_{\max}, \theta_0, \theta_1, K, a, \beta_0, C_1, n\}$ probability distribution (conditional on $H_s = 10^{-5.3}$).

A.2.3 Updated calibration procedure for the SSM-NAM-V6

For the SSM-v6 stress-dependent exponential taper model, the loglikelihood function for the Magnitude-Frequency-Distribution had to be updated to allow for non-zero values of zeta (ζ):

Equation 10 (and identically, the equation on step 5 in Implementation Box 2 - SRM *b-value model V5 training*), changes from:

$$LL(\gamma, \psi) = \sum_{i=1}^n \log [b(x_i, t_i) \log(10)] - \sum_{i=1}^n b(x_i, t_i) \log(10)(m_i - m_0)$$

to

$$LL(\gamma, \psi) = \sum_{i=1}^n \left[\ln \left(b + \frac{3}{2} \zeta(x_i, t_i) \times 10^{\frac{3}{2}(m_i - M_{\min})} \right) + \ln(\ln(10)) \right. \\ \left. - b \ln(10)(m_i - M_{\min}) - \zeta(x_i, t_i) \times (10^{\frac{3}{2}(m_i - M_{\min})} - 1) \right]$$

where γ is the parameter vector $\{\theta_0, \theta_1, \theta_2\}$, and ψ is the parameter vector $\{r_{\max}, \sigma\}$ which controls the incremental Coulomb Stress $\Delta C(x_i, t_i)$. From γ , the values of b and $\zeta(x_i, t_i)$ are calculated.

This updates the equation on step 3 in Implementation Box 2 - SRM *b-value model V5 training*:

$$b = \theta_0$$

$$\zeta(x_i, t_i) = \theta_1 e^{-\theta_2 \Delta C(x_i, t_i)}$$

In contrast to V5, in the V6 implementation of TNO, H_s is left free, such that $\{H_s, \sigma, r_{\max}, \theta_0, \theta_1, K, a\}$ are calibrated simultaneously for the activity rate model, and $\{H_s, \sigma, r_{\max}, \theta_0^m, \theta_1^m, \theta_2^m\}$ are calibrated simultaneously for both V6 magnitude models (note that θ_x refers to activity rate parameters while θ_x^m refers to magnitude model parameters). These calibrations can be combined to obtain the full $\{\sigma, r_{\max}, \theta_0, \theta_1, K, a, \theta_0^m, \theta_1^m, \theta_2^m\}$ probability distribution for each magnitude model.

A.3 References

EZK (2021). Verwachtingenbrief aan de Nederlandse Aardolie Maatschappij. DGKE-PDG/20330147

KEM (2020). KEM-10 research review, evaluation and interpretation of “Public seismic hazard and risk assessment model Groningen 2018-2019”.

NAM (2015a). An activity rate model of induced seismicity within the Groningen Field (Part 1). <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/8b6f2ff1-b98e-4148-a1db-bf06881579e5>

NAM (2015b). An activity rate model of induced seismicity within the Groningen Field (Part 2). <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/c906565b-6b54-4768-874b-b23b46b1ee5e>

NAM (2020). Operationele Strategieën voor het Gasjaar 2020-2021, including Appendix A: Seismic Hazard and Risk Assessment Groningen Field update for Production Profile GTS 2020, EP202002207545, March 13, 2020 (update April 9, 2020).

Tessella (2020). Software review report TNO PSHRA Software Assessment SodM.

TNO (2020a). Status of the TNO Model Chain Groningen per October 1, 2020 and recommendations for the public Seismic Hazard and Risk Analysis 2021. TNO2020_R11464

TNO (2020b). Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen, TNO2020_R11052

B On Bayesian inference of the exponentially tapered magnitude distribution

At the request of EZK this documentation is submitted for external review. The review response is attached to this report.

B.1 Introduction

For the induced seismicity in Groningen (and for seismicity in many other places worldwide), the bulk of the seismic risk is associated with earthquake magnitudes larger than the largest one observed so far (Huizinge, 2012, $M=3.6$). Hazard and risk assessment therefore relies heavily on an extrapolation of the magnitude distribution (MD) beyond the observed catalogue. The bulk of the MD is usually – also in Groningen - successfully described by the Gutenberg-Richter (G-R) relation, which is basically an exponential distribution of magnitudes, equivalent to a Pareto distribution for seismic moments. However, it is also widely recognized (Main, 1995, Kagan and Schoenberg, 2001) that the G-R scaling relation cannot be extrapolated to higher magnitudes indefinitely, because it should break down at some point, if only due to the finite dimensions of the seismogenic zone. Honoring this expected breakdown in hazard and risk assessment necessitates the use of a dedicated description of the high-end tail of magnitude distribution. This description can take a variety of forms, including a hard cutoff, represented by a maximum magnitude parameter M_{max} , or a smoother, exponentially tapered variation parameterized by a corner magnitude M_c (e.g., Kagan and Schoenberg, 2001).

In any case, the tail of the distribution is, almost by definition, but most certainly in practice, poorly sampled by observations, and therefore difficult to impossible to constrain from the data. Also in the Groningen file this was recognized early on. It was for this reason that the M_{max} workshop for Groningen (NAM, 2016) was organized: the tail of the magnitude distribution, represented by of the parameter M_{max} , is an epistemic uncertainty that cannot be inferred reliably from the Groningen earthquake catalogue.

However, with the introduction of a tapered MD for Groningen (NAM, 2019; Bourne and Oates, 2020), the authors claim to be able to do just that: to infer the location of a corner magnitude describing an exponential taper at the tail of the MD.

In the following sections we will present the magnitude model equations, describe Bayesian inference in the context of its application in the Groningen source model definition and subsequent hazard and risk assessment. Next we focus on the central role of the posterior predictive magnitude distribution in hazard and risk forecasting. We identify the inference+forecasting procedure as a statistical estimator that we can evaluate in terms of its bias and dispersion. With all the technical tools in place we can then study the specific examples of the inference of magnitude models with and without a taper. Finally, on the basis of these examples we can draw conclusions.

B.2 Magnitude distribution model with and without taper

The magnitude model (Bourne and Oates, 2020) used by NAM in the HRA2020 (NAM, 2020) can be written as:

$$P_M(m|b, \zeta) = 10^{-b(m-M_{\min})} \exp(-\zeta (10^{\frac{3}{2}(m-M_{\min})} - 1))$$

with $P_M(m|b, \zeta)$ the probability of exceedance (survival function of the MD) conditional on the Gutenberg-Richter b -value b , and exponential taper parameter ζ . M_{\min} is the minimum magnitude considered, commonly the lowest magnitude for which the catalogue of observed magnitudes is considered complete. The functional form is a bit different from Bourne and Oates (2020) because we express the distribution in magnitudes rather than in moments.

A ζ parameter setting of 0 yields the classical Gutenberg-Richter relation:

$$P_M(m|b, \zeta = 0) = 10^{-b(m-M_{\min})}$$

The ζ parameter is equal to the inverse of the “corner moment” (Bourne and Oates, 2020). Its interpretation is probably easier when it is expressed in terms of the corresponding *corner magnitude* M_c :

$$M_c = M_{\min} - \frac{2}{3} \log_{10} \zeta$$

and its inverse:

$$\zeta = 10^{\frac{3}{2}(M_{\min} - M_c)}$$

For any magnitude distribution with a generic parameter set θ the probability density function (PDF) $p_M(m|\theta)$ is related to the exceedance probability as :

$$p_M(m|\theta) = -\frac{dP(m|\theta)}{dm}$$

B.3 Bayesian inference of seismic source models

The seismic source model for Groningen hazard and risk assessment is a probabilistic model for the seismic activity as a function of time, space, and magnitude (TNO, 2020). The model consists of various components, among which the seismic rate model, which describes the spatiotemporal distribution of seismic event rate (number of earthquakes above a minimum magnitude M_{\min} per unit time and space), and the magnitude model, which describes the spatiotemporal variation of the distribution of the earthquake magnitudes. In this Appendix we focus on the magnitude distribution in isolation, which, within structure of current models, can be calibrated independently of the seismicity rate.

The Groningen seismic source models in general and the magnitude distribution in particular are calibrated using Bayesian model parameter inference. Bayesian inference makes use of Bayes' rule to update prior information on a set of model parameters by the new evidence supplied by a set of observations.

Let $p_M(m|\theta)$ be the PDF of the magnitude distribution for magnitude m , conditional on the model parameters θ . Given a set of observations of m , incorporated in the earthquake catalogue $D^{\text{obs}} = \{m_i^{\text{obs}}, \dots, m_N^{\text{obs}}\}$, the *likelihood* $L(\theta|D^{\text{obs}})$ of a parameters θ conditional on the observations is defined as the product of the probability density values of all observed magnitudes:

$$L(\theta|D^{\text{obs}}) = \prod_i^N p_M(m_i|\theta)$$

This expression for likelihood is the basis for finding the values of θ that maximize likelihood, i.e., the *maximum likelihood estimate* (MLE) $\hat{\theta}(D^{\text{obs}}) = \text{argmax} L(\theta|D^{\text{obs}})$. However, in the context of Groningen we aim to account for uncertainties as much as possible, including the uncertainties in the source model parameters. It is clear that in some neighborhood of $\hat{\theta}$ the likelihood of the model parameters is substantial. Also, the likelihood function may have secondary highs, corresponding to a multimodal distribution. In principle, in a fully probabilistic fashion, all possible values of θ must be taken into account, proportional to their associated likelihood. However, apart from the weighting induced by the likelihood, a secondary weighting is importance: a (probabilistic) measure on the space of θ . Such a measure determines the relative contributions of different values of θ in integrations over θ ¹. The Bayesian inference procedure provides exactly that.

Let $p_\theta(\theta|\pi)$ be the PDF of the *prior distribution* of θ , where π represents the essentially subjective prior information (or prior belief, prior assumptions) on θ that are taken into account² by the model developer or practitioner. The Bayesian inference procedure for the model parameters θ now defines the PDF of the *posterior distribution* $p_\theta(\theta|D^{\text{obs}}, \pi)$ as the normalized product of the likelihood and the prior:

$$p_\theta(\theta|D^{\text{obs}}, \pi) = \frac{L(\theta|D^{\text{obs}})p_\theta(\theta|\pi)}{\int L(\theta|D^{\text{obs}})p_\theta(\theta|\pi)d\theta}$$

From this definition it is clear that the posterior distribution $p_\theta(\theta|D^{\text{obs}}, \pi)$ is conditioned both on the observed data D^{obs} through the likelihood and the prior information π through the prior distribution.

¹ Here it is informative to realize that the likelihood by itself is not a PDF: choosing a different parameterization of the model, by a coordinate transformation of θ , does not change the value of the likelihood at a specific point, but it does change the shape of the likelihood function around it. In a PDF the shape and the values would change hand in hand to maintain an integral measure of probability.

² Usually the explicit reference to prior information (π) is left out, and the unconditional $p(\theta)$ is used to represent the prior distribution. Here we choose to make the conditioning role of the prior explicit to stress its importance in Bayesian inference.

We can interpret the Bayesian inference equation in two ways. On the one hand, the likelihood function acts as a modifier on a prior distribution to accommodate the new information obtained by the observations and to derive a posterior distribution. On the other hand, the prior distribution provides a probabilistic measure on the θ parameter space such that a likelihood function derived from a set of observations can be translated to a proper (posterior) probability distribution. The latter interpretation gives rise to the concept of the *maximum a posteriori* (MAP) estimate $\check{\theta}(D^{\text{obs}}, \pi) = \operatorname{argmax}_{\theta} p_{\theta}(\theta|D^{\text{obs}}, \pi)$, which – in contrast to the maximum likelihood estimate – depends on the prior information.

It should be noted that the choice of a uniform/constant prior in some suitable domain bounds is not an uninformative, objective, choice. Any choice of prior distribution introduces a degree of subjectivity. For example, a homogeneous distribution in one parameterization of θ corresponds to an inhomogeneous distribution in another parameterization, while the model performance does not depend on the specific parameterization. In fact, the combination of the choice of model parameterization and the choice of prior parameter distribution constitute the prior information imposed by either the model developer or practitioner. We acknowledge the importance of these choices with the parameter π .

In general, the prior information on the model parameters is relatively broad and smooth, and we expect the data to provide more detailed information through the likelihood function. The higher the number of data points, the more precisely defined the likelihood will be, and the smaller the influence of the prior. On the other side, the lower the number of data points, the more important the role of the prior will be. This matches our intuition: if we believe we have a fair 6-sided dice, and we roll a 4, there is little reason to update our belief that the dice is fair. However, if we roll it a thousand times, and we roll a 4 over 50% of the time, we would start to move away from our prior belief, and assume that the dice is somehow more primed to yield 4's compared to the other values. Bayesian inference provides a formal framework to capture this intuition.

What this means is that, especially in the case of a small dataset, the role of prior is important, and should not be underestimated. This will become apparent later in this document.

B.4 Forecasting using the posterior predictive magnitude distribution

Now that we have a full posterior distribution of the model parameters we can use it for a probabilistic forecast. This involves the *posterior predictive magnitude distribution* (PPMD), which is obtained by marginalizing the parameter dependence of the magnitude model over the posterior:

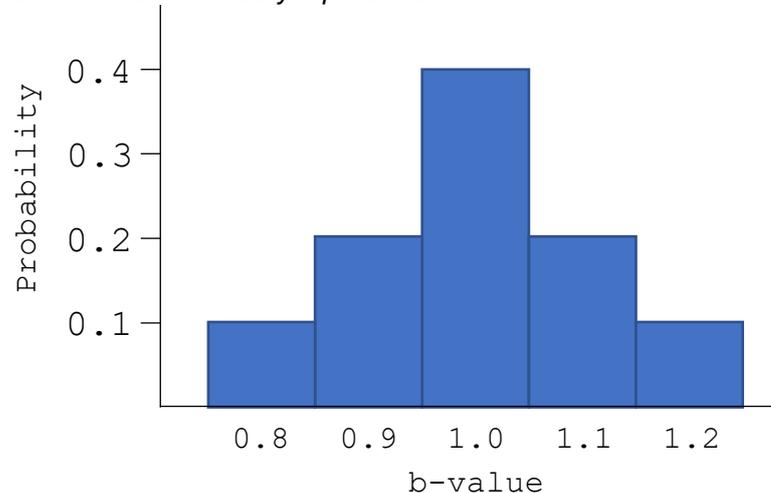
$$P_M(m|D^{\text{obs}}, \pi) = \int_{\theta} P_M(m|\theta) p_{\theta}(\theta|D^{\text{obs}}, \pi) d\theta$$

The PPMD is the expected magnitude distribution conditional on both the observations and the prior information. It is effectively this magnitude distribution that is used in hazard and risk assessment. As long as the ground motion and

fragility models do not explicitly depend on the source model parameters θ , which is unlikely and – at least in the current and previous generations of Groningen models – not the case, the marginalization/integration over θ can take place in an early stage, just after the source model inference, but before the evaluation of ground motion and building fragility effects.

Note that both TNO's grid-based integration approach and NAM's Monte Carlo based integration approach to hazard and risk assessment effectively make use of the PPDM. Since this appears to have caused significant confusion in the past, this is explained with a small example in the box below.

We consider a Gutenberg-Richter magnitude model parameterized with a single parameter, the “b-value” b , representing the slope of the magnitude distribution. For simplicity we discretize the parameter space coarsely in steps of 0.1. The result of a Bayesian inference is displayed in the figure below. If we want to obtain a forecast of the magnitude distribution, we do not simply evaluate a model with $b = 1.0$ (which in this particular case is both the MAP and the mean b-value). Instead, we evaluate a model with $b = 0.8$, and let this model contribute to the PPMD for 10%, we then evaluate a model with $b = 0.9$, and let it contribute to the PPMD for 20%. Similarly, each other model gets evaluated, and contributes to the PPMD according to its weight. In a Monte Carlo approach, the $b = 1.0$ model would on average get evaluated four times as often as the $b = 0.8$ model, while each evaluation then contributes equally to the PPMD and subsequent hazard and risk results. These approaches are equally valid and mathematically equivalent.



Another way to look at the PPMD is that it represents for each value of m a (point) estimate of the exceedance probability, conditional on the observations and the prior. As such, the procedure followed so far is a statistical point estimation process, embedded in the Bayesian inference framework. For the current context we will refer to this estimator as the *PPMD estimator*. The performance and reliability of an estimator can be evaluated by studying its bias and dispersion. Favorable properties for accuracy are low bias and low dispersion. However, we should note

that in the Bayesian parameter inference context a certain bias is to be expected and not necessarily problematic or undesired. In fact, the inclusion of the prior information in the inference actually amounts to a bias by design. As long as the observations/data do not sufficiently constrain the model parameters, the final result is influenced or even dominated by the chosen prior.

In the probabilistic framework, the observations D are considered as “just” a sample from the distribution that is to be inferred. It is clear that for other samples from the same distribution the estimator would or will provide different results. The performance of the estimator can be characterized by the distribution of the results for many samples. This can be achieved in a synthetic experiment where the “ground truth” distribution is known.

Let θ_g be the set of ground truth model parameters, and let $D = \{m_1, \dots, m_N\}$ be a synthetic catalogue of N independent samples from the ground truth magnitude distribution $P_M(m|\theta_g)$. Then the mean/expected PPMD, P_M^E , conditional on the ground truth, catalogue size N and prior information, is formally obtained by an N -dimensional marginalization integral:

$$P_M^E(m|\theta_g, N, \pi) = \int_{m_1} \dots \int_{m_N} P_M(m|D, \pi) \prod_i^N p_M(m_i|\theta_g) dm_1 \dots dm_N$$

In practice, this many-dimensional integral can be evaluated with a Monte Carlo sampling procedure, drawing a large number of synthetic catalogues from the ground truth.

The expected PPMD can be used to calculate the magnitude dependent bias of the PPMD estimator:

$$\text{bias}(m|\theta_g, N, \pi) = P_M^E(m|\theta_g, N, \pi) - P_M(m|\theta_g)$$

or its relative bias:

$$\text{rbias}(m|\theta_g, N, \pi) = \frac{P_M^E(m|\theta_g, N, \pi)}{P_M(m|\theta_g)}$$

Using the same Monte Carlo procedure also the magnitude-dependent dispersion of the PPMD estimator can be characterized, e.g., through the 5%-95% confidence bounds. In the following examples the theory described above is elaborated for magnitude models with and without an exponential taper.

B.5 Example: Bayesian inference of a magnitude model without taper

First, let us consider the inference of a magnitude distribution without a taper: a simple Gutenberg-Richter MD, with $M_{\min}=1.5$. We'll use this also as the ground truth model, with a b-value $b = 1.0$. As prior information we assign equal probability to all

b-values in the range $[0.4, 1.6]$, i.e., a truncated homogeneous distribution, which we will refer to as π^b .

We draw a synthetic catalogue ($N = 300$) from the ground truth MD, infer a posterior model parameter distribution and then integrate over this posterior distribution to obtain the PPMD. The results are shown in Figure 1.

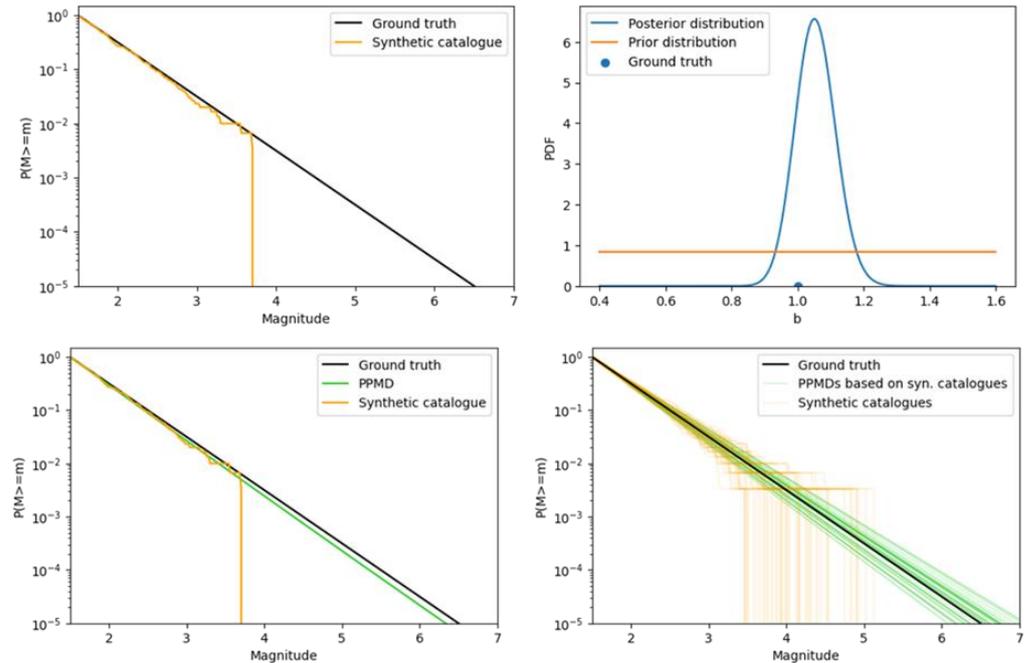


Figure 1: Model inference for a Gutenberg-Richter magnitude distribution, on a synthetic dataset of 300 samples from a ground truth with $b = 1.0$. Top left: Exceedance probability distribution (survival function) of ground truth and synthetic data. Top right: Bayesian inference of b-value. Bottom-left: Same as top left but including the posterior predictive model implied by the Bayesian inference. Bottom right: Experiment repeated for 50 catalogues, showing the distribution of the PPDM results.

The PPMD obtained from the synthetic catalogue (Figure 1, bottom left) lies a little below the ground truth. So in this particular instance, the result of the PPMD estimator happens to underestimate the probability of high magnitude events.

However, to characterize the performance of the PPMD estimator we're interested in its bias and dispersion. How do we expect it to perform in general, rather than on this specific synthetic catalogue?

To investigate this we draw 10,000 catalogues of 300 events each, and obtain a PPMD for each synthetic catalogue. We can then look at the distribution of PPMDs, which are all based on different realizations of data generated by the same ground truth. In the first 50 cycles, displayed in Figure 1, bottom right, the obtained PPMDs are sometimes below, and other times above the ground truth. When we look at all 10,000 realizations, the picture in Figure 2 emerges.

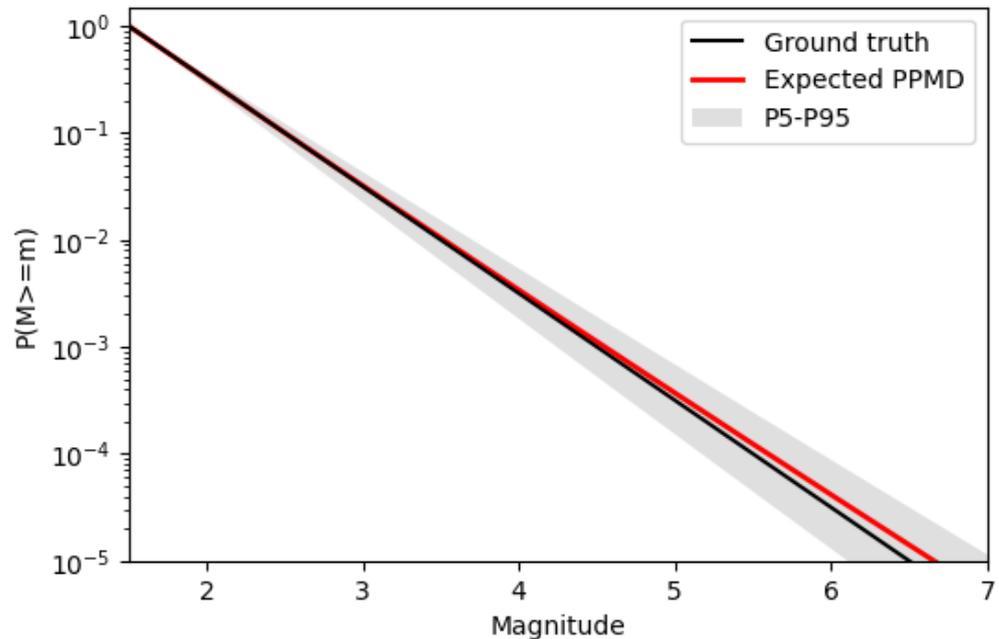


Figure 2: Characterization of the performance of the posterior predictive magnitude distribution (PPMD) estimator for a non-tapered magnitude distribution, conditional on a ground truth with $b = 1$, a sample size of 300, and a uniform prior in b . The bias of the estimator is the difference between the black and red lines. Due to the logarithmic scale it is to be interpreted as the relative bias. The grey band represents its dispersion by the 5%-95% percentile range.

In this Figure, the red line represents the expected PPMD $P_M^E(m|\theta_g \equiv b = 1, N = 300, \pi = \pi^b)$. It turns out that over the full magnitude range, we expect our estimator to slightly overestimate the probability of exceedance of a given magnitude. For example, in 57% of the cases, the PPDM overpredicts the probability exceeding M5.0, while 43% of the time, the PPDM underpredicts. The relative bias of exceeding M5.0 is 17%. The amount of expected overprediction will decrease with increasing catalogue size. In other words, we expect our Bayesian inference on a catalogue of 300 earthquakes to result in a slightly conservative forecast, and we expect to get closer to the ground truth with increasing catalogue size.

In pSHRA we ideally would like to use an MD that best represents the ground truth. When we don't have access to the ground truth (i.e. in every real world scenario), we have to rely on our inference procedure to represent the ground truth. In other words, the relation between the expected PPMD and its dispersion and the ground truth tells us whether we expect to have reliable pSHRA results.

B.6 Example: Bayesian inference of a magnitude model with taper

Next we study the performance of the estimation procedure on a magnitude distribution that does include a taper.

Again, we choose a prior, use a dataset to infer the posterior distribution, and integrate the posterior to get the posterior predictive magnitude distribution PPMD. We'll choose a uniform prior in b - ζ over the ranges b : [0.4; 1.6], ζ : [0; 0.003]³, which we will refer to as $\pi^{b\zeta}$. Note that this prior does include the possibility that there is no taper ($\zeta=0$), but it is considered just as likely (a priori) as $\zeta=0.001$, $\zeta=0.01373$, or any other value in the range up to 0.003.

We can see what happens if we use the same initial catalogue from the previous paragraph, when we only attempted to infer a b -value.

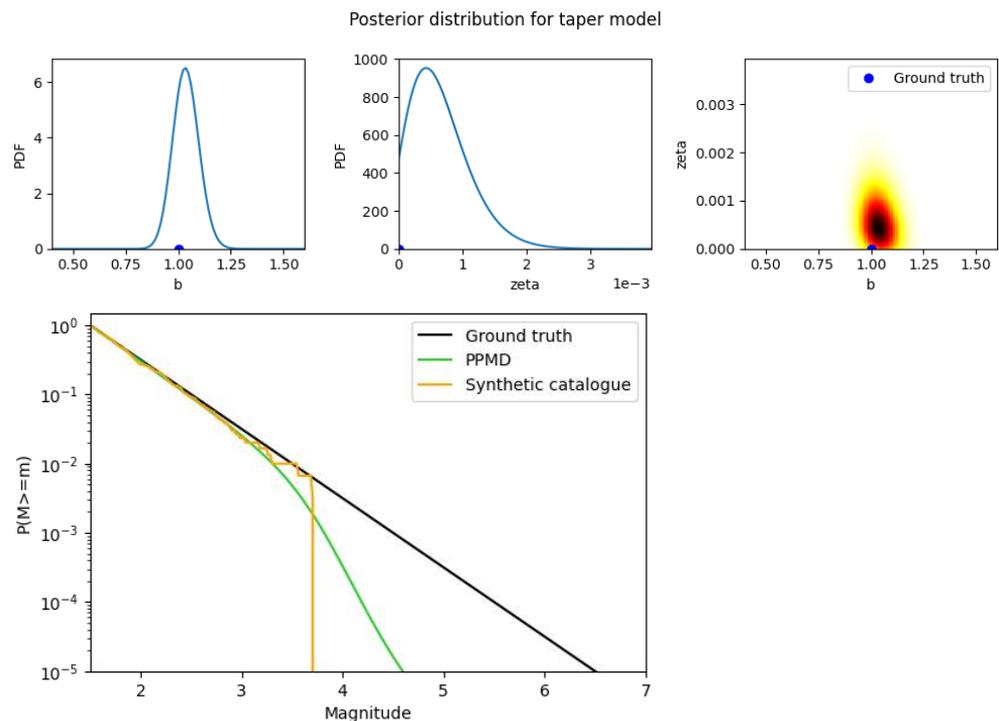


Figure 3: Posterior distribution (marginalized and joint) based on a uniform prior in b - ζ space (all points in joint space have equal prior probability). The resulting PPMD in relation to the synthetic catalogue and the ground truth is shown below.

This result is quite concerning. The posterior predictive magnitude distribution appears to fit the synthetic data quite well, but we know that this catalogue was actually obtained from a ground truth that did not contain a taper whatsoever. In this particular case, the probability of exceeding $M5.0$ is underpredicted by a factor of 147.4 (14740%). If we would not have known the ground truth, we would be completely unaware of this extreme mismatch, since the posterior distribution appears to be well-constrained and the posterior predictive magnitude distribution fits the data well.

Again, since we're interested in expected behavior, rather than anecdotal evidence from a single catalogue, we repeat this procedure many times, each time obtaining

³ We'll look at the effect of choosing another prior at a later stage.

and integration a posterior distribution⁴. The result, analogous to Figure 2, is displayed in Figure 4.

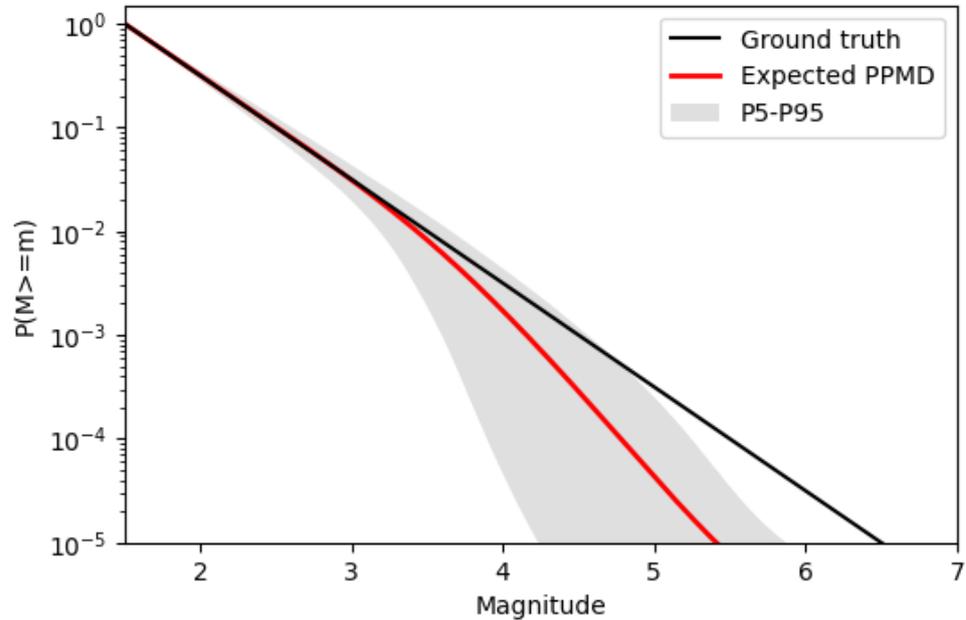


Figure 4: Characterization of the performance of the posterior magnitude distribution (PPMD) estimator for a tapered magnitude distribution conditional on a ground truth with $b = 1$, $\zeta = 0$ (i.e., $M_c = \infty$), a sample size of 300, and a uniform prior in both b and ζ . The bias of the estimator is the difference between the black and red lines. The grey band represents its dispersion by the 5%-95% percentile range.

In this Figure, the red line represents the expected PPMD estimator for a tapered magnitude distribution conditional on a ground truth with $b = 1$, $\zeta = 0$ (i.e., $M_c = \infty$), a sample size of 300, and a uniform prior in both b and ζ . It turns out that the behavior we observed for the single catalogue was no accident, but rather expected behavior for this parameter estimation procedure, when using this prior faced with such a relatively small dataset. In 97% of the cases, the posterior predictive magnitude distribution overpredicts the probability exceeding M5.0, while 3% of the time, the posterior predictive magnitude distribution underpredicts. The expected overprediction of exceeding M5.0 is a factor of 7.36. In comparison with the estimator used for a b-value only, characterized in Figure 2, both the relative bias and the dispersion are much larger.

This inference method, which appeared to work so well for estimating b-values, seems to break down when attempting to constrain a taper location.

Let's start by getting one thing out of the way: Bayesian inference is not 'broken'. The posterior distributions that we obtain are entirely valid. However, we should not forget what the posterior distribution is, namely the combination of the prior distribution, which then gets modulated with information obtained from the observations. The less data you have, the more your posterior distribution will

⁴ In this loop, we choose a more flexible prior range (always uniform) for zeta, to ensure the tail at higher zeta values is not truncated when a catalogue with a particularly low maximum observed magnitude is generated. This way, the posterior distribution of each individual catalogue is not influence/truncated by the choice of a tool limited prior.

resemble the prior. As will become clear in the following, the prior embedded in the estimator characterized in Figure 2, with a uniform prior in ζ , insists quite heavily on the presence of a taper, while the data can't rule it out.

So let's subsequently look at these two factors: the prior distribution and the data.

B.6.1 The role of the prior distribution

In the previous section, we chose a uniform prior in ζ : $[0; 0.003]$. This means we place equal a priori probability density on any value in this range. When we convert this PDF to an equivalent PDF in terms of corner magnitude, we see that this effectively means we have a strong a priori preference for low corner magnitudes.

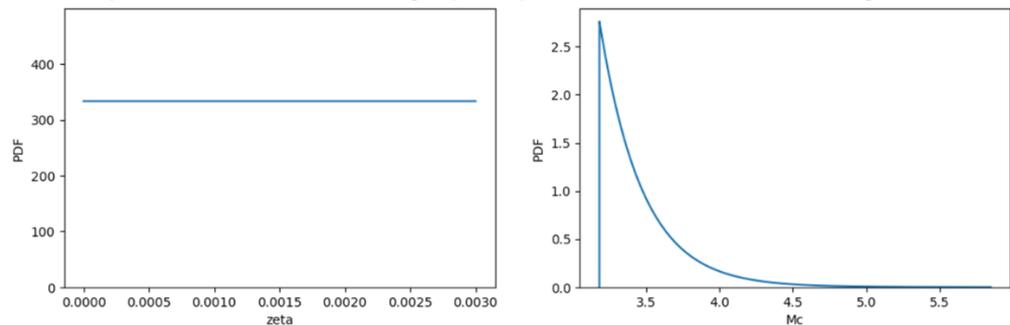


Figure 5: A uniform prior in zeta (from 0.0 to 0.003) is equivalent to an exponentially decaying prior in M_c .

This is not necessarily wrong, but it is at least important to realize (and justify) the choice of prior⁵. We could attempt to define a different prior, perhaps uniform in M_c . However, this choice also needs to be justified, and comes with its own challenges. A lower bound for M_c can easily be found (one could for example chose $M_c^{\text{low}} = M_{\text{min}}$), but an upper bound is much harder. Let's, for the sake of demonstration use a uniform prior for M_c : $[1.5, 7.5]$ to replace the uniform prior in ζ , and repeat the experiment to see the expected behavior of the Bayesian inference. The complete prior information will be referred to as π^{bM_c} .

For the first catalogue, we see that the posterior distribution for M_c is truncated at M7.5 due to the truncation of the prior. The higher we choose this upper bound, the more posterior weight gets put on higher corner magnitudes (we would also put more prior weight on these higher magnitudes after all), and the more the PPDM resembles a traditional Gutenberg-Richter distribution without taper. In this case, that happens to be the ground truth, so this might seem desirable, but the same would happen on a ground truth that does contain a taper.

⁵ The stress-dependent taper model from NAM has a different prior which – due to the stress-dependence – is a little more complicated. Nonetheless, the argument remains that any prior should be justified, especially when it heavily impacts the posterior.

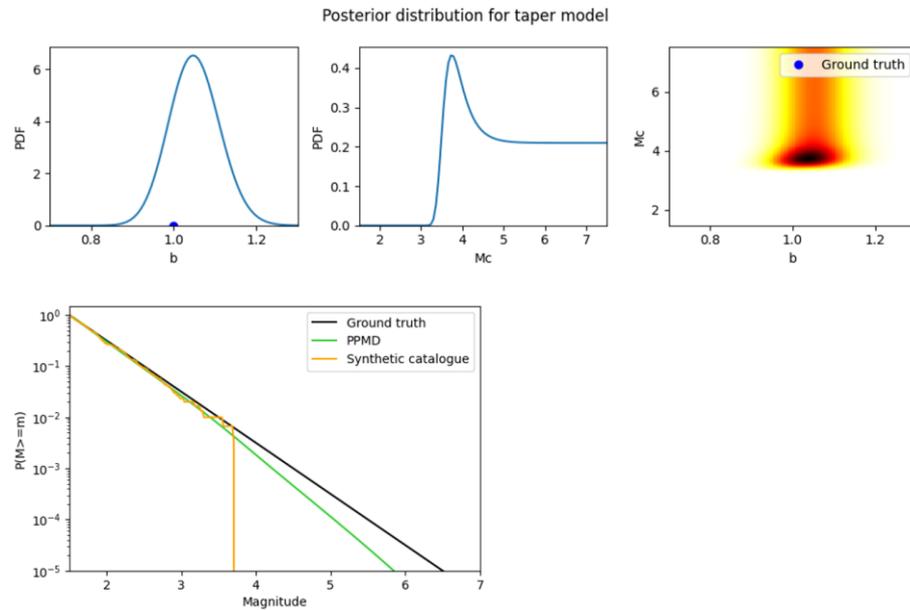


Figure 6: Posterior distribution (marginalized and joint) based on a uniform prior in b, M_c space (all points in joint space have equal prior probability). The resulting PPMD in relation to the synthetic catalogue and the ground truth is shown below.

We therefore compare the expected PPDM for ground truths with M_c in the set $\{3,4,5,6, \infty\}$, both with the prior $\pi^{b\zeta}$, uniform in b and ζ and prior π^{bM_c} , uniform in b and M_c .

All these cases, we run 10,000 iterations of sampling a catalogue from the ground truth ($N=300$), inferring a posterior model parameter distribution, and calculating the PPMD. The expected PPMDs for these cases are shown in Figure 7.

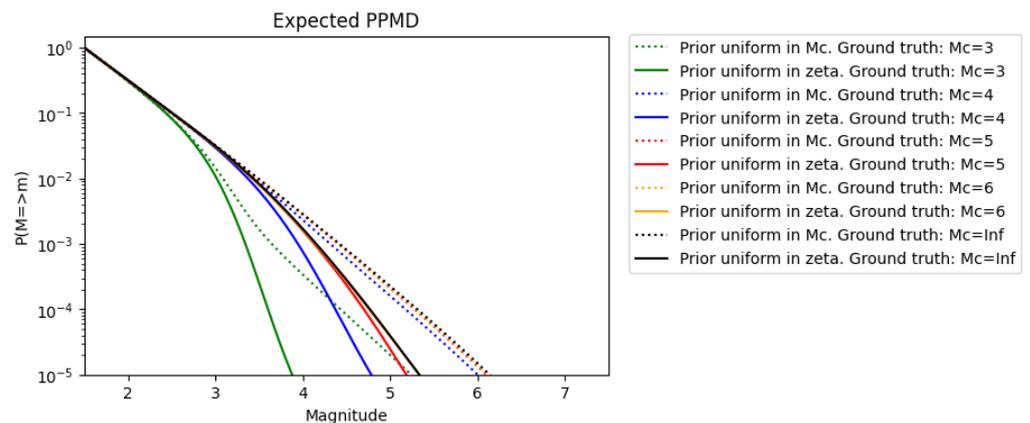


Figure 7: The expected PPMD (similar to Figures 2 and 4) for different ground truths and priors. Priors uniform in M_c are shown with a dotted line, priors uniform in ζ are shown with a solid line. The Expected PPMD curves for ground truth $M_c=6.0$ are underneath the $M_c=\infty$ curves. There is substantial difference between the solid and dotted lines of the same color, indicating the role that parameterization/prior distribution plays.

This demonstrates that the choice of prior has an extremely large impact on the posterior predictive magnitude distribution, even larger than the ground truth itself. In the following section, we will show the reason for this.

B.6.2 The role of the data set

As stated before, the posterior distribution is determined by the prior distribution and the data. In the previous paragraph, we have shown the dominant effect of the prior when attempting to infer the location of a (potential) taper in a dataset of 300 earthquakes. Why were we able to completely overwhelm our uniform prior in b -value, while we can't seem to get a hold on the taper location? The reason for this is in the nature of the parameter: the b -value describes a property of the body of the dataset, while the taper location is a property of the tail.

Let's look in more detail at one of the previous experiments, with ground truth: $b = 1.0$, $M_{\min} = 1.5$, $M_c = 5.0$, and a uniform prior in b and ζ ⁶. For each synthetic catalogue, we record the highest sampled magnitude (the highest 'observed' magnitude), the Maximum Likelihood Estimate (MLE) for the taper location, and the under/overprediction of M5.0 exceedance compare to the ground truth⁷.

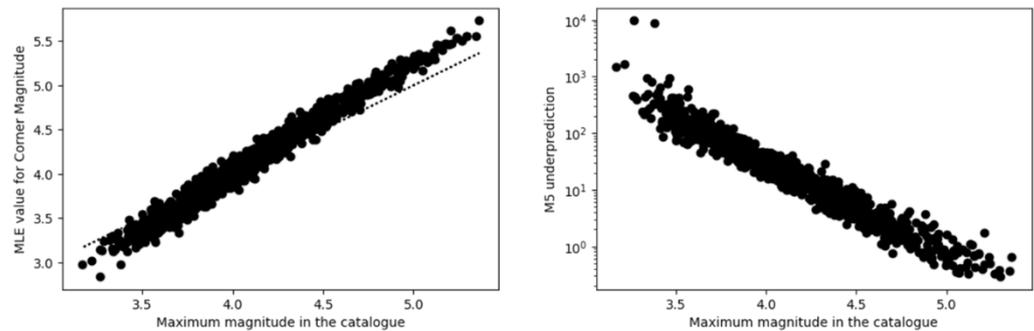


Figure 8: The clear relation between the highest observed magnitude in a catalogue and the MLE for M_c . Similarly, the relation between the M5.0 underprediction (PPMD M5.0 exceedance probability compared to ground truth). This illustrates the large role that chance plays in these small data sets.

The clear relation between the maximum magnitude in the catalogue and the inferred MLE value of M_c (which is nearly one to one), as well as between the maximum magnitude in the catalogue and the underprediction at M5.0 (which takes into account the full inferred posterior distribution of model parameters) shows a crucial problem: the posterior distribution (and therefore the posterior predictive magnitude distribution) is heavily impacted by the largest observed magnitude.

This can be understood by considering that the data will reject (assign very low likelihood to) all taper locations that are below the largest observed magnitude. Also, taper locations above the maximum observed magnitude will be evaluated increasingly unlikely, because the entire magnitude distribution will become broader, leading to a lower probability density over the range of past observations, and therefore to a lower total likelihood. Also, in the case of a prior uniform in ζ , the prior will also help to keep the corner magnitude as low as possible (see Figure 5), just a bit above the maximum observed.

⁶ The same behaviour appears regardless of ground truth.

⁷ Note that while the MLE value describes one point of the posterior distribution, the M5.0 exceedance under/overprediction is a property to PPMD, and therefore the entire posterior distribution of model parameters. On the other hand, the MLE value is independent of any subjectivities in the choice of parameterization/prior.

In a dataset of 300 earthquakes above magnitude $M=1.5$, and a corner magnitude in somewhere above, say, $M=3.0$ and up, the sampling of the tail of the distribution is extremely sparse. In fact, in a catalogue of this size, the tail model parameter is almost entirely characterized by the largest magnitude sampled. And since this magnitude may span several magnitudes (in the ground truth $b = 1.0$, $M_{\min} = 1.5$, $M_c = 5.0$, $M_{\max_observed}$ ranged between $M=3.0$ and $M=5.5$), the apparent taper location is in fact largely controlled by chance. If you happen to sample $M_{\max_observed}=3.0$, you'll infer a taper around $M_c=3.0$. If you happen to sample $M_{\max_observed}=5.0$, you'll infer a taper around $M_c=5.0$. The ground truth barely impacts this, as shown in Figure 9.

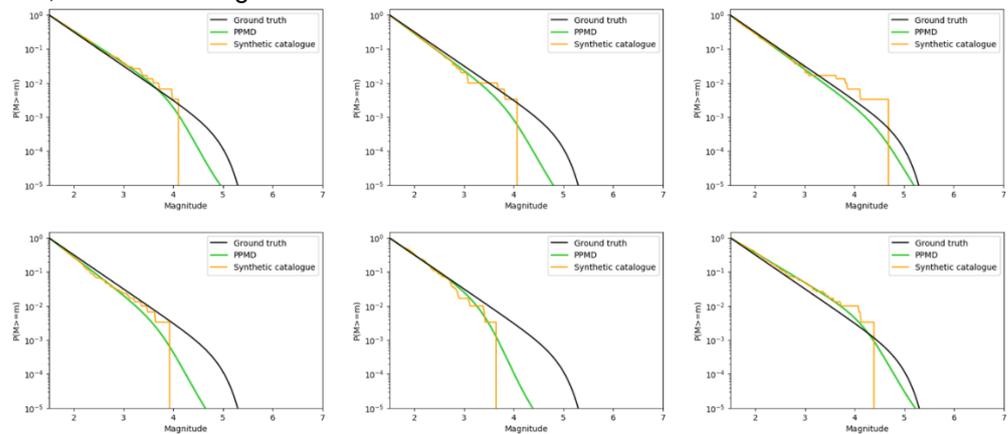


Figure 9: A number of catalogues sampled from the same ground truth ($M_c = 5.0$) and the resulting PPMD (based on uniform priors in b - ζ space). The PPMD is barely related to the ground truth, and is in fact almost entirely controlled by the prior and the largest observed magnitude(s).

B.7 Conclusion

Synthetic experiments confirm that Groningen induced seismicity (earthquake catalogue of around 300 earthquakes) has not nearly supplied sufficient data to constrain the high end tail of the magnitude distribution. This was already universally agreed upon for truncated tail model with a cutoff parameter M_{\max} . Instead, in Groningen hazard and risk assessment, M_{\max} is treated as an epistemic uncertainty. An expert elicitation workshop was organized to establish its distribution on the basis of expert opinion (NAM, 2016). Somehow, this insight has been largely ignored in response to the introduction of the smooth alternative for the tail description with an exponential taper characterized by a corner magnitude M_c . Regardless of the smoothness of the tail model (truncated or tapered) its parameter (M_{\max} or M_c) cannot be inferred from the data and should be treated as an epistemic uncertainty.

Our analysis does not imply that a tapered magnitude distribution is not applicable as a model in Groningen. However, if we want to represent the possibility of a taper in the Groningen pSHRA, it should be based on a sound procedure, properly taking into account the epistemic uncertainty. For instance, a taper location distribution could be included in the logic tree, with its weights determined through expert elicitation. In that case, the expert elicitation panel would also be able to assess

whether M_c should replace M_{\max} in the logic tree, or whether these two tail-describing parameters should both be considered simultaneously. Alternatively, either tail model could be subjected to a Bayesian inference procedure, but only with conscientious treatment of the epistemic uncertainty represented in the prior distribution of the tail model parameter(s). Also these prior distributions should ideally be determined by expert elicitation. In that case, a challenging aspect would be to make sure that the observations do not taint the proposed priors, as that would amount to using the same data twice.

B.8 Additional material provided with this appendix

B2: TNO's position on the applicability of NAM's SSM V6, expressed to the Ministry of Economic Affairs and KEM-subpanel (2021-03-02). This earlier but not widely distributed memo provides a TNO reaction to the NAM rebuttal and the NAM assurance panel review of earlier critique by TNO on the application of the exponential taper model.

B.9 References

- Bourne, S.J., Oates, S.J., (2020). *Stress-Dependent Magnitudes of Induced Earthquakes in the Groningen Gas Field*. Journal of Geophysical Research: Solid Earth. <https://doi.org/10.1029/2020JB020013>
- Kagan, Y. & Schoenberg, F. P. (2001). *Estimation of the Upper Cutoff Parameter for the Tapered Pareto Distribution*. Journal of Applied Probability. 38. 10.1239/jap/1085496599.
- Main, I. G. (1995) *Earthquakes as critical phenomena: Implications for probabilistic seismic hazard analysis*. Bulletin of the Seismological Society of America 85 (5): 1299–1308. doi: <https://doi.org/10.1785/BSSA0850051299>
- NAM, (2016). *Report on Mmax Expert Workshop. 8-10 March 2016*. <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/cef44262-323a-4a34-afa8-24a5afa521d5>
- NAM, (2019). *Evolution of induced earthquake magnitude distributions with increasing stress in the Groningen gas field*. <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/7ac27ead-c59e-418b-9d1f-3897a96021d7>
- NAM (2020). Operationele Strategieën voor het Gasjaar 2020-2021, including Appendix A: Seismic Hazard and Risk Assessment Groningen Field update for Production Profile GTS 2020, EP202002207545, March 13, 2020 (update April 9, 2020).
- TNO (2020). Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen, TNO 2020 R11052

C Period-to-period correlation model for public SHRA Groningen

C.1 Introduction

In the Status Report 2020, TNO (2020a) signaled points of concern related to inconsistencies between the design and the implementations of period-period correlation in the Groningen Ground Motion model. The inconsistencies had been noted earlier in a comparison study between the NAM and TNO implementations of the Groningen seismic hazard and risk model (TNO, 2019). In the Status Report it was noted that the topic still had to be discussed before giving an advice.

The discussion has subsequently taken place during an informational workshop organized by the developers of Groningen GMM V7, where TNO was present as a guest. In preparation for this workshop TNO had written a discussion paper (TNO, 2020b). In the discussion paper TNO advised to maintain the standing practice of applying the period-period correlation model both in the reference ground motion model and the site response amplification model. Unfortunately, the NAM development team took the opposite position and has advised to not apply any period-period correlation in the site response. This advice was subsequently adopted by the State Supervision of the Mines and ultimately it appeared as one of the model prescriptions in the Assignment letter for the public seismic hazard and risk assessment 2021 (EZK, 2021). The position of the model developers was explained in a review/tutorial document written by P. Stafford (2021).

The decision chain leading to the appearance of the model prescription, based on the advice by the model developers, has been very unfortunate, because the objections of TNO described in TNO (2020b) still stand. The explanation by Stafford (2021) has not addressed the major concerns adequately, as will be explained in the following section. The consequences of the incompatibility between ground motion and fragility functions are illustrated in the section “Effects on risk assessment”.

C.2 Response to “Critique” by P. Stafford

The “Critique ..” document by P. Stafford (2021), PS2021, is for a large part written in tutorial form, very clear in presentation and a welcome addition to the GMM documentation by Bommer et al. (e.g., 2018, 2019). However, on the topic of period-period correlation and the application and consequences for Groningen risk assessment, the document is not convincing. We provide pointwise a number of rebuttals and comments. TNO2020b stands for TNO (2020b). B&J stands for Baker and Jayaram (2008), p2p stands for period-to-period, FF stands for fragility framework.

- PS2021 asserts that the TNO2020b relies “extremely heavily” on the assumption that the B&J model is correct. That is not our point. We welcome any other model proposal with justification. So far, the B&J model is the only

model that was explicitly justified for p2p correlation in the Groningen GMM in Bommer et al. (2017). Therefore, the B&J is our point of reference. On the other hand, the FF has been derived on the basis of B&J, so the only way to maintain compatibility without changing the FF is indeed to reproduce B&J. Any deviations should be accompanied by a recalibration of the FF. When we refer to SaAvg values as being too low due to the reduced correlation structure, it is in the context of matching the FF.

- PS2021 suggests that our 10,000 realization Monte Carlo experiment could have been obtained more directly. Although that may be the case, we like to point out that due to the nonlinear relation between reference ground motions and site response amplification the effective correlation structure is certainly not trivial. The Monte Carlo approach is at least a very straightforward approach to obtain results quickly.
- The concept of “Correlation in the means” as introduced by the author is not what is being discussed here. The p2p correlation structure discussed here is by definition relative to a mean (ground motion, or amplification).
- The main justification offered by PS2021 for the lack of p2p correlation structure in the site response is based on (non-) ergodicity. However, in our understanding, the concept of ergodicity bears little relation with the p2p correlation structure, as it refers to the inference of local probability distributions from spatially distributed observations. The practice of “partially non-ergodic” models may help to improve model precision and reduce the dispersion, but does not necessarily affect, let alone remove, the p2p correlation structure of the remaining variability. As long as there is uncertainty/variability in the site amplification, it is bound to have some p2p correlation structure. It is interesting to find out what it looks like though. In fact, it can be studied from the Deltares site response analysis data.
- The Groningen model is referred to as “partially non-ergodic”. However, within each of the site response regions the model is as “ergodic” as any. Therefore, it seems that non-ergodicity cannot be used as an argument to reject the B&J correlation model.
- PS2021 claims that the B&J correlation model is appropriate for use in the fragility model, regardless of the correlation model in the GMM, because it would for that purpose represent spatial variability between zones. In the words of the author: “[...] we have highlighted the fact that the Baker & Jayaram (2008) model is not the appropriate target within the partially non-ergodic Groningen GMM. However, the Baker & Jayaram (2008) correlation model is appropriate to use in the fragility curve derivation because the fragility curves are not developed to be specific to individual site zones. Single curves for building classes are applied to all relevant buildings throughout the field, regardless of which site zone they are located within. It is therefore appropriate to use an ergodic correlation model for the purpose of deriving the fragility curves.” This justification is quite difficult to grasp, as it is not clear how the spatial variability between zones could be relevant for the p2p correlation structure at a single site within a single zone. Perhaps there is some confusion with a distinct application in the context of group/portfolio risk, where buildings with different vibration periods are also located at different locations. (This application is elaborated in the slide set included in PS2021). However, that application is not relevant for the current issue in the calculation of local personal risk.

- PS2021 misses the point TNO2020b makes on the effect of p2p correlation on SaAvg values, the incompatibility between the ground motion and fragility models, and resulting underestimation of seismic risk. This point is further elaborated in the following section.

To conclude, the document by Stafford (2021) does not adequately address the points of concern/objection by TNO (2020b), and therefore does not affect its message.

C.3 Effects on risk assessment

A risk calculation has been carried out using a hazard consistent with the NEN-NPR webtool of December 2018 (SSM V5, GMM V5) and the V6 fragility model. The Groningen region has been divided in 6236 grid cells that correspond to the NPR webtool and at the center of each grid cell, a building of each vulnerability class in the V6 fragility model was placed. The local personal risk (LPR) has been calculated twice for each cell and each vulnerability class: first, using a site response correlation model according to the (former) NAM HRA2020 implementation (Bourne et al., 2019), and second, using the Bommer et al. (2018, 2019) specification⁸.

Two maps showing the cells within which the vulnerability class URM4L exceeds the Meijdam norm (10^{-5}) are shown for both implementation in Figure 1. It is seen that the number of exceeding cells decrease from 605 with the (former) NAM implementation to 18 using the Bommer et al (2018, 2019) implementation.

URM4L,2020, max LPR=2.36e-05, no.cells>1e-5=605 URM4L,2020, max LPR=1.26e-05, no.cells>1e-5=18

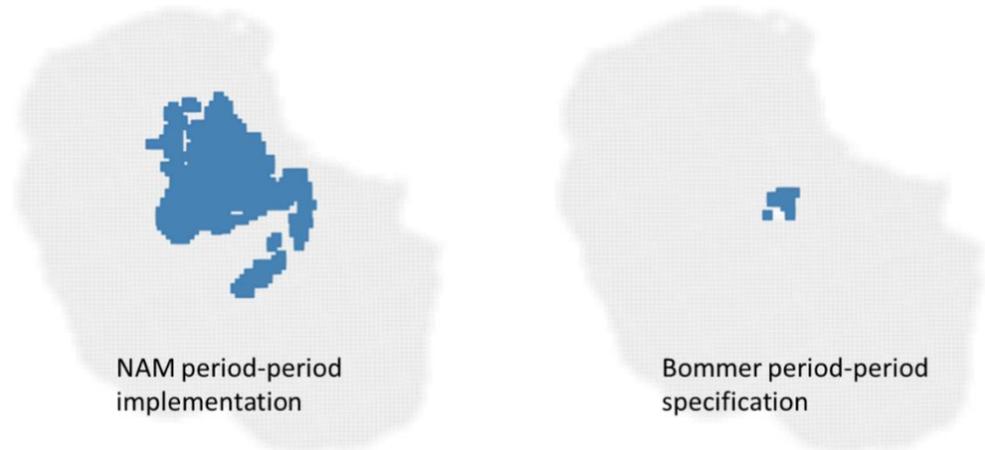


Figure 1 Grid cells where vulnerability class URM4L V6 exceeds Meijdam norm using the two period-period correlation models. The “NAM” and “Bommer” designations refer to the status at the HRA 2020.

⁸ Note that this analysis is not updated with the latest model versions. The result would be qualitatively the same.

The reduction in LPR is further investigated by plotting the ratio of LPR computed with the Bommer et al (2018, 2019) period-period specification to the NAM implementation for all cells and all building vulnerability classes, see figure 2. From this figure it is observed that the LPR values are always lower using the Bommer et al specification, in particular for LPR values from $1e-7$ and larger the reduction in risk is more than 20% for all building vulnerability classes. The reduction in this range can be as much as 80% for some grid cells and building classes. For the maximum LPR for URM4L the reduction is over 40%. Note that all dots represent a factor of risk underestimation. The blue dots do not necessarily contain actual buildings.

Note that in both cases the individual spectral components of ground motions have the same exceedance probabilities. Therefore, vulnerability classes that are dominantly sensitive to the excitation of a single spectral period, should yield exactly the same risk. The reason for the effects shown in Figures 1 and 2 is the incompatibility in period-period correlation structure between the ground motion and the fragility models as explained in TNO (2020b). The V6 fragility was calibrated using “hazard-consistent” data that honors (rightly so) the period-period correlations structure of the ground motions. As a result, the values of SaAvg, which are used as the independent variable in the regressions, are compatible with the period-period correlation structure. For similar ground motion levels of individual spectral periods at the high end of the ground motion distribution⁹, the associated SaAvg value is expected to be larger when the periods are more correlated, and smaller when the periods are less correlated. Therefore, if during risk assessment the ground motion periods are less correlated than during fragility calibration, the same levels of ground motions will be associated with lower values of SaAvg, and therefore, apparently, with a lower probability of collapse. However, this is merely an artifact of the incompatibility, and should therefore, obviously, be prevented.

⁹ The risk is predominantly associated with the relatively rare “above average” ground motions.

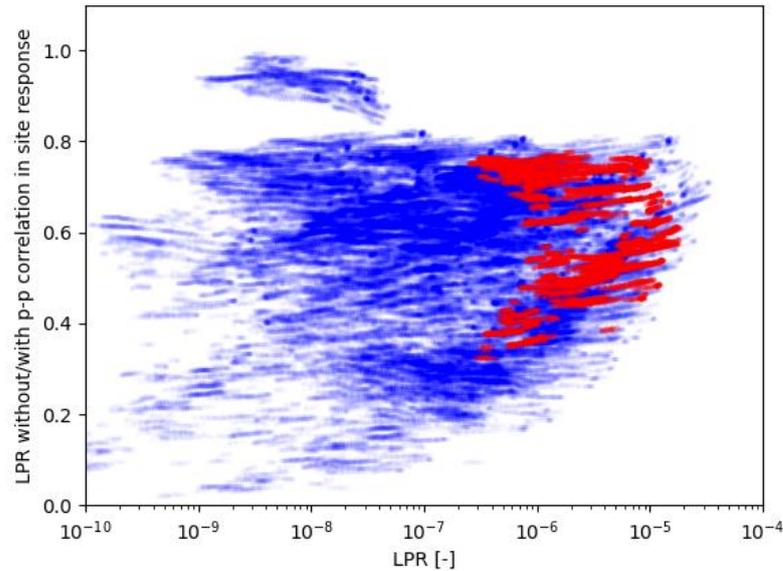


Figure 2 Ratio of Local Personal Risk (LPR) computed with the Bommer et al (2018,2019) specification of the period-period correlation (period-period correlation on NS_B reference level only) to the NAM implementation (period-period correlation on NS_B level and site response) plotted against the LPR as computed with the NAM implementation. Blue are all building vulnerability classes from the V6 fragility report, Red is URM4L. All values in this figure are artefacts of the incompatibility between the ground motion and fragility models.

If the ground motion model developers consider it necessary to change the period-period correlation structure of the ground motion model, then the fragility framework needs to be recalibrated as well. Otherwise, accidents like those illustrated in this section happen.

C.4 Conclusion

The neglect of period-period correlation in the site response as advised by the model developers should be rejected for two reasons major reasons:

- 1 **Insufficient justification.** The model developers claim that (1) the Baker and Jayaram (2008) model is not adequate for Groningen and that (2) the site response amplification model does not require a period-period correlation structure. The first claim may be correct, but is not supported by empirical evidence, while the justification in relation to ergodicity is inadequate. More importantly, the effective correlation structure that implicitly results at the surface is never shown, discussed or checked. The second claim even seems to defy the laws of physics. Any variation in the site response amplification is expected to show period-period correlation to some degree. The choice to neglect these site response correlations seems to be invoked to somehow mitigate the imposed Baker & Jayaram correlations at the reference level. However, there are no checks and balances to make sure the result is correct or adequate. In fact, information on the period-period correlation structure of the site response is available from the site response analyses used to calibrate the site response model. We advise to consider incorporation of these in the next GMM version.

- 2 **Incompatibility between ground motion and fragility models.** The fragility framework was calibrated based on the Baker & Jayaram period-period correlation model. In principle, the fragility framework may not be very sensitive to the period-period correlation structure of the GMM, especially when the fragility of a building is primarily sensitive to the ground motion amplitude at a single spectral period. However, due to the choice of SaAvg as the universal intensity measure for all vulnerability classes (see fragility model V5 to V6), the entire fragility framework has become very sensitive to the period-period correlation structure. The change in period-period correlation structure between the calibration of the fragility model, and the application in the pSHRA 2021 has led to an incorrect apparent decrease in risk levels that is an artifact of the incompatibility that should not be continued in future calculations. If the period-period correlation model should for some reason be changed, then also the fragility model should be recalibrated.

Instead, we advise to maintain the Baker and Jayaram (2008) period-period correlation model in the site response model. This ensures compatibility between the ground motion and fragility models and seems to be a reasonable choice in the absence of a better model.

C.5 Additional material provided with this appendix

C2: Document prepared by TNO for the 2020-10-29 GMM V7 workshop: On the period-to-period correlation structure of ground motion residuals in the Groningen GMM

C3: Critique of the TNO Report entitled "On the period-to-period correlation structure of ground motion residuals in the Groningen GMM" from 29th October 2020. Peter Stafford

C.6 References

Baker, J.W. and N. Jayaram (2008). Correlation of Spectral Acceleration Values from NGA Ground Motion Models, Earthquake spectra, February 2008.

Bommer, J.J., B. Dost, B. Edwards, P.P. Kruiver, P. Meijer, M. Ntinalexis, A. Rodriguez-Marek, E. Ruigrok, J. Spetzler & P.J. Stafford (2017). V4 Ground-Motion Model (GMM) for Response Spectral Accelerations, Peak Ground Velocity, and Significant Durations in the Groningen Field. Version 2, 3 June 2017, 525 pp.

Bommer, J.J., B. Dost, B. Edwards, P.P. Kruiver, P. Meijer, M. Ntinalexis, A. Rodriguez-Marek, E. Ruigrok, J. Spetzler & P.J. Stafford (2018). V5 Ground-Motion Model for the Groningen Field. Re-issue with assurance letter. Revision 1, 14 March 2018, 299 pp.

Bommer, J.J., B. Edwards, P.P. Kruiver, A. Rodriguez-Marek, P.J. Stafford, B. Dost, M. Ntinalexis, E. Ruigrok & J. Spetzler (2019). V6 Ground-Motion Model for

Induced Seismicity in the Groningen Field. With Assurance Letter. Revision 1, 19 December 2019, 193 pp.

Bourne, S.J., S.J. Oates, R. Scheefhals, P. Castellanos Nash, A. Mar-Or, T. Storck, P. Omid, J. van Elk (2019) A Monte Carlo method for probabilistic seismic hazard and risk analysis of induced seismicity in the Groningen gas field. 29 May 2019, 62 pp.

Crowley, H. and R. Pinho (2020), Report on the v7 Fragility and Consequence Models for the Groningen Field, March 2020, 71 pp.

EZK (2021). Verzoek tot voorstellen operationele strategie voor het gasjaar 2021-2022, Kenmerk DGKE-PDG/20330147.

NEN (2020). NPR 9998:2018+C1:2020/A1:2020 nl: Beoordeling van de constructieve veiligheid van een gebouw bij nieuwbouw, verbouw en afkeuren - Geïnduceerde aardbevingen - Grondslagen, belastingen en weerstanden. NEN, 1 August 2020, 40 pp.

Stafford, P. (2021). Critique of the TNO Report entitled "On the period-to-period correlation structure of ground motion residuals in the Groningen GMM" from 29th October 2020, NAM report, <https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/1a25e38e-5591-4a88-80ae-f8a4c6a7a5d5>

TNO (2019). Comparative analysis of the NAM and TNO implementation in the Groningen Seismic Hazard and Risk Assessment, TNO 2019 R11997.

TNO (2020a). Status of the TNO Model Chain Groningen per October 1, 2020 and recommendations for the public Seismic Hazard and Risk Analysis 2021. TNO 2020 R11464, October 9, 2020.

TNO (2020b). On the period-to-period correlation structure of ground motion residuals in the Groningen GMM, October 29, 2020.

D Model description FCM version TNO-2020

D.1 Introduction

Parallel to the development of the TNO Model Chain Groningen for the public Seismic Hazard and Risk Analysis (SHRA) Groningen, a method for assessing the safety of buildings in Groningen with regard to earthquakes was developed by TNO. The development of this methodology was commissioned by the Ministry of Economic Affairs and Climate Policy.

The assessment method is based on a subdivision of buildings into so-called typologies. A typology is a collection of buildings with the same seismic characteristics. The safety assessment is based on a risk calculation at typology level, taking into account the seismic threat in Groningen, and taking into account the strength properties and their variations within a typology. This approach is similar to the regional risk assessment provided by the public SHRA Groningen.

The risk calculation uses strength models for the typologies that adequately describe the properties of the typologies, including the variations present within a typology. In the definitions and elaboration of these strength models, calculations available from other studies (e.g. Crowley et al., 2019; Crowley and Pinho, 2020) which deemed suitable have been used, as well as variation studies carried out by TU Delft in the context of the development of the typology approach. TNO and TU Delft went through a number of steps in translating these calculation results into the model settings used in the risk calculation (TNO, 2021a; 2021b). In response to the results of TNO and TU Delft, the ACVG¹⁰ has assessed the method (ACVG, 2020) and formulated instructions for the implementation of the method in practice (ACVG, 2021).

The description of the properties and the model settings used for the assessment of the typologies are described in dedicated reports for each specific typology or typology group. (TNO, 2021c; 2021d; 2021e)

Starting point of the typology based safety assessment the assignment of a building to a single typology class, based on visual, non-destructive inspection. For this reason the used typology classification slightly deviates from the typology classification used in public SHRA Groningen.

The FCM version TNO-2020 translates the strength models developed for the above described typology approach to the typology classification used in the public SHRA Groningen.

¹⁰ ACVG, the Advisory Board Safety Groningen (Adviescollege Veiligheid Groningen), is an independent board of experts appointed by the minister of Economic Affairs and Climate Policy and provides advice on the safety of buildings and construction works in the context of the reinforcement operation in Groningen.
<https://adviescollegeveiligheidgroningen.nl>

D.2 Description FCM version TNO-2020

This version is based on FCM version NAM-V7. Updates concern the unreinforced masonry (URM) vulnerability classes: The proposed updates generally apply to all unreinforced masonry vulnerability classes and addresses the quantification and implementation of model uncertainty and adjustments of the median seismic capacity for unreinforced masonry.

D.2.1 Quantification model uncertainty

The recommendations from FEMA report P-58 (FEMA, 2012) are used to quantify the model uncertainty. The model uncertainties calculate the effect of the accuracy of the calculation models used. The FEMA system can be applied generically with regard to the model uncertainty. The parameters used are derived and substantiated specifically for the situation in Groningen.

The model uncertainty is represented by the coefficient of variation β_m .

FEMA-P58 stated:

“In this methodology, demand parameter dispersions are estimated based on judgment regarding the uncertainty inherent in response calculation.”

The model uncertainty β_m is specified as the summation of two contributions β_c and β_q :

$$\beta_m = \sqrt{\beta_c^2 + \beta_q^2}$$

Contribution β_c is related to the quality control during the design and construction of the building. It acknowledges the probability that the properties of the structural parts present differ from the properties specified in construction drawings and calculations. This depends, among other things, on the availability and quality of the drawings present and on the extent to which there are inspections to verify whether the quality corresponds. For new buildings this is determined by the assumption to what extent the building corresponds to the design. FEMA P-58 makes recommendations for the value for β_c (Table 12).

Table 12 Values for β_c from FEMA P-58 (FEMA, 2012)

Building Definition and Construction Quality Assurance	β_c
<p><i>Superior Quality, New Buildings:</i> The building is completely designed and will be constructed with rigorous construction quality assurance, including special inspection, materials testing, and structural observation.</p> <p><i>Superior Quality, Existing Buildings:</i> Drawings and specifications are available and field investigation confirms they are representative of the actual construction, or if not, the actual construction is understood. Material properties are confirmed by extensive materials testing.</p>	0.10
<p><i>Average Quality, New Buildings:</i> The building design is completed to a level typical of design development; construction quality assurance and inspection are anticipated to be of limited quality.</p> <p><i>Average Quality, Existing Buildings:</i> Documents defining the building design are available and are confirmed by visual observation. Material properties are confirmed by limited materials testing.</p>	0.25
<p><i>Limited Quality, New Buildings:</i> The building design is completed to a level typical of schematic design, or other similar level of detail.</p> <p><i>Limited Quality, Existing Buildings:</i> Construction documents are not available and knowledge of the structure is based on limited field investigation. Material properties are based on default values typical for buildings of the type, location, and age of construction.</p>	0.40

Contribution β_q is related to the quality and degree of completeness of the analytical model used. The choice of β_q is made based on understanding the sensitivity of the calculated response to structural properties such as strength, stiffness, deformation capacity and issues such as degradation due to the cyclic load. FEMA P-58 makes recommendations for the value for β_q which are shown in Table 13.

Table 13 Values for β_q from FEMA P-58 (FEMA, 2012)

Quality and Completeness of the Analytical Model	β_q
<p><i>Superior Quality:</i> The numerical model is robust over the anticipated range of response. Strength and stiffness deterioration and all likely failure modes are explicitly modeled. Model accuracy is established with data from large-scale component tests through failure.</p> <p><i>Completeness:</i> The mathematical model includes all structural components and nonstructural components in the building that contribute to strength or stiffness.</p>	0.10
<p><i>Average Quality:</i> The numerical model for each component is robust over the anticipated range of displacement or deformation response. Strength and stiffness deterioration is fairly well represented, though some failure modes are simulated indirectly. Accuracy is established through a combination of judgment and large-scale component tests.</p> <p><i>Completeness:</i> The mathematical model includes most structural components and nonstructural components in the building that contribute significant strength or stiffness.</p>	0.25
<p><i>Limited Quality:</i> The numerical model for each component is based on idealized cyclic envelope curves from ASCE/SEI 41-13 or comparable guidelines, where strength and stiffness deterioration and failure modes are not directly incorporated in the model.</p> <p><i>Completeness:</i> The mathematical model includes structural components in the seismic-force-resisting system.</p>	0.40

D.2.1.1 Model uncertainty FCM NAM-V7

For unreinforced masonry Eucenter (Crowley et al., 2017) stated:

- For β_c , the model uncertainty due to construction quality assurance is assessed as “average” and the value of dispersion is taken as 0.25. This is because, even though documents regarding the building design are available and material properties have been tested, those refer to a specific index building. For the other buildings in the class β_c will be larger and “superior” quality cannot be assigned.

- For β_q , the analytical model was judged as “superior” due to “the extensive cross validation of the LS-DYNA and ELS software” and a value of 0.1 has been assigned.[...] These software tools have been validated and/or calibrated for seismic analysis of Groningen buildings using the results of a large number of experimental tests.”

The combination of the two yields a value of 0.27 for β_m .

This is value for model uncertainty is adopted for the URM typology classes in the FCM NAM-V7 and earlier version.

D.2.1.2 Model uncertainty FCM TNO-2020

TNO recommends using a higher value for β_q for masonry buildings instead of 0.10.

This is due to the fact that the SDOF models used for the derivation of the vulnerability curves cannot be regarded as 'superior', because they are a simplification of the calibrated 3D models.

While TNO acknowledges that the 3D MDOF FEM models for index buildings are indeed of high quality, the approach followed to derive fragility functions is to match simplified SDOF models to these MDOF models. Fragility functions directly derived from the MDOF models show more variability than those derived from SDOF models. In addition, the many modeling assumptions made in the MDOF models are generally not validated by the more realistic shaking table tests. These uncertainties lead to larger variability as well. TNO (2021-Appendix C) demonstrated by comparing MDOF and SDOF fragility functions for selected index buildings, that using $\beta_q = 0.25$, which FEMA recommends for average models, leads to agreement of MDOF vs SDOF derived fragility functions.

By using a values of 0.25 for β_q , the model uncertainty β_m for all unreinforced masonry vulnerability classes is increased from 0.27 to 0.35.

D.2.2 Implementation model uncertainty

The model uncertainty is taken into account via a so-called logic tree with a 'Lower', 'Middle' and 'Upper' branch. The model-based fragility curve is considered as a 'best estimate', so it is implemented as a 'middle branch' in the logic tree. The model uncertainty is then implemented via the lower and upper branches.

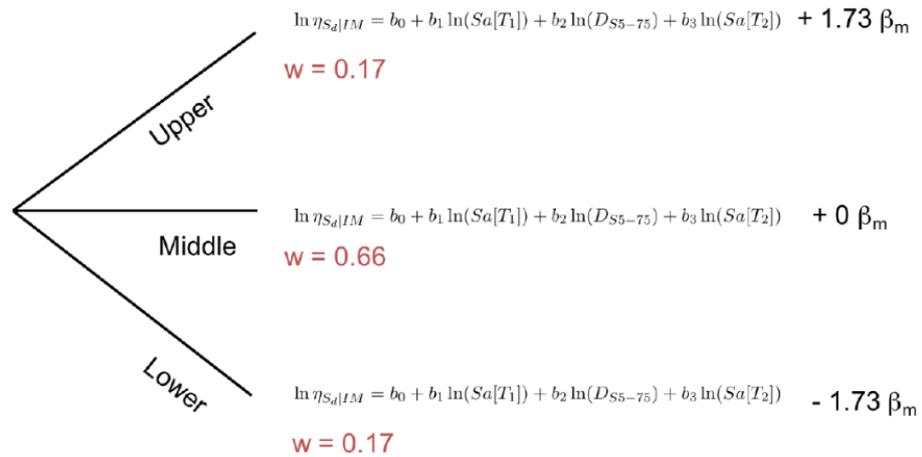


Figure 10 Logic tree for the fragility model taking into account model uncertainty β_m cf. Crowley et al. (2020)

It should be noted that the approach in NAM-V7 (Crowley et al, 2020) deviates from this; there, the model-based fragility and consequence curve is implemented as an 'upper branch' in the risk calculation, which leads to a lower risk. Crowley et al. (2020) justifies that the numerical models consistently underestimate the true strength. TNO finds this insufficiently substantiated and considers the methodology used to be insufficiently safe on this point.

A comparable approach was followed in NAM-V7 for the logic tree based uncertainty of the consequence model. The model-based consequence curve is also implemented as an 'upper branch' in the risk calculation, which leads to a lower risk.

TNO considers the model-based consequence curve as a 'best estimate', so it is implemented as a 'middle branch' in the logic tree.

D.2.3 Derivation of fragility curves

The vulnerability of a typology is described with a fragility curve and in that fragility curve all the effects of different failure mechanisms (including in-plane and out-of-plane failure) are included. Fragility curves indicate the relationship between an intensity measure and the probability of collapse and are determined by a median and a variation around the mean.

The derivation of fragility curves for a typology is based on complex 3D finite element models with material models calibrated to shake table tests, both at element level (walls in/out of the plane) and at building level. These models, in the form of NLTH (Non Linear Time History) calculations and hereinafter referred to as MDOF (Multi Degree of Freedom) models, describe all relevant failure mechanisms of a building up to and including complete collapse.

The fragility curves should include 3 types of variability: the variability in earthquake signals (signal-to-signal variation), the variability in parameters within buildings and the building-to-building variation.

The most advanced method to arrive at fragility curves would be based entirely on MDOF models. However, these MDOF models are very computationally intensive and to calculate a sufficiently large set of earthquake signals with sufficient variations of building parameters and for sufficiently different buildings within one typology, an enormous amount of calculations is required. This is not feasible. Therefore, an approach is used in which MDOF models for one or more reference buildings are used in the derivation of the vulnerability curves. Due to the intensive computation times, these MDOF models cannot be subjected to many earthquake signals. In order to deal with this, for each reference building an SDOF (Single Degree of Freedom) model was constructed and calibrated to the existing MDOF model.

The calibrated SDOF models are subjected to 200 tremor signals in order to determine the signal-to-signal variation. When using several reference buildings per typology, these together give a fragility curve.

A fragility curve derived via SDOF models only partially describes the distribution within a typology: the variation of properties within the building still has to be processed and the building-to-building variation around the reference building or the reference buildings within the typology still has to be processed. These are taken into account by adding an extra variation and applying a median shift if necessary. A correct median choice must ensure that the median seismic capacity in the fragility curve actually corresponds to the Groningen practice for the specific typology.

MDOF-based fragility curves with all variations are only available for a limited number of typologies, namely URM3L and URM4L and for farms (URM1F) (Arup, 2019a;2019b). This information is therefore combined with information from the literature. The MDOF vulnerability curves derived in (Arup, 2019a;2019b) are based on variations as found in the 'Exposure Database' (Arup, 2019c) and as such can be seen as a representative representation of the Groningen building stock. This concerns both variations in properties within a building and variations between buildings within a typology.

D.2.4 Median shift for within building variations and building-to-building variations

The median of the fragility curves for the signal-to-signal variation for SDOF models has a slightly different position than for MDOF models for reference buildings, sometimes higher sometimes lower. If, in addition to signal-to-signal variation, within building variation is also investigated with MDOF models, a median shift is observed for masonry buildings. When the signal-to-signal variation in SDOF is taken together with within building variation, the effect over several reference buildings is a median shift to lower intensity values of about 15%. When building-to-building variation is included, no significant median shift is observed beyond this 15%.

TNO considers the observed median shift to lower intensities as material property of masonry. Therefore a shift of -15% is therefore applied to all unreinforced masonry vulnerability classes.

D.2.5 Additional safety margins proposed by ACVG

ACVG (2021) proposed the use of additional safety margins. This concerns an additional median shift of the fragility curve and additional building-to-building variation. These safety margin apply to specific vulnerability classes and not all unreinforced masonry classes. Because the vulnerability classes of the public SHRA are not exactly similar, and the typology based safety assessment serves a different purpose, these additional safety margins are therefore not translated to FCM version TNO-2020.

D.2.6 Summary of model updates

The model updates and changes with regard to NAM-V7 are summarize in Table 14

Table 14 Model updates of TNO-2020. For comparison the NAM-V7 values are given.

vulnerability class	Model uncertainty β_q		Representative logic tree branch fragility curve		Representative logic tree branch consequence curve		Shift median seismic capacity	
	NAM-V7	TNO-2020	NAM-V7	TNO-2020	NAM-V7	TNO-2020	NAM-V7	TNO-2020
URM1F_B	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM1F_HA	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM1F_HC	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM2L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM3L	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM3M_B	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM3M_D	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM3M_U	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM4L	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM5L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM6L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM7L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM8L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM9L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM10	0.1	0.25	middle	middle	middle	middle	0%	-15%

D.3 References

ACVG (2020). Advies typologieaanpak, kenmerk ACVG202011-01, 17 november 2020. <https://adviescollegeveiligheidgroningen.nl/sites/default/files/2020-12/ACVG202011-01%20ACVG-advies%20Typologieaanpak.pdf>

ACVG (2021). Beoordeling aanvullingen TNO op typologieaanpak, kenmerk ACVG202105-01, 27 mei 2021. <https://adviescollegeveiligheidgroningen.nl/sites/default/files/2021-06/ACVG202105-01%20Beoordeling%20aanvullingen%20TNO%20op%20typologieaanpak.pdf>

- Arup (2019a). Typology Modelling Explicit MDOF Validation of Fragility Functions - URM3L - URM4L, 30 juni 2019
- Arup (2019b). Typology Modelling Explicit MDOF Validation of Fragility Functions - URM3L-URM1_F, April 2020
- ARUP (2019c). "Exposure Database V7: Data Documentation, Technical Report and Exposure Model", December 2019
- Crowley, H., Pinho, R. and Cavalieri, F. (2019). Report on the v6 Fragility and Consequence Models for the Groningen Field, March 2019.
- Crowley, H. & Pinho, R. (2020). Report on the Fragility and Consequence Models for the Groningen Field (version 7), March 2020.
- FEMA (2012). FEMA 58-1, Seismic Performance Assessment of Buildings, Volume 1, Methodology, FEMA, Washington DC, USA, 2012
- TNO (2021a). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen -Typologisch toedelen, TNO 2021 R11002B, 2 september 2021.
<https://www.nationaalcoordinatorgroningen.nl/documenten>
- TNO (2021b). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen - Achtergrond bij de methode, TNO 2020 R10628B, 2 september 2021.
<https://www.nationaalcoordinatorgroningen.nl/documenten>
- TNO (2021c). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen – Uitwerking van typologie METSELWERK1, TNO 2020 R10699B, 6 september 2021.
<https://www.nationaalcoordinatorgroningen.nl/documenten>
- TNO (2021d). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen – Uitwerking van typologie METSELWERK2, TNO 2020 R10700A, 3 september 2021.
<https://www.nationaalcoordinatorgroningen.nl/documenten>
- TNO (2021e). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen – Uitwerking van typologiegroep METSELWERK-C (typologieën METSELWERK5+6+7), TNO 2020 R10966A, 2 september 2021.
<https://www.nationaalcoordinatorgroningen.nl/documenten>

E Additional material

This report refers to addition material. For convenience these documents are added to this report. It contains the following documents:

- B2: TNO's position on the applicability of NAM's SSM V6, expressed to the Ministry of Economic Affairs and KEM-subpanel (2021-03-02)
- C2: Document prepared by TNO for the 2020-10-29 GMM V7 workshop: On the period-to-period correlation structure of ground motion residuals in the Groningen GMM
- C3: Critique of the TNO Report entitled "On the period-to-period correlation structure of ground motion residuals in the Groningen GMM" from 29th October 2020. Peter Stafford
- Review of the TNO 2021 Model Chain Groningen Report, Jean-Paul Ampuero and Eric-Jan Wagenmakers, December 28, 2021

TNO's position on the applicability of NAM's SSM V6

Date

Our reference
AGE 21-10.012

Page
1/5

For the Groningen Hazard and Risk Assessment 2020 (hereafter HRA 2020), as reported in NAM (2020a), NAM has developed an update to the Seismological Source Model. This update is described in (NAM 2019) and has informally been dubbed the 'SSM V6' source model. At the request of the Minister, TNO has advised on the suitability of the model updates for use in the HRA 2020 (TNO 2020, hereafter TNO Advice). TNO advises to revise the logic tree weights used by NAM, effectively 'turning off' the most impactful update in V6: the application of an exponential taper on the frequency-magnitude distribution.

After the TNO Advice, NAM has written an extensive rebuttal (NAM 2020b, hereafter NAM Rebuttal) and has asked its external Assurance Panel (which had also reviewed the update to V6) to comment on both documents. In addition, NAM and TNO have had a sequence of interactions through video-conferencing and informal exchange of memos, to discuss analysis methods and points of view and to clarify any potential misunderstanding.

We can summarize our conclusion by stating unequivocally that the TNO Advice of May 2020 still stands as formulated originally. In fact, our efforts to improve the communication of our findings during the discussions with NAM have helped to accentuate our argumentation. Neither the NAM Rebuttal nor the associated Assurance Panel Review report have presented reasons to deviate from our original advice. Both of these documents are built on misreading of our Advice and structurally misinterpreting the analysis methods therein.

Therefore the TNO advice regarding the model selection for the HRA 2021, as formulated in the Status Report (TNO, 2020b) is also in line with the conclusions of the TNO Advice of May 2020.

In the following memo, we briefly review the formal reaction to the TNO Advice by NAM. Then we summarize the final status of the discussions with NAM and finally discuss and conclude with the applicability of NAM's SSM V6.

1. Reactions TNO Advice

NAM Rebuttal

On June 30, 2020, TNO received the NAM Rebuttal document "*Discussion of the 'Advice by TNO on the seismological model for the Hazard and Risk Assessment for induced seismicity in Groningen'*". In the rebuttal NAM expresses an objection to almost every statement by TNO. Most objections, however, are based on misreading the advice and/or structurally misinterpreting the analysis methods therein.

Assurance Panel Review

On September 7, 2020, TNO received the Assurance Panel report: "*Discussion with Assurance Review of the 'Advice by TNO on the*

Date

Our reference
AGE 21-10.012Page
2/5

*seismological model*TM. Two out of three reviewers explicitly support the model and the rebuttal by NAM. De third reviewer (Main) does not do so explicitly, but he confirms a number of the observations made by NAM with regard to their own analysis.

It appears that all reviewers are, to some extent, guided by the narratives of the NAM Rebuttal, since they follow along with the misreading and misinterpretation by NAM. An obvious example is the false narrative by NAM that TNO bases its analysis on point estimates, which is paraphrased by all reviewers. In fact, TNO consistently states that the full posterior model parameter distributions should be used for evaluating model performance, in accordance with HRA practice, while NAM concentrates on the performance of individual model realizations. Also, the reviewers interpret and confirm the supplementary analyses by NAM without judging the (largely missing) relevance to the objections by TNO.

In summary, the Assurance Panel Review cannot be considered separate from the NAM Rebuttal and does not contribute any relevant arguments to the discussion.

2. Summary of the final status of the interactions between TNO and NAM

The following factual statements have been formulated in the preparation for the second online discussion session with NAM of September 9, 2020. The statements were used as guidance for the discussion. Here, we use them to structure the summary of our interactions with NAM according to our interpretations.

Statement 1: The NAM implementation of the source model (using synthetic earthquake catalogues) and the TNO implementation of the source model (using the mean activity rate density grid) are mathematically completely equivalent, apart from differences introduced due to finite sample sizes and finite grid resolutions.

This point turned out to be a common denominator in much of the critique of NAM to the analyses in the TNO Advice as expressed in the NAM-Rebuttal and subsequent interactions. However, the equivalence is an unambiguous mathematical reality. When producing a hazard or risk assessment based on the full posterior distribution of the source model parameters, you are *effectively* using the mean posterior seismic activity rate and frequency-magnitude distributions. The implementation by TNO is not an approximation, but simply a more efficient alternative to the implementation by NAM. Although this point kept coming up in the discussions it is not relevant for our objection to the applicability of the taper model. However, TNO argues that the scientific team from NAM does not acknowledge this *effective* mathematical equivalence. This directly impacts the interpretation of the model competition framework used by NAM (see Statement 2).

Statement 2: The model competition framework used by NAM to determine forecast performance and logic tree weights is not suited for use in HRA context.

Date**Our reference**
AGE 21-10.012**Page**
3/5

NAM evaluates the relative forecasting performance of their proposed model variations by staging an all-play-all forecasting competition. The score of each forecasting “match” is determined by the win/loss proportion of a large number of “games” between individual realisations of both competing model variations. The realisations are randomly selected from the posterior model parameter distributions of both model variations, which in turn are obtained by training on (roughly) the first half of the Groningen earthquake catalogue. The outcome (win or loss) of each game is determined by evaluating which of the two competing models provides a better forecast for the second half of the Groningen dataset. NAM then selects two model variations to appear in the logic tree and proposes to use their respective win percentages as branch weights.

According to TNO, this approach has two major flaws.

The first flaw is that the performance comparison of individual realisations has no pertinence on the actual application of the models in HRA practice. The Groningen HRA is performed using a full posterior distribution of model parameters and individual realisations are never used in isolation. Any relevant performance metric should therefore only be based on forecasts made by the full posterior model parameter distribution, not by individual realisations in isolation.

The second flaw is that the performance comparison is made only on a single earthquake catalogue, i.e. the Groningen catalogue of the past. This means that chance plays a large role in the outcome, and therefore in the selection of the logic tree weights. An important feature related to the role of “chance” is that in the training stage of the model variations, NAM has chosen to define the split between the first and the second half of the catalogue at January 1st, 2013. As a consequence, the largest magnitude of the catalogue, the M3.6 of Huizinge, August 16, 2012, is the last earthquake included in the first half. This means that the largest magnitude happens to be inside the training data. As a result, the magnitudes appearing in the second half of the dataset are all lower and do not come as a surprise. A model is not robust for HRA forecasts, if the inclusion or exclusion of the largest magnitude event in the training set (by choice or by coincidence) changes the logic tree weights significantly. This inclusion strongly favours models with a taper and therefore leads to a biased result as demonstrated by TNO in Experiment 3 of the TNO Advice.

This Experiment 3 has received a lot of attention in the NAM Rebuttal and consequently also in both the later interactions between TNO and NAM and the Assurance Panel Review. The major narrative of NAM – copied by the Assurance Panel – is that TNO mistakenly focuses on a single realisation (point estimate) whereas NAM uses the full distribution, while in fact the opposite is true, as explained above and also in the TNO Advice. Despite the large attention, also this point is not central to our objection to the applicability of the taper model. However, our findings refute the choice of logic tree weights by NAM (and according to the procedure applied by NAM).

Date

Statement 3: The inference-and-forecast procedure, applied to the taper model and a limited dataset for calibration, results in an inherent bias in the forecasted seismicity.

Our reference
AGE 21-10.012**Page**
4/5

In the early stages of development of the HRA models for Groningen it was recognized by NAM that it was not possible to constrain the high end tail of the frequency-magnitude distribution from the observations. The organisation of the Mmax expert elicitation workshop in 2016 has been a direct consequence of this recognition. The tail of the distribution is not or at least not sufficiently well sampled by the past observations, and therefore it lies in the domain of epistemic uncertainty. The common approach to deal with this uncertainty is to use a frequency-magnitude model that extends the scaling law of the smaller earthquakes, as inferred from the data, to higher magnitudes until some cut-off magnitude Mmax. The value of Mmax is fundamentally uncertain and therefore has been included as a discrete distribution in the logic tree with values and weights determined by an expert elicitation process.

With the taper model NAM has introduced an alternative mathematical formulation for the description of the tail of the magnitude distribution. It is presented along with a storyline that makes a lot of sense from a physical perspective and that is more elaborate than a simple, stationary cut-off at Mmax. However, this storyline does not make up for the fact that the sampling of the tail is simply insufficient to be able to calibrate it from past observations.

3. Discussion

The starting point in the NAM HRA model development was that the tail of the frequency-magnitude distribution cannot be determined from the observed earthquake catalogue. This has been confirmed by the Mmax workshop in 2016, where the tail frequency-magnitude was determined by "expert elicitation". Yet, the proposed taper model (SSM V6) departs from the above starting point by determining the tail of the frequency-magnitude distribution from the observed earthquake catalogue. The biggest objection of TNO is that it is not sufficiently convincingly substantiated why the tail of the frequency magnitude distribution can now (presumably) be determined from the observed earthquake catalogue. Unfortunately, the Assurance panel does not discuss this either.

TNO demonstrated – supported by various (counter-)examples - that the choice of a taper model in combination with a calibration (inference) of the SMM V6 model from the limited catalogue of past observations in Groningen leads to biased forecasts. The inference-forecast sequence results in a preference for frequency-magnitude distribution models that taper off quickly beyond the highest magnitude present in the training data. In their Advice and subsequent discussions with NAM, TNO has shown that this happens regardless of the underlying distribution (i.e. regardless of whether a taper is truly present). The effect is aggravated by the choice of model parameterization and prior parameter distribution chosen by NAM, which strongly favours low values of the corner magnitude of the taper.

4. Conclusion

Date**Our reference**
AGE 21-10.012**Page**
5/5

Both the NAM Rebuttal and the associated Assurance Panel Review report have presented no reasons to deviate from TNO's original advice. Both of these documents build on misreading our Advice and structurally misinterpreting the analysis methods therein. Irrespective of their willingness to achieve scientific consensus on this matter, NAM hasn't been successful in disproving our concerns. Hence, TNO still considers the use of the taper model in SSM V6 unsuitable for application in the hazard and risk analyses in Groningen.

5. References

NAM (Bourne, S.J., Oates, S.J.), 2019. Evolution of induced earthquake magnitude distributions with increasing stress in the Groningen gas field. NAM report.

NAM (Van Elk, J., Landman, A.J., Uilenreef, J., Doornhof, D.), 2020. Seismic Hazard and Risk Assessment Groningen Field update for Production Profile GTS - raming 2020. NAM report

NAM (Bourne, S., Oates, S., Van Elk, J., Doornhof, D.), 2020. Discussion of the "Advice by TNO on the seismological model for the Hazard and Risk Assessment for induced seismicity in Groningen", June 2020.

NAM (Bourne, S., Oates, S., Van Elk, J., Doornhof, D.), 2020. Discussion with Assurance Review of the "Advice by TNO on the seismological model", August 2020.

TNO 2020a. Advies vaststellingsbesluit Groningen gasveld 2020/2021. AGE 20-10.043

TNO 2020b. Status of the TNO Model Chain Groningen per October 1, 2020 and recommendations for the public Seismic Hazard and Risk Analysis 2021. TNO2020 R11464

On the period-to-period correlation structure of ground motion residuals in the Groningen GMM

Summary

TNO has identified an inconsistency between the NAM-HRA implementation and the original model specification of the GMM period-to-period correlation structure. The inconsistency has no impact on hazard metrics commonly communicated for the Groningen (hazard curves, hazard maps, UHS spectra). However, it does have a substantial impact on risk assessments, with observed differences in the order of a factor of 2 in LPR.

After examination of the relevant reports and literature, supported by a number of computational experiments, TNO recommends to maintain the implementation by NAM and to revise the specification in the upcoming GMM model version V7 accordingly.

Introduction

TNO (2019) has identified an inconsistency with regard to the period-to-period correlation structure of the spectral acceleration residuals in the Groningen ground motion model (GMM V5/V6) between the documented model specifications by Bommer et al. (2018, 2019), and the actual implementation by NAM in the HRA (Bourne et al., 2019).

On the one hand, the original specification prescribes a period-to-period correlation of the ground motion variability at the North Sea Base reference level, and does not specify a period-to-period correlation for the site response (i.e., to the site-to-site variability). On the other hand, the NAM implementation applies the reference level correlation structure to the site response amplification as well.

The implementation of the period-to-period correlation structure does not influence hazard metrics that are commonly communicated for the Groningen, such as hazard curves or maps for individual spectral acceleration periods, PGA or PGV. It also does not influence the evaluation of uniform hazard spectra (UHS), which are used as the definition of seismic forcing in the context of the NEN (2020) NPR9998 building code, as UHS spectra ignore the period-to-period correlations by definition (e.g., Lin and Baker, 2008). However, the correlation structure does have a non-trivial effect on the assessment of seismic risk in the Groningen context, as the current version of building Fragility Models applied in Groningen (e.g., Crowley and Pinho, 2020) are conditioned on the average spectral acceleration over a sequence of spectral periods.

Period-to-period correlation in the GMM specification and HRA implementation

The Groningen ground motion model (GMM) has been developed in a series of iterations with version designations ranging from V1 to V6 (Bommer et al., 2015a, 2015b, 2017, 2018, 2019).

In the first GMM versions (V1 and V2, Bommer et al., 2015a, 2015b), the period-to-period correlation structure was adopted from Akkar et al. (2014). The V2 report explicitly writes (Section 11.3) that these correlations be applied “[...] to the full variability and therefore to all the variability components listed in Eq. (11.2)”, where said equation contains both reference ground motion and site response (site-to-site) variabilities. A figure (Figure 11.1) with a flow chart representing the V2 implementation, however, fails to convey this prescription as it suggests the use of a period-to-period correlation only for the between-event variability. The V2 report also mentions that it is a “pending exercise [is] to explore whether this correlation matrix is consistent with Groningen V2 GMPE and recordings.”

This consistency exercise was subsequently done and reported in the V4 report (Bommer et al., 2017; the V3 report has not become publicly available). Figure 11.68 therein compares the empirical Groningen period-to-period correlations of residuals at the North Sea Base (NS_B) horizon to three models from literature including Akkar et al. (2014) and Baker and Jayaram (2008). For the V4 and later also V5 and V6 versions the model by Baker and Jayaram was adopted as a “perfectly reasonable and defensible choice”. Although the residuals correlations were compared at the NS_B reference horizon, the V4 report also clearly states that (p.262) “The calculation of the surface accelerations at other response periods at the same site needs to take account of the period-to-period correlation model”, also, in the Executive Summary (p. ix) is says, “The correlation coefficients, [are] to be applied to all components of variability [...]”.

In version V5 (Bommer et al., 2018) and V6 (Bommer et al., 2019), the choice for the correlation model of Baker and Jayaram (2008) is maintained. Section 6 of the V5 report and Section 5 of the V6 report, on *Model Summary and Implementation*, still state (p. 71 and 70, respectively) that “For the risk calculations, values of $Sa(T)$ calculated at a given location for different periods, T , must account for the period-to-period correlations of the residuals. The correlation coefficients, to be applied to all components of variability, for $Sa(T)$ at all 23 periods are exactly the same as those used in the V4 model [...]”. However, in the same sections, Tables 6.5 and 5.3, respectively, it is stated that the period-to-period correlations of spectral accelerations be employed at the NS_B horizon. Table 5.3 of the V6 report is included below as Figure 1. In addition, the elaborated algorithms for “Sampling Variability in Ground-motion Values for Risk Calculations” on pages 78-79 and 74, respectively, explicitly mention the use of a (period-to-period) correlation matrix for all variabilities in Steps 1 and 2, i.e., for the within-event, between-event and component-to-component variability, but not in Step 4, where the site-to-site variability associated with the site response is calculated.

Table 5.3. Correlations of residuals in the Groningen GMM

Symbol	Description	GM Parameter ¹	Horizon ²
ρ_{T_2T}	Period-to-period correlation of spectral accelerations	Sa at multiple T	NS_B

Notes: 1 – The ground-motion parameters to which it applies; 2 – Reference elevation at which employed.

Figure 1: Table 5.3 of the V6 GMM report by Bommer et al. (2019).

Therefore, the more explicit guidance to the application of the period-to-period correlation introduced in the V5 and V6 reports seems to introduce both an internal consistency within the reports, and a discontinuity with the earlier GMM versions.

The implementation of the V5 HRA implementation by NAM has been described to a high level of detail in Bourne et al. (2019). The same implementation has been studied by TNO for the comparison report TNO (2019), including reference to both C and Python computer codes. The implementation by NAM applies the period-to-period correlations to all components of variability, i.e., including the site-to-site variability, which therefore seems (partially) inconsistent with the V5 model specification. It appears that this situation is maintained in the V6 model version and the 2020 HRA implementation by NAM¹.

Experiment on effective correlation structure at the surface

To assess the effect of including or excluding the period-to-period correlations in the site-to-site variability we have generated suites of 10.000 surface ground motion samples for a range of magnitudes, a range of rupture distances and for all GMM logic tree branches. A selection of the results is shown in Figures 2 and 3. From the Figures it becomes clear that to maintain the Baker & Jayaram correlation structure it is important to include period-to-period correlations also to the site response. As expected, lack of period-to-period correlation in the site response leads to a decrease in correlation of the surface ground motions. On the other hand, assuming perfect/full period-to-period correlation in the site response increases the final correlation at the surface.

Figure 3 shows that as the magnitude increases, the correlation/decorrelation effects of the site-response become more pronounced. The reason is that the relative contribution of the site-to-site variability to the total variability increases with increasing reference ground motion.

¹ Although there is no equivalent to Bourne et al. (2019) for the 2020 HRA assessment, and TNO does not have the computer codes, TNO is still able to reproduce the NAM HRA results by maintaining the period-to-period correlation in the site response (see forthcoming comparison report).

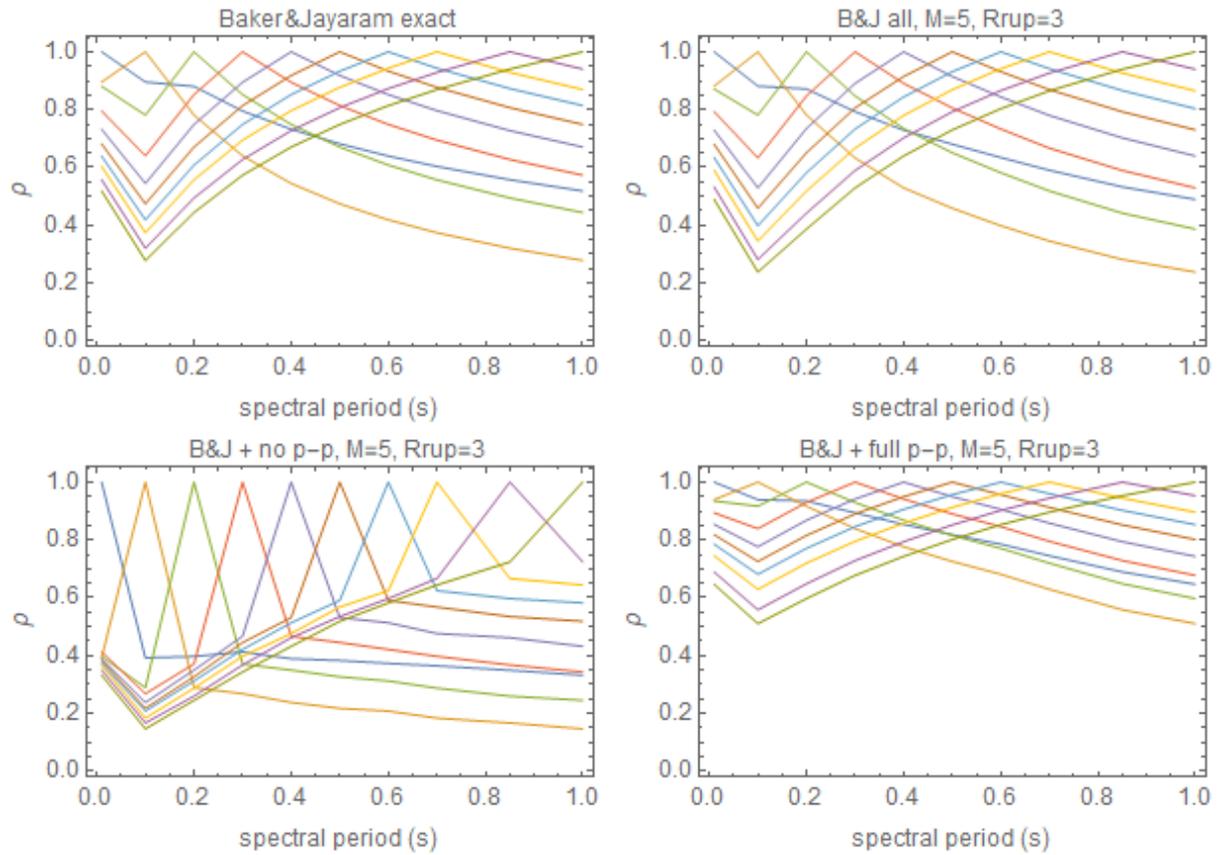


Figure 2: Period-to-period correlation structures. Each line represents a single line/column of the correlation matrix. The diagonal components have a value 1, which makes it easy to associate each curve to a specific row/column. Top left is the exact representation of the Baker and Jayaram (2008) correlation model. The three other figures display experimentally derived (sample) correlation structures for magnitude $M=5$, rupture distance $Rrup=3$, Central-Upper median GMM branch and Upper GMM phi-branch. Top right shows results of assuming the Baker and Jayaram model applies to site response. Bottom left assumes no period-to-period correlation in site-response while bottom right assumes full correlation in site-response.

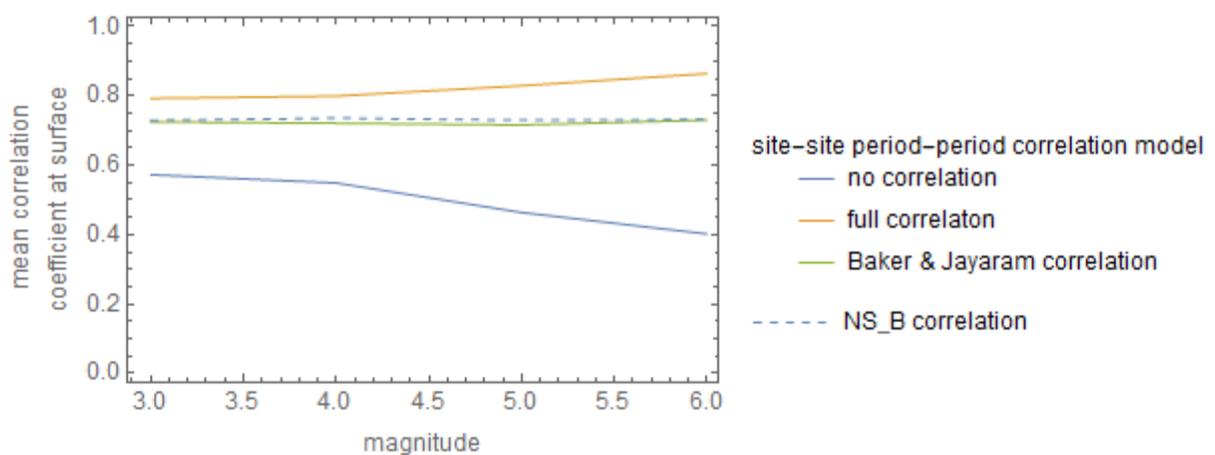


Figure 3: Variation of mean correlation coefficient (mean of all elements of the correlation matrix) for ground motion residuals at the surface as a function of magnitude, for rupture distance $Rrup=3$, Central-Upper median GMM branch and Upper GMM phi-branch. The correlation structure at the reference level is shown as well.

Period-to period correlation in Fragility Model

The period-to-period correlation structure of the Ground Motion Model has impact on the risk assessment through the building fragility model. The Groningen Fragility and Consequence Models (FCM) have been developed in a series of iterations with version designations ranging from V1 to V7 (e.g., Crowley and Pinho, 2020, on V7). In versions V6 and V7 the fragility models for all vulnerability classes have been conditioned on the geometric mean (“average”) spectral acceleration (SaAvg) over a sequence of 10 spectral periods (0.01s, 0.1s, 0.2s, 0.3s, 0.4s, 0.5s, 0.6s, 0.7s, 0.85s, 1.0s). As a result, due to the variance reducing effect of averaging, the internal period-to-period correlation structure is highly influential. Without correlation structure the variance of SaAvg residuals would be reduced by a factor of 10 relative to the average variance of its contributing periods. For perfect correlation, on the other hand, there would be no reduction at all (i.e., a factor of 1).

As part of the fragility model calibration a simplified model of a building is excited using a suite of 200 selected earthquake ground motion records (Crowley and Pinho, 2020). The records are selected to be “hazard-consistent”, in the sense that they are similar to the expected ground motions in Groningen according to number of criteria (Lin and Baker, 2015, Kohrangi et al., 2017). An important aspect of this is that the period-to-period correlation structure match the one expected in Groningen.

To check the period-to-period correlation structure of the records used for fragility model calibration we are able to make use of the original set of 200 records as supplied by personal communication (Crowley, 2019). The obtained records are subdivided in 4 groups corresponding to 4 return periods. Each group contains records that are individually scaled to have a specific, typical value of SaAvg for Groningen for that return period. Although the scaling procedure is useful and perfectly acceptable for the purpose calibration, it does affect the in-sample correlation structure of the records. The easiest way to both understand the effect of the scaling procedure and to make a clean comparison between calibration records and model, is to apply the same procedure to the model correlation structure.

Let Δ be the vector of logarithmic spectral acceleration residuals for N spectral components. The residual of the $\ln(SaAvg)$ is defined as:

$$\Delta_{\ln(SaAvg)} = \frac{1}{N} \sum_{i=1}^N \Delta_i.$$

According to the GMM residual vector Δ has an (approximately) normal distribution with covariance matrix C , that includes the variances of the individual components along the diagonal as well as the covariances for each component pair in the off-diagonal entries. Scaling a record translates as scaling individual spectral acceleration component by the same amount, or, alternatively, shifting their logarithms by the same amount. The collection of calibration records has been scaled in such a way that the sample mean of the $\Delta_{\ln(SaAvg)}$ vanishes. This can be formalized using linear “shifting” operator L , defined as:

$$L_{ij} = \delta_{ij} - \frac{1}{N},$$

with δ_{ij} the Kronecker delta (or unit diagonal matrix), that operates on the residual vector Δ :

$$\Delta^* = L \cdot \Delta$$

to arrive at the shifted residual vector Δ^* . The covariance matrix C^* of Δ^* can now be expressed in terms of the covariance matrix of C and shifting operator L :

$$C^* = L^T \cdot C \cdot L.$$

The “shifted” period-to-period correlation matrix R^* can then be derived from the covariance matrix by C^* using the formula:

$$R_{ij}^* = \frac{C_{ij}^*}{\sqrt{C_{ii}^* C_{jj}^*}}$$

without applying the Einstein convention.

We apply the above procedure to the period-to-period correlation matrix of the Baker and Jayaram (2008) model, basically assuming unit variance for all components. The “shifted” correlation matrix may then be compared to the sample correlation matrix of the hazard consistent calibration records. The result is shown in Figure 4. It turns out that there is a striking similarity. Apparently, the hazard-consistent record selection procedure has managed to capture the desired correlation structure, as it corresponds well to the correlation structure of the actual HRA implementation.

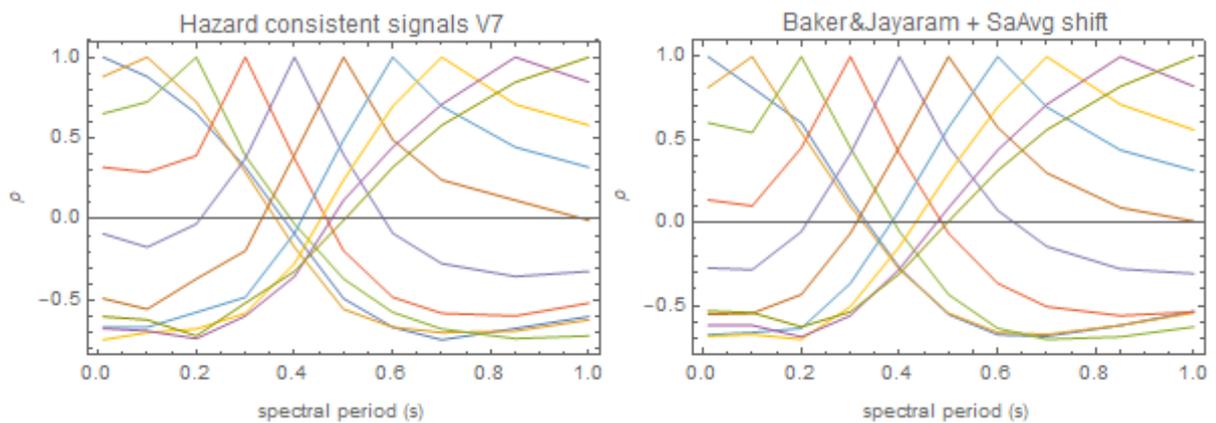


Figure 4: Correlation structures of the hazard-consistent signals used for the V7 fragility calibration (left) and the Baker and Jayaram (2008) model (right). Each line represents a single line/column of the correlation matrix. The diagonal components have a value 1, which makes it easy to associate each curve to a specific row /column. Both correlation structures are determined for residuals relative to the SaAvg residual, see main text for details. The correlation structures are very similar, meaning that the calibration signal represent the imposed Baker and Jayaram correlation structure very well.

Discussion

We have noted the inconsistency between the HRA implementation and the GMM model specification in versions V5 and V6 with regard to the period-to-period correlation structure. We have also noted that the lack of period-to-period correlation in the site response in V5 and V6 is a deviation from the earlier model versions. The fact that the NAM HRA implementation follows the approach of the earlier versions may well indicate that the implementation of the period-to-period correlations pre-dates the V5 model report and simply has not been updated.

Next we may ask the question whether the lack of period-to-period correlations in the site-to-site variability is intended. An alternative explanation might be that it is a mistake in the two reports. From a physical perspective, the lack of period-to-period correlation in the site response seems implausible. For example, sites with a relatively low elastic impedance are expected to show increased amplitudes for all spectral acceleration components. This seems to call for strong period-

to-period correlations. Resonances that might occur in the site response are also expected to influence period ranges with possibly positive and negative correlation coefficients. It is not clear how the correlation structure of the site response would compare to the correlation structure of the reference ground motions. In principle it is possible and probably defensible that the Baker & Jayaram correlation structure should apply at the reference (base rock) level rather than at the surface. However, assuming a vanishing correlation structure in the site response then seems unrealistic. In fact, the period-to-period correlation structure of the site response could probably be extracted from the simulation data that was used to calibrate the amplification functions.

In an analysis not reported within this memo, we have studied the effect of assuming no period-to-period correlation in the site response to risk assessment in Groningen. We found that this leads to more than 20% reduction in local personal risk for all vulnerability classes relative to the implementation according to NAM. For the vulnerability classes with the highest risk, a reduction up to 40% (i.e. close to factor of 2) is observed in the center region. An important cause for the strong reduction is that the intensity measure used in the V6 and V7 Fragility Models is defined as the average (geometric mean) of the spectral accelerations at 10 spectral periods. This averaging reduces the variability in the intensity measure relative to the variability of the contributing periods. However, this reduction effect is stronger if the variation of the ground motions for the individual periods are less correlated. Ignoring correlation in the site response therefore leads to lower variability in the intensity measure and ultimately to a lower risk assessment.

Based on our current state of information, TNO recommends to maintain the implementation by NAM and to apply the period-to-period correlations structure to all components of variability, i.e., including the site-to-site variability associated with the site response.

However, considering the ambiguity in the specification and implementation of the period-to-period correlations, as well as the sensitivity of risk metrics to this aspect of the GMM, TNO strongly recommends the discussion and resolution of this issue in the upcoming GMM workshop (November 2020) for versions V5, V6 and/or the upcoming version V7.

References

- Akkar, S., M.A. Sandikkaya & B.Ö. Ay (2014). *Compatible ground-motion prediction equations for damping scaling factors and vertical-to-horizontal spectral amplitude ratios for the broader Europe region*. Bulletin of Earthquake Engineering 12(1), 517-547
- Baker, J.W. and N. Jayaram (2008). *Correlation of Spectral Acceleration Values from NGA Ground Motion Models*, Earthquake spectra, February 2008.
- Bommer, J.J., P.J. Stafford, B. Edwards, B. Dost & M. Ntinalexis (2015a). *Development of Version 1 GMPEs for Response Spectral Accelerations and for Strong-Motion Durations*, Version 2, 21 June 2015, 294 pp.
- Bommer, J.J., B. Dost, B. Edwards, P.P. Kruiver, P. Meijers, M. Ntinalexis, B. Polidoro, A. Rodriguez-Marek & P.J. Stafford (2015b). *Development of Version 2 GMPEs for response spectral accelerations and significant durations from induced earthquakes in the Groningen field*. Version 2, 29 October 2015, 515 pp.
- Bommer, J.J., B. Dost, B. Edwards, P.P. Kruiver, P. Meijer, M. Ntinalexis, A. Rodriguez-Marek, E. Ruigrok, J. Spetzler & P.J. Stafford (2017). *V4 Ground-Motion Model (GMM) for Response Spectral Accelerations, Peak Ground Velocity, and Significant Durations in the Groningen Field*. Version 2, 3 June 2017, 525 pp.
- Bommer, J.J., B. Dost, B. Edwards, P.P. Kruiver, P. Meijer, M. Ntinalexis, A. Rodriguez-Marek, E. Ruigrok, J. Spetzler & P.J. Stafford (2018). *V5 Ground-Motion Model for the Groningen Field. Re-issue with assurance letter*. Revision 1, 14 March 2018, 299 pp.

- Bommer, J.J., B. Edwards, P.P. Kruiver, A. Rodriguez-Marek, P.J. Stafford, B. Dost, M. Ntinalexis, E. Ruigrok & J. Spetzler (2019). *V6 Ground-Motion Model for Induced Seismicity in the Groningen Field. With Assurance Letter*. Revision 1, 19 December 2019, 193 pp.
- Bourne, S.J., S.J. Oates, R. Scheefhals, P. Castellanos Nash, A. Mar-Or, T. Storck, P. Omid, J. van Elk (2019) *A Monte Carlo method for probabilistic seismic hazard and risk analysis of induced seismicity in the Groningen gas field*. 29 May 2019, 62 pp.
- Crowley, H. and R. Pinho (2020), *Report on the v7 Fragility and Consequence Models for the Groningen Field, March 2020*, 71 pp.
- Kohrangi, M., P. Bazzurro, D. Vamvatsikos & A. Spillatura (2017). *Conditional spectrum based ground motion record selection using average spectral acceleration*. *Earthquake Engineering & Structural Dynamics* 46(10), p.1667-1685
- Lin T. and J. Baker (2015). *Conditional Spectra*. In: Beer M., Kougoumtzoglou I., Patelli E., Au IK. (eds) *Encyclopedia of Earthquake Engineering*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36197-5_386-1
- NEN (2020). NPR 9998:2018+C1:2020/A1:2020 nl: *Beoordeling van de constructieve veiligheid van een gebouw bij nieuwbouw, verbouw en afkeuren - Geïnduceerde aardbevingen - Grondslagen, belastingen en weerstanden*. NEN, 1 August 2020, 40 pp.
- TNO (2019). *Comparative analysis of the NAM and TNO implementation in the Groningen Seismic Hazard and Risk Assessment*, TNO 2019 R11997.



NAM

Critique of the TNO Report entitled “On the period-to-period correlation structure of ground motion residuals in the Groningen GMM” from 29th October 2020

Peter Stafford

Editors Jan van Elk

Date January 2020

Contents

Overview	5
Comments on “Experiment on effective correlation structure at the surface”	6
Background to the Baker & Jayaram (2008) correlation model	7
Details of the Groningen GMM.....	11
Comments on “Period-to-period correlation in Fragility Model”	15
Comments on “Discussion”	17
Closing Remarks	19
References	21
Appendix A - Correlation Structure of Groningen Ground Motions Resolution of confusion around implementation in the HRA	23
Presentation by Peter J Stafford, Workshop on GMM with TNO, 5th November 2020.....	23

Overview

The present document provides a critical review of the above report relating to the period-to-period correlation structure of the Groningen GMM. The above report is referred to as the TNO T2T Report hereafter. The TNO T2T Report discusses a difference between the implementation specifications in Bommer *et al.* (2018, 2019) and the actual implementation in the NAM HRA, as described by Bourne *et al.* (2019). This difference was highlighted in 2019 and was addressed by NAM so that future risk estimates (using the V7 GMM, onwards) are consistent with the guidance in the Bommer *et al.* (2018, 2019) reports (and their corresponding update for the V7 GMM). However, the principal recommendation of the TNO T2T Report is that future implementations should revert to the inconsistent implementation previously adopted. The TNO T2T Reports attempts to justify this recommendation.

The TNO T2T Report draws its conclusions after assuming that an objective of the Groningen GMM is to reproduce the correlation structure implied by the Baker & Jayaram (2008) correlation model for surface spectral ordinates over the Groningen Gas field. That assumption is not correct, primarily, but not exclusively, because the Baker & Jayaram (2008) model is a generic ergodic model while the Groningen GMM is partially non-ergodic. The non-ergodic aspects of the Groningen GMM relate to the treatment of site effects, and the TNO T2T Report recommends that the generic ergodic correlation model of Baker & Jayaram (2008) be applied to this non-ergodic component of the Groningen GMM. There are a number of reasons why this is not appropriate, the most important of these are outlined in the present report.

The TNO T2T Report is brief and consists of four main sections. The first main section discusses the specification of the period-to-period correlation in the Groningen GMM, going all the way back to V1. There is little merit in discussing this section of the report as it has already been established that the explicit step-by-step guidance in the V6 report (Bommer *et al.*, 2019) should be followed. The second section presents an 'experiment on effective correlation structure at the surface', and specific comments related to this section are provided in what follows. The third section then moves to consider the 'period-to-period correlation in Fragility Model', and some brief comments are provided in response to this section later in the present report. The final section of the TNO T2T Report presents the 'Discussion' and raises some points that will also be responded to at the end of the present report.

Comments on “Experiment on effective correlation structure at the surface”

The TNO T2T Report presents an ‘experiment’ looking at the impact of making different assumptions regarding the period-to-period correlations within the site-to-site variability. They begin by making the assumption that the Baker & Jayaram (2008) correlation model is the ‘correct’ target that we should be striving to reproduce, and then consider the effect of using this correlation model for between-event, between-component, within-event and site-to-site variability (‘all’), not considering correlation within the site-to-site variability (‘no p-p’), and considering perfect correlation within the site-to-site variability (‘full p-p’) – with ‘all’, ‘no p-p’ and ‘full p-p’ corresponding to the labels in Figure 2 of the TNO T2T Report.

The results presented in Figures 2 and 3 of the TNO T2T Report are based upon 10,000 Monte Carlo simulation realizations, drawing from a known correlation matrix. However, the fact that the correlation matrix is known a priori means the same results can be obtained much more directly.

The recommendation of the TNO T2T Report is based extremely heavily upon the assumption that the ergodic Baker & Jayaram (2008) correlation model should be recovered when generating spectral ordinates at the ground surface in the partially non-ergodic Groningen GMM. They assert that not considering period-to-period correlation in the site effects leads to incorrect correlation structure in the Groningen GMM – because we don’t match the Baker & Jayaram (2008) correlations. The primary issue that undermines the recommendations of the TNO T2T Report is that the Baker & Jayaram (2008) correlation model is not the appropriate target for the Groningen GMM. The key conceptual barrier that does not appear to have been appreciated by the TNO T2T Report authors has already been alluded to above, namely that the Baker & Jayaram (2008) model is an ergodic correlation model, while the Groningen GMM is a partially non-ergodic model. Furthermore, the Groningen GMM is developed to provide inputs to a HRA that works with portfolios of buildings that are spatially aggregated, and where site response functions are specified for spatial site amplification zones. These features of the Groningen GMM also lead to reasons why the Baker & Jayaram (2008) correlation model is not the appropriate target.

It must be emphasised that, once one accepts that we expect to observe differences between the Groningen GMM correlation model and the Baker & Jayaram (2008) model, the investigations within the TNO T2T Report do not serve to help determine whether or not the treatment of period-to-period correlation within the Groningen GMM is appropriate or otherwise.

Background to the Baker & Jayaram (2008) correlation model

The logarithmic spectral amplitude, $\ln Sa(T; \mathbf{x})$, for a particular period, T , and at a given location, \mathbf{x} , can be expressed in terms of a mean model prediction and apparently random deviations away from this prediction. For a long time, it has been traditional within GMM development to use a ‘random effects’ regression approach (Brillinger & Preissler, 1984; 1985, Abrahamson & Youngs, 1992) where the total difference (residual) between a logarithmic observed spectral acceleration and the associated prediction is partitioned into two components. In the equation below, $\mu_{\ln Sa(T)}(rup, \mathbf{x})$ represents the predicted mean logarithmic spectral acceleration for a rupture scenario rup at a particular location \mathbf{x} . The two residual components are δB and $\delta W'(\mathbf{x})$, and these are assumed to be independent of each other. The δB represents a systematic earthquake effect (or a between event residual) that influences the observations at all sites for a given rupture scenario – it is therefore independent of location \mathbf{x} . The remaining term, $\delta W'(\mathbf{x})$, is particular to a specific location and represents the apparent randomness at location \mathbf{x} for this earthquake scenario. The $\delta W'(\mathbf{x})$ is also referred to as the within-event residual.

$$\ln Sa(T; \mathbf{x}) = \mu_{\ln Sa(T)}(rup; T, \mathbf{x}) + \delta B(T) + \delta W'(T; \mathbf{x})$$

The between-event residuals are assumed to be normally distributed with a mean of zero and standard deviation of $\tau(T)$, while the within-event residuals are normally distributed with a mean of zero and a standard deviation of $\phi(T; \mathbf{x})$. As a result, the logarithmic spectral acceleration values are also normally distributed according to:

$$\ln Sa(T; \mathbf{x}) \sim N[\mu_{\ln Sa(T)}(rup; T, \mathbf{x}), \tau(T)^2 + \phi(T; \mathbf{x})^2]$$

The Baker & Jayaram (2008) correlation model is constructed using the results of GMMs that have been developed consistently with the above framework. In this framework, the covariance between two logarithmic spectral ordinates at periods T_i and T_j , at the same spatial location, \mathbf{x} , can be defined as:

$$\text{cov}[\ln Sa(T_i; \mathbf{x}), \ln Sa(T_j; \mathbf{x})] = \rho_{\ln Sa(T_i), \ln Sa(T_j)} \sigma(T_i) \sigma(T_j)$$

Therefore, in order to derive an expression for the overall correlation among spectral ordinates at two different periods, T_i and T_j , Baker & Jayaram (2008) use the expression:

$$\rho_{\ln Sa(T_i), \ln Sa(T_j)} = \frac{\rho_{\delta B(T_i), \delta B(T_j)} \tau(T_i) \tau(T_j) + \rho_{\delta W'(T_i), \delta W'(T_j)} \phi(T_i) \phi(T_j)}{\sigma(T_i) \sigma(T_j)}$$

where the numerator is equivalent to the covariance, $\text{cov}[\ln Sa(T_i; \mathbf{x}), \ln Sa(T_j; \mathbf{x})]$. However, in actual fact, the model of Baker & Jayaram (2008) is actually based upon the assumption that $\rho_{\ln Sa(T_i), \ln Sa(T_j)} \approx \rho_{\delta W'(T_i), \delta W'(T_j)}$ due to the fact that $\phi(T)$ is typically much larger than τ (Carlton & Abrahamson, 2014).

It is important to note that the Baker & Jayaram (2008) model is based upon linear correlations among within-event residuals using an ergodic ground-motion database and without taking into account any systematic site effects. Therefore, the correlation model isn't targeted for application at any given site, but rather reflects generic implicit inter-period correlations between site effects. Sites are only characterised as a function of $V_{S,30}$ and a parameter accounting for deeper velocity structure (that plays an extremely weak role for the periods of interest in the Groningen GMM). To be clear, any site-specific resonances, or differences among locations of resonant peaks for sites with the same $V_{S,30}$ values are treated as aleatory variability in the Baker & Jayaram (2008) model.

The issue discussed in the TNO T2T Report relates to a situation in which a different variance decomposition framework is adopted. For consistency with the framework presented above, and for simplicity, we can extend the expressions above to be applicable for the model below:

$$\ln Sa(T; \mathbf{x}) = \mu_{\ln Sa(T)}(\text{rup}; T, \mathbf{x}) + \delta B(T) + \delta W(T; \mathbf{x}) + \delta C(T; \mathbf{x}) + \delta S(T; \mathbf{x})$$

Here, $\mu_{\ln Sa(T)}$ still represents the mean logarithmic prediction, the between-event residual remains δB (although its numerical value will be different), but now the within-event residual from the earlier presentation, $\delta W'(\mathbf{x})$, is split into $\delta W'(\mathbf{x}) \equiv \delta W(\mathbf{x}) + \delta S(\mathbf{x})$. Another new residual component is also introduced. The term $\delta C(T; \mathbf{x})$ is the component-to-component residual – this wasn't previously considered as the Baker & Jayaram (2008) framework is based upon a type of average over the horizontal

components¹. The final term, $\delta S(T; \mathbf{x})$, represents a systematic between-site residual, *i.e.*, the systematic, repeatable effects of site response at a particular location. All of these residuals are assumed to be independent of one another, and to have zero means. The standard deviations are τ for δB , ϕ_{SS} for δW , σ_{C2C} for δC , and ϕ_{S2S} for δS^2 .

Under this extended framework, the overall correlation between logarithmic response spectral ordinates is defined by:

$$\rho_{\ln Sa_i, \ln Sa_j} = \frac{\rho_{\delta B_i, \delta B_j} \tau_i \tau_j + \rho_{\delta W_i, \delta W_j} \phi_{SS,i} \phi_{SS,j} + \rho_{\delta C_i, \delta C_j} \sigma_{C2C,i} \sigma_{C2C,j} + \rho_{\delta S_i, \delta S_j} \phi_{S2S,i} \phi_{S2S,j}}{\sigma_i \sigma_j}$$

In the above formulation, the subscripts i and j simply represent the periods T_i and T_j and are introduced for brevity.

Using the above equation, the cases considered within the TNO T2T Report can be obtained by simply making different assumptions about the $\rho_{\delta S_i, \delta S_j}$ term. As the TNO T2T Report assumes that the other correlation terms are always the same, a simplified form of the above expression is:

$$\rho_{\ln Sa_i, \ln Sa_j} = \frac{\rho_{\delta NSB_i, \delta NSB_j} \sigma_{NSB,i} \sigma_{NSB,j} + \rho_{\delta S_i, \delta S_j} \phi_{S2S,i} \phi_{S2S,j}}{\sigma_{SUR,i} \sigma_{SUR,j}}$$

where the first three terms of the numerator of the more elaborate expression are all compressed into a single term that represents the covariance of residuals at the NSB horizon, $\rho_{\delta NSB_i, \delta NSB_j} \sigma_{NSB,i} \sigma_{NSB,j}$. The TNO T2T Report considered the cases:

- ‘all’: $\rho_{\delta S_i, \delta S_j} = \rho_{\delta NSB_i, \delta NSB_j}$;
- ‘no p-p’: $\rho_{\delta S_i, \delta S_j} = 0$; and,
- ‘full p-p’: $\rho_{\delta S_i, \delta S_j} = 1$.

¹ Similarly, if the fragility functions in the NAM HRA engine are based upon the geometric mean of the two horizontal components then the component-to-component variability would not be included.

² It is important to note that while the symbol ϕ_{S2S} is also employed within the Groningen GMM, the interpretation of this term is different here. In the current text, the ϕ_{S2S} follows the traditional convention where it represents site-to-site aleatory variability of systematic site terms – with these site terms corresponding to a unique spatial location. Within the Groningen GMM, ϕ_{S2S} actually represents epistemic uncertainty associated with elements of the site response model, and spatial variability of site response over site zones.

The standard deviations shown above are obtained from $\sigma_{NSB}^2 = \tau^2 + \phi_{S2S}^2 + \sigma_{C2C}^2$ and $\sigma_{SUR}^2 = \sigma_{NSB}^2 + \phi_{S2S}^2$. Hence, it is clear that the 'all' case should recover the original Baker & Jayaram (2008) correlation structure – if that is what is assumed as being the underlying model. Furthermore, it is clear that 'no p-p' will give lower correlations than the 'all' case, and that 'full p-p' will give higher correlations. The denominator doesn't change in the above cases, we simply obtain different values for the numerator.

This is all consistent with the results of the 'experiment' presented in the TNO T2T Report. However, the conclusions and recommendations drawn from the TNO T2T Report investigations are all predicated upon the assumption that the Baker & Jayaram (2008) model represents the target that the NAM HRA engine is trying to replicate. There are a number of reasons why this is not a sound assumption.

As the Groningen GMM does not invoke the Baker & Jayaram (2008) correlation model for $\rho_{\delta S_i, \delta S_j}$, but rather specifies that $\rho_{\delta S_i, \delta S_j} = 0$, the TNO T2T Report infers that inter-period correlations between spectral accelerations at the surface are too low, and that this leads to lower risk. As noted above, this conclusion is based upon the (incorrect) assumption that the correct risk result is obtained using the ergodic Baker & Jayaram (2008) correlation model throughout.

The following section provides more specific details about the characteristics of the Groningen GMM in order to emphasise the non-ergodic nature of the model and demonstrate why the Baker & Jayaram (2008) model is not adopted for the site response.

Details of the Groningen GMM

The formulations presented in the previous section are generic and serve to demonstrate the underlying framework adopted by Baker & Jayaram (2008), as well as the conceptual differences that arise when moving to explicitly consider systematic site effects. However, the actual framework implemented for the Groningen GMM is different to that presented above, and this does not appear to be appreciated from the TNO T2T Report.

For the Groningen GMM the logarithmic spectral ordinates are obtained by first making a prediction at the NS-B reference horizon, and then combining this prediction with site-specific (really, zone-specific) site response functions.

A particular level of spectral acceleration at the NS-B horizon is represented in the model as:

$$\ln Sa_{NSB}(T; \mathbf{x}) = \mu_{\ln Sa_{NSB}}(rup; T, \mathbf{x}) + \delta B(T) + \delta W_{NSB}(T; \mathbf{x}) + \delta C(T; \mathbf{x})$$

The subsequent level of spectral acceleration at the surface is obtained from:

$$\ln Sa_{SUR}(T; \mathbf{x}) = \ln Sa_{NSB}(rup; T, \mathbf{x}) + \mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x}) + \varepsilon_{SUR}(T; \mathbf{x})$$

where $\mu_{\ln AF}$ is the mean of the logarithmic site amplification for the zone containing location \mathbf{x} , and $\varepsilon_{SUR}(T; \mathbf{x})$ is the apparent randomness in the site response. Note that this mean site amplification is a function of the spectral acceleration level (the actual level, not the predicted level) at the NS-B horizon.

It is essential to recognise that $\mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x})$ is a location-specific (zone-specific within the NAM HRA) site amplification function, and that it describes the systematic site effects assumed to be relevant for motions propagating from NS-B to the surface in the relevant zone. This formulation does not include a term $\delta S(T; \mathbf{x})$ as we had in the previous generic formulation.

If we wished to explicitly include a $\delta S(T; \mathbf{x})$ term, then we could represent the zone-specific site amplification function as:

$$\mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x}) \equiv \mu_{\ln AF}(Sa_{NSB}; T) + \delta S(T; \mathbf{x})$$

where $\mu_{\ln AF}(Sa_{NSB}; T)$ represents some generic average amplification function (potentially the average site response over the entire Groningen field), and $\delta S(T; \mathbf{x})$ represents the deviation away from this generic average site response to account for the systematic effects specific to the zone hosting \mathbf{x} . However, whereas the $\delta S(T; \mathbf{x})$ terms in the previous formulations were treated as random variates, in the above formulation, where we are working in a partially non-ergodic framework, it is assumed that we know the value of $\delta S(T; \mathbf{x})$ for each zone. Of course, we do not know this site-specific deviation perfectly. So, the role of the ϕ_{S2S} terms within the Groningen GMM is to reflect the epistemic uncertainty in the $\delta S(T; \mathbf{x})$ values for each zone. It is for this reason that these ϕ_{S2S} values vary with zone, and depend upon the spatial heterogeneity and soil characteristics of the individual zones.

In this context, it is clear that there are two components to the site effects modelled within the Groningen GMM: the systematic effects associated with each site zone, $\mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x})$, and the apparent randomness, $\varepsilon_{SUR}(T; \mathbf{x})$. Crucially, whereas the previously presented framework considered (and the framework adopted by Baker & Jayaram, 2008) the $\delta S(T; \mathbf{x})$ terms as being unknown random variables within an ergodic approach, the partially non-ergodic approach followed in the Groningen GMM dictates that the bulk of the systematic site effects are assumed known.

To emphasise this further, consider again the representation where we make the distinction between the location-dependent $\mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x})$ and the location-independent $\mu_{\ln AF}(Sa_{NSB}; T)$. If we view $\delta S(T; \mathbf{x})$ as:

$$\delta S(T; \mathbf{x}) = \mu_{\ln AF}(Sa_{NSB}; T, \mathbf{x}) - \mu_{\ln AF}(Sa_{NSB}; T)$$

then, the $\delta S(T; \mathbf{x})$ values at periods T_i and T_j will have a relationship that depends upon the underlying site transfer function. To appreciate why this is the case, recall from Random Vibration Theory (RVT) that spectral amplitudes can be written as being proportional to the integral of the squared Fourier amplitude spectrum (FAS), *i.e.*, the zeroth spectral moment (see, *e.g.*, Stafford *et al.*, 2017):

$$Sa(T; \mathbf{x}) \propto \sqrt{m_0(T, \zeta; \mathbf{x})}$$

where,

$$m_0(T, \zeta; \mathbf{x}) = 2 \int_0^{\infty} |Y(f; T, \zeta, \mathbf{x})|^2 df$$

and $|Y(f; T, \zeta, \mathbf{x})|$ is the Fourier amplitude spectrum of the response of a single degree-of-freedom (SDOF) system with period T and damping ratio of ζ . This Fourier amplitude spectrum can be decomposed into the product of the FAS at the NS-B horizon, $|A_{NSB}(f; \mathbf{x})|$, the site transfer function³, $|S(f; \mathbf{x})|$, and the transfer function of the SDOF oscillator, $|H(f; T, \zeta)|$. That is, we can write:

$$|Y(f; T, \zeta, \mathbf{x})| \equiv |A_{NSB}(f; \mathbf{x})| |S(f; \mathbf{x})| |H(f; T, \zeta)|$$

The transfer function of the SDOF acts as a narrow-band filter with a bandwidth that depends upon the damping ratio. This transfer function will be the same for all sites and all earthquake scenarios. The site transfer function will be specific to a site/zone and reflects the particular layering and impedance contrasts at that location (Kramer, 1996). One could interpret the position-independent average amplification, $\mu_{ln AF}(Sa_{NSB}; T)$, as reflecting the average effects of the $|S(f; \mathbf{x})|$ terms over the entire Groningen field. However, the position-dependent amplification, $\mu_{ln AF}(Sa_{NSB}; T, \mathbf{x})$, will only reflect the location specific transfer function $|S(f; \mathbf{x})|$.

For the above reason, the $\delta S(T; \mathbf{x})$ that are implicit within the $\mu_{ln AF}(Sa_{NSB}; T, \mathbf{x})$ terms of the Groningen GMM are not random variates that should be sampled from some covariance matrix with an ergodic correlation structure. Rather, the particular layering and impedance contrasts that lead to the site-specific $|S(f; \mathbf{x})|$ determine the relationship among response spectral site effects.

In light of the above considerations, it is not correct to assert, as the TNO T2T Report does, that the Groningen GMM does not include period-to-period correlation in the site response. There is certainly inter-period correlation. However, this correlation is deterministically imposed through the site transfer functions and should be regarded as ‘correlation in the means’, *i.e.*, deterministic correlation between the $\mu_{ln AF}(Sa_{NSB}; T, \mathbf{x})$ functions from period to period, rather than as being some stochastic correlation that should be reflected within a covariance matrix.

The remaining site residual ε_{SUR} appears as an aleatory variability in the Groningen GMM framework presented above. However, these values represent the random aspects of site response that are not included within the Groningen site response functions. This includes the effects of motion-to-motion variability, as well as aspects of deviations away from the 1D site response that is assumed by the numerical modelling of site response. This component is not explicitly sampled within the step-by-step specifications in the V6 Groningen GMM. However, this is due to the fact that this element of variability is embedded within the ϕ_{SS} component of the model that is imposed at the NS-B horizon.

³ Here, we focus upon a linear system, and linear site response. The concepts extend directly to the consideration of nonlinear site response, but the mathematical details differ quite considerably. The simplification to consider the linear case here still allows the relevant points to be made.

In the Groningen GMM, when we have sampled from ϕ_{S2S} in the past⁴, we are effectively recognising our imperfect knowledge of $\delta S(T; \mathbf{x})$, and the manner in which this varies over a given site zone.

The present section has covered material that is not broadly appreciated within the engineering seismology community, and it is understandable that it may not have been fully appreciated by the authors of the TNO T2T Report. However, in light of the above presentation, it should be clear why the ergodic Baker & Jayaram (2008) correlation model is not used *in addition to* the correlation in the means already incorporated within the partially nonergodic Groningen GMM.

⁴ In the V7 model, the epistemic nature of this model component is more explicitly recognised and represented within the logic tree, rather than being sampled as part of the Monte Carlo treatment of aleatory variability.

Comments on “Period-to-period correlation in Fragility Model”

This section of the TNO T2T Report starts with some generic comments regarding the nature of the average spectral acceleration and the role that inter-period correlation plays in influencing the effective variance of the average spectral acceleration. This is referred to as the ‘averaging effect’ and comes up again in the final Discussion section of the TNO T2T Report. The section then proceeds to present an exercise in which the 200 ground-motion records that were used in the calibration of the fragility curves were analysed and were found to have the same correlation structure as the Baker & Jayaram (2008) correlation model. The results of this exercise led to the conclusion that “there is a striking similarity” and that “Apparently, the hazard-consistent record selection procedure has managed to capture the desired correlation structure”. These results are presented in a manner that suggests the investigators were not anticipating such a good level of agreement. However, this seems to point to an incomplete understanding of the actual record selection procedure.

The record selection procedure adopted by Crowley & Pinho (2020) follows advanced state of practice methods and is primarily based upon the algorithm first proposed by Jayaram *et al.* (2011). It is generally a challenge to select a suite of ground-motion records that have the appropriate target covariance (which is a conditional covariance in most advanced applications). This is particularly the case if the problem is approached in a combinatorial manner, *i.e.*, attempting to deterministically check for the unique combination of motions from a database that best matches the target conditional covariance matrix (and conditional mean). Indeed, earlier attempts to solve this problem were hampered by the computational demands of such a combinatorial approach as the databases from which to select ground motions from increased in size. For example, to exhaustively search for the best 20 ground motions from a database of 3000 candidate records, you would need to consider 1.345×10^{51} combinations – and Crowley & Pinho (2020) sought 200 motions from a larger database. The breakthrough presented in Jayaram *et al.* (2011) made this problem far more numerically efficient.

The Jayaram *et al.* (2011) method, which has subsequently been slightly refined by researchers from the same group to the form adopted by Crowley & Pinho (2020), starts by defining the target hazard-consistent distribution. Samples are then drawn from this conditional target distribution. For example, if one wished to select 200 records, as done by Crowley & Pinho (2020), then one would draw 200 samples from the target conditional distribution. Provided that the numerical method used for the sampling is robust, the statistics of this sample will very closely match those of the target distribution. The record selection then proceeds by searching the database of potential ground motions and identifying the individual record that best matches each of the 200 realizations. If a good agreement can be obtained between an individual candidate record and a realization drawn from the target distribution, and if this level of agreement is obtained for all 200 realizations, then the statistics of the selected suite of records will closely match those of the 200 realizations. And, as a direct consequence, the statistics of the selected suite of records will also match those of the target distribution.

Note that the above procedure is the first, and most fundamental, step of the record selection procedure, and subsequent optimizations can be applied to enhance the match of the selected records to the target. However, the key point to emphasise here is that the record selection procedure adopted by Crowley & Pinho (2020) is explicitly designed to ensure that a good match to the target distribution is obtained. As the target distribution is defined using the Baker & Jayaram (2008), it is therefore not at all surprising to find that the selected records have a correlation structure that is very similar to the predictions of this same model.

The exercise presented within the TNO T2T Report in this section is therefore simply a confirmation that the algorithm of Jayaram *et al.* (2011) works as intended. Given that this algorithm has been in use for the best part of a decade and that its adoption is essentially standard practice in earthquake engineering, there are no surprises in this section of the TNO T2T Report.

The only point worth emphasising with respect to this section is whether the Baker & Jayaram (2008) correlation model is an appropriate model for defining the target distribution for the fragility derivation. In the previous sections of the present report, we have highlighted the fact that the Baker & Jayaram (2008) model is not the appropriate target within the partially non-ergodic Groningen GMM. However, the Baker & Jayaram (2008) correlation model is appropriate to use in the fragility curve derivation because the fragility curves are not developed to be specific to individual site zones. Single curves for building classes are applied to all relevant buildings throughout the field, regardless of which site zone they are located within. It is therefore appropriate to use an ergodic correlation model for the purpose of deriving the fragility curves. This does not create an inconsistency. If the partially non-ergodic approach was also applied within the fragility model, then we would require bespoke fragility curves for each building class in each site zone. This is practically prohibitive and would likely lead to minimal impact upon the final results anyway.

Comments on “Discussion”

The final section of the TNO T2T Report discusses the previous set of results. The discussion raises questions and proposes potential explanations for the inconsistencies between the old NAM HRA implementation described in Bourne *et al.* (2019) and the implementation prescription in Bommer *et al.* (2018,2019). The first question raised is whether “the lack of period-to-period correlations in the site-to-site variability is intended”. As should be clear by now, the Groningen GMM does include period-to-period correlation, but it is deterministic correlation in the means consistent with the partially non-ergodic approach adopted, rather than the stochastic correlation that the TNO T2T Report advocates. It is also suggested that “an alternative explanation might be that it is a mistake in the two reports”. We can confirm, again, that the step-by-step guidance regarding the sampling of variability and the use of correlation models is as intended by the authors of the report. It is apparent from the differences in implementation, and the discussions and reports that have ensued, that the wording of that section can be made more explicit/verbose in the V7 report, and it is our intention to do this.

The discussion section then moves to explain why the TNO authors expect that correlation in site effects should be included. What they loosely describe with respect to elastic impedance and resonances is what has been explained in this report as being ‘correlation in the means’ – and, again, this form of correlation is included in the Groningen GMM. Presumably the authors of the TNO T2T Report would agree that these impedance and resonant site effects at a given site do not randomly change in a stochastic manner. If a site has a particular impedance signature that leads to certain spectral ordinates being consistently higher/lower than others, then this will not change with each earthquake scenario. This is precisely the effect that is already included within the Groningen GMM, and that would not be achieved through the stochastic approach being advocated by the TNO T2T Report.

The TNO T2T Report then reports on the results of some exploratory calculations that have been conducted to test the impact of either modelling the correlation in site response using the Baker & Jayaram (2008) model, or following the guidance of Bommer *et al.* (2018,2019). They report that lower risk is obtained in the latter case. They argue that an important contributor to this difference is the ‘averaging effect’ of using average spectral acceleration. It is true that a lower correlation results in a lower variance for the average spectral acceleration, as can be seen from the equation below:

$$\sigma_{\ln Sa_{avg}}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho_{\ln Sa(T_i), \ln Sa(T_j)} \sigma_{\ln Sa(T_i)} \sigma_{\ln Sa(T_j)}$$

However, this does not imply that the NAM HRA is therefore underestimating the true Local Personal Risk. As has been discussed throughout this report, correlation in site effects is considered within the NAM HRA through the partially non-ergodic treatment of site effects within the Groningen GMM. Furthermore, the

site variability that the TNO T2T Report is imposing correlation structure upon is actually epistemic uncertainty within the Groningen GMM rather than being a component of aleatory variability. The differences in risk results discussed in the TNO T2T Report should therefore simply be interpreted as being the result of following the implementation guidance in Bommer *et al.* (2018,2019) versus performing calculations in an inappropriate manner. Such investigations should not be used as the basis for making recommendations regarding the future approach to modelling ground-motion correlation for the hazard and risk analyses in the Groningen field.

Closing Remarks

This critique of the TNO T2T Report is longer than the TNO T2T Report itself, but the elaborate explanations that have been provided herein should hopefully resolve this issue. To summarise, the step-by-step implementation prescriptions in Bommer *et al.* (2018, 2019), and the relevant update to be released for the V7 model are not inadvertently missing a component of the correlation structure. The NAM HRA model code has been updated to be consistent with those prescriptions, and future calculations implementing the V7 model should also follow the relevant prescriptions. The supporting text in the V7 GMM report will be made more explicit in the hope of eliminating any potential points of confusion. It is likely that the explicit treatment of ϕ_{S2S} as an epistemic uncertainty in the V7 model will help in this regard as it will more clearly distinguish between the issues of the deterministic ‘correlation in the means’ and the stochastic elements of the aleatory variability.

Finally, it is also worth noting that there are additional reasons for desiring a lower correlation than implied by using the Baker & Jayaram (2008) model, and these primarily relate to the consideration of spatial issues within the NAM HRA. Further information on these points is provided in the slide pack that accompanies this report as an appendix.

References

1. Abrahamson, NA & R Youngs (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America*, **82**(1), 505-510.
2. Baker, JW & N Jayaram (2008). Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra*, **24**(1), 299-317.
3. Bommer, JJ, B Dost, B Edwards, PP Kruiver, P Meijer, M Ntinalexis, A Rodriguez-Marek, E Ruigrok, J Spetzler & PJ Stafford (2018). *V5 Ground-Motion Model for the Groningen Field. Re-issue with assurance letter*. Revision 1, 14 March 2018, 299pp.
4. Bommer, JJ, B Edwards, PP Kruiver, A Rodriguez-Marek, PJ Stafford, B Dost, M Ntinalexis, E Ruigrok & J Spetzler (2019). *V6 Ground-Motion Model for Induced Seismicity in the Groningen Field. With assurance letter*. Revision 1, 19 December 2019, 193pp.
5. Bourne, SJ, SJ Oates, R Scheefhals, P Castellanos Nash, A Mar-Or, T Storck, P Omid & J van Elk (2019). A Monte Carlo method for probabilistic seismic hazard and risk analysis of induced seismicity in the Groningen gas field. 29 May 2019, 62pp.
6. Brillinger, DR & HK Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data. *Bulletin of the Seismological Society of America*, **74**(4), 1441-1450.
7. Brillinger, DR & HK Preisler (1985). Further analysis of the Joyner-Boore attenuation data. *Bulletin of the Seismological Society of America*, **75**(2), 611-614.
8. Carlton, B & NA Abrahamson (2014). Issues and approaches for implementing conditional mean spectra in practice. *Bulletin of the Seismological Society of America*, **104**(1), 503-512.

9. Crowley, H & R Pinho (2020). Report on the V7 Fragility and Consequence Models for the Groningen Field. March 2020, 71pp.

10. Jayaram, N, T Lin & JW Baker (2011). A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. *Earthquake Spectra*, **27**(3), 797-815.

11. Kramer, SL (1996). *Geotechnical Earthquake Engineering*. Prentice Hall.

12. Stafford, PJ, A Rodriguez-Marek, B Edwards, PP Kruiver & JJ Bommer (2017). Scenario dependence of linear site-effect factors for short-period response spectral ordinates. *Bulletin of the Seismological Society of America*, **107**(6), 2859-2872.

Appendix A - Correlation Structure of Groningen Ground Motions Resolution of confusion around implementation in the HRA

Presentation by Peter J Stafford, Workshop on GMM with TNO, 5th November 2020

Correlation Structure of Groningen Ground Motions

Resolution of confusion around implementation in the HRA

Peter J Stafford, 5th November 2020

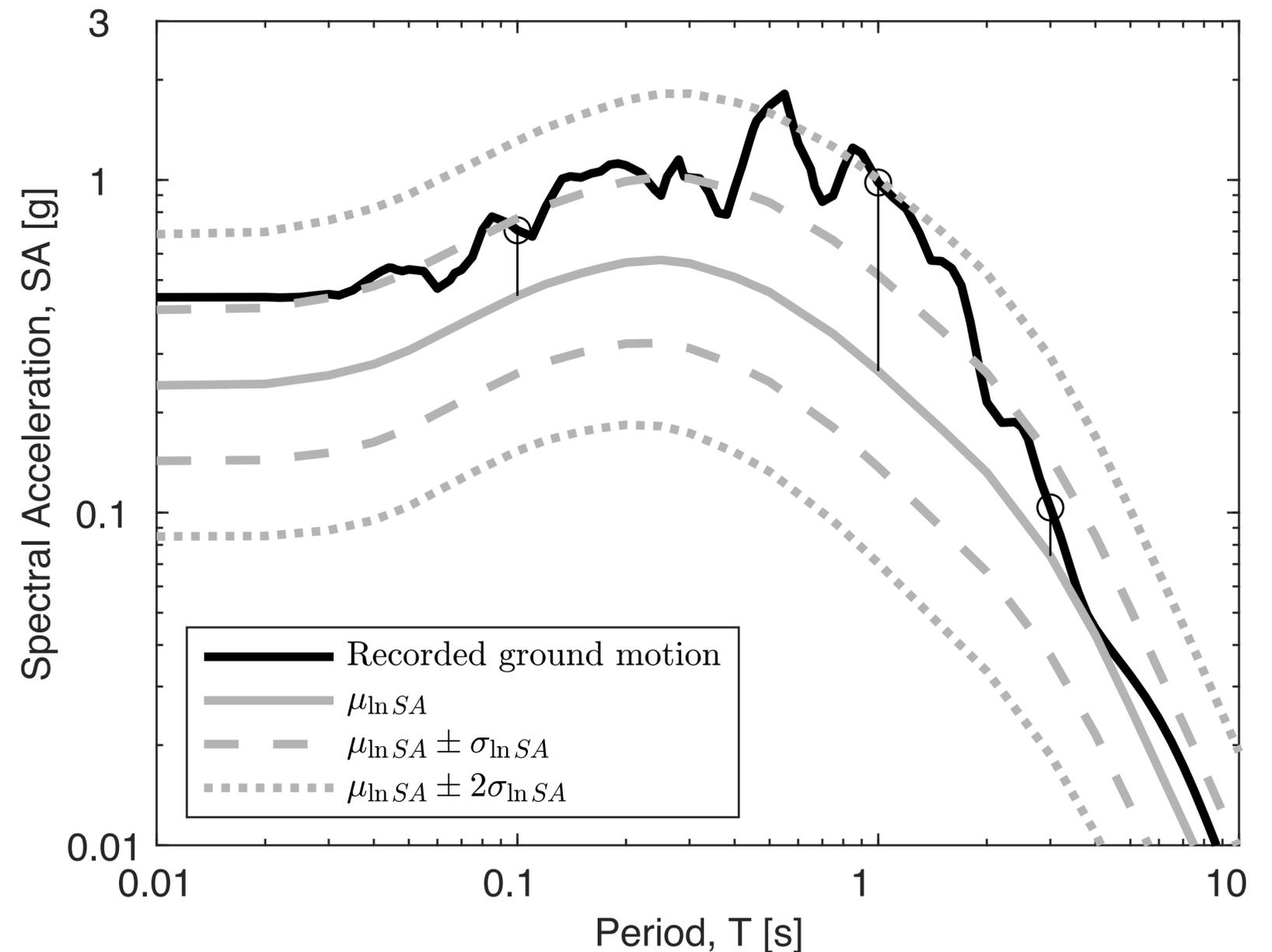
Outline

- What is inter-period correlation
- How inter-period correlation models are derived
- Requirements for GMMs for risk analysis
- Implications for correlation models
- Role of correlation models in ground-motion selection

What is Inter-period Correlation

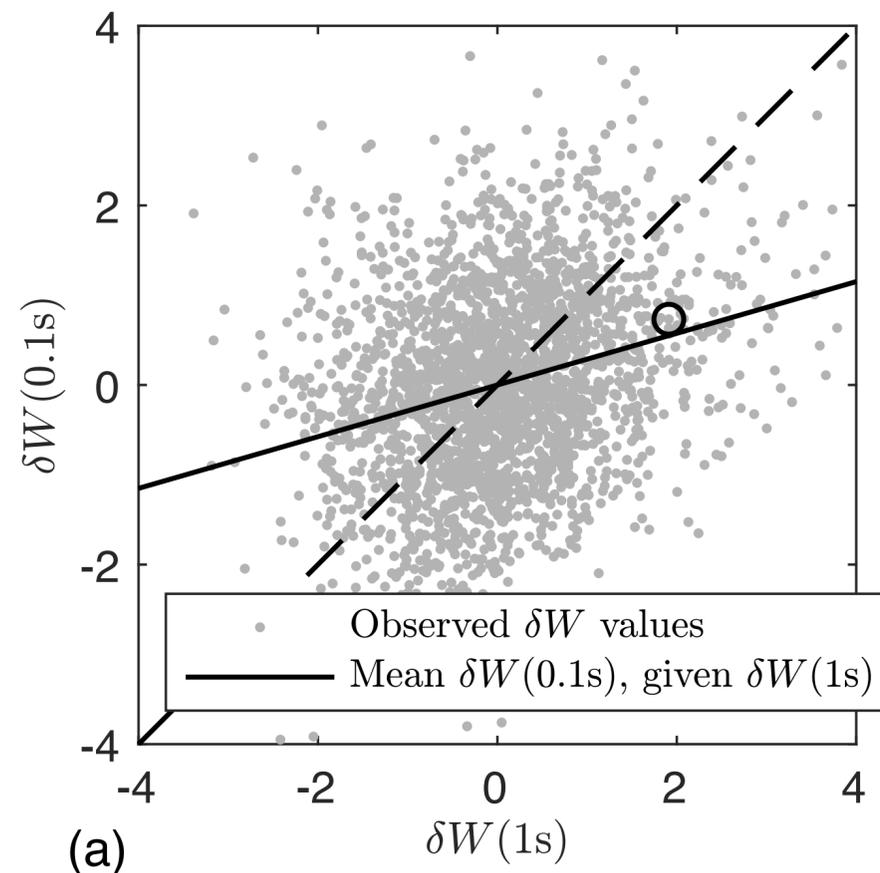
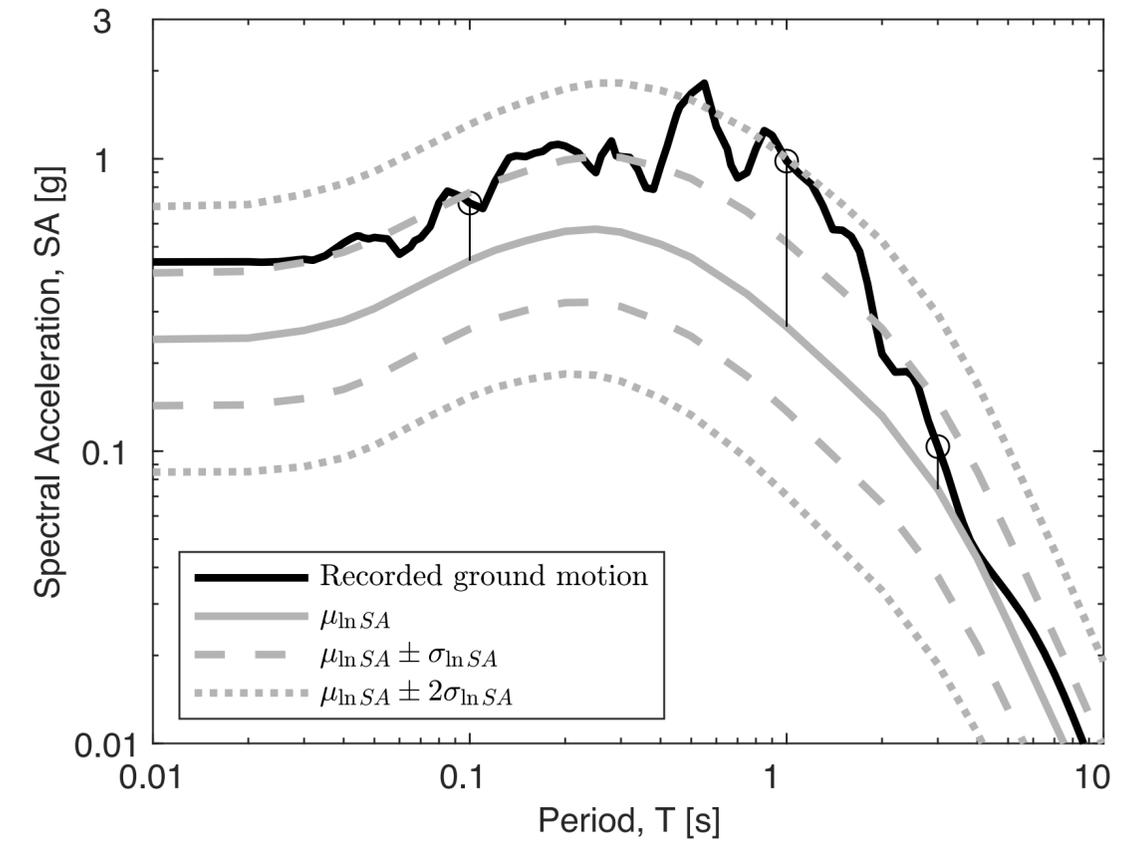
What is inter-period correlation

- An individual response spectrum exhibits spectral variability (or ‘peak-to-trough’ variability)
- Nearby response spectral ordinates tend to be similarly above/below the expected level from a ground-motion model (GMM)

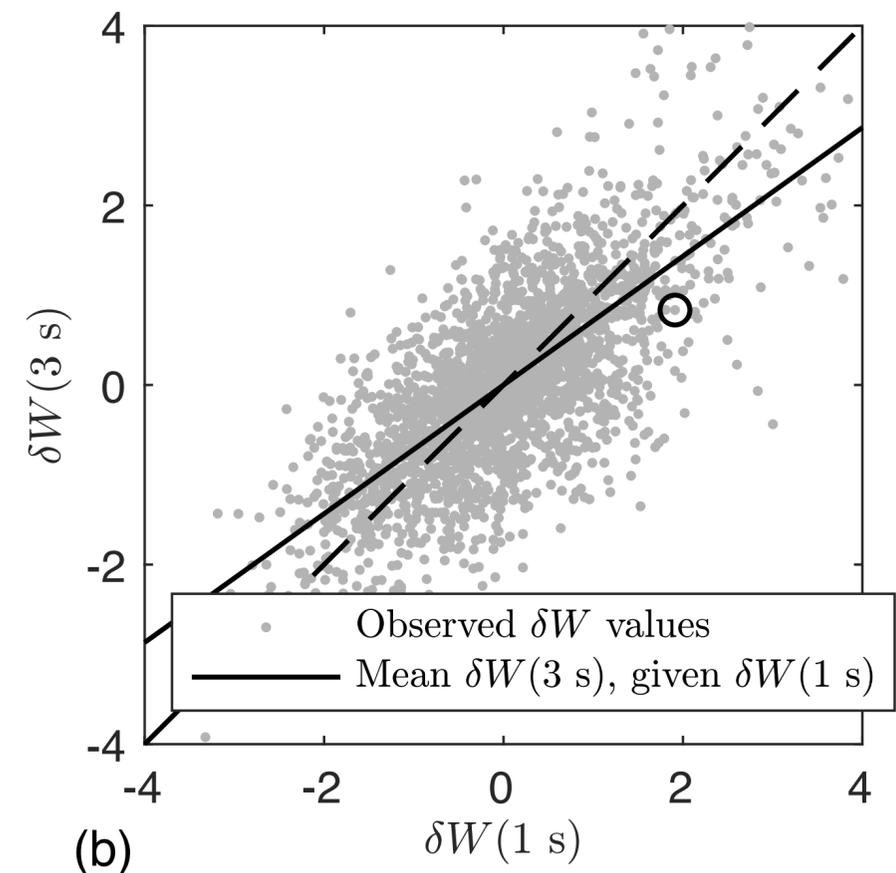


Inter-period correlation

- Collecting many response spectra from different ground motions allows for the computation of correlations between spectral ordinates at different periods
- In this particular figure, scatter plots of within-event residuals are shown



(a)



(b)

Deriving Inter-period Correlation Models

Deriving inter-period correlation models

- A GMM for response spectral ordinates is often represented as:

$$\ln Sa(T)_{ij} = \underbrace{\mu_{\ln Sa(T)}(rup, site)}_{\text{Mean}} + \underbrace{\delta B_i}_{\text{Residuals}} + \underbrace{\delta W_{ij}}_{\text{Residuals}}$$

- The dependence between two response spectral ordinates comes from:
 - Functional dependence (correlation in the means)
 - Stochastic dependence (correlation in the residuals)
- Residuals are computed from many recordings from many earthquakes (using ergodic ground-motion databases)

Deriving inter-period correlation models

- Correlations between the different types of residuals are obtained using Pearson's linear correlation
- The overall correlation is computed as a weighted combination of the correlations of the between and within-event residuals

$$\rho_{\ln Sa(T_i), \ln Sa(T_j)} = \frac{\rho_{\delta B(T_i), \delta B(T_j)} \tau(T_i) \tau(T_j) + \rho_{\delta W(T_i), \delta W(T_j)} \phi(T_i) \phi(T_j)}{\sigma(T_i) \sigma(T_j)}$$

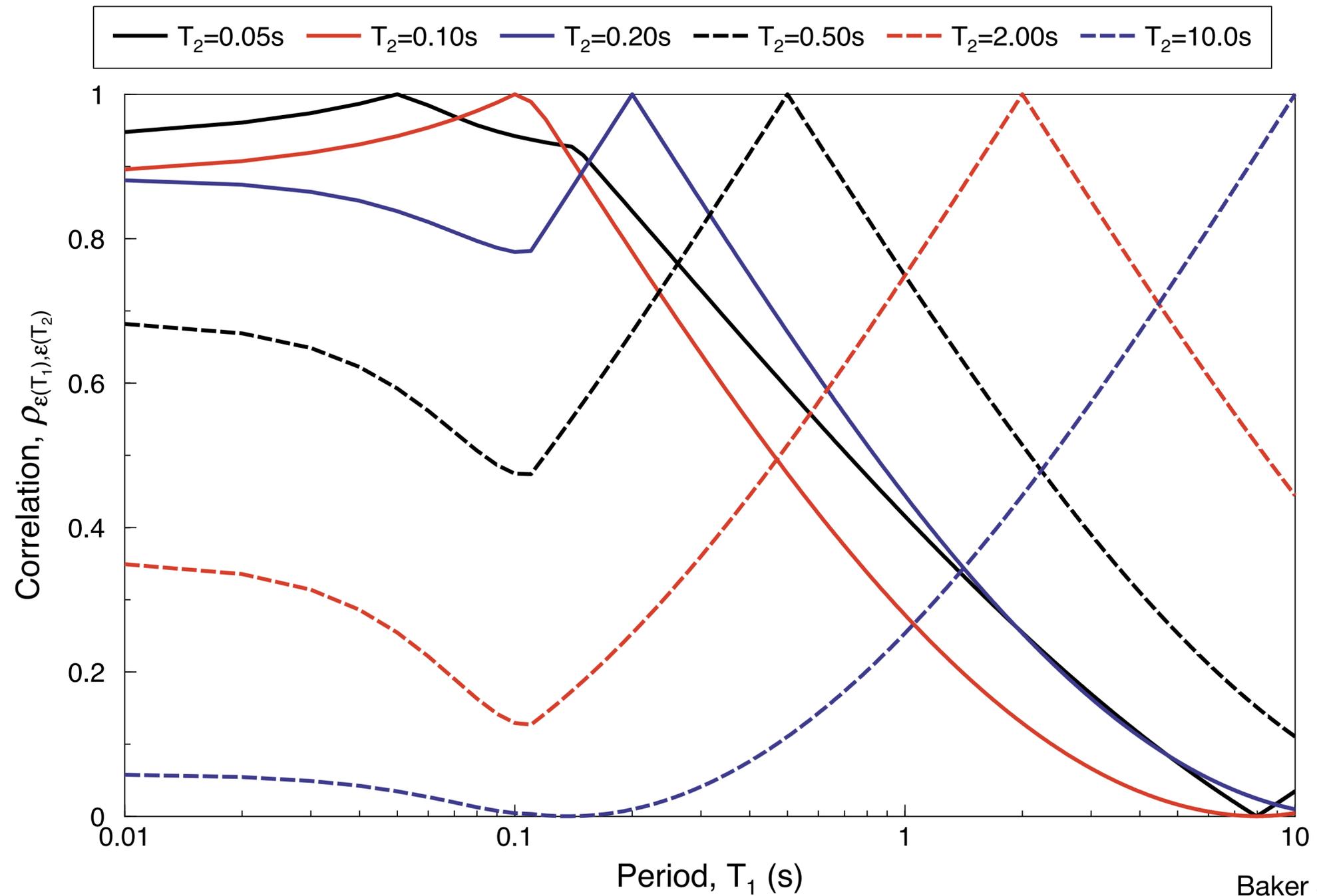
- Normally $\phi > \tau$, and so $\rho_{\ln Sa(T_i), \ln Sa(T_j)} \approx \rho_{\delta W(T_i), \delta W(T_j)}$

Deriving inter-period correlation models

- The actual parametric correlation models are derived by fitting a parametric equation to the empirically estimated correlations
- Correlations are bounded on $[-1, 1]$ and the standard errors of correlation estimates are a function of the correlation value
- Fitting is therefore done using a Fisher-z transformation $z = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right)$, where z is approximately normally-distributed, with standard errors of $1/\sqrt{n - 3}$ where n are the number of residual pairs used to estimate ρ

Inter-period correlations

- The Baker & Jayaram (2008) model is an example of a parametric model that has been derived in this manner



Issues that may not be appreciated

- The partitioning of residuals into between-event and within-event components is traditional, but not necessarily 'correct'
- An alternative approach would be to use:

$$\ln Sa(T)_{ij} = \mu_{\ln Sa(T)}(rup, site) + \delta B_i + \delta W_{ij} + \delta S2S_j$$

- The correlation in this case is:

$$\rho_{\ln Sa(T_i), \ln Sa(T_j)} = \frac{\rho_{\delta B(T_i), \delta B(T_j)} \tau(T_i) \tau(T_j) + \rho_{\delta W(T_i), \delta W(T_j)} \phi(T_i) \phi(T_j) + \rho_{\delta S2S(T_i), \delta S2S(T_j)} \phi_{S2S}(T_i) \phi_{S2S}(T_j)}{\sigma(T_i) \sigma(T_j)}$$

- This framework assumes that dependence among systematic site residuals is stochastic
- The δW terms contain effects that are related to both earthquake source, and site response. The δB and $\delta S2S$ terms only represent repeatable systematic deviations
- A site-specific ground-motion model would really be (and note δW still contains some random site effects):

$$\ln Sa(T)_{ij} = \mu'_{\ln Sa(T)}(rup, site) + \delta B_i + \delta W_{ij}$$

$$\mu'_{\ln Sa(T)}(rup, site) \equiv \mu_{\ln Sa(T)}(rup, site) + \delta S2S_j$$

- For any sites influenced by nonlinear site response, GMMs define NL response based upon median predictions of input motions

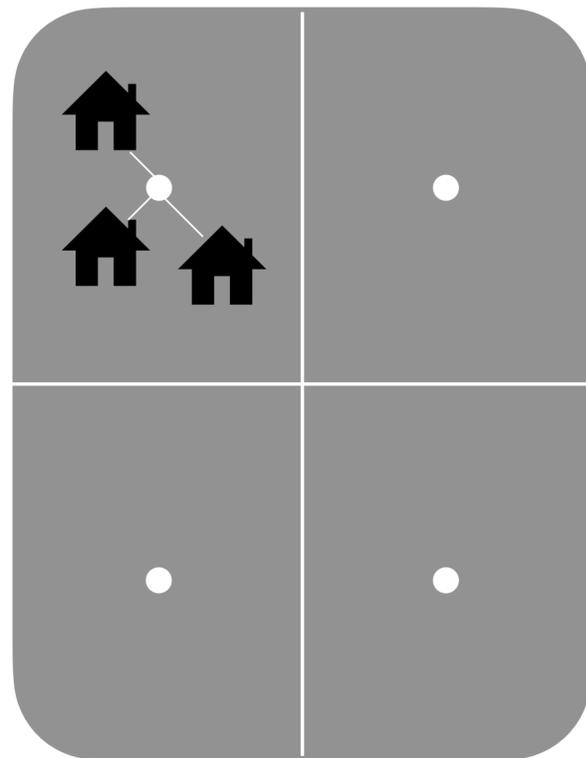
Requirements for a GMM in Risk Analysis

Correlation cases

For a general risk model there are various correlations among IMs that need to be considered

'Same'
buildings,
'same' position

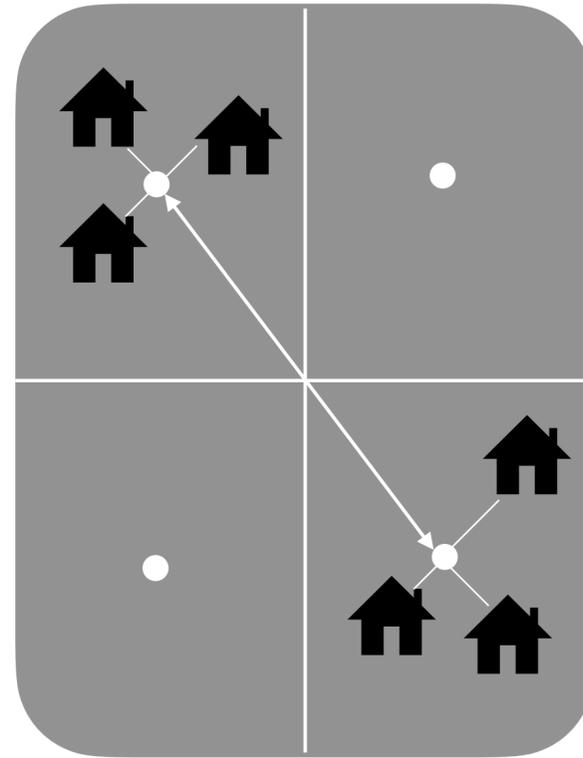
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_1, \mathbf{x}_1\}) = 1$$



No correlation
required

'Same' buildings,
different positions

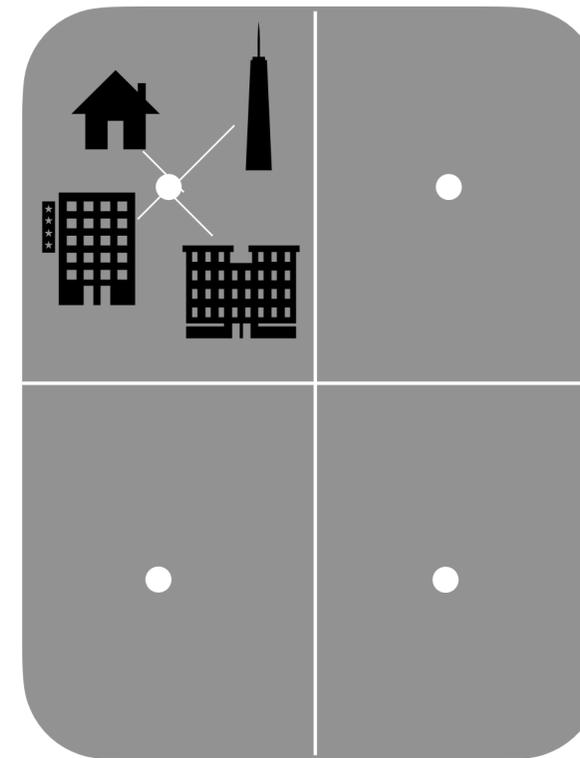
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_1, \mathbf{x}_2\}) < 1$$



Only spatial
correlation

Different
buildings,
'same' position

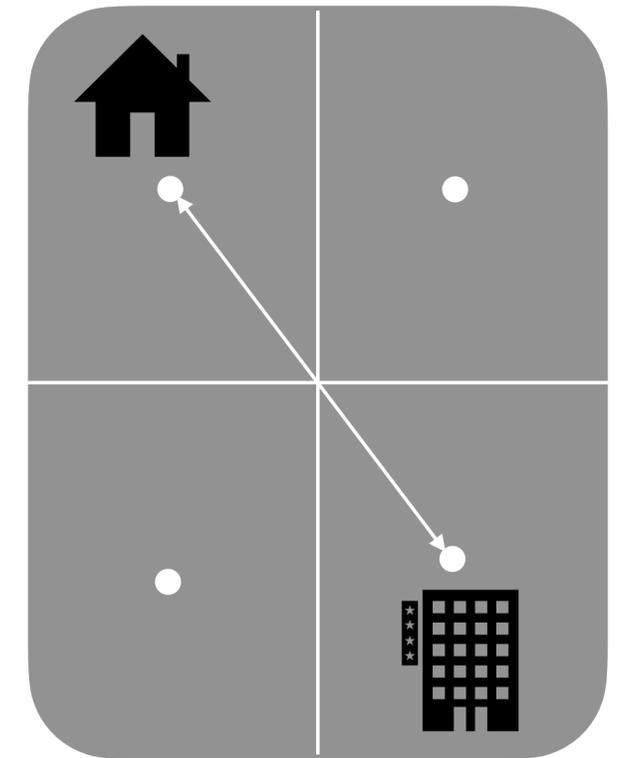
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_1\}) < 1$$



Only inter-IM
correlation

Different buildings,
different positions

$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_2\}) < 1$$



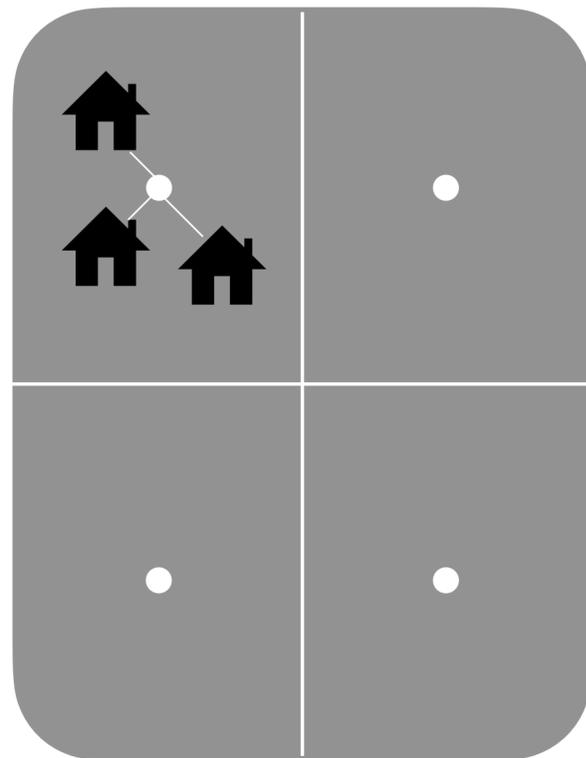
Spatial and inter-
IM correlation

Issues with treatment of correlation cases

For a general risk model there are various correlations among IMs that need to be considered

'Same'
buildings,
'same' position

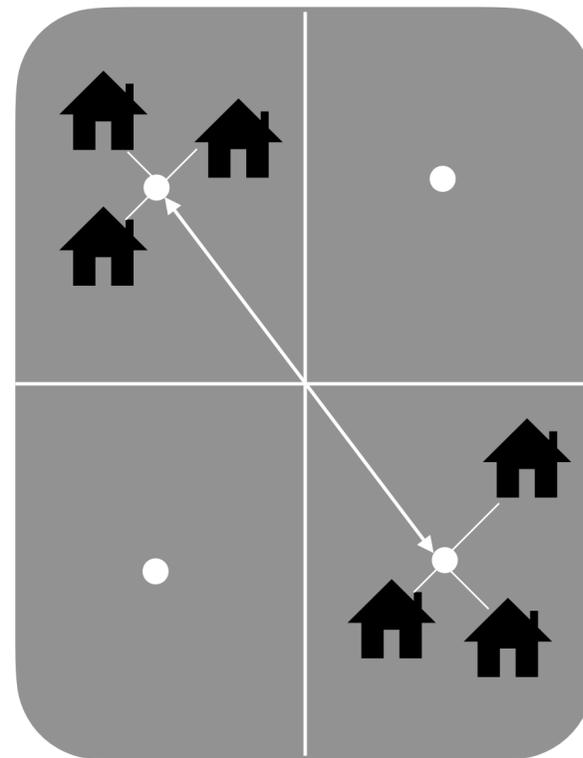
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_1, \mathbf{x}_1\}) = 1$$



All buildings obviously not
at the same location -
cannot really have perfect
correlation within cell

'Same' buildings,
different positions

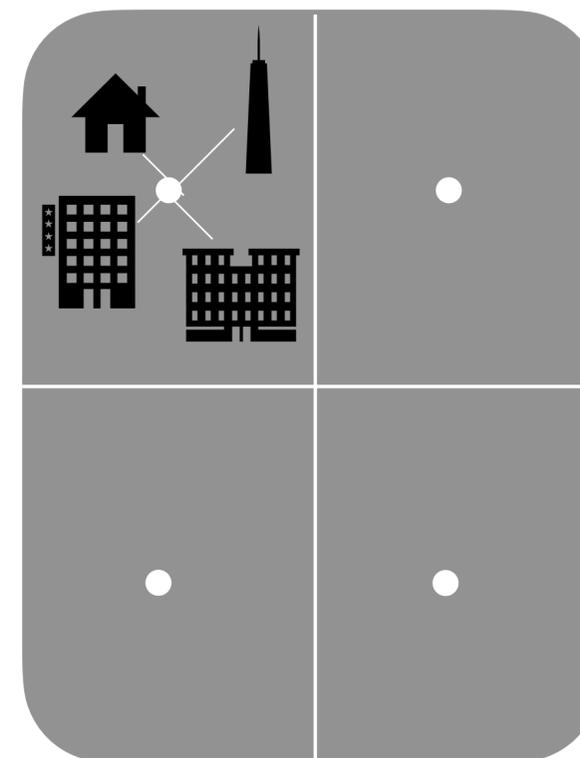
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_1, \mathbf{x}_2\}) < 1$$



Pure spatial correlation
from point-to-point ignores
spatial distribution of
buildings within cells

Different
buildings,
'same' position

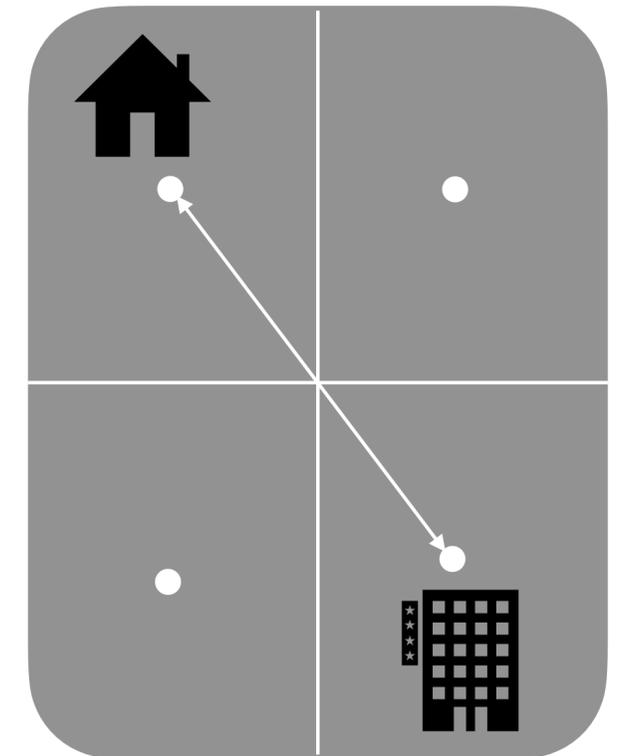
$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_1\}) < 1$$



Need to down-weight the
typical inter-period
correlation to reflect spatial
distribution

Different buildings,
different positions

$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_2\}) < 1$$



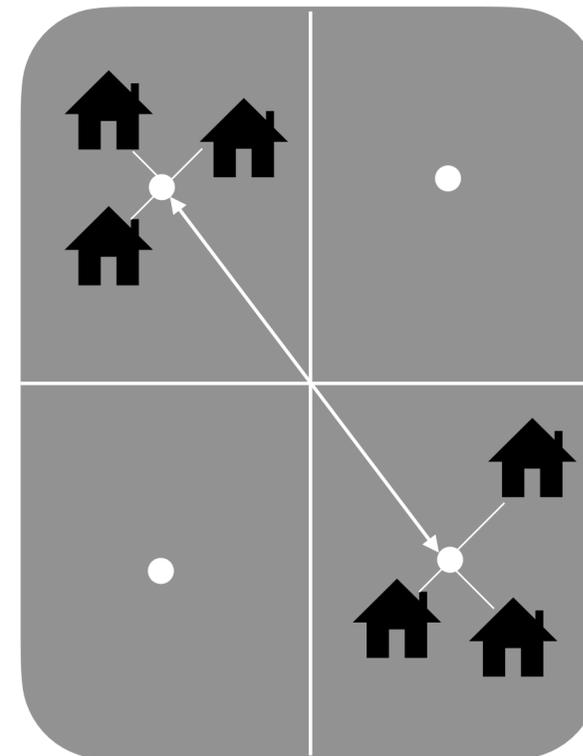
Have to consider possible
combinations of actual
locations for each building

Issues with treatment of correlation cases

- For the Groningen risk model, the use of *AvgSa* as the intensity measure for all building types means that these two scenarios become of interest
- All building types are effectively the same in terms of the IM used, so we have the same vector of periods in each case
- The two distinct cases are:
 - A single geographic cell with multiple structures receiving the same ground-motion. A single individual building experiences something like the Baker & Jayaram correlations, but that level of correlation is not appropriate for a spatial cell
 - Spatial correlations exist between spatial cells

'Same' buildings,
different positions

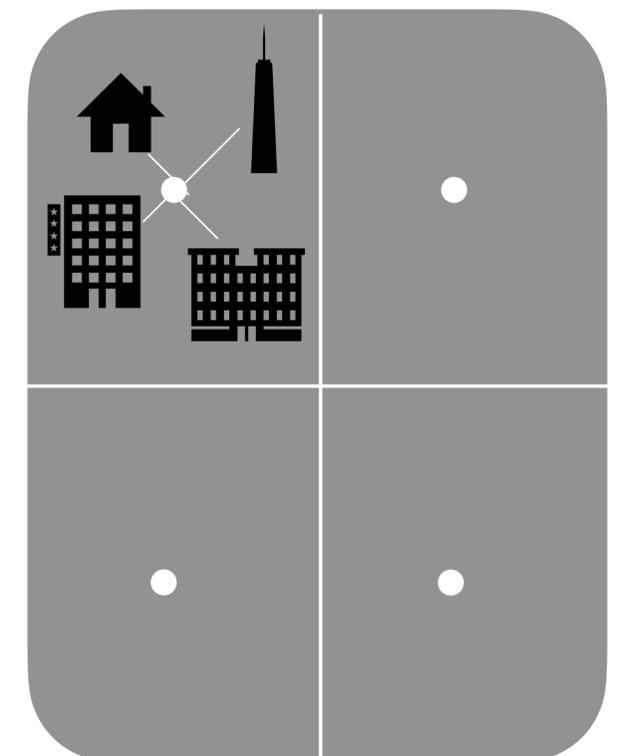
$$\rho(\{T, \mathbf{x}_1\}, \{T, \mathbf{x}_2\}) < 1$$



Pure spatial correlation
from point-to-point ignores
spatial distribution of
buildings within cells

Different
buildings,
'same' position

$$\rho(\{T, \mathbf{x}_1\}, \{T, \mathbf{x}_2\}) < 1$$



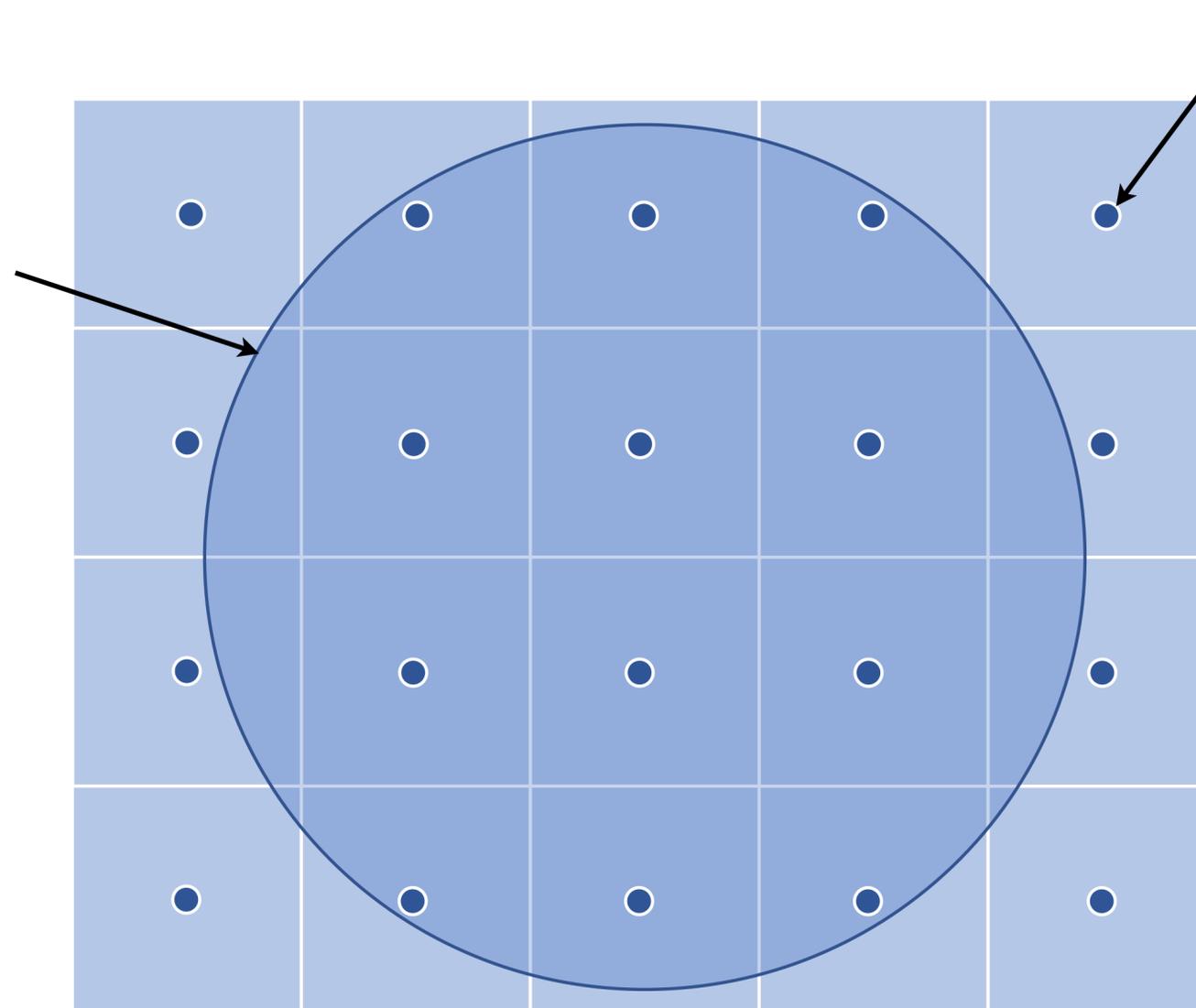
Need to down-weight the
typical inter-period
correlation to reflect spatial
distribution

Spatial discretisation of the field

- The consideration of LPR means that spatial correlation does not need to be explicitly modelled
- However, there is some implicit spatial correlation already included as a result of the spatial discretisation

Site zone

- Amplification models developed for a large number of zones
- Amplification assumed the same over the zone – but the variability in amplification accounts for spatial variability in site properties over the zone as well as uncertainties in the site profile



Grid point

- Hazard and risk calculations performed for grid points
- Each site zone can contain multiple grid points

Grid cell

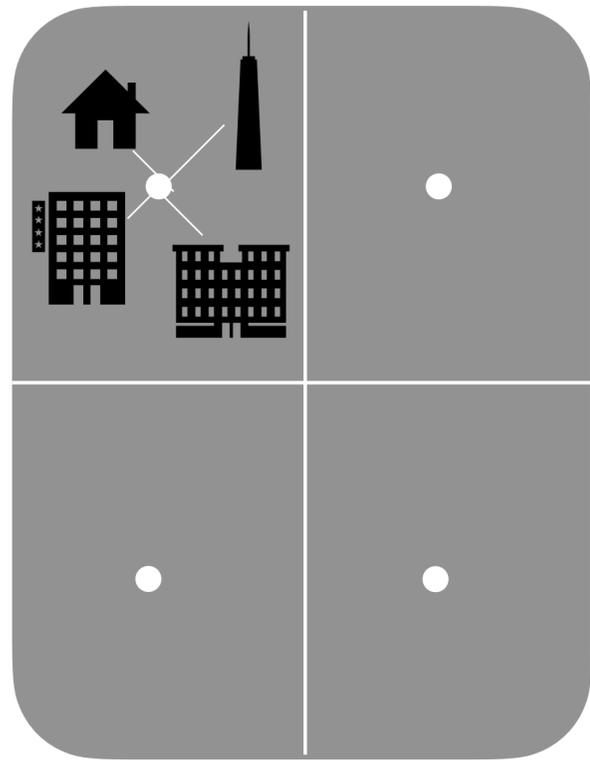
- Motions at each grid point represent those over the cell
- Typically $0.5 \times 0.5 \text{ km}^2$

Implications for Correlation Models

Effective inter-period correlations

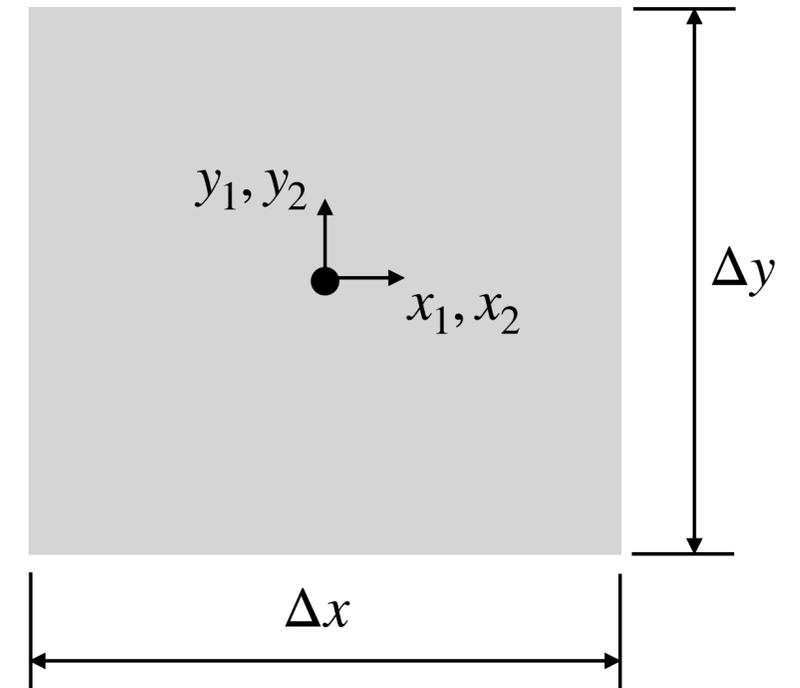
Different buildings,
'same' position

$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_1\}) < 1$$



Need to down-weight the
typical inter-period
correlation to reflect spatial
distribution

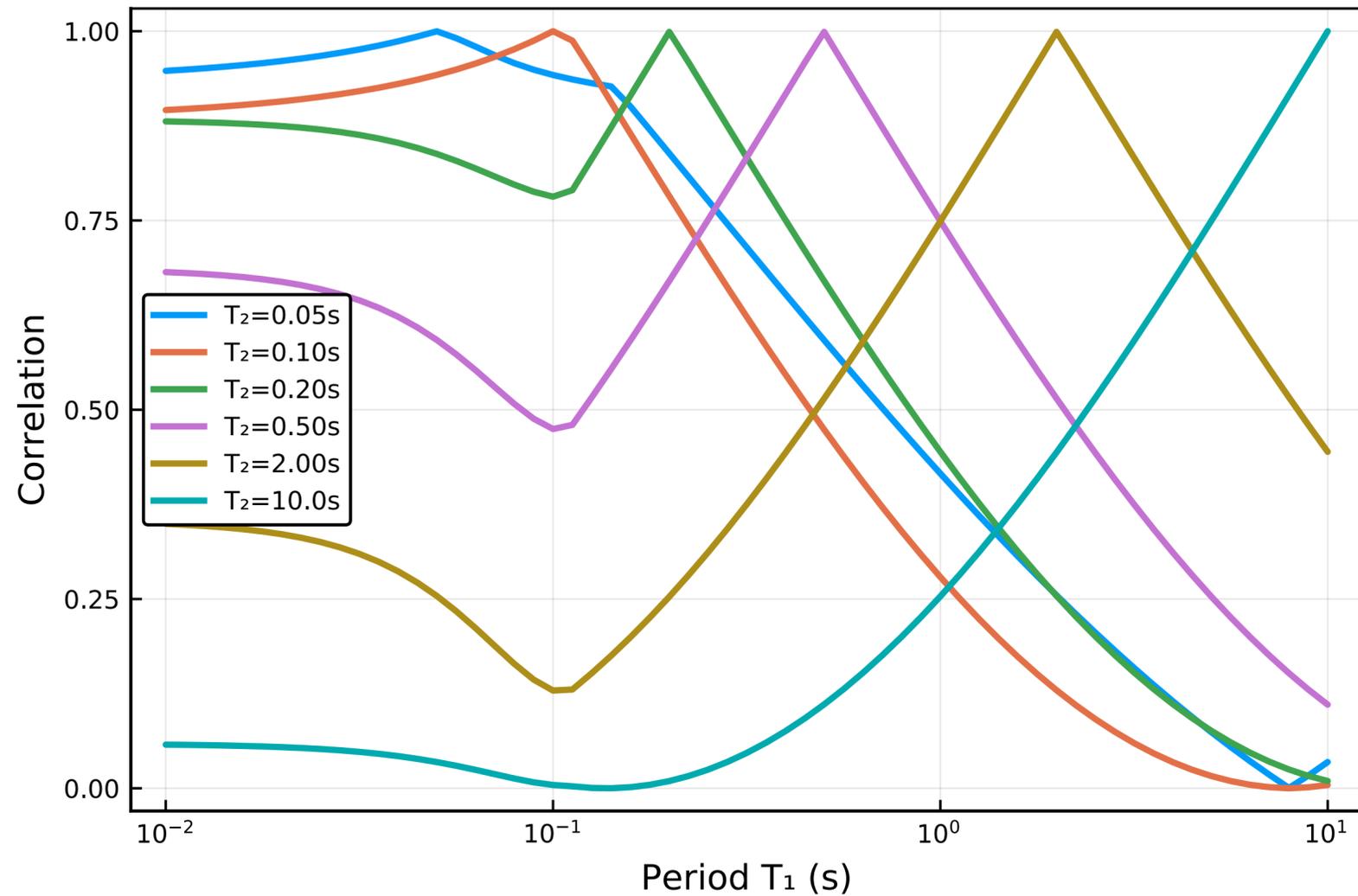
- Correlation models that are normally used represent correlations among co-located *IMs*
- However, this isn't really what we want
- Even ignoring the fact that buildings within a class will have different periods (which is normally accounted for in the fragility derivation), the spectral demands across buildings within a cell will vary due to their spatial separation



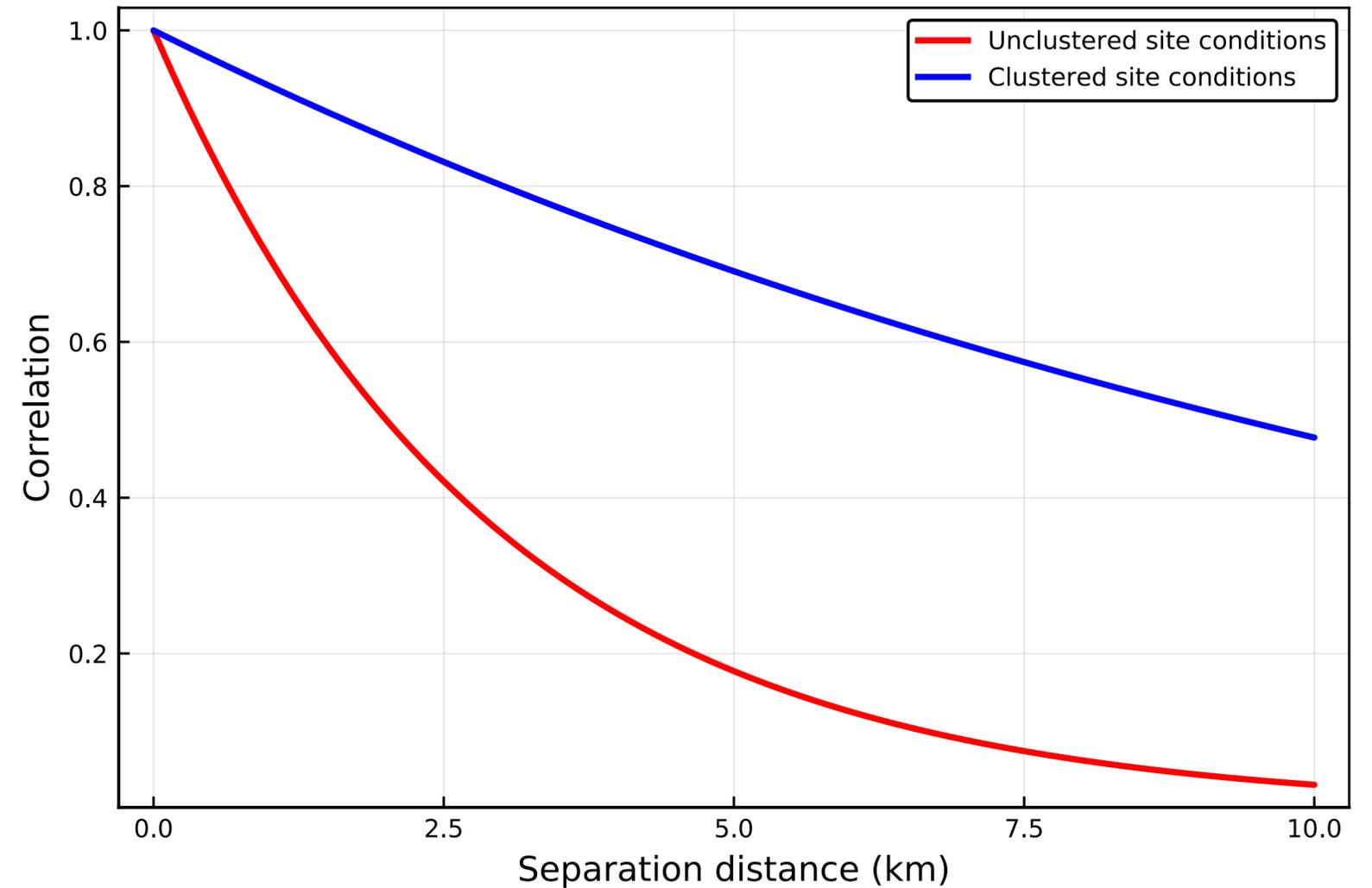
$$\rho_{eff}(T_1, T_2) = \frac{1}{\Delta x^2 \Delta y^2} \iiint \rho(\{x_1, y_1\}, \{x_2, y_2\}, T_1, T_2) dx_1 dx_2 dy_1 dy_2$$

Markovian approximation

Inter-period correlation from
Baker & Jayaram (2008), $\rho(T_1, T_2)$



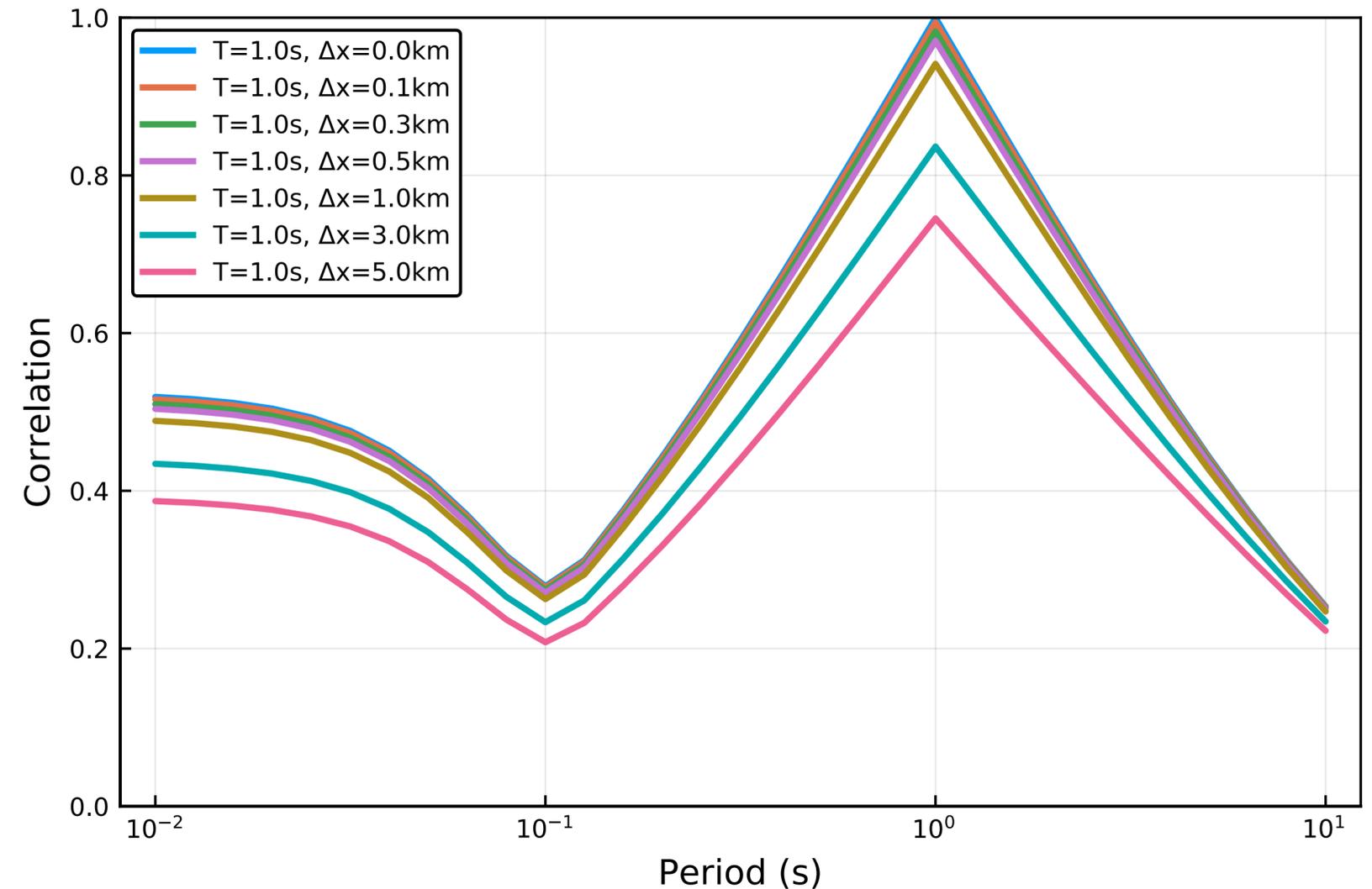
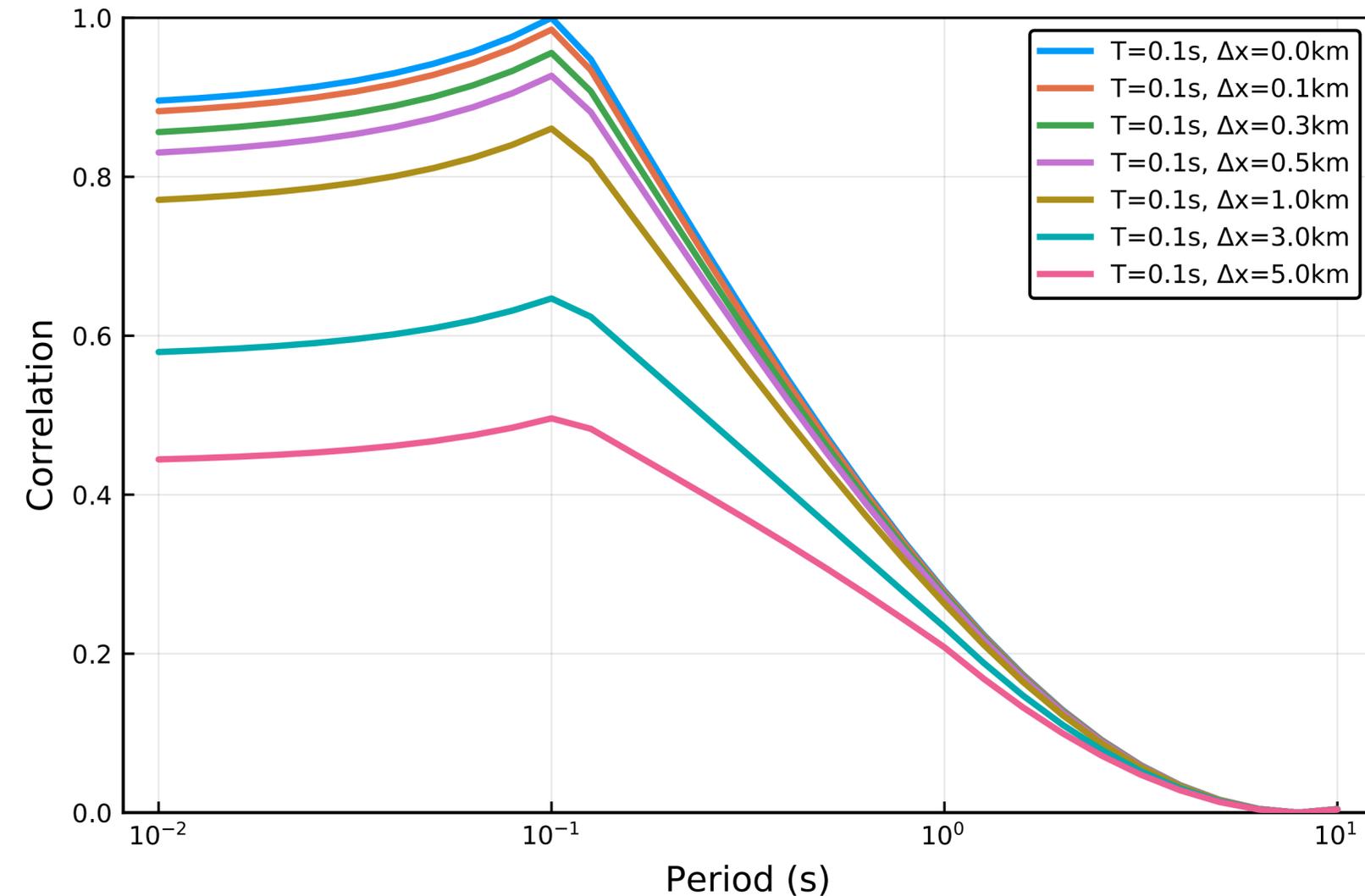
Spatial correlation of
Jayaram & Baker (2009), $\rho(\mathbf{x}_1, \mathbf{x}_2 | \max(T_1, T_2))$



- The correlation between IM_p at location \mathbf{x}_i and IM_q at location \mathbf{x}_j can be very well approximated by the inter- IM correlation for zero separation distance, and the spatial correlation for the two locations based upon the lowest frequency IM

$$\rho\left(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_2\}\right) \approx \rho(T_1, T_2) \times \rho(\mathbf{x}_1, \mathbf{x}_2 | \max(T_1, T_2))$$

Effective inter-period correlations



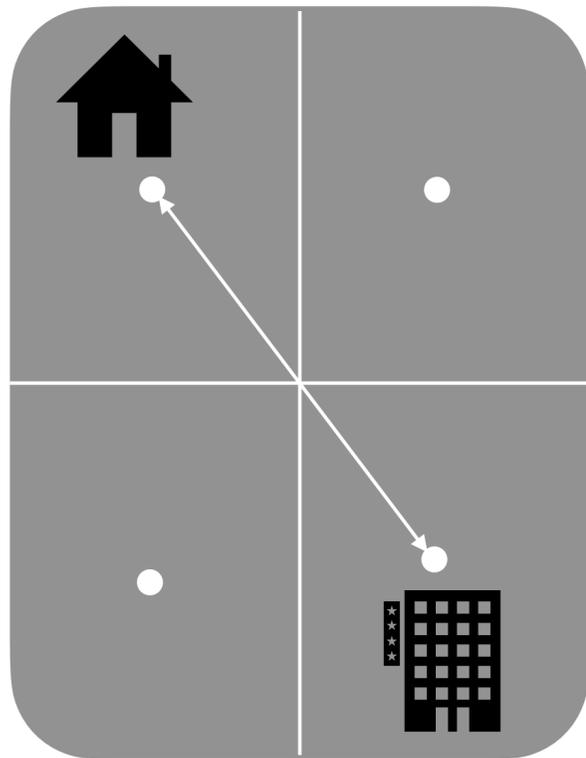
- Naturally, the larger the cell size, the greater the reduction in the inter-period correlation
- One needs to be working at a relatively high resolution in order to ignore this effect
- This reduced inter-period correlation also induces a slight increase in the ground-motion variability for the cell

$$\Delta\sigma(\mathbf{x}) \approx \sigma(\mathbf{x})\sqrt{1 - \rho_{eff}^2}$$

Effective inter-cell correlations

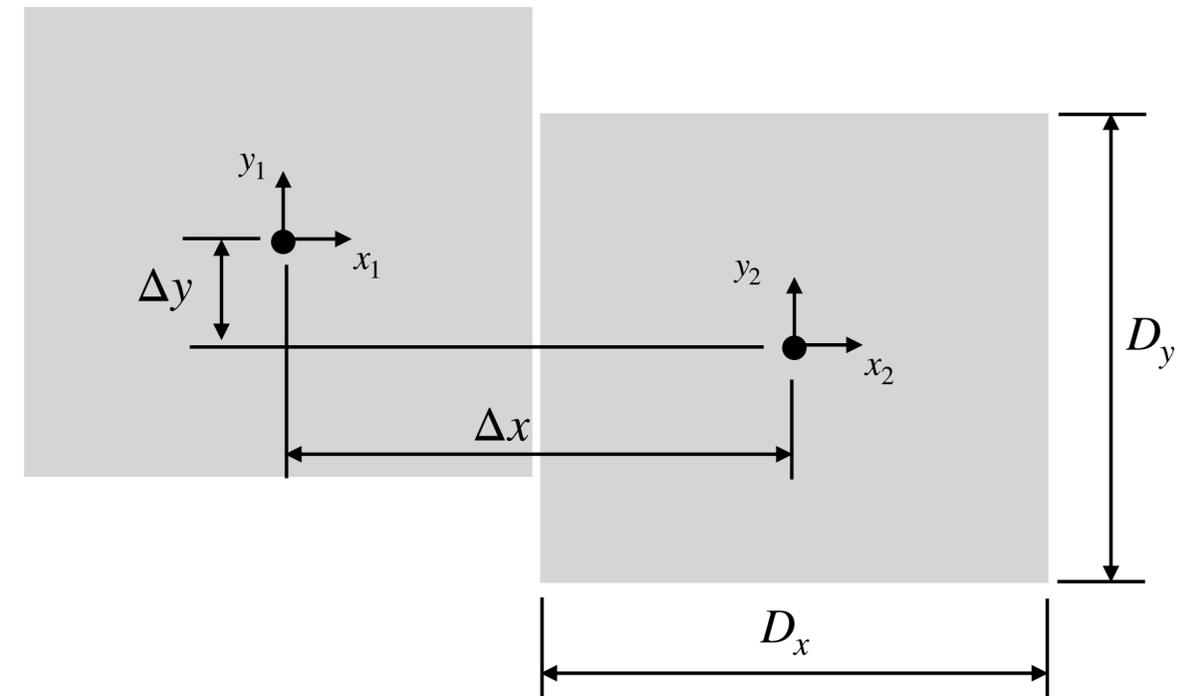
Different buildings,
different positions

$$\rho(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_2\}) < 1$$



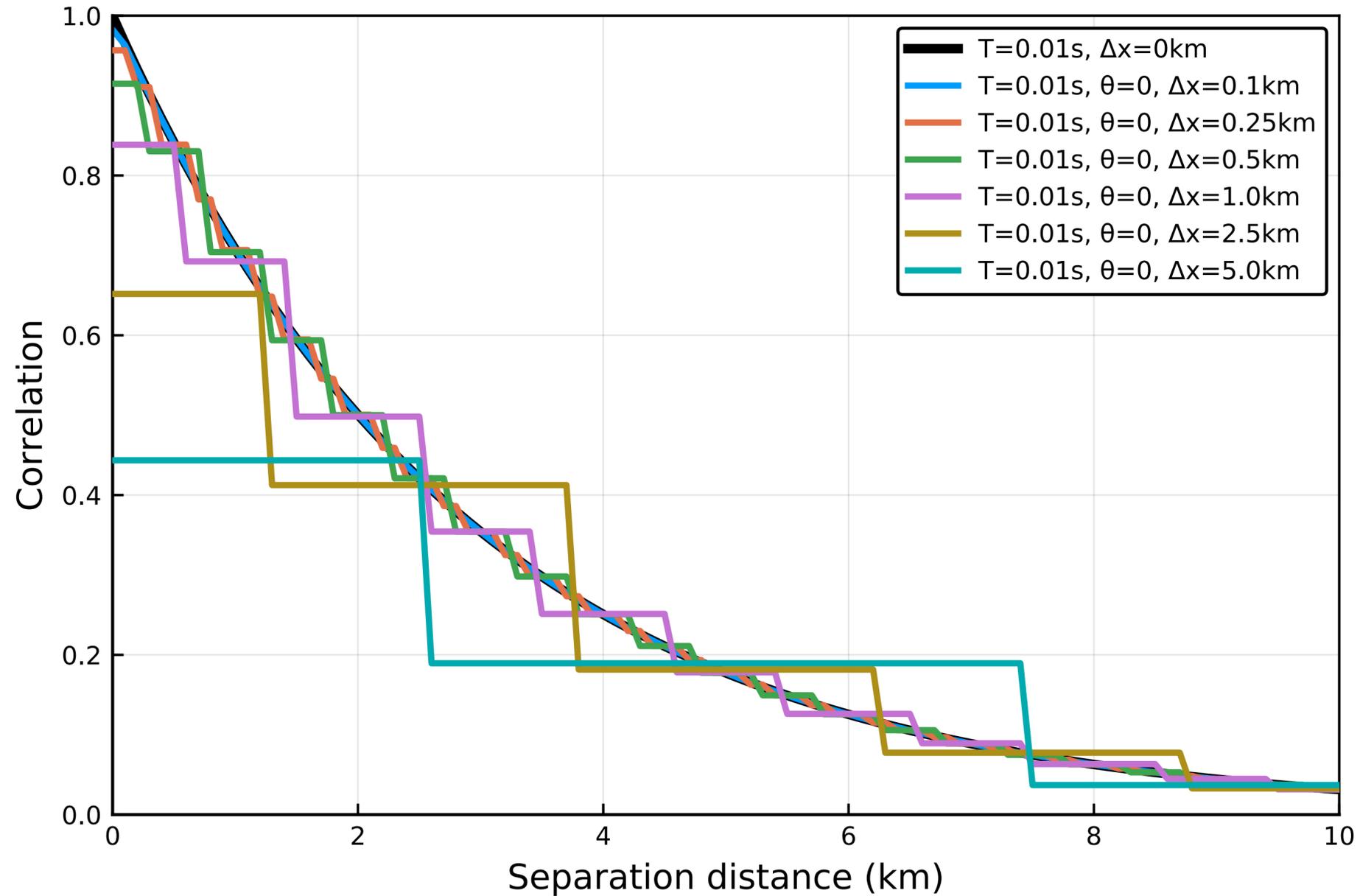
Have to consider possible
combinations of actual
locations for each building

- ◉ Similarly to the effective inter-period correlations, we make the extension of the spatial discretisation over offset cells
- ◉ This reflects the unknown location of buildings of different classes throughout the respective cells



$$\rho_{eff}(T_1, T_2) = \frac{1}{D_x^2 D_y^2} \iiint_{+=\Delta x, +=\Delta y} \rho(\{x_1, y_1\}, \{x_2, y_2\}, T_1, T_2) dx_1 dx_2 dy_1 dy_2$$

Effective inter-cell correlations



- Largest differences arise at short separation distances where the correlation is most important
- Typically a small reduction in correlation, but it varies with the relative sizes of the cells

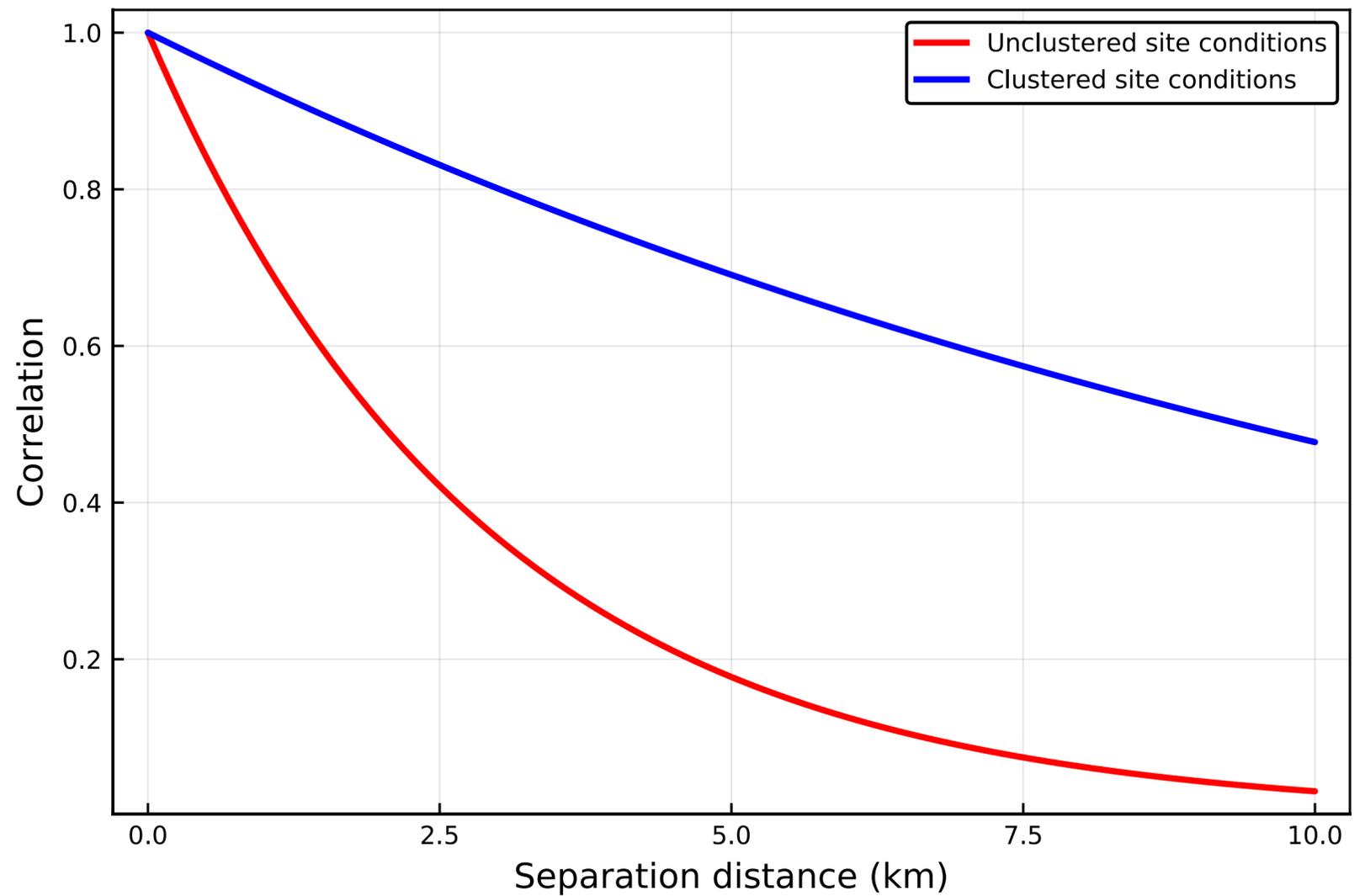
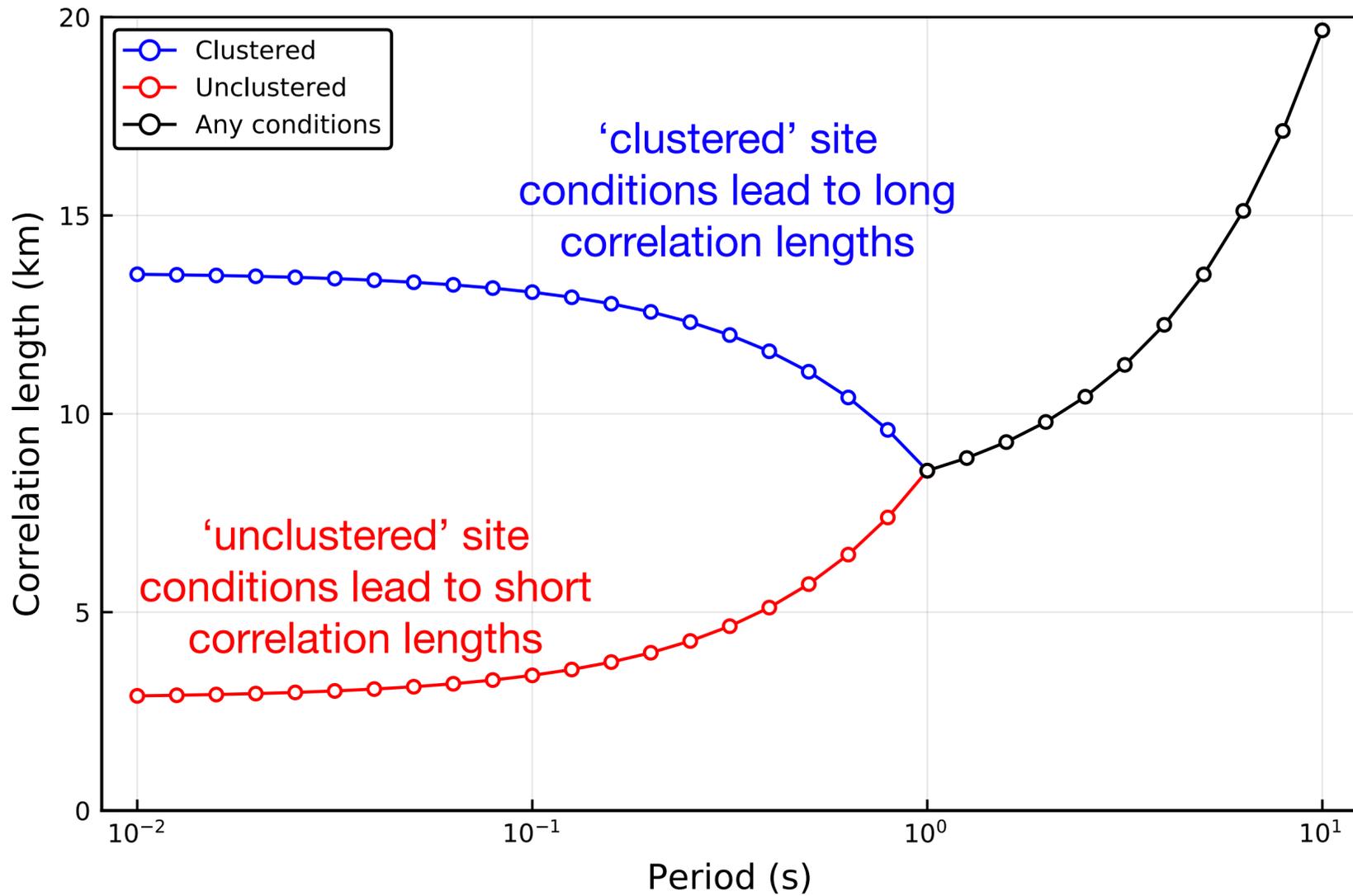
Site effects and spatial correlation

- Jayaram & Baker (2009) presented a now widely-adopted spatial correlation model
- They use an exponential model that has a different correlation length, r , for periods below 1.0 seconds, depending upon whether site conditions are ‘clustered’ or ‘unclustered’
- This approach was motivated by strongly different correlation lengths found in different regions (and soil conditions were held responsible)
- Their model is **ergodic** and can be expressed as:

$$\ln Sa(\mathbf{x}) = \mu_{\ln Sa}(\mathbf{x}; rup) + \delta_B + \delta_W(\mathbf{x}) \quad \delta_B \sim N(0, \tau^2) \quad \delta_W(\mathbf{x}) \sim MVN(\mathbf{0}, \phi^2(\mathbf{x})\mathbf{R}(\mathbf{x}))$$

$$\rho(\delta_W(\mathbf{x}_i), \delta_W(\mathbf{x}_j)) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{r}\right) = \exp\left(-\frac{\Delta}{r}\right) = \rho(\Delta)$$

Ergodic spatial correlation



Non-ergodic spatial correlation

- If we partition the variance so that we have systematic event and site effects then we rewrite the model as:

$$\ln Sa(\mathbf{x}) = \mu_{\ln Sa}(\mathbf{x}; rup) + \delta_B + \delta_{S2S}(\mathbf{x}) + \delta_{W_{es}}(\mathbf{x})$$

$$\delta_B \sim N(0, \tau^2)$$

$$\delta_{S2S}(\mathbf{x}) \sim MVN[\mathbf{0}, \phi_{S2S}^2(\mathbf{x})\mathbf{R}_S(\mathbf{x})]$$

$$\delta_{W_{es}}(\mathbf{x}) \sim MVN[\mathbf{0}, \phi_{SS}^2(\mathbf{x})\mathbf{R}_W(\mathbf{x})]$$

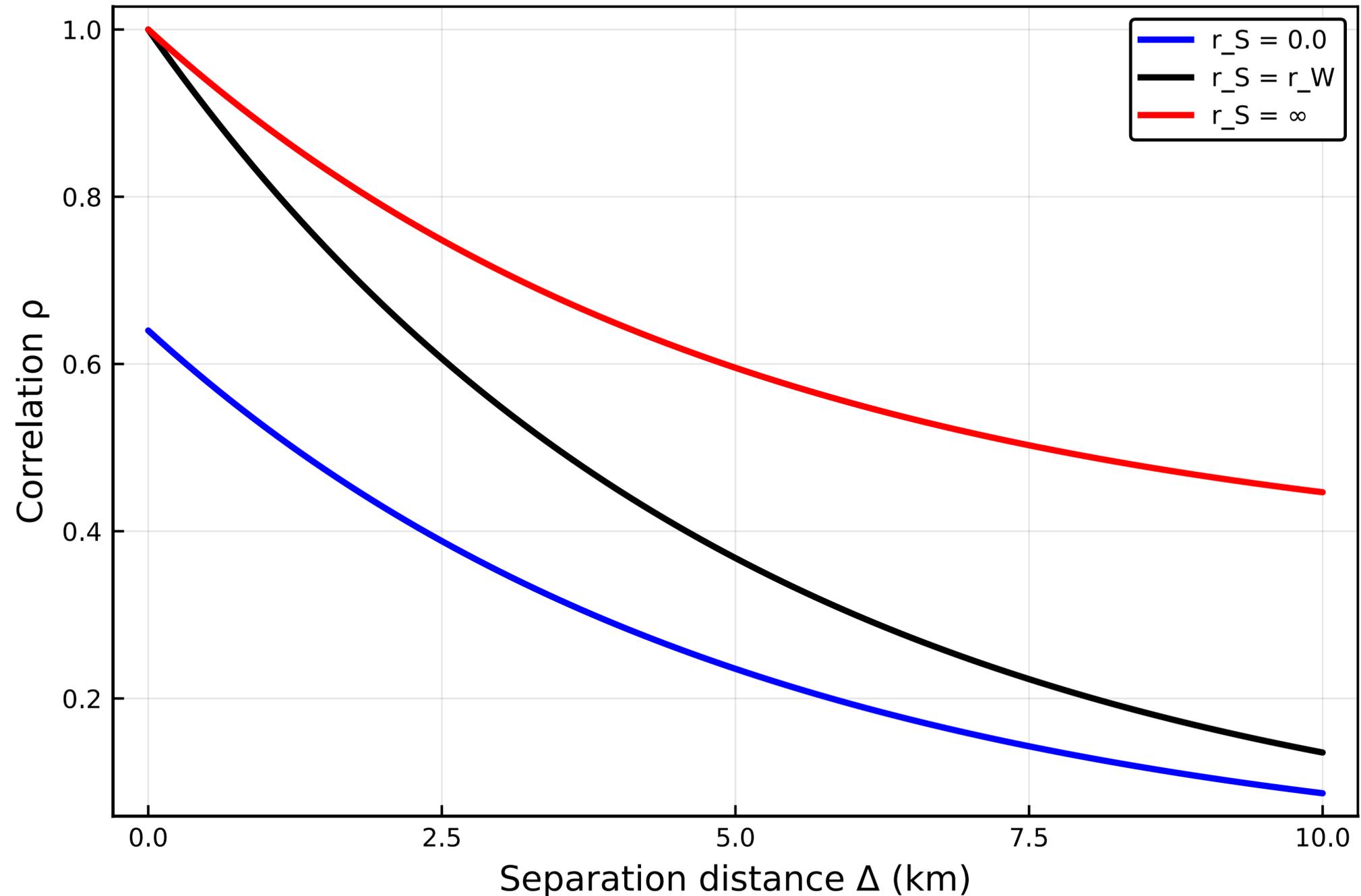
- The overall within-event spatial correlation in this case is represented by:

$$\rho[\delta_W(\mathbf{x}_i), \delta_W(\mathbf{x}_j)] = \frac{\rho_W(\mathbf{x}_i, \mathbf{x}_j)\phi_{SS}(\mathbf{x}_i)\phi_{SS}(\mathbf{x}_j) + \rho_S(\mathbf{x}_i, \mathbf{x}_j)\phi_{S2S}(\mathbf{x}_i)\phi_{S2S}(\mathbf{x}_j)}{\phi(\mathbf{x}_i)\phi(\mathbf{x}_j)} \quad \phi^2 = \phi_{SS}^2 + \phi_{S2S}^2$$

- Express the spatial correlations among δ_{S2S} terms and among $\delta_{W_{es}}$ terms as exponential models

$$\rho_S(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\Delta}{r_S}\right) \quad \rho_W(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\Delta}{r_W}\right)$$

Non-ergodic spatial correlation

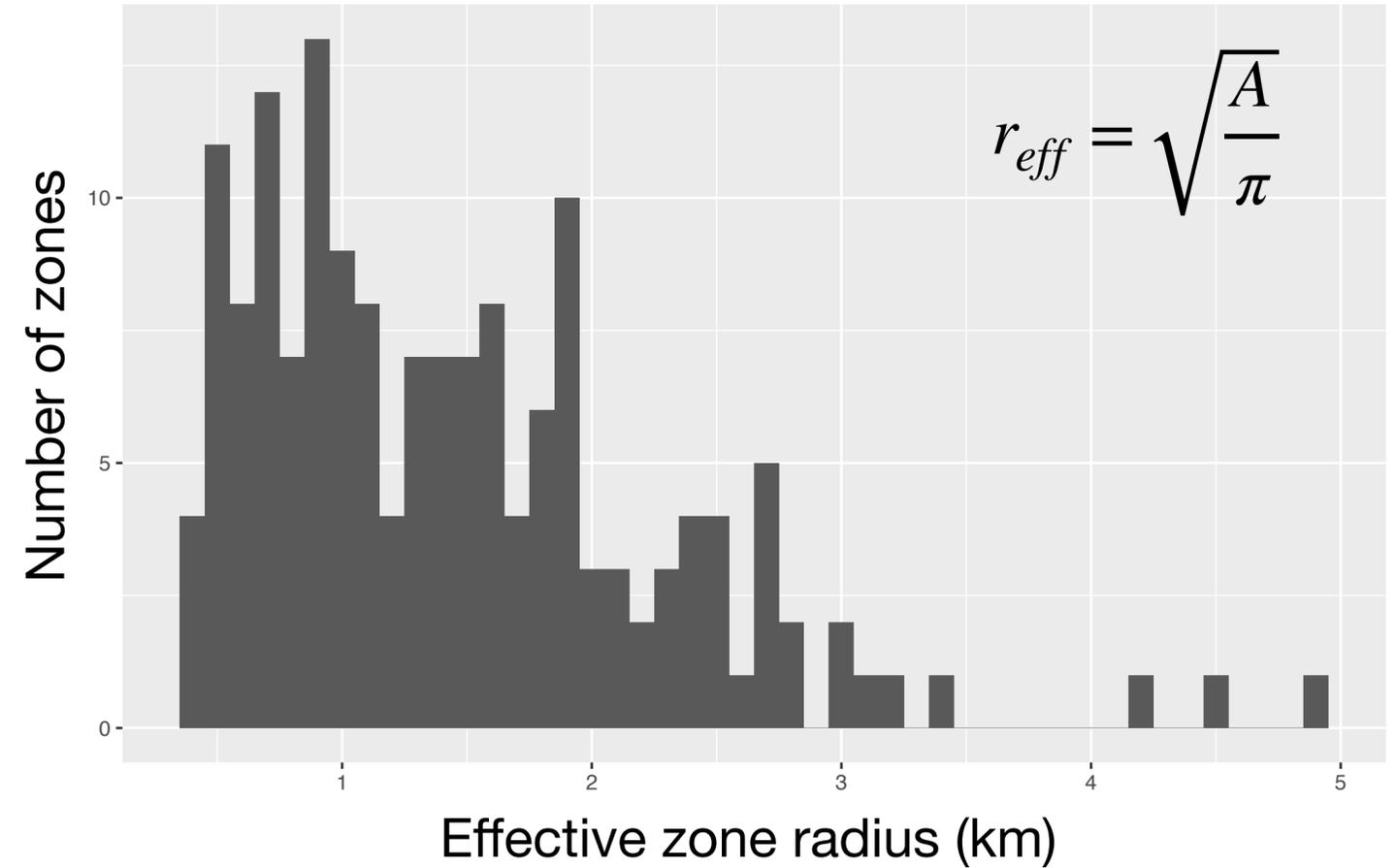
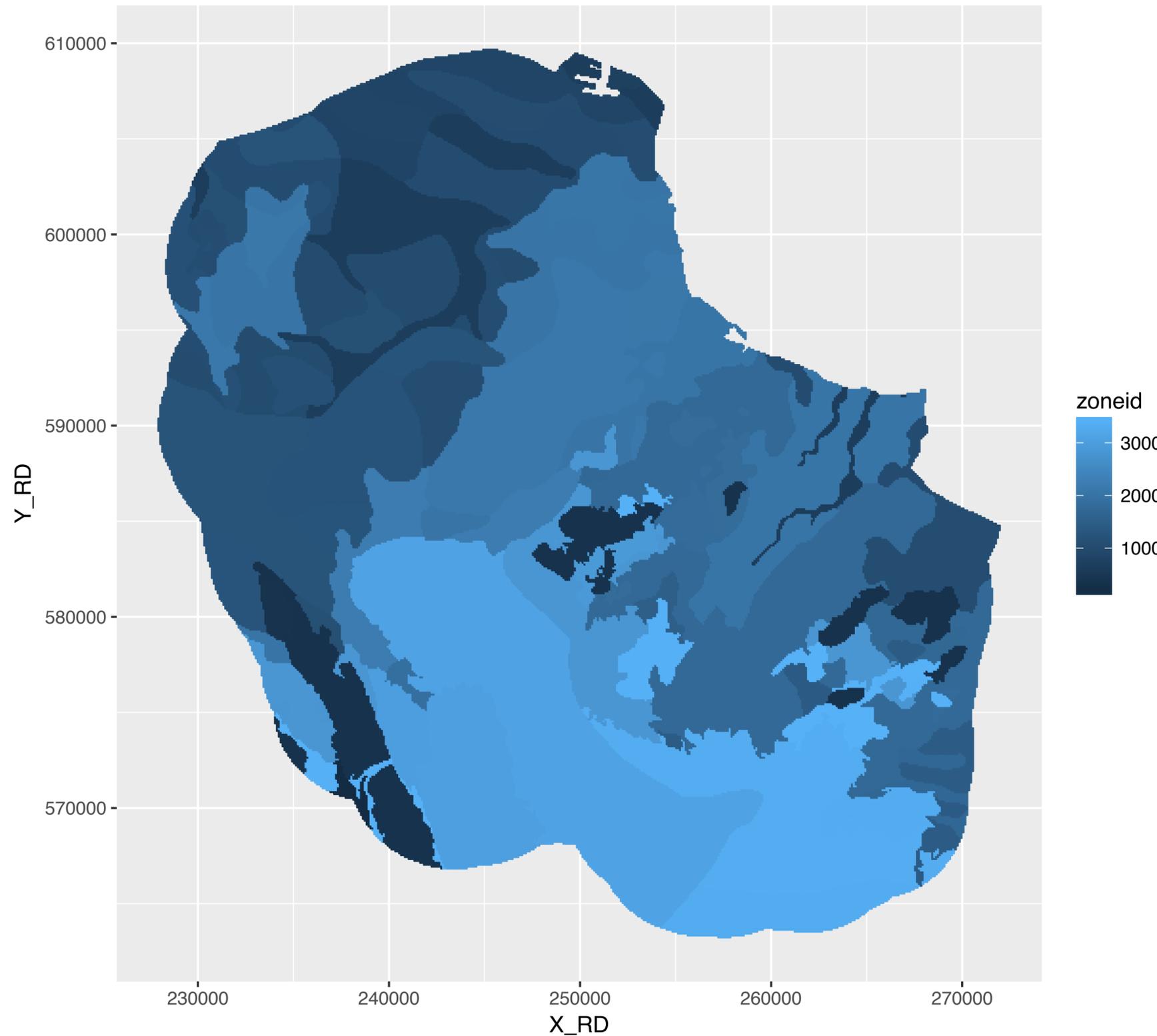


- We can therefore consider the ergodic correlation model as being contaminated by correlations among systematic site effects (epistemic terms) that have some correlation length r_S

$$\rho(\Delta) = \frac{1}{\phi^2} \left[\phi_{SS}^2 e^{-\frac{\Delta}{r_W}} + \phi_{S2S}^2 e^{-\frac{\Delta}{r_S}} \right]$$

- Consider limiting cases of $r_S \rightarrow \infty$ (perfectly correlated site conditions) and $r_S \rightarrow 0$ (completely random site conditions)
- Jayaram & Baker's clustered vs non-clustered results just reflect implicit r_S values for the regions they investigated
- Geographic cells in (or spanning) different site zones will have a loss of correlation

Implied spatial correlation



- Many site zones, each of a different size
- The implied correlation depends upon the size of these regions

Implied spatial correlation

Overall spatial correlation arises from:

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = \frac{\rho_{SS,ij} \phi_{SS}(\mathbf{x}_i) \phi_{SS}(\mathbf{x}_j) + \rho_{S2S,ij} \phi_{S2S}(\mathbf{x}_i) \phi_{S2S}(\mathbf{x}_j) + \rho_{c2c} \sigma_{c2c}(\mathbf{x}_i) \sigma_{c2c}(\mathbf{x}_j)}{\sigma(\mathbf{x}) \sigma(\mathbf{x})}$$

Partial correlation inside grid cell

$$\rho_{SS} = 1, \rho_{S2S} = 1, \rho_{c2c} = 0$$

Partial correlation among grid cells in the same zone

$$\rho_{SS} = 0, \rho_{S2S} = 1, \rho_{c2c} = 0$$

Perfect correlation inside grid cell

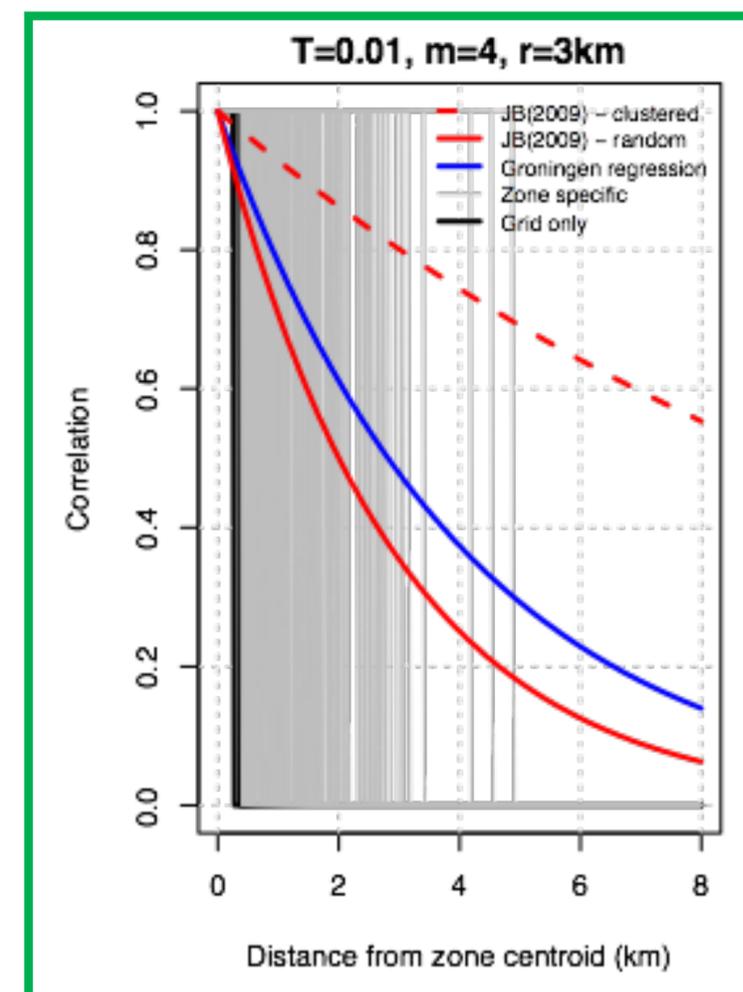
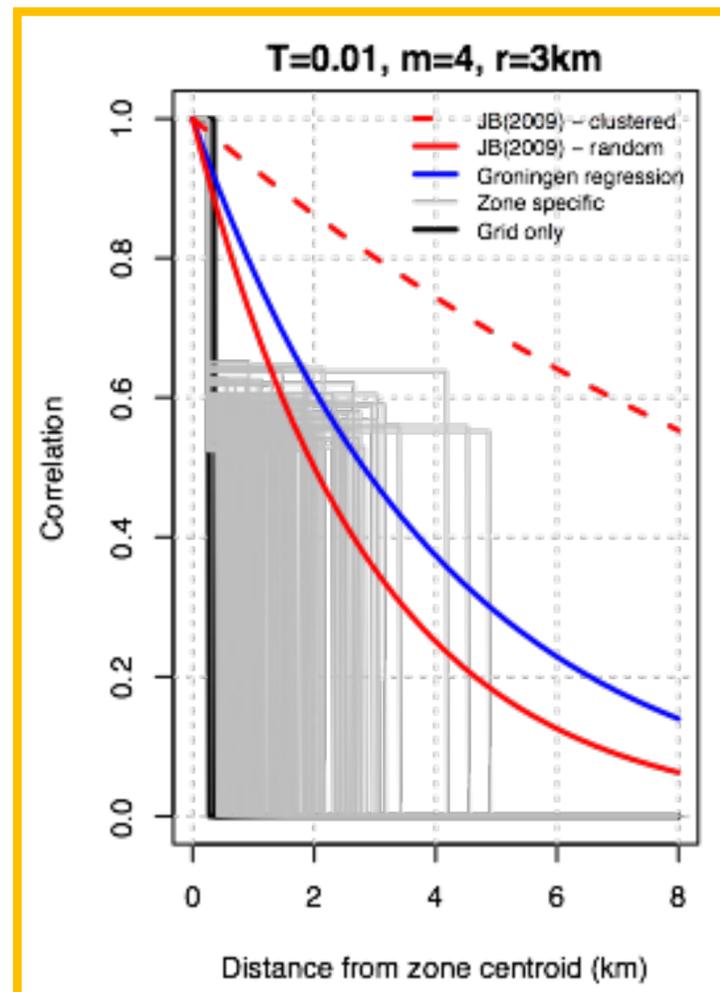
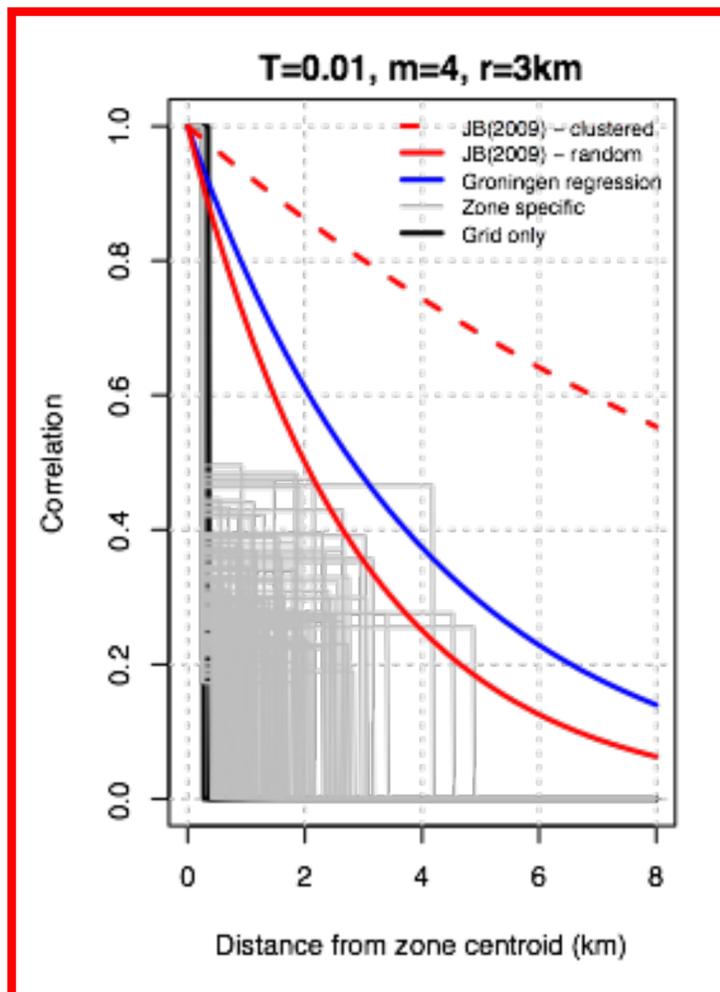
$$\rho_{SS} = 1, \rho_{S2S} = 1, \rho_{c2c} = 1$$

Partial correlation among grid cells in the same zone

$$\rho_{SS} = 0, \rho_{S2S} = 1, \rho_{c2c} = 1$$

Perfect correlation inside site zone, zero correlation outside

$$\rho_{SS} = 1, \rho_{S2S} = 1, \rho_{c2c} = 1$$



Perfect correlation case seems most appropriate as an interim measure while a field-specific correlation model is developed

Sampling components of variability

Amplification variability

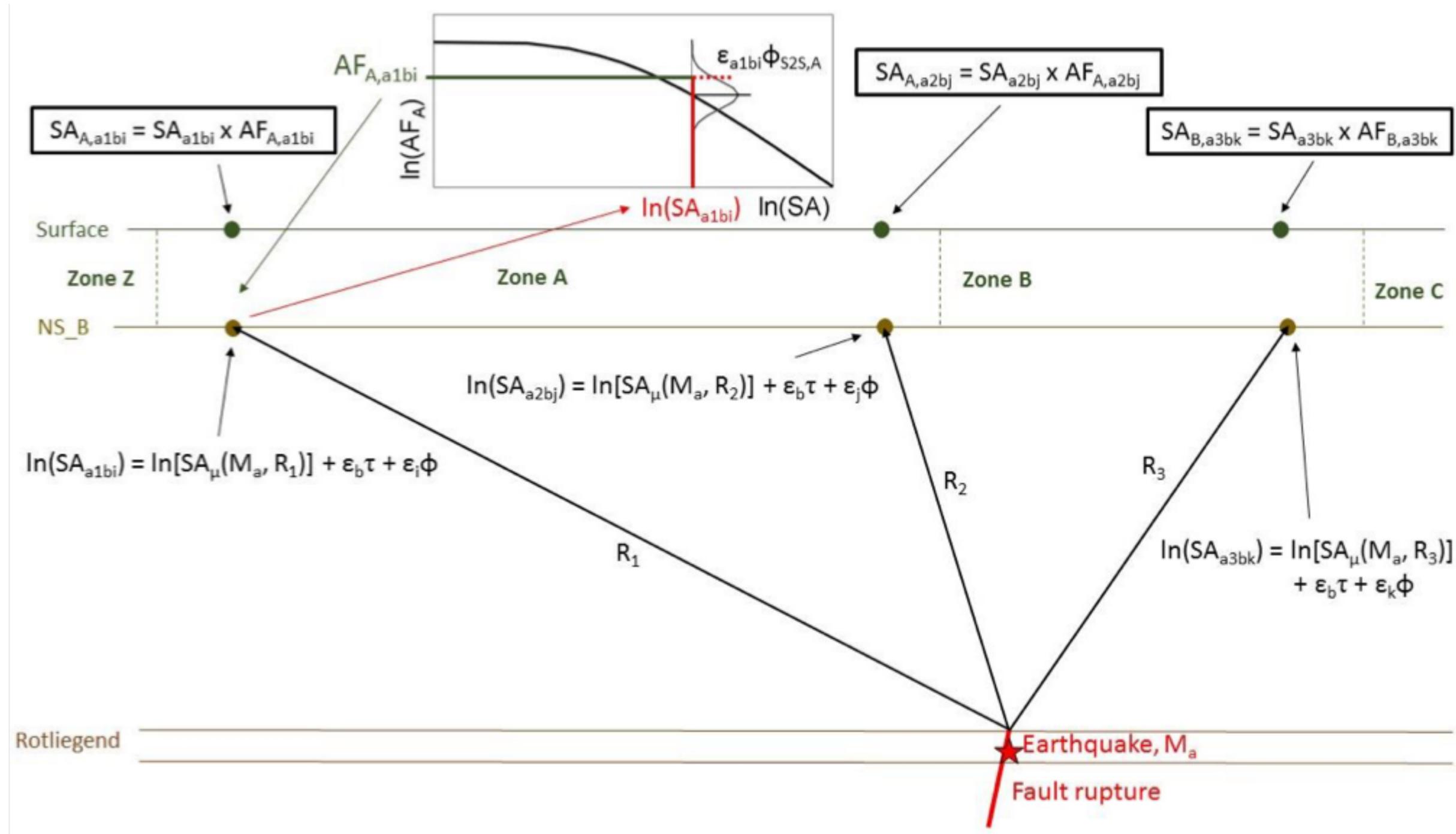
- Sample from ϕ_{S2S}
- Use same ε for all grid points within zone

NS-B Variability

- Include both ϕ_{SS} & σ_{C2C}
- Draw same ε for each grid point in the site zone
- No explicit inclusion of spatial correlation
- Correlated among IMs

Event term

- Generated for each event
- Constant for all locations for a given IM
- Correlated among IMs



Summary of Implications

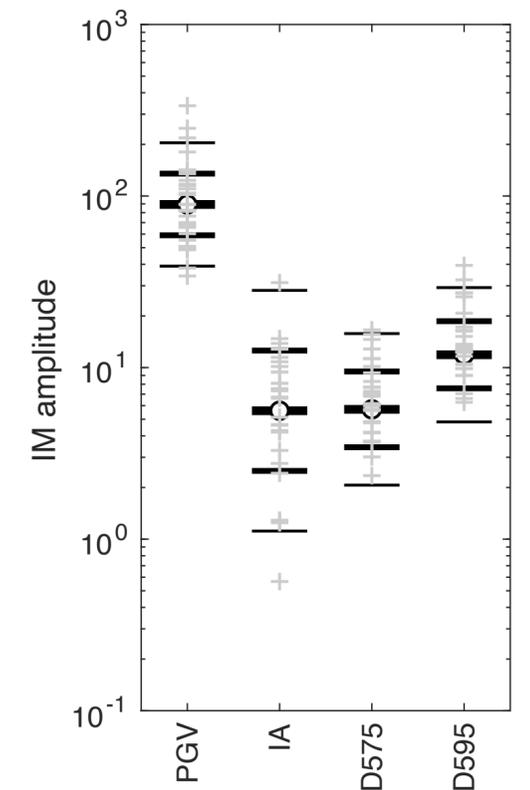
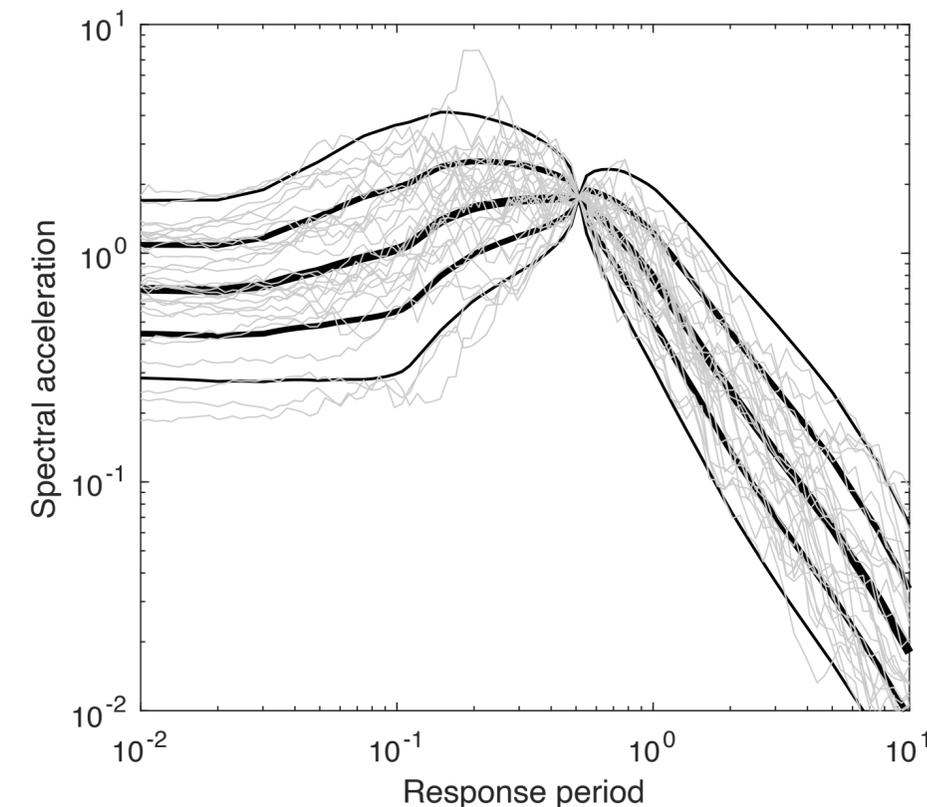
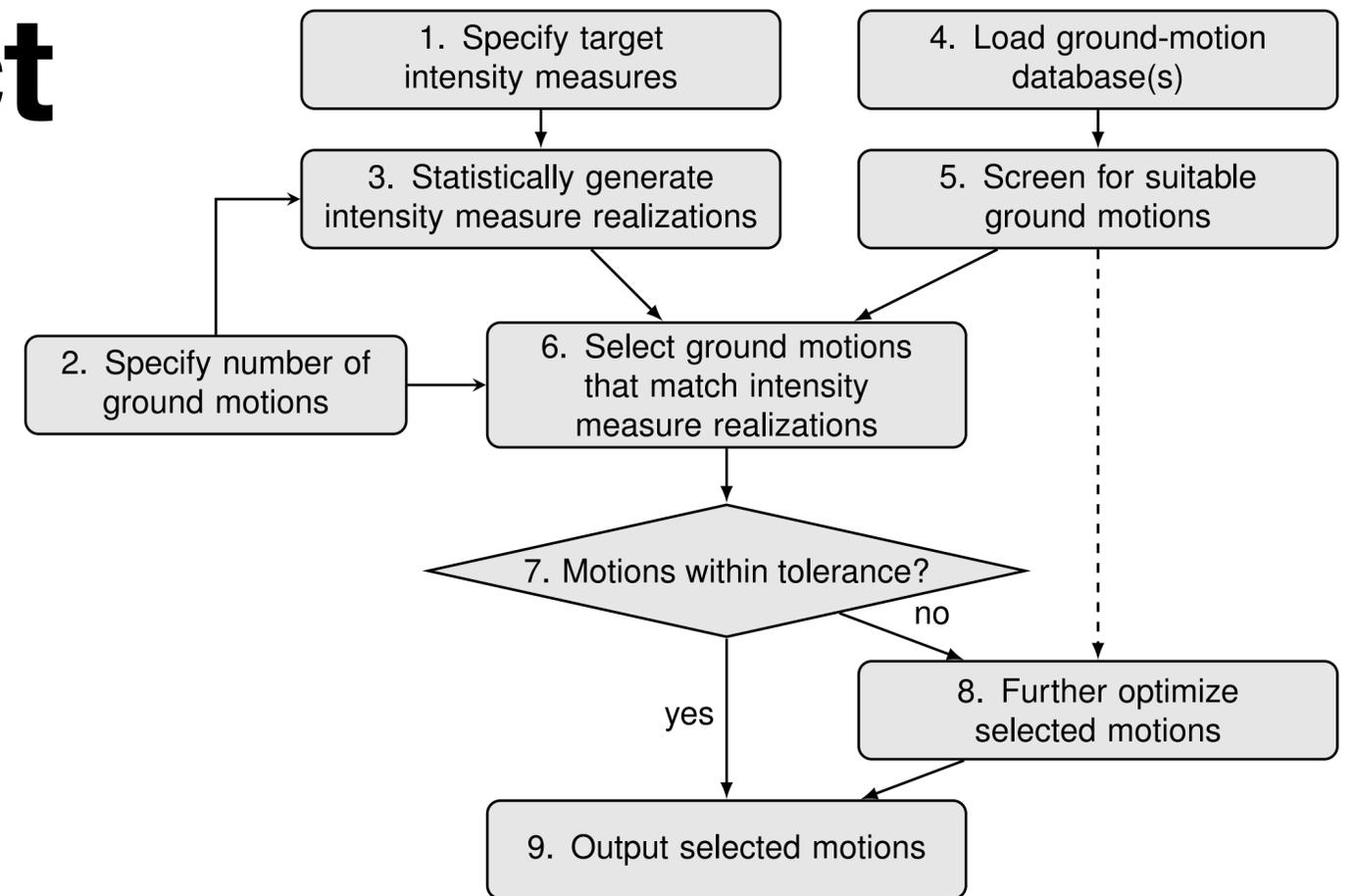
Response to “2020-10-29 TNO - P2P Correlation Note”

- Which implementation is conceptually ‘correct’: NAM or TNO?
 - Neither. The ‘correct’ solution would require a complete overhaul of the framework, as well as the development of components that don’t currently exist
- Which implementation has followed the instructions and intentions of the GMM team?
 - TNO (to my knowledge)
- Did the GMM team make a mistake in leaving out inter-period correlations of the site residuals?
 - No. It was done deliberately, for the following reasons:
 - Ergodic correlation models include period-to-period correlations related to site models that are not relevant to Groningen. We have developed zone-specific transfer functions and nonlinear site amp functions.
 - The Groningen site amplification functions are based upon transfer functions that impose systematic period-to-period correlation effects. Imposing this correlation and then also including full ergodic site correlations isn’t appropriate.
 - The Nonlinear site response is predicted as a function of GM *realisations* at NS-B, rather than median predictions, so correlated motions are propagated through the site amp functions
 - We want to account for partial loss of correlation to approximate spatial effects that are not explicitly modelled in the hazard and risk engine.
 - Site residuals are perfectly correlated spatially over a cell. Under the Markovian approach to modelling $\rho\left(\{T_1, \mathbf{x}_1\}, \{T_2, \mathbf{x}_2\}\right) \approx \rho(T_1, T_2) \times \rho(\mathbf{x}_1, \mathbf{x}_2 | \max(T_1, T_2))$, imposing some degree of spatial correlation has an equivalent effect of modifying the period-to-period correlation

Ground-motion selection

Comments on the impact of scaling

- Its no coincidence that the motions selected for the fragility calibration have the appropriate correlation structure
- The records are specifically selected to achieve this feature
- Scaling can distort inherent correlations *between intensity measures of different types*, but that doesn't matter because the selection algorithm only selects motions that match the target



Review of the TNO 2021 Model Chain Groningen Report

Jean-Paul Ampuero and Eric-Jan Wagenmakers

December 28, 2021

We have been tasked by TNO to review five specific questions related to their report “Status of the TNO Model Chain Groningen per October 1, 2021 and recommendations for the public Seismic Hazard and Risk Analysis 2022”, and as they relate to the two following documents:

1. Appendix A: Seismic Source Model Calibration;
2. Appendix B: Undesirable consequences of using a ‘data-calibrated’ exponential taper model in the context of Groningen hazard and risk analysis.

General review

Appendix A presents arguments in favor of making public the source model calibration. These arguments are clear and compelling. It also describes a Bayesian procedure to calibrate the source model. The description is complete (with references to a previous report) and the approach is mathematically rigorous. Overall the work presented in this appendix is of high scientific quality and no scientific objections to using this methodology were identified.

Appendix B confirms the concept that parameters related to the tapering of the tail of the magnitude-frequency distribution are poorly constrained by the existing data. This point is further substantiated by providing a very clear demonstration, through simulations of Bayesian inference of taper model parameters, of the crucial effect that the assumed prior distribution has on inferred parameters.

Specific questions

The five review questions are quoted below in italics and our review comments are provided immediately below each question.

1. *To what extent is a taper fundamentally different from Mmax and should they be used in conjunction?*

NAM’s smooth taper model and the sharply-truncated Mmax model play similar roles: they modify the Gutenberg-Richter earthquake magnitude distribution by reducing the rate of high magnitude events. The joint use of these similar model mechanisms introduces statistical redundancy and therefore demands strong empirical evidence before being adopted. Lacking such evidence, as is the case here, we believe that it is statistically prudent to consider these mechanisms separately, as different branches of the logic tree. Almost no data are available on the tails of the distribution, and this means that a proper model-selection method will prefer the simple account, in accordance with Ockham’s razor (e.g. see #4). The complete taper model has higher complexity: it is time-dependent, via the stress-dependence of its parameters.

It is conceivable to use the Mmax distribution derived from expert elicitation as part of the prior in Bayesian estimation of the taper model parameters. But that is not equivalent to use the taper and Mmax models in conjunction in the way proposed by NAM.

2. *How can a taper location be determined from physics-based evidence or models?*

For the Mmax model, this question was addressed by some participants in the expert elicitation Mmax-workshop. For instance, physics-based models have been proposed in which Mmax may be determined by constraints from fracture mechanics (a la Dempsey and Suckale) or by slip budget (a la Avouac).

Bourne and Oates (2020) presented the physical motivations of their taper model, including statistical physics models of failure in brittle materials (critical point theories). But there is an inconsistency between the proposed model and its physical motivation, which makes it difficult to look for further physics-based constraints on the model parameters. The inconsistency is detailed as follows. The proposed functional form of the stress-dependent parameters involves the difference between the Coulomb stress *change* and a parameter θ . The latter is analogous to a critical point or failure threshold and is assumed constant across the Groningen field. However, critical point theories would naturally involve the *total* Coulomb stress, not its change: a critical point for failure is reached when the total Coulomb stress reaches a threshold. Thus the relevant quantity is $(\text{Coulomb stress change}) + (\text{initial Coulomb stress}) - \theta$. The two last terms can be lumped into an effective parameter $\theta' = \theta - (\text{initial Coulomb stress})$. But, because the initial Coulomb stress is unlikely to be spatially uniform, it cannot be assumed that θ' is uniform across Groningen. This is in contrast with the uniform parameter θ in Bourne & Oates (2020). One could argue that θ approximates the spatial average of θ' , but that would not be satisfactory: heterogeneities of Coulomb stress changes and initial Coulomb stresses might be comparable in amplitude, thus there is no reason to account for one and ignore the other. Thus, the physical interpretation of the parameter θ in the current taper model is ill defined.

3. *Can a taper location be calibrated on the available data?*

In principle it is possible to use data to estimate a taper location. However, the data at hand are not in the tail of the distribution, which is where the taper is particularly relevant. In light of the sparsity of the data in the tail, it seems unwise to introduce additional mechanisms to model the behavior of this tail.

The tests conducted by TNO (section B.6) show that estimates of the taper location strongly depend on the prior. Their results confirm that the taper location is poorly resolved by the available data. This issue has also been discussed in the literature on statistical seismology (e.g. Zöller, 2013 <https://doi.org/10.1002/grl.50779>; Zöller and Holschneider 2016a <https://doi.org/10.1785/0220150176>).

4. *Can a logic tree weighing be based on pseudo-prospective performance?*

Using a pseudo-prospective performance test, Bourne and Oates (2020) found that the (stress-dependent) taper model ranked highest. However, the pseudo-prospective performance test is a non-standard procedure for model comparison and suffers from at least three important drawbacks:

(A) The pseudo-prospective performance test is biased in favor of the more complex model, as are most cross-validation-style methods. The reason is that the training data benefit the complex model more than they benefit the simple model. Consequently, cross-validation-style methods are generally not statistically consistent. In other words, if the simple model were (approximately) true, the pseudo-prospective performance test would never support it conclusively. To remedy this problem, out-of-sample performance ought to be assessed for all data points, not half of them (with the other half available for training). This is automatically

accomplished by Bayesian model selection (including BIC), as this can be interpreted as accumulative one-step-ahead prediction error, starting with predictions for the very first observation.

(B) The pseudo-prospective performance test uses a pairwise likelihood ranking procedure that we believe is as unusual as it is suspect. Assume a hypothetical scenario where the training set is so large that the posterior distribution is extremely narrow. Furthermore, assume that the predictive difference in likelihood for the validation data set is very small, with model A just edging out model B. The correct interpretation would be that both models perform about equally well. However, the ranking method would conclude that in nearly 100% of the posterior samples, model A outpredicts model B. This would be true, but it would also be irrelevant and a misleading statement concerning the models' relative performance. We worry that the ranking method inflates predictive performance differences that are in reality only minor.

(C) It is unclear to us why the pseudo-prospective performance test considers entire distributions of likelihoods. After observing the first half of the data, the posterior distribution can indeed be used to compute predictive performance for the second half of the data, but the parameter values are a nuisance factor and should be averaged over. The advantage of retaining the association with the parameter values is unclear (except for the possibility of applying the ranking method, which we feel is deeply suspect, see point B).

Another general concern is that the data are known, such that it is tempting to include knowledge of the data in the specification of the prior distribution.

5. *What are the (dis)advantages of integration over Coulomb stress fields?*

For the Coulomb-stress model, NAM uses the Maximum Likelihood Estimate, which implicitly assumes a uniform / non-informative prior, and TNO proposes to integrate over the posterior of the Coulomb-stress modeling. Without integration the risk is larger by ~50%.

From a Bayesian perspective, integrating is the proper way to proceed. Computational cost is not a concern: the integration approach is more computationally expensive, but still feasible.

The sensitivity to the priors is a potential concern, which could be assessed by a sensitivity analysis. So far TNO has considered a uniform prior on the Coulomb-stress parameters (σ , r_{\max} , H_s). Note that these parameters are strictly positive "Jeffreys quantities" in the sense of Tarantola (2015 <https://doi.org/10.1137/1.9780898717921>), meaning that other choices of parametrization (such as $X' = 1/X$ instead of X) are equally valid. To set a non-informative prior for a Jeffreys quantity X independent of the parametrization choice, it is preferable to set a uniform prior on $\log(X)$, which leads to a prior $\sim 1/X$ on X itself.