TNO PUBLIEK

Princetonlaan 6
3584 CB Utrecht
P.O. Box 80015
3508 TA Utrecht
The Netherlands

www.tno.nl

[redacted]

**TNO report**

**TNO2022_R11961 | Final report**

# Status of the TNO Model Chain Groningen per October 1, 2022 and recommendations for the public Seismic Hazard and Risk Analysis 2023

| | |
|---|---|
| Date | 18 October 2022 |
| Author(s) | [redacted] |
| Copy no | |
| No. of copies | |
| Number of pages | 44 (incl. appendices) |
| Number of appendices | 4 |
| Sponsor | Ministry of Economic Affairs and Climate Policy |
| Project name | Publieke SDRA Groningen |
| Project number | 060.51885 |

# Samenvatting


Dit rapport beschrijft de resultaten van de lopende modelontwikkeling met betrekking tot de SDRA en de status van de TNO Modelketen per 1 oktober 2022. Daarnaast geeft het een overzicht van de door TNO aanbevolen modelversies op basis van de laatste wetenschappelijke inzichten voor het uitvoeren van de publieke Seismische Dreigings- en Risico Analyse Groningen 2023 ten behoeve van het "Vaststellingsbesluit gasjaar 2023-2024".

**Overzicht modelontwikkeling**
De modelontwikkeling met betrekking tot de publieke Seismische Dreigings- en Risicoanalyse (pSDRA) Groningen is momenteel gericht op het ontwikkelen van een betrouwbaardere beschrijving van het toekomstige seismische gedrag na de beëindiging van de gasproductie. Om de prestaties van nieuwe modellen objectief te beoordelen is een Test- en Vergelijkingsraamwerk (TCF) geïmplementeerd om deze taak te ondersteunen. In het seismische bronmodel (SSM) is een snelheidsafhankelijk compactiemodel geïmplementeerd. Dit alternatief voor het momenteel gebruikte lineair-elastische compactiemodel kan wellicht een betere voorspelling geven van bevingen bij een constante of toenemende poriëndruk in het reservoir, die wordt verwacht als gevolg van het insluiten van het gasveld. Voorlopige resultaten uit de model prestatietests wijzen er niet op dat dit model significant beter presteert dan het huidige model. Naast deze ontwikkeling heeft TNO een alternatief magnitude-frequentie (b-waarde) model ontwikkeld dat de Gutenberg-Richter b-waarde beschrijft als een trapsgewijze functie van de reservoirdikte. Resultaten van de uitgevoerde prestatiebeoordeling wijzen op superioriteit van dit model vergeleken met de huidige en alternatieve magnitude-frequentiemodellen. Een pseudo-prospectieve beoordeling van het model in een Bayesiaanse context zoals het Test- en Vergelijkingsraamwerk is echter nog niet voltooid en er zijn nog enkele tests nodig om tot een weloverwogen voorstel van weegfactoren te komen voor gebruik in de pSDRA Groningen.

**Status TNO Modelketen Groningen**
De TNO Modelketen Groningen is ingericht om de publieke Seismische Dreigings- en Risico Analyse (pSDRA) 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 2022 omvat de modelversies gebruikt met ingang van de HRA/pSDRA Groningen 2019, en daarnaast alternatieve versies van het SSM en het FCM die door TNO geadviseerd zijn te gebruiken in de pSDRA 2023 Per 1 oktober 2022 zijn er drie nieuwe modelcomponent implementaties beschikbaar: GMM-V7, een magnitude-frequentiemodel met ruimtelijke afhankelijke b-waarde en een seismisch bronmodel gebaseerd op het Rate Type Compaction Model. 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 2023**

Op basis van de beschikbare informatie per 1 oktober 2022 beschouwt TNO onderstaande (sub)modelversies als het meest geschikt voor gebruik in de publieke SDRA Groningen 2023. 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 2023. 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 2023. Deze modelversie wordt gekenmerkt door het gebruik van een distributie van stress covariate velden en een magnitudeverdeling die begrensd wordt door een Mmax verdeling. TNO adviseert voorlopig om de geactualiseerde Mmax 2022 verdeling in overweging te nemen voor gebruik in de pSDRA Groningen, op voorwaarde dat de beschrijving van de staart van de frequentie-magnitude verdeling onderwerp zal zijn van nader onderzoek.
- TNO adviseert het gebruik van GMM versie NAM-V7 in de publieke SDRA Groningen 2023. Verbeteringen van deze nieuwe modelversie zijn onder meer een herziening van het dempingsmodel en de expliciete benadering voor gebouwen die op wierden staan.
- TNO adviseert het gebruik van FCM versie TNO-2020 in de publieke SDRA Groningen 2023. 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

This report describes the results of ongoing model development in relation to the SHRA and the status of the TNO Model Chain as of October 1, 2022. It also provides an overview of the TNO recommended model versions based on the latest scientific insights based on the latest scientific insights to run the public Seismic Hazard and Risk Analysis Groningen 2023 for the “Vaststellingsbesluit gasjaar 2023-2024”

**Overview model development**
Model development in relation to the public Seismic Hazard and Risk Analysis (pSHRA) Groningen is currently focussed on describing the future post production seismic behaviour more reliable than the current seismic source model (SSM). To objectively assess the performance of new models a Testing and Comparison Framework (TCF) was implemented to assist in this task. A rate dependent compaction model has been implemented in the seismic source model (SSM). This alternative to the currently used linear elastic compaction model may be able to better predict main shock events for constant or increasing reservoir pore pressure expected during shut-in. Provisional results of the performance tests of this model do not indicate that this model performs significant  better than the current model. In addition to this TNO has developed an alternative magnitude-frequency (b-value) model that describes the Gutenberg-Richter b-value as a step function of the reservoir thickness. results of the conducted performance assessment indicate superiority of this model compared to the current and alternative magnitude-frequency models. However, a pseudo-prospective model performance test in a Bayesian model inference context such as the Test  and Comparison Framework is not yet completed and some further testing is needed to come to a deliberate proposal of model weighting factors for use in the pSHRA Groningen.

**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, 2022 includes all the model versions used for HRA/pSHRA Groningen as of 2019, as well as alternative versions of SSM and FCM recommended by TNO for use in pSHRA 2023. As of October 1, three new model component implementations are available: GMM-V7, a magnitude-frequency model with spatial dependent b-value and an activity rate model based on the Rate Type Compaction Model. 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 2023**
Based on the available information as of October 1, 2022, TNO recommends the following versions of (sub) models as most suitable for use in the public SHRA Groningen 2023. 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 2023. 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 2023. This approach uses a full posterior distribution of stress covariate fields and uses a magnitude distribution that is truncated by a Mmax distribution. TNO preliminary advises to consider the updated Mmax 2022 distribution for use in pSHRA Groningen provided that the tail description of frequency-magnitude distribution will be subject of further research.
- TNO recommends the use of GMM version NAM-V7 in the public SHRA Groningen 2023. Improvements of this new model version include among other things a revision of the damping model and the explicit site amplification for buildings located on dwelling mounds (wierden).
- TNO recommends the use of FCM version TNO-2020 in the public SHRA Groningen 2023. 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

# 1 Introduction

As of 2021, the public 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 fulfil this task, TNO makes use of the internally developed model toolkit: the TNO Model Chain Groningen (TNO, 2019; 2020c).

This report describes the results of ongoing model development in relation to the SHRA of (nearly) finished components (chapter 2), the status of the TNO Model Chain per October 1, 2022 (chapter 3), as well as TNO’s recommendations for model components to be used in the public SHRA Groningen 2023 (chapter 4). The scope of this report 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 based on the latest scientific insights to run the public Seismic Hazard and Risk Analysis Groningen 2023 for the “Vaststellingsbesluit gasjaar 2023-2024”

# 2 Overview model development 2022

## 2.1 GMM-V7

TNO has implemented the GMM-V7 ground motion model (Bommer et al., 2022) in the TNO Model Chain Groningen. This model incorporates substantial changes with respect to the earlier model GMM-V6. A detailed overview of the implementation of these changes is provided in TNO (2022b). The new features that have the largest impact on the implementation include the expansion of the logic tree of epistemic uncertainties, the inclusion of dedicated amplification functions for ground motions on *wierden* (dwelling mounds), and the conditioning of the median branch weights in the logic tree on earthquake magnitudes. The latter feature actually leads to both conceptual and practical difficulties for determining the logic tree (epistemic) percentiles (i.e., P90) of the risk distribution (TNO, 2022b). As a result, TNO does not provide risk percentiles in combination with the GMM-V7 ground motion model. As an alternative to the use of P90 as a safety margin on the risk assessment, TNO suggest to apply a safety margin to the risk norm.

## 2.2 Test & comparison framework

The Testing and Comparison Framework (TCF) for the TNO Model Chain Groningen has been implemented and can be utilized within the traceability and reproducibility framework that is used for all hazard and risk calculations. At the moment, the TCF comprises nine statistical performance tests related to Seismological Source Models (SSMs). These tests can be divided into consistency tests and comparative tests. The tests within the TCF are largely based on the CSEP publications (Collaboratory for the Study of Earthquake Predictability, 2022). More information on the TCF is provided in Appendix A of this report.

## 2.3 Magnitude-frequency model with spatially dependent b-value

TNO has developed an alternative magnitude-frequency (b-value) model that describes the Gutenberg-Richter b-value as a step function of the reservoir thickness. The description of the model, including significance and performance assessments on Groningen data can be found in the manuscript by Kraaijpoel et al. (2022). The new model performs better than the hyperbolic tangent b-value model conditioned on stress that was used in the HRA 2020 and the pSHRA 2021 & 2022. Kraaijpoel et al. show that the maximum likelihood fits of both (currently used & new proposed) models, that employ an equal number of parameters, to the full catalogue result in a relative likelihood of about 50 in favour of the new thickness model. The manuscript, which has been accepted for publication, is provided as Appendix B.

## 2.4 Rate Type Compaction Model inspired SSM

A time-delay compaction model has been implemented in the seismic source model (SSM). This alternative to the currently used linear elastic compaction model may be able to better predict main shock events for constant or increasing reservoir pore pressure expected during shut-in. The compaction model used is the Rate Type isotach Compaction Model (RTiCM) (de Waal, 1986; Pruiksma et al., 2015). This model is the current state of the art for subsidence predictions in Groningen. The model parameters have recently been calibrated on surface subsidence data (NAM 2021). More information on the model implementation and model performance (TCF) is given in Appendix C of this report.

## 2.5 Event rate analysis pSHRA Groningen

The predicted earthquake event rate and the observed event rate in the Groningen gas field were analysed. We answered two questions:
1. How well has the SSM predicted the observed event rate in the past?
2. Are there clear signs that the SSM is underpredicting the observed event rate in recent years?

Subjected to these questions are the pSHRA calibration prescribed by EZK and provided by NAM, and TNO’s implementation of the SSM calibration.
Details of the analysis are given in Appendix D of this report.

In summary our analysis shows that:
1. The predicted pSHRA-2022 NAM calibration based event rate is significantly lower than the observed one. Moreover, it is unable to predict the correct number of events it was calibrated on.
2. It is very unlikely that the pSHRA-2022 NAM calibration based event rate predicts the observed data. These conclusions support and expound on the observation of SodM (2021).
3. The TNO implementation for SSM and accompanied calibration procedure outperforms the for pSHRA-2022 prescribed SSM implementation and calibration in predicting the total number and rate of earthquake events.
4. However, in the most recent recorded gas years we observe a large disparity between predicted and observed event rate that is highly unlikely to be labelled as variations within the model prediction bandwidth. This raises concerns on the SSM prediction performance for the post-production phase of the Groningen field.
5. In anticipation of a better alternative, we recommend to use the TNO implementation of the SSM.

## 2.6 Mmax II Workshop

The Second Workshop on Mmax for Seismic Hazard and Risk Analysis in the Groningen Gas Field has taken place in Amsterdam on June 13-17, 2022. The results and presentations have been made public in September 2022 (NAM, 2022). TNO has attended this expert elicitation workshop with six persons in total, of which two persons have presented.

Sander Osinga has presented, as a “resource expert” (for terminology see NAM, 2022), the concerns raised by TNO (e.g., TNO 2022a) with respect to the exponential taper that is present in the current pSHRA magnitude-frequency model. This taper relies on an additional parameter for the description of the tail of the magnitude distribution, which, to a large extent, reduces the effect of the Mmax parameter on the outcome. Like Mmax, this taper parameter cannot be constrained from the Groningen data. Hence, similar to Mmax, its value is largely determined by prior assumptions. Unlike Mmax, however, these prior assumptions have up until now not been scrutinized by expert elicitation. In fact, the prior assumptions have not been explicitly discussed or justified by the model developers.

Loes Buijze has presented, as a “proponent expert”, her work in cooperation with Université cote d’Azur, Nice, on dynamic rupture modelling in the Groningen Field. The outcome of these synthetic experiments led her to propose an Mmax range of 4.1-4.5 for ruptures that remain within the confines of the reservoir. In addition, she noted that, based on her modelling, the probability that the geomechanical conditions are such that earthquake ruptures may be able to break out of the confines of the reservoir is relatively small, in the order of a few percent.

TNO has studied the final report and findings of the Mmax expert panel, including the update Mmax distribution (NAM, 2022). We note that the panel ultimately did not address the issue concerning the exponential taper as raised by TNO. We consider this a missed opportunity. In fact, we are of the opinion that the fourth point of the workshop objectives (*4. To determine if the proposed Mmax distribution compatible with the existing PSHRA framework for Groningen, including the V6 seismological model and the logic tree.*, NAM, 2022, p.7) was not addressed in the report at all. A major question with regard to the recurrence of large earthquakes within the Groningen Field now remains open.

# 3 Technical Status per 1 October 2022

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 is 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, the model components 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. Below the implemented model versions are represented in a tabular format with reference to the model documentation, implementation report and model usage in past HRA/pSHRA's.

## 3.1 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.

### 3.1.1 *NAM-V6*

The SSM version NAM-V6 was used for the HRA 2020 and pSHRA 2021 & 2022. An overview of this model and the different sub models are listed in Table 1. 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 1 Overview of SSM version NAM-V6

| **SSM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| NAM-V6 | Bourne et al., 2019 | TNO, 2020c | | x | x | x |
| Sub models | | | | | | |
| NAM-Model calibration provided as input | | not part of HRA | x | x | x | x |
| Coulomb stress predictor for activity rate<br>Coulomb stress predictor for to magnitude distribution | | TNO, 2020c | | x | x | x |
| Activity rate | | TNO, 2020a | x | x | x | x |
| ETAS | | TNO, 2020a | x | x | x | x |
| MD: hyperbolic tangent b-value & Mmax distribution | | TNO, 2020c | | x | x | x |
| MD: single b-value & exponential taper & Mmax distribution | | TNO, 2020c | | x | x | x |

### 3.1.2 *TNO-2020*

SSM version TNO-2020 was proposed by TNO for usage in the public SHRA 2021 and 2022. An overview of this model and the different sub models are listed in Table 2. The TNO-2020 implementation is a selection of sub-models that were previously used in HRA and pSHRA runs and alternative implementations of sub-models based on the original documentation (Bourne et al., 2019). These implementations are extensively documented (TNO, 2020a; 2020c).

Table 2 Overview of SSM version TNO-2020

| **SSM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| TNO-2020 | Bourne et al., 2019; TNO, 2022a | 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 | x |
| ETAS | | TNO, 2020a | x | x | x | x |
| MD: hyperbolic tangent b-value & Mmax distribution | | TNO, 2020c | | x | x | x |

The main difference relative to version NAM-V6 is the removal of the tapered frequency-magnitude model (MD). An in-depth justification for this is provided in TNO, 2022a.
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. The calibration approach that was already implemented in the first public version of the TNO Model Chain (TNO, 2020a) is extensively described and assured by an external review panel (TNO, 2022a).

### 3.1.3 *All available SSM components*

Table 3 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 sequence version.

New implemented components are the magnitude-frequency model with spatial dependent b-value (Appendix B) and the activity rate model based on the Rate Type Compaction Model (Appendix C).

Table 3 Overview of all available SSM sub models & version

| **SSM: all sub models** | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|
| | | 2019 | 2020 | 2021 | 2022 |
| NAM-Model calibration provided as input | not part of HRA | x | x | x | x |
| TNO-Model calibration | TNO, 2020a | | | | |
| Single Coulomb stress conditioned to activity rate | TNO, 2020a | x | | | |
| Coulomb stress predictor for activity rate<br>Coulomb stress predictor for magnitude distribution | TNO, 2020c | | x | x | x |
| Coulomb stress distribution | TNO, 2020a | | | | |
| Activity rate | TNO, 2020a | x | x | x | x |
| Activity rate RTiCM | Appendix C | | | | |
| ETAS | TNO, 2020a | x | 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 | x |
| MD: single b-value & exponential taper & Mmax distribution | TNO, 2020c | | x | x | x |
| MD: spatially dependent b-value & Mmax distribution | Appendix B | | | | |

### 3.1.4 *Mmax distribution*

The MD models used in the past HRA/pSHRA versions that are truncated by an Mmax distribution applied the outcome of the Mmax expert elicitation of 2016 (NAM, 2016). In 2022 the Mmax distribution was revised by the same expert panel that proposed the initial distribution in 2016 (NAM, 2022).
Any discrete distribution of Mmax with accompanied weighting values can be applied within the pSHRA Groningen framework.

## 3.2 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.

### 3.2.1 *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 4 Overview of GMM version NAM-V6

| **GMM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| 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.

### *3.2.2 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 (e.g. TNO, 2022a). Remarkably the successor version NAM-V7 deviates from this correlation assumption to a full period-to-period correlation for the amplification/attenuation functions of the site response model.

Table 5 Overview of GMM version NAM-V6-2021

| **GMM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| NAM-V6-2021 | Bommer et al., 2019 | TNO, 2020c | | | x | x |

### *3.2.3 NAM-V7*

The GMM NAM-V7 model (Bommer et al., 2022) builds upon its predecessor, but differs substantially 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. Improvements of this model include a revision of the damping model and the explicit site amplification for buildings located on dwelling mounds (wierden).

The TNO implementation of GMM-V7 is reported in TNO (2022b). The most impactful feature of the model is the conditioning of the median branch weights in the logic tree on magnitude. As explained in TNO (2022b) this compromises the calculation of a logic tree risk distribution, and therefore, the determination of percentiles (such as P90) within this distribution.
If P90 and compatibility with earlier risk assessments is regarded important, the use of GMM NAM-V6 can be considered. As an alternative to the use of a risk percentile as a safety margin, a safety margin on the risk norm is recommended.

Table 6 Overview of GMM version NAM-V7

| **GMM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| NAM-V7 | Bommer et al., 2021 | TNO, 2022b | | | | |

## 3.3 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.

### 3.3.1 *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 7 Overview of FCM version NAM-V7

| **FCM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| NAM-V7 | Crowley and Pinho, 2020 | TNO, 2020a | | x | x | x |

### 3.3.2 *TNO-2020*

FCM version TNO-2020 was proposed by TNO (2020d) for usage in the public SHRA as of 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 TNO, 2022a.
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* commissioned by the Ministry of Economic Affairs and Climate Policy, and independently reviewed by the Advisory Board Safety Groningen (ACVG).

Table 8 Overview of FCM version TNO-2020

| **FCM** | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| version | | | 2019 | 2020 | 2021 | 2022 |
| TNO-2020 | TNO, 2022a | TNO, 2020a | | | | |

## 3.4 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). Already in 2021 this version was complemented by NAM with information on *wierden* required for GMM-V7. It is currently unclear if an updated version of the EDB is to be expected in 2023.

# 4 Recommendations for the public Seismic Hazard and Risk Analysis (SHRA) Groningen 2023

Based on the available information as of October 1, 2022, TNO recommends the following versions of (sub) models as most suitable for use in the public SHRA Groningen 2023 (Table 9). 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 qualitatively described in this chapter. More in depth analysis is available in the pSHRA Groningen 2022 (TNO, 2021) and the results of the KEM-09 research project (TNO, 2022c).

Table 9 Overview of recommended model versions public SHRA Groningen 2023

| model version | documentation | TNO software implementation | HRA | | pSHRA | |
|---|---|---|---|---|---|---|
| | | | 2019 | 2020 | 2021 | 2022 |
| **SSM** | | | | | | |
| TNO-2020 | Bourne et al., 2019, TNO, 2022a | 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 | x |
| ETAS | | TNO, 2020a | x | x | x | x |
| MD: hyperbolic tangent b-value & Mmax distribution | | TNO, 2020c | | x | x | x |
| **GMM** | | | | | | |
| NAM-V7 | Bommer et al., 2022 | TNO, 2022b | | | | |
| **FCM** | | | | | | |
| TNO-2020 | TNO, 2022a | TNO, 2020a | | | | |

## 4.1 Recommended SSM

TNO recommends the use of version TNO-2020 in the public SHRA Groningen 2023, as described in paragraph 3.1.2. This approach uses a full posterior distribution of stress covariate fields and uses a magnitude distribution that is truncated by a Mmax distribution.
The overall impact of this model version on the hazard and risk analysis is described in the pSHRA Groningen 2021 & 2022 (TNO,2021; 2022d).

TNO does not (yet) recommend the use of the new SSM activity rate model with the RTiCM implementation. Results of the TCF test suite show no distinct gain in model performance compared to the current (elastic) activity rate model. Besides the results of the comparative test, some of the consistency tests do not pass depending on the applied forecast period. The two models have been tested on periods of significant production, whereas the RTiCM is implemented with the post-production phase in mind. We therefore aim in the near future to extent the forecast period to the most recent event data, because in the last years the some parts of the field have transitioned into the post-production phase (pressure increase in conjunction with observed events). For further testing a pending update of the ETAS treatment in the forecast module is required.

With regard to the newly developed magnitude-frequency model with (reservoir) thickness-dependent b-value, results of the conducted performance assessment indicate superiority of this model compared to the current and alternative magnitude-frequency models. However, a pseudo-prospective model performance test in a Bayesian model inference context such as the Test and Comparison Framework is not yet completed and some further testing is needed to come to a deliberate proposal of model weighting factors for use in the pSHRA Groningen.

#### 4.1.1 *Calibration*

TNO recommends to include the most recent earthquake observations to perform the public SHRA Groningen 2023. 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. The TNO calibration approach is extensively described in TNO 2022a.
TNO emphasizes that the externally delivered calibration for use in pSHRA 2021 and 2022 could not be reproduced based on current NAM documentation. Event rate analysis of the NAM provided calibration results (Appendix D) show that it is very unlikely that the NAM calibrated event rate predicts the observed data. Therefore TNO considers the use of this NAM provided calibration result under no circumstance a feasible scenario within the pSHRA Groningen.

#### 4.1.2 *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.

#### 4.1.3 *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 Mmax. This is described in more detail in TNO, 2022a.

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& 2022, should not be used in the 2023 pSHRA Groningen.
This is substantiated by the proposal to use only the hyperbolic tangent Magnitude Model present in NAM-V6.

According to TNO two steps should be taken before considering the stress-dependent exponential taper model for usage in the pSHRA Groningen. The most important step is to propose a well-motivated informed prior distribution to derive the taper location. TNO sees no possibility to obtain such a distribution, but others may. TNO is open to assess such proposals. The second step is to perform a pseudo-prospective model performance test to determine a weighting for this specific model in the logic tree. The second step can only be executed after the first step is completed.

#### *4.1.4 Mmax distribution*

In the past several points of attention were raised by TNO and others concerning the 2016 Mmax distribution and its application in the pSHRA Groningen. Some concerned the appropriateness of the weighting of branches in view of recent (geomechanical) insights, like the *propagate significantly out of reservoir* and the *analogs* branches.
Others concerned the compatibility and implementation of the Mmax distribution with the tapered frequency magnitude distribution as part of the seismological model. Current Mmax implementation assumes an independence between the shape of the frequency magnitude distribution (FMD) and Mmax distribution. At the same time at least two different earthquake source mechanisms (induced & triggered) are distinguished within the Mmax distribution. These different source mechanisms do not necessarily share the same (shape of the) frequency magnitude distribution. Besides this aspect, the introduction of a data-calibrated tapered frequency magnitude distribution in the V6 seismological model resulted effectively in sidestepping any defined Mmax distribution, because a very dominant weighting of 80% in the logic tree is assigned to this tapered frequency magnitude distribution.

Unfortunately these compatibility aspects were only sparsely discussed during the workshop and not addressed in the final reporting of the 2022 Mmax workshop. Although new weights and logic tree structure were proposed to give substance to new scientific insights with respect to the occurrence of earthquakes outside the reservoir and the absent value of analogue fields, no effort has been made to describe the (for SHRA) equally important tail description of the frequency magnitude distribution.

A general observation from TNO is that the panel report (NAM, 2022) does not explicitly justify the update of the weights in the distribution in relation to the new data, modelling results and insights communicated during the workshop sessions, and in what way the workshop participants contributed to proposed distribution. As a consequence, the outcome of the 2022 Mmax workshop appears to be based on a weaker scientific foundation than the results of the 2016 Mmax workshop.

The update of the Mmax distribution relative to the previous distribution (NAM, 2016) itself appears to be in line with our perception of the presentations and discussions. And it appears that the final distribution accommodates all estimates presented during the workshop. Many proponent and resource experts showed that according to their analysis it would be difficult for earthquake ruptures to extend out of the reservoir. As a result, a shift to lower Mmax values is justified. Also, it

appears reasonable that there are no adequate analogues for Groningen seismicity among other induced seismicity cases around the world.
Despite the scientifically sparsely motivated conclusions and substantiation of the Mmax distribution in the report, TNO preliminary advises to consider the 2022 Mmax distribution for use in pSHRA Groningen – provided that the tail description of the frequency magnitude distribution will be subject of further research.

### 4.2 Recommended GMM

TNO recommends the use of GMM version NAM-V7 in the public SHRA Groningen 2023. This model version is a next iteration based on NAM-V6. Improvements of this model include among other things a revision of the damping model and the explicit site amplification for buildings located on dwelling mounds (wierden). The model was calibrated with the latest Groningen ground motion dataset. Although TNO is unable to judge the quality of the performed model calibration process due to the unavailability of the underlying (sub) models and data in the public domain, the large number of scientists from different institutions that contributed to the model development gives confidence in the quality of the newly developed model.

### 4.3 Recommended FCM

TNO recommends the use of FCM version TNO-2020 in the public SHRA Groningen 2023. 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’.

# 5 References

Arup (2021a). EDB V7.1 Cover Note. File reference 229746_031.0_NOT2101, 1 February 2021.

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 (2022). V7 Ground-Motion Model for Induced Seismicity in the Groningen Gas Field (Revision 1.1), 26 February 2022.

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. (2020). Report on the Fragility and Consequence Models for the Groningen Field (version 7), March 2020.

Collaboratory for the Study of Earthquake Predictability (2022): Theory of CSEP Tests — pyCSEP v0.6.0 documentation (cseptesting.org)

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. Pruiksma, A. Tsouvalas, KEM 02 Research Question 2 report, Ministry of Economic Affairs and Climate Policy, 1 April 2020.

Kraaijpoel, D., J. Martins, S. Osinga, B. Vogelaar, J. Breunese (2022). Statistical analysis of static and dynamic predictors for seismic b-value variations in the Groningen gas field. Accepted for publication in Netherlands Journal of Geosciences.

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, (2021). Groningen long term subsidence forecast. EP202008201822

NAM, (2022). Report on the Second Workshop on Mmax for Seismic Hazard and Risk Analysis in the Groningen Gas Field. 13-17 June 2022. https://nam-onderzoeksrapporten.data-app.nl/reports/download/groningen/en/77951661-552a-46bc-9f2e-f1580cd6abc3

Pruiksma, J. P., Breunese, J.N., van Thienen-Visser, K., and de Waal., J.A. (2015). Isotach formulation of the rate type compaction model for sandstone, International Journal of Rock Mechanics and Mining Sciences 78, June 2015, doi: 10.1016/j.ijrmms.2015.06.001.

SodM (2021). Beoordeling SodM halfjaarrapportage seismiciteit Groningen, overschrijding grenswaarde aardbevingsdichtheid & beving Garrelsweer. 9 dec 2021.

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.

TNO (2022a). Status of the TNO Model Chain Groningen per October 1, 2021 and recommendations for the public Seismic Hazard and Risk Analysis 2022, TNO 2021 R11742|update 2022, 28 April 2022.

TNO (2022b). Implementation of GMM-V7 in the TNO model chain, TNO 2022 R10801|final draft, 18 July 2022.

TNO (2022c). KEM9: Cumulative propagation effect of Groningen risk model component uncertainties on hazard and risk predictions, TNO 2021 R12442, 9 September 2022.

TNO (2022d). Publieke Seismische Dreigings- en Risicoanalyse Groningen gasveld 2022, TNO 2022 R10517, 25 maart 2022.

De Waal, J. A. (1986) On the rate type compaction behaviour of sandstone reservoir rock.

# 6 Signature

Utrecht, 18 oktober 2022 TNO

[redacted]

Head of department Advisory Group for Economic Affairs

# A Testing and Comparison Framework

The Testing and Comparison Framework (TCF) for the TNO Modelchain Groningen has been implemented and can be utilized within the traceability and reproducibility framework that is used for all hazard and risk calculations. At the moment, the TCF comprises a number of statistical performance tests related to Seismological Source Models (SSMs). These tests can be divided in two categories: 1) consistency tests, which compare the model predictions with the data, and 2) comparative tests, which compares the predictive performance of 2 different models. The tests within the TCF are largely based on the CSEP (Collaboratory for the Study of Earthquake Predictability, 2022)

All these tests are based on some shared principles:

1. The available data is divided into a calibration dataset and a testing dataset. This division is made in the time-domain, in such a way that the calibration data precedes the testing data.
2. An SSM is calibrated using the calibration data. For comparative tests, a set of SSMs is calibrated using the same calibration data.
3. The calibration is used to produce a forecast for the testing period. For comparative tests, the SSMs produce a forecast based on the same calibration data, and for the same forecast period.
4. The test in question is performed.

This general procedure allows us to ask the questions:

1. “How well does the model in question predict the seismicity in the testing period, if we calibrate it over the calibration period?”. This question is answered by performing consistency tests (see A.1 below)
2. “Which model does a better job of predicting the seismicity in the testing period?” This question is answered by performing comparative tests (see A.2 below).

## A.1 Consistency tests

Consistency tests confront the model forecasts with the data in order to assess whether the models are sufficiently probable given the data. In other words, the tests assess whether a model must be rejected based on the data, because it is judged to be sufficiently unlikely for the data to occur under the model (with a significance level $\alpha = 5\%$). In Figure A-1 a visual example of the results of all the comparative tests is given.

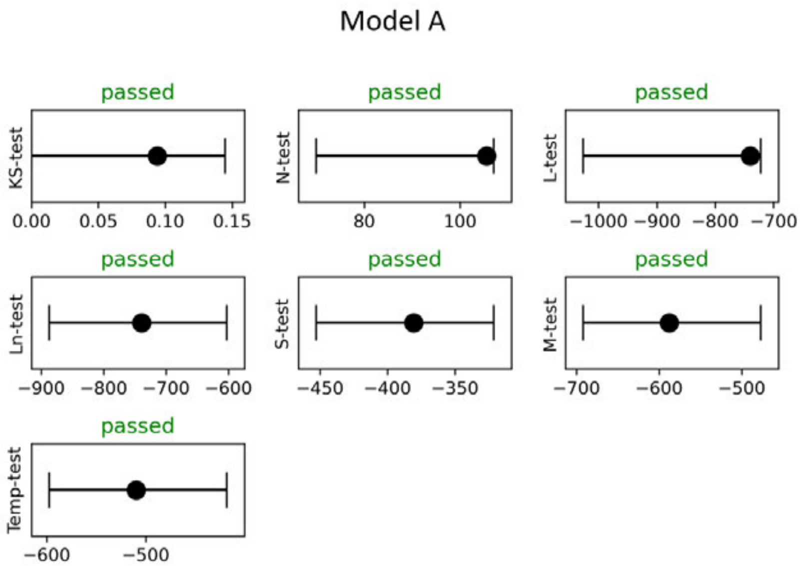


Figure A-1 Example of a result of the seven different consistency test. For each test the test name is indicated on the vertical axis. The horizontal axis represents the likelihood score of the test. The test result (likelihood) is indicated by the black dot. The 95% significance bandwidth is indicated by the horizontal line bounded by vertical markers.

### *A.1.1 Likelihood test (L-test)*

This test, which is described in Schorlemmer et al. (2007) is based on the full forecast in space, time and magnitude. The forecast consists of an X-Y-time-magnitude grid, where each bin is assigned an expected rate over the forecast period. When generating a large ensemble of catalogues from this forecast, each catalogue can be used to calculate a likelihood score for the model. The resulting distribution of likelihoods describes the expectation of likelihood scores if the model is generating the data. If the likelihood of the model under the actually observed catalogue falls within the central 95% of the likelihood range, the model passes the L-test. If the model falls in the bottom 2.5%, the model is rejected. If the model falls in the top 2.5%, it is not rejected but is marked as suspect. After all, the observed data is somehow more likely under the model than at least 97.5% of the catalogues that are generated by the model itself. This likely indicates that the model is underpredicting the total number of events, and that it will fail the N-test (see A.1.2).

### *A.1.2 Number test (N-test)*

This test, which is described in Schorlemmer et al. (2007) is based on the summation of the full forecast over space, time and magnitude. Once the model is summed over all dimensions, there only remains an expected rate $\lambda_{exp}$ over the forecast period. The range of the expected number of events under the model is given by a Poissonian distribution with rate parameter$\lambda_{expected}$ in combination with a significance level $\alpha$. When the number of observed events falls outside of the expected range, the N-test rejects the model.

### A.1.3 *Conditional Likelihood test (Ln-test).*

This test is described in Werner et al. (2011) and is a simple modification of the L-test. The rates in the forecast are first adjusted, such that their sum corresponds to the number of observed events. Subsequently, the test is performed as normal.

### A.1.4 *Magnitude test (M-test).*

This test is described in Zechar et al. (2010) and focusses on the consistency of observed event magnitudes with forecasted magnitudes. The test is performed in a fashion similar to the L-test, but the forecast is first summed over all dimensions except the dimension containing the magnitudes. As with the L-test, a range of ‘expected’ likelihood scores is generated, and the likelihood score of the model under the actually observed catalogue is compared to this range.

### A.1.5 *Spatial test (S-test).*

This test is described in Zechar et al. (2010) and focusses on the consistency of observed event locations with forecasted locations. The test is performed in a fashion similar to the L-test, but the forecast is first summed over all dimensions except the spatial dimensions. As with the L-test, a range of ‘expected’ likelihood scores is generated, and the likelihood score of the model under the actually observed catalogue is compared to this range.

### A.1.6 *Temporal test (Temp-test).*

This test is not described in the context of CSEP and focusses on the consistency of observed event timing with forecasted timing. Since CSEP tests were designed with natural seismicity in mind, a temporal variation in the forecast was not considered. However, for induced seismicity, our models often contain a temporal dimension. The addition of this test of the temporal dimension fits in the general philosophy of CSEP, and is added to establish insight in the temporal performance of the model.

The test is performed in a fashion similar to the L-test, but the forecast is first summed over all dimensions except for the time dimension. As with the L-test, a range of ‘expected’ likelihood scores is generated, and the likelihood score of the model under the actually observed catalogue is compared to this range.

### A.1.7 *Kolmogorov-Smirnov test (KS-test)*

This test is described in Rhoades et al. (2011) but is currently not part of the CSEP curriculum. We have added it to the TCF to provide additional insight.

As stated before, the forecast consists of an X-Y-time-magnitude grid, where each bin is assigned an expected rate over the forecast period. These expected rates form a distribution which can be confronted with the distribution of the observed events over these bins. For example, you would expect approximately 50% of the events to occur in the bins which together constitute 50% of the expected rate. Some variability around this is expected, and the amount of variability that is expected is known. The KS-test assesses whether the distribution of observed events over the bins is in line with the forecast.

## A.2 Comparative tests

Comparative tests are performed to assess whether one model performs significantly better during the forecast period than another model. Two tests, the T-test and the W-test, are applied to make this assessment. In Figure A-2 a visual example of the results of the two comparative tests is given.

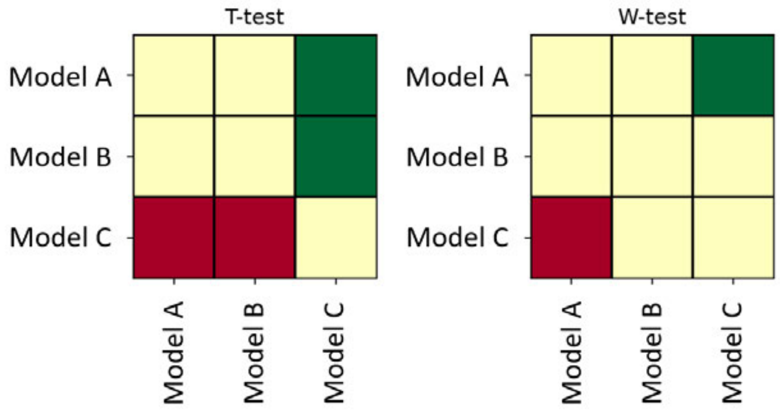


Figure A-2 Example of comparative test results. A green field indicates that the row model outperforms the column model. A red field indicates that the column model outperforms the row model.

### A.2.1 *T-test*

This test, described in Rhoades et al (2011), compares two models against each other. Suppose we have model A and model B, which predict a total number of events $\hat{N}_A$ and $\hat{N}_B$ respectively. If we write $X_i = \ln \lambda_A (k_i)$ and $Y_i = \ln \lambda_B (k_i)$, where $\lambda_A(k_i)$ is the rate of model A in the bin where event $k_i$ occurs, and $\lambda_B(k_i)$ is the rate of model B in the bin where event $k_i$ occurs, we can compute

$$I_N(A,B) = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (X_i - Y_i) - \frac{\hat{N}_A - \hat{N}_B}{N_{obs}}$$

where $I_N(A,B)$ is the mean sample information gain per earthquake. If $I_N(A,B)$ differs significantly (according to a classical paired T-test) from zero, the model with the lower likelihood can be rejected in favor of the other. Otherwise, the models perform similarly, and neither model can be rejected in favor of the other.

### A.2.2 *Wilcoxon signed-rank test (W-test)*

This test, described in Rhoades et al (2011), is a weaker version of the T-test, which can be applied if the differences between $X_i$ and $Y_i$ do not appear to be normally distributed. In that case, we test whether the median of $X_i - Y_i$ is sufficiently close (according to the Wilcoxon signed-rank test) to $(\hat{N}_A - \hat{N}_B)/N_{obs}$.

According to Rhoades et al. (2011), the T-test is dependable for larger numbers of observed events. For small numbers of observed events, the conservative approach would be to only accept one model in favor of another if this conclusion is supported by both the T-test and the W-test. It is difficult to define an exact value for which it is acceptable to accept one model in favor of another based on the outcome of only one of the tests. We therefore recommend to only accept one model in favor of another if this conclusion is supported by both tests, regardless of the number of observed events.

## A.3 References

Collaboratory for the Study of Earthquake Predictability (2022): Theory of CSEP Tests — pyCSEP v0.6.0 documentation (cseptesting.org)

Rhoades, D.A., Schorlemmer, D., Gerstenberger, M.C., Christophersen, A., Zechar, J.D., Imoto, M. (2011). Efficient testing of earthquake forecasting models. *Acta Geophysica 59(4), 728 - 747.* doi:10.2478/s11600-011-0013-5

Schorlemmer, D., Gerstenberger, M.C., Wiemer, S., Jackson, D.D., Rhoades, D.A. (2007). Earthquake Likelihood Model Testing. *Seismological Research Letters, 78(1), 17 - 29.* doi:10.1785/gssrl.78.1.17

Werner, M.J., Helmstetter, A., Jackson, D.D.; Kagan, Y.Y. (2011). High-Resolution Long-Term and Short-Term Earthquake Forecasts for California. *Bulletin of the Seismological Society of America, 101(4), 1630 - 1648.* doi:10.1785/0120090340

Zechar, J.D., Gerstenberger, M.C., Rhoades, D. A. (2010). Likelihood-Based Tests for Evaluating Space-Rate-Magnitude Earthquake Forecasts. *Bulletin of the Seismological Society of America, 100(3), 1184 –1195.* doi:10.1785/0120090192

# B Statistical analysis of static and dynamic predictors for seismic b-value variations in the Groningen gas field

Paper accepted for publication in Netherlands Journal of Geosciences.

# C Implementation of Rate Type isotach Compaction Model

In the wider context of future-proofing the public SHRA Groningen model chain for a period of shut-in, a rate dependent compaction model has been implemented in the seismic source model (SSM). This alternative to the currently used linear elastic compaction model may be able to better predict main shock events for constant or increasing reservoir pore pressure expected during shut-in. The compaction model used is the Rate Type isotach Compaction Model (RTiCM) (de Waal, 1986; Pruiksma et al., 2015). This model is the current state of the art for subsidence predictions in Groningen, and the model parameters have been recently calibrated on surface subsidence data (NAM 2021).

## C.1 Synopsis of RTiCM

The RTiCM provides an empirical description for the change of the compaction coefficient (slope of stress-strain curve) with stress rate as observed in laboratory experiments (de Waal, 1986; Pruiksma et al., 2015):

$$c_{\mathrm{m}} = c_{\mathrm{m,ref}} \left( \frac{\dot{\sigma_{\mathrm{ref}}}}{\dot{\sigma}} \right)^{b},$$

where $b$ is some to-be-fitted empirical constant, typically between 0.005 and 0.025. For a rate-dependent strain (compaction), the assumption is that the loading path follows a rate-dependent isotach given by the compaction coefficient above. A change in loading rate shifts the stress-strain path to another isotach. All isotachs intersect in one reference point at the beginning of loading. The vertical compaction is the sum of a direct instantaneous (i.e. elastic) compaction strain, and a creep compaction strain, as a function of stress and time is given by this set of equations:

$$\varepsilon_{\mathrm{total}} = \varepsilon_{\mathrm{s}} + \varepsilon_{d},$$

$$\varepsilon_{\mathrm{d}} = \mathrm{c_m} \sigma,$$

$$\varepsilon_{\mathrm{s}} = \dot{\sigma_{\mathrm{ref}}} \left( \mathrm{c_m} - \mathrm{c_{m,d}} \right) \left( \frac{\mathrm{c_m}}{\mathrm{c_{m,ref}}} \right)^{-1/b},$$

$$\mathrm{c_m} = (\varepsilon_{\mathrm{total}} + \sigma_{\mathrm{ref}} \, \mathrm{c_{m,ref}})/\sigma.$$

This set of equations relies on three material parameters ($b$, $\mathrm{c_{m,d}}$, $\mathrm{c_{m,ref}}$) and one state parameter ($\dot{\sigma_{\mathrm{ref}}}$). The compaction coefficient parameters are a reference coefficient $\mathrm{c_{m,ref}}$, corresponding to $\dot{\sigma_{\mathrm{ref}}}$, and the instantenous (elastic) coefficient $\mathrm{c_{m,d}}$. An explicit numerical scheme is used to solve above equations (see Pruiksma et al., 2015), assuming that the pressure history (stress history) is known, and that at $t = 0$, direct (i.e. elastic) and creep strains ($\varepsilon_d$ and $\varepsilon_s$, respectively) are zero.

## C.2 RTiCM compaction behavior

Two simple pressure reduction histories are used to illustrate the difference between the linear elastic compaction model and Rate Type compaction. One pressure history has a symmetric trapezoidal shape (solid black line in Figure C-1, top panel), representing an extreme end-member of extraction and full pressure recovery. The other pressure history has a small pressure recovery after an initial large pressure reduction (dashed black line, Figure C-1, top panel), representing a more realistic pressure path as may occur in the Groningen field. The (normalised) compaction response is shown in the panel below, where elastic compaction (red curves) follows the pressure histories. The Rate Type compaction history (blue curves) shows continued creep (caused by $\varepsilon_s$) during 'shut-in' after the initial pressure reduction. During the large pore pressure increase (solid blue line), there is less elastic decompaction. After the large pore pressure increase, no additional creep is observed. During the small pore pressure increase (blue dashed curve), we see that elastic decompaction is initially dominant. After (at constant pore pressure), the creep strain increases total compaction.

The reservoir compaction is linked to seismicity: In the SSM, compaction results in fault loading through Coulomb stress increase, and causes seismicity when a Coulomb failure threshold is reached. In its current implementation following Bourne & Oates (2017), the compaction is calculated as a linear elastic response to pressure change. The elastic modulus is given as a spatial static compaction coefficient grid, provided by NAM. Assuming that not only elastic, but all compaction strain causes fault loading, Rate Type compaction behaviour would result in delayed seismic activity from pore pressure reduction, even during periods of stable pore pressure. Moreover, creep-induced loading of faults and subsequent seismicity may continue while pore pressure increases by small amounts (i.e., $\varepsilon_s$ is larger than $\varepsilon_d$).

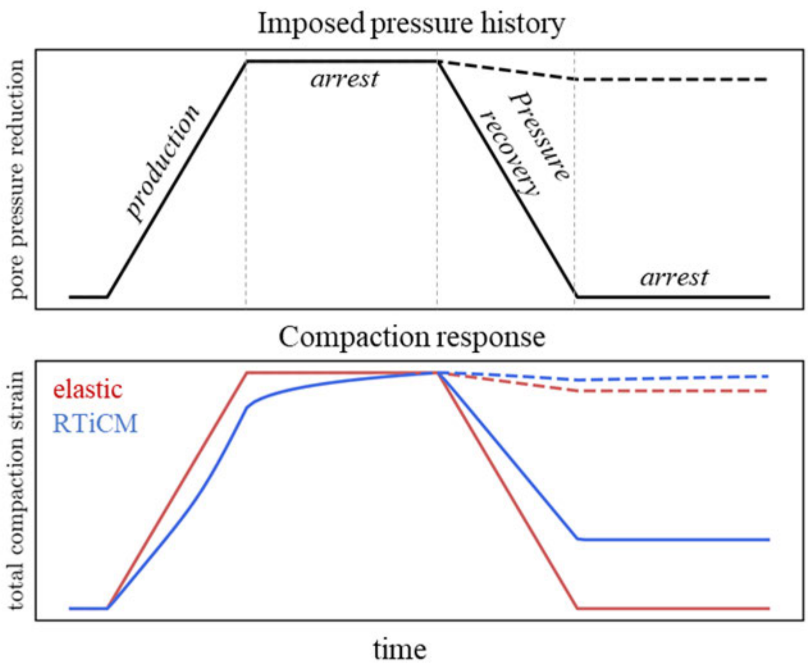


Figure C-1 Simple pressure histories (top panel), and the linear elastic (red curves) and Rate Type (blue curves) compaction responses. Note that the compaction curves have been normalised by their maximum value.

## C.3 Implementation in SSM

The RTiCM is implemented in the dynamic subsurface module (DSM) of the SSM, without substantially changing the SSM as developed by Bourne & Oates (2017). We assume that the total strain (elastic plus creep) results in loading of the faults, i.e., is proportional to incremental Coulomb stress change. For the implementation and relation to the activity rate model (ARM), we follow the precedent set by NAM (2018), where the authors implement a visco-elastic rheology to study seasonal effects on seismicity. In this visco-elastic model, equivalent to the RTiCM, the total strain is the sum of a direct elastic strain component and a viscous strain component. For a strain that is the sum of its constituents, the cumulative event rate obtained using an extreme threshold failure model (equation 69 in Bourne and Oates, 2017) is written as (NAM, 2018):

$$\Lambda = h\theta_0(\mathcal{M}_d\mathcal{M}_s - 1),$$

$$\mathcal{M}_d = e^{\theta_1\varepsilon_d},$$

$$\mathcal{M}_s = e^{\theta_2\varepsilon_s},$$

with $\theta_1$ and $\theta_2$ the (model) parameter terms for the elastic and plastic parts of the strain. $\mathcal{M}_d$ gives the instantaneous activity rate to a pressure change, $\mathcal{M}_s$ gives the activity rate related to longer-term reservoir creep. For $\mathcal{M}_s = 1$ (i.e., zero creep strain), the activity rate reverts back to the activity rate for a linear compaction model as is currently implemented in the SSM.

We use the same values for the model parameters $\theta_1$ and $\theta_2$ , so that creep and elastic strain have the same contribution to seismicity rate. The value for these model parameters is formulated the same as the equivalent model parameter in the linear compaction SSM by Bourne & Oates (2017). The direct and creep strains are calculated using the explicit numerical scheme of Pruiksma et al. (2015), using the known pressure grids. The reference stress that is required is given as the overburden pressure minus initial gas pressure: $\sigma_{\mathrm{ref}} = \rho\, g\, d_{\mathrm{top}} - p$, with $d_{\mathrm{top}}$ the depth of the top of the reservoir. A spatial top-reservoir depth map has been provided by NAM for this. The four RTiCM parameters are implemented in such a way that they can be treated as calibration parameters in the SSM calibration module. However, for now the best fitting RTiCM parameters published in NAM (2021) were used, since these were already calibrated on subsidence data. Key in NAM (2021) is the use of a spatial $\mathrm{c}_{\mathrm{m,prior}}$-field as a prior to which both compaction coefficient parameters are linked through multiplication factors $A$ (for $\mathrm{c}_{\mathrm{m,ref}}$) and $d$ (for $\mathrm{c}_{\mathrm{m,d}}$). Three $\mathrm{c}_{\mathrm{m,prior}}$ maps were calibrated in NAM (2021): An initially spatially uniform $\mathrm{c}_{\mathrm{m,prior}}$ value, and two maps with spatially varying values based on porosity and shear wave speed values. For all three maps, the best fit for the multiplication factors, and value $b$, are provided in NAM (2021). The $\mathrm{c}_{\mathrm{m,prior}}$ map with the best performance was the porosity map, which has been shared by NAM and used here for the Rate Type compaction calculations. For the comparison of elastic and Rate Type compaction models below, the elastic model uses this same $\mathrm{c}_{\mathrm{m,prior}}$ map instead of the standard compaction coefficient map used in the official pSHRA (see the KEM9 study, Figure 7 for comparison of both compaction coefficient grids).

## C.4 RTiCM simulation results

The outcome of a SSM with a RTiCM implementation has been compared with the SSM containing linear elastic compaction, using the TCF tests (described in Appendix A of this report). The linear elastic compaction simulations follow the advised EZK SSM settings for the public SHRA, with the exception of the TNO calibration module (i.e., the settings are similar to those used in the KEM-9 study; TNO, 2022). The RTiCM simulations settings differ only in the used compaction model. Both the linear elastic and RTiCM rate models have been used for three simulations (Table C-1):

Tabel C-1

| Calibration period | Corresponding forecast period (gas years) |
|---|---|
| 01-01-1995 until 31-12-2011 | 01-10-2012 until 30-09-2019 (2012-2018) |
| 01-01-1995 until 31-12-2013 | 01-10-2014 until 30-09-2019 (2014-2018) |
| 01-01-1995 until 31-12-2015 | 01-10-2016 until 30-09-2019 (2016-2018) |

The forecast period here involves full forecasting of the magnitude-frequency distributions, but event rates are forecasted up to 2050 in all simulations. These calibration and forecast periods were chosen so that there was sufficient observed earthquake data in the forecasting period to run the pseudo-prospective TCF CSEP tests.

The full forecast period did not venture into gas year 2019 or beyond (beyond 30-09-2019), because four observed earthquake events occurred in areas where pressure increased. For these areas, the linear elastic SSM forecast module predicts no seismicity (i.e., negative infinity loglikelihood). Therefore, the TCF testing tools, which use these observed events for testing, will automatically reject the linear elastic SSM. Although this is a clear indication for the urgency of future-proofing the SSM, we wish to test the SSM models on events that both models are able to predict.
We emphasize that the performance of the TNO SSM calibration module is not impeded by training events that occur in regions of increasing pore pressure: Such events obtain a finite loglikelihood by the aftershock model. In the forecast module, such events can be predicted through either a new main shock model (e.g., the RTiCM model presented here) or by adjusting the aftershock model in the forecasting module. Both options are in the process of being implemented.

### *C.4.1 Check of model parameter prior ranges*

The DSM and ARM parameters that are calibrated in the SSM calibration module are explored over a pre-defined parameter space (i.e., the prior). Since the DSM model has changed with an alternative expression for the compaction strain, the posterior probability distribution may have shifted in the model parameter space. This requires a change of the prior ranges. The posterior distribution for one of the RTiCM SSM simulations is shown in Figure C-2. Compared to the standard prior ranges used in the pSDRA with TNO calibration, only the parameter range of the elastic parameter *Hs_exp* (bottom right panel) has been expanded from 4.75 - 5.75 $\log_{10}$ ($MPa^{-1}$) to 4.75 - 6.75 $\log_{10}$ ($MPa^{-1}$).

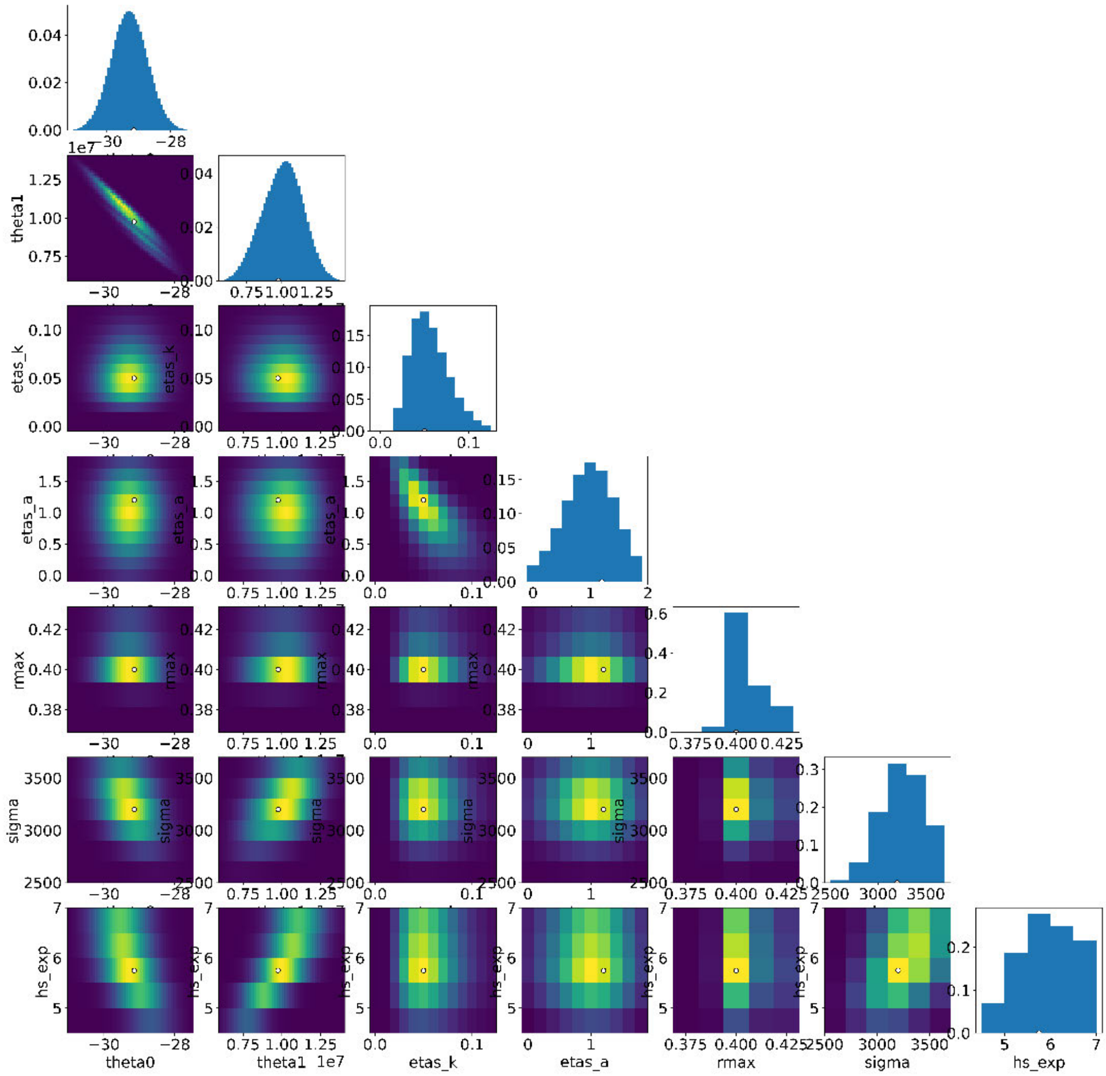


Figure C-2: Posterior distributions for the DSM and ARM parameters, for the RTiCM SSM simulation calibrated on data observed from 01-01-1995 until 31-12-2013. The white marker indicates the set of parameters with the highest probability.

### C.4.2 *Simulation results: annual event rate*

The field-wide annual event rates (per gas year) for the three pairs of model outcomes (Figure C-3) show that the RTiCM predicts a slower event rate decay beyond the 2013 production peak compared to the linear elastic model. This is best seen for the simulation for gas years 2012-2018 (top row, Figure C-3): in the period 2013-2019, the simulated yearly events by the RTiCM exceed the number of observed events by a large amount, whereas the overestimate by the linear elastic model is much less. This ‘early’ overestimate of the yearly event rates is suppressed when the calibration period includes the peak production year 2013 and some gas years beyond (e.g., simulation for gas years 2016-2018). Beyond 2020, when the transitions occurs to production arrest and shut-in, RTiCM predicts a higher annual event rate as well: For instance, 1 event/year is predicted beyond gas year 2040 for the elastic model, whereas the RTiCM model does not predict such low rates until at least 2050 (the end of the event rate forecasting period).

These first results are in line with the concept of the RTiCM model relative to the elastic model: Compaction is not fully instantaneous but contains a creep component, so that part of the seismicity caused by a pressure change occurs at later times.

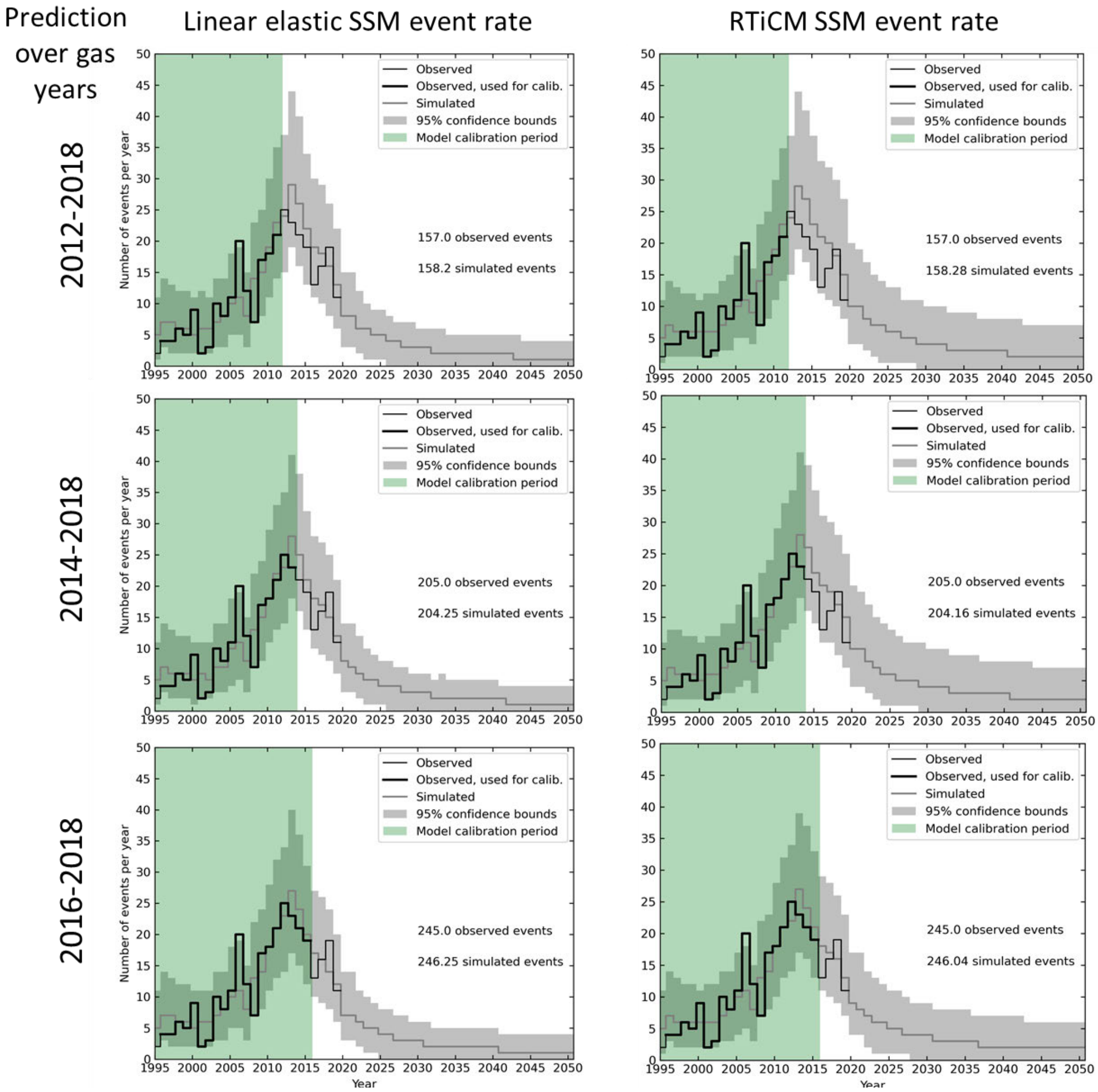


Figure C-3: Simulated field-wide event rates per gas year (gray curve) and the 95% confidence bounds (gray area) for the linear elastic SSM (left side) and the RTiCM SSM (right side), for all three calibration-forecasting periods (rows). Further shown are the model calibration period (highlighted in green), the observed event data used for calibration (thick black curve), and observed data outside the calibration period, up to end of gasyear 2018 (thin black curve).

### *C.4.3 Simulation results: TCF SCEP tests*

The RTiCM implementation in the SSM produces the results that were conceptually expected. Now we quantify if the RTiCM produces better results with respect to the linear elastic SSM, using the TCF. We test and compare the pairs of models that were calibrated over the same time period.

For the models that forecast gas years 2012-2018, the RTiCM forecast fails the consistency N-test, the L-test result is suspect (Figure C-4) – this is a likely result from overpredicting the number of events. The elastic model passes all tests. These results reflect the observation made previously: The RTiCM model tends to overestimate the number of events particularly for the shortest calibration period (up to 2012).

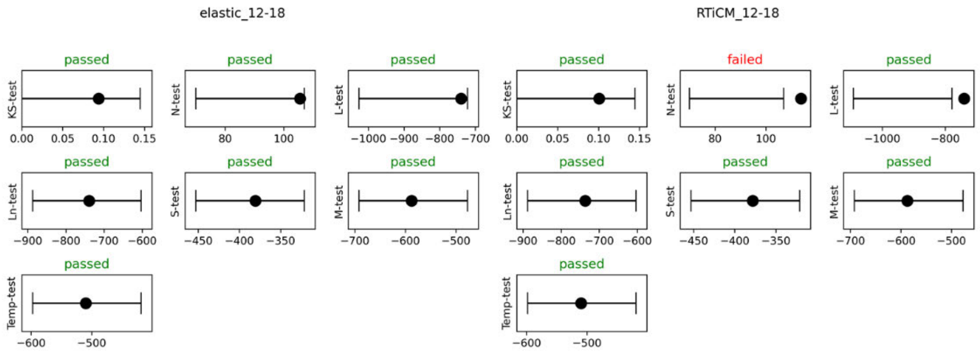


Figure C-4: Consistency test results for the models that forecast gas years 2012-2018.

For the forecast of gas years 2014-2018, both forecasts pass all consistency tests (Figure C-5). For the final forecast period (gas years 2016-2018), the elastic model passes all tests, whereas the RTiCM model fails the KS-test. This means that the RTiCM is unlikely to explain the data.

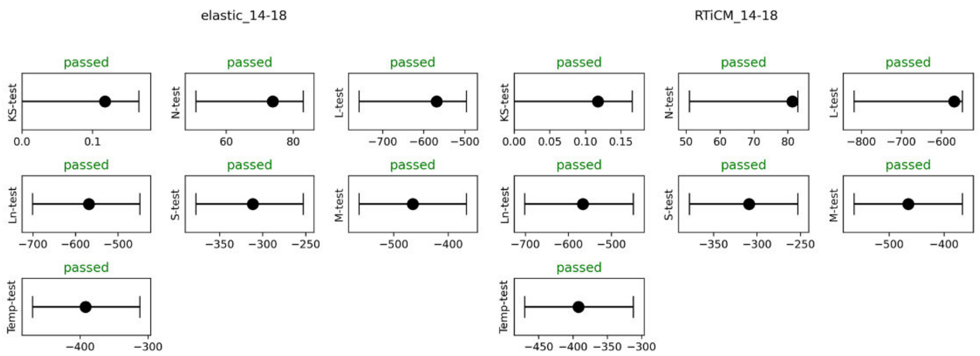


Figure C-5: Consistency test results for the models that forecast gas years 2014-2018.

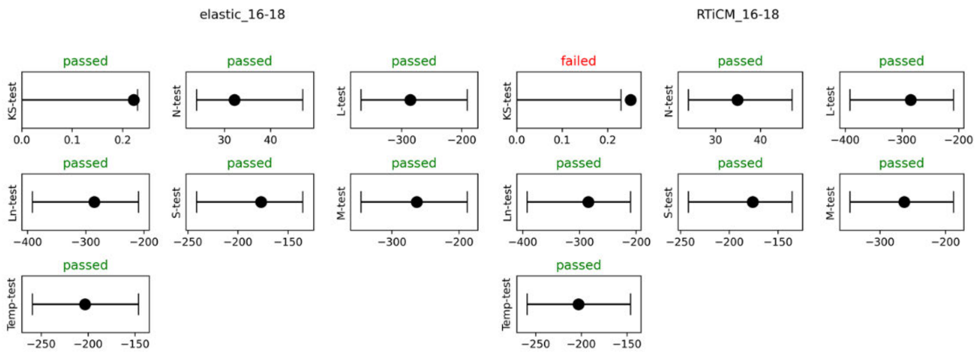


Figure C-6: Consistency test results for the models that forecast gas years 2016-2018.

Across the three forecast periods, the elastic model passes all consistency tests, whereas the RTiCM implementation fails some tests at some forecast periods. However, the comparative T- and W-tests (not shown here) for all three forecast periods do not indicate a significant winner (i.e., better model). The two models have been tested on periods of significant production, whereas the RTiCM is implemented with the post-production phase in mind. We therefore aim in the near future to extent the forecast period to the most recent event data, because in the last years the some parts of the fields have transitioned into the post-production phase. As described before, testing can commence once the pending update of the ETAS treatment in the forecast module has been achieved.

So far, near fault reservoir compaction (elastic or this rate dependent variant) has been used as a proxy for the true stress changes imposed on the fault planes and internal fault material. We acknowledge that subsurface rock material (reservoir and/or fault gouge) exhibit more complex behaviour than just linear elastic. The research question is if this complex material significantly influences the induced stresses and thereby the induced seismicity and seismic risk. TNO is committed to further pursue this research question.

## C.5 References

Bourne, S. J., & Oates, S. J. (2017). Extreme Threshold Failures Within a Heterogeneous Elastic Thin Sheet and the Spatial-Temporal Development of Induced Seismicity Within the Groningen Gas Field. Journal of Geophysical Research, 122(12). doi: 10.1002/2017jb014356.

TNO (2022). KEM9: Cumulative propagation effect of Groningen risk model component uncertainties on hazard and risk predictions, TNO 2021 R12442, 9 September 2022.

NAM, (2018). The influence of stress rates on induced seismicity rates within the Groningen gas field.

NAM, (2021). Groningen long term subsidence forecast. EP202008201822

Pruiksma, J. P., Breunese, J.N., van Thienen-Visser, K., and de Waal., J.A. (2015). Isotach formulation of the rate type compaction model for sandstone, International Journal of Rock Mechanics and Mining Sciences 78, June 2015, doi: 10.1016/j.ijrmms.2015.06.001.

De Waal, J. A. (1986) On the rate type compaction behaviour of sandstone reservoir rock.

# D Event rate analysis pSHRA Groningen

The predicted earthquake event counts and the observed event counts in the Groningen gas field were analysed. This analysis is motivated by the concerns put forward by SodM regarding the annual amount of predicted earthquakes, that they consider to be systematically lower to the observed number of earthquakes in the most recent years (SodM, 2021). SodM raises doubts on the validity of the current seismic source model (SSM-NAM-V6) used for said predictions. We answer two questions:

1) How well has the SSM predicted the observed event counts in the past?
2) Are there clear signs that the SSM is underpredicting the observed event counts in recent years?

Subjected to these questions are the pSHRA calibration prescribed by EZK and provided by NAM, and TNOs implementation of the SSM calibration. This section expands on the previously reported event count analysis by TNO (14 Jan 2022), using the pSHRA 2022 results and augmented with the seismicity observed during the two gas years (2020/2021 and 2021/2022).

## D.1 Approach

To answer the first question, two analyses were conducted: The first analysis is a simple comparison between the total number of observed and predicted (i.e., hindcasted) earthquake events over the calibration period. The observed earthquakes used are those from the publicly available KNMI catalogue *induced_earthquakes*. For the pSHRA-2022 with the NAM calibration settings prescribed by EZK (hereafter pSHRA-2022), calibration was performed with events recorded between 1st of January 1995 and 1st of January 2021 (TNO, 2022b). For the calibration of the pSHRA produced using TNOs implementation of the SSM calibration (hereafter TNO SSM), events between 1st of January 1995 and 13th of February 2022 were used. The same analysis was performed for pSHRA-2021 (TNO, 2021), where both models were calibrated over the same period (TNO, 2022a). As shown below, the main conclusions of this first analysis hold for pSHRA-2022, regardless of equal or differing calibration periods.

For the second analysis, the observed annual event count is checked against the predicted annual event count probability density. For this, we use observed event count data between the 1st of October 1994 and the 30th of September 2022 (28 gas years). The observed event count for a given year corresponds to a percentile range in the simulated probability distribution of earthquakes: We choose the middle to represent the interval. The probability distribution of the event count is provided as standard output from TNO Model Chain code base. In case the seismic source model is correct, the observed sample (i.e., the 28 data points representing the 28 gas years of observed earthquakes) is distributed uniformly over the percentiles: e.g., 14 years are expected to fall in the lower 50% and 14 years in the upper 50%. From this, we define the null hypothesis for our statistical test: the observed percentiles are drawn from a uniform distribution. The above defined null hypothesis is subjected to a one-sample 2-tailed Kolmogorov-Smirnov test (KS-test) which is based on the distance measure (the KS-statistic) between the empirical and theoretical cumulative distribution functions (CDF). The empirical sample is the 28-year series of percentiles obtained from comparing the observed earthquake events, discretised per gas year, with the simulated event rate (Figure D-1). The sample is compared with the theoretical continuous uniform distribution between 0 and 100%. The KS-test provides a p-value between 0 and 1, which represents the probability that the observed KS-statistic could is exceeded under the null hypothesis. A low p-value is thus indicative that the null hypothesis can be rejected.

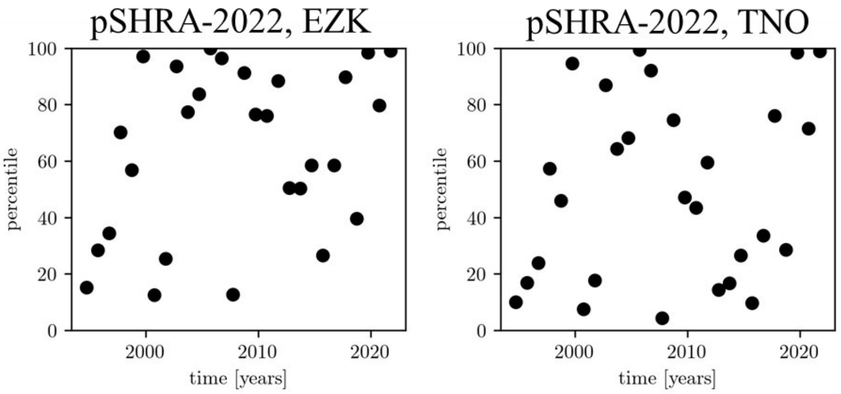


Figure D-1: Percentiles in which the observed event rate falls per gas year, for the pSHRA-2022 simulation (right) and for the TNO SSM implementation & calibration (left).

The results on an analysis on the time discretization of the event count (i.e., gas years or calendar years) can be found in TNO, 2022a, based on the pSHRA-2021 results.

The second question – are there indications for a recent disparity between observed and predicted event counts? – we discuss by means of the individual percentiles for the observed event counts of most recent years. This provides a more direct analysis and is less bias-prone then a KS-tests on a number of recent years.

## D.2 Results

### *D.2.1 Total number of events*

The total number of events between the 1st of January 1995 and 30th of September 2020 is 319. The prescribed pSHRA-2022 SSM implementation hindcasts 273 events (Figure D-2) – an underestimation of 46 events. The total event count resulting from the TNO-style SSM implementation & calibration hindcasts 329 events (Figure D-2), a small overestimate of 3 events relative to the observed number of events – this is mostly explained by conversion of the predicted event rates to round numbers. The observation that the pSHRA-2022 SSM calibration is unable to predict the number of events it is calibrated against, indicates a fundamental problem in the model calibration procedure.

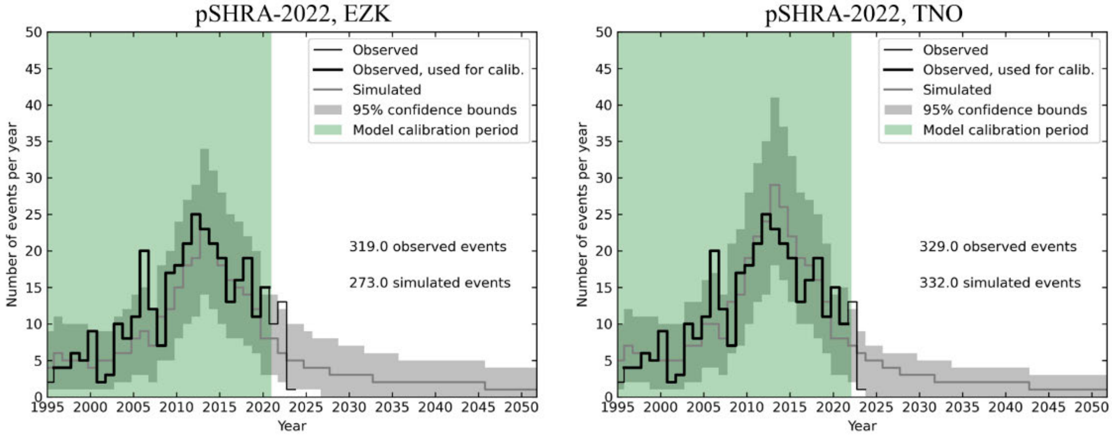


Figure D-2: Predicted and observed event counts. pSHRA-2022 results (left) and TNO SSM implementation and calibration (right), per gas year.

### *D.2.2 KS-test results for full event history*

The empirical and theoretical CDFs are shown in Figure D-3 for the pSHRA-2022 SSM implementation & calibration (left panel) and for TNO's SSM implementation & calibration (right panel). One may immediately observe that the empirical CDF of the pSHRA-2022 event count has a large offset with the uniform CDF. This empirical CDF lies 'below' the uniform CDF (i.e., at higher values on the horizontal axis), which indicates that the uniform CDF of the model underestimates the observed annual event count. The TNO-style empirical CDF fits better with the uniform CDF.

This has been quantified with the KS-tests, for which the p-values have been provided in Figure D-3: For the pSHRA-2022 event count, a p-value of 0.037 indicates that the uniform model CDF is very unlikely to explain the observed empirical CDF. The p-value of 0.861 for TNO's predicted event count suggest a significant likelihood that the empirical CDF is drawn from the uniform CDF.

From this, we may conclude that the pSHRA-2022 SSM with EZK settings fails to accurately predict the observed annual event count, and consistently underestimates it. In contrast, the TNO-style SSM calibration manages to predict the observed event count better.

The same test conducted on the pSHRA-2021 (TNO, 2022a) confirms the failure of the pSHRA-2021 with EZK settings, with a p-value of 0.036, whereas a p-value of 0.95 was found for the TNO SSM calibration.

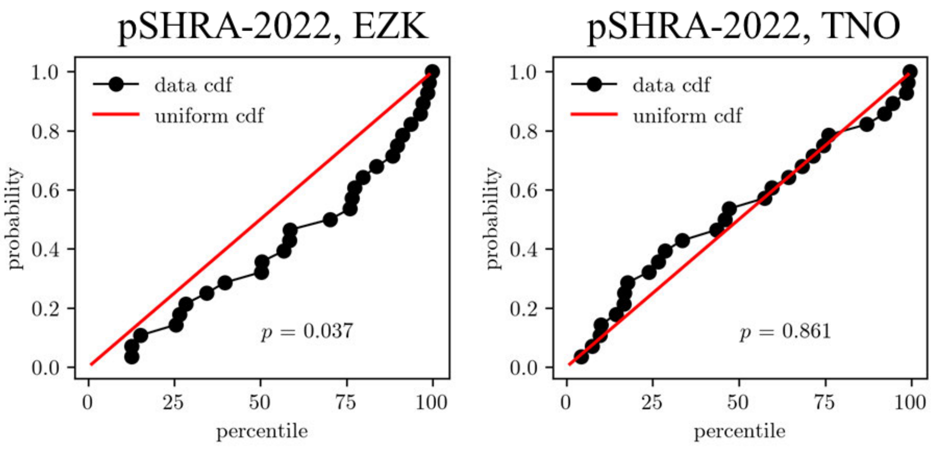


Figure D-3: Empirical CDF (black markers) and continuous uniform CDF (red line). Empirical CDF based on the pSHRA-2022 event counts (left) and on event counts from the TNO SSM calibration (right). p-values from the KS-test are given for both cases.

*D.2.3 Recently observed high event rates*

The statistical analyses above shows that the TNO-style SSM implementation outperforms the pSHRA-2022 calibration over the last 28 gas years. However, as highlighted by SodM, in recent years there is an underestimation of the event count by the SSM. This is to be expected from using the pSHRA-2022 with EZK-prescribed calibration, as it systematically underestimates the event counts. The question that arises is: Does the disparity persist for predictions based on the TNO-style SSM calibration?

Of the observed event counts in the last three gas years (2019-2021), two years fall in the percentiles 98.0% and 98.6% (gas years 2019/2020 and 2021/2022, respectively; Figure D-1). Whereas one such a high percentile can be ascribed to random variation – as may have occurred in gas year 2005/2006 (99.3% percentile) – two of such high percentiles in quick succession are highly remarkable and suspect. It is highly unlikely that these two variations are random variations within the model predictions. This observation indicates that the current model (Bourne & Oates style SSM with TNO-style calibration) either underestimates the mean event count or its dispersion in this transition stage to post-production of the Groningen gas field. The cause for this underestimate may reside in a few factors: a) The input data, primarily the annual pressure grids derived from reservoir models by NAM, require revision or lack sufficient calibration data. b) The current SSM misses a time delay between gas extraction (pressure reduction), compaction, and seismicity. c) Aftershocks activity from mainshock events may last for a longer period. This emphasises the need for continued model development for the post-production phase of the Groningen field.

Nonetheless, in absence of a thoroughly tested SSM model alternative, the TNO-style implementation of the SSM is clearly the better choice to the currently prescribed pSHRA-2022 SSM.

## D.3 Summary

For this analysis, the simulated event counts from the 2022 pSHRA (public Seismic Hazard and Risk Analysis; TNO 2022b) were used. They are comprised of the event count obtained from the SSM calibration supplied by NAM, currently recommended for the annual pSHRA Groningen, and TNO's recommended SSM calibration module.

### *D.3.1 pSHRA-2022 NAM calibration*

319 earthquake events were observed during the NAM calibration period between 1995 and 2019, whereas 273 events were hindcasted by NAM for the same time period. The NAM predicted event counts resulting from their calibration are unlikely to describe the event rate over a 28 year period of gas years, evidenced by a p-value of 0.037 from the statistical KS-test.

### *D.3.2 TNO calibration*

The hindcasted event count from the TNO implementation of the SSM and accompanied calibration procedure was tested against the observed data from 1994 to 2020 (27 gas years). We now obtain 332 predicted events against 329 observed. Over a 28 year period (gas years 1994 to 2021), the observed event count is likely explained by the predicted event count, evidenced by a p-value of 0.861. These test results speak in favour of the TNO calibration over the NAM calibration.

### *D.3.3 Recent disparity between observed and simulated gas years*

In the last three gas years, the observed event rate matched modelled percentiles above 98% twice (gas years 2019/2020 and 2021/2022). This is remarkable and suspect, indicating that the current seismic source model is not able to correctly predict annual event counts in the transition towards the post-production phase of the Groningen field. Some model improvements options were identified for model development.

### *D.3.4 Conclusion*

Our analysis shows that:
1) The predicted pSHRA-2022 NAM calibration based event rate is significantly lower than the observed one. Moreover, it is unable to predict the correct number of events it was calibrated on.
2) It is very unlikely that the pSHRA-2022 NAM calibration based event rate predicts the observed data. These conclusions support and expound on the observation of SodM (2021).
3) The TNO implementation for SSM and accompanied calibration procedure outperforms the for pSHRA-2022 prescribed SSM implementation and calibration in predicting the number and rate of earthquake events.
4) However, in the most recent recorded gas years we observe a large disparity between predicted and observed event counts that is highly unlikely to be labelled as variations within the model prediction bandwidth. This raises concerns on the SSM prediction performance for the post-production phase of the Groningen field.
5) In anticipation of a better alternative, we recommend to use the TNO implementation of the SSM.

## D.4 References

TNO (2021). Publieke Seismische Dreigings- en Risicoanalyse Groningen gasveld 2021. TNO2021 R10441, 24 maart 2021.

TNO (2022a). Analyse aantal waargenomen bevingen in Groningen in relatie tot modelverwachting en de rol van modelkalibratie. Kenmerk AGE 22-10.003, 14 januari 2022.

TNO (2022b). Publieke Seismische Dreigings- en Risicoanalyse Groningen gasveld 2022, TNO 2022 R10517, 25 maart 2022.

SodM (2021). Beoordeling SodM halfjaarrapportage seismiciteit Groningen, overschrijding grenswaarde aardbevingsdichtheid & beving Garrelsweer. 9 Dec 2021.