KEM subpanel advice on the TNO report "*Status of the TNO Model Chain Groningen per October 1, 2022 and recommendations for the public Seismic Hazard and Risk Analysis 2023*" (TNO2022_R11961)

Dr. Pierre-Yves Bard (*Univ. Grenoble*, France), Prof. Fabrice Cotton (*GFZ*, Germany), Prof. Ing. Iunio Iervolino (*Univ. Naples Federico II* and *IUSS Pavia*, Italy), Dr. Ir. Femke Vossepoel (*Delft Univ. of Technology*, The Netherlands), Prof. Stefan Wiemer (*ETHZ*, Switzerland)*

December 5, 2022

1. PREAMBLE

1.1. Context of this report

In late 2020, a subpanel of KEM (Knowledge Programme on Effects of Mining) was established by the Ministry of Economic Affairs and Climate (MEAC/EZK) of the Netherlands to follow the development of the public seismic hazard and risk assessment (the public SHRA) model for Groningen, accompanying the transition of the responsibility for the SHRA from NAM to TNO. The purpose of the subpanel is to advise State Supervision of Mines (SodM) and the Ministry of Economic Affairs and Climate on the scientific aspects of both the public SHRA model versions and its development. The specific goals of the subpanel are: (i) to review the studies having for their potential to be part of the public SHRA development, (ii) to advice on the proposed public SHRA versions and developments of TNO on a yearly basis; (iii) to report to the KEM panel about the public SHRA progress and development.

The KEM subpanel is composed of independent international experts with specific expertise on probabilistic hazard and risk assessment overall and on three main modules of risk analysis, namely: the seismological source model (SSM); the ground motion model (GMM); and the fragility and consequence model (FCM). All the subpanel members have contributed to the present advice and endorse the entire document.

On November 4, 2022, the KEM subpanel received from SodM the request to advice on the TNO-SHRA-status-report 2022 (TNO, 2022d) and SodM included in the request several specific questions to be addressed. They are related to quality assurance and control (QA/QC) and model validation/testing aspects as well as comments on the scientific soundness and practical usefulness of particular (sub)models in the TNO public SHRA model chain.

The present advice first addresses, in chapter 2, the specific questions raised by SodM. Then it discusses the general aspects of the public SHRA model of TNO, in chapter 3, while in chapters 4-6 there are the subpanel advices on the proposed seismic source model (SSM), ground motion model (GMM) and Fragility and Consequence model (FCM) versions proposed in the TNO Report, respectively. As the specific questions of SodM are partly concerned with general aspects and partly focus on specific (sub)model components of the public SHRA, some overlap in the document was unavoidable.

2. SPECIFIC QUESTIONS OF SodM

The set of questions from SodM that address quality Assurance and related issues, as well as selected model components, are answered in the following individually.

Question 1: Is the Quality Assurance and Quality Control (QA/QC)-procedure as applied to the report and its content sufficiently clear of appropriate quality and traceable?

The subpanel agrees that appropriate, clearly defined, transparent and well documented Quality Assurance and Quality Control (QA/QC) procedures are key to SHRA and essential to minimize the risk of errors, to enhance acceptance and

* Dr. Ipo Ritsema is the KEM subpanel secretary.

trust in the model by all stakeholders. It is also to note that the comments given on QA/QC procedures, and those indirectly contributing to it, do not imply that the KEM subpanel is, or it could be, responsible to ensure QA and QC, as it advises primarily on scientific aspects of model development.

In the KEM subpanel assessment, the guidelines for the QA/QC procedures are overall well-defined and documented by TNO in a collection of twenty-five documents (TNO, 2022c). TNO, as an agency, has in general extensive experience with QA/QC procedures and has carefully considered them for the public SHRA. The KEM subpanel is informed that compliancy audits take place on a yearly basis. However, the extent to which these QA/QC procedures have been applied systematically to all pSHRA documents, decision processes and model components is not always clear to the subpanel, and this is not well described in the TNO document. Missing are for example clear milestones of deliverables of the QA/QC process that could be traced, documentation of decision-making procedures applied, or access to internal review reports. So far, external reviews of model components or structured expert elicitation did rarely take place and were conducted somewhat planned ad-hoc. A planning of external reviews at a yearly basis or closely attached to model changes would improve the QA/QC. Lists of experts involved in the review, including their CVs, would make the procedure more traceable.

The external audit report provided to the KEM subpanel describes an evaluation of a spot check. One of the points of improvement identified in this report, is the management of documents and data. The audit report states that relevant information has been shared within the project team and is being documented, but the documentation is not easily traceable. This also complicates the evaluation of the QA/QC procedures. The audit report furthermore states that all documentation is in Dutch because the stakeholders are Dutch-speaking. However, this complicates the readability for non-Dutch speaking team members, as well as the evaluation of the progress and decisions by external experts.

Based on the available documentation, it is difficult for the KEM subpanel to verify to what extent the existing QA/QC procedure has been followed systematically and if so, how this happened in detail. It is also difficult for the subpanel to judge if these procedures are fully appropriate and mature, and how these procedures are approved and further developed. However, despite some deficiencies in the documenting how the QA/QC procedures have been applied in this model version, the subpanel does not consider this an issue which would question the validity of the entire proposed model version, although it is restated that introducing systematic peer-review is a critical success factor for public SHRA developments.

Based on the above, the panel would like to suggest five improvements to enhance QA/QC in the future.

- **R1.1:** in the first half of 2023, TNO would report in a meeting to the subpanel, SodM and EZK on their existing QA/QC procedures, also giving the opportunity to suggest improvements.
- **R1.2:** in future documents, TNO specifically comments on the QA/QC procedures and steps applied. It is suggested that any substantial change of the model should be accompanied by a decision-making process with clear decision gates (for example, Identify, Assess, Select, Define, Execute, Operate). Each decision gate should be accompanied by a review with relevant experts outside of the project team (typically internal experts, occasionally external).
- **R1.3:** TNO should allocate sufficient resources to QA/QC procedures, also allowing to involve external experts or expert panels where appropriate (and as part of the decision-making procedure). The independent review approach used by NAM can be worth to be considered to design the peer-review process.
- **R1.4**: Any documentation of information important for external review is written in English.
- **R1.5**: Innovations developed for public SHRA are published in international scientific peer-reviewed as much as possible and the public SHRA software is provided with manuals and made publicly available for reproducibility and what-if analysis from others.

Question 2: Is the Testing and Comparison Framework (TCF) sufficiently developed, complete, validated, and appropriate for its goal of testing and comparing any existing and new models in a structured and quantitative way to determine their appropriateness and readiness for possible adoption in the public SHRA? Any comments on the quality and QA/QC of the TCF?

The subpanel is convinced that TCF is an important contribution to accompany the model development, and this is highlighted already in the Nov. 2021 advice ('Validation and Testing', section 2.3). The validation and testing should include all components of the public SHRA, namely SSM, GMM, and FCM. Systematic testing and validation are needed because of three main reasons:

1. the complexity of the phenomena accounted for in the components of public SHRA requires modelling that is supported by both theoretical arguments as well as empirical counterparts, as much as possible;
2. model replacement and update should be quantitatively proven via comparison-supported arguments.
3. a suitable testing framework can form an important contribution to QA/QC.

Regarding empirical validation, empirical data in the Groningen field essentially refer to earthquake events, to which several variables (or functions) can be associated, most notably: magnitude, location, time, ground shaking at the recording sites, and ground motion intensity measures. All model components that depend on one or more of these variables or functions can be, in principle, tested against empirical observation. This mainly applies to the SSM, to the GMM, and to a much lesser extent to the FCM.

Overall, the subpanel appreciates and supports the efforts undertaken by TNO to develop and use a TCF as part the public SHRA model development. Appendix A of the TNO report and previous reports are a good summary of the tests used, these classes of tests are widely applied in the scientific community and the TNO team has shown an important level of understanding of the test, their applicability ranges, and limitations. While the consistency and comparison tests have mostly been applied in the literature to natural seismicity, they can be considered fit for purpose also for induced seismicity analysis. The selected TCF is therefore considered adequate for applications related to the SSM development of the public SHRA, the framework is flexible enough to be extended.

While the TCF has started to penetrate the TNO SSM model development efforts, its lacks a consistent and systematic application in model development; it also lacks benchmarking, community feedback and peer review. The KEM subpanel suggest that TNO in 2023 and 2024 focusses on further developing the TCM framework to enhance the openness, transparency, and usefulness. These include.

- **R2.1:** that the TCF and testing experiment (see R2.2) is discussed in a community workshop with selected experts, for example from the CSEP (https://cseptesting.org/) community, and that a benchmarking of the test against existing implementations is performed and/or openly documented.
- **R2.2:** TNO sets up a formal, fully prospective testing experiment for Groningen that establishes on a regular basis how different models are performing against the observed seismicity. This requires defining in advance all 'rules of the game', such as the study region, magnitude range, model calibration/updating strategies, spatial binning, etc.; the testing setup should be published in a peer reviewed journal and the test, as well as tests results, openly accessible (see also reply to Question #1).

The TNO comparative testing and validation efforts so far have almost exclusively focused on the SSM level, which is in-line with ongoing testing efforts worldwide that first of all focus on earthquake rates. With respect to GMM, only low-magnitude events data can be subjected to empirical testing/validation against ground motion data recorded in the Groningen area, and this so far has been addressed in specific KEM research questions, such as KEM 02-04 and 34 (https://kemprogramma.nl). Larger ground motions that dominate the risk are rare or non-existing in the Groningen data set so far recorded, hence prospective formal testing is implausible. Effort in this direction should be directed at testing and possibly improving the ground motion simulation technique used to generate the large ground motions, something that KEM-04 has addressed, but yet has not been considered by NAM.

Structural damage and fatality risk analysis likewise has to deal with rare events, and therefore, is also constrained by a lack of access to empirical data. Moreover, the scarce standardization of building construction impairs even further any testing/validation of FCM models. Nevertheless, analysis by means of comparison of different models is possible and advisable. In this respect the superiority of the 'typological approach' for the fragility modelling, with respect to the currently used models derived by NAM, or other literature on the topic, is not documented in sufficient detail, at least given the information referred to or provided to the subpanel. The KEM subpanel supports the efforts TNO has undertaken to test and verify some specific issues in the GMM/FCM interface, that is, the p2p correlation.

Finally, it is to recall that model updating based on new data (e.g., Bayesian updating) is an option alternative to model replacement, and one closely related also to the TCF. This option is so far insufficiently explored in the context of seismic risk analysis. The KEM subpanel speculates that this option could be beneficial for public SHRA for Groningen.

In addition to the recommendations related to the SSM testing, the subpanel recommends the following steps:

- **R2.3:** the testing/validation/comparison framework is not uniformly applied with the same level of in-depth effort in all the model components; TNO has put most of the effort in this direction on SMM, and on some specific issues of GMM/FCM interface; although extending into other domains is challenging, it is recommended continuing efforts into this direction.
- **R2.4:** it is recommended that validation/testing/comparison, whatever is most suitable for the case at hand, is considered as an integral part of the QA/QC process and hence integrated into the QA/QC guidelines.

Question 3: Are the various models in the model chain and the model outcomes, as well as the seismological model (further: SSM)- calibration (including the single Coulomb stress distribution predictor for activity rate), sufficiently validated, verified and reproducible?

The whole chain to estimate the risk is composed of three main modules (the already recalled SSM, GMM, and FCM) which are in turn composed of several (sub)models: semi-empirical based on observations/data gathered in Groningen, based on physics and mechanics developed either specifically for Groningen, or tuned/adjusted to Groningen. The actual implementation of each model may also imply some non-obvious choices which may significantly impact the results. The soundness of the chain of the three main modules is not based only on the individual validity of the various elements, but also on the internal consistency of the chain, i.e., of the links between the various elements.

The overall architecture of the present public SHRA model, as well as its main components, are inherited of the NAM achievements to develop from scratch a suite of components able to reach the ambitious risk estimation goal. As emphasized in the numerous NAM reports, these components have given rise to many publications in peer-reviewed journals, which constitute an indirect and not completely granted, yet meaningful indication of their scientific soundness. TNO has then implemented these components with their own coding and assembled them within a new software script architecture. This process has allowed to clarify several insufficiently detailed implementation issues through close exchanges with the NAM team, and it has also enabled to assess their reproducibility after several exchange loops. This important verification step established the soundness of the computational framework and of the computational implementation of model components. Missing is, however, open access to the codes and models (discussed in more detail in section 3), which would allow for a broader community evaluation and further enhance reproducibility.

As to the new developments or implementation variants proposed by TNO for some of these individual components, TNO made systematic efforts to investigate the associated changes through sensitivity studies and to document them in specific reports (e.g., TNO, 2022a,b), and sometimes to publish them in peer-reviewed journals (e.g., Kraaijpoel et al., 2022 for the spatial variations of b-value). These efforts with respect to the SSM are overall judged to be sufficient to ensure validation and reproducibility before being used as part of the public SHRA, although additional efforts to publish the model and to systematically test them would be welcome (see also response to question 2).

More difficult is the question if model components have been sufficiently *validated*. The term 'validation' usually refers to some form of comparison (or testing) of model forecasts with actual observations, that is, empirical data. The development of the TCF and its pseudo-prospective application to the SSM forecasts established confidence in the average SSM performance, for events that can be observed with a statistical meaningful number. It is very unlikely that it will be possible to truly validate the models for low probability events, given the limited duration of the observation period. However, as TNO has shown clearly, the risk figures relevant for safety concerns in Groningen entirely come from events that have not been observed yet in the field. The same may be said for the GMM; the GMM predictions for weak events (local magnitude below 3.4) are consistent with observations for the selected intensity measure (Sa-average or Sa_avg), even though the GMM underpredicts the very high frequency contents - for presently unknown reasons. On the other hand, as there are no data from higher magnitude events in the Groningen field, a true validation against field-specific data is – so far – impossible: some of the GMM model constituents may, or may not, fail when extrapolated to larger events (assumptions on non-linear behavior, damping, finite source geometry, etc.). Therefore, a true validation of the SSM or GMM models for events of larger magnitude is not possible, and it has also not been performed previously by NAM. As discussed also in reply to question #2, the FCM seem to lack attention with respect to SSM and GMM, except for the p2p correlation issue (to follow). The typological approach for the fragility modelling, although cannot be verified against observed damage data (because data about damages significant for the local personal risk do not exist in the field), should be still proven superior with respect to the NAM model (which has its own identified issues) or with respect to the state-of-the-art of building-class vulnerability modelling. Similar reasoning applies to the NAM consequence models.

Regarding the model outcomes, that is, the risk assessment, it is likewise very difficult to say they are fully validated, again because of the short observation duration and the control of risk estimates by low to very high probability events, which is typical in the case of seismic risk analysis. Each component of the model chain is scientifically acceptable, considered (by the model developers) validated when applied to observed events, and because the extrapolation to larger, non-observed events is based on state-of-the-art knowledge and practice. One issue that appears susceptible of improvement is the consistency between the different blocks: for instance, the intermediate-size sources (Mw between 3 and 4.5) that have been considered for the development of the GMM (geometry of the extended sources considered in the stochastic simulations), may not always be fully consistent with the 'within-reservoir ruptures' emphasized during the recent, June 2022 Mmax (i.e., maximum magnitude) workshop. In other respects, the fragility curves have been developed with a set of acceleration time histories based on the GMMV5 model, that may not be fully consistent with the changes in the p2p correlation considered in later GMM versions. Improving such a consistency would require however very significant efforts and resources; the recommendation is to start with sensitivity studies to decide whether it is needed to engage in medium term developments.

In summary, the KEM subpanel considers the model components sufficiently verified, reproducible, but not validated, since this is a nearly impossible task in SHRA. To further improve the model building process, the following is recommended.

- **R3.1:** check thoroughly and improve, if needed, the consistency between the different blocks (SSSM/GMM, GMM/FCM)
- **R3.2:** perform sensitivity studies to decide whether to engage in medium-term developments of the GMM to ensure consistency
- **R3.3:** develop a framework to assess adequacy of superiority with respect to the state-of-the-art of the used or proposed models which cannot be directly validated against data observed in Groningen events.

Question 4: Are the proposed alternative models for the model chain scientifically sound?

All alternative SSM models have been subjected to sanity checks, quality control, sensitivity studies and are documented with pros and cons with respect to the original NAM version. These model alternatives have been exposed for several years now to internal and sometimes external review, presented at workshops and meetings and in parts are published or under review. The reasons of their preference are well documented in the various TNO reports referenced in their October 2022 recommendations (TNO, 2022d). The change in the Maximum magnitude distribution is based on a structured expert elicitation process of the highest quality. Therefore, it is recommended to consider the proposed alternative models scientifically sound, representing the state of the art.

The KEM subpanel appreciates that TNO has proposed that some of the potential alternatives, although they would seem sound form a purely scientific viewpoint (e.g., spatial dependence of b-value in relation to the local reservoir thickness), are not yet recommended to be included, because all the TCF tests could not be completed yet, which is sensible.

With respect to the GMM V7, this process has also been undertaken by NAM and its external expert with great care, the new model is thoroughly verified and well documented. It has also been exposed to reviews and evaluations by TNO and by the KEM subpanel. Therefore, the GMM V7 also can be considered scientifically sound and state of the art.

With respect to FCM the view is shared with the TNO that there is room for improvement, possibly wide, on some aspects of the NAM models. In this respect, the so-called typological approach for the fragility assessment is being pushed forward for public SHRA. However, such an approach is not documented enough and, as far the KEM subpanel is aware of, has not been published in any scientific venue or undergone international/expert peer review. Given these considerations, it is not possible to positively answer, or even answer at all, to this question.

As already discussed in detail in the 2021 KEM subpanel advice, transparency, reproducibility, and openness are core objectives of the public SHRA under the lead by a public agency rather than industry. They are a key element of establishing a scientifically sound methodology. The KEM subpanel considers it very important that the TNO public SHRA engine, and the models, are fully documented and openly availed to the community (extending possible the created platform https://www.nlog.nl/publieke-sdra-groningen or EPOS data facilities used by DeepNL). Such dissemination strategy is part of the efforts that should be made to guarantee the quality of the models (by giving the possibility to others to detect errors). The suggestion to adopt an open-source license to the code (e.g., a Creative Commons BY 4.0 License) is kept in this advice as well. Likewise, the hazard and risk input data and documents should be made available so that interested scientists can reproduce all calculations. EZK should agree with TNO on the conditions for opening up the code and results, and mixed solutions are conceivable (some parts may only be available on justified requests, some data for example on building may not be released due to privacy issues, etc.).

Question 5: The fundamental goal of the public SHRA, is "to make the best possible, scientifically most sound, estimate of the seismic risk originating from past and present gas production from the Groningen gasfield".

a) Do the recommended model-versions by TNO sufficiently address all the sources of uncertainty related to the various models of the modelchain?

Although uncertainty modelling is not an issue solved once for all, given that uncertainty is essentially lack of knowledge (in this view the usual distinction between aleatory and epistemic uncertainty is not only pleonastic but possibly misleading), and therefore with the uncertainty and its required characterization are subjected to evolution, it can be said that the legacy of uncertainty consideration from the NAM model is adequate in its fundamental aspects. The results of KEM-09 also indirectly and generally confirm this.

With respect to the SSM, the KEM panel has for many years expressed the view that the NAM model is too narrow to capture the full epistemic uncertainty in processes that may lead to induced seismicity. The TNO model in that respect is largely based on the NAM model, with selected deviations that have a small influence on the uncertainty distribution. KEM-09 has been important to capture the uncertainty, but more of these findings should in the years to come finds its way into the SSM, for example the effect of the completeness magnitude, or spatial variability of b-values. In the medium term, the SSM uncertainty characterisation may be improved by using more physics-based, or hybrid, modelling approaches, but these have yet to be mature enough to be included in the SSM. In that sense, the current model represents somewhat narrowly the state of the art of the informed technical community, and if an expert elicitation workshop on the SSM were to take place, alternative models may be proposed.

The NAM team in charge of the ground motion prediction model, in particular V7, had already done an important and very complete work to characterise the uncertainties. There are therefore few avenues for improvement except a specific analysis of the uncertainty related to the choice of the p2p correlation model, the choice of models for calculating the non-linearity of soil amplifications, and the assumed geometries for the finite sources in the stochastic simulations (see the discussion of paragraph 5).

b) Are any a priori conservative choices implicitly or explicitly being made by selecting the TNO recommended model-versions?

The subpanel is not aware of explicit conservative assumptions in the public SHRA models recommended by TNO, although a specific detailed investigation with respect to this issue has, to the subpanel knowledge, not been carried out. In the discussions with TNO, the KEM subpanel has never noticed that conservative choices were considered, and the panel members have always warned against making them. As an example, the recommendation to adopt the new maximum magnitude distribution proposed at the end of the dedicated workshop last June seems not to represent a conservative choice.

There could possibly be some conservative choices in the FCM, in particular in the fragility modelling, inherited from the NAM modelling, this has been discussed by the subpanel in previous comments to public SHRA, yet needs further deepening.

In any case, it must be recalled that, given that the risk-driving events (i.e., those most contributing to 10E-5 local personal risk or LPR from risk disaggregation) are not yet observed in the field, conservative assumptions can only be checked in a purely theoretical basis, based on analogy with other situations.

c) Related to 5b), an example: The seismicity data measured at the Groningen gas field cannot confirm or reject the presence of a taper in the magnitude-frequency distribution. Therefore, SodM reasons that, by fully excluding the taper model from the model chain as proposed by TNO, possibly a conservative choice is being made, not resulting in the best possible estimate of the risk in Groningen. At the same time, the currently used 80/20 weight distribution in the logic tree does not seem justified either. It does not right to the aforementioned fact nor the concerns expressed by TNO in the report regarding the taper. Therefore, we think both the 'hyperbolic-tangent b-value and Mmax model' and the 'single b-value and exponential taper and Mmax model' should be considered in the model chain, with, for the time being, for example an equal weight distribution in the logic tree (until new scientific insights indicate otherwise or result in a different distribution of weights). How does the KEM subpanel

reflect on this specific case, and the by TNO proposed alternative model choice for the SSM? Any thoughts on the weight distribution?

The KEM subpanel does not consider the exclusion of a taper on the magnitude-frequency distribution a conservative choice. On the contrary, concern was expressed already by the KEM subpanel and echoed by TNO that the use of both, a taper on the magnitude-frequency distribution and conventional maximum magnitude distribution, may potentially lead to a systematic underestimation of the rate of larger events. Applying the distribution of maximum magnitudes represents itself a tapering process, and the reasoning for limiting the maximum magnitude relates to the plausibility of large ruptures to occur given the available faults, geometrical constraints, stress and loading regime, and observation elsewhere. In that sense, both tapers refer to similar and related processes, and hence do not well represent the PSHA concept of mutually exclusive logic tree branches. It is also noted that applying a taper is common in PSHA studies of natural or induced earthquakes.

At the recent maximum magnitude workshop, it was clear that the experts considered the maximum magnitude largely independently of the existence and shape of a taper on the frequency-magnitude distribution. They did not consider if the magnitude-frequency distribution taper may overlap or interfere with the Mmax distribution, a concern that adds an additional argument to remove the magnitude-frequency distribution taper, as proposed by TNO.

3. COMMENTS ON THE STATUS OF THE PUBLIC SHRA OF TNO

Below the KEM subpanel would like to comment briefly on the overall status of the public SHRA 2023, as presented by TNO.

The KEM subpanel remains convinced that the transition of the Groningen seismic hazard and risk assessment from an industry model to a public one is fundamentally the right choice. TNO has by now acquired the necessary expertise and tools to successfully operate the public SHRA and is in the process of establishing clear leadership in model development. The 2023 proposal for the public SHRA in that sense represents an important milestone on the transition, and while not all suggestions in the TNO model are established with the scientific rigor, QA/QC transparency and external review that the subpanel ultimately would like to see; however, the progress made is impressive and the process is considered on track on track.

However, the KEM subpanel is deeply troubled by one aspects of the model development: The KEM subpanel, in the last two years, has repeatedly insisted that a key strength of a public SHRA lies in its openness. Today, the underlying codes to compute the public SHRA are still not open, held back by legal arguments or lack of priority that it is not possible to comment on. To us, this delay in opening codes and models is not a minor nuisance, but a fundamental flaw, because it questions the very foundation that a public model in the FAIR context is built upon. Given the by now more than 2-year delay in opening up the public SHRA software and models, it is of concern that it may never happen, and this would invalidate much of the progress achieved. The advice of 2021 is repeated here:

FAIR, Reproducibility and Transparency. *While the process of opening up the SHRA software and input model is, in principle, undisputed, it is still pending on licensing agreements etc. It is important to enhance reproducibility and transparency, and to respect the FAIR principles, that the full model input files and the software needed to run the SHRA, along with appropriate documentation, is made available very soon to the scientific community. It is important that license issues are resolved soon; Creative Commons 4.0 Attribution (CC BY), would be a possible candidate. This license is also compatible with the recommendations of the EPOS consortium.*

Because the KEM subpanel considers this issue vitally important, the primary general recommendation to SodM and EKZ is:

- **OR 1:** the KEM subpanel recommends that the TNO public SHRA model proposal (with potential modifications) is only implemented and used if the software framework and underlying model comments are openly available, respecting the FAIR principles commonly applied to public funding. The KEM subpanel is not willing to review the next model proposal unless open access has been implemented.

In addition, the following aspects about five aspects of the overall model development are of concern and the subpanel has some suggestions about them:

- **OR 2:** the presentation of the proposed 2023 model by TNO is in parts confusing, repetitive and the writing and table entries difficult to follow. While this is partially owing to the complex version history, the subpanel thinks that with some effort a better representation of the currently used model, the changes proposed and the rational for these changes could be presented in a comprehensive way, backed up where needed by detailed appendices or ideally in peer reviewed journal articles.

- **OR 3:** risk sensitivity is important as part of a proposed change, and the KEM-09 framework developed by TNO contains the necessary tools. It is recommended that any model development should be primarily evaluated against the effect on the final results of the risk analysis (i.e., the local personal risk), and risk disaggregation, to evaluate their cost/benefit with respect to risk analysis. This is requested not to guide the decision, but to enhance process understanding, as a sanity check and to judge the relevance and impact of the proposed change. It was also advised, and support has been provided by the KEM subpanel, that TNO considers a series of alternative risk metrics to explore the implications of model changes/updates/replacements more deeply, but TNO has acknowledged these recommendations in a very limited manner, if any.

- **OR4:** all subpanel members are convinced, based on their experience that it is not sufficient for verification and sanity check of a public or commercial SHRA model to check each individual component. It is vital to also check the combination of all components, and hence also the impact of the combined changes on the final risk metrics of interest. It may happen that, in a complex model, while all individual components make sense, the combination does not. Therefore, it is suggested to apply and document sanity checks, risk sensitivities and quality control to the full SHRA model before it is used for decision making. This is particularly important in the 2023 model proposal since numerous changes are proposed.

- **OR 5:** in the advice on the 2022 public SHRA, the KEM subpanel defined a set of guiding principles for model development and for judging model matureness. It was then foreseen that these principles would be useful and used in the future model development, but the 2023 proposal to TNO does not refer to these principles, nor argue why TNO considers them not worth applying. It is recommended to consider using them in future proposals to better justify the choices made.

- **OR 6:** the KEM subpanel is concerned that TNO is not able to allocate sufficient resources to the public SHRA model development, to ensuring QA/QC, to publishing results etc. It is sensed a mismatch between the expectations of SodM, EKZ and the KEM subpanel and one side, and the resources available to TNO on the other side, and some of the questions the subpanel has been asked to respond to may stem ultimately from this mismatch. Therefore, it is suggested to review the requirements and resource needs soon.

4. STATUS OF SEISMOLOGICAL SOURCE MODEL VERSIONS (SSM)

Below some different components of the SSM model, as proposed by TNO in section 4 of their report (2022d), are discussed.

1. **Data and workflow for model calibration:** it is agreed that it is important to use all available data for calibrating the seismogenic source model. The fact TNO was not able to reproduce the NAM calibration process (although differences are not substantial), is concerning and it is agreed with TNO that it is time to progress to a fully open and reproducible calibration process. The TNO calibration approach is well documented and sensible, it also seems robust, thus it is suggested following the recommendation of TNO to use the public SHRA workflow to calibrate the model parameters, and their uncertainty.

2. **Coulomb stress field:** it is agreed with the TNO proposal to using a posterior distribution of conditioning parameters obtained from Bayesian inference from the observations. TNO has demonstrated in their proposal and past work (e.g., TNO, 2021) sufficiently that their approach is robust and capable to characterise the uncertainty distribution well, and the subpanel considers it ready for the use in the SSM.

3. **Magnitude model:** as discussed in section 2 (sub-section 5c), the KEM subpanel agrees that the use of a stress-dependent exponential taper is not justified and may introduce a bias. It is agreed that the use of only the hyperbolic tangent Magnitude Model, tapered by a broad Mmax distribution, is sufficient to capture uncertainties.

4. **Mmax distribution:** the maximum magnitude workshop conducted by NAM in June 2022 was attended by several KEM subpanel members and the presentations conclusions of the workshop have been shared. As discussed in section 2 already, the KEM subpanel considers the workshop of overall good quality, attended by many of the leading experts on earthquake physics and induced seismicity. The expert panel likewise was composed of recognised and independent experts. While it is agreed that the issue of the potential coupling between a taper and the Mmax distribution could have been more broadly discussed,

the Mmax workshop is a good example of the usefulness of expert elicitation in Probabilistic Seismic Hazard Assessment. The recommendations of the maximum magnitude expert panel are sound, well-reasoned and representing the advances in the state of the art since the last assessment in 2016. They cannot readily be improved by TNO nor should they be ignored. Therefore, the Mmax distribution as defined by the 2022 expert panel should be adopted in the 2023 public SHRA with no changes. The KEM subpanel is sceptical that further investigation of the tail of the Mmax distribution, as suggested in the TNO report, will help soon to reduce the inherently broad uncertainty in the Mmax distribution.

5. STATUS OF GROUND MOTION MODEL VERSIONS (GMM)

TNO has implemented the GMM-V7 ground motion model in the Groningen hazard and risk model chain. The V7 GMM is the final refinement, in the NAM intentions, in terms of implementing the framework that was established at the V4 and V5 stages of the model development. The KEM subpanel supports this implementation since several important improvements (Bommer et al., 2022a) have been incorporated with respect to the V6 GMM and earlier models (e.g. use of an expanded and revised database with additional, new recordings, correction for soil-structure interaction effects at B-stations with basements, new record processing procedures applied to the complete database, a refined field-wide site model including a new model for soil damping, refinements to the FAS inversions including a layered Q model, logic-tree branches for between-event ground-motion variability).

A comparison with reference data provided by the model development team shows that the implementation is convincingly accurate. Explicit implementation choices have been made to calculate the effective weights of the Median branches for the full logic tree distribution and the site-to-site variability is kept out of the logic tree. These choices do not impact the mean/expected risk levels but can affect the percentile estimates of the logic tree.

The V7 prediction of PGA (equivalent the spectral acceleration at 0.01 s) tends to be underestimated at short distances in the magnitude range of the data. This is not viewed by Bommer et al. (2022a) as a major concern for application of the model in the public SHRA (through the average spectral acceleration). This may however be an issue if PGA values are used in future risk models.

These underestimations are also a perplexing feature for the V7 GMM team, TNO GMM developers and the KEM subpanel. The favoured explanation from Bommer et al. (2022a) is that the underestimation of observed high frequencies might lie in the role of P-waves and that these underestimations will be lower for large earthquakes (which are controlling the risk). This working hypothesis will remain difficult to test.

An alternative explanation may be related to the choices made to compute the non-linearity of site response or the choices and the wave damping. This hypothesis could be explored (e.g., update of KEM04 non-linear computations, test of alternative choices to adjust the Darendeli (2001) and Menq (2003) damping models used by NAM to develop the V7 model, sensitivity study on this maximum non-linear depth).

Both in terms of the average LPR and in terms of the number of buildings exceeding the norm, the GMM-V7 induces a reduction of risk relative to the GMM-V6 for the model chain according to the TNO preferred model chain configuration (but not for the public SHRA model for which an increase is observed). According to the TNO analysis, this difference is most probably related to the period-to-period correlation model that was not present in the public SHRA official model chain configuration but was included in the TNO configuration. This result confirms the importance of p2p correlations related choices and interest to develop a scientifically acceptable p2p model specific to Groningen. In addition, there should be a consistency between the actual p2p correlation structure as a result of the V7 implementation, and the p2p correlation structure used for the development of the fragility curves.

It should be also noted that the "GMMV7 implementation" means the replication of the GMPEs (the final step of the simulation chain developed by the NAM V7 team) but not the implementation of the full sets of V7 models (e.g., underlying stochastic models) which may have offered the possibility to perform sensitivity analysis for a few debatable parameters and fully reproduce the V7 modelling process. Such implementation would be highly desirable in particular to ensure a consistency between the source models highlighted during the 2022 Mmax workshop (i.e., ruptures confined within the reservoir, with unusually large length/width ratios), and those used in the stochastic simulations.

In conclusion, the subpanel appreciates the important but necessary effort made in 2022 to implement the V7 version of the GMM model and to test the consistency of the results with the version developed by NAM. While some open issues remain, public SHRA from 2023 onwards should use the GMM V7. In 2023, efforts should take place to consolidate and explore the GMM V2 and p2p correlation models, and the following actions should be undertaken:

- **R5.1:** Developing a p2p correlation model based on the data acquired in Groningen
- **R5.2:** Evaluating alternative choices to compute the non-linearity of the amplifications and the effects of damping
- **R5.3:** Implementing the full sets of V7 models (including its underlying constituents), and starting to investigate whether the geometry of extended sources for moderate events between Mw 3 and 4.5 might significantly impact the surface ground motion predictions using the stochastic simulation approach

6. STATUS OF FRAGILITY AND CONSEQUENCE MODEL VERSIONS (FCM)

As noted already by the subpanel, it seems that, in general, FCM has generally received less attention than SSM and GMM in the process of public SHRA, although TNO propose a typological approach that is somewhat alternative to NAM's. It is underlined by the subpanel that fragility modelling, unlike SSM and GMM, is mostly based on numerical building modelling and simulations rather than observed (experimental) data; thus, it is much more dependent on working assumptions, an issue that makes it especially delicate. Moreover, the structural engineering need of structure-specific modelling is unfeasible given the size of population of the building stock in Groningen. Therefore, necessarily FCM needs very much simplified modelling of only index-building being representative of a wide class, with the definition of classes, their extensions being arbitrary. The intra-class variability of structural features and its reflection on the class-specific fragility, also requires attention as per the modelling in FCM V6-V7.

One of the key components of the risk assessment is the modeling of the volume loss in case of damage. It is also known that the single-degree of freedom systems – vastly used in V6 and V7 FCM – are not suitable to explicitly model the volume loss from displacement seismic response. Thus, the FCM put in place by NAM relies on semi-empirical relationships from the structural response to the volume loss. It is considered somewhat a priority to put under scrutiny the chain from response to fatality to find room for improvements.

More in general, several comments given already by the subpanel advice entitled 'KEM subpanel advice to MEA on the Groningen SHRA model components to be used in 2022' still apply to the FCM.

7. CONCLUSIONS AND SUMMARY

The KEM-sub panel was asked by SodM to comment on the model components proposed by TNO for the SHRA for the gas year 2023. It was asked to comment on the difficult questions of when a model component is to be considered scientifically ready for being included in the SHRA model and if the QA/QC procedures are adequate. The comments given above extensively answer to all questions raised and comment further on the model changes proposed by TNO. Below some comments on the procedure of giving advice and a summary of the subpanel recommendations are given.

7.1. Procedural comments

The subpanel would overall like to comment that compiling the advice has been challenging for us, for several reasons:

- the timelines are very tight, also considering that the end of the year is typically a busy period for the panel members due to their primary institutional roles;
- this year, numerous potential changes in the public SHRA are being considered, and these are also mixed with relevant questions (on quality assurance etc.) requiring a broad response;
- progress by TNO on publishing the overall software framework, as well as on model components, has been slower than hoped for by the subpanel, limiting the exposure to peer review, which in turn would have been beneficial with respect to answering the questions raised by SodM;
- the guiding principles for model selection recommended by the subpanel in the advice of November 2021, and in further and/or related interactions, were largely not used by TNO in their developments and thus are not

reflected in the report, making it more challenging to assess the model components for their suitability as part of the public SHRA.

The panel believes that the questions raised are responded to adequately and with sufficient depth and reflection. Hopefully the advice will be useful for SodM and EZK in their challenging task of overseeing the Groningen seismic risk assessment. Note that several of the recommendations we give are more related to the medium or even long-term model development procedures, rather than the actual 2023 implementation proposal. This is somewhat unavoidable since model development and implementation are a continuous process. The panel would welcome the opportunity to discuss early next year how the process of giving advice can be further optimised.

7.2. Summary of advice

In the preceding sections these choices and the rational for their implementation were addressed in detail. The advice to SodM and EZK, purely based on the scientific evaluation of model components, is summarised in Table 1. In summary, the advice follows in all model components the proposal of TNO.

SHRA model options		2019	2020	2021	n/a	2022				2023	
Nr.	Model and submodel version option (ref)	NAM/ HRA	NAM/ HRA	pSHRA	NAM/ V7	2022 tno	2022 kem	2022 sodm	2022 used	2023 tno	2023 kem
	Main model version options										
INF-NAM	Model version V5: SSM-NAM-V5, GMM-NAM-V5, FCM-NAM-V6 (NAM)	x									
INF-NAM	Model version V6: SSM-NAM-V6, GMM-NAM-V6, FCM-NAM-V7 (NAM)		x	x				x	x		
INF-NAM	Model version V7: SSM-NAM-V7, GMM-NAM-V7, FCM-NAM-V7 (NAM)										
INF-TNO	TNO 2021: SSM-TNO-2020, GMM-V6, FCM-TNO-2020 (TNO 2021 R11742)					x	x				
INF-TNO	TNO 2022: SSM-TNO-2020, GMM-V7, FCM-TNO-2020 (TNO 2022 R11961)									x	x
	SSM submodel version options										
SSM-Cal	NAM modelcalibration provided as input (na.a)	x	x	x				x	x		
SSM-Cal	TNO model calibration as input (TNO 2021 R11742, TNO 2022 R11961)					x	x			x	x
SSM-Coul	Coulomb stress predictor for activity rate (B&O 2018)	x									
SSM-Coul	Coulomb stress predictor for activity rate and magnitude distribution (B&O 2019)		x	x				x	x		
SSM-Coul	Coulomb stress distribution (TNO 2021 R11742)					x	x			x	x
SSM-Arate	Activity rate (Lin. Elastic compaction) (B&O 2018)	x	x	x		x	x	x	x	x	x
SSM-Arate	Activity rate (Rate Type isotach Compaction Model (de Waal 1986, Pruiksma, 2016, TNO 2022 R11961)										
SSM-ETAS	Epidemic type afterschock distribution (B&O 2018)	x	x	x		x	x	x	x	x	x
SSM-MD	Magnitude distribution: Constant B value model & Mmax distribution (TNO 2020 R11052)										
SSM-MD	Magnitude distribution: Inverse powerlaw B-value model & Mmax distribution (B&O 2018)	x									
SSM-MD	Magnitude distribution: Hyperbolic B-value model & Mmax distribution (B&O 2019)		x	x		x	x	x	x	x	x
SSM-MD	Magnitude distribution: Stress dependent B-value and exponential taper B-value model and Mmax dsitr. (B&O 2019)		x	x				x	x		
SSM-MD	Magnitude distribution: Spatially dependent B-valty & Mmax distr. (TNO 2022 R11961, Kraaijpoel et al 2022)										
SSM-LogT	Mmax distribution 2016 (NAM 2016)	x	x	x		x	x	x	x		
SSM-LogT	Mmax distribution 2022 (NAM 2022)									x	x
	GMM submodel version options										
GMM-V5	NAM-V5 (Bommer at al. 2017)	x									
GMM-V6	NAM-V6 (Bommer er al 2019)		x			x	x				
GMM-V6-2021	NAM-V6-2021 (Bommer er al 2017, p2p removed in SR)			x				x	x		
GMM-V7	NAM-V7 update, incl. Wierden (Bommer et al., 2021, TNO 2022 R1081)									x	x
	FCM submodel options										
FCM-V6	NAM (Crowley et al, 2017)	x									
FCM-V7	NAM (Crowley and Pinho, 2020)		x	x			x	x	x		x
FCM-TNO-2020	TNO typologies (TNO 2021 R11742, App. B)					x				x	
	Test and comparison framework										
TCF-2022	Framework for reproducable and traceable sensitivity analysis and testing (TNO 2022 R11961, TNO 2022 R12442)						x			x	x

Table 1: *Summary of the model component used in different generations of the Groningen SHRA and SHRA. Marked with 'x' are the model components used, the final column indicates in green the preferred choices argued for in this advice by the KEM subpanel. Orange colouring indicates that the advice differs from the suggestion of TNO for the 2022 model components.*

In the 2021 advice to EZK and SodM, the subpanel suggested that instead of selected updates to the SSM model, it may be preferential to delay this update in favour for a 'large' update that would also include GMM V7 and the new Mmax distribution, and indeed this was the eventual decision by EZK. The KEM subpanel suggests that now the time has come to implement the various changes in the model components, to avoid that the public SHRA model is outdated and not representing the state of the art any longer. Such a large update, however, is also challenging because interface issues may be detected. As mentioned before, it is recommended that the updated model results are very carefully checked and verified and evaluated for plausibility by internal as well as external experts. One critical step here is, as emphasized several times throughout this advice, to make the full model chain and computational codes, openly available to the community.

8. REFERENCES

Bommer JJ, Edwards B, Kruiver PP, Stafford PJ, Dost B, Ntinalexis M. (2018). V5 Ground-Motion Model (GMM) for the Groningen Field Re-issue with Assurance Letter. NAM, Assen: 2018.

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 Gas Field. Revision 1, 19 December 2019, 196 pp.

Bommer, J.J., B. Edwards, P. Kruiver, A. Rodriguez-Marek, P. Stafford, M. Ntinalexis, E. Ruigrok and B. Dost (2022a). Ground-Motion Model V7 for Induced Seismicity in the Groningen Gas Field (Update) With Assurance Letter, 282 pp.

Bommer, J.J., et al. (2022b). Report on the Second Workshop on Mmax for Seismic Hazard and Risk Analysis in the Groningen Gas Field (Main Report), 13-17 June 2022, The Infinity Building, Amsterdam, 1222 pages.

Bourne, S. J., & Oates, S. J. (2017a). 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: Solid Earth, 122, 10,299–10,320. https://doi.org/10.1002/2017JB014356

Bourne, S.J., & Oates, S. J. (2017b). Development of statistical geomechanical models for forecasting seismicity induced by gas production from the Groningen field. Netherlands Journal of Geosciences — Geologie en Mijnbouw |96 – 5 | s175–s182 | 2017 doi:10.1017/njg.2017.35

Bourne, S. J., Oates, S. J., & Van Elk, J. (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. Geophysical Journal International, 213(3), 1693-1700.

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. and Pinho R. (2020). Report on the Fragility and Consequence Models for the Groningen Field (Version 7). NAM, Assen.

Crowley, H., Pinho, R. and Cavalieri, F. (2019). Report on the v6 Fragility and Consequence Models for the Groningen Field, March 2019.

Darendeli, M. (2001). Development of a new family of normalized modulus reduction and material damping curves. Ph.D. Thesis, Dept. of Civil Eng., University of Texas, Austin, TX.

De Waal, J. A. (1986). On the rate type compaction behaviour of sandstone reservoir rock. PhD Thesis, Delft University of Technology, 1986, 164 pages. Available at http://resolver.tudelft.nl/uuid:b805782b-2eb4-4f72-98f4-f727c4ea9df0

KEM subpanel advice to MEA on the Groningen SHRA model components to be used in 2022, December 10, 2021, 11 pages.

Kraaijpoel D, Martins JE, Osinga S, Vogelaar B, and Breunese J. (2022). Statistical analysis of static and dynamic predictors for seismic b-value variations in the Groningen gas field. Netherlands Journal of Geosciences, Volume 101, e18. https://doi.org/10.1017/njg.2022.15

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

Menq, F.Y. (2003). Dynamic Properties of Sandy and Gravelly Soils, PhD Thesis, Department of Civil Engineering, University of Texas, Austin, TX.

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.

TNO (2020). Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen. TNO 2020 R11052, 72 pages.

TNO (2021). Status of the TNO Model Chain Groningen per October 1, 2021 and recommendations for the public Seismic Hazard and Risk Analysis 2022. Report TNO 2021 R11742| update 2022, by M.P.D. Pluymaekers, D.A. Kraaijpoel and S. Osinga, 15 October 2021, 55 pages.

TNO (2022a). KEM9: Cumulative propagation effect of Groningen risk model component uncertainties on hazard and risk predictions, Report TNO2021_R12442 by F. Aben, S. Osinga, D. Kraaijpoel & M. Pluymaekers. 24 June 2022, 33 pages.

TNO (2022b). Implementation of GMM-V7 in the TNO model chain, report TNO2022_R10801 by D.A. Kraaijpoel, P.M. Aben, and M.P.D. Pluymaekers., 18 July 2022, 28 pages.

TNO (2022c). SDRA Modelketen Groningen, internal report, TNO, version 1.3, 06-10-2021.

TNO (2022d). Status of the TNO Model Chain Groningen per October 1, 2022 and recommendations for the public Seismic Hazard and Risk Analysis 2023, Report TNO2022_R11961 by M.P.D. Pluymaekers, F.M. Aben, S. Osinga and D. Kraaijpoel. 18 October 2022, 44 pages.