[redacted]

Prof. Dr. Stefan Wiemer
Swiss Seismological Service, ETH Zürich
Sonneggstrasse 5, NO H61
8092 Zürich

[redacted]

www.seismo.ethz.ch

Zürich, Nov. 22 2015

Dear [redacted] dear [redacted]

The State supervision of Mines has appointed me in the summer 2015 as an expert reviewer and advisor on induced seismicity in the Groningen area. As part of this mandate I was asked to review the report *"Hazard and Risk Assessment for induced seismicity in Groningen (interim update November 2015)"* by NAM, dated November 7 2015. My review will be based primarily on the aforementioned report; however, it will also consider on the various supporting documents supplied by NAN as well the discussions during the two feedback meetings I attended.

Please note that my review expresses my own personal opinion as a seismologist with expertise on seismic hazard and risk assessment. It does not represent an official position of the Swiss Seismological Service or ETH Zurich. Please also note that, in compliance with your request, my review will be mostly focused on the 'big picture', not on detailed questions on individual subcomponents of the model or computations. While there are numerous suggestions in my review, I am well aware that some, or even the majority, will end up not being implemented. Opinions differ and overriding constraints may exist. Please consider my ten recommendations as suggestions from a somewhat outside perspective, which hopefully you will find helpful input for your difficult task. Finally, reviewers are by the very nature of their assignment often critical. I think it is important to state that overall I observe great and rapid progress made the past three years towards an quantitative assessment and management of the risk posed by induced earthquakes in the Groningen area.

I would agree if my review, or parts of it, are made available to other scientist or even publically if you consider this appropriate. In this case, I would appreciate to be informed beforehand.

With kind regards

Stefan Wiemer

1. Comments on the overall approach and current status

The dramatic increase in induced seismicity in the Groningen area poses a problem of great societal and economical relevance. In such situations, it is the role of science and engineering to produce the factual baseline for discussion and decision-making. Ideally, risk-cost-benefit analysis offers a transparent pathway to assemble and integrate relevant evidence to support such complex decision-making, considering also the uncertainties and knowledge gaps. In a recent review article, Fischhoff[1] outlined and discussed in some depth this kind of approach, based on selected case studies.

I fully support the path adopted by the operators (in short NAM from hereon) and State supervision of mines (SodM) to build up a fully probabilistic, state of the art (and partially beyond), time-dependent seismic hazard and risk model. The model that is now emerging, if successfully calibrated and partially validated, should in the near future enable risk analysis, risk assessment and risk management, including decision support for control and mitigation actions.

I also overall support the methodology adopted, which in essence decomposes the complex overall system into manageable components (the well-known bridge or train) and then calculating the overall system response. The model largely follows also the established best practice of representing uncertainties in formalized ways, although, as detailed below, there is a heterogeneous, partially inconsistent approach adopted across the different components of the overall model.

The operator invested substantial and much appreciated efforts into gathering over the past two years a wealth of geophysical, geotechnical and engineering data of relevance. These data are essential to calibrate the model components for a site-specific Groningen model, and the efforts are nicely starting to find their way into the model. In my assessment detailed below, the work has progressed at somewhat different speeds and with different rigor in the various model components, specifically, the seismological analysis and model building is somewhat lagging behind.

Recommendation 1: The overall approach adopted by the operator to the assessment of the risk posed by induced earthquakes in the Groningen area is, in my opinion, fully appropriate and almost without alternative. It should be continued and extended in the future, in an evolutionary sense.

The model version 2.0 as of November 2015 is clearly much more advanced and appropriate for decision making than version 1.0. The methodological approach has been refined, vast amounts of new data have been integrated, various sensitivities tested, etc. The complexity of the model and the number of (free) model parameters has also greatly increased. As a consequence, model verification, calibration and independent validation is becoming increasingly a challenge, one that is in my assessment only partially addressed so far. Also, the interdependencies between model components have not been fully explored. It is a well-known fact in the re-insurance business that combining all the best individual component of a complex risk model may in the end lead to a model that is in contradiction to the historical record of shaking and damages. Issues that have not yet been explored in many depths, for example, are the potential of double-counting uncertainties between ground motion, site amplification and building fragilities.

---

[1] B. Fischhoff, Science 350, aaa6516 (2015). DOI: 10.1126/science.aaa6516

It is also a fact that much of the model and model calibration has been done under great time pressure, and that some of the model components are less well tested and developed than others, as detailed below. The model is clearly at this stage 'work in progress', and will likely continue to evolve substantially in the next months. Not all model parts and decisions taken along the way are fully transparent and reproducible, and no independent verification of model components has take place. There is a certain degree of arbitrariness in decisions on parts of the logic tree model adopted, and on the weights assigned, which could potentially have a large impact on the overall hazard and risk levels.

Recommendation 2: The model in its current development stage is in my assessment not yet robust enough for drawing firm quantitative conclusions about the absolute level of hazard and risk, its spatial distribution and the effectiveness of risk reduction strategies. However, it is conceivable the model version 3.0 will have evolved such that informed decisions on the Winningplan 2016 are feasible.

2. Comments on process ownership and risk governance

Successful risk governance is an analytical-deliberative process in which field operators, independent risk analysts, regulators and stakeholders collaborate in managing risks. Key elements of this process are a clear separation of roles, the transparency of the process and the quality of the exchanges and communication. Ultimately, the acceptance of the risk assessment for decision-making transpires from the trust in the models and in the ones that have constructed the model. In the Groningen induced seismicity case, trust into the field operators has at least been partially been lost, and re-gaining trust inevitably will be a substantial effort and take a long time.

In the seismic risk assessment of Groningen, there is also in my assessment a potentially detrimental conflict of interest between the interest of NAM in operating the gas field ways that maximizes the economical gain and the recommendations from the seismic risk model on minimizing the risk. The potential impact of this conflict of interest is strongly amplified by the existing mistrust between the various parties involved. Added to this mix is the fact that the risk model is highly complex and computationally demanding, with countless model parameters, and numerous decisions to be made in all stages of the modeling development, verification, calibration and validation. Currently, it is impossible for SodM, TNO or KNMI to fully verify or reproduce the results of the computations risk; even understanding exactly what has been done and what the model sensitivities are is a challenge in such a complex and rapidly developing model.

In my assessment, based also on the experience of similar entrenched discussions on shale gas in the UK, and nuclear energy in Switzerland, it is well possible that the risk model will not be widely accepted as a basis for decision making because of this - perceived or real - conflict of interest, the complexity and irreproducibility of the model, and the partially unclear definition of roles of individuals and organizations. An additional layer of complexity is added by the fact that (1) many of the staff members have been employed at various stages in their careers by the various organizations (NAM, SHELL, KNMI, TNO, SodM), that (2) much of the primary data is collected by KNMI and TNO and that (3) the roles of external experts used by the project, by the external consultants and by state agencies are not always clearly defined. There are clear benefits to the collaborative process adopted in many risk governance projects. Structured feedback, and open discussions between all parties are vital components of a successful project. However, these need to be embedded in a framework that transparently establishes well-defined roles for all parties.

In my assessment, the acceptance of the Groningen seismic risk model would ultimately be much higher if the model and the process to build, maintain and update it would be owned, or at least co-owned, by an agency that has no obvious conflict of interest. For example, through a long-term contract the field operator, regulator and possibly other parties would all have full access to and co-ownership of the model and data, the model would be developed in a collaborative spirit. The partners would oversee and co-own the development roadmap of the model and guarantee its continuous review and quality assurance. External consultants conduct much of the work in any case, and this could be continued. However, the consultants should then report to an independent agency rather than NAM.

Recommendation 3: The process of risk governance overall should be reflected on beyond the technical aspects of risk assessment, in order to maximize the wider legitimacy of the work. In the current setup, the roles and interactions of involved parties, but also the ownership of the model overall, is in my view problematic.

3. Comments on the use of consultants, external experts for review and quality assurance

NAM is relying on a number of external experts to build the risk models. In my assessment this is a key element of the quality assurance and ultimate success of model building. NAM has attracted some of the leading experts and consultants in the field of ground motion prediction, local site response, geotechnical aspects and earthquake engineering. These experts have extensive experience in project management of similar projects, a further key ingredient for success. They seem to have established a good working relationship with staff at NAM. The one exception to this is the seismic source model (from gas production to hazard, chapter 3), where, to my knowledge, in-house experts conduct most of the developments.

As stated in the report in various places, and listed also in Appendix B of the report, a large group external experts has been involved in various stages of the process and for selected components of the modeling. This kind of external expert involvement is potentially highly useful to improve the quality of the work. However, I have three concerns related to the involvement of external experts:

- The use of experts is somewhat inconstant between the various components of the model. The GMPE and earthquake engineering on the one hand parts have established a somewhat formalized hazard and building fragility review team, and use them to a certain extend to capture also the epistemic uncertainties and to assign logic tree weights. However, the 'seismogenic source model' and - partially - the site characterization, seems to involve external experts to a smaller degree, and in a less formalized way.
- There is an apparent lack of clear rules and responsibilities of external experts, observers, guest and reviewers. Appendix B for example is a long list of people that had at some point interactions with the project (including myself), with no distinction and responsibility. The written reviews of external experts should be made openly available, and the way the expert feedback is used, or not used, justified. It is not clear if the feedback workshops are seen as expert elicitations, if they reflect also on 'ownership' of the model and how they specifically help to capture uncertainties.
- The choice of the experts is somewhat arbitrary. At least in theory it is possible that this choice biases the feedback and input received.

Recommendation 4: The use and roles of external experts should in the future be more formalized and applied consistently across the various model components. More expert feedback/elicitation especially on the 'sources' model seems appropriate.

4. Comments on the analysis of seismic data

A key element for enhancing process understanding and forecasting skill in the future is in my assessment a more advanced analysis of the micro-seismicity. In my opinion[2], there are numerous first-order questions of relevance that need to be addressed in much more detail than attempted so far. I list a few examples below:

1) Improving the precision and accuracy of hypocenters. Using advanced (re-) location techniques, applied to past data as well as in near-real time, this will also require a new minimum 1D and 3D seismic velocity model (P- and S-; separately) as well applying relative re-location algorithms. Potential benefits include:
    a. Improved spatial distribution of patterns of seismicity as input for rate density models, characterizing clustering etc.
    b. Ability to correlate faults and hypocenters, leading to an improved geo-mechanical understanding of the coupling between compaction, faults and seismicity.
    c. Ability to analyze with more confidence migration path of seismicity, identification of lineaments etc.
    d. Improved depth resolution, ability to detect activation of basement faults.
2) Targeted efforts to detect and locate smaller earthquakes, using for example automated template-matching approaches. Potential benefits include:
    a. Enhance the existing and future seismicity record by an order of magnitude in terms of the number of events, leading to much more robust statistical patterns (rate density, size distribution) and geo-mechanical interpretations.
    b. Improved ability to calibrate and validate forecast models.
3) A Groningen specific space-time model of completeness and homogeneity of magnitude reporting (using Gutenberg Richter based techniques, BMC, or PMC combined with noise analysis at the sites). Given the rapid changes in the network, such a space-time model of the reporting completeness will be important to be able to assess the relative earthquake size distribution, defined the activity rates with high resolution but also help in the optimization of the network performance.
4) Improved source characterization of future and past events, integrating also the local site amplifications, in order to determine in near-real time also moment magnitudes, corner frequencies and stress drops for small events, and revisit old magnitude and magnitude scales. This will lead to:
    a. Improved assessment of space-time patterns in the earthquake size distribution.
    b. Improved link of physical parameters to ground motion predictions (e.g., stress drop of events as a function of location,).

---

[2] There is an obvious potential bias in my stated opinion that the analysis of seismicity data and the seismogenic source model are most in need of additional work: It reflects my core competence and in other domains my own ignorance may prevent me from seeing the need for additional work.

5) Improved ability to assess focal mechanism of also smaller events.

   a. Improved link between faults and earthquakes, input for the geo-mechanical model.

6) Systematic search for slow/unusual events or non-volcanic tremor, possibly an indication of a-seismic motions, and potentially linked to the partitioning between seismic and a-seismic deformation, a poorly understood feature of the deformation.

7) Precursory activities as part of warning. Using the improved seismicity is may be possible to detect trends in micro-seismicity that could be indicative of upcoming larger (M>3) events. This may have potential for short-term (days to weeks) hazard and risk assessment.

This list is not very original; most of the proposed methods are based on well-established techniques documented in the literature. Much of this work is as far as I know already on the way or anticipated in the future, and the importance of addressing these questions is generally well established. In my assessment, the limits of the seismological analysis are the primary bottleneck limiting the development and validation of forecasting models.

There has also been a commendable effort to improve the seismic network in the Groningen area to an adequate monitoring level. In my assessment, however, the progress in addressed the aforementioned questions is too slow, given the urgency of the needs of the hazard and risk model. It is also substantially slower than progress in the GMPE/site and building parts of the model. This may be partially related to the fact that considerably fewer resources are available for seismological analysis. External consultants have been hired by NAM for GMPE and building parts, whereas the R&D on seismicity analysis resides largely with KNMI with less flexibility to substantially scale up the effort rapidly, in addition to the network building and hazard modeling also ongoing at KNMI.

Recommendation 5: The efforts related to analyzing and interpreting the seismological data should be prioritized and up-scaled substantially. This may require in addition efforts of groups outside of KNMI and NAM.

5. Comments on the seismogenic source model

The seismogenic source model is by requirement the most innovative part of the entire analysis chain. In the GMPE, site and building parts, the work is ambitious in scope but in essence based on existing state of the art and methodology. However, there is no precedence for such a complex time-dependent seismogenic source model that links production rates and future seismicity. The implications of the rate model is also the one that is most understandable to non-scientist, since the magnitude and frequency of events is easily understood and checking the model against observations is straightforward.

The seismogenic source model as it stands now is in many ways innovative and sophisticated; however, I question that it covers the requirement for PSHA to be exhaustive in capturing the epistemic and aleatory uncertainties. A key requirement of a PSHA is *"to represent the center, the body, and the range that the larger informed technical community would have if they were to conduct the study"*.

The current model implements one possible pathway to a rate forecast, and calibrates it to observations, achieving a good fit. Alternative implementations using, for example, largely statistical models, using rate and state approaches, using strain-rates rather than compaction and strain-thickness, using different parameterization of b-values, and accounting differently for the potential of 'non-linear', unexpected behavior of earthquakes could be imagined and are documented in the literature on induced earthquakes. These alternative models may all fit the observed rates equally well, but they may substantially differ in their forecasts and in the impact that production changes may have on future seismicity.

In addition, the seismogenic source model in its current state of documentation no fully reproducible and transparent. The compaction model, as well as the coupling of compaction to rates, is described not in sufficient detail in the available documentation, or only in somewhat outdated versions were not all recent decisions takes are well documented and well justified.

The spatial and temporal variability of b-values is clearly a driver of the risk, and its future evolution one of the key elements that could substantially alter the hazard and risk estimates. This is also a substantial source of uncertainty, given the limited physical understanding of the link between b and stress/strain/strength/structure, and given the limited data available in Groningen. Coupling the strain thickness to b-values as applied in Version 2.0 is an interesting idea, but has to my knowledge not been done in past hazard studies nor published in the peer reviewed literature. A number of alternative models exist (dependence of b-values on differential stresses, on faulting styles, only on space-time patterns, constant overall, etc.), a source of uncertainty that has in my assessment not been fully appreciated in Model 2.0.

I suspect that hazard and risk sensitivity estimates show currently a weak dependency on the seismogenic model also because the uncertainties have not been explored as systematically as done for the GMPE, site and building part. This is especially important since it has the potential to also underestimate the effect of changes in production rate on earthquake hazard and risk.

Recommendation 6: The seismogenic source model is in my assessment not diverse enough to satisfy the usual PSHA requirement of *capturing the center, body and range of the informed technical community*. Additional efforts are warranted to widen the model/uncertainty space.

The potential for using ensemble models

Ensemble models are a well-established tool in many areas of sciences, such as weather and climate forecasting. Recently, such methods have been adopted to include time-varying hazard and risk assessment of natural seismicity as well as induced seismicity in geothermal contexts. In all of these cases, ensemble models can indeed outperform each individual model, especially if the model weights are adjusted dynamically, based on the past performance of the entire model, or of model components. The Bayesian framework implemented by NAM for Model version 2.0 likely is similar in spirit; however, it appears to be used primarily to select models, not combine them. Ensemble forecast framework works best if a wider range of input models is used to make forecasts, and if it is used also in near-real time to update forecasts and weights on the fly as new data arrives and model performance is re-assessed automatically. A dynamically weighted ensemble model of a rage of models, some simple, some complex, would in essence replace the need for a logic tree to capture uncertainty in the time-dependent forecasting model. It would likely result in a more robust forecasting model overall.

Recommendation 7: The use and potential benefit of integrating ensemble models should be further explored and formalized.

Model validation using forecasting experiments

The principal challenge to the entire model, but specifically to the time-dependent seismogenic source model, is in my assessment the validation of the model against independent data. This is a common problem in many areas of science: Extrapolating a model outside of the range calibrated by data is challenging, and it is then when over-parameterization can make a big difference and where more simple models may in fact perform better than complex ones. In the case of Groningen, extrapolations into the future and outside of the magnitude range experienced are the key challenges for models.

In my assessment, a critical element of model validation and ultimately of building up trust in the model by scientists and the public would be a somewhat formalized, community accepted approach to test the model performance against data. This includes:

- The formalized use of pseudo-prospective test: how well do different models forecast the space-time-magnitude evolution of observed seismicity when given a limited learning period?
- The use of fully prospective, formalized and potentially independently conducted assessment of the forecast performance. This gold standard in testing is widely applied also in the medical industry: Any drug must be proven to work in so-called double-blind studies. The same standard could be applied here.

Formalized testing is first of all a powerful tool to analyze and understand the performance of models: Where and when do they do well, when do they fail? How well can we forecast future seismicity? There is a wide set of literature in the seismological community on forecast model evaluation, which is increasingly applied when evaluating especially time-depended forecasting models and 'Operational Earthquake Forecasting'. Much of this has been conducted in the spirit of the Collaboratory of Earthquake Predictability (CSEP; www.cseptesting.org). There are also applications to induced seismicity emerging in the literature.

Recommendation 8: Formalized, independently conducted testing of the future performance of seismicity forecast models should be considered as a key ingredient to improve model building, an element of model validation and an important component to build up confidence in the performance and reliability of the seismogenic source model.

6. Comments on the Ground Motions Prediction Model

One oft he major achievements of model 2.0 is clearly the development of a Groningen specific ground motion prediction model (in short 'GMPE model'). There has been a substantial investment of resources, resulting in substantial progress. Given the importance of the GMPE mode, reconfirmed in sensitivity analysis, these efforts are well justified. The work is conducted by some of leading scientists in the field and reviewed by an impressive group of international experts.

The key uncertainty in this domain is extrapolating to the ground motions of large magnitudes, which have not been observed in the instrumental record so far. To capture this uncertainty for the rock hazard, a logic tree with a number of stress drop values is adopted.

Key questions that in my view should be addressed with even more care in Version 3.0 of the GMPE model are:

- What is the evidence and how certain can one be that also induced earthquakes will be of lower stress drop, even if they would not fall with the reservoir and considering the fact that mostly data from injection (not depletion) related induced seismicity may (in some cases) have lower stress drops?

- Can it be already fully justified to only use a stochastic model, ignoring empirical GMPEs?
- Because designing and setting the weights of the logic tree for different stress drops has a large effect on hazard values, the process of defining branches and setting weights must be as transparent and independent as possible. How can this be done more independently and more formalized? Who is in the end responsible for setting the weights? What is the role of the international experts in assigning the branches and weights?
- Are there alternative ways to translate rock hazard to local site conditions, is the current approach capturing the considerable uncertainty well enough?

Two of these concerns related back to process of risk governance overall, and on the use of experts (my points 2 and 3).

Recommendation 9: Recognizing the importance of the GMPE logic tree and its weight to the overall hazard level, designing of the tree and setting these weights must be achieved as transparently and independently as feasible. The benefits of structured expert elicitation should be considered.

7. Comments on the Maximum Magnitude Assessment

Model Version 2.0 adopts a simple logic tree to express the large uncertainty of the maximum magnitude possible. This is a reasonable intermediate approach that more accurately reflects the fact that indeed there is a large uncertainty and widely varying opinions. It is also sensible that a workshop on the subject will be conducted in the preparation of Version 3.0. However, in my assessment there are potential problems in the procedure adopted:

- To what extend is it possible in a short workshop of experts only partially familiar with the regional context and with little familiarity to the Groningen seismogenic source model to draw meaningful conclusions and distill them into a logic tree?
- What is the ownership these experts will have on the Mmax model and how will their opinions be integrated in a formalized way to set the weights of an Mmax logic tree? How and by whom are the experts selected?
- Is it feasible to decouple the Mmax problem from the rest of the seismogenic source model, where this group of external experts has no involvement and ownership?
- Could there be useful preparatory work on Mmax estimates that will enrich/collect the empirical databases, present selected case studies or dedicated attempt to modeling? Do, for example, the extreme case such as the three Gasli M7 earthquakes in Uzbekistan, or the M5.4 Lorca earthquakes in Spain, offer meaningful insights?
- Can the potential of the seismicity to be triggered beyond the reservoir area be quantified, such as the potential to re-active faults in the basement?

Recommendation 10: The Mmax workshop is likely to be focused on highly controversial topics with widely varying opinions between exerts. It is important to prepare the workshop well, also considering if a more structured expert elicitation is needed or useful.

8. Comments assessing building fragilities

The work on characterizing the fragilities of buildings has in my assessment, greatly advanced in Version 2.0, owing to the considerable and well-directed efforts of the earthquake-engineering domain. Not being much of an expert in the domain myself, I am impressed by the efforts undertaken, which should clearly aid in quantifying the risk and target the retrofitting efforts. The role of external experts to review the model is also much appreciated for quality assurance, although my comments on the role and ownership of experts apply here also.

The only concern that I would like to comment on is an obvious one: Does the current fragility model capture the diversity of buildings? Are for example the impressive full-scale tests conducted in Pavia indeed representatives for this building typology, are the effects of imperfections, aging etc. well enough understood to draw firm conclusions? Are there additional validation experiments in the field, or numerical modeling, that can firm up the transferability of the results to the wider Groningen building stock?

9. Knowledge transfer from other fields

There are countless oil and gas fields on Earth that are experiencing subsidence similar to the Groningen area. Groningen is somewhat unique not because of the induced seismicity observed but because of the population density. However, there is only very limited information available on seismic events in these fields, because the relevant data does not exist or is not shared openly by operators and regulators. In my assessment, much could be learned if more case studies of high quality could be conducted. Because Shell and Exxon are major players in oil and gas extraction, they may be in a position to advocate the benefits of more transparency on induced earthquakes for all involved parties, and develop a strategy how validation of models can also be advanced by application to a wider range of case studies.