

*Ensemble Based Subsidence application to
the Ameland gas field – long term
subsidence study part two (LTS II)*



NAM

Contents

1	Summary.....	5
2	Introduction.....	8
2.1	Summary of part 1 (NAM, 2015a)	8
2.2	Objective of the Long Term Subsidence study part two (LTS II)	9
3	Regional Geology.....	11
4	Historic development and current state of the Ameland area	16
4.1	Ameland Oost.....	17
4.2	Ameland Westgat	17
4.3	Ameland-N07FA	17
5	Dynamic modeling and evaluation	19
5.1.1	Results assessment.....	22
6	Use of geodetic data.....	23
6.1	Innovations as recommended in LTS phase 1 (2013-2015)	23
6.2	Survey data and measurement noise.....	24
6.2.1	Levelling.....	24
6.2.2	GPS network (since 2006).....	24
6.2.3	GPS baselines (1993-2004)	25
6.2.4	InSAR.....	25
6.3	Idealisation noise.....	25
6.4	Outlier handling.....	26
7	Workflow confrontation of model results with measured data	28
7.1	General, description	28
7.2	CUPiDO	29
7.3	Calculate compaction	31
7.3.1	Upscaling	31
7.3.2	Compaction models.....	31
7.4	Calculate Subsidence.....	35
7.5	Model confrontation and conditioning	36
7.5.1	Red flag confrontation method	37
7.6	Demonstration of the ESIP workflow	38
7.6.1	Upscaling and interpolation steps.....	38
7.6.2	Subsidence calculation	39

7.6.3	ESIP-outcome reproducibility.....	42
7.6.4	Comparing ESIP output with known input	45
7.6.5	Effect of data density on model discrimination	48
7.6.6	Summary of ESIP testing.....	57
7.7	Comparison of AEsups with FE element numerical simulations: testing the influence of salt flow. 57	
7.8	The effect of the Ameland salt dome on the position the point of maximum subsidence with time.	68
8	Results of the confrontation.....	72
8.1	Selection of benchmarks	72
8.2	Selection of reservoir scenarios	73
8.3	Selection of area of interest for subsidence rate calculation.....	75
8.4	Selection of reservoir and overburden parameters.....	76
8.5	Results for the compaction models using an elastic heterogeneous overburden.....	79
8.5.1	Subsidence rates for the calculated volumes.....	83
8.6	Results for the compaction models using a viscous salt	84
8.6.1	Subsidence rates for the calculated volumes.....	90
8.7	Results for the compaction models using the Knothe influence function	91
8.8	Emergency stop scenarios.....	97
8.9	Discussion of the results.....	102
8.9.1	The impact of influence functions.....	102
8.9.2	The impact of compaction models on the results.....	103
8.9.3	The impact of reservoir scenario on the results.....	106
8.9.4	Discussing the χ^2/N values.....	107
9	Implications of the LTS II results on subsidence forecasts for the “Wadden fields”	117
10	Conclusions.....	119
11	References.....	122

1 Summary

The long term subsidence study part two (LTS II) is a follow-up of part one that was finished in 2015 (NAM, 2015a). The main objective of the LTS II project is implementation of the results obtained in LTS I into a stochastic workflow and its application on the Ameland field. If successful, this workflow could provide a new technical standard that may be used for future subsidence models and forecasts to be implemented in the “measurement and control cycles” as part of the “hand-on-tap-procedure”. A unique feature of this newly developed stochastic workflow is that it considers the uncertainty of both data and multiple subsurface models in the forecasting of subsidence resulting from gas production.

Results of the LTS II study demonstrate that a successful workflow was created that is able to confront model results with measurements in an objective way. The workflow incorporates the findings of the LTS I study work and is able to identify the most likely model factors like reservoir scenarios, compaction models and influence functions. Moreover, posterior (after confrontation with the data) probability distributions for the input data show redefined distributions for the input values of the data.

Multiple tests in NAM and TNO ensured the consistent usage of the new software code (ESIP) and proved the capability to produce meaningful results. A special synthetic “dummy” model was created to better understand the findings.

Several significant innovations were applied in the use of geodetic data. Observations from levelling and GPS techniques have been used in the format of spatio-temporal double differences. Uncertainties are described by a fully populated covariance matrix, that also takes shallow movements into account. Outlier removal has been implemented in a formal, conservative way. All these measures help getting the modelling closer to reality by reducing previously adopted assumptions and idealizations.

To describe the uncertainty in possible pressure scenarios for Ameland, 193 reservoir scenarios were created. Thirteen of these scenarios were selected with the aim to be representative for the full bandwidth of the possible 193 scenarios. It is assumed in the confrontation with the data that the prior probability is the same for all scenarios.

On top of the pressure variation we also varied compaction models and influence functions. Each of these factors has its own parameter uncertainty distribution of which the values are probed by Monte Carlo simulation.

Each parameter pick by the Monte Carlo procedure results in a geomechanical model member with a group of members defining the ensemble. Each member is compared with the data where the goodness of fit defines the probability and weight of the specific member. In LTS II we choose for presentation purposes to mimic the “Waddenzee” measurement and control cycle (e.g. NAM 2016a) by calculating the 6 year averages of the subsidence rate for a given area. Rather than deterministic scenarios we are now able to provide the expected weighted average case including confidence bounds for the ensemble to express the uncertainty.

The same workflow has been applied on “emergency stop scenarios”. This showed a clear reduction in the subsidence rates after a hypothetical stop in 1996 and a less clear reduction following a hypothetical stop in 2016. The latter can be explained by the small response of the pressures as the

gas field has produced most of the gas at this time. We therefore conclude that an emergency stop scenario in the Waddenzee for the “younger” fields like Nes and Moddergat would result in a significant decrease of the subsidence rate in the Waddenzee.

As described in the work plan (NAM, 2016) we selected TNO’s Red Flag methodology (Nepveu et al., 2010) to confront model results with data. This methodology is transparent because it will not change the input parameters that defined the prior ensemble. In hindsight this choice was correct because it provided insight in the most likely reservoir scenarios, compaction models and influence functions. Most likely reservoir scenarios are the scenarios that show no or low depletion in the aquifers. The most likely compaction model is the Time decay and RTCiM model. At present the Knothe influence function describes best the data, however it should be noted that the AEsims simulator that is used for the elastic and salt influence does not allow for combinations between salt and elastic layers having large contrasts in the values of the stiffness parameters. When this is improved on, we expect this combination to give lower values of the test statistic because it should match the spatial-temporal behavior of the subsidence better. The Knothe function lacks a temporal component.

The ability of the workflow to differentiate the various scenario-model combinations appears to be low for the actual Ameland model. However, on a synthetic model the process is capable to differentiate the original input model. Further discussion and investigation is required to understand why the process is not capable of differentiating between the real scenario-model combinations before this can be implemented in the official measurement and control cycle.

Reflecting on the objectives as defined in the workplan we conclude the following:

1. A workflow that can test the probability of possible subsidence models in an objective way, with the possible hypotheses resulting from the LTS I study taken into account

We have demonstrated a workflow capable of testing the probability of subsidence models incorporating the possible factors of influence as defined by LTS I.

2. To provide an ensemble of likely subsidence forecasts, each member with a different effect on the future subsidence above the Ameland field. A P50 should be presented from the total distribution.

An ensemble of results has been presented in this report. All statistical values can be derived from the probability distribution of the ensemble like confidence bounds, expectation case and P50 values.

3. An analysis comparing the new distribution of subsidence forecasts with the Ameland forecasts selected in the M&R cycle report over 2015 along with a description of the discrepancies and the possible consequences this may have for M&R 2016 cycle.

Selected geomechanical members based on the salt influence function were compared to the Geomec M&R 2015 results and we concluded that a good match between these models is found. However, the results obtained so far of the stochastic expectation case are very different from the results that follow from the deterministic expected case as presented in NAM (2016a). We have confirmed in LTS II that the reservoir scenario and compaction model used in the M&R cycle 2015

are the most likely reservoir scenario and compaction model. We therefore conclude that expected case for Ameland will remain the same for M&R 2016.

4. The impact of this result on the other gas fields that are part of the M&R cycle will be made clear in a qualitative sense and based on analogies. (Within the given LTS II time schedule it will be not possible to set up a similar project for the remaining fields.)

There are clear differences in the structural setting of the Wadden Fields like Nes and Moddergat. The Wadden Fields are bounded by large faults that prevent a connection to large lateral aquifers. Therefore, we expect less uncertainty arising from the possible reservoir scenarios for these fields. However, the constraint of possible scenarios by the data is poorer because of the lifetime of these fields. The last observation probably results in a larger uncertainty of the subsidence forecast.

Therefore, at present it is difficult to make a strong conclusion on the impact of the LTS II results on subsidence prediction in the Waddensee arising from the Wadden Fields. We firstly need to decide on certain conditions e.g. do we use all compaction models for the Waddensee or only one or two; certain methods for the confrontation e.g. Red Flag versus Ensemble Smoother and certain definitions for e.g. the test statistic.

5. NAM will start with the application of the LTS II workflow to the other Wadden Fields. The first results will be expected to appear in M&R report over the year 2017.

This will be executed under the premise that the LTS II workflow is accepted by the stakeholders.

6. An analysis of the effect of a hypothetical stop of production of the Ameland gas field at specific point(s) in time, including a statement on the analogy with the other Wadden Fields and the effectiveness of the “hand on the tap” principle.

The hand on the tap principle has been demonstrated for different ensembles. A production stop in the “Wadden Fields” would have an even more profound effect because these fields connect to a lesser extend to lateral aquifers.

7. A report that clarifies the effect of an overlying salt layer on the subsidence with a comparison between the results from the analytical AEsubs geomechanical model and the results from the Ameland Geomec model.

Agreement between the models was found. A chapter is included in this document that covers this comparison.

2 Introduction

This report describes the results of the Long Term Subsidence Study part 2, which is a follow up study of part 1 (NAM, 2015a). Both studies are aimed to better understand the observed temporal behavior of subsidence measurements above the Rotliegend gas reservoirs in the northern part of the Netherlands. In particular, the measurements above the Ameland field show a continuing surface subsidence even after the rate of reservoir pressure depletion has slowed down. This behaviour was not well understood and led to the introduction of a time dependent function between pressure change in the Rotliegend gas reservoir and subsidence in order to resolve the mismatch between model predictions and subsequent survey measurements. This allowed for a better fit, but the possible underlying physical mechanisms are not properly understood. This situation was unsatisfactory for NAM, SodM and other stakeholders like “de Waddenvereniging”. Part 1 of the study (NAM, 2015a) focused on the possible mechanisms that could explain such observations. The main conclusions from this study are listed below. In the report a short section is included on the regional geology and on the development history of the Ameland area.

2.1 Summary of part 1 (NAM, 2015a)

[1] The time dependent subsidence effect is real and not an artefact of noise and uncertainty in the geodetic data.

[2] Time dependent creep behaviour is observed and predicted to be associated with compaction of the sandstone in the gas reservoir, pressure diffusion and partial depletion of the aquifers as well as flow of the overlying salt. Salt flow in isolation appears not to be a plausible explanation for time dependent subsidence, while the compaction and pressure depletion models remain viable hypotheses within the possible uncertainty ranges.

[3] Deformation experiments of Rotliegend reservoir core material under in-situ conditions show that reservoir compaction involves a porosity-dependent element of in-elastic deformation through grain cracking and an elastic (reversible) element. The contribution of non-reversible inelastic strain increases with porosity.

[4] The subsidence modelling precision can significantly be improved by taking correlation structures in the surveillance data into account. By appropriately differencing the survey data, biases as well as complexities in covariance structures can be reduced. In addition, methods have been developed for identifying and handling outlier measurements, data reduction techniques for large geodetic data sets, as well as improvements to processing and including GPS data.

[5] An improved and more formal statistical method is proposed to validate and test the quality of subsidence predictions against the survey data. It is based on a Bayesian framework that can provide a coherent structure for the creation of initial models built on prior information, the objective updating of these models using collected geodetic data and the quantitative testing of future predictions.

2.2 Objective of the Long Term Subsidence study part two (LTS II)

Point 5 of the main conclusions recommends to formally test the quality of the subsidence prediction against the measurements using a Bayesian framework, while conclusion 2 clearly states that the observed diffusional subsidence behavior is probably the result of a mix of mechanisms and subsurface processes. Therefore, all of the mentioned mechanisms like salt creep, pressure diffusion, aquifer depletion and compaction creep should be incorporated in such a method.

This recommendation was in essence taken over by the regulator that defined a number of conditions for a follow up study. Based on the discussions a work proposal was written by NAM (NAM, 2016).

The main objectives of this proposal are:

1. A workflow that can test the probability of possible subsidence models in an objective way, with the possible hypotheses resulting from the LTS I study taken into account
2. To provide an ensemble of likely subsidence forecasts, each member with a different effect on the future subsidence above the Ameland field. A P50 should be presented from the total distribution.
3. An analysis comparing the new distribution of subsidence forecasts with the Ameland forecasts selected in the M&R cycle report over 2015 along with a description of the discrepancies and the possible consequences this may have for M&R 2016 cycle.
4. The impact of this result on the other gas fields that are part of the M&R cycle will be made clear in a qualitative sense and based on analogies. (Within the given LTS II time schedule it will be not possible to set up a similar project for the remaining fields.)
5. NAM will start with the application of the LTS II workflow to the other Wadden Fields. The first results will be expected to appear in M&R report over the year 2017.
6. An analysis of the effect of a hypothetical stop of production of the Ameland gas field at specific point(s) in time, including a statement on the analogy with the other Wadden Fields and the effectiveness of the “hand on the tap” principle.
7. A report that clarifies the effect of an overlying salt layer on the subsidence with a comparison between the results from the analytical AEsubs geomechanical model and the results from the Ameland Geomec model.

In order to meet these objectives a research consortium was set up with the organizational structure as pointed out in Figure 1.

LTS II project

1. Subsurface models, parameters and uncertainty ranges (NAM)
2. Geodetic data and uncertainty (TU Delft +NAM)

Confrontation of models with data. Work flow and testing on dummy model by TNO. Application to Ameland dataset by NAM

Project management by NAM

Technical steering committee

- TNO AGE
- SodM + technical advisor

Stakeholders:

- Com. Waddengas 2006

Com. Monitoring Bodemdaling Ameland

- Audit com. (MER)

Process management by NAM

Figure 1: Organizational structure of the LTS II project.

3 Regional Geology

The Ameland gas field area consists of a number of separate but adjacent early Permian Rotliegend fault blocks located in the area straddling the offshore M09 block, the western N07 block, and the Noord Friesland concession. These are part of the Ameland Bank which forms a wedge-shaped terrace on the north-eastern, down-thrown side of the Hantum Fault Zone. It opens towards the north and narrows towards the south with its apex in the vicinity of Kollum in North Friesland. The Ameland Bank structure is bounded to the east by the Lauwerszee Trough graben, and to the northwest by the Terschelling Basin (Figure 2).

The Rotliegend sequence is dissected by a complex fault pattern and generally dips towards the south, resulting in an overall rise in structural elevation to the north. The structural configuration is dominated by two sub-perpendicular intersecting fault trends: E-W and NNW-SSE. The E-W fault trend generally exhibits greater throws and more lateral continuity than the NNW-SSE trend, and therefore controls the distribution and shape of accumulations in the area. The result is a series of narrow, east-west oriented, en-echelon structures.

The major phase of structuration occurred during the mid-Jurassic rift and involved re-activation of many pre-Jurassic extensional faults. The re-activation of Carboniferous faults is, at least partially, responsible for Zechstein halokinesis.

The Upper Rotliegend Group was deposited on the southern margin of the Southern Permian Basin, where clastics from the Variscan mountains were deposited as discrete alluvial fans. These fans grade northwards into sands and silty shales and ultimately into red shales and evaporites, laid down in a sabkha and desert lake environment (Figure 2). The prevailing easterly to north-easterly winds moved sands from the alluvial fans and distributed them in a wide belt of dunes along the southern basin margin. These dune sands form the main reservoir for the Rotliegend Play of the Northeast Netherlands.

During deposition of the Rotliegend, the Ameland Block was located at the southern fringes of the Southern Permian Basin desert lake. The sediments point to a complex interplay of fluvial (wadi), aeolian, sabkha and lacustrine depositional processes under arid conditions. A sequence of alternating claystones and sandstones was deposited in mudflat to damp sandflat and aeolian depositional environments (Figure 3).

The Southern Permian Basin continued to subside throughout the Permian. By Late Permian times, the basin was periodically flooded by seawater and followed by intense evaporation. Consequently, the Zechstein sequence was deposited, comprising at least four depositional cycles, each showing the effects of increasing salinity with time. Cycles began with open marine conditions and the deposition of carbonates and shales. As evaporation continued and basin restriction increased, gypsum (now anhydrite), followed by thick sequences of halite were deposited in the center of the basin before incursion of marine conditions occurred, initiating the start of a new cycle. Carbonate platform and slope deposits could develop along the margins of the basin.

Continental environments prevailed again during the Triassic. Therefore, these Zechstein cycles were followed by deposition of clastics of the Lower Germanic Trias, comprised of alternating claystones and sandstones deposited in terrestrial, aeolian and lacustrine environments. Upper Triassic strata were deposited in shallow marine, lacustrine and evaporitic environments. During the Late Triassic, rim synclines developed which were filled with sandstones, shales and evaporites deposited in an aeolian/fluviatile/marine depositional environment. The development of Triassic rim-synclines and

the presence of collapse structures point to halokinesis of the Zechstein salt during this time. Due to erosion related to the Mid-Kimmerian tectonic event, the Triassic sequence is only preserved in depressions between salt ridges and domes.

Following the major erosional episode recorded by the Mid Jurassic Unconformity, deposition started again in the Late Jurassic. The area was flooded during the Early Cretaceous resulting in deposition of the marine Vlieland Shale and Holland Marl of the Rijnland Group.

The Rijnland Group is overlain by a thick package of open deep marine carbonates of the Chalk Group. Late Cretaceous inversion led to reactivation of the Zechstein salt diapirs, minor faulting and local erosion of the Chalk. The deposition of the fine-grained marine clastics of the North Sea Group along the Ternaard Salt Ridge was dominated by salt movement and inversion. This led to great thickness variations, especially in the Lower and Middle North Sea Group. See Figure 4 for a cross-section indicating structural variation in the overburden of the Ameland gas field.

Within the Ameland area, the Upper Rotliegend Group can be divided into the Ten Boer Claystone, Upper Slochteren Sandstone, Ameland Claystone and Lower Slochteren Sandstone (Figure 4). The Upper Slochteren Sandstone can further be subdivided into 6 reservoir units of which unit 4 is the most productive one.

The Ten Boer Claystone is a silty claystone, 90 to 100 meters thick, deposited in a mudflat setting. Towards the south, it contains locally some porous fluvial sand streaks at the base. However, within the Ameland area, no sands of significant thickness are expected, and as such, the Ten Boer Claystone is expected to act as both a lateral seal and as a top seal.

The main reservoirs of the Rotliegend Play are the dune / dry sandflat to damp sandflat sands of the Upper Slochteren Sandstone (ROSLU). The ROSLU, within the Ameland area, has been deposited in a predominantly damp to wet sandflat environment, grading towards the north in a wet sandflat to mudflat depositional environment. The gross thickness of the Upper Slochteren Sandstone is expected to be around 94 meters.

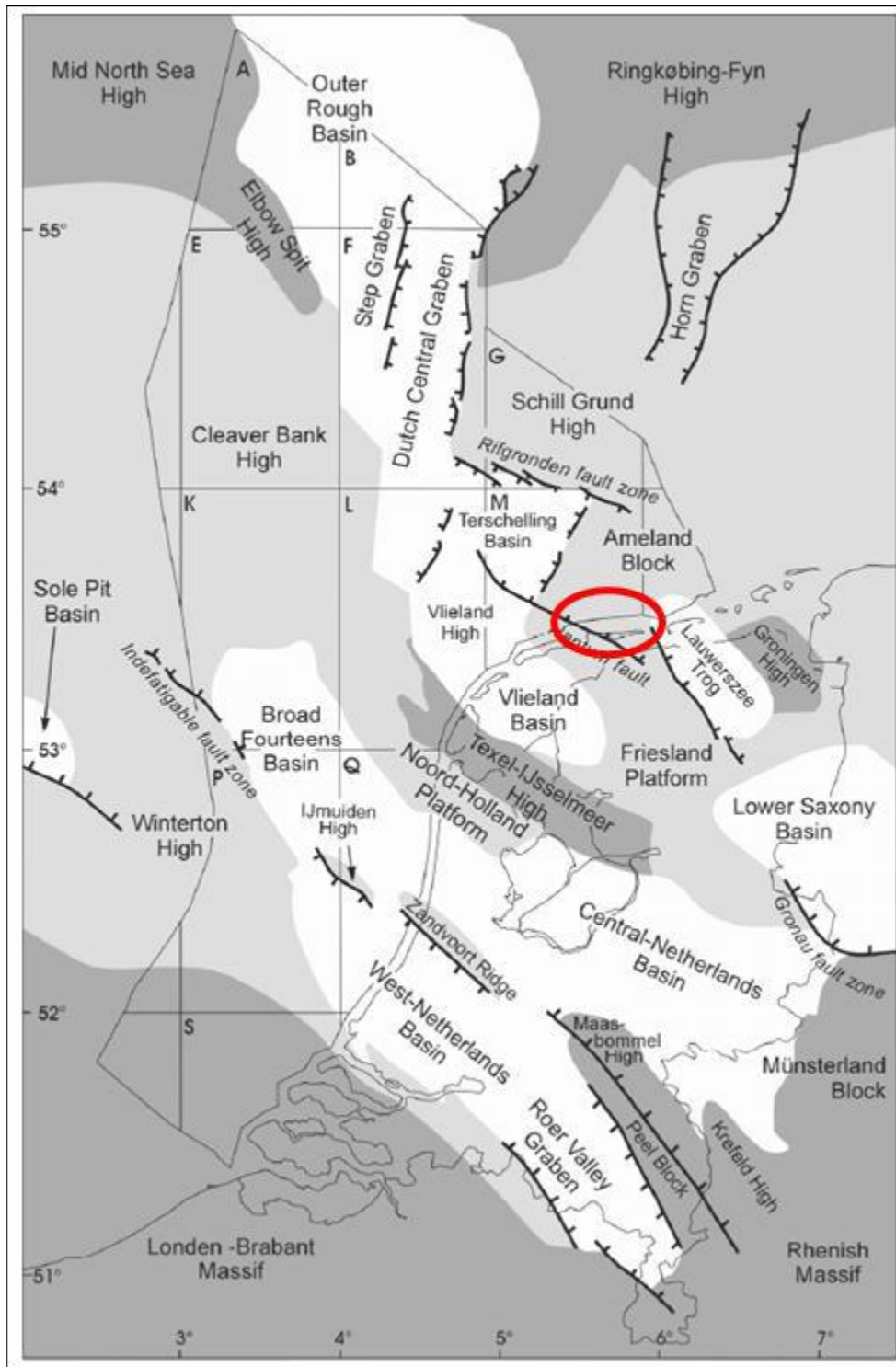


Figure 2: Structural elements in the Dutch subsurface during the Late Jurassic to Early Cretaceous. Shaded in white are basins whereas darker colours indicate progressively higher areas. The red circle indicates the location of the study area (after Van Adrichem Boogaert & Kouwe, 1993).

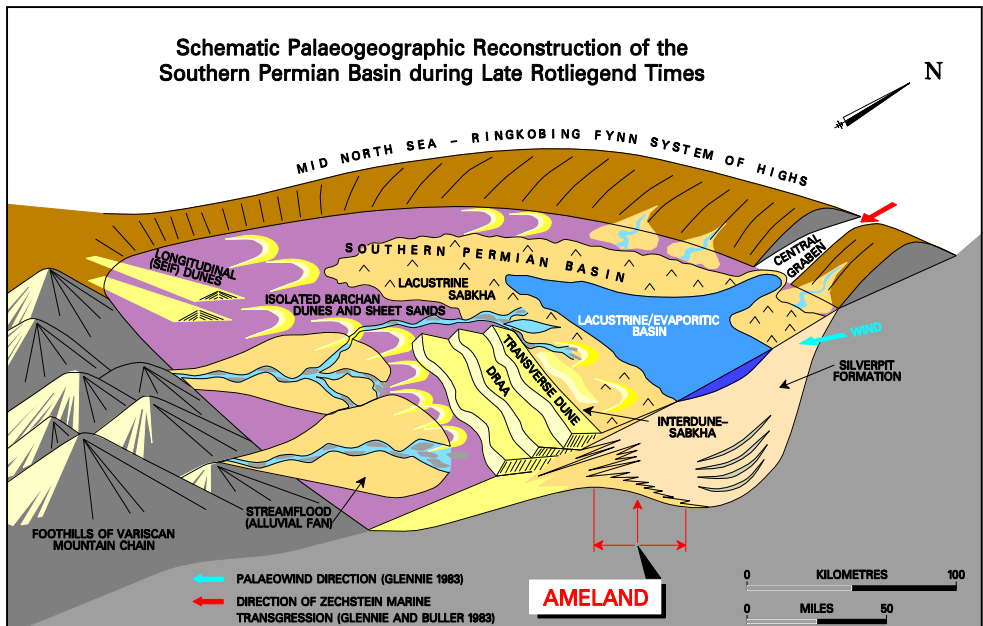


Figure 3: Conceptual model of the sedimentary environment during deposition of the Permian Rotliegend; the red circle indicates the location of the study area in this depositional setting; dashed box indicates transition from Slochteren to Silverpit formation (McKie, 2011).

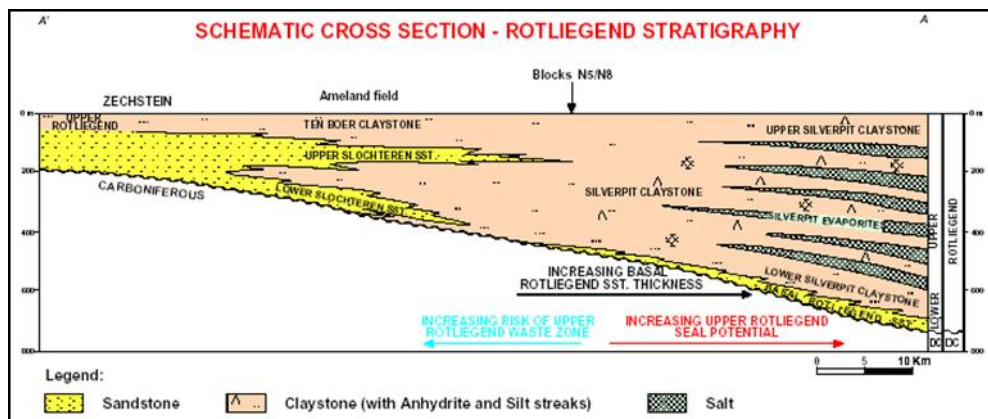


Figure 4: Conceptual cross-section through the Permian Rotliegend indicating the change of reservoir from south to north (McKie, 2011).

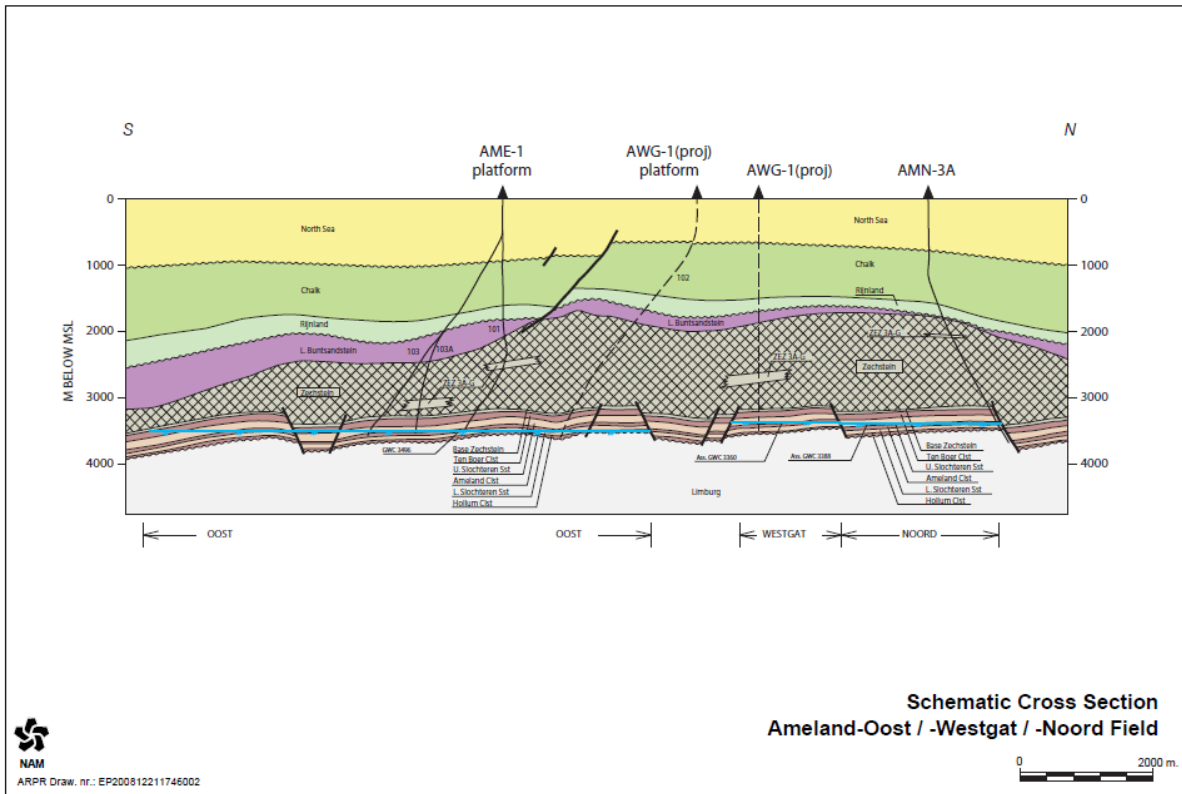


Figure 5: Cross section indicating structural variation in the overburden of the Ameland gas field.

4.1 Ameland Oost

Ameland-Oost is located in the Dutch part of Southern North Sea area. The field was discovered in 1964 by well AME-1 drilled from the Ameland Island. The field contains several intra field faults further sub-dividing the field into a number of producing blocks with limited connectivity across the faults (Figure 6). Production started in 1986 through wells drilled from the locations AME-1 and AWG-1, followed by producing wells drilled from the AME-2 location. Produced gas from AME-1 and AME-2 is delivered to AWG-1 platform by the connecting pipelines, where it is compressed and exported to the main export pipeline (NGT) and routed to the processing facility onshore in the Netherlands (Uithuizermeeden). After an initial stage where the gas flowed naturally from the reservoir through the facilities, a compression station was installed in 1994 to increase the recovery by lowering the AWG (and AME-1/AME-2) operating pressure. After various reconfigurations, currently the minimum pressure at which the wells are operating is around 10 bar.

16 wells were drilled to Ameland-Oost field, 2 of which were abandoned and 7 are currently producing gas. The remaining wells are not producing currently due to number of reasons, specific in each individual case.

4.2 Ameland Westgat

Ameland-Westgat, located north of Ameland-Oost, was first discovered in 1975 by well AWG-1. The field has 3 major faults dividing it into 4 major blocks with a proved connectivity across these faults. The Upper Slochteren is being produced from the AWG-1 platform complex. In 1993, the first development well AWG-107 was drilled in the reservoir block W10. The second development well AWG-108, also targeting block W10, was drilled in 1998 and found reservoir pressure higher than expected indicating the limited connectivity within the same block. It's currently believed that there might be additional faults within W10 block than cannot be seen on seismic. The third development well (AWG-101C) was drilled in block W30 in December 2007. This well showed the block to be depleted by some 185 bar in most of the high-permeability streaks. Due to technical issues while completing the well only Units1 to Unit3 are believed to be perforated.

3 wells were drilled to Ameland-Westgat field, 1 of which is currently not producing (AWG-101C, different options are considered to restore production from block W30) and 2 are currently producing gas: AWG- 107 and AWG-108.

4.3 Ameland-N07FA

Ameland-N07FA Field was discovered by well N07-2 (renamed to N07-FA101) in 1991. It is a Base Zechstein tilted fault block, bounded to the north, east and west by prominent faults, with throws of 100-200 m, and dipping to the south. The structure lies immediately east of the Ameland-Westgat field. The Upper Slochteren reservoir has been developed from the AWG-1 platform complex. In Q3 1997 a mono-pile structure was placed to provide the structure for a mini-satellite platform for development of Ameland-N07FA together with M09-FB. A Field Review has shown that the gas volumes in M09-FB are too small to warrant a development of the block. This has resulted in a different development concept for Ameland-N07FA, now with an extended reach well from the AWG platform (AWG-110). Note that the AWG-110 well found the Ameland-N07FA reservoir depleted by some 60 bar compared to the N07-2 discovery well. As the only 2 producing fields in the vicinity are

Ameland-Oost and Ameland-Westgat, it is likely that the Ameland-N07FA structure is connected to either one of them. There is 1 production well drilled into Ameland-N07FA field: AWG-110.

Accumulations M09-FA, M09-FB, Nes-Noord and Ameland-Noord have not been developed yet. Based on the Ameland-N07FA development well (AWG-110) results, minor depletion of M09-FA accumulation cannot be excluded, while Nes-Noord, Ameland-Noord and M09-FB fields are believed to be at virgin pressure.

5 Dynamic modeling and evaluation

In order to assess future production and development options, a 3D simulation model has been built for the Ameland area, modeling both the flow of gas and water as well as the adjacent aquifers. Also the nearby, as yet undrained, accumulations are included in this model to assess the remaining development opportunities. Based on the observations made in Ameland so far, depletion of M09-FA cannot be excluded and hence this has been assumed to be slowly depleting. In contrast M09-FB, Ameland Noord and NSN have been modeled as remaining at virgin pressure as given the current understanding, it is highly unlikely that these accumulations are depleted.

Based on production data and pressure observations this model has been matched to a best fit for future production predictions. The same model is used every year for the M&R cycle updates. For the subsidence evaluation, the remaining uncertainty in the pressures in the producing areas is very low, due to the direct measurements that are available. However, especially on the edges of the field and in the adjacent aquifers where little data is available, large uncertainties with respect to the pressure in these areas remain, which does influence subsidence calculations and evaluations. Hence for this exercise the simulation model approach has been adapted to generate multiple scenarios that provide a match to the historic production and observed pressures, but do generate different pressure predictions for those areas where no direct pressure measurements have been collected.

In order to generate these scenarios, the main parameters that influence the pressure of the Rotliegendes reservoir away from the existing drainage points have been selected and varied using an automated history match routine that is normally used to find the best fit. These parameters are:

- Total GIP of the Ameland field (GIIP)
- Permeability in gas leg of the reservoir (K_gas)
- Permeability in water leg of the reservoir (K_aqf)
- Residual gas in aquifer (Sgr)
- End-point relative water permeability (Kwrg)
- Fault transmissibility to the East of Ameland-Oost (Fault_81, Figure 9)
- Fault transmissibility to the South of Ameland-Oost (Fault_64 and Fault_105, Figure 9)

	GIIP		K_Gas		K_Aqf		Sgr		Kwrg		Fault_105		Fault_81		Fault_64	
	min	max	min	max	min	max	min	max	min	max	min	max	min	max	min	max
Very wide	0.95	1.05	0.01	100	0.01	100	0.5	1.5	0.5	1.5	0	1	0	1	-	-
Wide	0.96	1.04	0.04	25	0.04	25	0.6	1.4	0.6	1.4	0	1	0	1	-	-
Mid	0.97	1.03	0.1	10	0.1	10	0.7	1.3	0.7	1.3	0	1	0	1	-	-
Narrow	0.98	1.02	0.2	5	0.2	5	0.8	1.2	0.8	1.2	0	1	0	1	-	-
Selective	0.98	1.02	0.2	10	0.001	10	0	2	0.8	1.2	0	1	0	1	0	1

Figure 8: Ranges in multiplication factor for the various parameters that have been studied.

Within these parameter ranges, visualized in Figure 8, individual estimates for each parameter were chosen and the model was run using these parameters, both over the history of the Ameland field as well making a prediction into the future. By comparing the differences between the simulated historical pressures and the actual measured pressures, one can obtain a numerical indicator how close the model represents the real measurements. This parameter is defined as sum of all the

squared differences between observed and measured pressured and expressed as the RMS (Root Mean Squared) value.

As GIIP and permeability were only varied over the total field, one can see that in some cases the residual error as calculated by the program would very quickly increase suggesting a very poor match or even non convergence of the model. Hence based on the initial results some fine tuning of the parameter range has been done leading to the range in the lowest row of the table above.

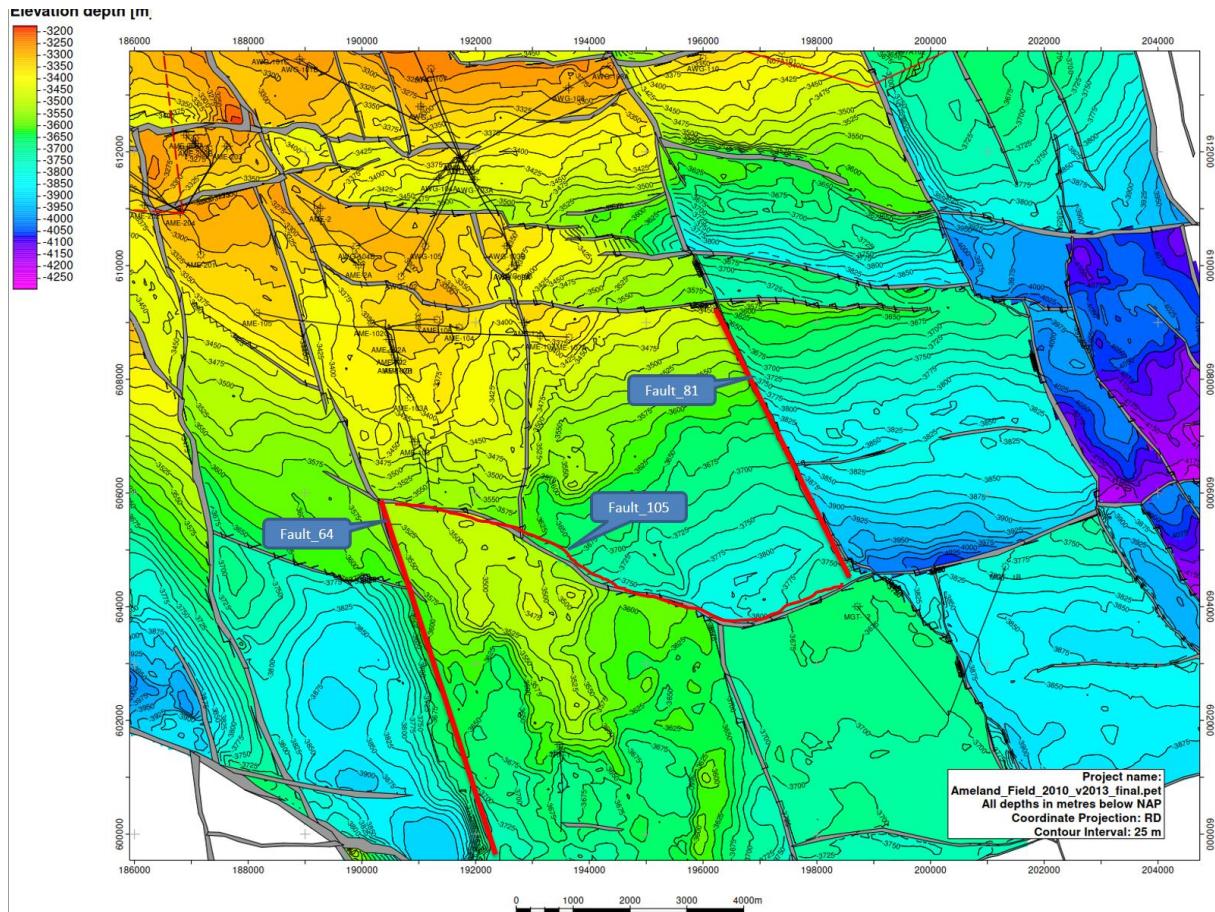


Figure 9: Top of structure map of the Ameland area with faults 64, 81 and 105 indicated by the red lines.

However even within that range, some models would lead to a poorer match. The RMS values of all the 193 reservoir models is shown in Figure 10, from the better models (RMS<45) a selection of 13 models has been chosen for the ESIP calculation (Figure 10).

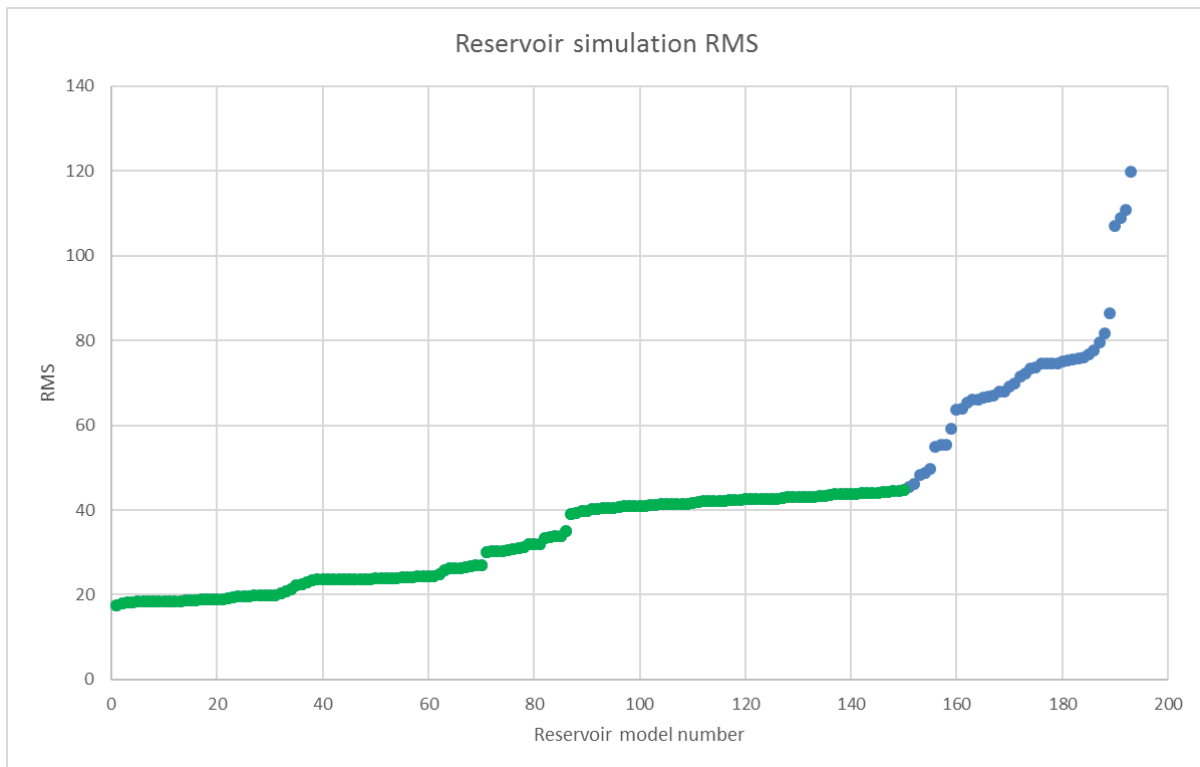


Figure 10: RMS for each Mores model. The selection for the ESIP calculation has been made from the reservoir models with RMS < 45 (green dots).

It should be noted that a number of further variables (local variation of GIIP multipliers, intra field faulting intra field permeability variation) exist, that ensure that in these scenarios the model could be history matched without creating large changes to the pressures levels away from the center area. Rather than increasing the number of parameters for the routine to match these individual cases, in such cases the poorer history match was accepted as a viable scenario for predicting the overall pressure in the region. Based on the fact that even the closest model shows an RMS of 20 and provides a reasonable history, there is quite some confidence that models with an RMS up to 60 can be matched and even some models with higher RMS's probably still can be matched quite reasonably. The approach has been chosen to just assume that models could be matched up front and only if subsidence analysis would point to specific models that have a higher RMS, to spend time investigating these cases and see how the match quality could improve without affecting the pressures away from the field.

The data generation of such an exercise is rather large and hence the real simulated models were not stored to avoid running out of space. To aid analysis of the consequences of varying the above parameters on the pressure distributions, 'dummy wells' were defined in those areas of the model that do not contain any real wells (Figure 11).

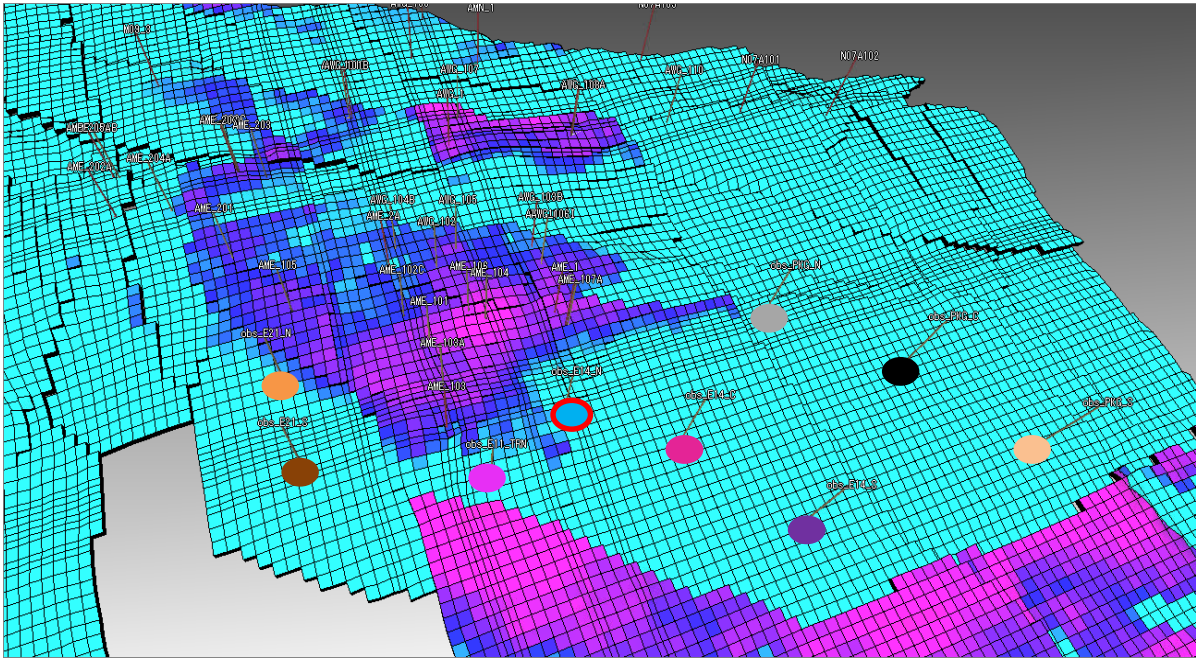


Figure 11: Overview of 'dummy' wells introduced to observe aquifer response.

These dummy wells do not interfere with the flow pattern of the simulation, but do generate a table that shows the pressure that would be observed in these wells if they were real. Hence they provide a good insight how the pressure in those remote areas varies over time as we change parameters.

5.1.1 Results assessment

For all of the scenarios, pressure grids were generated and provided to the geomechanics team as input to the subsidence routines. At a later stage to test the validity of the so called emergency stop, pressure scenarios have been generated assuming the Ameland production would have stopped in 1996 and a second set where no more production after 2016 is assumed.

Assessing the results, it is clear that the quality of the pressure match on an observation point by point basis in the Ameland wells is predominantly governed by the gas permeability multiplier and the GIIP (which is why the GIIP range had to be narrowed down). This is exactly as one would expect. In contrast, the pressure in the 'pseudo' aquifer observation wells over history is mostly influenced by the following parameters

- a) Aquifer permeability: the larger the aquifer permeability, the more depletion is observed
- b) The faults separating the aquifer from the gas bearing blocks
- c) Whether or not residual gas is present. The presence of residual gas has 2 effects;
 - a. Due to the presence it will influence the relative permeability of the water, reducing the ability of the aquifer to flow towards the depleted gas bearing sands
 - b. The gas present in the aquifer will have a strong dampening effect on the pressure response, maintaining high pressures in the area.

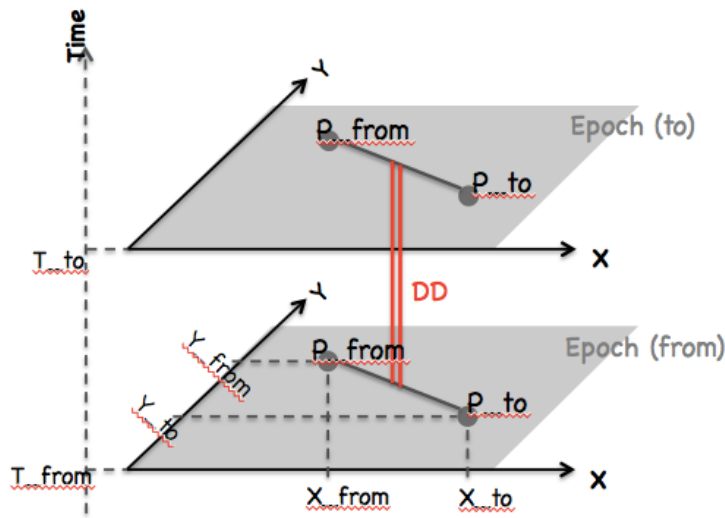
Hence overall one could say that a scenario with a low aquifer permeability, sealing faults or with significant residual gas will ensure high pressures in today's aquifer, and in order to generate depletion of the aquifer 'today', one would need a scenario with high aquifer permeability, no residual gas and non-sealing faults.

6 Use of geodetic data

Levelling and GPS (Global Positioning System) data have been prepared for geomechanical modelling following the new approach that has been proposed in LTS I project. Data processing has been carried out by Delft University of Technology (van Leijen et al., 2017) in close cooperation with NAM.

6.1 Innovations as recommended in LTS phase 1 (2013-2015)

1. The observations from different survey techniques are not combined prior to geomechanical modelling in a single dataset. Handling separate datasets for levelling and GPS avoids smoothness assumptions that would be required for their alignment.
2. Spatio-temporal double differences are used instead of temporal single differences. The confrontation of measurements and geomechanical model predictions takes place at the level of changes of surface positions between two points in space and two epochs in time (see Figure 12). Thus, no assumptions on a stable reference point are necessary. Note that the original recommendation from LTS-1, which was more restrictive, has been revised (Samiei-Esfahany and Bähr, 2017).



$$DD_OBS = \underbrace{(H_{P_{to}}^{T_{to}} - H_{P_{from}}^{T_{to}})}_{\Delta H^{T_{to}}} - \underbrace{(H_{P_{to}}^{T_{from}} - H_{P_{from}}^{T_{from}})}_{\Delta H^{T_{from}}}$$

H_p^T is the height of benchmark P at epoch T

Figure 12: Concept of double differences (TU Delft, 2016).

3. The stochastic properties of the geodetic observations are described by their full covariance matrix. Compared to previous approaches this yields a more realistic assessment of uncertainties.
4. Individual outliers are identified by a more formal and objective approach. As the assumption of an a priori geomechanical model is unavoidable for this purpose, the threshold is chosen in a way that only the most obvious outliers are rejected. Thus, subjective expert assessment is reduced to the necessary minimum. "Tuning" the measurements towards a previous model prediction is largely avoided.

- The discrepancy between measurements and geomechanical model predictions originates not only from uncertainty of geodetic observations (measurement noise) and model errors but also from shallow movements. These movements, like building settlements, cannot be attributed to measurement imperfection and are not considered in the geomechanical model either. Thus, they are subsumed by the so-called idealisation noise, which is now taken into account.

6.2 Survey data and measurement noise

Figure 13 provides an overview of the datasets used in the study.

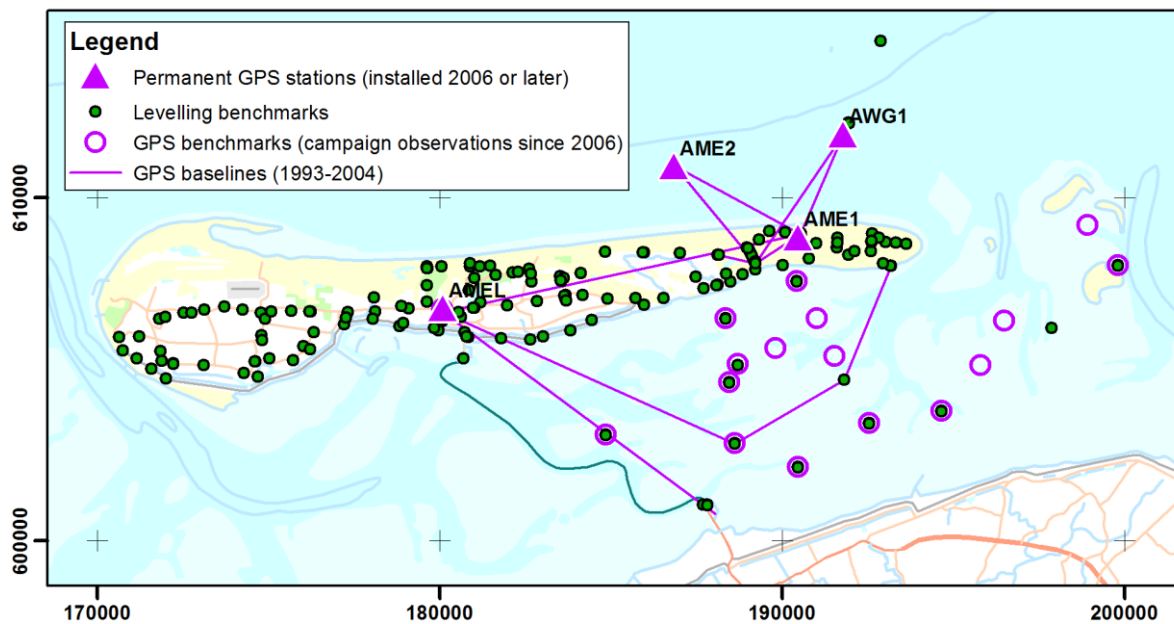


Figure 13: Spatial distribution of geodetic observations used in this study.

6.2.1 Levelling

Levelling data from 436 benchmarks observed in 26 campaigns between 1986 and 2014 have been used in this study. For the first time, optical levelling data have been complemented by hydrostatic levelling observations. This technique of measuring height differences by means of a water-filled tube has been applied by Rijkswaterstaat in the study area until 2002. Thus, also benchmarks in the Waddenzee can be connected to the levelling network on the island. To quantify the measurement uncertainty, standard models have been used for both optical and hydrostatic levelling.

6.2.2 GPS network (since 2006)

The more recent GPS observations originate from a network of permanent monitoring stations established in 2006. Four of them have been used in LTS II (see Table 1). The network is extended campaign-wise by the benchmarks in the Waddenzee. Benchmarks on the platforms AWG1 and AME2 have been observed in several campaigns before installing continuously operating stations in 2014. From the campaign data as well as the stations on the two platforms, only the vertical component is used. The observed horizontal displacement is considered unreliable for these points due to the survey setup.

Table 1: Continuous GPS stations used in this study.

Name	Location	Installation	Used components
AME1	Production location on Ameland	2006	x, y and z
AME2	Production platform	2014	only z
AMEL	Nes village	2014	x, y and z
AWG1	Production platform	2014	only z

The data from the four continuously operating stations have been sampled yearly and corrected for temperature effects, atmospheric loading as well as annual and semi-annual harmonics. The latter correction compensates for systematic errors that are inherent to the GPS system. The measurement uncertainty is modelled by the stochastic model proposed in LTS I (Williams 2015), which comprises the uncertainty of determining the antenna position and a setup error. This setup error of 1.5 mm is applied to the campaign data only and accounts for the yearly mounting of the equipment.

The previously mentioned concept of spatio-temporal differencing has been applied to the GPS data in a modified way. NAM obtains the processed data in form of displacements with respect to an assumedly stable reference network of currently more than 10 stations. Their stability is closely monitored in yearly intervals. The spatial differences are formed with respect to a virtual reference station representing the whole network of reference stations. For reliability purposes, the monitored benchmarks in the Waddenzee consist of clusters of three. Observations from all three benchmarks of a cluster have been used to maximise the information content for the modellers.

6.2.3 GPS baselines (1993-2004)

Prior to 2006, only few GPS observations have been carried out by NAM in the study area, mainly to establish a connection from the island to the platforms AWG1 and AME2. Benchmarks have been occupied and processed in pairs (baseline). Only the vertical component is available from these measurements. Since very little is known about the processing approach, a conservative standard deviation of 11 mm has been assumed for these height differences.

6.2.4 InSAR

In spite of the availability of InSAR (Interferometric Synthetic Aperture Radar) data from Ameland, these have not been used in the study for multiple reasons. The spatial coverage in the eastern part of the island, where the gas field is situated, is very sparse. In addition, the quality of the available measurements is poor due to lack of stable Radar reflectors (well-founded buildings, rock outcrops, etc.) in that area. This circumstance together with pending imperfections in stochastic modelling would require a very conservative approach for uncertainty modelling. After all, there would have been little (if any) added value to this project while the effort of data preparation would have been disproportionately large.

6.3 Idealisation noise

This noise component describes displacements of the measurement points due to shallow movements like building settlement or soil compaction. Together with the measurement noise, it models the uncertainty of geodetic observations when confronted with geomechanical predictions. The model parameters for idealisation noise have been calibrated in an onshore study area in LTS phase 1 (Samiei-Esfahany and Bähr, 2015) and revised recently (Samiei-Esfahany and Bähr, 2017). The model differentiates between two components:

- *Temporal component*: This subsumes all shallow effects that are correlated in time and uncorrelated in space. It can be considered as an autonomous movement of an individual benchmark that has nothing in common with the behaviour of neighbouring benchmarks. An example would be settlement of the individual building a benchmark is attached to.
- *Spatio-temporal component*: This subsumes all shallow effects that are correlated in both space and time. It can be considered as a coherence in movement with neighbouring benchmarks, whereas the level of coherence decreases with distance. An example would be the compaction of a shallow peat layer.

A determining factor for the temporal component is the type of foundation of a benchmark. The autonomous movement of a deeply founded (“underground”) benchmark may be negligible, whereas the settlement of a benchmark attached to a poorly founded building may be not. A discrimination between different types of benchmark is not possible, because a classification would be very subjective and because an extended model estimation would suffer from too few observations in too many classes. Hence, the autonomous movement of all benchmarks is considered to be reasonably explained by a single model. This assumption is extended to GPS permanent stations, which are attached to buildings and production platforms and thus fall into the stability range of levelling benchmarks. The only exception was made for offshore benchmarks in the Waddenzee, which are form class on its own. They all have a practically identical layout (a several metres long pole driven into the ground) and are placed in clusters of three. The levellings between benchmarks within the same cluster constitute an excellent opportunity to estimate model parameters for the temporal component.

Another factor is the type of soil. Peat grounds are more prone to natural compaction than sandy soil, and the associated idealization noise can be expected to be higher. This applies in particular to the spatio-temporal component. Since the study area from LTS I is dominated by peat grounds and the soil on Ameland and below the Waddenzee is dominated by sand, it can be expected that the model “peat based” parameters from (Samiei-Esfahany and Bähr, 2017) would be too pessimistic for LTS II. To account for this, the spatio-temporal component has been assumed zero in all modelling calculations with the Ameland data.

6.4 Outlier handling

Outliers that are detectable by the closed loop property of levelling networks have been excluded from the dataset. So have been outliers that could be identified from relative movements within a cluster of three (GPS) benchmarks in the Waddenzee (van Leijen et al., 2017). The 2006 observations of those Waddenzee benchmarks that have been placed in that year have been excluded as well. They are suspected of significant settlement during the first year after placement. Another exclusion applies to two hydrostatic levelling campaigns (279H08 and 279H09). For these, contradictory indications existed regarding the year of observation (1987 or 1988). It was not possible to resolve this inconsistency.

Apart from the aforementioned cases there are also types of outliers that cannot be identified without assumptions on the geomechanical model. An example is the confusion of a benchmark with a nearby benchmark during surveying, which can yield jumps of several decimeters in the time series. To detect these types of outliers (“abnormal behavior”, “identification errors”, “disturbances”), the approach proposed in LTS phase 1 has been applied to the levelling data: An a priori geomechanical

model prediction is subtracted from time series of height differences between pairs of benchmarks. The resulting residuals are tested in a geodetic framework. A largely relaxed threshold accounts for the uncertainty of the potentially imperfect model assumptions. Thus, 9 observations have been identified as outliers (van Leijen et al., 2017).

7 Workflow confrontation of model results with measured data

7.1 General, description

A probabilistic workflow was developed by TNO (2016) with the objective to confront model results with the measured data. Basically it automates the choice of model aspect plus the accompanying input variables and confronts these results with the measured data. A measure of the fit to the data is obtained by calculating the quality of the match of the model with the data. The quality is expressed by a probability value to each model member. The tool is named **ESIP (Ensemble-based Subsidence Interpretation and Prediction)**. Next to delivering a probabilistic workflow, ESIP's objective is also to provide an objective description of the statistical outcome and to provide steer in reducing the uncertainty by incorporating the geodetic data. ESIP incorporates TNO's semi-analytical simulator AEsups (Fokker and Orlic, 2006). A semi-analytical approach is necessary in this project because it allows for fast computing of results. Typical computing times for the Ameland numerical model, based on a finite element approach, are around 4 hours, which would take more than 4 years before one ensemble is finished. AEsups produces an ensemble within one day for the elastic case and two weeks for a case that includes a viscous salt layer.

Nomenclature

The ESIP procedure uses a number of words that deserves more explanation and definition. A description of these can be found below:

- Compaction model: Four compaction models have been implemented in the ESIP tool: RTCiM, Time decay, Bi-Linear and Linear
- Reservoir scenario: pressure input that results from the Mores simulator. The Mores simulator is not part of the ESIP module but ESIP handles the pressure output from Mores
- Member: a model member represents a single run of the geomechanical model with certain choices for the influence function, compaction model and reservoir scenario
- Scenario-model combination: A specific combination of a reservoir scenario, compaction model and influence function
- Ensemble: a collection of members that result from a Monte Carlo analysis on a specific scenario-model combination.
- Influence function: function which translates the reservoir compaction to surface subsidence. The influence functions used are Greens functions applied in AEsups and Knothe.

ESIP requires the following input:

1. Geodetic data.
2. An ensemble of reservoir flow model simulation results (pressure *scenarios* through time) and a prior ensemble of geomechanical realisations or members being built on top of the pressure scenarios
3. Input of geomechanical parameters for the different compaction models.

From the reservoir scenarios, ESIP calculates compaction and subsidence using the TNO geomechanical simulator AEsups. The geodetic data are processed into double difference data and a corresponding covariance matrix. The modelled ensemble is confronted with the data. A schematic representation of the ESIP workflow steps is provided in Figure 14.

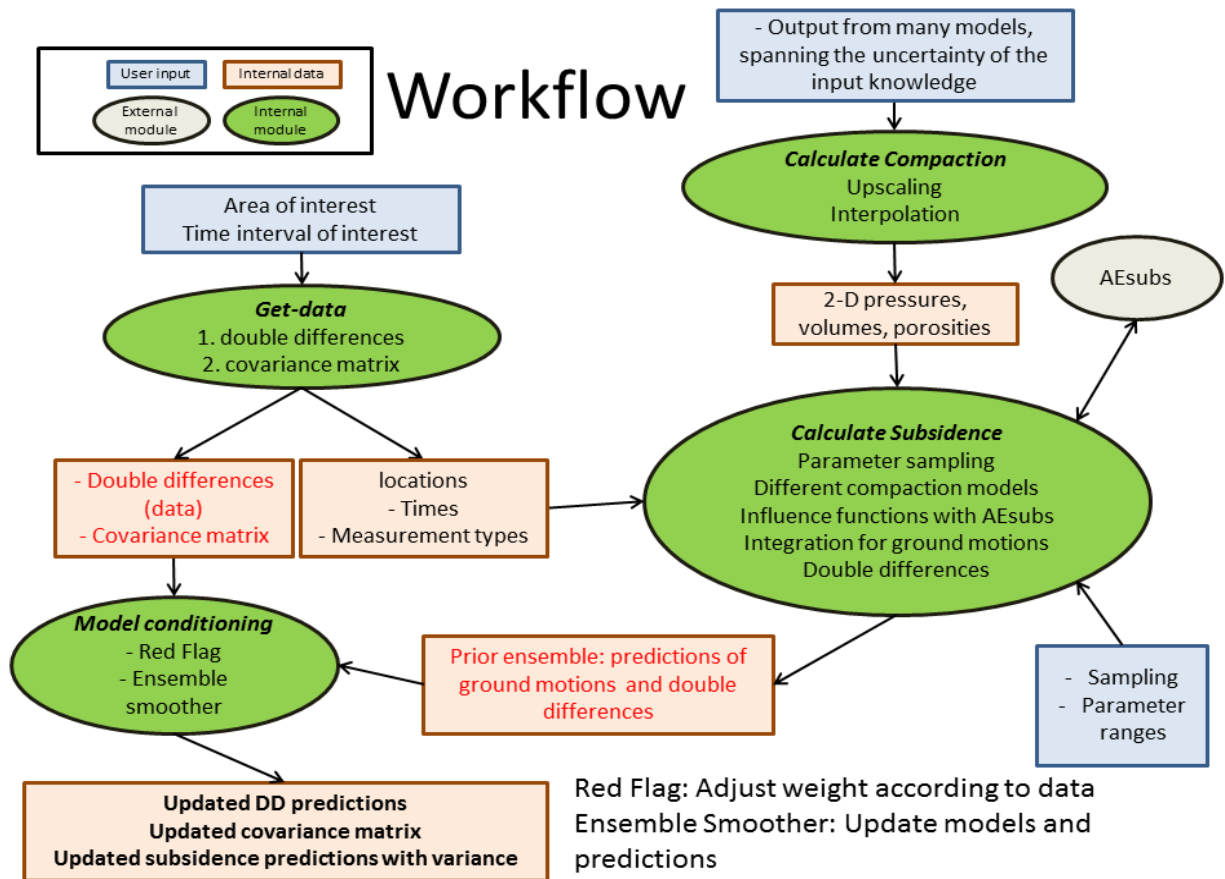


Figure 14: Schematic representation of the ESIP workflow (TNO, 2016).

The ESIP workflow contains 4 internal modules that form the core of the program.

7.2 CUPiDO

The module CUPiDO (Connecting Undifferenced Points in Deformation Observations; working title “getdata”) has been created by TU Delft as an interface between geodetic data and the ESIP workflow. It selects observations for a specified area and time period of interest. The output is a non-redundant set of double differenced displacements (see Figure 12) with an appropriate stochastic description in form of a fully populated covariance matrix (see Figure 15). Figure 16 shows the standard deviations of all double differences as a function of distance and time.

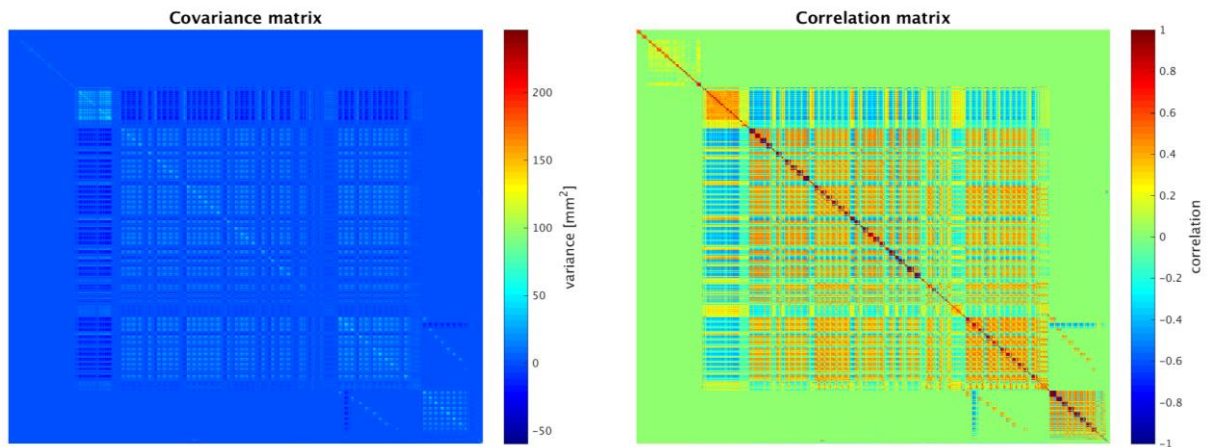


Figure 15: Covariance matrix (left) of the double difference observations used in the project. The off-diagonal elements describe the stochastic interdependence. The correlation matrix (right) is a normalized form of the covariance matrix, which better visualizes the correlation structure.

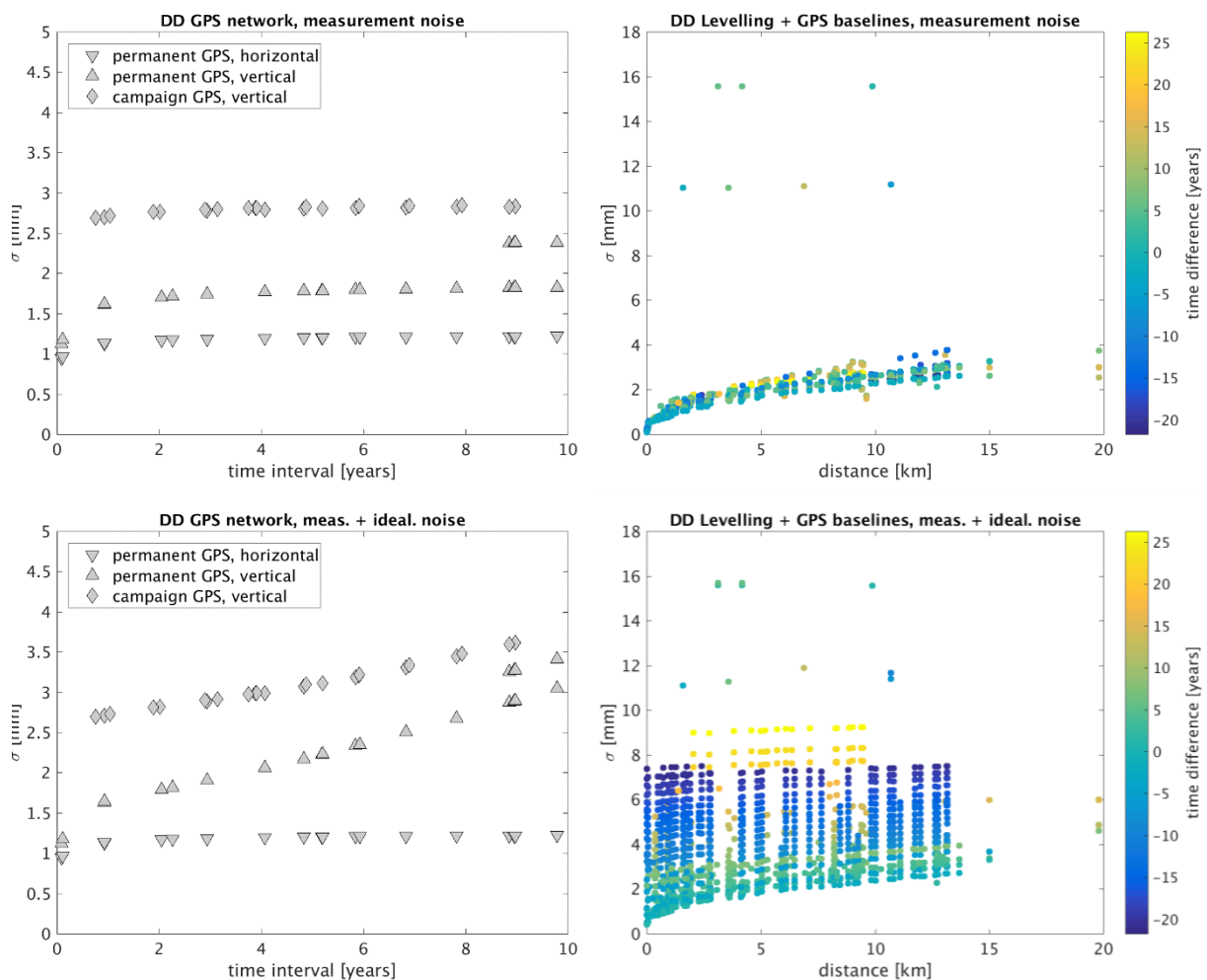


Figure 16: Standard deviations of the individual double difference observations. Whereas this visualization does not convey the full stochastic information, it provides a good impression of the precision of the individual double differences. The upper row shows the contribution of the measurement noise, which is related to the uncertainty of the measurement itself. The lower row displays the full uncertainty budget. This also comprises the idealisation noise, which accounts for movements of shallow soil layers. The figures on the left hand cover GPS network observations (permanent station and campaigns 2006-2016), and the figures on the right hand side cover levelling (1986-2014) and GPS baseline observations (1993-2004).

7.3 Calculate compaction

The calculation of the compaction involves a number of steps that are summarised in the next paragraphs.

7.3.1 Upscaling

The reservoir model Mores uses a fine pressure grid that needs to be up-scaled in both vertical and horizontal directions before the pressure information can be used as an input to the geomechanical models. The goal of up-scaling is that a fine compaction grid produces the same amount of compaction as the up-scaled grid. The vertical upscaling is performed firstly followed by a horizontal up-scaling over a coarser grid.

The *vertically* averaged pressures P_{av} , volumes V_{tot} , porosities ϕ_{av} and grid block center positions X_{av} and Y_{av} for one vertical column of the 3-D grid can be expressed as:

$$P_{av} = \frac{\sum_{k=1}^n C_{mk}(\phi_k) \cdot P_k \cdot V_k}{C_m(\phi_{av}) \cdot V_{tot}} \quad (1)$$

$$V_{tot} = \sum_{k=1}^n V_k \quad (2)$$

$$\phi_{av} = \sum_{k=1}^n \phi_k \cdot \frac{V_k}{V_{tot}} \quad (3)$$

$$X_{av} = \sum_{k=1}^n X_k \cdot \frac{V_k}{V_{tot}} \quad (4)$$

$$Y_{av} = \sum_{k=1}^n Y_k \cdot \frac{V_k}{V_{tot}} \quad (5)$$

$C_{mk}(\phi_k)$ in equation (1) corresponds to the uni-axial compaction coefficient of one grid block as a function of its porosity, pressure, and time. After the first step of vertical upscaling we have a 2D irregular fine grid for P_{av} , ϕ_{av} , V_{tot} at upscaled positions (X_{av}, Y_{av}) .

The next step involves the horizontal upscaling of the vertically up-scaled grid. ESIP allows for a flexible entry of the desired grid size but typically the size varies between 500 and 1000 m. Basically the horizontal upscaling follows the same equations as used for the vertical up-scaling. The end product is a 2D coarser irregular grid.

7.3.2 Compaction models

ESIP allows for input into four different compaction models that are briefly described in TNO (2016).

7.3.2.1 Linear elasticity

The simplest model that is available is the linear compaction model of which the uni-axial compaction can be described by:

$$V_{comp}(x, y, t) = C_m(x, y) \cdot V(x, y) \cdot dP(x, y, t) \quad (6)$$

The C_m can be entered as a single value or as a function of porosity and will translate values of the C_m into a stiffness value. Paragraph 8.4 shows how the dependency of the C_m and Young's modulus with the porosity is assessed for the Ameland case.

7.3.2.2 Bi-linear elasticity

This is a model that introduces a different C_m value after a critical depletion pressure. It is based on the observations of the temporal subsidence behaviour by NAM where a slow subsidence rate during the first years of production is followed by a faster subsidence rate. This rate can be translated in a $V_{comp_{pre}}$ and $V_{comp_{post}}$, which are defined by the following relationships

$$V_{comp_{pre}}(x, y, t) = C_{m_{pre}}(x, y) \cdot V(x, y) \cdot (P_0(x, y) - P(x, y, t)) \quad (7)$$

$$V_{comp_{post}}(x, y, t) = C_{m_{pre}}(x, y) \cdot V(x, y) \cdot (P_0(x, y) - P_{trans}(x, y)) + C_{m_{post}}(x, y) \cdot V(x, y) \cdot (P_{trans}(x, y) - P(x, y, t)) \quad (8)$$

where P_0 and P_{trans} respectively define the virgin pressure and the transition pressure where the compressibility changes from the lower $C_{m_{pre}}$ to a higher $C_{m_{post}}$ value.

7.3.2.3 Time-decay compaction model

The observation of a delayed, slowly accelerating subsidence at the onset of pressure depletion in combination with continuing subsidence after depletion has ceased, is consistent with a time lag (time decay) process where the subsidence response to reservoir compaction is asymptotic, with a characteristic time decay constant. Processes of this type are fundamental and commonplace throughout the natural world; they are the signature of non-equilibrium dynamical systems. The archetype of processes in this class is the familiar diffusion or heat equation. Time decay type models have been proposed as explanations for subsidence delay in the past. Houtenbos [pers. comm., 2006] proposed a simple empirical time decay relationship between 'subsidence volume' and the mass of gas produced. A number of issues in the physical reasoning led to a rejection of this proposal by NAM and SodM. It was observed at the time though, that transfer functions of this type did appear to provide a satisfactory temporal match to subsidence data and that they are characteristic of a diffusive, and therefore physically reasonable, process. A distantly related time dependent process was contained within the Rate Type Compaction Model [RTCM] (de Waal, 1986), which also sought to explain observed subsidence delay above a number of reservoirs.

Time-decay compaction can be modeled using a convolution of a linear relationship between pressure depletion and reservoir rock compaction with an exponential time decay function:

$$V_{comp}(x, y, t) = C_m(x, y, t) \cdot V(x, y, t) \cdot dP(x, y, t) *_t \frac{1}{\tau} \exp\left[\frac{-t}{\tau}\right] \quad (9)$$

Here, t , is time, $*_t$, is the convolution operator with respect to time and, τ , is a time decay constant.

7.3.2.4 Rate type isotach model

Traditional uniaxial experiments are conducted at one single constant loading rate. De Waal (1986) ran a series of experiments changing the loading rate during the experiments. The basis for the experiments originated from observation done on soft soil and chalks (Bjerrum, 1967) that indeed show a compaction behaviour that is dependent on loading rate. De Waal (1986) demonstrated that a faster loading rate led to stiffer response of the rock sample.

TNO (2013) further investigated the application of this model on cemented rock using the laboratory experiments from de Waal as a starting point (De Waal, 1986). Figure 17 shows the Standard Linear Solid (SLS) model, the simplest form of the general isotach model, as a spring-dashpot system (TNO, 2013).

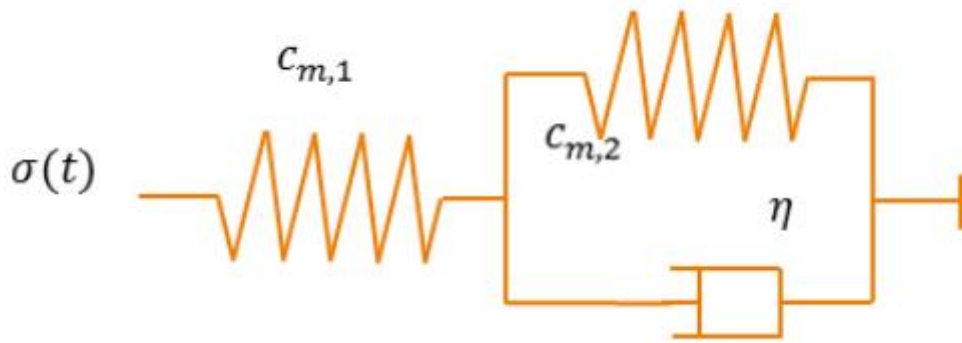


Figure 17: Representation of the SLS model.

De Waal (1986) ran a series of experiments changing the loading rate during the experiments where he demonstrated that a faster loading rate led to stiffer response of the rock sample. These experimental observations might explain the subsidence behaviour at the onset and arrest of production, which can be seen as changes in loading rate of the reservoir rocks. At a change of the loading rate, a first direct strain response is recorded followed by a more gradual response. Therefore, total strain is the sum of a direct part and a time dependent secular part:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_d + \boldsymbol{\varepsilon}_s \quad (10)$$

The rate type isotach compaction in **ESIP** is derived from an explicit Euler finite-difference scheme keeping a constant time step Δt . To calculate the compaction of one grid block grid (x, y) the applied numerical scheme can be divided into 5 steps as follows (TNO, 2016):

1) From the current effective vertical stress $\boldsymbol{\sigma}'(t)$ and strain $\boldsymbol{\varepsilon}(t)$, calculate the creep strain rate as:

$$\dot{\boldsymbol{\varepsilon}}_s(t) = \left(\frac{\boldsymbol{\varepsilon}(t) - \boldsymbol{\varepsilon}_0}{\boldsymbol{\sigma}'(t)} - C_{m,d} \right) \dot{\boldsymbol{\sigma}}'_{ref} \left(\frac{\boldsymbol{\varepsilon}(t) - \boldsymbol{\varepsilon}_0}{\boldsymbol{\sigma}'(t) \cdot C_{m,ref}} \right)^{-1/b} \quad (11)$$

The vertical effective stress is derived from the reservoir depth and the mean density ρ_{mean} of the subsurface up to the reservoir top \mathbf{z}_r as:

$$\boldsymbol{\sigma}'(t) = (\rho_{mean} \cdot \mathbf{g} \cdot \mathbf{z}_r) - \mathbf{P}(t) \quad (12)$$

At t_0 , that is at the onset of pressure depletion/production, the direct elastic strain $\epsilon_d(t_0)$ and secular or creep strain $\epsilon_s(t_0)$ are both considered equal to zero, and thus total strain $\epsilon(t_0)$ is set to zero.

The reference total strain is expressed as:

$$\epsilon_0 = -C_{m_{ref}} \cdot \sigma'_{ref} \quad (13)$$

with the reference vertical effective stress $\sigma'_{ref} = \sigma'(t_0)$.

Three material parameters ($C_{m_{ref}}, C_{m_d}, b$) and one state parameter (σ'_{ref}) are needed to compute the rate type compaction. The material parameters $C_{m_{ref}}$ and C_{m_d} are respectively the reference and direct compaction coefficients, where $C_{m_{ref}}$ is the compaction coefficient corresponding to the pre-depletion loading rate, and thus by definition quite high. Parameter C_{m_d} is dedicated to map out the direct effect at the change of loading rate. In the scenario of the change of loading rate due to the onset of pressure depletion, C_{m_d} is expected to be low in order to mimic the stiff response of the reservoir rocks. The ranges and source for the values of these parameters are described in paragraph 8.4).

2) The second step of the Euler scheme consists in calculating the increase in creep strain as:

$$\Delta\epsilon_s = \dot{\epsilon}_s(t) \cdot \Delta t \quad (14)$$

and update the creep strain as:

$$\epsilon_s(t_{+1}) \rightarrow \epsilon_s(t) + \Delta\epsilon_s \quad (15)$$

3) The time is updated as $t_{+1} \rightarrow t + \Delta t$

4) Following a linear stress-strain relationship one can calculate the direct elastic strain as:

$$\epsilon_d(t + \Delta t) = C_{m_d} \cdot (\sigma'(t + \Delta t) - \sigma'_{ref}) \quad (16)$$

5) Finally one can calculate the total cumulative strain as:

$$\epsilon(t + \Delta t) = \epsilon_s(t + \Delta t) + \epsilon_d(t + \Delta t) \quad (17)$$

And the total cumulative compaction as:

$$V_{comp}(t + \Delta t) = -\epsilon(t + \Delta t) \quad (18)$$

with V the grid block net volume, assumed constant over time. Clearly, accounting for changes in grid block net volume will not significantly change the compaction. After this last fifth step the workflow returns to the first step for the next time step.

Again it is important to note that the present rate type isotach compaction model is attempting to mimic the delay and persistence in subsidence rates at the onset and arrest of production, by only considering the reservoir compaction and assuming a purely elastic linear response of the rocks surrounding the reservoir rocks. The creep of a possible visco-elastic salt layer on top of the reservoir might also contribute to the non-linearity in the subsidence.

7.4 Calculate Subsidence

The compaction of the reservoir deforms the overburden as well resulting in the development of a subsidence bowl at the surface. The translation of reservoir compaction to subsidence requires the involvement of an influence function or Green's function. Geertsma (1973) provided a linear analytical influence function based on the nucleus of strain concept and assuming the over and under layers surrounding the reservoir as elastically homogeneous. Fokker and Orlic (2006) developed a semi-analytical method for generating influence functions for a multi-layer model, for layers with different elastic properties. The semi-analytical approach has the advantage of combining relatively complex sub-surface settings with small computational times, making it suitable for inversion and data assimilation exercises. The TNO tool in which this methodology is incorporated is called AEsups and is part of the ESIP workflow. AEsups is a program that calculates the subsidence profile for a single compacting "nucleus" with given depth, and for a given elasticity profile. The single-nucleus subsidence profile can be integrated over an input field mesh to achieve a subsidence profile. The application calculates the subsidence for a model with different horizontal layers with each a different elasticity modulus and Poisson's ratio.

Salt flow can be modelled in AEsups by giving a layer a non-zero linear viscosity. The assumption of linear viscosity is supported by both laboratory observations (Spiers et al., 1986; Spiers and Carter, 1998) as field observations.

There are a number of simpler, non-physical, influence functions that have been proposed in the literature that are placed under the umbrella of geometrical-integration methods (e.g. Sroka et al. 2011). These methods are mainly developed to describe the effect volume extraction of the subsurface by room and pillar mining on the surface subsidence but can be applied to a compacting reservoir as a result of depletion as well (e.g. Hejmanowski and Sroka, 2000). The Knothe function is a widely applied influence function of this type and therefore implemented in ESIP. The benefit of this method is the freedom to fit the influence angle to the observations (φ in Figure 18).

The Knothe influence function is a functional form which is dependent on the reservoir depth (H) and the angle(φ) (Cai et al. 2014). Here φ is the Monte Carlo Parameter (Figure 18).

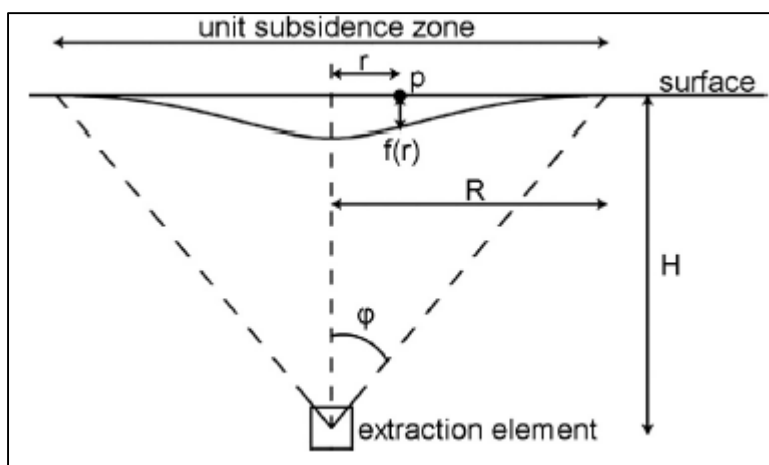


Figure 18: Knothe influence function.

The function implemented in ESIP follows Knothe's theory: $g_v = \frac{1}{R^2} e^{-\pi \frac{r^2}{R^2}}$

where R is the influence radius as $R = D \cdot \tan(\varphi)$, with D the reservoir depth and φ the influence angle. The influence angle is the only parameter that can be varied in the Monte Carlo procedure and is not a function of time.

A comparison with the subsidence bowl shape generated with Geertsma and van Opstal and the Knothe influence function is shown Figure 19. For Knothe an influence angle of 55 degrees has been used as the ESIP calculation returns this as the angle which fits best to the modelled subsidence. The Knothe influence function returns a steeper subsidence bowl.

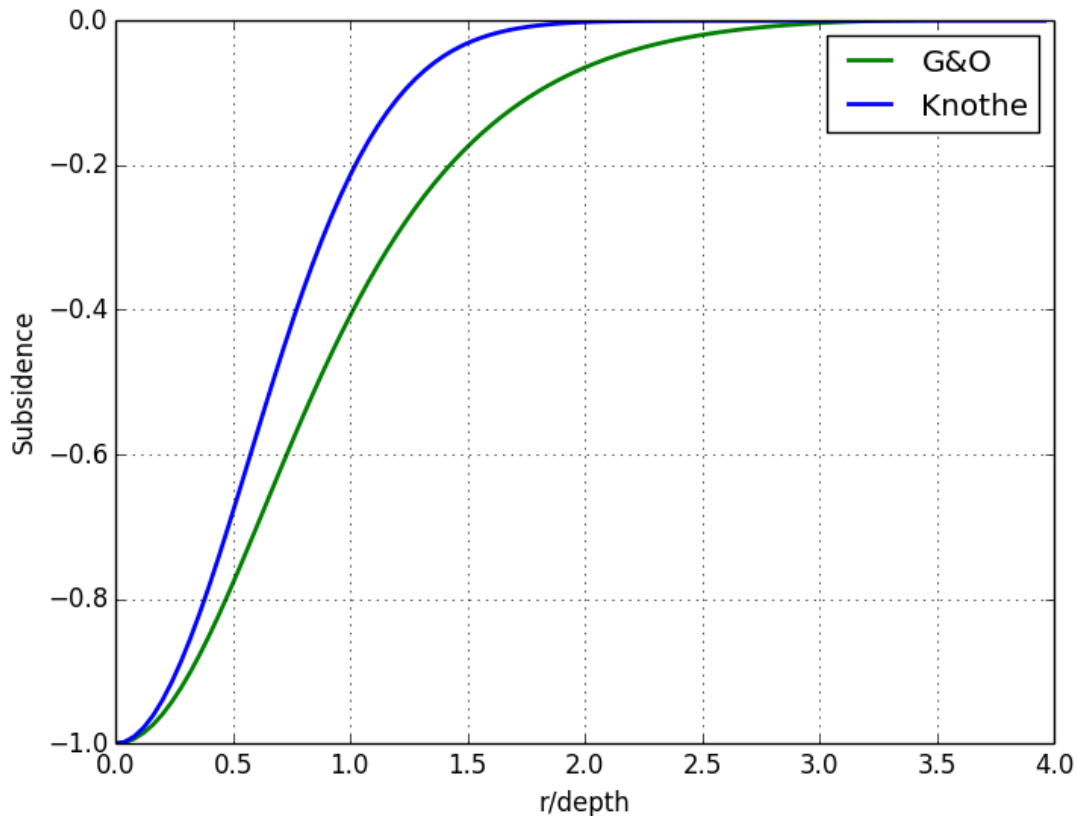


Figure 19: Comparison with the shape of the Knothe subsidence bowl using an influence angle of 55 degrees and a Geertsma and van Opstal calculation.

7.5 Model confrontation and conditioning

The model confrontation is the heart of the Bayesian ESIP workflow and confronts the a-priori defined model results with the data including its variance/covariance structure. The difference between the measurements and the data is the measure to calculate the likelihood of the specific member. To do so the model member produces double-difference results at the locations of the benchmarks. The goodness of fit between the model results and data is expressed by the value of χ^2/N .

In ESIP there are two methods available with the goal to refine the predictions:

1. The “Red-Flag” confrontational method (Nepveu et al., 2010)
2. The Ensemble Smoother approach (e.g. Fokker et al. 2016).

The LTS II study results are only based on the Red-Flag approach. The ensemble smoother is a methodology that mixes and conditions the output of individual model results to create new model results finally iterating to a narrower distribution of possible realisation. The mixing of the member solutions makes a mapping to the original reservoir scenarios and geomechanical model members less transparent. The aim of LTS II is to demonstrate the tool and provide more insight in plausible scenarios and input values to the models and influence functions. Therefore, only a description of the Red Flag model is presented in this report.

To arrive at the best posterior ensemble to be used in subsidence predictions, the ensemble smoother method is perhaps a better method, because it will collapse the prior model ensemble around the best scenarios, models and input values. We refer to TNO (2016) for a description of the ensemble smoother.

7.5.1 Red flag confrontation method

The Red Flag method (Nepveu et al., 2010) was designed to enable decision-making when a threshold in a process is about to be exceeded. The methodology combines measurements with prior information and updates the probability of each model member after a confrontation step with the data. With this updated probability the probability of the member ensemble can be calculated.

For a particular realisation \mathbf{r} of the ensemble \mathbf{dd}_r^{prior} , the Red-Flag approach defines the mismatch function as:

$$\chi^2 = (\mathbf{dd}_r^{prior} - \mathbf{dd})^T \mathbf{C}_{dd}^{-1} (\mathbf{dd}_r^{prior} - \mathbf{dd}) \quad (19)$$

with \mathbf{dd} the data vector and \mathbf{dd}_r^{prior} the prior model. The lower the value of the mismatch function, the better the match of the vector of prior double difference predictions with the measurements. In this formulation, \mathbf{C}_{dd} represents the covariance matrix of the measurements. The Bayesian probability of a particular realization \mathbf{r} of the ensemble is given by:

$$P(\mathbf{dd}_r^{prior} | \mathbf{dd}) = \frac{P(\mathbf{dd}_r^{prior}) \cdot P(\mathbf{dd} | \mathbf{dd}_r^{prior})}{\sum_{i=1}^{Ne} P(\mathbf{dd}_i^{prior}) \cdot P(\mathbf{dd} | \mathbf{dd}_i^{prior})}. \quad (20)$$

With:

$$P(\mathbf{dd} | \mathbf{dd}_r^{prior}) = \exp\left[-\frac{\chi^2}{2N}\right], \quad (21)$$

with N the number of double differences. The calculation scheme of the confrontation with the geodetic data measured on e.g. T1 and T2 is explained in detail in TNO, 2016.

7.6 Demonstration of the ESIP workflow

Based on discussions with the steering committee, several tests were set up to test and demonstrate the functionality, effectiveness and reproducibility of the ESIP software. The aim of these tests was to answer the following questions:

- How does the outcome of the upscaling and interpolation procedures compare to existing methods currently used by NAM? (7.6.1)
- How does the calculated subsidence from different compaction models compare to those from existing methods currently used by NAM? (7.6.2)
- How consistent are the outcomes of the ESIP software when run by different users on different machines? (7.6.3)
- What probability distributions for the different compaction models does ESIP return when applied on synthetic input data with a known compaction model? (7.6.4)
- What is the impact of data density (spatial and temporal) variations on ESIP's ability to discriminate between the likelihood of certain pressure depletion and compaction models? (7.6.5)

Here, we describe the testing process for answering these questions. We start with a test aimed to verify the quality of upscaling, then a test of the output consistency. These tests are followed by an experiment/discussion of the output distributions provided by ESIP compared to the known input, and an experiment to verify the influence of data density on model discrimination.

7.6.1 Upscaling and interpolation steps

The pressure data as received typically consists of pressure values for a number of reservoir layers, and often uses a grid size that is unpractically small for compaction and subsidence calculations. One of the first steps in the ESIP workflow is therefore to transform the pressure data into manageable grid sizes and time steps. To verify the quality of the upscaling procedure, and check if the results are consistent with procedures previously used, we used raw pressure data from the 'Dummy model' and upscaled these using ESIP and NAM's equivalent scripts in parallel.

Figure 20 shows the pressures after upscaling from a 100x100m grid size with 15 distinct vertical layers with varying thickness, to a one-layer format with a lateral grid block size of 500x500m. To quantify the difference between the NAM and ESIP upscaling, we look at the sum of the pressure times the volume for all the grid blocks. Since the compaction is related to the volume times pressure depletion, the sum of pressure times volume will be a proxy for the compaction and subsidence and should not differ too much between upscaling procedures. See Table 2 for the obtained values. In the exercise here, we find a match to within 0.1% between the methods (main contributions to the difference comes from the edges of the upscaled grid, which originates from different grid-centres in both methods), clearly well below the uncertainty levels of both pressure and volume, and demonstrating the agreement between the upscaling procedures.

To also establish if the procedure has actually correctly upscaled the properties from the raw input, we calculated the same sum of volume times pressure over the fine input data. For the pre-upscaled input data, the total sum was 2.16262×10^{13} , while for the post-upscaled process we obtain an almost identical value (difference 0.0014%), indicating that the upscaling of the pressure data will have very little impact on the accuracy of the calculated compaction volume.

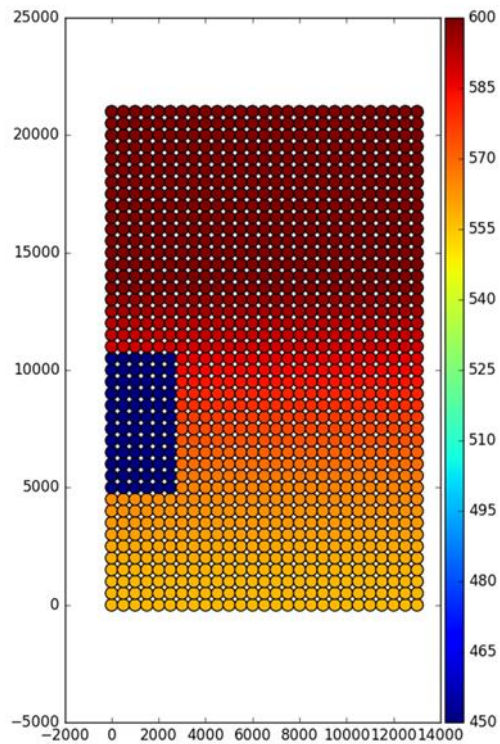


Figure 20: Upscaled pressures in a dummy model.

Table 2: Sums of volume times pressure over the grids for different upscaling procedures, and pre-upscaling.

Data set	sum(V*P) for all grid cells [m ³ bar]	Relative deviation from input
NAM upscaling	2.16519×10^{13}	0.1%
ESIP upscaling	2.16262×10^{13}	<0.01%
Raw input	2.16262×10^{13}	0%

7.6.2 Subsidence calculation

We continued with the upscaled data from the previous section to verify that the differences in the resulting subsidence as determined by ESIP and existing NAM scripts are negligible. The NAM scripts used, apply a Geertsma and van Opstal (1973) calculation with a rigid basement at 5 km. The subsidence is calculated for a linear, time decay and a rate type compaction model. The results from both workflows are presented in Figure 21 (linear compaction), Figure 22 (time decay compaction) and Figure 23 (rate type compaction), showing an excellent agreement.

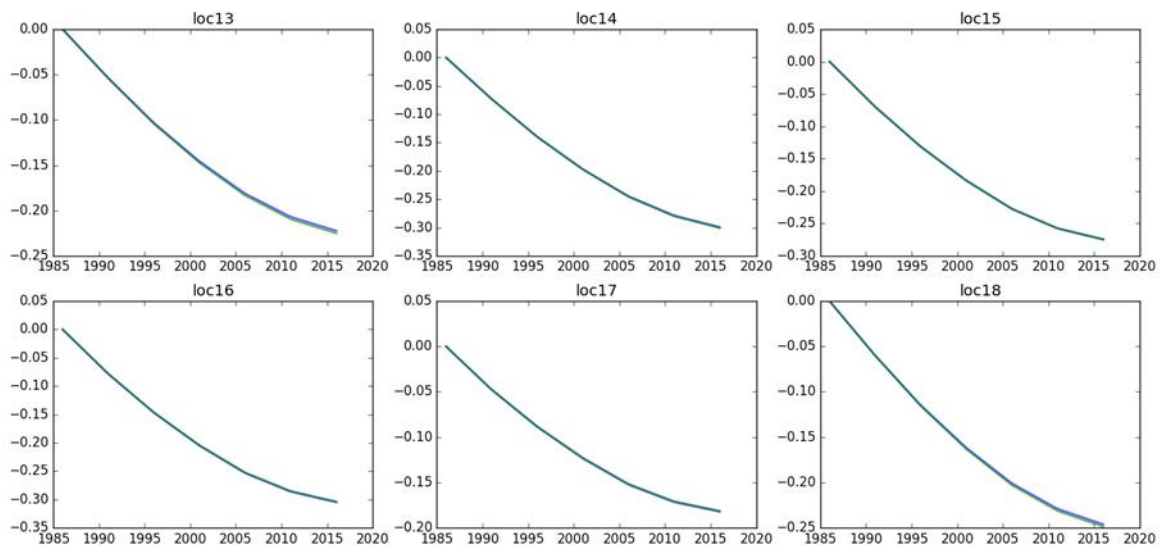
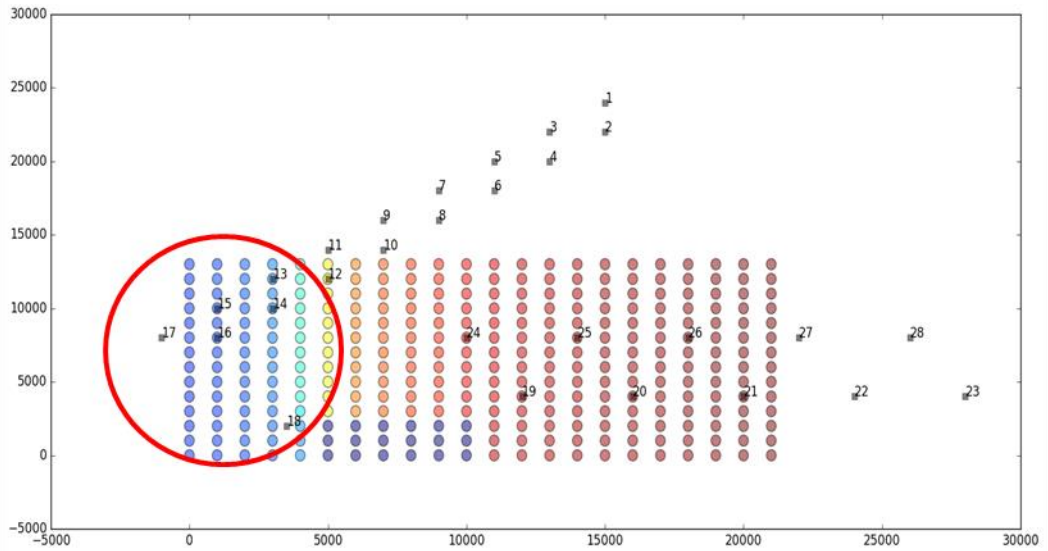


Figure 21: Calculated subsidence using a linear compaction model from NAM scripts (green) and ESIP (blue). The red circle in the top figure highlights the location of the benchmarks for which the subsidence is shown.

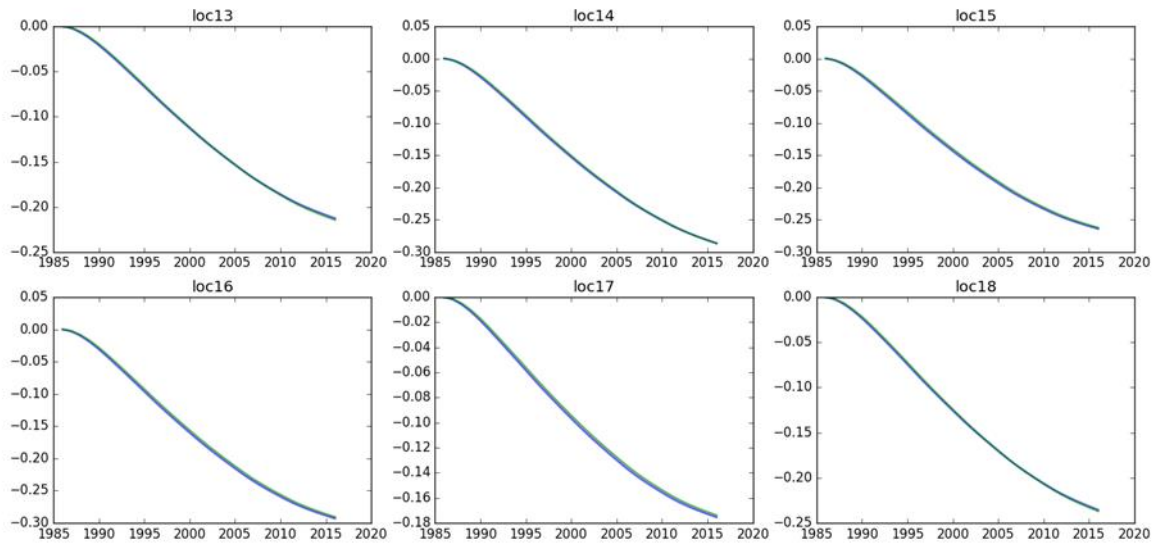
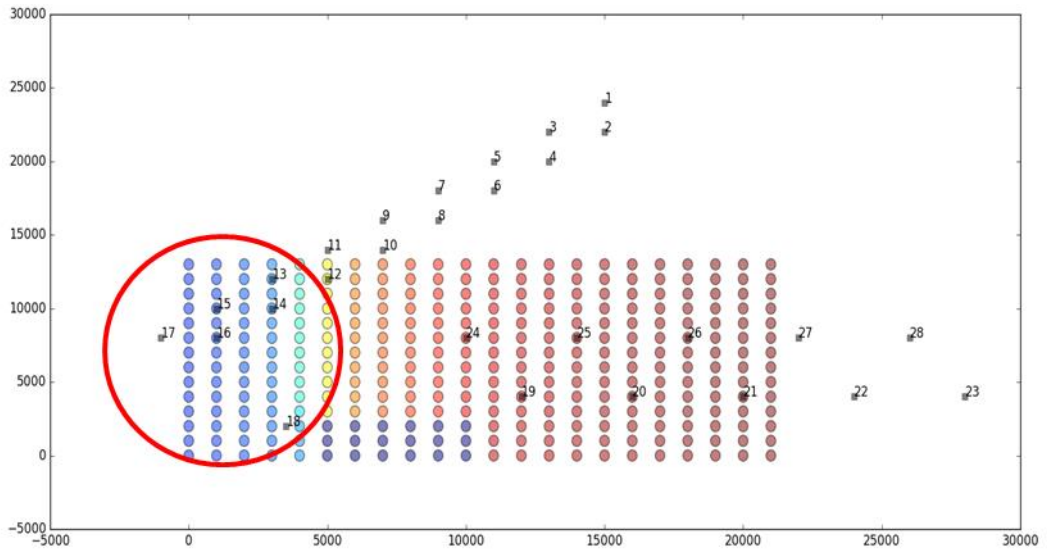


Figure 22: Calculated subsidence using a Time decay compaction model from NAM scripts (green) and ESIP (blue).

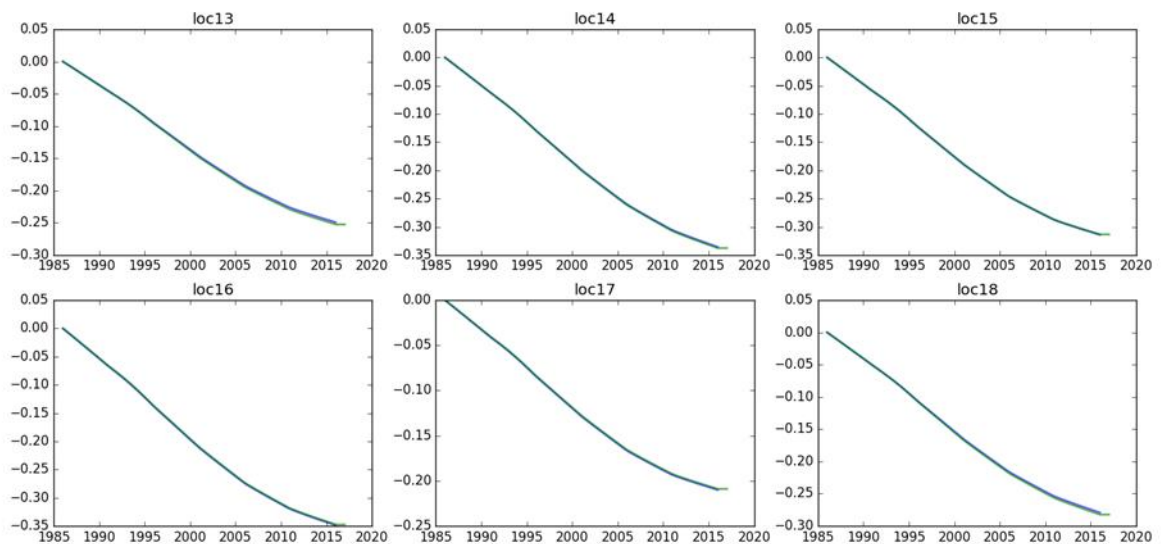
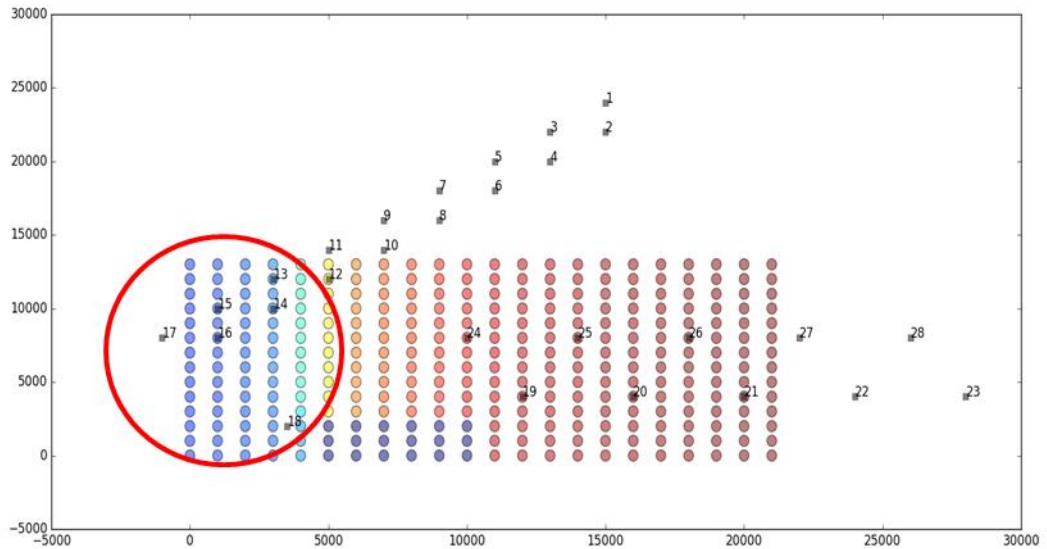


Figure 23: Calculated subsidence using a rate type compaction model from NAM scripts (green) and ESIP (blue).

7.6.3 ESIP-outcome reproducibility

The aim of the reproducibility test is to establish that, given the complexity of the software, the outcome of the software is not affected when changing either the operator or the physical computer it is installed on.

To perform the test, a set of synthetic double difference data were generated to serve as geodetic input data, together with a total of 76 different reservoir scenarios. Two different operators, at two different computers (one at TNO, the other at NAM), used the identical geodetic and pressure input data, but ran the software independently, starting from the upscaling of the input data all the way through to the confrontation steps.

In these independent runs, four compaction models were tested [linear, bilinear, time decay, rate type]. On each pressure model, 15 different sets of parameters (generated by the MC approach, which draws all input parameters from a uniform distribution) were tested for each compaction

model. The total number of ensembles tested per compaction model was therefore 76 times 15 = 1140, bringing the total number of ensembles for all four compaction models to 4560.

The double differences generated for these 1140 ensembles per compaction model are confronted with the synthetic input double difference data, making a weighing of the individual ensembles within a compaction model possible. Figure 24 shows the resulting (sorted) weights per compaction model for the two runs. The distributions of the weights are close to identical for the runs. It should be noted that the non-infinite number of MC runs adds some random deviations between the runs. Only if one were to run the same test for an infinite number of MC runs, identical distributions should be expected.

In Figure 25, we compare the two sets in terms of double differences and prior and posterior weighted averages. In green, the double differences as calculated for all 1140 tested (linear compaction) models. The result for one of the runs (NAM or TNO) is in light green, the other in darker green. For both sets, an average (prior to the data confrontation) is shown in blue (again, light and dark blue to distinguish between the two runs), indicating both runs have generated similar data sets (spread in green lines) with a near-identical average (blue lines). After confrontation, the different models get a weight assigned based on their resemblance to the data, resulting in a posterior average that are indistinguishable (red and cyan lines). Figure 25 shows this for the linear compaction model. The same can be done for the bilinear, time decay and rate type compaction models.

From the scatter in the plot of model vs data, a $\frac{\chi^2}{N}$ value can be determined using:

$$\frac{\chi^2}{N} = \frac{1}{N} (\mathbf{dd}_r^{prior} - \mathbf{dd})^T \mathbf{C}_{dd}^{-1} (\mathbf{dd}_r^{prior} - \mathbf{dd}) \quad (22)$$

To quantify a goodness of fit. The resulting $\frac{\chi^2}{N}$ values from these three different runs (one by TNO, one by NAM) are listed in Table 3 from the output of the red flag confrontation method, for all four compaction models that were tested. The good agreement between the different operators demonstrates the reproducibility in all cases.

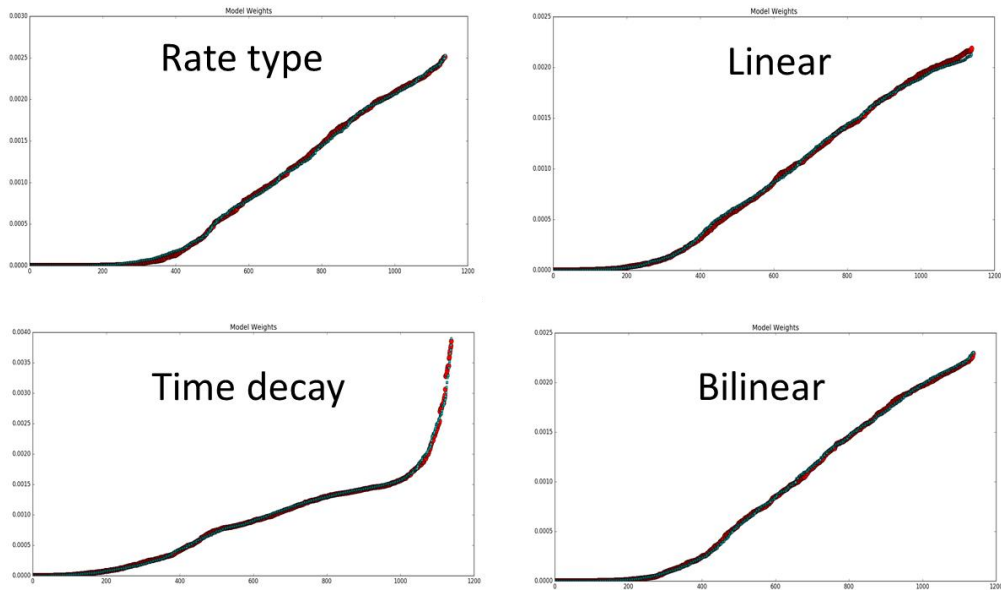


Figure 24: Model weight distributions from two different ESIP runs (NAM, TNO) on the same input data. Runs were performed on the same (synthetic) geodetic input and the same 76 pressure models. The weights have been sorted per compaction model from small to large (x-axis indicates member number, total of 1140 members per compaction model).

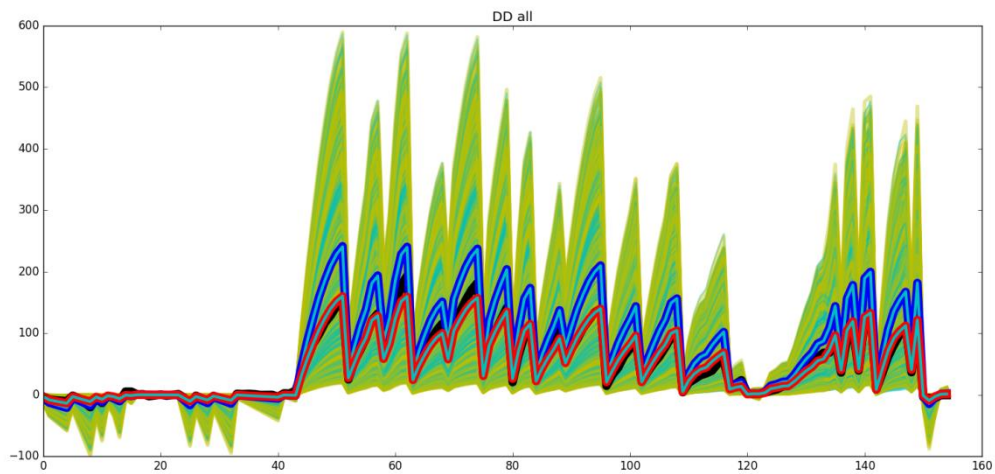


Figure 25: Synthetic input double differences (black line) compared to model double differences from all linear compaction model members (light green and dark green reflecting the difference in run, by either TNO or NAM). Prior weighted averages for both runs (light and dark blue) as well as the posterior weighted averages (red and cyan) are both indistinguishable between the two runs. Posterior average shows close agreement to the input data.

Table 3 χ^2/N results between different runs (TNO, one by NAM), for Red Flag confrontation of the synthetic data with four compaction models (each value based on a total of 1140 members; 15 sets of compaction model parameters for 76 different pressure scenarios).

Operator	Confrontation method	Linear	Bilinear	Time decay	Rate Type
TNO	Red Flag	1.053	1.049	1.126	1.050
NAM	Red Flag	1.053	1.050	1.118	1.050

7.6.4 Comparing ESIP output with known input

ESIP was developed to assess the ranges of possible surface subsidence based on uncertainty in input data. Using the dummy model, we are in a position where we can generate a synthetic subsidence data set, and compare the output distributions with the known input compaction model. Because the uncertainty in the input data (which is also present in the synthetic data) is taken into consideration in the ESIP process, the output will never be a one-model selection. Instead, it will produce a range of models each having a probability assigned to it. It should be expected, however, that the known input lies within the range of models with reasonable probability. We discuss the results here, and use the outcome of the tests here to create an understanding and awareness of the workflow output.

For the comparison we used one reservoir scenario to generate synthetic double differences (and subsidence) using a bilinear compaction model with known parameters $C_{m, pre}$, $C_{m, post}$ and P_{trans} . We then applied the ESIP workflow, which generates double differences for a range of compaction model parameters and confront these with the synthetic data, resulting in a weighing, or probability distribution of all the tested models. Two different setups are presented here.

In the first test, we generated synthetic data using a bilinear model where the two compaction coefficients, $C_{m, pre}$ and $C_{m, post}$, are very close to one another, making the model effectively close to a linear compaction model. A total of 1000 models were tested in the ESIP process, of which the model weights after confrontation with the synthetic data is shown in Figure 26, plotted against the three compaction model parameters that were varied among the 1000 members. The parameter values as used for generating the synthetic data are marked by the blue lines.

One can see a clear distribution of the model weights centred around the input value for $C_{m, pre}$, and to a lesser extend for $C_{m, post}$, whereas for the transition pressure there appears to be no preference for a certain value. Although on first glance these results might not be a convincing and successful demonstration of the workflow, these results are to be expected for the model used here. The non-discriminating power in P_{trans} is the result from the near-identical values used for $C_{m, pre}$ and $C_{m, post}$, meaning that the location of P_{trans} has only little effect on the resulting compaction and subsidence.

The difference in the distributions of C_m parameters also has its origin in the synthetic data. Because the transition pressure chosen in the synthetic data only comes into effect later on in the data, there

are relatively few double differences affected by a change in $C_{m, post}$, whereas $C_{m, pre}$ has a higher impact on the match between model and data, hence a narrower distribution.

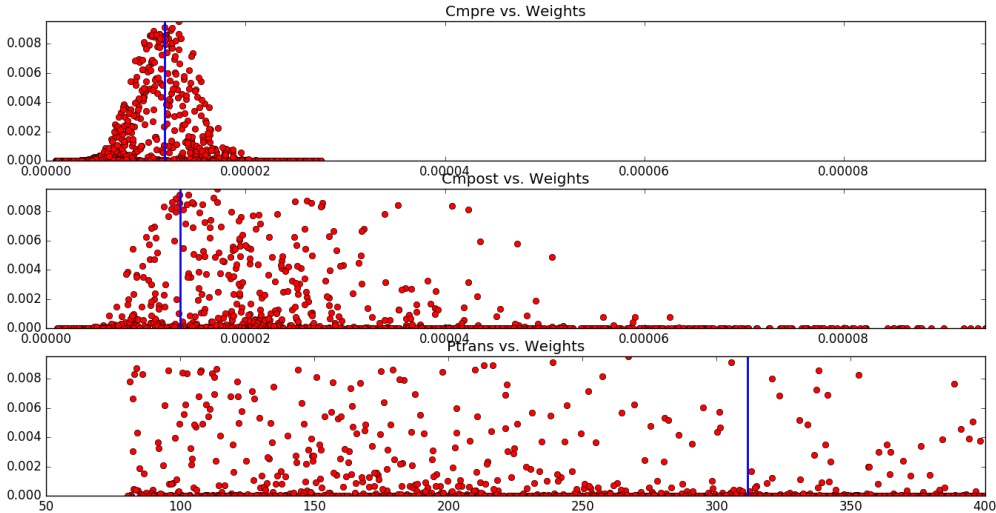


Figure 26: Model weights resulting from the confrontation with synthetic data, plotted against the compaction model parameters. The input parameter values for the synthetic data are marked by the blue lines.

To assess that the models with higher weight assigned in Figure 26 also follow the synthetic input data (i.e. they are actually good models describing the data), we plotted the model compaction (C_m times dP) versus the pressure in Figure 27. The comparison is commonly shown as a subsidence vs time plot, but to show the behaviour of the compaction model only, we eliminate all the contributory factors coming from the pressure and elasticity parts of the models, by plotting the compaction instead of the subsidence. The red line is the synthetic data. All other lines show the compaction predicted by each model tested, their colour represents the model weight as determined by the ESIP procedure. Because of the low contrast in $C_{m, pre}$ and $C_{m, post}$, the data in red shows almost a linear compaction behaviour. The best models in yellow, even those where P_{trans} is very different from that used in the synthetic data, all lie around the input compaction data.

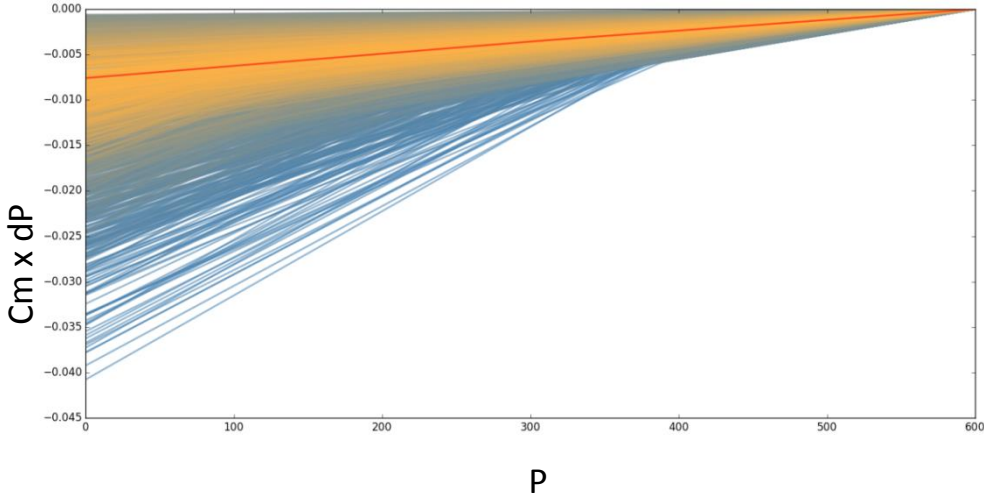


Figure 27: Model compaction values as a function of pressure. The red line is the synthetic data, and the other lines the tested models. Their colour represents the model weight as in Figure 26, yellow being the higher probability.

The test was performed a second time with a bilinear compaction model as input now having more contrasting pre and post compaction coefficients. The resulting model weights after confrontation are plotted again against the model parameters in Figure 28. As a result of the higher contrast in the compaction coefficients, it now does matter what transitional pressure is used for a model to have a good resemblance to the data, resulting in the presence of structure in the weight distribution of P_{trans} , which was absent in Figure 26. The models with a high weight now has a value for each of the three compaction model parameters around the known input values.

Note however that there are many models with a very low (close to zero) weight, even at locations around the input parameters. These exist because the weights can actually not be decoupled to the different parameters as the figure suggest, but are the result from a set of parameters. Because there is a large contrast in parameters in this test, it can happen that even though one parameter is close to the input, say $C_{m,pre}$ is close to the blue line, the other parameters might be too far off, resulting in a poor match to data and correspondingly low weight.

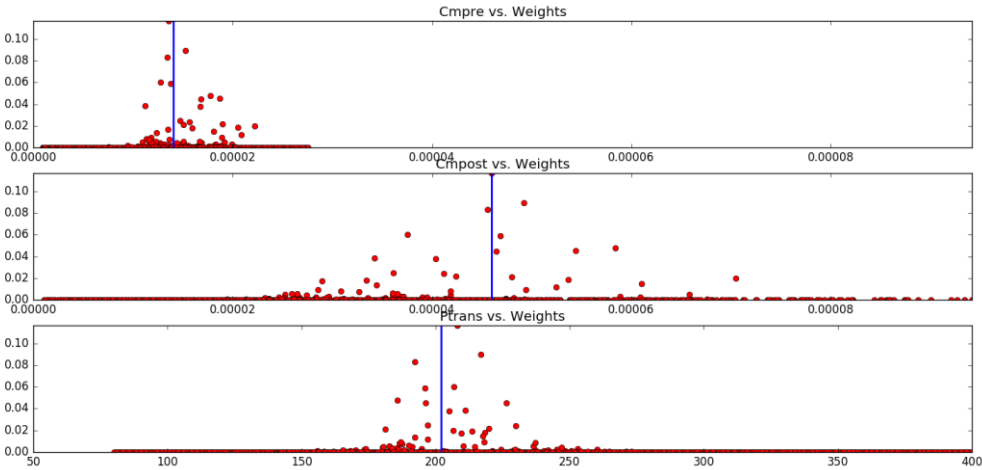


Figure 28: Model weights resulting from the confrontation with synthetic data, plotted against the compaction model parameters. The input parameter values for the synthetic data are marked by the blue lines. Compared to the results in Figure 26, there is now a large contrast in the two compaction coefficients of the input.

Like for the previous set, we also generated a compaction versus pressure plot, in Figure 29. Because there now is contrast in compaction coefficient pre and post the transitional pressure, the input data (red line) exhibits a clear kink around P_{trans} . Again, the data with the highest weights after confrontation (yellow) follow the synthetic input data very well.

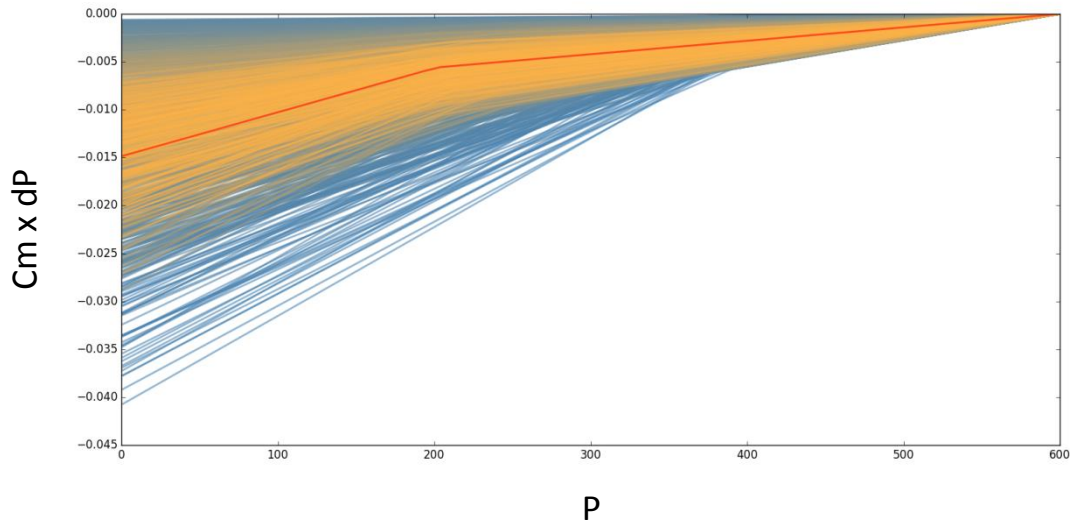


Figure 29: Model compaction values as a function of pressure. The red line is the synthetic data, and the other lines the tested models. Their colour represents the model weight as in Figure 28, yellow being the higher probability.

In summary, the models with a high weight after the confrontation with the geodetic data in ESIP do describe the data well. When plotting the assigned model weights against the model parameter values, the known input values lie within the regions of highest weights. Note that the width of these distributions depends on the impact the particular parameter has on the match to the data; the more impact, the narrower the constrained/distribution of weights for this parameter will be. Additionally, note that the weights can strictly speaking not be decoupled into the individual parameters that make up the model, meaning that also low weights can exist at or around the true parameter value in these type of plots.

7.6.5 Effect of data density on model discrimination

Since most of the area of the subsidence bowl lies below the Waddenzee and only part of the subsidence bowl on the Ameland island, the subsidence data is not evenly distributed in the lateral sense. Most of the levelling benchmarks are located directly above the gas reservoir, while there are only a few GPS stations directly above the lateral aquifers. The distribution of the benchmarks has raised concern in the past about the ability of the data handling to discriminate between different reservoir scenarios.

An example is the ability to discriminate between a model where the lateral aquifers do deplete in one model and not in another. While these models can predict a very different subsidence volume, and correspondingly very distinct subsidence rates, they could both fit the subsidence bowl above the gas very well, i.e. the part where most of the geodetic data lie. Because of the high data density in this area, a very different fit between model and data on the few benchmarks above the aquifer might not result in a clear discrimination between these two subsidence bowls, even though they can largely vary in terms of subsidence rate predictions.

Since ESIP carries not only the variance per double difference but also the covariance between double differences, it is claimed that it should be able to still make the discrimination between models to some degree, even though the models only give different predictions on a very small portion of double differences. The idea is that, even though there is a very high density of double differences above the island, these get a relatively lower individual effective weight through the

covariance matrix since they carry more dependence on one another, compared to points that lie more isolated in the Waddensee area.

We tested the effect of the data density above the aquifer by running the ESIP model multiple times. We varied both the data density and the aquifer depletion in these models. In terms of data density, we once used a high density above the gas and once after removing some of these data, while keeping the amount of data above the aquifer constant. In terms of the pressure models for generating the input geodetic data, we used one being a pressure model with significant aquifer depletion. Both this pressure model as well as a pressure model with very little aquifer depletion are then tested in the ESIP workflow, with the hypothesis that, independent of the data density, the pressure model used for generating the synthetic benchmark data is the one that gets higher weights assigned.

The synthetic data is generated for a bilinear compaction model. The confrontation was done for 1000 members per reservoir scenario, so a total of 2000 members per tests.

If the data density does indeed not result in the method favouring certain models, then the high depletion – low density and high depletion – high density tests would give similar weight distributions, both in favour of the high aquifer depletion model, and similar for the two low depletion tests.

The final pressures for the two pressure models (dummy model) used for the confrontation are shown in Figure 30. The locations of the benchmarks are shown in Figure 31, where the green polygon highlights the benchmarks used in the high data density cases, while the red polygon highlights those for the low data density case.

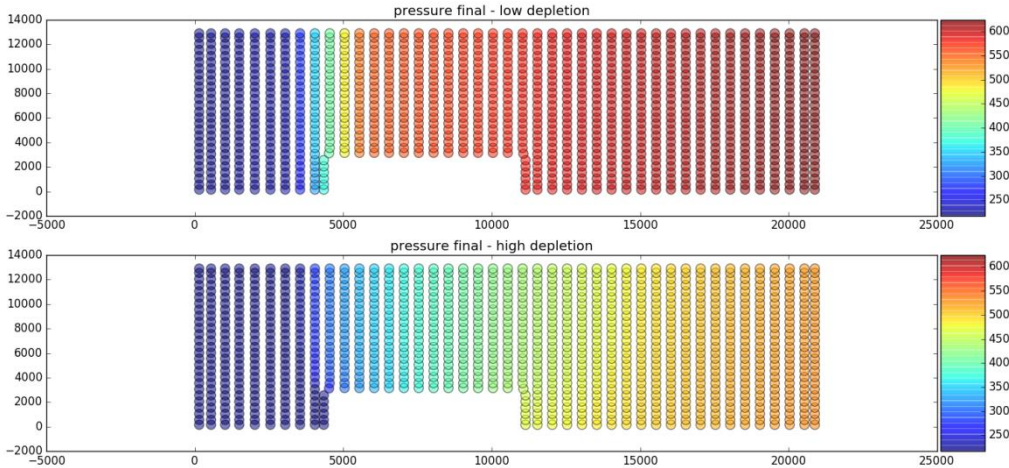


Figure 30: Two pressure models used in the tests here. The top figure shows low pressure depletion in the aquifer, while the bottom figure shows high depletion in the aquifer. Depletion in the gas reservoir are comparable. Pressure model with the high aquifer depletion was used to generate synthetic double difference data.

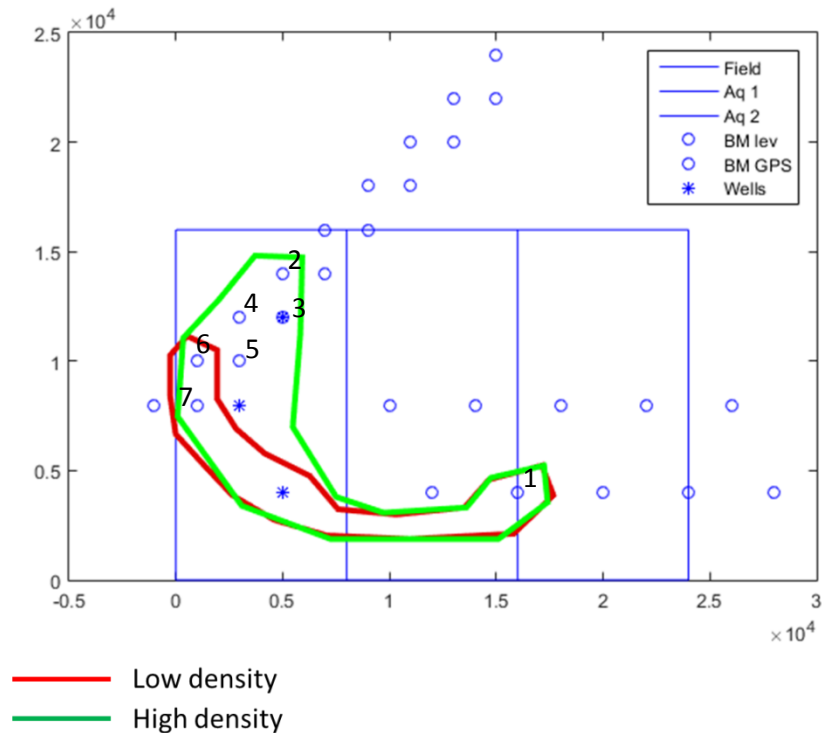


Figure 31: Benchmarks used in the tests. Green polygon bonds all benchmarks used for the case of high data density above the reservoir, while the red polygon encloses the benchmarks for the low data density case. Numbers in the figure label the benchmarks as used in the high data density case from 1 to 7.

Here we test the workflow using synthetic data that was generated using the high aquifer depletion pressure model. This model would give rise to subsidence above the aquifer, where there is however only one benchmark. The synthetic data is generated using a bilinear compaction model; once with high compaction coefficients to amplify the difference between the two pressure models above the aquifer, and once with low compaction coefficients to do the test for models that only have subtle differences above the aquifer.

The first set of synthetic data was generated using a bilinear compaction model with relatively large compaction coefficients, to amplify the differences between the two different reservoir models above the aquifer.

We start with the **low data density case**, where making the distinction between different pressure scenarios is expected to be easier, given the relatively large number of points above the aquifer.

Similar to the plots presented in Section 7.6.4, Figure 32 shows the model weights after confrontation for the low data density case, plotted against the compaction model parameters. In blue, the values used for generating the synthetic data are shown. Clearly, the weights from the full ensemble (containing models based on both the high and low aquifer depletion models), demonstrating that the workflow does return models around the known input.

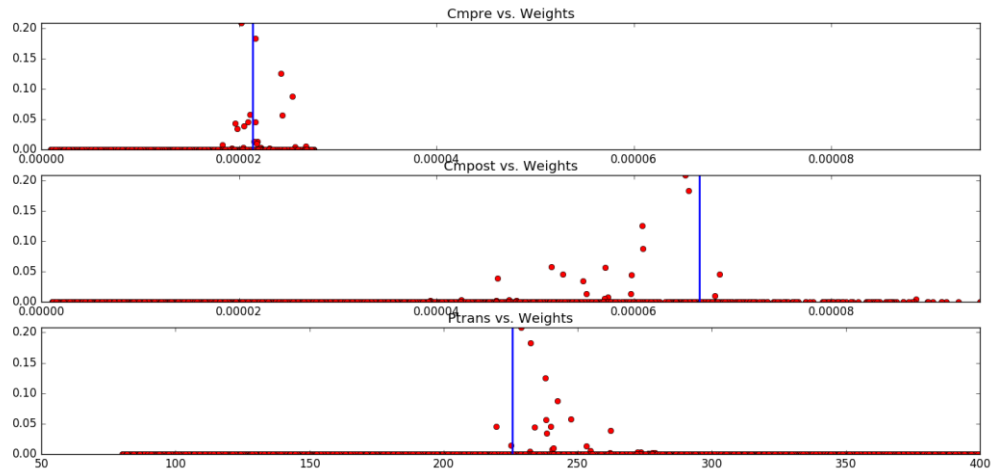


Figure 32: Model weights plotted against the compaction model parameters (low data density). The blue lines indicate the parameters used for generating the synthetic data.

The same model weights are plotted in Figure 33, now against ensemble number. The first 1000 ensembles use the high aquifer depletion model (i.e. the same that was used for generating the synthetic benchmark data), while varying the compaction model parameters, and the second 1000 models use the low aquifer depletion model. All models that have a weight assigned to them that is visible on the scale of the plot belong to the set that used the high aquifer depletion pressure scenario, i.e. the best subsidence models after confrontation seem to all be based on the correct pressure model. The sum of the weights of the first thousand members is >0.99 , and the sum of the weights of the second thousand members is 10^{-12} (total sum being 1); evidently allocating most weight to the members that used the same pressure scenario that was used for generating the synthetic input data. As expected, a distinction in the likelihood of these two pressure scenarios can be made, with the low data density.

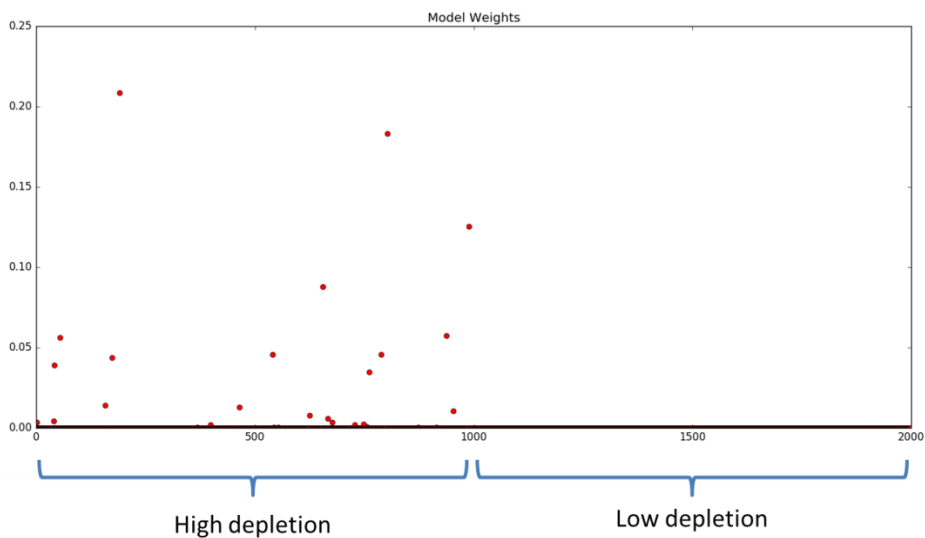


Figure 33: Model weights plotted for all the ensembles tested (low data density, synthetic input data generated using the high aquifer depletion). The first 1000 ensembles use the high aquifer depletion pressure model, whereas the second set of 1000 models use the low aquifer depletion model. Sum of the weights of the first thousand members is 0.99, whereas the sum of the second thousand members gives 10^{-12} .

The same bilinear compaction model was used to generate the synthetic data for the larger number of benchmarks used in the **high data density case**. As in the low data density case, we confronted the synthetic data with 1000 ensembles using the high aquifer depletion pressures, and 1000 ensembles using the low aquifer depletion scenario. The resulting weights are plotted as a function of the compaction model parameters in Figure 34, which similar to Figure 32, demonstrates that also with the high density case the best models seem to be using compaction model parameters similar to those used for the input.

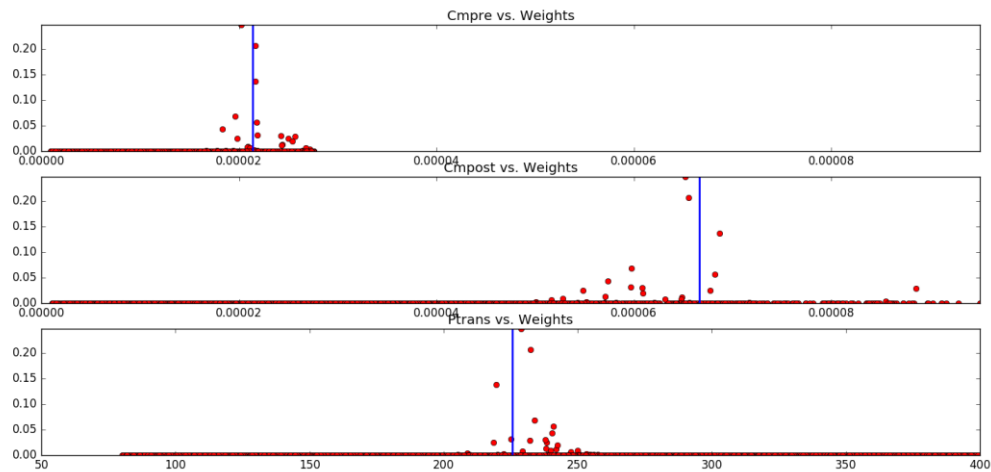


Figure 34: Model weights plotted against the compaction model parameters (high data density). The blue lines indicate the parameters used for generating the synthetic data.

More importantly, Figure 35 shows that one can still discriminate between the two reservoir scenarios, as the models with the highest weights or probability are based on the pressure scenario with the high depletion. The sum of the weights of the first one thousand members is now still >0.99 , with the sum of the second still negligible with a sum of 10^{-6} , demonstrating that still a clear distinction is possible.

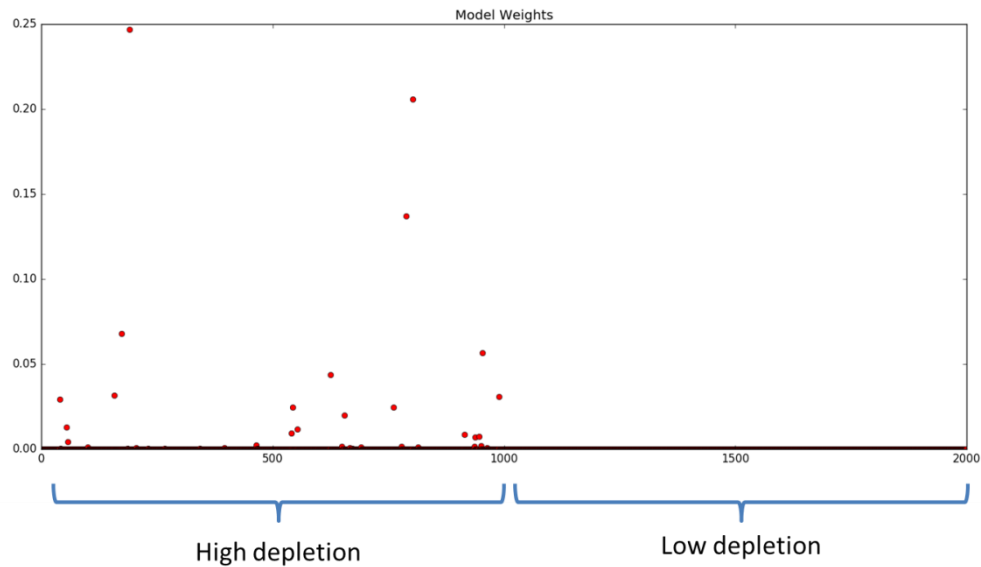


Figure 35: Model weights plotted for all the ensembles, similar to Figure 33, now using a high data density (relatively fewer points above the aquifer). A similar distinction between pressure depletion models is seen as for the low data density case. Sum of the weights of the first thousand members is 0.99, whereas the sum of the second thousand members gives 10^{-6} .

The above shows that for a more extreme case, the workflow can still discriminate between the two different pressure scenarios, regardless of the distribution of the benchmarks. We do the same, but now using a compaction model on the synthetic data that gives less compaction, making the difference above the aquifer in terms of subsidence smaller.

In the second test we use an extreme case where the compaction coefficients are small, such that the subsidence above the aquifer resulting from members with or without aquifer depletion will be more similar. The member weights after confrontation with the synthetic data are shown in Figure 36 for both the low data density and high data density cases, indicating both setups return member weight distributions around the known compaction parameter input.

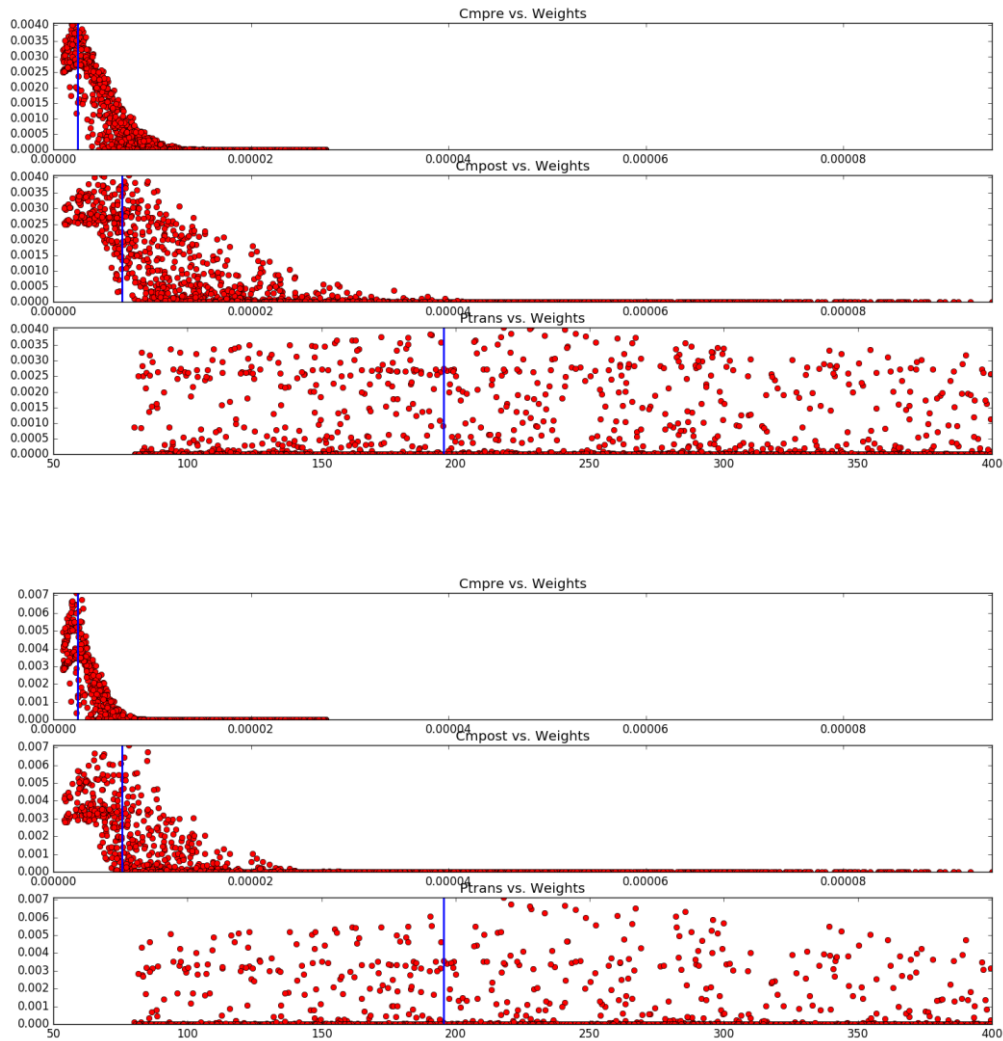


Figure 36: Member weights plotted against the compaction model parameters. Top figure is when using the low density of benchmarks, bottom for using a high density of benchmarks. The blue lines indicate the parameters used for generating the synthetic data, showing the low compaction coefficients used. Both high and low data density cases return similar distributions, both around the parameters used in the synthetic data.

In Figure 37, the same weights for the low data density test are shown but now plotted against the member number. Again, the first one thousand members used the same pressure member that was used for generating the synthetic data (with high aquifer depletion) and the second half used the pressure model with the low aquifer depletion. Clearly, the best models all fall in the first half of the set, which are the members that used the same pressure scenario as that was used for generating the synthetic data. The sum of the weights of the first half is 0.54, where that of the second half is 0.46, showing a small but visible distinction based on the average weight, for this extreme case that shows very little compaction. Figure 38 shows the same but now for the high data density case. Again, the best models lie within the first half, meaning the best models are still those that use the correct pressure scenario. The average weights when using the high data density, however, are much closer for both pressure scenarios, as the total sum of the high depletion scenario is 0.49, and the low depletion scenario is 0.51, showing on average almost no difference. The distinction based on the average weight is therefore almost not present based on the average weights, since there are

also models with the high depletion case that have a much lower weight (those that result in too much compaction).

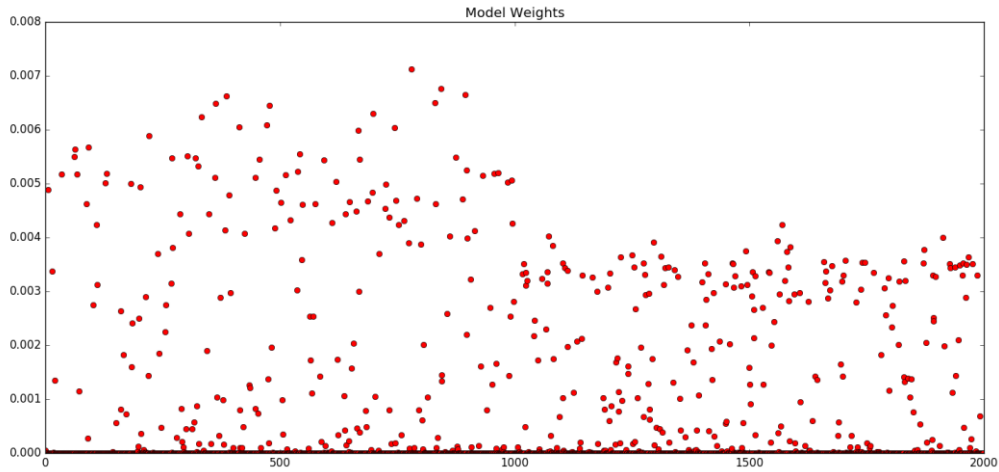


Figure 37: Same as data as in Figure 36 for the low data density case, but now plotted against the member number. The first half of the set used the same pressure scenario as was used for the data generation (high aquifer depletion), while the second half used the other pressure scenario. Clearly, the models with the highest weight all fall in the first half of the set, all using the same pressure scenario as was used for the input data. The sum of the weights of the first thousand members is 0.54, while the sum of the weights of the second set of one thousand members is 0.46, indicating still a small but visible discrimination.

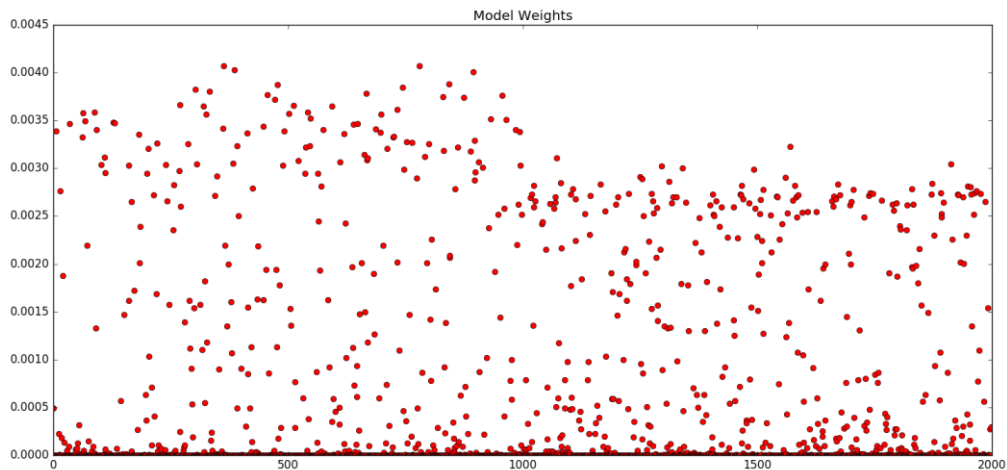


Figure 38: Same weights as in Figure 36, for the high data density case, but now plotted against the member number. The first half of the set used the same pressure scenario as was used for the data generation (high aquifer depletion), while the second half used the other pressure scenario. Still, the models with the highest weight all fall in the first half of the set, all using the same pressure scenario as was used for the input data. However, in this case the sum of the weights of the first thousand members is 0.49, while the sum of the weights of the second set of one thousand members is 0.51, showing that a discrimination based on the average weight per pressure scenario is in this extreme case difficult.

The total weighted average, though, still describes the data well, and still takes more weight from the models that used the high depletion scenario. The effect is illustrated in Figure 39 and Figure 40. Figure 39 shows the subsidence on six locations, location 1 being above the aquifer, for all the

members tested in the high data density case. The colouring indicates the member's weight; yellow being high, blue being low. In dark blue is the prior average, and in red the posterior average, the latter closely resembling the data (black), also above the aquifer (location 1). The spread is however very large, as some of the members using the high depletion case can predict very large subsidence. In the confrontation, these members will receive a very low weight.

The same data is plotted in Figure 40, but now only plotting the data from members that used the low depletion scenario (i.e. the pressure model that was not used in the input data). The spread of data at location 1 is much smaller, and all these members seem to describe the data at location 1 well, meaning that on average all these models can give good predictions, but at the same time none of them are very good, suggested by the lack of yellow lines. All the best models, with the highest weights, still come from the set of members using the high depletion scenario.

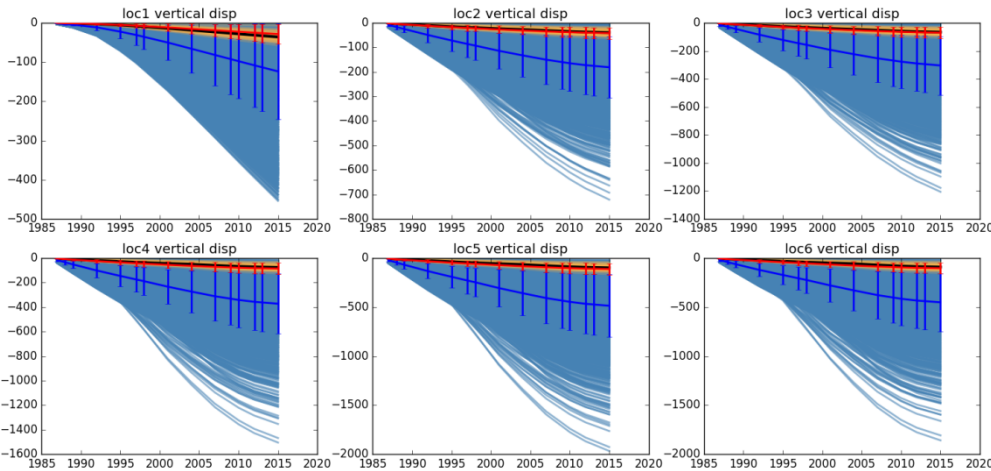


Figure 39: Subsidence on three locations, high data density, all members. Locations of the benchmarks are indicated in Figure 31.

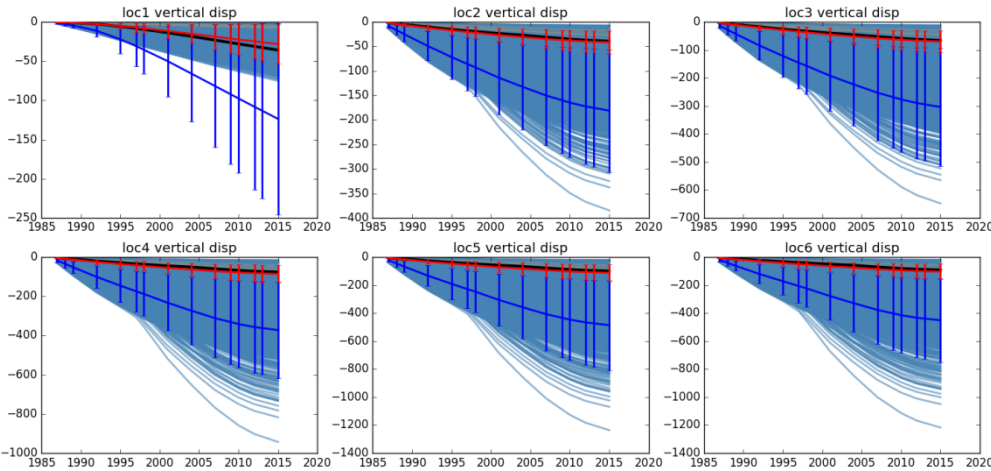


Figure 40: Subsidence on some locations, high density data, only 'wrong pressure members'. Locations of the benchmarks are indicated in Figure 31.

In summary, the tests here demonstrated that for a model with significant subsidence above the aquifer, a clear distinction can be made between different aquifer depletion scenarios with both a low and high data density above the gas column (away from the aquifer). The distinction can be made both based on the best models, and on the sum of all weights over all the members for each pressure scenario (see Figure 33 and Figure 35). When the input data was generated to the extreme case where little compaction is expected, and correspondingly little difference in subsidence with or without aquifer depletion, the members that gained highest weight after data confrontation still belonged to the set of members using the correct pressure depletion model, and a visual distinction is possible. Based on the average weights, however, the discrimination disappears for the high data density case. This is because besides the high weight members in the high depletion scenario, there are also scenarios (with high compaction) that have a much lower weight than that are present in the low aquifer depletion case, pulling the average down. With the low data density, there is still a small distinction possible based on the average weights.

7.6.6 Summary of ESIP testing

A range of tests were devised to demonstrate the ESIP workflow and to get acquainted with the output. The tests described in Section 7.6.1 and Section 7.6.2 compare the upscaling, interpolation and subsidence calculations that are performed in the ESIP workflow with those from methods used by NAM, to verify differences in methods are negligible. Section 7.6.3 demonstrates that the output of the complex piece of ESIP software does not change when used by different individuals or on different machines. Section 7.6.4 compares the output, especially the model weights after confrontation, with a known input of synthetic data, to on one side demonstrate that the known input parameters fall indeed within the ranges where model weights are higher, and on the other hand to help create an understanding for how to interpret the output. In the last test, Section 7.6.5, we verify that, through the incorporation of the covariance matrix, areas with a low density of subsidence data, such as those in the Waddenzee, do not incite a particular favour towards certain pressure scenarios that give a good fit above the island (with high data density), leaving little distinction about the aquifer depletion. The test demonstrated that, even for an extreme case, distinctions between pressure scenarios can be made based on the weight distribution.

7.7 Comparison of AEsups with FE element numerical simulations: testing the influence of salt flow.

To demonstrate that the analytical solutions of AEsups are able to approach the results that can be obtained from a numerical FE element simulator, we use the Ameland FE model as a reference. This model is described in NAM (2011) and can be regarded as the state-of-the-art geomechanical model for Ameland to provide the subsidence volume contribution caused by the Ameland production in the measurement and control as part of the “Hand aan de Kraan” procedure. The model contains a grid that represents the geology of the Ameland area.

The linear viscous salt behavior in AEsups was tested in the past and results (Figure 41) were published by Fokker and Orlic (2006). In the LTS II report we will demonstrate that AEsups is also able to produce solutions similar to FE models based on the Ameland area.

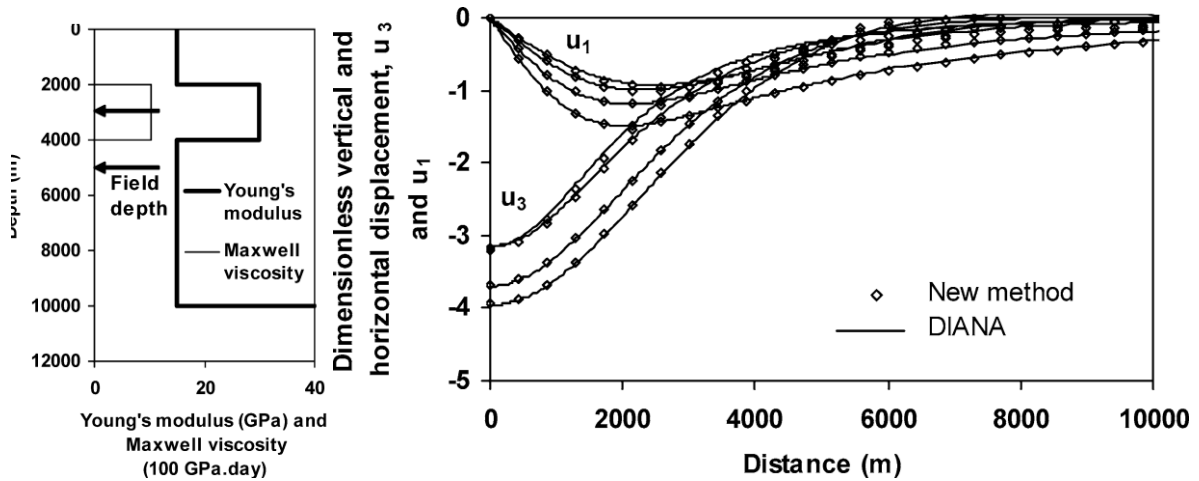


Figure 41: Validation of the AEsubs for subsidence prediction with the finite element simulator DIANA, using the visco-elastic profile presented in the left figure and the center of compression at a depth of 3,000 m. Profiles are shown for zero time, 95 days, 440 days, and 950 days. Vertical displacements are increasing with time; horizontal displacements are decreasing (from Fokker and Orlic, 2006).

We selected one reservoir scenario for making the comparison of the results obtained for AEsubs and the FE model. The particular model has a good quality history match. The model's pressure depletion profile in the year 2016 is plotted in Figure 42. Since we are interested in comparing the subsidence profiles from the different approaches and not the compaction that results from different compaction models, we only consider one compaction model. The model used is a time decay compaction model with a C_m multiplier of 1.2 and a time-constant of 7 years, i.e. what is used in the most recent measurement and control cycle publication (NAM, 2016a).

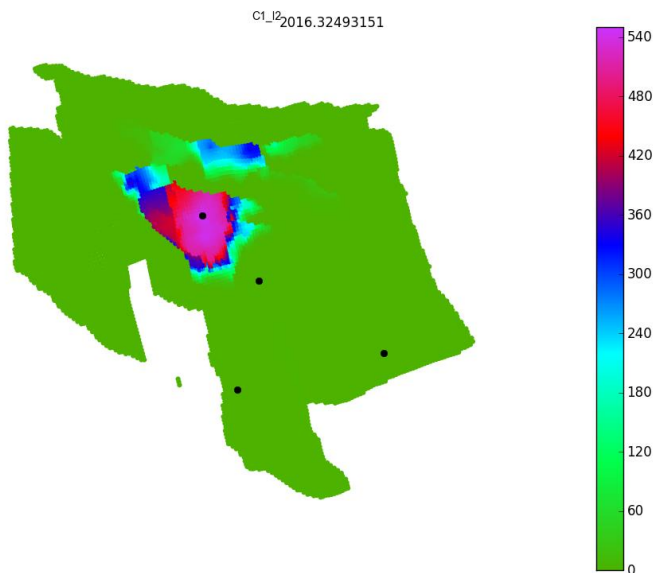


Figure 42: Pressure depletion in 2016 according to the model used for the AEsubs vs Geomec comparison.

The main geomechanical layers of the FE model are (NAM, 2011) listed in Table 4. This table presents the average value taken from the FE model. The Maxwell viscosity in this table is equal to the creep factor used in Geomec (pers. Comm. P. Fokker (TNO) and P. Fokker (Shell)). In Geomec vertical strain rate is defined by:

$$\dot{\epsilon}_{33} = D(\sigma_{33} - \sigma_{11})^n \quad (23)$$

With n being 1 for linear viscosity and for the condition that radial strain (ε_{11} and ε_{22}) is zero and $\sigma_{11} = \sigma_{22}$.

Maxwell viscosity is defined by:

$$2\mu\dot{\varepsilon}_{ij} = \frac{\mu}{\eta} \left(\sigma_{ij} - \frac{1}{3} \sigma_{kk} \delta_{ij} \right) \text{ and for } ij = 33 \quad (24)$$

$$\dot{\varepsilon}_{33} = \frac{1}{2\eta} \left(\sigma_{33} - \frac{1}{3} (\sigma_{11} + \sigma_{22} + \sigma_{33}) \right) \text{ and } \sigma_{11} = \sigma_{22} \quad (25)$$

$$\dot{\varepsilon}_{33} = \frac{1}{3\eta} (\sigma_{33} - \sigma_{11}), \quad (26)$$

$$\text{Therefore } D = \frac{1}{3\eta} \quad (27)$$

$$\text{In Geomec: } D = A \cdot \exp \left(-\frac{Q}{RT} \right) \quad (28)$$

with A the “creep factor” [$\text{MPa}^{-1}\text{s}^{-1}$], Q the activation energy [J mol^{-1}], R the gas constant [$\text{J mol}^{-1}\text{K}^{-1}$] and T the temperature [K].

Table 4 Average values of the finite element model

Layer name	Top layer [m]	Mean E [GPa]	Mean ν	Viscosity PaS
North Sea	0	2	0.3	-
Chalk	1000	10	0.25	-
Cretaceous/Jurassic	1500	16	0.25	-
Zechstein	2000	30	0.35	$6.7 \cdot 10^{17}$
Ten Boer Claystone	3200	40	0.2	-
Rotliegendes Sandstone (reservoir)	3300	8	0.2	-
Limburg	3410	40	0.2	-

AEsubs makes it possible to include the effects of salt viscosity for subsidence calculations without having to use finite element models, however it does not allow for a combination of a viscous layer together with large contrasts and variability in layer stiffness. To make a comparison to the Geomec simulations, we used a simplified four-layer model in AEsubs, of which the properties are shown in Table .

In order to include the effect of the viscous layer over time, AEsubs generates a range of influence functions that together describe the evolution of the surface motion resulting from a subsurface strain event over time. Rather than only having an instantaneous effect on the surface deformation, a subsurface strain event now continues to deform the surface for a certain amount of time, given by the convolution over time of the subsurface strain with the time dependent influence function.

In total, AEsubs generates the influence function at 21 different time steps. Figure 43 shows the influence function at three moments in time (using the model parameters in Table 1Table 5). The first is the instantaneous influence function ($t=0$) and behaves in the same fashion as when no viscous layer is included. The other two curves show the surface deformation resulting 20 years and 2500 years after a reservoir strain occurred.

Note that the deepest point, at $x=0$, increases over time, as expected for a viscous layer, and that the subsidence volume is (within the analytical precision) preserved. The shape of the influence functions is mainly controlled by the combination of Young's moduli, Poisson's ratios and depths of the various layers used, whereas the viscosity controls the time at which these influence functions take effect. The final influence function in this case, for example, comes into effect at a time 2500 years after a subsurface strain happens. When increasing the viscosity by a factor of two, the time of all the influence functions also increase by a factor of two, while leaving the shape unchanged.

Adding more than four layers, in an attempt a better resemble the variety that is present in the Geomec model, resulted in unrealistic influence functions, where the subsidence volume is no longer preserved. Among these layers, the contrast in especially the Young's modulus cannot be too large, and will be smaller than the contrasts in the Geomec simulations. A further deviation from the Geomec model is the depth of the reservoir. As a rule of thumb, AEsups requires the separation between interfaces to be more than 10-15% of their depth, when the layer properties change significantly (pers. comm. P. Fokker, TNO). Hence, in order to maintain the resemblance to the salt layer thickness in the comparison here, the reservoir cannot be closer than 300-400m from the top of Layer 3. To guarantee stable and realistic influence functions, the reservoir is placed at a depth of 3600m. An example of unrealistic influence functions is plotted in Figure 44, generated when moving the reservoir up to 3300m.

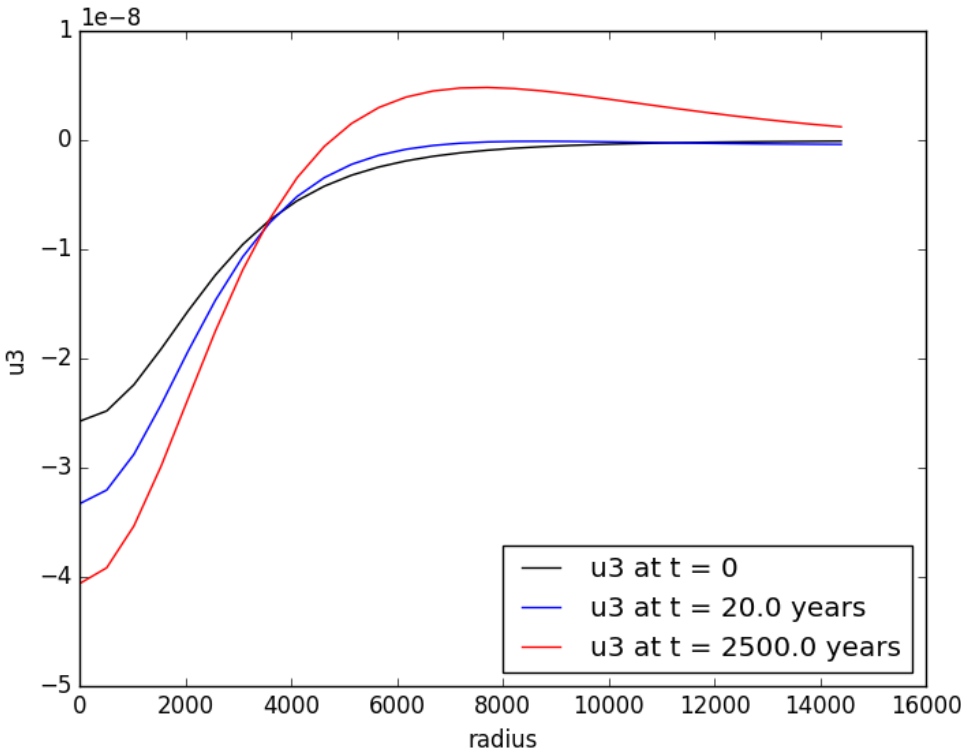


Figure 43: Vertical influence functions for the model parameters as in Table , plotted for three different times after strain nucleation.

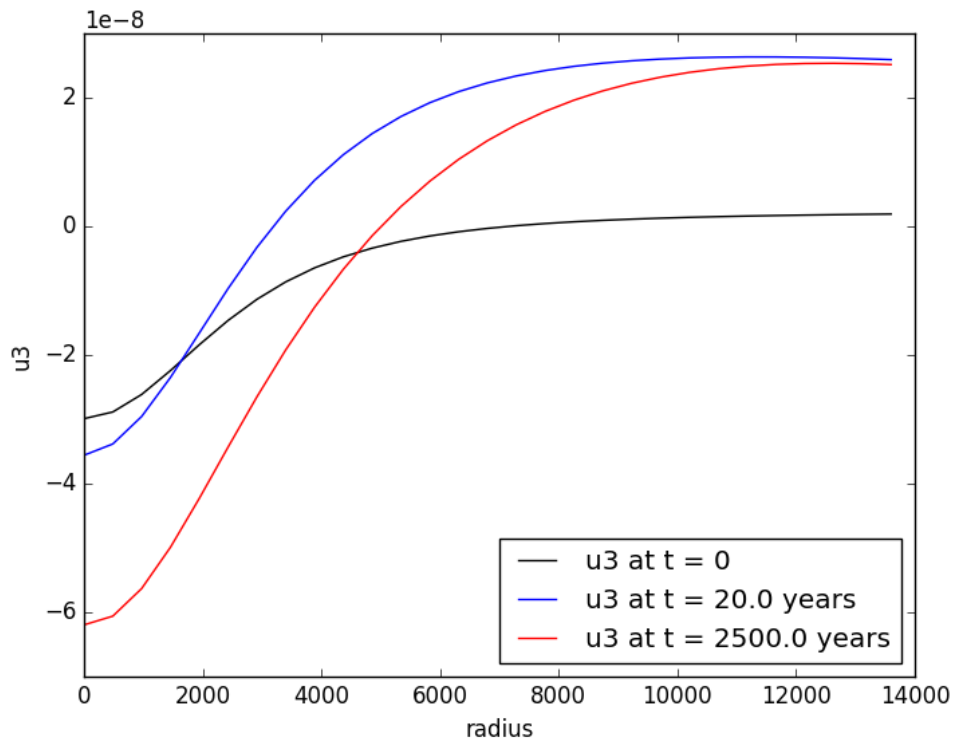


Figure 44: Example of unrealistic influence functions, generated when interfaces are placed too close to one another. In the case plotted here, the reservoir was placed at 3300m depth, compared to the 3600m in Table . The subsidence volume is not preserved over time, and there is permanent deformation still at a distance 14km away from the centre.

Table 5 Layer properties as used in the AEs subs calculation.

Layer	Top depth (m)	Young's modulus (GPa)	Poisson's ratio	Viscosity (Pa s)
1	0	8	0.25	-
2	2000	8	0.35	7×10^{17}
3	3100	20	0.25	-
4	4200	20	0.25	-

For the comparison to Geomec, we will use the parameters from Table 5. The depths of the viscous layer (Layer 2) is comparable to the average top of the Zechstein in the Geomec model, and the top of Layer 3 is similar to that of the top of the Ten Boer. The properties of the four layers are kept as close to those in Geomec at similar depths where possible.

Figure 46 compares the surface deformation in 2016 across the subsidence bowl as calculated by AEs subs and Geomec. Clearly, the deepest point in the subsidence is different between the AEs subs model (with Cm factor of 1.2) and the output from Geomec.

Two significant differences between the setup of the AEs subs and Geomec models, however, can account for such a difference. One of these is the reservoir depth, which, as discussed, is at an average depth of 3300m in Geomec and limited to 3600m in AEs subs because a shallower position of the reservoir will lead to unstable results for the subsidence. The second significant difference is the

lack of significant contrast in layer stiffness in the AEsups model. Both of these will result in a deeper subsidence bowl when using the properties in the Geomec model.

To get a quantitative handle on the contribution of these differences on the depth of the subsidence bowl, we used AEsups to generate influence functions for different reservoir depths and properties. We did this for four models with parameters as indicated in Table 6. Model 1 uses layer properties that still produce a reliable influence function when adding viscous properties to Layer 2, and is the model used for comparison to Geomec. Model 2 is the same as Model 1 apart from the reservoir depth, so will provide an illustration of the effect of the reservoir depth limitation. Model 3 has the same reservoir depth as Model 1, but has a higher contrast of layer stiffness, and a higher stiffness below the reservoir than is possible when also using a viscous layer. Model 4 combines the two alterations, having both a shallower reservoir and higher stiffness contrast than Model 1.

Table 6: Model parameters used in AEsups to verify the effect of the limitations on the depth of the subsidence bowl. Property values that deviate from those in Model 1 are highlighted in bold red.

		Layer 1	Layer 2	Layer 3	Layer 4	Reservoir
Model 1	Top depth (m)	0	2000	3100	4200	3600
	Poisson's ratio	0.25	0.35	0.25	0.25	
	Young's modulus (GPa)	8	8	20	20	
Model 2	Top depth (m)	0	2000	3100	4200	3350
	Poisson's ratio	0.25	0.35	0.25	0.25	
	Young's modulus (GPa)	8	8	20	20	
Model 3	Top depth (m)	0	2000	3100	4200	3600
	Poisson's ratio	0.25	0.35	0.25	0.25	
	Young's modulus (GPa)	5	8	30	40	
Model 4	Top depth (m)	0	2000	3100	4200	3350
	Poisson's ratio	0.25	0.35	0.25	0.25	
	Young's modulus (GPa)	5	8	30	40	

Differences in influence functions from contrast and reservoir depth are shown in Figure 45. Both the difference in reservoir and stiffness contrast can individually account for a difference of more than 20% in the deepest point of the influence function, and when combined (Model 4) this is more than 50%. In order to make a fair comparison between the AEsups and Geomec models, which is the aim of the exercise here, it is justified to increase the Cm factor from 1.2 to 1.8 (a 50% increase) to compensate for the limitations encountered in AEsups with a viscous layer.

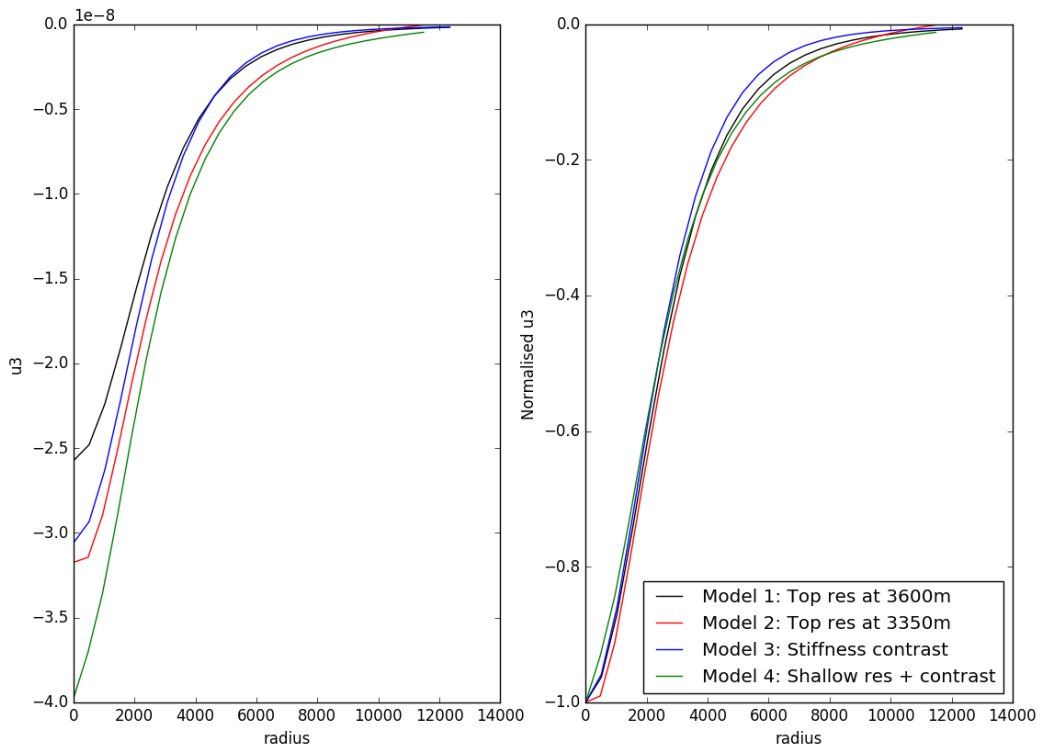


Figure 45: Influence functions (u_3 being vertical) generated by AEsubs for the models in Table 6. Left graph shows the nucleus of strain, right graph the normalised values to allow comparison between the shape of the influence functions.

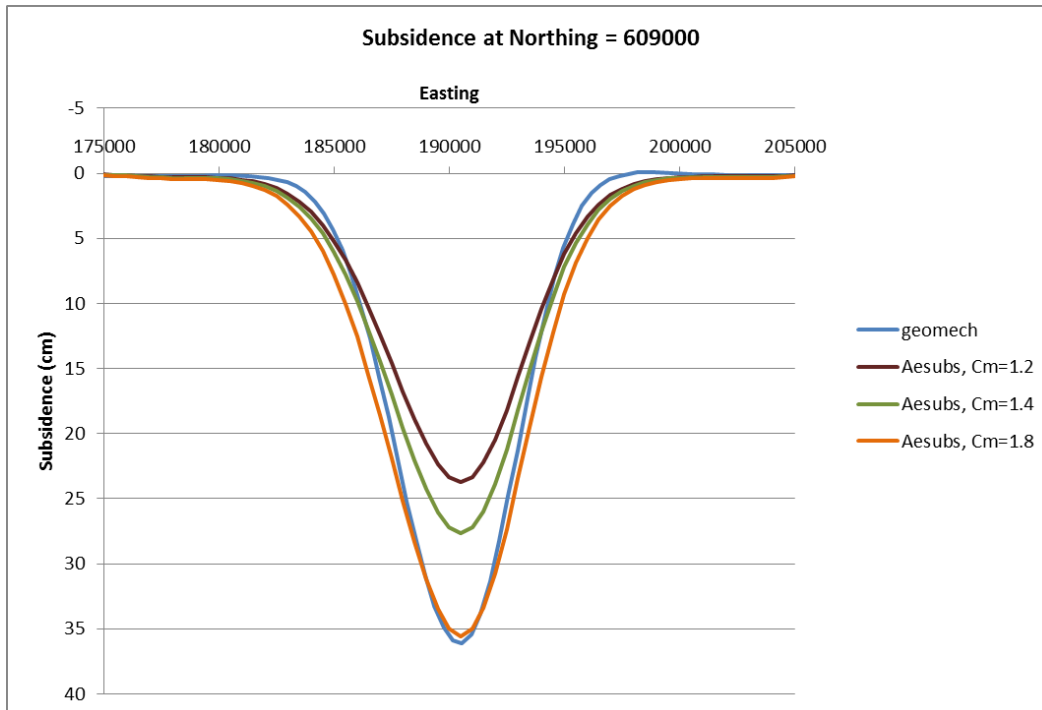


Figure 46: Comparison of the calculated subsidence in 2016 through a cross section through the deepest point in the subsidence bowl. The salt viscosity used in AEsubs is $7e17$ PaS similar to the Geomec salt viscosity. To compensate for the lack of stiffness contrast and for the limitation to the depth of the reservoir, the C_m factor in the Time decay model was increased by 50% from the original $C_m=1.2$, to $C_m=1.8$.

The results from AEsups with an increased C_m factor are included in Figure 46, which demonstrates that, when using the C_m value to compensate for model limitations, the depth of the subsidence bowl from AEsups using the parameters in Table 6 become almost identical to that resulting from the more complex Geomec modelling. The evolution of the deepest point of the subsidence bowl over time is plotted in Figure 47, while the contours at several moments in time are plotted in Figure 48. The results from both AEsups and Geomec show good agreement in the development of the maximum subsidence over time, and similarly the shape of the bowl shape shows good resemblance, apart from some minor offsets in the location of some contour lines. A more detailed contour plot for the subsidence in the year 2016 is shown in Figure 49.

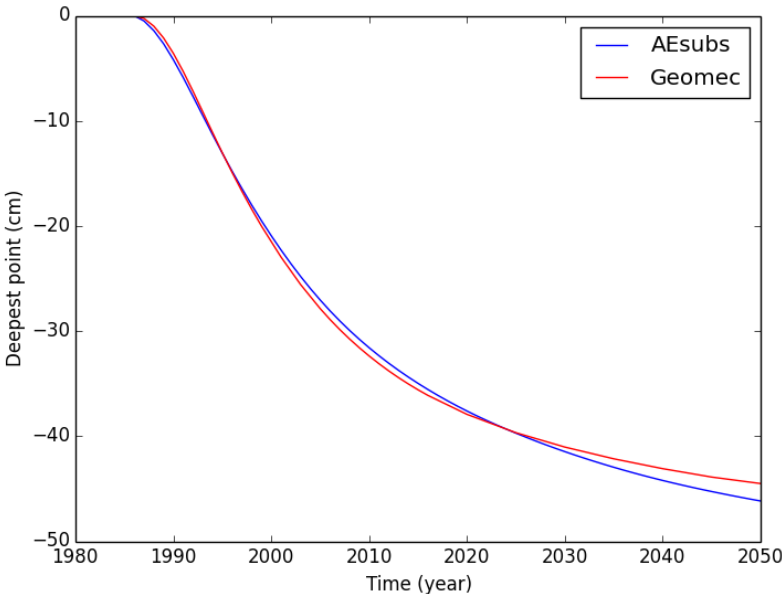


Figure 47: Evolution of the deepest point in the subsidence bowl from Geomec and AEsups (AEsups with C_m factor of 1.8).

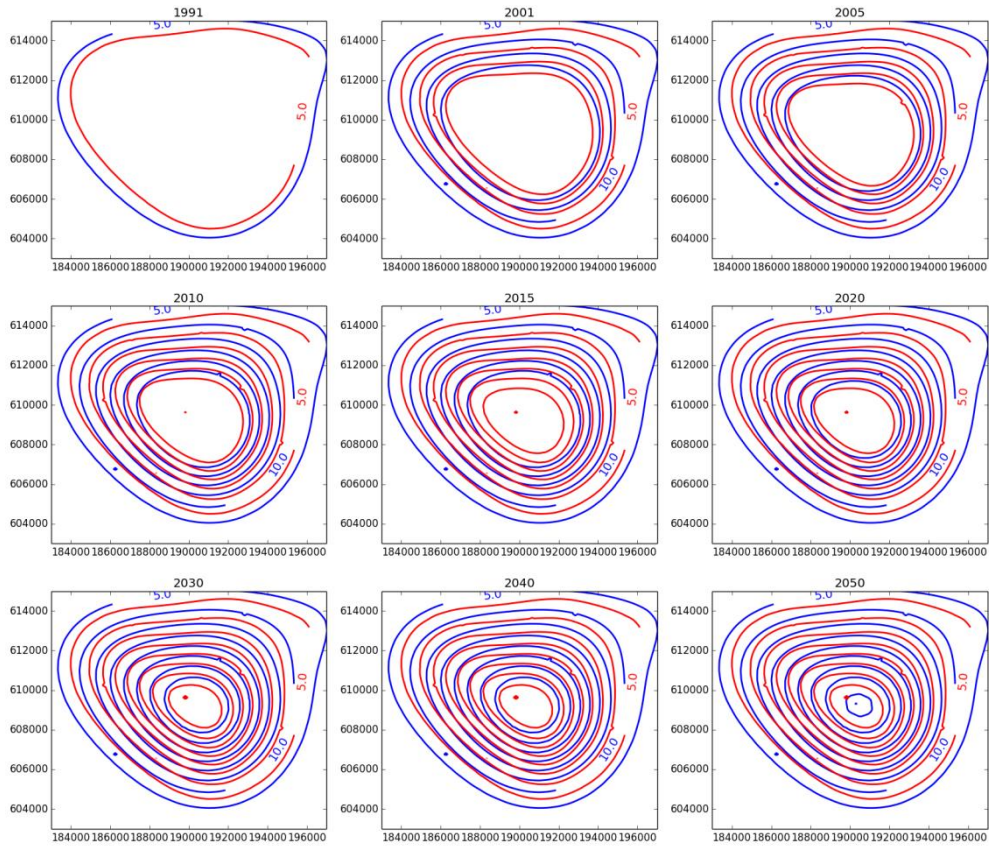


Figure 48: Subsidence contours through time, for Geomec (red) and AEsubs (blue).

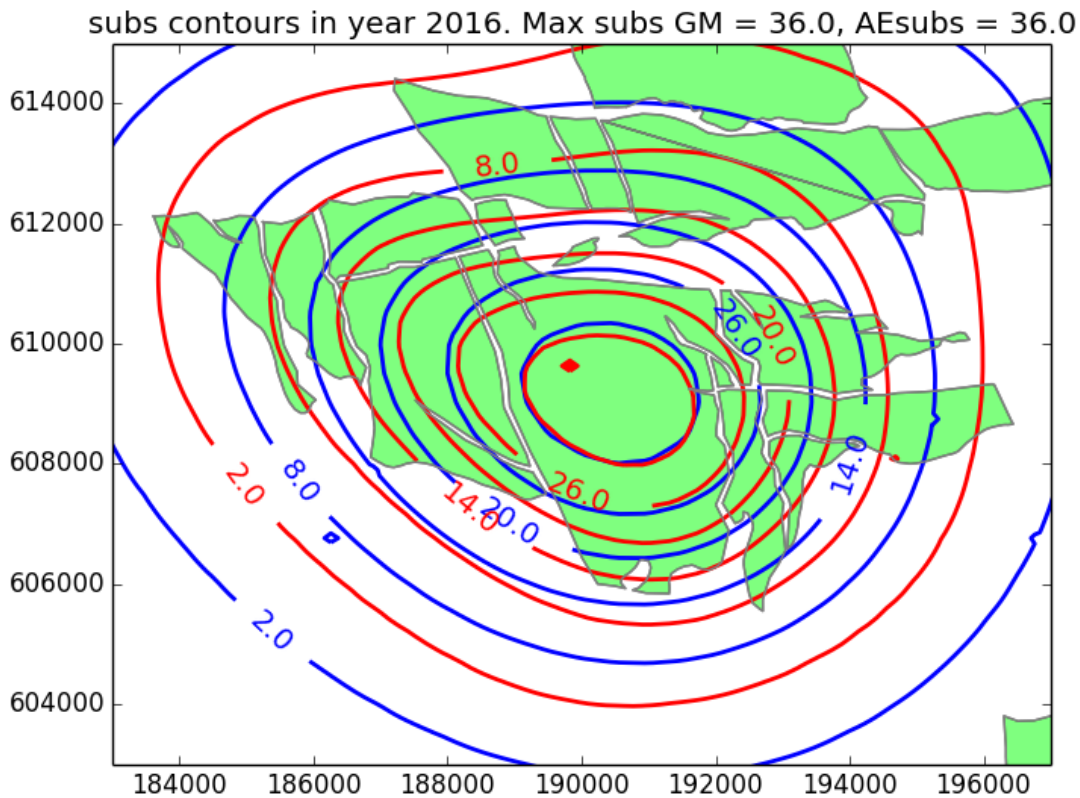


Figure 49: Detailed subsidence contours above the Ameland field, generated by Geomec (red) and AEsups (blue).

The shape does show a reasonable match keeping in mind the simplified layering in AEsups compared to Geomec. The contour pictures through time in Figure 48 demonstrate as well that the domal shape of the Zechstein salt has a negligible influence on the position of the bowl in time, a topic that will be discussed in more detail in paragraph 7.8.

The results from running the AEsups calculation using parameters derived from the Geomec model show that both methods produce very similar subsidence predictions under similar circumstances. The resemblance between the two outcomes indicates that the AEsups calculations will be able to generate subsidence bowls of similar shape to Geomec, after including a viscous salt layer. The results here prove that the ESIP workflow can produce models that have a steep, deep bowl shape like the Ameland subsidence bowl.

As an additional test, we improve the resemblance one step further by slightly adjusting the viscosity in AEsups, while remaining well within the acceptable viscosity values (and thereby also within the distribution used in the future viscosity distributions in the ESIP workflow). The resulting subsidence across the subsidence bowl for the year 2016 are shown in Figure 50 and Figure 51 for a salt viscosity ranging from 5×10^{16} Pa s to 7×10^{17} Pa s, the latter being the one that is comparable to what is used in the Geomec calculations above. Judging from Figure 51, which shows the subsidence normalized to its deepest point, the steepness of the bowl as obtained by the Geomec model is reproduced using AEsups with a viscosity within the range mentioned above. Models with a salt viscosity exceeding 3×10^{17} Pa s, in combination with the other layer properties generate a slope that is slightly less steep, whereas a viscosity lower than 3×10^{17} Pa s results in a very similar gradient of the flanks of the bowl

when compared to Geomec. These models however produce a bulge at the edges of the subsidence bowl that should be regarded as a model artefact because heave has not been observed by the geodetic measurements.

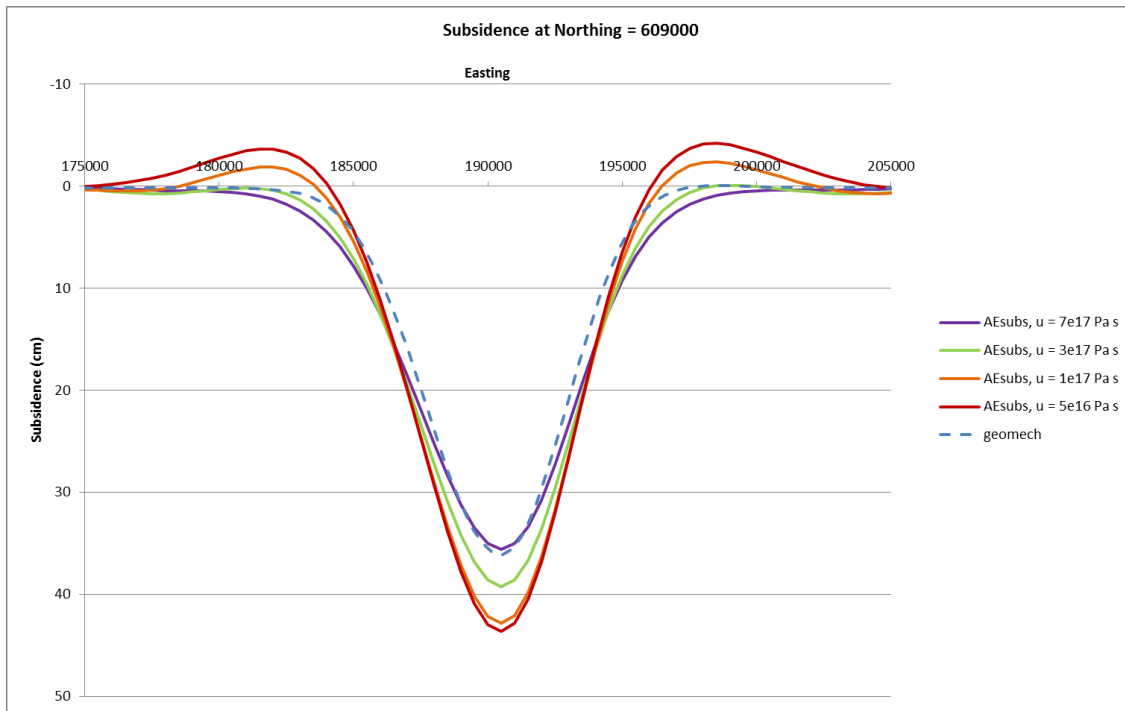


Figure 50: Calculated subsidence in 2016 across the subsidence bowl for a range of viscosity values (Cm factor is kept at 1.8).

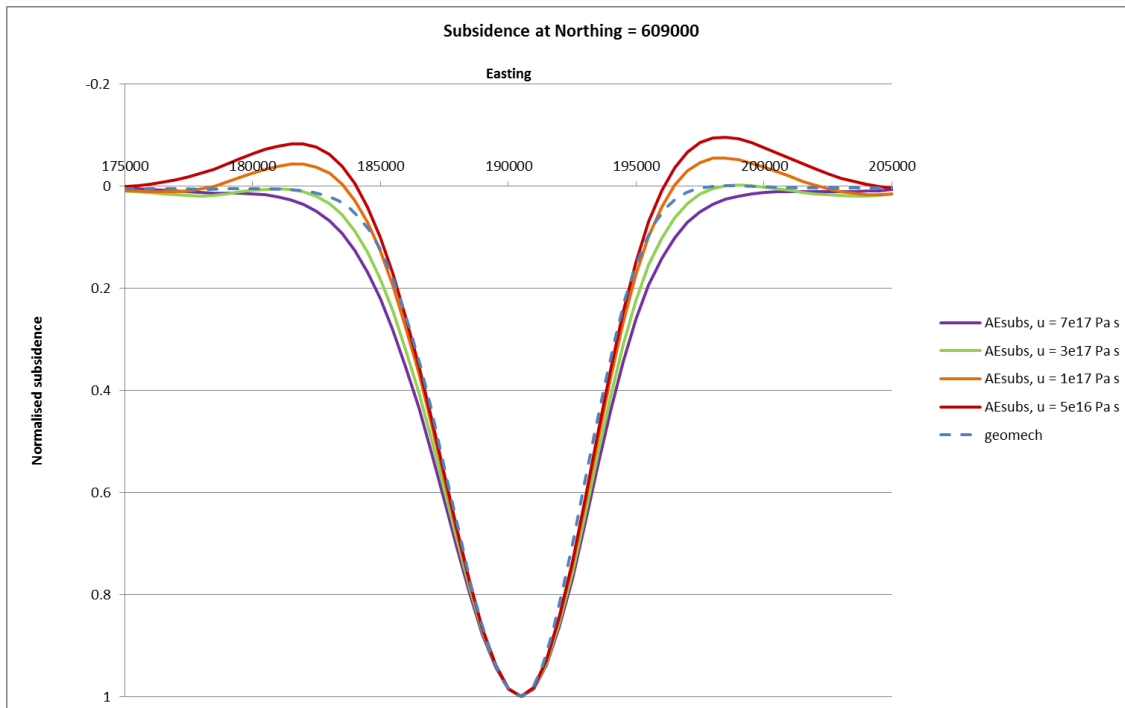


Figure 51: Same as Figure 50 but now showing the subsidence normalized to the deepest point, to allow better comparison of the shape of the bowl.

In summary, we demonstrated here that the subsidence as calculated by the complex Geomec model can be approached with a much simpler AEsubs calculation, even within the restrictions encountered by the inclusion of a viscous layer in AEsubs. When using Geomec-derived parameters, and altering the C_m value to compensate for the restrictions imposed on the reservoir depth and stiffness contrast in AEsubs, we obtain subsidence results very much alike those in Geomec in both the temporal and spatial response. By changing the viscosity and C_m by a small fraction well within the parameter uncertainties, we can obtain an even closer resemblance to the Geomec output. These results demonstrate that, within a reasonable range of parameters around those of their counterparts in the Geomec model, a subsidence bowl very similar in both time and shape are generated within the ESIP workflow.

7.8 The effect of the Ameland salt dome on the position the point of maximum subsidence with time.

A concern was raised during a meeting on 24-11-2016 with the steering committee that the structure of the salt dome above the Ameland would have a significant impact on the subsidence bowl position with time, which cannot be addressed by the layer cake model in AEsubs. If this would be the case, the chosen semi-analytical approach could be regarded as not adequate enough. These concerns were based on conclusions made in the report by TNO (TNO, 2011).

A further investigation with the Ameland Geomec model demonstrates however that a finite element model scenario that allows for viscous flow of the salt will produce similar results for the position of the deepest point of the subsidence model through time as a scenario with only an elastic behavior. Figure 52 shows the results of the two scenarios for the simulated year 2016 and Figure 53 shows the position of the point of maximum subsidence through time. The figure demonstrates that the location of the deepest points of the two scenarios are similar. Still, the bowl shape and therefore also the value for the maximum subsidence is affected by the deformation of the viscous salt resulting in narrower and steeper subsidence bowl when compared to a full elastic model. Depletion increases to the NW giving an explanation for the shift in the deepest of the subsidence bowl with time.

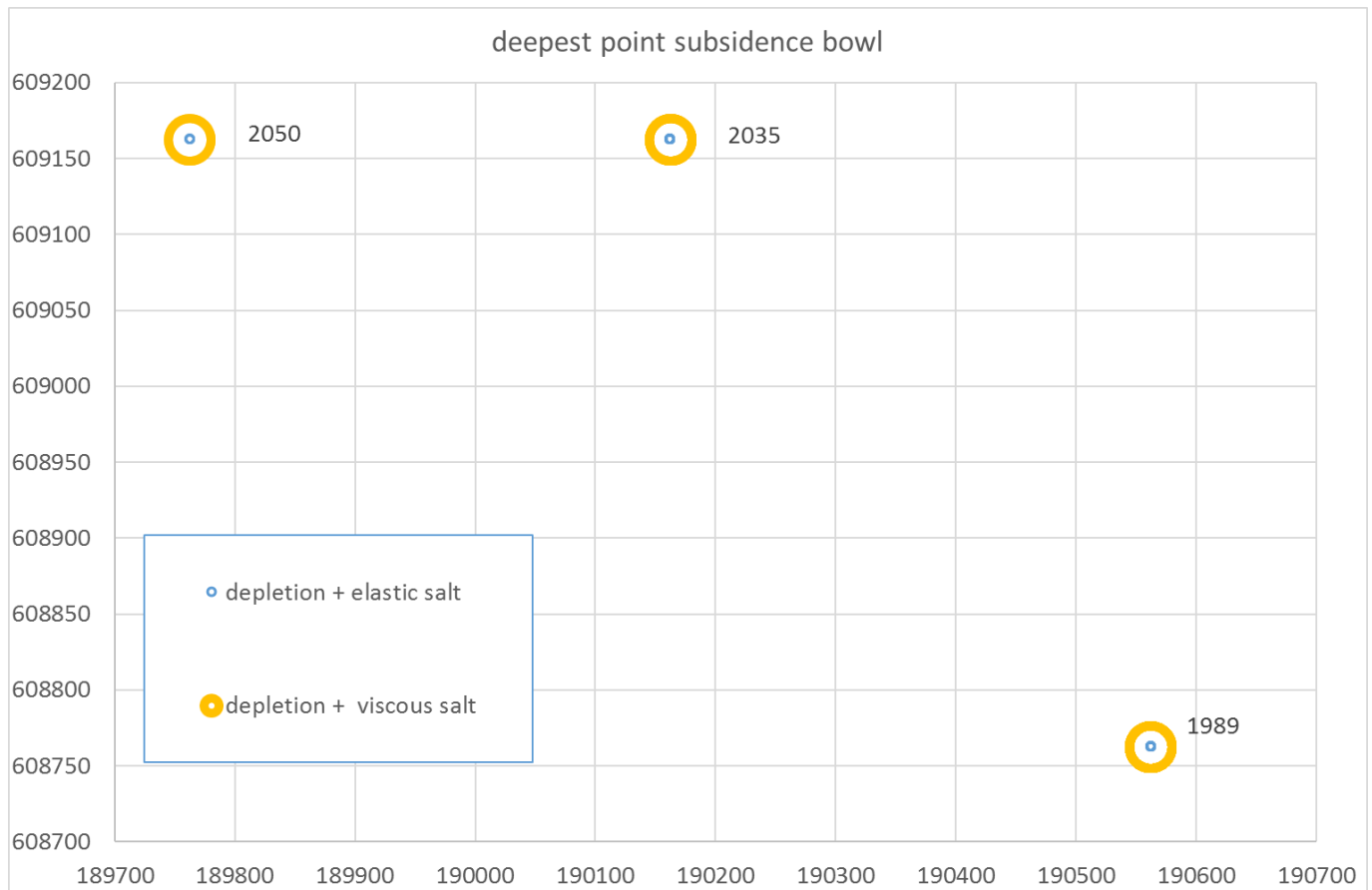


Figure 53: Location of the deepest points with time. Comparison between a full elastic model and a model with a viscous behaviour for the salt layer.

An explanation of the approximate shift of 1 km to the west and 400m to the north can be found in the development of the pressures with time. Figure 54 shows the area that is affected by the pressure evolution.

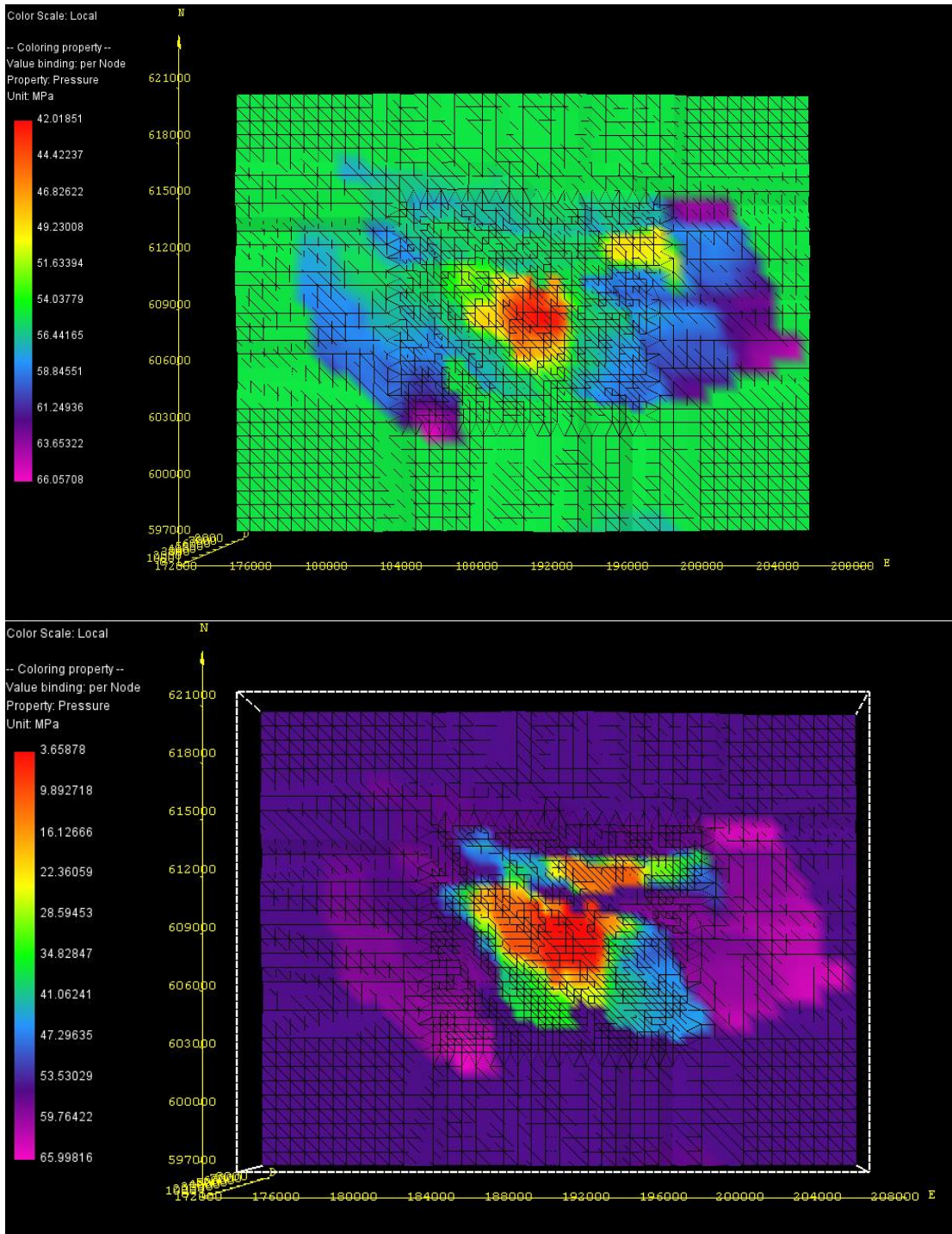


Figure 54: Depletion maps from the history matched Mores model. Top picture shows the pressure for 1992 while the bottom picture shows the results for the year 2050.

8 Results of the confrontation

This chapter describes the results of the confrontation of the selected reservoir scenarios and geomechanical models and members with the geodetic data. Based on discussions with the steering committee it has been decided to present the results in a format as close as possible to the results presented in the measurement and control documents (e.g. NAM 2016a) for the Waddenzee area. Rather than presenting subsidence bowls and profiles, average subsidence rates will be presented generated from calculated subsidence volumes. Each model member will give a forecast and from the confrontation with the geodetic data we can infer the likelihood of that specific member depending on the mismatch with the data. Knowing the likelihood, a cumulative probability density plot can be made and P50 and confidence bounds can be inferred from it. These values provide a description of the quality of model predictions.

The layout of this chapter is structured around the different influence functions (AEsubs, Knothe) and different assumptions for the behavior of the overburden (elastic vs viscous salt). The first part of this chapter is a description of the used geodetic data, reservoir scenarios and geomechanical parameters in these calculations.

8.1 Selection of benchmarks

The benchmarks selected for the confrontation between modelled and measured double differences is based on a polygon which covers the influence area of the Ameland subsidence bowl. The geodetic data that can be affected by the production from the Nes field (SE corner) have been excluded (Figure 55). All red dots in the figure contain data that is being processed by the CUPiDO tool in ESIP.

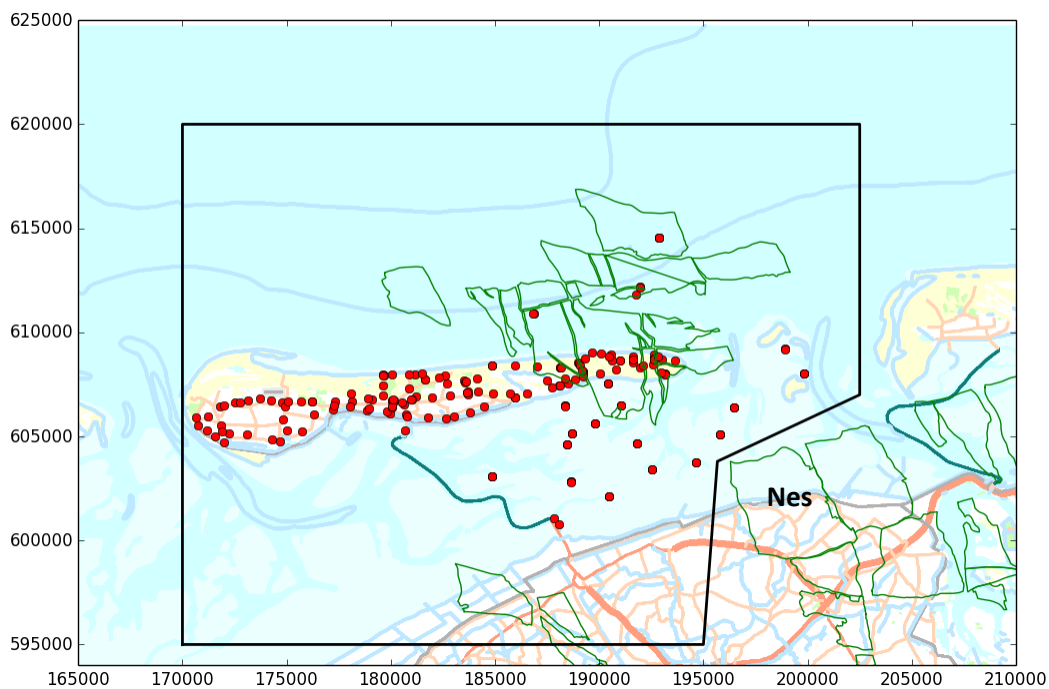


Figure 55: Selected benchmarks with in the selection polygon (black line).

8.2 Selection of reservoir scenarios

Around 200 reservoir scenarios resulted from the automated history match workflow with Mores. The main parameters that were varied are fault transmissivity, permeability, and residual gas. A scenario selection procedure was required to limit the amount of scenario's and as well calculation time. This procedure is acceptable because many scenarios are more or less alike. The selection of scenarios is mainly based on the transmissivity of the faults because this parameter determines which block is able to deplete and can cause subsidence. For each fault configuration, scenarios were selected that represent the depletion ranges within that configuration. This procedure resulted in the selection of 13 reservoir scenarios that are a good representation of the total temporal and spatial depletion span given by the total set of 193 scenarios. The selected models are shown in Figure 56.

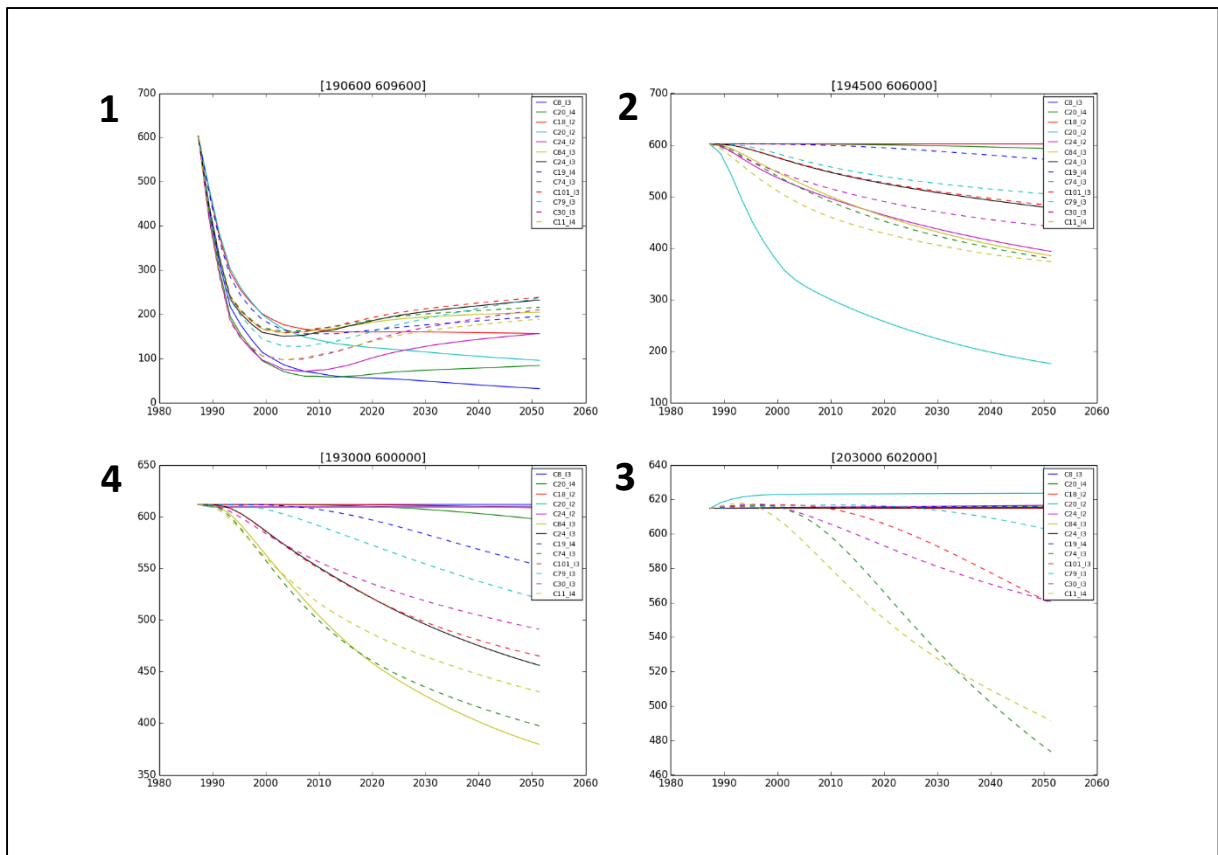
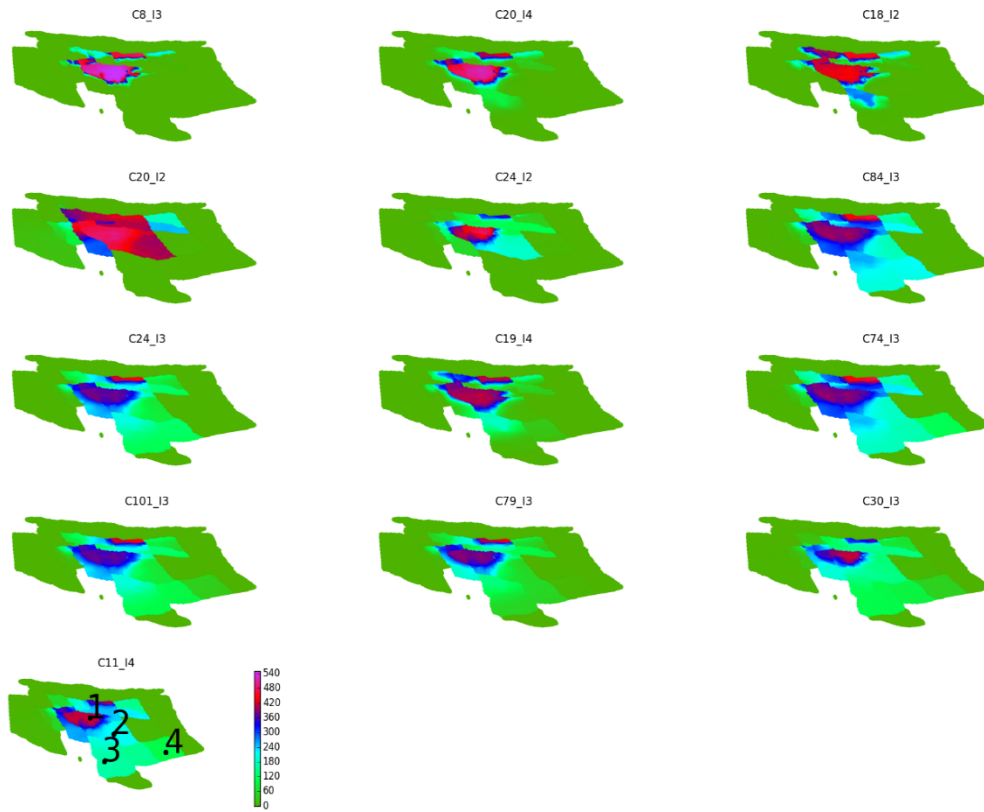


Figure 56: Selected reservoir models for the ESIP calculation The top figure shows the lateral depletion for the year 2040, the bottom the figure show the pressure vs time for the four points shown in the top figure (last plot).

8.3 Selection of area of interest for subsidence rate calculation

As mentioned in the introduction of this chapter the output is presented as subsidence rate. The subsidence rate is calculated from the modelled subsidence volume generated per year in a specific area of interest and divided by the surface area. For all the models the subsidence rate vs. time is calculated for the selected area presented in Figure 57.

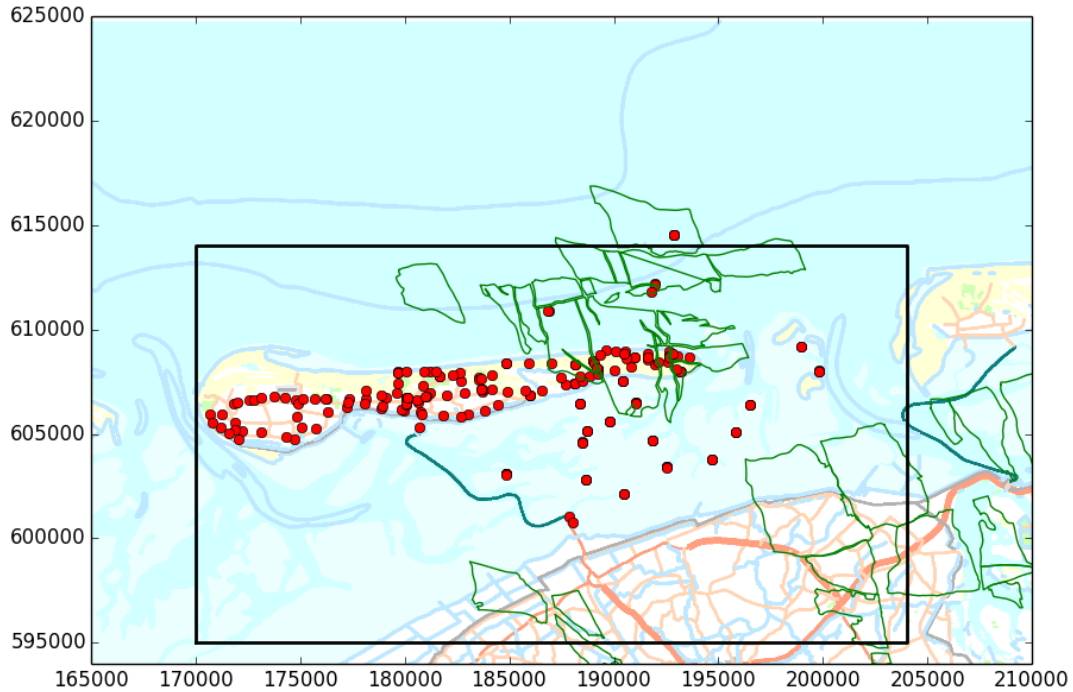


Figure 57: Area for which the subsidence rate is calculated.

Rather than presenting the subsidence rate for each year, also here we decided to stay as close as possible to the calculation schemes as defined for the Waddensee. This calculation scheme provides a 6 years moving average. The average rate (rate (Y)) for each year is calculated from subsidence rates 3 years before and 2 years after that specific year.

$$\text{rate (Y)} = [S(Y-3) + S(Y-2) + S(Y-1) + S(Y) + S(Y+1) + S(Y+2)] / 6 \text{ [mm/year]}.$$

For all the members the goodness of fit (χ^2/N) and the probability of each member has been calculated. The expectation weighted model case has been calculated by:

$$R_{exp} = \sum_{i=0}^{n_{member}} P_i R_i$$

Where R is the rate per member, P the probability of that member and R_{exp} is the weighted model rate. Note that the sum of the probabilities is 1. An example of the rate distribution is shown in Figure 58.

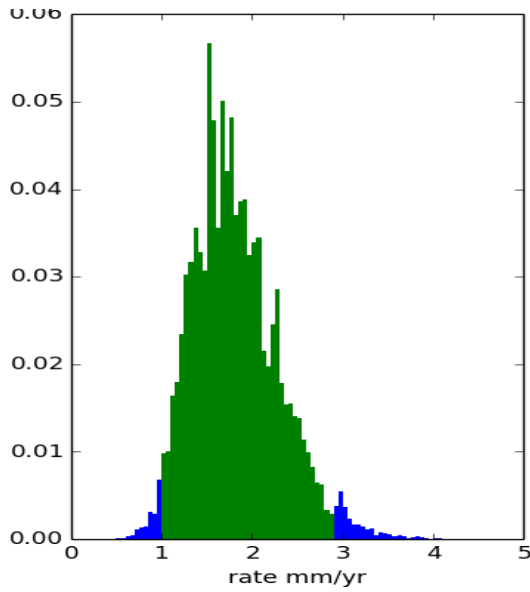


Figure 58: Example of the PDF for the subsidence rate, the green area shows the area within the 2.5 and 97.5% probability.

8.4 Selection of reservoir and overburden parameters

The elastic overburden and its properties is compiled following the Ameland Geomec model, which contains 7 layers with different stiffness properties. The Geomec stiffness values are used to define the mean of the input distributions of which the ranges are presented in Table 7. The range for the stiffness of the reservoir rock is defined by the measured laboratory data from Ameland. First cycle tests on core plugs show a stiffness that correlates with porosity using a fixed value of 0.2 for the Poisson's ratio to convert the measured uni-axial compressibility values to a Young's modulus (Figure 59a). A better approach for getting values for both Young's modulus and Poisson's ratio is to invert to these parameters via a numerical model using the available stress-strain curves. Results from this inversion are presented in Figure 59b.

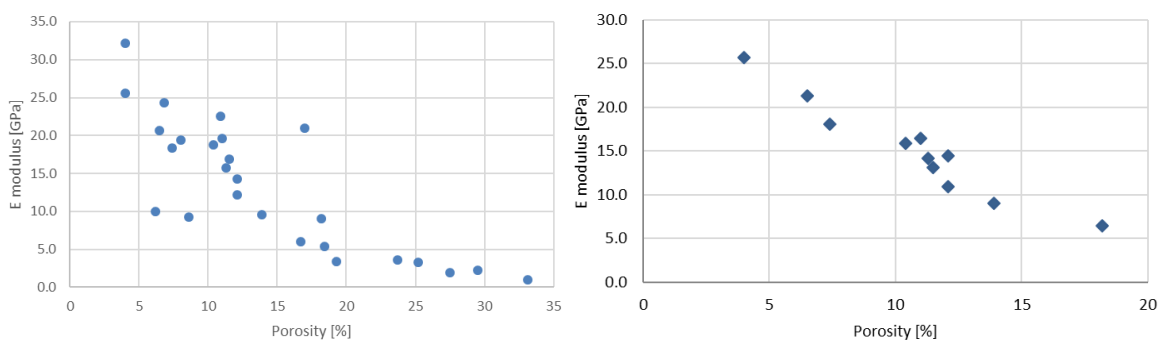


Figure 59: a) Estimated E modulus for all Ameland core plug data b) Inverted elasticity data from available stress strain curves.

From Figure 59b a relation between the porosity φ and C_m has been derived which has been used in ESIP. This equation of this regression line is:

$$C_m = 754 \cdot \varphi^3 - 142.8 \cdot \varphi^2 + 10.1 \cdot \varphi + 0.18$$

This line is shown in Figure 60 as the blue solid line. The range for the C_m values in ESIP is based on all available core plug experiments from the Rotliegendes formation in the North Netherlands, (Figure 60, grey dots). The dashed lines are the result of multiplication factors (C_m factor) of 0.3 and 4 to the C_m, ϕ relation (blue solid line).

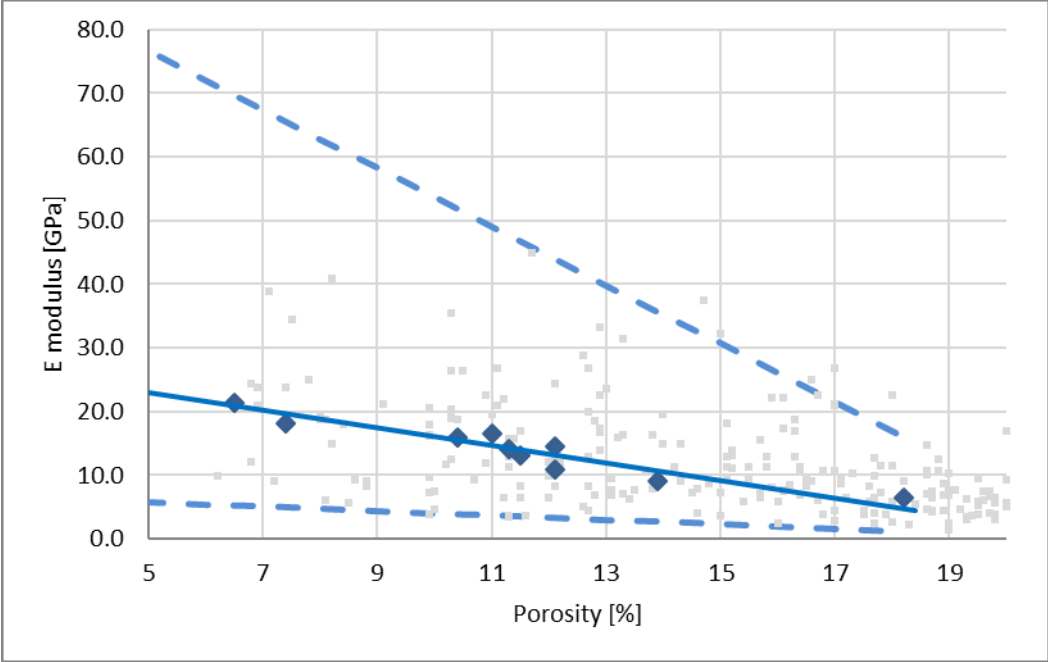


Figure 60: Ameland (blue dots) and other Rotliegendes Elasticity Moduli (grey dots) (converted from C_m core plug experiments).

The upscaled porosities in ESIP range from 8% to 17% (Figure 61).

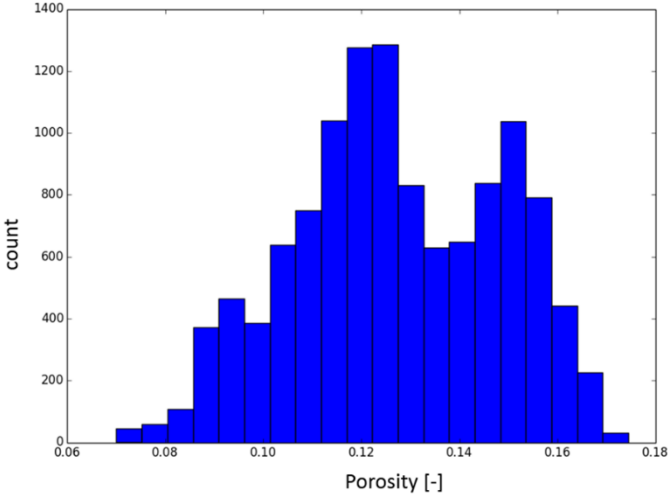


Figure 61: Upscaled porosities from the Mores model.

There are no NAM or Shell laboratory data available for geomechanical units other than the reservoir. Stiffness properties for the other layers are based on sonic data. The direct dynamic values are used in the derivation of the stiffness profile in Geomec that has been used to provide the mean

values in Table 7. The dynamic values of the stiffness are closest to an undrained stiffness that best represents the mechanical behaviour of the rock regarding the low permeability of the rocks and the small strains and stresses that are induced by the compaction of the reservoir on these over- and underburden. The ranges are chosen arbitrarily, making sure that a wide enough range is used in the prior ensemble.

Table 7 Layer properties of the under- and overburden in the geomechanical model.

Layer name	Top layer [m]	Mean E [GPa]	Range E [GPa]	Mean ν	Range ν (\pm)
North Sea	0	2	1 – 7	0.3	0.2 – 0.4
Chalk	1000	10	5 – 15	0.25	0.15 – 0.35
Cretaceous/Jurassic	1500	16	6 -26	0.25	0.15 – 0.35
Zechstein	2000	30	20 - 40	0.35	0.25 – 0.45
Ten Boer Claystone	3200	40	20 - 60	0.2	0.1 – 0.3
Rotliegendes Sandstone (reservoir)	3600	Provided by compaction grid (C_m , ϕ relation)			
Limburg	3700	40	20 - 60	0.2	0.1 – 0.3

For each compaction and over- and underburden parameter, a parameter distribution is given and from this a range of models is generated to be confronted with the data. To generate the parameter distribution a Latin Hypercube Monte Carlo sampling has been used (TNO, 2016). The input parameters for the different compaction models are: Linear(C_m), Timedecay (C_m , Time constant), Bilinear ($C_{m,pre}$, $C_{m,post}$, P_{trans}) and RTCIM($C_{m,d}$, $C_{m,ref}$, b , $\dot{\sigma}'_{ref}$). For each over- and underburden layer the Young's modulus (E) and Poisson's ratio (ν) is sampled. In case of the inclusion of a linear viscous salt layer, the Maxwell viscosity is sampled as well.

Table 8 Compaction model parameter ranges for uniform distributions.

Model	Parameter (unit)	Minimum	Maximum	source
Linear	C_m factor	0.3	4	Core plugs
Bilinear	$C_{m,pre}$	0.3	1.7	Observation of subsidence data (Ameland) and core plugs
	$C_{m,post}$	1	4	Observation of subsidence data (Ameland) and core plugs
	P_{trans} (bar)	150	350	Observation of subsidence data (Ameland) and core plugs
Time decay	C_m factor	0.3	4	Core plugs
	Time constant (years)	2	20	Inversion of Wadden subsidence data (NAM, 2016a)
Rate type	$C_{m,d}$	0.3	1.2	Inversion of subsidence data (Wadden and Groningen, NAM, 2015c) and core plugs
	$C_{m,ref}$	1.3	4	Inversion of subsidence data (Wadden and Groningen, NAM, 2015c) and core plugs
	$\log(\dot{\sigma}'_{ref})$	-4	-2.5	Inversion of subsidence data (Wadden and Groningen, NAM, 2015c) and core plugs
	b	0.01	0.1	Inversion of subsidence data (Wadden and Groningen, NAM, 2015c) and core plugs

For each compaction and over- and underburden parameter a parameter distribution is given and from this a range of models are generated to be confronted with the data. To generate the parameter distribution a Latin Hypercube Monte Carlo sampling has been used.

8.5 Results for the compaction models using an elastic heterogeneous overburden

In this paragraph the results of the ESIP calculation is described for the case with a heterogeneous elastic overburden using the elastic influence function in AEsibs. Calculations are performed for 4 compaction models. For each of the 13 reservoir scenarios and each compaction model 200 draws with the Monte Carlo simulation were executed (e.g. this results in $200 \cdot 13 \cdot 4 = 10400$ calculations). Figure 62 presents for each compaction model the member probability corresponding to the values of the chosen compaction parameters. Maximum values of the distribution are found within the range of the chosen input distribution (Table 8). This is however less well defined for the $C_{m,post}$ of the Bi-linear and $C_{m,ref}$ of the RTCiM, where values appear to have the highest probability close to the end points of the distributions.

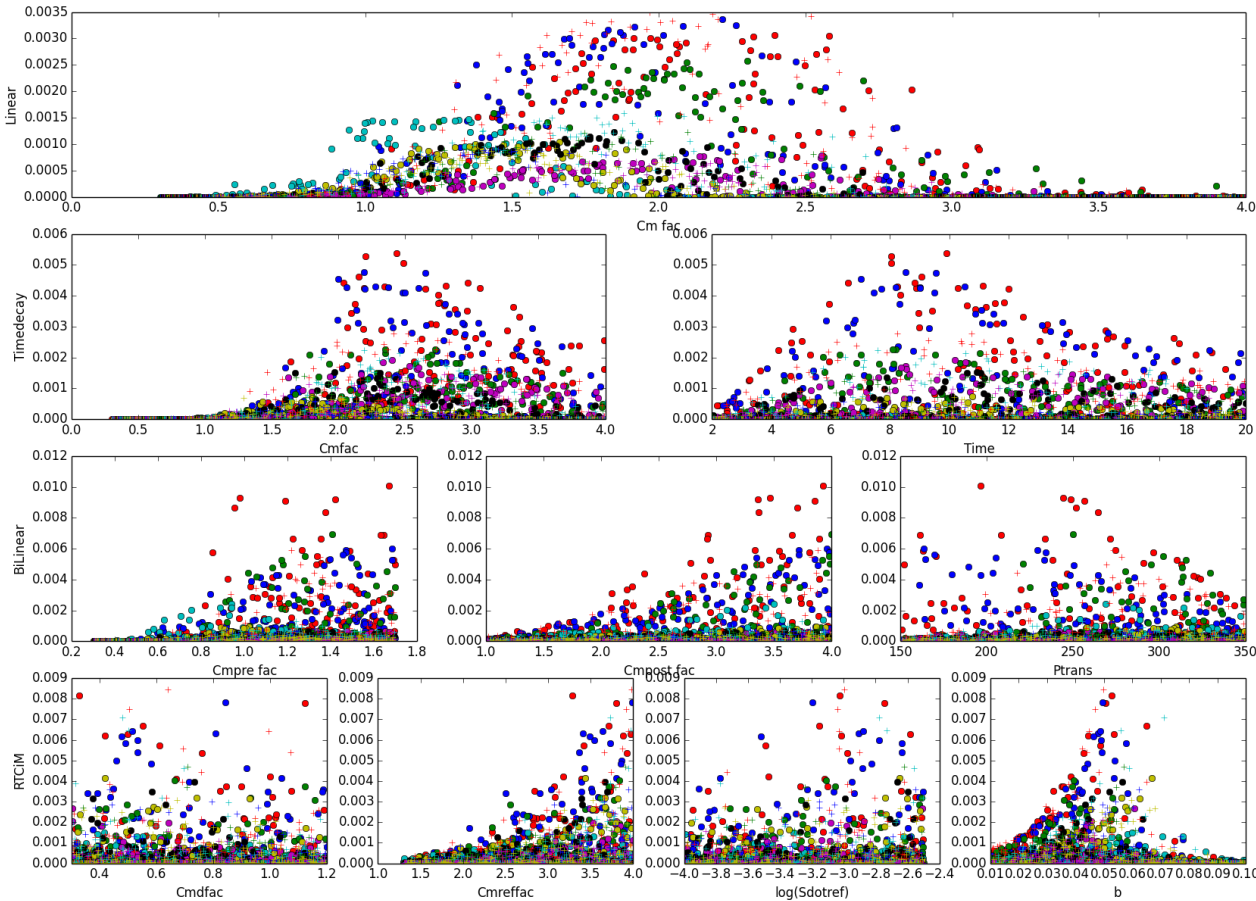


Figure 62: probability per member, the colors indicate the different reservoir models.

Both the Time-Decay and linear model members show that the highest probability members have a C_m factor in the range of 2-2.5. This is not consistent with the factor of 1.2 applied in the Geomec model and is the result of the wider subsidence bowl arising from the elastic overburden and the

deeper position of the reservoir at the location of the gas field. Within the range of possible realizations, it was not possible to find a specific member that matches all data points within a bandwidth of 2 cm. This is illustrated in Figure 63 for the members of each compaction model having the highest probability after the confrontation with the geodetic data. This figure presents a cross-plot of the model double differences (y-axis) versus the measured double-difference (x-axis). The plots demonstrate that especially the higher measured double differences are not matched by the modeled values, in particular when the linear compaction model is used. The Time-Decay model does a reasonable job matching most of the data within the bandwidth of ± 2 cm indicated by the black lines around the red line.

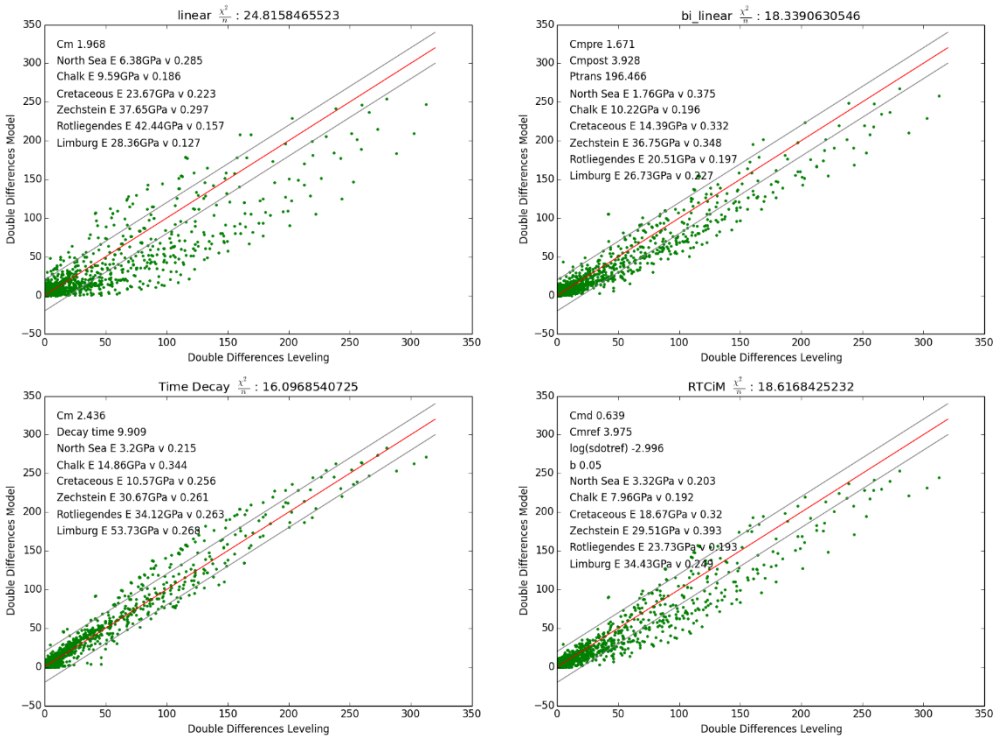


Figure 63 : Measured vs modelled subsidence for the best member per compaction model, the model parameters used are shown in each graph.

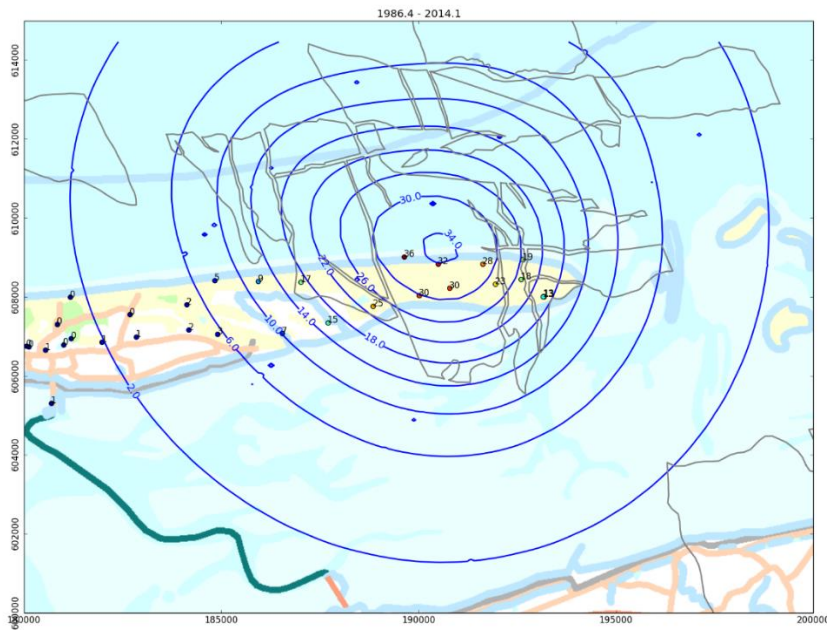


Figure 64 : Subsidence (cm) between start of production and 2014 for the best Time decay model.

This fit can be illustrated as well by comparing the contour lines with the measured data (MOVE3 data). This is done for the Time Decay member having the highest probability and for the year of 2014 (Figure 64). The model shows a good match with the data in the center of the bowl, but the fit is less good going towards the edges of the bowl. The fact that higher model values do not show up in the cross-plot is because of the difference between the double differences in the ESIP workflow versus the maximum absolute subsidence since 1986 used in Figure 64.

Figure 65 shows the results for several benchmarks in time. For each compaction model the modelled results for the members with the highest probability are presented .

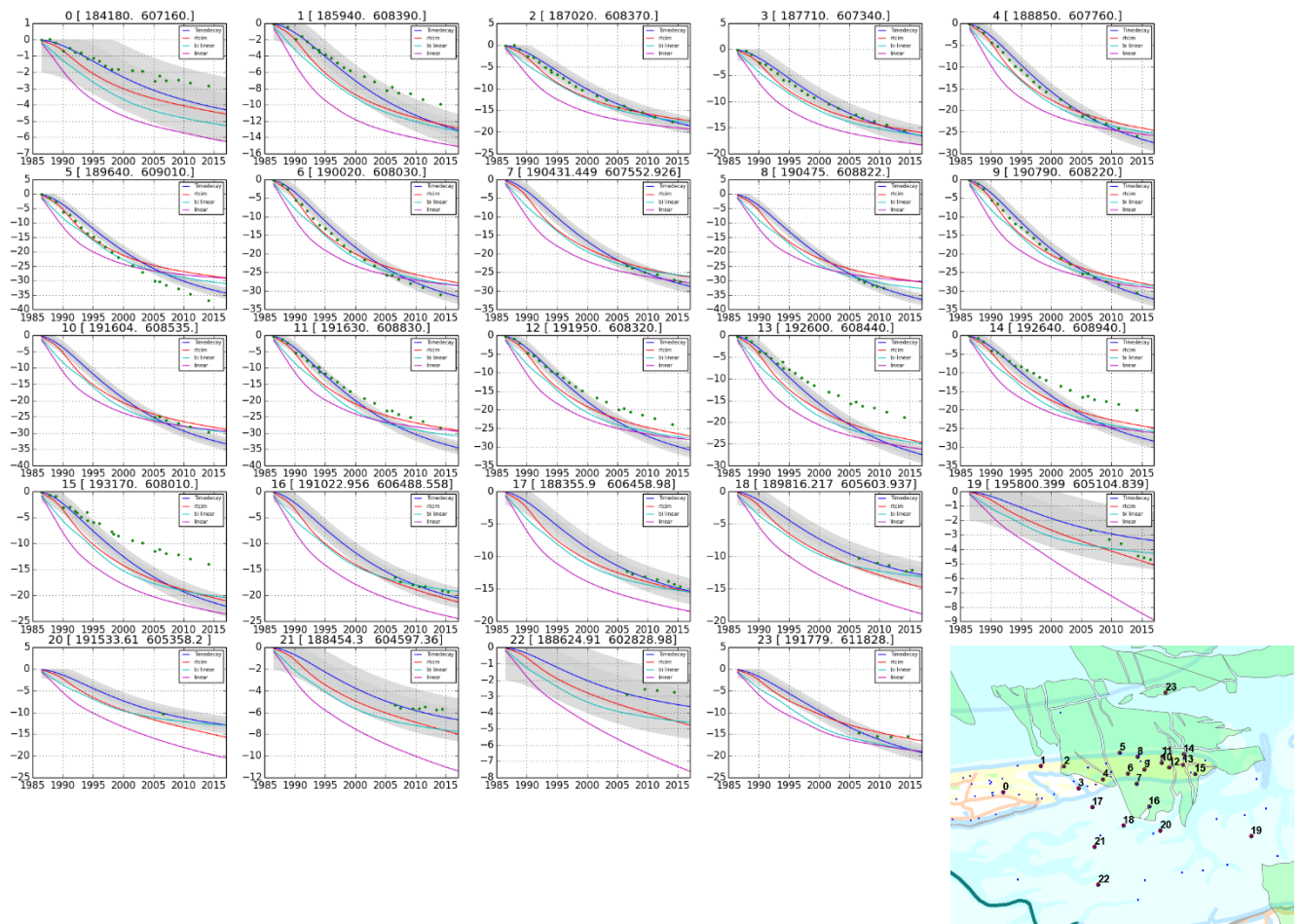


Figure 65 : Subsidence vs time at certain benchmark locations(bottom) for the model members having the highest probability and for the reservoir scenario with the least aquifer depletion.

The results above indicate that in general the best model members overestimate the measured subsidence. Again this is likely to be the result of the elastic boundary conditions. It illustrates moreover that the actual bowl is narrower and therefore an additional condition should be found that could influence the shape of the bowls. The LTS-1 study already showed that the viscous behavior of the salt is the prime candidate (e.g. Marketos et al. 2016), but LTS II now also provides the opportunity to study the impact of different reservoir scenarios and to address the impact of depleting aquifer blocks on the bowl shape. Therefore, an assessment was made of the likelihood of the various reservoir scenarios by summing all member probabilities per reservoir model. Ranking of the reservoir models is shown in Figure 66. Not surprisingly the reservoir scenario with the highest probability is the scenario with least amount of depletion in the aquifers (top left scenario in Figure 66). This result confirms the conclusion of NAM (2011) and TNO (2011) stating that the models without aquifer depletion fit best to the observed data. In the last 5 years some signs of aquifer depletion were detected by the GPS benchmarks in the Waddenzee and new reservoir scenarios including moderate aquifer depletion in the blocks south and southwest of the gas field are seen as the most probable scenarios. The challenge is now whether this conclusion can be confirmed by further calculations in this study when a viscous salt layer or the Knothe influence function are added in the calculations.

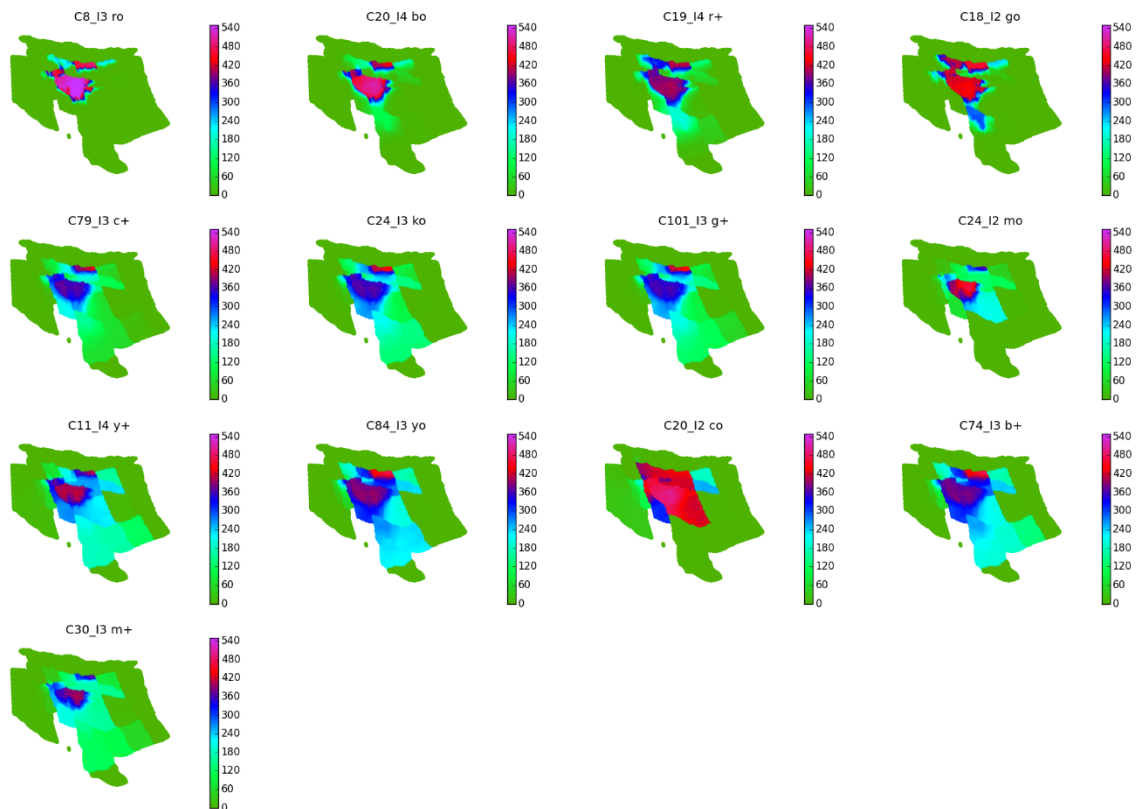


Figure 66 : Ranking of the scenarios based on the sum of all member probabilities per scenario. Best scenario is the top left scenario. Going from left to right and top to bottom the least probable scenario can be found in the bottom left corner of the picture.

8.5.1 Subsidence rates for the calculated volumes

For all members of the Monte Carlo simulation the subsidence volume versus time is calculated in a large area around the field (Figure 57) according to the method described in paragraph 8.3. This calculation results in Figure 67 that is similar to figures used in the measure and control documents (NAM, 2016a) for the Waddensee. The area selected for the calculation of the rates is however different. The 6-year averaged subsidence rates are presented for all 10400 members. From the probability density function (Figure 68) and cumulative density function the expectation case (solid black line) and the 95% confidence bound are calculated and presented by the dotted black lines. The figure illustrates that 95% of all scenarios can be found within a relatively narrow range.

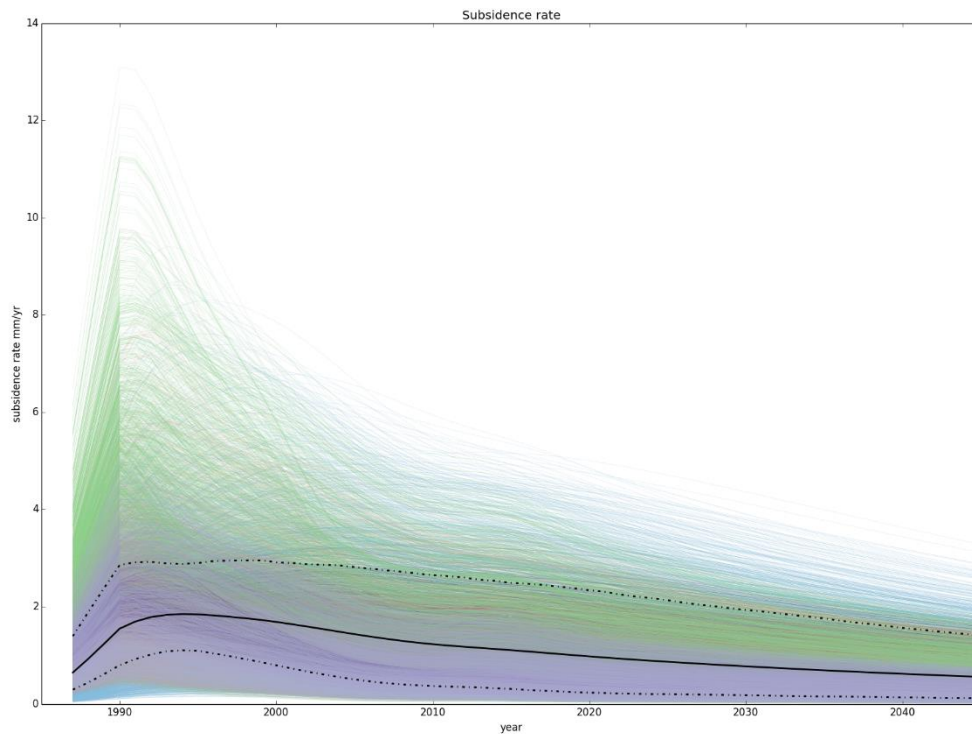


Figure 67 : Subsidence rate vs time, the black line is the expected subsidence rate, the dotted lines are at 2.5 and 97.5% cumulative probability. Colours represent different compaction models (Time decay=blue, RTCiM=red and Linear=green and Bi-linear=purple).

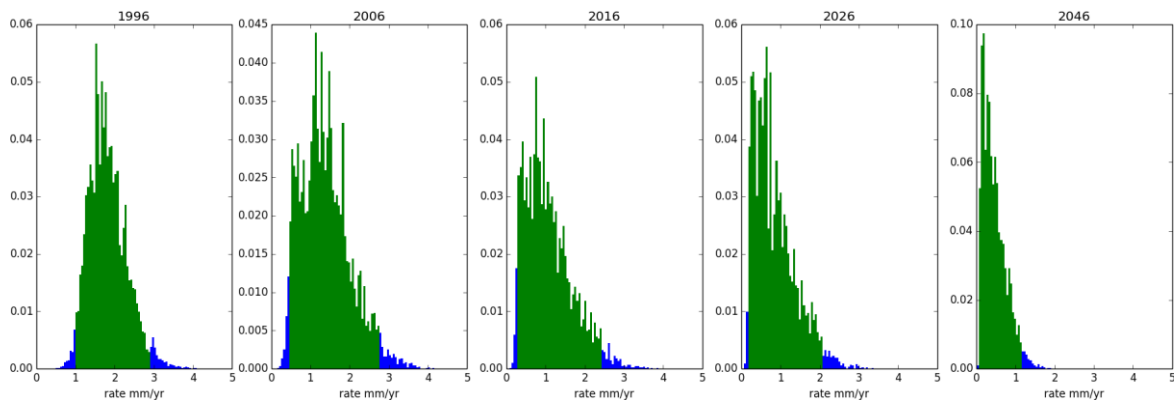


Figure 68: Probability density function for subsidence rates of certain years. The green area are the rates within the 2.5 - 97.5 % probability range (dotted lines of previous graph).

8.6 Results for the compaction models using a viscous salt

AEsubs is used to generate the functions to translate compaction to subsidence when including a viscous salt layer. Due to restrictions, discussed in Section 7.7, we limit the AEsubs model to four subsurface layers, to guarantee stable influence functions. The parameter ranges are presented in Table 9. The ranges for the compaction model parameters are the same as those used in the simulations of the elastic heterogeneous overburden in Table 8. The only difference here is that we did not run the rate type compaction model, because of time restrictions. Where it is technically

possible to run the rate type compaction model in combination with a viscous layer in AEsups, the testing of one member took about 30 min (due to necessary refinement of the time steps), compared to ‘only’ 7 min when confining the runs to just linear, bilinear and time decay compaction models. Finishing the workflow for as many members as was done in the elastic runs in the previous section would have taken more than a month. Even without the rate type model, the duration is still significant, hence we also restricted the run to 50 members per compaction model, per reservoir scenario (in total testing 1950 members; factor of 5 fewer than in the elastic runs)

Both the Poisson’s ratio and Young’s modulus are taken from a uniform distribution with the limits as given in Table 9. The narrow ranges are necessary to prevent AEsups from generating unrealistic influence functions where subsidence volumes are no longer preserved. A similar contrast between Layers 1, 2 and Layers 3, 4 was used in the comparison of AEsups results with Geomec (see Section 7.7). For the salt layer, a Poisson’s ratio of 0.35 is assumed, in line with data (Fokker, 1995), around which we allow a small distribution. For other layers, we restricted the Poisson’s ratio to the realistic range of 0.2 to 0.3.

Although the lack of variability in stiffness contrast in the restricted salt models can, to some extent, be compensated for in the compaction model parameters, it means that only those model parameters that form a particularly good combination with the particular set of stiffness parameters used will gain a higher probability in the ESIP workflow. If, for example, it was possible to increase the stiffness contrast, then generally models that generate a lower compaction would gain some likelihood, whereas the stiffness contrast can be lowered, models would require a higher compaction to describe the data (see discussion around Figure 45). Ideally, one would want to probe a larger range of models, including a range of stiffness contrasts as well as salt viscosity, such that no particular compaction model or pressure scenario gains any preference. Unfortunately, as discussed, this is currently not possible with AEsups, so for now we restrict the layer properties to those in Table 9.

Since realistic values for salt viscosity span several orders of magnitude [Marketos et al, 2016], we use a log-uniform distribution for the viscosity of Layer 2, with the upper and lower limits as in Table 9.

Table 9: Elastic parameter ranges used in ESIP workflow when including a viscous layer.

Layer	Top depth, fixed (m)	Poisson ratio min	Poisson ratio max	Young’s modulus min (GPa)	Young’s modulus max (GPa)	Viscosity min (Pa s)	Viscosity max (Pa s)
Layer 1	0	0.2	0.3	7.5	8.5	-	-
Layer 2	2000	0.34	0.36	7.5	8.5	1×10^{16}	1×10^{19}
Layer 3	3100	0.2	0.3	19.5	20.5	-	-
Layer 4	4200	0.2	0.3	19.5	20.5	-	-

The weights for all the members, normalised per compaction model, are plotted in Figure 69 against the compaction model parameters. Like the output for elastic models presented in Figure 62, the maxima of the prior distributions lie within the ranges for the input parameters, see Table 8, for the linear and time decay models. Compared to the elastic heterogeneous models, the centre of the distributions for the compaction coefficients appear to have shifted to slightly lower values when using the layer properties in Table 9, but also the distributions have become narrower.

Note that, as discussed above, the salt runs presented here had to be limited over the ranges of the elastic properties in the model. Had it been possible to run the workflow also including a higher stiffness in the underburden, then models combining those higher stiffnesses with a lower compaction coefficient would have had some reasonable probability, higher than for similar C_m factors in Figure 69, and the resulting distributions would have been wider. Following the same logic, if the model parameters had been limited to only higher stiffnesses then used now, then the distribution of the C_m factors would shift to a lower value. This illustrates that, at the moment, the parameter distributions are to some extent dependent on the limited layer properties used.

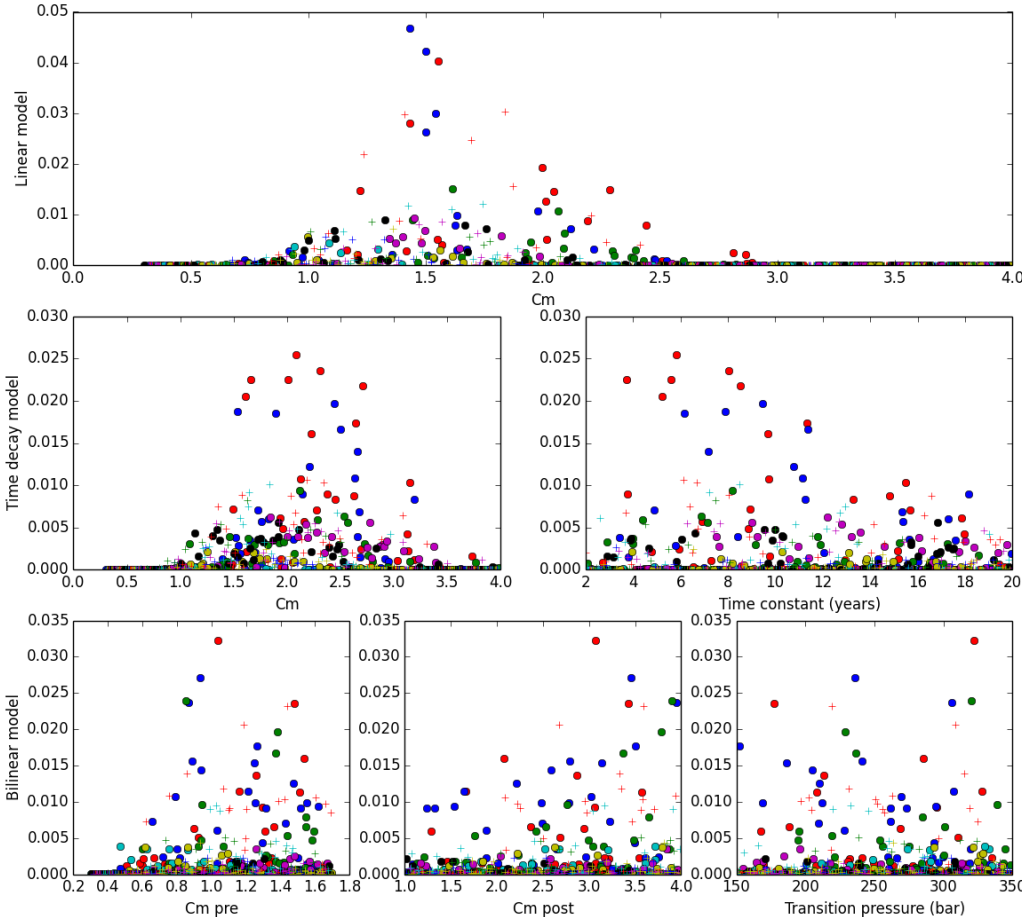


Figure 69: Probabilities (normalized per compaction model) for the compaction model parameters. Each colour represents a different reservoir pressure scenario.

A similar plot is made for the member weights, now presented as a function of the viscosity of the salt layer, in Figure 70. The members with the highest weights all use a viscosity in the range 1×10^{17} - 1×10^{18} Pa s, which agrees well with the value of 6.7×10^{17} Pa s used in the Geomec model. Both the linear and bilinear members show a narrower distribution than those from the time decay model, likely resulting from the additional time dependency; models with faster salt but slower time decay may give a similar time dependence for the subsidence as those with slower salt but faster time decay, widening the range of viscosities.

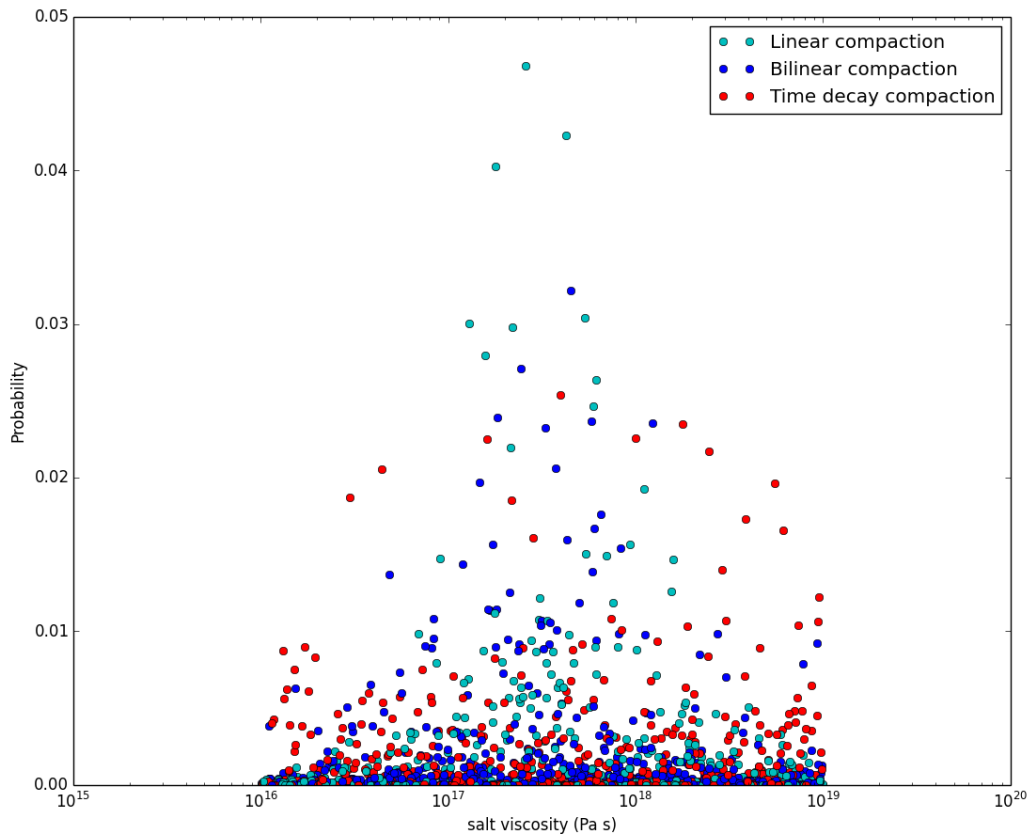


Figure 70: Model probability plotted versus the viscosity of the salt layer.

Figure 71 compares the measured and modelled double differences for the member with the highest weight for each of the three compaction models tested. The red lines are $y=x$, and the grey lines show the ± 2 cm bandwidth. The linear compaction model seems to show the largest differences between model and data. The bilinear and time decay models perform better in the sense that the modelled data lie closer to the measurements in the cross plots, also reflected by the lower values of χ^2/N .

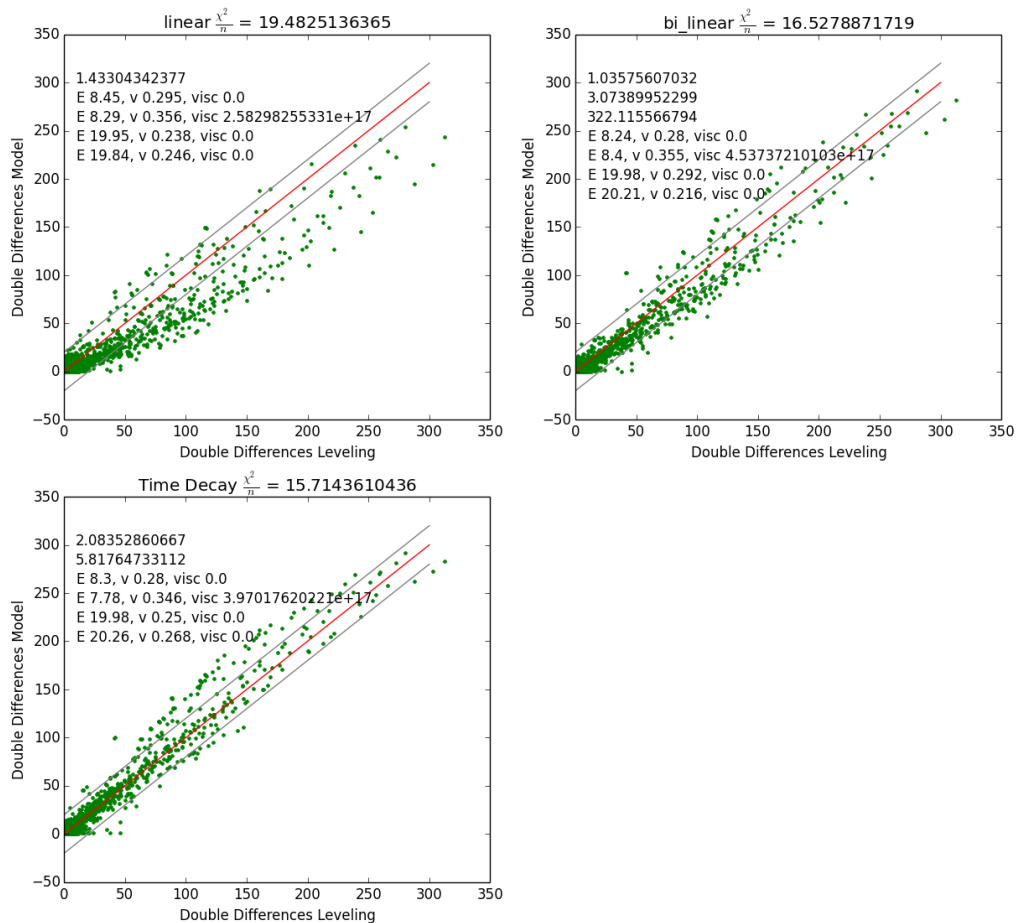


Figure 71: Measured versus modelled double differences for the members with the highest weight for each compaction model.

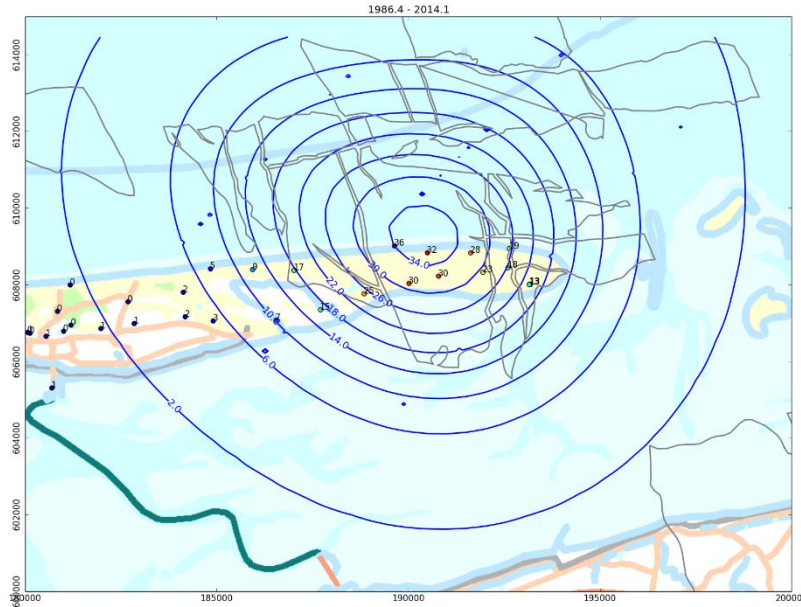


Figure 72: Subsidence (cm) between start of production and 2014 for the best Time decay model.

The fit of the best salt influence function member is illustrated by comparing the contour lines with the measured data. This is done for the Time Decay member having the highest probability and for the year of 2014 (Figure 71). The model shows a good match to the data in the center of the bowl but the fit is less good going towards the edges of the bowl. The fact that higher model values do not show up in the cross-plot is because of the difference between the double differences in the ESIP workflow versus the maximum absolute subsidence since 1986 used in Figure 71.

Figure 73 presents the results for several benchmarks in time presenting the model results for the members of each compaction model having the highest probability.

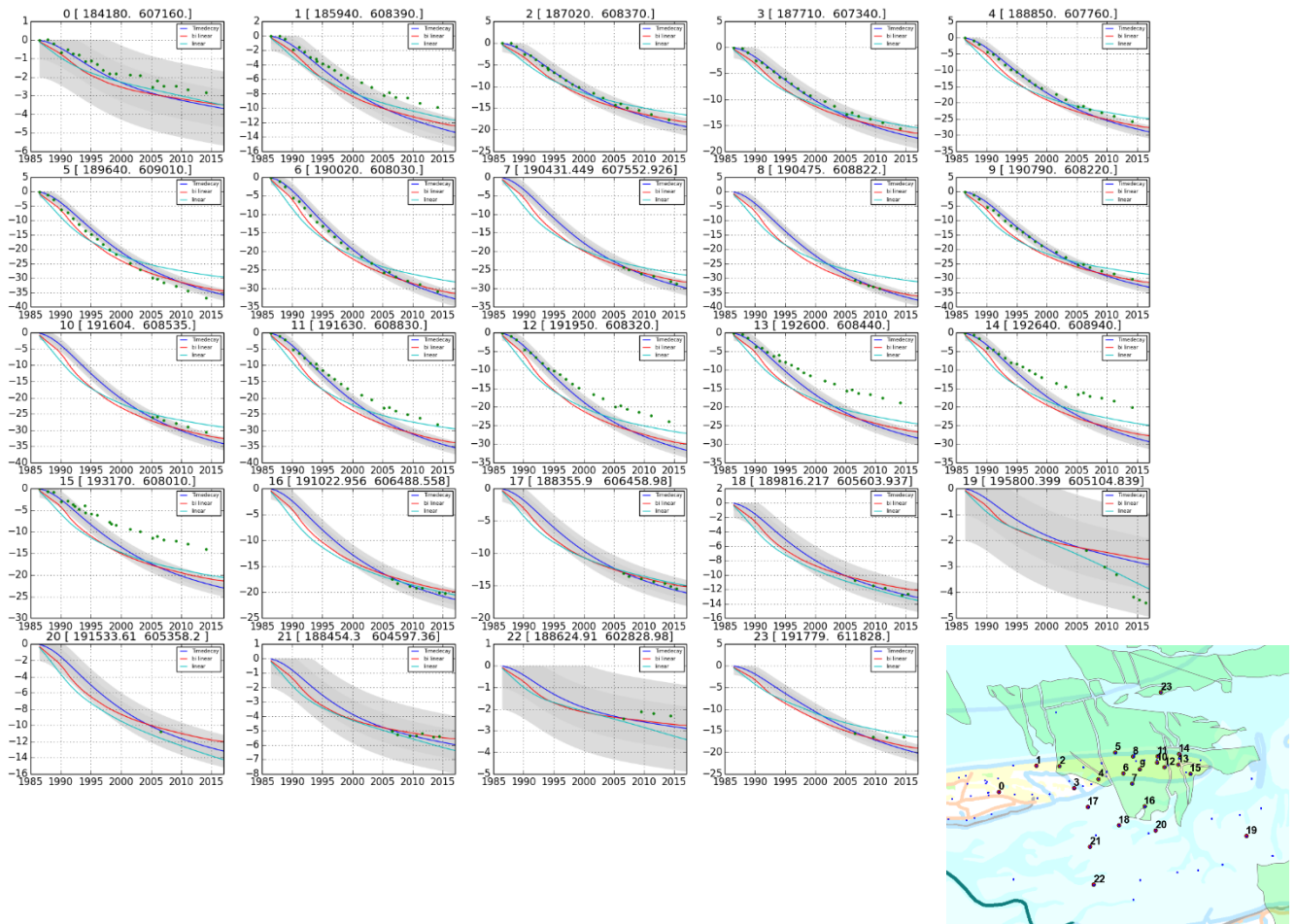


Figure 73: Subsidence vs time at certain benchmark locations(bottom) for the model members having the highest probability and for the reservoir scenario with the least aquifer depletion.

8.6.1 Subsidence rates for the calculated volumes

The subsidence volume over the area defined in Figure 57 was determined for all the members using the viscous layer, which was converted to an average subsidence rate. The results are shown in Figure 74. Using the individual member’s weights, a weighted average in subsidence rate is plotted in black, with a 95% confidence interval marked by the dash-dot lines. The transparency of the rates from each member is dependent on the member’s weight, making the least likely model the most transparent. Although some of the members result in very high average subsidence rates, the 95% confidence bound defines a relatively narrow range. Figure 75 shows the probability distributions of the subsidence rates for a number of years.

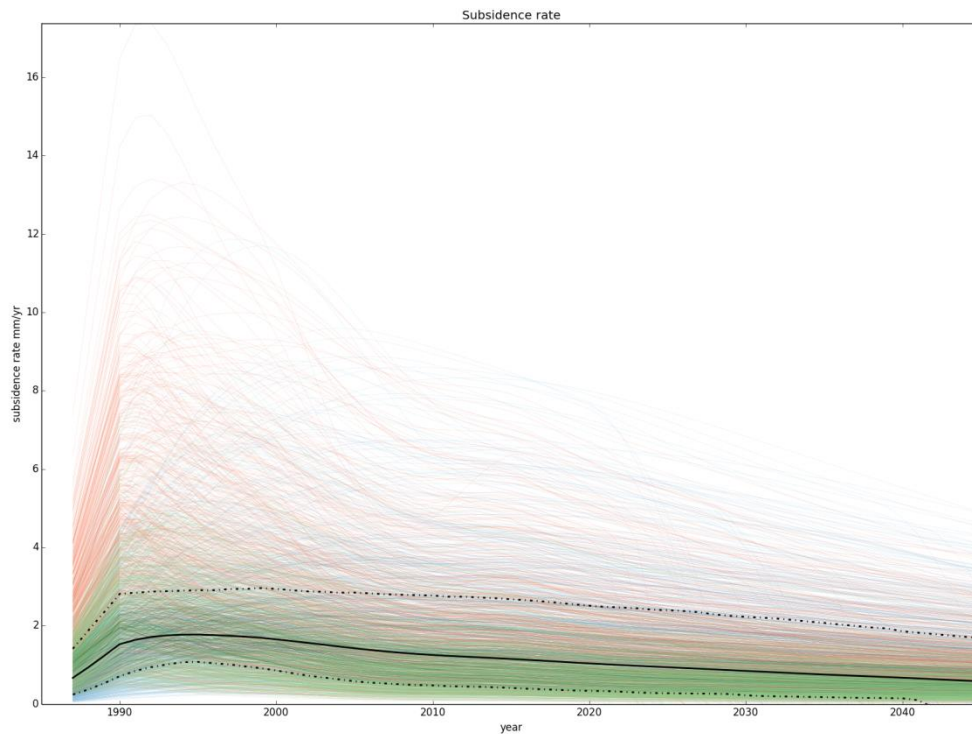


Figure 74: Subsidence volumes in the area defined in Figure 57, converted to subsidence rates, resulting from ESIP with a viscous salt layer. Colours represent different compaction models (Time decay=blue, linear=red and bilinear=green), the transparency reflects the member probability.

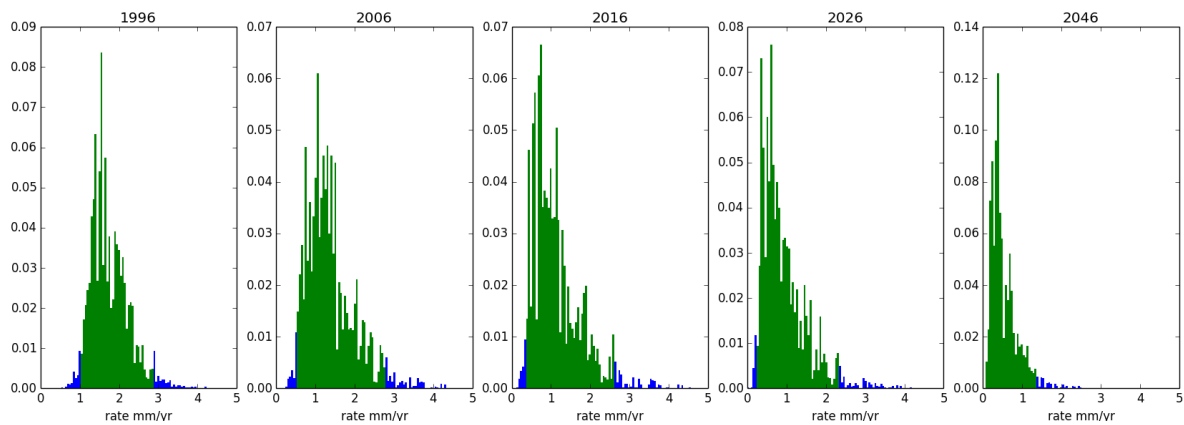


Figure 75: Probability density plots of the subsidence rate for certain set of years. Green data lies within the 95% confidence interval, data in blue are the 2.5% on either side of the distribution.

8.7 Results for the compaction models using the Knothe influence function

The Knothe influence function uses the influence angle φ to calculate the subsidence from a compaction source. This angle is varied between 30 and 70 degrees where a smaller angle results in a narrower subsidence bowl. Figure 76 shows the results for the probability of this range for the most likely reservoir scenario according to the previous elastic calculations. Members having an influence function of around 55° show the highest probability. Note that the influence angle in the Knothe

function cannot be varied with time while an influence angle that results from the viscous salt does vary with time. The angle of 55° can therefore only be regarded as an average optimum angle.

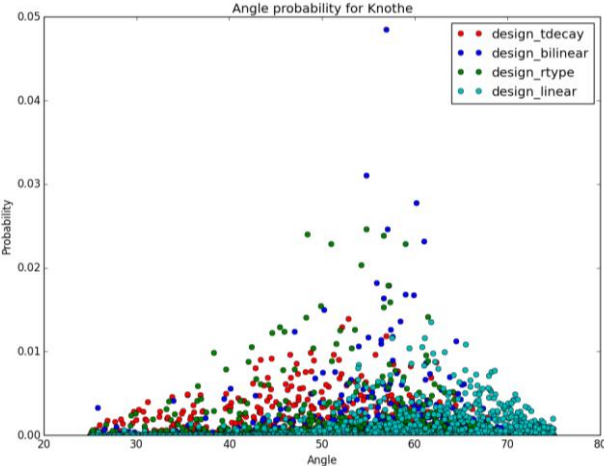


Figure 76: Knothe influence angle versus model probability.

150 members for each of the thirteen reservoir scenarios were generated using Knothe’s influence function. The uncertainty ranges used for the compaction model parameters are the same as for the elastic influence function (paragraph 8.5). The probability for each member per compaction model is shown in Figure 77, which can be compared to Figure 62. In general, the most probable C_m factors have lower values when compared to the elastic case. This is expected because the levelling points in the center of the bowl can be matched by the narrower shape of the bowl with C_m factors closer to the laboratory measurements.

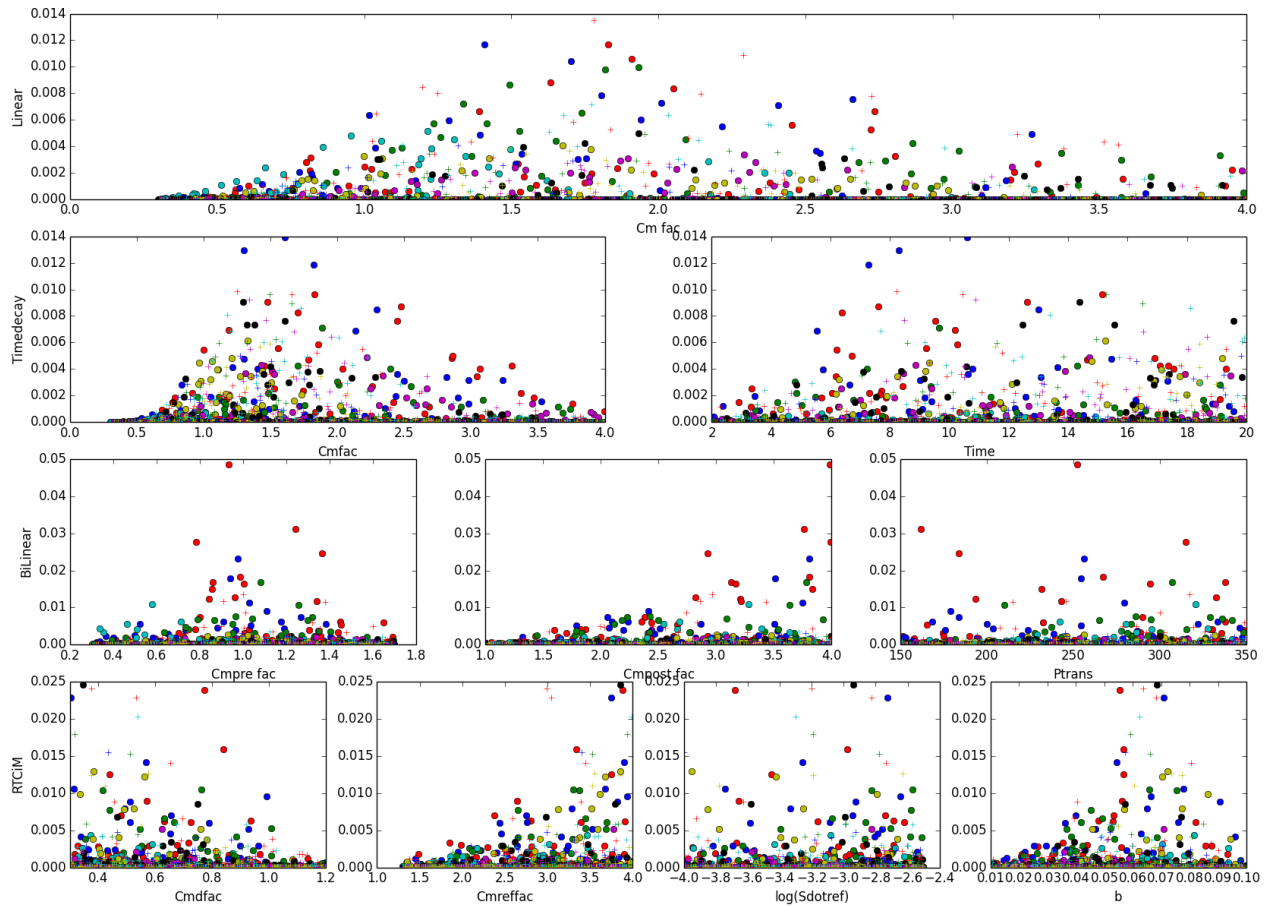


Figure 77: Probability per member for each parameter per compaction model for the Knothe influence function.

Within the range of possible realizations, it is not possible to find a specific member that matches all data points within a bandwidth of 2 cm. This is illustrated for each compaction model in Figure 78 for the members with the highest probability after the confrontation with all geodetic data. This figure presents a cross-plot of the model double-difference on the y-axis versus the measured double differences on the x-axis. Also this figure can be compared to the elastic case and it is observed that the Knothe influence function leads to a slightly better fit to the data expressed by both the narrower ranges as well as the lower values for the χ^2/N values.

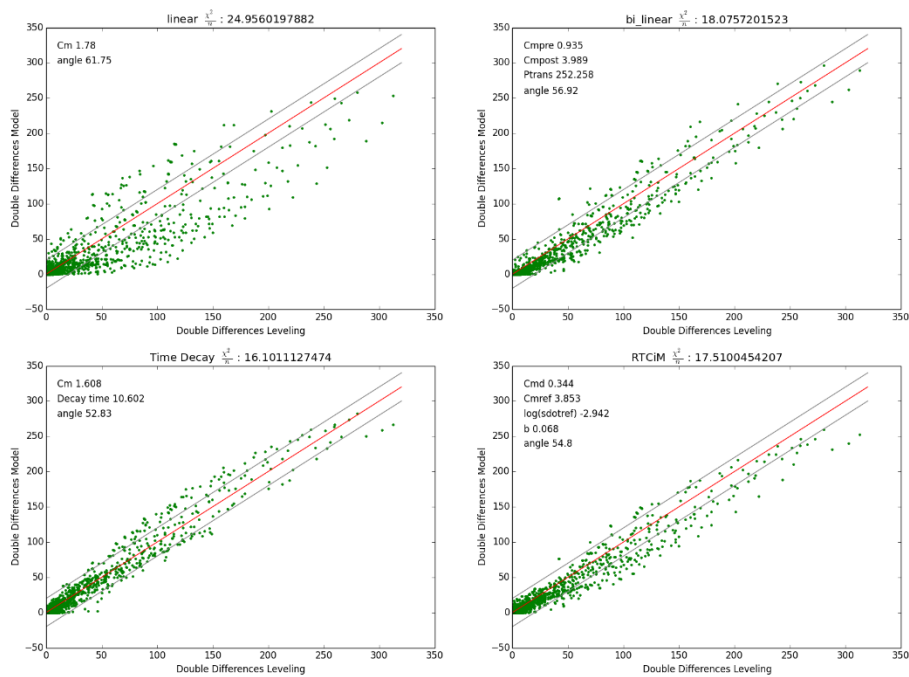


Figure 78: Measured vs Modelled double differences for the best member per compaction model, the model parameters used are shown in each graph.

The fit to the data can be illustrated as well by comparing the 2014 contour lines of the modelled results with the measured data. This is done for the RTCIM member having the highest probability (Figure 79). When compared to the elastic case (Figure 63), the model results for the Knothe influence function show a better match to all data, an observation that results directly from having the possibility to narrow the bowl. The fact that higher model double difference values do not show up in the cross-plot (Figure 78) is because of the difference between the double difference values used in the ESIP workflow versus the maximum absolute values for the measured subsidence since 1986 used in Figure 79.

Figure 80 presents the results for several benchmarks in time presenting the model results for the members of each compaction model having the highest probability. Also this figure shows that the model results are closer to the data when compared with the results presented in Figure 65.

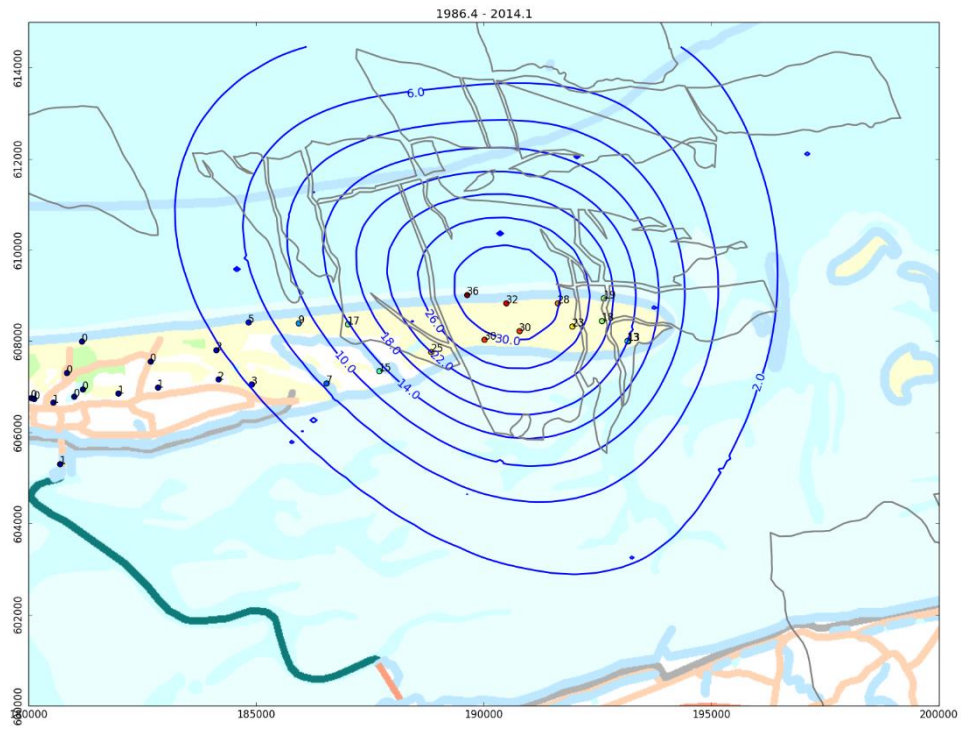


Figure 79: Subsidence (cm) between 1986 and 2014, modelled (contours) and measured (points).

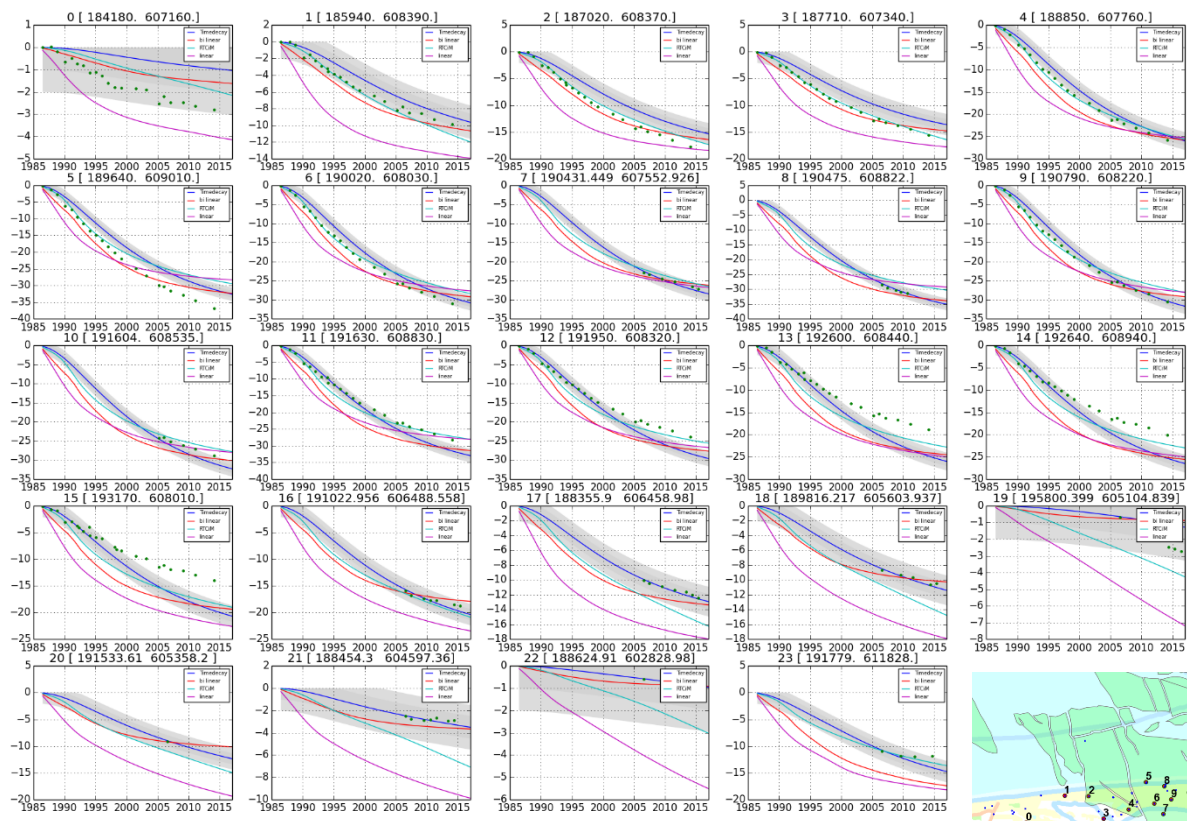
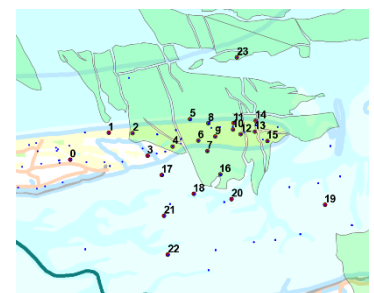


Figure 80: Subsidence vs time for the best compaction models.



For all members of the Knothe influence function the subsidence rate is calculated for the same area as for the AEsubs elastic influence function and viscous salt. The subsidence rate is calculated for a 6 yearly average and shown in Figure 81. Comparing the subsidence rate with the rates that result from the elastic AEsubs model we conclude that on average the subsidence rates are lower for the Knothe influence function.

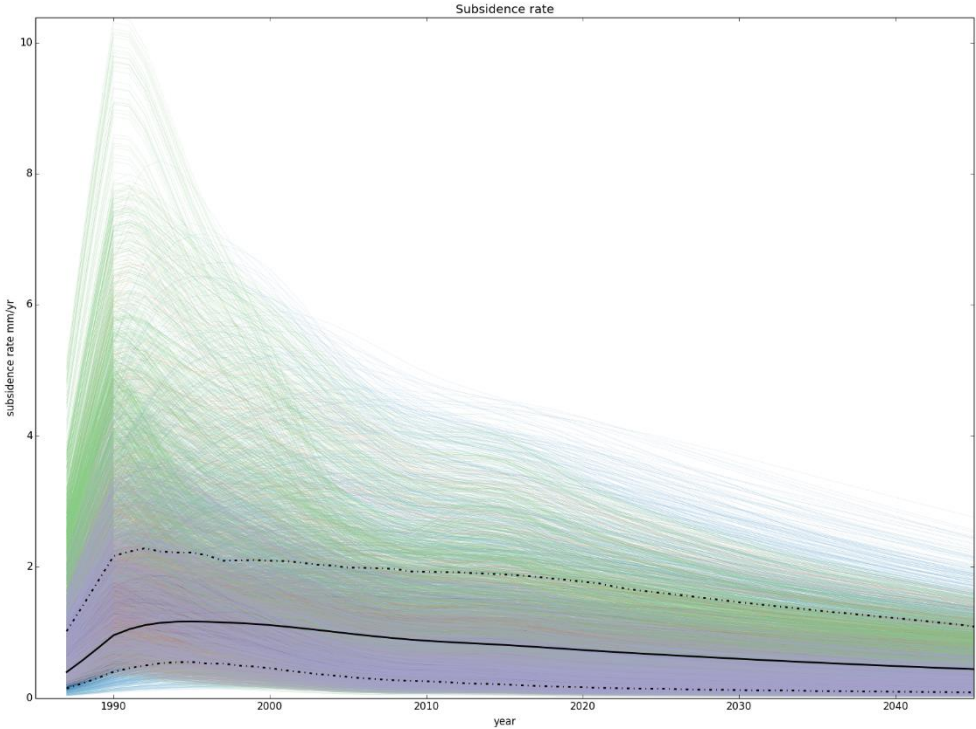


Figure 81: Subsidence rate vs time for the Knothe influence function, the black line is the expected subsidence rate, the dotted lines are at 2.5 and 97.5% cumulative probability.

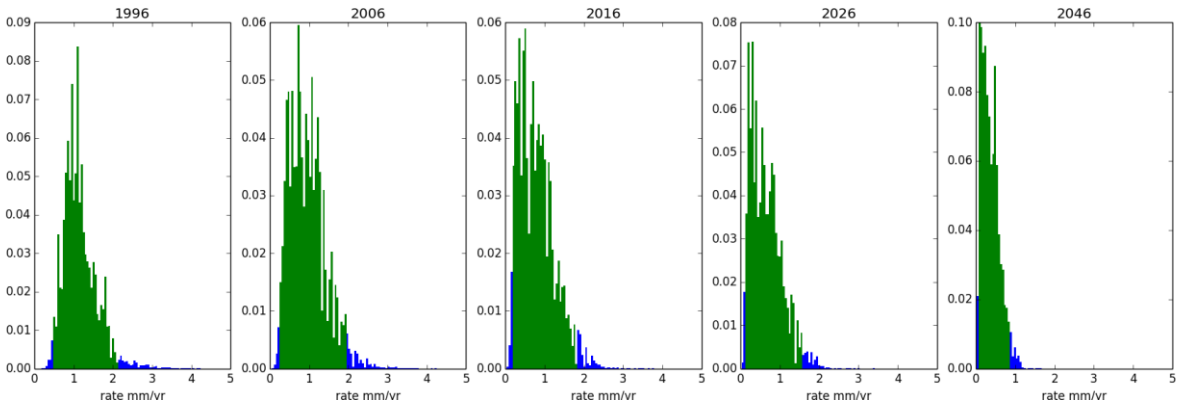


Figure 82: Probability density function for subsidence rates of certain years. The green area are the rates within the 2.5 - 97.5 % probability range (dotted lines of previous graph).

8.8 Emergency stop scenarios

The hand-on-the-tap procedure relies on the mechanism that when production is stopped the subsidence rate will decrease. The ESIP tool can be used to investigate whether this assumption is correct. Therefore, the years 1996 and 2016 have been selected to simulate a full production stop in Ameland while the model allows further pressure equilibration in time. The effect of the emergency stop in 1996 on the reservoir pressures is shown in Figure 83 and in Figure 84. Location 1 presents the response of the gas field, with increasing pressures for those scenarios with active aquifers. Active aquifers (location 2,3,4) will continue to deplete (to equilibrium) also after the production stop because of the pressure difference with the gas field.

Emergency stop in 1996

The impact of the production stop in 1996 on the subsidence rates for all members is visualized in Figure 85. The black line in this figure is the expectation subsidence rate, the dotted lines are at 2.5% and 97.5% probability defining the 95% confidence bound. When compared to the normal production scenario we can observe a clear distinction of the expected rate that results from the emergency stop scenarios when compared to the normal production cases. It can be observed that the deviation with the normal production scenario occurs earlier than 1996. This is explained by the 6 year moving average that is used in the definition of the subsidence rate to mimic the conditions of the measurement and control cycle for the “Waddenzee”.

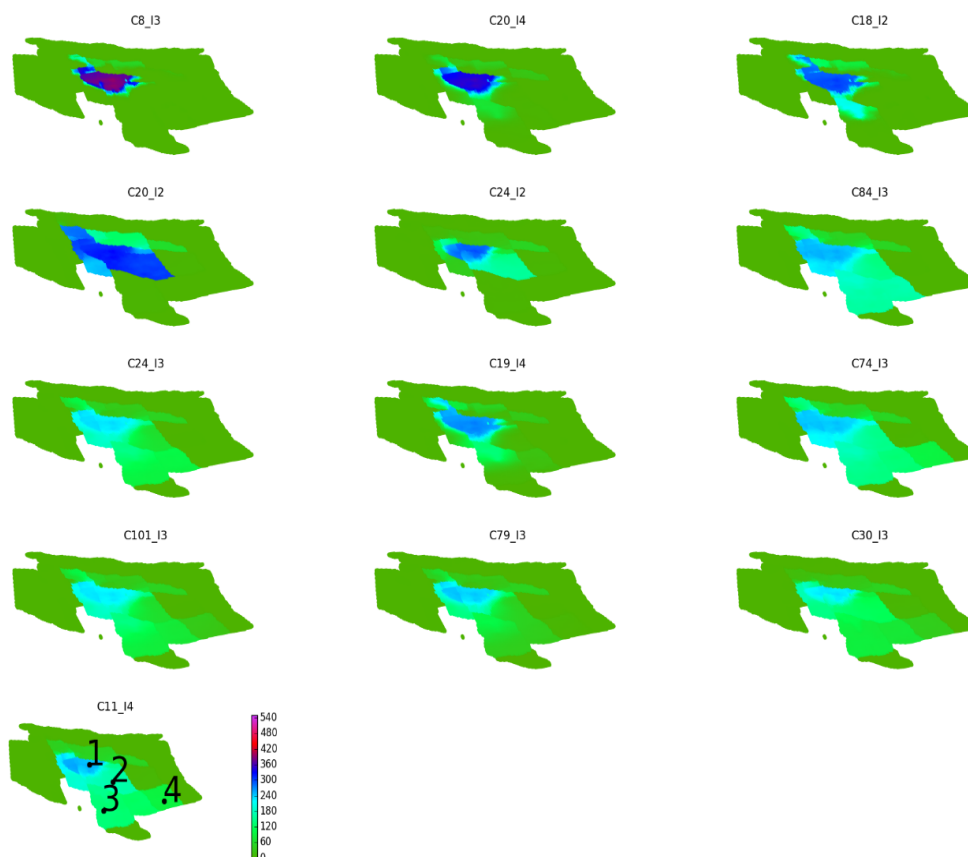


Figure 83: Lateral pressure depletion distribution for the selected reservoir models for the emergency stop in 1996 case for 2040.

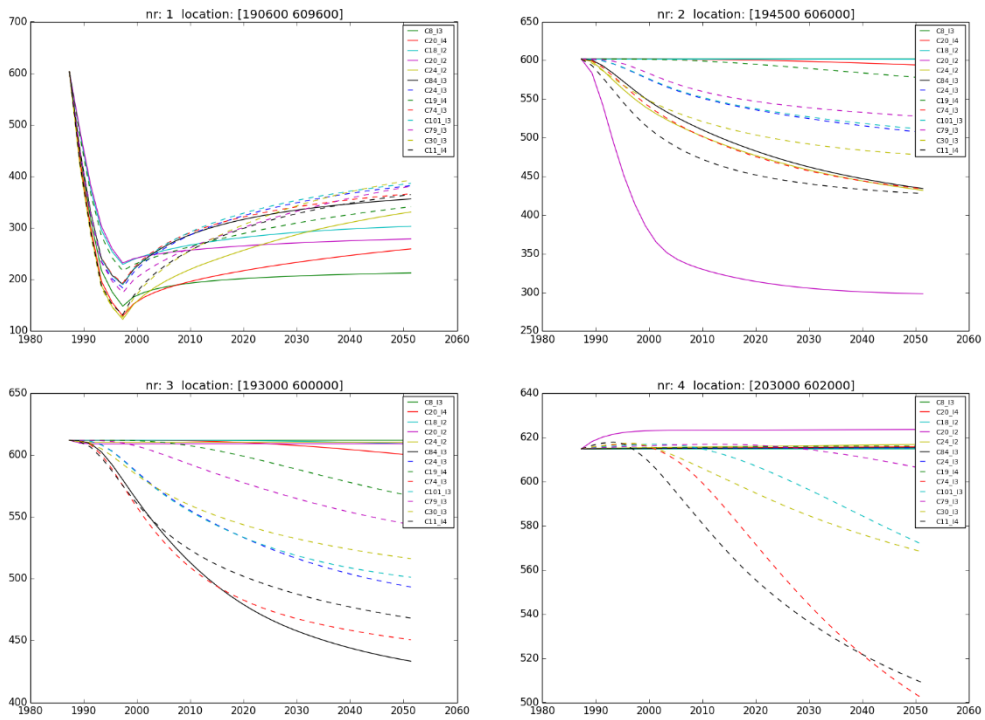


Figure 84: Pressure vs time for the emergency stop 1996 scenario at the location specified in previous graph.

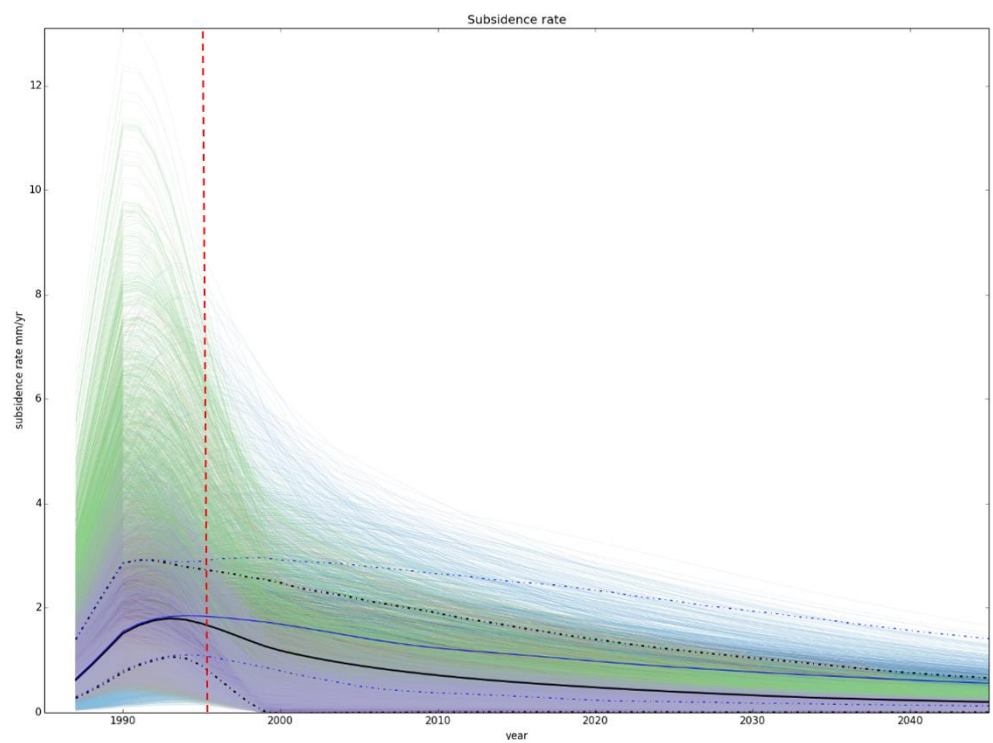


Figure 85: Subsidence rate vs time, the black line is the expectation subsidence rate, the dotted lines are at 2.5 and 97.5% probability. The blue lines are showing the weighted average and confidence bounds for the elastic model with no emergency stop. Red dashed line indicates the timing of the emergency stop.

A similar exercise has been executed for the Knothe influence function (Figure 86). Both the subsidence rates and the conclusions based on the deviation with the normal production situation are similar to that what has been stated for the elastic AEsups influence function.

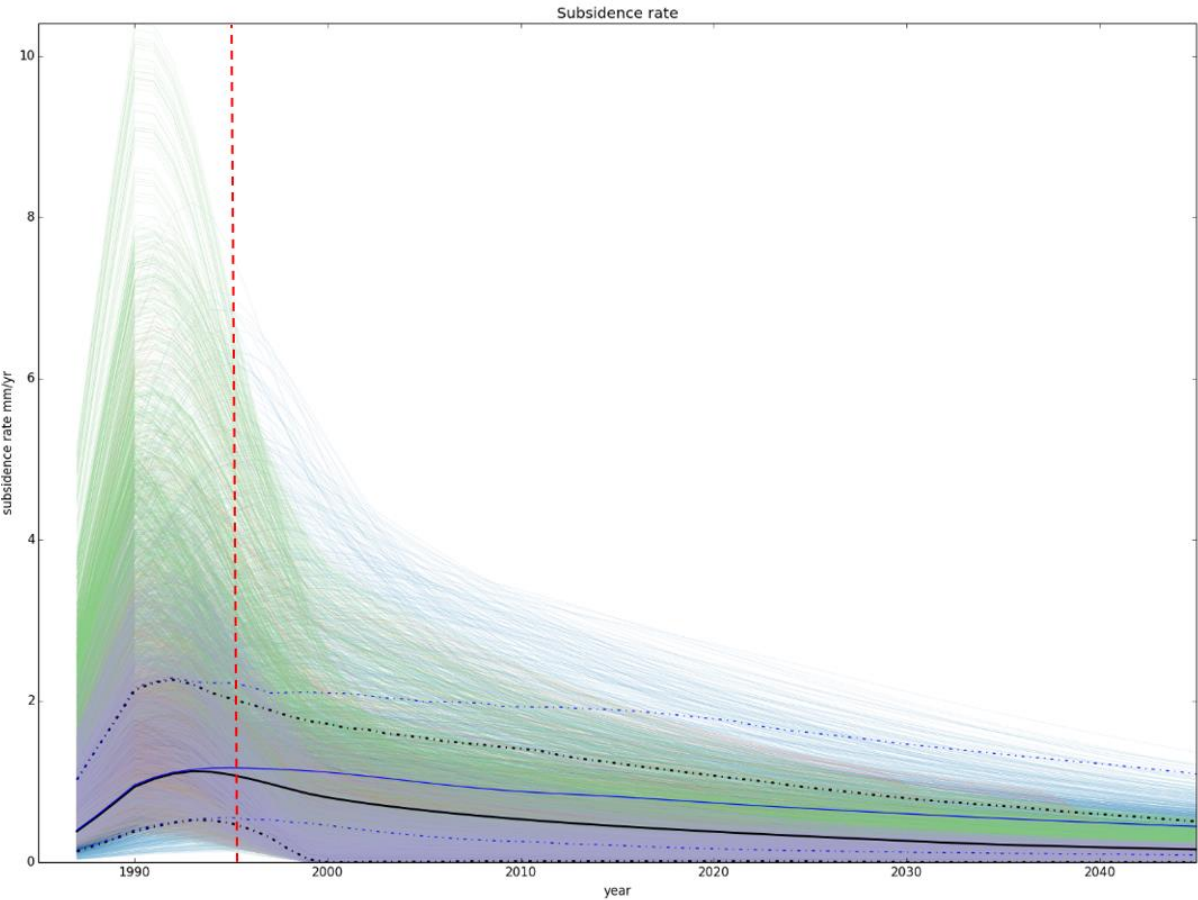


Figure 86: Subsidence rate vs time, the black line is the expectation subsidence rate, the dotted lines are at 2.5 and 97.5% probability. The blue lines are showing the weighted average and confidence bounds for the Knothe model with no emergency stop. Red dashed line indicates the timing of the emergency stop.

A similar exercise was performed for the Salt influence function (Figure 87). Both the subsidence rates and the conclusions based on the deviation with the normal production situation are similar to that what has been stated for the elastic AEsups and Knothe influence functions.

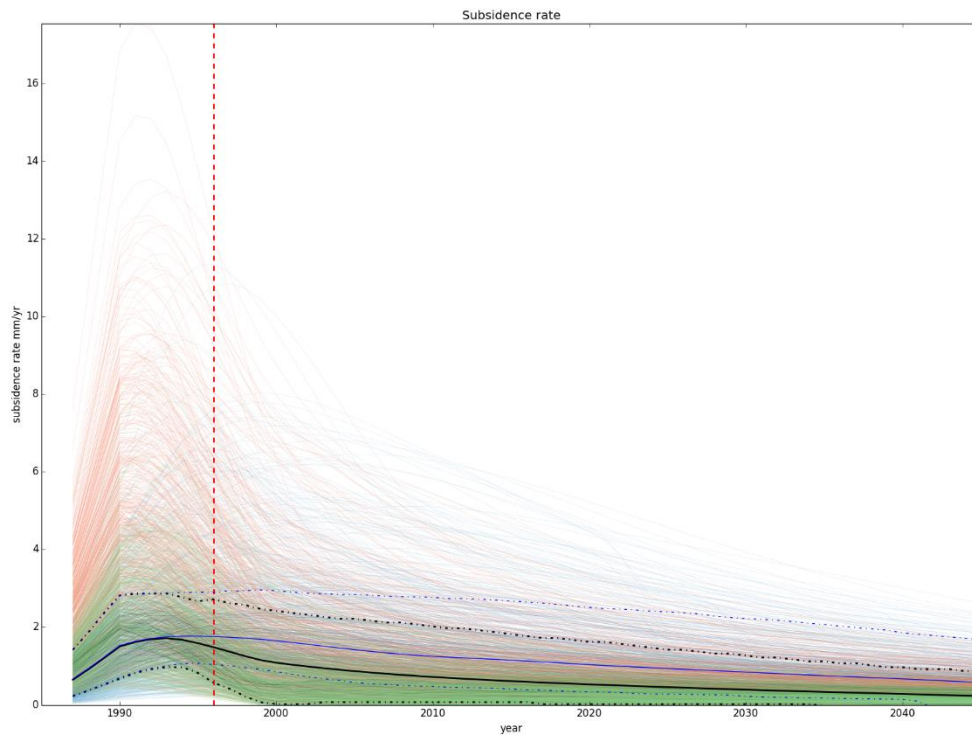


Figure 87: Subsidence rate vs time, the black line is the expectation subsidence rate, the dotted lines are at 2.5 and 97.5% probability. The blue lines are showing the weighted average and confidence bounds for the viscous salt model with no emergency stop. Red dashed line indicates the timing of the emergency stop.

Emergency stop in 2016

The effect of an emergency stop in 2016 on the reservoir pressures is shown in Figure 88 and Figure 89. The effect of this hypothetical stop on depletion compared to the original production scenario is small as the field has already produced the majority of all the gas (Figure 90). The small deviation is mainly caused by pressure equilibration of the aquifers with the gas field. This small effect is also visible in the small deviation of the subsidence rates when compared to the cases without a production stop. We only show the results for the elastic case.

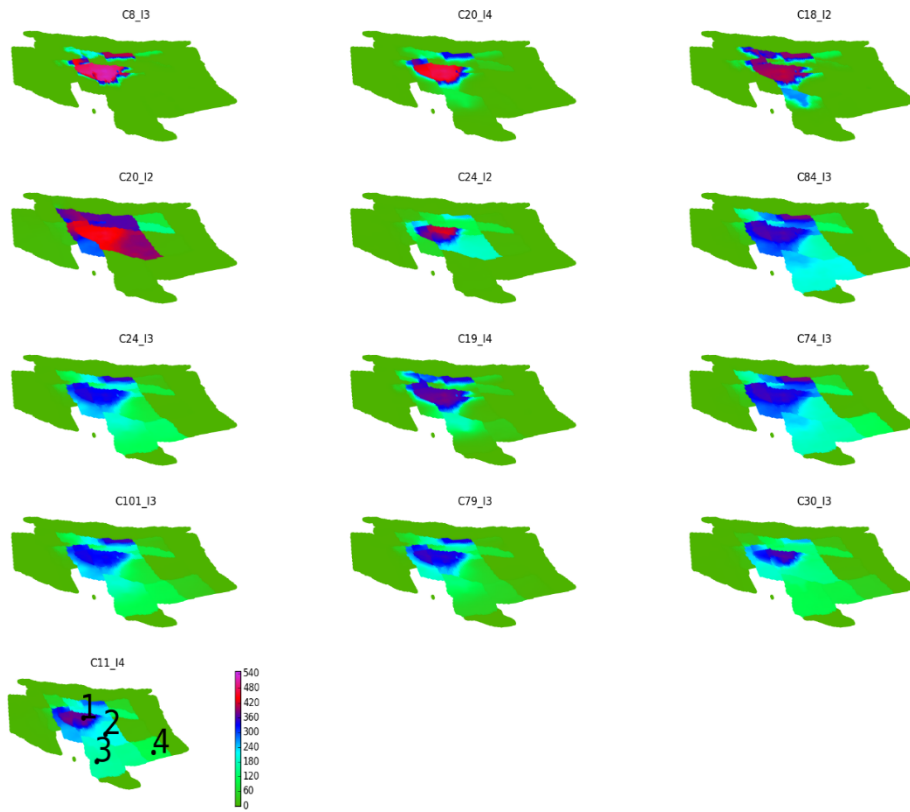


Figure 88: Lateral pressure depletion distribution for the selected reservoir models for the emergency stop in 2016 case for 2040.

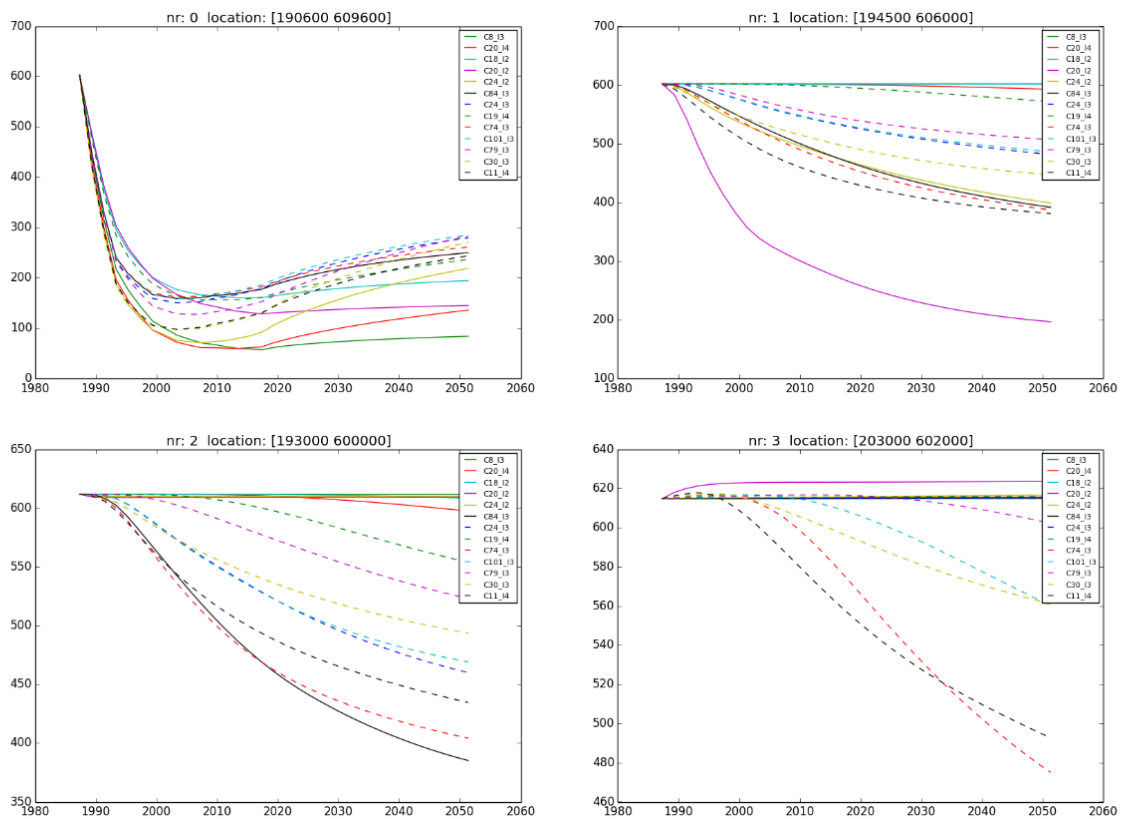


Figure 89: Pressure vs time for the emergency stop 2016 scenario at the location specified in previous graph.

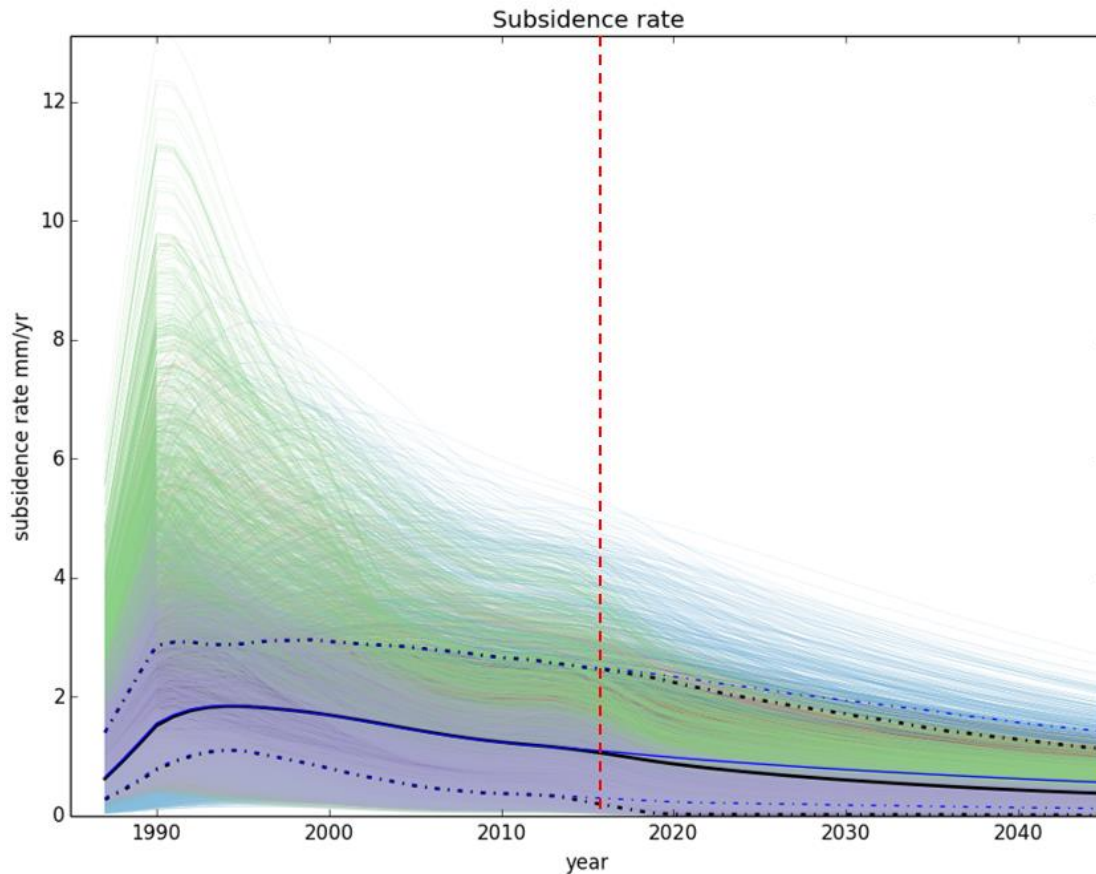


Figure 90: Subsidence rate vs time, the black line is the expectation subsidence rate, the dotted lines are at 2.5 and 97.5% probability. The blue lines are showing the expected case and confidence bounds for the elastic model with no emergency stop. Red dashed line indicates the timing of the emergency stop.

In conclusion: the emergency stop scenarios show a clear response to a hypothetical stop of the Ameland gas production in 1996. This provides confidence in the application of the hand-on-the-tap procedure as defined for the production of gas from the ‘young’ Waddenzee fields (Nes, Moddergat, Lauwersoog-C, -West and -Oost and Vierhuizen-Oost).

8.9 Discussion of the results

The ESIP workflow has been used to confront model results with data. Besides parameter uncertainty, also compaction model uncertainty and reservoir (including aquifer) pressure uncertainty is introduced in the setup of the prior ensembles. In the previous paragraphs the results show clear differences between the controlling factors: the influence functions, reservoir scenarios and compaction models. This paragraph discusses the impact of the uncertainty per factor. We reflect at the end of this paragraph on the high χ^2 values that result from the confrontation.

8.9.1 The impact of influence functions

To illustrate further the impact of the influence functions on the results we plotted the weighted mean of the modelled double differences against the measured double difference in one cross plot (Figure 91). This figure is compiled using the results for only the Time-decay model, but similar results can be produced for the other compaction models, leading to the same conclusion. The RMS values, shown in the legend are a measure of the mismatch between data and model and show a

lower value for the salt influence function and the lowest value for the Knothe influence function. Purely based on the measures for the mismatch like the χ^2 and RMS one would conclude that the Knothe influence function would be the best choice. The drawback of this function is however the lack of a physical underlying basis. As described earlier, the combination of a stiff burden around the reservoir and a layer of viscous salt would in our opinion lead to similar results, but at this moment AEsubs faces some restrictions for this combination and for one the Knothe influence would be the best influence function to be used in further calculations.

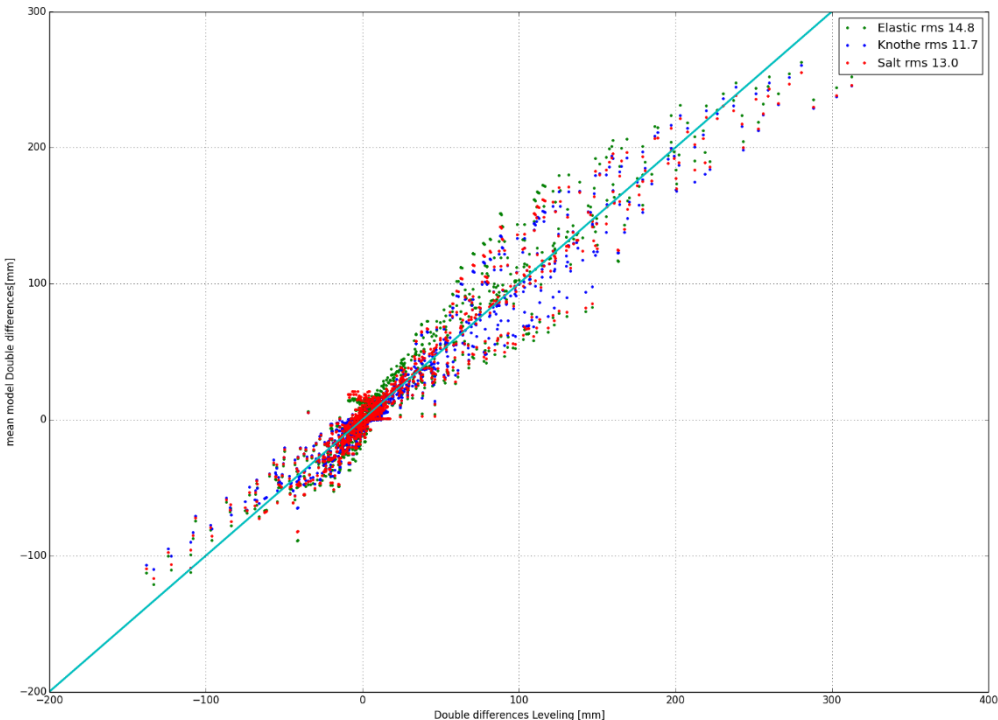


Figure 91: mean model double differences vs data double differences for the Time-decay model for each influence function.

8.9.2 The impact of compaction models on the results

Four compaction models are used in the confrontation with the data. The number of parameters is part of the definition for the χ^2/N but has a negligible impact if the number of double differences is large like in our case. It is therefore likely that a more complex model like the RTCiM model should honor the data better and return lower values for the χ^2/N than simpler models like a linear and Time decay compaction model. From the models result we conclude that both the Time decay and RTCiM model return similar values for the χ^2/N . The Bi-linear model also shows a relative low value for the χ^2/N , but we know from its definition that this model is not able to physically match compaction without depletion, a mechanism that is likely to occur in the Ameland field. Therefore, we recommend to use RTCiM and Time-decay compaction models in follow-up studies.

The impact of both the different compaction model and the elastic, viscous salt and Knothe influence function is visualized in Figure 92. This figure shows the probability of the total subsidence volume after total production of the gas in the Ameland field. Overall we observe that the Knothe influence function on average leads to lower volumes for the subsidence with the lowest volumes for the bi-

linear model. The figure shows as well that the probability density functions of both the Time-decay and RTCiM are quite comparable. Larger volumes are predicted for the elastic influence function in AESubs reflecting the wider subsidence bowls that arise from this influence function.

The calculated subsidence rate for the Knothe influence function is in general lower than found with the AESubs models. Also the subsidence volume calculated with the Knothe influence function is lower (Figure 92). This can be explained by the fact that the subsidence bowl using the Knothe influence function is steeper, this is shown in Figure 93 where an east-west subsidence profile over the Ameland Island is plotted for the best RTCiM model using resp. the Knothe and AESubs influence functions for the viscous salt influence function the Time decay compaction is used. The subsidence bowl of the AESubs model is wider and results in a higher compaction and subsidence volume. In general, the Cm values for the AESubs elastic influence function are higher than the members using other influence functions.

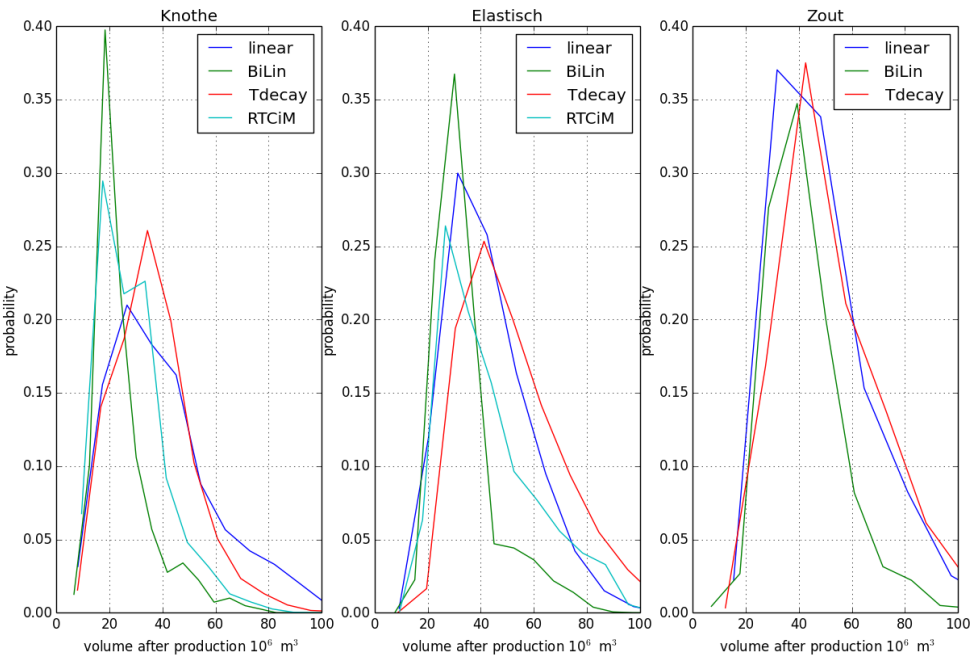


Figure 92: Volume distribution plot for the different influence functions.

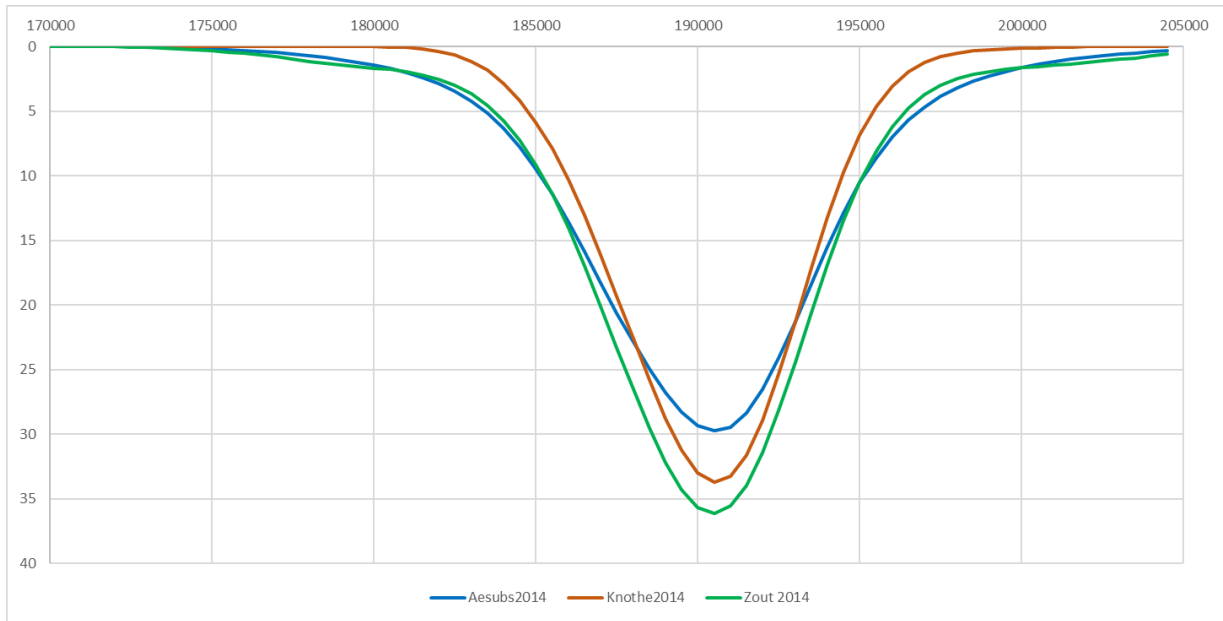


Figure 93: W-E subsidence profile over the island (Y=609000) for the best RTCiM model using the different influence functions.

The combined effect of the Time decay compaction model parameter variation, reservoir pressures and the different influence functions is shown in Figure 94. Significant impact on the rates results from the Knothe influence function. The salt and elastic influence functions are giving more or less similar results as the effect of a stiff underburden or a viscous salt on the rate is more or less the same.

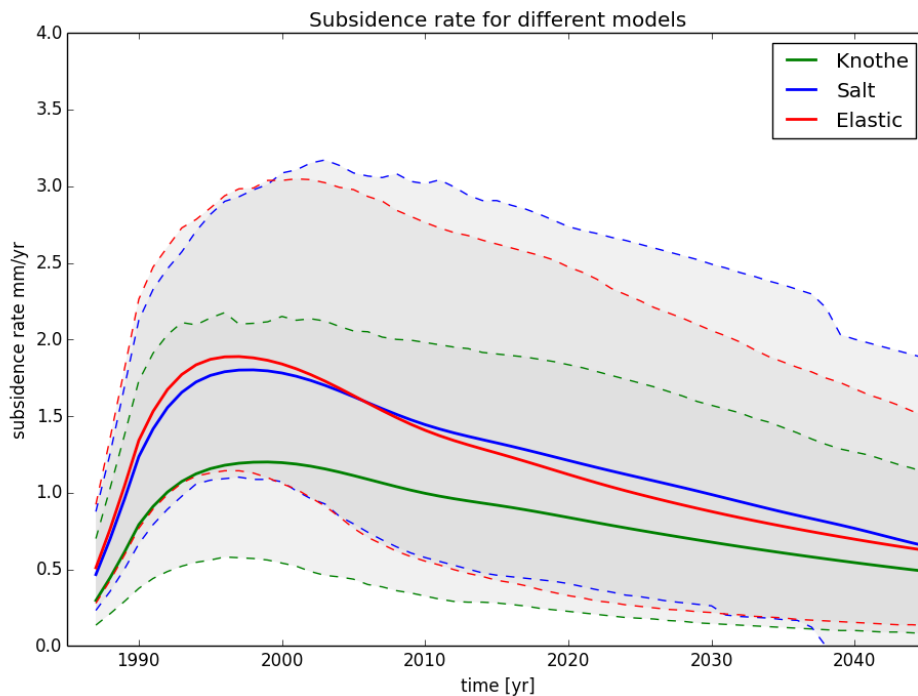


Figure 94: Subsidence rate including uncertainty for the Time-decay for the different influence function.

8.9.3 The impact of reservoir scenario on the results

Thirteen reservoir scenarios were selected to be used in the ESIP calculation. Table 10 shows the χ^2/N values for each reservoir model and influence function. In this table the results for the different compaction models were aggregated (excluding the RTCiM as this model was not run with the viscous salt influence function). The RE models are sorted starting with the best model (C8_I3). Table 11 shows the χ^2/N values per compaction model and influence function pointing out that the RTCiM and Timedecay have better (lower) χ^2/N values.

Table 10: Overview of all χ^2/N values per influence function and reservoir model.

Re Model	Elastic	Knothe	Viscous salt	Average
'C8_I3'	16.3	15.9	15.6	15.9
'C20_I4'	16.5	16.0	16.1	16.2
'C19_I4'	17.4	16.8	16.9	17.0
'C79_I3'	17.7	16.5	17.4	17.2
'C18_I2'	17.8	17.2	17.4	17.5
'C101_I3'	18.8	16.8	18.3	18.0
'C24_I2'	18.4	17.7	17.9	18.0
'C30_I3'	18.7	16.8	18.4	18.0
'C24_I3'	18.8	16.8	18.5	18.0
'C11_I4'	20.3	17.6	19.7	19.2
'C84_I3'	20.4	17.7	19.7	19.3
'C74_I3'	20.8	18.0	20.1	19.7
'C20_I2'	22.5	21.1	21.6	21.7
	18.80	17.30	18.28	

Table 11: Overview of all χ^2/N values per compaction model and influence function.

Compaction model / Influence function	Bi-linear	Linear	RTCiM	Time-decay
Elastic	20.5	25.5	20.7	17.3
Knothe	20.3	25.5	18.7	16.5
Viscous salt	18.2	21.1		16.6

8.9.4 Discussing the χ^2/N values

We have shown a relative good match of the AEsups models with the Geomec model and as well we observe from the cross plots, presented earlier in this chapter the reasonable match with the data. Still the values for the χ^2/N for even the best member are around 15, implying a “mismatch” of the model result of around 4 times the standard deviation of the data. Besides the goodness of fit, the χ^2/N values for the various scenario-model combinations only vary over a small range. This influences the power to make a distinction between the scenarios. The high values were discussed in the December progress meeting with the technical steering committee and it was assumed that the precise measurements could explain the poor goodness of fit. It was proposed to include the variance/covariance matrix of the models in the equation but the result was not as expected. Indeed, the absolute values were reduced with some 10% but not to the expected value of around 1. Also, posterior probability of the possible scenario-model combinations showed even less variation. Subsequent multiple meetings and discussions between TNO, NAM and TU-Delft did not lead to a clear explanation and understanding of the observation described above.

Nevertheless, 4 reasons were identified to possibly explain the high values of χ^2/N :

1. Behavior of the covariance.
2. Outliers having a large effect on the values.
3. Uncertainty of the models.
4. Models unable to describe the measurements.

A mixture of these reasons would explain the behavior as well. Each of the identified reasons will be discussed in more detail in the paragraphs below.

8.9.4.1 Behavior of the covariance.

As a result of the redistribution of the differences between data and model through the covariance matrix, it is difficult to visualize if there are any specific double differences, areas or epochs that add to the high χ^2/N value. In an attempt to somehow dissect the matrix and allow some form of visual inspection, we do a rough verification of the influence of covariance matrix by looking at the variances (values on the diagonal of the covariance matrix) and the covariance separately.

The weighted covariance matrix for the posterior model ensemble is calculated by:

$$C_{jk}^{post} = \frac{1}{1 - \sum_{i=1}^{Ne} w_i^2} \sum_{i=1}^{Ne} w_i (dd_{ij}^{prior} - \mu[dd_j^{post}])(dd_{ik}^{prior} - \mu[dd_k^{post}])$$

C_{jk}^{post} is the posterior covariance between the j^{th} location and the k^{th} location with the weighted mean for the posterior double differences calculated by the sum of the weighted (w_i) prior modelled double differences for each member of the ensemble Ne :

$$\mu[dd^{post}] = \sum_{i=1}^{Ne} w_i dd_i^{prior}$$

In Figure 95 we plotted for a set of models (salt in AEsups with time decay compaction) compared to the data. In the top graph, the absolute difference between the posterior weighted average of the models and the data is plotted in black. A dashed blue line can be seen, having values typically around 3mm, being the standard deviation of the data (see also Figure 16). These standard

deviations are taken as the square root of the values on the diagonal of the covariance matrix of the data, which are the data variances. The same is done for the model covariance matrix, from which the standard deviations are shown in the dashed red lines. By using the diagonal values only, we are ignoring the covariance intensities for now.

It is clear by comparing the black and the blue lines that the model is frequently a multiple of data-standard deviations away from the data. When including the uncertainty of the model, and combining the two standard deviations (of data and model), the green line suggests that given the data and model uncertainty (and ignoring the covariance), the two agree reasonably well.

The bottom graph illustrates the effect of including the model uncertainty. The data plotted is the absolute difference between data and model (black line in top graph), divided by the standard deviation. The red data uses only the data standard deviation here, while the blue data includes both the data and model standard deviations. Where the red dots occasionally reach a value of 10, the blue data are consistently lower, lying within two data+model-standard deviations.

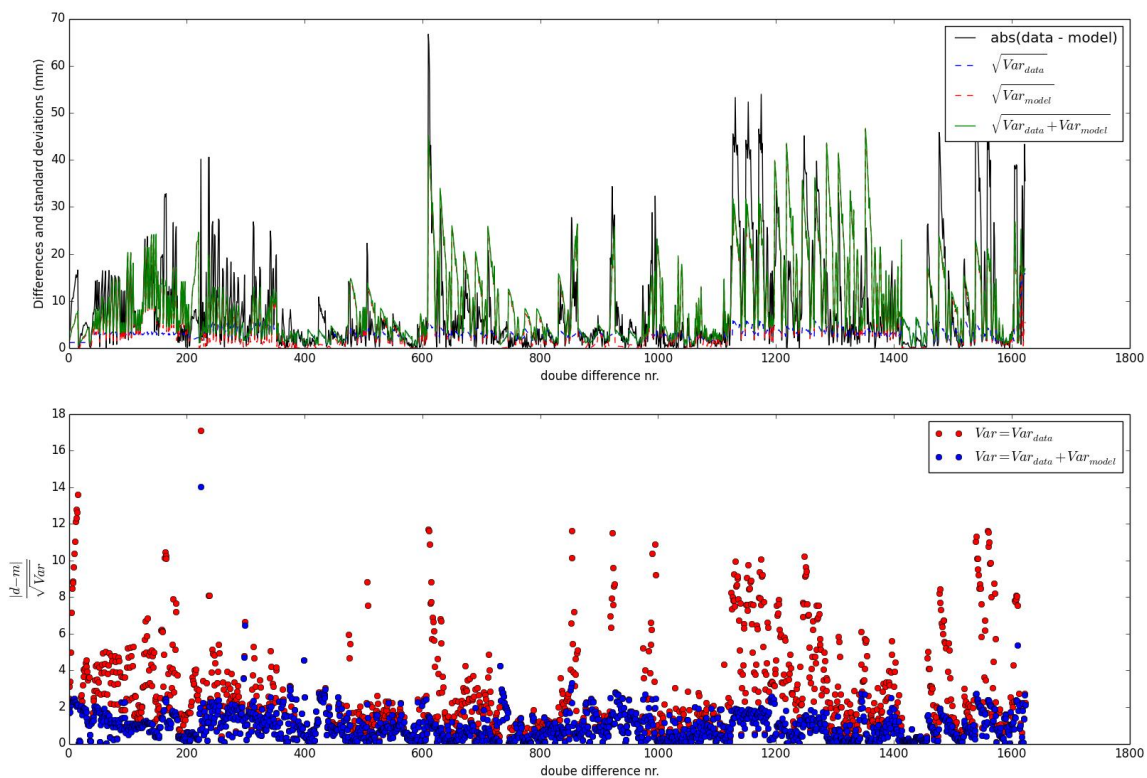


Figure 95: Comparing data with the weighted average of a set of models. Top graph shows the absolute difference between data and model in black. The blue dashed line is the standard deviation of the data, i.e. using the values on the diagonal of the covariance matrix of the data. The red dashed line is the standard deviation from the models (i.e. from diagonal values of the covariance matrix of the models). Green is the combined standard deviation. Bottom graph is the normalised differences of the data. The red dots are the absolute differences divided by the standard deviation of the data, the blue are the absolute differences divided by the combined standard deviation of data and model.

Figure 96 illustrates the reduction of the normalised differences, by presenting the data in histogram form. In red, the differences between data and model are normalised using the standard deviation from the data covariance matrix only, while the blue data applies the combined standard deviation of data and model. The inclusion of the model uncertainty clearly reduces the bandwidth. The data is measured very accurately, reflected in the low standard deviation of the data only, but adding the

larger uncertainty from the model shows a reasonable match between the data and model, considering the individual variances (still not including the covariance). Using the averages of the squared normalised differences as a test statistic, T^2/N , we obtain $T^2/N = 11.9$ when including only the data uncertainty, and $T^2/N = 1.6$ when also considering the model uncertainty. In other words, when, in the crude investigation to the effect of the covariance performed here, only the variances are considered, a low test statistic and equally good match to the data within uncertainty is obtained.

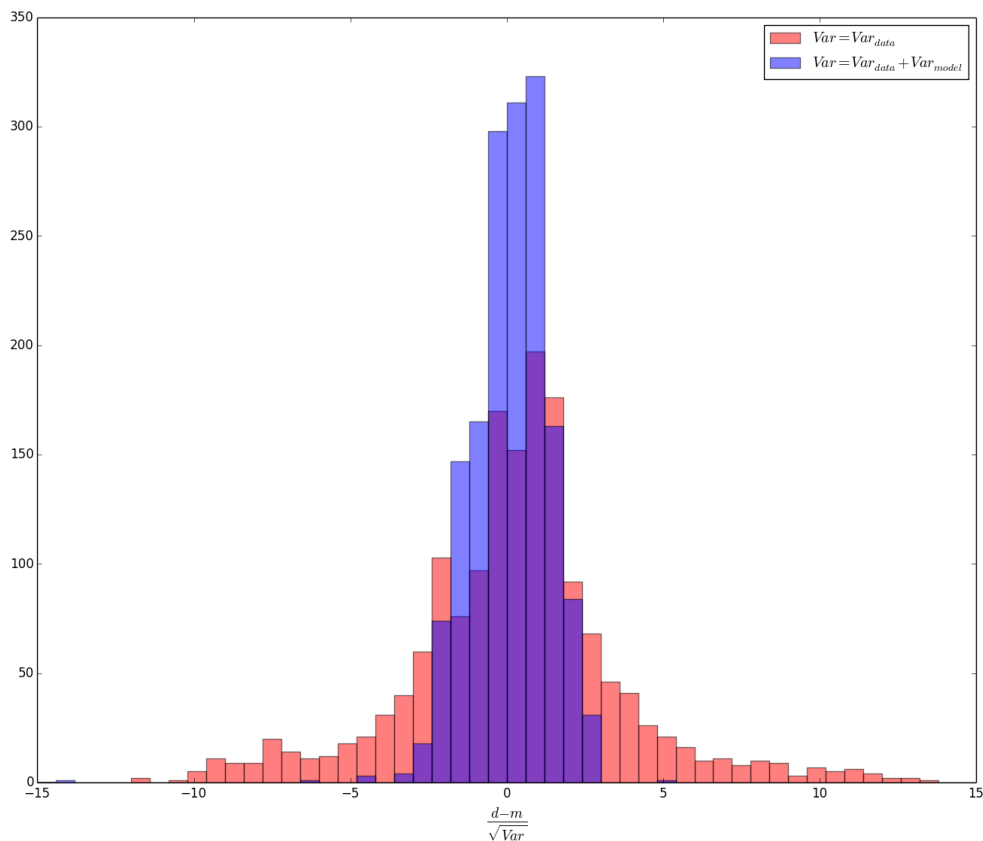


Figure 96: Histogram of the normalised differences between data and model. In red, the normalised differences are generated using only the variance of the data, while in blue the variance of both data and model are used.

However, when now using the full covariance matrix, the effect of the addition of the model uncertainty becomes much less pronounced. To be able to do a similar visualization as above but also including the off-diagonal components, we diagonalise the covariance matrix in order to decorrelate the double differences. We use the eigenvalue decomposition of the inverse of the covariance matrix, using $\Lambda = V^T \cdot C^{-1} \cdot V$, where V contains the eigenvectors of C^{-1} , and Λ is a diagonal matrix where the diagonal is populated with the eigenvalues of C^{-1} . This way, all information from the covariance matrix is contained, and the resulting χ^2 values would be the same as when using the full matrix calculations. The only difference is that the intensity is rotated to a space that is made up out of linear combinations of double differences, and not the actually observed double differences anymore, making a link to the individual data impossible. For the resulting test statistic, however, it is all the same.

The results are shown in Figure 97, where now the difference between data and model is rotated using the eigenvectors, and the variance (Var) is taken from the diagonals of the diagonalised matrix Λ ($\sigma_i = 1/\sqrt{\Lambda_{i,i}}$), which are the eigenvalues of C^{-1} . The effect on the distribution when incorporating the model variances is now much smaller, and accordingly the total test values are also similar. Calculating the test statistic in the same way as above for the variances (being the averages of the squared normalised differences, or the x-values of the histogram), will in the case of the diagonalised matrix be the actual χ^2/N values, since the full covariance information is included. The values obtained are now $\chi^2/N = 16.6$ when only using the covariance matrix of the data, and $\chi^2/N = 14.53$ when also including the covariance of the models. Comparing these test statistics to those calculated in the same way but excluding the covariance information (the T^2/N values), it seems the covariance has a very large impact on the values of the test statistic. Moreover, it seems that when using the full covariance matrix, the inclusion of the covariance of the model has a reduced impact.

Because further visualization of the covariance matrix is difficult, it is not possible to attribute the differences to any particular double differences, or to determine what particular aspect of the covariance causes this distinct increase in test value.

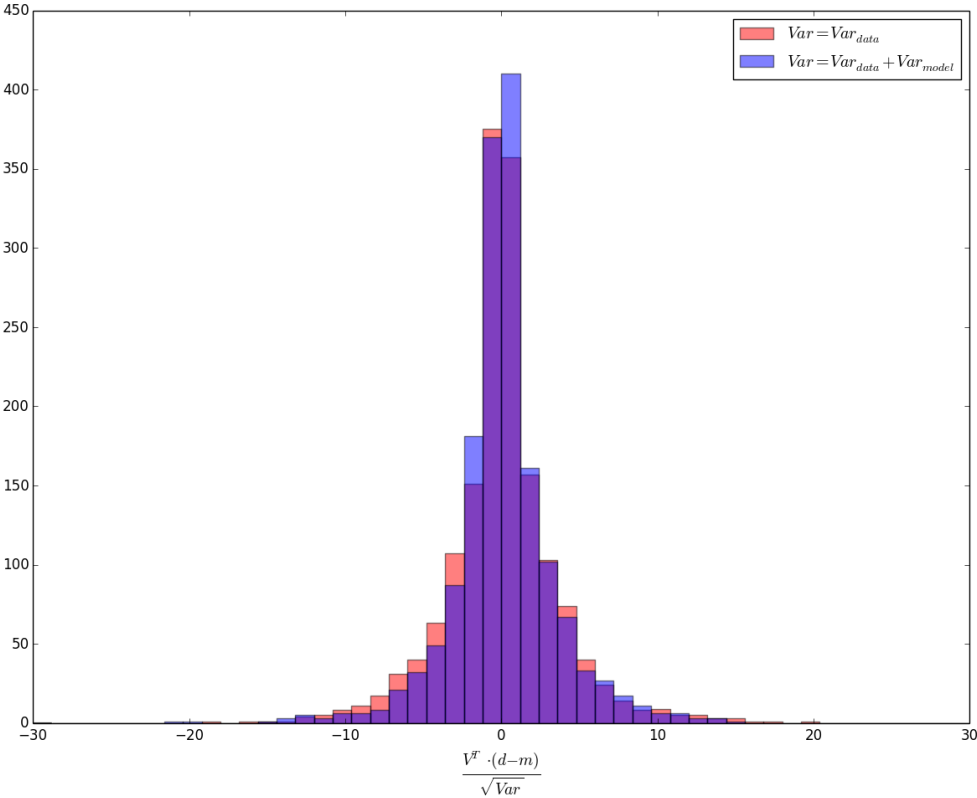


Figure 97: Similar histograms as in Figure 96, but now produced using the full covariance matrix through decomposition to the eigenvalues. The variances used are now taken from the values on the diagonal matrix Λ .

The different test parameters, based on only the variances as well as on the full covariance matrix, are listed in Table 12 for three different subsets of members, being ones that used an influence function that included a salt layer, split into three different compaction models (time decay, linear and bilinear). The three models all have a different test value based on the variances only (the values

on the diagonals of the covariance matrix), however when including the full covariance matrix, all models seem to be of equal probability. Additionally, the test statistic changes from a reasonably low value (close to one) when not including the covariances, and increases to large values when the covariance is included. While such a large effect is not immediately obvious when comparing the measured and modelled double differences, we cannot exclude that some issues exist within the covariance matrices, either the data covariance or the model, or both.

Table 12: Values of test statistics based on the variances only (T^2/N) and total covariance (χ^2/N). Covariance matrices used are the combination of the covariance of the data and the covariance of the model.

	T^2/N (based on variances)	χ^2/N (based on full covariance matrix)
Time decay	1.56	14.53
Linear	3.77	14.66
Bilinear	2.67	14.42

8.9.4.2 Outliers having a large effect on the values.

Because of the square of the error in the test statistic, single outliers can make up a large portion of the total test statistic. In this section, we apply a simple test to verify if (a small set of) outliers could account for the high values of the test statistics observed. It cannot be considered a statistically rigorous quality assessment, because it does not take the correlation structure into account. It will therefore not be used in the actual workflow for removing outliers, but this test is suitable to provide a first insight into the contribution of outliers towards the test statistic.

The left histogram of Figure 98 shows the distribution of the normalized differences between data and model, where the normalization is made by their combined standard deviations. This simplified, variance-only based distribution has an acceptable width where most data is contained within two standard deviations. The average of the differences-squared divided by their variance (which is same as a χ^2/N when all off-diagonal values of covariance are zero, i.e. the T^2/N test in the previous section), is 1.56, which would suggest a reasonable agreement between data and model if covariances were indeed zero. Outside the 5-standard deviation boundaries (red dashed lines in the histogram), we find only three out of the 1624 double differences, which we will consider as outliers in this test.

The graph on the right in Figure 98 shows effectively the same, but now including the full covariance information. To obtain the histogram, we again first diagonalise the covariance matrix to its eigenvalues and transform the differences accordingly, such that all information is preserved (similar to what was done in Figure 97). A direct comparison between individual double differences is no longer possible, but the distribution should give an impression of the level of resemblance of the data to the model, only in a rotated space. The χ^2/N value of the data set can now also be calculated by taking the average of the normalized differences in the transformed space (on the x-axis), which is the same as calculating the full covariance multiplication with the differences between data and model, but easier to visualize.

For the full data set, the $\chi^2/N = 14.53$, which suggests that, even though the data seems to be in good agreement if only variances could be considered, the resemblance between data and model appears worse when including all covariances. After removing the three double differences identified in the

variance-based histogram on the left of Figure 98, the effect after re-diagonalising the covariance matrix after removal of the three double differences, is significant but not very large, as the χ^2/N value drops from 14.53 to 10.40. So excluding some potential outliers does not take the χ^2/N value considerably closer to one. Removing more points does not result in a fast further reduction of the test value.

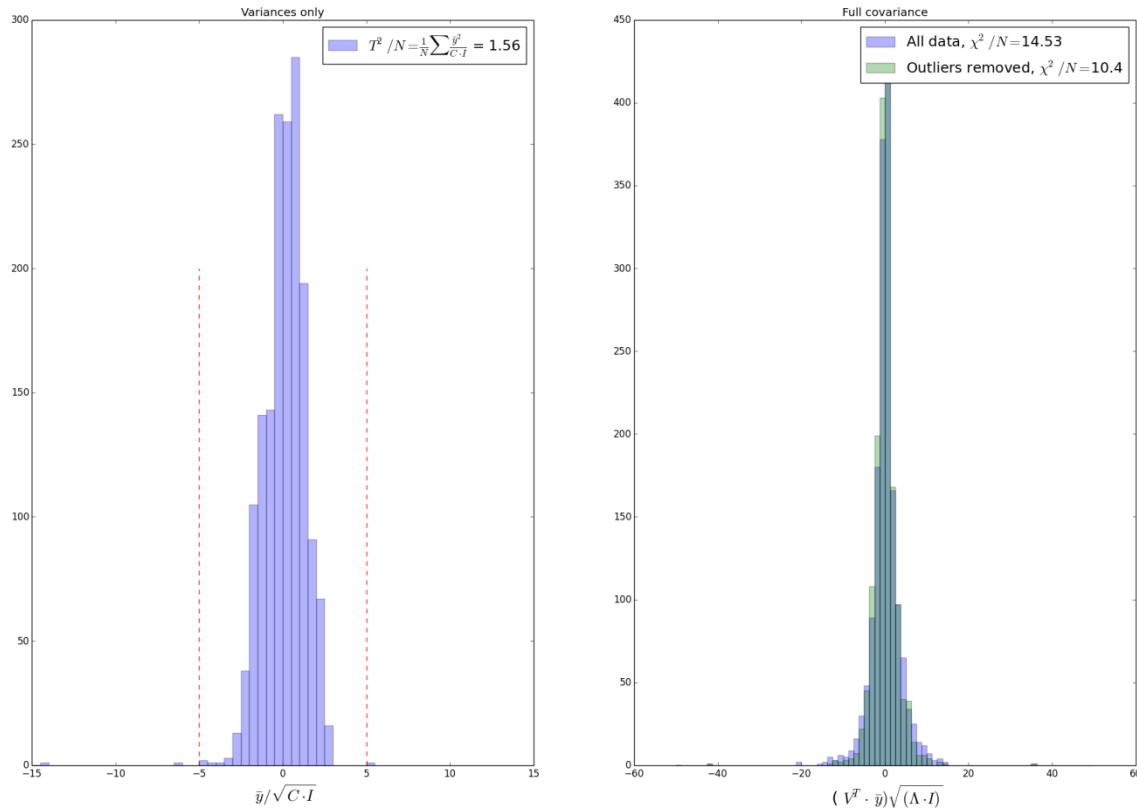


Figure 98: Normalised difference data (left), where \bar{y} is difference between data and (weighted average of) model, C is the covariance matrix of data and model. The figure on the right shows the same, but now after diagonalising the covariance matrix to make visualisation possible. The normalised difference data after diagonalising the covariance matrix to its eigenvalues are then such that all covariances are also accounted for.

The effect of the outlier removal is listed for three different compaction models in Table 13. The test values shown are after the removal of the three outliers. The same outliers (double differences) were removed in all three data sets for fair comparison. In fact, the three outliers that were removed appeared outside the 5-standard deviation mark for all three sets, so could equally be classified as an outlier in all sets. Comparing the data to that for the full sets in Table 12, there is only a bulk shift in the χ^2/N , without increasing any distinction between the three compaction models. In all three cases, the outlier removal does not result in a χ^2/N much closer to one, indicating that outliers alone cannot explain the high test values observed here.

Table 13: Values of test statistics based on the variances only (T^2/N) and total covariance (χ^2/N). Covariance matrices used are the combination of the covariance of the data and the covariance of the model. Data used is the same as in Table 12, but now after outlier removal.

	T^2/N (based on variances)	χ^2/N (based on full covariance matrix)
Time decay	1.40	10.40
Linear	3.58	10.51
Bilinear	2.48	10.29

8.9.4.3 Uncertainty of the models.

In the December progress meeting, it was discussed that the covariance/variance of the model results should be included as well in the χ^2/N definition. The recommendation was implemented but did only manage to reduce the χ^2/N values by a small amount. The explanation of this unexpected result is still lacking and should be investigated before any further implementation of this workflow with the aim to assess subsidence rates in the Waddenzee.

The test statistics shown above in Table 12 and Table 13 already include the covariance matrix of the models as well as those of the data, and still have high χ^2/N values. We did the same as above in Table 12, but now excluding the covariance of the models, to illustrate the effect of the addition of the model uncertainty. The results are listed in Table 14, showing higher values for both the variance-only based test statistic (T^2/N) and for the full-covariance based test statistic (χ^2/N). The effect of the reduction of the T^2/N (test statistic without covariance information) is illustrated in Figure 96: since the uncertainty on the data (ignoring the covariance) is very small, the difference between data and model is relatively large in terms of the number of standard deviations. Including the uncertainty of the models, which is larger than that of the data, the normalised difference between data and model becomes a narrower distribution around zero. The reduction of the distribution of the normalised differences based on the total covariance, like in Figure 97, is however much less well pronounced.

What can further be noted when comparing Table 14 with Table 12, is that the spread in χ^2/N values is reduced when including the covariance of the models. Such a reduction in the spread would imply a loss of differentiation between the models.

The changes in the χ^2/N values for the different setups (covariance of data only, covariance of model and data, covariance of model and data and three outliers removed) are summarized in Table 15

Table 14: Values of test statistics based on the variances only (T^2/N) and total covariance (χ^2/N). Data used is the same as in Table 12, but now only using the covariance matrix of the data, and not including the covariance matrix of the models.

	T^2/N (based on variances)	χ^2/N (based on full covariance matrix)
Time decay	11.9	16.6
Linear	31.5	21.1
Bilinear	19.4	18.1

Table 15: Summary of χ^2/N values for the different cases tested, for three different compaction models.

	Cov = CovD	Cov = CovD + CovM	Cov = CovD + CovM, DD removed (3 out of 1624)
Time decay	16.6	14.5	10.4
Linear	21.1	14.6	10.5
Bilinear	18.1	14.4	10.3

8.9.4.4 Models unable to describe the measurements.

It can be questioned whether the models used can adequately fit the data. A total mismatch of the models with the data would explain a high value for the test statistic. From the cross-plots presented in this chapter we already concluded that the members having lowest χ^2/N values fit with the data within a 2 cm range. An alternative visualization is presented by Figure 99, where a histogram of the data-model differences is presented for the scenario with low aquifer depletion and a combination of Time decay compaction and viscous salt above the reservoir. The standard deviation from this distribution is close to 11 mm.

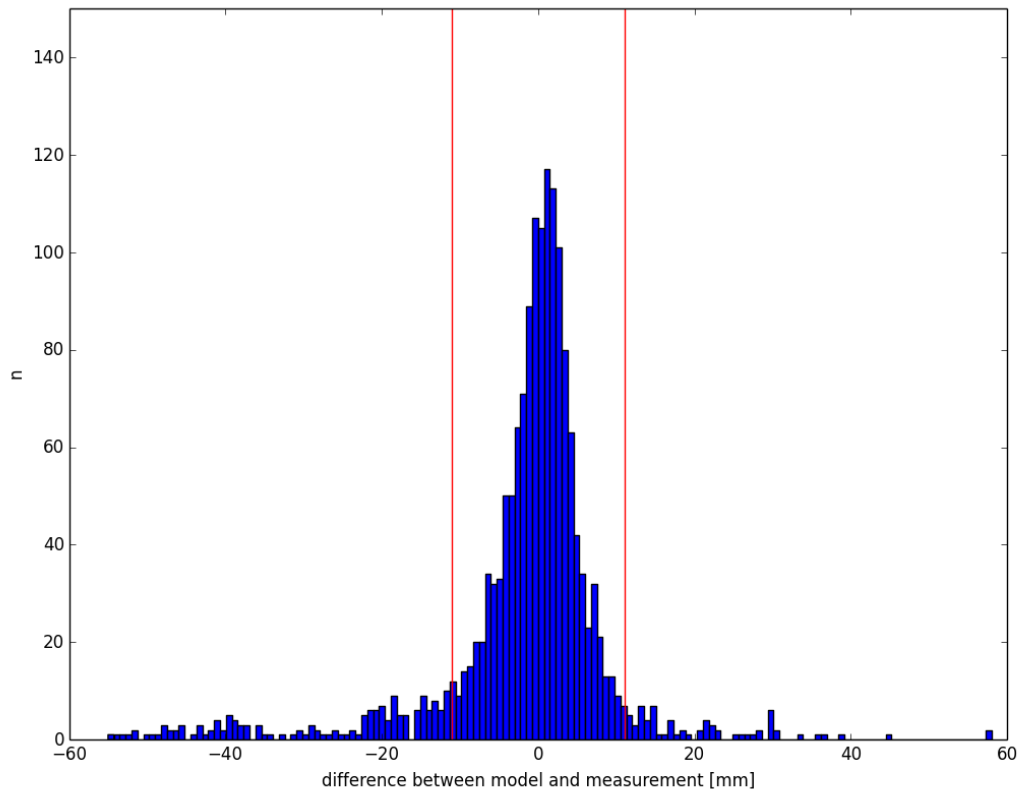


Figure 99: Histogram of the measured - modelled subsidence of the best salt model. The empirical standard deviation from this histogram is 11 mm, which is where the red lines are placed.

The results of the Geomec model used in the measurement and control cycle, based on a similar reservoir scenario have been processed as well to make a double-difference comparison possible. The histogram of that result is presented in Figure 100. This figure closely resembles Figure 99 and shows actually a larger standard deviation for the Geomec model. We conclude therefore that the AEsups results are close to the Geomec model and even closer to the data. To reduce the χ^2/N from

a model perspective would require an alternative model that needs to perform better than a finite element model.

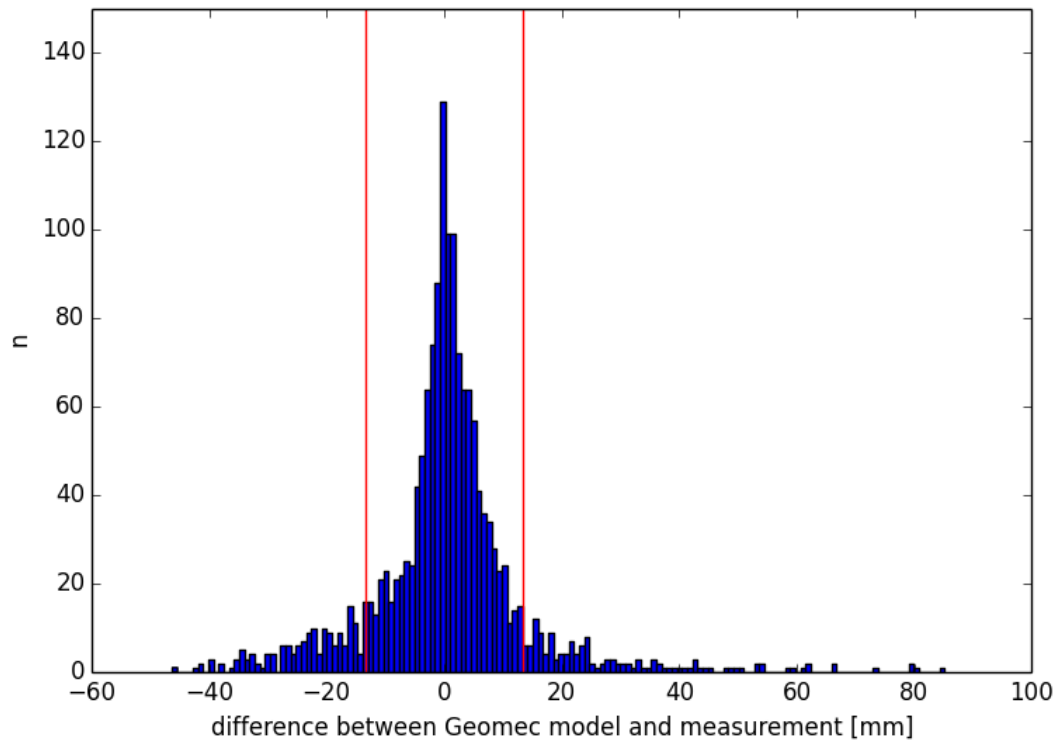


Figure 100: Histogram of the measured - modelled subsidence of the Geomec model. The empirical standard deviation from this histogram is 11 mm, which is where the red lines are placed.

To demonstrate the clear differences in results using the 13 different reservoir scenarios and the ability of some members to match the measured subsidence, we selected a number of benchmarks and for these benchmarks we choose a double difference that covers the youngest period over a reasonable time span. The period of 2007-2014 is in that sense a logic choice as the measurements in the Waddenzee started around the year of 2007. The model results are visualized (Figure 101) for the best member (lowest χ^2/N , blue dots), the weighted average of the ensemble for the individual reservoir scenarios (green dots) and the 95% confidence range. The top 4 graphs represent benchmarks located in the Waddenzee that are more indicative for the differences in aquifer depletion of the 13 reservoir scenarios. The bottom 4 graphs show the results for benchmarks selected on the island, note that the models with the least aquifer depletion (C8_I3 and C20_I4) fits best to the data. The bottom 4 graphs represent the values for model and data on benchmarks located on the island of Ameland. The results from the different scenarios show less variation in the subsidence at these locations, which is expected. These benchmarks reflect the pressure behavior in the gas part that is better constrained by the well data and therefore surrounded by less uncertainty.

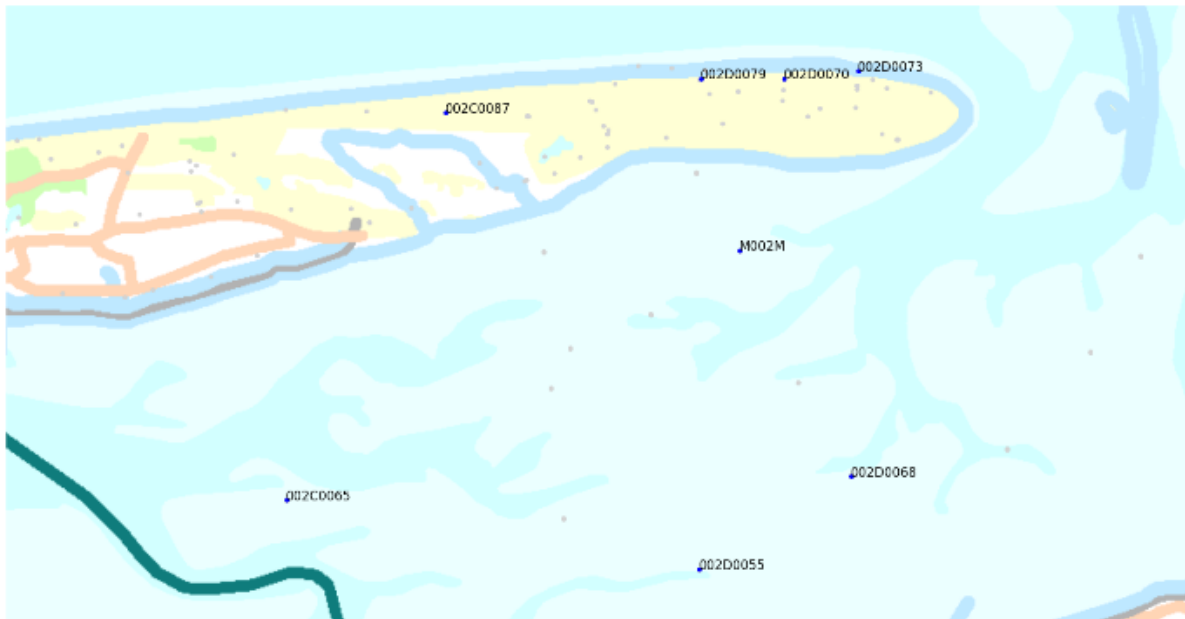
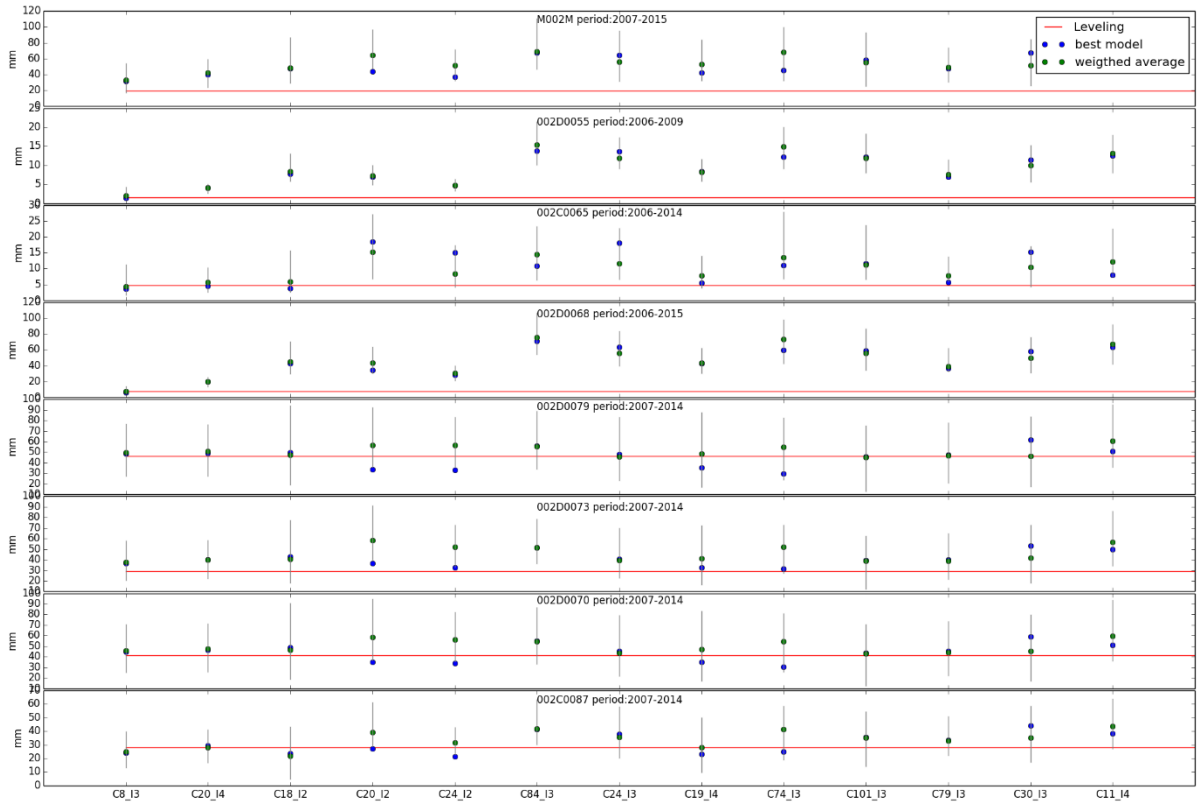


Figure 101: Double differences (top figure) for the Time decay compaction model using viscous salt per reservoir scenario at a selection of benchmarks (see bottom map for the location of the bench marks) for the period 2007 – 2014. The red line indicates the double difference of the observation, the blue dot the double difference of the best model and the green dot representing the weighted average of the individual ensembles. Grey vertical lines represent a 95% uncertainty band.

9 Implications of the LTS II results on subsidence forecasts for the “Wadden fields”

One of the main objectives of the LTS II study is to test a stochastic workflow that can be applied to the measurement and control cycle for the Waddenzee. We conclude that, rather than providing a deterministic expected scenario (NAM, 2016b), we are now able to demonstrate the capability to provide a weighted distribution with only the data deciding on the outcome. This statement is maybe too bold, as some subjective choices will always have an effect on the outcome. Parameter ranges can be different than assumed in the LTS II calculations, compaction models may be improved or new models and influence function may be identified and selected in future applications of the workflow. In LTS II we conservatively used an equal probability for all scenarios whereas it is noted that all the reservoir scenarios have a history match RMS themselves when comparing Mores model results to the production and pressure data. This RMS could be used as a priori knowledge, redefining prior distribution and also impacting the posterior distributions.

From the discussion in the previous paragraph we concluded as well that the values for the test statistic are high and show a small variation amongst the different scenario-model combinations. This implies that also less probable reservoir scenarios will have a large impact on the weighing of the results. For the possible subsidence rates this implies a large uncertainty bandwidth and relative high subsidence rates for the statistical expectation case.

This is illustrated by Figure 102 where the upper bound of the blue surfaces represents the expectation case that follows from the total distribution arising from all reservoir scenarios using a Time decay compaction model and an elastic influence function. The differentiation in colors represents the relative contribution of each reservoir scenario to the expectation case. The M&R report concluded that the deterministic model based on low aquifer depletion reservoir scenario fits the data best. This conclusion is confirmed by the LTS II workflow with C8_I3 and C20_I4 having the highest probability (Figure 103) but the other reservoir scenarios still have a relative large impact on the total weighted average leading to higher values for the stochastic expectation case when compared to M&R expectation case.

The main issue with this conclusion is that the derived stochastic expectation case is not close to scenario-model combination that fits best to the data. Also in the discussion paragraph we showed that different definitions of the test statistic would lead to very different results. Without a satisfying explanation on the role of the covariance matrix of data and model on the test statistic, the conclusion is that it is too early to make a direct comparison with the measurement and control cycle prediction possible nor we can make a firm statement on the impact of this workflow on the Wadden Fields. There are however clear differences in the structural setting of the Wadden Fields like Nes and Moddergat. The Wadden Fields are bounded by large faults that prevent a connection to large lateral aquifers. Therefore, we expect less uncertainty arising from the possible reservoir scenarios for these fields. However, the constraint of possible scenarios by the data is poorer because of the lifetime of these fields. The last observation probably results in a larger uncertainty of the subsidence forecasts.

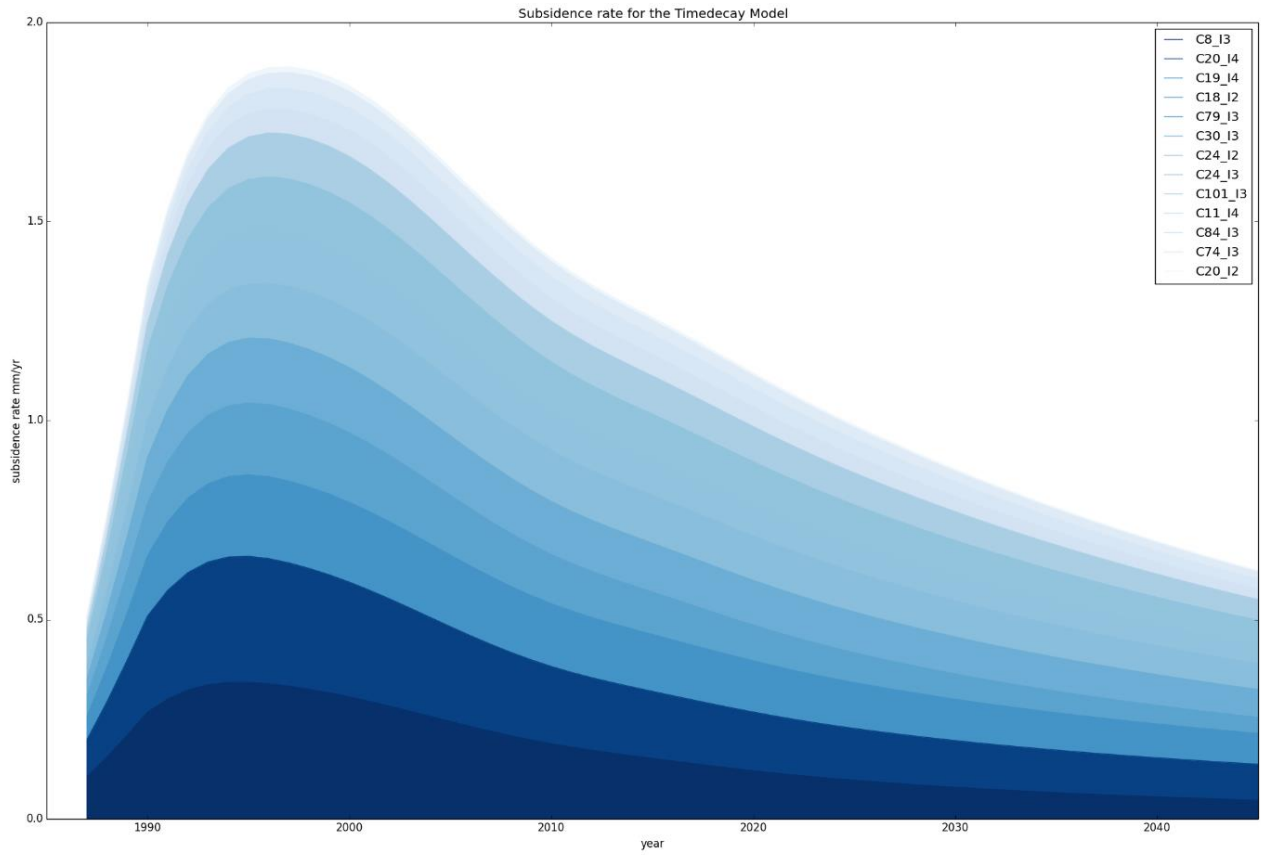


Figure 102: Subsidence rate, with the contribution per reservoir model.

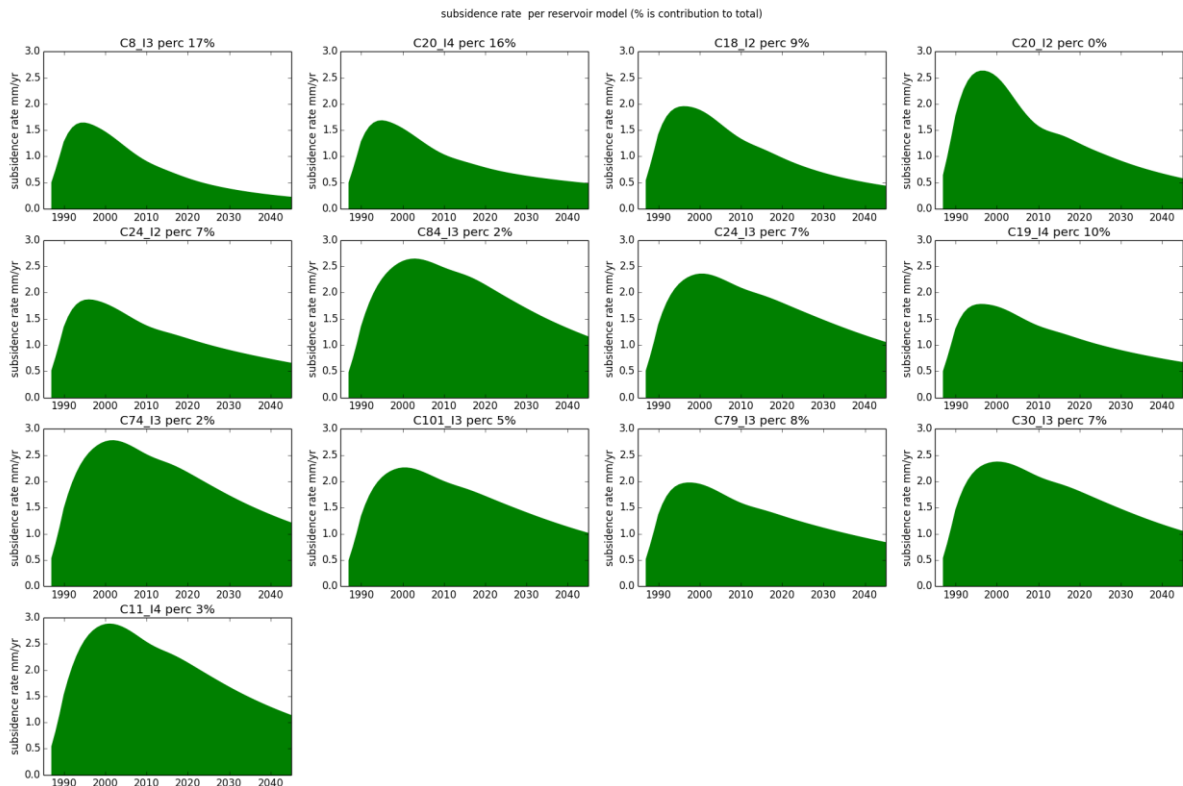


Figure 103: Subsidence rates for the Time decay model per reservoir scenario and using the elastic influence function, the percentage in the title is the part which this reservoir model contributes to the total.

10 Conclusions

In the LTS II study a workflow was created to allow confrontation of model results with measurements in an objective way. This workflow incorporates the findings of the LTS I study and is able to identify the most likely model factors like reservoir scenarios, compaction models and influence functions. Moreover, posterior (after confrontation with the data) probability distributions for the input data show redefined distributions for the input values of the data, demonstrating the effect of the confrontation on the probability of certain input values.

Multiple tests between NAM and TNO ensured the consistent usage of the workflow software code (ESIP) and proved the capability to produce meaningful results. A special synthetic “dummy” model was created to better understand the findings. Comparison between the semi-analytical approach of AEsups including a viscous salt layer with the finite element results from Geomec shows only small differences between the two approaches. The use of geodetic data is characterized by several significant innovations. Observations from levelling and GPS techniques have been used in form of spatio-temporal double differences. Outlier removal follows a more formal and less subjective approach. Uncertainties are described by a fully populated covariance matrix, that also takes shallow movements into account. All these measures help getting the modelling closer to reality by reducing previously adopted assumptions and idealizations.

To describe the uncertainty in possible pressure scenarios for Ameland, 193 reservoir scenarios for Ameland were created. Thirteen of these scenarios were selected with the aim that this selection represents the full bandwidth of the possible 193 scenarios. It is assumed in the confrontation with the data that the prior (prior to confrontation) probability is the same for all scenarios.

On top of the pressure variation we also varied compaction models and influence functions. Each of these factors has its own parameter uncertainty distribution of which the values are probed by the Monte Carlo calculation.

Each parameter pick by the Monte Carlo procedure results in a geomechanical model member with a group of members defining the ensemble. For presentation purposes we choose to mimic the “Waddenzee” measurement and control cycle (e.g. NAM 2016a) by calculating the 6 year averages of the subsidence rate for a given area. Rather than deterministic scenarios we are now able to provide the expected weighted average case including confidence bounds for the ensemble to express the uncertainty.

The same workflow was applied on the emergency stop scenarios, resulting in a clear expression on the subsidence rates following a hypothetical stop in 1996 and a less clear expression if the stop was made in 2016. The latter can be explained by the small response of the pressures as the gas field has produced most of the gas at this time. We therefore conclude that an emergency stop scenario in the Waddenzee for the “younger” fields like Nes and Moddergat would result in a significant reduction of the subsidence rate in the Waddenzee.

As described in the work plan (NAM, 2016) we selected TNO’s Red Flag methodology (Nepveu et al., 2010) to confront model results with data. This methodology is transparent because it will not change the input parameters that defined the prior ensemble. Most likely reservoir scenarios are the scenarios that show no or low depletion in the aquifers. The most likely compaction model is the Time decay and RTCiM model.

At present the Knothe influence function describes the data best, but it should be noted that the AEsubs simulator that is used for the elastic and salt influence does not allow for combinations between salt and elastic layers having large contrasts in the values of the stiffness parameters. When this is corrected, we expect this combination to give lower values of the test statistic because it should match the spatial-temporal behavior of the subsidence better. The Knothe function allows for a fit to a steeper subsidence bowl but lacks a temporal component that can describe a possible change in steepness of the bowl.

The ability of the workflow to differentiate the various scenario-model combinations appears to be low for the actual Ameland model. However, on a synthetic model the process is capable to differentiate the original input model. Further discussion and investigation is required to understand why the process is not capable of differentiating between the real scenario-model combinations before this can be implemented in in the official measurement and control cycle.

Reflecting on the objectives as defined in the workplan we conclude the following:

1. A workflow that can test the probability of possible subsidence models in an objective way, with the possible hypotheses resulting from the LTS I study taken into account

We have demonstrated a workflow capable of testing the probability of subsidence models incorporating the possible factors of influence as defined by LTS I.

2. To provide an ensemble of likely subsidence forecasts, each member with a different effect on the future subsidence above the Ameland field. A P50 should be presented from the total distribution.

An ensemble of results has been presented in this report. All statistical values can be derived from the probability distribution of the ensemble like confidence bonds, expectation case and P50 values.

3. An analysis comparing the new distribution of subsidence forecasts with the Ameland forecasts selected in the M&R cycle report over 2015 along with a description of the discrepancies and the possible consequences this may have for M&R 2016 cycle.

Selected geomechanical members based on the salt influence function were compared to the Geomec M&R 2015 results and we concluded that a good match between these models is found. However, the results obtained so far of the stochastic expectation case are very different from the results that follow from the deterministic expected case as presented in NAM (2016a). We have confirmed in LTS II that the reservoir scenario and compaction model used in the M&R cycle 2015 are the most likely reservoir scenario and compaction model. We therefore conclude that expected case for Ameland will remain the same for M&R 2016.

4. The impact of this result on the other gas fields that are part of the M&R cycle will be made clear in a qualitative sense and based on analogies. (Within the given LTS II time schedule it will be not possible to set up a similar project for the remaining fields.)

There are clear differences in the structural setting of the Wadden Fields like Nes and Moddergat. The Wadden Fields are bounded by large faults that prevent a connection to large lateral aquifers. Therefore, we expect less uncertainty arising from the possible reservoir scenarios for these fields. However, the constraint of possible scenarios by the data is poorer because of the lifetime of these fields. The last observation probably results in a larger uncertainty of the subsidence forecast.

Therefore, at present it is difficult to make a strong conclusion on the impact of the LTS II results on subsidence prediction in the Waddenzee arising from the Wadden Fields. We firstly need to decide on certain conditions e.g. do we use all compaction models for the Waddenzee or only one or two; certain methods for the confrontation e.g. Red Flag versus Ensemble Smoother and certain definitions for e.g. the test statistic.

5. NAM will start with the application of the LTS II workflow to the other Wadden Fields. The first results will be expected to appear in M&R report over the year 2017.

This will be executed under the premise that the LTS II workflow is accepted by the stakeholders.

6. An analysis of the effect of a hypothetical stop of production of the Ameland gas field at specific point(s) in time, including a statement on the analogy with the other Wadden Fields and the effectiveness of the “hand on the tap” principle.

The hand on the tap principle has been demonstrated for different ensembles. A production stop in the “Wadden Fields” would have an even more profound effect because these fields connect to a lesser extend to lateral aquifers.

7. A report that clarifies the effect of an overlying salt layer on the subsidence with a comparison between the results from the analytical AEs subs geomechanical model and the results from the Ameland Geomec model.

Agreement between the models was found. A chapter is included in this document that covers this comparison.

11 References

Bjerrum, L., 1967, Engineering Geology of Norwegian Normally Consolidated Marine Clays as Related to the Settlements of Buildings, *Geotechnique*, 17, pp 81-118.

Breunese, J.N., Eijs, R.M.H.E. van, Meer, S. de & Kroon, I.C. (2003) Observation and prediction of the relation between salt creep and land subsidence in solution mining - The Barradeel case. Solution Mining Research Institute Fall meeting 2003, October 5-8, Chester, United Kingdom, abstract.

Cai, Y., Verdel, T., Deck, O. (2014) On the topography influence on subsidence due to horizontal underground mining using the influence function method. *Computers and Geotechnics* 61. Pp. 328–340

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

<http://repository.tudelft.nl/islandora/object/uuid:b805782b-2eb4-4f72-98f4-f727c4ea9df0?collection=research>

Fokker, P.A. (1995) The behaviour of salt and salt caverns. Doctoral Thesis.

<http://repository.tudelft.nl/islandora/object/uuid:6847f8e4-3b09-4787-be02-bcce9f0eed06?collection=research>

Fokker, P.A., Orlic, B. (2006) Semi-Analytic Modelling of Subsidence. *Mathematical Geology*, Vol. 38, No. 5, July 2006. DOI: 10.1007/s11004-006-9034-z

Fokker, P.A., Wassing, B.B.T., van Leijen, F.J., Hanssen, R.F., Nieuwland, D.A. (2016) Application of an ensemble smoother with multiple data assimilation to the Bergermeer gas field, using PS-InSAR. *Geomechanics for Energy and the Environment*, Vol. 5, No. March, 2016, p. 16-28.

Geertsma, J. (1973) , Land Subsidence Above Compacting Oil and Gas Reservoirs, *J. Petr. Tech.*, pp.734-744.

Geertsma, J. and van Opstal, G. (1973). A Numerical Technique for Predicting Subsidence Above Compacting Reservoirs, Based on the Nucleus of Strain Concept. *Verh. Kon. Ned. Geol. Mijnbouwk. Gen.*, 28, pp. 63-78.

Hejmanowski.R., Sroka A. (2000) Time-space ground subsidence prediction determined by volume extraction from the rock mass: Land subsidence : proceedings of the sixth international symposium on Land subsidence : Ravenna 24–29 September 2000. Vol. 2: Measuring and monitoring theory and modeling, 2000. pp. 367–375.

Knothe S. (1953) Równanie ostatecznie wykształconej niecki osiadania. *Archiwum Górnictwa i utnictwa*, tom 1, z.1, 1953. (Equation of a finally developed depression. In: *Mining and Metallurgy Archives*, vol. 1, bul. 1)

van Leijen F., S. Samiei-Esfahany, H. van der Marel, R. F. Hanssen (2017): Uniformization of Geodetic data for deformation analysis – Contribution to the research project: Second phase of the long-term subsidence study in the Wadden Sea Region (LTS2). Delft University of Technology.

Marketos, G., C. J. Spiers, and R. Govers (2016) Impact of rock salt creep law choice on subsidence calculations for hydrocarbon reservoirs overlain by evaporite caprocks, *J. Geophys. Res. Solid Earth*, 121, 4249–4267, doi:10.1002/2016JB012892.

McKie, T. (2011), A comparison of modern dryland depositional systems with the Rotliegend Group in the Netherlands, in *The Permian Rotliegend of the Netherlands*, J. Grötsch and R. Gaupp (eds.)

NAM (2011) Subsurface Technical Report Subsidence Modeling of Ameland Fields. UIE Report No: EP201105208617

NAM (2015a) Wadden Sea Long term Subsidence Studies –Overview report. EP201506209625

<http://feitenencijfers.namplatform.nl/download/rapport/2ca6c8d8-c0d4-4c10-8460-672f93b4cdaa?open=true>

NAM (2016a) Werkplan Long Term Subsidence Fase II (LTS II). Bijlage 4 in Gaswinning vanaf de locaties Moddergat, Lauwersoog en Vierhuizen Resultaten uitvoering Meet- en regelcyclus 2015. EP201602210605

<http://www.commissiener.nl/projectdocumenten/00000929.pdf?documenttitle=MR%20rapportage%20%202015.pdf>

NAM (2016a) Gaswinning vanaf de locaties Moddergat, Lauwersoog en Vierhuizen. Resultaten uitvoering Meet- en regelcyclus 2015.

<http://www.commissiener.nl/projectdocumenten/00000929.pdf?documenttitle=MR%20rapportage%20%202015.pdf>

NAM (2016b) Gaswinning vanaf de locaties Moddergat, Lauwersoog en Vierhuizen: Integrale beoordeling en samenvatting van de monitoringresultaten over 2015.

<http://feitenencijfers.namplatform.nl/download/rapport/dcf5fe13-a1aa-4dea-84e5-dff484326f68?open=true>

NAM (2016c) Technical Addendum to the Winningsplan Groningen 2016 Production, Subsidence, induced Earthquakes and Seismic Hazard and Risk Assessment in the Groningen Field. Part II- Subsidence

<https://www.sodm.nl/binaries/staats-toezicht-op-de-mijnen/documenten/publicaties/2016/06/21/4--technical-addendum-to-the-winningsplan-groningen-2016-chapter-6/4-technical-addendum-to-the-winningsplan-groningen-2016-chapter-6-15042016.pdf>

Nepveu M., Kroon, I. C., Fokker, P. A. (2010) Hoisting a Red Flag: An Early Warning System for Exceeding Subsidence Limits. *Mathematical Geosciences*. February 2010, Volume 42, Issue 2, pp 187–198

Samiei-Esfahany S., H. Bähr (2015): Research and Development Project for Geodetic Deformation Monitoring – Contribution to the research project: “Long-term study on anomalous time-dependent subsidence in the Wadden Sea Region”. NAM report EP201505216980.

<http://feitenencijfers.namplatform.nl/download/rapport/6af45eca-b02f-4844-8adf-7835b5e0ee55?open=true>

Samiei-Esfahany S., H. Bähr (2017): Erratum to the report: "Research and Development project for Geodetic Deformation Monitoring" – Revision of recommendations from the project: "Long-term study on anomalous time-dependent subsidence in the Wadden sea region". NAM report EP201701215912.

Spiers, C.J., Urai, J.L., Lister, G.S., Boland, J.N., Zwart, H.J.,(1986) The influence of fluid–rock interaction on the rheology of salt rock. EUR 10399 EN, Office for Official Publications of the European Communities, Luxembourg, 131 pp.

Spiers, C.J. & Carter, N.L.(1998) Microphysics of rocksalt flow in nature. In: M. Aubertin & H.R.Hardy (Eds.) The Mechanical Behaviour of Salt: Proceedings of the Fourth Conference Series on Rock

Sroka A., Tajdus K. and Preusse A. (2011) Calculation of Subsidence for Room and Pillar and Longwall Panels, 11th Underground Coal. Operators' Conference, University of Wollongong & the Australasian Institute of Mining and Metallurgy, 2011, 83-90.

TNO (2011) Toetsing van de belasting op de gebruikruimte in de kombergingsgebieden in Pinkegat en Zoutkamperlaag door bodemdaling ten gevolge van gaswinning onder de Waddenzee. TNO-060-ur-2011-020351c

TNO (2013) A general framework for rate dependent compaction models for reservoir rock. TNO 2013 R11405

TNO (2016) Ensemble-based Subsidence Interpretation and Prediction ESIP: Technical Reference Manual. TNO report-R11278

van Adrichem Boogaert, H. A., and W. F. P. Kouwe (1993-1997b), Late Permian litho-chronostratigraphic chart for the wester offshore of the Netherlands, edited by TNO. <http://www.nlog.nl/nl/pubs/atlas/doAtlas/resources/fig211.pdf>

Williams S. (2015): Description of GPS uncertainties within the Long Term Study on Anomalous Time-Dependent Subsidence. National Oceanographic Centre, Liverpool, UK.