FORCE is a cooperating forum for sustainable recovery (SR), improved exploration (IE), and energy efficiency & environment (EEE) conducted by oil and gas companies and authorities in Norway.

Guideline for Ensemble Data Sharing

Published	Revised
12/10/2023	As this is the first edition of this guideline document, the aim is to update and adapt this document after its first year in use - June 2024
Prepared by	
On behalf of the Integrated Reservoir Modelling Network group in FORCE (alphabetical order): Alexander Kovalevich (ConocoPhillips), Chris Townsend (Orec), Eugene Dahl (Wintershall Dea Norge, formerly with ConocoPhillips), Hazim Lee (TotalEnergies), Jon Sætrom (Resoptima / Halliburton), Katherine Waite (Equinor), Konstantin Monastyrev (Petoro), Marcin Bartnik (PGNiG / Vår Energi), Marine Seignole (AkerBP), Per Olav Eide Svendsen (Equinor), Remus Gabriel Hanea (Equinor), Sandor Voelgyi (Wintershall Dea Norge), Sophie Haseldonckx (NPD), Stian Haland (AkerBP).	
Contact information	Contacts
E-mail: postboks@npd.no (NB – public) Telephone: +47 51 87 60 00	Sophie Haseldonckx (NPD) & FORCE IRM group

List of contents

1 Introduction

What data is shared when ensembles are built currently varies greatly between licenses. Some share very little, others all data available. Both the sharing of too little and the sharing of too much data often leads to licensees being unable to execute their see-to-duty and to constructively participate in the decision-making process. This document aims to set out guidelines as to what licensees should share and in what format. During the process of writing the document it has become clear that collaboration between parties is the key to understanding rather than proscribed sharing and checklists.

A cross-functional group have come together under the FORCE umbrella to set-out guidelines which are agreeable from both an operator as well as a partner perspective. Companies involved have been AkerBP, ConocoPhillips, Equinor, Petoro, PGNiG, Resoptima, TotalEnergies, Vår Energi and Wintershall Dea Norge with facilitation from the Norwegian Petroleum Directorate.

As each field / modelling project has different needs, this document is set out as a guideline rather than a standard. This document is not a best practice guide for building ensembles and hence does not elaborate on this topic. It assumes that standard quality control (QC) products used in model building / history match / prediction are provided and does not list these. As this guideline does not provide an exhaustive list of suggestions, it should not be used as a limitation to what can be shared. Reference is also made to the collaboration agreement which each JV has with regards to what is to be shared in a JV.

One chapter makes general suggestions for partner interaction with regards to ensembles. Progress is often dependant on both parties understanding the capability and needs of the other. This is most easily achieved through ongoing dialog, hence using for a for interaction is encouraged. To facilitate dialog and equal interpretation of this document, terminology used in the document is defined.

Three possible levels of data sharing are presented in this guideline.

- Level 1 - results from the ensemble are to be **reviewed**.
- Level 2 - results from ensemble are to be **analysed**.
- Level 3 - results from ensembles are to be **regenerated** in-house.

These 3 different levels of data sharing are each outlined in their own chapter.

As this is the first edition of this guideline document, the aim is to update and adapt this document after its first year in use (June 2024). If you have feedback, please contact a member in the workgroup (listed above). It is also possible to send feedback to postboks@npd.no with „Feedback – Guideline for ensemble data sharing" in the subject box [1]. In addition, the people involved in creating this document will actively gather feedback within their companies.

[1] Please be aware that our correspondence is a case document which will be registered in NPDs postale journal. Case documents and journals are public except as otherwise provided by statute or by regulations pursuant thereto. Any person may apply to NPD for access to case documents and journals. Information that is subject to a duty of confidentiality by or pursuant to law is exempted from access.

2 Terminology/ Abbreviations

One of the challenges in communication is that several commonly used terms are used with different meanings in traditional and ensemble modelling. The alphabetical list below tries to define the most common of these as a starting point for discussion and specifies how these are meant to be read in the context of this document.

Aggregated results	Results per ensemble and/or case.
Aggregation	The result produced by aggregating data across multiple entities, frequently across all realizations in an ensemble. Frequently representing the calculation of statistical representations.
Case	One or more ensembles that belong together, or build on each other, e.g. in a history matching context (≠ realisation or scenario)
Ensemble	A collection of realisations that belong together and combine to one output.
Geological concept	A mental picture of the reservoir which is converted into a quantitative representation describing structural and stratigraphic elements. Conceptual models usually incorporate rules about possible geometries and successions of facies that can be included in a geological scenario.
Iteration	Each ensemble generated during conditioning while moving from prior to posterior represents an iteration.
Mean/expected /Std/P90/P50/P10	Statistical representations of a distribution, frequently calculated across all realisations of an ensemble or all iterations of a case or similar. The arithmetic mean is often called the expected value and is widely used in decisions. Percentiles: P90: 90% of the outcomes for a given metric will be higher than this value. P50: 50% of outcomes for a given metric will be higher than this value, 50% of outcomes will be lower. P10: 10% of outcomes for a given metric will be higher than this value. Note! Other industries define this the other way round – here oil industry convention is used - that is for a numeric distribution, P90 is in the lower end, P10 is in the higher end.
Model setup	The model setup is the mathematical description of the subsurface understanding at any one point in time and must be conditioned to (real) data to produce results (= realisation). It is the workflows, configuration, input data (sometimes). The collaborative product that the team makes and maintains. Synonyms in some contexts: Setup, template project, workflow (≠ grid)
Objective function	The objective function minimises the misfit between chosen observed data during history matching to arrive at the posterior. A different objective function is used during forecasting to determine for example the optimum drainage strategy.

Posterior	The posterior describes the state of the input parameter distributions after history matching or other automated conditioning has been performed.
Prior	The prior describes the state of the input parameter distributions before history matching or other automated conditioning has been performed.
Raw Result	Results per realisation.
Realisation / Ensemble member	The individual members of an ensemble. One deterministic outcome of realising the model.
Reference case	One or more ensembles which the current case is compared/referenced. E.g. last major milestone. (≠ best technical case)
Reference Realisation	A specific realisation used as a reference point for other realizations. (≠ best technical case, not a most likely case).
Scenario	A specific set of input assumptions and/or constraints represented by a single ensemble member / group of members / a full ensemble. Different scenarios are used to represent deterministic values for parameters or sets of parameters.
Sensitivity	A specifically designed run to determine the sensitivity of one or more parameters. A sensitivity run can be a full case, a single iteration, groups of single realisations, or a single realisation. In normal mode, the ensemble member/realisations are not sensitivities.

3 Collaboration in the partnership

When new modelling projects (new builds or updates of existing model setups) are kicked off, it is highly recommended for the partnership to discuss how collaboration will take place and which products can and will be provided. These discussions should begin as early as possible to allow all parties to align expectations and plan for relevant involvement. Regular follow up is recommended and initial plans can and should be revisited and adjusted as necessary during the work.

Important topics to discuss include:

- Formats for collaboration (regular meetings, workshops, sessions in front of workstations, status reports, …)
- Feedback loops
- What products the operator can provide
- What information and/or products partners require and how they will use these
- Clarify terminology used in the project to avoid misunderstandings.

From these discussions an agreement on the relevant level of data sharing for the ongoing project (see descriptions for levels 1, 2 & 3 in this document) should be reached including - ideally - a list of expected products to be delivered. Note that this list should be updated if/when relevant changes occur.

Early alignment allows for better planning – both for operators and licence partners. Active collaboration and regular follow up discussions help to build trust, enable good discussions in the partnership and ensure that partners have an opportunity to give feedback during the modelling process.

4 LEVEL 1 - Review

4.1 Aim of Level 1

At this level results from the ensemble are to be **reviewed** only. Summaries / overviews / maps of input data of the ensemble input are provided together with the main results and the operators "digested" analysis.

Level 1 is a very common scenario where a partner company is able to make decisions based on the provided and documentation.

Documentation can be a written report, but also presentation material from formal or informal JV meetings or other data visualisation tools (e.g. webviz) which are accessible to the JV companies. A section below is dedicated to what data should be shared as data (rather than a visualisation). These are data normally covered under the JV collaboration agreement.

4.2 Common documentation practice

The following should be included in the documentation (but not limited to):

- State where the model set-up is appropriate to be used / where it cannot be used. (When is the model set-up predictive?)
- Set up of (prior) ensemble
    - What and how are modelling inputs etc used in the modelling and why?
        - Wells
        - Seismic
        - Log picks
        - Concepts
        - Relative permeability
        - etc
    - Which inputs are not included? Why?
    - List of uncertain parameters (ranges, distributions, correlations and human readable variable names).
    - Description/Justification of the uncertain parameters selected ranges, distributions, correlations for uncertain parameters with respect to concepts and/or available measurements.
    - Demonstrate how the geological concept(s) is represented in the model/workflow
    - Specify which control HM simulation model is run on.
    - Demonstrate how the simulation model is initialized and if it is dynamically stable.
    - Any other QC products agreed on in the JV

    - Present which parameters have most impact on the ensemble results, focus on PRIOR results (e.g. GIIP/STOIIP/FOPT/FGPT/FWPT/Pressure @start/end simulation per field or per well, depending on the decision (see 1st point above)).
    - Reference to the "parent ensemble" (if present)
- Description of simulation deck – both history and prediction if used
    - Ranges of operational uncertainties, such as uptime, constraints, drilling schedule should be specified.
- Objective function – History Match
    - Describe the Objective function set-up, give a list of the parameters used and their reasoning
        - Type of data used (production data, RFTs/DST, seismic data, tracers, etc.)
        - Weighting on the different type of data (if used – both data type and in time)
        - Describe the methodology behind the set-up (e.g. localization)
        - Contribution per well and data type
        - Impact of the input parameters and their sensitivities
        - Model error (input errors) (for the simulation runs)
    - Describe the evolution of Objective function with the distinct ensemble iterations
- Objective function – Prediction (if applicable)
    - Describe the Objective function set-up, give a list of the parameters used and their reasoning
        - Type of data used (e.g. Field cumulative oil, least water volume, number of wells)
        - Weighting on the different type of data (if used)
        - Describe the methodology behind the set-up (e.g. localization)
        - Contribution per well and data type
        - Impact of the input parameters and their sensitivities
        - Model error (input errors) (for the simulation runs)
    - Describe the evolution of Objective function with the distinct ensemble iterations
- QC of posterior as relevant for the model set-up
    - Selected standard QC plots static and dynamic responses as agreed in JV – prior vs. Posteriori (specifics should be agreed on in the partnership as early as possible. This should be available without investment in software) such as
        - Cross-sections, average maps (e.g poro-perm-facies, thickness, depth surfaces, field/contacts outline)
        - 3D parameters - static and dynamic (e.g. saturation, and pressure)
        - Spread/coverage (typical RE plots) including transition from HM to Prediction if relevant
            - Observed data vs realisations - spread
            - Distribution changes and interpretations
            - Predictive power (in history and future setup)- blind test approach
        - Mismatch between observed and simulated
            - Relevant levels (e.g. well, field) to be agreed in JV
            - (seismic / gravimetry/ tracers…)
    - Correlation of input parameters vs. output properties (e.g. OWC vs. FOPT), trends seen in data - Understanding more important than matching

        - What obs impacted what parameters?
        - Why the distributions changed?
    - Present which parameters have the most impact on the modelling, focus on posterior results (GIIP/STOIIP/FOPT/FGPT/FWPT/Pressure @start/end).
    - Convergence (or stagnation?) of uncertain parameters from PRIOR to POSTERIOR
    - Discuss fudges/modification put in the simulation model WPI multipliers etc.
    - Comparison between previous ensemble and current ensemble
- Include agreed QC products used in model building / history match / prediction to provide "sanity" check for the ensemble (e.g. material balance, transition to prediction period)
- Explain presented aggregated results. Present the profiles that is input to the decision.
- Selection of ensemble subsets
    - If there has been any selection within the ensemble, it should be clearly stated how the samples have been selected. (e.g. subset of 50 members from an ensemble of 500).
    - Specify if it is a subset or a new ensemble

4.3 To be provided as data

The following should be provided as (raw) data in tabulated format:

- Initial inplace volumes per segment/sand/ etc (to be decided in JV discussion) for each realisation for prior and posterior (if relevant)
- Cumulative well /group/field production all phases (yearly)

5 LEVEL 2 - Analysis

5.1 Aim of Level 2

At this level results from the ensemble are to be **analysed** independently. Underlying (raw) data is provided to be able to cross-check inputs and do their own analysis of the results. Companies can perform in-house filtering of the ensemble members (clustering). Individual companies can in this way use mean/P50 or another measure for decision making if preferred.

The expectation is that if a level 2 is requested that there is active feedback to the operator on own analysis done by partners.

The data shared is in machine readable format which the receiver can adapt to use in standardized workflows like webviz/IRMA/others to visualize/analyse ensembles.

5.2 To be provided as data

The following should be provided as (raw) data in tabulated format:

- volumes per segment/sand/ etc (to be decided in JV discussion) for each realisation for prior and posterior (if relevant)
- Cumulative well /group/field production all phases (monthly suggested as baseline, but frequency to be agreed within JV)
- Input / Parameterization matrix (variable spreadsheet) including seeds where appropriate.

- 3D grid with parameters (e.g. start of history match, end of history match/start prediction and end of prediction (3 time steps)).
  - Assets with very long histories and or 4D data can consider more frequent timesteps.
- Simulation files (data deck + include files)
- Specific agreed grid properties that are not exported as standard to the simulation deck (e.g. flow regions, facies/rock type or fracture density)

6 LEVEL 3 - Regenerate

This is the highest level where the aim is to be able to **regenerate** ensembles in-house and the receiver can choose to modify the ensemble to create their own using their own interpretations, running with different ranges on input parameters, or performing sensitivities that the operator did not focus on.

This topic is fraught with both technical complexity as well as possible legal issues and has for these reasons not been included in this first edition. The intention is that the guidelines for this will be worked up and published at a later date.