Data Assimilation
and its applications
Inverse Problem – Conceptual understanding

- The forward problem can be conceptually formulated as follows:
  Model parameters → Data

- The inverse problem - relates the model parameters to the data that we observe:
  Data → Model parameters

- The transformation from data to model parameters (or vice versa) is a result of the interaction of a physical system with the object that we wish to infer properties about.
Some examples

<table>
<thead>
<tr>
<th>Physical system</th>
<th>Governing equations</th>
<th>Physical quantity</th>
<th>Observed data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth's gravitational field</td>
<td>Newton’s law of gravity</td>
<td>Density</td>
<td>Gravitational field</td>
</tr>
<tr>
<td>Earth's magnetic field (at the surface)</td>
<td>Maxwell’s equations</td>
<td>Magnetic susceptibility</td>
<td>Magnetic field</td>
</tr>
<tr>
<td>Seismic waves (from earthquakes)</td>
<td>Wave equation</td>
<td>Wave-speed (density)</td>
<td>Particle velocity</td>
</tr>
</tbody>
</table>
Key elements for successful solution?
Cooking book
NEED / CHALLENGE
APPETITE
‘FUNDAMENTAL’ KNOWLEDGE
COOKING SKILLS
DOMAIN KNOWLEDGE
type of layers
PASSIONATE / CREATIVE PEOPLE
PASSION / CHEMISTRY
cross fertilization
WILLINGNESS TO MIX & EXPERIMENT
List of recipes

- Optimal interpolation
- Kriging
- Variational methods
- Ensemble methods
- Hybrid methods
Data Assimilation – Main ingredients

- Two sources of information about the true state of the nature:
  - **Model** (abstraction of reality in terms of a set of differential equations)
  - **Measurements** (measure of certain quantities of interest)
  - **Uncertainties** are present in both worlds.
  - **Prior knowledge** (expert opinion)
The goal: An optimal estimate of the truth based on the combination of both uncertain sources of information.

“I think you should be more explicit here in step two.”
The goal: An optimal estimate of the truth based on the combination of both uncertain sources of information.
A different flavour for everyone / every challenge

Did you notice how a country-specific cuisine tasted differently in said country and abroad?

- Chinese food tastes like Indonesian in Netherlands and like Vietnamese in France.

- Italian pizza you have at your local Italian restaurant is rarely the same as the one you have in Italy.

Foods are tailored to meet the specific preferences of each country.
A different flavour for everyone / every challenge

“The podiatrist wants jam on his toast, the psychiatrist wants nuts on his cereal, the plastic surgeon wants no wrinkles on her bacon, and the fertility doctor wants his eggs frozen.”
Model and its uncertainties

- Climate, Air, and Sustainability
- Geoscience uncertainties
- Uncertainties in physics
- Different scales
- Reaction rates
- Grain dimensions
- Unmodelled physics
- Different scales
Observations/Measurements and uncertainties

**Petroleum Geosciences**
- Uncertainties

**Climate, Air and Sustainability**
- Water level
- Inzinking
- Density and velocities
- ...
We solve different problems with the same approach (cross-fertilization)

Climate, Air and Sustainability
Fluid dynamics

- Estimate essential parameters measuring directly not feasible
- Model calibration
- Integrating information from geological elements of different scales
- Optimal estimates for the dynamical parameters
- Optimal estimates for the dynamical parameters
- Predicting peaks of ozone high concentrations
- Optimize dredging cycle fuel cost, cycle time
- Redo developing the plan
  - Optimize production strategies
  - Optimize well locations
Probabilistic Data Assimilation – Bayes’ rule

\[ P(x|y) = \frac{P(y|x)P(x)}{P(y)} \]

\[ P(x|y) \propto P(y|x)P(x) \]

- $P(x|y)$: Posterior probability
- $P(x)$: Prior probability
- $P(y|x)$: Likelihood of observations, given a model
- $P(y)$: Probability of observations

Bayes’ Rule

Sequential Methods

Variational Methods

Kalman Filter

Adjoint based methods

Ensemble Kalman Filter
Classical Kalman Filter Steps

1) Forecast Step
   Based on Model

   $x^f(t_{k+1})$ and $P^f(t_{k+1})$

2) Analysis Step
   Combining forecast and measurements weighted by Kalman Gain

   Measurement
   $y^o(t_{k+1})$

   $x^a(t_{k+1})$ and $P^a(t_{k+1})$

   $x^a(t_k)$ and $P^a(t_k)$

$K$
System and the measurements:
\[ x^f(t_{k+1}) = M(x^f(t_k)) + w(t_k) \]
\[ y^o(t_{k+1}) = H(t_{k+1})x^f(t_{k+1}) + v(t_{k+1}) \]
\[ w \sim N(0, Q) \]
\[ v \sim N(0, R) \]

1) Forecast step:
\[ x^f(t_{k+1}) = E(x^f(t_{k+1})) = M(t_k)x^a(t_k) \]
\[ P^f(t_{k+1}) = E[(x^f(t_{k+1}) - x^f(t_{k+1}))(x^f(t_{k+1}) - x^f(t_{k+1}))^T] \]

2) Analysis step:
\[ x^a(t_{k+1}) = x^f(t_{k+1}) + K(t_{k+1})(y^o(t_{k+1}) - H(t_{k+1})x^f(t_{k+1})) \]
\[ P^a(t_{k+1}) = E[(x^a(t_{k+1}) - x^a(t_{k+1}))(x^a(t_{k+1}) - x^a(t_{k+1}))^T] \]
\[ K(t_{k+1}) = P^f(t_{k+1})H(t_{k+1})^T[H(t_{k+1})P^f(t_{k+1})H(t_{k+1})^T + R(t_{k+1})]^{-1} \]

Calculates only the first statistical moments: mean and covariance
Non-classical Kalman Filters

• Classical Kalman Filter assumes:
  • Linearity for the model operator and observation operator.
  • Gaussian distribution for the statistics of the error distribution.

• But in reality, this is usually not the case

• Remedies:
  • The Extended Kalman filter
    Was used in the Apollo missions, but it is not practical for complex systems because of computational burden.
  • Ensemble Kalman filter and adjoint based methods can be used with a nonlinear model and nonlinear measurement model.
Ensemble Kalman Filter

- **Advantages**
  1. Can be used for nonlinear models.
  2. Fairly simple to implement.
  3. No need to go into the details of the forward model.
  4. Computational advantages (lower rank covariances)

- **Disadvantages**
  1. It is very sensitive to the “good” knowledge of the statistics.
  2. Requires a large number of members of the ensemble to converge to the real parameter.
Ensemble Kalman Filter

\[ \bar{x}(t_{k+1}) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^f(t_{k+1}) \]

- Initial state \( x_0 \)

Represent the uncertainties in \( x_0 \) using an ensemble of \( N \) states \( \xi_i^a(t_0) \)

Propagate each ensemble using the original model

\[ \xi_i^a(t_k) = \xi_i^f(t_{k+1}) + K(t_{k+1})[y^o(t_{k+1}) - H(t_{k+1})\xi_i^f(t_{k+1}) + v_i(t_{k+1})] \]

\[ P_f(t_{k+1}) \approx P_e^f(t_{k+1}) = E[(\bar{x}(t_{k+1}) - x^f(t_{k+1}))(\bar{x}(t_{k+1}) - x^f(t_{k+1}))^T] \]

\[ P_a^f(t_{k+1}) \approx P_e^a(t_{k+1}) = E[(\bar{x}(t_{k+1}) - x^a(t_{k+1}))(\bar{x}(t_{k+1}) - x^a(t_{k+1}))^T] \]
Time

True state

Measurements with errors

Initial state with errors

Time
Measurements with errors

Model prediction with errors

True state

Time
Time

Updated estimate with errors

Measurements with errors

True state

Model prediction with errors
Time

New model prediction with errors

Measurements with errors

True state

Updated estimate with errors

Time
Measurements with errors

True state

Updated estimate with errors

Time
$\min J(x, u)$

$x$ represents the state variables, in our case pressure and saturation

$u$ represents
- the reservoir model parameters that we want to estimate in the history matching, or
- the control parameters that we want to optimally set in the field development plan
Variational methods – the principle

$$\min J(x)$$
Variational methods – the Jacobian, the malefactor

- We need to calculate the gradient (Jacobian)

\[
\frac{dJ(x,u)}{du} = \frac{\partial J}{\partial u} + \frac{\partial J}{\partial x} \frac{\partial x}{\partial u}
\]

Derivatives with respect to the parameters

Derivatives with respect to the state variables

adjoint

- \( u \) may easily represent 100s of variables, but worse
- \( x \) may represent millions of variables, for each time step!

Options to calculate the Jacobian:
  - Numerical differentiation: computationally not feasible in our case
  - Adjoint method: computationally efficient, but requires significant programming efforts
Challenges

• ...
The TOPAZ model system

- TOPAZ3: Atlantic and Arctic
  - HYCOM + EVP sea-ice model
  - 11-16 km horizontal resolution (800 x 880)
  - 22 hybrid layers
- EnKF
  - 100 members
- Observations
  - Sea Level Anomalies (CLS)
  - Sea Surface Temperatures (NOAA)
  - Sea Ice Concentr. (AMSR, NSIDC)
  - Sea ice drift (CERSAT) [asynchronous]
  - Argo T/S profiles (Coriolis)
- Runs weekly, 10 days forecasts
  - ECMWF forcing
- Exploited at met.no since March 2008
Analyse the ocean circulation, sea-ice and biogeochemistry. Provide real-time forecasts to the general public and industrial users.
Case studies – Highly nonlinear dynamics

2D variables (400 x 600 grid cells)
- Barotropic pressure
- u/v velocity
- Ice concentration
- Ice thickness

3D variables (400 x 600 x 22 grid cells)
- Temperature
- Salinity
- u/v current
- Layer thickness

TOTAL: 27,600,000 variables

Sea level anomalies (satellite, radar altimeters)
- Non-linear function of state variables
- 100,000 observations every week

Sea-surface temperature (satellite, optical)
- 8,000 observations every week

Sea-ice concentrations (satellite, microwave)
- 40,000 observations every week

TOTAL: 148,000 measurements
An unified numerical weather forecasting operational system

Global Environmental Multiscale (GEM) Forecasting & Modelling System 2011-2021

- Regional and Mesoscale Forecast (24-48 h, 10-15 km) & Data assimilation
- Medium-range Forecast (240 h, 10 to 35 km) & Data assimilation

Middle Atmosphere Model & Data assimilation

- Multi-Seasonal Forecast
- Monthly Forecast

Limited-Area Model 0-24h 0-2.5km & Data assimilation

- Ensemble Forecast
- Regional Climate Model

Micro-meteorology (10m-1km)

Canadian Meteorological Center, Weather prediction Division
Reservoir management workflow benchmark study
Peters et al., 2010, SPE J.

- Synthetic case
- 44500 active grid cells with 4 values in each grid
- relative perms, initial OWC, vertical transmissibility
- 10 or 20 year production history
- 20 producers and 10 injectors
- time-lapse seismic

Can we optimize the oil production (maximize NPV) over 30 years period?
Reservoir management workflow benchmark study

- Estimation of different properties
- Non-linear dynamics
- Two distinct data types
- Different simulators used by the participants

Successes:

- Use of the EnKF as a history matching method was a common factor among the best performers
- Updating models and production strategies more frequently improves the forecast of the final realized NPV.
Seismic History Matching of Fluid Fronts

Trani et al. 2011, submitted to SPE Journal

- Synthetic case based on Brugge field
- 20000 active grid cells with 2 values in each grid cell
- 14 years of production
- 17 producers and 10 injectors
- A re-parameterization of time-lapse seismic into front arrivals times (no extra inversion required)

What is the added values of a new parameterization for time-lapse seismic?
Seismic History Matching of Fluid Fronts

- Non-linear dynamics
- Two distinct data types
- MORES simulator

Successes:
- The new re-parameterization is a success
- No extra inversion required
- Improved match for both production and seismic data => improved forecast skills
Roswinkel Field Case
Joint HM subsidence and well data

Wilschut et al., SPE 141690

- Heavily faulted gas field in NE-Netherlands
- 35 possible compartments
- GIIP 24.6 bcm
- Production period 1980-2005
- 9 leveling subsidence campaigns
- Max. subsidence 17 cm

Can we identify compartmentalization based on both subsidence and production data?
Roswinkel Field Case
Joint HM subsidence and well data

- Estimation of fault properties
- Moderately non-linear
- Two distinct data types
- Simulator: IMEX coupled with geomechanical model

Successes:
- Added value of second data type
- EnKF can also be used as diagnostic tool
Infrared remote sensing of atmospheric composition and air quality: towards operational applications
Conclusions

• Combination between the model and measurements

• Estimation and forecast tool under uncertainties.

• It is very sensitive to the right description of the uncertainties

• Data assimilation is a successful recipe/solution for a lot of different types of applications