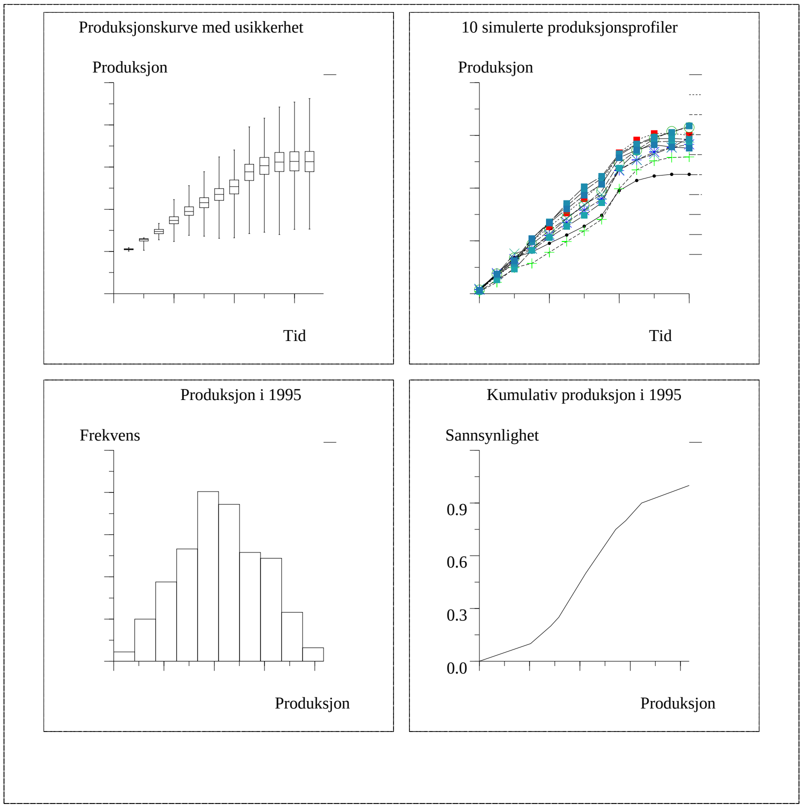
Designing and analyzing computer experiments



Thore Egeland, Norwegian Computing Center
March 9, 1995. [1]

[1] Available from: http://www.nr.no/home/SAND/thore/

Contents

- **Motivation.**
  Academic and commercial background.
- **Experimental design. Weighing example.**
- **The computer is our laboratory.**
- **What's special about computer experiments?**
- **What's special about reservoir simulations?**
- **What we have tried to achieve:**
  - **Please consider replacing 'one–at–time–plans'. Why?**
  - **Please consider replacing computer program by response surface.**
  - **Please consider using experimental design as a way of structured problem approach.**
  - **Examples.**

Norsk Regnesentral (NR)

NR Established 1952

Non-profit applied research foundation

Situated at University of Oslo campus

70 scientists

Internationally known for:

- **First european Univac 1100**
- **SIMULA**
- **Automatic mapping**
- **Labor union projects**
- **Geostatistics**

Ultimate goal: Net Present Value

Weighing apples and oranges

History

DESIGN of experiments.
Maximum Information at Minimum Cost:

- **Agriculture 1920.**
- **Medicine, chemistry,...**
- **Computer experiments 1970.**
- **Reservoir simulations 1980–.**

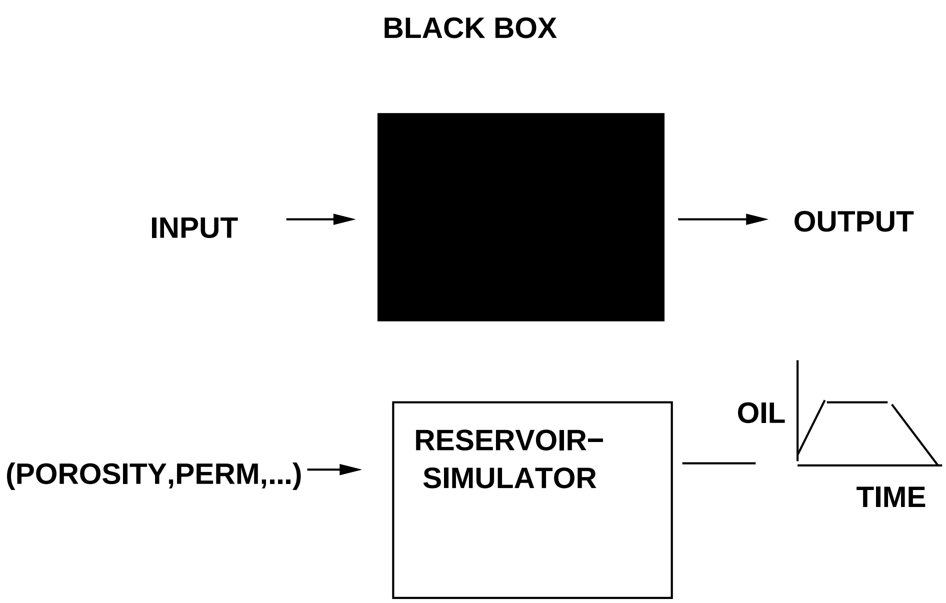
NR activity since 1989

Petroleum:

- **Uncertainty in production profiles.**
- **Automatic history matching.**
- **Rate optimization.**

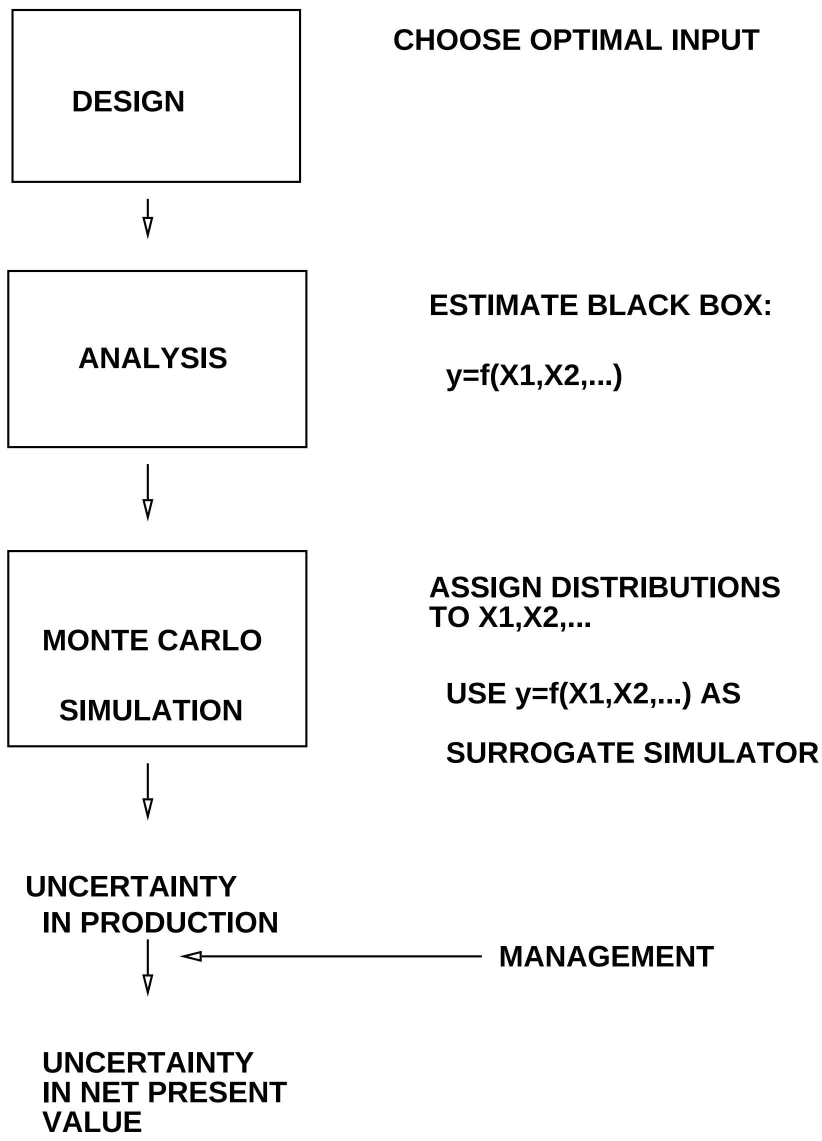
Projects for roughly 3 500 000 NOK.
(Commercial software developed: DECISION).

Basic model



BLACK BOX
INPUT
OUTPUT
(POROSITY,PERM,...)
RESERVOIR−
SIMULATOR
OIL
TIME

Basic approach



What's special about computer experiments?

The experiments are deterministic, i.e., no noise.

Identical input and seed produce identical output.

Replications do not make sense without varying seed.

What's special about reservoir simulations?

- **The applications we have in mind are extremely cpu–demanding.**
  One run may require 10-20 hours.
- **There are many input variables.**

Typically: 5–15 input variables. Each defined at three levels.

Example: $3^{13} = 531441$ **possible runs.**

Time and money for 50 runs.

Goal: select optimal runs.

What we have tried to achieve

Please consider replacing 'one–at–time–plans'. Why?

Input variables:

$$\textbf{Porosity} = \begin{cases} -1 & \textbf{Best guess} \\ 1 & \textbf{Optimistic} \end{cases}$$

$$\textbf{Permeability} = \begin{cases} -1 & \textbf{Best guess} \\ 1 & \textbf{Optimistic} \end{cases}$$

$$\textbf{Faults} = \begin{cases} -1 & \textbf{Best guess} \\ 1 & \textbf{Optimistic} \end{cases}$$

Simplified version of (previous?) approach in the oil industry:

Porosity	Permeability	Faults	Oil
-1	-1	-1	y_1
1	-1	-1	y_2
-1	-1	-1	y_3
-1	1	-1	y_4
-1	-1	-1	y_5
-1	-1	1	y_6

Problems:

- **Porosity effect,** $y_2 - y_1$**, calculated for unbalanced comb. of remaining variables.**
- **More accurate estimates available.**
- **No possibility to detect, say, Porosity–Permeability interaction**

Fractional Factorial plans **solve the problems.**

Factorial plan. Alternative

Porosity	**Permeability**	**Faults**	**Oil**
1	**-1**	**-1**	y_1^*
-1	**1**	**-1**	y_2^*
-1	**-1**	**1**	y_3^*
1	**1**	**1**	y_4^*

Porosity effect:

$$(y_1^* - y_2^* - y_3^* + y_4^*)/2$$

Estimate based on four; not two runs.

Four factorial runs may well provide more information than 6 one–at–a–time runs.

Factorial plan. Figure

Example

Based on cooperation with Conoco (Aberdeen).

Response variables:

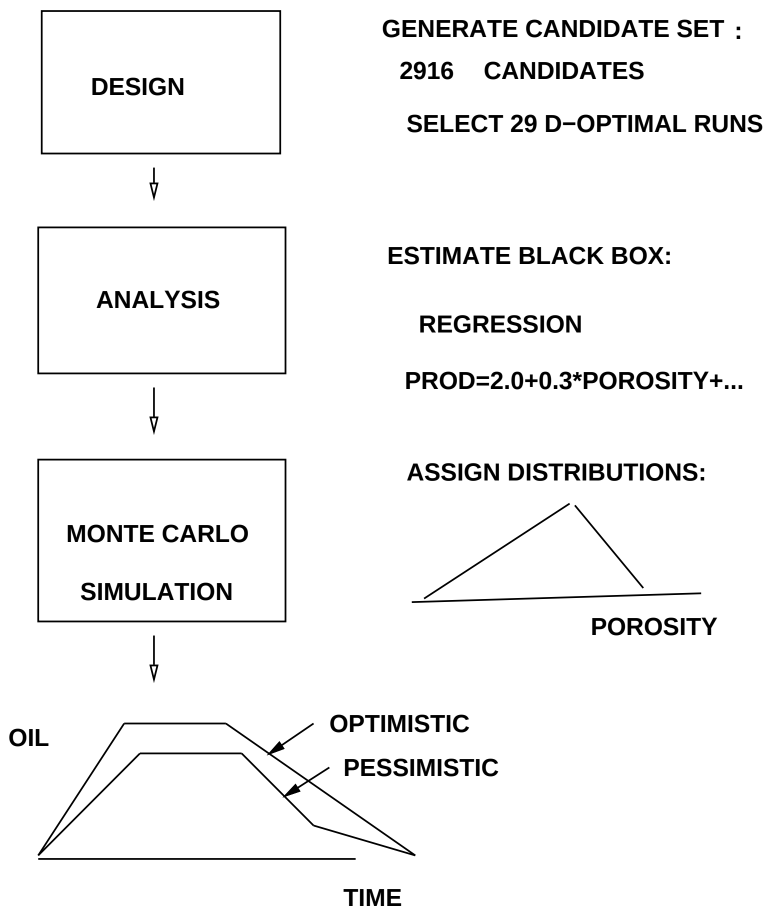
$$
\begin{aligned}
y_k &= \textbf{oil produced (cumulative)}, \\
k &= 1995, 1996, \ldots 2017.
\end{aligned}
$$

Input variables describing reservoir:
6 variables on 3 levels,
2 variables on 2 levels.

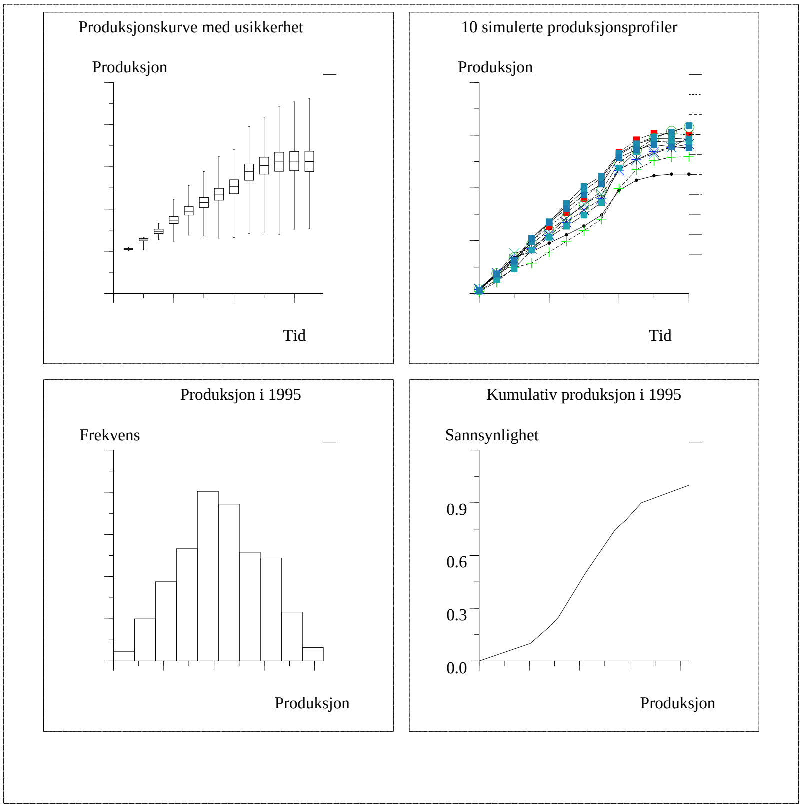
Possible number of runs:

$$3^6 \cdot 2^2 = 2916$$

Approach


DESIGN
GENERATE CANDIDATE SET :
2916 CANDIDATES
SELECT 29 D−OPTIMAL RUNS
ANALYSIS
ESTIMATE BLACK BOX:
REGRESSION
PROD=2.0+0.3*POROSITY+...
MONTE CARLO
SIMULATION
ASSIGN DISTRIBUTIONS:
POROSITY
OIL
OPTIMISTIC
PESSIMISTIC
TIME

Results


Produksjonskurve med usikkerhet
Produksjon
Tid
10 simulerte produksjonsprofiler
Produksjon
Tid
Produksjon i 1995
Frekvens
Produksjon
Kumulativ produksjon i 1995
Sannsynlighet
0.9
0.6
0.3
0.0
Produksjon

Summary

- **The computer is a laboratory and methods used to design, say, chemical experiments apply.**
- **Experimental design may save time and money.**
- **A good design does not require advanced data analysis. A bad designed experience is a challenge for an experienced statistician.**