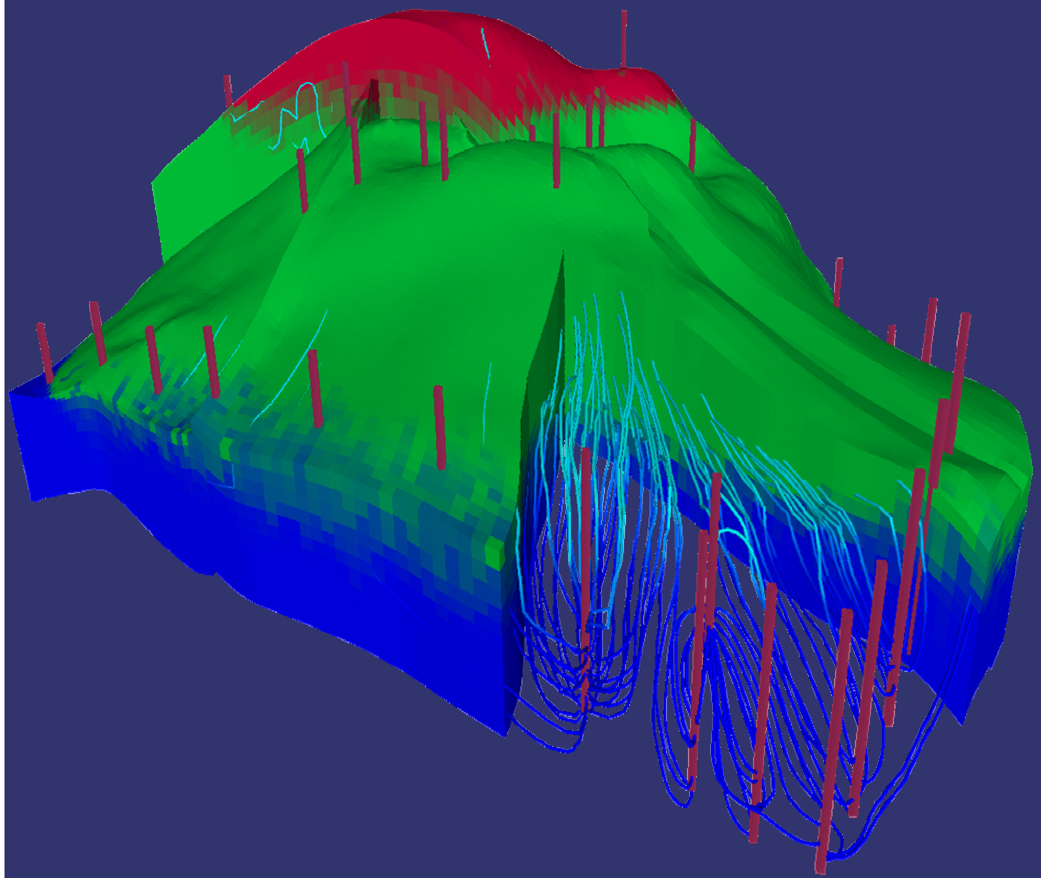
STATISTICAL ANALYSIS OF RESERVOIR SIMULATORS OUTPUTS: SOME MODELLING ISSUES



Atelier GdR Mascot Num

IFP-LYON

Daniel Busby (IFP Rueil Malmaison)

Mathieu Feraille (IFP Rueil Mamaison)

Ingénierie de réservoir : Objectifs et enjeux

- Exploitation d'un gisement de gaz ou de pétrole pour l'optimisation des rendements économiques

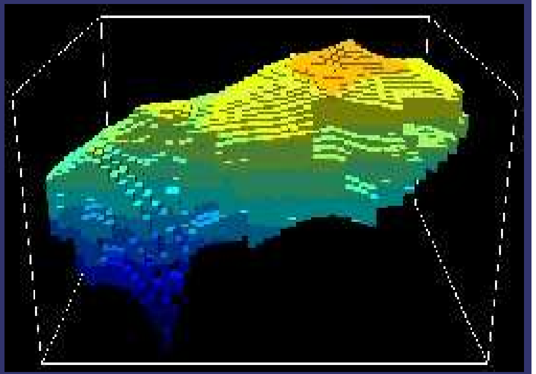
- Enjeux économiques : coûts d'exploitation et de production importants
  - Plusieurs dizaines ou centaines de puits - prix pour un puits ($ 1-10M)
  - Récupération assistée – injection de fluide pour augmenter la production

➢ Nécessité de comprendre, caractériser et prévoir le comportement dynamique du gisement

  ➨ Recours à une modélisation numérique des phénomènes dans le gisement

- Problèmes :
  - Les gisements les plus complexes restent à exploiter
  - Plusieurs km², enfouis dans de grandes profondeurs ou difficiles d'accès
  - Propriétés du gisement très mal connues et peu caractérisées

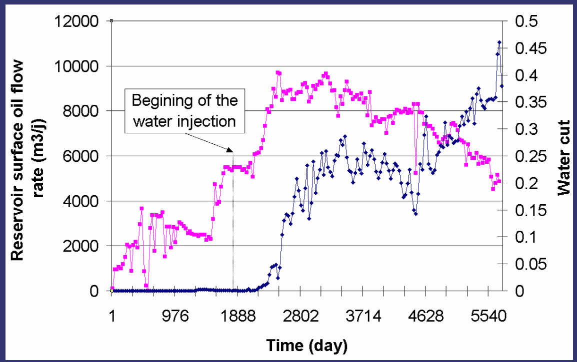
Ingénierie de réservoir : Objectifs et enjeux



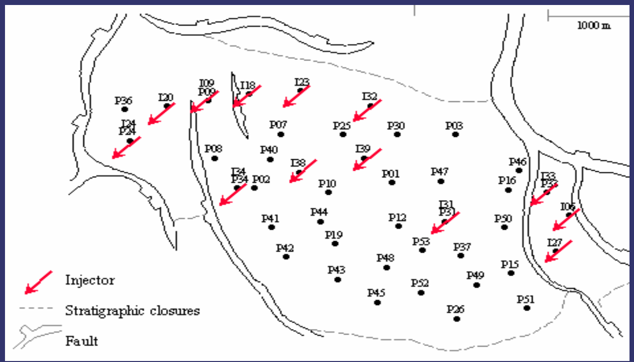
- Construction d'un modèle numérique de simulation de la production d'hydrocarbures en fonction des paramètres caractérisant les gisements :

$$f : IR^K \rightarrow IR$$

- Utilisation du simulateur f pour :
  - Comprendre le comportement dynamique du gisement
  - Prévoir la production future du gisement
  - Décider de l'intérêt économique d'un gisement
  - Construire et optimiser le schéma de production du gisement



Estimation de la production future

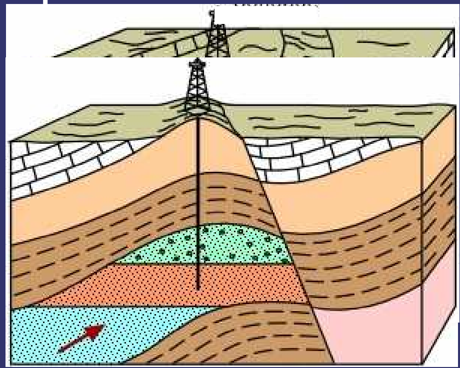


Emplacement des puits

Ingénierie de réservoir : Objectifs et enjeux

Notion d'incertitudes

- Modèle de simulation impliquant un grand nombre de paramètres potentiellement incertains :

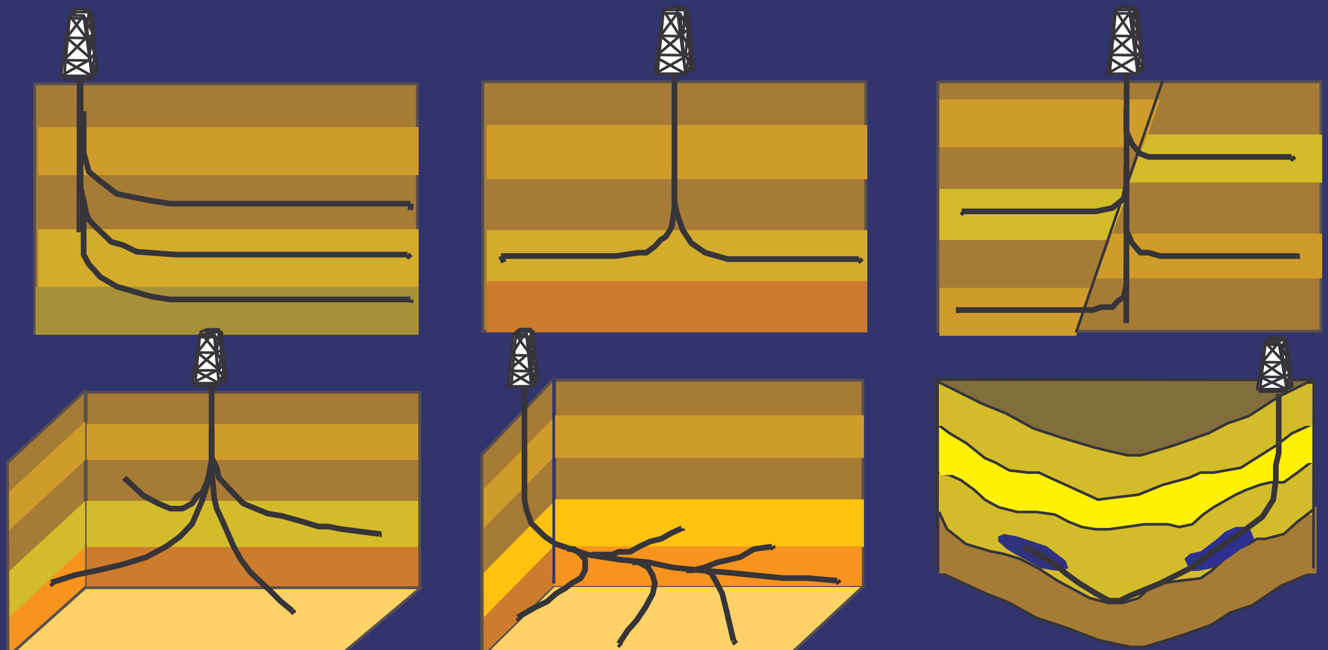


  – Paramètres liés au gisement
    - Géométrie du réservoir : épaisseur, limites, force d'aquifère, présence de faille, etc.
    - Remplissage du réservoir : plusieurs milliers de mailles à renseigner en propriétés pétrophysiques (porosité, perméabilité)
    - Propriétés des fluides eau/huile/gaz : niveau des contacts entre les fluides, viscosité, saturations, PVT
    - Interactions roches/fluides : perméabilités relatives
      – Capacité d'un fluide à se déplacer gênée par la présence d'un autre fluide
    - Puits & production : Indice de productivité (IP), effet pariétal (skin)
      – Modification de la perméabilité aux abord du puits - lié au forage

Ingénierie de réservoir : Objectifs et enjeux

Notion d'incertitudes

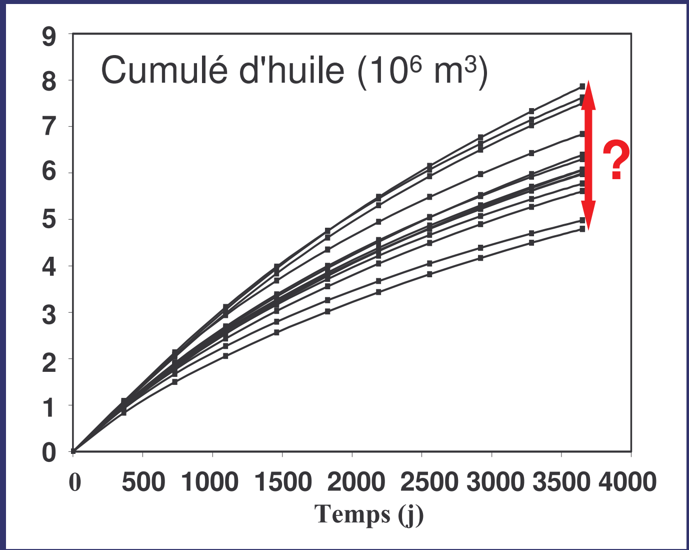
– Paramètres liés aux options de production :
  - techniques de récupération : injection, type d'injection (eau, gaz, eau+gaz, vapeur, polymères)
  - puits : nombre, type, emplacement, architecture
  - débits d'injection, de production



Ingénierie de réservoir : Objectifs et enjeux

Évaluation des incertitudes – Quels objectifs ?

- Propagation des incertitudes des paramètres du modèle sur les prévisions de production $f(x),\ x \in IR^k$



➢ Ces incertitudes sont-elles influentes sur les prévisions de production ?

➢ Comment quantifier ou réduire leur impact ?

✓ Diagnostiquer les risques venant d'une mauvaise connaissance du réservoir

✓ Optimisation des schémas de développement

Ingénierie de réservoir : Objectifs et enjeux

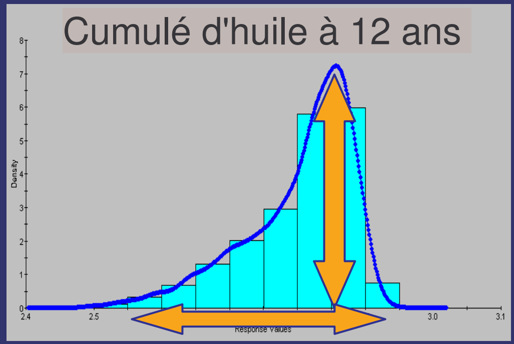
Évaluation des incertitudes – Approche classique

- Évaluations de production obtenues par simulation déterministe
  - ➢ Incertitude sur les paramètres d'entrée permet de considérer la sortie du simulateur comme une variable aléatoire

➨ Emploi de techniques statistiques classiques

– Prédiction de la réponse moyenne et du risque induit par le contexte incertain par échantillonnage Monte-Carlo (Walstrom et al., 1967)

– Temps de simulation très importants :
  - de quelques heures à quelques jours



Évaluation des risques

↔

Plusieurs milliers de simulations

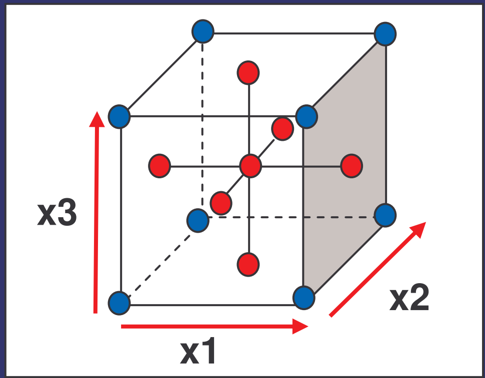
➨ Construction d'un modèle approché du simulateur pour une évaluation des risques par échantillonnage Monte-Carlo

Ingénierie de réservoir : Approches utilisées

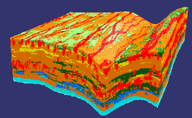
- Méthodes et techniques utilisées:
  - Plans d'expériences,
  - Modélisation de surface de réponses (régression, krigeage),
  - Arbre de décisions,
  - Formalisme bayésien.

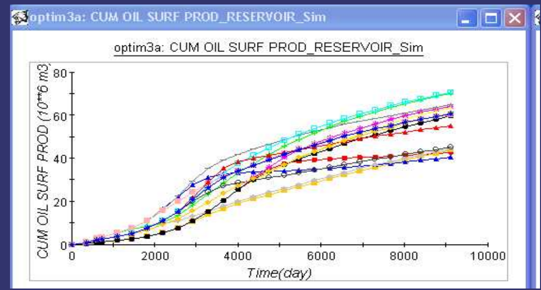
Ingénierie de réservoir : Approches utilisées

plan d'expériences



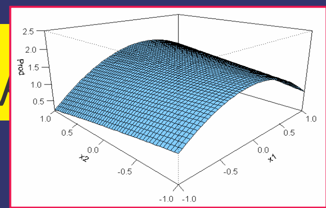
n simulations réservoir

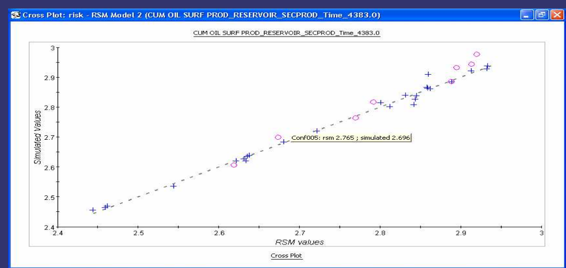




Surface de réponse

$$CumOil = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 \cdots + \beta_{11} x_1^2 + \cdots$$





Contrôle Qualité

Utilisation RSM pour analyses futures

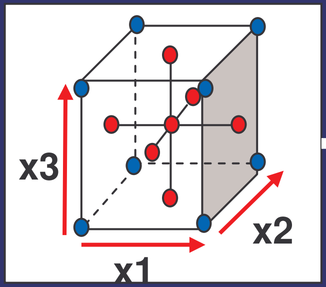
dans le domaine

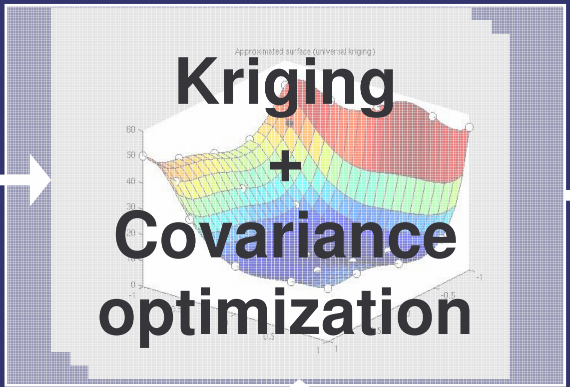
des paramètres incertains...

Ingénierie de réservoir : Approches utilisées

More advanced methods for non linear response

First experimental design

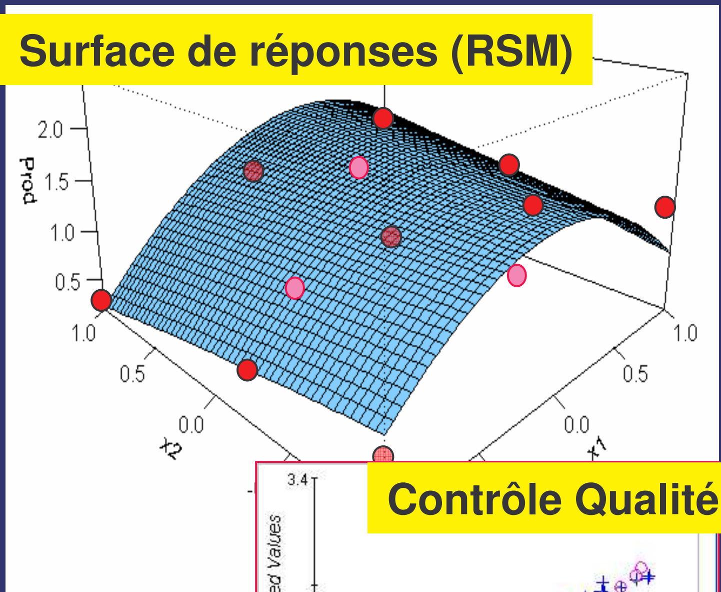




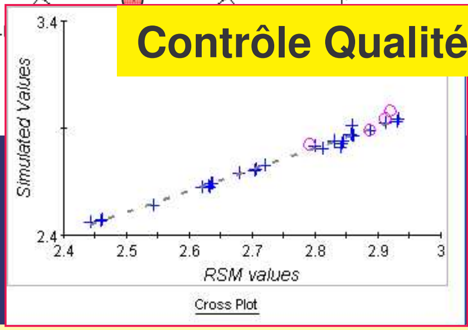
Dynamic training by adding new points

Ingénierie de réservoir : Approches utilisées

Surface de réponses (RSM)

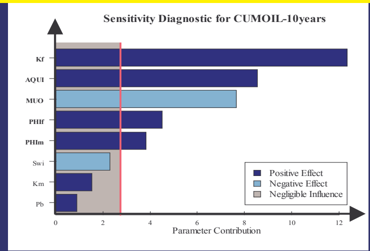


Contrôle Qualité



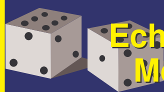
$$CumOil = 28.2$$
$$+5.6kf + 3.2AQUI - 2.7Muo + 2.1PHIf$$
$$+1.8PH\,Im - 1.2kf^2 - 0.5Muo^2 + 0.28Kf.PHIf$$

Analyse de sensibilité

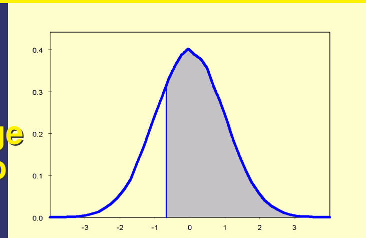


Sensibilité sur prévisions de production

Analyse de risque



Echantillonage Monte-Carlo



Prévisions de production **probabilisées**

Optimisation

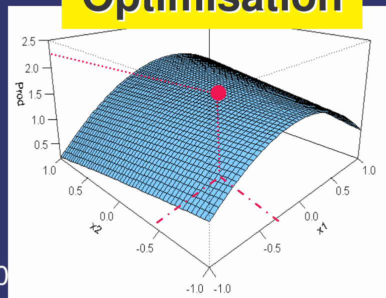
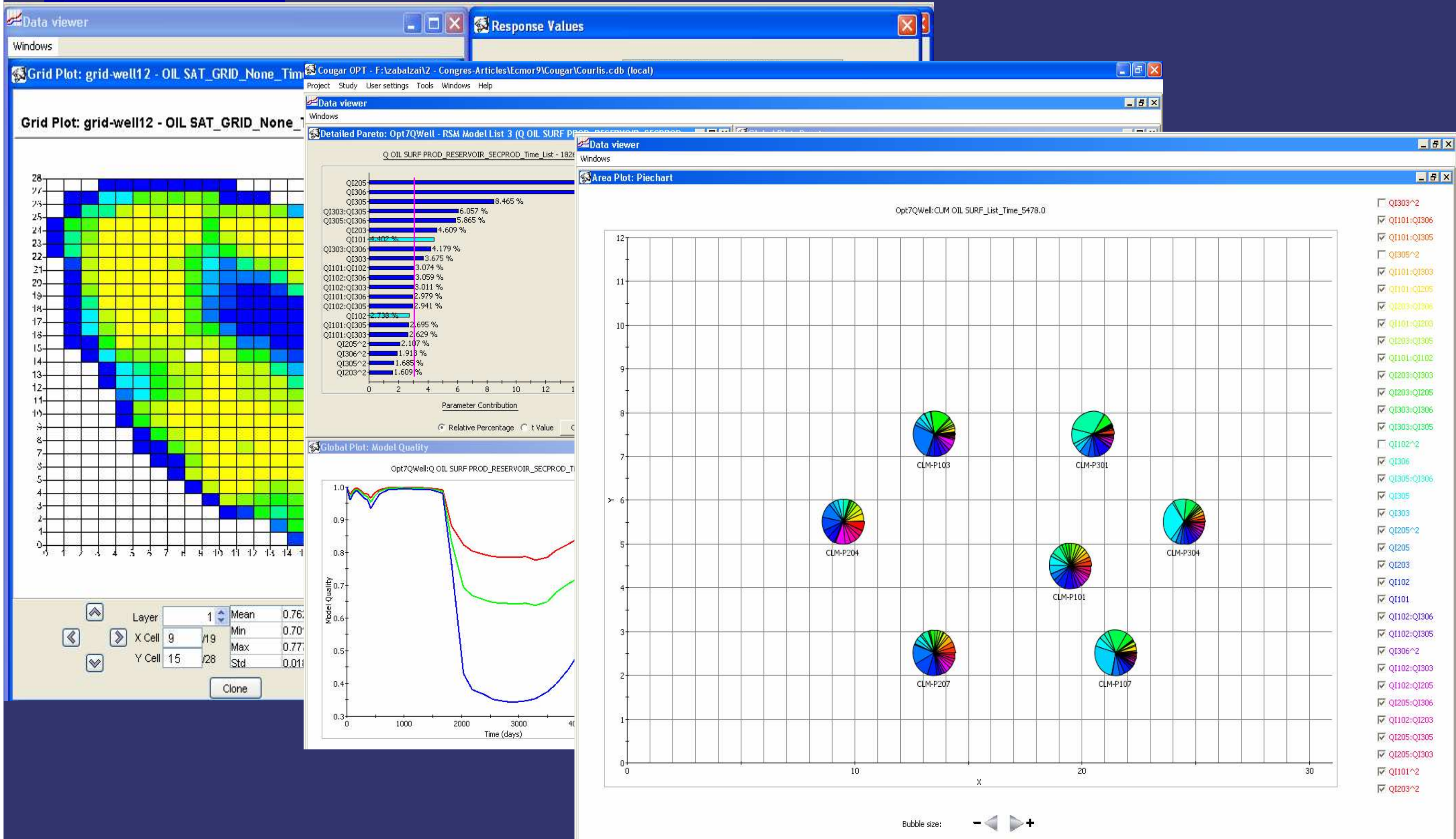
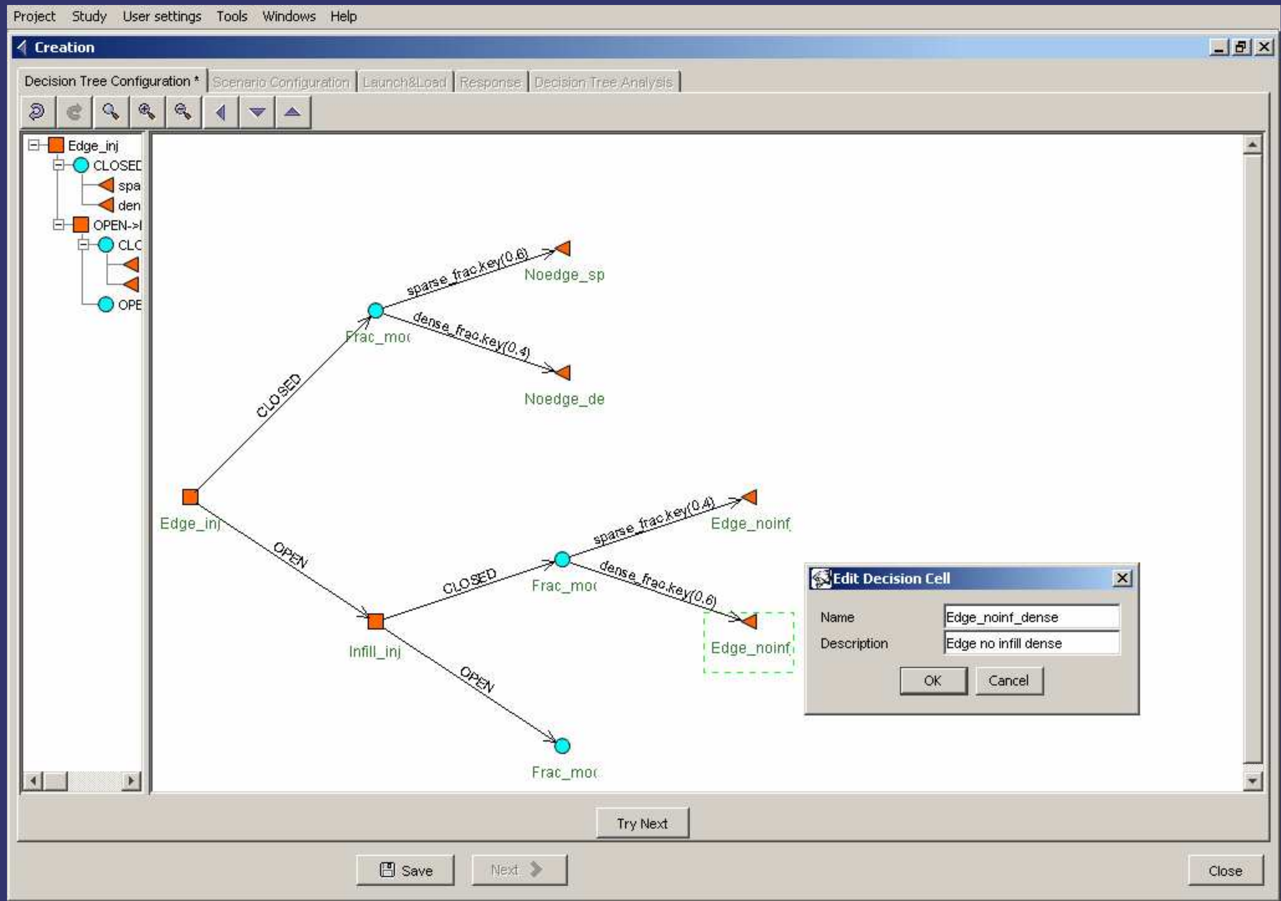


Schéma de développement **optimal**

Ingénierie de réservoir : Approches utilisées



Ingénierie de réservoir : Approches utilisées



Introduction

- Oil Reservoir simulators are used to model very complex physical systems and to solve complicated multiphase flow equations
- Many of the simulator inputs are highly uncertain
- Uncertainty Analysis, Recovery process optimization (under uncertainty), and model calibration (history matching) are fundamental tasks of reservoir engineers
- Because the time for a single simulation can be very long, response surface/emulators methods are very appealing...

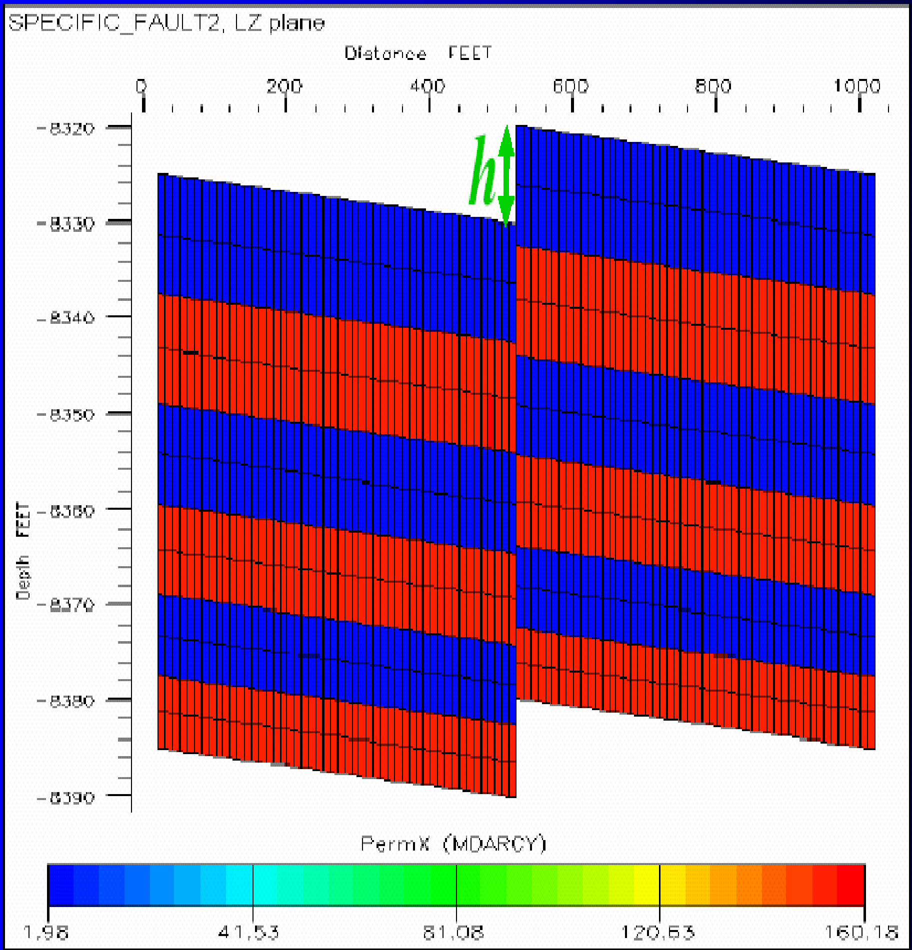
Statistical analysis of simulators outputs

- Classical Experimental Design methods based on regression are sometimes not accurate enough to model response surface, particularly if our objective is optimization (or history matching)
- Non parametric regression methods such as kriging or Bayesian approaches seems more suitable to deal with these problems
- There are however modelling issues that make these approaches difficult to use, particularly for non statisticians

Main Topics of Discussion

1. Model Selection Issues in non-parametric regression
2. Are Gaussian (stationary) models always a good choice?
   What can we do if the response is clearly non-stationary?
3. How reliable are error estimates when the statistical model assumptions are wrong?
   Are there any more robust error estimates? (Bootstrap etc...)
4. How many points to obtain "reliable" approximations?
5. What can we do if our input uncertain parameters are very big?

Example from reservoir forecasting: IC Fault Model



Synthetic 2-D Model

- 2 Wells: 1 Injector, 1 Producer,
- 6 Layers:

  1,3,5 Poor Sand kp (blue)

  2,4,6 Good Sand kg (red)
- 1 Fault

Three Uncertain Inputs:

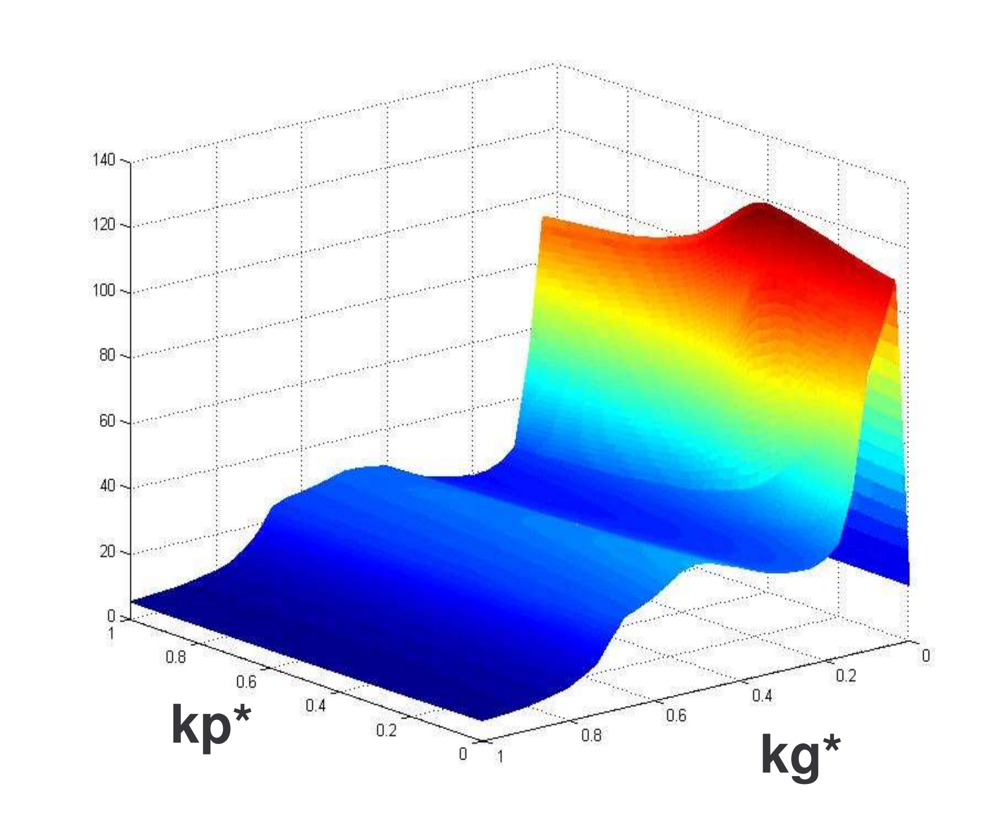
1. fault throw: h [0,50],
2. good permeability: kg [100,200],
3. poor permeability: kp [0,50]

IC Fault Model

Characteristics of this model problem:

1. Very simple model presenting a very complicated response surface as shown in the figure:

2. Used as a test case to show that history matching methods can be very unreliable; In fact the objective function has many local minima



Oil production rate after 10 years vs. Kg and Kp
at a fixed high value of h (obtained with 1000 simulations)

Non Parametric Regression

In some cases there is a need to model complicated non-linear response surfaces: classical experimental design techniques are not sufficiently accurate

In the statistical design and analysis of computer experiments an output s(x) is modelled as a realization of a stochastic process:

$$s(x) = h(x)^T \beta + Z(x)$$

where:

$$h(x) \equiv \{h_0, h_1(x), \ldots, h_L(x)\}$$

and Z(x) is a Gaussian process with

zero mean and a certain covariance function

$$R(x, y) \equiv \sigma^2 C(x, y)$$

1. Modelling Issues

- Choose the regression functions h(x)

$$h(x) \equiv \{x_1, x_2, x_3, \ldots, x_1 x_2, \ldots, x_1^2, \ldots.\}$$

- Choose the correlation function
  c(x,y): exponential
  Matern...

$$c(x,y) = \exp\left(-\sum_{i=1}^{N} \left(\frac{|x_i - y_i|}{\lambda_i} \right)^p \right)$$

$$c(x,y) = \prod_{i=1}^{d} \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{2\sqrt{\nu}\,|x_i - y_i|}{\lambda_i^m} \right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}\,|x_i - y_i|}{\lambda_i^m} \right)$$

Are there any methods for automaticly selecting regression functions and correlation functions (BIC, AIC or others...) ?

2. The stationarity assumption

Response Surfaces such as **ICFM** are clearly non-stationary.

We have estimated by maximum likelihood estimation (MLE) the correlation lengths cl(h),cl(Kg),cl(Kp) of the correlation function

$$c(x,y) = \exp\left(-\sum_{i=1}^{3} \left(\frac{|x_i - y_i|}{\lambda_i} \right)^2 \right)$$

for the IC Fault Model using a maximin LHD with 50 simulations we obtain 2 maxima of the likelihood function

	cl(h)	cl(kp)	cl(kg)	
Local Max 1	1.5	1.5	0.05	Same probability
Local Max 2	1.5	0.21	0.25	

2. The Stationarity Assumption

- Estimates of correlation lengths using maximum likelihood estimation can have multiple equiprobable solutions
- Why? each "stationary" region corresponds to different local maxima of the likelihood function
- Thus the process is non stationary and estimate of hyperparameters is unstable
- Note that this also occurs when using a Bayesian formalism

How to handle this type of non-stationarity?

non-stationary random fields? domain decomposition approaches? ...

3. Variance Estimation

$$MSE(x) = \tilde{\sigma}^2 \left[1 - \left[r^T(x), h^T(x) \right] \begin{bmatrix} R & H \\ H^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} r(x) \\ h(x) \end{bmatrix} \right]$$

The estimated variance or more generally the posterior distribution is also very sensitive to our modelling assumptions

$$IRMSEest = \sqrt{\frac{\sum_{n=1}^{N} (MSE(x_n))}{N}}$$

$$IRMSEemp = \sqrt{\frac{\sum_{n=1}^{N} (s(x_n) - f(x_n))^2}{N}}$$

Comparison of the integrated RMSE empirical (N=1000) and estimated (using kriging variance) for ICFM using maximin LHD with 100 points

	Full Domain	h>40	h<40
IMSEemp	40%	75%	9%
IMSEest	50%	50%	50%

3. Variance estimation

As shown in our numerical experiments on IC Fault model the estimated variance is globally accurate, but can be very inaccurate locally when the response is non-stationary

Are there any more robust approaches to quantify approximation errors (Cross-Validation, Bootstrap...)?

4. Size of the Experimental Design

Depending on the function to emulate and on the input uncertainty range, the number of points necessary to obtain an emulator with a given accuracy can have huge variations

ICFM Integrated RMSE for a kriging emulator on different domains

	1 LHD size 100 on the restricted domain h<40	1 LHD size 100 on the full domain 0<h<60
IMSEemp	7%	40%
IRMSEest	7%	50%

4. Size of the Experimental Design

Sequential Experimental Designs are needed

Stopping criteria?

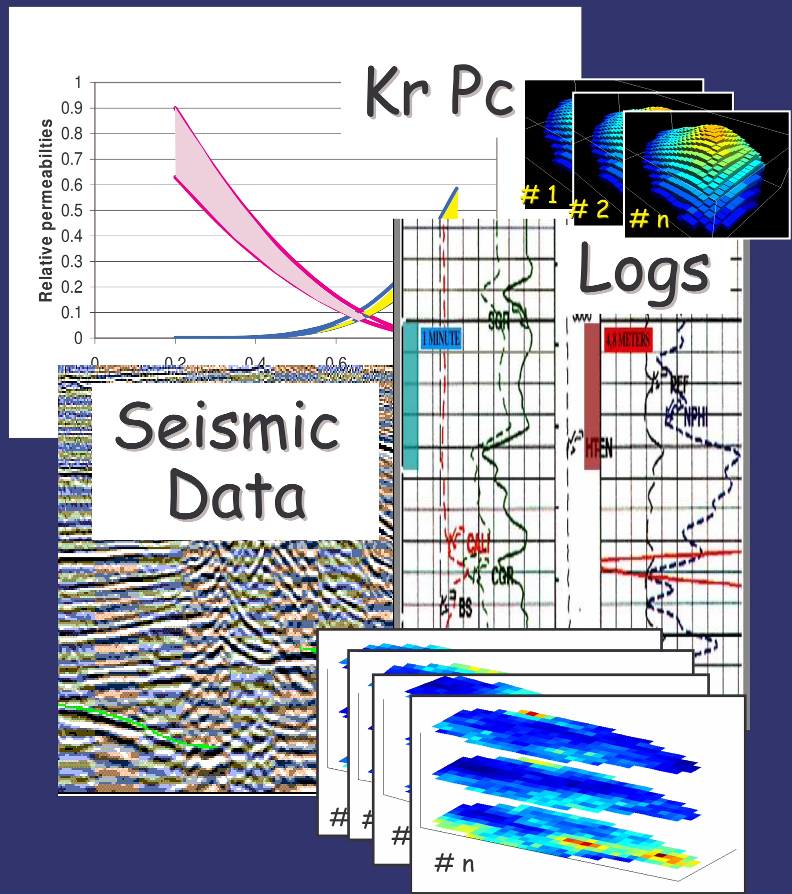
Can we define a complexity
measure of a simulator output
f(x)?

5. When the input is a stochastic process

- A huge number of uncertain parameters will induce a huge amount of simulations to obtain a reliable emulator. Which is unfeasible in reservoir modeling.

To be able to use emulators in these cases the dimension of the problem has to be reduced

What are the current methods to reduce the dimension of a random field (PCA...)? In which conditions are these methods reliable?



References

1) "The design and analysis of computer experiments" Santner, T.J., B.J. Williams, and W.I. Notz (2003), . Springer-Verlag, New York

2) "Bayesian calibration of computer models. M.C. Kennedy & A. O.Hagan

Journal Royal Statistical Society, Series B, 63, Part 3, 2001, pp. 458-45

3) “Hierarchical non-linear Approximation for Experimental Design and Statistical Data Fitting” 2007 SIAM Journal of Scientific Computing

D. Busby, C. Farmer, A. Iske

4) "Adaptive Experimental Design for non-linear Modeling"

Application to quantification of risk for real field production

2006 ECMOR X, C. Scheidt, D.Busby, I. Zabalza-Mezghani, M. Feraille,D. Collombier

5) "The correct Kriging variance estimated by bootstrapping" Den Hertog, D., J.P.C. Kleijnen, and A.Y.D. Siem (2006), . Journal of the Operational Research Society, 57, no. 4, pp. 400-409