

Quelques leçons tirées de mes expériences de simulation

Claude Mayer

Santa Café 30 janvier 2017

Plan

- Résumé de mes expériences
 - Réalisations personnelles
 - Etudes de réalisations existantes
- Premières leçons - tirées au fur et à mesure
- Considérations plus avancées
 - Approche inverse, relation avec l'approche bayésienne
 - Méthode de Monte-Carlo
 - Simulations multi-échelles
- La modélisation / simulation en pratique
 - Les étapes successives d'un projet
 - Sources possibles d'erreurs (*modélisation / exploitation*)
 - Le problème de la sous-détermination
 - Contrôle qualité et critères de fiabilité
- Conclusion

Contextes d'entreprises

- 1971 – 1974 : SEMA
 - Simulation « abeilles » Port du Havre (rapport sur étude tierce)
 - Réseau Télex Armée de l'Air (Etude, réalisation, exploitation)
 - Pénurie de pétrole en France (Modèle linéaire vs. simulation)
- 1974 – 1980 : SCHLUMBERGER « WIRELINE »
 - « GLOBAL » : Simulation inverse de mesures de puits de pétrole (Développement, promotion, implantation, publications)
 - Simulation de réservoirs (étude de systèmes existants)
- 1982 – 1987 : CETIS (créateur et PDG)
 - « ULTRA » : Extension de « GLOBAL », développements et vente d'un système inverse d'interprétation de mesures
 - « AUTOMOD » : Inversion double multi-échelles, publications, vente
- Autres produits développés et vendus, hors simulations :
RORSCHACH, LOGIDENT, HYPERTIME, VISUAL PLANNING

Simulation remorquage Port du Havre

(Compte-rendu d'une étude → chapitre d'un livre)

Port du Havre - Le remorquage : Les Abeilles



Il n'est pas obligatoire. Toutefois, pour des raisons de sécurité, le Commandant de Port peut imposer un ou plusieurs remorqueurs. La compagnie de remorquage les Abeilles dispose d'une flotte de neuf remorqueurs de 3000 à 5 000 chevaux (40 à 65T de traction) équipés de puissants moyens de lutte contre l'incendie.

Simulation remorquage Port du Havre



Le petit poussait...

Le problème (1970)

- Optimiser le rapport performance/coût du remorquage en variant le nombre d'« abeilles »
- Forte concurrence (Rotterdam, Anvers...)
 - *Performance (pour le client)* : temps d'attente, prix du remorquage
 - *Coût (pour le Port)* : ~ nombre d'abeilles (très cher...)
- Préparer l'avenir
 - Scénarios *d'évolution de la demande* (trafic, taille des navires, complexité, exigences...)
 - Scénarios *d'évolution de la concurrence (?)*

Simulation

- Développer un modèle de trafic
 - Nombre et types d'abeilles
 - Caractéristiques de la **demande** (*taille des navires, nombre et profil des arrivées, objectifs d'attente...*)
 - **Contraintes** (*types d'abeilles, règles, priorités, incidents, emplacements disponibles...*)
- Résultats attendus
 - **Temps d'attente** moyen, min, max, par type de navires...
 - **Simulations d'hypothèses** : trafic, nombre d'abeilles
 - **Simulations de rentabilité**

Caractéristiques de cette simulation

- Modèle direct
 - Prévoir le trafic selon une demande en entrée
- Ajustements et tests
 - Par « *history matching* » (*trafic passé connu*)
- Exploitation sous diverses hypothèses
 - Pour un objectif de trafic et de coût donné, *multiplier les simulations* pour obtenir un dimensionnement « optimal » du système
- Exploitation probabiliste
 - Les arrivées de navires correspondent à des *lois de probabilités* et doivent être traitées comme telles (*Monte-Carlo*)

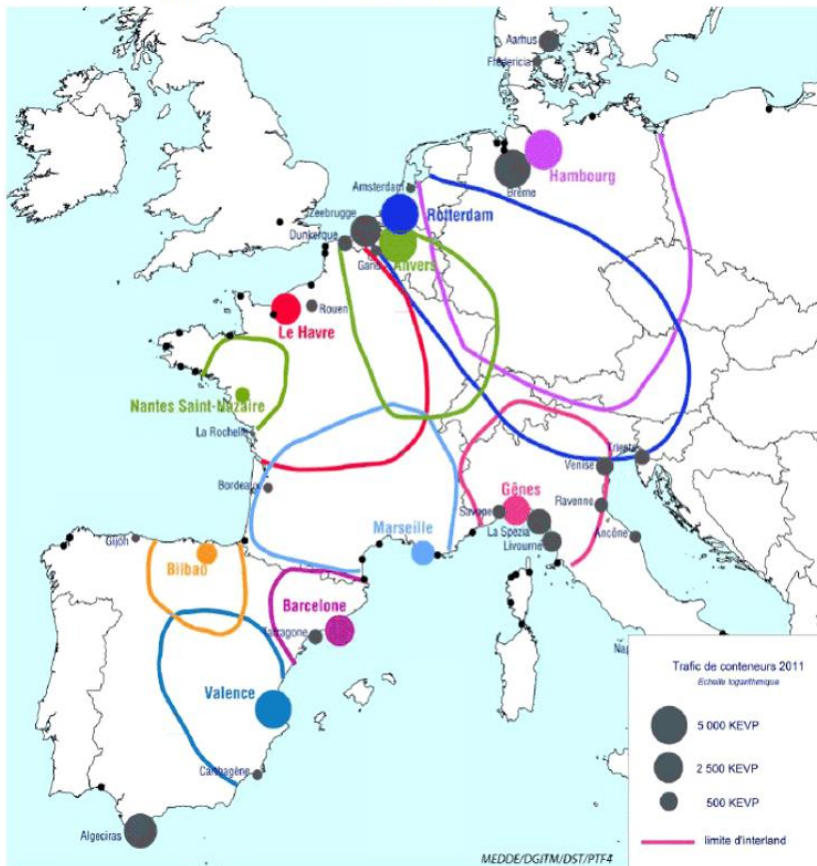
Pourquoi une simulation ?

- Aucun modèle numérique n'était possible
 - Problème non linéaire
 - Objets hétérogènes
 - Hypothèses nombreuses
 - Règles nombreuses
- Expérimentation par maquette irréalisable
- Tests rapides d'hypothèses multiples
- Coût de la simulation << Bénéfice escompté

Résultats (→ 2015)

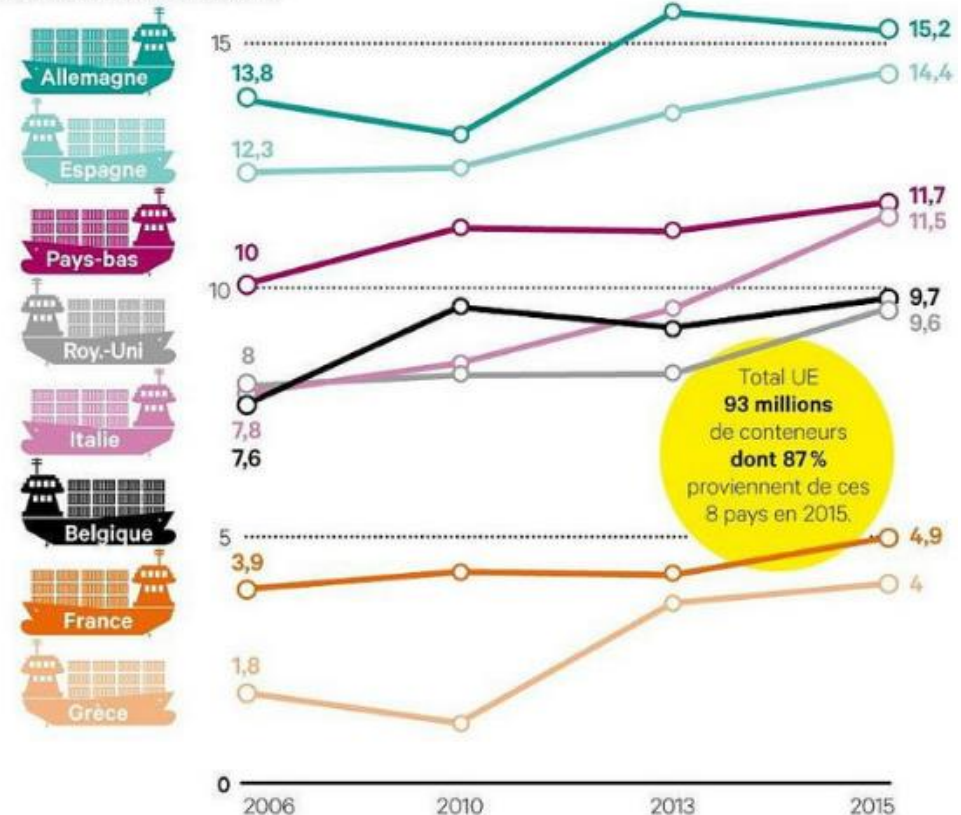
Malgré cette excellent étude, les ports français sont en constant recul relativement à leurs concurrents européens

HINTERLANDS DES PRINCIPAUX PORTS FRANÇAIS ET EUROPÉENS



Ports à conteneurs en Europe : la France stagne

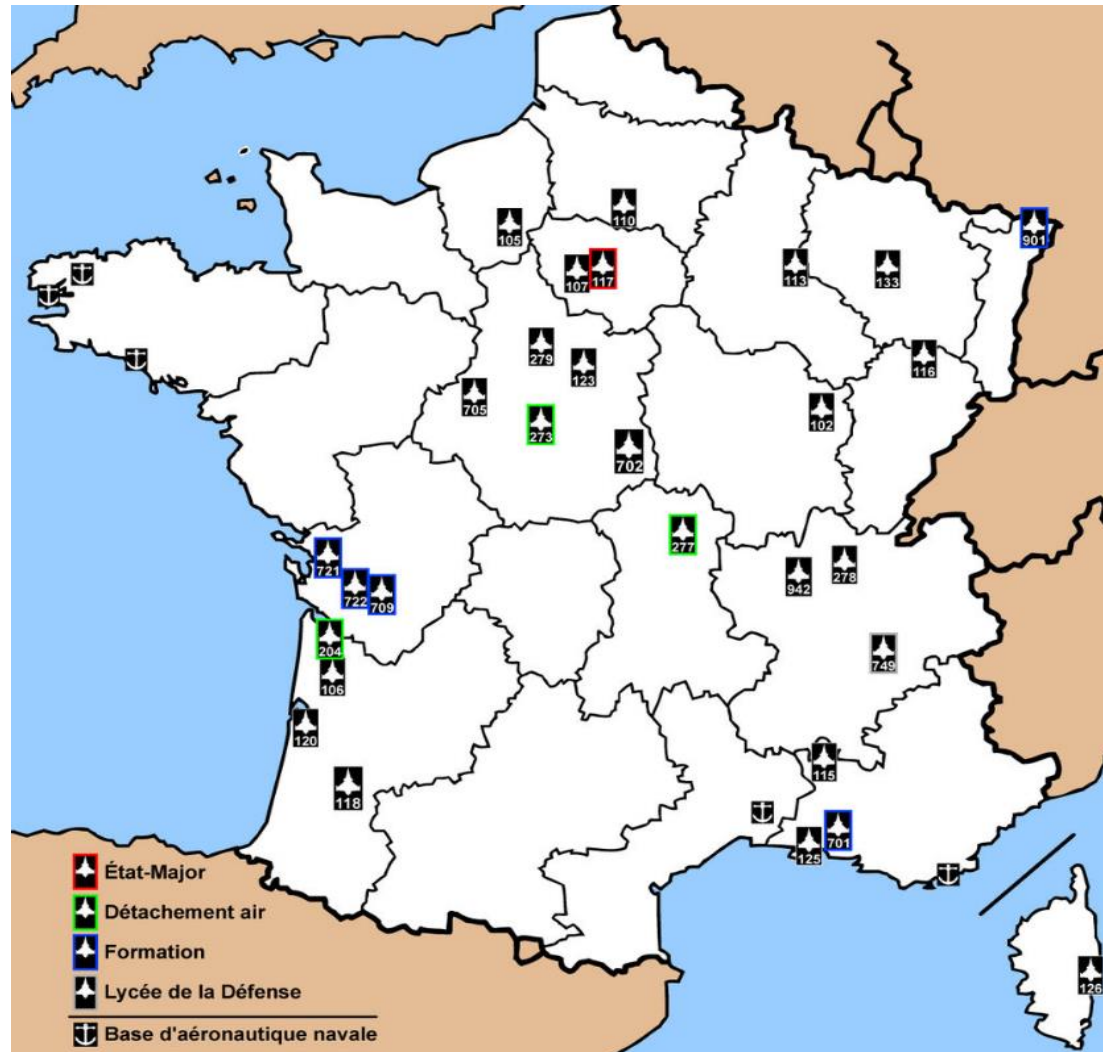
En millions de conteneurs EVP



STTA – Simulation du Télex AA

- Services Techniques Télécom Armée de l’Air
- Réseau de télétransmissions par Télex (1971)
 - Messages point à point entre Bases Aériennes
 - Lignes privées (*SAT*), Nœuds privés
 - Sortie sur ruban de papier perforé
 - 5 niveaux de priorité (*1 à 4, 5 = « Flash »*)
- Le réseau « *actuel* » 1971 est-il bien dimensionné ?
 - Y a-t’il des fragilités ? (*destruction de lignes...*)
 - Que se passe-t-il selon divers *scénarios de trafic* ?
 - Objectif : *Temps de transmissions par priorités*
 - Moyens, Mini, Maxi

La carte des bases aériennes



Etude STTA - déroulement

- *1^{ère} conclusion* : le réseau est largement *surdimensionné* par rapport au *trafic ordinaire*
 - *Les généraux en retraite de la SAT vendaient des lignes doublées aux généraux d'active de l'Armée de l'Air*
- Je décide de tester le réseau dans des configurations plus « *musclées* »
- *J'effectue une enquête* pour déterminer des hypothèses de trafic en temps de conflit
 - Même si ce n'était pas la demande explicite du client, je m'imaginais que l'Armée était faite pour faire la guerre...
 - Mes interlocuteurs pensent plutôt « *exercice* » ou « *manœuvre* », et à l'extrême limite si j'insiste : « *crise* »...
 - Je suis très bien reçu : ces officiers sont des gens très *civils*...
- A défaut d'hypothèses de trafic fournies par mes interlocuteurs, *je propose et fais approuver des hypothèses qui me semblent raisonnables*

Simulation et résultats

- Le modèle représente fidèlement l'état actuel du réseau (*pas de simplification sur la structure*)
- *Programmation intégrale* (*pas d'outils disponibles à cette époque*)
- L'objectif essentiel de la simulation est de tester les performances du réseau en cas de *période intense* (*crise, conflit*) et en cas de *dégradation du réseau*
- La simulation a fourni la réponse du réseau à diverses *hypothèses*, à l'intention des décideurs
 - Pour fournir des temps *moyens, mini et maxi*, il a fallu procéder à de nombreuses simulations selon des *lois probabilistes* d'émission des messages (*Méthode de Monte-Carlo*)
- Les *résultats* confirment le bon dimensionnement des *lignes* (*elles étaient doublées*), mais le *maillage* peut être insuffisant en cas de *destruction de nœuds*

Ce que j'avais retenu de ces études

- En 1970, la « *Recherche Opérationnelle* » à la SEMA était encore largement basée sur des *modèles linéaires*
- Cependant, dans de nombreux cas (*en fait, presque toujours*) *la réalité n'est pas linéaire*
- Dans ces cas-là, la simulation numérique est *le meilleur moyen (en général le seul)* pour comprendre un système et pour réfléchir aux conditions de son exploitation
- De plus, cette approche est extrêmement *flexible* et permet de tester toutes sortes *d'hypothèses*
- *J'étais devenu un adepte, et j'étais donc mal préparé à la mission qui va suivre...*

Prévision de pénurie de pétrole (1972)

- **Objectif de l'étude** : quelles conséquences pour l'Industrie française d'une pénurie *importante* et *durable* de l'approvisionnement en pétrole ?
- **Moyen d'étude imposé** : les matrices de Leontief (système linéaire)

Matrices de Leontief

- Ce sont des tableaux permettant de croiser:
 - Pour chaque type de produit, l'utilisation que sa fabrication exige en autres produits, matières premières, énergie

	Biens nécessaires		
	<i>X</i>	<i>Y</i>	<i>Z</i>
Produire une unité de <i>X</i> demande :	0,3	0,4	0,1
Produire une unité de <i>Y</i> demande :	0,5	0,2	0,6
Produire une unité de <i>Z</i> demande :	0,1	0,3	0,1

- Il existe autant de matrices de Leontief que de secteurs d'étude, c'est le principe qui compte
- L'application *linéaire* de ces matrices permettait de traiter le problème posé, en supposant que toutes ces interactions restent *linéaires sur le domaine de l'étude*

Limites de la linéarité en cas de pénurie

- En cas de pénurie « *importante* » et « *durable* », l'hypothèse linéaire *ne tient pas*
 - *Les industries employant du pétrole comme source d'énergie tentent de basculer vers d'autres types d'énergie*
 - Le gaz (reconversion souvent possible des chaudières)
 - L'électricité d'origine non thermique
 - En recul mais possible : charbon, bois...
 - *Les industries fabriquant des produits dont certains composants intègrent beaucoup de pétrole peuvent leur substituer des composants plus économes*
 - Par ex : matières plastiques → métaux...
 - Composants dont la fabrication consomme moins d'énergie
- J'ai proposé de réaliser une simulation avec scénarios
- Devant un refus net, j'ai préféré quitter le service

Simulation vs modèles linéaires (1970s)

- En faveur du linéaire :
 - Les *mathématiques* en sont comprises depuis des siècles
 - Ce sont les *méthodes traditionnelles*
 - Il peut exister des *solution explicites*
 - Les études, étant linéarisées, sont allégées, *moins coûteuses*
 - La linéarité peut constituer une *bonne approximation* pour les petits domaines
- Contre le linéaire (et pour la simulation) :
 - Le monde n'est pas linéaire, et *l'illusion de linéarité* peut conduire à de graves erreurs dès qu'on s'écarte du domaine de validité (*et on ne connaît même pas les limites de ce domaine !!!*)
 - Le non-linéaire est *plus exact et plus simple à modéliser*, car on reproduit ce qu'on voit au lieu de se tordre le cou à faire des hypothèses simplificatrices
 - La simulation permet de *suivre pas à pas* le déroulement des processus simulés
 - Le non-linéaire est *l'avenir*, car la puissance croissante des machines lui permet de se développer dans les *domaines les plus complexes*
 - La modélisation linéaire consiste souvent à *chercher la clé sous le réverbère*

Interprétation de mesures pétrolières de puits

Schlumberger « Wireline services »

1974 - 1980

Exploration pétrolière

Principes du « logging »

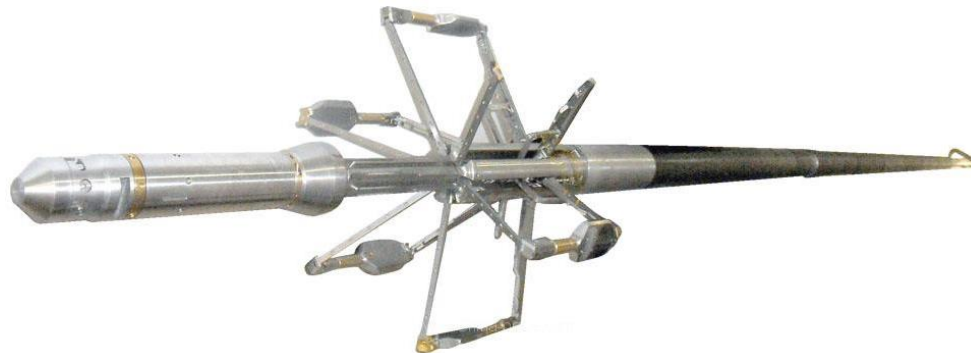
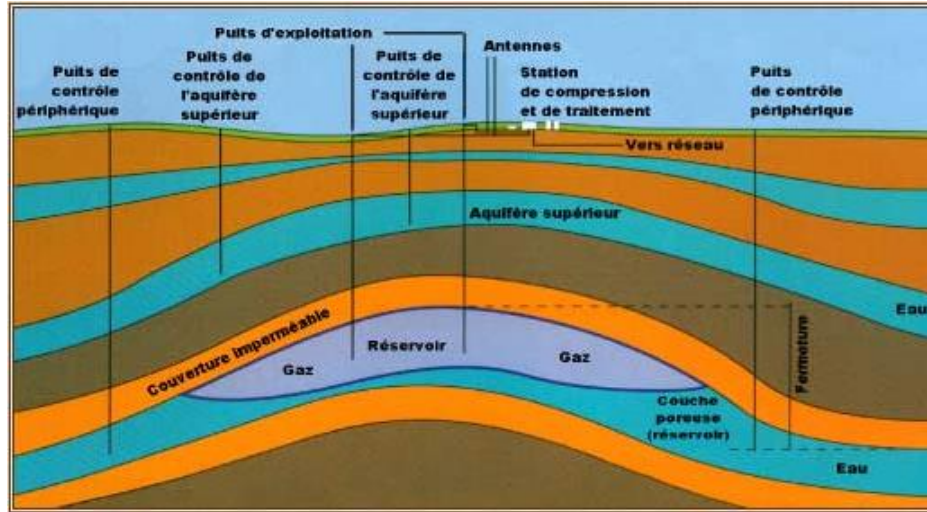
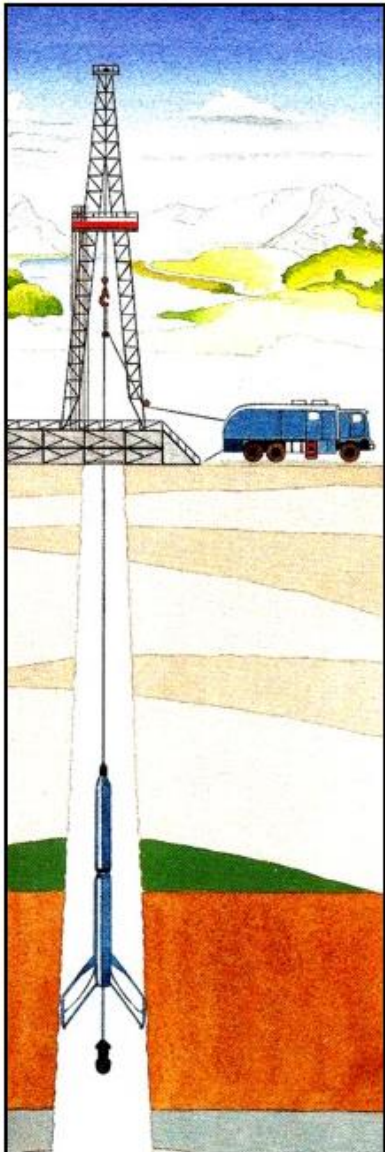
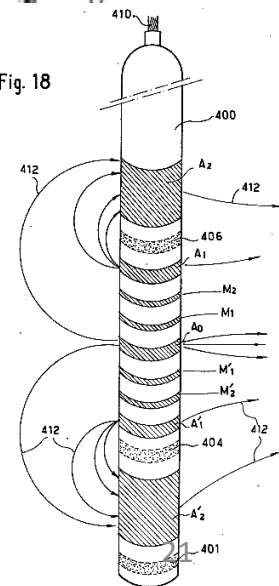


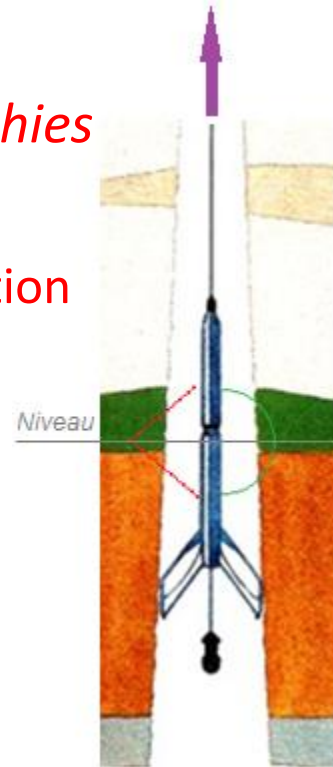
Fig. 18



Exploration pétrolière

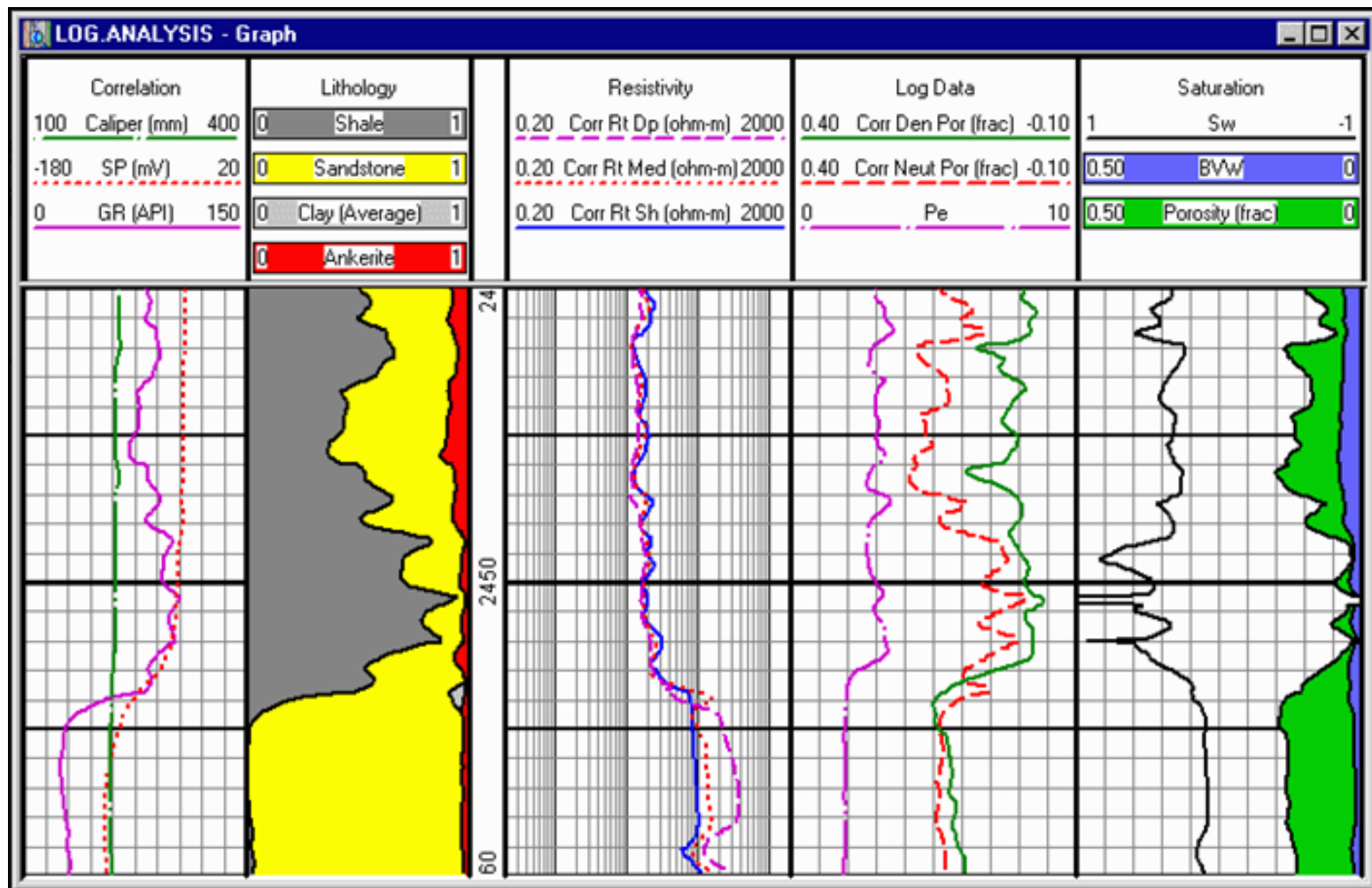
Principes du « logging »

- Mesures dans un puits d'exploration (« *logs* » ou « *diagraphies différées* ») par des « outils » de *diverses technologies* :
 - Réception de rayons gamma après avoir bombardé la formation
 - Par des *neutrons lents*
 - Par des *rayons gamma*
 - Vitesse et formes d'ondes sonores traversant la formation
 - Conductivité électrique
 - Radioactivité naturelle
 - Résonance magnétique nucléaire (IRM)
- Interprétation : à chaque *niveau de profondeur*, on doit estimer :
 - La *composition solide* de la roche (*lithologie*) : minéraux, argiles...
 - La *porosité* et si possible la *perméabilité*
 - La *teneur des pores en fluides* : eau, huile, gaz



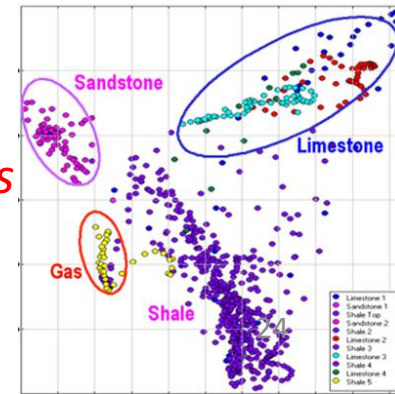
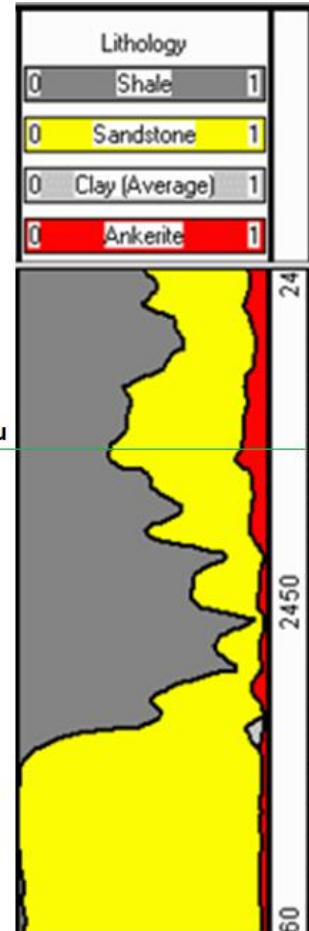
Une impression typique

Logs et résultats



Interprétation « classique »

- A chaque niveau de profondeur, on mesure :
 - $A = (A_1, \dots, A_n)$ (les *données* : lectures des outils)
- On recherche pour ce niveau :
 - $X = (X_1, \dots, X_k)$ ($k \leq n$) (les *résultats* recherchés)
 - ex : $X_1 = \% \text{ calcaire}$, $X_2 = \text{porosité}$, $X_3 = \text{teneur en huile}$...
- Approche « classique » (la seule connue avant 1976)
 - On recherche une *heuristique* $A \rightarrow X$
 - Mais il n'existe *aucune relation fonctionnelle* directe $X = f(A)$
 - On recourt à de nombreuses *abaques* et « *cross-plots* »
 - Cette approche est *rigide* et *non adaptée* à toute suite de logs
 - Elle n'est pas adaptée aux *lithologies complexes* (mica...)
 - Elle tient mal compte des *connaissances a priori*
 - Elle n'utilise pas pleinement toutes les mesures disponibles
 - Elle tient mal compte des *incertitudes de mesure* (cutoffs)
 - Elle n'est pas adaptable aux *futurs outils issus de la R&D*



Modèle direct

- Soient **A** les données (*mesures outils*), **X** les inconnues (*la formation*)
- On recherche une procédure $A \rightarrow X$
 - Mais on a vu qu'*il n'existe pas* de telle procédure autre qu'une heuristique complexe, peu flexible et peu fiable
 - Cette heuristique est « *l'approche classique* »
- En revanche il existe un « *modèle direct* » $X \rightarrow A$
 - *C'est le sens « naturel » : Cause $\rightarrow$ Effet*
 - Il s'agit de la *simulation des « réponses »* des outils
 - Chaque outil est étudié théoriquement et en laboratoire et l'on sait *prédire sa réponse* à une formation **X**
 - C'est le « *modèle direct* » $X \rightarrow A$
- *NB : Le modèle direct $X \rightarrow A$ est disponible dans le sens opposé à l'approche classique $A \rightarrow X$.*
 - *(De fait, l'approche classique utilise par morceaux le modèle direct mais de manière partielle et rigide)*

Approche inverse

« Méthode globale »

- Comment utiliser le *modèle direct* $X \rightarrow A$ pour résoudre *dans toute sa généralité* le *problème posé* $A \rightarrow X$?
- C'est l'approche inverse, par *essais et erreurs*
 - Essais : on teste une hypothèse X_0 sur le résultat
 - Simulation : on applique le *modèle direct* qui donne A_0 , la mesure attendue si X_0 était la solution
 - Erreur : on établit une *distance (écart)* entre A et A_0 : $d(A, A_0)$: plus l'écart est grand, plus on est loin de la solution
 - Optimisation : on effectue une série de tests $X_1, X_2...$ et on simule à chaque fois, jusqu'à ce que la suite $A_1, A_2...$ *converge vers un optimum* proche de A
 - Cela revient à *minimiser l'écart* $d(A, A_n)$
- J'ai pu démontrer que cette solution « optimale » était aussi *la plus probable*, sous des hypothèses très générales tenant compte des *incertitudes de mesure* et des *connaissances géologiques locales*
 - La fonction d'écart d doit alors tenir compte des *contraintes a priori sur les résultats* et des *incertitudes de mesures*

Schéma de l'approche inverse

Essais / Erreurs

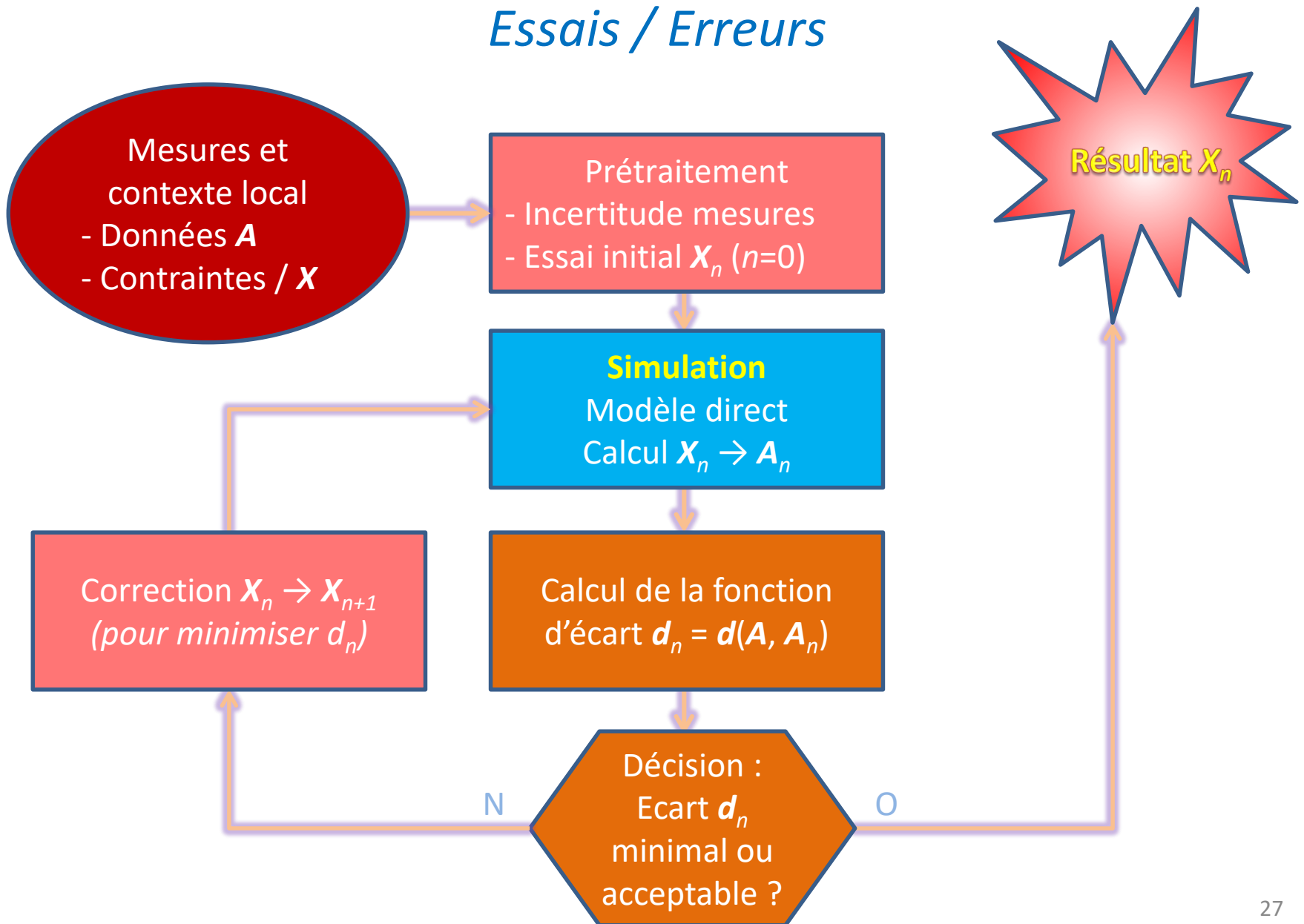
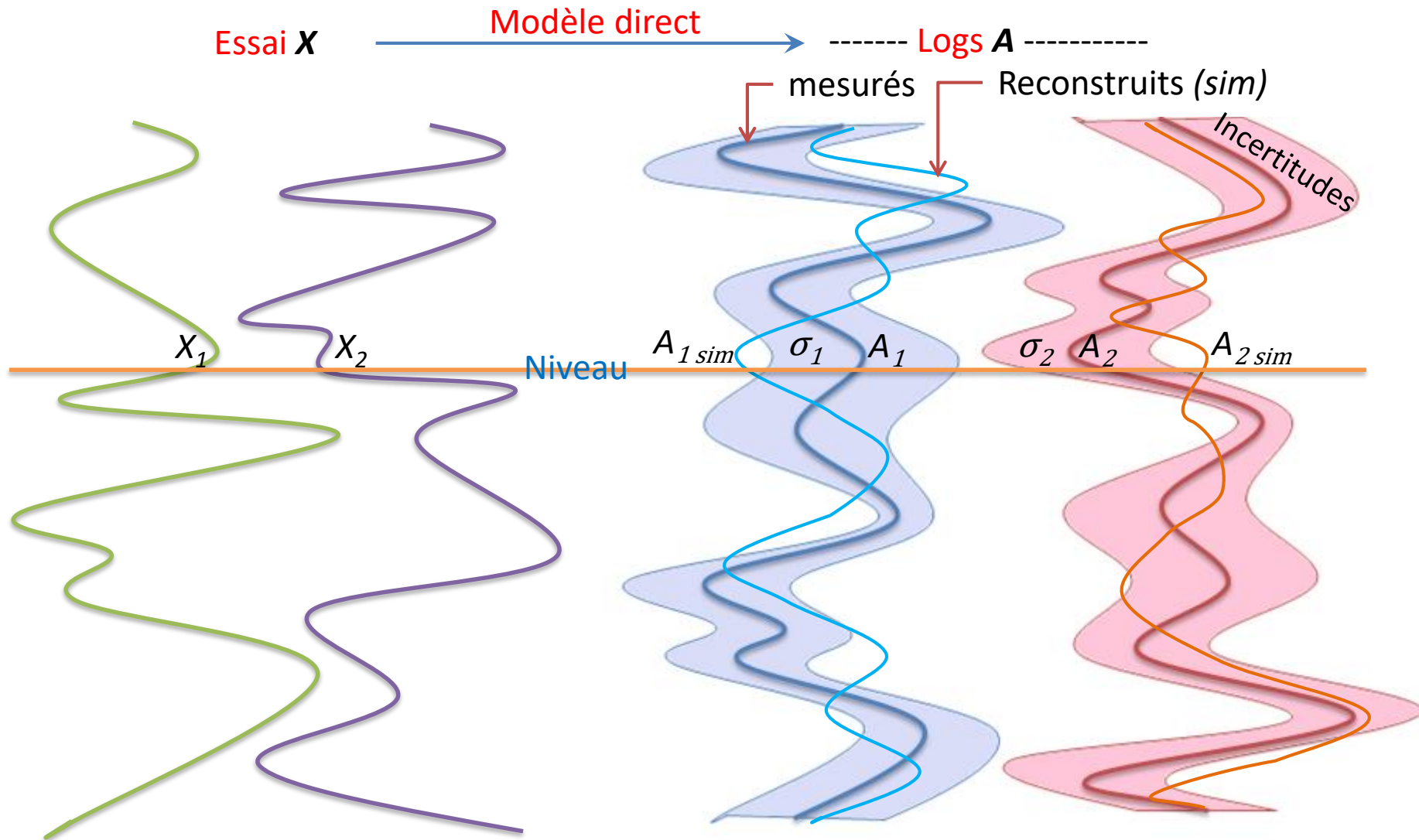


Schéma du processus et Contrôle qualité



$$d(A, A_{sim}) = \text{Somme des } [(A_i - A_{i sim}) / \sigma_i]^2 \text{ pour tous les logs } i + \text{pénalités}$$

« Méthode globale » : Avantages *par rapport à l'approche classique*

- Flexibilité par rapport à la *suite d'outils* utilisée (*standard*, *spéciaux* ou issus de la *recherche future*)
 - Pour chaque outil i , il suffit d'établir sa *fonction de réponse* $X \rightarrow A_i$ et sa *sensibilité* σ_i aux erreurs de mesure : simulation de la réponse de l'outil (*établie par la R&D*)
 - Chaque outil est introduit comme un *simple terme additif* dans la fonction d'erreur d
- On intègre ainsi *toutes les connaissances* sur la réponse des outils
- On tient compte des *incertitudes sur les mesures* σ_i , dépendant elles-mêmes des conditions de mesure
- Tous les outils sont traités *globalement* et simultanément
- Flexibilité par rapport au contexte géologique local : on peut faire varier les *composantes que l'on veut du vecteur de résultats* X en fonction du type de formation attendue
- On prend en compte des *contraintes a priori* $g(X)=0$ (*environnement géologique*)
- On obtient la *solution bayésienne la plus probable* (*maximum likelihood*)
- On peut également estimer l'*incertitude sur le résultat* X et l'intégrer dans des estimations de réserves

Considérations générales sur la simulation par *approche inverse*

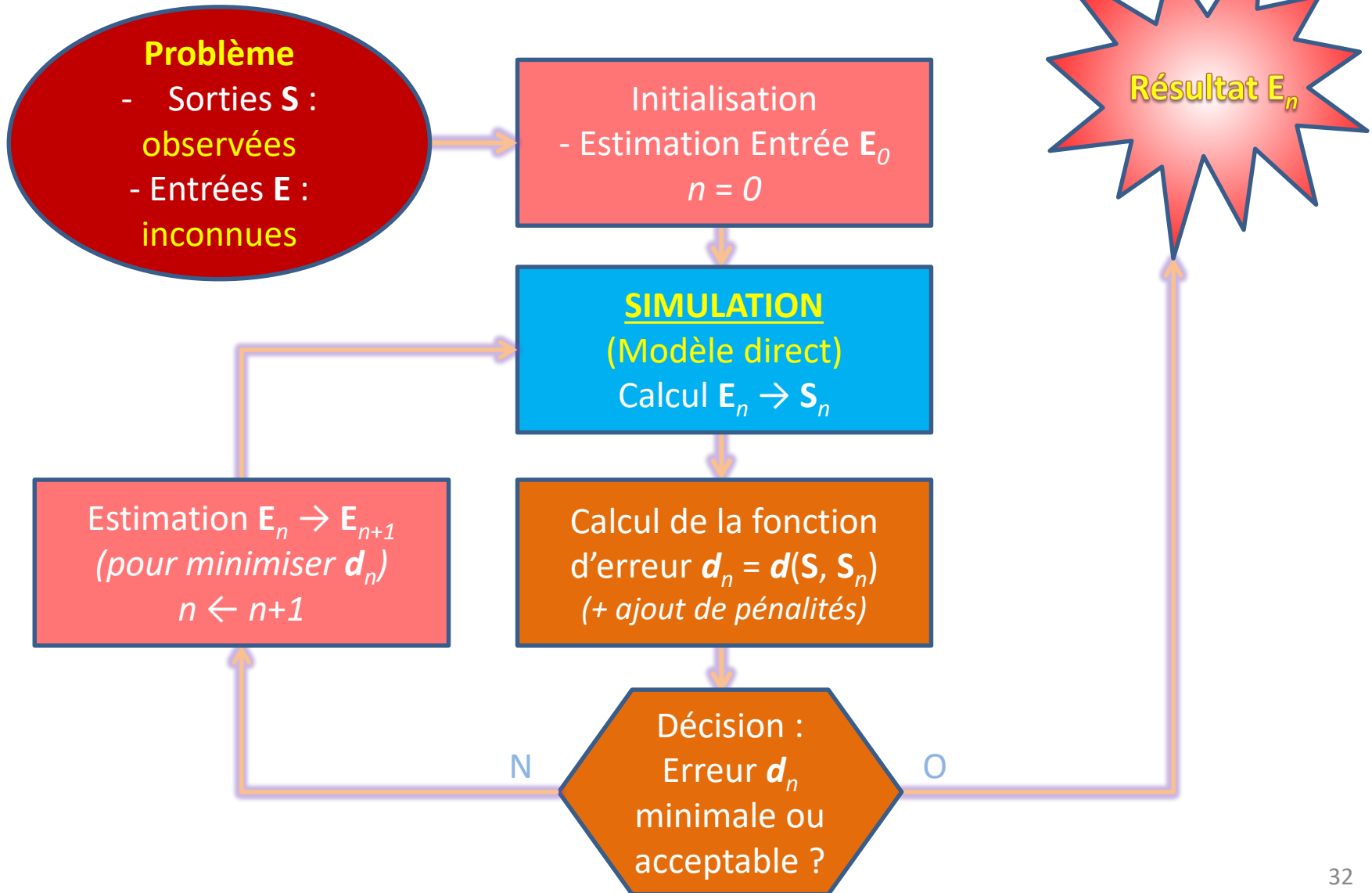
Itérations des simulations d'un modèle direct pour en rechercher les entrées connaissant les sorties

Variables d'une simulation

- Une simulation est constituée :
 - D'un modèle M
 - Ce modèle dépend de paramètres $p : M(p)$
- A l'exécution
 - Elle admet des entrées E
 - Elle réalise une *fonction* dont les sorties sont S
 - $S = [M(p)](E)$ ou encore $M(p) : E \rightarrow S$
- Une telle modélisation constitue le *modèle direct*
- NB : Toutes les entités ci-dessus peuvent être des *nombres*, des *vecteurs* ou des *entités non numériques*

Approche inverse générale

Essais / Erreurs



Objectifs approches directe vs. inverse

- **Modèle direct** (sens de la simulation) $(E, p) \rightarrow S$
- **Approche directe** : on simule un comportement
 - Le modèle est direct et s'applique tel quel à E et p
 - *Objectif : Prévoir le comportement S du système (futur ou extrapolé)*
- **Approche inverse** (*le sens de la simulation reste direct*)
 - 1. Recherche de E connaissant p et S (*itérations sur E*)
 - *Objectif : Diagnostic, données inconnues, recherche des causes...*
 - 2. Recherche de p connaissant E et S (*itérations sur p*)
 - *Optimisation des paramètres d'un système, mise au point modèle*
- **Cas mixtes, plus ambitieux**
 - Recherche des paramètres ET des résultats !
 - Recherche de p et S connaissant E (*itérations sur p et S*)
 - Ou recherche de p et E connaissant seulement S
 - Problème de *sous-détermination ? (voir plus loin)*

Choix du *sens* du modèle direct

- On cherche *E* connaissant *S*
- Une approche inverse, itérative, serait *inutile* et *trop lourde* si l'on savait modéliser *directement* le comportement $S \rightarrow E$
- Mais cela n'est pas toujours possible, en particulier s'il y a une relation de *causalité* dans le sens $E \rightarrow S$
- Exemples de causalité $E \rightarrow S$
 - *Etat de la nature* $\rightarrow$ *Observables*
 - *Nature d'une panne, d'une maladie* $\rightarrow$ *Symptômes*
 - *Structure d'un système* $\rightarrow$ *Propriétés émergentes (observables)*

Bayes et l'approche inverse

- Lorsque l'approche inverse a pour objectif de calculer les **entrées cachées** E d'un système dont on **connaît les sorties** S (*observables*), on est dans le contexte-type de **l'approche bayésienne : recherche des causes**
- La philosophie de l'approche bayésienne est de tenir compte, pour estimer E , des **probabilités de transition** $\mathcal{P}(S|E)$ et des probabilités *a priori* $\mathcal{P}(E)$
- L'approche inverse générale *met en œuvre* cette philosophie
 - Les probabilités de transition sont simulées par le **modèle direct** $E \rightarrow S$
 - Les probabilités a priori sur E peuvent être rendues par des termes de la **fonction d'erreur** d qu'il faut minimiser (*d +grand si probabilité a priori faible*)
- Dans le cas particulier de la « *méthode globale* » pétrolière, les hypothèses adoptées représentent mathématiquement une **approche bayésienne stricte (= formule de Bayes)**
 - La fonction d tient compte de manière probabiliste des connaissances a priori sur les résultats
 - L'optimum de la fonction d correspond à un « *maximum likelihood* » sur E , au sens de Bayes

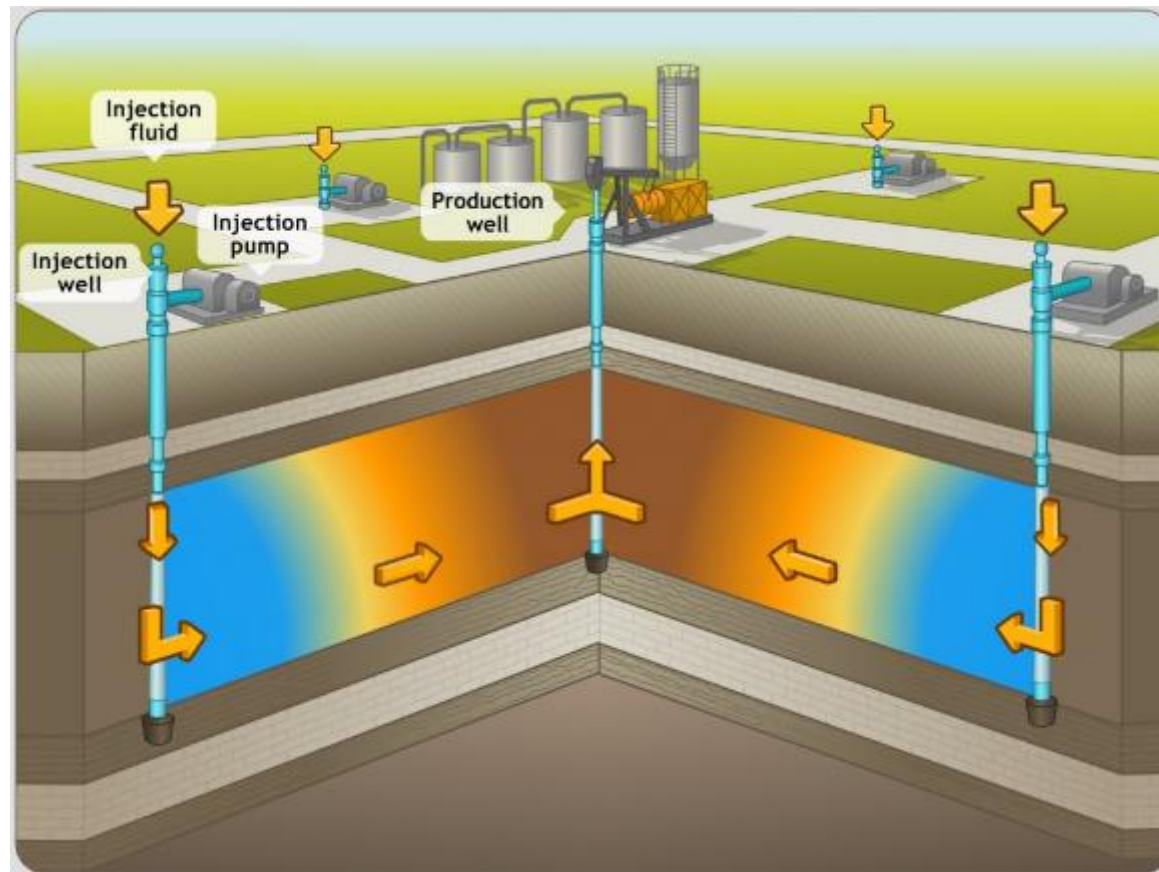
Simulation de réservoirs pétroliers

Objectifs

- Un « *réservoir* » est un champ pétrolifère contigu
- Il peut s'étendre sur de vastes surfaces (*zones, pays...*)
- Il comporte des poches de pétrole et de gaz, qu'il faut produire *sans laisser de poches abandonnées*
- La géologie est *très complexe* et doit être *estimée en 3D* ;
l'écoulement des fluides (eau, huile, gaz) est également complexe
- L'objectif est de tester l'efficacité de diverses *configurations de production*
 - Nombre et localisation des *puits de production*
 - Nombre et localisation des *puits d'injection*
 - Procédés de *recouvrement secondaire* : injection de vapeur ou de solvants, fracturation hydraulique...

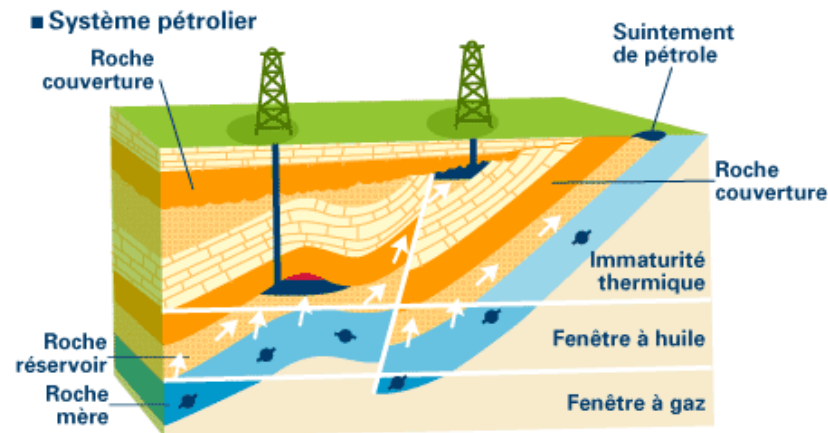
Simulation de réservoirs pétroliers

Principe de l'injection

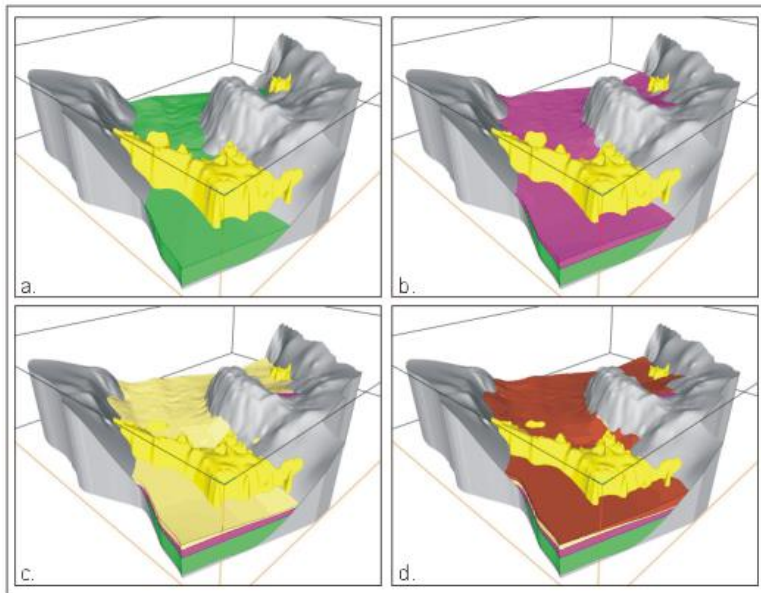


Eau, vapeur, polymères, CO2...

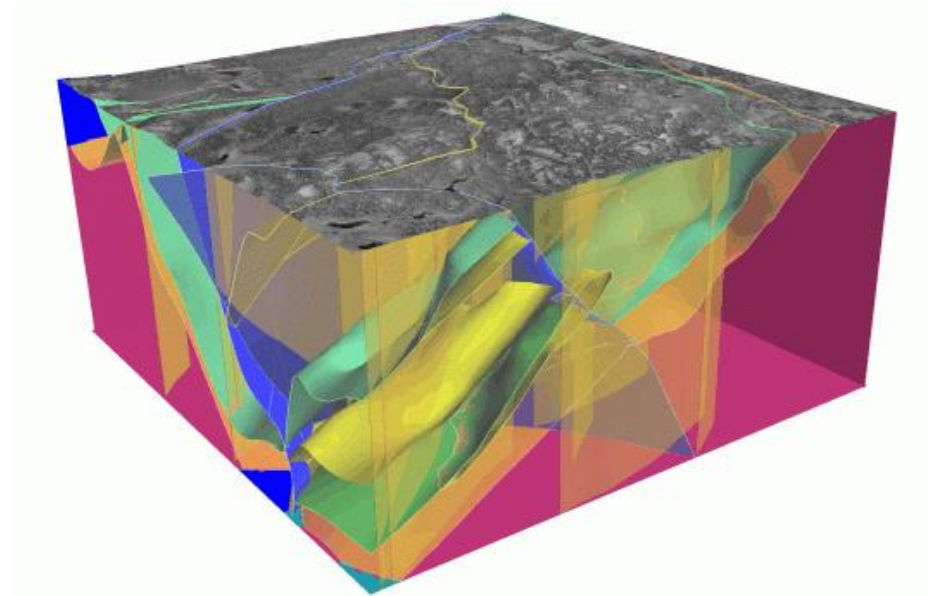
Modélisation de réservoirs



Un réservoir pétrolier

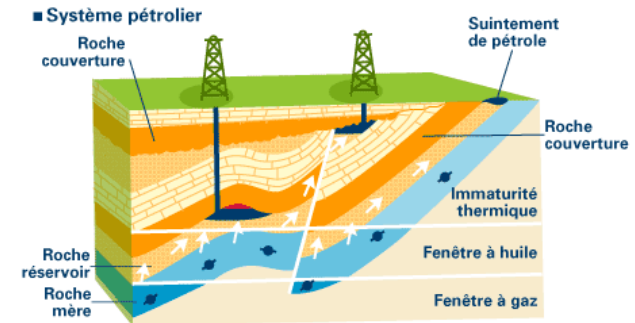


Coupes sismique 3D



Etapes principales d'un projet de simulation de réservoir

- **Modélisation géologique 3D**
 - Exploitation des *données de puits*
 - Exploitation de la *sismique 3D*
 - *Données de production* existantes
- **Simulation** : maillage (*éléments finis*) et exploitation des équations d'écoulement hydraulique dans le modèle supposé
- « **History matching** » : mise au point du modèle et de ses paramètres pour s'accorder aux *données existantes*
- **Exploitation directe** selon les *configurations de production envisagées*
- Utilisation de la **méthode de Monte-Carlo** en cas de paramètres mal connus ou aléatoires (*voir plus loin*)
- *Un tel projet est fortement multidisciplinaire et peut s'étendre sur de nombreuses années*





Questions à ...

**Mickaële Le Ravalec-Dupin, Expert
Modélisation en Ingénierie de réservoir à
IFPEN**

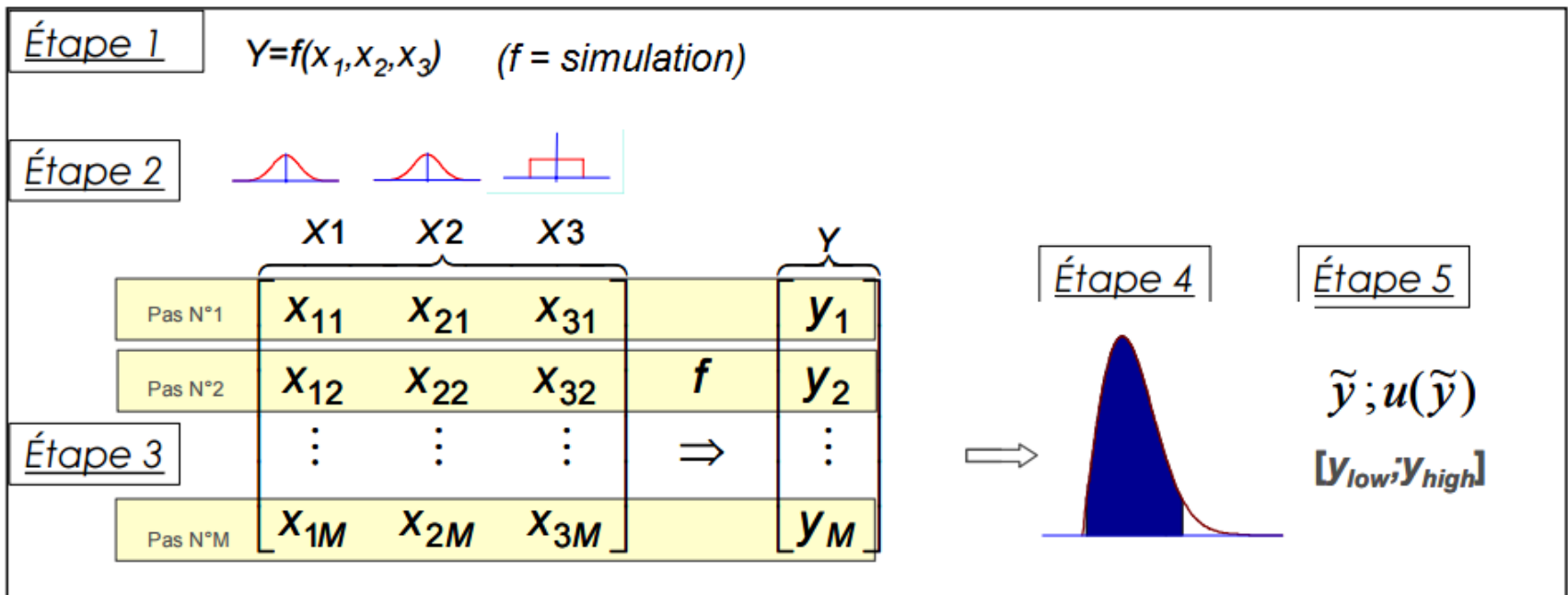
- **À quoi sert la modélisation en exploration-production ?**
- **Mickaële Le Ravalec-Dupin :** Les formations géologiques souterraines sont des milieux très hétérogènes. Pour les représenter, on utilise un modèle géologique. C'est une grille 3D reproduisant **l'hétérogénéité du milieu naturel**, dans laquelle on **simule l'écoulement des fluides**. Cette connaissance peut être utilisée pour placer de nouveaux puits sur un gisement déjà exploité ou pour optimiser les techniques de récupération améliorée du pétrole (EOR), comme l'injection de CO₂ ou de vapeur. L'enjeu est d'augmenter le taux de récupération, sachant que l'on extrait aujourd'hui en moyenne seulement 35 % du pétrole des réservoirs.
- **Sur quelles techniques s'appuie-t-elle ?**
- **M. Le R.-D. :** Toute la difficulté consiste à construire un modèle aussi représentatif que possible de la réalité. Pour cela, nous nous appuyons sur notre connaissance des milieux sédimentaires et l'ensemble des données collectées sur le terrain : mesures sur carottes extraites des puits, mesures sismiques, mesures de pression, de débit, etc.
- À partir d'une **représentation initiale de la formation géologique**, nous **modélisons la récupération des hydrocarbures et prédisons les profils de production**. Ces informations sont une aide précieuse à la prise de décision des compagnies exploitantes. Toutefois, la connaissance du sous-sol étant imparfaite, les prédictions sont entachées d'incertitudes. Des techniques spécifiques permettent de rendre les modèles plus prédictifs en les contraignant à reproduire les données de production mesurées sur le terrain.
- **Les incertitudes sont également analysées et quantifiées par la mise en œuvre d'approches statistiques.** Ces techniques sont implémentées dans le simulateur de réservoir PumaFlow™ et le logiciel EasySense/Cougar, commercialisés par **Beicip-Franlab**.

Simulations « Monte-Carlo »



- *Contexte probabiliste*
 - Certaines *entrées* ou *paramètres* sont *aléatoires*
 - Leurs *lois de variation* sont supposées connues (*intervalles de variation, lois gaussiennes, de Poisson...*)
 - Les *résultats* sont donc aussi aléatoires ou dispersés
- *On effectue de nombreuses simulations*
 - En faisant *varier* à chaque fois les entrées et/ou les paramètres selon leurs lois respectives (*tirages aléatoires*)
 - On analyse les variations correspondantes des *résultats*
 - *Moyennes, écarts-types, extrema, intervalles de confiance*
 - *Graphiques*

Schéma Monte-Carlo



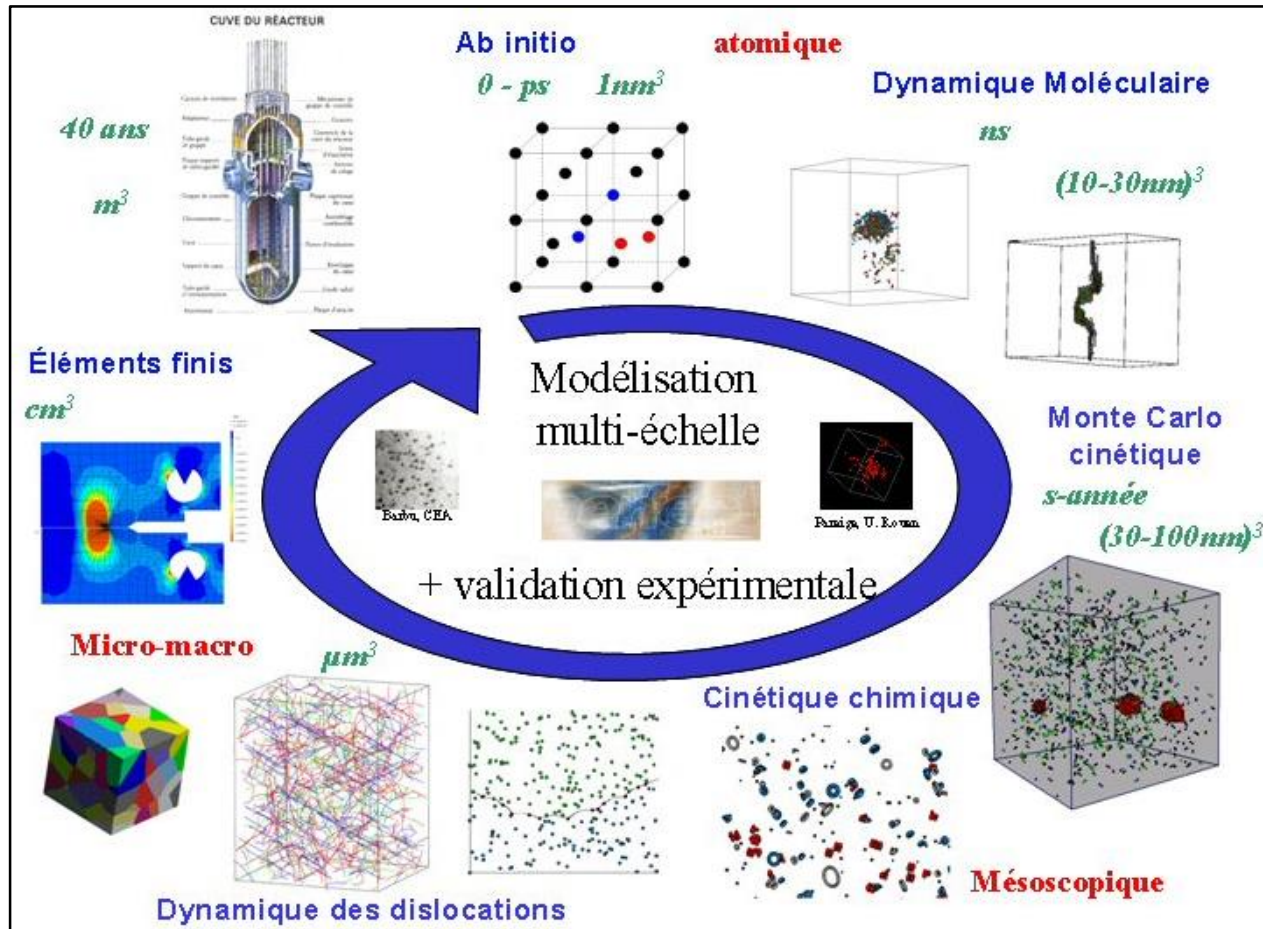
Les différentes étapes de la méthode de Monte Carlo

Modélisation multi-échelles

- Un système peut être étudié à différentes échelles spatiales ou temporelles (*souvent liées*)
 - **Economie** : macro- et micro-
 - **Sociologie** : sociétés, classes, cellules, individus...
 - **Systèmes physiques** : global, local, moléculaire...
- Au lieu d'un modèle général, trop complexe :
 - On simule des sous-modèles **pour chaque échelle**
 - Les résultats globaux (*émergents*) d'un niveau sont communiqués au niveau supérieur
 - En effet les propriétés émergentes d'une échelle relèvent de l'échelle supérieure

Exemple de simulation multi-échelles

Estimation de la durée de vie des centrales nucléaires



Projet intégré européen « **PERFECT** » dans le cadre du **6ème PCRD EURATOM**

Prévoir l'évolution dans le temps des matériaux des réacteurs nucléaires en développant un ensemble d'outils de simulation numérique multi-échelles

Mise en pratique

Projet de modélisation / simulation

Etapes d'une modélisation/simulation

- 1) Enquête, documentation
- 2) Modélisation initiale, hypothèses (*explicitées*)
- 3) Choix d'un outil → Prototype de simulation
- 4) En parallèle, récolte de données expérimentales (pour *l'History Matching*)
- 5) Mise au point initiale sur *History Matching*
- 6) Cas des paramètres inconnus ou aléatoires
- 7) Exploitation (directe, inverse, Monte-Carlo...)
- 8) **Validation, Contrôle qualité, diagnostics**

1) Enquête, documentation

- Relevé des **agents**
- Relevé des **interactions** (logiques, fonctionnelles, diff^{fl}es...)
- Relevé des **paramètres** à ajuster (*d'agents, d'interactions*)
- Faire **schémas, graphes, tableaux**, et les étudier
- Utilisation d'outils de **Data Mining** (*Big Data*) ?
- **Sources** :
 - Observation
 - Réflexion
 - Théories éventuellement applicables
 - Etudes terrain
 - Interviews
 - Lectures (sur le sujet ou problèmes analogues)
- **Outils ?** (*statistiques, graphiques, data mining...*)

2) Modélisation initiale, hypothèses

- Elaborer un modèle programmable
 - Ni trop simpliste (faux), ni trop riche (ambigu, sous-déterminé)
 - Relever par écrit toutes les **hypothèses simplificatrices**
 - Ce relevé sera **indispensable** pour tout retour en arrière sur la modélisation, suite au **contrôle qualité**
- Types d'hypothèses simplificatrices
 - **Suppression d'agents** supposés insignifiants
 - **Regroupement d'agents** semblables → **agents composites**
 - *(Ex: un foyer, une ville, un nuage... sont des agents composites)*
 - Simplification des **caractéristiques des agents**
 - *(Ex: réduire un individu à quelques comportements schématiques)*
 - Simplification ou suppression **d'interactions supposées négligeables** *(sans effet supposé sur le résultat global)*
 - Paramètres mal connus **fixés à des valeurs de compromis...**

3) Choix d'un outil, prototypage

- Types d'outils
 - Outils génériques (*multi-agents, automates cellulaires, éléments finis...*)
 - Outils dédiés à un type d'application (*Mécanique, hydraulique, électronique, processus continus...*)
 - Développement direct *from scratch* (*programmation*)
- Représenter le modèle choisi dans le langage de l'outil (*ou développer le logiciel*)
- Mettre au point sur un jeu d'essai initial
 - Entrées, paramètres...

4) Récolte de données expérimentales

- Relever l'historique du comportement passé réel, observé, de ce système
 - *Date de chaque observation*
 - *Entrées, sorties observées*
 - *Paramètres mesurés ou certains*
 - *Remarques particulières (données manquantes,...)*
- Enregistrer cet historique, et l'enrichir plus tard après chaque exploitation contrôlée

5) Mise au point initiale (History Matching)

- Exploiter le modèle sur les *données historiques*
- *Ajustement des paramètres* pour obtenir un fit
 - Soit *manuellement*, si c'est possible
 - Soit par itérations (*approche inverse*)
- *Toutes les données* historiques doivent être retrouvées avec un *jeu unique* de paramètres
- En cas d'échec, *remettre en cause le modèle !*
 - Revenir à la liste des *hypothèses simplificatrices* effectuées
 - Tester tour à tour la *levée de certaines hypothèses* : étudier la *sensibilité* à ces hypothèses

6) Cas de paramètres inconnus ou aléatoires

- Certains paramètres importants peuvent être mal connus ou variables
 - Exemple: le taux de vapeur d'eau dans les prévisions climatiques
- On peut agir de trois manières principales
 - Retenir une *valeur fixe de compromis*
 - Tenter de les optimiser par une *approche inverse*
 - *Mais cela exige de disposer de très nombreuses observations*
 - Estimer leur variabilité et simuler leurs variations par la méthode de *Monte-Carlo* : on obtient ainsi un domaine probabiliste de variations des résultats
 - *Mais cela exige d'avoir au moins une estimation probabiliste de ces paramètres*

7) Exploitation

- ✓ Une fois réunis tous ces préliminaires, on exploite la simulation à chaque besoin
 - Approche directe : à chaque besoin on introduit les entrées **E** dans le simulateur, et l'on obtient les sorties **S**
 - Approche inverse : Les sorties étant connues (« observables »), on estime les entrées inconnues par optimisation
 - *Optimisation bayésienne* : si l'on introduit *des probabilités a priori* sur les entrées
 - Méthode de Monte-Carlo : on probabilise certaines entrées et/ou paramètres, et on exploite les différentes combinaisons
 - Les résultats sont délivrés en termes probabilistes eux aussi : *moyennes, intervalles de confiance, extrema...*
- ✓ On peut aussi modifier certains éléments du modèle
 - Test d'hypothèses en adaptant le modèle à certaines *modifications envisagées du système réel*

8) Validation, Contrôle Qualité, Diagnostics

- Un modèle, même bien réglé, peut parfois conduire à des prévisions **erronées** ou **fantaisistes**
 - *Voir pourquoi slides suivantes*
- Un « bon modèle » doit être au minimum :
 - **Stable dans le temps**
 - Il ne faut pas avoir à réajuster les paramètres après chaque exploitation (*si infirmée par l'observation*)
 - Un modèle **instable** est **toujours faux**, même s'il est réglé pour reproduire toutes les données passées
 - **Peu sensible aux petites erreurs** sur les données ou paramètres
 - Un modèle **trop sensible** est suspect de donner des prévisions erronées
 - Il existe des **tests de sensibilité** (*voir plus loin*)

Validité et précision

Les prévisions et extrapolations d'une simulation sont-elles fiables ?

Sources possibles d'erreurs

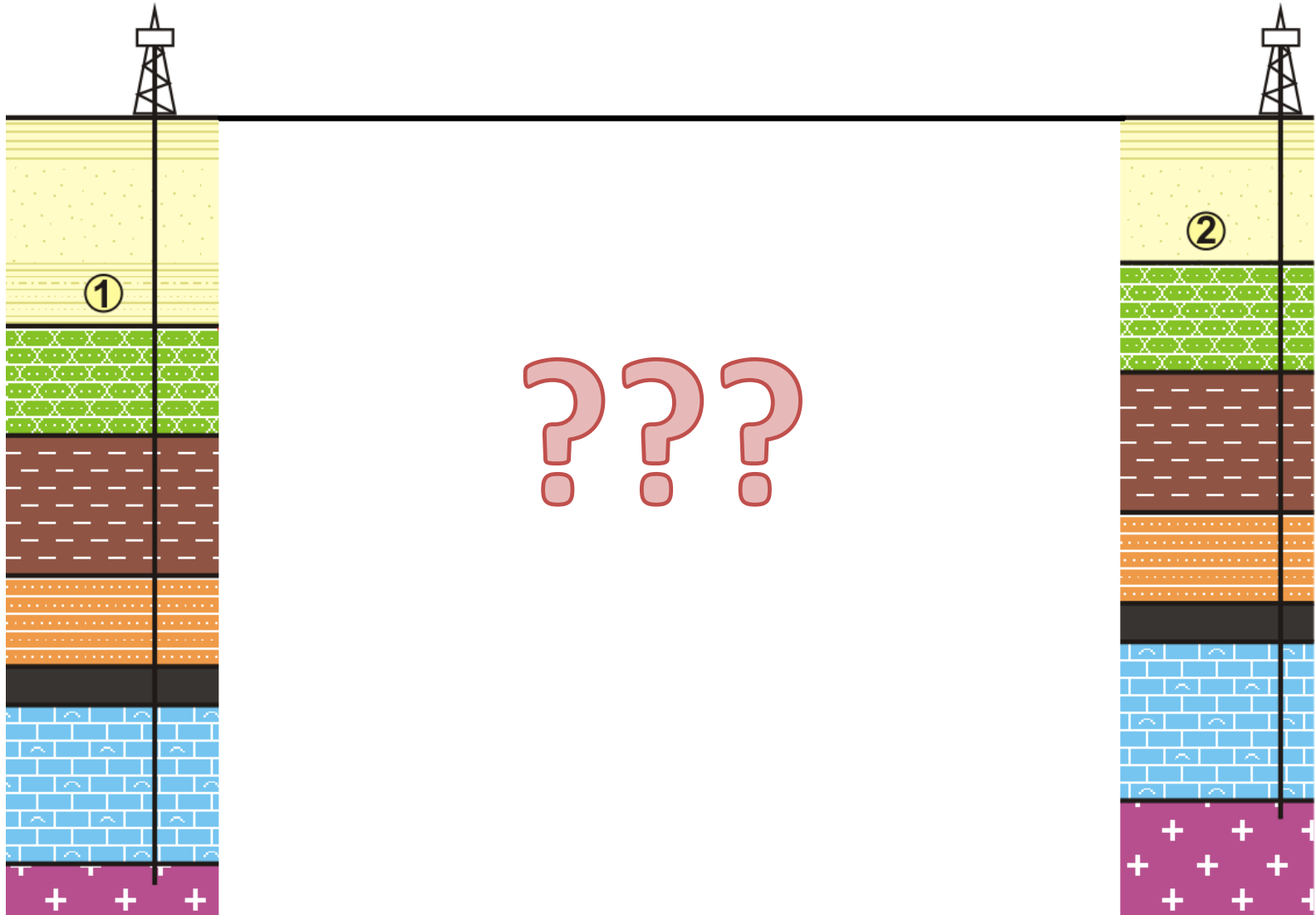
- Modèle qualitatif erroné
 - Souvent trop *simpliste*
 - Attention à *l'illusion du rasoir d'Occam* ! (*la réalité peut être plus complexe que l'apparence*)
 - Ex : Comportements trop simplifiés de certains agents (*humains...*)
 - Modèle météorologique sans certains accidents de terrain, etc.
- Modèle ayant trop de paramètres indépendants
 - Par exemple modèle trop riche, avec trop de *paramètres indépendants*
 - *On peut lui faire dire n'importe quoi* (voir plus loin, la *sous-détermination*)
- Valeurs de certains paramètres inconnues ou inexactes
 - Paramètres variables supposés constants
 - Données d'entrée ignorées, supposées constantes
- Hypersensibilité du modèle à certaines données
 - « *Signaux faibles* » mal pris en compte
- Erreurs sur les interactions
 - Equations appliquées hors de leur *domaine de validité*
 - Interactions faibles *ignorées mais significatives*

Risques de la sous-détermination

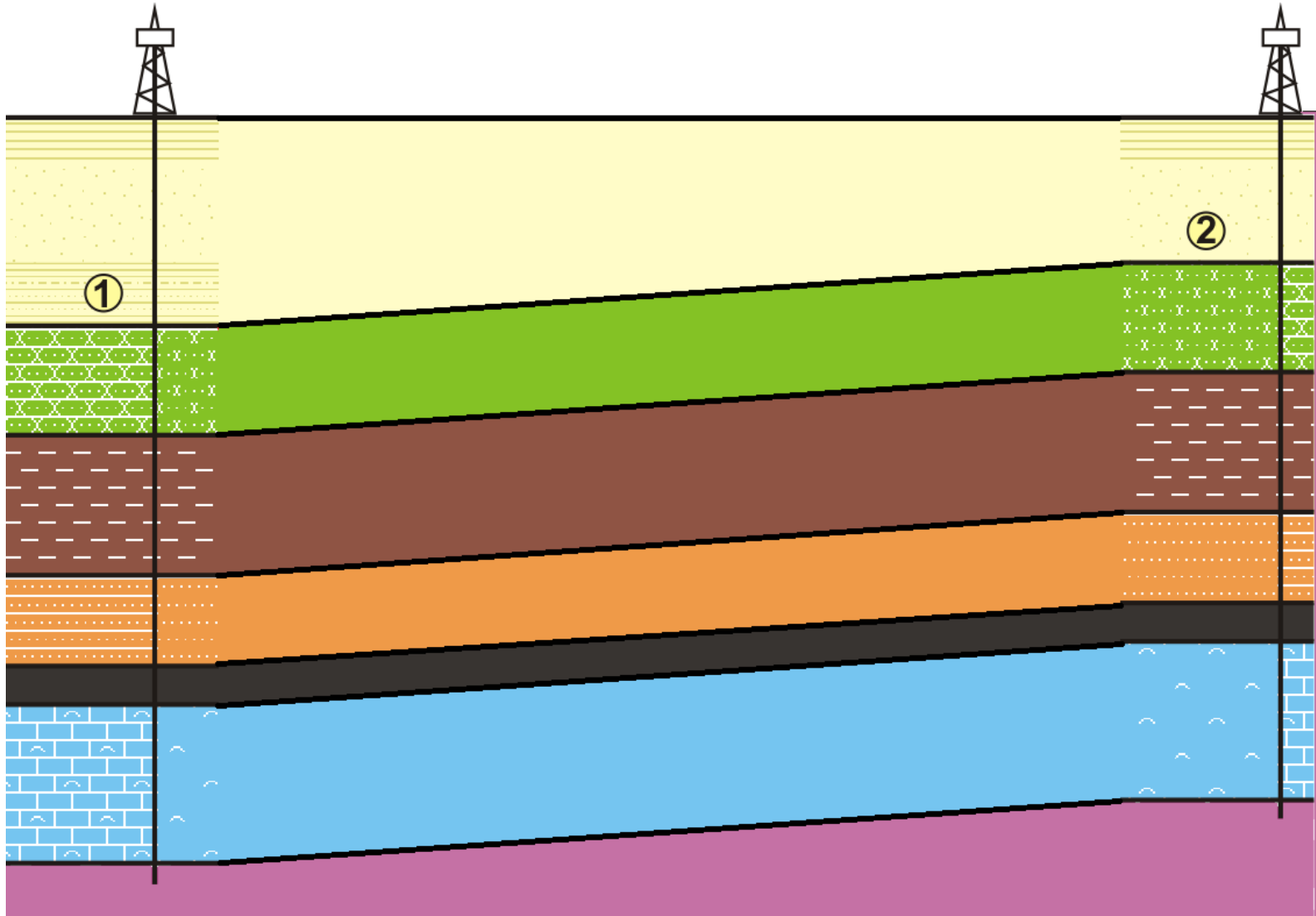
Exemples

Théorie

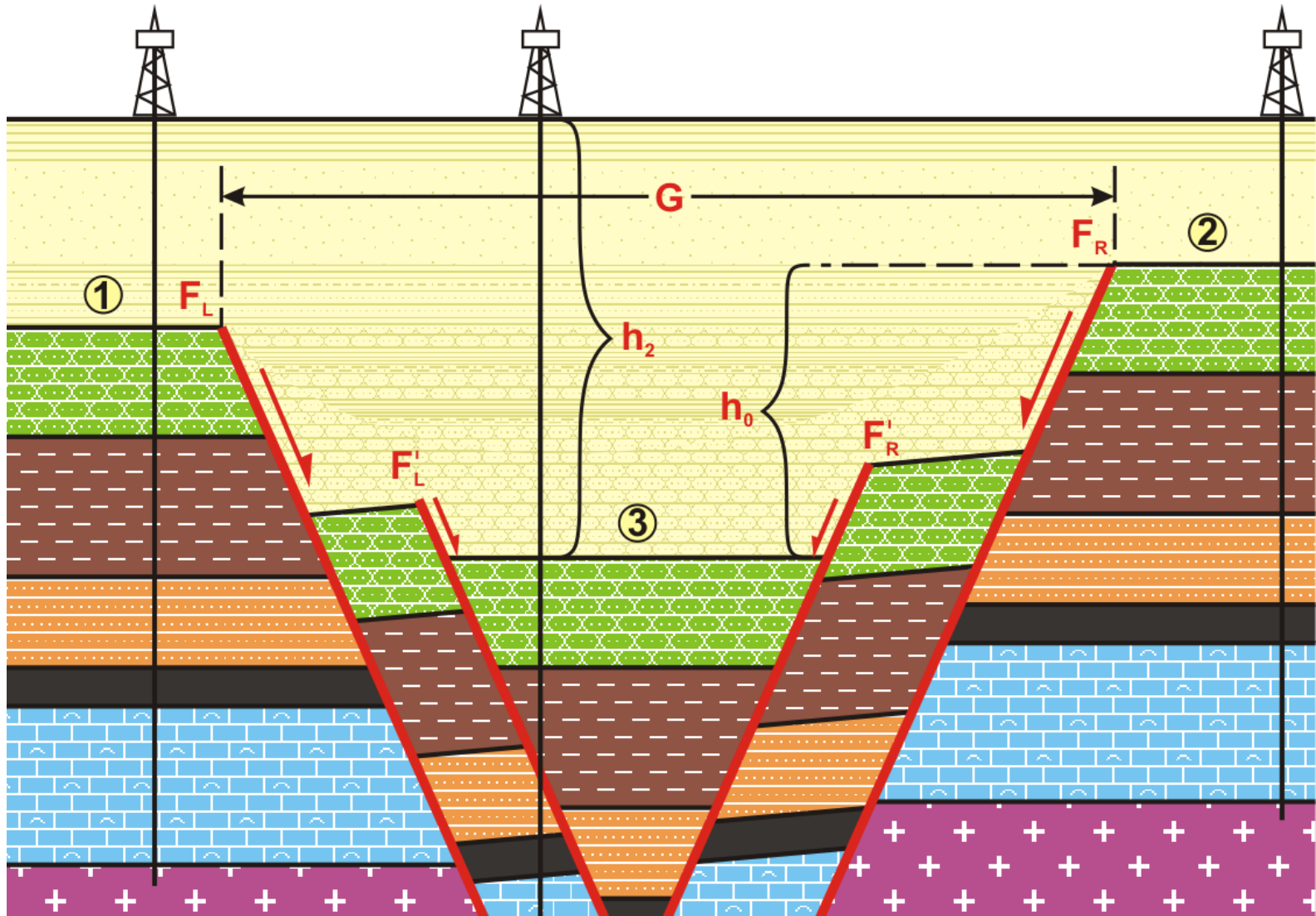
Formation géologique explorée par 2 puits



Hypothèse la plus raisonnable (*Occam*)



La réalité du terrain !



La sous-détermination

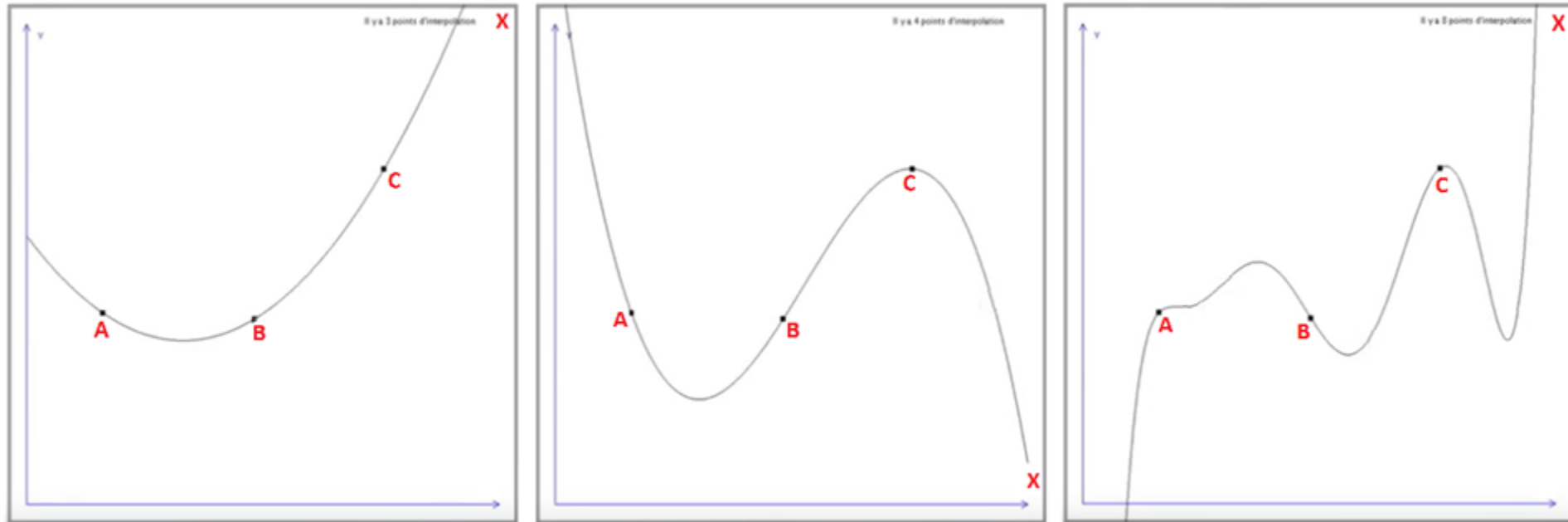
- Définition : plus d'inconnues que d'éléments connus
- Éléments connus
 - Les *entrées mesurées* **E** d'une nouvelle simulation
 - Toutes les *données mesurées* ou observées (*Historique*)
 - Tous les *paramètres* dont on est scientifiquement certain
 - Toutes les *contraintes connues* diminuant les degrés de liberté entre les inconnues (*données ou paramètres*)
- Inconnues
 - Toutes grandeurs (par exemple, les sorties **S**) que l'on veut *calculer ou estimer*
 - Tous les *paramètres indépendants* que l'on veut *ajuster* par History Matching
- Mais ce décompte peut être très difficile !
 - Contraintes floues, approximations, estimations probabilistes, inégalités...
- Cause fréquente de sous-détermination
 - Modèle très riche → trop de paramètres inconnus et indépendants

Effet de la sous-détermination

- Un système sous-déterminé peut être ajusté pour « *fitter* » *n'importe quelles données*
- Les résultats extrapolés peuvent être *totalelement erronés*
- Lorsqu'il y a beaucoup de paramètres, il est *nécessaire* d'estimer si on est dans un contexte sous-déterminé !
- Dans les cas où le calcul est difficile, seule une *étude de sensibilité* permet de détecter une éventuelle sous-détermination

Exemple : sous-détermination d'une interpolation polynômiale

Interpolation polynomiale - Extrapolation



- **A, B, C** : *Données* de l'History Matching
- **X** : *Inconnue calculée (extrapolée)* par le modèle
- *Exemples d'ajustements* du modèle par 3, 4, ou 8 *paramètres (coefficients)*
- *Hypersensibilité* de l'inconnue aux données et paramètres
 - Des variations *infimes* peuvent donner des résultats *divergents*

Critères de qualité d'un modèle

- Reproduction des données connues (*History Matching*)
- Retrouver les résultats établis par la théorie
- Stabilité dans le temps
 - Il n'est pas nécessaire de *réajuster les paramètres* après chaque nouvelle simulation *confrontée à une observation*
- Stabilité par rapport aux incertitudes sur les données
 - *Faible sensibilité* : les résultats doivent *peu varier* lorsqu'on fait varier les *entrées* ou les *paramètres* à l'intérieur de leurs *domaines d'incertitude*
- Comparaison de modèles
 - *Si deux modèles se contredisent, l'un au moins est faux*

Exemple : incertitudes sur le climat

- Selon le climatologue Hervé Le Treut
 - *Membre du GIEC, de l'Académie des sciences, etc.*
- « Les *divergences quantitatives* posent encore un problème pour la gestion [des] risques. La même fourchette d'*incertitude de 1.5 à 4.5°C* de réchauffement global pour un doublement du CO2 reste d'actualité, comme le montre une comparaison récente des *modèles disponibles*. »

Les causes d'incertitudes selon Le Treut

- « On peut distinguer trois familles de problèmes :
 - La première est *intrinsèque* au système climatique lui-même: il n'est tout simplement pas un système *entièrement prévisible*.
 - Une deuxième source d'incertitudes correspond aux *simplifications inévitables* apportées dans la construction des modèles. La difficulté de la représentation des *nuages* en est un exemple.
 - Un troisième facteur vient *limiter la portée pratique* des modèles : en dépit des progrès affichés dans ce domaine, ils ne représentent toujours *qu'une partie du système* climatique complet. »
- *NB : On retrouve ici une partie des « sources possibles d'erreurs » listées précédemment*

Exploitation d'un modèle

- *La bonne exploitation d'un modèle* passe *toujours* par de *multiples simulations*
- *Elles sont nécessaires pour :*
 - *Mise au point du modèle* par ajustements
 - Tests de *fiabilité* et de *sensibilité*
- *Et en fonction des objectifs recherchés :*
 - Test de diverses *hypothèses d'exploitation*
 - *Approche inverse* (boucle d'optimisation)
 - *Monte-Carlo* (traitement des aléas)

Conclusion (« leçons tirées... »)

- La modélisation / simulation est *l'outil par excellence* pour *l'étude et la prévision* des systèmes complexes
- Une modélisation / simulation est un projet *vaste et ambitieux*
 - Plusieurs hommes-années en général
 - Equipes pluridisciplinaires (*avec obtention d'un consensus*)
 - Coût élevé
- Chaque nouvelle modélisation est un *projet original*
 - Faible réutilisabilité des études existantes et des approches précédentes
 - Chaque système doit être étudié *avec un œil neuf* et comporte une forte dose *d'innovation*
- Il existe de nombreuses *stratégies d'exploitation* d'une simulation
 - *Test d'hypothèses, approche inverse, Monte-Carlo, Multi-échelles...*
- Il est important de bien *valider le modèle (fidélité, stabilité)* afin de pouvoir *faire confiance* à ses prédictions futures
 - Un modèle *instable* peut être *faux* ou *sous-déterminé* et *doit être retravaillé*
 - Pour cela de *multiples exploitations* de la simulation sont nécessaires

MERCI !...