



---

# Quantification des incertitudes dans la chaîne d'alerte sismique

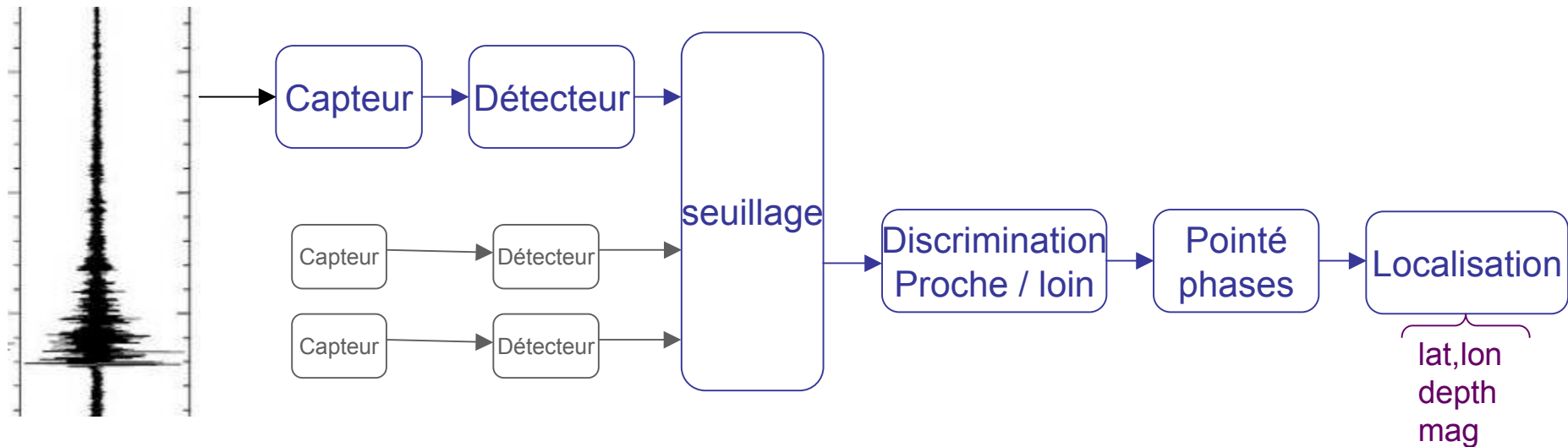
Stéphane BRUNO

*IMPEC - 13 Octobre 2008*

# Mission du laboratoire



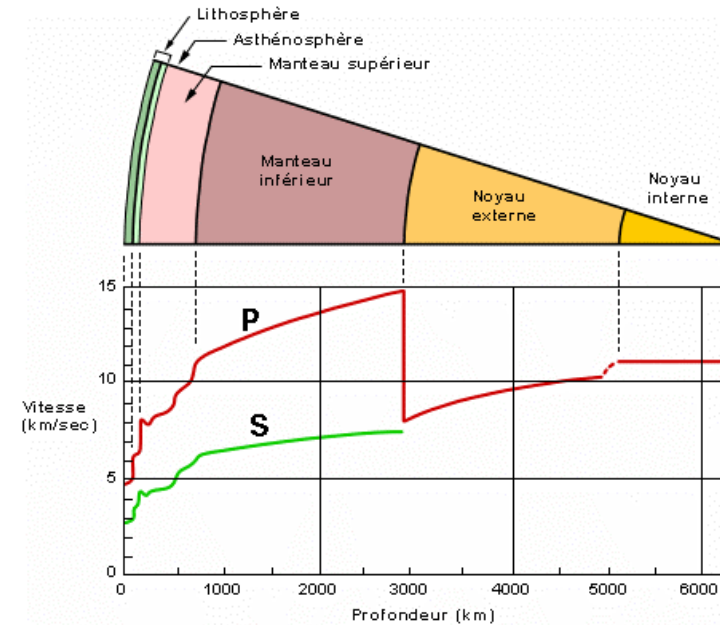
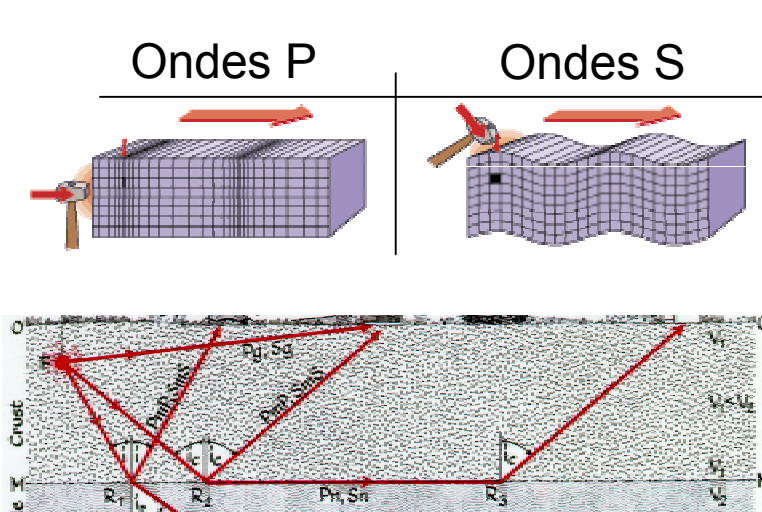
- LDSO : Laboratoire de détection sismique opérationnelle
  - Surveillance permanente des événements sismiques
  - Fournir le plus vite possible une information circonstanciée
- Réseau de stations sismiques et chaîne de traitement associée



# La chaîne de traitement sismique

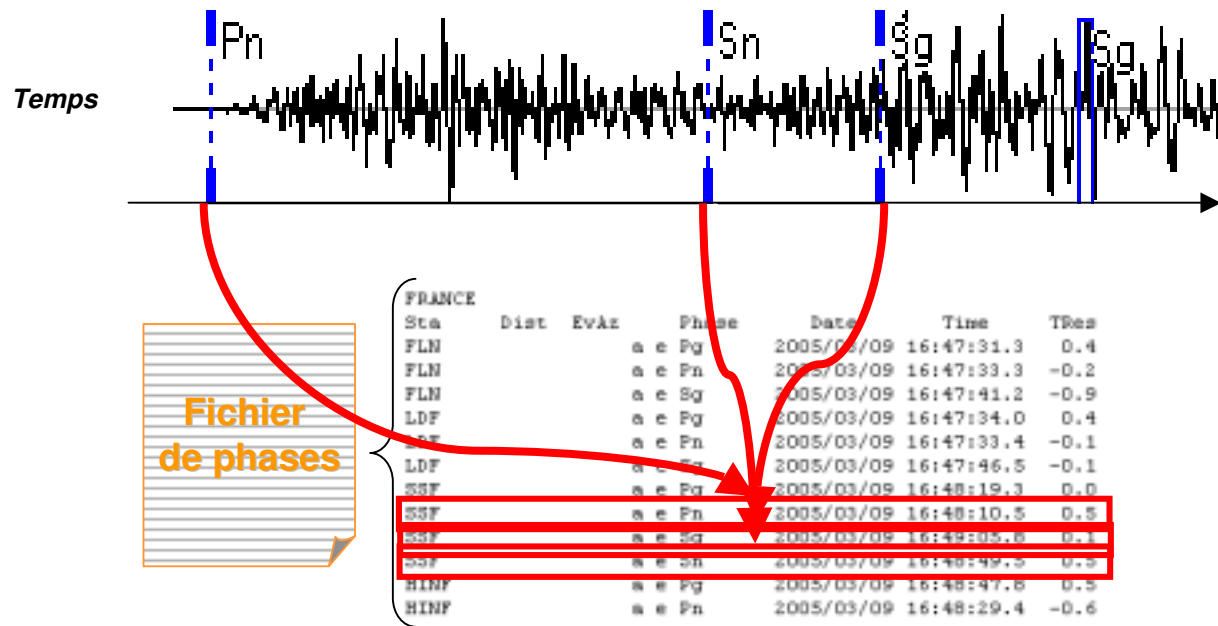


- Exploite la différence de vitesse des différentes ondes



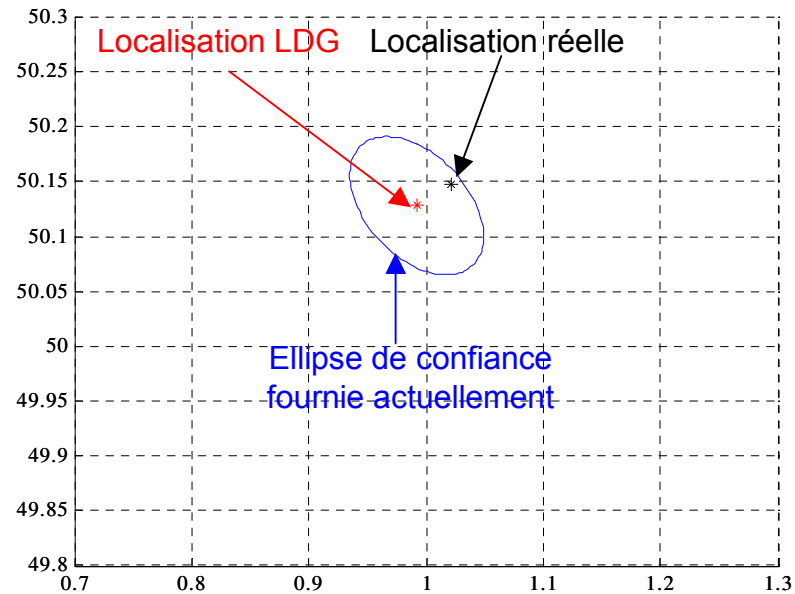


On *pointe* les différentes arrivées d'onde



- Puis il faut résoudre le système d'équation :
  - $\text{Heure\_arrivée} = f \{ (\text{lat}, \text{lon}, \text{prof}, \text{time}), \text{vitesse} / \text{type\_onde} / \text{couche} \}$

## Résultat actuel...



- Ellipse de confiance définie par l'incohérence des temps d'arrivées des phases sismiques
- Explosions de mines en Manche, réalisées par la marine
  - 18% dans l'ellipse à 95%

# D'où vient l'erreur de localisation ?



Fichier de phases

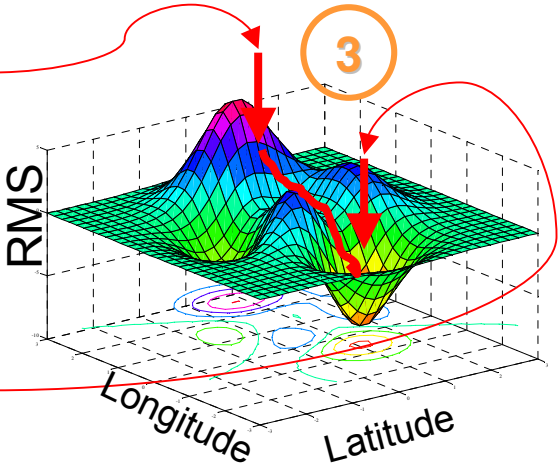
FRANCE	Sta	Dist	EvAz	Phase	Date	Time	TRes
	FLN			a e Pg	2005/03/09	16:47:31.3	0.4
	FLN			a e Pn	2005/03/09	16:47:33.3	-0.2
	FLN			a e Sg	2005/03/09	16:47:41.2	-0.9
	LDF			a e Pg	2005/03/09	16:47:34.0	0.4
	LDF			a e Pn	2005/03/09	16:47:33.4	-0.1
	LDF			a e Sg	2005/03/09	16:47:46.5	-0.1
	SSF			a e Pg	2005/03/09	16:48:19.3	0.0
	SSF			a e Pn	2005/03/09	16:48:10.5	0.5
	SSF			a e Sg	2005/03/09	16:49:05.8	0.1
	SSF			a e Sn	2005/03/09	16:48:49.5	0.5
	HINF			a e Pg	2005/03/09	16:48:47.8	0.5
	HINF			a e Pn	2005/03/09	16:48:29.4	-0.6

1

Prélocalisation

Modèles de terrain

Localisation



```

3          !number of sub-layers
18.       ! limit distance between
3         ! numero of the refractor
3         ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

2 Résultats : latitude, longitude, profondeur  
GAP, RMS



---

## Question

# Comment obtenir une information sur la dispersion possible de la localisation

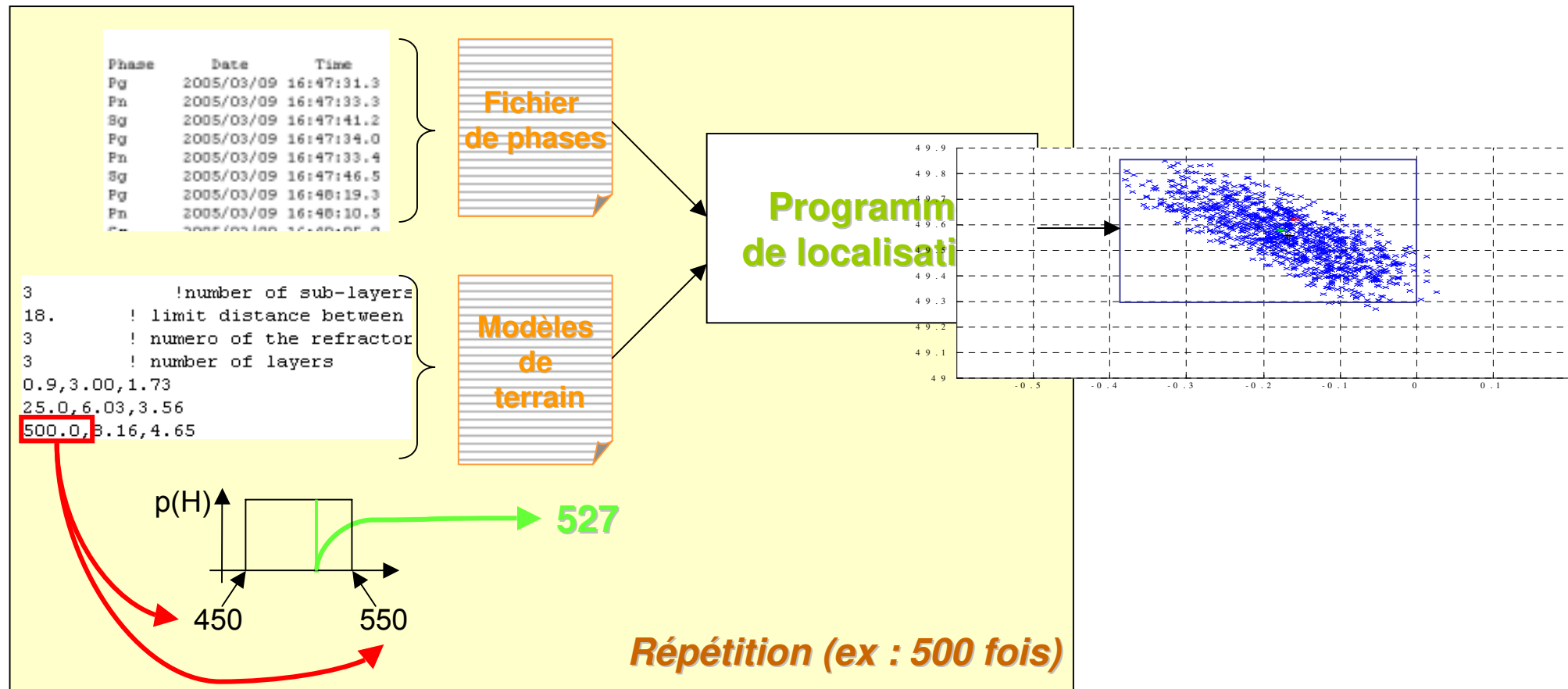
*Méthode combinatoire*

Simuler toutes les combinaisons de paramètres

# Première formulation du problème



- Comment évolue la localisation selon l'erreur possible sur chacun des paramètres utilisés par le modèle
  - Exemple : déviation de la latitude selon l'incertitude sur la profondeur d'une couche géologique



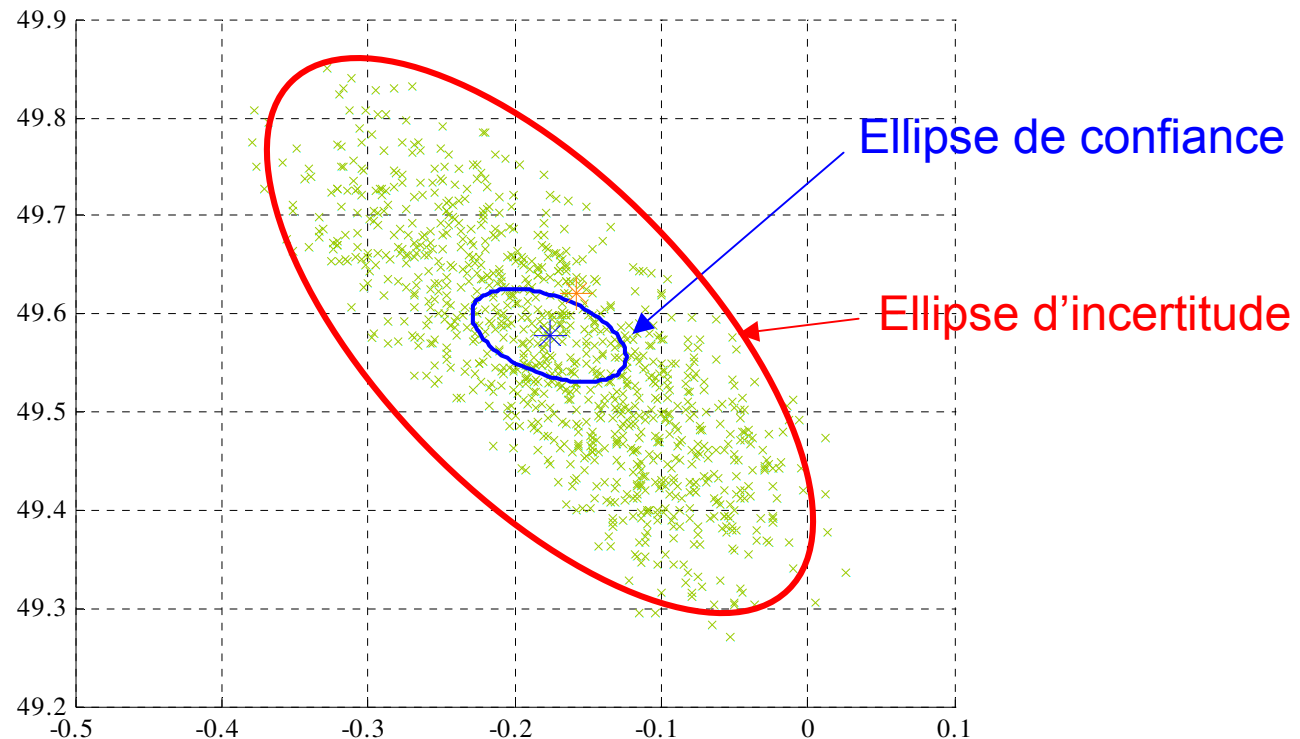


# Où peut se situer la localisation réelle ?

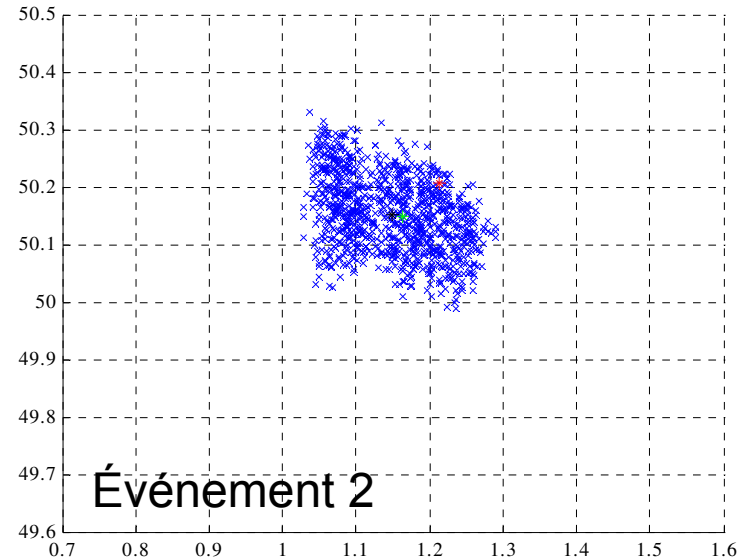
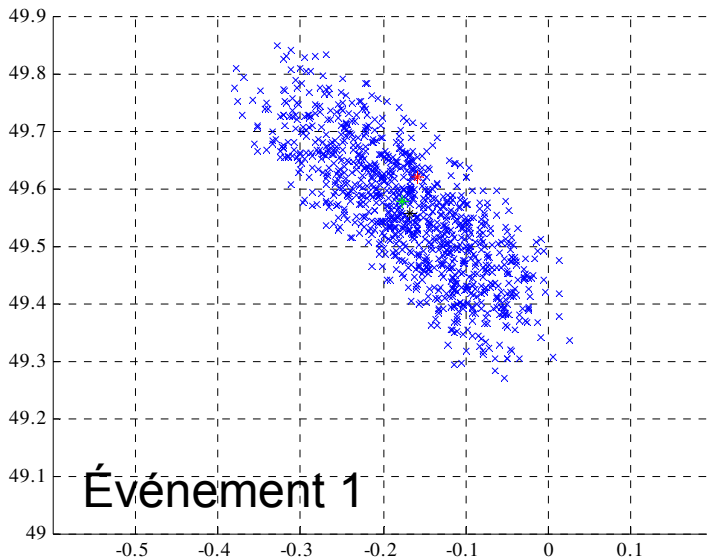


Dispersion des localisations en considérant une erreur

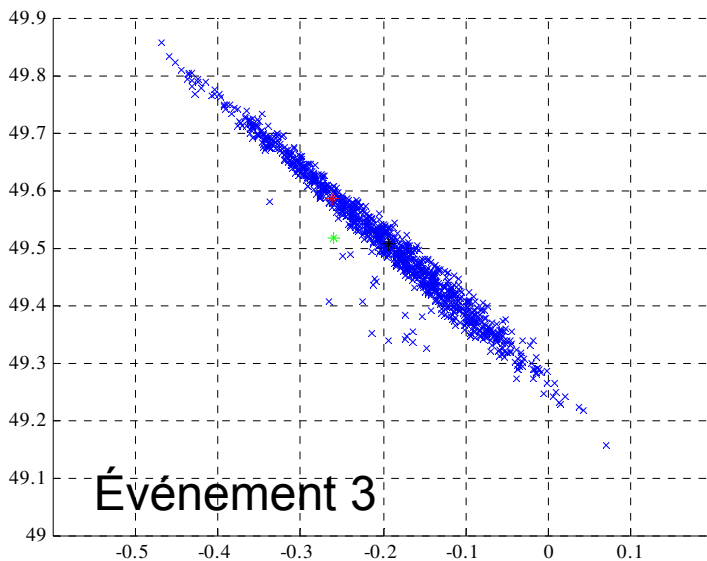
- De +/- 0.5s sur les pointés de phases
- De +/- 10% sur les modèles de terrain



# Méthode directe : Exploration des combinaisons de paramètre



*Perturbation des fichiers d'arrivées et des modèles de terrain par une valeur tirée sans un intervalle défini selon une loi uniforme (0.5s temps, 10% modèles)*



## Les informations apportées

- Définit une zone de localisation certaine de l'événement
- Apporte une confiance à la localisation proposée
- Calcul direct trop coûteux / opérationnel



---

## Deuxième méthode

Estimation de l'ellipse d'incertitude à partir des paramètres de données associés à l'événement

### *Méthode combinatoire*

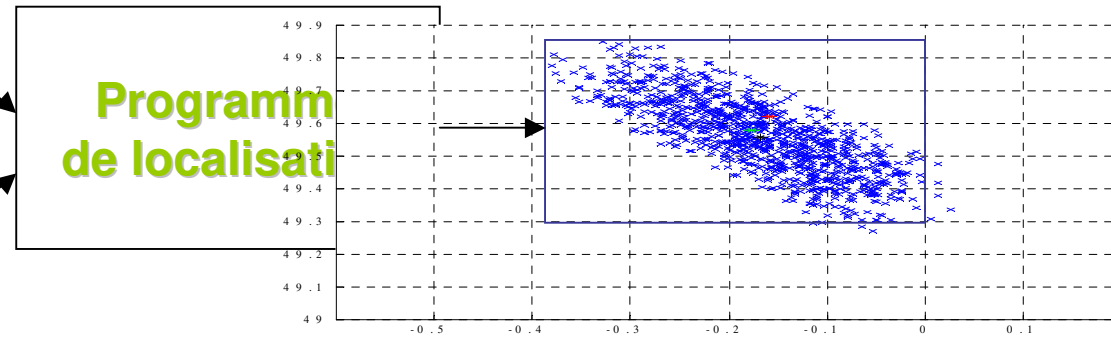
Évaluer l'interaction entrée – sortie pour chacun des paramètres et calculer uniquement les cas conduisant aux plus fortes déviations

# Relation latitude vs modèle



```

3          !number of sub-layers
18.        ! limit distance between
3          ! numero of the refractor
3          ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

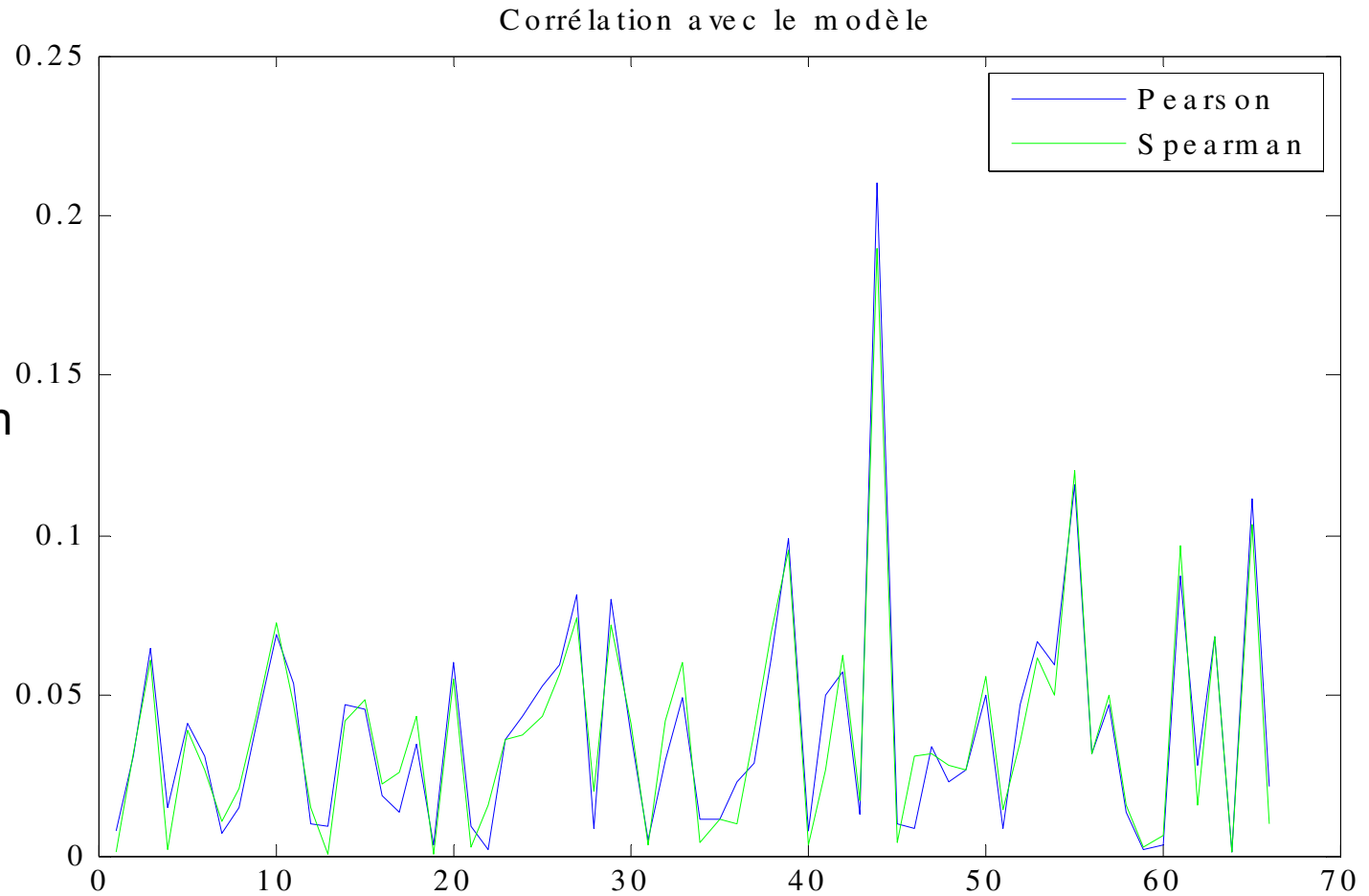


Sta	Dist	EvAz	Phase	Date	Time	TRes
FRANCE						
FLN			a e Pg	2005/03/09	16:47:31.3	0.4
FLN			a e Pn	2005/03/09	16:47:33.3	-0.2
FLN			a e Sg	2005/03/09	16:47:41.2	-0.9
LDF			a e Pg	2005/03/09	16:47:34.0	0.4
LDF			a e Pn	2005/03/09	16:47:33.4	-0.1
LDF			a e Sg	2005/03/09	16:47:46.5	-0.1
SSF			a e Pg	2005/03/09	16:48:19.3	0.0
SSF			a e Pn	2005/03/09	16:48:10.5	0.5
SSF			a e Sg	2005/03/09	16:48:05.8	0.1
SSF			a e Sn	2005/03/09	16:48:49.5	0.5
HINF			a e Pg	2005/03/09	16:48:47.8	0.5
HINF			a e Pn	2005/03/09	16:48:29.4	-0.6

# Pas de corrélation entre la valeur du modèle et la latitude



Coefficient  
de corrélation

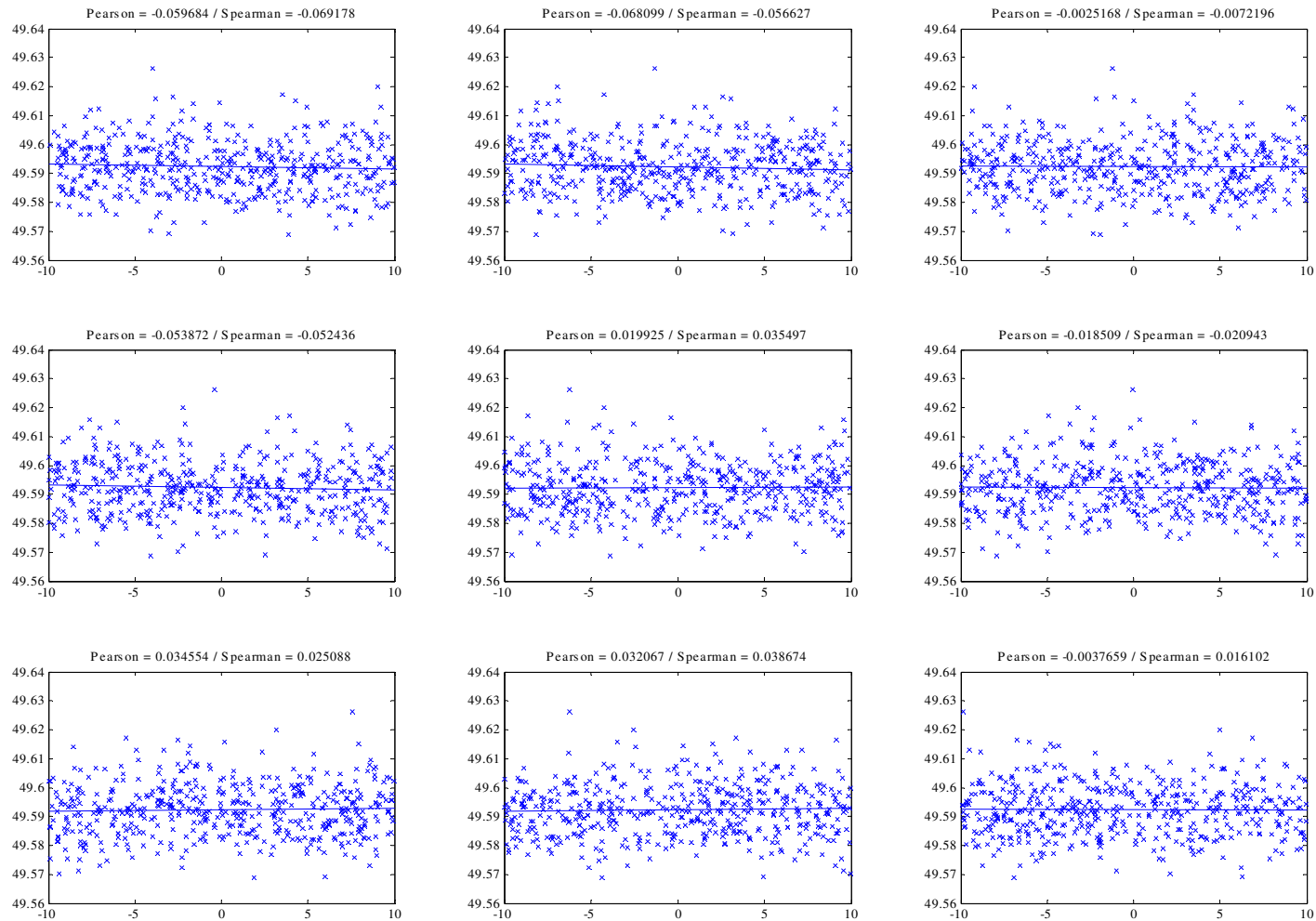


```
3          !number of sub-layers
18.        ! limit distance between
3          ! numero of the refractor
3          ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
```

# Relation latitude vs modèle



- Pas de dépendance linéaire visible, ou alors elle est noyée dans la dépendance aux autres variables

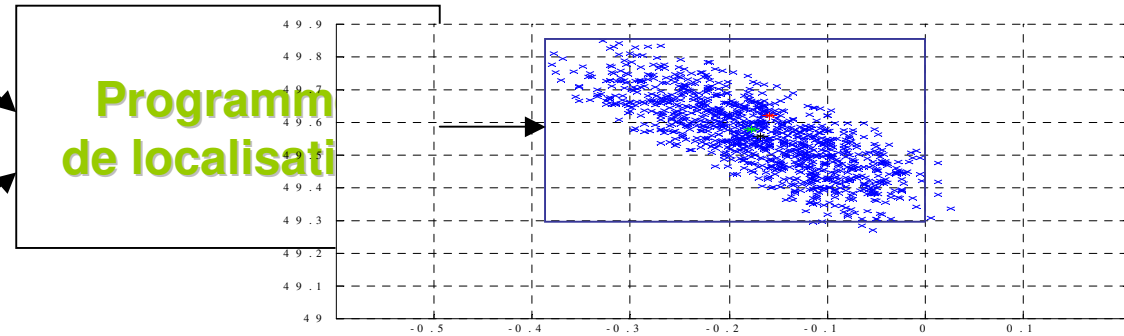


# Relation latitude vs Phase



```

3          !number of sub-layers
18.       ! limit distance between
3        ! numero of the refractor
3        ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

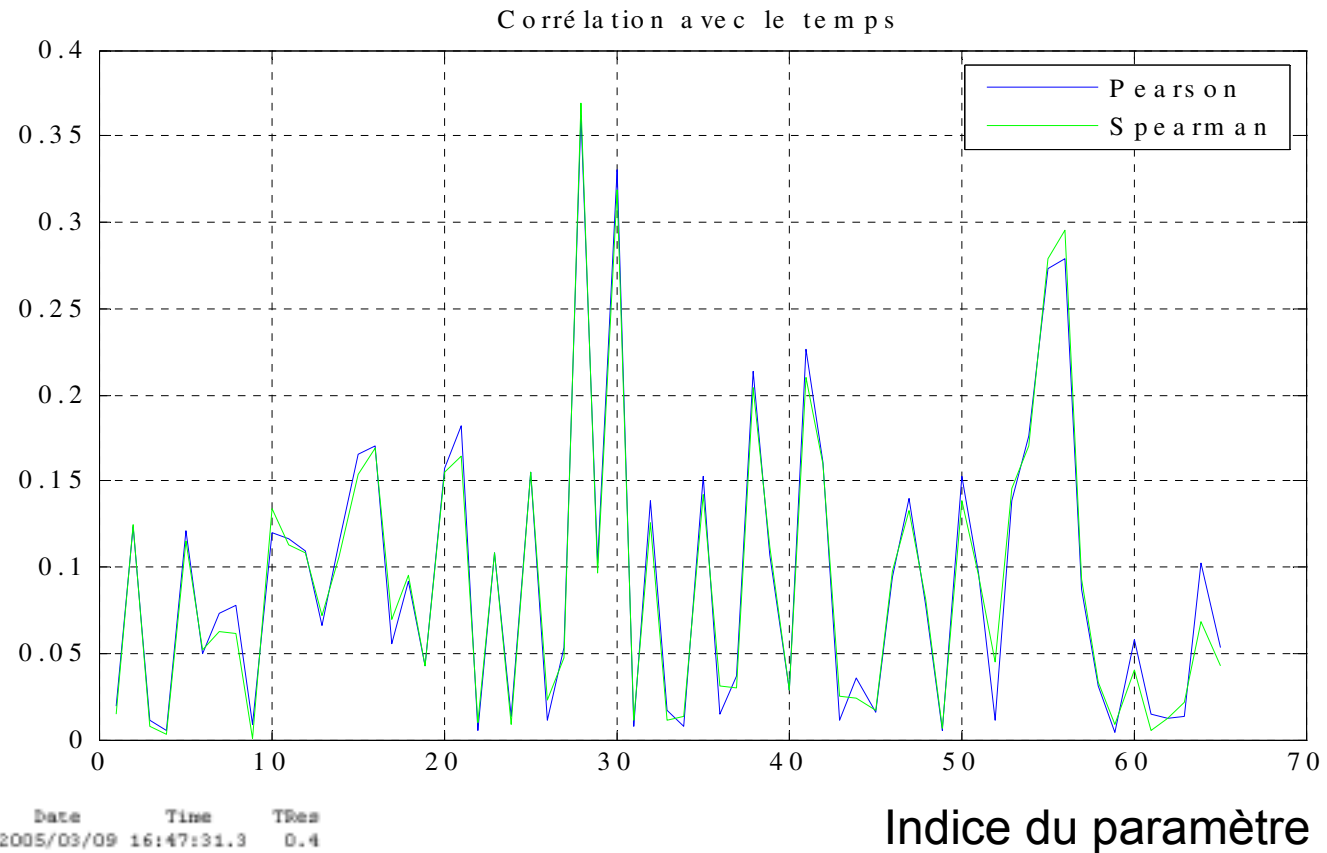


FRANCE	Sta	Dist	EvAz	Phase	Date	Time	TRes
	FLN			a e Pg	2005/03/09	16:47:31.3	0.4
	FLN			a e Pn	2005/03/09	16:47:33.3	-0.2
	FLN			a e Sg	2005/03/09	16:47:41.2	-0.9
	LDF			a e Pg	2005/03/09	16:47:34.0	0.4
	LDF			a e Pn	2005/03/09	16:47:33.4	-0.1
	LDF			a e Sg	2005/03/09	16:47:46.5	-0.1
	SSF			a e Pg	2005/03/09	16:48:19.3	0.0
	SSF			a e Pn	2005/03/09	16:48:10.5	0.5
	SSF			a e Sg	2005/03/09	16:48:05.8	0.1
	SSF			a e Sn	2005/03/09	16:48:49.5	0.5
	HINF			a e Pg	2005/03/09	16:48:47.8	0.5
	HINF			a e Pn	2005/03/09	16:48:29.4	-0.6

# Pas de corrélation entre le pointé temps et la latitude



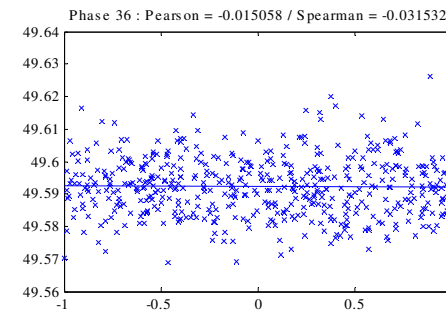
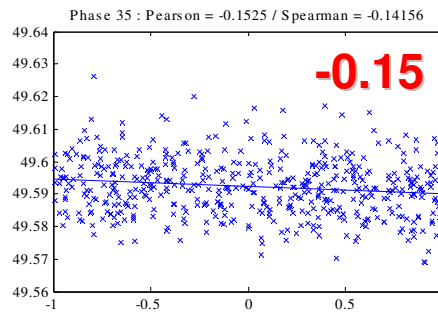
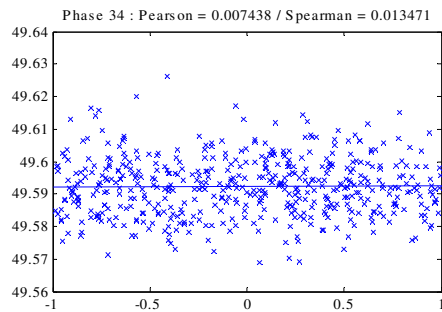
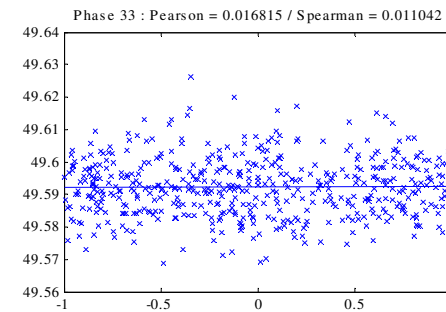
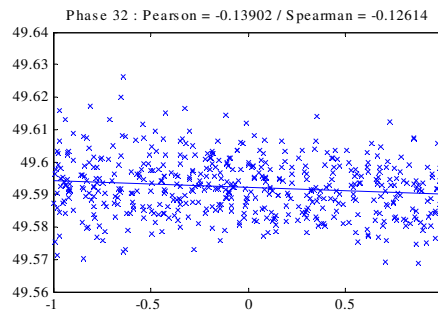
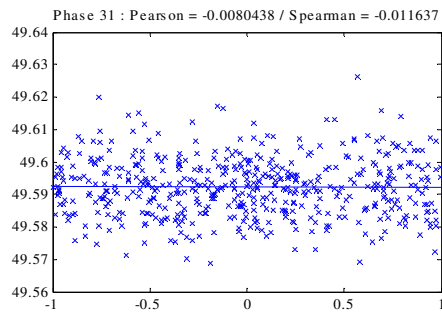
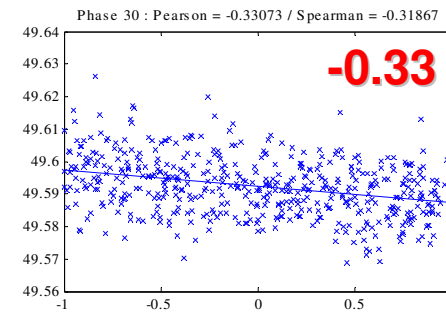
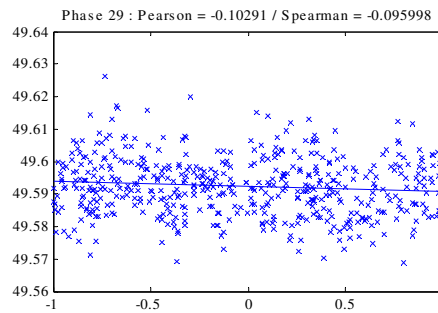
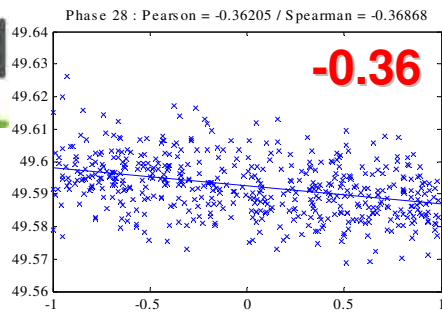
Coefficient de corrélation



FRANCE	Sta	Dist	EvAz	Phase	Date	Time	TRes
	FLN			a e Pg	2005/03/09	16:47:31.3	0.4
	FLN			a e Pn	2005/03/09	16:47:33.3	-0.2
	FLN			a e Sg	2005/03/09	16:47:41.2	-0.9
	LDF			a e Pg	2005/03/09	16:47:34.0	0.4
	LDF			a e Pn	2005/03/09	16:47:33.4	-0.1
	LDF			a e Sg	2005/03/09	16:47:46.5	-0.1
	SSF			a e Pg	2005/03/09	16:48:19.3	0.0
	SSF			a e Pn	2005/03/09	16:48:10.5	0.5
	SSF			a e Sg	2005/03/09	16:49:05.8	0.1
	SSF			a e Sn	2005/03/09	16:48:49.5	0.5
	HINF			a e Pg	2005/03/09	16:48:47.8	0.5
	HINF			a e Pn	2005/03/09	16:48:29.4	-0.6



# Relation latitude vs Phase



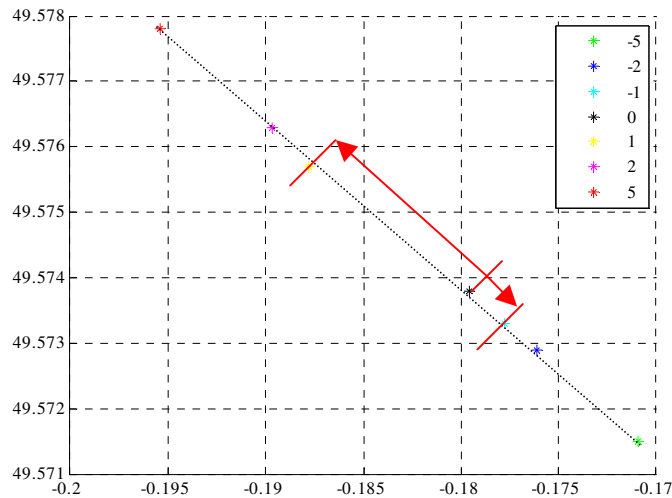
## Décomposition en sous problèmes élémentaires

---

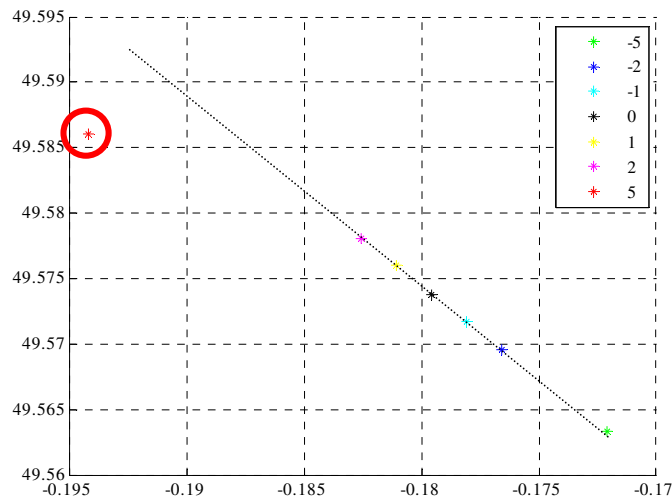


- Certains paramètres ont une influence linéaire non négligeable sur la déviation de la localisation
- Étudions maintenant comment se traduit cette influence autour de la localisation centrale
- Pour calculer facilement une déviation maximale

# Décomposition en sous problèmes élémentaires

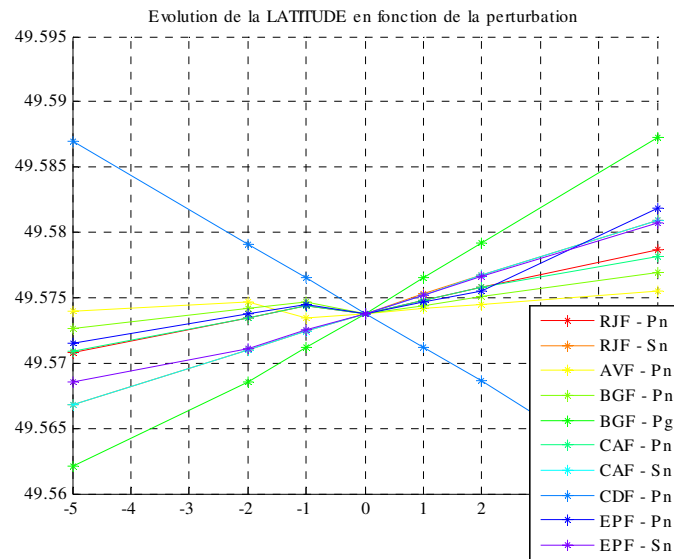


- Symétrie azimutale
- Symétrie imparfaite en distance

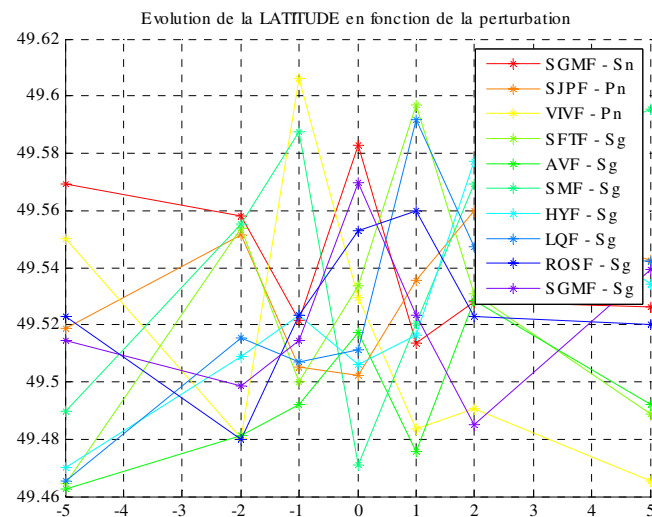


- Linéarité imparfaite
  - Une phase trop perturbée n'est plus considérée par le programme de localisation

# Décomposition en sous problèmes élémentaires



- Déviation prévisible de la latitude en fonction de la perturbation d'entrée

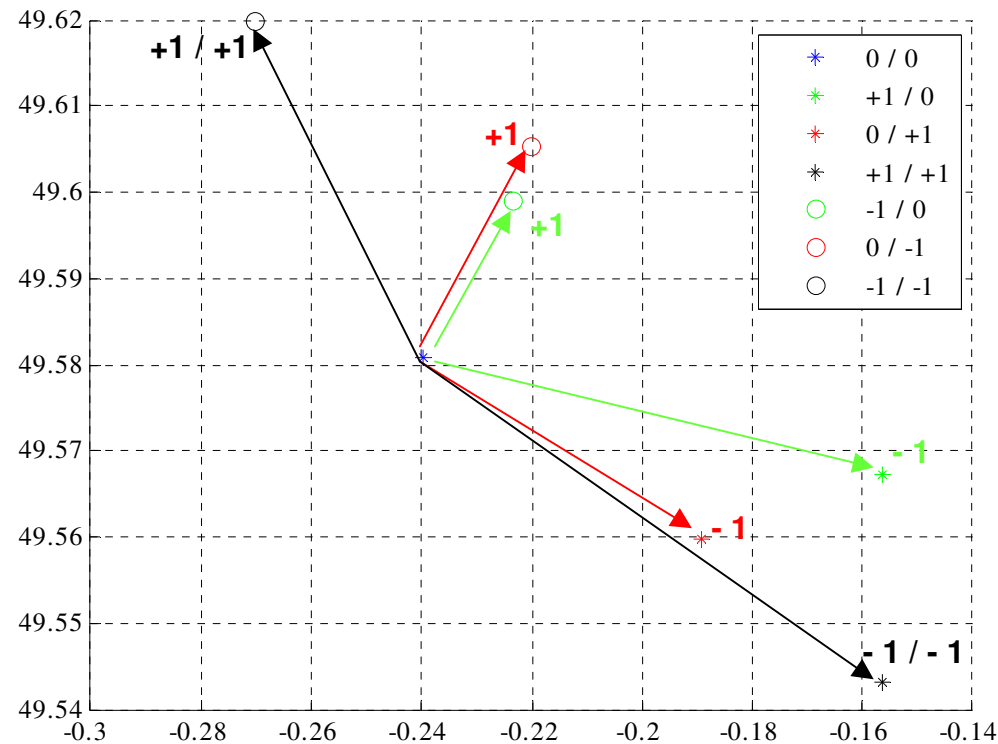


- Pas forcément applicable à tous les événements

# Décomposition en sous problèmes élémentaires



- Non linéarité par rapport aux déviations élémentaires :



- La déviation provoquée par la perturbation de 2 phases n'est pas identifiable à partir des 2 déviations élémentaires.

## Conclusion partielle

---



- Déterminer l'influence entrée-sortie du modèle pour un événement est difficile
  - Caractère non linéaire... et loi non évidente
  
- Même si on sait le faire pour un événement sismique, chaque événement est différent
  - nb d'entrées et influence  $\neq$  pour chaque événement sismique
  
  - Il n'est pas garanti (loin de là !) que l'on puisse déterminer des généralités sur les phases ou paramètres importants
  
  - Idem concernant le mode d'influence (relation entrée-sortie)



*Identifier des points caractéristiques est un problème difficile car chaque événement sismique entraîne un jeu de données différent*

### **Troisième méthode**

Construire un modèle de génération de l'ellipse d'incertitude à partir des paramètres de couverture de l'événement

Méthode AI

# Estimation des quantiles pour différents événements, lois fixes



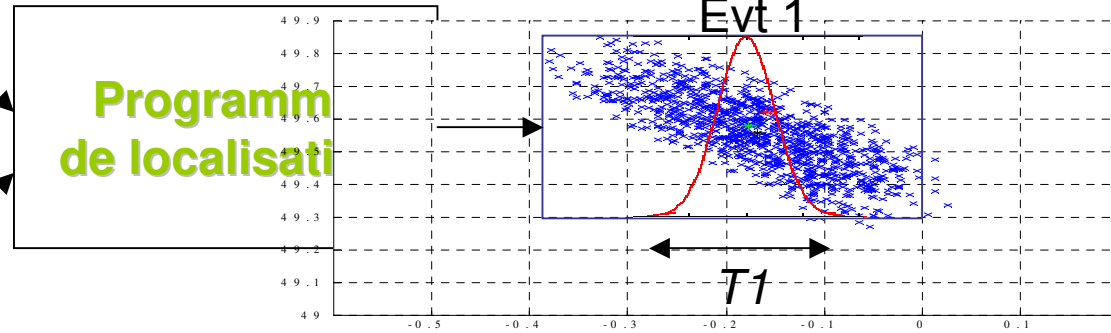
Phase	Date	Time
Pg	2005/03/09	16:47:31.3
Pn	2005/03/09	16:47:33.3
Sg	2005/03/09	16:47:41.2
Pg	2005/03/09	16:47:34.0
Pn	2005/03/09	16:47:33.4
Sg	2005/03/09	16:47:46.5
Pg	2005/03/09	16:48:19.3
Pn	2005/03/09	16:48:10.5
Pn	2005/03/09	16:48:07.8

Evt 1  
Fichier de phases

```

3          !number of sub-layers
18.       ! limit distance between
3        ! numero of the refractor
3        ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

Evt 1  
Modèles de terrain



Phase	Date	Time
Pg	2005/03/09	15:47:02.3
Pn	2005/03/09	15:47:02.3
Sg	2005/03/09	15:47:04.0
Pg	2005/03/09	15:47:02.4
Pn	2005/03/09	15:47:02.3
Sg	2005/03/09	15:47:02.3
Pg	2005/03/09	15:47:02.3
Pn	2005/03/09	15:47:02.3

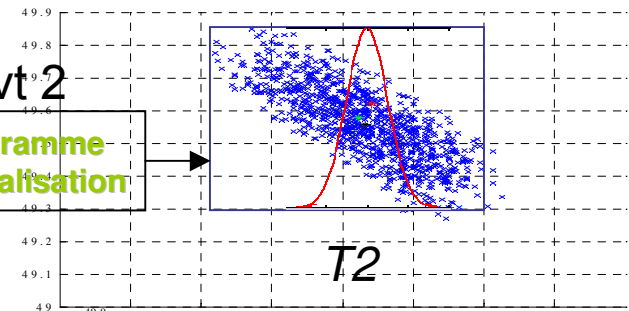
Fichier de phases

```

3          ! number of sub-layers
18.       ! limit distance between
3        ! numero of the refractor
3        ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

Modèles de terrain

Evt 2  
Programme de localisation



Phase	Date	Time
Pg	2005/03/09	15:47:02.3
Pn	2005/03/09	15:47:02.3
Sg	2005/03/09	15:47:04.0
Pg	2005/03/09	15:47:02.4
Pn	2005/03/09	15:47:02.3
Sg	2005/03/09	15:47:02.3
Pg	2005/03/09	15:47:02.3
Pn	2005/03/09	15:47:02.3

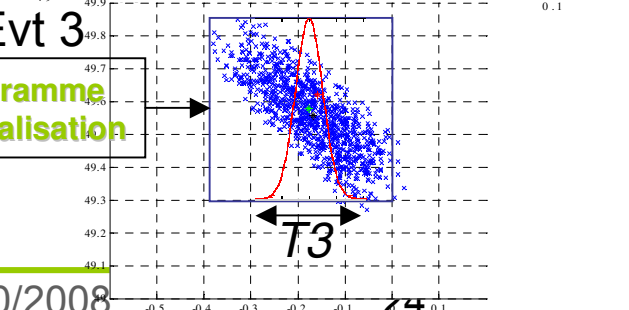
Fichier de phases

```

3          ! number of sub-layers
18.       ! limit distance between
3        ! numero of the refractor
3        ! number of layers
0.9,3.00,1.73
25.0,6.03,3.56
500.0,8.16,4.65
    
```

Modèles de terrain

Evt 3  
Programme de localisation





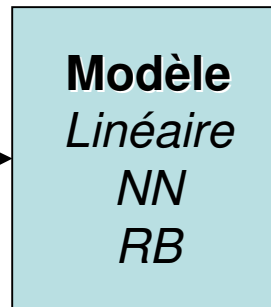
# Changement de point de vue



- On considère la dispersion comme une variable aléatoire fonction des paramètres de couverture

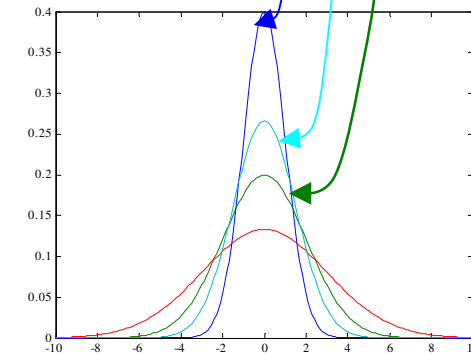
Paramètres de couverture

*Evt1 ( gap<sub>1</sub>, dist<sub>1</sub>, ... nb\_Pg<sub>1</sub> )*  
*Evt1 ( gap<sub>2</sub>, dist<sub>2</sub>, ... nb\_Pg<sub>2</sub> )*  
...  
*Evt1 ( gap<sub>N</sub>, dist<sub>N</sub>, ... nb\_Pg<sub>N</sub> )*



Quantiles (issus de MC)

*Q99<sub>1</sub>*  
*Q99<sub>2</sub>*  
...  
*Q99<sub>N</sub>*



**Paramètres de couverture :**

*Gap, + grand angle entre 2 stations*

*Distance entre épicentre estimé et station la + proche*

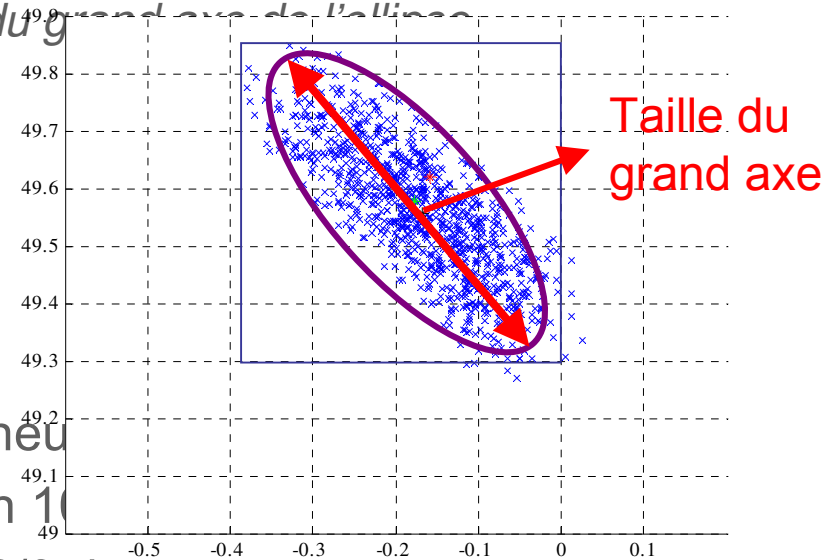
*Nombre de stations*

*Nombre de phase Pn, Pg...*

## En pratique...



- Apprentissage du quantile de dispersion des localisations
  - Approximé par l'ellipse contenante
    - Réduit le nombre de simulations (20 pts par événement)
    - évite le choix de la loi de dispersion
    - Pas de différence de complexité physique
    - Quantité d'intérêt : *taille du grand axe de l'ellipse*

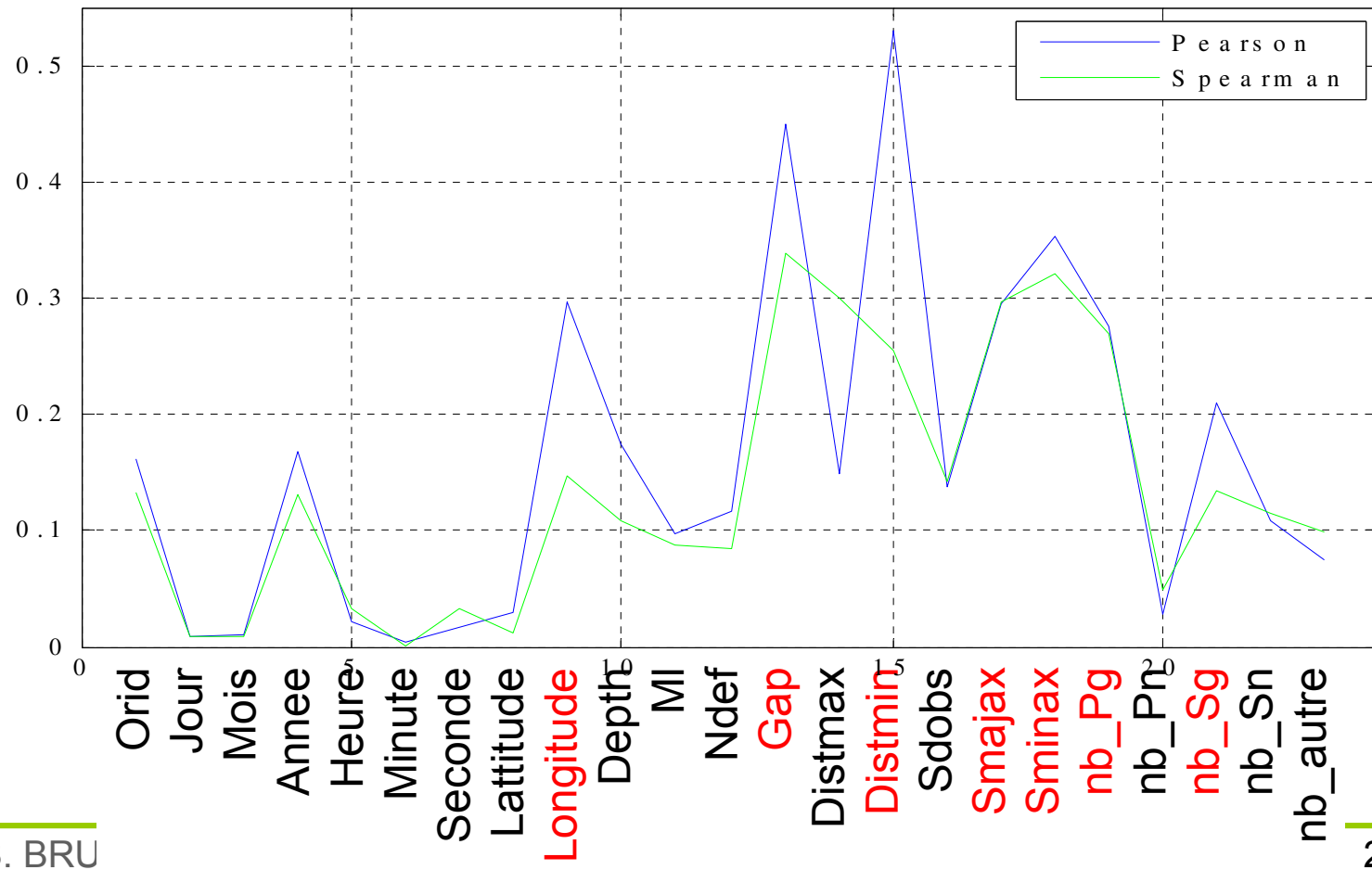


- Apprentissage automatique
  - Utilisation d'un réseau de neurones
  - Discrétisation des tailles en 10 classes
  - Résultat acceptable dans 2/3 des cas
  - Mais coté *obscur* incompatible avec injection de connaissance analyste

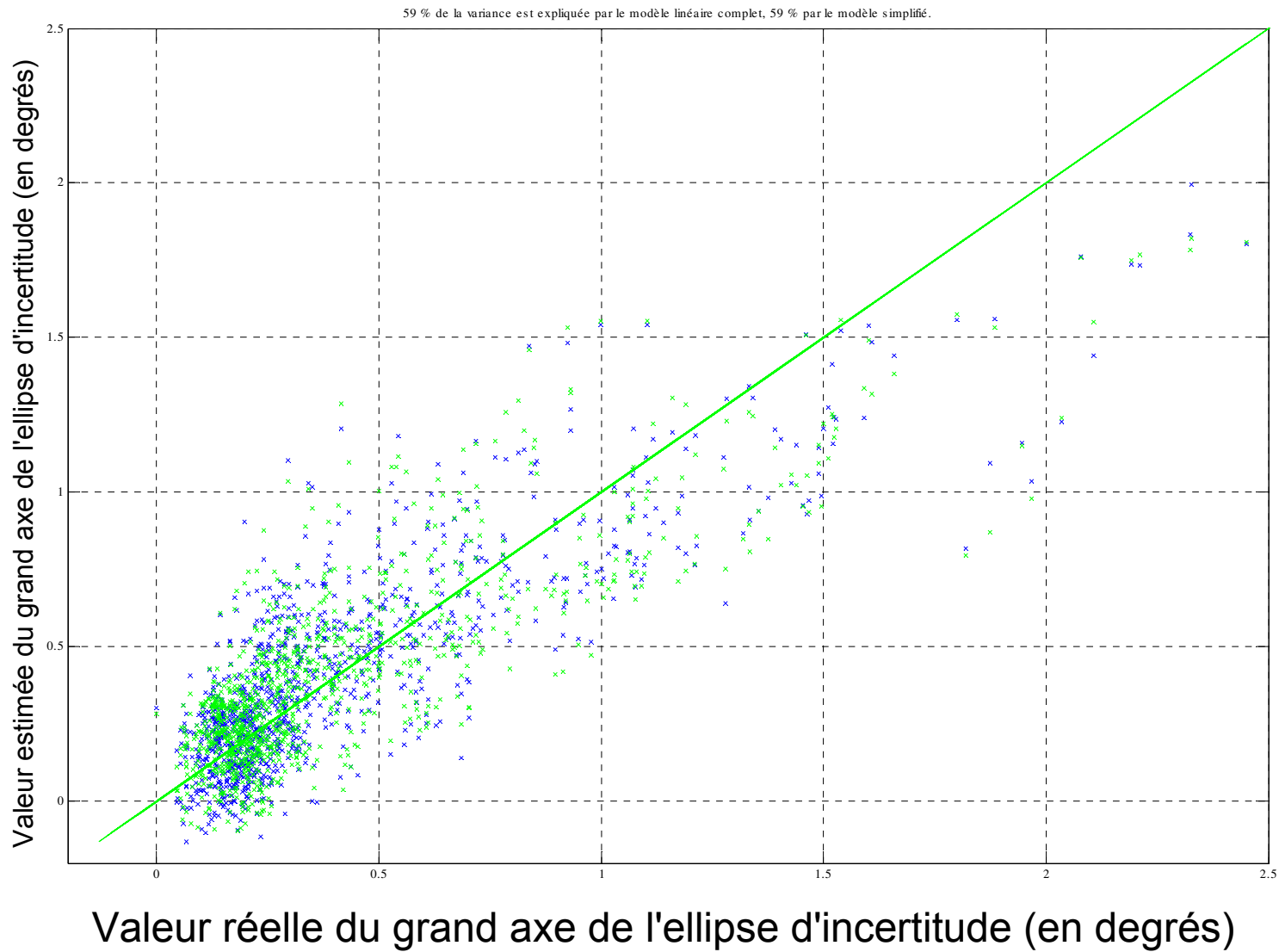
## Corrélation entre l'ellipse et les paramètres de couverture



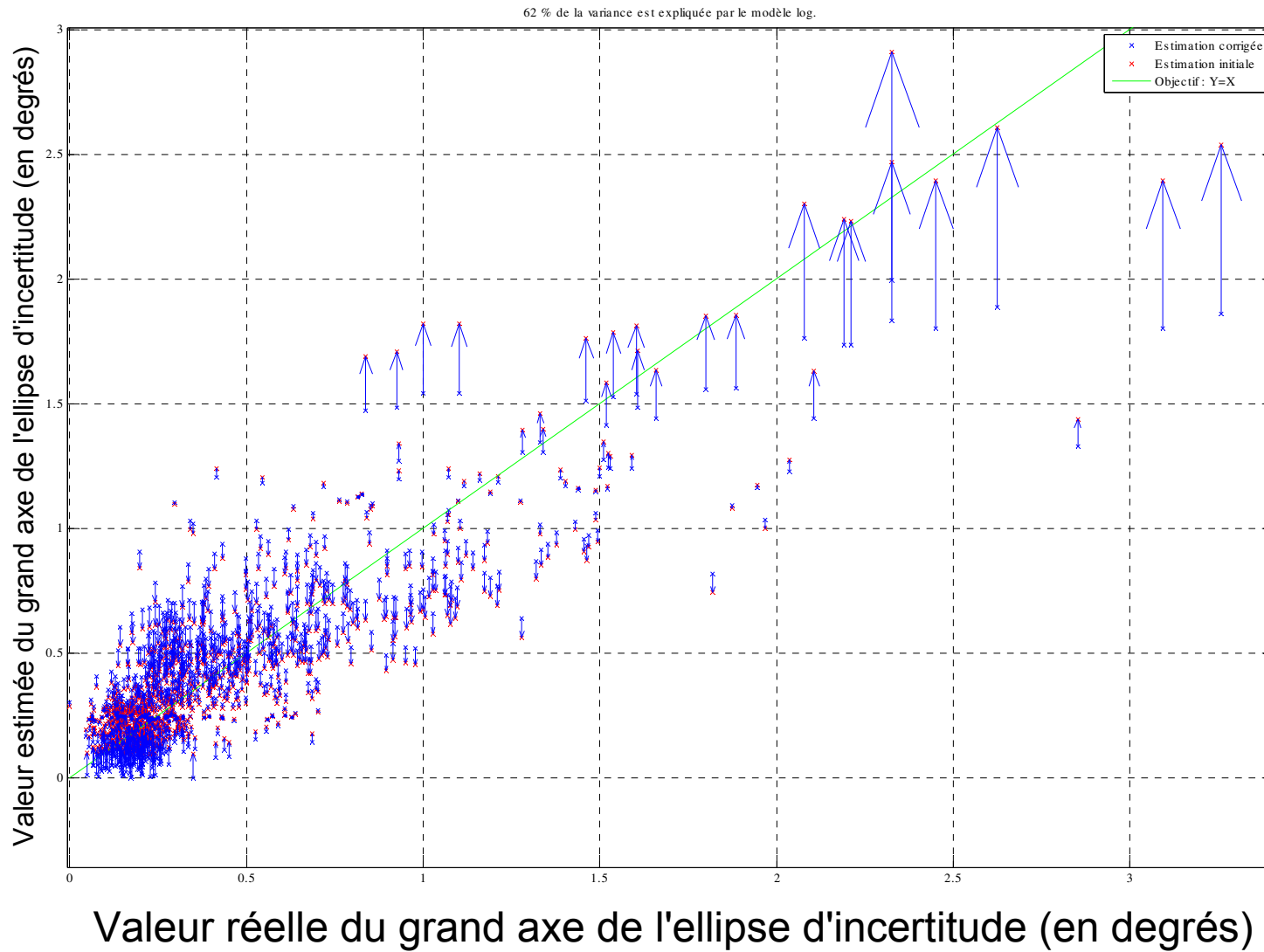
- + corrélés : Longitude, gap, distmin, smajax, sminax, nb\_Pg, nb\_Sg
- - corrélés : MI, nb\_Pn, latitude
- Distmax corrélé quadratiquement



# Régression linéaire multiple



# Correction logarithmique



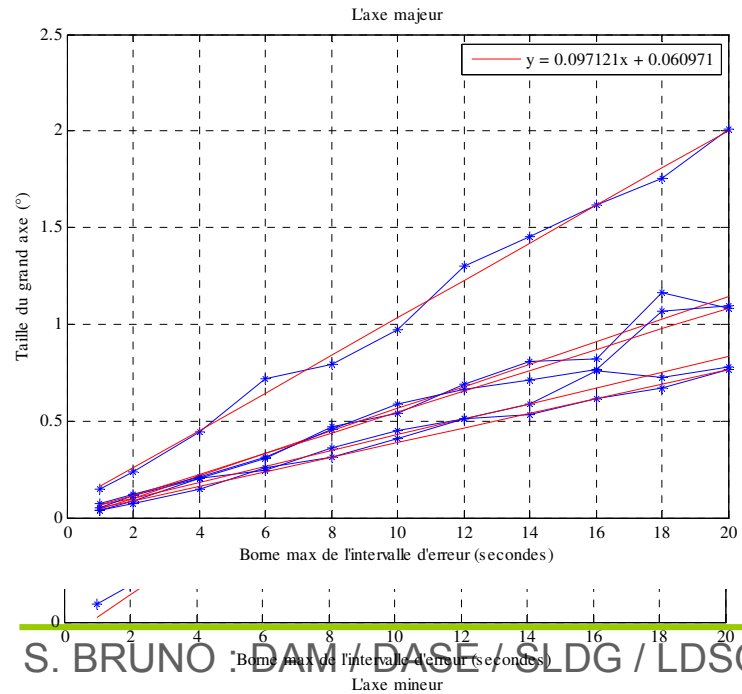
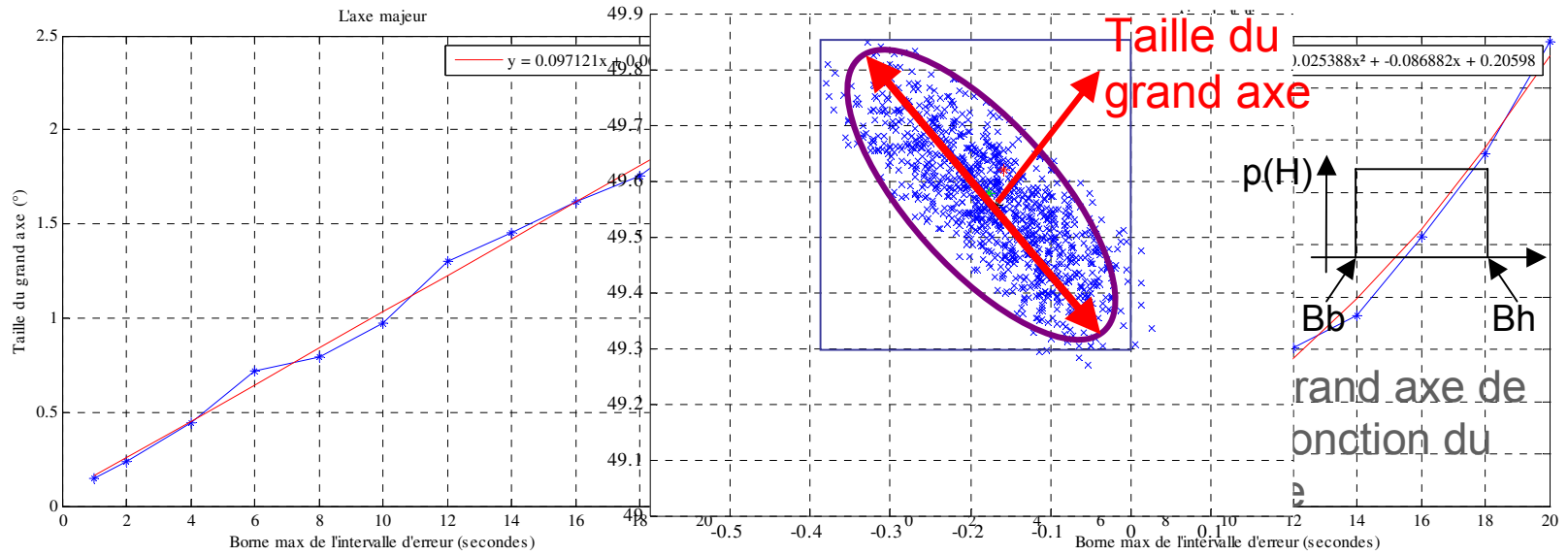
# Modèle de régression

---



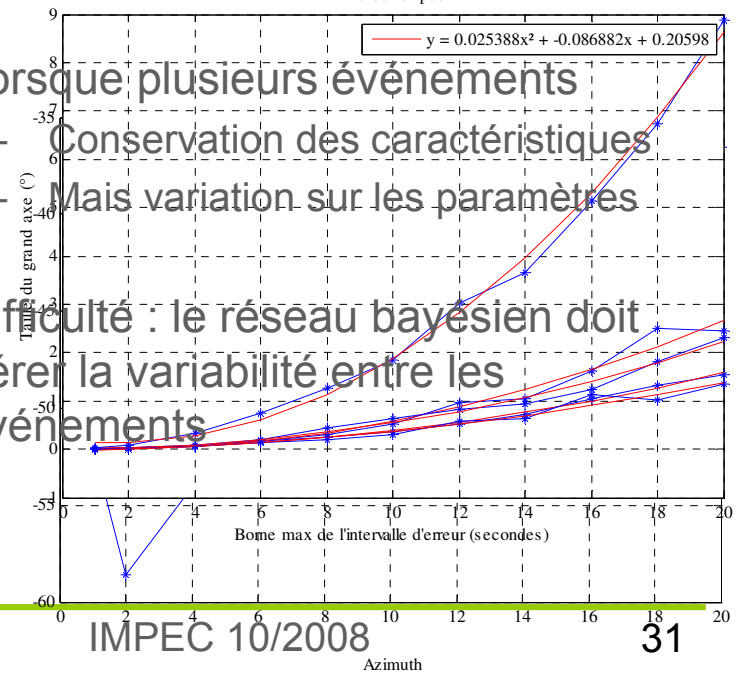
- L'erreur moyenne est de
  - 40% pour la régression linéaire (globale ou partielle),
  - 12% pour la régression log
- Soit environ 5 km pour un axe moyen de 45km.
  
- 95% des erreurs inférieures à 9km
  
- Il reste à réaliser un *modèle de modèles* :
  - Comment évolue ce modèle en variant les paramètres d'erreur ?
    - loi, support
    - différence finie

# Illustration de la variabilité entre les événements



- Lorsque plusieurs événements
  - Conservation des caractéristiques
  - Mais variation sur les paramètres

- Difficulté : le réseau bayésien doit gérer la variabilité entre les événements



S. BRUNO : DAM / DASE / SLDG / LDSO

IMPEC 10/2008 31

# Modèle de régression

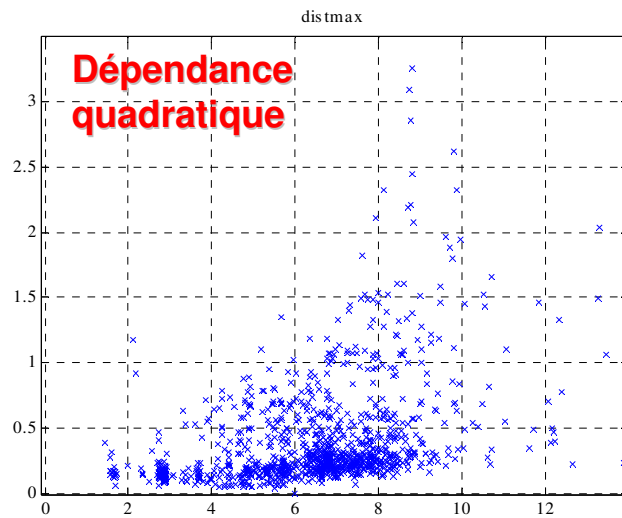
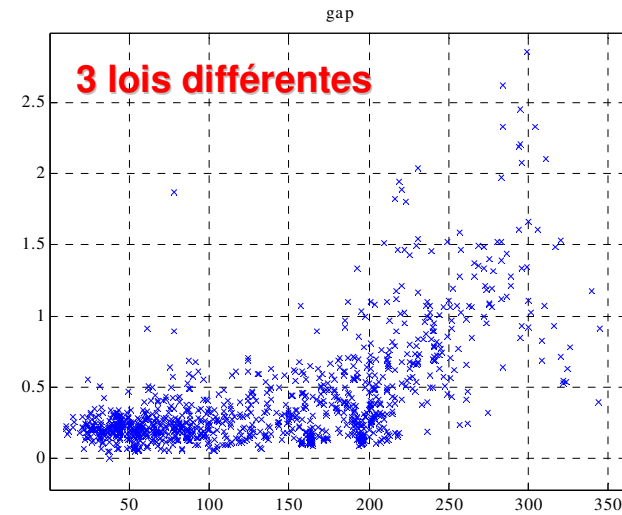
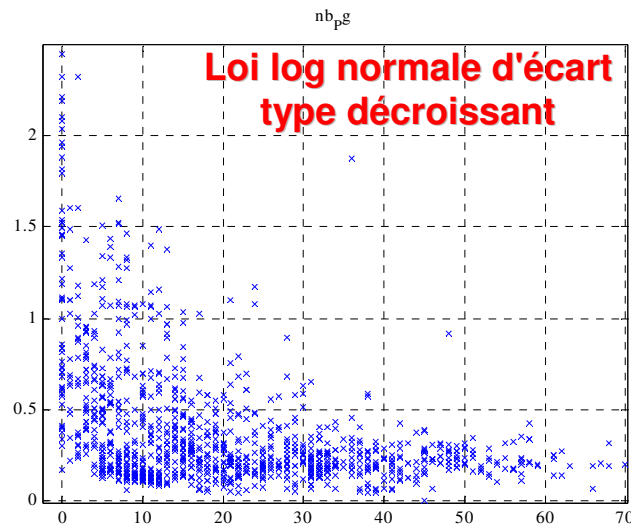
---



- L'erreur moyenne est de
  - 40% pour la régression linéaire (globale ou partielle),
  - 12% pour la régression log
- Soit environ 5 km pour un axe moyen de 45km.
  
- 95% des erreurs inférieures à 9km
  
- Il reste à réaliser un *modèle de modèles* :
  - Comment évolue ce modèle en variant les paramètres d'erreur ?
    - loi, support
    - différence finie
  
- La régression linéaire n'est pas une modélisation performante des distributions observées



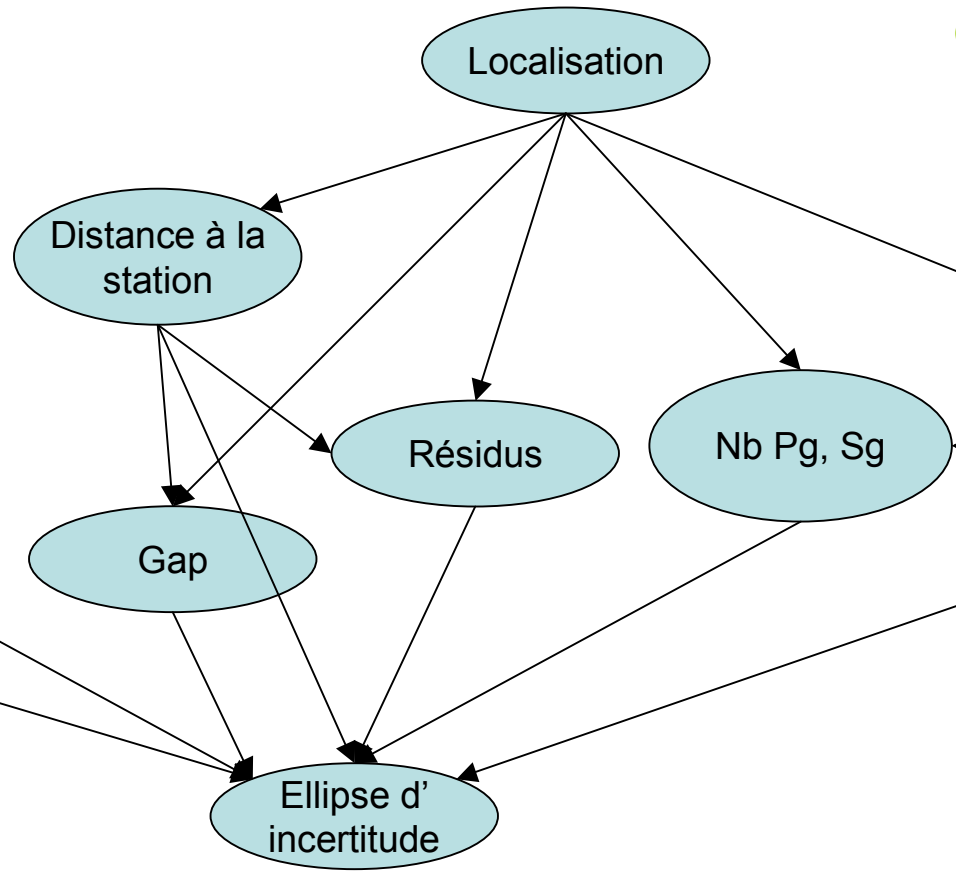
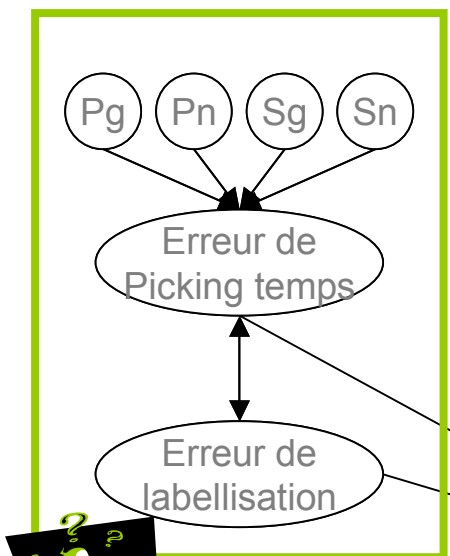
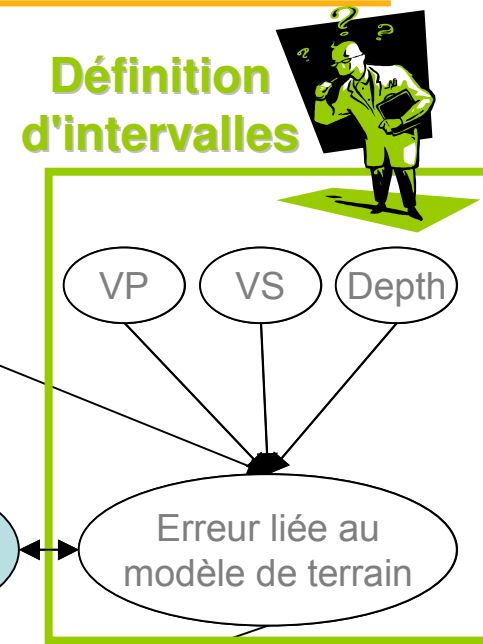
# Exemple de distributions entrées - sorties



Mauvaise qualité de la régression linéaire

Amélioration possible par un modèle plus souple

# Graphe de dépendances de l'incertitude de localisation

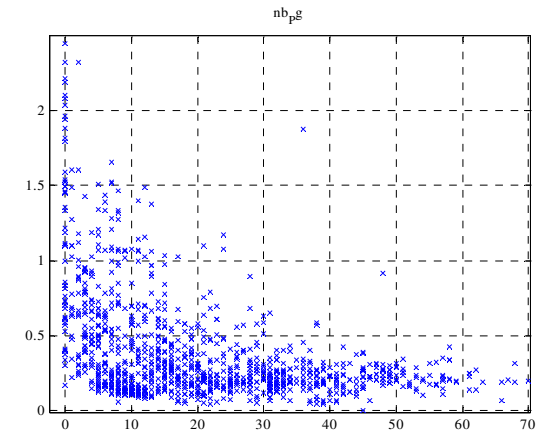


## Conclusion

---



- Bonne performance de l'estimation de l'ellipse d'incertitude à partir des paramètres de couverture
- Reste à ajouter les autres sources d'erreur
  - Erreur de labellisation
  - Erreur sur le nombre de couches
- Utiliser une modélisation plus performante
  - Réseau Bayésien
  - Modèle crédibiliste / possibiliste



# Questions ?

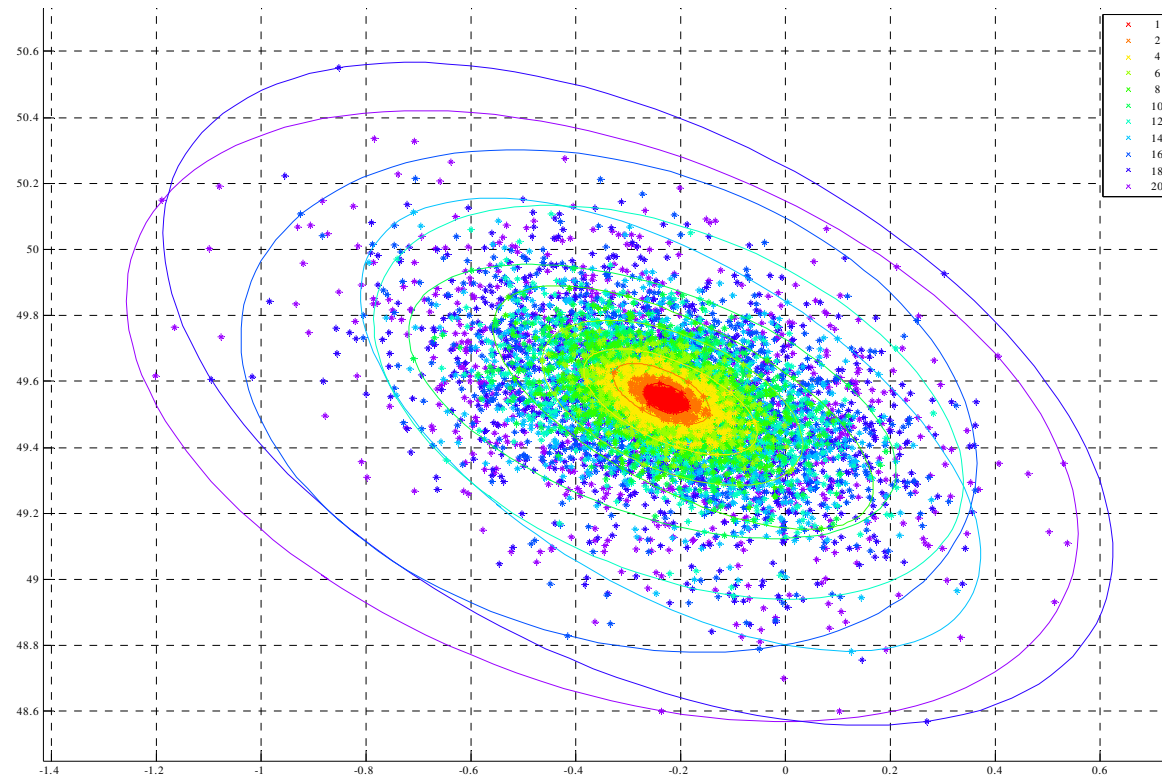
---



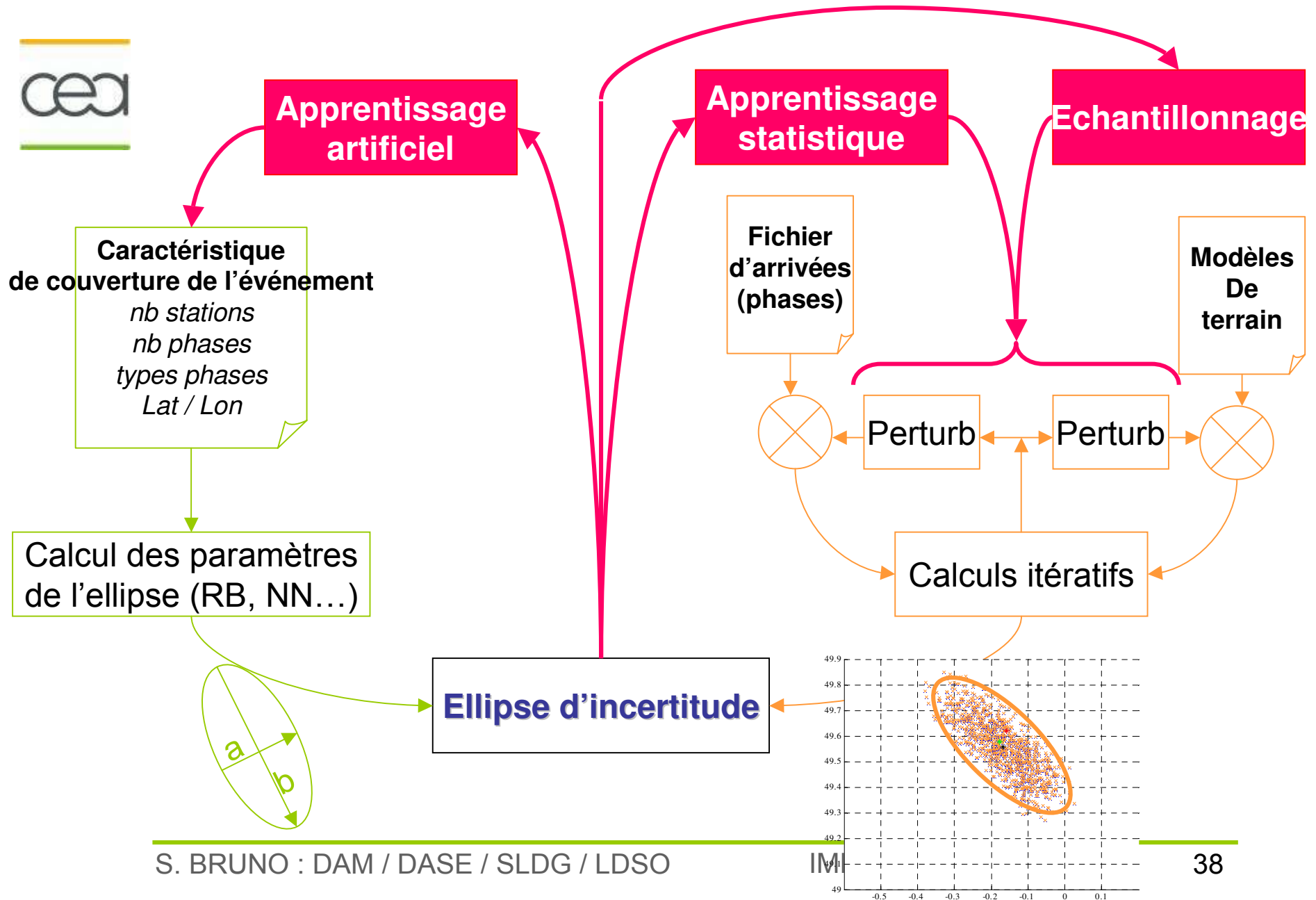
## Pour autant, le comportement global peut être calculé



- Pour autant, règles simples lorsqu'on s'intéresse au phénomène dans sa globalité
  - dépendance de l'ellipse à une borne d'erreur
  - Toutes les phases perturbées par une loi uniforme de même support
    - On fait varier les bornes du support



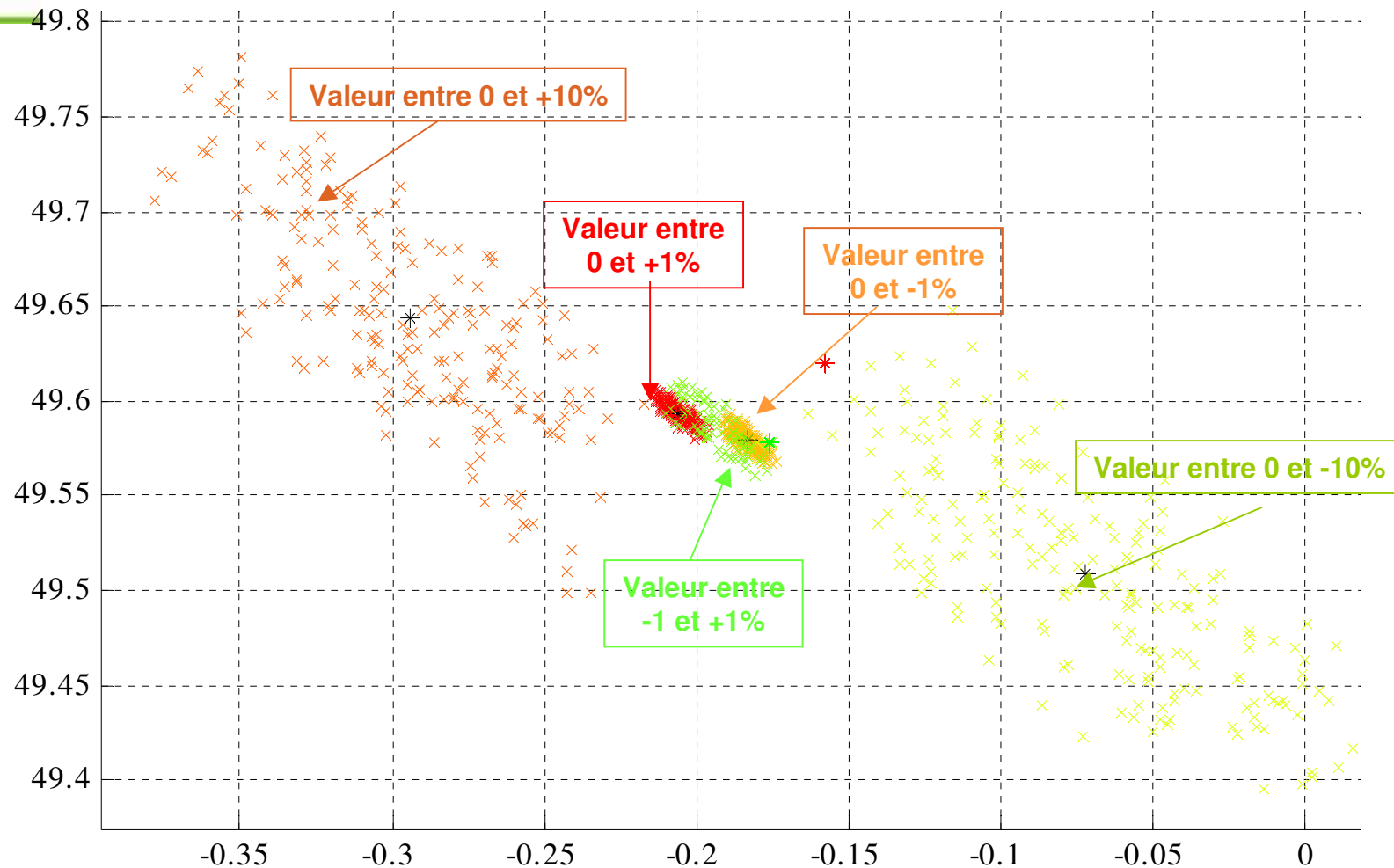
# Conclusion : deux méthodes complémentaires de calcul



## Exemple de difficulté de décomposition en sous-problèmes



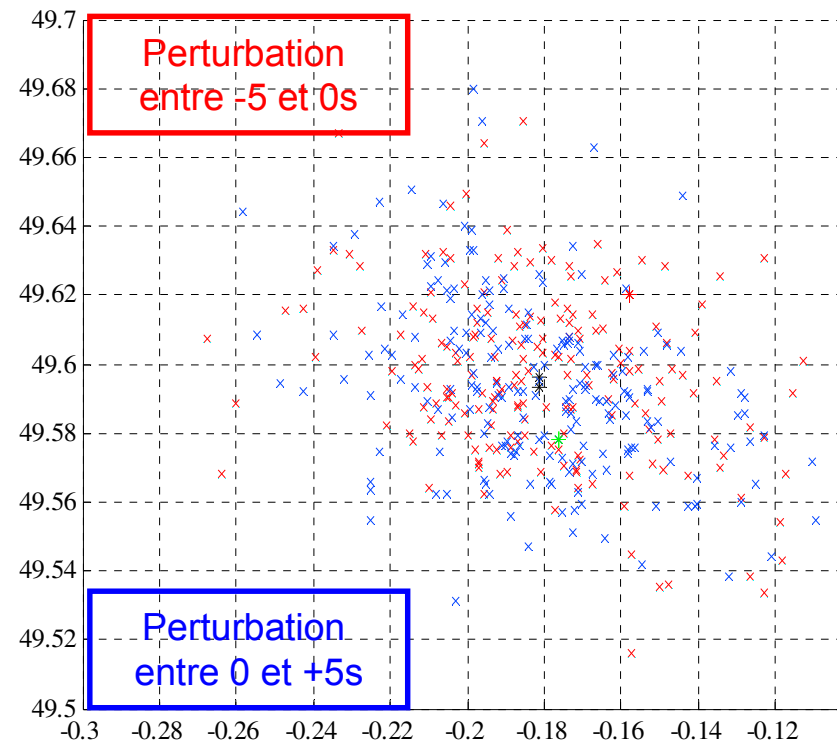
- Perturbation de tous les paramètres des modèles de terrain
- Comportement globalement symétrique



## Exemple de difficulté de décomposition en sous-problèmes



- Perturbation aléatoire de toutes les phases
  - Valeur choisie aléatoirement entre les deux bornes selon une loi uniforme. Valeur différente pour chaque phase.
- Pas d'effet de signe





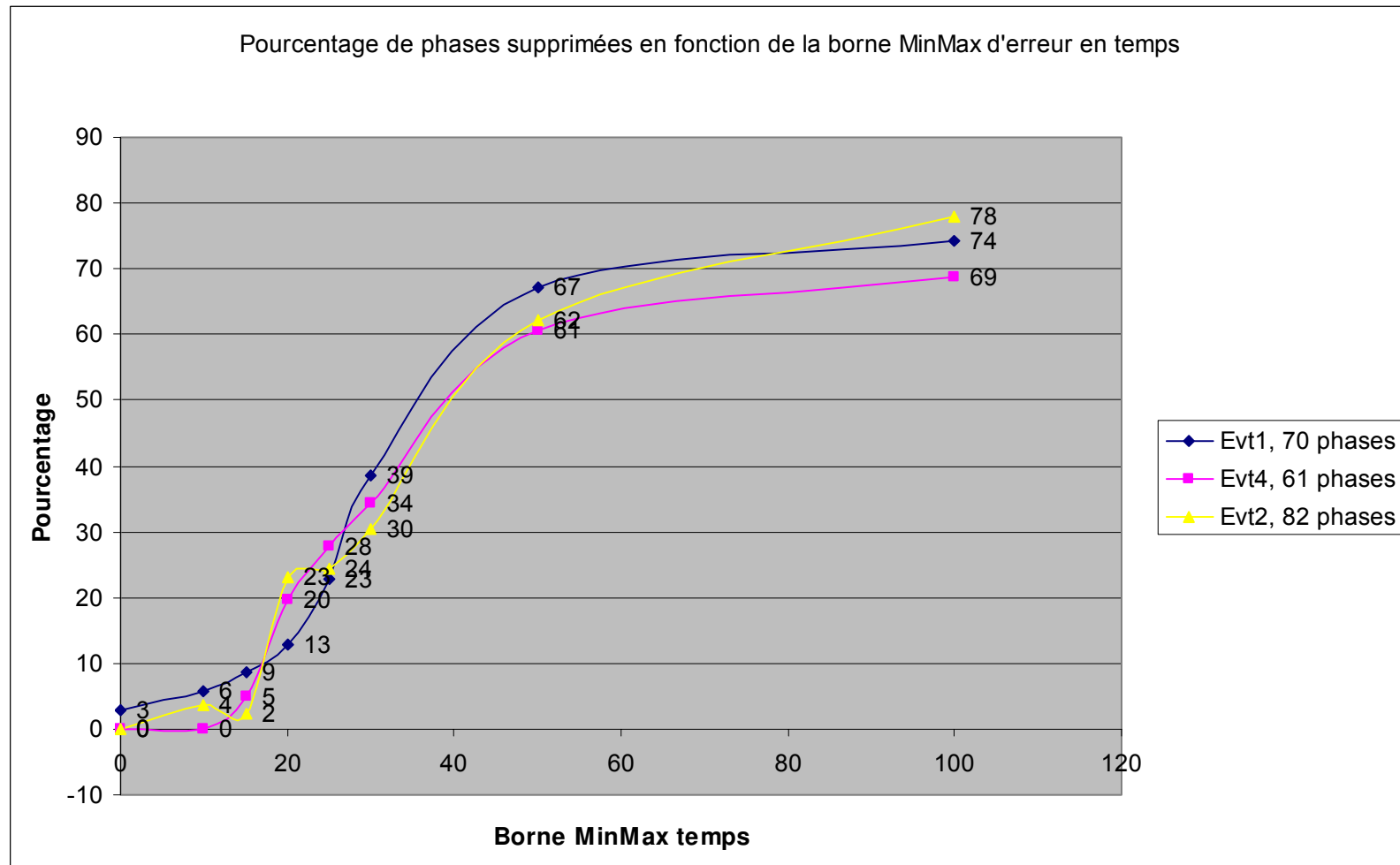
## Quels sont les paramètres influents

---



- La perturbation d'une  $P_n$  entraîne t'elle la même variation de localisation que la même perturbation d'une  $S_n$ ,  $P_g$ ,  $S_g$  ?
  - Pondération par type dans le programme de localisation
  - Dépend de la redondance des évènements
- Les points les plus éloignés sont-ils atteints pour
  - tous les  $P_g$  cohérents entre eux mais incohérents avec les  $S_g$
  - Tous les  $P_g$  incohérents entre eux et identiquement aux  $S_g$
- Influence relative entre incertitudes en temps et incertitudes en modèle

# Dépendance linéaire...mais gestion interne non linéaire !



- On ne peut pas décomposer le problème en éléments simple

# Démarche retenue et engagée

---



- Définition des paramètres entachés d'erreur et quantification de celle-ci
- Technique Monte-Carlo : exploration de l'espace
  - On considère que les valeurs de référence ne sont plus les valeurs vraies, mais qu'il s'agit du centre d'intervalles de longueur définie et qui contiennent la valeur vraie
  - On réalise des combinaisons aléatoires de ces valeurs de paramètres et on calcule la localisation associée
  - En réalisant suffisamment de simulations, on considère que les paramètres réels ont été approchés et donc que la localisation réelle est contenue dans le nuage de points simulés.
- Analyse des paramètres importants
  - ex : la variabilité due à l'imprécision sur  $V_p$  est peut-être négligeable devant la variabilité de la localisation liée à  $\Delta t(P_g)$ .
- Construction d'un modèle d'incertitude à partir de ces données
  - Difficulté liée à la non répétabilité des événements

## Difficultés liées au contexte sismique

---



- Chaque événement est différent
  - Nombre de phases variable
  - Proportion variable de chaque type de phase
  - Erreur du modèle de terrain local
- Pas de possibilité d'expérience
  - Comparaison entre modèle et réalité (quasi) impossible !
- Choix de ne pas entrer dans l'algorithmique
  - Effets d'interaction entre les différentes entrées
  - Comportement non linéaire