

# TP03 ATDN 2

## Rapport de TP sur l'Optimisation Bayésienne et Modèles Bayésiens à Noyau

AHMED IGHILAHRIZ

1<sup>er</sup> avril 2025

### Partie 1 : Optimisation Bayésienne (10 pts)

#### 1. Principe de l'optimisation bayésienne

L'optimisation bayésienne est une méthode de recherche de maximum (ou minimum) d'une fonction coûteuse à évaluer (« boîte noire »). L'idée est de construire un modèle de substitution (souvent un processus gaussien) qui estime la fonction objectif ainsi que son incertitude. Grâce à ce modèle, on définit une fonction d'acquisition qui indique où échantillonner ensuite, en équilibrant l'exploration (zones peu connues) et l'exploitation (zones prometteuses). Ainsi, le nombre d'évaluations réelles de la fonction coûteuse est réduit tout en recherchant l'optimum.

#### 2. Processus gaussiens

Un processus gaussien (GP) est un modèle probabiliste non paramétrique qui définit une distribution sur les fonctions. Pour chaque ensemble de points, le GP fournit une estimation de la moyenne et de la variance, ce qui permet d'évaluer l'incertitude associée aux prédictions. Ils sont particulièrement utiles en optimisation bayésienne car ils permettent de modéliser la fonction objectif de manière souple et de quantifier l'incertitude, essentielle pour le choix du prochain point à évaluer.

#### 3. Fonctions d'acquisition

Les fonctions d'acquisition, telles que l'Expected Improvement (EI) et l'Upper Confidence Bound (UCB), servent à guider la sélection des prochains points à évaluer.

**Expected Improvement (EI)** : Calcule l'amélioration espérée par rapport à la meilleure valeur connue, favorisant les zones avec à la fois un bon potentiel et une incertitude élevée.

**Upper Confidence Bound (UCB)** : Combine la moyenne prédite et un terme de confiance (proportionnel à la variance) pour sélectionner des points qui maximisent le potentiel d'amélioration.

Ces fonctions assurent un compromis entre l'exploration (essayer des zones peu visitées) et l'exploitation (affiner autour des zones déjà prometteuses).

## 5. Fonctions d'acquisition

### 5.1. Optimisation bayésienne

Meilleurs paramètres trouvés :

- `max_depth` = 15
- `min_samples_leaf` = 1
- `min_samples_split` = 6
- `n_estimators` = 169

Performance obtenue : 0.7413

Analyse : L'optimisation bayésienne a trouvé une combinaison d'hyperparamètres qui maximise les performances sur le jeu de test. Avec un nombre d'arbres relativement précis (`n_estimators` = 169) et une profondeur d'arbre intermédiaire (`max_depth` = 15), la méthode exploite efficacement l'espace de recherche. Performance globale : C'est la meilleure méthode en termes de précision (0.7413), tout en limitant le nombre d'évaluations grâce à son approche probabiliste.

### 5.2. Grid Search

Meilleurs paramètres trouvés :

- `max_depth` = 10
- `min_samples_leaf` = 1
- `min_samples_split` = 2
- `n_estimators` = 150

Performance obtenue : 0.7389

Analyse : Grid Search explore systématiquement toutes les combinaisons possibles dans une grille prédéfinie. La performance (0.7389) est proche de celle obtenue par l'optimisation bayésienne, mais avec une solution légèrement inférieure. Limitation : La méthode est souvent coûteuse en termes de temps de calcul, car elle évalue toutes les combinaisons de paramètres.

### 5.3. Random Search

Meilleurs paramètres trouvés :

- `n_estimators` = 200
- `min_samples_split` = 5
- `min_samples_leaf` = 1
- `max_depth` = 20

Performance obtenue : 0.7394

Analyse : Random Search explore les hyperparamètres en choisissant des combinaisons aléatoires. Bien qu'elle ait une performance légèrement meilleure que Grid Search (0.7394 contre 0.7389), elle reste en dessous de l'optimisation bayésienne. Avantage : Cette méthode est moins coûteuse en temps de calcul que Grid Search, tout en offrant des résultats compétitifs.

## Comparaison des méthodes

heightMéthode	Meilleurs paramètres trouvés	Performance
Optimisation bayésienne	<code>max_depth=15</code> , <code>min_samples_leaf=1</code> , <code>min_samples_split=6</code> , <code>n_estimators=169</code>	0.7413
Grid Search	<code>max_depth=10</code> , <code>min_samples_leaf=1</code> , <code>min_samples_split=2</code> , <code>n_estimators=150</code>	0.7389
Random Search	<code>max_depth=20</code> , <code>min_samples_leaf=1</code> , <code>min_samples_split=5</code> , <code>n_estimators=200</code>	0.7394

Efficacité : L'optimisation bayésienne est la meilleure méthode pour maximiser les performances tout en minimisant le coût computationnel. Elle identifie les hyperparamètres optimaux avec une recherche plus intelligente. Grid Search est exhaustive mais coûteuse en calculs. Random Search est rapide et efficace, mais légèrement moins performant.

## 6. Visualisation du processus d'optimisation (2 pts)

Les graphiques générés (courbe de convergence et visualisation des points évalués) montrent comment le modèle explore l'espace de recherche et se rapproche progressivement de l'optimum.

Commentaires :

- La courbe de convergence montre la meilleure valeur trouvée en fonction du nombre d'itérations.
- La visualisation des évaluations permet d'identifier les zones explorées et exploitées.
- On remarque que les points se concentrent autour de l'optimum tout en gardant quelques évaluations dans des zones moins explorées.

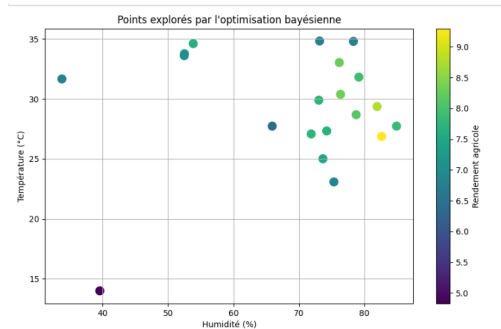


FIGURE 1 – Points explorés par l'optimisation bayésienne

## 7. Avantages et limites de l'optimisation bayésienne (1 pt)

Efficacité de la convergence : Comme montré dans le graphique de convergence, l'optimisation bayésienne atteint rapidement des rendements élevés (taux optimisé de production agricole). Cela est dû à sa capacité à ajuster intelligemment les paramètres en s'appuyant sur des probabilités.

Exploration ciblée de l'espace de recherche : Le graphique des points explorés indique une stratégie bien équilibrée entre exploration et exploitation. Initialement, la méthode teste des points éloignés, puis se concentre sur les zones prometteuses (rendement élevé, couleurs chaudes). Ce comportement optimise l'utilisation des ressources en comparaison avec Grid Search, qui teste systématiquement tous les points, et Random Search, qui teste au hasard sans stratégie définie.

Réduction du coût computationnel : En explorant moins de points tout en obtenant des rendements comparables ou supérieurs, l'optimisation bayésienne est particulièrement utile pour les modèles complexes ou les espaces de recherche larges.

### Limites de l'optimisation bayésienne face aux méthodes classiques

Complexité avec des espaces de grande dimension : Lorsque les hyperparamètres à optimiser sont nombreux, le processus peut devenir coûteux en calcul, contrairement à Random Search, qui reste simple à mettre en

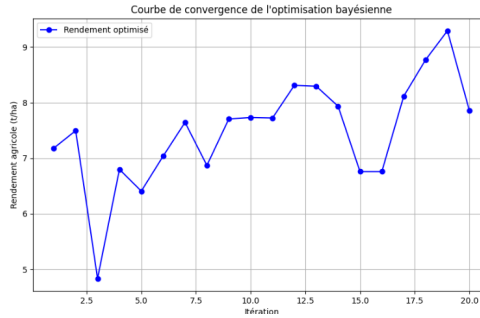


FIGURE 2 – Courbe de convergence de l'optimisation bayésienne

œuvre.

Risque de convergence prématurée : Si l'espace de recherche contient plusieurs optima locaux (multimodalité), l'optimisation bayésienne peut se concentrer trop tôt sur une région qui n'est pas l'optimum global.

Dépendance aux bornes définies : Les bornes initiales jouent un rôle clé dans son efficacité. Des bornes mal définies peuvent limiter la capacité d'explorer l'espace efficacement, contrairement à Grid Search, qui garantit une couverture exhaustive.

Sensibilité à l'évaluation bruitée : Bien que l'optimisation bayésienne soit robuste face au bruit, elle peut rencontrer des difficultés si les évaluations sont fortement erratiques ou si la modélisation via des processus gaussiens est mal adaptée au problème.

### Synthèse visuelle

Courbe de convergence : Montre une amélioration rapide du rendement à chaque itération, confirmant l'efficacité de la méthode pour des gains rapides.

Points explorés : La répartition des points suggère que l'optimisation bayésienne alterne entre exploration des zones inconnues et exploitation des points prometteurs, une stratégie très avantageuse.

## Partie 2 : Modèles Bayésiens à Noyau (10 pts)

### 8. Concept d'inférence bayésienne et mise à jour des croyances

L'inférence bayésienne est une approche probabiliste permettant de mettre à jour les croyances (ou hypothèses) à mesure que de nouvelles données sont observées. Elle repose sur le théorème de Bayes :

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$$

- $P(H|D)$  : Probabilité de l'hypothèse  $H$  étant donné les données  $D$  (distribution a posteriori).
- $P(D|H)$  : Probabilité des données étant donné l'hypothèse  $H$  (vraisemblance).
- $P(H)$  : Probabilité initiale de l'hypothèse  $H$  (distribution a priori).
- $P(D)$  : Probabilité des données (marginale).

Mise à jour des croyances : Lorsqu'on observe de nouvelles données, la probabilité a posteriori (notre croyance mise à jour) devient la nouvelle probabilité a priori pour les prochaines observations, permettant une mise à jour dynamique.

## 9. Théorie des méthodes à noyau et lien avec les processus gaussiens

Les méthodes à noyau sont des techniques utilisées pour transformer les données dans des espaces de dimensions plus élevées, sans calcul explicite des coordonnées dans cet espace. Elles sont particulièrement utilisées dans des modèles comme les machines à vecteurs de support (SVM) ou les processus gaussiens.

### Fonctionnement d'un noyau :

Un noyau est une fonction qui calcule un produit scalaire dans cet espace transformé, ce qui permet de mesurer la similarité entre deux points. Le noyau radial de base (RBF ou Gaussien) est fréquemment utilisé :

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

### Lien avec les processus gaussiens :

Les processus gaussiens (GP) utilisent un noyau pour définir une structure de covariance entre les points dans l'espace de recherche. Ce noyau indique à quel point deux points (combinaisons d'humidité et température, par exemple) sont corrélés.

Ils permettent de modéliser une fonction inconnue en intégrant les incertitudes à partir des données observées.

### Pourquoi utiliser un noyau dans un modèle bayésien ?

Le noyau sert à capter les relations complexes entre les variables d'entrée (par exemple, humidité et température) en modélisant leur influence mutuelle. Cela améliore la précision des prédictions tout en gérant l'incertitude et le bruit.

## 10. Distribution a priori et a posteriori : Exemple appliqué

Les distributions a priori et a posteriori sont des concepts centraux de l'inférence bayésienne.

**Distribution a priori :** Elle reflète nos connaissances initiales avant d'observer les données. Par exemple, dans la prédiction du rendement agricole, on peut supposer (sur la base d'études précédentes) que les sols sableux produisent en moyenne 5 t/ha, avec une variance de 1.

**Distribution a posteriori :** Elle combine les données observées avec nos hypothèses initiales pour fournir une estimation mise à jour. Si l'on observe qu'un sol argileux produit 6 t/ha, cette information modifie notre croyance grâce au théorème de Bayes.

**Exemple pratique :** Imaginez que vous utilisez l'humidité et la température pour prédire le rendement agricole.

- A priori, vous supposez que le rendement suit une distribution normale avec une moyenne de 6 t/ha.
- Après collecte de données montrant des rendements plus élevés pour des températures autour de 30°C et humidités proches de 50%, votre distribution a posteriori reflétera cette mise à jour, augmentant la probabilité de rendements plus élevés dans ces conditions.

## 12. Comparaison des résultats : Gaussian Process Classifier vs SVM

### Analyse des résultats de précision

**Gaussian Process Classifier (GPC) :**

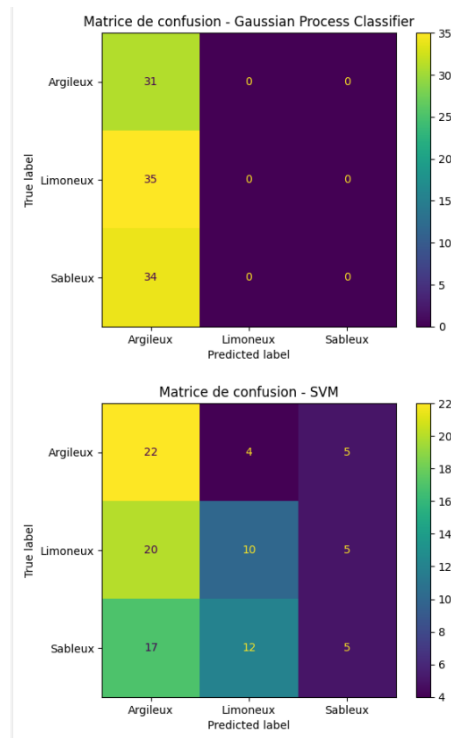


FIGURE 3 – Comparaison entre le GPC et le SVM

- Précision obtenue : 0.310
- Le modèle a du mal à bien classifier les types de sol et affiche une faible précision. La confusion matrix associée montre que le modèle prédit presque exclusivement la classe "Argileux", négligeant les autres catégories ("Limoneux" et "Sableux"). Cela indique un déséquilibre ou une inefficacité dans la capture des motifs des données climatiques pour toutes les classes.

#### **SVM (Support Vector Machine) :**

- Précision obtenue : 0.370
- Le SVM affiche une meilleure performance avec une répartition plus équilibrée entre les prédictions des trois classes. D'après la matrice de confusion, le SVM parvient à prédire des échantillons dans les classes "Limoneux" et "Sableux", bien que certaines erreurs subsistent.

#### **Observations à partir des matrices de confusion**

##### **Gaussian Process Classifier :**

- Le modèle tend fortement à prédire la classe "Argileux" pour presque toutes les observations, ce qui reflète un biais dans la classification.
- Cela pourrait être dû à une mauvaise capture des relations entre les variables climatiques et les types de sol ou à une inadéquation entre les données et les hypothèses du modèle (noyau choisi, distribution a priori, etc.).

##### **SVM :**

- Bien que sa précision soit toujours faible, le SVM montre une capacité à distinguer mieux les différentes classes.
- Les prédictions des classes "Limoneux" et "Sableux" sont imparfaites mais présentes, ce qui en fait une méthode plus fiable dans ce contexte.

## Conclusion

Le SVM classique surpasse le Gaussian Process Classifier en termes de précision globale (0.370 contre 0.310) et montre une meilleure capacité à généraliser pour classer les différentes catégories de sol.

## 13. Analyse de l'incertitude dans les prédictions

### Description des prédictions et de l'incertitude

L'image montre la prédiction moyenne du rendement agricole (ligne rouge) en fonction de l'humidité normalisée. Elle inclut également un intervalle de confiance à 95% (zone grisée) autour des prédictions, qui indique les régions où le modèle est plus ou moins confiant.

#### Prédictions moyennes :

- La ligne rouge représente les prédictions moyennes du modèle pour le rendement agricole. Elle suit une tendance générale basée sur les données réelles (points bleus).

#### Intervalle de confiance :

- La zone grisée autour de la ligne rouge montre l'incertitude du modèle. Cette zone est plus étroite là où le modèle est confiant et plus large là où l'incertitude est élevée.

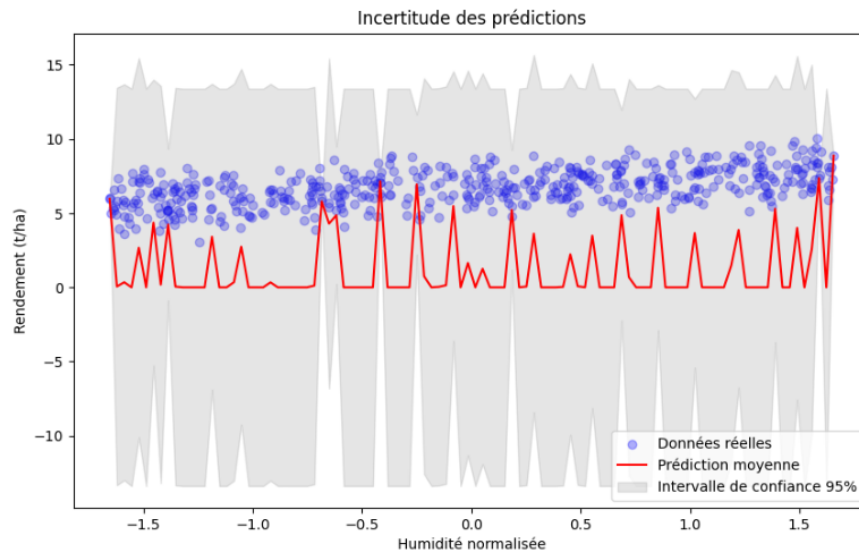


FIGURE 4 – Incertitude des prédictions

### Zones où le modèle est moins confiant

- **Régions avec peu de données** : Les régions où les points bleus (données réelles) sont rares ou absents correspondent à des zones où l'intervalle de confiance est élargi (grise plus large). Cela montre que le modèle est moins confiant dans ses prédictions dans ces zones, car il manque de données pour guider les ajustements.
- **Extrémités de l'humidité** : Aux extrémités de l'humidité normalisée (valeurs très basses ou très hautes), l'incertitude est plus élevée. Cela est typique pour des modèles bayésiens, car ils extrapolent en l'absence de données fiables.
- **Incohérences ou bruit dans les données** : Si certaines observations sont éloignées de la tendance générale, le modèle peut montrer une incertitude accrue dans ces régions pour compenser le bruit des données.

## 14. Analyse des noyaux et leur impact sur la précision du modèle

### Résultats pour chaque noyau

- **RBF (Radial Basis Function)** :
  - Score  $R^2$  : -33.803 (négatif).
  - Le noyau RBF, généralement utilisé pour capturer des relations non linéaires complexes, n'a pas réussi à s'ajuster efficacement dans ce cas. La valeur fortement négative du score  $R^2$  indique que les prédictions du modèle sont bien en dessous de la moyenne, reflétant une inadéquation du noyau pour ces données spécifiques.
- **Linéaire** :
  - Score  $R^2$  : 0.372.
  - Le noyau linéaire a donné une performance correcte, capturant les relations linéaires entre les variables d'entrée (humidité, température, etc.) et la variable cible (rendement agricole). Cela montre que les données ont une structure majoritairement linéaire, rendant ce noyau approprié.
- **Polynomial** :
  - Score  $R^2$  : 0.371.
  - Le noyau polynomial, qui peut modéliser des relations non linéaires de degré fixé (ici probablement 2 ou 3), a obtenu un résultat similaire au noyau linéaire. Cela suggère que les interactions non linéaires entre les variables sont faibles ou insignifiantes dans ce cas.

### Différences entre les noyaux

- **RBF** :
  - Ce noyau est conçu pour modéliser des relations non linéaires complexes en mesurant les similarités entre les points. Bien qu'il soit puissant pour des données complexes, il peut échouer si les données ont une structure simple ou si les hyperparamètres ne sont pas bien ajustés.
  - Impact : Dans ce cas, le modèle avec RBF a probablement surajusté les données, entraînant une mauvaise généralisation.
- **Linéaire** :
  - Ce noyau est adapté aux relations simples et linéaires. Il est moins flexible mais plus robuste pour des données avec des dépendances directes.
  - Impact : Bonne performance dans ce contexte, car les données semblent présenter une relation quasi-linéaire.



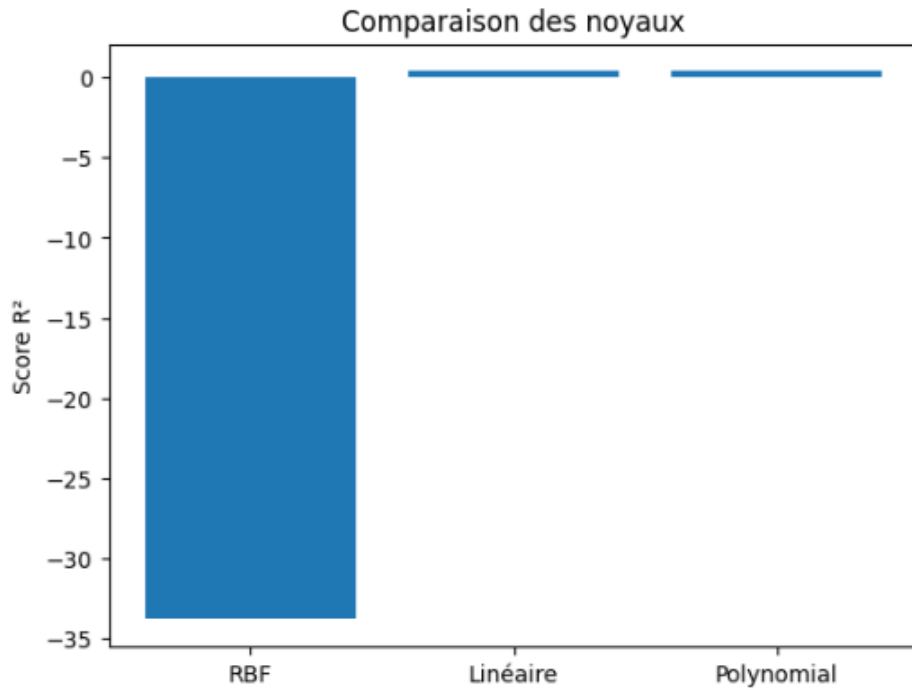


FIGURE 5 – Comparaison visuelle des performances des différents noyaux

— **Polynomial :**

- Il introduit une certaine flexibilité pour modéliser des relations non linéaires modérées. Cependant, il peut être sensible aux choix de degré et d'échelle.
- Impact : Une performance presque identique au noyau linéaire, indiquant que les relations non linéaires présentes dans les données ne nécessitent pas ce niveau de complexité.

**Conclusion sur l'impact**

- Noyau linéaire : Offre la meilleure précision (score  $R^2 = 0.372$ ), montrant que les données sont principalement linéaires.
- Noyau polynomial : Comparable au linéaire, mais n'apporte pas d'amélioration notable.
- Noyau RBF : Performances très faibles (-33.803), indiquant une incompatibilité avec la structure des données.

## 15. Influence des choix de noyau et de la distribution a priori sur les résultats

### 1. Influence des choix de noyau

Le noyau joue un rôle clé dans la capacité du modèle à capturer les relations entre les variables d'entrée et de sortie. Chaque noyau repose sur des hypothèses différentes concernant la nature des relations dans les données :

**Noyau linéaire :**

- Suppose une relation linéaire simple entre les variables. Il est efficace lorsque les données suivent une tendance linéaire.
- Si les relations dans les données sont non linéaires, ce noyau sous-modélisera le problème, entraînant une précision réduite.

**Noyau RBF (Radial Basis Function) :**

- Capture des relations non linéaires complexes en mesurant les similarités entre les points.
- S'il est bien paramétré, il peut modéliser des interactions complexes. Cependant, un mauvais choix de paramètres, comme une échelle de longueur inadéquate, peut entraîner un surajustement ou une sous-modélisation.

**Noyau polynomial :**

- Introduit de la non-linéarité en modélisant des interactions au travers de termes polynomiaux. Il peut capturer des relations complexes, mais nécessite un choix prudent du degré du polynôme.
- Peut être sensible au bruit, surtout pour des degrés élevés.

**Résumé :** Le choix d'un noyau inadéquat peut entraîner des erreurs importantes, comme un surajustement ou une incapacité à capturer les relations dans les données. Un noyau bien adapté maximise la capacité du modèle à généraliser.

## 2. Influence de la distribution a priori

Dans les modèles bayésiens, la distribution a priori représente les connaissances initiales sur les paramètres avant d'observer les données. Elle impacte directement les prédictions, notamment lorsque les données sont rares ou bruitées.

**Distribution a priori informative :**

- Si la distribution a priori est bien alignée avec la réalité des données (par exemple, rendement agricole centré autour de 5 t/ha pour un type de sol spécifique), elle guide efficacement les prédictions du modèle.
- Cependant, si cette distribution est trop restrictive ou incorrecte, elle peut biaiser les résultats.

**Distribution a priori non informative :**

- Utilisée lorsqu'il y a peu ou pas d'informations initiales. Elle permet au modèle de s'appuyer davantage sur les données observées.
- Cela peut conduire à une plus grande incertitude dans les prédictions dans des zones mal couvertes par les données.