



**HAL**  
open science

## Problèmes inverses et test bayésien d'adéquation du modèle

Pierre Palud, Pierre-Antoine Thouvenin, Pierre Chainais, Emeric E. Bron,  
Franck Le Petit

► **To cite this version:**

Pierre Palud, Pierre-Antoine Thouvenin, Pierre Chainais, Emeric E. Bron, Franck Le Petit. Problèmes inverses et test bayésien d'adéquation du modèle. 29° Colloque sur le traitement du signal et des images, GRETSI - Groupe de Recherche en Traitement du Signal et des Images, Aug 2023, Grenoble, France. pp.697 - 700. hal-04479586

**HAL Id: hal-04479586**

**<https://hal.science/hal-04479586>**

Submitted on 27 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Problèmes inverses et test bayésien d'adéquation du modèle

Pierre PALUD<sup>1,2</sup> Pierre-Antoine THOUVENIN<sup>1</sup> Pierre CHAINAIS<sup>1</sup> Emeric BRON<sup>2</sup> Franck Le PETIT<sup>2</sup>  
I. BEŠLIĆ<sup>3</sup> L. EINIG<sup>3</sup> M. GARCIA SANTA-MARIA<sup>3</sup> J. H. ORKISZ<sup>3</sup> A. ZAKARDJIAN<sup>3</sup> S. BARDEAU<sup>3</sup>  
J. CHANUSSOT<sup>3</sup> M. GERIN<sup>3</sup> J. R. GOICOECHEA<sup>3</sup> P. GRATIER<sup>3</sup> V. V. GUZMAN<sup>3</sup> F. LEVRIER<sup>3</sup> K. ÖBERG<sup>3</sup>  
N. PERETTO<sup>3</sup> J. PETY<sup>3</sup> A. ROUEFF<sup>3</sup> E. ROUEFF<sup>3</sup> A. SIEVERS<sup>3</sup> A. HUGHES<sup>3</sup>

<sup>1</sup> Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

<sup>2</sup> LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 92190 Meudon, France

<sup>3</sup> Consortium ORION-B (<https://www.iram.fr/~pety/ORION-B/team.html>)

**Résumé** – La résolution d'un problème inverse est basée sur le choix d'un modèle, dont la qualité affecte fortement la pertinence des estimateurs. Ce travail présente une méthode d'identification des cas où le modèle utilisé ne permet pas de reproduire les observations de façon vraisemblable, ce qui traduit un choix de modèle peu pertinent. Cette méthode, simple à implémenter et à interpréter, est basée sur un calcul de valeur-p bayésienne. Nous proposons d'utiliser la vraisemblance comme statistique de test, ce qui ne nécessite aucune hypothèse sur la distribution du bruit ou sur la régularité du modèle direct. L'absence de calcul exact de la valeur-p correspondante est compensée par une modélisation des incertitudes inhérentes à son évaluation numérique.

**Abstract** – Inference procedures in inverse problems are based on the choice of models, whose qualities greatly affect estimations relevance. This work presents an identification method for cases where the model used cannot reconstruct observations, which permits detecting potential modelling errors a posteriori. This method, simple to implement and interpret, is based on a Bayesian p-value. We propose to use the likelihood as test statistic as it is applicable for any noise law and forward model. The absence of exact computation of the p-value is compensated by accounting for the uncertainties associated to its numerical evaluation.

## 1 Introduction

La résolution d'un problème inverse repose sur le choix d'un modèle. La qualité de ce modèle affecte grandement la pertinence des estimations pour un utilisateur, notamment dans des problèmes sur données réelles. Il est donc crucial de pouvoir évaluer cette qualité dans l'absolu, et son impact sur les estimations. La plupart des travaux dans la littérature comparent plusieurs modèles pour déterminer lequel est le meilleur [1]. Les travaux permettant d'évaluer un modèle individuellement s'appuient en général sur l'espérance du log de la densité de la distribution prédictive a posteriori [2, chapitre 7]. Cette évaluation est généralement réalisée sur les données ayant servi à définir la loi a posteriori, ce qui induit un biais optimiste. Plusieurs termes de correction existent, basés sur de la validation croisée [3]. Ces corrections nécessitent beaucoup de calculs et les valeurs obtenues sont difficilement interprétables pour des fonctions de vraisemblances complexes.

Dans de nombreuses disciplines, la capacité d'un modèle à reproduire les observations est souvent vérifiée avec un test du  $\chi^2_\nu$  à  $\nu$  degrés de liberté [4]. Ce test simple, peu coûteux et interprétable, n'a statistiquement de sens que pour les problèmes inverses basés sur un modèle direct linéaire sous hypothèse de bruit additif gaussien. Il est pourtant fréquemment utilisé avec des modèles directs non-linéaires [5].

Nous proposons une méthode générale d'identification d'incompatibilité entre observations réelles et reconstruites à l'aide du modèle d'observation. Cette méthode repose sur le calcul d'une valeur-p bayésienne [2, chapitre 6], interprétable et tenant compte de nombreuses sources d'incertitudes. Elle est peu coûteuse en temps de calcul, peut s'appliquer dans un cadre bayésien ou d'optimisation, quels que soient le modèle direct ou la loi du bruit dans le modèle d'observation.

La section 2 présente l'approche proposée. La section 3 en propose deux illustrations : sur un cas simple où le test du  $\chi^2$  est adapté, puis sur un problème inverse en astrophysique [6].

## 2 Test d'hypothèse bayésien

### 2.1 Rappels en inférence bayésienne

Nous considérons un problème inverse général. Les observations sont notées  $\mathbf{Y} = (\mathbf{y}_n)_{n \in \llbracket 1, N \rrbracket} \in \mathbb{R}^{N \times L}$ , où  $N$  est le nombre d'observations individuelles  $\mathbf{y}_n$ , par exemple des pixels, et  $L$  le nombre de composantes par observation. Les paramètres à estimer sont notés  $\Theta = (\theta_n) \in \mathbb{R}^{N \times D}$ , avec un vecteur  $\theta_n \in \mathbb{R}^D$  par observation. Nous supposons pouvoir prédire des observables  $\mathbf{y}_n$  à partir d'un paramètre  $\theta_n$  grâce à un modèle direct connu  $\mathbf{f} : \mathbb{R}^D \rightarrow \mathbb{R}^L$ . Un modèle d'observation combine modèle direct  $\mathbf{f}$  et modèle de bruit. Il permet de définir une fonction de vraisemblance  $\pi(\mathbf{Y}|\Theta)$ . Les réalisations des observables  $\mathbf{y}_n$  sont supposées indépendantes :

$$\pi(\mathbf{Y}|\Theta) = \prod_{n=1}^N \pi(\mathbf{y}_n|\theta_n) \quad (1)$$

Cette hypothèse n'est pas strictement nécessaire mais permet des diagnostics plus détaillés dans certaines applications telles que dans la partie 3.2. L'ajout d'une information a priori  $\pi(\Theta)$  à la vraisemblance définit une distribution a posteriori. Grâce au théorème de Bayes, la distribution a posteriori s'écrit

$$\pi(\Theta|\mathbf{Y}) \propto \pi(\mathbf{Y}|\Theta)\pi(\Theta) \quad (2)$$

Une approche bayésienne permet de quantifier naturellement l'incertitude associée à des estimateurs ponctuels.

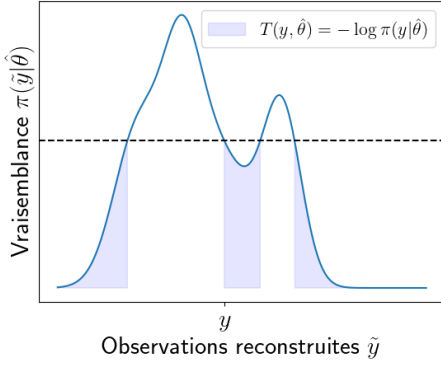


FIGURE 1 : Application de la valeur-p (6) pour un estimateur  $\hat{\theta}$  avec la statistique proposée (5) dans un cas de vraisemblance multimodale.

L'évaluation du résultat d'une inférence sur les paramètres  $\Theta$  nécessite de comparer les véritables observations  $\mathbf{Y}$  avec les observations reconstruites, notées  $\tilde{\mathbf{Y}}$ . Cette comparaison est effectuée avec la distribution prédictive a posteriori :

$$\pi(\tilde{\mathbf{Y}}|\mathbf{Y}) = \int \pi(\tilde{\mathbf{Y}}|\Theta) \pi(\Theta|\mathbf{Y}) d\Theta \quad (3)$$

Bien que nous favorisons ici une approche probabiliste, la distribution prédictive a posteriori peut être utilisée avec un estimateur  $\hat{\Theta}$ . Le posterior est alors remplacé par un dirac en  $\hat{\Theta}$ , et la distribution prédictive a posteriori réduite à la fonction de vraisemblance  $\pi(\tilde{\mathbf{Y}}|\hat{\Theta})$ .

En pratique, un calcul d'estimateur à partir de la distribution a posteriori correspond généralement à une intégrale incalculable, hormis pour des cas simples, par exemple pour un modèle direct linéaire et une distribution a priori conjuguée. Deux approches principales contournent cette difficulté : l'approche variationnelle [7] et l'approche par échantillonnage. Nous favorisons ici cette dernière, en utilisant des algorithmes de Monte-Carlo par chaîne de Markov (MCMC) [8].

## 2.2 Valeur-p bayésienne et choix de statistique

Le test statistique considéré est basé sur l'hypothèse nulle selon laquelle le modèle d'observation peut reconstruire les observations  $\mathbf{Y}$ . Les observations reconstruites par le modèle sont notées  $\tilde{\mathbf{Y}}$ . Si l'hypothèse nulle est vraie, alors les prédictions  $\tilde{\mathbf{Y}}$  devraient être aussi vraisemblables que les observations réelles  $\mathbf{Y}$ , au sens d'une statistique  $T : (\mathbf{Y}, \Theta) \mapsto T(\mathbf{Y}, \Theta)$ . Par définition, une valeur-p correspond à la probabilité d'obtenir des valeurs  $T(\tilde{\mathbf{Y}}, \Theta)$  plus étonnantes que  $T(\mathbf{Y}, \Theta)$ , c'est-à-dire (càd) à la mesure de l'ensemble  $I = \{(\tilde{\mathbf{Y}}, \Theta) \mid T(\tilde{\mathbf{Y}}, \Theta) \geq T(\mathbf{Y}, \Theta)\}$ , en supposant l'hypothèse nulle vraie. Si la valeur-p obtenue est inférieure à un seuil  $\alpha$  fixé au préalable (typiquement 0.05 ou 0.01), alors l'hypothèse nulle peut être rejetée avec un niveau de confiance de  $1 - \alpha$ .

Le test du  $\chi^2$  habituellement utilisé pour un estimateur  $\hat{\Theta}$  suppose un modèle direct  $\mathbf{f}$  linéaire et une vraisemblance  $\mathbf{Y}|\hat{\Theta} \sim \mathcal{N}(\mathbf{f}(\hat{\theta}_n)_{n=1}^N, \sigma^2 I_{NL})$ . La statistique de test considérée est alors

$$T(\mathbf{Y}, \hat{\Theta}) = \frac{1}{\sigma^2} \|\mathbf{Y} - \mathbf{f}(\hat{\theta}_n)_{n=1}^N\|_2^2, \quad (4)$$

qui suit une loi du  $\chi_{NL}^2$ . La valeur-p de ce test  $\mathbb{P}_{T(\tilde{\mathbf{Y}}) \sim \chi_{NL}^2} [T(\tilde{\mathbf{Y}}) \geq T(\mathbf{Y}) | \mathbf{Y}]$  est facilement calculable.

Pour les problèmes inverses basés sur des modèles directs non-linéaires ou un bruit non gaussien, une statistique de test menant à un calcul de valeur-p exact aussi simple n'existe pas. Nous proposons d'utiliser la neg log vraisemblance comme statistique de test, en nous inspirant des méthodes d'évaluation de modèles [3] :

$$T(\tilde{\mathbf{Y}}, \Theta) = -\log \pi(\tilde{\mathbf{Y}}|\Theta). \quad (5)$$

En particulier, dans le cas d'un estimateur ponctuel, d'un modèle direct linéaire et d'un bruit additif gaussien, la statistique (4) du test du  $\chi^2$  correspond, à une constante additive près, à la neg log vraisemblance. La Fig. 1 illustre cette statistique  $T$  pour une vraisemblance complexe et un estimateur  $\hat{\Theta}$ .

La valeur-p considérée peut être encore généralisée en remplaçant un estimateur  $\hat{\Theta}$  par la distribution a posteriori. La valeur-p bayésienne [2, chapitre 6] permet de considérer des incertitudes sur les paramètres  $\Theta$  en intégrant sur l'espace des paramètres selon la distribution prédictive a posteriori :

$$\begin{aligned} p &= \mathbb{P} [T(\tilde{\mathbf{Y}}, \Theta) \geq T(\mathbf{Y}, \Theta) | \mathbf{Y}] \\ &= \int \mathbb{1}_I(\tilde{\mathbf{Y}}, \Theta) \pi(\tilde{\mathbf{Y}}|\Theta) \pi(\Theta|\mathbf{Y}) d\Theta d\tilde{\mathbf{Y}}. \end{aligned} \quad (6)$$

Avec la statistique  $T$  de (5), la valeur-p (6) mesure l'ensemble des couples  $(\tilde{\mathbf{Y}}, \Theta)$  moins vraisemblables que les couples  $(\mathbf{Y}, \Theta)$ . L'hypothèse (1) permet de définir l'hypothèse nulle pour chaque observation individuelle  $\mathbf{y}_n$  avec un ensemble  $I_n = \{(\tilde{\mathbf{y}}_n, \Theta) \mid T(\tilde{\mathbf{y}}_n, \Theta) \geq T(\mathbf{y}_n, \Theta)\}$  associé. Une carte  $\mathbf{p} = (p_n)_{n=1}^N$  de valeurs-p est alors définie par

$$\begin{aligned} p_n &= \mathbb{P} [T(\tilde{\mathbf{y}}_n, \Theta) \geq T(\mathbf{y}_n, \Theta) | \mathbf{Y}] \\ &= \int \mathbb{1}_{I_n}(\tilde{\mathbf{y}}_n, \Theta) \pi(\tilde{\mathbf{y}}_n|\Theta) \pi(\Theta|\mathbf{Y}) d\Theta d\tilde{\mathbf{y}}_n. \end{aligned} \quad (7)$$

Cette définition est générale en ce qu'elle ne requiert aucune hypothèse particulière sur la forme du problème.

## 2.3 Calcul de valeur-p et incertitudes

Les valeurs-p  $\mathbf{p}^{(t)} = (p_n^{(t)})_{n=1}^N$  peuvent être évaluées numériquement pendant la phase stationnaire d'un échantillonnage de la distribution a posteriori avec un algorithme MCMC, càd après la phase de burn-in de durée  $T_{\text{BI}} \geq 0$  :

$$\hat{p}_n^{(t)} = \frac{1}{t - T_{\text{BI}}} \sum_{t'=T_{\text{BI}}+1}^t \mathbb{1}_{I_n}(\tilde{\mathbf{y}}_n^{(t')}, \Theta^{(t')}). \quad (8)$$

L'Algorithme 1 réalise ce calcul avec une mise à jour (màj) en ligne des valeurs-p. Dans le cas d'un estimateur  $\hat{\Theta}$ , la chaîne de Markov est remplacée par  $\Theta^{(0)} = \dots = \Theta^{(T_{\text{MC}})} = \hat{\Theta}$ .

Tenir compte des incertitudes inhérentes à l'évaluation des valeurs-p avec des chaînes de Markov de taille finie rend le test plus robuste. Puisqu'une valeur-p est une probabilité, nous associons à chaque  $p_n$  une loi de probabilité sur  $[0, 1]$ . La probabilité de rejet associée au niveau de confiance  $\alpha$  correspond ainsi à la fonction de répartition évaluée en  $\alpha$ ,  $\mathbb{P}[p_n \leq \alpha]$ . En utilisant un seuil  $\delta \leq 0.5$ , si cette probabilité de rejet se situe dans  $[\delta, 1 - \delta]$ , alors plus d'échantillons sont nécessaires pour prendre une décision avec le niveau de confiance exigé. Le modèle est rejeté lorsque  $\mathbb{P}[p_n^{(t)} \leq \alpha] > 1 - \delta$  et ne l'est pas quand  $\mathbb{P}[p_n^{(t)} \leq \alpha] < \delta$ .

---

**Algorithme 1** : Algo. MCMC avec calcul de valeur-p
 

---

**Entrée** : Paramètre de départ  $\Theta^{(T_{\text{BI}})}$ , noyau  $\mathcal{K}$ ,

Durées  $T_{\text{BI}}$  et  $T_{\text{MC}}$

**Init.** : carte de valeurs-p  $\hat{p}^{(T_{\text{BI}})} = \mathbf{0}_N$ ,  $\mathbf{u}^{(T_{\text{BI}})} = \mathbf{0}_N$

/\* Après burn-in \*/

**pour**  $t = T_{\text{BI}} + 1, \dots, T_{\text{MC}}$  **faire**

/\* échantillonnage de  $\mathcal{K}$  et calcul de vraisemblance \*/

$\Theta^{(t)}, \left(\pi(\mathbf{y}_n | \Theta^{(t)})\right)_{n=1}^N = \mathcal{K}(\Theta^{(t-1)})$

/\* m à j des valeurs-p \*/

$\tilde{\mathbf{Y}}^{(t)} \sim \pi(\cdot | \Theta^{(t)})$

$u_n^{(t)} = u_n^{(t-1)} + \mathbb{1}_{I_n}(\tilde{\mathbf{y}}_n^{(t)}, \Theta^{(t)})$  pour  $n \in \llbracket 1, N \rrbracket$

**fin**

$\hat{p}^{(T_{\text{MC}})} = \frac{1}{T_{\text{MC}} - T_{\text{BI}}} \mathbf{u}^{(T_{\text{MC}})}$

**Sortie** :  $\{\Theta^{(t)}\}_{t=T_{\text{BI}}+1}^{T_{\text{MC}}}$ , carte de valeurs-p  $\hat{p}^{(T_{\text{MC}})}$

---

Pour quantifier l'incertitude sur le calcul de la valeur-p avec (8), nous proposons de le considérer comme un échantillon d'une distribution binomiale de paramètres  $t - T_{\text{BI}}$  et  $p_n$ . La loi conjuguée est la loi Beta. En choisissant une distribution a priori Beta(1, 1), càd une loi uniforme sur  $[0, 1]$ , pour tout  $t > T_{\text{BI}}$ , on obtient pour  $p_n$  une distribution a posteriori Beta avec des paramètres explicites :

$$p_n^{(t)} \sim \text{Beta}\left(1 + N_n^{(t)} \hat{p}_n^{(t)}, 1 + N_n^{(t)} (1 - \hat{p}_n^{(t)})\right), \quad (9)$$

avec  $N_n^{(t)} = t - T_{\text{BI}}$  lorsque les échantillons sont supposés indépendants. Cette hypothèse est satisfaite si l'on considère un estimateur  $\hat{\Theta}$ . Elle n'est en revanche pas vérifiée pour une chaîne de Markov sur les  $\Theta$  dont les éléments sont corrélés. La distribution a posteriori sur  $p_n$  aura dans ce cas tendance à sous-estimer les incertitudes. Pour compenser cette corrélation, nous utilisons  $N_n^{(t)} = \text{ESS}^{(t)}$ , avec l'ESS la taille d'échantillon effective (*effective sample size*) [9] sur  $\Theta$ . Cette approche permet de tenir compte des erreurs associées à l'évaluation numérique de la valeur-p, ce qui rend le test plus robuste.

## 3 Applications

### 3.1 Gaussienne bi-dimensionnelle

L'équivalence de la valeur-p (6) et d'un test du  $\chi^2$  a été établie dans le cas d'un estimateur ponctuel, d'un modèle direct linéaire et d'un bruit additif gaussien. Nous considérons un problème inverse tel que  $N = 1$  et  $L = 2$ , avec une observation  $\mathbf{y} \in \mathbb{R}^2$ . Le bruit choisi est gaussien à variance unitaire. Le modèle direct est l'identité  $\mathbf{f} = \text{id}_2$ . Nous considérons trois estimateurs  $\hat{\theta}^{(i)} \in \mathbb{R}^2$  tels que les distances  $\|\mathbf{y} - \hat{\theta}^{(i)}\|$  valent 2.9, 2.4 et 2.1, respectivement. Les niveaux de confiance  $\alpha$  et  $\delta$  sont respectivement fixés à 0.05 et 0.1. La Fig. 2 illustre la convergences des valeurs-p estimées avec (8) et issues de (9) vers les valeurs théoriques. Avec  $T_{\text{MC}} = 10^3$  échantillons indépendants tirés de la posterior prédictive  $\pi(\tilde{\mathbf{y}} | \hat{\theta}^{(i)})$ , le calcul (8) amène à un rejet de  $\hat{\theta}^{(1)}$  et de  $\hat{\theta}^{(2)}$  et à un non-rejet de  $\hat{\theta}^{(3)}$ . Les

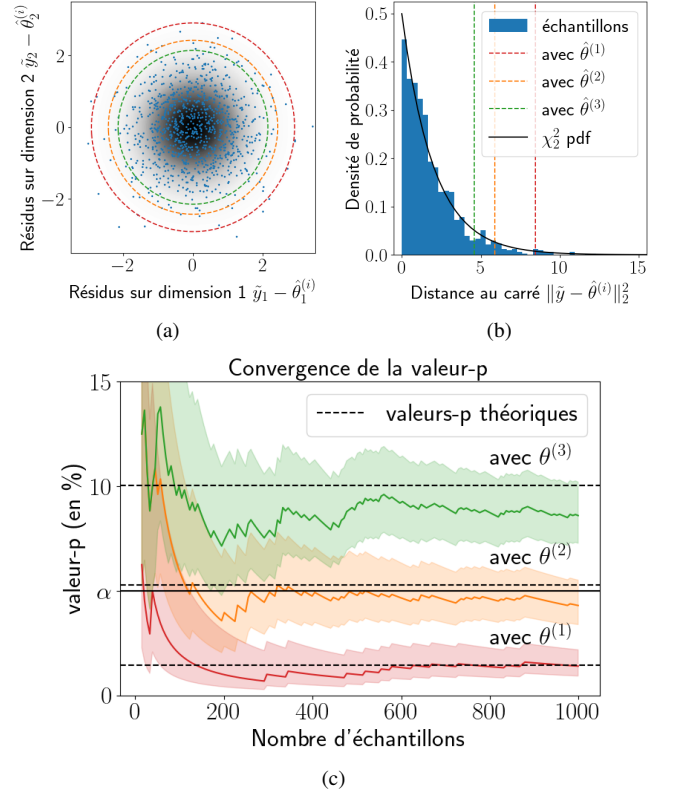


FIGURE 2 : (a) Fonction de densité des résidus  $\tilde{\mathbf{y}} - \hat{\theta}$  (en niveaux de gris) et  $T_{\text{MC}} = 1000$  échantillons indépendants  $\tilde{\mathbf{y}} \sim \pi(\tilde{\mathbf{y}} | \hat{\theta}^{(i)})$ . Les lignes de niveaux correspondent aux distances  $\|\mathbf{y} - \hat{\theta}^{(i)}\|$  associées aux estimateurs. (b) Distribution des  $\|\tilde{\mathbf{y}} - \hat{\theta}\|_2^2$ , suivant une loi  $\chi_L^2$ . Les barres verticales correspondent aux lignes de niveaux de la Fig. (a). (c) Convergence de l'estimateur (8) de la valeur-p et des intervalles de crédibilité à 90% issus de (9).

décisions prises avec la valeur-p calculée sont cohérentes avec celles des valeurs-p théoriques pour  $\hat{\theta}^{(1)}$  et  $\hat{\theta}^{(3)}$ . Pour  $\hat{\theta}^{(2)}$  en revanche, le rejet est à tort : la valeur-p théorique est supérieure au seuil  $\alpha$ . Le rejet est ici dû aux incertitudes inhérentes au calcul numérique de la valeur-p. La quantification d'incertitudes avec (9) fournit une probabilité de rejet  $\mathbb{P}[p_n^{(t)} \leq \alpha] = 0.83$ . Puisque  $\mathbb{P}[p_n^{(t)} \leq \alpha] \in [\delta, 1 - \delta]$ , cette insuffisance est identifiée et la situation est considérée comme indéterminée. Avec  $T_{\text{MC}} = 10^4$ , on obtient  $\mathbb{P}[p_n^{(t)} \leq \alpha] = 0.096 < \delta$ . L'indétermination est levée et  $\hat{\theta}^{(2)}$  n'est pas rejeté, ce qui est cohérent avec la valeur-p théorique correspondante. Le test est donc robuste à l'usage d'une méthode Monte Carlo pour calculer les  $p_n$ .

### 3.2 Problème inverse astrophysique

L'approche est appliquée à un problème astrophysique synthétique décrit dans [6]. La carte des paramètres à estimer  $\Theta$  contient  $N = 10 \times 10$  vecteurs de paramètres  $\theta_n \in \mathbb{R}^D$  avec  $D = 4$ . Le modèle direct  $\mathbf{f}$  est un réseau de neurones approchant un code astrophysique, le code Meudon PDR [10], fortement non-linéaire [11]. Dans ce cas particulier,  $\mathbf{f}$  calcule un  $\mathbf{y}_n$  contenant  $L = 10$  observables uniquement à partir de

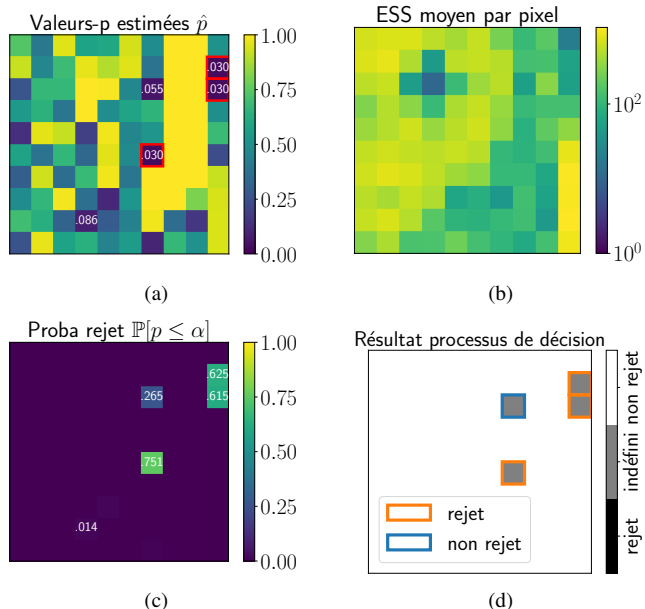


FIGURE 3 : (a) Carte de valeurs-p calculée avec (8). Les 8 valeurs les plus basses sont indiquées. Les pixels encadrés sont sous le seuil  $\alpha = 0.05$ . (b) Carte de l’ESS par pixel. (c) Probabilité de rejet de l’hypothèse nulle, calculée avec (9). Les 8 valeurs les plus hautes sont indiquées. (d) Décision prise avec l’approche bayésienne : aucun pixel n’est rejeté, mais 4 cas sont indéterminés. Les rectangles indiquent la décision correspondant à une chaîne plus longue.

$\theta_n$ . Il est donc ici possible d’utiliser des  $N_n^{(t)} = \text{ESS}_n^{(t)}$  différents pour chaque pixel. Le modèle d’observation combine censure, bruit additif gaussien et bruit multiplicatif log-normal. La neg log vraisemblance est donc complexe et ses valeurs sont difficilement interprétables, ce qui rend les méthodes d’évaluation de modèle [2, chapitre 7] peu adéquates. L’Algorithme 1 est intégré à l’échantillonneur proposé dans [6].

La Fig. 3 illustre l’analyse pour une chaîne de  $T_{\text{MC}} = 5 \times 10^3$  échantillons, dont  $T_{\text{BI}} = 10^3$  de burn-in. Le modèle utilisé pour générer les données et réaliser les inversions est identique. On considère un niveau de confiance  $\alpha = 0.05$  et un seuil sur les probabilités de rejet  $\delta = 0.2$ . La probabilité de rejeter l’hypothèse nulle alors qu’elle est vraie est donc de 5%. La régularisation spatiale décrite dans [6] est ici faible pour pouvoir considérer les pixels grossièrement indépendants. Avec l’estimateur (8), 3 pixels sur les 100 sont rejetés, et 2 autres sont proches de ce seuil. Quatre de ces valeurs-p (en haut à droite) correspondent à des ESS faibles. Les incertitudes associées sont donc grandes. Quand elles sont prises en compte avec (9), ces 4 pixels requièrent une chaîne de Markov avec plus d’échantillons. Une chaîne de  $T_{\text{MC}} = 10^5$  échantillons lève l’indétermination. Le modèle est alors rejeté pour 3 pixels. Le test présenté est donc cohérent avec les propriétés statistiques générales de test d’hypothèse et applicable pour des vraisemblances complexes.

## 4 Conclusion

Nous avons proposé une méthode d’analyse de résultats de problèmes inverses permettant d’identifier des cas où obser-

vations reconstruites et réelles sont incompatibles, ce qui est suffisant pour rejeter un modèle. Cette méthode est basée sur un test d’hypothèse par valeur-p bayésienne. Elle tient compte des incertitudes liées au modèle ainsi qu’au calcul de valeur-p à partir d’échantillons. Elle peut être appliquée dans une approche bayésienne ou par optimisation, quel que soit le modèle d’observation. L’approche proposée, simple à implémenter et à interpréter, ne nécessite aucune hypothèse sur le problème considéré.

## Remerciements

Ce travail est soutenu par un projet 80IPrime du CNRS, et par le Programme National « Physique et Chimie du Milieu Interstellaire » (PCMI) du CNRS/INSU, INC/INP, co-financé par le CEA et le CNES. Nous remercions également la Chaire IA Sherlock ANR-20-CHIA-0031-01 portée par P. Chainais, ainsi que par le programme national d’investissement d’avenir ANR-16-IDEX-0004 ULNE et la Région Hauts-de-France.

## Références

- [1] N. FRIEL et J. WYSE, “Estimating the evidence - a review”, *Statistica Neerlandica*, t. 66, n° 3, p. 288-308, 2012.
- [2] A. GELMAN, J. B. CARLIN, H. S. STERN et al., *Bayesian Data Analysis*, 3<sup>e</sup> éd. New York : Chapman et Hall/CRC, juill. 2015.
- [3] A. VEHTARI, A. GELMAN et J. GABRY, “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”, *Stat Comput*, t. 27, n° 5, p. 1413-1432, sept. 2017, arXiv :1507.04544 [stat].
- [4] R. WU, E. BRON, T. ONAKA et al., “Constraining physical conditions for the PDR of Trumpler 14 in the Carina Nebula”, *A&A*, t. 618, A53, oct. 2018.
- [5] R. ANDRAE, T. SCHULZE-HARTUNG et P. MELCHIOR, *Dos and don’ts of reduced chi-squared*, arXiv :1012.3754 [astro-ph, physics :physics, stat], déc. 2010.
- [6] P. PALUD, P.-A. THOUVENIN, P. CHAINAIS et al., *Efficient sampling of non log-concave posterior distributions with mixture of noises*, arXiv :2211.12915 [stat], nov. 2022.
- [7] D. M. BLEI, A. KUCUKELBIR et J. D. MCAULIFFE, “Variational Inference : A Review for Statisticians”, *Journal of the American Statistical Association*, t. 112, n° 518, p. 859-877, avr. 2017.
- [8] M. A. PEREYRA, P. SCHNITER, E. CHOZENOUX et al., “A Survey of Stochastic Simulation and Optimization Methods in Signal Processing”, *IEEE Journal of Selected Topics in Signal Processing*, t. vol. 10, n° n° 2, pp. 224-241, nov. 2015, Publisher : IEEE.
- [9] J. S. LIU, *Monte Carlo Strategies in Scientific Computing* (Springer Series in Statistics). New York, NY : Springer New York, 2004.
- [10] F. LE PETIT, C. NEHME, J. LE BOURLOT et al., “A Model for Atomic and Molecular Interstellar Gas : The Meudon PDR Code”, *The Astrophys. J. Supp. Series*, t. 164, n° 2, p. 506-529, juin 2006.
- [11] L. EINIG, P. PALUD, J. CHANUSSOT et al., “Réduction de modèles multi-physiques par réseaux de neurones”, Soumis au GRETSI 2023.