

Analyse de données bio-médicales

Recherche reproductible et modélisation mathématique

Li-Thiao-Té Sébastien

Concours CRCN INRAE

Plan

Parcours scientifique

Travaux

Intégration

Parcours scientifique

- 2005-2009 : Doctorat en Mathématiques Appliquées, ENS Cachan et Institut Pasteur.
Caractérisation et Détection de Signaux Protéiques en Spectrométrie de Masse
- 2009-2010 : Post-doctorat dans l'Équipe Statistique et Génome, UMR 518 AgroParisTech/INRA
Estimation du nombre d'espèces en métagénomique
- 2010- : Maître de conférences en mathématiques appliquées, LAGA UMR 7539, Université Paris 13

Parcours scientifique

2010- : Maître de conférences en mathématiques appliquées,
LAGA UMR 7539, Université Paris 13

- 2014-2018 : traitement d'images biomédicales avec F Dibos, M Luong, JM Rocchisani, DV Tran (doctorant)
- 2014-2018 : traitement du signal en spectroscopie RPE avec Y Frapart, DN Tran (doctorant)
- 2018-2022 : traitement d'images et modèles EDP pour les maladies inflammatoires intestinales avec H Zaag, J Chaussard, X Treton, E Ogier-Denis, S Al Ali (doctorante)
- nov 2021 : Habilitation à Diriger des Recherches Analyse de données bio-médicales. Recherche reproductible et modélisation mathématique.

Parcours scientifique

Responsabilités :

- 2014-2018 : comité de direction du programme interdisciplinaire Imageries du Vivant (800 000 euros de financement, Pres SPC, Universités P5, P7 et P13)
- depuis nov 2018 : co-responsable avec H Zaag de l'équipe Mathématiques pour la Biologie et les Images
- depuis juin 2021 : membre du Comité national pour la Science Ouverte (collège publications)

Plan

Parcours scientifique

Travaux

Intégration

2. Recherche reproductible

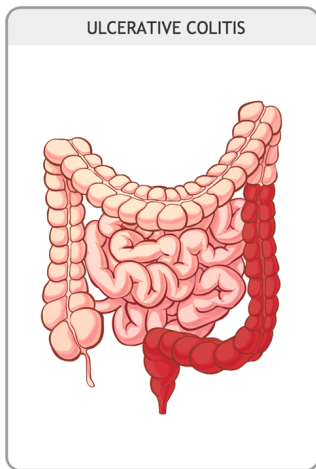
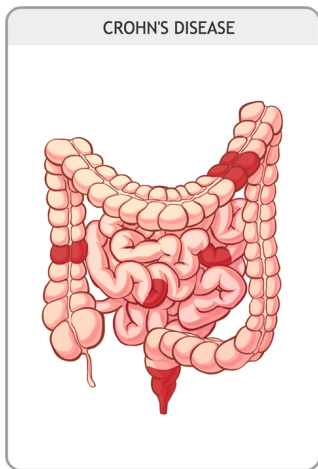
Ensemble d'outils et de méthodes de travail :

- clarté et rigueur scientifique
- collaborations interdisciplinaires
- valorisation et transfert vers les applications

Mise en oeuvre concrète :

- Plateforme RPE : Matlab + VM du Cloud IDV + GitHub
- Hôpital Bichat : Python + Docker + PlmLab

Maladies Inflammatoires Chroniques de l'Intestin



Collaboration avec H Zaag, J Chaussard, X Treton (H Bichat-Beaujon), E Ogier-Denis (Inserm)

Coloscopie



Projet de thèse de S Al-Ali

- modifications du protocole médical
- analyse d'image : détection des saignements et des ulcères
- modélisation de la répartition des lésions par EDP
- sévérité de la maladie = vitesse de propagation

Analyse d'images

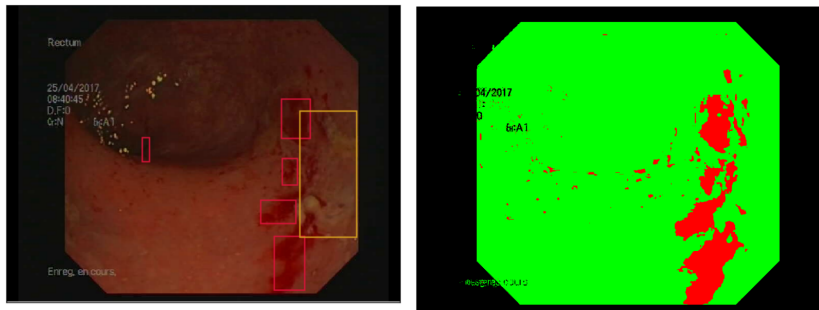


Figure – Détection des saignements

Analyse d'images

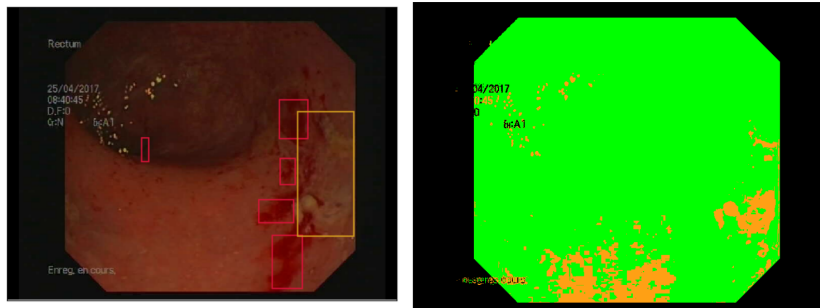


Figure – Détection des ulcères

Localisation des lésions

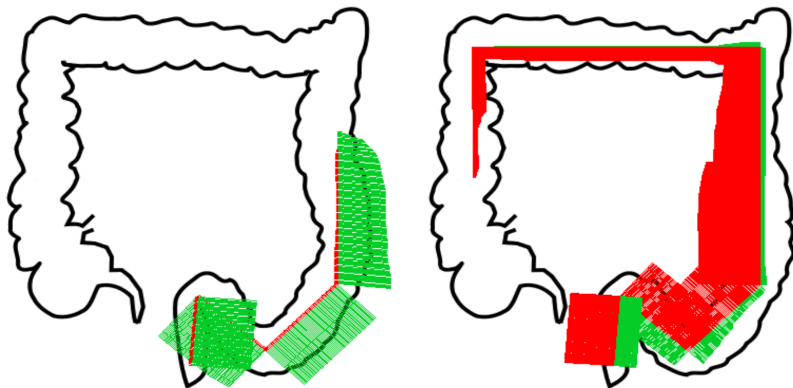


Figure – Répartition des lésions chez deux patients de même score UCEIS=5

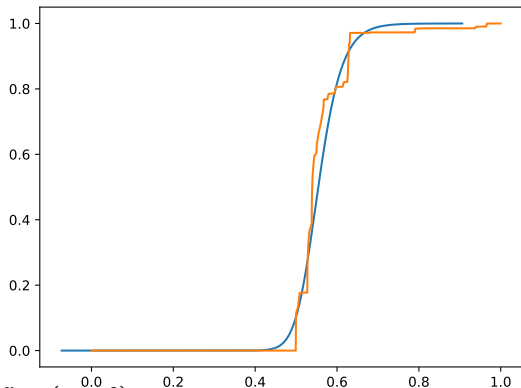
Modélisation par EDP

Équation de Fisher-Kolmogorov-Petrovski-Puskinov

$$\frac{\partial}{\partial t} u - D \frac{\partial^2}{\partial s^2} u = u(1 - u), \quad t \geq 0, s \in \mathbb{R} \quad (1)$$

- fronts d'onde à vitesse $2\sqrt{D}$
- recalage pour le problème inverse (dilatation + translation)
- vitesse et temps d'invasion

Modélisation par EDP



Code chunk 1: «fkpp (part 2)»

```
imin = np.argmin([fkpp_interp(p,d) for d in D])
_ = fkpp_interp(p,D[imin], 'fkpp-recalage.pdf')
v = 2 * np.sqrt(D[imin])
t = (1-np.trapz(p,dx=Xp[1]-Xp[0])) / v
print('Vitesse ', np.round(v,3), 'Temps d\'invasion', np.round(t,2))
```

Interpret with python3

```
Vitesse 0.028 Temps d'invasion 19.99
```

Plan

Parcours scientifique

Travaux

Intégration

Projet WHAT SOW



5 postures :

- assis
- debout
- coucher ventral
- coucher latéral droit
- coucher latéral gauche

- [1] Laurianne CANARIO et al. "Fusion de données vidéo et d'accéléromètre pour la détection automatique de l'activité posturale des truies en case de mise bas". In : [Journées de la Recherche Porcine 53 \(2021\)](#), p. 111-112.
- [2] Mathieu BONNEAU et al. "Predicting sow postures from video images : Comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups". In : [Biosystems Engineering 212 \(2021\)](#), p. 19-29.

1. Conception des expériences

- c'est un travail d'analyse de données
- qui fournit des invariants : translation, rotation pour la réduction de dimension, éviter le sur-apprentissage

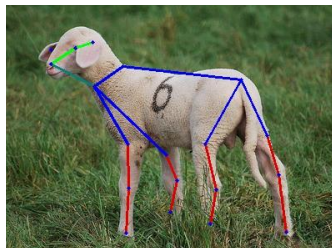
Proposition de projets

- instrumentation : technologie des capteurs
- nombre de sujets avec ses contraintes pratiques, réglementaires, éthiques
- optimisation du protocole expérimental
positionnement des capteurs, de la caméra
taux d'échantillonnage, observation sur une fenêtre.

2. Analyse d'images et traitement du signal

Vision par ordinateur :

- reconnaissance de poses / gestes
- reconnaissance des émotions

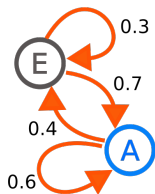


Proposition de projets

- segmentation
- détection d'objet pour la tête et les membres
- recalage sur template déformables
- utilisation de motifs tracés sur le harnais ou le dos

3. Modélisation et analyse de données

Modélisation du comportement par une chaîne de Markov



- transitions entre postures
- l'accéléromètre indique des changements de posture
- le flux optique dans la vidéo aussi

Ensuite : relier les paramètres de la chaîne de Markov au bien-être ou au stress de l'animal.

Travail avec P Gloaguen : variables latentes, trajectoires et comportement

Autoencodeurs variationnels

Intérêt partagé :

- GIBBS : workshops en 2018 et 2021
- SOLSTIS : Groupes de travail "State of the R" et "Autoencodeurs variationnels"
- + intérêt personnel

Pour le projet WHAT SOW : bien-être en fonction

- du comportement (chaîne de Markov)
- des données brutes
- des autres variables omiques

Questions méthodologiques

- sélection de variables : noeuds stables par les invariants ?
- ordre des exemples dans l'apprentissage, réduction de variance, échantillonnage d'importance ?

Merci pour votre attention

Postdoctorat AgroParisTech/INRA

Pour une espèce d'abondance λ_i , le nombre de fragments observés est une variable aléatoire de loi de Poisson

$$f(X_i|\lambda_i) = \frac{\exp^{-\lambda_i} \lambda_i^{X_i}}{X_i!}$$

La répartition des espèces, c'est-à-dire la distribution des λ_i est un mélange de lois exponentielles

$$f(\lambda) = \sum_q \alpha_q f_q(\lambda) \quad \text{où} \quad f_q(\lambda) = \mu_q e^{-\mu_q \lambda}$$

A priori :

$$\alpha \sim \text{Dirichlet}(\vec{a})$$

$$\pi_q \sim \text{Beta}(b_q, c_q)$$

$$Z \sim \text{Multinom}(\alpha)$$

$$X|Z \sim \text{Geom}(\pi_q)$$

Postdoctorat AgroParisTech/INRA

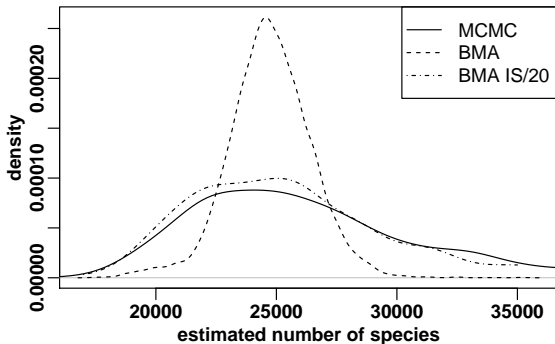


Figure – Distribution a posteriori du nombre d'espèces.

Thèse de Tran DV, imagerie biomédicale

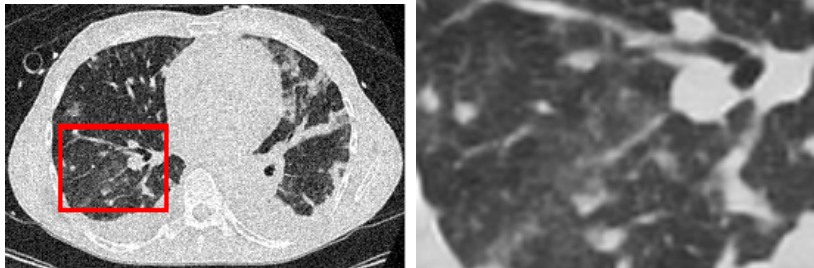


Figure – Débruitage et super-résolution d'une image de scanner du thorax.

Thèse de Tran DV, imagerie biomédicale

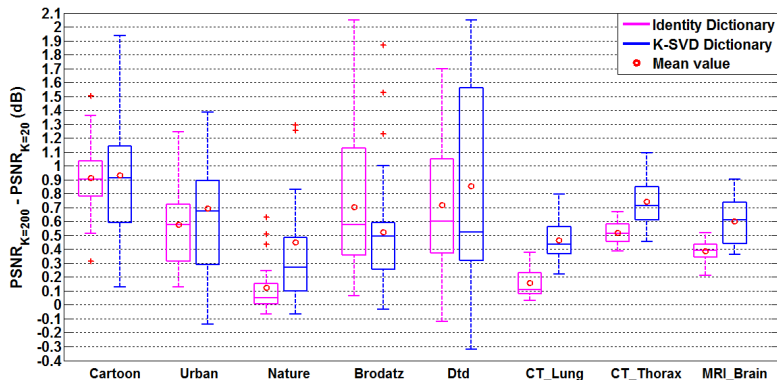


Figure – Gain de performance avec un dictionnaire plus grand.

Thèse de Tran DV, imagerie biomédicale

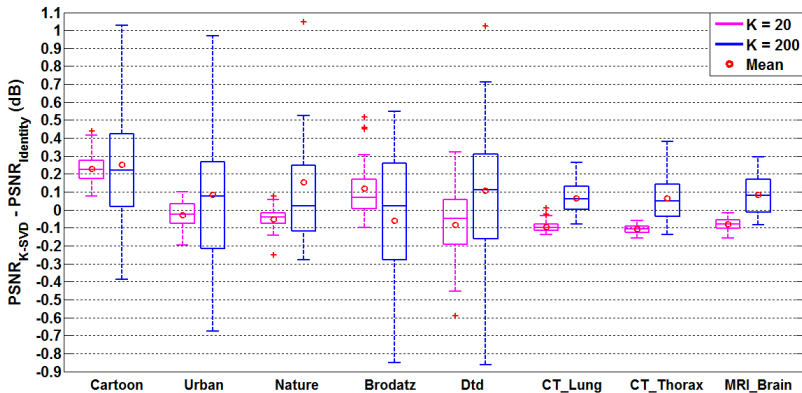


Figure – Gain de performance suivant la méthode de construction du dictionnaire.

Thèse de Tran DN, spectroscopie RPE

