

Livret de rapports

Ibrahim SEYDI

Contents

1 Ier Trimestre – Mai, Juin, Juillet, Aout 2025	3
--	----------

Ier trimestre

CHAPTER 1

Ier Trimestre – Mai, Juin, Juillet, Aout 2025

Rapport de la semaine du 12/05/2025

Travaux/Tâches effectués

- Récupération ordinateur au service IT
- Réunion Adel
- Progression inscription ADUM
- Préparation et pitch thèse + séminaire technique risque sismique
- Création Github privé (+ permet également d'archiver les articles, accessible par tous)
- Lecture article PSHA classique
- Début de rédaction de résumé critique
- Recherche bibliographique sur les méthodes zoneless

Problèmes rencontrés / Questions en suspens

- Copie de la convention de cotutelle SIGNÉE
- Copie de l'attestation d'assurance responsabilité civile
- Connexion à distance avec ordi EDF
- Config Authenticator

Idées surgies

- Piste axe de recherche sur "amélioration des modèles de récurrence", tentative de modélisation bayésienne (priors informatifs, ...) ; Etude de la loi de GR sur la magnitude (le paramètre qui semble supplanter les autres)
- Piste méthodes zoneless (cf recherche biblio)

Lectures et ressources

- Titre : *A probabilistic seismic hazard map for the metropolitan France*

Type : Article

Lien ou référence : Dans Github

Notes : Article de référence sur l'approche standard PSHA ; Gloria Senfaute co-autrice

Étapes suivantes

- Finaliser inscription ADUM
- Finir analyse critique écrit
- Fixer axe de recherche à suivre
- Remplir docs pour forfait jour et envoyer à Bruno pour signature
- Continuer à suivre travail de Adel sur processus de Hawkes
- Se faire une culture sismo (docs Gloria)
- Formaliser mathématiquement (chap 2/3 du papier de l'approche classique PSHA),
- Deadline(s) à venir : ...

Notes libres

- Manque de productivité vendredi, un peu de fatigue, quelques lacunes sur le sommeil (à améliorer)
- ...

Rapport de la semaine du 19/05/2025

Travaux/Tâches effectués

- Réunion Adel
- Réunion avec Pierre ARROUCAU et Guillaume DANIEL pour discuter de l'orientation de la thèse et incorporation de la thèse dans SIGMA 3/WP1 (livrables à fournir et réunions avec eux tous les 2/3 mois)
- Accès à SoftID
- Premier jour à APT, réunion avec Sophie et Julien : formalisme du problème + mini cours BNP
- Formalisme modèle
- Test du KDE sur un exemple simple (carré unitaire), sans recherche d'optimisation
- Recherche bibliographique (dossier avec différents articles, thèses, cours)
- Lecture document écrit par étudiant (cf mail Sophie 'Ressources')
- Pseudo code Stick Breacking

Problèmes rencontrés / Questions en suspens

- Pas de connexion dans bureau APT donc partage compliqué + pas étudiant donc pas de compte Eduroam

Idées à explorer

- Pistes pour faire de l'inférence : Chinese Restaurant Process + Gibbs ; VI ?
- Eduroam coté EDF

Lectures et ressources

- Titre : *Density Estimation Using Nonparametric Bayesian Methods*
Type : Article
Lien ou référence : https://openscholarship.wustl.edu/art_sci_etds/1507/
Notes : Rapport réalisé par un étudiant
- Titre : *Bayesian Nonparametrics*
Type : Livre
Lien ou référence : ...
Notes : Prêter par Sophie

Étapes suivantes

- Améliorer le KDE (choix de la fenêtre)
- Implémenter Stick Breacking (soit proposer une évaluation du choix de K, soit fixer un seuil pour la taille du bâton, à partir duquel on sort de la boucle (utiliser plutôt un while))
- Rédaction fiche Bayésien NP
- Clarifier et poser sur papier, pistes pour faire de l'inférence
- Lire biblio
- Finir analyse papier PSHA

Notes pour moi-même

- Avancer progressivement, ne pas hésiter à commencer doucement avec cas jouet ou exemples simples (à l'instar de TS)
 - Echanger avec collaborateurs dès que doute ou besoin (à l'instar de TS)
 - Jeudi 29 férié
-

Rapport de la semaine du 26/05/2025

Travaux/Tâches effectués

- Re-travail du KDE spatial (gaussiennes dans les quatre coins + CV pour avoir fenêtre optimal)
- Lecture articles
- Début rédaction fiche Bayésien NP
- Ecriture Stick-Breacking avec seuil τ et début implémentation DP par Stick-Breacking

Problèmes rencontrés / Questions en suspens

- Eclaircissement sur le pk de l'étirement du KDE quand les localisations des points étaient générées aléatoirement ?
- Cross Valid avec OpenTURNS ?
- Pb avec fonction Beta de OpenTURNS

Idées à explorer

- Modèles de diffusion/SBM (cf article envoyé par Merlin, essayer de voir dans quelle mesure c'est exploitable (idée pour un peu plus tard encore))

Lectures et ressources

- Titre : *Bayesian nonparametrics and the probabilistic approach to modelling*
Type : Article
Auteur : Zoubin Ghahramani
Notes : Auteur a pas mal bossé sur Bayésien non paramétrique
- Titre : *dirichletprocess: An R Package for Fitting Complex Bayesian Nonparametric Models*
Type : Article
Lien ou référence : ...
Auteur : Gordon J. Ross
Notes : Auteur qui a pas mal étudié sur Hawkes et DP
- Titre : *Dirichlet Process*
Type : Cours
Auteur :
Notes :

- Titre : *Lecture 23: Bayesian Nonparametrics: Dirichlet Processes*
Type : Article
Auteur : Eric P. Xing
Notes : Auteur a pas mal bossé sur DP
- Titre : *Nonparametric Bayesian methods: a gentle introduction and overview*
Type : Article
Auteur : Steven N. MacEachern
Notes : Reprendre partie 2.4 et 2.6 (et p-e 3) pour fiche BNP
- Titre : *Non-parametric Bayesian Methods*
Type : Cours
Auteur : Zoubin Ghahramani
Notes : Excellent !

Étapes suivantes

- Finir implémentation Stick-Breaking (avec NIW)
- Finir Fiche Bayésien NP (partie)
- Lire articles sur Inférence et commencer à faire
- Rédiger proprement inférence dans Fiche Bayésien NP
- Régler ptits pbs sur KDE

Notes à moi-même

- Venue si possible à Agro mercredi

Rapport des semaines 02/06/2025, 09/06/2025 et 16/06/2025

Travaux/Tâches effectués

- Régler ptits pbs sur KDE/SB
- Avancement fiche Bayésien NP
- Lecture articles
- Travail sur première approche pour faire de l'inférence (Neal 2000)
- Travail avec Sophie sur comment intégrer un prior informatif (Formalisme d'une piste)
- Séminaire journeé utilisateurs OT
- Code DPMM avec prior informatif
- Avancement analyse critique du papier standard PSHA
- Rdv avec Sismologues Pierre ARROUCAU et Guillaume DANIEL
- Étude des caractéristiques d'un prior NIW

Problèmes rencontrés / Questions en suspens

- Lien entre densité d'une distribution d'évènements et intensité d'un processus Poissonien ? (cf mail Sophie)

Lectures et ressources

- Titre : *Markov Chain Sampling Methods for Dirichlet Process Mixture Models*
Type : Article
Auteur : Radford M. Neal
Notes : Article fondamental pour inférence (Gibbs)
- Titre : *Nonparametric Bayesian methods : a gentle introduction and overview*
Type : Article
Auteur : Steven N. MacEachern
Notes : pour fiche BNP

Étapes suivantes

- Finir analyse critique sur papier approche standard
- Faire apparaître des lignes de niveaux sur simu
- Subplots pour comparaisons entre lignes de niveaux et f_0 (catalogue)
- Calcul distances entre \tilde{f}_0 et f et entre f et f_0 (en fonction de α , et des hyperparamètres du NIW)
- Simu avec prior non informatif et comparer en fonction de α
- Colormap/Heatmap de f , f_0 , \tilde{f}_0
- Extension de la piste actuelle à un cadre plus complexe
- Deadline à venir : 18/06/25 - Réunion avec P.ARROUCAU et G.DANIEL
- Deadline à venir : 27/06/25 - Abstract travaux pour Luis et Sismologues

Notes pour moi-même

- Inclure Anne Dutfoy dans boucle des personnes intéressées par la thèse
-

Rapport de la semaine du 23/06/2025

Travaux/Tâches effectués

- Faire apparaître des lignes de niveaux sur simu DPMM
- Subplots pour comparaisons entre lignes de niveaux et f_0 (catalogue)
- Calcul distances entre \tilde{f}_0 et f et entre f et f_0 (en fonction de α)
- Colormap/Heatmap de f , f_0 , \tilde{f}_0
- Simu avec prior non informatif et comparer en fonction de α
- Biblio pour vérif que notre piste n'existe pas déjà dans la nature

Problèmes rencontrés / Questions en suspens / Idées à explorer

- Mauvais domaine de définition de la densité du DPMM. Deux pistes de solutions :
 1. Réécriture de la densité avec constante de normalisation (qui est compliquée). Constante peut être approchée ? Estimateur de MOnte-Carlo ? Constante de normalisation dépend des paramètres il me semble donc si utilisation d'une méthode itérative ça risque de poser pbs ? Méthodes SMC peut aider ici (permet de bosser avec des densités non normalisées, et peut donner des estimations de la constante de normalisation) ? Méthode pour tirer dans un domaine bornée ?
 2. Considérer la densité a posteriori traditionnelle et contraindre les paramètres du mélange (centres et covariances des gaussiennes) pour qu'ils soient dans une zone bornée. Dans inférence : pénaliser la vraisemblance du modèle pour être dans support d'intérêt ? ; Sorte de rejet lorsque l'on est pas dans support d'intérêt ?
2/ -> garde intacte la forme du posterior (conjugaison éventuelle)

Étapes suivantes

- Mettre au propre le code
- Modifier code pour avoir DPMM avec renormalisation des gaussiennes tronquées
- Regarder les marginales
- Finir analyse critique sur papier approche standard
- Écrire lien entre densité et intensité d'un processus (cf mail Sophie)
- Finir calcul de la conjugaison de la NIW avec Normale

- Extension de la piste actuelle à un cadre plus complexe
- Deadline à venir : 02/07/2025 - RDV avec toute la team rapprochée

Notes pour moi-même

- Biblio pour vérif que notre piste n'existe pas déjà dans la nature
-

Rapport des semaines 30/06/2025, 07/07/2025

Travaux/Tâches effectués

- Mise au propre du code
- Réunion avec la team rapprochée : Mise en clair des problématiques, enjeux, objectifs, pistes exploratoires
- Restructuration du code/github
- Avancement sur analyse papier PSHA standard
- Analyse du code de Julien
- Upgrade au plan premium d'Overleaf
- Création d'une carte stratégique mentale

Problèmes rencontrés / Questions en suspens / Idées à explorer

- Renormalisation du DPMM informatif sur $[0, 2]^2$, puis suppression de cette renormalisation (on s'en fiche pour l'instant)

Étapes suivantes

- Finir restructuration codes (faire des classes, ...)
- Faire Docstring à l'image de codes phebus
- Réaliser tout ce que je fais avec des zones irrégulières (from code phebus)

Notes pour moi-même

- Continuer à faire de la biblio pour vérif que piste n'existe pas dans la nature

Rapport des semaines 14/07/2025, 21/07/2025

Travaux/Tâches effectués

- Fin restructuration codes
- EM pour approximation du mélange d'uniformes par mélange de gaussiennes
- Écriture de la méthodo réalisée jusqu'ici
- Études contre model Merlin, discussions perspectives thèses avec Sophie
- Docstrings codes

Problèmes rencontrés / Questions en suspens / Idées à explorer

- ADUM (ENCORE ET TJR)

Étapes suivantes

- Appliquer méthode avec des zones irrégulières (from code phebus)
- Commencer inférence (Gibbs et VI)
- Lire autour de HDP et DDPM
- Comparaison KDE, DPMM standard, DPMM informatif
- Lire autour de GP pour estimation log densité (Logistic Gaussian Process/Cox)
- Finir analyse papier standard PSHA
- Modifier et finir Carte mentale
- Deadline(s) à venir : CA - 14/08/2025

Notes pour moi-même

- Continuer à faire de la biblio pour vérif que piste n'existe pas dans la nature

Rapport des semaines 28/07/2025, 04/08/2025, 08/09/2025

Travaux/Tâches effectués

- Application DPMM informatif sur exemple avec des zones irrégulières
- Re-travail du code
- Lecture autour de HDP
- Points avec Adel, lecture rapport de stage, commentaires sur rapport
- Avancement Fiche Bayésien NP
- Travail sur inférence du modèle informatif, ébauche -> inspiration de l'algo 2 de Neal 2000
- Avancement ADUM
- Fin de stage d'Adel : rapport + slides
- Formation OpenTurns (08/09/10 septembre)
- Début de lecture sur inférence du paramètre de concentration
- ...

Problèmes rencontrés / Questions en suspens / Idées à explorer

- ADUM (ENCORE ET TJR)
- Questionner Adel sur différents méthodes de simu des Hawkes (existence d'une méthode de simu horizontale (conceptuellement parlant) ?)

Lectures et ressources

- Titre : *Bayesian nonparametrics and the probabilistic approach to modelling*
Auteur : Zoubin Ghahramani
Type : Article
Lien ou référence : Dans Github
Notes : Cool (à revoir pour relever points intéressants pour la suite)
- Titre : *Hierarchical Dirichlet Processes*
Type : Article
Lien ou référence : <https://people.eecs.berkeley.edu/~jordan/papers/hierarchical-dp.pdf>
Notes : Papier fondateur HDP

- Titre : *Hierarchical Dirichlet Processes*
Type : Article
Lien ou référence : https://www.stat.berkeley.edu/~aldous/206-Exch/Papers/hierarchical_dirichlet.pdf
Notes : Papier technique (approche développée) fondateur HDP
- Titre : *Prior selection for the precision parameter of Dirichlet Process Mixtures*
Type : Article
Lien ou référence : Sur Github
Notes : Choix du α dans DPMM, paramètre de concentration ; Lecture à poursuivre ; Fait l'étalage des approches existantes avant proposition d'une nouvelle approche
- Titre : *Markov Chain Sampling Methods for Dirichlet Process Mixture Models*
Auteur : Radford M. Neal
Type : Article
Lien ou référence : Dans Github
Notes : Semble être référence imtemporelle pour Gibbs pour DPMM ; Pillier sur lequel repose Gibbs pour DPMM informatif

Étapes suivantes

- Finir rédaction sur HDP (CRF)
- Lire sur DDPM, rédiger idées autour
- Mettre de l'ordre dans l'organisation du github
- Implémentation inférence Gibbs pour DPMM informatif
- Lire autour de VI pour DPMM
- Modifier distance L^2 -> Utiliser méthode de calcul d'intégrale
- Lire autour du choix du paramètre de concentration
- Comparaison KDE, DPMM standard, DPMM informatif sur données simulées
- Lecture autour de DDPM
- Application DPMM standard à données réelles
- Lire autour de GP pour estimation log densité (Logistic Gaussian Process/Cox)
- Deadline(s) à venir : ... - ...

Notes pour moi-même

- Continuer à faire de la biblio pour vérif que piste n'existe pas dans la nature
 - Relire approche de discréétisation pour comprendre comment erreur à pu arriver
 - Suivi cours de Vincent Guigne, Deep Learning avec Pytorch
 - Suivre demande de justificatif de formation OpenTurns
 - ...
-