

# Curriculum Vitæ

## État civil et situation actuelle

Pierre VANDEKERKHOVE  
54 ans, né le 28 novembre 1969 à Melilla (Espagne)  
Français, marié, 1 enfant.

Membre du LAMA, UMR-CNRS 8050  
Université Gustave Eiffel,  
5 Boulevard Descartes,  
44420 Champs-sur-Marne, France.

Tel :+353 89 613 9603

Courriel : pierre.vandekerkhove@univ-eiffel.fr

Page web <http://perso-math.univ-mlv.fr/users/vandekerkhove.pierre>

**Cursus et diplômes** Maître de conférences hors classe (section 26 du CNU) depuis septembre 1998 au Laboratoire d'Analyse et de Mathématiques Appliquées (UMR-CNRS 8050) de l'université Paris-Est Marne-la-Vallée (renommée Gustave Eiffel).

## FORMATION ET TITRES

**2007** Habilitation à diriger des recherches de l'université Paris-Est Marne-la-Vallée.

Titre : *Contribution à l'étude statistique des modèles à données manquantes et apprentissage statistique autour des modèles markoviens.*

Rapporteurs : P. Bertail (Nanterre), B. G. Lindsay (Penn State university, USA), E. Moulines (ENST).

Jury : P. Bertail (Nanterre), J.-F. Delmas (ENPC), P. Del Moral (INRIA Bordeaux), E. Gassiat (Orsay), M. Hoffmann (Paris-Est), D. Lambertson (Paris-Est), E. Moulines (ENST).

**1997-98** Post-doc à l'université d'Économie de Pavie (Italie) dans le cadre du réseau européen Training and Mobility Researcher.

**1997** Doctorat en Statistique de l'université Montpellier II, sous la direction de X. Milhaud.

Titre : *Identification de l'ordre des processus ARMA stables et contribution à l'étude statistique des chaînes de Markov cachées.*

Rapporteurs : M. Dufflo (Marne-la-Vallée), E. Gassiat (Évry).

Jury : D. Bakry (Toulouse), A. Berlinet (Montpellier), D. Dacunha-Castelle (Orsay), G. Ducharme (Montpellier), M. Dufflo (Marne-la-Vallée), E. Gassiat (Évry), L. Miclo (Toulouse), X. Milhaud (Montpellier).

**1992** DEA de Biostatistique de l'université Montpellier II, sous la direction de X. Milhaud et J. H. Friedman (Stanford).

Titre : *Sensibilité de l'algorithme ACE (Alternating Conditional Expectation).*

**1991** Maîtrise de Mathématiques Appliquées, option Topologie Algébrique, de l'université Montpellier II.

## EXPÉRIENCES PROFESSIONNELLES

**2014** Consultant chez Predictix (startup Big Data, Atlanta, USA).

**2013** Lecturer au Georgia Institute of Technology (Atlanta, USA).

**2007-2009** Enseignant de Mathématiques à l'ESIEE.

- 2002-2006** Enseignant de Statistique à l'école des Ponts.
- 1996-97** ATER (demi-poste) à l'université Paul Sabatier de Toulouse.
- 1994-96** Allocataire-Moniteur à l'université Paul Sabatier de Toulouse.
- 1993-94** Allocataire-Moniteur à l'université Montpellier II.
- 1992** Stage de DEA de 3 mois effectué à l'INRA de Montpellier.

#### PRODUCTION SCIENTIFIQUE

- 28 publications.
- 2 articles soumis et 4 en cours de rédaction.
- 2 preprints non publiés.
- 6 conférences internationales sur invitation.
- 3 séminaires dans des laboratoires étrangers.
- 11 conférences nationales.
- 34 séminaires effectués dans diverses universités françaises depuis 1997.
- 1 exposé au CIRM accessible en vidéo sur Youtube  
<https://www.youtube.com/watch?v=BK9HUNachHs>.

#### PROJETS FINANCÉS

- Chaire d'excellence : ACTIONS (1.65 Mio Euro). Partenaires : BNP Paribas CARDIF, Institut des Actuaire et l'Institut Louis Bachelier.
- Projet IMPT : Order and Disorder in a Turbulent Ocean (73 k Euro). IMPT : Institut des Mathématiques pour la Planète Terre (INSMI, CNRS).

#### CONFÉRENCES ET SÉMINAIRES RÉCENTS

- MLISTRAL Conference, CIRM, Marseille, 27 Septembre 2022.
- Séminaire Parisien de Statistique, IHP Paris (France), 27 Juin 2022.
- Institut Mathématique de Marseille, Marseille (France), 31 Janvier 2022.
- MiMo : workshop on Mixture models, Rouen (France), 8-9 Avril 2021.
- ISI 63rd (virtual) World Statistics Congress, 13 Juillet 2021.
- Workshop on "Estimation of entropies and other functionals : Statistics meets information theory", Cambridge (UK), 9-11 Septembre 2019.
- Advances in Finite Mixtures and other Non-regular Models, Guillin (China), 12-16 Août 2018.

#### ORGANISATION DE CONFÉRENCES

Organisation de **MHC 2021** (Mixture, Hidden Markov Models, Clustering), en collaboration avec Elisabeth Gassiat (Paris Sud, Saclay) et Hajo Holzmann (Marburg, Allemagne). Dates : 2-4 juin 2021. Lieu : institut de Mathématique d'Orsay, amphithéâtre Yoccoz.

<https://www.imo.universite-paris-saclay.fr/mhc2020/>

- Journée SESO 2015, Statistiques pour les énergies renouvelables, fin juin 2015.
- Journée SESO 2014, Statistiques pour les énergies renouvelables, 27 juin 2014.
- Journée de rencontre Mathématique Labex Bézout GeorgiaTech, 18 décembre 2013.
- Journées de Statistique de Marne, 22-24 juin 2011.
- Journées internationales de Pau les 23-24 juin 2008.
- Journées de Statistique de Marne les 13-14 janvier 2005.

# Exposé des activités d'enseignement et de recherche

## RECONNAISSANCES ET DIFFUSION SCIENTIFIQUE

- Délégation CNRS en 2023-2024 (Gustave Eiffel).
- Délégation CNRS en 2012-2014 (GeorgiaTech, USA).
- CRCT de 6 mois en 2010-2011 (Paris-Est).
- Délégation CNRS en 2009-2010 (Télécom ParisTech).
- Délégation CNRS en 2004-2005 (Stanford).
- Membre de l'ANR BigMC entre 2009-2012.
- Prime d'Encadrement Doctoral et de Recherche entre 2006 et 2010.
- Membre extérieur du comité de sélection de l'UTC de Compiègne en 2011.
- Membre extérieur du comité de sélection d'Orléans en 2009.
- Membre extérieur de la commission de spécialistes de l'UTC entre 2005 et 2007.
- Vice président de la commission de spécialistes (collège B) de Marne-la-Vallée entre 2002 et 2006.
- Lecture ou arbitrage d'articles pour : *Ann. Statist.*; *Electron. J. Statist.*; *J. Non-param. Statist.*; *J. Statist. Planning and Inference*; *Scand. J. Statist.*; *Information and Inference*; *Biometrical Journal*; *Bernoulli*; *Stoch. Process. Appl.*; *ESAIM P&S*; *Statistica Sinica*; *Statistics, CRAS Paris*; *Maghreb Math. Rev.*

## CO-ENCADREMENT (POST)-DOCTORAL (2012-18)

- Rodrigue Nguyepe Zoumpe : étudiant en thèse au ISyE (Nicoletta Serban, Georgia-Tech).
- Jagan Paddidri : post-doc au dépt. Material Science (David McDowell, GeorgiaTech).
- Heiko Werner : étudiant en thèse (Hajo Holzmann, université de Marburg).

## DOMAINES DE RECHERCHE

Mes travaux de recherche portent sur trois domaines distincts :

- les algorithmes stochastiques : critères de convergence, ergodicité.
- les méthodes de Monte Carlo : chaînes de Markov, processus adaptatifs, contrôle et optimisation des vitesses de convergence, réduction de la variance, analyse non-paramétrique des vitesses de convergence en grande dimension.
- les modèles à données manquantes : modèles de Markov cachés, mélanges semi-paramétriques, mélanges semi-paramétriques de regression, problèmes inverses, tests, applications en Biologie, la Fiabilité et l'Actuariat.

## COLLABORATIONS NATIONALES ET SUJETS

- Laurent Bordes (université de Pau) : mélanges semi-paramétriques et Chaînes de Markov Cachées (CMC).
- Cristina Butucea (ENSAE, Gustave Eiffel) : mélanges semi-paramétriques.
- Didier Chauveau (université d'Orléans) : algorithmes MCMC et algorithmes de type EM pour les mélanges semi-paramétriques.
- Céline Delmas (INRAE Toulouse) : modèles de mélange avec application aux puces ADN.
- Gersende Fort (CNRS-Institut Mathématique de Toulouse) : algorithmes MCMC adaptatifs.
- Olivier Lopez (ISUP, Sorbonne Université) : Test d'indépendance.

- Xavier Milhaud (Aix-Marseille) : tests pour des mélanges semi-paramétriques.
- Eric Moulines (Télécom ParisTech) : algorithmes MCMC adaptatifs.
- Stéphane Mottelet (université de technologie de Compiègne) : mélanges semi-paramétriques.
- Denys Pommeret (Marseille-Luminy) : tests pour des mélanges semi-paramétrique.
- Pierre Priouret (Paris 6) : algorithmes MCMC adaptatifs.
- Yahia Salhi (ISFA, Lyon 1) : tests pour des mélanges semi-paramétriques.

#### COLLABORATIONS INTERNATIONALES ET SUJETS

- Eric Feron (GeorgiaTech, USA) : contrôle du trafic aérien.
- Paolo Giudici (université de Pavie, Italie) : modèles graphiques et CMC.
- Hajo Holzmann (université de Marburg, Allemagne) : mélanges semi-paramétriques.
- David Hunter (Penn State university, USA) : mélanges semi-paramétriques.
- Nikolai Leonenko (Université de Cardiff, Ecosse) : estimation de l'entropie.
- Florian Maire (UCD, Dublin, Irlande) : MCMC filamenteux et sparses.
- David McDowell (GeorgiaTech, USA) : valeurs extrêmes et matériaux.
- Rodrigue Nguéye Zumpe (GeorgiaTech, USA) : mélanges topographiques.
- Jagan Padbiri (GeorgiaTech) : mélanges et matériaux.
- Tobias Rydén (université de Lund, Suède) : modèles graphiques et Chaînes de Markov Cachées.
- Pierre Tarrès (université d'Oxford, UK) : bandit à deux bras.
- Heiko Werner (université de Marburg, Allemagne) : mélanges semi-paramétriques.

#### VISITE DE LABORATOIRES ÉTRANGERS

- **Atlanta** : school of aerospace, collaboration avec Eric Feron (2013-2015).
- **Atlanta** : school of mechanical engineering, collaboration avec David McDowell (2012-2013).
- **Oxford** : institut de Mathématiques d'Oxford (UK), sur invitation de Pierre Tarrès (2 semaines en janvier 2006).
- **Stanford** : département de Statistique de l'université de Stanford (Sequoia Hall, États Unis) sur invitation de Susan Holmes (2 semaines sur septembre et octobre 2005).
- **Cambridge** : département du Signal de Cambridge (UK) sur invitation d'Arnaud Doucet et de Christophe Andrieu (2 semaines en septembre 1999).
- **Lund** : centre des sciences Mathématiques de Lund (Suède) sur invitation de Tobias Rydén (2 semaines en janvier 1998).
- **Pavie** : département d'Économie et d'études quantitatives de l'université de Pavie (Italie) dans le cadre de mon post-doc (1 an 1997-98).

#### PARTICIPATION À DES JURY DE THÈSE

- Rapporteur de la thèse de Mr. Heiko Werner (Marburg University, Allemagne), soutenance le 26/09/2018.
- Rapporteur de la thèse de Mr. Diaa Al Mohammad (UPEMC), soutenance le 17/11/2016.
- Membre du jury de thèse de Mr. Rodrigue Nguéye Zumpe (GeorgiaTech, USA), soutenance le 20/05/2015.

#### RESPONSABILITÉS ADMINISTRATIVES

- 2001-05** Membre élu du bureau de Laboratoire Marne-la-Vallée/Créteil.
- 2000-06** Co-responsable de la bibliothèque.
- 2000-06** Correspondant de la SMAI.

## ENSEIGNEMENTS RÉCENTS

- 2023**– Cours et TP sous R de classification **L3-pro** (20 heures).
- 2023**– Cours de probabilité en amphithéâtre **L2** (36 heures).
- 2023**– TD de probabilité **L2** (36 heures).
- 2023**– TD de suites et fonctions **L1** (36 heures).
- 2021-2023** Cours et TD de Calculus **L1** (72 heures).
- 2020-2021** Soutien en distanciel de Calculus **L1** (24 heures).
- 2020-2021** Soutien en distanciel de Méthodologie **L1** (24 heures).
- 2019-2023** TD de Statistique en **L3** (36 heures).
- 2019-2021** TD de Probabilités et Statistique en **L1** Sc. Eco. (25 heures).
- 2016-2019** Cours et TP de datamining sous R en **L3** (50 heures+25heures de projet tutoré).
- 2016–19** Cours et TD de Calculus en **L1** (72 heures).
- 2016–19** Cours et TD d'Algèbre Linéaire **L1** (72 heures).
- 2013** Cours de Time Series Analysis au niveau Graduate (GeorgiaTech, 28 heures).
- 2011** Cours d'Intégration en **L3** Maths-Info (36 heures).
- 2011** TD de probabilité en **L3** Maths-Info (36 heures).
- 2009-11** Cours (amphithéâtre) d'Analyse en **L1** (36 heures).
- 2006-09** Cours, en partenariat avec J.F. Delmas (ENPC), sur les *Modèles Aléatoires et Applications* en **Master 2** Mathématiques et Applications (20 heures).
- Thèmes :
- analyse des zones codantes dans l'ADN ;
  - chaînes de Markov cachées ;
  - algorithme EM ;
  - contrôle optimal ;
  - processus de sauts, file d'attente ;
  - recuit simulé, algorithme de Hastings Metropolis.
- 2006-09** TD de *Processus Stochastiques* en **Master 1** Mathématiques et Applications (24 heures).
- 2002-09** TD de *Statistique mathématique* en **Master 1** IMIS et Mathématiques et Applications (18 heures).
- 2003-09** Cours (amphithéâtre) et TD de *Calculus* en **L1** (60 heures).
- 2006-09** TD d'*Algèbre Linéaire* en **L1** (36 heures).
- 2008-11** Cours et TD de *Probabilité* en **L2** option Maths-Info (45 heures).

## ENSEIGNEMENTS DÉJÀ EFFECTUÉS

- TD et TP de *Statistique Empirique et Analyse des Données* (40 heures) sous SAS et MATLAB en **Master 1 actuariat**, IMIS, Mathématiques et Applications durant 5 années.
- Cours et TD de *Probabilité* (45 heures) en **L3-Maths** durant 6 années.
- Cours et TD de *Probabilité* (20 heures) pour la filière **Ingénieur 2000** durant 2 années.
- TD de *Probabilité* (18 heures) en **L2** option Physique durant 3 années.
- TD de *Calcul différentiel* (36 heures) en **L2** option Physique durant 2 années.
- TD d'*Analyse* (36 heures) en **L2** option Physique durant 2 années.
- TD d'*Algèbre linéaire* en **L2** option Maths durant 4 années.

## ENSEIGNEMENTS EN ÉCOLE D'INGÉNIEUR

**2002-06** Cours et TP (Scilab) de *Statistique* (28 heures) en deuxième année à l'École Nationale des Ponts et Chaussées (ENPC). Participation à la rédaction du polycopié de cours.

**2007-09** Cours (amphithéâtre) et TD d'*Analyse* ( $2 \times 30$  heures) en première année de prépa intégrée à l'École Supérieure d'Ingénierie en Électronique et Électrotechnique (ESIEE).

**2007-08** Cours et TD de mise à niveau en *Analyse* (45 heures) pour les intégrants en 3<sup>ème</sup> année à l'ESIEE.

**Logiciels :** R, Excel, Tableau, SAS, Matlab, Splus, Scilab.

### BILAN DES TRAVAUX DE RECHERCHE DE CES CINQ DERNIÈRES ANNÉES

- **Agrégation locale de noyaux markoviens pour l'échantillonnage de lois à bruit évanescent.** En collaboration avec F. Maire (université de Montréal). Dans ce travail nous nous intéressons au problème de l'échantillonnage de lois de probabilité dont la masse est portée par un support de dimension significativement inférieure à l'espace ambiant (lois filamenteuses, support de type variété, etc.). De telles loi apparaissent en statistique Bayésienne lorsque l'on traite des problèmes inverses (détection/repérage d'une source acoustique), ou en biologie moléculaire pour estimer les paramètres de modèles de transfection (introduction d'acides nucléiques dans des cellules eukaryotiques), correspondant à des systèmes d'EDO, au moyen de méthodes Bayésiennes. Pour ce type de problèmes nous introduisons des algorithmes de Metropolis Hastings revisités dont la loi instrumentale à l'instant  $t$  dépend de la position courante de l'algorithme (permettant un apprentissage local d'une stratégie d'échantillonnage efficace au sens où elle vise localement le support de la loi et non le vide ambiant). Nous montrons que ces algorithmes admettent comme loi invariante la mesure cible et qu'ils convergent dans certaines situations plus vite que les algorithmes classiques. Pour évaluer numériquement les performances de nos algorithmes nous utilisons le critère de l'entropie proposé par Chauveau et Vandekerckhove (2007, 2013, 2014). Nous avons énormément travaillé sur cet article depuis 2017 (27 pages pour l'article principal et 20 pages pour le supplementary material). Il a été accepté il y a quelques mois dans SIMODS (SIAM Journal on Mathematics and Data Science).
- **Estimation adaptative en norme-sup pour des mélanges semiparamétriques conditionnels.** En collaboration avec H. Werner, H. Holzmann (université de Marburg). Dans ce travail nous proposons une méthode permettant d'estimer un mélange constitué par deux modèles de régression. Dans l'un des modèles la loi conditionnelle de l'observation sachant la covariable est inconnue mais supposée symétrique par rapport à un paramètre de localisation, alors que dans l'autre modèle cette loi conditionnelle est spécifiée à un paramètre d'échelle près. Précisons que les paramètres de localisation et d'échelle sont aussi autorisés à dépendre de la covariable. Nous donnons des conditions d'identifiabilité pour notre modèle et en déduisons un  $M$ -estimateur local des paramètres qui converge en norme-sup au taux optimal usuel pour des classes Hölderiennes ad. hoc. Nous introduisons aussi une version adaptative de nos estimateurs basée sur la méthode de Lepski. Les bornes en norme-sup que nous obtenons montrent que notre  $M$ -estimateur local permet d'estimer globalement les paramètres fonctionnels du modèle. Par ailleurs nous montrons par simulations les performances de notre méthode à taille d'échantillon fini et illustrons son intérêt en l'appliquant à des données réelles issues de la bioinformatique. L'article associé à ce travail est paru dans Electronic Journal of Statistics en 2019.

- **Sur les tests pour les mélanges de contamination.** En collaboration avec X. Milhaud, D. Pommeret (Université Aix Marseille) et Y. Salhi (Université Lyon 1, ISFA). Nous nous intéressons spécifiquement dans cette série de travaux aux modèles de contamination. Ces modèles sont des mélanges à deux composantes dont l'une est connue et l'autre inconnue, la proportion du mélange étant elle aussi inconnue. Nous avons traité pour ce modèle la question du test d'appartenance de la composante inconnue à une famille paramétrique de lois. La méthode consiste à comparer, jusqu'à un ordre parcimonieux, deux à deux les coefficients de la densité inconnue exprimée dans une base de Hilbert, pour une estimation de la proportion  $p$  du mélange entièrement basée soit sur l'hypothèse paramétrique de référence soit sur une hypothèse relâchée de type symétrie simple. Un article sur ce travail est paru dans *Electronic Journal of Statistics* en 2019.

Dans un deuxième travail nous nous intéressons au test de comparaison de la composante inconnue dans un problème à double échantillon. La technique employée repose à nouveau sur l'hypothèse de symétrie des densités inconnues et la comparaison parcimonieuse de leur décomposition estimée dans des bases de Hilbert. Un article sur ce travail est paru dans *Journal of Statistical Planning and Inference* en 2022.

Dans le troisième travail nous revisitons entièrement notre méthodologie de test dans le cas du double échantillon afin de traiter les situations où les composantes inconnues ne sont pas symétriques. La difficulté dans le cadre général, voir Patra et Sen (2016), est qu'il n'existe pas d'estimateur  $\sqrt{n}$ -consistant de la proportion du mélange ce qui empêche toute procédure de test basée sur une construction "plug-in" de la statistique de test. Pour palier ce problème nous proposons un test du type Cramer Von Mises relâché, les fonctions de répartition inversées n'étant pas régularisées comme dans Patra et Sen (2016), dont nous établissons les lois asymptotiques sous  $H_0$  (égalité des composantes contaminantes) et  $H_1$  (différence entre les lois contaminantes), avec une dérive proportionnelle à  $n$  sous  $H_1$ . Nous montrons les bonnes performances, en terme de niveau et puissance, de notre procédure de test pour des lois admettant différents types de support ( $\mathbb{R}$ ,  $\mathbb{R}^+$ ,  $\mathbb{N}$ ). Enfin nous appliquons notre procédure de test à des jeux de données liés à l'excès de mortalité durant les premières semaines de la pandémie de COVID 19 en Europe. Nous comparons plus spécifiquement si l'impact du COVID 19 a été semblable entre pays pris deux à deux dans un panel de pays européens. Un article sur ce travail vient d'être accepté dans *Bernoulli*. Notons que tous nos algorithmes sont implémentés dans un package R appelé `admix` (<https://cran.rstudio.com/web/packages/admix/index.html>).

- **L'estimateur par plus proches voisins de l'entropie : un outil adapté à l'évaluation des algorithmes MCMC en grande dimension.**

La plupart des méthodes (adaptatives) de Monte Carlo par Chaîne de Markov (A)MCMC ont en général des vitesses de convergence inconnues vers leur loi cible. Afin de mieux appréhender l'impact réel de stratégies de simulation sur la convergence nous proposons une méthode, basée sur de la simulation en parallèle d'algorithmes, utilisant la distance de Kullback comme critère. L'estimation de l'entropie interne et croisée se fait à chaque itération au moyen de la méthode dite des plus proches voisins. Ces estimateurs ont été essentiellement utilisés, jusqu'à une époque récente, en dimension un alors que des besoins d'estimation en dimension plus grande se font sentir dans des domaines comme les neurosciences ou la biologie des systèmes. Le problème qui apparaît lorsque la dimension devient grande est que la vitesse de convergence de l'estimateur est fortement ralentie (biais clairement notable). Pour contrer ce pro-

blème nous proposons une approche originale basée sur un “échantillon asymptotique” de référence permettant de réduire drastiquement le biais en dimension modérée couplée à une solution basée sur l’ACP (pour ne garder que les sous-espaces discriminants permettant d’examiner la convergence) en dimension plus grande. Notons que tous nos algorithmes sont implémentés dans un package R, appelé **EntropyMCMC** (<https://cran.r-project.org/web/packages/EntropyMCMC/index.html>), utilisant intensivement les avantages du codage haute performance (en parallèle). Une version révisée de ce travail vient être soumise.

#### TRAVAUX PUBLIÉS OU À PARAÎTRE

1. Bakry, D., Milhaud, X. and Vandekerkhove, P. (1997). Statistics of Hidden Markov chains with finite state space. The nonstationary case. *C. R. Acad. Sci. Paris*, Série I, 203–206.
2. Vandekerkhove, P. (1998). Simulated annealing with a sequential estimator of the energy. *C. R. Acad. Sci. Paris*, Série I, 1003–1006.
3. Chauveau, D. and Vandekerkhove, P. (1999). Un algorithme de Hastings-Metropolis avec apprentissage séquentiel. *C. R. Acad. Sci. Paris*, Série I, 173–176.
4. Giudici, P., Rydén, T. and Vandekerkhove, P. (2000). Likelihood-Ratio Tests for Hidden Markov Models. *Biometrics*, **56**, 742–747.
5. Chauveau, D. and Vandekerkhove, P. (2001). Algorithmes de Hastings Metropolis en interaction. *C. R. Acad. Sci. Paris*, Série I, 881–884.
6. Chauveau, D. and Vandekerkhove, P. (2002). Improving convergence of the Hastings-Metropolis Algorithm with a learning proposal. *Scand. J. Statist.*, **28**, 13–29.
7. Bordes, L. and Vandekerkhove, P. (2005). Statistical inference for Partially Hidden Markov Models. *Communications in Statistics*, **34**, 1081–1104.
8. Vandekerkhove, P. (2005). Consistent and asymptotically normal estimates for hidden Markov mixtures of Markov models. *Bernoulli*, **11**, 103–129.
9. Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two component mixture model. *Ann. Statist.*, **34**, 1204–1232.
10. Bordes, L., Delmas, C. and P. Vandekerkhove. (2006). Semiparametric estimation of a two-component mixture model where a component is known. *Scand. J. Statist.*, **33**, 733–752.
11. Chauveau, D. and Vandekerkhove, P. (2007). A Monte Carlo estimation of the entropy for Markov chains. *Methodology and Computing in Applied Probability*, **9**, 133–149.
12. Bordes, L., Chauveau, D. and Vandekerkhove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Computational Statistics and Data Analysis*, **51**, 5429–5443.
13. Bordes, L. and Vandekerkhove, P. (2010). Semiparametric two-component mixture model when a component is known : an asymptotically normal estimator. *Math. Meth. Statist.*, **19**, 22–41.
14. Tarrès, P. and Vandekerkhove, P. (2012). On ergodic two-armed bandits. *Ann. Appl. Probab.*, **22**, 457–476.
15. Fort, G., Moulines, E., Priouret, P., Vandekerkhove, P. (2012). A simple variance inequality for U-statistics of a Markov chain with applications. *Statist. Probab. Letters*. **82**, 1193–1201.
16. Vandekerkhove, P. (2012). Estimation of a semiparametric mixture of regression model. *J. Nonparam. Statist.* , **25**, 181-208.

17. Chauveau, D. and Vandekerkhove, P. (2013). Smoothness of Metropolis-Hastings algorithm and application to entropy estimation. *ESAIM P&S*, **17**, 419-431.
18. Bordes, L., Kojadinovic, I. and Vandekerkhove, P. (2013). Semiparametric estimation of a mixture of two linear regressions where one component is known. *Electronic Journal of Statistics*, p. 2603-2644.
19. Fort, G., Moulines, E., Priouret, P., Vandekerkhove, P. (2014). A central limit Theorem for adaptive and interacting Markov chains. *Bernoulli*, **20**, 457-485 .
20. Butucea, C. and Vandekerkhove, P. (2014). Estimation in semiparametric mixtures of symmetric distributions. *Scand. J. Statist.*, **41**, 227-239.
21. Chauveau, D. and Vandekerkhove, P. (2014). Simulation Based Nearest Neighbor Entropy Estimation for (Adaptive) MCMC Evaluation, *JSM Proceedings, Statistical Computing Section. Alexandria, VA : American Statistical Association*, 2816-2827.
22. Vandekerkhove, P., Padbidri, J.M. and McDowell, D.L. (2014). Integrated Cumulative Error (ICE) distance for mixture model selection : Application to extreme values in metal fatigue problems. *Electronic Journal of Statistics*, 8, p. 3141-3175.
23. Butucea, C., Nguype Zumpe, R. and Vandekerkhove, P. (2017). Semiparametric topographical mixture models with symmetric errors. *Bernoulli*, **23**, 825-862.
24. Pommeret, D. and Vandekerkhove, P. (2019) . Semiparametric density testing in the contamination model. *Electronic Journal of Statistics*, **13**, p. 4743-4793.
25. Werner, H., Holzmann, H. and Vandekerkhove, P. (2019). Adaptive estimation in the sup-norm for conditional semiparametric mixtures. *Electronic Journal of Statistics*, **14**, p. 1816-1871
26. Milhaud, X., Pommeret, D., Salhi Y. and Vandekerkhove, P. (2022). Semiparametric two-sample mixture components comparison test. *Journal of Statistical Planning and Inference*. **216**, p. 135-150.
27. Maire, F. and Vandekerkhove, P. (2022). Markov Kernels Local Aggregation for Noise Vanishing Distribution Sampling. *Accepté dans SIAM journal on Mathematics and Data Science*.
28. Milhaud, X. Pommeret, D., Salhi, Y. and Vandekerkhove, P. (2022). Two-sample contamination model test. *Accepté dans Bernoulli*.

#### ARTICLES SOUMIS [S] OU EN COURS DE RÉDACTION [C]

- S1 Chauveau D. and Vandekerkhove, P. (2022). The Nearest Neighbor entropy estimate : an adequate tool for high dimensional adaptive MCMC evaluation.
- S2 Milhaud, X., Pommeret, D., Salhi Y. and Vandekerkhove, P. (2022). Contamination based K-sample clustering clustering.
- C1 Milhaud, X., Lopez, O. and Vandekerkhove, P. (2022). Semiparametric contamination model Independence test.
- C2 Lacour, C. and Vandekerkhove, P. (2022). Semiparametric Markovian contamination process.

#### PROJETS ANR

1. Programme ANR BigMC (Issues in large scale Monte Carlo), 2009–2012.  
Porteuse du projet : Gersende Fort (CNRS, ENST Paris-Tech).  
Pôles impliqués : ENST Paris-Tech, Dauphine, Ecole des Ponts Paris-Tech.

2. Programme ANR Mixture (mélanges semiparamétriques). Soumis sans succès en 2006, 2008 et 2009.  
Porteur du projet : Pierre Vandekerkhove (Université Paris-Est).  
Pôles impliqués : Pau, Paris-Est, Génopôle d'Evry, Paris Sud, Orléans.

### PROPOSAL NASA ET NSF

1. Titre : CODEREDHAT–CONflict DETECTION and RESolution Discovery for Historical Air Traffic.  
Porteur du projet : Eric Feron. Date de dépôt : 02/28/2014.  
Contribution : modèle de mélange de régression multivarié où la donnée cachée est la teneur des discussions entre pilotes.
2. Titre : SMARTBiD–MAGNet–Statistical Models And Reduction Techniques for Big Data from Multimodal Air and Ground Networks.  
Porteur du projet : Eric Feron. Date de dépôt : 09/30/2014.  
Contribution : modèles de Markov caché pour filtrage de grand nuages de points résumés au moyen de mélanges gaussiens.
3. Titre : CRISP–Critical Resilient Interdependent Infrastructure Systems and Processes.  
Porteur du projet : Eric Feron. Date de dépôt : 03/20/2015.