

**Programme pluriformations**

**Titre :**

**Création de STAR  
pôle Statistique de Rennes**

**département scientifique concerné :**

**101  
Mathématiques et leurs interactions**

**Etablissement :**

**Ecole Normale Supérieure de Cachan  
(Antenne de Bretagne)**

## I - Structuration et moyens du programme pluriformations faisant l'objet d'une demande de reconnaissance

### I.1 - Caractéristiques de la demande de reconnaissance d'un PPF par le ministère

Établissement faisant cette demande : ...**ENS Cachan**..... **XX** à titre principal

Autre(s) établissement(s) présentant également une demande pour le même PPF :

.....  à titre principal  à titre secondaire  
 .....  à titre principal  à titre secondaire

**Intitulé du programme pluriformations** (ne pas utiliser de sigle) :  
 Création de STAR, pôle de la Statistique Rennaise

#### Département(s) scientifique(s) et secteurs disciplinaires de la MSTP :

- Département scientifique principal : .....1..... et éventuellement secondaires : .....
- Secteur disciplinaire principal : .....101 et éventuellement secondaires : .....

#### Responsable

M./Mme	Nom	Prénom	Corps-Grade	Section du C.N.U. ou de l'organisme
_M_	_PIERRE_	_MICHEL_	_PU_	_26è_

J'autorise la diffusion de mon nom sur internet (annuaire des PPF).

#### Mots-clefs MSTP (cf. nomenclature) :

_Statistique_	_	_	_
_	_	_	_

#### Mots-clefs libres :

_	_
---	---

#### Coordonnées officielles du PPF :

- Localisation et établissement : **Antenne de Bretagne de l'ENS Cachan**
- Numéro, voie : Campus de Ker Lann
- Boîte postale :
- Code Postal et ville : 35170 - BRUZ
- Téléphone : 02 99 05 93 13                      Télécopie : 02 99 05 93 28
- Adresse électronique : michel.pierre@bretagne.ens-cachan.fr

**Date et signature du responsable de la demande :**

**Partie à remplir par le responsable de l'établissement demandeur :**

Je donne mon accord à la présente demande de reconnaissance par le ministère d'un programme pluriformations.

Sous réserve de l'accord de la direction de la recherche, la direction du PPF serait assurée par :

**M PIERRE Michel**.....

Nom et prénom du responsable de l'établissement demandeur (**établissement principal**) :

**DUPAS Claire**.....

Qualité : ...Directrice de l'ENS Cachan.....

Date :

Signature :

Nom et prénom du responsable de l'établissement demandeur (**établissement secondaire**, le cas échéant) :

.....

Qualité : .....

Date :

Signature :

## I.2 - Moyens matériels et financiers

### I.2.1 Liste des achats d'équipement souhaités ou programmés pour le prochain contrat

Descriptif et nombre	Coût unitaire	Source de financement (ministère, EPST ou EPIC à préciser ...) (1)	Coût total
3 SERVEURS	10 Keuros		30 Keuros
LOGICIELS Statistique			6 Keuros
LIVRES/REVUES	2,5 Keuros/an		10 Keuros
<b>Tota</b>			46 Keuros

(1) Préciser si les financements sont demandés (D) ou acquis (A). Faire apparaître les éventuels cofinancements prévus.

La justification des financements demandés au ministère pour les achats d'équipements doit être explicitée dans le dossier scientifique.

## I.2.2 Budget prévisionnel du PPF pour le prochain contrat

	NATURE DES FINANCEMENTS	2006	2007	2008	2009	Moyenn e annuelle
Crédits demandés au ministère  (direction de la recherche)	Crédits scientifiques	26 Keuros	26 Keuros	26 Keuros	26 Keuros	26 Keuros
	Vacations	?				

N.B. :

- Si des financements sont demandés pour l'unité par l'intermédiaire de plusieurs établissements d'enseignement supérieur et de recherche, la **répartition** de ce financement entre les différents établissements partenaires sera mentionnée dans ce tableau.
- Les dotations contractualisées sur 4 ans sont **récurrentes**. Toutefois l'établissement pourra faire apparaître, à titre indicatif, les modalités de réalisation du programme demandé ainsi que l'échéancier.
- Dans le dossier scientifique, la **justification des besoins** doit être présentée, de même qu'éventuellement leur modulation au cours du contrat.

Pour information :

Crédits attendus dans le cadre du CPER		2006	Remarques éventuelles
Part Etat	Ministère		
	EPST ou EPIC		
Part Collectivités locales			
<b>Total</b>			

**REMARQUE** : dans sa séance du 27 mai 2005, le Conseil Scientifique de l'ENS Cachan a attribué une aide pour la préparation de ce projet et pour le démarrage dès 2005 d'actions concertées en Statistique sur le site de Rennes, à hauteur de 5000 Euros.

Comme il sera indiqué explicitement dans le compte-rendu de cette même séance, le conseil scientifique de l'ENS Cachan soutient fortement la mise en place de ce projet.

### I.3 - Liste des équipes participant au programme ou utilisatrices des équipements collectifs décrits dans le dossier

Etablissement de rattachement principal	Label et n° (le cas échéant)	Intitulé de l'équipe	Nom et prénom du responsable	Département scientifique concerné
ENS Cachan		Equipe de recherche en mathématique	PIERRE Michel	1
	IRMAR, UMR 6625	rattachée à l'Institut de Recherche Mathématique de Rennes (IRMAR)		
Université de Rennes1	IRMAR	Equipe Statistique au sein de l'équipe "Processus Stochastiques et Statistique"	YAO Jian-Feng	1
Université de Rennes2		Laboratoire de Statistique de Rennes2	MATZNER-LOBER Eric	1
ENSAI	CREST-ENSAI	Laboratoires de Statistique et Modélisation et de Statistique d'Enquêtes	DELECROIX Michel	1
Agrocampus Rennes		Laboratoire de Mathématiques Appliquées	PAGES Jérôme	1
IRISA	UMR 6074	Plusieurs projets (voir plus loin au sein du projet scientifique)	BASSEVILLE Michèle	9

## II - Dossier scientifique

### II.1 - Rapport scientifique concis et utilisation des crédits des quatre années précédentes (en cas de demande de renouvellement du PPF)

### II.2 - Projet scientifique pour la période 2006-2009

Le projet scientifique et ses effets structurants doivent être explicités dans cette annexe.

Les participants au PPF et les publications significatives seront mentionnés.

Préciser comment la demande de PPF se situe par rapport aux projets des structures fédératives de l'établissement (demande indépendante, interne, associant plusieurs structures fédératives).

## **Titre : Création de STAR, pôle Statistique de Rennes**

### *Plan du dossier*

1. *Objectif général*
2. *Situation actuelle*
3. *Le projet proprement dit*
  - *Orientation Statistique à l'antenne de Bretagne de l'ENS Cachan*
  - *Création d'une équipe "Statistique" au sein de l'IRMAR*
  - *Développement d'un programme commun de recherche en Statistique*
  - *Mise en place d'actions communes*
  - *Impact escompté*
4. *Demande de financement*

**1. Objectif général** : Fédération et Structuration de la recherche en Statistique sur le site Rennais avec la création d'un pôle pour la mise en commun des moyens et des compétences en vue du développement d'un programme de recherche coordonné en Statistique. Il rassemblera les équipes de Statistique des Universités de Rennes1 et Rennes2, de l'ENSAI, d'Agrocampus, de l'ENS Cachan ainsi que des équipes de l'IRISA. Il s'articulera autour de l'IRMAR, qui a vocation à développer et piloter les recherches en mathématiques sur le site rennais et qui accueillera en son sein une grande partie des chercheurs du pôle.

### **Rappel des sigles :**

**IRMAR** : Institut de Recherche Mathématique de Rennes, UMR 6625, avec comme tutelles le CNRS, l'Université de Rennes1, l'Ecole Normale Supérieure de Cachan et l'**INSA-Rennes**, Institut National de Sciences Appliquées

**ENSAI** : Ecole Nationale de la Statistique et de l'Analyse de l'Information, située sur le campus de Ker Lann près de l'Antenne de Bretagne de l'ENS Cachan. Les activités de recherche relèvent du **CREST**, Centre de Recherche en Economie et en Statistique.

**Agrocampus** : Institut National d'Enseignement Supérieur et de Recherche Agronomique et Agro-alimentaire de Rennes

**IRISA** : Institut de Recherche en Informatique et Systèmes Aléatoires, situé sur le campus de Beaulieu, UMR 6074 avec partenariat de l'INRIA, du CNRS, de l'Université de Rennes1 et de l'INSA-Rennes.

**2. Situation Actuelle** : dans les grandes lignes, elle se caractérise par

- un point fort qui est *l'existence sur Rennes d'une activité importante de recherche en Statistique*, répartie sur 6 établissements
- un point faible constitué précisément par la dissémination des équipes sur le site sans une réelle coordination jusqu'ici et où l'IRMAR joue un rôle trop peu important malgré sa mission de pilotage de la recherche en mathématique sur Rennes.

Rappelons les **forces en présence** (voir en annexe la liste nominative des équipes)

*Le laboratoire de Statistique de Rennes2*, l'équipe la plus nombreuse avec 14 membres permanents, dont 4 professeurs et 10 maîtres de conférences. Cette équipe vient de se renouveler de façon significative avec le recrutement depuis 2001 de cinq jeunes sur des postes de maîtres de conférences (dont un à ce mouvement. Le statut de ce laboratoire est local (reconnaissance par le conseil scientifique de l'université de Rennes2). Les instances ministérielles l'incitent à rejoindre l'IRMAR.

*Le laboratoire de Statistique et Modélisation du CREST-ENSAI* avec 8 membres dont 2 professeurs, 6 maîtres de conférences et *le Laboratoire de Statistique d'Enquêtes du CREST-ENSAI* avec 3 chercheurs. Tous deux sont sur le campus de Ker Lann où l'ENSAI a été délocalisée en 1994. Ses membres sont rattachés au CREST dont la direction et l'essentiel de son effectif sont sur le site parisien. Plusieurs départs sont prévus à court terme, compte tenu des habilitations qui viennent d'être soutenues.

*Le laboratoire de Mathématiques Appliquées d'Agrocampus* avec 6 membres (1PR, 5 MC), laboratoire de statut local au sein d'Agrocampus.

*L'équipe Processus Stochastiques et Statistique de l'IRMAR* : elle comprend 4 statisticiens (3PR et 1 MC). Notons que Arnaud Debussche (PR Ker Lann) est rattaché à 50% à cette équipe.

*Plusieurs équipes de l'IRISA* développent et/ou utilisent la Statistique : on peut citer les projet ARMOR, SISTHEM, ASPI, DISTRIBCOM (commun avec l'ENS Cachan), TEXMEX, VISTA et METISS.

Il n'y a pour l'instant *aucun statisticien à l'Antenne de Bretagne de l'ENS Cachan*. Cependant, des formations en Statistique y sont offertes avec le concours de collègues de l'ENSAI et de l'Université de Rennes1 (pour l'agrégation, en particulier). D'autre part, le département de mathématiques s'attache à orienter les élèves vers des études doctorales en Statistique. Une élève de la toute première 1<sup>ère</sup> promotion de 1996 vient juste d'être recrutée comme maître de conférences dans cette discipline ; un élève entré sur le concours 3<sup>ème</sup> année est maintenant CR CNRS sur ce profil. Plusieurs autres sont engagés dans cette voie et vont bientôt se présenter sur le marché des recrutements.

**La coordination est déjà amorcée**. Depuis quelques mois, une concertation s'est mise en place entre les différents partenaires. Rappelons quelques étapes de sa progression :

- A la rentrée 2004, le laboratoire de Rennes 2 et le laboratoire de mathématiques appliquées d'Agrocampus Rennes ont mis sur pied un programme de travail commun
- Des Journées STAR "de la Statistique Rennaise" ont été organisées en octobre dernier à l'ENSAI à l'initiative du laboratoire CREST-ENSAI et de l'IRMAR. Elles ont été un temps fort pour les équipes rennaises de Statistique.
- Une réflexion a eu lieu au sein de l'équipe de Rennes 2, à l'instigation de la direction du département "Mathématiques et leurs Interactions" à la MSTP, pour étudier une collaboration plus formalisée avec l'IRMAR
- Dans le cadre de la préparation du plan quadriennal 2006-2009 de l'ENS Cachan, des réunions de travail ont eu lieu sur le site de Ker Lann, à l'initiative du département de mathématiques, pour analyser les possibilités de projets communs en Statistique : elles ont rassemblé les directeurs de l'IRMAR, de l'ENSAI, de l'antenne de Bretagne de l'ENS Cachan, des laboratoires de Statistique de l'ENSAI et de Rennes2 et plusieurs membres de ces organismes, puis des représentants d'Agrocampus et de l'IRISA.
- Des réunions de réflexion ont eu lieu sur le site de Beaulieu pour étudier les collaborations Rennes1/IRMAR/Rennes2
- Le développement et l'organisation de la Statistique Rennaise ont été à l'ordre du jour *de quatre conseils scientifiques de l'IRMAR* où des décisions ont été prises concernant les projets annoncés plus loin.

*La mise en place de ce projet fait suite à ces réunions préparatoires et prend en compte les suggestions qui ont été émises et les décisions qui ont été prises.*

### **3. Le Projet**

Il présente des aspects scientifiques, mais aussi des aspects structurels et organisationnels. Enumérons les diverses actions prévues.

**3.a. Ouverture d'une nouvelle orientation « Statistique » à l'antenne de Bretagne de l'ENS Cachan** : celle-ci est proposée dans le plan quadriennal 2006-2009 en cours de finalisation où est demandée la création d'un poste de professeur en Statistique, suivie de la création d'un poste d'AGPR et/ou Maître de conférences sur la période. Elle est, en particulier, motivée par l'énorme demande, largement insatisfaite à ce jour, de jeunes chercheurs en Statistique sur le territoire national. L'ENS, et particulièrement son antenne de Bretagne qui est en pleine expansion, se doit de jouer un rôle dans cet effort national hautement encouragé et appelé par nos instances. L'environnement rennais rappelé ci-dessus, la proximité de l'ENSAI, le potentiel de jeunes chercheurs constitué par les élèves du département de mathématiques, l'implication forte du département dans le Master rennais en Mathématiques, son orientation pluridisciplinaire et appliquée, tout cela fait de l'antenne de l'ENS un *endroit stratégique* pour développer cette orientation en Statistique.

Elle se traduira aussi par un encouragement des élèves normaliens et magistériens vers des études doctorales en Statistique, par une collaboration intensifiée avec nos voisins de l'ENSAI, par un apport aux autres équipes de l'antenne d'une compétence explicitement et vivement réclamée (équipes BIOMIS, projet commun DISTRIBCOM avec l'IRISA, projet transversal « puces communicantes », voir en fin du point 3.c ci-dessous...), et conduira à une visibilité accrue des mathématiques à l'antenne.

**3.b. Création d'une nouvelle équipe "Statistique" au sein de l'IRMAR.** L'institut regroupe actuellement 8 équipes. La Statistique n'est représentée que par 4 membres permanents. D'un autre côté, la nombreuse équipe de Rennes2 n'a pour l'instant qu'un statut local au sein de l'université de Rennes2 et nos instances (ministère, CNRS) les incitent fortement à s'associer à l'IRMAR.

Il a été décidé, d'un commun accord entre l'équipe de Rennes2 et le conseil scientifique de l'IRMAR que l'IRMAR accueille les membres de l'équipe de Rennes2 qui le désirent. Ceci pourrait se faire très vite et il sera proposé que la tutelle de l'IRMAR s'élargisse à l'université de Rennes2, selon le souhait des instances ministérielles.

Cet accueil est également offert aux membres des équipes de l'ENSAI et d'Agrocampus qui le désirent. Une étude sera faite sur les modalités d'une telle association, compte tenu de la tutelle de deux autres ministères. Des conventions ad hoc inter-établissements pourront en préciser la teneur.

Bien sûr, les futurs statisticiens de l'ENS seront d'office membres de l'IRMAR comme c'est le cas actuellement des tous les mathématiciens.

Une équipe "Statistique" bien identifiée sera donc créée à l'IRMAR qui réfléchira à une nouvelle organisation globale de ses équipes relevant du domaine stochastique.

*Dernière minute : le lundi 23 mai, le conseil scientifique de l'IRMAR a approuvé la création d'une telle équipe et d'une représentation de tous les établissements ci-dessus au sein du conseil.*

**3.c. Développement d'un programme commun de recherche en Statistique** concernant toutes les équipes ci-dessus. Il sera coordonné autour des grands thèmes qui sont déjà parmi les points forts des recherches en Statistique rennaise : Statistique non-paramétrique (ENSAI, Rennes2), Statistique des processus (Rennes1, ENSAI), Analyse des données, modèles linéaires et applications (Rennes2, Agrocampus), Modélisations appliquées en analyse d'image et traitement du signal (Rennes1, IRISA).

Des interactions croisées auront lieu au-delà des localisations ci-dessus et d'autres thèmes ou sous-thèmes pourront se dégager après la mise en commun des compétences et les échanges qui se créeront. Dans l'immédiat, l'accent sera mis sur les actions de recherche suivantes.

**Statistique des données fonctionnelles.** Dans de nombreux domaines, les quantités mesurées ne sont plus seulement des éléments de l'espace  $R^T$ , mais des objets plus complexes (courbes, images...) identifiés sous le nom de "données fonctionnelles" (Ramsay et Silverman, 1997). L'analyse statistique de ces données fonctionnelles est un domaine émergent en Statistique comme en témoigne, entre autres, l'accroissement de la littérature sur le sujet. Son essor se situe tant au niveau de la recherche fondamentale qu'à celui des applications en entreprise.

L'équipe se préoccupera dans un premier temps des données représentant des fonctions, qui sont souvent des courbes. Citons par exemple

- évolution de la température au cours du temps (physique)
- évolution du pH au cours du temps (industrie agro-alimentaire)
- évolution de l'absorbance en fonction de la longueur d'ondes (biologie)
- évolution des taux d'intérêt au cours du temps (économétrie).

Les méthodes classiques de Statistique seront d'une part adaptées à ces données ; d'autre part, des approches nouvelles seront développées : choix adéquats de normes, sélection de parties de courbes, ...

Les domaines d'application privilégiés sont les données issues de l'industrie alimentaire. Le laboratoire de Rennes 2 a eu un contrat de recherche avec le centre de recherche de DANONE en 2004.

*Le responsable de ce thème est A. Guyader (Rennes2).*

**Analyse des données multidimensionnelles.** Les méthodes d'analyse des données multidimensionnelles sont d'utilisation courante dans tous les domaines afin de prendre en compte le plus d'éléments possibles dans une action donnée. Dans cette perspective, la disponibilité d'outils de synthèse de vastes ensembles de données numériques est indispensable.

Les représentations euclidiennes, qui permettent de visualiser d'un seul coup d'œil de complexes écheveaux de ressemblance et d'associations, jouent un rôle essentiel. Le problème est bien résolu pour les ensembles de données homogènes. Mais les problématiques que l'on voit apparaître aujourd'hui, aussi bien dans le monde de la recherche que dans le monde professionnel, nécessitent de pouvoir analyser simultanément des données de types différents, en particulier, des variables qualitatives, quantitatives et textuelles.

Il importe aussi de disposer d'outils gérant automatiquement des structures de tableaux complexes : par exemple, dans l'analyse d'un questionnaire, l'outil doit prendre en compte la structure en thèmes et en sous-thèmes dans l'ensemble des questions, ce qui renvoie à l'analyse simultanée de tableaux possédant une dimension en commun.

L'objectif est, d'une part, de mener une réflexion théorique sur les données et leur mode de représentation, d'autre part, de mettre à la disposition des utilisateurs de nouveaux outils, avec le souci de l'expérimentation sur le terrain.

Il faut noter que, malgré les énormes retombées potentielles, ce thème est peu représenté en milieu universitaire français et il est important que les composantes Rennaises coordonnent leurs efforts en ce sens.

Signalons aussi la place un peu particulière de l'analyse sensorielle (au sens de qualités perçues des produits, gustatives ou autres) : les réflexions méthodologiques déjà menées ont conduit à un enrichissement des problématiques applicatives qu'elles ont donc devancées, alors que l'inverse est plus fréquent.

*L'animateur du groupe est Jérôme Pagès (Agrocampus).*

**Modèle linéaire.** Cette approche méthodologique fait ses preuves dans des domaines d'applications privilégiés comme les sciences de la vie, l'agronomie, l'analyse sensorielle, la qualité de la vie, les sciences de l'environnement.

En amont du recueil même des données, les méthodes de planification expérimentale forment un sujet de prédilection en matière de Statistique appliquée aux sciences de la vie. Il s'agit de définir les modalités de construction de plans optimaux en analyse sensorielle respectant des contraintes spécifiques. La prise en compte de coûts expérimentaux dans la planification des essais motivent également l'activité de recherche. Ainsi, il s'agit de travailler sur la construction de plans d'échantillonnages intégrant de manière optimale des informations auxiliaires et définissant des stratégies optimales d'estimation en présence de données incomplètes.

Mentionnons d'autre part la résolution de problèmes de multi-colinéarité des prédicteurs, la mise en évidence de prédicteurs latents et l'estimation de dimensionnalité d'un modèle, qui rejoignent partiellement les questions *de Statistique des données fonctionnelles* évoquées plus haut.

*L'animateur du groupe de travail est D. Causeur (Agrocampus).*

**Statistique des processus et algorithmes associés.** La statistique des processus s'est rapidement développée ces dernières années suite à une demande croissante en modélisation.

D'une part, nous développerons le fondement d'une nouvelle classe de diffusions à coefficients markoviens et exogènes. Ces modèles ont des applications dans divers domaines impliquant des séries d'observations dont la dynamique peut changer au cours du temps (dynamique temporelle hétérogène).

D'autre part, nous nous intéresserons au problème de l'estimation paramétrique et plus particulièrement lorsque les diffusions sont observées en des instants discrets. Ici des algorithmes originaux seront proposés pour le calcul non trivial de la vraisemblance des observations.

Par ailleurs, une activité centrale de plusieurs équipes de l'IRISA concerne le traitement de l'information délivrée par des capteurs (signaux et images), sur la base de méthodes d'inférence statistique associées à une modélisation des phénomènes physiques sous-jacents. Il s'agit d'une problématique en plein renouvellement, en raison notamment de la diffusion massive de capteurs de toutes natures dans des secteurs variés, et de l'émergence de normes (sécurité, environnement) de plus en plus nombreuses.

La statistique des processus, et notamment l'estimation de paramètres et la prise de décision, lorsque les processus sont observés en des instants discrets, est au cœur de ces activités. La conception, la mise en oeuvre numérique, voire l'analyse mathématique, d'algorithmes originaux pour la simulation des processus, et/ou le calcul non trivial de la vraisemblance (ou d'autres fonctions) des observations, à des fins d'estimation de paramètres, d'analyse d'événements rares, de détection et de diagnostic d'événements caractéristiques ou au contraire indésirables, font partie des contributions majeures des équipes Armor, Aspi, Distribcom, Metiss, Sisthem, Vista.

*Ces thèmes sont animés dans le cadre de diverses collaborations IRMAR/IRISA.*

**Modèles spatiaux pour l'analyse des mesures de mouvement dans une séquence d'images.** Les mesures de mouvement dans une séquence d'images véhiculent des informations fondamentales sur le contenu des images. Elles constituent ainsi un passage obligé pour l'interprétation et l'exploitation automatique des images. La modélisation de ces mesures spatiales sont d'une nature aussi variée que la nature elle-même et, d'autre part, les mesures sont généralement très bruitées.

Dans une collaboration établie depuis trois ans entre l'équipe VISTA de l'IRISA et l'équipe de Statistique de l'IRMAR, une nouvelle classe d'auto-modèles ainsi que les premiers résultats de modélisation ont été validés.

Une thèse co-dirigée sur ce sujet sera soutenue en 2005. Plusieurs stages ont aussi été encadrés dans le même domaine.

Cette thématique sera approfondie et fera l'objet d'une collaboration entre les deux sites.

*Ce thème est animé par Jian-Feng Yao(Rennes1) et Patrick Bouthémy (IRISA).*

**Modélisations statistiques non et semi paramétriques.** De très nombreuses études de données portent fondamentalement sur les modélisations des « fonctions de régression », qui traduisent l'impact des valeurs prises par un ensemble de variables dites « explicatives » sur des variables d'intérêt, dans le but d'analyser et quantifier le modèle obtenu à des fins décisionnelles, ou qui permettent de prédire les valeurs futures de séries temporelles, les variables causales par rapport au présent étant alors les variables du passé.

Le classique « modèle linéaire généralisé » utilisé de façon universelle, jusqu'alors, dans les études appliquées s'avère inadapté dans un certain nombre de situations pratiques. Les travaux entrepris dans le thème ici analysé portent sur le développement de méthodologies qui permettent une estimation performante des fonctions de régression sous des hypothèses de modélisation moins contraignantes, donc dans un cadre non-paramétrique (ce qui amène à des problèmes d'estimation fonctionnelle), ou dans le cadre de modèles semi-paramétriques introduits pour rendre plus performants sur petits échantillons les résultats obtenus (ce qui amène à une estimation conjointe « paramètre vectoriel-fonction » classique dans les approches de « réduction de dimension »).

Les techniques utilisées doivent logiquement s'imposer progressivement vis à vis de l'utilisation du « modèle linéaire généralisé » dans les secteurs d'analyse de données recourant classiquement à celui-ci, en techniques actuarielles par exemple (tarification automobile, analyse de sinistralité,...). Elles nécessitent pour cela des études algorithmiques préliminaires supplémentaires qui en rendent l'emploi plus convivial, et l'utilisation plus aisée. Elles débouchent naturellement, d'un autre point de vue, sur des études de régression sur données fonctionnelles, et une amélioration des méthodes de prédiction de séries temporelles, comme la « courbe de charge » de l'E.D.F..

Enfin, l'approche semi paramétrique, qui généralise la modélisation de Cox, est désormais utilisée dans le cadre d'études de données censurées, dans la recherche d'estimations performantes de modèles adaptés à l'épidémiologie ou les essais cliniques. C'est là un autre axe spécifique d'études et développements méthodologiques apparaissant dans le thème.

*Ce thème est animé par Michel Delecroix (ENSAI).*

**Statistique d'enquêtes.** La statistique d'enquête se développe dans deux directions bien présentes au laboratoire du Crest/Ensaï.

L'une consiste à étudier l'algorithmique et l'asymptotique des échantillonnage contraints, c'est à dire assujettis à restituer avec une variance d'échantillonnage nulle, ou, au plus, avec un écart en  $O_p(1/n)$ , certaines quantités connues dans la population.

L'autre consiste à ajuster un système de pondérations dans des cas analogues. Ces dernières techniques sont susceptibles d'extension à la statistique inférentielle classique, particulièrement dans le cas non paramétrique (estimer une médiane d'une distribution connaissant sa moyenne par exemple).

*Ce thème est animé par Jean-Claude Deville (ENSAI).*

**Collaborations pluridisciplinaires à l'antenne :** même si aucune recherche en Statistique n'a encore lieu sur l'antenne, plusieurs projets en cours dans d'autres domaines bénéficieraient d'un apport de compétence en Statistique. Leur prise en compte nécessite sans doute que l'interface soit assurée par un Statisticien en poste, mais une collaboration pourrait déjà être initialisée avec les forces Rennaises existantes. Citons quelques exemples de "besoins":

- DISTRIBCOM (projet commun IRISA/ENS Cachan) : évaluation statistique de l'encombrement généré par des applications offertes sur le web (pour définir un overbooking efficace par exemple).
- BIOMIS : évaluation statistique de l'activité cellulaire (activité électrique, expression d'un gène) dans des vastes réseaux de cellules sur puce (en particulier quand les seuls moyens de vérification sont destructifs).
- "Puces communicantes" : Etude statistique de la mobilité des individus et de la répartition spatiale et temporelle des interconnexions dans les réseaux sans fils (en vue d'algorithmes de communication performants)

### **3.d Mise en place d'actions communes**

Organisation annuelle de Journées STAR de la Statistique à l'instar de celles organisées en 2004. Elles le seront alternativement sur chacun des sites. Elles élargiront leur audience en visant un impact international et privilégieront les thèmes ci-dessus. En 2005, elles auront lieu à Agrocampus.

Mise en place et animation de groupes de travail et de séminaires réguliers sur les thèmes ci-dessus. Il sera veillé à ce que chaque site soit concerné, mais que les actions restent communes. Ceux déjà prévus à ce jour sont les suivants :

- un séminaire Statistique (au sens large), comprenant les aspects de modélisation aléatoire, alternativement sur les sites de Beaulieu (Rennes 1) et Villejean (Rennes 2)
- sur le site de l'ENSAI : poursuite du séminaire trois fois par mois,
- sur le site d'Agrocampus : groupe de travail "Modèle linéaire" les premiers jeudis du mois
- sur le site de Rennes 2 : groupe de travail "Statistique des données fonctionnelles" le jeudi toutes les deux semaines
- sur le site de l'IRISA : les séminaires et groupes de travail réguliers des projets se poursuivront et seront ouverts à l'ensemble de la communauté.
- sur le site de l'ENS : à terme des groupes de travail y seront organisés. Dans un premier temps, il sera mis sur pied un programme d'invitation de conférenciers extérieurs sur des sujets généraux touchant à la Statistique, et s'adressant, en particulier, aux jeunes en Master ou même en licence.

Mise sur pied d'un programme d'invitation de visiteurs étrangers de durée de 1 à 3 mois. Ils seront accueillis sur l'un des sites, mais interféreront avec l'ensemble des sites. Ils assureront des conférences/mini-cours.

*Formation* : Augmentation de l'offre de cours en Statistique dans le Master rennais de Mathématiques et dans l'Ecole doctorale MATISSE (Mathématiques, Informatique, Signal, Electronique). Propositions de stages aux élèves normaliens et magistériens, à divers stades de leur formation, et de sujets de thèses

Utilisation des moyens existants de financement pour accueillir des postdocs et co-encadrer des thèses multi-pays.

Développement de collaborations industrielles, d'encadrement de thèses CIFRE : plusieurs des établissements sont déjà impliqués dans de telles actions. Un effort systématique sera fait pour les augmenter et pour attirer les propositions de collaborations. La création du pôle et l'accroissement de la visibilité du potentiel en Statistique devrait le favoriser. Ceci rejoint la politique de l'IRMAR qui veut intensifier ses actions contractuelles.

Des démarches seront entreprises pour attirer des jeunes sur des postes de chercheurs CNRS ou des chercheurs plus confirmés (ceci va dans le sens du poste de Directeur de 1<sup>ère</sup> classe CNRS affiché en Statistique sur cinq laboratoires dont l'IRMAR).

Mise en commun des moyens de calcul et logiciels : chaque site a bien sûr son propre environnement informatique. Une mise en commun des moyens sera encouragée par une aide à la communication entre les sites, par l'ouverture d'accès aux différents serveurs d'un site à l'autre, par le partage de logiciels.

### ***3.e Impact escompté***

La création de ce pôle entraînera une *très grande visibilité de la Statistique sur Rennes* : l'équipe de Statistique ainsi créée au sein de l'IRMAR devrait atteindre 34 membres permanents. Il faut y ajouter les membres non permanents qui s'y associeront (doctorants, visiteurs) ainsi que la participation de chercheurs de l'IRISA. Cela constituera l'un des pôles numériquement les plus importants en France, avec une visibilité accrue par le regroupement sous un même pôle, une identification unique, et une volonté affichée de travailler tous ensemble.

Ce pôle a donc un rôle national, voire international, à jouer à côté du pôle parisien ou des centres de Province comme Toulouse, Grenoble, Marseille, Montpellier, Lille,... où la Statistique est présente de longue date ou plus récemment.

Une autre force est la diversité des thèmes présents qui balaient l'essentiel des problématiques de la Statistique avec un ancrage très fort sur les applications et des orientations exceptionnellement diversifiés pour un même site grâce à la présence d'établissements aussi spécifiques et renommés que l'ENSAI, Agrocampus, l'IRISA et la présence des deux universités rennaises : c'est aussi une chance pour l'ENS de pouvoir piloter cette opération.

**4. Demande de financement.** Comme il est explicité ci-dessus, un des volets essentiels du projet est la fédération, la structuration et le développement d'actions communes entre différentes équipes qui disposent déjà d'environnements de travail opérationnel. Les besoins en financement concernent donc surtout le soutien au fonctionnement des nouvelles actions envisagées, ainsi qu'un complément d'équipement.

AIDE AU FONCTIONNEMENT - TOTAL : 26 Keuros/an = 104 Keuros

Organisation annuelle des Journées STAR : 8 Keuros/an=32 Keuros

Invitation de chercheurs étrangers de haut niveau à raison de 3 mois par an : 3x 3 Keuros/an=36 Keuros

Organisation des séminaires et groupes de travail sur les différents sites (environ 5) et séjours des conférenciers de 1 jour à une semaine : 6 Keuros/an x4=24 Keuros

Organisation de 4 workshops de 2 jours : 3 Keuros x4=12 K euros

EQUIPEMENT - TOTAL : 46 Keuros

Acquisition de serveurs et outils de calcul sur les différents sites et de logiciels (36 Keuros)

Aide à l'acquisition de livres et de revues de Statistiques : 2,5Keuros :an= 10 Keuros/an.

## ANNEXE.

### Situation actuelle : Liste des membres des équipes

#### Laboratoire de Statistique de Rennes 2.

[http://www.uhb.fr/sc\\_sociales/labstats/](http://www.uhb.fr/sc_sociales/labstats/)

Jacques BENNASSENI	Professeur
Mohamed BENNANI-DOSSE	Maître de Conférences
Gildas BROSSIER	Professeur
Michel CARBON	Professeur (détaché à l'ENSAI)
Pierre-André CORNILLON	Maître de Conférences
Marc DAVID	Maître de Conférences
Dominique DEHAY	Professeur
Magalie FROMONT	Maître de Conférences
Arnaud GUYADER	Maître de Conférences
Agnès HAMON	Maître de Conférences
Eric MATZNER-LOBER	Maître de Conférences
Myriam MAUMY	Maître de Conférences
Alain MOM	Maître de Conférences
Nouveau ?	Maître de Conférences
Kairou MOUMOUNI	ATER

#### Laboratoire de Statistique et Modélisation, CREST-ENSAI

<http://www.ensai.com/fr/maj-e/c1a2i6861/recherche/crest/crest.htm>

Michel CARBON	Professeur (en détachement de Rennes2)
Jean-Yves DAUXOIS	Maître de Conférences INSEE
Michel DELECROIX	Professeur INSEE
Pierre DRUILHET	Maître de Conférences INSEE
Marian HRISTACHE	Maître de Conférences INSEE
Salim LARDJANE	Maître de Conférences (CMC)
Valentin PATILEA	Maître de Conférences (CMC)
Denys POMMERET	Maître de Conférences (CMC)
Sophie SENCEY	Attaché
Vincent LEFIEUX	Doctorant
David VARRON	Doctorant

#### Laboratoire de Statistique d'Enquête, CREST/ENSAI

<http://www.ensai.com/fr/maj-e/c1a2i6861/recherche/crest/crest.htm>

Jean-Claude DEVILLE	Inspecteur Général de l'INSEE
---------------------	-------------------------------

Mohamed EL HADJ TIRARI  
Guillaume CHAUVET

Chargé de Mission ENSAI  
Attaché INSEE et Doctorant

### **Laboratoire de Mathématiques Appliquées d'Agrocampus Rennes**

[http://www.agrocampus-rennes.fr/scripts/fr/C\\_equipe/si/math/](http://www.agrocampus-rennes.fr/scripts/fr/C_equipe/si/math/)

David CAUSEUR	Maître de Conférences
Ludovic DESTAMPES	Maître de Conférences
François HUSSON	Maître de Conférences
Radwan JALAM	Maître de Conférences
Jérôme PAGES	Professeur
Nouveau ?	Maître de Conférences
Elisabeth MORAND	Doctorante

### **Université de Rennes1/INSA et IRMAR, équipe de Statistique**

<http://www.math.univ-rennes1.fr/irmar/>

Philippe BERTHET	Maître de Conférences
Jean DEHAYES	Professeur
Bernard DELYON	Professeur
Jian-Feng YAO	Professeur
Benoîte De SAPORTA	ATER
Pierre ALLIOT	Doctorant

Cette liste va évoluer car une réorganisation est en cours dans l'équipe actuelle "Processus Stochastiques et Statistique"

IRISA : nous indiquons ci-dessous les projets concernés par la Statistique, soit dans leur méthodologie, soit dans leurs applications, et à des degrés divers. Nous indiquons à chaque fois le titre du projet et le nom du responsable.

Equipe SISTHEM: inférence statistique pour la surveillance d'intégrité de structures ; Michèle BASSEVILLE

Equipe ASPI : applications statistiques des systèmes de particules en interaction ; François LEGLAND

Projet DISTRIBCOM : algorithmes itératifs et distribués pour la gestion de réseaux et le traitement du signal en télécommunications ; Albert BENVENISTE ; (projet commun avec l'ENS Cachan, co-dirigé par Claude JARD)

Projet ARMOR : architectures et modèles de réseaux ; Gerardo RUBINO

Projet VISTA : vision spatio-temporelle et apprentissage ; Patrick BOUTHEMY

Projet METISS : modélisation et expérimentation pour le traitement des informations et des signaux sonores ; Frédéric BIMBOT

D'autres projets accueillent aussi des experts en Statistique comme le projet TEXMEX (techniques d'exploitation des documents multimédias) et le projet SYMBIOSE (systèmes et modèles biologiques, bioinformatique et séquences).

### **ENS Cachan, Antenne de Bretagne**

<http://www.bretagne.ens-cachan.fr/>

Comme il a été dit, il n'y a pas actuellement de Statisticien à l'antenne de Bretagne de l'ENS Cachan. Notons cependant que Arnaud DEBUSSCHE est rattaché à 50% à l'équipe "Processus Stochastique et Statistique" de l'IRMAR. L'équipe de mathématiques de l'antenne de l'ENS se compose de 2 professeurs, une maître de conférences, un AGPR, un ATER, 3 doctorants, et deux chargés de recherche CNRS de l'IRMAR y ont géographiquement rattachés.