

## 6 Resumen de la investigación

### 7 Antecedentes

#### 7.1 Estado actual del conocimiento

En los últimos tiempos ha adquirido un vigoroso desarrollo el conjunto de técnicas conocidas como de *Aprendizaje Automático*, que pueden ser encontradas en la literatura científica bajo distintos denominaciones (Statistical Learning, Machine Learning, Classification, Pattern Recognition, etc.). Desde la óptica de un estadístico, podría bien decirse que los objetivos del aprendizaje automático no son otros que los problemas más clásicos de la estadística (Regresión, Clasificación, etc.), pero con la peculiaridad de hacer más fuerte uso de procedimientos algorítmicos, computacionales, que permitan trabajar bajo modelizaciones más amplias y flexibles. Es incluso también evidente que en parte de la literatura en esta temática se plantean a menudo algoritmos basados en heurísticas, que no siempre van acompañados de un tratamiento teórico riguroso (ver Hastie et al. (2001)).

En una descripción muy somera de esta óptica, al menos en el caso del llamado aprendizaje supervisado, podrían señalarse los siguientes elementos:

Se dispone de un conjunto de  $n$  observaciones de variables  $(X_i, Y_i)$ , en las que las  $X$  son variables *explicativas* (observables) y las  $Y$  son las variables a explicar.

Se “entrena” a partir de estos  $n$  datos un algoritmo (es decir, se elige una función  $f$  dentro de una clase funcional determinada  $\mathcal{F}$ ) que permita predecir el valor de  $Y$  con el menor error posible (de acuerdo a cierto criterio de penalización del error). Las distintas selecciones de la clase  $\mathcal{F}$  y de los criterios de penalización dan lugar a métodos muy diferentes, que conforman una panoplia muy variada de técnicas. En los casos en que existe un sustento teórico riguroso que permita probar resultados de consistencia u optimalidad, la misma viene de la teoría de leyes límite uniformes, desarrollada, entre otras referencias notables por Vapnik: Es así que nociones de entropía, dimensión de Vapnik-Chervonenkis (VC) y otros conceptos afines aparecen como una forma de medir la “talla” de la clase de modelos  $\mathcal{F}$ , permitiendo probar, cuando dicha “talla” es moderada, que las leyes límite uniforme requeridas son válidas (ver Vapnik(1996)).

Para ser más precisos:

El “error” cometido al estimar  $Y$  por  $f(X)$ , donde  $f$  es la función elegida tras el entrenamiento empírico, tendrá dos componentes:

- (a) Un error estadístico puro; es la diferencia entre la performance de  $f(X)$  (el modelo elegido a partir de los  $n$  datos) y  $f^*(X)$  que es el óptimo elemento que se puede obtener en  $F$ , si se conoce la ley exacta de los datos.
- (b) Un error modelístico puro o de aproximación, que es la diferencia entre  $f^*(X)$  (el óptimo elemento de  $f$ ) y el óptimo universal  $f^{**}(X)$  (el que se obtiene cuando  $F$  es la clase de todas las funciones medibles).

Una clase  $\mathcal{F}$  “no muy grande” (en términos de dimensiones VC, por ejemplo), permitirá que el error estadístico puro tienda a cero cuando  $n$  tienda a infinito. Una clase  $\mathcal{F}$  “suficientemente grande” permitirá que el error de aproximación sea suficientemente pequeño (ver Cucker y Smale (2002)).

El problema central reside entonces en la justa elección de la clase  $\mathcal{F}$ , de manera de proveer al mismo tiempo los requerimientos de (a) y (b). Por ejemplo, Redes Neurales, CART, MARS, SVM, son algunos de los nombre frecuentes en la terminología técnica que establecen distintas aproximaciones a este enfoque, en los cuales se postulan algoritmos que apuntan a encontrar la solución de compromiso entre el error estadístico y el de aproximación (no en todos los casos existen demostraciones rigurosas que apoyen las propiedades de los algoritmos, en algunos casos se dispone exclusivamente de la performance empírica de los mismos frente a bases de datos profusamente estudiadas) (ver Hastie et al. (2001)).

## 7.2 Del equipo investigador en relación a la propuesta

El equipo del proyecto está integrado por Ernesto Mordecki (responsable) Gonzalo Perera, Gustavo Guerberoff e Isabel Cañette (investigadores de la Universidad de la República), Felipe Cucker (City University of Hong-Kong), Soledad Torres (Universidad de Valparaíso) Antoine Lejay (INRIA, Francia). Raúl Tempone (Universidad de Austin, Texas), y Federico De Olivera (estudiante de la Maestría en Ingeniería Matemática).

Como se verá en la sección siguiente, el objetivo general del proyecto es la producción teórica de métodos de solución de problemas estadísticos basados en el aprendizaje automático. Este objetivo se materializa en cuatro líneas de trabajo interconectadas, a saber: Fundamentos de aprendizaje automático, Aprendizaje automático en telecomunicaciones Aprendizaje automático en predicción de eventos climáticos. Métodos numéricos adaptivos en ecuaciones diferenciales estocásticas. A continuación se listan entonces las publicaciones de los integrantes del proyecto en las temáticas referidas.

Aspirot L., Bermolen P, Belzarena P., Ferragut A., Perera G., Simon M. Quality of Service Parameters and Link Operating Point Estimation Based on Effective Bandwidth (por aparecer en Performance Evaluation) (2003).

Bellanger L., Perera, G. High-level exceedances of non-stationary processes and irregular sets. Comptes Rendues, Académie des Sciences, Paris, t. 328, Série I, p. 337-342 (1999).

Bellanger L., Perera, G. Compound Poisson Limit Theorems for high-level exceedances of some non-stationary processes. Bernoulli, Vol. 9, No. 3, 497-515 (2003).

Cañette, Isabel Blind nonparametric regression. Can. J. Stat. 29, No.2, 173-189 (2001)

- Cheung, Dennis; Cucker, Felipe Learning from rounded-off data. *Inf. Comput.* 182, No.1, 1-13 (2003).
- Cucker, Felipe; Smale, Steve On the mathematical foundations of learning. *Bull. Am. Math. Soc., New Ser.* 39, No.1, 1-49 (2002).
- Cucker, Felipe; Smale, Steve Best choices for regularization parameters in learning theory: On the bias-variance problem. *Found. Comput. Math.* 2, No.4, 413-428 (2002).
- De Olivera, Federico. Clasificación y predicción del fenómeno de “El Niño”. Pasantí a. Licenciatura en Estadística. Facultad de Ciencias Económicas y de Administración (2003)
- Durañona and Perera. Extreme wind speed distribution of WSW direction for Montevideo, Uruguay. *Proceedings of the 11th Conference on Wind Engineering*, June 2-5, 2003, Lubock, Texas, USA, pp. 229–236.
- Fierro, Raúl; Torres, Soledad The Euler scheme for Hilbert space valued stochastic differential equations. *Stat. Probab. Lett.* 51, No.3, 207-213 (2001).
- Ghattas, B., Perera, G. A Memory-Restricted Learning Algorithm, *Prépublications IML, CNRS, Marseille* (2004).
- Ma, Jin; Protter, Philip; San Martín, Jaime; Torres, Soledad Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* 12, No.1, 302-316 (2002).
- Lejay, Antoine Weak solution of semi-linear PDE, BSDE and homogenization. *Monte Carlo Methods Appl.* 7, No.3-4, 265-272 (2001)
- Ma, Jin; Protter, Philip; San Martín, Jaime; Torres, Soledad Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* 12, No.1, 302-316 (2002).
- K-S. Moon, A. Szepessy, R. Tempone and G. Zouraris, Convergence Rates for Adaptive Weak Approximation of Stochastic Differential Equations, *Stochastic Analysis and Applications* (to appear)
- Pechiar J., Perera G., Simon M. Effective bandwidth estimation and testing for Markov sources. *Performance Evaluation* 945,p. 1-19. (2001).
- San Martín, Jaime; Torres, Soledad Euler scheme for solutions of a countable system of stochastic differential equations. *Stat. Probab. Lett.* 54, No.3, 251-259 (2001).
- Severov D.N., Mordecki E., Pshennikov V.A. SST anomaly variability in Southwestern Atlantic and El Niño/Southern oscillation *Advances in Space Research* 33 (2004) 343–347.

- A. Szepessy, R. Tempone and G. Zouraris. Adaptive weak approximation of stochastic differential equations, *Comm. Pure and Appl. Math.*, pp. 1169-1214, Volume 54, Issue 10, 2001, New York.

## 8 Objetivos de la investigación

Podemos destacar un objetivo general de la investigación consistente en comenzar el trabajo de investigación en una temática nueva en el país, el aprendizaje automático en procesos estocásticos, que se materializa a través de cuatro líneas de trabajo concretas

Fundamentan esta propuesta la carencia señalada en la exposición sobre el estado actual del conocimiento consistente en la escasa consideración que se ha dado al caso en que  $X$  es dimensionalmente enorme, o, más precisamente, infinito dimensional. Por ejemplo, en Ghattas-Perera (2004) se estudia una versión muy primaria del problema: técnicas de aprendizaje asistidas por expertos cuando  $X$  (llamado a menudo el input) es de dimensión muy alta, inspirados en problemas de ruteo en redes de telecomunicaciones. En dicho trabajo, más que aportarse la solución, se esboza claramente cuál es el problema. Si el input ocupa un enorme espacio en memoria es inviable confiar en el principio de "entrenar bien gracias a grandes masas de datos", porque las grandes masas de datos de gran dimensión resultan inmanejables. En la literatura existen abundantes referencias a procedimientos adaptativos, secuenciales, que buscan implementar los cálculos paso a paso, como forma de lidiar con la alta dimensionalidad de los inputs (en la web de COLT, citada en las referencias, es posible hallar muy numerosas referencias al respecto). Sin embargo, a menudo estos procedimientos son inestables frente a variaciones en la estructura probabilística de los datos y están completamente referidos a una aproximación algorítmica concreta.

En términos más tradicionalmente estadísticos, suelen ser escasamente robustos, en la medida que usan al máximo la información específica disponible de manera de instrumentar cálculos de manera eficiente y económica. El balance entonces entre la economía del esfuerzo de cálculo y la solidez de los resultados, es a nuestro juicio un desafío pendiente frente a datos de altas dimensiones. Los comentarios anteriores adquieren un nivel superlativo si "altas dimensiones" deviene "dimensión infinita". Uno de los ejemplos más sobresalientes del caso en que  $X$  es infinito dimensional, lo constituye el caso en que  $X$  es un proceso estocástico, con una cierta estructura probabilística, e  $Y$  es un funcional de  $X$ .

Este problema, ha sido a nuestro juicio, insuficientemente abordado desde la óptica del aprendizaje automático y es un verdadero melting pot de infinidad de

problemas de Estadística de Procesos Estocásticos, Control Estocástico, Aprendizaje Automático, etc. Este grupo de trabajo tiene como objetivo integrar bajo la metodología del aprendizaje automático en dimensión infinita diversos problemas en procesos estocásticos.

Es por esto que proponemos el estudio del aprendizaje automático en procesos estocásticos mediante el abordaje de cuatro direcciones complementarias, cada una con sus objetivos concretos determinados:

**Fundamentos en aprendizaje automático** Se trata de obtener principios generales, resultados estructurales, estimaciones de error estadístico y de aproximación, para el problema de aprendizaje supervisado, cuando el input es infinito dimensional (o su proyección a dimensiones muy altas), con la peculiaridad de resultar en criterios prácticamente implementable sobre los ejemplos concretos que se consideran en este proyecto (es decir, resultados estructurales generales, pero que brinden condiciones suficientes razonablemente aplicables en ejemplos específicos). Entre otras posibles líneas de abordajes cabe señalar la extensión a procesos del enfoque bayesiano no-paramétrico (ver Zribi y Ghorble (2003)).

**Aprendizaje en telecomunicaciones.** Una señal que es un proceso estocástico, y un red de datos es el ámbito donde confluyen muchas señales. El problema consiste en tomar una decisión con la información de las señales recibidas, modeladas mediante procesos estocásticos, tales como la aceptación de las tareas demandadas a la red (lo que en la jerga técnica se denomina CAC, Control Acceptancy Criteria), así como, una vez ingresadas las tareas a la red, seleccionar la ruta óptima a afectar a la tarea desde su origen a su destino (lo cual en la jerga técnica se denomina "routing") (ver, por ejemplo, Daskalaki & Mac Gregor (2004), Holmberg et al (2004)).

El grupo ARTES, conformado por ingenieros eléctricos y matemáticos, en el que trabaja uno de los integrantes de este proyecto (Gonzalo Perera) viene estudiando sistemáticamente este tema (ver, por ejemplo, Aspirot et al., Pechiar et al.), desde la óptica de la utilización de técnicas de grandes desvíos para la resolución de dichos problemas en base a minimización de probabilidades de pérdida de trabajos ingresados. Lo que se propone aquí es una visión alternativa: la instrumentación algorítmica de un mecanismo óptimo de CAC o routing, a través del entrenamiento por evaluación durante una fase de aprendizaje del funcionamiento de la red. La dimensionalidad del problema puede considerarse o bien infinita o bien finita pero de tala astronómica, de ahí la importancia del punto antecedente.

**Aprendizaje en predicción de fenómenos climáticos** A partir de estudios estadísticos sobre fenómenos climáticos realizados, se intentará abordar la problemática de la predicción de fenómenos climáticos a partir de datos masivos. En particular, se estudian dos situaciones concretas de predicción a partir de datos climáticos:

- (a) Predicción del fenómeno del Niño (El Niño Southern Oscillation - ENSO) a partir de datos de anomalías en las temperaturas en las corrientes marinas de Humboldt, Malvinas y Brasil. Este trabajo es continuación de Severov et al. (2004), en donde se estudia la relación del fenómeno con las temperaturas en el Atlántico (Malvinas y Brasil). En De Olivera (2003) se propuso una reducción en la dimensionalidad de los datos para la predicción del fenómeno.
- (b) Determinación de la proyección finito dimensional óptima y de previsión de eventos extremos sobre procesos climáticos, a partir del análisis de trazas históricas, como continuación de Durañona-Perera (2003). El análisis de vientos extremos, como el emprendido con las denominadas sudestadas en el trabajo aquí referido, es nuevamente un estudio de un gran desvío, frente a registros de datos prácticamente on-line (en nuestro país, hay estaciones de registro de vientos reteniendo los últimos 10 minutos de cada hora). Mas allá de la diferencia considerable de contexto, hay similitud significativas, en términos matemáticos, entre este problema y el de redes de datos, por lo cual se propone utilizar aquí un abordaje similar. Concretamente, el objetivo específico sería diseñar un índice de riesgo de extremos en los vientos en las distintas direcciones (o al menos en las más significativas), basado en el aprendizaje automático sobre la historia previa. Un trabajo con dicho enfoque fue realizado para la modelización de altos picos de polución por ozono en distintas regiones de Francia y algunas técnicas tradicionales de aprendizaje (CART, por ejemplo) fueron utilizadas en mixtura con procedimientos típicos de tratamiento de datos extremos (ver Bel et al.(1999), Bellanger (1999), Bellanger & Tomassone (2004), Ghattas (1999 a), b), c) y 2000),Ghattas et al (2000)), con la participación de uno de los integrantes del equipo de nuestro proyecto en la obtención de resultados teóricos utilizados en dichos trabajos (ver Bellanger & Perera (1999, 2003)). Aquí se propone el abordaje del problema en su totalidad, lo cual es altamente factible que requiera resultados teóricos adaptados al problema específico, como ocurriera en el antecedente citado.

### **Métodos numéricos adaptivos en ecuaciones diferenciales estocásticas**

El objetivo se centra en obtener cotas a priori para el error en métodos numéricos para resolver ecuaciones diferenciales estocásticas con saltos, tanto del tipo clásico. Aquí aparecen también los dos errores, el error modelístico debido a que la simulación no es exacta, no es posible simular en forma exacta la solución de una ecuación diferencial estocástica, y el error estadístico, que se debe a la imposibilidad de simular infinitas muestras. Los trabajos a realizarse están inspirados en A. Szepessy et al. (2001). En principio, en esta dirección está planteado el análisis de ecuaciones diferenciales estocásticas con saltos, y, en segundo lugar, las ecuaciones diferenciales con saltos retrógradas.

## 9 Metodología y plan de trabajo

Los métodos de trabajo son los habituales en la investigación matemática, es decir, el estudio, el intercambio con otros investigadores y la reflexión personal y colectiva. En consecuencia, se requiere destinar tiempo y recursos a viajes al exterior de los miembros del equipo humano así como invitar a investigadores distinguidos a realizar estancias en Uruguay para discutir sobre los temas a consideración.

Se debe destacar que cada uno de esos temas es objeto de trabajo intenso en diversos centros matemáticos de otros países y que la investigación de frontera requiere, por lo tanto, esa interacción personal, que no es posible reemplazar por otros medios a los que también habremos de recurrir como es normal (intercambio de publicaciones, diálogo electrónico, etc.).

## 10 Cronograma de actividades y resultados esperados

Las actividades a realizar también son las habituales en la investigación matemática: el estudio personal, el diálogo y la reflexión conjunta. Se debe destacar que esta investigación se desarrolla, en nuestra Universidad, en el marco de los Seminarios permanentes de Probabilidad y Estadística Matemática, que tienen lugar en el Centro de Matemática, en el Laboratorio de Probabilidad y Estadística, y en el Instituto de Estadística de la Facultad de Ciencias Económicas. Ello permite la participación regular, el contacto y el estudio de jóvenes investigadores en formación. La relación con colegas del exterior tiene lugar a través de visitas mutuas, y de la participación en reuniones científicas.

Un objetivo concreto a destacar es el dictado de un curso de Postgrado en la temática general a cargo de Gonzalo Perera, en primer semestre del 2005 (esto debería ser la ocasión además de acercar nuevos estudiantes de posgrado a la temática).

La planta física para el desarrollo del proyecto es la del Centro de Matemática de la Facultad de Ciencias, que cuenta con lugares de trabajo adecuados. Los otros dos aspectos básicos son la Biblioteca y el equipamiento informático.

La Biblioteca de Matemática está localizada en los locales de las Facultades de Ciencias y de Ingeniería, siendo operada de manera conjunta tanto en lo referente a adquisiciones, conexiones con bases de datos, personal técnico, préstamo y otros servicios. Si bien actualmente cuenta con la mayoría de las revistas necesarias para el desarrollo de estas investigaciones, hay una carencia significativa en materia de libros y también de algunos títulos de revistas. En consecuencia, una parte significativa de los fondos del proyecto estará destinada a la adquisición de libros y revistas directamente relacionados con los temas de investigación.

Existe equipamiento informático, que incluye los equipos localizados en el Centro de Matemática de la Facultad de Ciencias y en el Instituto de Matemática y Estadística "Rafael Laguardia" de la Facultad de Ingeniería, pertenecientes

al Laboratorio de Probabilidad y Estadística. A los efectos de la realización del proyecto, se requiere complementar los equipos y además adquirir software adecuado para el tratamiento de algunos problemas, como la simulación de procesos estocásticos de parámetro continuo, que es muy exigente en ese plano.

En lo que refiere a la financiación solicitada al proyecto, se financiará la participación de estudiantes avanzados de grado, y de posgrado, mediante la contratación y dedicación compensada.

Las compras bibliográficas se realizarán en cuanto estén disponibles los fondos, por los mecanismos usuales de la Biblioteca de Matemática del Centro de Matemática.

En lo que respecta a los seminarios, se planifica organizar el “Seminario de Probabilidad y Estadística” en forma semestral, como una actividad coordinada de la Facultad de Ciencias (sede del proyecto), junto con las Facultades de Ingeniería y la de Ciencias Económicas y Administración.

Se planea además continuar con el “Seminario de Estudio en Probabilidad y Estadística” en la Facultad de Ciencias durante todos los 4 semestres del proyecto.

## 11 Impacto del proyecto

### 11.1 Como generador de conocimientos

Como suele ocurrir normalmente con los trabajos de investigación en Matemática, es difícil evaluar a priori el verdadero impacto y los verdaderos beneficios que pueden derivarse de cumplir con los objetivos del proyecto propuesto. Al mismo tiempo, dado que se trata de un trabajo de investigación, es muy posible que el itinerario real sea diferente al trazado al formular el proyecto y que, por lo tanto, el impacto y los beneficios difieran de los que podemos prever actualmente.

Desde nuestra óptica presente, en cada uno de los capítulos que hemos enumerado se debe esperar que:

- Avanzar en la comprensión general de la temática del aprendizaje, consolidando un equipo humano, integrado básicamente por uruguayos, que trabaje en la temática.
- En el área de las telecomunicaciones, implementar técnicas robustas de ruteo y aceptación, aumentando el conocimiento tecnológico específico sobre este ámbito.
- En el área estadística, producir métodos de predicción de fenómenos climáticos que mejoren (en cuanto a la probabilidad de acierto) los métodos existentes, contribuyendo a conformar un grupo de trabajo local altamente calificado e interdisciplinario.
- En cuanto a los métodos adaptivos, producir algoritmos adaptivos con errores controlables a priori para ecuaciones diferenciales estocásticas con

saltos, así como, en un segundo plano, para ecuaciones diferenciales estocásticas con saltos retrógradas.

## **11.2 Como formador de recursos humanos**

Más allá de los aspectos estrictamente técnicos, el impacto del proyecto en nuestro ambiente académico también implicaría las siguientes consecuencias:

- Permitiría incorporar jóvenes (estudiantes de postgrado, especialmente) a temas de frontera en la investigación matemática.
- Ayudaría a mantener la interacción entre la comunidad matemática uruguaya y nuestros colegas de otros países, aspecto central de la calidad del trabajo científico en nuestra disciplina.

## **11.3 Como factor fortalecedor de la institución en que se desarrolla el proyecto**

- El proyecto contiene un fuerte énfasis en temas multidisciplinarios y de aplicaciones no triviales a otras disciplinas y ayudaría a intensificar el relacionamiento académico - tanto en la investigación como en la enseñanza - con otras áreas del conocimiento.
- Tendría un impacto positivo en cuanto a la adquisición de material bibliográfico. Debe destacarse que contar con material actualizado es una condición necesaria para el éxito de cualquier investigación científica en estas disciplinas.

## **12 Factibilidad de alcanzar los resultados con las facilidades actuales y el apoyo solicitado**

Los objetivos a alcanzar son de diversa índole. Algunos son muy concretos, y el proyecto coadyuvaría a su realización, como el dictado de cursos de posgrado. En este caso, la financiación del proyecto permitiría incluir en el dictado del curso la participación de docentes del extranjero. En el plano científico de producción de conocimiento, los resultados que se plantean son en su mayoría continuación de resultados ya obtenidos por el equipo de investigación, lo que hace presumible su realización. Tanto la presencia de bibliografía actualizada, la visita de profesores del extranjero, como la formación de recursos humanos son elementos catalizadores, que elevan la calidad de la producción científica, dándole posibilidades de continuidad.

## **13 Multidisciplinariedad**

El proyecto presentado, si bien está centrado en el área de la Probabilidad y Estadística, tiene elementos de multidisciplinariedad provenientes de: (i)

la informática dado que el “learning” es estudiado desde ambas áreas, y la propuesta tiene un énfasis importante en la producción de algoritmos concretos para la resolución de problemas; (ii) las ciencias naturales y en la ingeniería, en donde se propone aplicar la metodología estadísticas relacionadas con el learning, (iii) el análisis numérico, dado que en los problemas de aproximación numérica de ecuaciones diferenciales estocásticas los “datos” no provienen de la realidad sino de la simulación computacional.

## **14 Justificación de los recursos**

Se tomó el criterio de priorizar la financiación de aquellas actividades y adquisiciones que, siendo importantes para el proyecto, no estuviesen aseguradas por otras fuentes de financiamiento: en particular la compra de libros y revistas, la visita de profesores extranjeros (incluyendo únicamente las estadías) y la formación de recursos humanos.

### **14.1 Justificación de equipos solicitados**

No se solicita compra de equipos.

### **14.2 Justificación de viajes y estadías**

Se solicita financiación para viabilizar la visita al país por períodos breves de cuatro investigadores (dos de ellos de uruguayos trabajando en el extranjero) asociados al proyecto, mediante el pago de viáticos.

### **14.3 Justificación de los requerimientos del equipo de investigación**

#### **14.3.1 Extensión horaria**

No se solicitan extensiones horarias.

#### **14.3.2 Complemento salarial**

No se solicitan complementos salariales.

#### **14.3.3 Personal nacional a contratar**

Se solicita la contratación de un joven durante toda la realización del proyecto, con la idea de poder apoyarlo en su formación de posgrado. La contratación se realizará mediante una llamada del Consejo de la Facultad de Ciencias.

#### 14.3.4 Personal administrativo y de apoyo

No se solicita apoyo específico. El personal administrativo de la Facultad de Ciencias que da apoyo administrativo a los proyectos está contemplado en el canon del 24.000 pesos uruguayos.

#### 14.3.5 Colaboradores extranjeros

No se solicita

### Referencias bibliográficas

- Aspirot L., Bermolen P, Belzarena P., Ferragut A., Perera G., Simon M. Quality of Service Parameters and Link Operating Point Estimation Based on Effective Bandwidth (por aparecer en Performance Evaluation) (2003).
- L. Bel, L. Bellanger, V. Bonneau, G. Ciuperca, D. Dacunha-Castelle, C. Deniau, B. Ghattas, M. Misiti, G. Oppenheim, J.M. Poggi, R. Tomassone (1999) " Eléments de comparaison de prévisions statistiques des pics dozone " Revue de Statistique Appliquée. Vol. XLVII (3), 7-25.
- Bellanger L. Statistique de la pollution de l'air. Méthodes mathématiques. Applications au cas de la région parisienne. Thèse doctorat, Université Paris XI, (1999).
- Bellanger L., Perera, G. High-level exceedances of non-stationary processes and irregular sets. Comptes Rendues, Académie des Sciences, Paris, t. 328, Série I, p. 337-342 (1999).
- Bellanger L., Perera, G. Compound Poisson Limit Theorems for high-level exceedances of some non-stationary processes. Bernoulli, Vol. 9, No. 3, 497-515 (2003).
- Bellanger L., Tomassone R. Trend in high tropospheric ozone levels. Application to paris monitoring sites. Statistics, June 2004, vol. 38, iss. 3, pp. 217-241(25) Taylor and Francis Ltd.
- Cañette, Isabel Blind nonparametric regression. Can. J. Stat. 29, No.2, 173-189 (2001)
- Cheung, Dennis; Cucker, Felipe Learning from rounded-off data. Inf. Comput. 182, No.1, 1-13 (2003).
- COLT (Computational Learning Theory): <http://www.learningtheory.org>
- Cucker, F., Smale. S: On mathematical foundations of learning. Bulletin of the American Mathemaical Society 39, No. 1, 1-49 (2002).

- Cucker, Felipe; Smale, Steve Best choices for regularization parameters in learning theory: On the bias-variance problem. *Found. Comput. Math.* 2, No.4, 413-428 (2002).
- Daskalaki, Sophia; MacGregor Smith, J. Combining routing and buffer allocation problems in series-parallel queueing networks. *Ann. Oper. Res.* 125, 47-68 (2004).
- De Olivera, Federico. Clasificación y predicción del fenómeno de “El Niño”. Pasantí a. Licenciatura en Estadística. Facultad de Ciencias Económicas y de Administración (2003)
- Durañona and Perera. Extreme wind speed distribution of WSW direction for Montevideo, Uruguay. *Proceedings of the 11th Conference on Wind Engineering*, June 2-5, 2003, Lubock, Texas, USA, pp. 229–236.
- Fierro, Raúl; Torres, Soledad The Euler scheme for Hilbert space valued stochastic differential equations. *Stat. Probab. Lett.* 51, No.3, 207-213 (2001).
- B. Ghattas (1999a) Prédiction des pics de ozone par arbres simples et agrégés par bootstrap, *Revue de Statistique Appliquée*, Vol. XLVII (2), pp. 61-80.
- B. Ghattas (1999b) Prédiction par arbre de classification, *Mathématiques Informatique et Sciences Humaines*, N 142, pp. 31-49.
- B. Ghattas (1999c) Méthodes non paramétriques pour la prédiction de l’ozone, *OCEANIS*, N 24 .
- B. Ghattas (2000) Agrégation d’arbres de classifications *Revue de Statistique Appliquée*, Vol. XLVIII (2), pp. 85-98.
- Ghattas, B., Perera, G. A Memory-Restricted Learning Algorithm, *Prépublications IML, CNRS, Marseille* (2004).
- B. Ghattas, L. Mary, P. Renzy, D. Robin, (2000) ” Prédiction de l’ozone dans l’aire Métropolitaine Marseillaise, par des méthodes non paramétriques”, *Pollution Atmosphérique*.
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. *The elements of statistical learning. Data mining, inference, and prediction.* Springer Series in Statistics. New York, NY(2001).
- Holmberg, Kaj; Di Yuan. Optimization of internet protocol network design and routing. *Networks* 43, No.1, 39-53 (2004).
- Lejay, Antoine Weak solution of semi-linear PDE, BSDE and homogenization. *Monte Carlo Methods Appl.* 7, No.3-4, 265-272 (2001)
- Ma, Jin; Protter, Philip; San Martín, Jaime; Torres, Soledad Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* 12, No.1, 302-316 (2002).

- Ma, Jin; Protter, Philip; San Martín, Jaime; Torres, Soledad Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.* 12, No.1, 302-316 (2002).
- K-S. Moon, A. Szepessy, R. Tempone and G. Zouraris, Convergence Rates for Adaptive Weak Approximation of Stochastic Differential Equations, *Stochastic Analysis and Applications* (to appear)
- Pechiar J., Perera G., Simon M. Effective bandwidth estimation and testing for Markov sources. *Performance Evaluation* 945,p. 1-19. (2001).
- San Martín, Jaime; Torres, Soledad Euler scheme for solutions of a countable system of stochastic differential equations. *Stat. Probab. Lett.* 54, No.3, 251-259 (2001).
- Severov D.N., Mordecki E., Pshennikov V.A. SST anomaly variability in Southwestern Atlantic and El Niño/Southern oscillation *Advances in Space Research* 33 (2004) 343-347.
- A. Szepessy, R. Tempone and G. Zouraris. Adaptive weak approximation of stochastic differential equations, *Comm. Pure and Appl. Math.*, pp. 1169-1214, Volume 54, Issue 10, 2001, New York.
- Vapnik, V. (1996) *The Nature of Statistical Learning Theory*. Springer, NY.
- Zribi, Mourad; Ghorbel, Faouzi. An unsupervised and non-parametric Bayesian classifier. *Pattern Recognit. Lett.* 24, No.1-3, 97-112 (2003).