

GÉOSTATISTIQUE



La pratique de la Géostatistique est, d'abord, une occasion de rencontres: rencontres entre des champs d'application variés et parfois inattendus, entre des problématiques sans cesse renouvelées, et également une confrontation entre des objectifs et contraintes purement techniques d'une part et les exigences sociales, économiques, environnementales d'un monde complexe d'autre part. Autrement dit, tout en étant fiers de ce que le néologisme "géostatistique" jadis forgé à l'École des mines ait trouvé droit de cité dans le Petit Larousse, il est satisfaisant d'observer, au fil des ans, que l'immuable définition qu'en donne le dictionnaire s'éloigne de plus en plus de la réalité, et que notre discipline trouve à s'exprimer bien au-delà de la simple estimation des gisements miniers. De fait, dans tout domaine où des jeux de données numériques présentent une organisation spatiale ou temporelle, la Géostatistique a les outils pour apporter un éclairage original, à la fois constructif et sans concession. Il semble que cet aspect transversal et non-conformiste de la Géostatistique constitue désor-

mais son caractère dominant au regard des optionnaires, et nous ne manquons pas dans le futur de justifier cette appréciation. Ainsi, chaque année, la diversité des vœux des étudiants constitue une chance exceptionnelle de tester des méthodes nouvelles et de parcourir des domaines nouveaux, et la garantie d'insuffler un surplus de dynamisme à l'équipe encadrante. Mais la mise en œuvre d'une Géostatistique de qualité exige en permanence d'assurer un équilibre, parfois délicat, entre des exigences souvent contradictoires: garantir une rigueur théorique indispensable à la fiabilité des résultats tout en conservant un point de vue pragmatique et réaliste afin que les conclusions abstraites trouvent à s'appliquer sur le terrain. Sans oublier une indispensable déontologie, dans des domaines où souvent les contraintes économiques ou environnementales soumettent le géostatisticien à des pressions qui ne relèvent plus de la science ou de la technique...

Ouverture et équilibre: c'est dans cet esprit que nous avons continué à proposer à la promotion 2012 un voyage

de deux semaines en Guyane où, dans le contexte inhabituel et parfois tourmenté d'un DOM, les visites à des laboratoires, à des industriels et à des organismes institutionnels ont permis tout à la fois d'élargir l'horizon des optionnaires et de susciter un échange ouvert et fructueux avec nos interlocuteurs. Partie intégrante de la scolarité, la mission en Guyane constitue pour les optionnaires la phase d'initiation à la réalité du terrain.

Enfin, le souci d'ouverture s'est à nouveau exprimé au niveau des sujets de l'option. Outre la variété des champs d'application, une multiplicité de méthodes statistiques est employée. Nous avons ici l'illustration que notre démarche méthodologique peut trouver à s'appliquer dans de multiples domaines et cela souligne le caractère généraliste de l'option Géostatistique et Probabilités Appliquées, tant en ce qui concerne les champs applications abordés que les méthodes mathématiques mises en jeu.

Hans WACKERNAGEL

GEOSTATISTICS



The aim of Geostatistics is to study quantitative phenomena that are structured in space and/or time. Engineers are almost inevitably faced with problems related to geostatistical techniques such as the evaluation of natural reserves, the analysis of time series, cartography, etc., and, broadly speaking, any processing of "regionalized variables" according to the terminology of G. Matheron, the founder of geostatistics.

The Geostatistics and Applied Probability Minor gives priority to probabilistic models and statistical methods, and in particular focusses on their application to the processing of spatial and temporal data.

The training in this Minor is aimed essentially at providing students with a critical mastering of some of the tools which they might need to use during their internship. As an introduction to "geostatistics in the field", the Minor provides an opportunity of entering into contact with companies and personnel working with geostatistics in fields of application that correspond as much as possible to the themes of particular interest to the students' year. It is essential that the students acquire a sense of balance between an empirical approach and a mathematical

approach to a problem, a sense of balance which the third-year internship will put into practice in real terms.

The presentations this year again reflect the great variety of themes and applications that are encountered in this branch of applied statistics.

Hans WACKERNAGEL

GÉOSTATISTIQUE *GEOSTATISTICS*

Mardi 30 juin *Tuesday 30th June*

■ L106



Ricardo
CARRIZO VERGARA

9h-10h

Modèles de fonctions de covariance spatio-temporelles multivariées non-symétriques pour des variables climatologiques

Multivariate asymmetrical models of spatiotemporal covariance functions for climatological variables

Ces dernières années les approches géostatistiques ont été beaucoup utilisées pour l'étude du comportement des variables climatologiques. En général, ces variables présentent de fortes conditions d'interdépendance spatio-temporelle qui réclament à être décrites par des modèles de fonctions de covariances pour des fonctions aléatoires multivariées. Une grande variété de modèles ont déjà été proposés, parfois avec de bons résultats. Malgré tout, la plupart des modèles disponibles actuellement sont relativement simples (conditions de séparabilité et/ou symétrie) et ils n'arrivent pas à décrire de manière optimale les conditions de mouvement des fronts climatologiques qui, dans bien de cas, pourraient être facilement décrites en définissant une vélocité.

Le but de ce stage est d'étudier des modèles de covariance non-symétriques pour des variables de température, radiation et humidité, devant comprendre un paramètre de vélocité ou de mouvement afin de mettre en corrélation les variables liées selon ces conditions de mouvement. Il faut que ces modèles soient suffisamment riches pour décrire le phénomène de manière satisfaisante et à la fois suffisamment simples pour être déterminés par un nombre raisonnable de paramètres. On cherche aussi à développer des méthodes d'estimation de ces paramètres, en particulier pour la vélocité, et de les confronter à la riche base de données disponible à l'INRA, afin d'aider à la simulation et à la prédiction des conditions météorologiques des régions françaises.

Over the past years geostatistics has been frequently used for the study of the behavior of climatological variables. Generally these variables show strong spatio-temporal interdependency, which need to be described by models of covariance functions for multivariate random functions. A lot of models have been already developed, which sometimes yield good results. Most of these models are relatively simple (with separable and/or symmetrical conditions) and they cannot describe optimally the movement conditions of climatological fronts which, in many cases, could be easily described by defining a velocity parameter. The objective is to search and to study asymmetrical covariance models for temperature, radiation and humidity variables, which must considerate a velocity or another parameter describing the movement. These models must be rich enough for describing these phenomena satisfactorily, yet simple enough to be determined by a reasonable quantity of parameters. We also seek to develop estimation methods of these parameters, especially for the velocity, and to contrast them with the rich data base that INRA has accumulated for the simulation and prediction of the meteorological conditions of several French regions.

INRA
Avignon



François
PACAUD

10h-11h

Analyse de séries temporelles de production de puits pétroliers

Time-series analysis of production of oil wells

IDMOG est une start-up qui développe des solutions de Digital Oil Field (DOF) qui permettent à des producteurs d'hydrocarbures d'avoir à disposition en temps réel les données de production de différents champs d'hydrocarbures (débits des fluides, pressions, températures, etc.). Un problème récurrent pour l'analyse des données est cependant le manque d'instrumentation des champs, les

IDMOG is a start-up specialized in the development of Digital Oil Field (DOF) solutions, which allow oil producers to have production data available in real-time (such as flow rates, pressures, temperatures, etc.). A recurrent problem is the lack of metering in oil-fields. Wells are indeed barely well-metered. Thus, operators need to test regularly the oil wells to calibrate some measures

puits étant rarement dotés de capteurs de débit précis. Pour palier ce problème les puits sont régulièrement testés par des opérateurs pour avoir une calibration de certaines mesures (débits des fluides sortants, pression de tête du puits, etc.).

L'objectif de ce stage est d'agrèger les données disponibles pour essayer de mettre en place un Virtual Flow Meter (VFM) qui permettrait d'avoir une estimation du débit en temps réel. Dans un premier temps, des algorithmes de fouilles de données sont développés pour détecter automatiquement les outliers et les régimes de fonctionnement de chaque puits. Un modèle est ensuite établi pour pouvoir analyser de manière multivariée les différentes séries temporelles à disposition, par exemple par filtrage de Kalman. Finalement ces outils seront utilisés pour essayer de résoudre le problème de back-allocation, afin de disposer d'une estimation en temps réel du débit.

such as output flow rate, well head pressure, and so on. These measurements are used afterward to have an estimation of the working conditions for each well. The goal of this internship is to aggregate available data in order to develop an algorithm allowing to calculate oil flow rate in real-time. First, a data-mining algorithm has been developed to detect automatically outliers and detrimental situations. A model has then been established to find multivariate relations between available time series in order to apply a Kalman filter. Eventually those tools will be used to attempt to solve a back-allocation problem to have a real time estimate of the oil flow rate.

ID-MOG
Bordeaux

11h-12h

Estimation des ressources récupérables multi-variables par Krigeage Disjonctif discret

Estimation of multivariate recoverable resources by Discrete Disjunctive Kriging

Dans une exploitation minière sélective, le critère de sélection d'un bloc dépend de la valeur de son contenu. Si en complément du métal principal, la prise en compte de seuils sur un sous-produit ou un pénalisant est nécessaire pour définir cette valeur, le critère de sélection est multi-variable (par exemple la proportion d'argile est un critère important pour la lixiviation en tas, une teneur en arsenic trop élevée classe le matériel en stérile). Les méthodes d'estimation des ressources récupérables disponibles actuellement sont mono-variables et les cas multi-variables sont évalués via des simulations conditionnelles (méthode de Monte-Carlo). Une extension multi-variable du krigeage disjonctif discret est actuellement étudiée : l'approche proposée est d'établir le modèle de changement de support sur des images d'entraînement. Elle se décompose donc en deux étapes :

- Création d'images d'entraînement par simulation stochastique pour calculer la table du changement de support discret ;
- Krigeage Disjonctif déduit de l'analyse factorielle de la table de contingence simulée.

Elle permet de combiner la flexibilité des simulations non conditionnelles pour la modélisation empirique du phénomène et l'estimation par krigeage disjonctif du tonnage et du métal récupérés. L'objectif du stage est d'évaluer la pertinence opérationnelle de cette approche sur un gisement en cours de développement et pour lequel des données multi-variables et à différentes échelles de reconnaissance sont disponibles. Cette méthode prenant en compte plusieurs critères pour l'estimation des ressources récupérables doit permettre une évaluation de gisements complexes plus efficace que la méthode de Monte-Carlo traditionnelle.

In selective mining, the selection criterion of a block is the value of its content. If this value does not only depend on the main metal but also on limits on co-products or/and deleterious elements, a multivariate selection criterion should be used (e.g. shale proportion for static lixiviation, arsenic for special waste). However the methods of estimation of recoverable resources currently available only handle a univariate criterion (a cutoff grade) and multivariate cases are estimated via conditional simulations using Monte-Carlo methods. A multivariate extension of the Discrete Disjunctive Kriging is studied: the proposed approach is to use training images to establish the model for the change of support. It consists of two stages:

- The creation of training images by stochastic simulation to compute the table of the change of support
- Discrete Disjunctive Kriging using the model derived from the factorial analysis of the contingency table.

It combines the flexibility of the non-conditional simulations for the empirical modeling of the phenomenon and the efficiency of Disjunctive Kriging to estimate ore and metal. The objective of the internship is to evaluate the approach on a deposit under development. This method can be applied for the estimation of the recoverable resources on complex deposits where bloc selection results from several criteria and it should be more efficient than the method of Monte Carlo.

AREVA
La Défense



Lotfi
SLIM

Goël
AMAR

13h30-14h30

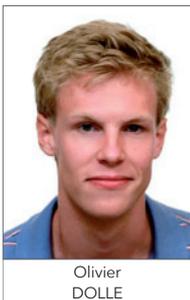
Machine learning et architecture des réseaux de neurones

Machine learning and architecture of neural networks

Les développements combinés du perceptron, des processeurs graphiques et d'importantes bases de données ont rendu possible l'élaboration de réseaux de neurones artificiels multicouches. Ces architectures utilisées dans les algorithmes de machine learning ont déjà prouvé leur supériorité sur les techniques d'apprentissage usuelles bien qu'elles manquent encore de fondements mathématiques. En particulier, encore aucun résultat ne permet d'estimer selon la nature et la quantité de données à traiter quels nombres de neurones et de couches utiliser pour obtenir le meilleur résultat possible. L'objectif de ce stage est d'étudier, d'une part, la possibilité d'optimisation non brutale de l'architecture de ces réseaux en comparant les résultats ainsi obtenus avec ceux issus d'autres méthodes, et d'apporter, d'autre part, des justifications aux choix effectués pour les réglages des paramètres de tels réseaux.

The recent developments of the perceptron, graphics processing units and data banks have allowed for the elaboration of deep neural networks. Commonly used in machine learning algorithms, these architectures have already demonstrated their superiority over all usual learning techniques although they still lack mathematical justifications. In particular, no result exists to estimate the number of neurones and layers to use to obtain the best possible result, depending on the kind and the quantity of data to analyze. The aim of this internship is to both study the possibility of an optimization technique that would not require to test all possible configurations, and to explain the choices made to fine-tune the parameters of neural networks.

Ecole Normale Supérieure
Paris

Olivier
DOLLE

14h30-15h30

Analyse du comportement de navigation sur un site web

Browsing behavioral analysis

Pour de nombreuses entreprises dont l'activité est intrinsèquement liée à internet, la compréhension et l'exploitation des données qu'elles collectent apporte une plus-value importante tant du point de vue marketing que du point de vue métier. Il s'agit de comprendre à partir de données brutes de log comment les utilisateurs utilisent véritablement un site et ainsi de reconnaître certains comportements (acheteur, flaneur, droit-au-but, etc). Pour ce faire, on dispose de toutes les données générées par les visites des clients sur le site et ce chaque jour. Il faut dans un premier temps trier cette énorme quantité de données à analyser, et travailler sur ces données ainsi que sur les métadonnées à disposition afin de créer des variables qui permettent de caractériser au mieux un type de navigation. Différents modèles, notamment à base de chaînes de Markov ou de modèles auto-régressifs, doivent ensuite être testés sur leurs capacités de prédiction (quand le client va-t-il partir? va-t-il acheter? si oui, à quel moment?). Ces modèles, de par leurs paramètres, peuvent ainsi s'avérer déterminants pour caractériser une session et donc pouvoir classer les utilisateurs. Enfin, l'objectif à terme est d'intégrer cette analyse comportementale dans un moteur de recommandation, d'autres applications comme la segmentation marketing sont également envisagées.

Many web-based companies struggle to use the data they collect each time a user comes on their websites. These data conceal a lot of information that the companies could use for marketing purposes or even to understand how their activities are evolving. The goal is to understand, from raw log values, the user's experience of the website and thus find patterns in their browsing behavior. To do so we have at our disposal a huge amount of data generated everyday by the client's clicks on the website. We first have to work on the data and on the metadata to create new features that can enable us to characterize browsing behaviors. Markov chain and auto-regressive models are then tested to see if they fit the data and if they can predict accurately a user's behavior (when will a user log off? will he buy? When?). These models parameters can also be important features to characterize a session and therefore enable to classify the users. In the end we wish to integrate this browsing analysis into a recommender system. Another possibility would be using this analysis to solve market-segmentation issues.

Vente-Privée.com
Saint-Denis

15h30-16h30

Optimisation de la distribution de catalogues de promotion par des méthodes statistiques

Optimization of the distribution of promotional catalogues by statistical methods

Afin d'améliorer l'influence de la promotion dans la grande distribution, il est nécessaire d'informer les clients de ces offres par le biais des catalogues. Le catalogue papier est encore le principal média publicitaire utilisé : Carrefour distribue 91 opérations promotionnelles par an à une moyenne de 14 millions d'exemplaires par opération ! Il est donc nécessaire d'optimiser cette distribution en ciblant les zones géographiques où il est pertinent de distribuer tout ou partie des opérations. Les méthodes actuellement utilisées sont relativement simples et basées uniquement sur des critères de proximité au magasin et de parts de marché réalisées dans les zones considérées appelées IRIS (résultant d'un découpage national établi par l'INSEE).

L'objectif de ce stage est d'établir un modèle (géo)statistique utilisant les données dont nous disposons et permettant de redéfinir les zones de distribution au niveau national. Il s'agit dans un premier temps de considérer les données propres à chaque magasin (surface, chiffre d'affaires, appartenance à un centre commercial...) afin d'évaluer le rayonnement du magasin dans sa région. Dans un second temps, il convient d'établir un modèle (géo)statistique permettant d'évaluer l'attractivité que Carrefour a dans chaque zone géographique, en considérant l'ensemble des magasins environnants, leur rayonnement, la durée d'accès... Ce modèle doit être un outil d'aide à la décision pour établir une cartographie des nouvelles zones de distribution.

In order to improve the performances of the special offers in retailing, it is necessary to make these offers known by customers thanks to catalogues. Paper version of catalogues is the most used media: 91 offer campaigns per year are distributed by Carrefour with about 14 million catalogues per offer! Consequently, Carrefour needs to optimize the distribution by targeting geographic areas where it is interesting to distribute catalogues corresponding to all or part of the campaigns. Actual methods are simple and based only on the distance to shops and the market share of Carrefour in the different considered areas called IRIS.

The objective is to develop a (geo)statistical model based on different data types with which it may be possible to redefine distribution areas of catalogues in France. First, we can consider data relative to the shops (like surface area, turnover and location in a shopping centre...) to estimate their influence in their areas. We have then to create a (geo)statistical model in order to estimate Carrefour's attractiveness in each IRIS by considering all the shops near the area, their influence, the travel time to the shops... This model should help to map the distribution areas in France.

CARREFOUR
Massy



Jean-Bernard
EYTARD

16h30-17h30

Classification automatique avec effets spatiaux pour la valorisation de prix de marché de véhicules

Automatic classification of spatial effects for car market price valuation

Autobiz est une jeune entreprise technologique française, créée en 2004, spécialisée dans l'analyse des marchés de l'automobile dans 20 pays. Elaboré en partenariat avec le laboratoire SAMM (Université Paris 1), le système de valorisation de prix de marché a été conçu entre 2007 et 2011 autour d'un référentiel descriptif de véhicules. Ceci a notamment permis de mettre en place une prédiction du prix d'un véhicule d'occasion, basée essentiellement sur sa référence (modèle, cylindrée, carburant,...) et ses caractéristiques propres de kilométrage et âge.

Ce système montre aujourd'hui des limites face aux nouveaux enjeux internationaux de l'entreprise. L'entreprise souhaite en effet évaluer un nombre plus important de références de véhicules, tout en

Autobiz is a French IT company, founded in 2004, that specializes in the analysis of car markets in different countries (20 countries in 2015). Between 2007 and 2011, a car pricing tool was developed in partnership with the SAMM laboratory (University Paris 1). And a large vehicles database was created and used in that purpose. This pricing tool could predict the price of an used vehicle by taking into account only its national characteristics : model, engine size, fuel, mileage, age... But today, this tool can no longer support the company's recent wish to become a major actor on the international market. This means that the company needs to estimate the price of a larger number of vehicles, while facilitating the price comparison between different markets.



Mehdi
GMIRA

facilitant la comparaison de prix entre différents marchés internationaux, dont les référentiels sont distincts. Le but du stage est donc de créer un nouveau système de cotation, qui fasse abstraction des référentiels de véhicules et utilise à leur place un référentiel dynamique de caractéristiques communes.

En premier lieu, il conviendra alors de mettre en place une classification automatique avec prise en compte des effets spatiaux. Le choix des variables influentes et la recherche de données aberrantes découlera de cette première étape.

La mise en place d'un modèle de prédiction du prix à partir de la classification proposée sera alors une seconde tâche. L'utilisation de méthodes paramétriques (modèles linéaires, non-linéaires, perceptrons multicouches,...) sera envisagée tout comme celle de méthodes non-paramétriques (arbres de régression, régression par noyaux,...).

The goal of this internship is to develop a new pricing tool, which no longer uses national characteristics, but instead takes into account more general and international characteristics.

The first step is to build a classification tool that takes into account the spatial distribution of vehicles. The development of this tool will include features selection and outliers. The second step is to build a predictive tool, using the classification tool first developed. The use of both parametric methods (Linear Models, Non-Linear, multi-layer perceptrons, and non-parametric methods (regression trees, stones regression, ...) will be considered.

**AUTOBIZ et Laboratoire
SAMM, Univ. Paris 1**
Paris

Mercredi 1^{er} Juillet Wednesday 1st July

■ L106



Driss
AHMIDOUCH

9h-10h

Optimisation de portefeuille avec des fonds de Private Equity

Portfolio optimization with private equity funds

Les investissements en Private Equity occupent aujourd'hui une proportion importante des portefeuilles des investisseurs institutionnels cherchant des alternatives aux actifs en cotation publique. Ces investisseurs profitent d'une réglementation flexible, d'un capital conséquent et d'un horizon de temps suffisamment long pour détenir ces actifs illiquides.

Cela étant, l'état actuel des connaissances sur la façon d'optimiser quantitativement des parts dans des fonds de private equity dans le cadre d'un portefeuille global est limité. La difficulté vient du fait que les parts dans ces fonds ne peuvent être vendues qu'à des instants précis. De plus, ces parts ne sont pas achetées instantanément. En effet, l'investisseur prend initialement un engagement en capital et le verse par la suite. Le calendrier et les montants versés sont des variables aléatoires. Cela complique la théorie d'optimisation standard de portefeuille où l'investisseur peut rééquilibrer les poids de tous les actifs dynamiquement.

Ainsi, l'objectif est de créer un modèle d'optimisation tenant compte des restrictions précédentes et utilisant deux classes d'actifs : les actions liquides et les fonds de Private Equity ainsi que les obligations sans risque. Les résultats consistent en un engagement en capital et une stratégie de portefeuille. Une comparaison est également faite avec le modèle d'optimisation standard pour examiner les effets de l'illiquidité sur l'allocation d'actifs.

Private Equity investments have become increasingly weighty in a portfolio context for institutional investors seeking to diversify their benefits from traditional stocks and bond holdings. These investors enjoy regulatory freedom, a sufficient amount of capital and a long enough time frame to hold illiquid investments. That being, the current understanding on how to quantitatively optimize private equity funds in an overall portfolio is quite limited. The difficulty comes from the fact that private equity participations can only be sold at certain times. Moreover, stakes in these funds are not bought instantly. Rather, the investor makes an initial capital commitment and disburses capital later on. The timing and the size of the capital are random variables. This complicates the standard portfolio optimization theory where it is assumed that the investor can rebalance the weights of all assets dynamically. The purpose is to create an optimization model taking into account the previous restrictions and using two asset classes: liquid stocks and private equity funds, and liquid risk-free bonds. The results are a combined commitment and a portfolio strategy. We also compares the results with the standard benchmark to analyze the effects of illiquidity on asset allocation.

AMETHIS FINANCE
Paris

10h-11h

Développement d'outils d'estimation des risques de marché

Development of tools for market risk estimation

L'arbitrage en finance de marchés est l'utilisation des incohérences des prix des instruments financiers pour réaliser des profits. Ces opérations assurent l'égalité des prix entre les différents marchés et participent à leur efficacité. Pour cela, il faut développer différentes stratégies d'investissement sur tous les marchés financiers et se couvrir en conséquence contre les risques pris par ces positions. Le défi est donc d'identifier les risques reliés à chaque arbitrage et d'être capable de définir des mesures de ces risques et de créer des outils qui permettent à la fois de suivre en temps réel l'exposition de l'entreprise sur les marchés financiers et de la contrôler. Le deuxième volet de ce travail d'option consiste au montage d'une « Business Unit » sur les options américaines sur des indices « equity », des « futures » sur indices et sur des « ETF ». Cette BU permettra un meilleur « pricing » des options américaines et une meilleure gestion des risques grâce à une meilleure anticipation de la volatilité implicite.

Arbitrage is the practice of taking advantage of a price difference between two or more different markets, the profit being the difference between the market prices. This activity is mandatory to be able to handle the efficiency of markets.

The objective of this study is to develop adequate tools to identify and manage the market risk that could occur once the arbitrage strategies are applied upon financial markets all over the world. In the second part of this work will consist in the development of a Business Unit to price American options on equity indexes, futures on indexes and ETF, using different models (SABR, Black and Scholes,...) to catch the implied volatility.

**ABC Arbitrage
Asset Management**
Paris



Hamza
BAZI

11h-12h

Mise en place de stratégies de trading de volatilité

Implementation of trading strategies on volatility

Les marchés dérivés, principalement les marchés d'options, sont par nature illiquides. Lorsqu'un investisseur veut traiter une option spécifique, il y a très peu de chance qu'un autre investisseur ait l'idée inverse sur le même produit au même moment. Les marchés dérivés requièrent des «teneurs de marché» (market maker) pour assurer cette liquidité. L'objectif de mon stage chez Mosaic Finance, jeune entreprise de market making, est dans un premier temps de gérer les portefeuilles d'actions sur les stocks anglais et d'assurer la liquidité de ces marchés. Dans un second temps, il s'agit de développer des stratégies de trading sur la volatilité, telles que la dispersion. Une stratégie de dispersion consiste par exemple à arbitrer statistiquement un écart de volatilité entre deux valeurs fortement corrélées, ou entre la volatilité d'un indice et d'un sous panier de stocks.

Derivative markets, mainly option markets, are by nature illiquid: when an investor wants to buy a specific option, there is very little chance another investor has the exact opposite idea on the same product at the same time. Derivative markets require "market makers" who can provide counterparty to any investor's request.

The objective of this internship at Mosaic Finance, a young market making firm, is to manage the portfolios on English stocks, and to provide liquidity to those markets. Trading strategies will be developed on volatility, such as dispersion. A dispersion strategy works by statistically arbitrating the difference in volatility between two highly correlated stocks, or between an index and a basket made of stocks which contribute to the index.

MOSAIC FINANCE
Paris



Grégoire
MAIN de BOISSIERE

13h30-14h30

Etude quantitative du risque systémique de marché

Quantitative study of systemic market risk

L'étude du risque d'effondrement généralisé du marché financier sous l'effet de la faillite d'une banque a été étudiée. Après examen de publications scientifiques spécialisées en finance nous avons procédé à l'analyse plus détaillée

This study concerns the risk of widespread collapse of the financial market as a result of the failure of a bank. First specialized scientific publications in finance were analyzed in order to become familiar with the concepts. Then the risk



Kassem
BENABDERRAZIK

des mesures de ce risque selon différentes approches (Marginal Expected Shortfall, Δ Cover... et les modèles mathématiques associés : modèle de GARCH pour la prévision des séries temporelles, modèle de la régression par quantile). Puis ces mesures de risque ont été implémentées en MATLAB et testées sur une base de données des 100 plus grands établissements financiers américains, afin de retrouver les résultats obtenus dans la littérature. Parallèlement, on a implémenté des mesures de performance de fonds d'investissements fondés sur l'analyse de moments d'ordre supérieurs à deux des distributions.

measures proposed by different authors were analyzed in more detail : Marginal Expected Shortfall, Δ Cover ... as well as the mathematical models associated: GARCH model for predicting time series, model of quantile regression. These risk measures were implemented on MATLAB and tested on a database constituted of the 100 largest US financial institutions to find the results obtained by the authors of the articles. At the same time, investment fund performance measures based on the analysis of higher-order moments were applied to two time series.

OBC NEUFLITZE
Paris



Valeriy
GUSMAN

14h30-15h30

Méthodes récentes de calcul de sensibilités pour différentes classes d'actifs

Recent methods of sensibilities calculation for various classes of products

Aujourd'hui la diversité des produits financiers, ainsi que les modèles pour les traiter sont très divers. Au même moment la vitesse du calcul est un des facteurs les plus importants. Dès lors il existe un besoin fort d'algorithmes génériques d'évaluation des instruments exotiques qui soient rapidement adaptables aux nouveaux produits. Une solution était proposée par LexiFi basée sur un langage de scripts qui permet de décrire des contrats complexes de façon simple et générique. En sus il faut calculer les dérivées partielles de ces produits pour pouvoir se prémunir contre des risques différents. Le but du stage est donc d'implémenter un algorithme de calcul des "grecs" (dérivées partielles), qui doit pouvoir marcher quel que soit le modèle choisi (Black-Scholes, Heston, Volatilité locale) et avec une fonction pay-off quelconque (discontinue, path-dependent, cross-asset). La majeure partie du stage consiste en l'analyse des algorithmes existants tels que la méthode de ratio de vraisemblance ou la méthode des trajectoires (pathwise) et ses adaptations à l'architecture et des besoins de l'application "Apropos".

Nowadays the diversification of financial products and corresponding quantitative models is very wide. At the same time the computation speed is one of the most important factors. Therefore there is a need for numerical solutions to treat exotic contracts in a generic and fast way, which should be easily adaptable to new products. An application was proposed by LexiFi based on a special script language which helps to describe a wide range of financial contracts in a general way. Additionally one needs to calculate the partial derivatives in order to hedge different sources of risk. The goal of this work is to implement an algorithm of the greeks (partial derivatives) calculation. It should be independent from the model (Black-Scholes/ Heston/ Local Volatility) and should work with various classes of Pay-off functions (discontinuous, path-dependent, cross-asset). The major part of the internship will be the analysis of existing algorithms like path-wise method or likelihood ratio method, studying their advantages and disadvantages, and finally developing and implementing the algorithm which will be coherent with the goals of rapidity, accuracy and versatility as well as with the structure of the LexiFi application "Apropos".

LexiFi
Boulogne-Billancourt



Wenyao
ZHANG

15h30-16h30

Structuration d'indices : développement de stratégies quantitatives d'investissement dans les actifs croisés

Index Structuring: development of cross-asset quantitative investment strategies

La structuration d'indices consiste à développer des stratégies quantitatives d'investissement dans les actifs croisés (actions, changes, revenu fixe, et commodités), pour des clients institutionnels, des banques privées et des offices de familles.

Index Structuring consists in designing Quantitative Investment Strategies and cross assets (equities, FX, Fixed Income, and Commodities) for Institutional Clients, Private Banks and Family Offices. This work contributes to the development of new

Ce travail contribue au développement de nouvelles stratégies, de présentations de commercialisation et de modélisation, etc. Etant intégré dans les nouveaux projets ainsi que dans les pipelines de trading existants, il a permis de surtout se focaliser sur des stratégies d'investissement systématique pour les primes de risques cachés, comme la dividende et la volatilité dans le marché d'actions. Les catégories de stratégies de Momentum, Carry, et Value / Mean Reversion sont considérées, qui exigent l'utilisation de diverses techniques quantitatives et statistiques.

strategies, marketing presentations, modeling, etc. being integrated in new projects as well as in the existing trade pipeline. Main focus is on developing systematic investment strategies on hidden asset risk premiums, such as dividends and volatility in equity markets. Strategy categories of Momentum, Carry, and Value / Mean Reversion are considered, which require the use of various quantitative and statistical techniques.

NOMURA

London (Great Britain)

16h30-17h30

Analyse de données clients issues des réseaux sociaux

Data analysis for client information gathered from social networks

Les métiers de l'assurance reposent sur le transfert d'un risque entre l'assuré et son assureur moyennant une compensation financière, la prime. Cette prime est calculée en tenant compte notamment de la probabilité d'occurrence de l'événement : plus elle est élevée, plus la prime est importante. Relativement à un sinistre donné, cette probabilité d'occurrence varie d'un individu à l'autre : il est plus risqué de fournir une assurance vie à un individu de 90 ans qu'à un individu de 20 ans. Les assureurs cherchent donc à connaître au mieux leurs clients afin d'évaluer de la manière la plus exacte possible le risque qu'ils représentent. Les réseaux sociaux leur fournissent pour cela une richesse de données personnelles jusqu'alors inimaginable. Ces données sont déjà utilisées pour lutter contre la fraude à l'assurance en relevant les incohérences entre ce qui a été déclaré et ce qui est dit sur Facebook ou Twitter. Leur potentiel est cependant loin d'être exploité : elles permettraient non seulement un raffinement des modèles de calcul de primes, mais également un marketing plus personnalisé grâce à des segmentations de clientèle plus fines et pertinentes.

Encore faut-il les analyser. Il est en particulier nécessaire d'étudier les corrélations statistiques entre tel comportement - un « j'aime » sur Facebook par exemple - et l'occurrence de tel sinistre. On commencera a priori par le secteur de l'assurance automobile sur lequel on a les données les plus nombreuses et les plus variées. Le but final est de développer un ou deux indicateurs statistiques nouveaux à l'aide des données des réseaux sociaux identifiées comme pertinentes et des données structurées internes à l'entreprise. Ces indicateurs permettront de mieux quantifier le risque représenté par le client.

The insurance business hinges on an individual - the policy-holder - hedging against a certain risk by transferring it to another party - namely the insurer. In order to do so he gives the insurer a financial compensation called a premium. Calculating this premium factors in, in particular, the probability of the event occurring : the higher the probability, the higher the premium. Relative to a given damage, this probability varies from one individual to the other : it is riskier to provide a 90-year-old with life insurance than it is to do so with a 20-year-old. Insurers thus strive to know their clients as best as they can in order to assess as accurately as is possible the risk they pose. Social networks provide them with a hereto unbelievable wealth of personal information to that end. Such data is already used to fight against insurance fraud by spotting any discrepancies there might be between what has been declared and what has been posted on Facebook or Twitter. Yet, the applications go way beyond what is being done at present: such data would not only allow for more refined premium-calculation models but they could also give rise to a more personalised marketing approach as the customer-base is micro-segmented in more relevant categories.

However, the data has to be analysed. Special heed must be paid to studying the statistical correlations that arise between a given type of behaviour - a Facebook "like" for instance - and the occurring of a given damage. The place to start with seems to be the car insurance sector because the information there is extensive and varied. The final aim is to develop one or two new statistical indicators encompassing the data from social network that has proven of interest and the pre-existing structured data internal to the company. These indicators should allow for a better assessment of the risk a client poses.

GROUPAMA Assicurazioni

Rome (Italie)



Céline
JACKSON