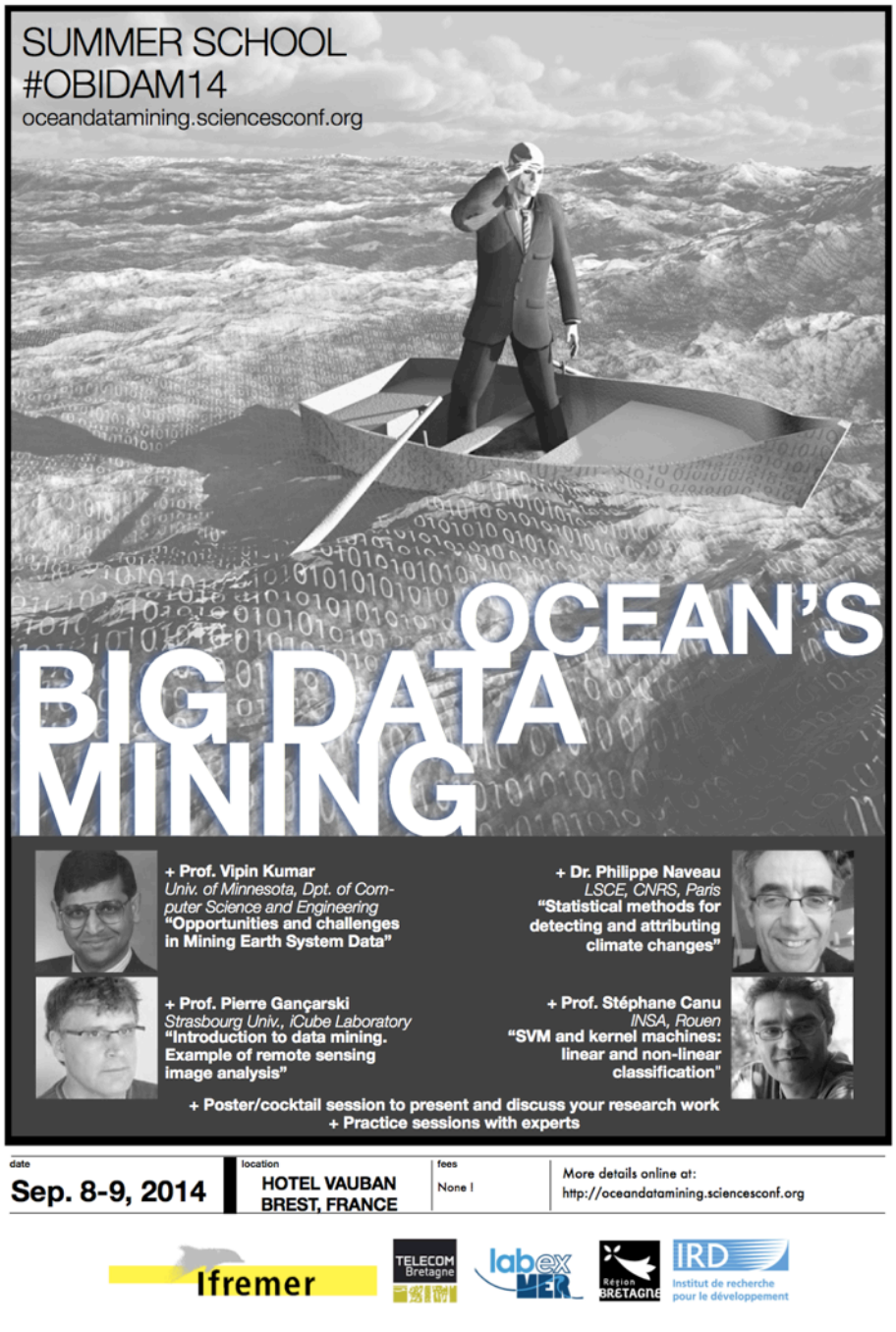
"FOUILLE DE GRANDES BASES DE DONNEES OCEANIQUES: NOUVEAUX DEFIS ET SOLUTIONS"

Compte-rendu factuel de l'école d'été OBIDAM14 organisée par l'Ifremer, le CNRS et Telecom Bretagne.



Cette école d'été s'est déroulée les lundi 8 et mardi 9 septembre 2014, à Brest, à l'hôtel Vauban.

L'école avait pour thème les nouveaux défis posés par l'analyse des données marines. En effet, les bases de données marines, alimentées par les satellites et les robots autonomes sous-marins comme les flotteurs du réseau Argo, sont de plus en plus grandes (plusieurs dizaines de gigaoctets et teraoctets) et changent d'heure en heure. Cette augmentation spectaculaire de la dimension à laquelle s'ajoute une complexité grandissante des données rend difficile leur exploitation avec les outils standards. Or c'est à partir de l'analyse des données que les chercheurs pourront réaliser de nouvelles découvertes scientifiques sur la dynamique des océans à grande et petite échelles, et les changements climatiques régionaux et globaux.

Heureusement, il existe des solutions à ces nouveaux problèmes. En formant les scientifiques de la communauté de recherche en océanographie physique à ces solutions, cette école d'été devrait contribuer à lever les verrous d'analyse et permettre de défier ce nouvel océan de données.

Le premier objectif de l'école était de se placer dans la continuité de la conférence "TIC et mer: nouveaux défis et solutions" organisé le 26 novembre 2013 au centre Ifremer de Brest (90 participants, plus de détails ici: http://wwz.ifremer.fr/bigdata) en se focalisant sur un thème spécifique. La conférence de novembre avait abordé trois thèmes structurant les nouveaux défis posés par l'analyse des données marines: l'inter-opérabilité, le stockage et la fouille. L'école d'été a permis 2 jours de cours et exercices sur le thème de la fouille des bases de données marines.

Le second objectif de l'école était de rassembler des acteurs des communautés française et internationale, du monde de la recherche océanographique et informatique autour du thème de la fouille de données marines.

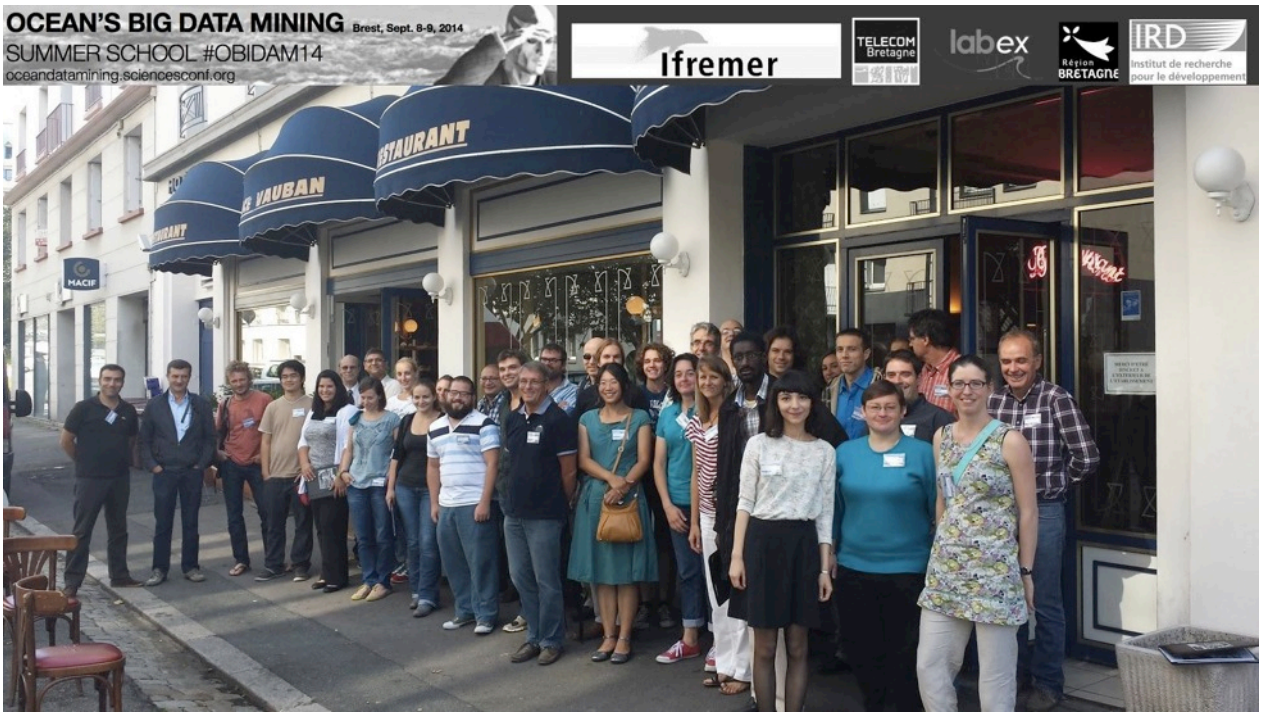
Parmi les 41 participants, 32 étaient affiliés à un laboratoire français et 9 à un laboratoire étranger (Angleterre, Pologne, Hollande, Danemark, Sénégal, Etats-Unis). Ce rapport ¾ nationaux, ¼ internationaux paraît équilibré étant donné le format relativement court de l'école (2 jours) qui pouvait décourager l'investissement dans un long voyage. Parmi les 32 chercheurs affiliés en France, 23 venaient de la communauté locale Brestoise et 9 d'autres laboratoires en France. Quantitativement, ce petit tiers de participation de la communauté française en dehors de Brest est compensé qualitativement par le fait que les principaux laboratoires d'océanographie physique nationaux avaient envoyé au moins un participant[1]. Notons enfin que des 23 "locaux", 14 avaient participé à la conférence introductive de novembre 2013, ce qui confirme l'intérêt de la communauté locale pour la thématique de l'école. Nous pourrons également noter le très bon équilibre entre permanents (22) et étudiants (19 dont 11 postdocs et 8 doctorants). Ces derniers ont par ailleurs constitué l'essentiel des présentations de posters (10 posters: 2 permanents, 4 postdocs, 4 doctorants). Enfin à la question "Quelle est votre communauté de recherche principale ?" les 41 participants ont répondu de la manière suivante:

[1] Notons la présence du LSCE, LOCEAN, LOV, LEGOS et CLS. L'absence du LEGI peut s'expliquer par l'orientation vers la modélisation du laboratoire, plus que vers l'observation.

Océanographie physique	16
Fouille de données	15
Sciences de l'environnement	8
Non identifiés	2

où l'on peut voir que les ¾ des participants étaient issus à parts égales des deux communautés principalement visées et que le ¼ des participants appartenaient à des communautés scientifiques voisines.

Au regard de la distribution des origines scientifiques et géographiques des participants, le comité scientifique[2] a le sentiment d'avoir atteint son objectif de rassembler des chercheurs locaux et étrangers à Brest issus des communautés océanographie physique et fouille de données.



[2] Le comité scientifque était constitué par ordre alphabétique de: R. Fablet, P. Lenca, G. Maze, H. Mercier et J.-F. Piollé à qui s'ajoutait F. Cudennec et T. Le Toullec pour l'organisation.

Retour sur les interventions et le déroulement de l'école d'été

Les premiers participants sont arrivés dès le dimanche 7 à l'hôtel Vauban. Lundi 8 vers 8h00 les membres du comité ont préparé la salle de restaurant de l'hôtel pour accueillir les 3 cours prévus pour cette première journée. Entre 9h00 et 10h00, les participants étaient invités à venir s'enregistrer pour retirer le conférencier marqué du logo de l'école et contenant le programme détaillé des interventions et une clé USB avec le matériel numérique nécessaire à la session de travaux pratiques prévu pour le mardi. Tous les participants inscrits se sont bien présentés.

Après un mot de bienvenue par Guillaume Maze qui a remercié les sponsors de l'école et les invités, le premier intervenant **Vipin Kumar** a donné le premier cours intitulé: "**Opportunities and challenges in mining Earth system data**", o*pportunités et défis dans l'analyse des données du système Terre*. Vipin Kumar est actuellement William Norris Professor et directeur du département Ingénierie et sciences informatiques de l'université du Minesota aux Etats-Unis. Ces recherches actuelles portent sur les méthodes de fouille de données, le calcul haute performance et leurs applications dans les domaines du climat, des écosystèmes et du biomédical. Il dirige un projet NSF de 5 ans (10M$) intitulé "Comprendre le changement climatique - une approche centrée sur les données" dont l'objectif est de repousser les limites de la recherche informatique en sciences de l'environnement. V. Kumar a reçu nombres de distinctions, écrit plus de 300 articles de recherche et édité ou écrit 11 livres.

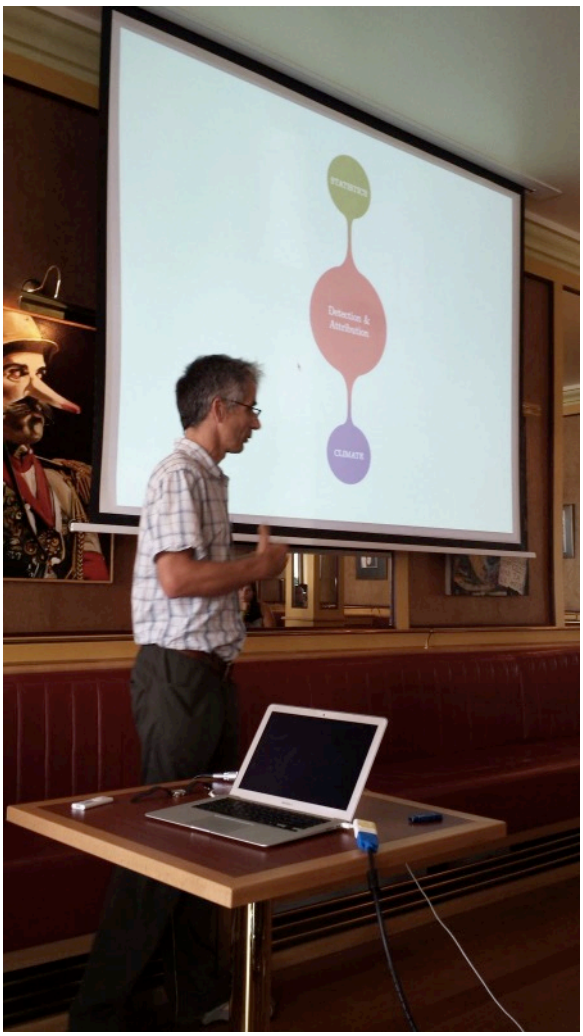
La présentation de V. Kumar est disponible en ligne sur le site web de l'école ici:
http://oceandatamining.sciencesconf.org/conference/oceandatamining/program/OBIDAM14_Kumar.pdf

V. Kumar a commencé son intervention par une brève introduction aux méthodes de fouille de données (classification, clustering, etc .. Puis il a parlé plus spécifiquement des opportunités offertes par ces méthodes pour les sciences de l'environnement. En effet, comme les organisateurs de l'école d'été l'ont déjà souligné, Vipin nous rappelle que les bases de données en sciences de l'environnement explosent en complexité et taille. Cette évolution est à rapprocher de l'explosion des bases de données liées à l'usage social d'internet et auquel se confrontent les géants du web comme Google, Twitter ou Facebook pour générer des revenus. Pour générer de la connaissance et de nouvelles découvertes, notre communauté doit elle aussi se confronter au "Big Data". V. Kumar nous détaillera ensuite 3 exemples d'application des méthodes de fouilles aux données environnementales: le suivi des changements des

écosystèmes, l'identification automatique de téléconnections atmosphériques et la caractérisation des tourbillons océaniques. Les deux premiers exemples reposent sur des méthodes peu employées en océanographie physique, en particulier la théorie des "graphes". Cette dernière semble très intéressante pour découvrir de nouvelles connections entre les régions océaniques et mieux comprendre celles déjà identifiées. Le troisième exemple est une démonstration des principes de reconnaissance de forme objective pour la caractérisation des tourbillons océaniques, en particulier leurs interactions (fusion, séparation) qui est une caractéristique inaccessible avec les méthodes employées jusqu'à présent.

Après un déjeuner pris dans la même salle au Vauban et la photo de groupe, les cours ont repris comme prévu à 14H00. Herlé Mercier a fait la présentation et modéré l'intervention de **Philippe Naveau** qui s'intitulait "**Statistical methods for detecting and attributing climate changes**", *méthodes statistiques pour la détection et l'attribution des changements climatiques*.

La présentation de P. Naveau est disponible en ligne sur le site web de l'école ici:
http://oceandatamining.sciencesconf.org/conference/oceandatamining/program/OBIDAM14_Naveau.pdf



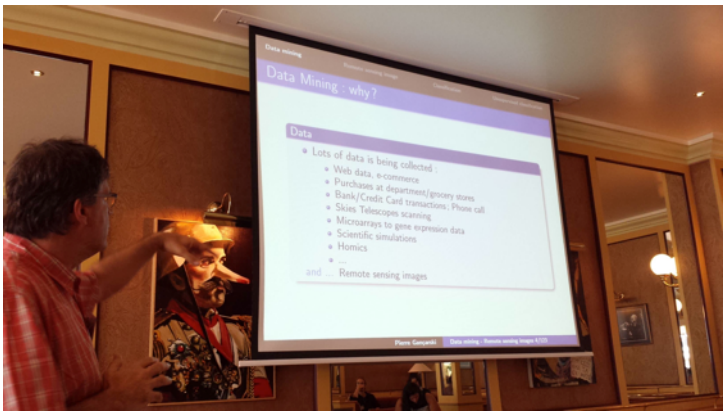
P. Naveau est chercheur CNRS au Laboratoire des Sciences du Climat et de l'Environnement (LSCE, UMR 8212) à Paris. Philippe a obtenu son doctorat en statistiques de l'université du Colorado en 1998. Après 3 ans au NCAR (Boulder, Colorado) et 3 ans de professorat au département de mathématiques appliquées de l'université du Colorado, il a rejoint le CNRS et le LSCE en 2004 où ses recherches portent sur les statistiques en sciences de l'environnement et plus particulièrement l'analyse des évènements extrêmes dont il a choisi de nous parler ce lundi.

Après une présentation des caractéristiques de la variabilité naturelle du climat, P. Naveau a définit les notions de "**détection**" et "**attribution**" selon l'IPCC. Il nous a ensuite décrit en détails deux approches statistiques classiques pour la détection/ attribution: la régression linéaire (problème de détection) et l'estimation de la fraction de risque attribuable ou FAR (problème d'attribution). Nous retiendrons d'une part que les difficultés de détection sont en grande partie liées à notre manque de connaissance de la variabilité du système climatique terrestre à toutes ses échelles de temps et d'espace. Donc détecter les changements climatiques repose sur une

meilleure observation et compréhension de la variabilité naturelle du système. Nous retiendrons également que pour attribuer des changements à des causes, il est nécessaire de parfaitement bien poser le vocabulaire et de distinguer les causalités nécessaires de celles suffisantes. Attribuer les changements climatiques repose sur notre capacité à décomposer et casser les chaînes de causalité et donc en grande partie à pouvoir générer des simulations numériques climatiques réalistes pour passer d'une approche probabiliste observationnelle à interventionniste.

Comme prévu, l'intervention de P. Naveau s'est terminée à 15h30. Après une pause café, Ronan Fablet a présenté **Pierre Gançarski** et modéré son intervention intitulée: "**Introduction to data mining. Example of remote sensing image analysis**", *introduction à la fouille de données, exemple de l'analyse d'images en télédétection*. Pierre Gançarski est professeur au département d'informatique de l'université de Strasbourg. Il conduit ses recherches au sein du laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (iCube, UMR 7357) sur la fouille de données par apprentissage multi-classes et non-supervisé.

La présentation de P. Gançarski est disponible en ligne sur le site web de l'école ici:
http://oceandatamining.sciencesconf.org/conference/oceandatamining/program/OBIDAM14_Gancarski.pdf



La première partie du cours est une introduction à la discipline de la "fouille" des données. Elle consiste à extraire des informations intéressantes (non-triviales, implicites, inconnues jusqu'alors, utiles) d'une collection de données. Pour cela, il y a deux types de méthodes: les descriptives et les prédictives. Les **méthodes prédictives** utilisent un ensemble de variables pour prédire des valeurs de ces variables dans le futur ou des valeurs de variables inconnues. Par exemple, les algorithmes de **classification supervisée** (attribuer un attribut de classe à une donnée) sont des méthodes prédictives qui cherchent à modéliser la relation entre des valeurs d'attributs et celle de la classe à prédire. Déterminer si une transaction bancaire est légitime ou frauduleuse est un exemple de classification, tout comme la classification des types de sols (étendue d'eau, espace urbain, forêt, etc …) à partir des données satellites. Les **méthodes descriptives** quant à elles, visent à trouver dans les données des structures ("patterns") interprétables par l'homme.

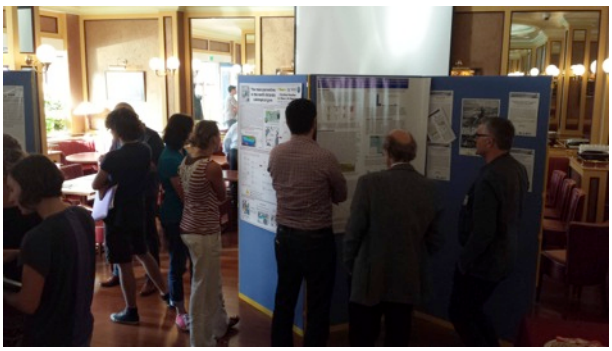
Après cette introduction, P. Gançarski nous a décrit les caractéristiques des images obtenues par télédétection (aériennes et satellites) et pour lesquelles il illustrera l'usage de méthodes de classification supervisée et non-supervisé. Nous retiendrons que la **classification supervisée** est une méthode de fouille prédictive qui consiste à construire un modèle qui à chaque donnée associe une catégorie - ou une classe, ou un label - à partir d'une base de donnée d'apprentissage qui contient déjà des associations entre données et classes. Dans cette catégorie nous trouverons la méthode des plus proches voisins, les arbres de décisions et

les méthodes basées sur les hyperplans (voir présentation de S. Canu). **La classification non-supervisé**, ou clustering, est une méthode descriptive qui permet d'identifier des groupes d'objets de telle manière que les objets d'un groupe seront similaires (ou liés) tout en étant différents (ou non-liés) aux objets des autres groupes. Pour cela on utilise des techniques de partitionnement, comme les k-moyennes, ou le regroupement hiérarchique.

Après l'intervention de P. Gançarski, nous avons installé les 10 posters présentés par les participants, pendant que le cocktail/buffet se mettait en place. A partir de 18h00, nous avons pu discuter et échanger sur les nouveaux travaux et les problèmes que rencontre les participants. La liste des posters est donnée ci-desous. La première journée s'est achevée après la session poster, vers 20h00.

Liste des posters présentés à OBIDAM14:

1. Chapman Chris (LOCEAN, Paris): *The Detection of Jets in the High Resolution Simulations of the Southern Ocean Using Wavelets and Higher Order Statistics.*
2. Epure Elena (CRI, Paris): Intention Mining: *Discovering Intentional Processes from Ocean's Big Data.*
3. Faghmous James (UM, Etats-Unis): *Mesoscale Ocean Eddies from Satellite Altimetry: Methods, Data, and Applications.*
4. Feucher Charlène (LPO, Brest), G. Maze, H. Mercier: *The main pycnocline in the North Atlantic subtropical gyre from Argo data.*
5. Gueye Abdou Karim (UASZ, Sénégal), S. Janicot, P. Braconnot, A. Lezine, C. Hély-Alleaume: *Analysis of climate simulations over Africa at 6Ky, 4Ky and pre-industrial periods by using self organizing Maps. Example of northern summer*.
6. Karagali Ioanna (DTU Wind Energy, Denmark): *SST Diurnal Variability from 6 years of Geostationary SST retrievals.*
7. Maes Christophe (LPO, Brest): *Replacing in situ observation from sea cruises in the oceanic dynamics.*
8. Polanco-Martinez Josué (BC3, Bilbao): *The R package "ReadObsHistWeaDat" (Reading Observational Historical Weather Data").*
9. Sauzède Raphaëlle (LOV, Villefranche), H. Claustre, C. Jamet, H. Lavigne, J. Uitz: *Calibration of in situ fluorescence profiles using a neural network: a first step in the development of a 3D global climatology of phytoplancton communities.*
10. Tandeo Pierre (Lab-STICC, Brest), P. Ailliot, R. Fablet, J. Ruiz, F. Rousseau, B. Chapron: *The Analog Ensemble Kalman Filter and Smoother.*



Mardi matin, la deuxième et dernière journée de l'école a commencé à 9h00. Philippe Lenca a présenté l'intervenant, Stéphane Canu, et modéré le cours intitulé "SVM and kernel machine: linear and non-linear classification", *machines à vecteurs supports et noyaux: classification linéaire et non-linéaire*.

La présentation de S. Canu est disponible en ligne sur le site web de l'école ici:
http://oceandatamining.sciencesconf.org/conference/oceandatamining/program/OBIDAM14_Canu.pdf

Stéphane Canu est ancien directeur et professeur du LITIS (Laboratoire d'informatique, du traitement de l'information, et des systèmes) et du département d'information technologique de l'Institut National des Sciences Appliquées (INSA) de Rouen. Il a obtenu son doctorat en commande des systèmes en 1986 de l'université de Compiègne et son HDR en 1997 de l'université Paris VI. Après avoir fondé et dirigé jusqu'en 2002 le département d'information technologique de l'INSA, S. Canu a passé un an dans le groupe d'intelligence artificiel du ANU/NICTA à Camberra. S. Canu est un expert des algorithmes d'apprentissage automatique, méthodes de régression et machines à cœur dont il a choisi de nous présenter les principes mathématiques pour cette partie d'OBIDAM14 qui se voulait une mise en pratique des méthodes d'apprentissage.

Parmi les méthodes de **classification supervisée** on trouve les **machines à vecteurs de support** (SVM). S. Canu nous les a d'abord présenté mathématiquement puis nous a permis de les tester pendant une séance de travaux dirigés qui a duré toute la journée. Nous avons abordé les SVMs pour la classification linéaire le matin, et non-linéaire l'après-midi. Les machines à vecteurs supports sont aussi appelées les **séparateurs à vaste marge**. Elles sont de nature prédictive et s'utilisent dans les situations où l'on cherche à modéliser la séparation entre deux classes dans un jeu de données labellisées. Si la séparatrice peut être une droite, la

classification est linéaire. Nous retiendrons que les SVMs reposent sur deux principes qui permettront de traiter les problèmes de classification linéaire ou non et de reformuler le problème comme une optimisation quadratique. Le premier principe est celui de **marge maximale.** Dans le cas linéaire, il existe une multitude de droites pouvant séparer deux classes, il faut donc un critère pour en sélectionner une. La ligne de décision optimale, ou frontière de séparation, est celle qui maximise la marge, la distance entre les données et la séparatrice. Seules les données les plus proches de la séparatrice jouent un rôle; on les appelle **vecteurs supports** (d'où le nom de la méthode). Les autres données n'interviennent pas dans la détermination, ce qui a son importance pour le coût numérique de calcul avec de grandes bases de données. Par ailleurs, il existe bien sûr des distributions de données pour lesquelles la séparatrice ne peut pas être une droite (on dit non-linéairement séparable). Pour traiter ces problèmes plus complexes, les SVMs s'appuient sur un second principe qui consiste à **changer d'espace de représentation des données pour un espace, peut-être de dimension supérieure, dans lequel le problème sera linéairement séparable**. Cela se fait par l'intermédiaire des fonctions mathématiques dites noyaux, ou *kernel* en anglais.

L'école s'est terminée vers 17h00 avec la fin des travaux dirigés et le départ des participants.

Conclusion

OBIDAM14 a permis de rassembler des océanographes physiciens et des statisticiens spécialistes des méthodes de fouille des données. Sur la forme, nous retiendrons l'intérêt des deux communautés à se rencontrer et interagir ainsi que l'attrait assez fort de la communauté européenne qui aura fait le déplacement jusqu'à Brest pour seulement 2 jours d'école. Sur le fond, la communauté des océanographes semble particulièrement en demande de nouvelles méthodes pour comprendre et analyser ses bases de données de plus en plus complexes. Les statisticiens semblent quant à eux très intéressés de découvrir un nouveau domaine d'application à leurs méthodes. Plus de collaborations sont donc envisageables. Nous retiendrons que les intervenants ont donné la preuve aux participants de la plus value apportée par ces méthodes de fouille, par rapport aux analyses statistiques plus classiques de la communauté, pour faire de nouvelles découvertes sur la dynamique et la structure des océans. Mais les synergies entre les outils standards pour la détection/attribution et les outils de fouille pour la prédiction/description restent encore à formaliser et être mieux appréhendés par les deux communautés scientifiques concernées. Le comité scientifique espère ainsi qu'OBIDAM14 ne sera qu'une étape dans ce processus.

Annexe: Liste des participants

Ailliot	Pierre	pierre.ailliot@univ-brest.fr
Canu	Stéphane	stephane.canu@insa-rouen.fr
Chapman	Chris	chris.chapman.28@gmail.com
Charria	Guillaume	guillaume.charria@ifremer.fr
Choury	Anna	anna.choury@gmail.com
Cisek	Malgorzata	gosiak@iopan.gda.pl
Epure	Elena	elenavepure@gmail.com
Fablet	Ronan	ronan.fablet@telecom-bretagne.eu
Faghmous	James	jfaghm@gmail.com
Feucher	Charlène	cfeucher@ifremer.fr
Gaillard	Fabienne	fabienne.gaillard@ifremer.fr
Gançarski	Pierre	gancarski@unistra.fr
Goszczko	Ilona	ilona_g@iopan.gda.pl
Gueye	Abdou Karim	akgueye@univ-zig.sn
He Guelton	Liyun	liyun.he@gmail.com
Karagali	Ioanna	ioka@dtu.dk
Kumar	Vipin	kumar@cs.umn.edu
Le Couls	Sarah	sarah.lecouls@cfto.fr
Le Goff	Clément	clement.legoff@telecom-bretagne.eu
Le Toullec	Tristan	tristan.letoullec@univ-brest.fr
Lenca	Philippe	Philippe.Lenca@telecom-bretagne.eu
Livina	Valerie	valerie.livina@npl.co.uk
Loubrieu	Thomas	Thomas.Loubrieu@ifremer.fr
Maes	Christophe	christophe.maes@ird.fr
Maudire	Gilbert	gilbert.maudire@ifremer.fr
Maze	Guillaume	gmaze@ifremer.fr
Mercier	Herlé	Herle.Mercier@ifremer.fr
Naveau	Philippe	naveau@lsce.ipsl.fr
Odaka	Tina	Tina.Odaka@ifremer.fr
Piollé	Jean-François	Jean.Francois.Piolle@ifremer.fr
Piron	Anne	anne.piron@ifremer.fr
Polanco-Martinez	Josué	josue.m.polanco@gmail.com
Puentes	John	John.Puentes@telecom-bretagne.eu
Lopez Radcenco	Manuel	manuel.lopezradcenco@telecom-bretagne.eu
Rusciano	Emanuela	rusciano@univ-brest.fr
Sauzède	Raphaëlle	raphaelle.sauzede@obs-vlfr.fr
Simon	Guillaume	gfsimon@gmail.com
Soulas	Julie	julie.soulas@telecom-bretagne.eu
Tandeo	Pierre	pierre.tandeo@telecom-bretagne.eu
Thao	Soulivanh	sthao@cls.fr
Timko	Patrick	p.timko@bangor.ac.uk

Annexe: Communication

L'annonce de l'école s'est faite par les canaux spécifiques et habituels des deux communautés de recherche concernées. Pour sensibiliser un publique plus large à la thématique de l'école, nous avons fait appel aux services de communications du centre Ifremer de Brest et de Telecom Bretagne qui ont relayés l'annonce sur les canaux appropriés. Nous retiendrons 3 éléments marquants: la parution d'un article dédié à l'école dans le journal "le Telegrame" de Brest (13/09/14), la publication du communiqué de presse (08/09/14), et l'usage des réseaux sociaux, en particulier Twitter où les messages concernant l'école pouvaient être suivis avec le hashtag #obidam14 (cliquer pour accéder à la page web).

Données marines.
Nouveaux défis à relever

Les lundi 8 et mardi 9 septembre, l'Ifremer, le CNRS, via ses laboratoires brestois et Telecom Bretagne ont organisé à Brest une école d'été internationale sur le thème des nouveaux défis qui sont posés par l'analyse des données marines. Une quarantaine de chercheurs et d'ingénieurs en informatique, en statistiques et en sciences de l'environnement, venant de France, d'Angleterre, de Pologne, de Hollande, du Danemark, du Sénégal et des États-Unis y ont participé.

Les bases de données marines, alimentées par les satellites et les robots autonomes sous-marins comme les flotteurs du réseau Argo, sont de plus en plus grandes (plusieurs dizaines de gigaoctets et teraoctets) et rapidement évolutives (elles changent d'heure en heure). Cette augmentation spectaculaire de la dimension et de la complexité des données rend difficile leur exploitation avec les outils standards. Or, c'est à partir de l'analyse des données que les chercheurs pourront réaliser de nouvelles découvertes scientifiques sur la dynamique des océans, à grande et petite échelles, et les changements climatiques régionaux et globaux.

Il existe cependant des solutions à ces nouveaux problèmes. En formant les scientifiques de la communauté de recherche en océanographie physique à ces solutions, cette école d'été visait à contribuer à lever les verrous d'analyse et à permettre de défier ce nouvel océan de données.

Davantage de renseignements sur le site Internet, http://oceandatamining.sciences-conf.org/et les réseaux sociaux avec le hashtag : #obidam14

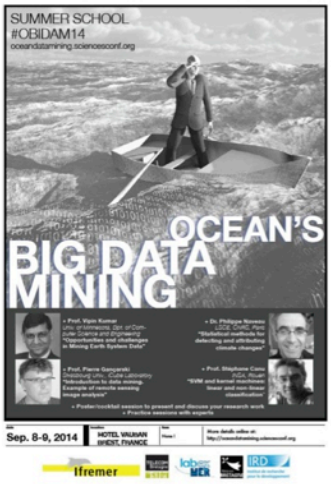
Figure 1 Article paru dans le journal "le Télégramme" du 13 septembre 2014.

Communiqué de presse

Plouzané, le 8 septembre 2014

Défier un océan de données

Ecole d'été « Ocean's Big Data mining »
Lundi 8 et mardi 9 septembre 2014
Hôtel Vauban à Brest



Les 8 et 9 septembre à Brest, l'Ifremer, le CNRS et Telecom Bretagne organisent une école d'été internationale sur le thème des nouveaux défis et solutions concernant l'analyse des données marines. Une quarantaine de chercheurs et ingénieurs en informatique, statistiques et sciences de l'environnement, y participeront.

Gérer des bases de données de plus en plus complexes...
Les bases de données marines sont de plus en plus grandes (plusieurs dizaines de gigaoctets et teraoctets[1]) et rapidement évolutives (les données sont collectées d'heure en heure). Ces évolutions sont dues à la multiplication des plateformes autonomes de mesures *in-situ* (par exemple les flotteurs du réseau Argo[2]) et à l'amélioration des capacités de mesure des satellites de l'océan à haute résolution comme le futur satellite SWOT[3].

... pour répondre aux défis scientifiques !
Cette augmentation spectaculaire de la dimension et de la complexité des bases de données rend difficile leur exploitation avec les méthodes et outils standards. Il existe pourtant des solutions avec lesquelles la communauté de recherche en océanographie physique n'est pas familière. En formant les scientifiques à ces solutions, cette école d'été devrait contribuer à lever les verrous d'analyse. L'objectif, à terme, est de répondre à de nombreux défis scientifiques, comme par exemple, l'acquisition de la connaissance sur la dynamique des océans à grande et petite échelles, ou le développement des indicateurs globaux et régionaux des changements climatiques.

En savoir plus : http://oceandatamining.sciencesconf.org/

[1] Un teraoctet (To) équivaut à 1000 milliards d'octets ou 1000 gigaoctet (Go)
[2] 3000 flotteurs profilants (petits robots autonomes) mesurent la température et la salinité depuis la surface jusqu'à 2000 mètres de profondeur sur l'ensemble des océans. L'un des deux centres mondiaux qui recueille et distribue leurs données, est le centre Coriolis au Centre Ifremer Bretagne.
[3] dont le CERSAT au Centre Ifremer Bretagne centralisera les données.

Communication Ifremer Bretagne : Johanna Martin – 02 98 22 40 05 - johanna.martin@ifremer.fr
Contacts presse Ifremer Paris : Thomas Isaak/ Marion Le Foll – 01 46 48 22 40/42 – presse@ifremer.fr

Figure 2 Communiqué de presse publié par l'Ifremer le lundi 8 septembre 2014