

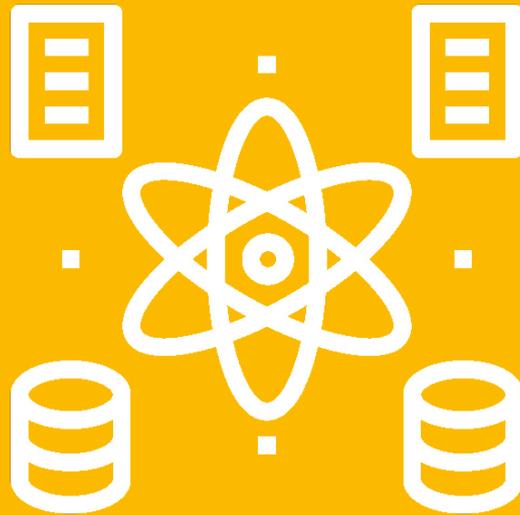
# Master 2

Sciences, Technologies, Santé

2024-2025

*Mathématiques et applications*

## Data Science



M2 DS

*Apprentissage*

*Contrat Pro.*



CONNAISSANCES  
université  
angers

---

# SOMMAIRE

# CONTENUS

**03**

Contacts de la formation

**04**

Calendriers de la formation

**06**

Présentation de la formation

**09**

Volumes horaires et évaluations

**11**

Contenu des enseignements  
Semestre 3

**20**

Contenu des enseignements  
Semestre 4

Sommaire interactif  
pour revenir au sommaire  
cliquer sur 



---

# CONTACTS

**Sandrine TRAVIER** : *Directrice adjointe à la Pédagogie*  
[sandrine.travier@univ-angers.fr](mailto:sandrine.travier@univ-angers.fr)

**Gilles STUPFLER** : *M1 Responsable pédagogique et Président du Jury*  
Tél. : 02 41 73 50 27  
[gilles.stupfler@univ-angers.fr](mailto:gilles.stupfler@univ-angers.fr)  
Bureau : bâtiment I, porte 115

**Fabien PANLOUP** : *M2 Responsable pédagogique et Président du Jury*  
Tél. : 02 41 73 50 29  
[fabien.panloup@univ-angers.fr](mailto:fabien.panloup@univ-angers.fr)  
Bureau : bâtiment I, porte 221

**Sandrine HERGUAIS** : *Gestion de la scolarité et des examens*  
Tél. : 02 41 73 54 85  
[sandrine.herguais@univ-angers.fr](mailto:sandrine.herguais@univ-angers.fr)

**Charlotte BROSSET** : *Alternance*  
Tél. : 02 41 73 52 17  
[re.sciences@contact.univ-angers.fr](mailto:re.sciences@contact.univ-angers.fr)

## Scolarité - Examens

Bâtiment A, Rez-de-chaussée  
Horaires d'ouverture  
8h30 – 12h30  
13h30 – 17h00  
Du lundi au vendredi



## Semestre 1

Rentrée des Alternants	Lundi 02 septembre 2024
Rentrée globale et début des cours	Lundi 09 septembre 2024
Vacances d'automne 	Du samedi 26 octobre 2024 au dimanche 03 novembre 2024
Vacances de Fin d'année 	Du samedi 21 décembre 2024 au dimanche 05 janvier 2025
Fin des cours du 1 <sup>er</sup> semestre	Vendredi 10 janvier 2025
Examens 1 <sup>er</sup> semestre   Session 1	Du lundi 13 janvier 2025 au vendredi 17 janvier 2025
Jury 1 <sup>er</sup> semestre   Session 1	Mardi 4 février 2025

## Semestre 2

Début des cours	Lundi 20 janvier 2025
Fin des cours du 2 <sup>ème</sup> semestre	Vendredi 28 février 2025
Examens 2 <sup>ème</sup> semestre   Session 1	Lundi et Mardi 10/11 mars 2025
Examens 1 <sup>er</sup> semestre   Session 2	Mercredi 12 au jeudi 13 mars 2025
Soutenance de Projet Annuel	Non alterants 13/14 mars 2025 Alternants 22/23 mai 2525
Stage non alternants	A partir du Lundi 17 mars 2025
Soutenance de Stage / Alternance	28/29 août, 1/2 septembre 2025
Jury 1 <sup>er</sup> semestre   Session 2 Jury 2 <sup>ème</sup> semestre   Session 1	Lundi 29 septembre 2025

Planning susceptible de modifications





# PRÉSENTATION DE LA FORMATION

Les mathématiques sont depuis longtemps partie prenante du monde socio-éco-nomique. « L'Étude de l'impact socio-économique des mathématiques en France » (Mai 2015) montre une contribution des mathématiques à hauteur de 15 % du PIB et 9 % des emplois en France. L'intégration de compétences mathématiques est désormais un choix stratégique du développement d'une entreprise, en France et dans le monde, dans tous les secteurs d'activité et particulièrement ceux à forte croissance.

**La data science** désigne une discipline à l'interface entre modélisation mathématique, statistique et informatique, née de la nécessité croissante de traiter et d'exploiter les données du big data. La data science est désormais l'outil essentiel d'aide à la décision dans des domaines d'activités extrêmement variés : banque, finance, assurance, e-commerce et grande distribution, communication et marketing, santé, agro-alimentaire, aéronautique et défense, internet des objets et télécoms, etc..

**Le parcours Data Science (DS)** du master mention Mathématiques et Applications est une formation de niveau un (Bac+5). Il vise à former des cadres data-scientists capables de mettre en œuvre les techniques mathématiques et informatiques inhérentes à ce métier, possédant en outre des compétences métiers propres aux secteurs d'activités visés par le choix d'options.

## Objectifs, compétences et capacités visés par la formation

Le master Data Science (DS) est une formation en deux ans qui débouche sur le diplôme national de Master mention Mathématiques et Applications, parcours Data Science. A l'issue de la formation, le ou la diplômé(e) possède :

**Des compétences en mathématiques.** Le diplômé maîtrise les méthodes de l'analyse statistique et algorithmique de l'information, qu'elles soient classiques (data-mining, classification, ..) ou spécifiques au big data (machine learning, ..), ainsi que les méthodes associées d'optimisation. Le diplômé est en capacité de modéliser les processus aléatoires, de les simuler et de confronter le modèle aux données statistiques du problème.

**Des compétences en informatique.** Le diplômé du master dispose des compétences relatives aux outils classiques de data management et d'analyse des données (R, Python, SQL,..), et plus spécifiques aux données de grandes dimensions et/ou non structurées : plateforme Hadoop, HDFS, NoSQL, etc..

**Des aptitudes professionnelles.** Le diplômé du master sait définir un projet, ses objectifs et son contexte, les modalités de réalisation et d'évaluation des actions, leurs priorités et leur planning. Il sait utiliser les technologies de l'information et de la communication, effectuer une recherche d'information à partir de sources diverses, organiser une veille technologique. Il a une bonne capacité de communication, une bonne maîtrise de l'anglais scientifique et technique (certification), et sait rédiger clairement une note ou un rapport selon des supports de communication adaptés.

**Des compétences métiers.** Suivant le choix d'option, le diplômé possède des aptitudes spécifiques au traitement et à l'analyse des données de la biologie (données omiques notamment), ou celles des entreprises dé-

diées au management du risque ou/et à l'optimisation des actions.

## Débouchés visés et rémunération des diplômés

Les diplômés du master Data Science (DS) occupent des emplois de cadre Data scientist (ou Data miner, Data analyst, etc..) couvrant un large spectre de secteurs d'activités, touchant aussi bien au domaine industriel qu'à celui du tertiaire. Sont plus particulièrement visés :

– Option données biologiques : secteurs de la santé, de la pharmacie, la bio-industrie, l'agro-industrie, les laboratoires de type Inserm, Cirad, Institut Pasteur, CEA, etc

..

– Option données numériques : tout secteur tel que sociétés de conseil en gestion des entreprises ou spécialisées en datamining, de la grande distribution, les grandes banques et compagnies d'assurances, etc..

L'enquête interne à 30 mois sur la promotion 2019-2020 montre une excellente insertion professionnelle (91% en emploi CDI ou CDD) pour un salaire médian brut annuel de 31980 €.

De même, l'enquête interne à 6 mois sur la promotion 2021-2022 relève un taux d'intégration professionnelle à 77 % avec un salaire médian brut annuel plus élevé de 34600 E (soit environ 2880 € brut par mois).

Le master est essentiellement à visée professionnelle. Des possibilités sont cependant offertes aux très bons étudiants d'une poursuite d'étude en doctorat par une thèse, en particulier dans le cadre d'une convention CIFRE avec une entreprise ou en partenariat avec un laboratoire de sciences appliquées. En moyenne (sur les 3 dernières années), environ 12% des étudiants poursuivent des études doctorales.

**i** **Enquête pour les diplômés du master ingénierie Mathématique dont le master DS est une évolution. Exemples de stages ou d'emplois récents : Renault DI-RC, INRA, iAdvize, CPAM Maine et Loire, LNE, Sodi-france-Netapsys, Atmo Poitou Charentes, Airparif, DAS-MMA, SFR, Ineox, AVISIA, Metro Cash & Carry France, Tronico groupe Alcen, Imbrication, Institut de Cancérologie de l'Ouest, CHU-Angers, Axens IFP Technologies, Coheris, BRED Banque populaire, Enza Zaden, Institut Gustave Roussy, Groupe PSA, CT Ingénierie, Harmonie Mutuelle, Banque de France, VYV Care IT, Capgemini, Sopra-Steria, Sigma Informatique, Socomec SAS,...**

## Public visé, pré-requis, conditions d'admission, tarification

Le master Data Science est à capacité limitée et l'admission en première (M1-DS) et deuxième année (M2-DS) du master est sélective. En dehors des qualités académiques, le recrutement accorde une part importante à la motivation des candidats. La validation du M1-DS entraîne l'admission de droit en M2-DS.

**Formation initiale.** Le M1-DS s'adresse en priorité aux étudiants diplômés d'une licence de Mathématiques, d'une licence de Mathématiques et informatique appliquées aux sciences humaines et sociales (parcours Mathématiques appliquées), ou d'un niveau équivalent. Sous réserve d'un niveau suffisant en mathématique, les dossiers d'étudiants ayant eu un parcours plus dirigé vers l'informatique sont considérés avec intérêt.

**Formation continue.** Les candidatures relevant de la formation continue sont les bienvenues et considérées avec la plus grande attention. Les tarifs sont variables et communiqués suite à l'étude du dossier. Un ensemble de financements est disponible (se renseigner auprès de la Direction de la Formation Continue).

Les pré-requis à l'entrée du M1-DS couvrent



le programme classique d'une Licence de Mathématiques ou tout autre programme européen ou international équivalent. **Outre l'aptitude mathématique, une connaissance et une pratique de base d'outils de programmation sont recommandées** (Python, Scilab, C et/ou R par exemple). L'acquis du certificat C2i-1 d'informatique est apprécié. Le diplôme de master visant des compétences en anglais certifiées de niveau B2 du CECRL, un niveau minimal B1 dans cette langue est recommandé à l'entrée du M1-DS. L'enseignement se faisant essentiellement en français, **un niveau minimal B2 certifié en français est exigé des étudiants internationaux à l'entrée du M1-DS.**

### Des raisons de choisir le master Data Science (DS) de l'UFR

La qualité de la recherche française en mathématiques fondamentales et appliquées est mondialement reconnue : suivant les critères choisis, elle se place du premier au troisième rang mondial. Dans ce cadre, la recherche en mathématique du laboratoire LAREMA de l'Université d'Angers (62 membres dont 40 permanents), associé au CNRS et partenaire du Labex Lebesgue, se distingue par son excellence. Aussi, **les étudiants du Master DS ont l'assurance de bénéficier des connaissances et des innovations les plus récentes, celles à même d'intéresser les entreprises dans leur stratégie compétitive.**

L'Université d'Angers est reconnue par ses résultats en matière de taux de réussite et d'insertion professionnelle de ses 22 700 étudiants : sur ces deux points, les enquêtes ministérielles la situent aux toutes premières places des universités françaises. **Les étudiants du Master DS ont l'assurance d'un accompagnement performant de leurs études jusqu'à leur insertion professionnelle.** Angers et son agglomération, ville étudiante par excellence (plus de 38 000 étudiants), apparaît dans les enquêtes nationales comme étant l'une des villes les plus attractives dans ce domaine : vitalité de la vie étudiante, qualité et modération

des prix du logement étudiant, facilité de déplacement, cadre de vie, etc.. En résumé, **les étudiants du Master DS ont l'assurance d'une qualité de vie et d'étude particulièrement propice à leur réussite et à leur insertion professionnelle.**



# VOLUMES HORAIRES - ÉVALUATIONS

## SEMESTRE 3

30 ECTS

UE	Matières	Volumes horaires					ECTS	Coef.	Contrôle des connaissances			
		CM	TD	TP	Auto. Et.	Tot.			1 <sup>ère</sup> session		2 <sup>ème</sup> session	Durée CT
									Assidus	D.A.		
1	Statistique en Grande Dimension et Apprentissage ❶	24	16	16	-	56	6	6	CT-0,5 CC-0,5	CT	CT	3h
2	Introduction à l'analyse du signal en temps discret	16	12	8	-	36	4	4	CT-0,5 CC-0,5	CT	CT	2h30
3	Séries Chronologiques et Applications	24	12	16	-	52	6	6	CT-0,5 CC-0,5	CT	CT	2h30 - SI 2h - S2
4	Informatique pour le Big Data ❶	24	0	32	-	56	6	6	CC	CT	CT	2h30
<b>Culture d'Entreprise</b>												
5	Anglais	0	0	16	-	16	2	2	CC	Oral	Oral	30mn
	Communication	0	0	6	-	6	1	1	Oral	Oral	Oral	1h
<b>Métiers 1 EC obligatoire + 1 EC d'Option au choix</b>												
6	<b>Obligatoire</b> : Apprentissage par renforcement et modèles génératifs 1	6	0	8	-	14	3	3	CC	CT	CT	2h
	<b>Option 1</b> : Données biologiques : Introduction à l'analyse de survie	8	0	10	-	18	2	2	CC	CT	CT	2h
	<b>Option 2</b> : Apprentissage par renforcement et modèles génératifs 2	8	0	8	-	16	2	2	CC	CT	CT	2h
	<b>Total option 1</b>	102	40	112	0	254	30					
	<b>Total option 2</b>	102	40	110	0	252	30					

❶ Note plancher de 8 à effectuer sur l'UE



### Conditions de validation du semestre 3

Admis·e si moyenne générale > ou= 10 et  
UE1 > ou= à 8 et  
UE4 > ou= à 8

CT = Contrôle Terminal  
CC = Contrôle Continu

P = Validation en Présentiel  
DA = Dispensé d'Assiduité

Auto. Et. = Autonomie Étudiant



## SEMESTRE 4

## 30 ECTS

UE	Matières	Volumes horaires					ECTS	Coeff.	Contrôle des connaissances			
		CM	TD	TP	Auto. Et.	Tot.			1 <sup>ère</sup> session		2 <sup>ème</sup> session	Durée CT
									Assidus	D.A.		
<b>Culture d'Entreprise</b>												
1	Conférences professionnelles et création d'entreprise	23	0	0	-	23	0	0	P	-	-	-
	3-PE	0	8	0	4	12	0	0	P	-	-	-
	Projet Annuel Tutoré	0	3	0	53	56	7	7	rapport + soutenance	-	-	-
<b>Métiers 1 EC obligatoire + 1 EC d'Option au choix</b>												
2	<b>Obligatoire</b> : Traitement d'images	12	0	12	-	24	3	3	CC	-	-	-
	<b>Option 1 Bio</b> : Inférence de réseaux biologiques	8	0	8	-	16	2	2	CC	-	-	-
	<b>Option 2 Éco</b> : Économétrie de l'Évaluation	14	12	0	-	26	2	2	CC	-	-	-
<b>Choix de 1 parmi 2 entre stage ou alternance</b>												
3	Stage	0	0	0	-	0	18	18	Rapport + soutenance	-	-	-
	Alternance											
	<b>Total option Bio</b>	43	11	20	57	131	30					
	<b>Total option Éco</b>	49	23	12	57	141	30					

<b>Total année option Bio</b>	<b>145</b>	<b>51</b>	<b>132</b>	<b>57</b>	<b>385</b>
<b>Total année option Éco</b>	<b>151</b>	<b>63</b>	<b>122</b>	<b>57</b>	<b>393</b>



### Conditions de validation du semestre 4

Admis-e si moyenne générale  $\geq 10$  et Si Conférences professionnelles validées par la présence



### Conditions de validation de l'année

Admis-e si  $(\text{Semestre 3} + \text{Semestre 4}) / 2 \geq 10$  et conditions de validation des semestres 3 et 4 validées



Pas de DA et pas de Session 2 au semestre 4

CT = Contrôle Terminal  
CC = Contrôle Continu

P = Validation en Présentiel  
DA = Dispensé d'Assiduité

Auto. Et. = Autonomie Étudiant



# CONTENU DES ENSEIGNEMENTS

## SEMESTRE 3

Les noms des responsables et intervenants sont donnés sous réserve de modifications.

### BLOC MATHÉMATIQUES - UE OBLIGATOIRES

#### UE1

#### STATISTIQUE EN GRANDE DIMENSION ET APPRENTISSAGE

Statistics for Big Data and Machine Learning

Responsable Fabien Panloup  
[fabien.panloup@univ-angers.fr](mailto:fabien.panloup@univ-angers.fr)

#### PRÉREQUIS

Notions et contenus

Algèbre linéaire et analyse (licence mathématiques L3). Statistique inférentielle (S1-UE3-DS). Optimisation non linéaire (S1-UE2-DS). Datamining et classification (S2-UE1-DS). Manipulation standard de R et de Python.

Compétences et capacités :

- Maîtriser les compétences enseignées dans les modules de statistique inférentielle et de datamining et classification du MI, particulièrement tout ce qui concerne les modèles de régression linéaire et les méthodes de classification.
- Maîtriser les bases de l'algèbre linéaire, du calcul matriciel, de l'analyse et de l'optimisation, en particulier : recherche des valeurs propres et singulières, la notion de produit scalaire (sur l'exemple de  $\mathbb{R}^n$  et de  $L^2$ ), les propriétés des espaces de Hilbert de fonctions ou encore l'optimisation sous contraintes.
- Maîtriser les notions de loi conditionnelle/espérance conditionnelle.
- Maîtriser les bases de l'optimisation déterministe (descente de gradient et extensions).
- Avoir une connaissance minimale des langages R et Python (syntaxe, manipulation élémentaire, calcul matriciel, gestion des graphiques, etc.).

#### CONTENUS

Bases mathématiques de l'apprentissage statistique. Bases de la grande dimension

(« fléau » de la dimension...). Régression en grande dimension (motivation par l'exemple, théorie du Lasso, seuillage, sparsité, inégalités classiques, régressions basées sur les composantes principales, PLS). Méthodes de pénalisation (généralisation du Lasso (Elastic-Net, Group-Lasso), sélection de variables, Ridge, etc.). Classification et apprentissage en grande dimension (SVM, Logistic-Lasso, forêts aléatoires, k plus proches voisins, ACP à noyaux, Boosting, Gradient Boosting). Introduction aux réseaux de neurones et au deep learning. Bases de l'optimisation stochastique. Large mise en pratique avec R ou Python.

#### COMPÉTENCES

- Avoir un recul professionnel sur la notion de grande dimension. Comprendre par l'exemple les raisons théoriques mais aussi les raisons techniques qui différencient les méthodes statistiques classiques des méthodes statistiques en grande dimension.
- Comprendre la théorie du Lasso et d'une manière générale la méthodologie consistant à pénaliser une estimation par des contraintes sur les coefficients de la régression. Avoir une vision claire des objectifs visés par les méthodes de pénalisation (estimation ou sélection de variables).
- Savoir réduire la dimension des observations par des méthodes classiques d'analyse des données (par exemple l'ACP) et savoir exploiter cette réduction de dimension pour les régressions en grande dimension ou possédant de la multicollinéarité.
- Savoir ce qui relie et ce qui différencie le Lasso des méthodes de pénalisation basées sur des normes différentes (Elastic-Net, Ridge, etc.). Connaître les points forts et points faibles de chacune des approches pour en permettre une mise en pratique pertinente sur des jeux de données réelles.
- Connaître les méthodes simples d'apprentissage automatique appliquées à la classi-



fication (comme les  $k$  plus proches voisins) et les méthodes plus sophistiquées (comme les forêts aléatoires). Avoir une vision d'ensemble de la théorie des SVM appliqués à la classification.

- Connaître les bases du boosting (Adaboost/Gradient Boosting, XGBoost).
- Comprendre le fonctionnement et savoir mettre en oeuvre un algorithme d'apprentissage par deep learning (perceptron et réseaux de neurones convolutionnels).
- Être capable de mobiliser et exploiter les méthodes étudiées dans ce module dans des cas pratiques sous R ou Python, en particulier les méthodes de régression et de sélection de variables en grande dimension.

## BIBLIOGRAPHIE

- Christophe Giraud, Introduction to High-Dimensional Statistics (2015).

### UE2

## INTRODUCTION À L'ANALYSE DU SIGNAL EN TEMPS DISCRETS

*Introduction to Discrete-Time Signal Processing*

Responsable **David Rousseau**  
[david.rousseau@univ-angers.fr](mailto:david.rousseau@univ-angers.fr)

## PRÉREQUIS

*Notions et contenus*

Espace euclidien, produit scalaire (licence mathématiques L3); séries de Fourier (licence mathématiques L2).

*Compétences et capacités*

Maîtriser le calcul des coefficients d'une série de Fourier, la décomposition d'une fonction sur une base orthogonale ; connaître et savoir appliquer le théorème de Parseval et les inégalités de Bessel ; maîtriser le langage de programmation Python.

## CONTENUS

Analyse du signal, motivations, exemples  
Séries de Fourier, Transformée de Fourier  
Filtres FIR (Réponse Impulsionnelle Finie) et IIR (Réponse Impulsion Infinie) – Impulsion – convolution  
Transformation en Z – Application à la stabilité des filtres  
Problèmes d'aliasing

Application à des signaux temporels de natures diverses

## COMPÉTENCES

- Maîtriser les calculs de transformation de Fourier et de Fourier Inverse.
- Savoir tester la stabilité d'un filtre.
- Savoir concevoir un filtre correspondant à des objectifs donnés
- Savoir utiliser Python et des bibliothèques de type `scipy.signal` pour mettre en pratique la conception des filtres.

## BIBLIOGRAPHIE

- Tisserand, É., Pautex, J. F., & Schweitzer, P. (2009). Analyse et traitement des signaux-2e éd.: Méthodes et applications au son et à l'image. Dunod.

### UE3

## SÉRIES CHRONOLOGIQUES ET APPLICATIONS

*Time Series and Applications*

Responsable **Frédéric Proïa**  
[frederic.proia@univ-angers.fr](mailto:frederic.proia@univ-angers.fr)

## PRÉREQUIS

*Notions et contenus*

- Bagage mathématique (algèbre, probabilités, etc.) de Licence et de Master 1.
- Modules de Statistique et de Modèles de régression du M1.
- Manipulation standard de R et de Python.

*Compétences et capacités*

- Savoir manipuler les opérateurs du calcul des probabilités (espérance, variance, covariance, etc.) appliqué aux variables et aux vecteurs aléatoires.
- Maîtriser les notions de statistique inférentielle du S1 et de régression linéaire du S2.
- Maîtriser les bases de l'algèbre linéaire et de l'analyse concernant les espaces vectoriels, les produits scalaires et la projection orthogonale (sur l'exemple de  $\mathbb{R}^n$  et de  $L^2$ ).
- Avoir une connaissance minimale des langages R et Python (syntaxe, manipulation élémentaire, gestion des graphiques, etc.).

## CONTENUS

Décomposition des séries chronologiques



(modèles additif et multiplicatif, filtrage linéaire). Processus aléatoires discrets (bruit blanc, loi d'un processus, espace  $L^2$ ). Stationnarité (sens fort, sens faible). Processus linéaires (opérateurs chronologiques, bruit d'innovation, causalité, ACF/PACF). Étude détaillée des processus ARMA. Inférence dans les processus ARMA. Introduction à la non-stationnarité (ARIMA-SARIMA) et à l'hétéroscédasticité (ARCH-GARCH). Application à la prévision (meilleur prédicteur linéaire, lissages exponentiels). Introduction à la statistique spatiale (krigeage). Large mise en pratique avec R et Python.

## COMPÉTENCES

- Être capable de reconnaître visuellement une série à composantes additives ou multiplicatives, avec tendance et/ou périodicité, et d'éliminer les tendances par un filtrage approprié.
- Savoir étudier la stationnarité d'un processus discret et savoir la tester en pratique. Comprendre la différence entre stationnarité forte et stationnarité au second ordre.
- Savoir calculer la fonction d'autocorrélation d'un processus stationnaire. Savoir construire son autocorrélation partielle par un algorithme approprié.
- Acquérir une approche théorique solide d'un processus ARMA. Savoir établir des conditions sur ses paramètres permettant d'étudier ses propriétés.
- Savoir estimer l'ACF et la PACF d'une série chronologique stationnaire, en proposer une modélisation ARMA adaptée par comparaison avec le comportement théorique des autocorrélations.
- Savoir étendre la définition des processus ARMA aux processus ARIMA, SARIMA, ARCH et GARCH ainsi que leur mise en pratique.
- Connaître les principaux tests d'hypothèses adaptés au cadre ARMA/ARIMA/SARIMA (tests de stationnarité, de significativité des paramètres, d'autocorrélation, etc.).
- Avoir un recul professionnel sur la modélisation d'une série, savoir comparer plusieurs approches sur la base de critères adaptés aux objectifs de l'étude.
- Être capable de mobiliser toutes ces compétences dans des cas pratiques avec R et Python.
- Connaître les rudiments de la statistique

spatiale par l'intermédiaire du krigeage.

## BIBLIOGRAPHIE

– P. Brockwell, R. Davis. Time Series: Theory and Methods. Seconde édition. Springer Series in Statistics, 1991.



## UE4

### INFORMATIQUE POUR LE BIG DATA

Big Data

Responsable **Jacquelin Charbonnel**

[jacquelin.charbonnel@univ-angers.fr](mailto:jacquelin.charbonnel@univ-angers.fr)

intervenants **Marie Badreau, Mohamed Benzarga, Jacquelin Charbonnel, David Rousseau**

### PRÉREQUIS

*Notions et contenus*

Bases de données relationnelles (S2-UE4-DS). Programmation (S1-UE5-DS). Statistique (S2-UE1-DS).

*Compétences et capacités*

Connaissances de base des modèles relationnels. Connaissances en statistique et des langages de programmation.

### CONTENUS

#### Concepts et les enjeux du Big Data

Technologies du Big Data :

- Architecture et composants de la plateforme Hadoop.
- Modes de stockage (NoSQL, HDFS).
- Principes de fonctionnement de MapReduce.
- Programmation concurrente (générateurs, multi-threading, multi-processing, asynchronisme).
- Web services.
- Nasse de data visualisation.

Gérer les données structurées et non structurées :

- Principes de fonctionnement HDFS.
- Importer des données externes vers HDFS.
- Réaliser des requêtes SQL avec HIVE.
- Utiliser PIG pour traiter la donnée.

Les méthodes d'analyse des données pour le Big Data (en liaison avec S3-UE1-DS) :

- Les méthodes d'exploration.
- Segmentation et classification.
- Estimation et prédiction.
- Implémentation des modèles.
- Méthodes de réduction de dimensions.

Introduction au MLOps : Mise en production ou déploiement d'un modèle de Machine

Learning.

- Aspects techniques d'un déploiement.
- Mise en pratique : déploiement d'une API Web avec Flask.
- Maintenance post-déploiement : logging, model drift, qualité des données.
- Prérequis d'un déploiement réussi : versionnage, tests unitaires, bonnes pratiques de code, etc.
- Vue d'ensemble des métiers et logiciels liés au MLOps.

### COMPÉTENCES

- Comprendre les concepts et l'apport du Big Data par rapport aux enjeux métiers.
- Comprendre l'écosystème technologique nécessaire pour réaliser un projet Big Data.
- Acquérir les compétences techniques pour gérer des flux de données complexes, non structurés et massifs.
- Visualisation d'espaces latents, implémenter des modèles d'analyses statistiques pour répondre aux besoins métiers.
- Comprendre l'intérêt du déploiement d'un modèle de Machine Learning
- Avoir une vue d'ensemble des points d'attention d'une mise en production, avant, pendant et après sa réalisation
- Acquérir une compétence technique de base pour créer une API Web avec Flask ; mettre en pratique quelques aspects du MLOps.

## BLOC CULTURE D'ENTREPRISE - UE OBLIGATOIRES

UE5

### ANGLAIS

English

Responsable **Philippe Torrès**

*philippe.torres@univ-angers.fr*

Intervenant **Bruno Menan**

*bruno.menan@univ-angers.fr*

### PRÉREQUIS

*Notions et contenus*

Les bases de la langue anglaise.

*Compétences et capacités*

Maîtriser le niveau B1 du CECRL (dit « d'utilisateur indépendant »).

### CONTENUS

– Permettre aux étudiants de continuer à travailler les cinq compétences en langue - Compréhension écrite et orale, expression écrite et orale, et interaction orale - à travers des supports authentiques (articles, documentaires, documents audio et vidéo d'internet, graphiques...) et des activités variées (exercices de compréhension, d'expression écrite, jeux de rôle, débats, présentations orales...).

– Etoffer les connaissances lexicales.

– Améliorer la prononciation et revoir certains points de langue le cas échéant.

– Valider à l'issue du 3ème semestre un niveau B2 du CECRL (par le biais d'une certification).

### COMPÉTENCES

On vise, à l'issue du 2ème voire 3ème semestre du Master, le niveau de compétence B2 du Cadre Européen Commun de Référence pour les Langues (CECRL) qui est résumé comme suit : « Peut comprendre le contenu essentiel de sujets concrets ou abstraits dans un texte complexe, y compris une discussion technique dans sa spécialité. Peut communiquer avec un degré de spontanéité et d'aisance tel qu'une conversation avec un locuteur natif ne comporte de tension ni pour l'un ni pour l'autre. Peut s'exprimer de façon claire et détaillée sur une grande gamme de sujets, émettre un avis sur un sujet d'actualité et exposer les avantages et les inconvé-

nients de différentes possibilités. »

UE5

### COMMUNICATION – PRÉPARATION À L'INSERTION PROFESSIONNELLE

*Communication – Professional Integration Program*

Responsable **Fabien Panloup**

*fabien.panloup@univ-angers.fr*

Intervenants **Intervenants du SUIO-IP et/ou autres intervenants professionnels**

### PRÉREQUIS

Compétences de l'UE S2-UE5 communication.

### CONTENUS

– Techniques de communication - Préparation au travail en groupe : Les fonctions de l'animateur ; Conduite de réunions - Techniques de rédaction des documents - Techniques de présentation orale - La recherche d'emploi : Lettres de candidature ; Curriculum vitae ; Entretien de candidature.



### APPRENTISSAGE PAR RENFORCEMENT ET MODÈLES GÉNÉRATIFS 1

Reinforcement Learning and Generative Models 1

Responsable **Sylvain Lamprier**

[sylvain.lamprier@univ-angers.fr](mailto:sylvain.lamprier@univ-angers.fr)

Intervenant **Sylvain Lamprier**

#### PRÉREQUIS

Solide bases d'apprentissage statistique, en particulier supervisé

- Notions sur l'apprentissage profond
- Connaissances en programmation NumPy et PyTorch

#### CONTENUS

Dans le contexte de l'apprentissage automatique, l'apprentissage supervisé est souvent privilégié pour sa capacité à prédire des étiquettes ou des valeurs cibles à partir de données d'entraînement étiquetées. Cependant, ces données sont souvent difficiles à obtenir, et les modèles qui en résultent sont limités par la qualité et le domaine de cette supervision. Lorsque l'étiquetage est trop complexe ou trop coûteux à collecter, au regard de la variété des situations auxquelles l'agent décisionnel peut être confronté, l'apprentissage par renforcement offre une alternative prometteuse. L'apprentissage par renforcement, contrairement à l'apprentissage supervisé, n'exige pas une supervision explicite pour chaque exemple de données. Au lieu de cela, l'agent apprend au travers d'interactions avec son environnement, sur la base de récompenses collectées en fonction de l'utilité de ses actions pour la tâche visée. Cette capacité à apprendre par l'expérience rend l'apprentissage par renforcement applicable à un bien plus large spectre d'applications, et permet de connecter plus finement les objectifs d'apprentissage optimisés avec les objectifs souhaités pour le modèle résultant.

L'apprentissage par renforcement est particulièrement bien adapté aux problèmes de prise de décision séquentielle, où les actions prises à un moment donné affectent les ré-

compenses futures. Par exemple, dans le domaine de la robotique, un robot peut apprendre à naviguer dans un environnement en effectuant des actions et en observant les récompenses associées à ces actions, telles que la distance parcourue ou la minimisation des collisions. C'est également le cas pour des problèmes de génération de données séquentielles, tels que la génération de langue par exemple, où les modèles de générations pré-appris de manière auto-supervisée à prédire le prochain mot en fonction des précédents peuvent être spécialisés pour des tâches particulières par apprentissage par renforcement, permettant de considérer des scores globaux que les séquences produites (e.g. ChatGPT a été finetuné par apprentissage par renforcement sur des préférences d'utilisateurs).

Ce cours vise à couvrir les principaux concepts et méthodes de ce paradigme d'apprentissage en pleine expansion, en mettant l'accent sur les différentes approches et leurs applications pratiques. Il pourra se structurer de la manière suivante :

1. Introduction à l'Apprentissage par Renforcement
  - Comparaison avec l'apprentissage supervisé
  - Processus de Décision de Markov (MDP)
  - Méthodes Tabulaires de Planification
  - Algorithme de Bellman
  - Policy Iteration et Value Iteration
2. Méthodes basées sur des Valeurs
  - Q-Learning, Sarsa
  - Deep Q-Networks (DQN)
3. Apprentissage par Policy Gradient et Méthodes Actor-Critic
  - Policy Gradient
  - Méthodes Actor-Critic: A2C, TRPO, PPO,
4. Apprentissage Hors Politique et Méthodes pour Actions Continues
  - Méthodes hors politique
  - Méthodes pour actions continues (DDPG, SAC, TD3)

## COMPÉTENCES

- Comprendre les différences de paradigme entre apprentissage supervisé et apprentissage par renforcement,
- Savoir formaliser des problèmes décisionnels sous la forme de problèmes d'apprentissage par renforcement
- Savoir implémenter les principaux algorithmes (et les faire fonctionner sur des environnements simples GYM)



## BLOC MÉTIERS - UE OPTIONNELLES (un choix sur deux)

UE6

### OPTION 1

#### DONNÉES BIOLOGIQUES : INTRODUCTION À L'ANALYSE DE LA SURVIE

Introduction to Survival Data Analysis

Responsable Mikael Escobar-Bach  
[mikael.escobar-bach@univ-angers.fr](mailto:mikael.escobar-bach@univ-angers.fr)  
Intervenant Mikael Escobar-Bach,

#### PRÉREQUIS

*Notions et contenus*

«Statistique» et «Modèles de régression» (cours du MI DS).

*Compétences et capacités*

Maîtriser les concepts généraux des probabilités, des statistiques et des modèles de régression vus en MI.

Le problème des données manquantes est fréquemment rencontré dans un grand nombre de domaines d'applications statistiques. Dans le cadre de la biologie, il n'est pas rare que l'acquisition de données de suivi soit en partie incomplète : pour un essai thérapeutique, le temps de rémission puis de guérison d'un patient peut être trop long et donc jamais observé, on parle alors de données censurées.

Dans ce cours, nous introduirons la méthodologie (fonction de survie, risque instantané, ...) et les outils statistiques (estimateur de Kaplan-Meier, Nelson-Aalen, ...) qui permettent d'appréhender ce genre de problèmes. Une implémentation sur R sera aussi proposée.

UE6

### OPTION 2

#### APPRENTISSAGE PAR RENFORCEMENT ET MODÈLES GÉNÉRATIFS 2

Reinforcement Learning and Generative Models 2

Responsable Sylvain Lamprier  
[sylvain.lamprier@univ-angers.fr](mailto:sylvain.lamprier@univ-angers.fr)  
Intervenant Sylvain Lamprier

#### PRÉREQUIS

*Notions et contenus*

- Solide bases fondamentales de l'apprentissage statistique, supervisé et par renforcement
- Notions en modélisation probabiliste et statistique
- Compétences en programmation Numpy et PyTorch
- Connaissances souhaitables en modélisation de données complexes, telles que du texte ou des images

#### CONTENUS

Les méthodes génératives profondes suscitent depuis récemment un très grand intérêt dans la société, pour les impressionnantes capacités que ces outils (e.g., Dall-E, ChatGPT, SORA) démontrent, en terme de qualité de contenu produit et d'interprétation des besoins de l'utilisateur. L'objectif de ce cours est de découvrir les concepts fondamentaux sous-jacents au fonctionnement de ces applications puissantes, en commençant par quelques fondamentaux sur l'apprentissage génératif, et les méthodes modernes pour guider cet apprentissage (VAE, VQ-VAE, GAN, Diffusion, apprentissage contrastif, etc.). L'idée de l'apprentissage génératif est de reproduire des distributions de données dont on observe des échantillons en apprentissage, ce cours se confronte à cet objectif ambitieux pour des distributions de données complexes, tels que des images ou du texte. Nous examinerons notamment comment les transformeurs révolutionnent la génération de texte en capturant des dépendances long terme, permettant d'envisager un très large spectre d'applications guidées



par le langage. Nous explorerons comment l'apprentissage par imitation peut être utilisé pour générer des données en imitant des comportements ou des styles spécifiques. Enfin, nous verrons comment l'apprentissage par renforcement peut guider la génération de contenu créatif en optimisant les actions pour maximiser une récompense donnée.

### 1. Concepts Fondamentaux de l'Apprentissage Génératif

- Exploration des principes de base de l'apprentissage génératif.
- Présentation des grandes familles d'algorithmes d'apprentissage génératif: VAE (et VQ-VAE), GANs, modèles de diffusion, apprentissage contrastif, etc.

### 2. Architectures neuronales profondes pour la prise en compte de dépendances long terme

- Différentes familles d'architectures profondes : MLP, CNNs, RNNs, Transformers
- Applications pour la générations de données complexes
- Modèles de fondation guidés par le langage: techniques d'adaptation (in-context learning, LORA, mixture d'experts, etc.)

### 3. Apprentissage par Imitation et par Renforcement pour la Génération de Contenu

- Exploration de l'utilisation de l'apprentissage par renforcement pour guider la génération de contenu, en optimisant les actions pour maximiser une récompense donnée (e.g., RLHF)
- Apprentissage par imitation hors ou en ligne (BC, AWR, GAIL, iQL, etc.)

## COMPÉTENCES

- Maîtriser les différents paradigmes de l'apprentissage génératif (optimisation bayésienne)
- Connaître les grandes familles d'architectures profondes adaptées aux données complexes
- Comprendre les enjeux de l'apprentissage par renforcement dans le domaine de la génération de données
- Savoir implémenter et appliquer les principaux algorithmes d'apprentissage ou d'adaptation des modèles génératifs modernes



# SEMESTRE 4

## BLOC CULTURE D'ENTREPRISE - UE OBLIGATOIRE

UE1

### CRÉATION D'ENTREPRISE - PRÉPARATION À L'INSERTION PROFESSIONNELLE

Responsable **Fabien Panloup**  
*fabien.panloup@univ-angers.fr*  
Intervenants **SUIO-IP**

#### CONTENUS

L'enseignement dédié à la création d'entreprise vise à fournir aux étudiants les éléments méthodologiques nécessaires. Il s'appuiera sur l'intervention de professionnels et sollicitera un travail personnel des étudiants sous forme de capsule.

#### Conférences Professionnelles

Responsable **Fabien Panloup**  
*fabien.panloup@univ-angers.fr*  
Intervenants **Extérieurs**

#### CONTENUS

Des exposés et des mini-cours seront proposés par des professionnels à la fois pour présenter le métier de data scientist mais aussi pour compléter la formation par des introductions à des thèmes de développement souvent récents.

#### Projet Annuel Tutoré

*Annual Training Project*

Responsable **Fabien Panloup**  
*fabien.panloup@univ-angers.fr*

#### CONTENUS

Les étudiants réalisent en groupe un projet annuel entre septembre et février avec une soutenance à mi-projet, sous l'encadrement d'un tuteur enseignant. La réalisation du projet s'efforcera de calquer le déroulement de la gestion de projets dans le monde socio-économique. Dans une première phase, après avoir circonscrit le problème, un cahier des charges sera réalisé et un calendrier sera défini pour la réalisation. La seconde

phase du projet consistera au traitement du problème en utilisant les concepts, méthodes et outils appropriés et satisfaisants au cahier des charges. Le projet annuel met en oeuvre l'ensemble des compétences théoriques et pratiques du master, sur un thème en lien avec l'option métier.

#### COMPÉTENCES

- Autonomie.
- Capacité à travailler en équipe.
- Capacité à mettre en oeuvre les contenus du master.
- Capacité à rédiger un rapport.
- Capacité à réaliser une soutenance.
- Capacité créative et de recherche.



### TRAITEMENT D'IMAGES

Responsable **David Rousseau**  
*david.rousseau@univ-angers.fr*

#### PRÉREQUIS

*Notions et contenus*

- Contenus des cours de -S3-M2 DS « Statistique en Grande Dimension et Apprentissage » et « Bases de Données Avancées-Big Data ».

*Compétences*

- Traitement des données avec Python.  
- Bases sur les réseaux de neurones.

#### CONTENUS

Cet enseignement a pour objectif d'introduire les concepts principaux du traitement d'images pour les mettre en oeuvre via des outils logiciels libres standards (Scikit image Python, imageJ, ilastik, Knime, ...) sur des questions variées. A l'issue du cours, les étudiants devront être capables de construire un « pipeline » de traitement d'images pour réaliser de l'apprentissage machine basé expert ou par réseau de neurones profonds.

#### BIBLIOGRAPHIE

– Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms (p. 288). The MIT Press.  
– Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more. O'Reilly Media.

## UE OPTIONNELLES (un choix sur deux)

UE2

### OPTION 1 BIO

#### INFÉRENCES DE RÉSEAUX BIOLOGIQUES

Responsable **Olivier Goudet**  
*olivier.goudet@univ-angers.fr*

#### PRÉREQUIS

*Notions et contenus*

- Contenu du cours S2-M1 DS « Introduction à la génomique ».
- Contenus des cours de S3-M2 DS « Statistique en Grande Dimension et Apprentissage » et « Bases de Données Avancées-Big Data ».

*Compétences*

- Traitement des données avec Python.
- Modèles de régression multivariée.
- Modèles graphiques.

#### CONTENUS

Cet enseignement se focalise principalement sur l'inférence de réseaux et la transcriptomique. Il a pour objectif de renforcer les compétences en analyse de données génomiques et d'introduire les réseaux biologiques.

Cet enseignement sera dispensé sous forme de CM et de TP de mise en pratique sur des données réelles. La réalisation d'un projet de traitement de données transcriptomiques et d'inférence de réseaux sera proposée à l'issue du cours.

#### COMPÉTENCES

Introduction aux réseaux biologiques | Méthodes d'inférences de réseaux | Causalité | Méthode de validation.

#### BIBLIOGRAPHIE

- Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms (p. 288). The MIT Press.
- Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more. O'Reilly Media.

UE2

### OPTION 2 Éco

#### ÉCONOMÉTRIE DE L'ÉVALUATION

*Econometric Techniques*

Responsable **Christophe Daniel**  
*christophe.daniel@univ-angers.fr*

#### PRÉREQUIS

*Notions, contenus et compétences*

- Modèles micro-économiques du travail-Econométrie des données transversales et de panel.

#### CONTENUS

**1** Théories microéconomiques :

- Des inégalités et des discriminations (hommes-femmes).
- De l'éducation (rendements de l'éducation ou de l'expérience), négociations collectives.
- Du choix des conditions de travail (théories des différences compensatrices, de la segmentation, et des salaires hédoniques).
- De la satisfaction au travail.

**2** Passage des modèles théoriques aux modèles économétriques : fonctions de gains (Mincer, élargies aux conditions de travail...), équation de choix des conditions de travail, biais d'endogénéité et de sélection, méthodes de correction (Variables instrumentales et méthode d'Heckman de l'inverse du Ratio de Mill), modèle à deux régimes, décomposition de Blinder-Oaxacca, de Machado-Mata, indices de satisfaction...

**3** Applications sur données transversales : estimations et tests

Méthodes d'estimation : MCO, Moindres Carrés Pondérés, Moindres Carrés Généralisés, probit et logit (odd-ratios, modèles simples ou bivariés), Variables instrumentales, Heckman, estimation par quantiles, Principal Component Regression et Partial Least Squares.

**4** Évaluations contrefactuelles des politiques publiques de l'emploi :

Approche expérimentale (groupe test, groupe de contrôle), différences-en-différences, score de propension, bootstrap...



### UE3

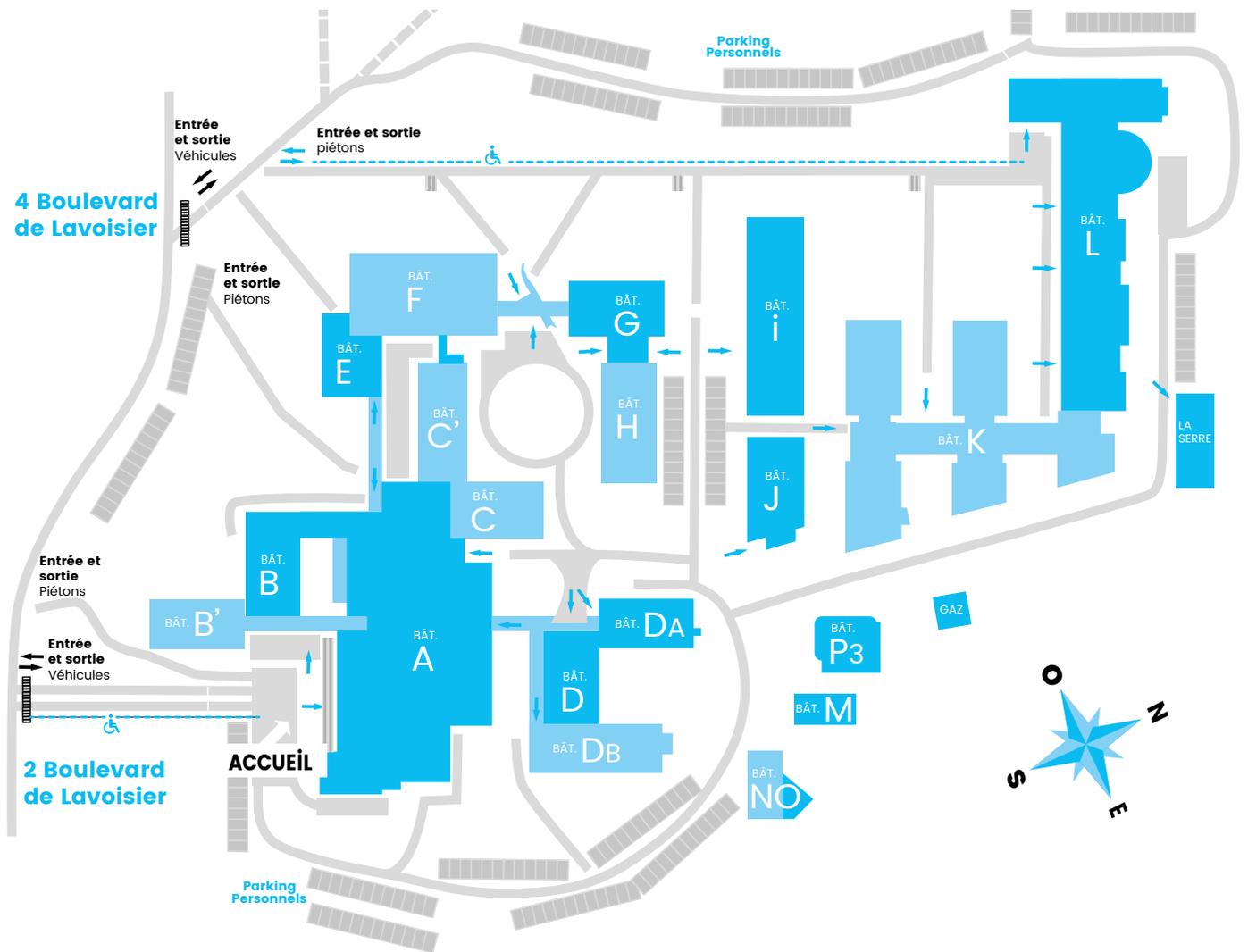
#### STAGE OU ALTERNANCE

*Stage de 5 à 6 mois en entreprise ou alternance sur l'année*

L'étudiant devra accomplir un stage en entreprise ou dans un laboratoire de recherche sur un sujet en adéquation avec les objectifs de la formation. Pour les alternants, la période de stage s'effectue sur les périodes de l'année scolaire définies dans le calendrier ci-joint. Pour les étudiants en formation initiale, il s'agira d'un stage de 22 à 26 semaines environ, à partir de la mi-mars. Les étudiants sont suivis par un tuteur entreprise et par un tuteur université.

Ce stage est finalisé par la rédaction d'un rapport d'une trentaine de pages et par une soutenance qui se déroule début septembre. La soutenance se déroule devant un jury constitué de membres de l'équipe pédagogique du master, en présence d'un ou plusieurs membres de la structure dans laquelle s'est déroulé le stage (possibilité de visioconférence pour ces derniers). L'évaluation prend en compte l'avis des tuteurs entreprise, la qualité scientifique et rédactionnelle du rapport ainsi que la prestation de l'étudiant lors de la soutenance, tant dans la présentation que dans les réponses aux questions.

N.B. Les stages peuvent se prolonger jusqu'à la veille de la date de jury si nécessaire.



- A** Scolarité | Accueil | Enseignement (Amphi A à E) | Administration
- B** Enseignement biologie
- B'** Enseignement biologie
- C** Enseignement chimie
- C'** Recherche
- D** Enseignement physique
- Da** Enseignement physique
- Db** Recherche
- E** Enseignement biologie
- F** Enseignement biologie | Recherche
- G** Enseignement géologie | informatique
- H** Enseignement informatique | Recherche
- i** Enseignement mathématiques | Recherche
- J** Enseignement chimie
- K** Recherche
- L** Enseignement transversaux | Enseignement (Amphi L001 à L006)

Impression Service Reprographie UA

