

## FICHE SAÉ 6-EMS-01

# Modélisation statistique pour les données complexes et le Big Data – Étude de cas 2

Une SAÉ (Situation d'Apprentissage et d'Évaluation) est une tâche **authentique** consciemment organisée pour permettre le développement de compétences. Elle demande de **réaliser** une production (matérielle ou immatérielle) **proche** de celles exigées d'un **professionnel**. Cette production doit faire sens pour l'étudiant. Elle nécessite de sa part de choisir les ressources internes et externes à mobiliser et à combiner, mais aussi de se **situer** régulièrement (en cours de route et en fin de tâche) par rapport à l'**objectif** qu'il s'est fixé en s'autoévaluant ou en bénéficiant du regard de ses enseignants, experts et pairs.

*F. Georges et M. Poumay, LabSET-Ulège, 2020*

### INFORMATIONS GENERALES

Année	2023/2024
Semestre	Semestre 6
BUT	Science des Données
Parcours	*Science des données : exploration et modélisation statistique
UE Concernée(s)	S6UE4 - Modéliser les données dans un cadre statistique
Pondération de la SAÉ dans l'UE	0.5 dans la SAÉ
Objectif de la SAÉ	<p>Il s'agit d'étudier la capacité des pays à faire face au changement climatique, et de faire une étude prospective visant à analyser les évolutions jusqu'à maintenant, pour envisager les évolutions futures.</p> <p>*L'étudiant doit choisir une thématique parmi les suivantes : démographie, social, économique, géophysique, ou un mélange entre ces thématiques. Cette thématique devra être formulée explicitement.</p> <p>*L'étudiant doit choisir des données en fonction de sa thématique. Si les données définies par la thématique ne sont pas satisfaisantes, il pourra adapter sa problématique après avoir expliqué cette évolution.</p>
Compétence(s) ciblée(s)	Analyser statistiquement les données
Niveau	Niveau 3. Mettre en œuvre l'analyse exploratoire de données complexes

<b>Situation</b>	<p>Vous devez fournir un outil permettant d'explorer facilement les enjeux internationaux dus au changement climatique, et comparer différentes zones du globe pour mettre en correspondance les risques encourus et les capacités à s'adapter.</p> <p>On fera notamment une modélisation d'indicateurs pour prévoir leur comportement à venir, une étude de corrélation entre les indicateurs temporels, une analyse de données pour classer les pays par ressemblance.</p>
<b>Apprentissage(s) critique(s) associé(s)</b>	<p>- AC32.01   Prendre conscience des différences entre des outils statistiques pour choisir le plus adapté</p> <p>- AC32.02   Saisir l'importance de la mise en œuvre de méthodes adaptées à des domaines et des données spécifiques (Statistique spatiale, temporelle, prévisions ...)</p> <p>- AC32.03   Prendre conscience des limites des méthodes classiques pour l'analyse des données complexes (données massives, données mal structurées...)</p> <p>– AC34.01EMS   Comprendre l'intérêt des approches statistiques pour la fiabilisation, la validation, les incertitudes, les imprécisions des données</p> <p>– AC34.02EMS   Comprendre l'intérêt de la problématique métier pour réaliser la modélisation</p> <p>– AC34.03EMS   Viser la réalisation d'un processus de modélisation dans son ensemble</p> <p>– AC34.04EMS   Prendre conscience des différences entre les modèles pour choisir le plus adapté</p> <p>– AC34.05EMS   Prendre conscience de la nécessité d'utiliser des moyens spécifiques pour analyser les données massives ou les flux de données</p>
<b>Ressource(s) mobilisée(s)</b>	– R6.01   Big Data : enjeux, stockage et extraction
<b>Heures encadrées</b>	30.5, dont 16.5 sont de l'autonomie encadrée
<b>Heures en autonomie planifiées</b>	40h
<b>Durée de la SAÉ</b>	8 semaines
<b>Nombre d'étudiants concernés</b>	15 étudiants
<b>Equipe pédagogique / académique</b>	Sylvie Viguière-Pla

## CONSIGNES

<b>Sujet</b>	<b>Une analyse des enjeux liés au changement climatique</b>
<b>Outil(s) à utiliser</b>	<b>R, SAS, Word, plus tout autre logiciel qui sera jugé utile</b>
<b>Exemples de sites de téléchargement des données</b>	<p>Les données sont à télécharger depuis des sites WEB. Voici une liste non exhaustive de sites où se trouvent les données à récupérer (un choix sera à faire selon la pertinence).</p> <p>Données d'évolution des principaux indicateurs climatiques (CO2, Global temperature, ice, Sea level) :</p> <p><a href="https://climate.nasa.gov/vital-signs/carbon-dioxide/">https://climate.nasa.gov/vital-signs/carbon-dioxide/</a></p> <p>Autre site pour le niveau des mers :</p> <p><a href="https://research.csiro.au/slrwavescoast/sea-level/">https://research.csiro.au/slrwavescoast/sea-level/</a></p> <p>Altimétrie :</p> <p><a href="http://www.aviso.altimetry.fr/fr/donnees.html">http://www.aviso.altimetry.fr/fr/donnees.html</a></p> <p>Informations bibliographie sur des thèmes environnementaux :</p> <p><a href="http://side.developpement-durable.gouv.fr/PDLO/doc/SYRACUSE/215165">http://side.developpement-durable.gouv.fr/PDLO/doc/SYRACUSE/215165</a></p> <p>Le climat de la France :</p> <p><a href="https://www.ecologie.gouv.fr/sites/default/files/ONERC_Climat_France_XXI_Volume_4_VF.pdf">https://www.ecologie.gouv.fr/sites/default/files/ONERC_Climat_France_XXI_Volume_4_VF.pdf</a></p> <p>Le site de l'Observatoire national sur les effets du réchauffement climatique :</p>

<https://www.ecologie.gouv.fr/observatoire-national-sur-effets-du-rechauffement-climatique-onerc>

Groupe de l'ONU travaillant sur le changement climatique :  
<https://www.ipcc.ch/>

Agence de protection de l'environnement aux Etats-Unis :  
<https://www.epa.gov/>

Données ouvertes de la banque mondiale :  
<https://donnees.banquemondiale.org/>

National academies :  
<https://www.nationalacademies.org/>

Données de la NASA :  
<https://mydasdata.larc.nasa.gov/global-climate-change>

Données du GISS (NASA) :  
<https://data.giss.nasa.gov/>

Spécialistes du changement climatique à la NOAA (agence météo américaine) :  
<https://www.climate.gov/>

The globe program de la NOAA :  
<https://www.globe.gov/globe-data/visualize-and-retrieve-data>

Réchauffement et salinité des mers :  
[https://www.nodc.noaa.gov/OC5/3M\\_HEAT\\_CONTENT/](https://www.nodc.noaa.gov/OC5/3M_HEAT_CONTENT/)

Données brutes NOAA  
[https://data.nodc.noaa.gov/woa/DATA\\_ANALYSIS/DATA/](https://data.nodc.noaa.gov/woa/DATA_ANALYSIS/DATA/)

Changement climatique Canada :  
<https://www.canada.ca/fr/environnement-changement-climatique.html>

Données sur les catastrophes dans le monde (météorologiques, géologiques, industrielles et climatiques, s'inscrire et c'est gratuit) :  
<https://public.emdat.be/>

Données du programme Copernicus :  
<https://data.marine.copernicus.eu/products>

Global carbon Atlas, pour des données d'émission de GES par origine et par pays :  
<https://globalcarbonatlas.org/fr/>

Données Eurostat d'émission de GES :  
<https://ec.europa.eu/eurostat/fr/web/environment/database>

Données d'émission de GES statista :  
<https://fr.statista.com/statistiques/955310/repartition-geographique-des-emissions-de-co2-monde-par-pays/>

Autres données d'émission de GES :  
<https://www.polytechnique-insights.com/tribunes/planete/comprendre-les-emissions-mondiales-de-gaz-a-effet-de-serre/>

Données du ShiftProject :  
<https://theshiftproject.org/portail-de-donnees/>

Fonds de carte pour la France :  
<https://www.insee.fr/fr/information/2115000>

Fonds de carte monde et autres :  
<http://eductice.ens-lyon.fr/EducTice/recherche/geomatique/veille/sig/Jeux-de-donnees/>

Il faut traiter au moins 10 indicateurs parmi les suivants :

- \*Température moyenne au-dessus de la Terre
- \*Température moyenne au-dessus des océans
- \*Élévation moyenne et par océan/mer du niveau des océans en mm/an
- \*Nombre et intensité des ouragans
- \*Surface de la banquise au Pôle Nord en Septembre
- \*Épaisseur de la banquise au Pôle Nord en Septembre
- \*Quantité de CO2 émis depuis la date de début de l'étude au niveau mondial
- \*Quantité de CO2 émis depuis la date de début de l'étude au niveau de chaque pays

	<p>*Quantité de CO2 équivalent émis depuis la date de début de l'étude (le CO2 équivalent représente l'agrégation des effets de toutes les molécules de gaz à effet de serre, y compris le CO2, en tenant compte des caractéristiques de chaque molécule) au niveau mondial</p> <p>*Quantité de CO2 équivalent émis depuis la date de début de l'étude au niveau de chaque pays</p> <p>*PIB de chaque pays</p> <p>*Population de chaque pays</p> <p>*Indice de Gini de chaque pays</p> <p>*Taux de mortalité infantile</p>
<b>Missions</b>	<ol style="list-style-type: none"> <li>1. Revue documentaire : quelle problématique, quelles données, pour quelles analyses ?</li> <li>2. Rassemblement des données sous forme de fichiers exploitables.</li> <li>3. Révision éventuelle des objectifs à atteindre en fonction des données exploitables.</li> <li>4. Analyses statistiques et interprétation. Ces analyses comporteront une dimension évolution temporelle et une dimension cartographique.</li> <li>5. Bâtir une dataviz Rshiny pour exposer un choix des résultats. Eventuellement l'application pourra se mettre à jour avec les données de flux.</li> <li>6. Synthèse et conclusion.</li> </ol>
<b>Accompagnement</b>	Cette SAE est encadrée à hauteur de 30.5 heures dont 4 sont réservées aux oraux d'évaluation. Elle prévoit 23.5h d'autonomie non encadrée.
<b>Livrables</b>	<p>Livrable 1 – Rapport 1 exposant la problématique, les sites de données repérés, la description de ces données, avec quelques statistiques et représentations graphiques pour fixer les idées, et les méthodes envisagées, au plus tard le 28 janvier 2024</p> <p>Livrable 2 – Rapport 2 avec au moins une analyse temporelle et une analyse multivariée (ACP, AFCM, Classification), à remettre au plus tard le 11 février 2024</p> <p>Livrable 3 – Rapport 3 final à remettre au plus tard le 22 mars 2024</p> <p>Livrable 4 – Diaporama et Dataviz présentés pour l'oral au plus tard le 22 mars 2024</p> <p>Livrable 5 – Une page individuelle par étudiant et non par groupe, où chaque étudiant précisera son apport au travail et à la rédaction des rapports, à déposer au plus tard le 24 mars 2024</p> <p>Livrable 6 – Le badge, à déposer au plus tard le 24 mars 2024.</p>
<b>Planning (sous réserve de modifications dues à des impondérables)</b>	<ul style="list-style-type: none"> <li>- Semaine du 15 janvier 2024, 6h encadrées</li> <li>- Semaine du 22 janvier 2024, 8h dont 2 en autonomie</li> <li>- Semaine du 29 janvier 2024, 6h dont 2 en autonomie</li> <li>- Semaine du 5 février 2024, 6h dont 2 en autonomie</li> <li>- Semaine du 26 février 2024 (sans apprentis), 6h en autonomie</li> <li>- Semaine du 4 mars 2024 (sans apprentis), 6h en autonomie</li> <li>- Semaine du 11 mars 2024, 6h dont 1.5 en autonomie</li> <li>- Semaine du 18 mars 2024, 6h dont 2h pour les oraux d'évaluation le 22 mars 14h-16h, présentation de 10 minutes, questions 6 minutes.</li> </ul>

## EVALUATION

<b>Par groupe</b>	<p>Rapport 1 : coefficient 1</p> <p>Rapport 2 : coefficient 1</p> <p>Dataviz : coefficient 3</p> <p>Rapport final : coefficient 5</p>
<b>Individuelle</b>	<p>Participation (note évaluant l'implication de chacun dans tous les travaux) : coefficient 1</p> <p>Page individuelle et badge : coefficient 1</p> <p>Oral et diaporama : coefficient 3</p>