

## Description de projet

### Hiver 2022

Le projet exige la prévision de la demande d'électricité dans une région gérée par le *Electric Reliability Council of Texas* (ERCOT), une organisation de transmission régionale qui dessert une grande partie de l'état du Texas.

On peut trouver des informations à propos d'ERCOT au <https://ercot.com/>.

Les données de 8 régions ont été assemblées dans l'objet R `ercotdata`. Les données pour la région entière ERCOT sont aussi incluses.

**Les données ont été sauvegardées dans `ercotdata.RData` disponible dans Zone Cours.**

L'objet `ercotdata` est de classe R `timeSeries`.

Après avoir chargé l'objet dans R, voir `names(ercotdata)` et la carte *weather zone* au <https://www.ercot.com/news/mediakit/maps/> pour identifier les 8 régions qu'il inclut.<sup>1</sup>

**Chaque équipe sera attribué une région unique.**

**Le point de départ pour toutes les équipes est les données horaires de la demande fournies dans le `ercotdata`.**

**Le code R des étudiants devra lire le fichier `ercotdata.RData`.**

Toute manipulation des données doit être faite dans le code R soumis par les étudiants. Cette dernière façon de procéder *laisse les traces* nécessaires à la réplication.

Les données doivent être traitées à un niveau **journalier**.

L'objectif est la prévision de la **somme de la demande horaire pour jour  $t + 1$** .

La prévision est à court-terme, i.e. **pour  $h = 1$  à partir de données disponibles au jour  $t$ , et ceci pour chaque jour de l'année.**

Au moins un *bon* candidat parmi les méthodes naïves, de lissage, de régression linéaire et des séries chronologiques doit être inclus dans les comparaisons. Les méthodes nonparamétriques seront seulement présentées à la 11e séance et ne seront pas utilisées dans le projet.

Des données historiques de la température et d'autres variables explicatives potentielles sont disponibles (sans frais). Voir e.g. <https://www.ncdc.noaa.gov/cdo-web/search>.

Des prévisions faites par ERCOT ou d'autres ne peuvent pas servir d'entrants aux modèles. Les modèles de prévision des étudiants doivent utiliser des données brutes et gratuites.

La demande horaire est disponible du 1er janvier 2012 au 31 décembre 2021. Il n'est pas nécessaire d'utiliser toutes les données disponibles. Chaque équipe doit bien argumenter ses choix selon les conditions dans sa région.

Les données sont fournies en GMT, mais les *jours* doivent être définis en heure locale. Une manipulation du temps sera requise avant que la somme de la demande horaire puisse être calculée et extraite pour chaque jour.

---

<sup>1</sup>ERCOT fournit la demande horaire dans les régions du bassin en format `.xlsx` ou `.xls`, un fichier par année. Le format des fichiers et les étiquettes attribuées ont changé au cours des années. Extraire les données pose plusieurs défis. Habituellement, le projet exige que les données soient extraites directement du site de l'organisation. Exceptionnellement, les données sont fournies comme un objet R afin qu'une plus grande proportion du temps des étudiants soit dédiée aux autres composantes du projet.

La **Partie I** du rapport écrit doit être soumise **au plus tard à midi le 10 février 2022**.  
Le rapport couvre :

- une introduction
- une analyse exploratoire des données
- une évaluation des méthodes naïves
- une description des variables explicatives possibles en indiquant clairement la source des données et fournissant des statistiques sommaires

Inclus dans la soumission :

- la Partie I du rapport, en format `.pdf` :
  - avec UN MAXIMUM DE 12 PAGES, incluant tous les tableaux, figures et annexes;
  - utilisant une police Times New Roman, 12 pt, double interligne, format de page “lettre” avec des marges d’un minimum de 1 pouce.

**3 points (sur 30) sont automatiquement enlevés si les consignes ci-dessus ne sont pas respectées.**

- les fichiers `.csv` des données historiques pour les variables explicatives;
- tous les codes R, en format ASCII comme fichiers `.R`, nécessaires pour reproduire les résultats dans le rapport. Voir les exigences pour le code R à la page 3.

La **Partie II** du rapport écrit doit être soumise **au plus tard à midi le 17 mars 2022**.  
Le rapport couvre :

- l’évaluation des méthodes de lissage
- l’évaluation des méthodes de régression

Inclus dans la soumission :

- la Partie II (seulement) du rapport, en format `.pdf` :
  - avec UN MAXIMUM DE 10 PAGES, incluant tous les tableaux, figures et annexes;
  - utilisant une police Times New Roman, 12 pt, double interligne, format de page “lettre” avec des marges d’un minimum de 1 pouce.

**4 points (sur 35) sont automatiquement enlevés si les consignes ci-dessus ne sont pas respectées.**

- tous les codes R, en format ASCII comme fichiers `.R`, nécessaires pour reproduire les résultats dans le rapport. Voir les exigences pour le code R à la page 3.

La **Partie III** du rapport écrit doit être soumise **au plus tard à midi le 14 avril 2022**.  
Le rapport couvre :

- l'évaluation des méthodes ARIMA/SARIMA
- des comparaisons globales
- une recommandation pour les utilisateurs
- une conclusion

Inclus dans la soumission :

- la Partie III (seulement) du rapport, en format `.pdf` :
  - avec UN MAXIMUM DE 7 PAGES, incluant tous les tableaux, figures et annexes;
  - utilisant une police Times New Roman, 12 pt, double interligne, format de page “lettre” avec des marges d’un minimum de 1 pouce.

**4 points (sur 35) sont automatiquement enlevés si les consignes ci-dessus ne sont pas respectées.**

- tous les codes R, en format ASCII comme fichiers `.R`, nécessaires pour reproduire les résultats dans le rapport. Voir les exigences pour le code R ci-dessous.

=====

### IMPORTANT

Le code R soumis par les étudiants devra :

- (1) commencer avec la lecture du fichier `ercotdata.RData`;
- (2) contenir au plus 70 caractères par ligne;
- (3) avoir des fonctions proprement définies. Tous calculs exécutés dans les fonctions doivent seulement nécessiter des objets qui sont passés dans les arguments de la fonction.

Si le code soumis ne satisfait pas ces trois exigences,

- il ne sera pas *debuggé* afin de déterminer la nature (mineure/majeure, compréhension, codage, mauvais calculs, etc) des erreurs dans les solutions fournies;
- toute erreur dans les solutions fournies sera supposée le résultat d’une faute de compréhension grave et sera pénalisée comme telle.

Seulement les *packages* R utilisés dans les codes R fournis par l’enseignante sont permis.

Il est strictement interdit d’utiliser tout autre *package* R dans le projet de semestre.

=====

La partie écrite du projet est évaluée sur 100 points selon la distribution suivante :

Partie I: 30 points

Partie II: 35 points

Partie III: 35 points

La partie écrite du projet représente 25% de la note finale dans le cours.

=====

The partie présentation du projet représente 5% de la note finale dans le cours.

Les étudiants feront la présentation de leur projet le 12 avril, 2022.

Des informations supplémentaires concernant les exigences de la présentation ainsi que la grille de correction seront fournies dans un autre document plus tard.

=====

Bon travail et bonne chance!