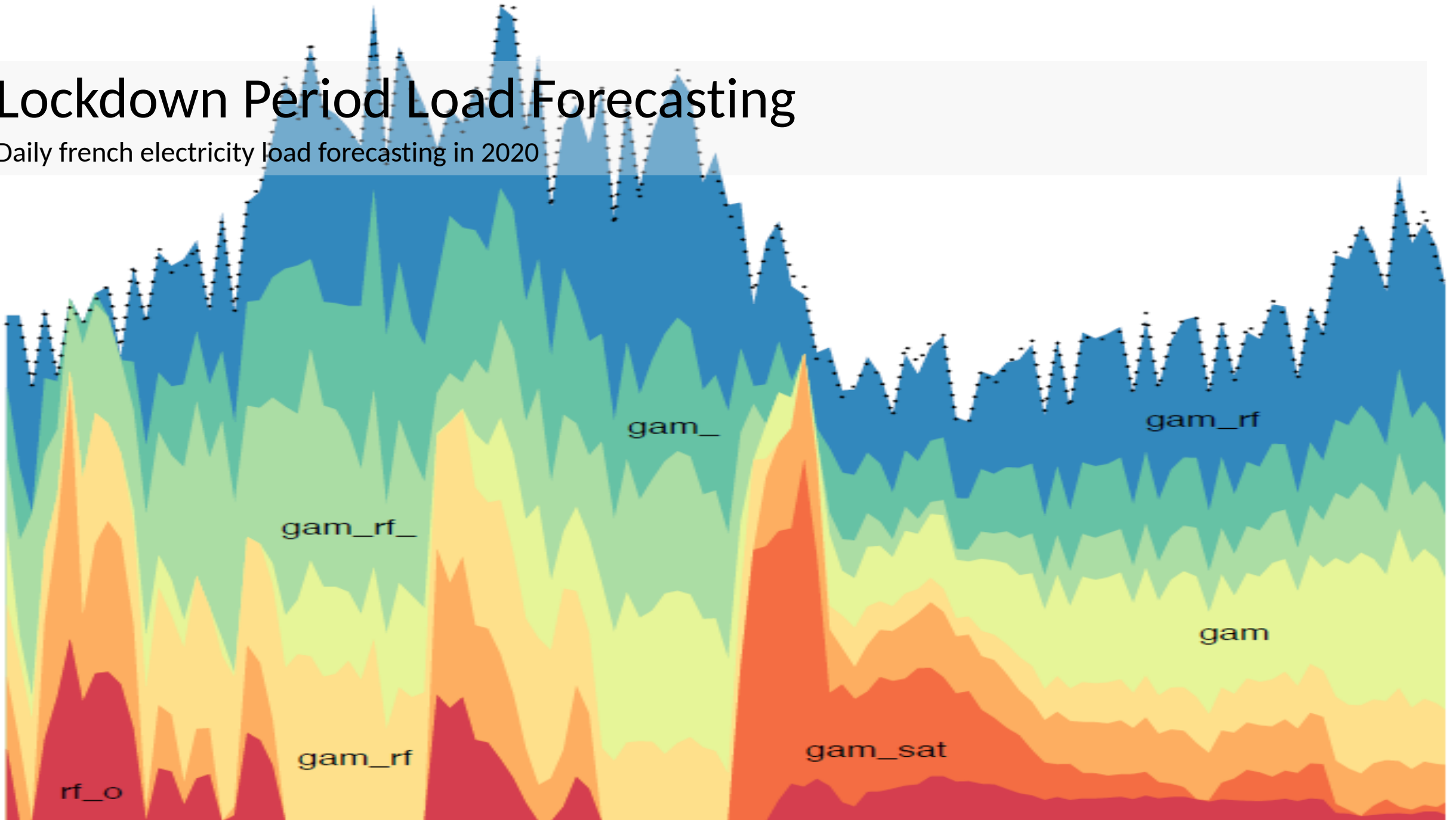


Lockdown Period Load Forecasting

Daily french electricity load forecasting in 2020



Objectifs du cours

- Effectuer une analyse descriptive des données
- Enrichir les données proposée par de l'open data: identifier les données pertinentes, nettoyer et fusionner les données
- Développer un modèle de prévision
 - Choisir un modèle
 - Sélectionner/construire les variables pertinentes
 - Valider le modèle: simulation, bootstrap, public score...
- Obtenir le meilleur score (plus faible erreur de prévision) sur le private set, en respectant les contraintes imposées
- Présenter ses résultats de manière intelligible et reproductible

Organisation

- Travail en groupes de 2 étudiants
- 2 soumissions par groupe autorisées par jour
- Séances de travaux qui couvriront:
 - Manipulation/visualisation/fusion de données avec R
 - Faire son package R/travail avec git
 - Modèles de prévision: régression linéaire, modèles GAM, série temporelle, ML
 - Validation de modèle

Attendus & Notation

- Un rapport final en Rmarkdown /10 (max 10 pages)
- Une présentation orale /9
- Obtenir le meilleur score (plus faible erreur de prévision) sur le private set /1

Critères de notation:

- qualité de présentation
- justification scientifique des choix de modèles, des transformations des données
- qualité du code R (idéalement un package R permettant de mettre en œuvre votre modèle)

Soutenances:

- 15 min de présentation, répartition équitable du temps par personne
- 5 min de questions
- l'ordre de passage est ici:
 - <https://docs.google.com/spreadsheets/d/1Jiaazb6qgpvo-RLi3d7rn36HImD3YT5rxWf2wUKP-ak/edit?usp=sharing>

<https://www.kaggle.com/t/37134625938e4af6b2583e09f3d975d7>

<https://www.math.u-psud.fr/~goude/>

Nom des équipes:

<https://docs.google.com/spreadsheets/d/1vylwyzFZoXUkFUv49ApTfk93UBGpTQAPpdmHo8c0VzM/edit?usp=sharing>