

Projet de fin de semestre

Planifier une tâche de calcul coûteuse en fonction de critères multiples dont environnementaux (en binôme)

Contexte : sensibilisation à la responsabilité numérique, calcul intensif, données ouvertes

Objectifs pédagogiques : travail en groupe, modélisation de données à l'aide des structures vues en cours, découpage fonctionnel, prise en compte de formats de fichiers.

Les centres de calcul intensif (simulation, optimisation, apprentissage...) ont une consommation électrique considérable. Leur puissance se situe généralement entre 1 et 15MW en Europe en 2021, 30MW pour le plus important référencé dans le monde¹.

L'impact environnemental de cette consommation dépend des moyens de production d'électricité utilisés, lesquels varient selon les régions et au fil du temps.

L'objectif du projet est de concevoir et de programmer un logiciel proposant plusieurs scénarios d'affectation pour une tâche de calcul intensif en respectant des critères environnementaux.

I Contexte : production électrique en France

1 Variabilité de la production électrique et des ses impacts

Du point de vue environnemental, nous considérerons la production de gaz à effet de serre (*exprimée en grammes équivalent CO₂ : eqCO₂*) comme l'impact principal, qui est aussi le mieux documenté, même si l'impact sur les ressources ou la bio-diversité sont non négligeables.

Le mix² électrique français émet en moyenne 55g eqCO₂/kWh, mais cette moyenne cache une variabilité spatiale et temporelle dépendant de la disponibilité des moyens de production. Il faut donc en tenir compte pour planifier la tâche de calcul. La figure 1 ci-contre illustre cette variation pour l'été 2022 en France métropolitaine.

Par ailleurs, la production et la consommation devant être équilibrées à tout moment sur le réseau, il existe des flux entre les régions et les pays. Lorsque le solde des flux est positif, les moyens de production sur le territoire ne suffisent pas à satisfaire les besoins. Il peut alors être judicieux de renoncer à exécuter une tâche non prioritaire. Dans le cas de gros consommateurs, cette baisse de consommation volontaire peut être contractualisée et est appelée effacement (par exemple, le CERN³ prévoit de réduire de 25 % sa puissance cet hiver).

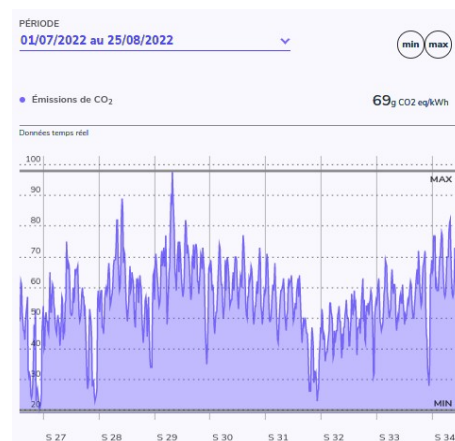


fig.1 : variation des émissions de CO₂ de la production électrique - France métropolitaine, été 2022
(source : rte-france.com/eco2mix).

2 Données d'évolution de la production

Les données temporelles montrant l'évolution de l'utilisation des moyens de production par région sont fournies par RTE, l'opérateur du Réseau de Transport de l'Électricité en France. RTE fournit à la fois des données historiques, des données temps réel et des prévisions de production. Des exemples de visualisation des données en temps réel (plus détaillées que celles que nous allons utiliser) peuvent être trouvées aux adresses suivantes :

<https://app.electricitymap.org/zone/FR>

<https://www.rte-france.com/eco2mix/les-donnees-regionales#consommation-et-production>

Les données prévisionnelles n'étant pas accessibles avec le niveau de détail qui nous intéresse, nous utiliserons à la place des données réelles historiques, celles concernant l'année 2021, qui ont fait l'objet d'un pré-traitement par les enseignants pour faciliter leur lecture. Elles sont téléchargeables sur madoc, dans la section "mini-projet".

1 <https://www.top500.org/lists/top500/2022/06/> (les données ne tiennent pas forcément tous compte des systèmes auxiliaires, notamment une partie du refroidissement, qui peut demander le double de la puissance allouée au calcul).

2 Le mix électrique désigne la répartition de la production électrique selon les moyens de production.

3 CERN : Conseil Européen pour la Recherche Nucléaire

Le fichier détaille la production, en MWh⁴, pour une heure donnée et pour une région, selon le type de système de production, (voir figure 2).

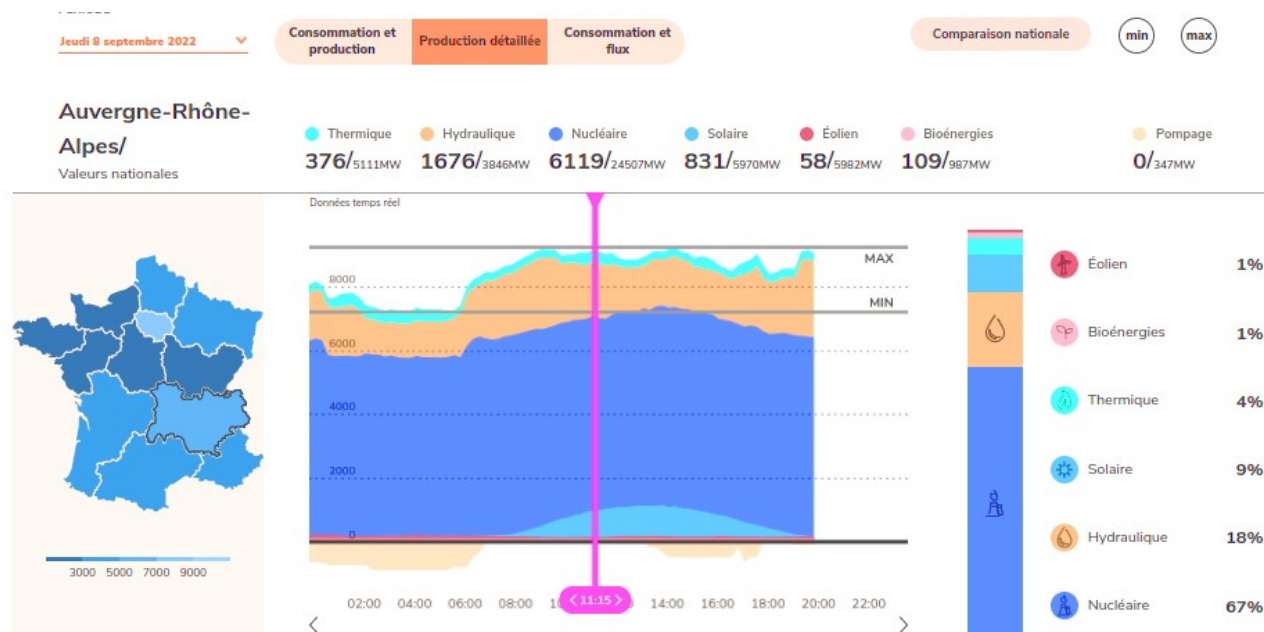


fig.2 : Répartition de la production le 08/09/2022 à 11h15 pour la région Auvergne Rhône Alpes (source : eco2mix)

Chaque ligne présente les données suivantes, pour une durée d'une heure :

- code de la région
- horodatage : mois, jour, heure
- production thermique
- production nucléaire
- production éolienne
- production solaire
- production hydraulique
- production de bioénergies
- solde des flux sortants et entrants : il est négatif en cas d'exportation, positif en cas d'importation.

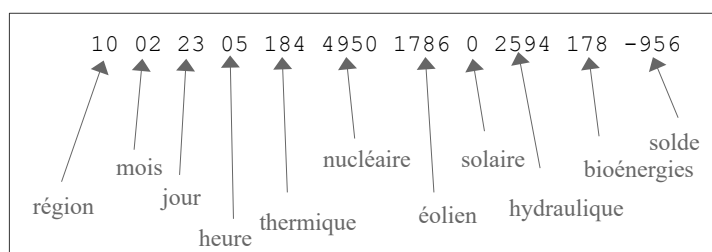


fig.3 : une ligne du fichier energieFrance2021.txt

3 Estimation des émissions en g eqCO₂/kWh

Il existe deux moyens principaux de compter les émissions : en se basant sur le coût à la production (c'est la méthode retenue par RTE pour le suivi sur son site) ou en analyse en cycle de vie (ACV), qui intègre en plus les coûts de construction, de maintenance et de démantèlement ainsi que ceux d'extraction et de transport du minerai, puis les met en rapport avec la production durant la durée de vie estimée du moyen de production. La différence entre les deux types de méthode est surtout notable pour les moyens bas carbone, dont la production n'engendre quasiment pas d'émissions directes.

Les valeurs communément utilisées sont les médianes des études retenues par le GIEC⁵ ; les études sur le parc français donnent des résultats plus bas pour certains moyens de production, mais ne changent pas les ordres de grandeur.

Le fichier de production fourni par RTE que nous utilisons regroupe sous une seule rubrique "Thermique" la production charbon, fioul et gaz (respectivement 820, 777 et 490g eqCO₂/kWh). En se fondant sur la répartition annuelle de la production de chacun de ces moyens, les émissions thermiques sont estimées à environ 560g eqCO₂/kWh.

Afin de pouvoir mettre à jour facilement des valeurs pour varier les hypothèses, les coûts respectivement du thermique, du nucléaire, de l'éolien, du solaire, de l'hydraulique et des bioénergies sont donnés dans un fichier texte nommé couts_emissions.txt (figure 4). Il s'agit à chaque fois du coût d'un gramme équivalent CO₂ par Kilowatt heure : g eqCO₂/kWh,

560 12 11 27 24 230

fig. 4 : couts_emissions.txt

4 1 MWh représente environ 5 mois de consommation électrique d'un français en moyenne. (159TWh Résidentiel annuel https://www.cre.fr/content/download/24881/file/Observatoire_detail_T3_2021.pdf pour 67,8 millions d'hab.)

5 Valeurs médianes IPCC 2014 https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_annex-iii.pdf
GIEC : Groupe d'experts intergouvernemental sur l'évolution du climat

4 Nom des régions

Le fichier texte `regions.txt` donne la correspondance entre le code des douze régions et leur nom.

II Objectif

Aider à planifier une tâche décrite dans un fichier texte. Ce fichier énonce :

- nom de la tâche
- nom du fichier de résultats
- durée de la tâche (en heures)
- horodatage début période (mois, jour, heure)
- horodatage fin période (mois, jour, heure)
- coût maximum moyen de chaque heure en g eq CO₂

```
grokalkul
sortie_grokalkul.txt
50
5 3 0
5 10 14
21.5
2 11 12
```

fig.5 : exemple de fichier tâche

Le coût moyen d'une heure est la moyenne des coûts de production de l'énergie

Par exemple, pour la ligne de la figure 3, le coût moyen est de

$$[(184 * 560) + (4950 * 12) + (1786 * 11) + (0 * 27) + (2594 * 24) + (178 * 230)] / (184 + 4950 + 1786 + 0 + 2594 + 178) \\ = 29,4 \text{ g eq CO}_2$$

- régions où peut être réalisée la tâche

La tâche doit se dérouler dans **une seule région**, de manière **continue**, dans la **période** énoncée.

Pour chacune des heures du déroulement de la tâche, la région doit être **productrice d'énergie** (solde négatif) et le **coût moyen** doit être inférieur au coût maximum moyen.

Votre programme lira les données du fichier `energieFrance2021.txt`, du fichier `couts_emissions.txt` et d'un fichier de description d'une tâche et énoncera, dans un fichier de résultats, la liste des planifications possibles, cette liste étant ordonnée selon le coût croissant.

Une planification est définie par

- la région où le calcul est planifié
- l'horodatage (mois, jour, heure) de la première heure de calcul
- l'horodatage de la dernière heure de calcul
- le coût du calcul (somme des coûts moyens de chaque heure)

2 2 11	5 7 9 5 7 10 5 3 15	5 9 10 5 9 11 5 5 16	812,84 820,39 884,92
région	première heure	dernière heure	coût (kg eq CO ₂)

fig.6 : extrait de `sortie_grokalkul.txt`

La figure 6 est un extrait du fichier de résultats `sortie_grokalkul.txt`.

La tâche peut être planifiée

- dans la région 2, du 7 mai 9h au 9 mai 10h, pour un coût de 812,84
- dans la région 2, du 7 mai 10h au 9 mai 11h, pour un coût de 820,39
- dans la région 11, du 3 mai 15h au 5 mai 16h, pour un coût de 884,92

Le résultat peut être visualisé à l'aide du logiciel `resultat_vers_SVG` (téléchargeable sur madoc) utilisable avec la ligne de commande :

```
./resultat_vers_SVG -i result.txt -o result.svg
```

où `result.txt` est un fichier de résultats respectant le format de la figure 6

et `result.svg` est un fichier visualisable.

Ce logiciel est utilisable sous Linux, dans les salles de TP.

III Résultats attendu (dépôt sur madoc)

– **un ou plusieurs fichiers sources** C++⁶ constituant votre logiciel⁷.

Les fichiers sources seront documentés comme vus en tp : auteurs, description de contenu ; pour chaque fonction ou procédure : rôle, précondition, jeu de tests le cas échéant, etc.

Un programme principal doit montrer que votre logiciel fonctionne bien.

Vous êtes encouragé à fournir au moins un sous-algorithme fonctionnant de manière récursive.

Si le projet n'est pas complet, le programme principal utilisera les fonctionnalités que vous avez réussi à développer afin d'en montrer le bon fonctionnement.

– au moins **trois fichiers de tâches** dans lequel figurent des données avec lesquelles votre logiciel fonctionne.

– **un compte-rendu** d'environ deux pages (environ 1000 mots) dans lequel vous indiquerez les noms des auteurs.

Vous y expliquerez les structures de données que vous avez définies, un ou deux algorithmes qui vous ont paru intéressants, les difficultés que vous avez rencontrées et que vous avez su (ou pas) résoudre, les limites de votre logiciel.

Le compte-rendu doit permettre d'évaluer dans quelle mesure vous avez su réaliser le logiciel demandé.

Le compte-rendu doit être bien écrit, agréable à lire, être composé de parties, commencer par une introduction et terminer par une conclusion.

Il n'y a ni logo, ni sauts de page.

6. Date limite

SAMEDI 24 DÉCEMBRE 2022, 20H

Remarques

Ce projet, bien que basé sur des données et préoccupations réelles, ne doit pas être considéré comme une solution réaliste du fait des nombreuses simplifications et contraintes pédagogiques qu'il intègre.

À titre d'exemple de simplifications : la capacité, la disponibilité et l'efficacité des centres de calculs ne sont pas prises en compte bien qu'elles varient beaucoup : <https://www.top500.org/lists/green500/2022/06/>). Pour ce projet, considérer que tous les centres ont une puissance de 1 MW.

Du point de vue du coût environnemental du numérique, le cas du calcul intensif est une exception du fait de la part importante que représente la consommation électrique ; pour la grande majorité des autres usages de l'informatique, le matériel représente au moins les trois quarts du coût, suivi par l'infrastructure, devant la consommation (en particulier sur un mix bas carbone comme celui de la France). L'optimisation prioritaire est alors l'allongement de la durée de vie des matériels et le bon dimensionnement des infrastructures.

⁶ .cpp

⁷ **Ne pas** rendre de fichier exécutable (.exe).