

## Projet de fin de semestre

### Planifier une tâche de calcul coûteuse mais découposable en fonction de critères multiples dont environnementaux (en quadrinôme)

**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 monde<sup>1</sup>.

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 des scénarios d'affectation d'une tâche de calcul intensif sur différents serveurs en respectant des critères environnementaux.**

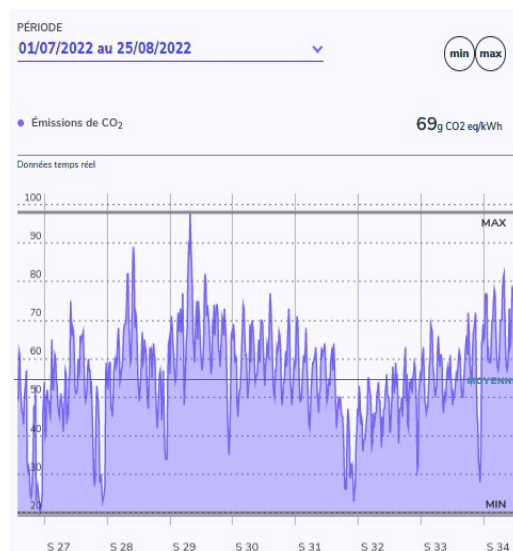
## I CONTEXTE DU SUJET

### 1 VARIABILITÉ DE LA PRODUCTION ÉLECTRIQUE ET DE SES IMPACTS

Du point de vue environnemental, nous considérerons la production de gaz à effet de serre (*exprimée en grammes équivalent CO<sub>2</sub> par kWh : g eqCO<sub>2</sub>/kWh*) comme l'impact principal, qui est aussi le mieux documenté, même si ceux sur les ressources et la bio-diversité sont non négligeables.

Le mix<sup>2</sup> électrique français émet en moyenne 55g eqCO<sub>2</sub>/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 ci-contre illustre cette variation pour l'été 2022 sur la France métropolitaine (source : [eco2mix](https://eco2mix.org/)).

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 CERN prévoit de réduire de 25 % sa puissance cet hiver*).



### 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>

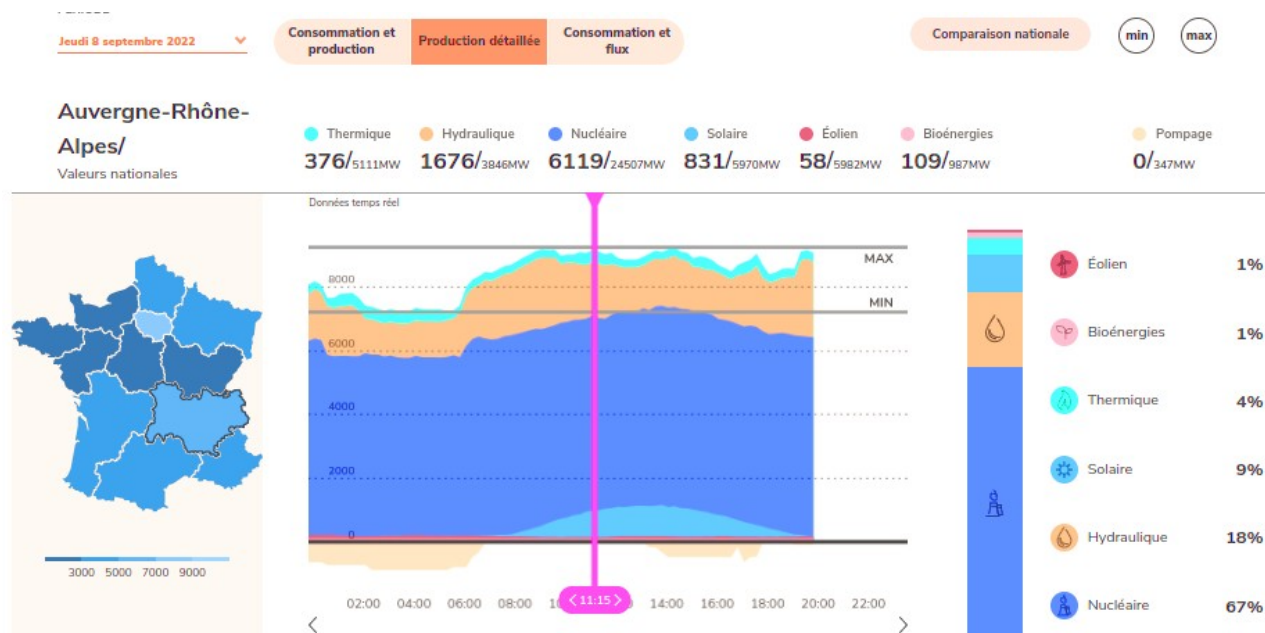
<https://www.rte-france.com/eco2mix/les-donnees-regionales#consommation-et-production>

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 rubrique projet.

1 <https://www.top500.org/lists/top500/2022/06/> (attention, les chiffres 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.

Le fichier détaille la production en MWh<sup>3</sup> pour une heure donnée et pour une région, selon le type de système de production, comme illustré dans la figure suivante.



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

### 3 ESTIMATION DES ÉMISSIONS EN g eqCO<sub>2</sub>/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 GIEC<sup>4</sup> ; 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 eqCO<sub>2</sub>/kWh*). En se basant sur la répartition annuelle de la production de chacun de ces moyens, on peut estimer à environ 560g eqCO<sub>2</sub>/kWh les émissions pour ce regroupement.

Afin de pouvoir mettre à jour facilement ces valeurs pour varier les hypothèses, elles seront données dans un fichier texte nommé `couts_production.txt` représenté ci-contre.

Émissions en g eqCO<sub>2</sub>/kWh

Thermique	560
Bioénergies	230
Solaire	27
Hydraulique	24
Nucléaire	12
Éolien	11

#### 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 (cela varie beaucoup en réalité : <https://www.top500.org/lists/green500/2022/06/>). On considérera que tous les centres ont une puissance de 1 MW (= 1000kW).

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.

3 1 MWh représente environ 5 mois de consommation électrique d'un français en moyenne. (159TWh Résidentiel annuel pour 67,8 millions d'habitants, cf. [https://www.cre.fr/content/download/24881/file/Observatoire\\_detail\\_T3\\_2021.pdf](https://www.cre.fr/content/download/24881/file/Observatoire_detail_T3_2021.pdf))

4 Valeurs médianes IPCC 2014 [https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc\\_wg3\\_ar5\\_annex-iii.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_annex-iii.pdf)

## II RÉALISATION

### 1 RÉSULTAT ATTENDU

#### 1.1 Un rapport d'analyse (à rendre dès la deuxième semaine) indiquant :

- les structures de données utilisées ainsi que le découpage fonctionnel
- la répartition des tâches entre les membres du groupe

#### 1.2 Un programme C++, recevant comme paramètres en ligne de commande :

- le fichier de données de production, le fichier des régions et le fichier des coûts,
- un fichier de description de la tâche (au format donné page suivante),

afin de générer une liste de scénarios d'exécution conformes aux critères prévus pour la tâche.

Un scénario reprend les informations identifiant la tâche puis une liste des plages horaires d'exécution (possiblement discontinues) ainsi que leur localisation et leur coût, puis le coût total et le temps écoulé entre le début et la fin de la tâche.

#### 1.3 Un rapport final, de 2-3 pages, précisant :

- l'évolution des structures et sous-programmes vis à vis de l'analyse rendue préalablement ;
- les limitations du programme si l'objectif du projet n'a pas été intégralement atteint, les difficultés rencontrées ;
- le mode d'emploi.

#### 1.4 Des jeux de tests pertinents et suffisants, sous forme de fichiers texte d'entrées (cf. Résultat 1.2) et d'un fichier texte README expliquant l'intérêt et la façon d'utiliser ces différents fichiers d'entrées.

#### 1.5 Une démonstration durant la dernière séance de TP, prise en compte dans la notation individuelle.

### 2 FORMAT DU FICHIER DE DONNÉES DE PRODUCTION :

Code région	Date	Heure	Thermique	Nucléaire	Eolien	Solaire	Hydraulique	Bioénergies	Échanges physiques
1	01-01	00:00:00	836	0	10	0	6	238	18738
2	01-01	00:00:00	210	18038	54	0	24	130	-12654
3	01-01	00:00:00	250	0	88	0	150	78	4774
4	01-01	00:00:00	614	15502	64	0	28	98	-7654
5	01-01	00:00:00	2292	9212	784	0	2	240	152
6	01-01	00:00:00	2174	18436	402	0	2000	248	-12616

Chaque ligne du fichier de production fourni sur Madoc donne la production en MWh pour une heure donnée et pour une région, selon le type de système de production. La colonne échanges physiques indique le solde des flux sortants et entrants sur la période : le solde est négatif en cas d'exportation, positif en cas d'importation. Le fichier est trié par heure puis par régions, sans trous ni doublons. Un fichier texte regions.txt disponible sur Madoc donne la correspondance entre le code des 12 régions et leur nom.

### 3 DESCRIPTION DES TÂCHES ET DES CONTRAINTES

Les tâches de calcul devront être décrites dans un fichier contenant les informations suivantes :

	Nom de l'information	Exemple de valeur
A	identifiant de la tâche	145
B	nom de la tâche	Simulation#23_modèle_de_croissance_forêt
C	durée estimée en heures <sup>5</sup>	8
D	date et horaire de départ minimum	01/01 0:00
E	date et horaire de terminaison maximale	01/06 0:00
F	coût moyen maximum du Mix toléré	300
G	coût marginal maximum du Mix toléré	400
H	coût du moyen de production significativement actif le plus polluant	
H	pourcentage de production marginale minimum pour considérer la contrainte précédente	4
	car un minimum de thermique peut être nécessaire pour la stabilisation du réseau sans pour autant être le moyen appelé pour satisfaire une consommation additionnelle	
I	pourcentage d'importation régional toléré	20
J	pourcentage d'importation national toléré	5
	au delà duquel on considère qu'il y a un pic de consommation et on reporte la tâche	
K	régions de calcul possibles	1 2 3 4 5 6
	liste de codes de région	

<sup>5</sup> les centres de calculs étant supposés de performances équivalentes (et d'une puissance de 1MW pour simplifier), la durée peut-être estimée en heure, plutôt qu'en exaflops qui seraient divisés par la performance de calcul (Rmax) du centre de calcul.

#### 4 EXEMPLE

Code région	Date	Heure	Thermique	Nucléaire	Eolien	Solaire	Hydraulique	Bioénergies	Échanges physiques	production	coût Moyen	Coût marginal sup à 4%
1	2021-01-01	00:00:00	836	0	10	0	6	238	18 738	1 090	479,96	560
			76,70 %	0,00 %	0,92 %	0,00 %	0,55 %	21,83 %	94,50 %			
2	2021-01-01	00:00:00	210	18 038	54	0	24	130	-12 654	18 456	19,78	12
			1,14 %	97,74 %	0,29 %	0,00 %	0,13 %	0,70 %	-68,56 %			
3	2021-01-01	00:00:00	250	0	88	0	150	78	4 774	566	287,12	560
			44,17 %	0,00 %	15,55 %	0,00 %	26,50 %	13,78 %	89,40 %			
4	2021-01-01	00:00:00	614	15 502	64	0	28	98	-7 654	16 306	33,96	12
			3,77 %	95,07 %	0,39 %	0,00 %	0,17 %	0,60 %	-46,94 %			
5	2021-01-01	00:00:00	2 292	9 212	784	0	2	240	152	12 530	116,36	560
			18,29 %	73,52 %	6,26 %	0,00 %	0,02 %	1,92 %	1,20 %			
6	2021-01-01	00:00:00	2 174	18 436	402	0	2 000	248	-12 616	23 260	66,56	560
			9,35 %	79,26 %	1,73 %	0,00 %	8,60 %	1,07 %	-54,24 %			
									-9 260	72 208		
									-12,82 %			

Le tableau ci dessus illustre en couleurs des exemples de calculs intermédiaires réalisés sur l'extrait du fichier de production de la page précédente afin de déterminer si une région peut satisfaire les contraintes de la tâche donnée en exemple à la page précédente pour une heure donnée.

En jaune est indiquée la part d'un type de production dans le mix régional (ex. 26,5 % de l'électricité produite dans la région 3 est d'origine hydraulique).

En orange la part des importations dans la consommation régionale si la région est importatrice et la part de la production exportée si la région est exportatrice (ex. la région 5 importe 1,2 % de sa consommation alors que la région 2 exporte 68,56 % de sa production).

En vert, la production locale (donc ne tenant pas compte des importations) à une région, le coût moyen de la production locale, et le coût le plus élevé parmi les moyens assurant localement plus de 4 % de la production.

En bleu, le cumul des importations et productions de toutes les régions, ainsi que la part de l'importation dans la consommation nationale si la France est importatrice ou la part de la production exportée si elle est exportatrice. (l'exemple ne concerne ici que 6 régions sur 12, les 12,82 % d'exportation ne sont donc pas vraiment pertinents pour évaluer le critère J.)

La tâche donnée en exemple donne les 6 régions possibles (liste K), mais pour chaque heure, l'application des contraintes en élimine plusieurs. Pour la première heure (01/01/2021 à minuit), voici le résultat de l'application des critères :

**Critère F** - La région 1 n'est pas retenue car son coût moyen est supérieur à 300

**Critères G et H** - les régions 1, 3, 5 et 6 ne sont pas retenues car le coût marginal (le plus élevé) des moyens participant à au moins 4 % (H) de la production est supérieur à 400 (G). Si dans la région 5 les valeurs de production Thermique et Bioénergies avaient été échangées, le coût marginal retenu aurait été celui de bioénergie et la région aurait été conservée.

**Critère I** - les régions 1 et 3 ne sont pas retenues car elles importent plus de 20 % de leur électricité (94,5 % et 89,4 % respectivement) ; on ne peut donc pas se fier aux valeurs de leur production pour prendre une décision<sup>6</sup>.

**Critère J** - On ne peut pas l'utiliser sur cet exemple car les informations sur les 12 régions sont absentes, mais si l'importation nationale était supérieure à 5 %, alors aucune des régions ne serait retenue pour cette tranche horaire.

Pour le créneau étudié (01/01 0:00) seuls les centres de calcul des régions 2 et 4 peuvent être utilisés. Afin de planifier la tâche qui dure 8h, il faudra explorer la suite du fichier et trouver d'autres créneaux satisfaisant les contraintes.

#### Plusieurs modes d'allocation à prendre en compte

L'objectif dans tous les modes est de terminer la tâche au plus tôt, tout en respectant les contraintes. La tâche peut être interrompue et reprise plus tard. Trois modes d'allocation sont considérés : **mono-région**, **parallèle** et **séquentiel**.

Dans le mode *mono-région*, il faut proposer un scénario pour chaque région de la liste K pour laquelle c'est possible. Dans notre exemple, les scénarios pour les régions 2 et 4 commenceraient avec le créneau du 2021/01/01 0:00, mais celles des autres régions commenceraient plus tard.

Centre-Val\_de\_Loire 8  
1 1 0 2 19.78 1 1 1 2 19.66 1 1 2 2 19.79 1 1 3 2 19.92 1 1 4 2 19.36 1 1 5 2 19.70 1 1 6 2 19.25 1 1 7 2 19.66  
Normandie 8  
1 1 0 4 33.96 1 1 1 4 35.20 1 1 10 4 35.31 1 1 11 4 35.11 1 1 12 4 35.05 1 1 13 4 35.17 1 1 14 4 35.23 1 1 15 4 35.01  
...

Avec le mode *parallèle*, il faut proposer un seul scénario qui utilise les créneaux valides disponibles au plus tôt, en répartissant sur plusieurs régions simultanément si possible. Dans notre exemple, les 8 heures de calcul pourraient être effectuées en seulement 4h dans les régions 2 et 4

PARALLELE 8  
1 1 0 2 19.78 1 1 0 4 33.96 1 1 1 2 19.66 1 1 1 4 35.20 1 1 2 2 19.79 1 1 3 2 19.92 1 1 4 2 19.36 1 1 5 2 19.70

Avec le mode *séquentiel*, il faut proposer un seul scénario qui choisit parmi les créneaux valides disponibles au plus tôt dans la liste des régions celui au coût le plus bas, mais sans pouvoir répartir sur plusieurs régions simultanément.

Ces modes d'exécution ne donnent pas forcément des résultats optimaux, mais ont été choisis pour simplifier les algorithmes.

<sup>6</sup> En utilisant le détail des flux entrants par territoire, on pourrait leur affecter le coût moyen pour le territoire (c'est ce qui est fait dans <https://app.electricitymap.org/zone/FR>), mais pour simplifier ce n'est pas demandé dans le projet.