

Estimation de l'empreinte carbone d'une heure de calcul sur un cœur CPU ou sur un GPU

Méthodologie

[Janvier 2024]

Par **Labos 1point5**

Sommaire

Introduction	3
Estimation de l'empreinte carbone d'un service de calcul	4
Estimation des unités fonctionnelles	11
Incertitudes	13
Annexes	15
Références	21

RÉDACTION

David Benaben (UMR BFP, Bordeaux Metabolome, INRAE, Bordeaux)

Françoise Berthoud (UMS GRICAD, Université Grenoble Alpes, Grenoble)

Gaël Guennebaud (Inria, GDS EcoInfo, Bordeaux)

Anne-Laure Ligozat (LISN, ENSIIE, Orsay)

Sophie Valcke (UMR CECI, CNRS, Cerfacs, Paris)

RELECTURE

Marion Avet (iEES Paris, INRAE, Paris)

Marianne De Paepe (MICALIS, INRAE, Jouy-en-Josas)

P.-E. R. (CNRS)

Céline Serrano (Inria, Paris)

INTRODUCTION

Les infrastructures de recherche (IR), en particulier les centres de calcul, sont des instruments mutualisés utilisés par les scientifiques. La méthodologie proposée dans ce document vise à permettre aux utilisateurs de ces centres d'évaluer l'empreinte carbone de leurs calculs.

Cet objectif général sous-tend deux questions :

- Comment **estimer** l'empreinte environnementale d'un service donné dans son ensemble ?
- Comment **répartir** cette empreinte entre les différents utilisateurs ?

Cette deuxième question passe par le choix d'une clé d'allocation conduisant à l'estimation d'un facteur d'émission (FE) qui représente la quantité de gaz à effet de serre (GES), en kilogramme de carbone équivalent (kgCO_2e), correspondant à une unité d'utilisation du service. Par exemple, dans le contexte du calcul intensif, un choix possible pour ce FE est la quantité de CO_2e émis par une heure de calcul sur un cœur d'un *central processing unit* (CPU) ou sur un *graphics processing unit* (GPU). L'utilisateur de l'infrastructure connaissant le nombre de cœurs utilisés et le nombre d'heures de calcul effectuées sur le cluster pourra alors s'attribuer l'empreinte carbone correspondante.

Le groupe de travail *Infrastructures de recherche : calcul* de l'équipe Empreinte du groupement de recherche (GDR) *Labos Ipoint5* propose ici d'intégrer dans l'outil *GES Ipoint5* une estimation des FE pour les infrastructures de calcul nationales et les mésocentres.

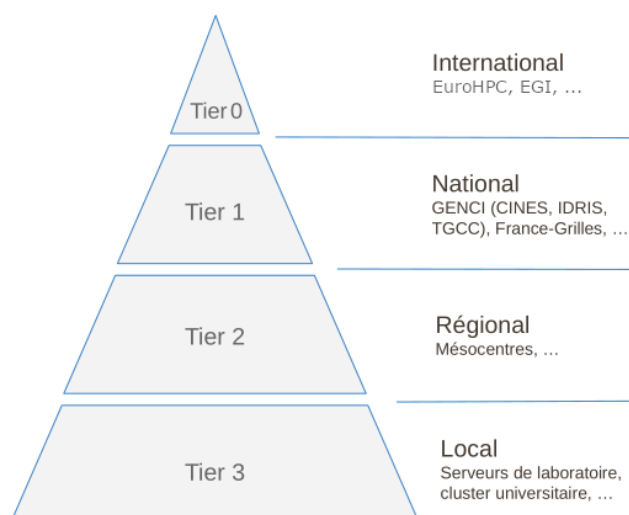
Ce document a pour objectif d'explicitier la méthodologie et les choix effectués en commençant par l'estimation de l'empreinte carbone d'un service de calcul lui-même et, dans un second temps, la stratégie d'allocation entre les utilisateurs mise en place.

La méthode a été éprouvée auprès de centres nationaux du Grand équipement national de calcul intensif (GENCI) et de mésocentres en 2023.

ESTIMATION DE L'EMPREINTE CARBONE D'UN SERVICE DE CALCUL

L'objectif est de définir une méthodologie commune à appliquer aux différents services de calcul intensif de manière à assurer une cohérence satisfaisante entre les résultats.

INFRASTRUCTURES PRISES EN COMPTE



Dans un premier temps, seront intégrés les calculateurs du GENCI hébergés au Centre informatique national de l'enseignement supérieur (CINES) à Montpellier, à l'Institut du développement et des ressources en informatique scientifique (IDRIS) du CNRS à Orsay et au Très grand centre de calcul (TGCC) du CEA à Bruyères-le-Châtel.

Les mésocentres régionaux du réseau MesoNET feront l'objet d'une évaluation séparée basée sur la même méthodologie. Cette méthodologie peut également être appliquée aux calculateurs thématiques ou mutualisés (en bio-informatique, en sciences du climat, etc.)

Une infrastructure de calcul est généralement composée de plusieurs partitions, chaque partition comprenant des nœuds (serveurs) de nature différente. Chaque serveur est lui-même composé d'un ou plusieurs CPU et peut également contenir plusieurs GPU. Un CPU est composé de plusieurs cœurs de calcul. Ces serveurs de calcul sont associés à des serveurs de service et de stockage dédiés. L'ensemble est hébergé dans une salle (centre de données) équipée en systèmes d'alimentation électrique et de refroidissement.

INTRODUCTION À LA MÉTHODOLOGIE

La méthodologie utilisée ici est partiellement basée sur le référentiel par catégorie de produit (RCP) pour les services numériques (1) proposé par l'ADEME en juillet 2021, lequel s'appuie sur la recommandation L.410 de l'Union Internationale des télécommunications (ITU). Le principe général repose sur une analyse simplifiée du cycle de vie d'un service numérique, qui ne considère que les émissions de GES.

Pour rendre compte de l'impact environnemental d'un service numérique, comme de n'importe quel service ou produit, la première étape consiste à définir l'unité fonctionnelle de base qui sera évaluée. Elle permet notamment de comparer les impacts environnementaux d'un même service à deux époques différentes, sur la base d'une unité commune. Cette unité reflète la fonction que le service apporte à l'utilisateur.

Dans le cas présent, la fonction d'un service de calcul intensif est de fournir une certaine quantité de calcul. Il a été choisi de représenter cette quantité par l'heure.cœur équivalent (ou heure.GPU équivalent) dont la définition précise sera donnée au paragraphe *Estimation des unités fonctionnelles*. Dans certains contextes, comme par exemple celui des mésocentres, cette notion de cœur équivalent CPU est difficile à définir, et celle-ci pourra donc être approchée par la notion de cœur CPU.

Deux unités fonctionnelles différentes seront donc considérées selon la nature de la partition :

- calculer pendant 1 heure sur 1 cœur équivalent (CPU),
- calculer pendant 1 heure sur 1 GPU équivalent.

Remarque : les valeurs obtenues sont des moyennes sur le fonctionnement des plateformes pendant une période significative.

FRONTIÈRE DU SYSTÈME

Les frontières du système permettent de délimiter l'ensemble des fonctions prises en compte dans le calcul.

L'évaluation environnementale de ce service (calculer pendant 1 heure sur 1 cœur CPU ou 1 GPU) prendra en compte obligatoirement les fonctions suivantes :

- les calculs,
- le stockage temporaire pendant le calcul,
- l'administration système des plateformes de calcul et de stockage temporaire,
- l'alimentation électrique sécurisée des équipements électriques,
- le refroidissement des équipements (serveurs et onduleurs).

Et si possible les fonctions suivantes :

- l'activité du personnel impliqué dans l'installation et l'exploitation du service (personnel technique et administratif),
- la construction et la maintenance du bâtiment hébergeant les équipements nécessaires au service.

Le stockage durable des données, le transfert des données vers ou depuis l'infrastructure et les utilisateurs du service (leurs équipements, missions, déplacements domicile-travail, etc.) sont exclus du système considéré.

Dans l'éventualité où les fonctions optionnelles sont évaluées, il conviendra de fournir les résultats avec et sans ces postes de manière à pouvoir agréger les résultats entre eux.

PÉRIMÈTRE

Le périmètre permet de définir les équipements qui sont pris en compte pour la réalisation des fonctions telles que définies ci-dessus et pour chaque équipement (ou un ensemble) les phases du cycle de vie considérées parmi : la fabrication (incluant l'extraction des matières premières), le transport, l'usage et la fin de vie.

Le périmètre inclut tous les équipements nécessaires pour rendre le service de calcul. Pour chaque équipement, les phases de fabrication, transport et usage seront prises en compte. L'usage inclut principalement la consommation électrique, mais aussi les fuites de fluides frigorigènes et la combustion du gasoil des groupes électrogènes. Les pièces détachées liées à la maintenance des équipements peuvent être prises en compte.

Si l'empreinte du personnel est prise en compte : les déplacements domicile-travail, les déplacements professionnels (missions) et les postes informatiques (fabrication, transport, usage) seront évalués.

Si l'empreinte du bâtiment est prise en compte : la construction et la maintenance avec une période d'amortissement de 50 ans feront partie du périmètre.

Dans cette version de la méthodologie, il a été décidé d'exclure la fin de vie, compte tenu de son très faible impact relatif compte tenu des circonstances de traitement des équipements électriques et électroniques (DEEE) professionnels en France. La récupération de la chaleur fatale est exclue car hors périmètre.

Les tableaux ci-dessous reprennent ces éléments.

Équipement informatique (calcul, stockage, réseau, etc.)		
Obligatoire/facultatif	Équipement	Détails
Obligatoire	Serveurs de calcul, frontends, workload manager	
Obligatoire	Serveurs d'administration et annexes	Supervision, statistiques, LDAP, portail web, hyperviseurs, etc.
Obligatoire	Réseau local (entre nœuds et autres serveurs et stockage)	Ethernet, Infiniband
Obligatoire	Stockage temporaire (lié aux calculs)	
Obligatoire	Pièces détachées (maintenance)	Remplacement des disques défectueux, etc.
Obligatoire	Autres équipements	Racks, câbles, etc.

Système d'alimentation		
Obligatoire/facultatif	Équipement	Détails
Obligatoire	Transformateurs	
Obligatoire	Onduleurs	
Obligatoire	Batteries	
Obligatoire	Groupe électrogène	Combustion de carburant comprise.

Système de refroidissement		
Obligatoire/facultatif	Équipement	Détails
Obligatoire	Système de refroidissement	Groupe eau froide, refroidisseurs adiabatiques, <i>direct liquid cooling</i> , etc.
Obligatoire	Fuites des fluides frigorigènes	Phase d'usage

Bâtiment		
Obligatoire/facultatif	Équipement	Détails
Facultatif	Construction du bâtiment	
Facultatif	Entretien du bâtiment	

Activité du personnel		
Obligatoire/facultatif	Équipement	Détails
Facultatif	Consommations & équipements du personnel	Poste de travail informatique, consommation chauffage, consommation climatisation
Facultatif	Déplacements professionnels	
Facultatif	Trajets domicile travail	

Pour faciliter la compréhension, l'interprétation et le suivi temporel des résultats, une fiche de résultats (voir annexe 6) sera complétée.

LES RÈGLES D'ALLOCATION

Pour chaque élément (équipement ou système) considéré dans le périmètre, une règle d'allocation est définie pour chaque étape du cycle de vie de l'élément afin de quantifier la part attribuée à la fourniture du service numérique de calcul.

L'allocation des phases de fabrication et de transport sont établies en général sur un critère physique, par exemple :

- la fabrication des équipements est rapportée au temps de vie des équipements,
- la consommation électrique des équipements froids (si globalisés sur d'autres services) est rapportée à la proportion de consommation électrique des équipements propres au service par rapport à l'ensemble des équipements.

Cette règle d'allocation permet de calculer un pourcentage d'affectation.

LES DONNÉES

Période considérée

Le bilan des émissions de GES porte sur une année calendaire complète d'activité. Dans la réalité, compte tenu de l'évolution des infrastructures et/ou des équipements, il est rare d'avoir une stabilité du système pendant une année complète. Dans ce cas, les données seront collectées sur la période stable la plus longue possible, puis extrapolées sur une année. Il faudra veiller à ce que les données correspondent toutes à la même période de référence (inventaire du matériel, consommations électriques, nombre d'heures de calcul).

Données primaires

Les données primaires sont directement issues de mesures :

- surfaces, nombre de racks, serveurs, nœuds, cœurs, etc.,
- nombre d'heures de calcul réalisées (effectives),
- consommation d'électricité,
- quantité de carburant utilisé,
- quantité de gaz réfrigérant,
- etc.

Données secondaires

Les données secondaires ne sont pas directement mesurées. C'est le cas général des données relatives à la fabrication des équipements. Les sources suivantes sont utilisées (indiquées par ordre de fiabilité décroissant) :

- **Ecodiag Data**¹ : pour la fabrication des équipements numériques (PC, portables, écrans, serveurs), base de données construite notamment à partir des données des constructeurs corrigées,
- **Boavizta**²,
- **ADEME** : pour de nombreux types d'équipement, Base Empreinte^{®3} et guide sur les technologies numériques (2).

Les données primaires et secondaires permettront d'alimenter, dans le fichier de résultats Excel, l'onglet d'inventaire de cycle de vie (ICV) global.

¹ Voir la base de données accessible ici : <https://ecoinfo.cnrs.fr/ecodiag-calcul/> et ici : <https://gitlab.inria.fr/guenneba/ecodiag-data>.

² Voir ici : <https://datavizta.boavizta.org/> et ici : <https://github.com/Boavizta/environmental-footprint-data>.

³ Voir la base de données disponible ici : <https://base-empreinte.ademe.fr/>.

L'ICV recense l'ensemble des postes du périmètre tel que défini plus haut, avec pour chaque poste :

- le type d'équipement
- le modèle,
- le nombre d'unités,
- la durée de vie (les valeurs varient suivant la machine ainsi que le centre et évoluent dans le temps) : pour les équipements de calcul, la durée de vie couramment associée est celle de la machine (de 4 à 8 ans, voire plus selon les politiques de site et les durées de garantie des équipements), les équipements d'alimentation électrique sont usuellement amortis sur 15 à 20 ans, les batteries sur 5 ans, les équipements de refroidissement sur 15 à 20 ans, et le bâtiment est considéré amorti sur 50 ans,
- le pourcentage d'affectation : pour les équipements mutualisés entre plusieurs services uniquement (le pourcentage d'affectation est de 100 % sinon),
- la consommation électrique en mégawattheure par an (MWh/an) pour 1 an (facultatif),
- le poids, la surface ou l'unité qui correspond au FE des GES (facultatif),
- le FE des GES,
- le total des GES pour ce poste (calculé à partir des informations précédentes),
- l'incertitude,
- la source de la donnée secondaire (le cas échéant).

Le FE (en kgCO₂e) correspond à la quantité de GES émise par le flux ou l'équipement considéré (typiquement pour une unité de production, par kg d'équipement ou de matière, ou encore par kWh électrique).

Exemples

FE de la production d'électricité (FE_{elec}) : Labos lpoint5 a fait le choix d'utiliser la valeur de la Base Carbone[®] de l'ADEME (base de données officielle pour les bilans GES), et notamment pour la France continentale le FE mix moyen⁴. Le FE_{elec} varie d'une année sur l'autre et influence fortement la part globale prise par la consommation électrique des infrastructures de calcul. En 2020, pour la France continentale, le FE mix moyen est de 0,0599 kgCO₂e/kWh.

FE du gaz réfrigérant R410a : le fonctionnement des systèmes de génération de froid peut occasionner des fuites de fluides frigorigènes (R134a, R410a, R22, R407c). Il est possible de récupérer le type et volume de gaz rechargé pour les systèmes de génération de froid (en litres) correspondant (hors charge additionnelle des équipements neufs et diminués d'éventuelles éliminations de fluide comme déchets) aux fuites de fluides frigorigènes. L'information est disponible dans les rapports de maintenance. Les FE des gaz sont dans la Base Empreinte[®]. Pour le R410a, le FE est de 1 924 kgCO₂e/kg.

Complétude des données

Il est fréquent de ne pas trouver les données pour certains modèles d'équipement de froid mais des extrapolations au prorata de la masse des équipements sont possibles à partir de modèles proches.

ESTIMATION DE L'EMPREINTE CARBONE DE L'INFRASTRUCTURE

L'empreinte carbone (EC) sur l'année est alors la somme de tous les postes (notés X) :

$$EC = \sum_x \%affectation_x \left(\frac{GES_{fabricationX}}{durée\ de\ vie_x} + FE_{elec} * Consommation_x \right) \\ + GES_{gaz\ réfrigérant} + GES_{carburant} \\ [+GES_{personnels} + GES_{batiment}]$$

Remarque : les GES des postes entre crochets sont facultatifs et les GES dus aux gaz réfrigérants ($GES_{gaz\ réfrigérant}$) et au carburant ($GES_{carburant}$) sont à calculer le cas échéant (selon les infrastructures existantes).

Dans une perspective de réduction des émissions de GES en valeur absolue, c'est cette empreinte carbone de l'infrastructure qui est l'indicateur de suivi pertinent (3). Le facteur d'émission (FE), calculé par la suite, permet aux utilisateurs d'estimer l'empreinte de leur calcul en répartissant l'empreinte carbone au prorata du volume de calcul effectué.

⁴ Voir le récapitulatif des choix effectués par Labos lpoint5 concernant les FE accessible ici : https://apps.laboslpoint5.org/static/carbon/FacteursEmission_GESlpoint5_Juin2021.pdf

On conservera autant que possible la décomposition de cette empreinte par phase du cycle de vie des équipements (fabrication, usage) et par type d'équipement (serveur de calcul, refroidissement, etc.) pour permettre une analyse plus fine des résultats.

ESTIMATION DES FACTEURS D'ÉMISSIONS PAR INFRASTRUCTURE

Le FE de l'heure.cœur CPU ou de l'heure.GPU est estimé en divisant l'empreinte carbone obtenue ci-dessus pour une année par le nombre d'heures de calcul effectives réalisées pendant cette période sur l'infrastructure.

***Remarque :** on considère ici le nombre d'heures effectives et non pas le nombre d'heures théoriques (nombre d'heures total théoriquement disponible sur le cluster de calcul si tous les cœurs étaient utilisés tout le temps), parce que cela permet de répartir l'empreinte des phases d'inactivité sur les heures de calcul effectives. De manière générale, dans ce contexte d'attribution des impacts, il est important que la somme des GES attribués à l'ensemble des utilisateurs soit égale à l'empreinte carbone de l'infrastructure.*

Deux FE sont estimés ici (selon les infrastructures) :

- **1 heure.cœur CPU** (en gCO₂e) : correspond à calculer pendant 1 heure sur 1 cœur CPU dans les conditions moyennes d'utilisation de l'infrastructure,
- **1 heure.GPU** (en gCO₂e) : correspond à calculer pendant 1 heure sur 1 GPU dans les conditions moyennes d'utilisation de l'infrastructure.

Ici, et dans le reste de ce document, 1 GPU représente une portion d'un nœud équipé de plusieurs GPU et d'un ou plusieurs CPU. Par exemple, pour une partition équipée de nœuds de 4 GPU et 2 CPU, 1 heure.GPU correspond à l'utilisation d'un quart d'un de ces nœuds pendant une heure, soit l'équivalent de 1 GPU plus 0,5 CPU.

ESTIMATION DES UNITÉS FONCTIONNELLES

Les FE, précédemment obtenus par infrastructure et par heure, sont essentiellement des indicateurs d'efficacité. Il est important de bien comprendre que ces facteurs sont composés pour l'essentiel de l'amortissement des GES émis lors de la production des différents équipements et de la consommation électrique dite statique des équipements informatiques (consommation qui ne dépend pas de la présence d'un job ou non). Utiliser ces FE par infrastructure pourrait amener à une fausse concurrence entre les différentes infrastructures, concurrence qui pourrait entraîner une obsolescence accélérée et donc une augmentation des impacts absolus, ce qui est contraire à l'objectif de réduction.

Pour éviter cet écueil, ces FE par infrastructure seront invisibilisés dans *GES Ipoint5* et les différentes infrastructures de calcul seront vues comme un unique service mutualisé. Autrement dit, on ne cherche plus à attribuer les GES d'une infrastructure en particulier aux utilisateurs de cette infrastructure, mais plutôt à attribuer les GES de l'ensemble des infrastructures à l'ensemble des utilisateurs.

Dans ce contexte, un FE par heure.cœur CPU (ou par heure.GPU) ne serait pas satisfaisant. En effet, la fonction d'un service de calcul intensif est de fournir une certaine quantité de calcul. Or, la puissance de calcul d'un cœur CPU (ou d'un GPU) dépend de sa micro-architecture, sa fréquence, des jeux d'instructions, etc. C'est pour tenir compte de cette diversité que les unités fonctionnelles proposées sont exprimées en heures équivalentes :

- calculer pendant 1 heure sur 1 cœur équivalent CPU,
- calculer pendant 1 heure sur 1 GPU équivalent.

Ici, 1 heure.cœur équivalent (respectivement 1 heure.GPU équivalent) est définie relativement à la puissance de calcul d'un cœur (respectivement 1 GPU) de référence. Pour une architecture donnée, le coefficient permettant de passer d'une 1 heure.cœur réelle à 1 heure.cœur équivalent reflète le ratio de performance entre l'architecture donnée et l'architecture de référence pour un code représentatif. De tels coefficients ont été mesurés par GENCI pour certaines architectures matérielles en prenant comme référence un cœur Intel Sandy Bridge 2,7 GHz et un GPU NVIDIA V100 pour les partitions scalaires (CPU) et accélérées (GPU) respectivement à partir de benchmarks applicatifs significatifs utilisés lors des appels d'offre des supercalculateurs. Ces coefficients sont rappelés en annexe.

Ces FE moyens sont estimés de la manière suivante (la formule est la même pour les heure.cœur équivalent et les heure.GPU équivalent, mais on estime bien deux facteurs différents, $FE_{cœur}$ ou FE_{GPU}) :

$$FE = \frac{\sum_i EC_i}{\sum_i \alpha_i \times nb\ heures_i}$$

Où α_i correspond au coefficient de puissance de calcul relative et EC_i à l'empreinte carbone absolue de l'infrastructure i .

Les utilisateurs de *GES Ipoint5* doivent ainsi fournir le nombre d'heure.cœur et heure.GPU pour chacune des infrastructures i . Les émissions de GES attribuées à un utilisateur sont ainsi obtenues de la manière suivante :

$$E_{calcul} = FE_{cœur} \times \sum_{CPUs\ i} \alpha_i \times nb\ heures_i + FE_{GPU} \times \sum_{GPUs\ i} \alpha_i \times nb\ heures_i$$

Notons que ces facteurs d'équivalence ne sont représentatifs que de codes exploitant parfaitement les capacités des différentes architectures matérielles. Pour un code sous-optimal, le coefficient qui serait réellement observé serait très largement inférieur. Ce mode d'attribution des impacts pénalise un code qui sous-exploite les capacités des architectures les plus récentes et efficaces. Cette propriété est plutôt la bienvenue car elle va à l'encontre de l'obsolescence accélérée.

Comme mentionné dans le paragraphe *Estimation de l'empreinte carbone d'un service de calcul*, dans le cas des mésocentres, cette notion de cœur équivalent CPU est difficile à définir à cause de la plus grande hétérogénéité au sein même des plateformes. En effet, cela rend difficile la séparation de l'empreinte carbone et des heures réalisées par type de nœud, ainsi que le benchmark de chaque architecture. On propose donc d'adopter dans un premier temps pour les mésocentres une méthodologie simplifiée consistant à considérer l'ensemble des mésocentres comme un service unifié indépendant des centres nationaux et ayant ses propres facteurs moyens $FE_{cœur}$ et FE_{GPU} sans distinction des facteurs de puissances de calculs, c'est-à-dire en prenant $\alpha_i=1$ pour tous les mésocentres dans les équations précédentes.

INCERTITUDES

Tout au long de ce calcul d'émissions, de nombreuses approximations et incertitudes s'accumulent. Il est important de signaler aux utilisateurs ce niveau d'imprécision. De manière classique, ce niveau d'incertitude peut être appréhendé en modélisant chaque quantité par une variable aléatoire avec un **écart-type relatif** (σ). Pour simplifier l'analyse, on néglige la consommation de carburant et les fuites de fluide frigorigène, et on commence par la partie fabrication des équipements. En supposant une distribution normale de l'erreur (dans ce contexte une distribution log-normale serait plus appropriée), pour la propagation de l'erreur il convient de déterminer non seulement l'écart-type des entrées, mais aussi la corrélation entre les différentes variables. Dans un souci de simplicité, on supposera que l'incertitude au sein d'un groupe d'équipements de nature semblable est parfaitement corrélée, c'est-à-dire que si le facteur d'un des équipements de ce groupe se trouve être significativement sous-évalué (respectivement surévalué), alors les facteurs des autres équipements du groupe sont eux aussi sous-évalués (respectivement surévalués). En d'autres termes, les erreurs commises pour les équipements d'un même groupe ne se compensent pas entre elles. Au contraire, l'incertitude entre ces groupes est quant à elle supposée être parfaitement décorrélée et peut être modélisée par des variables aléatoires dites indépendantes. Autrement dit, une surévaluation sur un groupe est ici supposée être probablement partiellement compensée par la sous-évaluation d'un autre groupe. Cette approche à deux niveaux est également celle retenue dans l'outil *GES lpoint5*⁵.

Les formules de calcul à la fois pour la partie fabrication, consommation électrique et le total sont fournies et détaillées en annexe.

Exemple

En attendant que cet exercice soit réalisé de manière systématique, on propose d'estimer un premier ordre de grandeur de ces incertitudes pour un cas moyen supposé représentatif. Pour ce scénario, on suppose aussi que les incertitudes relatives sur la fabrication F et la durée de vie d sont identiques au sein d'un groupe i . Selon les hypothèses d'incertitudes et de répartition des émissions de fabrication du tableau ci-dessous, on obtient une incertitude de l'ordre de plus ou moins ($\pm$) 34 % pour σ_{fab} .

Groupe	Incrtitude (écart-type)		Répartition des émissions de fabrication	Incrtitude combinée (fabrication et durée de vie)
	Fabrication	Durée de vie		
Équipements informatiques	30 %	25 %	85 %	39 %
Onduleurs	50 %	25 %	3 %	56 %
Batteries	40 %	25 %	2 %	47 %
Groupe et transformateurs	50 %	25 %	1 %	56 %
Refroidissement	50 %	25 %	7 %	56 %
Autre	80 %	25 %	2 %	84 %

⁵ Pour plus d'informations sur cette approche, voir la page *Calcul des incertitudes* de la documentation méthodologique de *GES lpoint5*, accessible ici : <https://apps.laboslpoint5.org/ges-lpoint5>.

En supposant des émissions dues à parts égales à la fabrication et à la consommation (c'est plus ou moins ce que l'on observe avec le mix électrique français), une incertitude de 5 % sur la consommation (due aux extrapolations) et de 10 % sur l'intensité carbone du mix électrique⁶, on obtient **une incertitude totale de l'ordre de ± 18 %**. Celle-ci monte à ± 21 % si l'incertitude sur la fabrication des équipements informatique est portée à ± 40 %.

Cette analyse rapide montre que malgré les incertitudes importantes sur les facteurs d'émissions, l'incertitude finale reste contenue.

⁶ D'après l'ADEME.

ANNEXES

ANNEXE 1 : MODÈLE DE FEUILLE D'ESTIMATION DE L'EMPREINTE CARBONE

Disponible dans le cloud *Labos Ipoint5*.

ANNEXE 2 : COMMENT OBTENIR LE NOMBRE D'HEURES EFFECTIVES ?

Le logiciel de gestion de cluster peut renvoyer le nombre d'heures effectives (réservées) par les utilisateurs du cluster de calcul sur une année.

Par exemple, via la commande Slurm (sreport) pour l'année 2020 :

```
sreport -t HourPer cluster Utilization Start=2020-01-01 End=2021-01-0
```


ANNEXE 3 : COEFFICIENTS DE CONVERSION α_i EN HEURE.CŒUR ÉQUIVALENT

Partition	Spécifications	Coefficient	Source
Partitions scalaires (CPU)			
Curie	E5-2680, 32nm, 2.7-3.5GHz, AVX	1	GENCI
Ada		1	GENCI
Occigen Haswell	E5-2690v3, 22nm, 2.6-3.5GHz, AVX-2	1,4	GENCI
Occigen Broadwell	E5-2690v4, 14nm, 2.6-3.5GHz, AVX-2	1.5	GENCI
Joliot-Curie SKL	Platinum 8168, 14nm, 2.7-3.7GHz, AVX-512	2	GENCI
Joliot-Curie ROME	EPYC 7H12, 7&14nm, 2.6-3.3GHz, AVX-2	2	GENCI
Jean Zay CSL	Gold 6248, 14nm, 2.5-3.9GHz, AVX-512	2	GENCI
Adastra GENOA	EPYC 9654, 5&6nm, 2.4-3.7GHz, AVX-512	2,4	GENCI
Joliot-Curie AFX64	ARMv8.2A, 1,8 GHz, SVE-512 (à voir en 2024)		GENCI
Coefficients génériques (fréquence de base f en GHz)			
x86-64	AVX	$0,37*f$	
Intel	AVX-2	$0,58*f$	
Intel	AVX-512	$0,77*f$	
AMD	AMD Zen 2 ou 3	$0,77*f$	
AMD	AMD Zen 4	$1*f$	
Partitions accélérées (GPU)			
Jean Zay V100	Nvidia - Tflops : 8FP32, 16FP64, 125TF32	1	GENCI
Jean Zay A100	Nvidia - TFlops : 10FP32, 20FP64, 150TF32, 300TF64, 600I8	2	GENCI
Adastra MI250x	AMD - Tflops : 50-100FP64, 380FP16	4	GENCI

ANNEXE 4 : DÉTAILS DES CALCULS D'INCERTITUDES

Fabrication

En suivant la méthodologie à deux niveaux introduite dans le paragraphe *Incertainitudes*, le calcul des émissions de fabrication peut être réécrit de la façon suivante :

$$E_{fab} = \sum_i E_i = \sum_i \sum_j n_{ij} \frac{F_{ij}}{d_{ij}}$$

Où i représente un groupe dans lequel les équipements j sont supposés avoir une incertitude corrélée, n_{ij} dénote la quantité d'équipement j du groupe i avec F_{ij} le FE de fabrication, et d_{ij} la durée de vie respective. L'incertitude relative des émissions de fabrication E_{fab} , peut alors être calculée par :

$$\sigma_{fab} = \frac{1}{E} \sqrt{\sum_i \left(\sum_j E_{ij} \sqrt{\sigma_{F_{ij}}^2 + \sigma_{d_{ij}}^2} \right)^2}$$

Ici, on suppose qu'il n'y a pas d'incertitude sur les quantités n_{ij} , et que les incertitudes sur le FE F_{ij} et la durée de vie d_{ij} sont indépendantes. Cette formule peut être utilisée directement dans les feuilles de calcul des émissions.

Consommation électrique

La part liée à la consommation électrique est le simple produit de deux quantités indépendantes :

$$E_{conso} = F_{elec} C_{elec}$$

L'incertitude relative est donc obtenue par la somme des variances relatives des deux quantités :

$$\sigma_{conso} = \sqrt{\sigma_{F_{elec}}^2 + \sigma_{C_{elec}}^2}$$

Incertainitude totale

L'incertitude relative totale est finalement obtenue par une somme des variances pondérées par le pourcentage relatif p de la part de fabrication et consommation :

$$\sigma = \sqrt{\sigma_{fab}^2 p_{fab}^2 + \sigma_{conso}^2 p_{conso}^2}$$

Analyse simplifiée

Dans le paragraphe *Incertainitudes*, on a réalisé une analyse simplifiée à partir d'hypothèses moyennes supposées être représentatives. Pour cela, on a supposé que les incertitudes relatives sur les FE F_{ij} et durées d_{ij} sont constantes au sein d'un groupe i (et donc indépendantes de j). Comme on s'intéresse uniquement à l'incertitude relative, on a de plus supposé que les émissions de fabrications somment à 1 ($E_{fab}=1$). L'incertitude σ_{fab} sur la fabrication se simplifie alors en :

$$\sigma_{fab} = \sqrt{\sum_i (\sigma_{F_i}^2 + \sigma_{d_i}^2) E_i^2}$$

Ici, E_i représente donc le pourcentage des émissions de fabrication du groupe i .

ANNEXE 5 : EXEMPLES

UMS Grenoble Alpes recherche infrastructure de calcul intensif et de données (GRICAD)

Estimation de l’empreinte carbone d’1 heure.cœur de calcul (4).

Génopole Toulouse (GenoToul)

Estimation de l’empreinte carbone de GenoToul pour l’année 2019 (5).

ANNEXE 6 : FICHE RÉSULTATS

Estimation de l'empreinte carbone de 1 heure.cœur CPU ou 1 heure.GPU

Centre :

Année de référence :

GES total pour l'infrastructure pour l'année de référence (en tonnes CO₂e) :

Résultat 1heure.cœur CPU (en gCO₂e/h) et incertitude :

Résultat 1h.GPU (en gCO₂e/h) et incertitude :

Taille de la machine (nombre de cœurs CPU et/ou nœuds GPU et configuration, stockage) :

Nombre d'heures de calcul effectives pour l'année de référence :

Méthode utilisée pour l'estimation des heures de calcul réalisées :

Taux moyen d'occupation des serveurs de calcul sur la période de référence :

Mix électrique (en kgCO₂/kWh) :

Durée de vie retenue pour les serveurs de calcul :

Power usage effectiveness (PUE) de la salle des serveurs de calcul :

Personnel (en nombre d'équivalent temps plein) :

Complétez les valeurs des éléments pris en compte dans votre calcul en le décomposant (pour 1 heure.cœur CPU ou 1heure.GPU) :

Requis	Décomposition de l'estimation pour 1 heure.cœur	gCO ₂ e/h
X	Fabrication serveurs de calcul	
X	Usage serveurs de calcul (électricité)	
X	Fabrication serveurs de stockage	
X	Usage serveurs de stockage (électricité)	
X	Fabrication autres serveurs (<i>front-ends, workload manager, administration, etc.</i>)	
X	Usage autres serveurs (<i>front-ends, workload manager, administration etc.</i>)	
X	Fabrication équipements réseau	
X	Usage équipements réseau	
X	Fabrication équipements (alimentation électrique, batterie et groupe électrogène)	
X	Usage équipements (alimentation électrique, batterie et groupe électrogène)	
X	Usage groupe électrogène (carburant)	
X	Fabrication équipements de refroidissement	
X	Usage équipements de refroidissement (électricité)	
X	Usage équipements de refroidissement (fuite des fluides frigorigènes)	
	Fabrication autres équipements (racks, câbles, etc.)	
	Fabrication pièces détachées (maintenance des serveurs de calcul et de stockage)	
	Construction et entretien du bâtiment	
	Personnel (équipements terminaux : fabrication, usage)	
	Personnel (déplacements domicile-travail)	
	Personnel (missions)	
	Personnel (installation de la machine)	

RÉFÉRENCES

1. Bureau Veritas, APL - datacenter, DDemain, ADEME. Référentiel méthodologique d'évaluation environnementale des services numériques [Internet]. 2021 p. 29. (Faits et Chiffres). Disponible sur: <https://librairie.ademe.fr/produire-autrement/6022-referentiel-methodologique-d-evaluation-environnementale-des-services-numeriques.html>
2. ADEME. Technologies numériques, information et communication (TNIC). Guide sectoriel 2012 [Internet]. 2012 p. 145. (Hors collection). Disponible sur: <https://librairie.ademe.fr/changement-climatique-et-energie/3526-technologies-numeriques-information-et-communication-tnic-guide-sectoriel-2012.html>
3. Guennebaud G. Empreinte carbone des heures de calculs: limites et paradoxes. In 2022. Disponible sur: <https://hal.science/hal-03947594/>
4. Berthoud F, Bzeznik B, Gibelin N, Laurens M, Bonamy C, Morel M, et al. Estimation de l'empreinte carbone d'une heure.coeur de calcul. 2020; Disponible sur: <https://hal.science/hal-02549565v5>
5. Newsletter #35: special carbon footprint issue [Internet]. GenoToul Bioinfo. 2021. Disponible sur: <https://bioinfo.genotoul.fr/index.php/news/newsletter-35special-carbon-footprint-issue/>

