

APPENDICE 3

GLOSSAIRE

COPRESIDENTS, ÉDITEURS ET EXPERTS

Coprésidents de la Réunion d'experts sur les méthodologies intersectorielles pour l'analyse des incertitudes et la qualité des inventaires

Taka Hiraishi (Japon) et Buruhani Nyenzi (Tanzanie)

CHEF DE REVISION

Richard Odingo (Kenya)

AUTEURS

Milos Tichy (République tchèque) et Simon Bentley (Australie)

REVISEURS

Roberto Acosta (Secrétariat de la CCNUCC), Simon Eggleston (Royaume-Uni), Ian Galbally (Australie), Katarina Mareckova (République slovaque), Thomas Martinsen (GIEC/OCDE), Jos Olivier (Pays-Bas), Jim Penman (Royaume-Uni), et Kristin Rypdal (Norvège)

Table des matières

APPENDICE 3 GLOSSAIRE

INDEX DU GLOSSAIRE.....	A3.4
A3.1 INTRODUCTION	A3.6
A3.1.1 Sélection des rubriques	A3.6
A3.1.2 Énoncé des rubriques	A3.6
A3.2 GLOSSAIRE.....	A3.7
REFERENCES.....	A3.22

Index

A		E	
Analyse de la sensibilité	A3.7	Écart type	A3.12
Analyse de l'incertitude	A3.7	Écart type de la population <i>voir</i> Écart type	
Aplatissement	A3.7	Échantillon	A3.12
Assurance de la qualité (AQ)	A3.7	Échantillon aléatoire simple	A3.12
Asymétrie	A3.7	Echantillonnage hypercube latin	A3.12
Autocorrélation	A3.7	Élasticité	A3.12
Autocovariance	A3.7	Erreur	A3.12
B		Erreur aléatoire	A3.12
Biais	A3.8	Erreur type de la moyenne	A3.12
Bonnes pratiques	A3.8	Erreur systématique	A3.12
C		Erreurs systématiques et aléatoires	A3.12
Catégorie de source clé	A3.8	Espérance mathématique	A3.13
Centile	A3.8	Estimateur	A3.13
Centile d'échantillon	<i>voir</i> Centile	Estimateur sans biais	A3.13
Centile de population	<i>voir</i> Centile	Estimation	A3.13
Coefficient de corrélation	A3.8	Exactitude	A3.13
Coefficient de variation	A3.9	Exhaustivité	A3.14
Cohérence	A3.9	F	
Comparabilité	A3.9	Facteur d'émission	A3.14
Confiance	A3.9	FDP <i>voir</i> Fonction de densité de probabilité	
Contrôle de la qualité (CQ)	A3.9	Fonction de densité de probabilité—FDP	A3.15
Corrélation	A3.10	Fonction de répartition	A3.15
Covariance	A3.10	Fonction de distribution cumulative <i>voir</i> Fonction de répartition	
D		I	
Diagramme décisionnel	A3.10	Incertaince	A3.15
Distribution de probabilité	A3.10	Indépendance	A3.15
Distribution log-normale	A3.10	Intervalle de confiance	A3.15
Distribution normale	A3.11	Intervalle sigma	A3.16
Distribution triangulaire	A3.11	L	
Distribution uniforme	A3.11	Loi des grands nombres	A3.16
Données sur les activités	A3.11		

M		Résidu	A3.19
Médiane	A3.16	S	
Méthode de Monte Carlo	A3.16	Sensibilité	A3.19
Mode	A3.16	Série temporelle	A3.19
Mode de population	<i>voir</i> Mode	Statistique	A3.19
Modèle	A3.17	Statistiques	A3.19
Modèle linéaire	A3.17	T	
Modèle non linéaire	A3.17	Technique bootstrap	A3.19
Moments (de variable aléatoire)	A3.17	Tendance	A3.19
Moyenne	A3.17	Théorème de la limite centrale	A3.20
Moyenne arithmétique	A3.18	Transparence	A3.20
O		V	
Opinion d'expert	A3.18	Valeur espérée	A3.20
P		Valeur extrême	A3.20
Paramètres de population	A3.18	Validation	A3.20
Population	A3.18	Variabilité	A3.20
Précision	A3.18	Variabilité aléatoire	A3.20
Probabilité	A3.18	Variance	A3.21
Propagation des incertitudes	A3.18	Variance de moyenne d'échantillon	A3.21
R		Vérification	A3.21
Régression linéaire	A3.19		

APPENDICE 3 GLOSSAIRE

A3.1 INTRODUCTION

Le Glossaire constitue une référence utile pour les compilateurs d'inventaires et les décideurs et contient des termes statistiques généraux et des termes ayant un sens particulier dans le contexte des inventaires d'émissions.

A3.1.1 Sélection des rubriques

La sélection des termes et de l'énoncé des rubriques vise essentiellement à :

- Distinguer entre des termes ayant des sens différents lorsqu'ils sont utilisés dans le contexte de la compilation des inventaires de gaz à effet de serre et lorsqu'ils sont utilisés dans un sens technique, statistique ou mathématique—comme le terme « cohérence », par exemple ;
- Fournir une notation unifiée pour les termes fondamentaux (pour la plupart statistiques) considérés comme essentiels pour la présentation pratique des inventaires ;
- Définir d'autres termes facilitant la compréhension et le développement de directives en matière de bonnes pratiques relatives aux incertitudes dans les inventaires nationaux.

A3.1.2 Énoncé des rubriques

Le Glossaire a une approche pragmatique et fournit un ou plusieurs types de définitions suivants pour chaque rubrique. Le premier type, « Définition pour les inventaires », concerne toute définition établie spécialement pour les inventaires. Dans certains cas, des exemples illustrent des sens spécifiques relatifs à la compilation des inventaires. Le second type, « Définition statistique », explique la définition statistique ou mathématique courante pour un terme donné. Ici aussi, dans certains cas, des exemples illustrent l'application de ces sens à l'utilisation des inventaires. Le dernier type concerne des définitions fournies par d'autres sources, y compris des définitions existantes du SBSTA ou du GIEC convenues par l'Organe subsidiaire de conseil scientifique et technologique (SBSTA) de la Convention-cadre des Nations unies sur les changements climatiques (CCNUCC) (indiqués par CCCC/SBSTA/1999/6 Add. 1), les *Lignes directrices du GIEC pour les inventaires nationaux de gaz à effet de serre—Version révisée 1996*, et l'Organisation internationale de normalisation (ISO). Les clauses qui entrent dans le cadre des définitions suivantes : moyenne arithmétique, espérance mathématique, population, distribution des probabilités, variable aléatoire, statistique et incertitudes et incluent une définition indiquée par [7] sont extraites de la publication *Guide to the Expression of Uncertainty in Measurement* et ont été reproduites avec l'autorisation de l'Organisation internationale de normalisation (ISO). Cette publication de l'ISO peut être obtenue auprès de toute organisation membre ou directement auprès du Secrétariat central, ISO, Case Postale 56, 1211 Genève 20, Suisse. Le copyright demeure la propriété de l'ISO.

Les définitions présentées dans ce Glossaire ne sont pas strictes au sens mathématique ou statistique complet du terme. La plupart des définitions statistiques figurant dans le présent document sont données dans le contexte d'une inférence statistique fréquentielle « classique », bien que l'on reconnaisse l'existence d'autres théories d'inférence statistique. Comme pour tout ouvrage de référence, la compréhensibilité, la clarté, la précision et la concision ont donné lieu à des compromis, et, à cette fin, on a utilisé un minimum de notation mathématique.

A3.2 GLOSSAIRE

ANALYSE DE LA SENSIBILITÉ

Définition statistique : L'analyse de la sensibilité est l'étude d'un algorithme d'un modèle pour déterminer son degré de sensibilité (ou de stabilité) aux variations de ses données d'entrée ou hypothèses sous-jacentes. Elle s'effectue en variant les valeurs d'entrée ou les équations du modèle et en observant

comment les valeurs de sortie du modèle varient en conséquence. Cette analyse de sensibilité peut avoir pour but :

- L'observation des valeurs de sortie correspondant aux variables d'entrée situées dans des plages « raisonnables » ; et
- Le calcul des approximations de différences finies pour les élasticités et les sensibilités requises par certaines méthodologies pour l'étude de la propagation des erreurs dans un système.

ANALYSE DE L'INCERTITUDE

Définition statistique : L'analyse de l'incertitude d'un modèle vise à fournir des mesures quantitatives de l'incertitude des valeurs de sortie dues aux incertitudes dans le modèle lui-même et dans ses valeurs d'entrée, et à examiner l'importance relative de ces facteurs.

APLATISSEMENT

Définition statistique : L'aplatissement est une mesure de l'aplatissement d'une fonction de densité de probabilité (FDP). C'est une fonction simple de deux moments :

L'aplatissement est donné par : $\gamma = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^2}$ où μ_2 et μ_4 sont le deuxième et quatrième moments centraux

centrés de population. Pour la distribution normale, l'aplatissement égale 3. L'aplatissement de l'échantillon a une définition correspondante, avec des moments d'échantillon remplaçant les moments de population ; il est très sensible aux points « aberrants ».

ASSURANCE DE LA QUALITE (AQ)

Définition pour les inventaires : Les mesures d'assurance de la qualité (AQ) incluent un système planifié d'exams effectués par un personnel ne participant pas directement à la compilation/au processus d'élaboration de l'inventaire afin de vérifier que les objectifs de qualité des données ont été atteints, de s'assurer que l'inventaire représente la meilleure estimation possible des émissions et des puits dans l'état actuel des connaissances scientifiques et des données disponibles, et de renforcer l'efficacité du programme de contrôle de la qualité (CQ).

ASYMETRIE

Définition statistique : L'asymétrie est une mesure de l'asymétrie d'une FDP. C'est une simple fonction de deux moments de la FDP, indiquée par : $\gamma = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$ où μ_2 , μ_3 , et σ^2 , sont des moments centraux

centrés. Les distributions symétriques ont $\gamma = 0$. Le même terme est souvent utilisé pour l'asymétrie d'échantillon, auquel cas les deux moments de population sont remplacés par des moments d'échantillon.

AUTOCORRÉLATION

Définition statistique : Coefficient de corrélation calculé pour deux éléments successifs de données dans une série temporelle.

Exemple : Les nombres d'animaux observés pour deux années successives présentent en général une autocorrélation élevée lorsque la durée de vie des animaux est sensiblement supérieure à deux ans.

AUTOCOVARIANCE

Définition statistique : Covariance calculée pour deux éléments de données successifs dans une série

temporelle.

BIAIS

Définition pour les inventaires : Erreur systématique de la méthode d'observation, dont la valeur est le plus souvent inconnue. Elle peut être due à l'utilisation de matériel de mesure mal calibré, à la sélection des éléments à partir d'une population erronée ou à la place privilégiée accordée à certains éléments d'une population, etc.

Définition statistique : Différence entre la valeur espérée d'une statistique et le paramètre qu'elle estime. Voir **Estimateur sans biais**.

Exemple : L'estimation d'émissions fugitives dues au transport et à la distribution du gaz uniquement par des mesures des fuites des pipelines à haute/moyenne pression peut être à l'origine d'un biais si l'on ne tient pas compte des fuites dans le réseau de distribution à basse pression (bien plus difficiles à mesurer).

BONNES PRATIQUES

Définition pour les inventaires : Les bonnes pratiques sont un ensemble de procédures visant à assurer que les inventaires d'émissions de gaz à effet de serre sont exacts en ceci qu'ils ne sont systématiquement ni surestimés ni sous-estimés autant qu'on puisse en juger, et que les incertitudes sont réduites autant que possible.

Les bonnes pratiques concernent le choix des méthodes d'estimations appropriées au contexte national, l'assurance de la qualité et le contrôle de la qualité au niveau national, la quantification des incertitudes, et l'archivage et la présentation des données à des fins de transparence.

CATÉGORIE DE SOURCE CLÉ

Définition pour les inventaires : Une catégorie de source clé est prioritaire dans le système d'inventaire national car son estimation a un effet significatif sur l'inventaire total des gaz à effet de serre directs pour ce qui est du niveau absolu des émissions, de la tendance des émissions ou des deux. (Voir Chapitre 7, *Choix de méthode et recalculs*).

CENTILE

Définition statistique : Le $k^{\text{ème}}$ centile ou centile de population est la valeur qui sépare la $k^{\text{ème}}$ partie inférieure de l'intégrale de la FDP—c'est-à-dire l'intégrale d'une queue de FDP à partir du $k^{\text{ème}}$ centile vers les densités de probabilité inférieures.

Le $k^{\text{ème}}$ centile de population ($0 \leq k \leq 100$) d'une population ayant une fonction de distribution $F(x)$ est égal à z lorsque z satisfait $F(z) = k/100$.

Le $k^{\text{ème}}$ centile d'échantillon est une approximation pour le centile de population qui est dérivé d'un échantillon. C'est la valeur au-dessous de laquelle le pourcentage k de l'observation est situé.

COEFFICIENT DE CORRÉLATION

Définition statistique : Nombre entre -1 et +1 donnant une indication de l'interdépendance de deux variables observées ensemble. Une valeur +1 signifie que les variables ont une relation linéaire directe parfaite; une valeur -1 signifie qu'il y a une relation linéaire inverse parfaite; et une valeur 0 signifie qu'il n'y a pas de relation linéaire. Il est défini comme la covariance des deux variables divisée par le produit de leurs écarts types.

COEFFICIENT DE VARIATION

Définition statistique : Le coefficient de variation, v_x est le rapport de l'écart type de population, σ_x , et de la moyenne, μ_x , où $v_x = \sigma_x/\mu_x$. Il indique souvent également le coefficient de variation de l'échantillon, qui est le rapport de l'écart type de l'échantillon et de la moyenne d'échantillon.¹

COHÉRENCE

Définition pour les inventaires : Cohérence signifie qu'un inventaire sera cohérent au plan interne pour tous ses éléments sur plusieurs années. Un inventaire est cohérent si l'on utilise les mêmes méthodologies pour l'année de référence et les années suivantes et si l'on utilise des ensembles de données cohérents pour estimer les émissions ou les absorptions par les sources ou les puits. Dans certains cas, indiqués aux paragraphes 10 et 11 de CCCC/SBSTA/1999/6 Add.1, on peut estimer qu'un inventaire utilisant des méthodologies différentes pour des années différentes est cohérent s'il a été recalculé avec transparence conformément aux bonnes pratiques.

Définition statistique : Un estimateur statistique pour un paramètre est dit cohérent s'il tend vers le paramètre lorsque la grandeur de l'échantillon utilisé pour l'estimation augmente—c'est-à-dire que la précision est améliorée par un nombre d'observations plus élevé.

COMPARABILITÉ

Définition pour les inventaires : On entend par comparabilité le fait que les estimations des émissions et absorptions présentées par les Parties dans les inventaires doivent être comparables entre les Parties. A cet effet, les Parties devront utiliser les méthodologies et formats convenus par la Conférence des Parties (CDP) pour estimer et présenter les inventaires. L'affectation des diverses catégories de sources/puits devra suivre les divisions figurant dans les *Lignes directrices du GIEC pour les inventaires nationaux de gaz à effet de serre—Version révisée 1996*, au niveau de son résumé et des ses tableaux par secteurs.

CONFIANCE

Définition pour les inventaires : On utilise le terme « confiance » pour représenter la confiance accordée à une mesure ou une estimation. Le fait d'avoir confiance dans les estimations d'un inventaire ne rend pas ces estimations plus exactes ou plus précises; mais contribuera éventuellement à établir un consensus pour savoir si les données peuvent être utilisées pour résoudre un problème. [6]* Cette utilisation du terme confiance diffère considérablement de son sens statistique dans le terme Intervalle de confiance.

CONTRÔLE DE LA QUALITÉ (CQ)

Définition pour les inventaires : Le contrôle de la qualité (CQ) est un système d'activités techniques systématiques destinées à mesurer et contrôler la qualité de l'inventaire pendant son élaboration. Le système de CQ vise à :

- (i) Fournir des contrôles systématiques et cohérents destinés à assurer l'intégrité, l'exactitude et l'exhaustivité des données ;
- (ii) Identifier et corriger les erreurs et omissions ;
- (iii) Documenter et archiver le matériel de l'inventaire et consigner toutes les activités de CQ.

Les activités de CQ incluent des méthodes générales telles que des contrôles d'exactitude pour l'acquisition des données et les calculs, et l'utilisation de procédures standard approuvées pour le calcul des émissions, les mesures, les estimations des incertitudes, l'archivage et la présentation des informations. Des activités de CQ de niveau supérieur incluent des examens techniques des catégories de source, des données sur les activités, des facteurs d'émission et des méthodes.

¹ « Coefficient de variation » est le terme qui est fréquemment remplacé par « erreur » dans une déclaration telle que « l'erreur est de 5 % ».

* Voir références (p. A3.22).

CORRÉLATION

Définition statistique : estimateur statistique donnant une indication de l'interdépendance de deux quantités. Voir Coefficient de corrélation.

COVARIANCE

Définition statistique : La covariance entre deux variables est la mesure de l'interdépendance entre deux variables.

La covariance d'échantillons appariés de variables aléatoires X et Y est calculée à l'aide de la formule suivante :

$s_{xy}^2 = \frac{1}{n} \sum_i^n (x_i - \bar{x})(y_i - \bar{y})$ où $x_i, y_i, i = 1, \dots, n$ sont des éléments dans l'échantillon et \bar{x} et \bar{y} sont des moyennes de l'échantillon.

DIAGRAMME DECISIONNEL

Définition pour les inventaires : Un diagramme décisionnel est un organigramme décrivant les étapes ordonnées spécifiques à suivre pour établir un inventaire ou un composant d'inventaire conformément aux principes des bonnes pratiques.

DISTRIBUTION DE PROBABILITE

Définition statistique : Fonction indiquant la probabilité pour qu'une variable aléatoire prenne une valeur donnée ou appartienne à un ensemble de valeurs donné. La probabilité pour l'ensemble complet de valeurs de la variable aléatoire est égale à 1. [7]*

DISTRIBUTION LOG-NORMALE

Définition statistique : La distribution log-normale est une distribution asymétrique, qui commence à zéro, s'élève jusqu'à un maximum avant de redescendre plus progressivement vers l'infini. Elle est reliée à la distribution normale : X a une distribution log-normale si $\ln(X)$ a une distribution normale.

La FDP de la distribution log-normale est donnée par :

$$f(x) = \frac{1}{\sigma_l x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu_l)^2}{2\sigma_l^2}}, \text{ for } 0 \leq x \leq \infty$$

Les paramètres requis pour spécifier la fonction sont : μ_l la moyenne de la transformation log-naturel des données; et σ_l^2 la variance de la transformation log-naturel des données. Les données et informations utilisables par le compilateur d'inventaire pour déterminer les paramètres d'entrée sont : moyenne = μ ; variance = σ^2 ; et les relations :

$$\mu_l = \ln \frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}$$

et

$$\sigma_l = \sqrt{\ln \left(\frac{\sigma^2}{\mu^2} + 1 \right)}$$

* Voir références (p. A3.22).

DISTRIBUTION NORMALE

Définition statistique : La distribution normale (ou gaussienne) a la FDP indiquée dans l'équation suivante et est définie par deux paramètres (la moyenne μ et l'écart type σ) :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ for } -\infty \leq x \leq \infty$$

DISTRIBUTION TRIANGULAIRE

Définition statistique : Une fonction de distribution triangulaire symétrique a une FDP

$$\begin{aligned} f(x) &= 2(x-a) / \{(b-a)(m-a)\} \text{ lorsque } a \leq x \leq m \text{ et } a < m \leq b \\ &= 2(b-x) / \{(b-a)(b-m)\} \text{ lorsque } m \leq x \leq b \text{ et } a \leq m < b \\ &= 0 \text{ ailleurs,} \end{aligned}$$

où les paramètres qui spécifient la distribution sont la valeur minimum a , la valeur maximum b , et la position la plus probable (c'est-à-dire le mode) m , sous réserve de $a \leq m \leq b$.

DISTRIBUTION UNIFORME

Définition statistique : Une variable aléatoire à distribution uniforme ou rectangulaire reste située dans une plage dans laquelle toutes les valeurs ont la même probabilité. Si les limites supérieures et inférieures de la plage sont a et b respectivement, la FDP est une fonction aplatie de a à b (les deux paramètres définissant la FDP).

La FDP d'une distribution uniforme est donnée par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{ailleurs} \end{cases}$$

où

$$\mu = \frac{a+b}{2}$$

est la moyenne et

$$\sigma^2 = \frac{(b-a)^2}{12}$$

est la variance.

DONNÉES SUR LES ACTIVITÉS

Définition pour les inventaires : Données sur l'ampleur d'activités humaines génératrices d'émissions ou d'absorptions se produisant pendant une durée donnée. Dans le secteur de l'énergie par exemple, les quantités totales de combustible brûlé constituent des données annuelles sur les activités pour les sources de combustion, et le nombre total d'animaux élevés, par espèces, constitue des données annuelles sur les activités pour les émissions de méthane provenant de la fermentation entérique. (*Lignes directrices du GIEC—Version révisée 1996* [9]*)

ECART TYPE

Définition statistique : L'écart type de la population est la racine carrée positive de la variance. Il est estimé par l'écart type d'échantillon qui est la racine carrée positive de la variance d'échantillon.

ECHANTILLON

Sens statistique : Un échantillon est un ensemble fini d'observations tiré d'une population.

ECHANTILLON ALEATOIRE SIMPLE

Définition statistique : Echantillon de n éléments choisis dans une population de sorte qu'il y ait la même probabilité de choix pour chaque échantillon possible.

ECHANTILLONNAGE HYPERCUBE LATIN

Définition statistique : L'échantillonnage hypercube latin est une technique de sélection de valeurs pour des entrées dans des exécutions de modèle sur ordinateur par stratification de la plage de chaque entrée du modèle, et sélection des valeurs d'entrées pour toute la plage des entrées du modèle.

ÉLASTICITÉ

Définition statistique : L'élasticité (ou sensibilité normalisée) est une mesure de la réponse d'une quantité à une variation dans une autre quantité associée. L'élasticité d'une quantité Y qui est affectée par des variations dans une autre quantité X est définie comme le pourcentage de variation dans Y divisé par le pourcentage de variation dans X qui a causé la variation dans Y .

ERREUR

Définition statistique : En matière de statistiques, le terme « erreur » est un terme général qui désigne la différence entre la valeur observée (mesurée) d'une quantité et sa « vraie » (mais en général inconnue) valeur et n'a pas la connotation péjorative de faute ou d'erreur d'étourderie.

ERREUR ALEATOIRE

Voir Erreurs systématiques et aléatoires.

ERREUR TYPE DE LA MOYENNE

Définition statistique : Terme fréquemment utilisé pour indiquer l'écart type d'échantillon de la moyenne.

ERREUR SYSTEMATIQUE

Définition statistique : Voir Erreurs systématiques et aléatoires.

ERREURS SYSTEMATIQUES ET ALEATOIRES

Définition statistique : Une erreur systématique est la différence entre la valeur vraie, mais en général inconnue, d'une quantité mesurée et la valeur observée moyenne qui serait estimée par la moyenne d'échantillon d'un ensemble infini d'observations. L'erreur aléatoire d'une mesure individuelle est la différence entre une mesure individuelle et la valeur susdite de la moyenne d'échantillon.

ESPÉRANCE MATHÉMATIQUE

Définition statistique : 1. Pour une variable aléatoire discrète X prenant les valeurs x_i avec les probabilités p_i , l'espérance mathématique est $\mu = E(X) = \sum p_i x_i$; et 2. Pour une variable aléatoire continue X ayant la fonction de densité de probabilité $f(x)$, l'espérance mathématique, si elle existe, est $\mu = E(X) = \int x f(x) dx$, l'intégrale étant étendue sur l'intervalle (les intervalles) de variation de X . [7]*

ESTIMATEUR

Définition statistique : Un estimateur est une formule spécifiant comment calculer une valeur d'estimation d'échantillon d'un paramètre de population à partir des données échantillonnées. Par exemple, les facteurs d'émission sont souvent estimés comme les moyennes d'échantillons d'ensembles de mesures. Il peut y avoir plusieurs estimateurs pour un paramètre de population, et, en général, chaque estimateur a ses propres propriétés d'échantillonnage, parmi lesquelles la cohérence et l'absence de biais sont parmi les plus importantes.

Des exemples d'estimateurs ponctuels incluent la moyenne \bar{x} qui est un estimateur utilisé couramment pour la valeur espérée (moyenne), et la variance de l'échantillon s^2 , qui est un estimateur couramment utilisé pour la variance.

ESTIMATEUR SANS BIAIS

Définition statistique : Un estimateur sans biais est une statistique dont la valeur espérée est égale à la valeur du paramètre estimé. On notera que ce terme a un sens statistique spécifique et qu'une estimation d'une quantité calculée à partir d'un estimateur sans biais peut ne pas avoir de biais au sens statistique, mais peut avoir un biais au sens plus général du terme si l'échantillon a été affecté par une erreur systématique inconnue. Par conséquent, en termes statistiques, un estimateur avec biais peut signifier une faiblesse de l'évaluation statistique des données collectées, et non pas au niveau des données elles-mêmes ou de leur méthode de mesure ou de collecte. Par exemple, la moyenne arithmétique (moyenne) \bar{x} est un estimateur sans biais de la valeur espérée (moyenne).

ESTIMATION

Définition statistique : L'estimation est l'évaluation de la valeur d'une quantité ou de son incertitude par l'affectation de valeurs d'observation numériques dans une formule d'estimation, ou estimateur. Les résultats d'une estimation peuvent être exprimés comme suit :

- Une estimation ponctuelle qui fournit un chiffre pouvant être utilisé comme approximation pour un paramètre (par exemple, l'écart type de l'échantillon qui estime l'écart type de la population), ou
- Une estimation par intervalle spécifiant un niveau de confiance.

Exemple : Une déclaration de type « On estime que l'émission totale est de 100 kt et son coefficient de variation de 5 % » est basée sur des estimations ponctuelles de l'écart type et moyen de l'échantillon, alors qu'une déclaration de type « L'émission totale est située entre 90 et 110 kt avec une probabilité de 95 % » exprime les résultats de l'estimation sous forme d'intervalle de confiance.

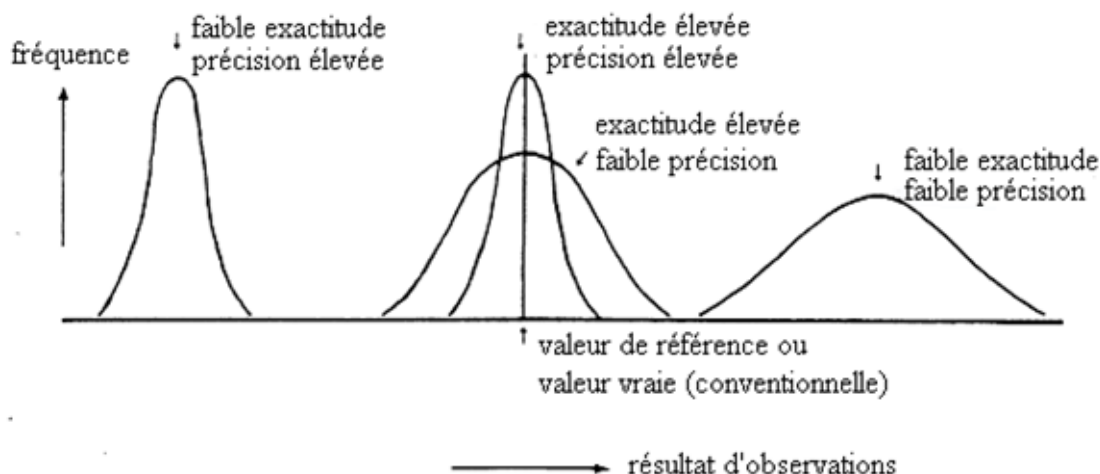
EXACTITUDE

Définition pour les inventaires : L'exactitude est une mesure relative de l'exactitude d'une estimation d'émission ou d'absorption. Les estimations doivent être exactes en ceci qu'elles sont systématiquement ni supérieures ni inférieures aux émissions ou absorptions vraies, autant qu'on puisse en juger, et que les incertitudes sont réduites autant que possible. On utilisera des méthodologies appropriées conformes aux directives sur les bonnes pratiques pour promouvoir l'exactitude dans les inventaires. (CCCC/SBSTA/1999/6 Add. 1)

* Voir références (p. A3.22).

Définition statistique : L'exactitude est un terme général qui décrit le degré selon lequel l'estimation d'une quantité n'est pas affectée par un biais résultant d'une erreur systématique. On doit la distinguer de la précision, comme illustré à la Figure A3.1.

Figure A3.1 Exactitude et précision (d'après [3]*)



EXHAUSTIVITÉ

Définition pour les inventaires : On entend par exhaustivité le fait qu'un inventaire couvre toutes les sources et puits, ainsi que tous les gaz inclus dans les *Lignes directrices du GIEC pour les inventaires nationaux de gaz à effet de serre—Version révisée 1996*, en plus des autres catégories existantes de sources/puits pertinentes spécifiques aux Parties individuelles (et qui peuvent donc ne pas être incluses dans les *Directives du GIEC*). Exhaustivité signifie également la couverture géographique complète des sources et des puits d'une Partie.²

FACTEUR D'ÉMISSION

Définition pour les inventaires : Coefficient qui associe les données sur les activités au volume du composé chimique qui est la source d'émissions ultérieures. Les facteurs d'émission sont souvent basés sur un échantillon de données de mesures, moyennées pour établir un taux d'émission représentatif pour un niveau d'activités donné dans des conditions de service données. (*Lignes directrices du GIEC—Version révisée 1996* [9]*)

FDP

Voir Fonction de densité de probabilité.

* Voir références (p. A3.22).

² Selon les instruments de ratification, il s'agit de l'acceptation, approbation ou adhésion à la Convention par une Partie donnée.

* Voir références (p. A3.22).

FONCTION DE DENSITE DE PROBABILITE—FDP

Définition statistique : Une fonction de densité de probabilité (FDP) est une fonction mathématique qui caractérise le comportement probable d'une population. C'est une fonction $f(x)$ qui spécifie la probabilité relative pour qu'une variable aléatoire continue X prenne une valeur proche de x , et elle est définie comme la probabilité pour que X prenne une valeur entre x et $x+dx$, divisé par dx , dx étant un nombre infiniment petit. Un ou plusieurs paramètres sont nécessaires pour spécifier complètement la plupart des FDP.

La probabilité pour qu'une variable aléatoire continue X soit entre les valeurs a et b est indiquée par l'intervalle de la FDP, $f(x)$, pour la plage entre a et b .

$$\Pr(a \leq x < b) = \int_b^a f(x)dx$$

La FDP est la dérivée (lorsqu'elle existe) de la fonction de distribution :

$$f(x) = \frac{dF(x)}{dx}$$

Dans la pratique, la FDP utilisée est choisie parmi un nombre relativement petit de FDP standard et la principale tâche statistique consiste à estimer ses paramètres. Par conséquent, dans le cas des inventaires, l'indication des FDP utilisées est un élément nécessaire dans la documentation d'une évaluation d'incertitudes.

FONCTION DE REPARTITION

Définition statistique : Une fonction de répartition ou une fonction de distribution cumulative $F(x)$ pour une variable aléatoire X spécifie la probabilité $\Pr(X \leq x)$ pour que X soit inférieure ou égale à x .

FONCTION DE DISTRIBUTION CUMULATIVE

Voir Fonction de répartition.

INCERTITUDE

Définition statistique : Une incertitude est un paramètre, associé au résultat de mesure qui caractérise la dispersion des valeurs qui pourrait être raisonnablement attribuée à la quantité mesurée (la variance d'échantillon ou le coefficient de variation, par exemple). [7]*

Définition pour les inventaires : Terme général et imprécis qui désigne l'absence de certitude (dans les composants d'un inventaire) résultant de tout facteur causal tel que des sources et puits non identifiés, un manque de transparence, etc.

INDÉPENDANCE

Définition statistique : Deux variables aléatoires sont dites indépendantes s'il y a absence complète d'association entre les variations de leurs valeurs d'échantillons. Le coefficient de corrélation est l'indication de l'absence d'indépendance entre deux variables aléatoires utilisée le plus souvent.

INTERVALLE DE CONFIANCE

Définition statistique : Un intervalle de confiance est la fourchette à l'intérieur de laquelle on est à peu près certain de trouver la valeur vraie d'une quantité. Le niveau de certitude est exprimé par la probabilité, dont la valeur est liée à la grandeur de l'intervalle. C'est une des façons d'exprimer l'incertitude (voir Estimation).

* Voir références (p. A3.22).

Dans la pratique, un intervalle de confiance est défini par une valeur de probabilité, par exemple, 95 %, et des limites de confiance de chaque côté de la valeur moyenne \bar{x} . Dans ce cas, les limites de confiance $L1$ et $L2$ seraient calculées à partir de la fonction de densité de la probabilité, pour laquelle il y avait une probabilité de 95 % pour que la valeur vraie de la quantité estimée par \bar{x} soit située entre $L1$ et $L2$. $L1$ et $L2$ sont le 2,5 centile et le 97,5 centile respectivement.

Exemple : « Une émission est entre 90 et 100 kt avec une probabilité de 95 %. » Une telle déclaration est possible lorsque l'intervalle de confiance est calculé (les valeurs numériques dans cet exemple sont choisies arbitrairement).

INTERVALLE SIGMA

Définition statistique : Un intervalle sigma est un intervalle de confiance symétrique centré sur la moyenne et s'étendant c fois l'écart type de chaque côté.

LOI DES GRANDS NOMBRES

Définition statistique : Théorème mathématique qui formalise l'idée souvent évoquée selon laquelle une moyenne devient une meilleure approximation de la moyenne lorsque le nombre d'observations augmente.

MEDIANE

Définition statistique : La médiane ou médiane de population est une valeur qui divise l'intégrale d'une FDP en deux moitiés. Pour des FDP symétriques, elle est égale à la moyenne. La médiane est le 50^e centile de population.

La médiane d'échantillon est un estimateur de la médiane de population. C'est la valeur qui divise un échantillon ordonné en deux moitiés égales. S'il y a $2n + 1$ observations, la médiane est le $(n + 1)^{\text{ème}}$ membre de l'échantillon ordonné. S'il y a $2n$, la médiane est au milieu entre le $n^{\text{ème}}$ et le $(n + 1)^{\text{ème}}$.

METHODE DE MONTE CARLO

Définition pour les inventaires : L'analyse Monte Carlo effectue les calculs de l'inventaire de nombreuses fois par ordinateur, en choisissant chaque fois aléatoirement (par ordinateur) les facteurs d'émission, les paramètres de modèle et les données sur les activités incertaines dans la distribution d'incertitudes spécifiée initialement par l'utilisateur. Les incertitudes relatives aux facteurs d'émission et/ou aux données sur les activités sont souvent importantes et peuvent ne pas avoir de distribution normale. Dans ce cas, les règles statistiques utilisées habituellement pour combiner des incertitudes deviennent très approximatives. L'analyse Monte Carlo peut résoudre ce problème en générant une distribution des incertitudes pour l'estimation de l'inventaire qui est cohérente avec les distributions des incertitudes des entrées relatives aux facteurs d'émission, aux paramètres du modèle et aux données sur les activités.

MODE

Définition statistique : Les distributions peuvent avoir un ou plusieurs modes. Dans la pratique, on rencontre en général des distributions à un seul mode. Dans ce cas, le mode ou mode de population d'une FDP est la mesure d'une valeur centrale autour de laquelle des valeurs échantillonnées à partir d'une distribution de probabilité tendent à être situées et est, généralement parlant, la valeur ayant la probabilité d'occurrence la plus élevée.

Le mode d'échantillon est un estimateur pour le mode de population calculé en subdivisant l'échantillon en sous-classes égales, en comptant le nombre d'observations dans chaque classe et en sélectionnant le point central de la classe (ou des classes) contenant le plus grand nombre d'observations.

MODELE

Définition statistique : Un modèle est une abstraction quantitative d'une situation réelle, qui peut simplifier ou ignorer certaines caractéristiques pour se concentrer plus particulièrement sur ses éléments plus importants.

Exemple : la relation indiquant que les émissions égalent un facteur d'émission multiplié par un niveau d'activités est un modèle simple. On utilise souvent le terme « modèle » pour désigner la réalisation sur ordinateur d'une abstraction de modèle qui calcule un ensemble de valeurs de sorties pour un ensemble de valeur d'entrées—par exemple, des modèles climatiques mondiaux numériques.

MODELE LINEAIRE

Définition statistique : Une variable y est dite reliée linéairement aux (ou est une fonction linéaire des) variables x_1, x_2, \dots si y peut être exprimé par la formule $y = b_0 + b_1x_1 + b_2x_2 + \dots$ où les termes b sont des nombres constants.

Le fait que la fonction soit considérée comme linéaire ou non peut varier selon le contexte de son application.

Exemple : Une émission E est en général exprimée comme le produit d'un facteur d'émission F et d'un niveau d'activité A . Si F est une constante fixe et E varie uniquement lorsque A varie, E est reliée linéairement à A . Cependant, lorsque F et A sont considérés comme des variables (lorsqu'on applique par exemple les équations de propagation d'erreur pour estimer la variance de E en tant que fonction des variances et covariance de A et F) E n'est pas une fonction linéaire de F et A .

MODELE NON LINEAIRE

Définition statistique : Un modèle est dit non-linéaire si la relation entre ses entrées et sorties est non linéaire. (Voir Modèle linéaire)

MOMENTS (DE VARIABLE ALEATOIRE)

Définition statistique : Un moment de population d'une variable X relative à une constante donnée α est défini comme la valeur espérée de la variable aléatoire $(X - \alpha)^k$, c'est-à-dire $E(X - \alpha)^k$. Lorsque α égale la moyenne de population, μ , le moment $E(X - \mu)^k$ est dit le $k^{\text{ème}}$ moment centré de X . Les moments sont importants car les calculs statistiques sont en général basés sur les moments de la FDP plutôt que sur la FDP elle-même. La moyenne et la variance sont les moments les plus courants.

La moyenne de l'échantillon est le premier moment autour de zéro et la variance est le deuxième moment centré. L'asymétrie et l'aplatissement sont deux fonctions de moments centrés utilisées fréquemment qui caractérisent la forme de la FDP.

Les moments d'échantillon sont des estimateurs de moments de population. Le moment d'échantillon de $k^{\text{ème}}$ ordre est la moyenne arithmétique de la $k^{\text{ème}}$ puissance de la différence entre les valeurs observées et leur moyenne.

MOYENNE

Définition statistique : La moyenne, moyenne de population, espérance mathématique ou valeur espérée, est, en termes généraux, une mesure centrale autour de laquelle des valeurs échantillonnées à partir d'une distribution de probabilité tendent à être situées. La moyenne de l'échantillon ou moyenne arithmétique est un estimateur pour la moyenne. C'est un estimateur sans biais et cohérent de la moyenne de population (valeur espérée), qui est lui-même une variable aléatoire avec sa propre valeur de variance. La moyenne de l'échantillon signifie la somme des valeurs divisée par le nombre de valeurs :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (x_i, i = 1, \dots, n \text{ sont des éléments d'un échantillon}).$$

MOYENNE ARITHMÉTIQUE

Définition statistique : Somme des valeurs divisées par le nombre de valeurs. [7]*

OPINION D'EXPERT

Définition pour les inventaires : Jugement qualitatif et quantitatif bien documenté et soigneusement réfléchi, apporté en l'absence de preuves d'observations sans équivoque, par une personne ou des personnes ayant une expertise démontrable dans le domaine considéré.

PARAMETRES DE POPULATION

Définition statistique : Paramètres de la distribution de probabilité qui caractérisent la population. Les moments sont les paramètres de population utilisés le plus fréquemment—par exemple, la moyenne et l'écart type pour la distribution normale. Une quantité utilisée pour décrire la distribution de probabilité d'une variable aléatoire. [7]*

POPULATION

Définition statistique : La population est la totalité des éléments observés. Dans le cas d'une variable aléatoire, on estime que la distribution de probabilité définit la population de cette variable. [7]*

Exemple : toutes les expériences ou événements imaginables d'un type donné.

PRECISION

Définition pour les inventaires : Précision est l'inverse d'incertitude au sens où plus une chose est précise, moins elle est incertaine.

PROBABILITE

Définition statistique : Une probabilité est un nombre réel dans l'échelle de 0 à 1 associé à un événement aléatoire. ([7]*, C.2.1) La probabilité peut être interprétée de plusieurs façons. Selon une interprétation, la probabilité a la nature d'une fréquence relative (c'est-à-dire la proportion de tous les résultats correspondant à un événement), alors que selon une autre interprétation, la probabilité est la mesure d'un degré de certitude. La probabilité pour qu'un événement aléatoire E se produise est souvent indiquée par $\Pr(E)$. Les probabilités peuvent aussi être exprimées sous forme de pourcentages. La *Théorie des probabilités* est un domaine mathématique développé à partir de bases axiomatiques, dont les résultats sont sous-jacents à l'inférence statistique.

PROPAGATION DES INCERTITUDES

Définition statistique : Les lois pour la propagation des incertitudes spécifient comment combiner algébriquement les mesures quantitatives de l'incertitude associées aux valeurs d'entrée avec les formules mathématiques utilisées pour la compilation des inventaires, pour obtenir des mesures de l'incertitude correspondantes pour les valeurs de sortie. Voir Chapitre 6, *Quantification des incertitudes en pratique*, et Appendice 1, *Base conceptuelle pour l'analyse de l'incertitude*.

* Voir références (p. A3.22).

* Voir références (p. A3.22).

REGRESSION LINEAIRE

Définition statistique : Une régression linéaire fournit un moyen d'ajuster une ligne droite à un ensemble de points de données observées, en tenant compte des effets de la variabilité des observations.

Exemple : Si des observations d'émissions sont tracées par rapport à des niveaux d'activité correspondants, la pente de la ligne ajustée par une régression linéaire fournit une estimation du facteur d'émission approprié. La technique peut servir également à estimer la tendance linéaire pour une quantité variable avec le temps.

RESIDU

Sens statistique : Pour une valeur observée dont le comportement est modélisé par un modèle statistique, le résidu est la différence entre la valeur observée et la valeur prévue par le modèle, par exemple par régression linéaire. Le résidu est donc le composant d'une observation qui ne peut pas être expliqué par le modèle.

SENSIBILITE

Définition statistique : La sensibilité est la mesure du degré de réponse d'une quantité par rapport à une variation dans une autre quantité associée. La sensibilité d'une quantité Y qui est affectée par des variations dans une autre quantité X , est définie comme la variation dans Y divisé par la variation dans X qui a entraîné la variation dans Y .

SERIE TEMPORELLE

Définition statistique : Une série temporelle est une série de valeurs qui sont affectées par des processus aléatoires et qui sont observées à des points temporels successifs (mais généralement équidistants).

STATISTIQUE

Définition statistique : Une statistique est une fonction de variables aléatoires d'échantillon. [7]*

STATISTIQUES

Définition statistique : Les statistiques peuvent désigner, au sens large, une compilation de données, souvent relatives à des activités humaines, ou dans un sens plus spécifique, le domaine scientifique portant sur le traitement numérique systématique de données obtenues à partir de groupes d'éléments.

TECHNIQUE BOOTSTRAP

Définition statistique : Terme désignant un type de méthodes statistiques à calculs intensifs sur ordinateurs, utilisant généralement un ré-échantillonnage répété à partir d'un ensemble de données pour évaluer la variabilité d'estimations de paramètres.

TENDANCE

Définition pour les inventaires : La tendance d'une quantité mesure sa tendance relative pendant une durée donnée, une tendance positive indiquant l'augmentation de la quantité, et une valeur négative indiquant sa diminution. Elle est définie comme le rapport de variation de la quantité pendant la durée, divisé par la valeur initiale de la quantité, et est généralement exprimée sous forme de pourcentage ou de fraction.

* Voir références (p. A3.22).

THÉORÈME DE LA LIMITE CENTRALE

Définition statistique : Nom général pour la catégorie de théorèmes mathématiques/statistiques, selon lesquels, de façon très générale, la moyenne arithmétique de n variables distribuées indépendamment et aléatoirement tend vers une distribution normale lorsque n tend vers l'infini. Ceci est vrai pour les distributions sous-jacentes des variables que l'on rencontrera probablement dans la pratique, et certainement pour toute distribution que l'on pourra rencontrer dans le contexte des inventaires de gaz à effet de serre. Pour les inventaires, le théorème permet d'interpréter les variances combinées des émissions totales (qui sont la somme des émissions sectorielles). De plus, dans certaines conditions, le théorème de la limite centrale peut justifier l'approximation selon laquelle les émissions totales d'un inventaire ascendant ont une distribution normale.

TRANSPARENCE

Définition pour les inventaires : On entend par transparence le fait que les hypothèses et méthodologies utilisées pour un inventaire doivent être expliqués clairement pour faciliter la reproduction et l'évaluation de l'inventaire par les utilisateurs des données présentées. La transparence des inventaires est fondamentale à la réussite du processus pour la communication et l'examen des données.

VALEUR ESPÉRÉE

Définition statistique : Voir Moyenne.

VALEUR EXTRÊME

Définition statistique : Les valeurs extrêmes d'un échantillon sont les valeurs maximales et minimales de l'échantillon. La théorie statistique des valeurs extrêmes se rapporte aux estimations des distributions de ces valeurs extrêmes pour de grandes valeurs de n .

VALIDATION

Définition pour les inventaires : Validation est l'assurance d'une démarche et de fondements sans défaut. Dans le contexte des inventaires d'émissions, la validation utilise la vérification pour s'assurer de la compilation correcte de l'inventaire conformément aux instructions et directives sur la présentation. Elle vérifie la cohérence interne de l'inventaire. Au sens légal du terme, validation signifie confirmer ou approuver officiellement un acte ou un produit. [6]*

VARIABILITE

Définition statistique : Désigne les différences observées attribuables à l'hétérogénéité vraie ou la diversité vraie d'une population. La variabilité est le résultat de processus intrinsèquement aléatoires ou dont la nature et les effets ont des répercussions mais ne sont pas connus. En général, la variabilité ne peut pas être réduite par d'autres mesures ou études, mais elle peut être caractérisée par des quantités telles que la variance d'échantillon. [6]*

VARIABLE ALEATOIRE

Définition statistique : Variable pouvant prendre une des valeurs d'un ensemble de valeurs spécifiées et à laquelle est associée une distribution de probabilité. Une variable aléatoire qui ne peut prendre que des valeurs isolées est dite « discrète ». Une variable aléatoire pouvant prendre n'importe quelle valeur dans

un intervalle fini ou infini est dite « continue ». [7]*

VARIANCE

Définition statistique : La variance ou variance de population est un paramètre de FDP qui exprime la variabilité de la population. C'est le deuxième moment central d'une variable aléatoire. La variance d'échantillon est définie comme une mesure de dispersion qui est la somme du carré des écarts d'observations par rapport à leur moyenne, divisé par le nombre d'observations diminué d'une unité. [7]*

$$s^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$$

VARIANCE DE MOYENNE D'ÉCHANTILLON

Définition statistique : La moyenne d'un échantillon de population est elle-même une variable aléatoire qui a son propre comportement caractéristique et sa propre variance. Pour ces moyennes d'échantillons, l'estimation appropriée de la variance n'est pas la variance d'échantillon, qui estime la variabilité associée à une seule valeur simple, mais une valeur inférieure, égale à la variance d'échantillon divisé par la taille de l'échantillon.

VERIFICATION

Définition pour les inventaires : On entend par vérification l'ensemble des activités et procédures qui peuvent être mises en œuvre pendant la planification et l'élaboration, ou au terme de la l'élaboration d'un inventaire et qui peuvent contribuer à établir sa fiabilité pour les applications prévues de l'inventaire. En général, on utilise des méthodes externes à l'inventaire pour vérifier la vérité de celui-ci, y compris des comparaisons avec des estimations effectuées par d'autres organismes ou avec des mesures d'émission et d'absorption obtenues à partir de concentrations atmosphériques ou des gradients de concentrations de ces gaz. [6]*

* Voir références (p A3.22).

REFERENCES

- [1] Brandt S. (1970). *Computational Methods in Statistics and Data Analysis*. North-Holland, Amsterdam, Pays-Bas.
- [2] Wonnacott R. et Wonnacott T. (1990). *Introductory Statistics for Business and Economics*. John Wiley, New York, États-Unis.
- [3] Zijp W.L. (1987). *Treatment of measurement of Uncertainties*. ECN-194, Petten, Pays-Bas.
- [4] Martin B.R. (1971). *Statistics for Physicists*. Academic Press, Londres, Royaume-Uni.
- [5] Kendall M.G. et Stuart A. (1966). *The Advanced Theory of Statistics*. Charles Griffin & Co.Ltd, Londres, Royaume-Uni.
- [6] Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) (1998). *Meeting Report: Managing Uncertainty in National Greenhouse Gas Inventories*, Rapport de la Réunion de Paris, 13-15 octobre 1998, GIEC/OCDE/AIE Programme des inventaires nationaux des gaz à effet de serre.
- [7] Organisation internationale de normalisation (ISO) (1993). *Guide to the expression of uncertainty in measurement*. ISBN 92-67-10188-9, ISO, Genève, Suisse.
- [8] Kotz, S. et Johnson, N.L. (1988). *Encyclopaedia of Statistical Sciences*. 9 volumes, John Wiley & Sons, New York, États-Unis.
- [9] Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) (1997). *Lignes Directrices du GIEC pour les inventaires nationaux de gaz à effet de serre—version révisée 1996*. Volume I. *Instructions de présentation*. J.T. Houghton et al., GIEC/OCDE/AIE, Paris, France.