

**UNE MÉTHODE DE RÉCONCILIATION DES DONNÉES  
D'ENQUÊTES POUR ÉVALUER LA DYNAMIQUE SPATIALE DE  
L'EMPLOI, CANADA 1987-2008 : APPLICATION À L'ENQUÊTE  
SUR LA POPULATION ACTIVE (EPA)**

Jean Dubé, Ph. D.

Inédit / *Working paper*, n° 2010-1

**Centre - Urbanisation Culture Société**

**INRS**  
Université d'avant-garde

[www.ucs.inrs.ca](http://www.ucs.inrs.ca)



**Une méthode de réconciliation des  
données d'enquêtes pour évaluer  
la dynamique spatiale de  
l'emploi, Canada 1987-2008 :  
application à l'Enquête sur la  
population active (EPA)**

Jean Dubé, Ph.D.

Cette recherche a été subventionnée par le Fonds  
québécois de recherche sur la société et la culture  
(FQRSC)

Institut national de la recherche scientifique  
Centre - Urbanisation Culture Société  
Montréal

Mars 2010

Jean Dubé, Ph.D  
Laboratoire d'analyse spatiale et d'économie régionale (LASER)  
Centre - Urbanisation Culture Société  
Institut national de la recherche scientifique  
Courriel : [jean.dube@ucs.inrs.ca](mailto:jean.dube@ucs.inrs.ca)

*Inédits*, collection dirigée par Mario Polèse :  
[mario.polese@ucs.inrs.ca](mailto:mario.polese@ucs.inrs.ca)  
Centre - Urbanisation Culture Société  
Institut national de la recherche scientifique  
385, rue Sherbrooke Est  
Montréal (Québec) H2X 1E3

Téléphone : (514) 499-4000  
Télécopieur : (514) 499-4065

[www.ucs.inrs.ca](http://www.ucs.inrs.ca)

# TABLE DES MATIÈRES

LISTE DES FIGURES .....	II
INTRODUCTION .....	3
CARACTÉRISATION DE L'ENQUÊTE SUR LA POPULATION ACTIVE .....	5
MÉTHODE ET DONNÉES .....	9
Minimisation de l'entropie croisée .....	9
Données de l'EPA, 1987-2008.....	10
EXAMEN DES RÉSULTATS.....	15
CONCLUSION.....	23
BIBLIOGRAPHIE .....	25

## Liste des tableaux

1 : Seuil minimal de divulgation des données d'emploi de l'EPA par province.....	6
2 : Écarts entre le nombre d'emplois par secteur productif (SP) en fonction de l'agrégation géographique .....	11
3 : Répartition du nombre d'emplois selon les secteurs productifs .....	13
4 : Répartition de l'emploi total (somme des secteurs productifs) selon l'agrégation géographique .....	16

## Liste des figures

1 : Évolution de l'emploi pour le secteur productif « Mines, Pétrole, Gaz - 21 » pour certaines RE, 1987-2008 .....	6
2 : Évolution de l'emploi pour le secteur « Services d'utilité publique - 22 » pour certaines RE, 1987-2008 .....	7
3 : Évolution de l'emploi du SP « Services d'utilité publique - 22 » pour la Côte-Nord de la Nouvelle-Écosse, 1987-2008 .....	19
4 : Évolution de l'emploi du SP « Extraction minière et extraction de pétrole et de gaz – 21 » pour la région de Montréal, 1987-2008.....	19
5 : Évolution de l'emploi du SP « Services d'utilité publique - 22 » pour Parklands Nord (Manitoba), 1987-2008 .....	20
6 : Évolution du nombre d'emplois total pour deux RE de la Colombie-Britannique, 1987-2008 .....	21

## **Résumé**

Cet article présente une application de la méthode d'estimation de la répartition spatiale et temporelle de l'emploi au Canada. La méthode de minimisation de l'entropie croisée permet de réconcilier des données qui sont, a priori, divergentes des totaux agrégés lorsque la désagrégation est suffisamment importante. À partir des données d'emploi provenant de l'enquête sur la population active en fonction des régions économiques (RE) et des secteurs productifs (SP), des matrices (rectangulaires) de répartitions spatiales de l'emploi sont obtenues pour l'ensemble des années disponibles (1987-2008). Cependant, les totaux marginaux, obtenus en sommant les éléments, sont différents des totaux provinciaux et canadiens. L'article montre comment il est possible de réconcilier les données d'enquêtes faisant état de règles d'arrondis et de règles de confidentialité dans le but d'obtenir une source de données spatio-temporelle permettant d'utiliser, au meilleur des connaissances, des données incomplètes.

## **Mots clés :**

Entropie croisée, Données d'enquêtes, Emploi, Répartition spatiale.

JEL : C61, C83, J82, R12

---

## **Abstract**

This paper presents an application of an estimation method for calculating coherent spatial employment distributions for Canada over time. The cross-entropy minimization method allows for data reconciliation when disaggregation causes divergence in total distributions. Using Statistics Canada Labour Force Survey data by economic region and industry class, spatial employment distribution (rectangular) matrices were obtained over the entire time period (1987-2008). However, the sums of regional entries are different from provincial or Canada-wide totals. The paper proposes a procedure for reconciling survey data in cases where confidentiality and rounding rules are apply in order to obtain a coherent spatial-temporal data set even where the original data are incomplete.

## **Key Words:**

Cross-Entropy, Survey data, Employment, Spatial Distribution.

JEL : C61, C83, J82, R12





# Introduction

En science régionale et en économie géographique, les chercheurs s'intéressent à l'étude des comportements de l'économie à un niveau de désagrégation important. Or, l'étude des phénomènes à l'échelle des petites régions est souvent compliquée par le fait que peu de données et d'informations existent pour ce niveau de désagrégation. Ce constat est encore plus frappant lorsqu'on cherche à effectuer des analyses à la fois spatiale et temporelle. De plus, lorsque certaines données sont disponibles, elles font l'objet de plusieurs restrictions vu leur faible fiabilité. Cette problématique complique l'analyse qui peut être faite, mais introduit également une réserve quant à l'utilisation de méthodes quantitatives et le recours à la modélisation. Il devient donc important de trouver une façon de contourner le problème du manque de données ou de leur manque de fiabilité au niveau régional.

Les données de sondage, par exemple, sont souvent sujettes à certaines restrictions lorsque le niveau de désagrégation est important. Certaines règles de confidentialité et/ou d'arrondissements peuvent rendre incomplètes les données disponibles. Ces règles de divulgation ne sont pas sans importance lorsqu'il est question de mettre en relation des variables d'intérêts ou de calculer des taux de croissance. Dans un tel contexte, les relations estimées peuvent être fallacieuses ou encore variables selon la segmentation retenue. Évidemment, ces lacunes peuvent être atténuées en optant pour des méthodes d'agrégation différentes, en changeant l'approche de modélisation ou encore en exprimant la variable dépendante de sorte qu'elle soit moins exposée aux problèmes potentiels.

Une alternative consiste à évaluer les données manquantes (confidentielles ou arrondis) de sorte à obtenir une nouvelle distribution qui ne soit plus sujette aux différentes règles d'arrondissements et de confidentialité. L'article présente une application de la méthode de minimisation de l'entropie croisée (MinXEnt) dans le but d'obtenir une nouvelle distribution spatiale et temporelle de l'emploi à partir de données provenant de l'enquête sur la population active. À partir d'une matrice de répartition (rectangulaire) de l'emploi par région économique (RE) et par secteur productif (SP), des différentes contraintes reliées aux totaux provinciaux agrégés et des règles d'arrondissements et de confidentialité, il est possible d'obtenir une nouvelle distribution qui soit plus près de la réalité tout en assurant un minimum de divergence avec la distribution d'origine. Les matrices de répartitions spatiales de l'emploi sont estimées<sup>1</sup> pour l'ensemble des années disponibles (1987-2008) et sont ensuite synthétisées et analysées à la fois sur la dimension spatiale et sur la dimension temporelle.

---

<sup>1</sup> Le terme « estimation » prend ici une tournure différente de celle employée dans les modèles statistiques. Il s'agit de l'expression d'une nouvelle distribution de l'emploi.

L'article est divisé en quatre parties. La première présente l'enquête sur la population active ainsi que ses caractéristiques dont la périodicité de l'enquête et les règles de confidentialité et de suppression des données de sondage. La seconde partie présente la méthode de la minimisation de l'entropie-croisée ainsi que les données disponibles et utilisées pour faire les estimations. La troisième partie présente les résultats d'estimation et met en relief les différences entre les données obtenues et celles de l'enquête sur la population active. Finalement, une brève conclusion clôture l'article.

## CARACTÉRISATION DE L'ENQUÊTE SUR LA POPULATION ACTIVE

L'Enquête sur la population active (EPA) est une enquête menée chaque mois auprès d'environ 54 000 ménages canadiens dans le but de dresser le portrait détaillé du marché du travail (Statistique Canada, 2009). Cette enquête permet de calculer certains indicateurs fondamentaux tels que le taux de participation au marché du travail, le taux d'emploi et le taux de chômage au Canada, par province, mais également par région économique (RE) et par région métropolitaine de recensement (RMR). Elle permet également de connaître la distribution de l'emploi en fonction des classes industrielles (dénotées dans l'article par secteurs productifs – SP) en plus d'évaluer le nombre moyen d'heures travaillé et le nombre d'emplois par type de profession. Elle reste la principale source pour suivre l'évolution de l'emploi au Canada entre les recensements.

L'EPA consiste en une interview pour un échantillon de la population en âge de travailler (15 ans et plus). Un ménage ciblé est suivi pendant 6 mois et chaque mois, un sixième de l'échantillon est renouvelé (chevauchement) afin d'assurer un taux de réponse intéressant tout en contrôlant pour les coûts de l'enquête. Les ménages sont sélectionnés à partir d'une méthode d'échantillonnage à deux degrés (Statistique Canada, 2008). En première étape, un échantillon de régions géographiques, désignées par le terme unité primaire d'échantillonnage (UPE), est sélectionné. Par la suite, pour chacune des UPE, un échantillon de ménage est sélectionné. Comme toutes enquêtes, l'EPA demeure sensible aux erreurs d'échantillonnages<sup>2</sup> et aux erreurs non reliées à l'échantillonnage<sup>3</sup>.

Les données peuvent être volatile pour de petites régions géographiques, telles que les RE, et fragiliser l'analyse chronologique des données à fréquence mensuelle. Pour cette raison, les analyses à l'échelle des RE sont souvent effectuées avec des moyennes annuelles ou encore en utilisant des séries désaisonnalisées. Même si ces approches n'éliminent pas la totalité des problèmes liés à la volatilité des données pour de petites RE, elles permettent néanmoins d'en contrôler une certaine partie.

L'EPA possède également des seuils de publication des données. En fonction de la province, certaines valeurs sont supprimées de la base de données lorsque jugées trop peu nombreuses ou susceptibles de révéler certaines informations et contrevenir à la règle de confidentialité (Tableau 1). Les données publiées sont également arrondies à la centaine les plus près. Ces détails, bien que de peu d'importance au niveau des provinces, peuvent représenter un défi de taille pour ceux qui désirent utiliser les données à l'échelle des RE. Les petites RE font état d'une forte variation

---

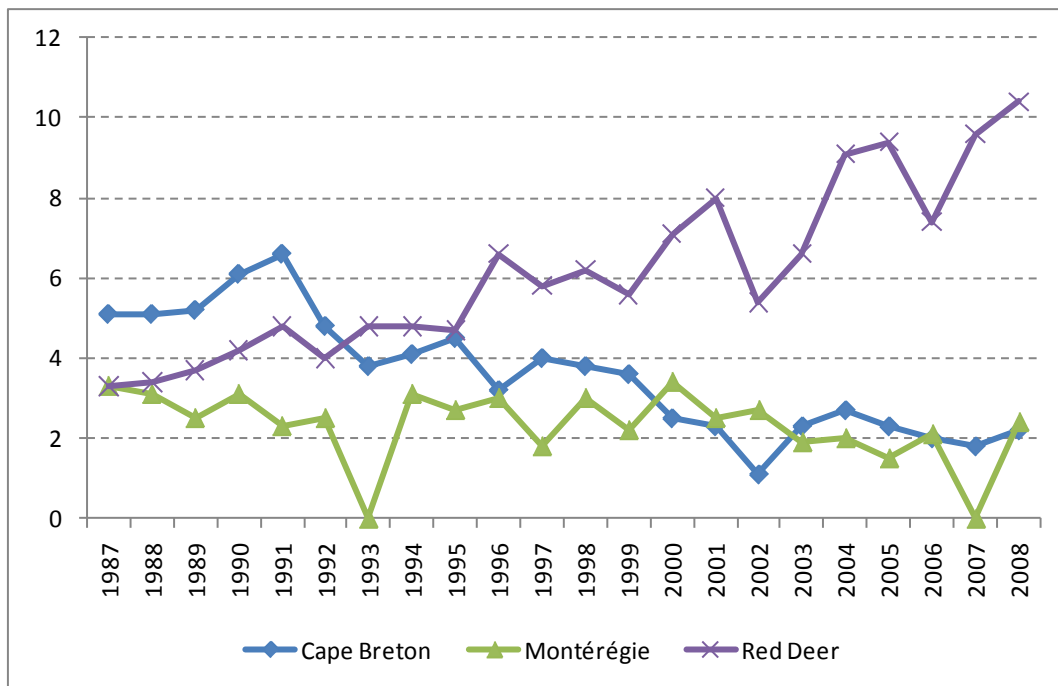
<sup>2</sup> L'estimation est différente de ce que l'on peut obtenir lorsque l'ensemble de la population est interviewé. De plus, elle peut varier en fonction de l'échantillon sélectionné.

<sup>3</sup> Les personnes interviewées peuvent ne pas donner la réponse exacte ou encore approximer la réponse. Il s'agit de toutes erreurs humaines reliées au sondage.

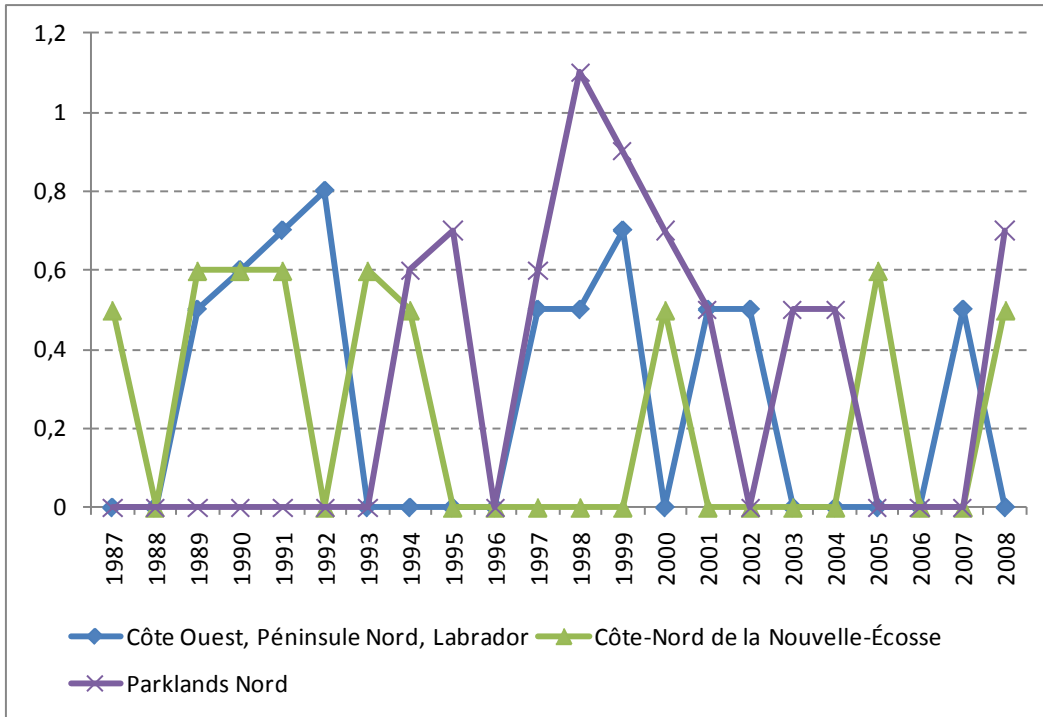
dans les estimations, même annuelle (Figure 1). De plus, les estimations sont susceptibles de montrer certains comportements hasardeux tels que la présence de « faux zéros » dans une série chronologique importante (Figure 2).

**Tableau 1 : Seuil minimal de divulgation des données d'emploi de l'EPA par province**

Province	Emplois	Province	Emplois
Terre-Neuve	500	Ontario	1 500
Île-du-Prince-Édouard	200	Manitoba	500
Nouvelle-Écosse	500	Saskatchewan	500
Nouveau-Brunswick	500	Alberta	1 500
Québec	1 500	Colombie-Britannique	1 500



**Figure 1 : Évolution de l'emploi pour le secteur productif « Mines, Pétrole, Gaz - 21 » pour certaines RE, 1987-2008**



**Figure 2 : Évolution de l'emploi pour le secteur « Services d'utilité publique - 22 » pour certaines RE, 1987-2008**

Ces quelques possibilités peuvent introduire des effets non désirables lorsqu'il est question d'étudier la croissance de l'emploi au niveau régional. La troncature des données, de par un nombre considérable de valeurs *zéro*, implique une mauvaise évaluation des taux de croissance et, par conséquent, risque d'introduire des éléments perturbateurs supplémentaires dans un modèle statistique. Dans ce cas, il peut être souhaitable d'obtenir une meilleure estimation du nombre d'emplois.

La méthode de minimisation de l'entropie croisée (MinXEnt) permet d'estimer une distribution spatiale et/ou temporelle de l'emploi qui soit cohérente avec l'ensemble des informations émanant de l'EPA. Elle s'appuie sur des fondements théoriques rigoureux tout en faisant preuve de transparence et en recourant au principe d'objectivité scientifique (Kapur et Kesavan, 1992).



## MÉTHODE ET DONNÉES

### MINIMISATION DE L'ENTROPIE CROISÉE<sup>4</sup>

La méthode de minimisation de l'entropie croisée (MinXEnt) a l'avantage d'utiliser une distribution existante et d'incorporer un ensemble des contraintes, provenant du respect des totaux marginaux, des règles d'arrondies et des règles de confidentialité, dans un problème d'optimisation visant à obtenir une nouvelle distribution. Il s'agit, à partir d'une distribution a priori, de minimiser la divergence entre les éléments d'origine et la nouvelle distribution (a posteriori) tout en respectant l'ensemble des contraintes imposées par le système. La forme de la fonction objectif assure que la nouvelle distribution obtenue (le résultat) est unique (Kullback, 1959; Kapur et Kesavan, 1992). La minimisation de l'entropie croisée repose sur la mesure de l'information de Kullback et Leibler (1951). Cette mesure propose de calculer la divergence entre deux distributions : i) une a priori<sup>5</sup>, dont les éléments sont notés  $q_{ij}$  et ii) une a posteriori<sup>6</sup>, dont les éléments sont notés  $p_{ij}$ .

De façon générale, le problème d'optimisation sous contrainte cherche à minimiser la mesure de Kullback et Leibler ( $I(p|q)$ ) entre deux distributions en fonction de différentes contraintes dont une de non-négativité des éléments et une plus générale permettant d'inclure l'ensemble des autres contraintes du système<sup>7</sup> (équation 1).

$$\begin{aligned} \text{Min } I(p|q) &= \sum_i \sum_j p_{ij} \log \left( \frac{p_{ij}}{q_{ij}} \right) \\ s / c : & \\ p_{ij} &\geq 0 \\ \sum_i \sum_j p_{ij} g_{ijr} &= a_r \end{aligned} \tag{1}$$

La solution à ce problème procure une distribution qui soit la plus près possible de la distribution de départ tout en respectant les contraintes imposées au système (Dubé et Lemelin, 2005). L'optimum répond au principe scientifique de Laplace, aussi désigné par le principe d'objectivité scientifique, qui stipule que lorsqu'aucune information supplémentaire ne permet de croire que des contraintes additionnelles peuvent être posées au problème d'optimisation, alors la nouvelle distribution obtenue est celle qui a le plus de chance de se réaliser.

---

<sup>4</sup> La méthode est présentée de façon rapide dans le présent article. Cependant, le lecteur intéressé par plus de détails pour consulter Theil (1971, 1967), Dubé et Dupéré (2004) et Dubé et Lemelin (2005).

<sup>5</sup> Qui ne respecte pas nécessairement les contraintes naturelles et celles posées par les règles imposées à divulgation des données.

<sup>6</sup> Qui assure que la nouvelle distribution estimée respecte les contraintes imposées.

<sup>7</sup> Une façon alternative de présenter le problème d'optimisation est proposée par Dubé et Dupéré (2004, 2006).

L'application de la méthode MinXEnt repose sur l'utilisation d'une table de contingence<sup>8</sup>, une par année, qui expose la répartition spatiale de l'emploi dans les régions économiques (RE) en fonction des secteurs productifs (SP). Cette matrice rectangulaire de répartition, dont la dimension est de  $I \times J$ , est composée des  $I$  RE en ligne ainsi que des  $J$  secteurs productifs en colonnes. Il s'agit des données désagrégées provenant de l'EPA. Les contraintes du système proviennent également de l'EPA, en fonction de leur comportement par rapport aux données canadiennes.

#### DONNÉES DE L'EPA, 1987-2008

L'Enquête sur la population active (EPA) découpe le territoire canadien (à l'exception des territoires) en 69 régions économiques (RE) et 16 secteurs productifs (SP), définis par le système de classification des industries d'Amérique Nord (SCIAN<sup>9</sup>). La présence de valeurs *zéro* dans la matrice de répartition de l'emploi par année introduit un défi de taille pour ceux désirant utiliser cette source de données. La distribution de l'emploi (les cellules) par SP et RE montre un écart important par rapport aux totaux marginaux (Tableau 2). Ces écarts sont particulièrement marqués pour les provinces du Québec, de l'Alberta et de la Colombie-Britannique, où la règle de divulgation minimale est fixée à 1 500 emplois.

La somme des emplois par SP en fonction des RE donne un total différent de la somme des emplois par SP en fonction des provinces (Tableau 3). Par contre, en considérant la répartition de l'emploi total, la somme à l'échelle des RE, des provinces et du Canada donne un portrait plutôt semblable, ce qui suggère que les données d'emploi par SP par province peuvent être considérées comme des contraintes du système d'optimisation. De la même façon, il est possible d'établir, à quelques centaines près, la contrainte des totaux marginaux pour l'emploi total par RE. Les écarts peuvent venir des règles d'arrondissements des valeurs, effectuées à la centaine près (un chiffre après la décimale).

---

<sup>8</sup> La présentation formelle du cadre d'analyse, la table de contingence, est explicitée dans Dubé (2003), Dubé et Dupéré (2004) et Dubé et Lemelin (2005).

<sup>9</sup> [www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2002/naics-scian02l-fra.htm](http://www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2002/naics-scian02l-fra.htm)



Tableau 2 : Écarts entre le nombre d'emplois par secteur productif (SP) en fonction de l'agrégation géographique

	Terre-Neuve		Nouvelle-Écosse		Nouveau Brunswick		Québec		Ontario	
	Somme par RE	Écart vs province	Somme par RE	Écart vs province	Somme par RE	Écart vs province	Somme par RE	Écart vs province	Somme par RE	Écart vs province
1987	186,9	3,3	357,7	1,5	278,9	1,1	2 988,4	33,7	4 886,5	8,9
1988	197,1	2,7	372,4	1,4	290,4	0,7	3 053,9	27,6	5 073,1	9,9
1989	203,8	2,6	380,2	1,3	296,5	0,3	3 104,0	24,8	5 192,4	6,4
1990	204,7	2,3	385,0	0,4	298,5	1,8	3 115,0	25,2	5 191,0	3,2
1991	202,5	2,1	380,2	0,4	293,7	1,5	3 056,7	27,8	5 011,5	5,7
1992	192,8	2,3	367,4	1,4	296,1	0,7	3 010,3	28,1	4 929,3	3,4
1993	191,3	2,5	365,5	0,9	298,9	0,9	3 007,1	23,8	4 933,3	4,9
1994	190,9	2,8	372,0	0,5	297,9	0,7	3 067,8	26,8	5 008,4	5,2
1995	192,3	2,2	374,5	1,3	306,6	0,8	3 109,6	25,7	5 094,3	5,8
1996	185,1	2,2	375,3	1,5	305,2	0,4	3 111,3	18,5	5 158,3	8,9
1997	186,2	2,1	380,1	1,8	308,8	0,3	3 147,5	25,3	5 279,1	12,1
1998	190,1	2,5	394,9	0,4	314,5	0,8	3 232,9	24,3	5 445,3	7,9
1999	197,9	3,0	402,5	1,6	324,5	0,9	3 308,2	19,9	5 626,5	10,2
2000	195,3	2,9	410,3	1,1	330,4	0,7	3 382,7	20,2	5 811,0	6,0
2001	201,0	2,8	414,0	1,2	328,9	1,1	3 414,6	25,5	5 920,8	5,2
2002	204,6	2,7	421,8	1,2	342,7	0,4	3 540,4	29,5	6 024,7	6,3
2003	209,6	2,5	430,3	1,1	342,4	0,7	3 605,6	23,3	6 206,9	6,2
2004	212,2	2,1	441,2	0,8	349,1	1,1	3 655,0	25,3	6 309,2	7,4
2005	211,3	2,6	442,5	0,6	349,2	1,3	3 689,4	27,8	6 392,1	5,6
2006	212,8	2,8	441,0	0,8	355,3	0,1	3 743,8	21,6	6 488,7	4,0
2007	215,0	2,2	446,2	1,4	362,1	0,6	3 829,4	22,2	6 588,5	5,4
2008	218,1	2,3	452,2	1,0	365,5	0,5	3 858,3	23,4	6 682,7	4,7

Tableau 2 (suite) : Écarts entre le nombre d'emplois par secteur productif (SP) en fonction de l'agrégation géographique

	Manitoba		Saskatchewan		Alberta		Colombie-Britannique	
	Somme par RE	Écart vs province	Somme par RE	Écart vs province	Somme par RE	Écart vs province	Somme par RE	Écart vs province
1987	501,2	4,1	460,8	1,1	1 171,1	16,4	--	--
1988	503,6	2,7	462,7	0,3	1 207,0	15,3	--	--
1989	508,7	3,9	455,7	0,6	1 235,3	16,0	--	--
1990	511,5	2,2	453,7	0,4	1 258,8	18,1	--	--
1991	503,7	3,0	452,5	0,8	1 267,8	16,7	--	--
1992	497,5	2,5	447,2	1,0	1 266,6	13,4	--	--
1993	502,0	1,6	447,9	0,5	1 275,3	13,6	--	--
1994	504,4	3,3	453,2	1,4	1 310,1	14,7	--	--
1995	513,1	3,5	457,6	0,6	1 352,9	12,0	1 770,0	15,8
1996	514,5	2,8	455,8	1,1	1 395,6	9,4	1 806,0	10,4
1997	522,9	2,6	466,3	-0,1	1 441,9	9,4	1 848,9	11,5
1998	530,6	3,6	470,2	0,3	1 500,8	9,1	1 846,9	11,5
1999	538,6	2,8	470,9	0,6	1 532,4	11,3	1 877,7	16,6
2000	550,4	1,9	473,2	0,5	1 574,8	9,0	1 918,8	12,4
2001	552,3	2,0	459,9	0,5	1 620,3	10,6	1 908,7	12,8
2002	564,7	2,5	467,3	1,3	1 658,3	12,5	1 948,3	16,8
2003	567,4	2,8	475,3	0,7	1 706,6	10,1	2 001,2	13,3
2004	574,7	1,9	479,7	0,1	1 748,6	8,8	2 052,4	10,3
2005	577,6	2,7	483,6	-0,2	1 777,6	6,7	2 122,0	8,5
2006	583,0	4,0	491,4	0,1	1 864,5	6,1	2 179,2	16,3
2007	595,2	1,4	501,2	0,8	1 953,0	6,4	2 255,7	10,7
2008	604,6	2,1	512,3	0,7	2 006,6	6,6	2 299,2	15,0

**Tableau 3 : Répartition du nombre d'emplois selon les secteurs productifs**

	Somme des secteurs productifs			Emploi total		
	Régions économiques	Provinces	Canada	Régions économiques	Provinces	Canada
1987	12 262,4	12 332,5	12 333,1	12 333,1	12 333,2	12 333,0
1988	12 649,5	12 710,1	12 709,6	12 709,7	12 709,5	12 709,6
1989	12 939,9	12 995,8	12 996,1	12 996,4	12 996,2	12 996,2
1990	13 032,8	13 086,4	13 086,3	13 086,3	13 086,4	13 086,4
1991	12 799,6	12 857,6	12 857,5	12 857,6	12 857,4	12 857,4
1992	12 678,1	12 730,9	12 730,8	12 730,9	12 731,0	12 730,9
1993	12 743,8	12 792,5	12 792,7	12 792,7	12 792,7	12 792,7
1994	13 003,5	13 058,9	13 058,8	13 059,0	13 058,7	13 058,7
1995	13 228,1	13 295,8	13 295,5	13 295,3	13 295,3	13 295,4
1996	13 366,0	13 421,2	13 421,3	13 421,2	13 421,4	13 421,4
1997	13 640,3	13 705,3	13 705,9	13 706,4	13 706,1	13 706,0
1998	13 985,7	14 046,1	14 046,1	14 046,1	14 046,3	14 046,2
1999	14 339,5	14 406,4	14 406,6	14 406,4	14 406,9	14 406,7
2000	14 709,3	14 764,0	14 764,1	14 764,0	14 764,3	14 764,2
2001	14 884,1	14 945,8	14 946,2	14 946,3	14 946,2	14 946,2
2002	15 237,4	15 310,6	15 310,5	15 310,3	15 310,5	15 310,4
2003	15 611,3	15 672,0	15 672,2	15 672,1	15 672,5	15 672,3
2004	15 889,0	15 946,8	15 946,9	15 946,8	15 947,0	15 947,0
2005	16 113,5	16 169,1	16 169,5	16 169,8	16 169,6	16 169,7
2006	16 428,3	16 484,1	16 484,3	16 484,4	16 484,4	16 484,3
2007	16 815,8	16 866,9	16 866,3	16 866,2	16 866,3	16 866,4
2008	17 069,6	17 125,9	17 125,9	17 125,5	17 125,9	17 125,8

Il est donc possible d'établir, à partir de ces constats, les contraintes à imposer au système d'optimisation (équation 1). Les contraintes se résument par : 1) le total de l'emploi par SP par province, 2) le total de l'emploi par RE, 3) les règles d'arrondissements fixées par Statistique Canada et 4) les règles de confidentialité/suppression des données sur la publication des données. Néanmoins, les totaux marginaux peuvent varier entre le total des lignes (RE) et des colonnes (SP). Il est donc nécessaire d'effectuer certains compromis sur l'imposition des contraintes. Pour le bien de l'exercice, les distributions marginales de l'emploi par SP par province sont utilisées comme contraintes principales dans le problème d'optimisation. Les distributions marginales de l'emploi par RE peuvent donc varier selon un seuil de tolérance prédéterminé<sup>10</sup>.

---

<sup>10</sup> En général, le seuil de tolérance est identique à celui de la règle d'arrondissement. Cependant, certains écarts sont plus élevés pour certaines années (1997, 2000 et 2007) où les seuils de tolérance ont été augmentés à 0.2 (1997 et 2007) et à 0.3 (2000). Le seuil de tolérance est également relâché pour les années antérieures à 1995 puisque les données ne sont pas disponibles pour les RE de la Colombie-Britannique.

## EXAMEN DES RÉSULTATS

Les matrices de répartition de l'emploi, en fonction des secteurs productifs et des régions économiques, permettent de prendre compte des règles d'arrondissements et de confidentialités tout en respectant les distributions marginales. Les règles de Statistique Canada ainsi que le total marginal des lignes (emploi total par RE) sont utilisées comme indicateurs de la dispersion et guident les ajustements qui peuvent être effectués sur les éléments de la matrice rectangulaire. Elles permettent d'assurer une cohérence des résultats tout en minimisant la « distance » entre la distribution de l'emploi de l'EPA et la nouvelle distribution estimée. En tout, 22 matrices de distribution spatiale de l'emploi sont estimées, soit une par année. Les résultats sont ensuite compilés dans une nouvelle base de données permettant de dresser le portrait de la répartition spatio-temporelle de l'emploi au Canada entre 1987 et 2008 selon la définition de l'EPA.

Les résultats sont imposants en termes de volume de données (chaque matrice est composée de 69 lignes et 16 colonnes), ce qui rend difficile la présentation détaillée des résultats<sup>11</sup>. Sans présenter formellement l'ensemble des tableaux de données, certains constats généraux peuvent être présentés. Le premier constat est, que les nouvelles distributions permettent d'obtenir des résultats, par année et par province, qui sont cohérents avec les données globales de l'EPA (Tableau 4).

Le second constat concerne les cellules non nulles a priori, qui varie très peu par rapport aux données provenant de l'EPA. De par les contraintes imposées au système, la plupart des cellules n'enregistrent que très peu ou pas de modifications bien que les règles d'arrondissements à la centaine peuvent influencer le nombre d'emplois estimé. Par exemple, l'EPA suggère que le nombre d'emplois reliés au secteur de l'agriculture, de la forêt et de la chasse et de la pêche (secteur 11) en 1987 pour la région économique de Toronto est de 7 500, alors que l'application de la méthode MinXEnt suggère que ce nombre est plutôt de 7 525.

Le troisième constat concerne les cellules dont les valeurs sont *zéro* dans la matrice de distribution originale (EPA). L'application de la méthode MinXEnt permet d'obtenir, à la lumière des règles de confidentialité fixée par province, une distribution du nombre d'emplois par secteur qui soit nettement plus cohérent. En tirant quelques exemples, il appert que la méthode procure des séries chronologiques dont les valeurs permettent de calculer des taux de croissance pour chacune des années disponibles (Figures 3, 4 et 5). Un calcul qui était, jusque-là, impossible pour plusieurs SP dans plusieurs RE.

---

<sup>11</sup> Les résultats détaillés sont disponibles auprès des auteurs sur demande.

Tableau 4 : Répartition de l'emploi total (somme des secteurs productifs) selon l'agrégation géographique

	Terre-Neuve		Nouvelle-Écosse			Nouveau-Brunswick			
	Région économique		Province	Région économique		Province	Région économique		Province
	EPA	MinXEnt	EPA	EPA	MinXEnt	EPA	EPA	MinXEnt	EPA
1987	194,6	190,2	190,3	356,7	359,2	359,0	278,9	280,0	280,3
1988	204,7	199,8	199,7	371,1	373,8	373,7	290,4	291,1	291,0
1989	211,2	206,4	206,3	379,0	381,5	381,6	296,5	296,8	296,8
1990	212,4	207,0	206,9	384,2	385,4	385,3	298,5	300,3	300,3
1991	210,3	204,6	204,7	379,2	380,6	380,6	293,7	295,2	295,2
1992	200,3	195,1	194,9	366,5	368,8	368,9	296,1	296,8	296,9
1993	199,4	193,8	193,8	364,8	366,4	366,5	298,9	299,8	299,9
1994	198,9	193,7	193,5	371,2	372,5	372,6	297,9	298,6	298,6
1995	200,3	194,5	194,4	373,6	375,8	375,9	306,6	307,4	307,5
1996	192,8	187,3	187,5	374,4	376,8	376,9	305,2	305,6	305,5
1997	193,5	188,3	188,3	378,8	381,9	382,0	308,8	309,1	309,1
1998	198,9	192,6	192,4	393,9	395,3	395,3	314,5	315,3	315,2
1999	207,0	200,9	201,0	401,4	404,1	404,0	324,5	325,4	325,5
2000	203,7	198,2	198,0	409,1	411,4	411,4	330,4	331,1	331,2
2001	208,6	203,8	203,8	412,7	415,2	415,2	328,9	330,0	330,1
2002	212,9	207,3	207,2	420,7	423,0	422,9	342,7	343,1	343,1
2003	219,9	212,1	212,3	429,3	431,4	431,2	342,4	343,1	343,1
2004	220,7	214,3	214,3	440,2	442,0	442,2	349,1	350,2	350,1
2005	220,2	213,9	214,1	441,5	443,1	443,1	349,2	350,5	350,5
2006	221,8	215,6	215,7	440,0	441,8	441,8	355,3	355,4	355,4
2007	222,6	217,2	217,1	444,9	447,6	447,6	362,1	362,7	362,8
2008	227,1	220,4	220,3	450,5	453,2	453,2	365,5	366,0	366,2

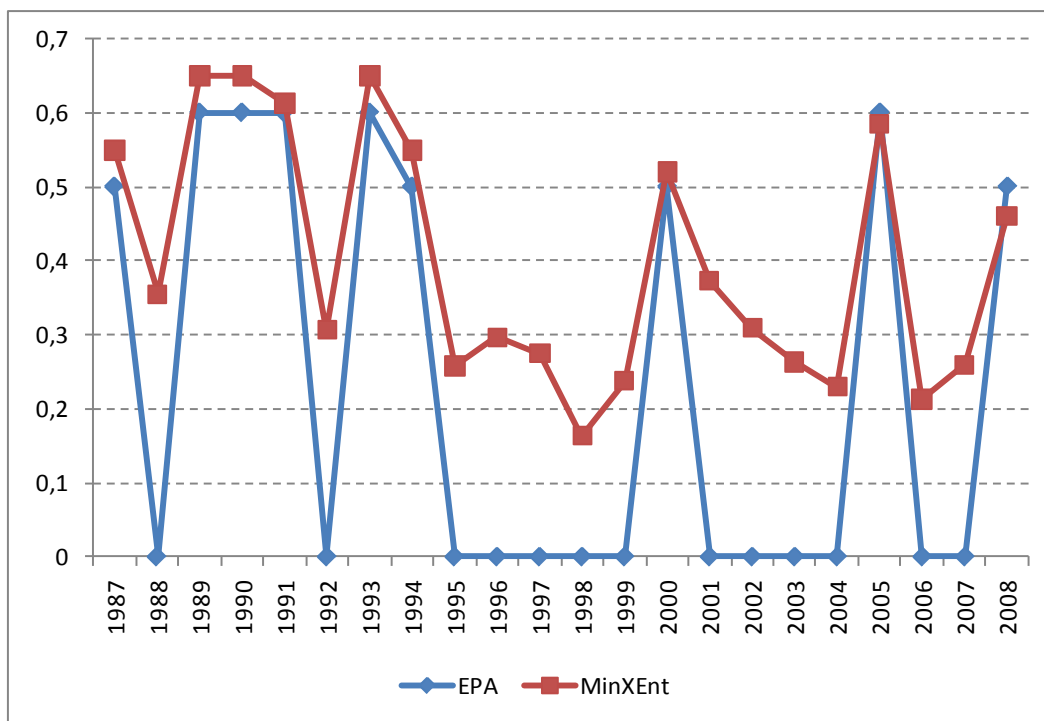
Tableau 4 (suite) : Répartition de l'emploi total (somme des secteurs productifs) selon l'agrégation géographique

	Québec		Ontario			Manitoba				
	Région économique		Province	Région économique		Province	Région économique			Province
	EPA	MinXEnt	EPA	EPA	MinXEnt	EPA	EPA	MinXEnt		EPA
1987	2 988,4	3 022,1	3 022,1	4 886,5	4 895,4	4 895,6	501,2	505,3	505,2	
1988	3 053,9	3 081,5	3 081,4	5 073,1	5 083,0	5 083,1	503,6	506,3	506,3	
1989	3 104,0	3 128,8	3 128,8	5 192,4	5 198,8	5 199,0	508,7	512,6	512,7	
1990	3 115,0	3 140,2	3 140,3	5 191,0	5 194,2	5 194,1	511,5	513,7	513,8	
1991	3 056,7	3 084,5	3 084,4	5 011,5	5 017,2	5 017,1	503,7	506,7	506,8	
1992	3 010,3	3 038,4	3 038,6	4 929,3	4 932,7	4 932,9	497,5	500	499,9	
1993	3 007,1	3 030,9	3 030,9	4 933,3	4 938,2	4 938,0	502	503,6	503,8	
1994	3 067,8	3 094,6	3 094,8	5 008,4	5 013,6	5 013,6	504,4	507,7	507,7	
1995	3 109,6	3 135,3	3 135,3	5 094,3	5 100,1	5 100,0	513,1	516,6	516,5	
1996	3 111,3	3 129,8	3 129,8	5 158,3	5 167,2	5 167,1	514,5	517,3	517,4	
1997	3 147,5	3 172,8	3 172,8	5 279,1	5 291,2	5 291,4	522,9	525,5	525,6	
1998	3 232,9	3 257,2	3 257,5	5 445,3	5 453,2	5 453,3	530,6	534,2	534,2	
1999	3 308,2	3 328,1	3 328,1	5 626,5	5 636,7	5 636,7	538,6	541,4	541,4	
2000	3 382,7	3 402,9	3 402,8	5 811,0	5 817,0	5 817,1	550,4	552,3	552,3	
2001	3 414,6	3 440,1	3 440,2	5 920,8	5 926,0	5 926,2	552,3	554,3	554,3	
2002	3 540,4	3 569,9	3 569,9	6 024,7	6 031,0	6 031,4	564,7	567,2	567,2	
2003	3 605,6	3 628,9	3 628,8	6 206,9	6 213,1	6 213,2	567,4	570,2	570,3	
2004	3 655,0	3 680,3	3 680,5	6 309,2	6 316,6	6 316,5	574,7	576,6	576,6	
2005	3 689,4	3 717,2	3 717,3	6 392,1	6 397,7	6 397,7	577,6	580,3	580,3	
2006	3 743,8	3 765,4	3 765,4	6 488,7	6 492,7	6 492,7	583	587	587	
2007	3 829,4	3 851,6	3 851,7	6 588,5	6 593,9	6 593,8	595,2	596,6	596,5	
2008	3 858,3	3 881,7	3 881,7	6 682,7	6 687,4	6 687,3	604,6	606,7	606,7	

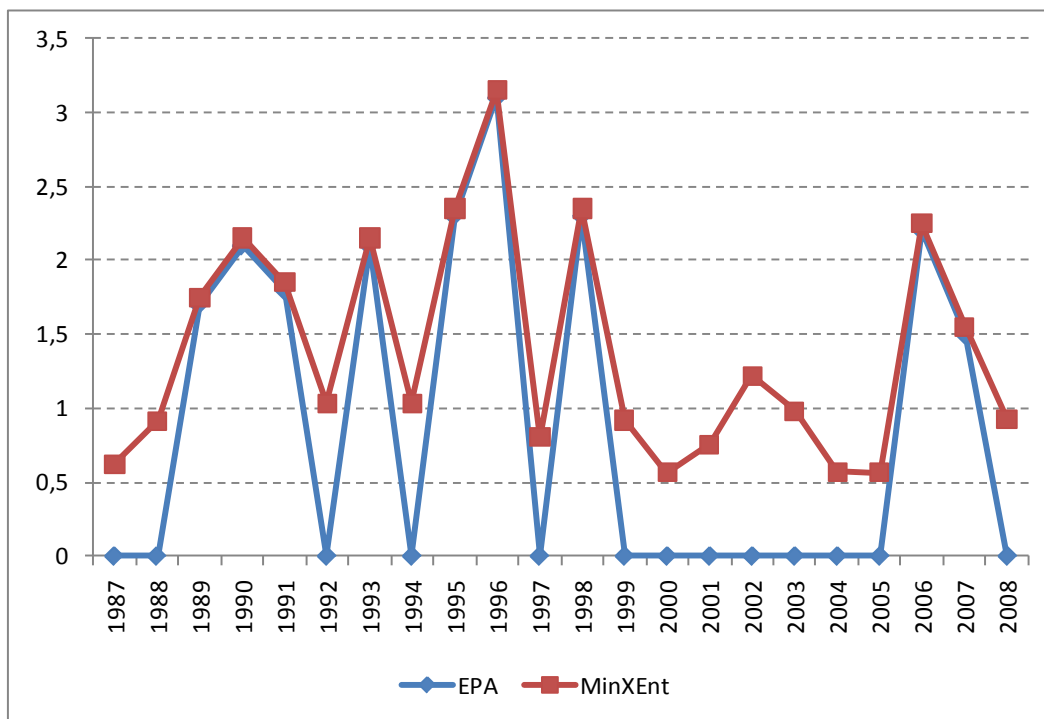
Tableau 4 (suite et fin) : Répartition de l'emploi total (somme des secteurs productifs) selon l'agrégation géographique

	Saskatchewan			Alberta			Colombie-Britannique		
	Région économique		Province	Région économique		Province	Région économique		Province
	EPA	MinXEnt	EPA	EPA	MinXEnt	EPA	EPA	MinXEnt	EPA
1987	460,8	461,9	461,9	1 171,1	1 187,5	1 187,7	0,0	1 377,6	1 377,7
1988	462,7	463	462,8	1 207,0	1 222,3	1 222,3	0,0	1 434,7	1 434,6
1989	455,7	456,3	456,4	1 235,3	1 251,3	1 251,3	0,0	1 508,3	1 508,3
1990	453,7	454,1	454,2	1 258,8	1 276,9	1 276,8	0,0	1 559,6	1 559,6
1991	452,5	453,3	453,3	1 267,8	1 284,5	1 284,4	0,0	1 577,6	1 577,5
1992	446,6	448,2	448	1 266,6	1 280,0	1 280,0	0,0	1 617,3	1 617,2
1993	447,9	448,4	448,5	1 275,3	1 288,9	1 288,7	0,0	1 668,0	1 668,0
1994	453,2	454,6	454,5	1 310,1	1 324,8	1 324,5	0,0	1 743,1	1 743,2
1995	457,6	458,2	458	1 352,9	1 364,9	1 364,9	1 770,0	1 785,8	1 785,6
1996	455,8	456,9	456,8	1 395,6	1 405,0	1 405,1	1 806,0	1 816,4	1 816,4
1997	466,3	466,2	466,2	1 441,9	1 451,3	1 451,4	1 848,9	1 860,4	1 860,5
1998	470,2	470,5	470,5	1 500,8	1 509,9	1 509,9	1 846,9	1 858,4	1 858,4
1999	470,9	471,5	471,6	1 532,4	1 543,7	1 544,0	1 877,7	1 894,3	1 894,4
2000	473,2	473,7	473,5	1 574,8	1 583,8	1 584,0	1 918,8	1 931,2	1 931,3
2001	459,9	460,4	460,3	1 620,3	1 630,9	1 630,9	1 908,7	1 921,5	1 921,6
2002	467,3	468,6	468,3	1 658,3	1 670,8	1 670,8	1 948,3	1 965,1	1 965,0
2003	475,3	476	476,1	1 706,6	1 716,7	1 716,7	2 001,2	2 014,5	2 014,7
2004	479,7	479,8	479,7	1 748,6	1 757,4	1 757,5	2 052,4	2 062,7	2 062,7
2005	483,6	483,4	483,5	1 777,6	1 784,3	1 784,4	2 122,0	2 130,5	2 130,5
2006	491,4	491,5	491,6	1 864,5	1 870,6	1 870,7	2 179,2	2 195,5	2 195,5
2007	501,2	502	501,8	1 953,0	1 959,4	1 959,4	2 255,7	2 266,4	2 266,3
2008	512,3	513	512,7	2 006,6	2 013,2	2 013,3	2 299,2	2 314,2	2 314,3

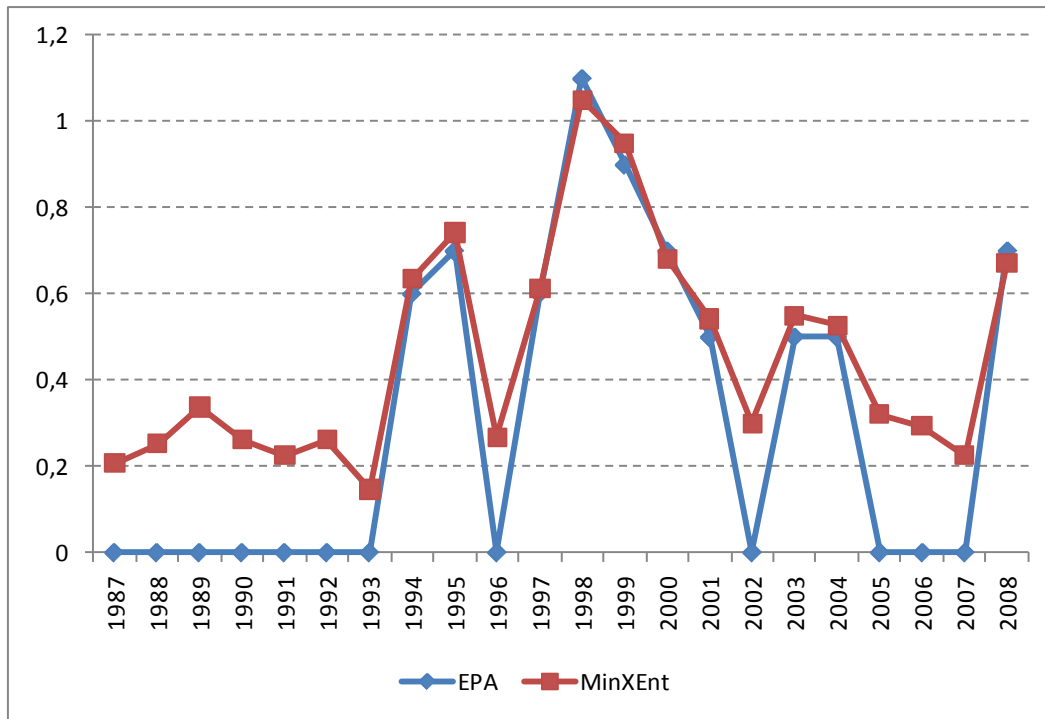




**Figure 3 : Évolution de l'emploi du SP « Services d'utilité publique - 22 » pour la Côte-Nord de la Nouvelle-Écosse, 1987-2008**



**Figure 4 : Évolution de l'emploi du SP « Extraction minière et extraction de pétrole et de gaz - 21 » pour la région de Montréal, 1987-2008**



**Figure 5 : Évolution de l'emploi du SP « Services d'utilité publique - 22 » pour Parklands Nord (Manitoba), 1987-2008**

Le quatrième constat concerne les données d'emploi pour les RE de la Colombie-Britannique entre 1987 et 1994. Les données d'emploi provenant de l'EPA ne sont pas disponibles pour ces RE bien que le total soit disponible pour l'ensemble de la province. En utilisant les données de répartition de 1995 comme distribution a priori et en retirant les contraintes liées aux règles d'arrondissements et de confidentialité pour le RE de cette province, il est possible d'obtenir une nouvelle distribution par année pour chaque SP et chaque RE (Figure 6). Bien que sensible à la distribution a priori sélectionnée<sup>12</sup>, l'approche a l'avantage de fournir des données d'emploi pour l'ensemble des années.

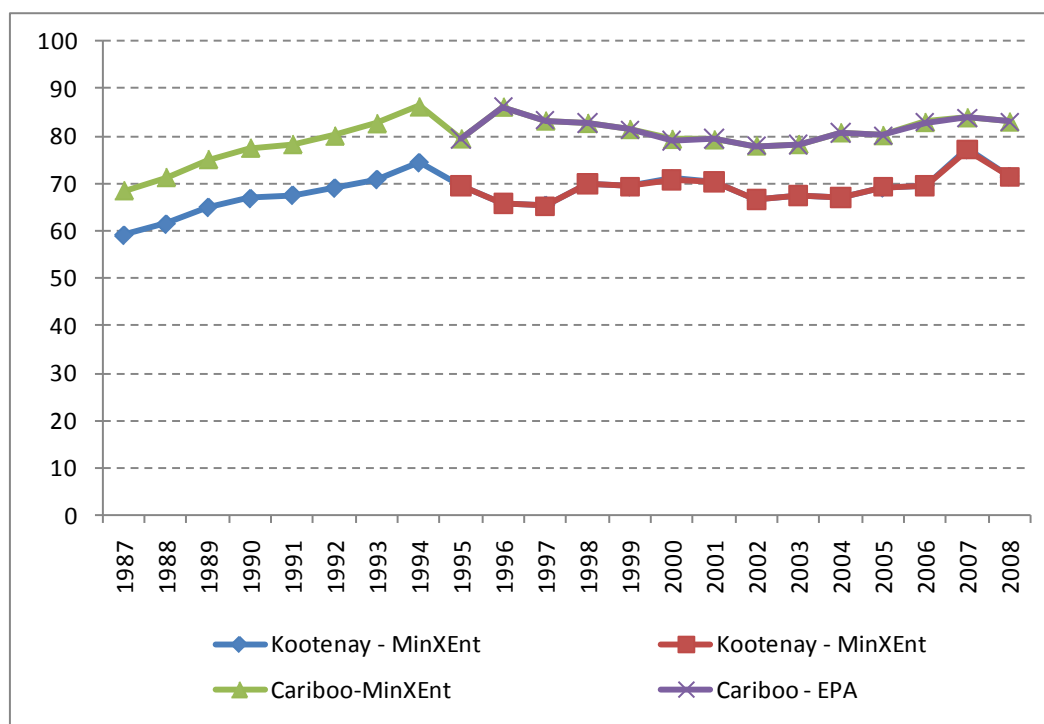
Au final, l'application de la méthode MinXEnt permet d'obtenir une nouvelle base de données de type panel qui respecte les contraintes des totaux par SP par province et qui élimine les problèmes d'arrondissements ainsi que ceux de confidentialités. À défaut de pouvoir utiliser les données « réelles »<sup>13</sup> de l'emploi, il s'agit d'une estimation qui soit intéressante globalement.

<sup>12</sup> L'approche suppose que la distribution de l'emploi en 1995 reflète bien la distribution de l'emploi pour les années manquantes antérieures.

<sup>13</sup> Évidemment, la donnée réelle d'emploi n'est pas disponible à moins d'effectuer un recensement de l'emploi. Nous considérons, pour le bien de l'exercice, que le nombre d'emplois de la province, pour les SP, est le nombre réel d'emplois. Les données de l'EPA ont l'avantage d'être rajustées tous les cinq ans afin que les estimations reflètent la réalité des recensements. Une répartition de l'emploi par SP serait, fort probablement, plus juste avec les données de l'Enquête sur la rémunération et les heures de travail (EERH) qui est basée sur des fichiers administratifs. Cependant, l'EERH a le désavantage d'exclure, a priori, les travailleurs autonomes, diminuant ainsi le nombre d'emplois.

Évidemment, la volatilité des distributions demeure un problème entier qui ne peut être réglé simplement avec l'application de la méthode de façon indépendante chaque année. De plus, certaines valeurs demeurent basées sur la distribution d'une année donnée.

Une application plus globale permettant de considérer à la fois les dimensions temps et espace serait potentiellement une approche intéressante, mais plus complexe étant donné que la matrice à estimer serait à trois dimensions (régions économiques, secteurs productifs et temps). À défaut de pouvoir développer une telle approche, une méthode de lissage des données chronologiques, permettant d'éliminer une partie de la volatilité des séries, peut être intéressante. Cependant, une telle approche ne peut assurer une cohérence avec les données globales de l'EPA, ce qui constitue une faiblesse inhérente et représente une limite à son utilisation. C'est d'ailleurs pour cette raison que les résultats n'ont pas été ajustés pour cette volatilité temporelle des séries et que les résultats sont présentés à partir des ajustements spatiaux uniquement. Les résultats sont donc les meilleurs qui sont possibles d'être obtenus.



**Figure 6 : Évolution du nombre d'emplois total pour deux RE de la Colombie-Britannique, 1987-2008**



## Conclusion

Les données d'emploi provenant de l'Enquête sur la population active (EPA) permettent d'étudier l'évolution de l'emploi au Canada pour un ensemble prédéterminé de régions géographiques, en l'occurrence les 69 régions économiques (RE) définies par Statistique Canada. Cependant, certaines faiblesses sont inhérentes à cette source de données. La première est que la distribution du nombre d'emplois par secteur productif (SP) par région économique (RE) est différente de la distribution de l'emploi par SP et par province. Des règles de confidentialité et d'arrondissements font en sorte qu'il est impossible d'avoir des distributions cohérentes lorsqu'on s'intéresse à des géographies de taille plus fine. La seconde est une présence importante de valeurs *zéros*, fragilisant ainsi le calcul des taux de croissance de l'emploi par SP pour certaines RE.

Une façon de contourner ces deux faiblesses consiste à appliquer la méthode de minimisation de l'entropie croisée (MinXEnt) qui permet, à la lumière des données existantes, d'obtenir une nouvelle distribution de l'emploi par SP et par SE qui satisfait aux différentes contraintes fixées. Cette méthode permet d'obtenir, pour chacune des années, une distribution de l'emploi qui satisfait les répartitions marginales de l'emploi par SP par province en plus d'éliminer le problème des données manquantes. L'application de la méthode MinXEnt a également l'avantage de fournir une estimation du nombre d'emplois pour certaines RE où ces données ne sont pas disponibles pour certaines années (Colombie-Britannique). Au final, l'objectif est de montrer qu'il est possible, à partir des données de sondage qui sont irréconciliables au départ, d'obtenir une nouvelle distribution spatiale et temporelle de l'emploi qui soit cohérente aux estimations globales.

Cependant, il laisse encore certaines questions en suspend. La méthode MinXEnt ne peut régler la totalité des problèmes liés à l'EPA. La volatilité des données d'enquête, même à fréquence annuelle, ne peut être réglée avec l'application de cette méthode de façon indépendante à la dimension temps. Une application plus globale permettant de considérer à la fois les dimensions temps et espace serait potentiellement une approche intéressante, mais qui reste à développer.



## Bibliographie

- Dubé, J. 2003. *Estimation des flux d'échanges interrégionaux par la méthode de minimisation de l'entropie-croisée*. Mémoire de maîtrise codirigé par A. Lemelin et B. Decaluwé, Québec : Université Laval, 106 p.
- Dubé, J. et Dupéré, M. 2006. Une méthode de répartition pour déduire les flux d'échanges économiques, *Bulletin économique du transport au Québec*, Ministère des Transports du Québec, juillet, 37 : 24-28.
- Dubé, J. et Dupéré, M. 2004. *Expérimentation de la méthode d'entropie croisée pour l'estimation d'échanges économiques à partir de flux de transport*, Collection Études et recherches en transport, ministère des Transports du Québec, 38 p.
- Dubé, J. et Lemelin, A. 2005. Estimation expérimentale des flux d'échanges interrégionaux par la méthode de minimisation de l'entropie croisée, *Revue canadienne de sciences régionales*, 28(3) : 513-534.
- Kapur, J.N. et Kesavan, H.K. 1992. *Entropy Optimization principles with applications*, Academic Press, Inc., San Diego, CA, 405 p.
- Kullback, S. 1959. *Information Theory and Statistics*, John Wiley & Sons Inc., Canada, 395 p.
- Kullback, S. et Leibler, R. 1951. On information and sufficiency, *Annals of Mathematical Statistics*, 22 : 79-86.
- Statistique Canada. 2009. *Produits et services de l'Enquête sur la population active*, no. 71-544-X au catalogue, 19 p.
- Statistique Canada. 2008. *Méthodologie de l'Enquête sur la population active au Canada*, no. 71-526-X au catalogue, 122 p.
- Theil, H. 1967. *Economics and information theory*, Rand McNally & Company, Chicago, Studies in mathematical and managerial economics, 7, 488 p.
- Theil, H. 1971. *Principles of econometrics*, A Wiley/Hamilton Publications, John Wiley & Sons, Inc., 736 p.