

Série d'apports énergétiques

Mandat confié en 2005 au groupe d'experts

Résumé et rapports des experts

Novembre 2005

Référence: CD Comité des experts 2005

Contexte et mandat 2005

HQP sommaire

P. Rasmussen (rapport), Université du Manitoba

M. Slivitzky et A.C. Favre (rapport), INRS

J.F. Angers (rapport), Université de Montréal

Constitution d'un Comité d'expert en hydrologie et en statistiques pour évaluer la série d'apports énergétiques 1943-2004 d'Hydro-Québec

Contexte

En 1998 et en 2001, le comité d'experts mandaté par Hydro-Québec pour revoir les données historiques d'apports énergétiques a conclu que la moyenne historique des apports énergétiques de 1943 à 2000 demeurait le meilleur indicateur à long terme de la prévision des apports énergétiques. Compte tenu de l'ajout des années 2001 à 2004 à la série de données disponibles en 2000 et du fait que les méthodes d'analyse statistique ont continué de progresser, Hydro-Québec souhaite consulter des experts pour valider la série chronologique 1943-2004 et déterminer si la moyenne historique des apports énergétiques demeure le meilleur indicateur des apports énergétiques à long terme.

Mandat du groupe de travail

Déterminer, sur la base des données historiques des apports énergétiques 1943-2004,

- Si, outre le changement de moyenne résultant de l'ajout des données de 2001 à 2004, la moyenne historique des apports énergétiques calculée sur la base des données historiques 1943-2004 reste toujours le meilleur indicateur de la moyenne à long terme des apports énergétiques annuels.
- Quel est le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2% (écart actuellement utilisé -64 TWh sur deux ans et -98 TWh sur quatre ans).

Conclusion d'Hydro-Québec Production

Novembre 2005

En 2005, Hydro-Québec a mandaté des experts en hydrologie et en statistique pour évaluer la chronique 1943-2004 des apports énergétiques et déterminer, sur la base des données historiques des apports énergétiques 1943-2004,

- si, outre le changement de moyenne résultant de l'ajout des données de 2001 à 2004, la moyenne historique des apports énergétiques calculée sur la base des données historiques 1943-2004 reste toujours le meilleur indicateur de la moyenne à long terme des apports énergétiques annuels
- quel est le «déficit» par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2 % (écart actuellement utilisé - 64 TWh sur deux ans et -98 TWh sur quatre ans)

Ces experts ont formalisé la chronique sous forme de différents modèles statistiques dont ils se sont servis pour comparer leurs résultats à l'ensemble ou à différentes périodes de la chronique et estimer les déficits cumulés d'apports énergétiques.

Parmi les conclusions des experts, nous retenons que :

- la moyenne de la chronique 1943-2004 des apports énergétiques est un bon indicateur à long terme (pour les cinq à dix prochaines années) de la valeur énergétique des apports
- les déficits cumulés associés à une probabilité de dépassement de 2% (64 TWh sur deux ans et 98 TWh sur quatre ans) sont prudents et comparables aux résultats obtenus par les différents modèles statistiques utilisés pour représenter en tout ou en partie la chronique 1943-2004 des apports énergétiques
- le choix d'un modèle statistique sur la base des données disponibles reste un exercice difficile étant donné la faible taille de l'échantillon (62 ans) et les imprécisions inhérentes aux observations

A la lumière de ce qui précède, HQP conclut que la valeur moyenne de la chronique 1943-2004 représente le meilleur indicateur des apports énergétiques à long terme et que l'estimé des déficits cumulés d'apports énergétiques de 64 TWh sur deux ans et de 98 TWh sur quatre ans est encore valable.

Analyse des apports énergétiques 1943-2004 d'Hydro-Québec

Peter Rasmussen, Ph.D., P.Eng.
175 King's Drive
Winnipeg MB
R3T 3G3

(204) 474-6960
rasmusse@cc.umanitoba.ca

Le 26 août 2005

Introduction

Ce rapport résume mes principales conclusions concernant la modélisation statistique des apports énergétiques de 1943 à 2004 ainsi que leur prévision pour la période de 2005 à 2009. Plus particulièrement, je vais essayer de répondre aux deux questions suivantes qui nous ont été posées par Hydro-Québec:

1. Est-ce que la moyenne historique (1943-2004) est le meilleur indicateur de la moyenne (la prévision) pour les cinq à dix prochaines années?
2. Quel est le déficit par rapport à la moyenne associé à une probabilité de dépassement égale à 2%?

La série d'apports énergétiques considérée pour cette analyse apparaît dans la figure 1 de l'annexe. De manière visuelle, on observe trois régimes différents, correspondant approximativement aux périodes 1943-1964, 1965-1983 et 1984-2004. En particulier, on peut noter que la période 1965-1983 était en moyenne beaucoup plus humide que les deux autres périodes. La réponse aux deux questions dépend étroitement des hypothèses relatives à ces régimes. Plus spécifiquement, pour répondre aux questions on doit d'abord se demander quel modèle statistique peut décrire adéquatement les régimes observés.

Mon analyse est basée sur les hypothèses suivantes:

1. Les données ne contiennent pas d'erreurs autres que celles dues à la reconstitution de certaines données.
2. La série est stationnaire.

Ces hypothèses sont importantes et nécessaires pour notre analyse. Les premières données de la série ont été partiellement reconstituées et contiennent donc des incertitudes (erreurs). Cependant, je considère que les données contiennent de l'information pertinente et pour cette raison doivent être incluses dans l'analyse malgré les erreurs inévitables dues à la reconstitution. Il importe de souligner que l'on suppose qu'il n'y a pas d'erreurs *systématiques*. La stationnarité implique entre autre que la moyenne et l'écart-type du processus naturel n'ont pas changé au cours de la période, ce qui n'empêche pas d'observer de longues périodes de faible ou de forte hydraullicité.

Dans mon rapport de 2001, j'avais recommandé de distinguer la moyenne globale et la moyenne locale de la série. Dans ce contexte, la moyenne globale est la moyenne à long terme (la moyenne du processus stationnaire) qu'on pourrait estimer directement à partir de la série observée et la moyenne locale est le *niveau typique* d'une période donnée, qu'on estimera par exemple à l'aide d'une régression non paramétrique ou par une moyenne mobile. Alors que la moyenne globale estimée comme la moyenne de la série observée constitue la meilleure estimation du niveau général des apports, la moyenne locale (moyenne des dernières années) serait peut-être un meilleur indicateur de la moyenne pour les cinq à dix prochaines années (question 1). Notre analyse préliminaire vise à répondre à cette question.

Analyse

Avant de répondre aux deux questions qui nous ont été posées, j'ai effectué quelques analyses dans le but de guider le choix du modèle statistique le plus adéquat pour représenter les données. Plus spécifiquement, j'ai considéré les sous-questions suivantes:

1. Quelle est la meilleure distribution des données?
2. Est-ce qu'un modèle de type bruit blanc ou de type AR(1) permet d'expliquer les séquences observées de faible et de forte hydraulicité?
3. Est-ce que la forme de la fonction d'autocorrélation observée s'explique par la présence de régimes secs et humides?
4. Y-a-t-il un modèle qui peut adéquatement représenter les régimes secs et humides?
5. Parmi un ensemble de possibilités, quel est le meilleur modèle de prévision?

Dans ce qui suit, je fais la synthèse des principales conclusions pour chaque question.

Distribution des données

La figure 2 de l'annexe montre que les apports énergétiques de 1943 à 2004 sont assez bien représentés par une distribution normale. Des tests statistiques plus rigoureux ne permettent pas de rejeter l'hypothèse d'une distribution normale. Cependant, on observe un certain écart dans la queue gauche. Cet écart peut être réduit si l'on adopte une distribution non paramétrique, estimée par exemple avec la méthode des noyaux. La figure 3 montre l'ajustement obtenu avec la méthode non paramétrique des noyaux. Chacune des deux hypothèses – normale et non paramétrique – peut être justifiée. On compare plus tard les résultats obtenus avec les deux méthodes pour l'estimation du déficit de période de retour de cinquante ans.

Évaluation des modèles bruit blanc et AR(1)

Dans l'analyse des données hydrologiques annuelles, les modèles de type bruit blanc et AR(1) sont souvent préconisés. Ces modèles sont relativement simples et sont appropriés lorsqu'une série ne contient aucune persistance (bruit blanc) ou une persistance à court terme (AR(1)). Avant de considérer des modèles plus complexes, on devrait d'abord démontrer l'inadéquation des modèles simples.

Chacun des deux modèles a été estimé en utilisant des techniques conventionnelles. On a ensuite procédé à la génération d'un grand nombre de séries de longueur 62 (nombre d'années d'observation). Pour chaque série, on peut calculer le déficit maximum et le surplus maximum pour différentes durées. À partir des séries simulées, on peut alors déterminer la probabilité des déficits et des surplus observés. Si un modèle mène à la conclusion que les déficits ou surplus observés sont très peu probables ($<10\%$), on doit rejeter le modèle.

Les figures 4 à 7 montrent les résultats de l'analyse pour les modèles bruit blanc et AR(1) sous forme de boxplot et de tableaux de probabilité. On peut noter qu'il est peu probable qu'un bruit blanc soit à l'origine des observations, car les surplus sur 15 ans et sur 20 ans sont peu probables. Un surplus sur 15 ans plus extrême que celui observé ne se produit qu'avec une probabilité de 7% alors que le surplus observé sur 20 ans ne se produit qu'avec une probabilité de 3%. On rejette donc l'hypothèse d'un bruit blanc. Le modèle AR(1) est un peu plus plausible selon les critères considérés, bien que la probabilité du surplus sur 20 ans reste assez faible (11%).

Fonction d'autocorrélation

La fonction d'autocorrélation est donnée dans la figure 8 en annexe. On observe des autocorrélations positives d'ordre 1 et 2, alors que des corrélations négatives sont observées pour les ordres de 16 à 25. On peut se demander si ces valeurs sont le résultat des trois régimes différents dont on a fait mention plus haut.

Afin de répondre à cette question, on a de nouveau fait appel à la simulation. On a généré 2000 séries de longueur 62. Pour chaque série, les 22 premières données ont été tirées d'une loi normale de moyenne et écart-type déterminés à partir des observations de 1943 à 1964, les 19 données suivantes ont été tirées d'une loi normale de moyenne et écart-type représentatives de la période 1965-1983, et les 21 dernières données d'une loi normale représentative de la période 1984-2004. Pour chaque série générée, on a calculé la fonction d'autocorrélation. On a finalement pris la moyenne des 2000 fonctions d'autocorrélation. Le résultat est donné dans la figure 8. On conclut que l'hypothèse de trois régimes différents est tout à fait en accord avec la fonction d'autocorrélation observée.

Modèle pour représenter les régimes secs et humides

Afin de mieux représenter les séquences de faible et de forte hydraulicité, nous avons considéré le modèle Markov caché. Dans ce modèle, on fait l'hypothèse d'une chaîne de Markov à deux états. Il est possible de considérer plus que deux états mais pour le cas actuel, la quantité de données ne le justifie pas. Les deux états représentent les régimes secs et humides. L'état n'est pas explicitement observé, d'où le nom Markov caché. Selon les hypothèses du modèle, l'observation pour une année donnée est tirée d'une loi normale dont les paramètres dépendent de l'état de l'année. Pour un modèle à deux états, l'ensemble des observations est donc une combinaison de tirages de deux lois normales. La persistance des régimes secs et humides est déterminée par la matrice de probabilité de transition associée à la chaîne de Markov caché. Le modèle Markov caché contient ainsi huit paramètres (2 moyennes, 2 écarts-types et 4 probabilités de transition) qui peuvent être estimés à l'aide de la technique EM (Expectation Maximization).

Les figures 9 et 10 montrent les résultats de l'analyse des déficits et des surplus obtenus avec le modèle Markov caché. Les valeurs observées sont généralement très raisonnables selon ce modèle, à part la moyenne minimum d'une durée de trois ans. On note d'ailleurs dans la série historique la présence de trois années consécutives de très faible hydraulicité (1961-64).

Le modèle Markov caché sera considéré dans ce qui suit comme une alternative au modèle AR(1). Il est difficile de conclure qu'un des modèles est meilleur que l'autre.

Comparaison des modèles de prévision

Afin d'effectuer la prévision la plus précise possible, j'ai comparé la performance de différents modèles statistiques de prévision. La comparaison a été basée sur la racine carrée de l'erreur quadratique moyenne de prévision (REQM). Avec chaque modèle, j'ai effectué une prévision de l'apport énergétique de chaque année de la période 1948-2004 en utilisant seulement l'information du passé. La REQM est calculée comme suit:

$$REQM = \left[\frac{1}{57} \sum_{i=1948}^{2004} (\hat{A}_i - A_i)^2 \right]^{1/2}$$

ou $\hat{A}_i$ est la prévision pour l'année i et A_i est l'observation pour la même année. La sommation commence en 1948 afin de pouvoir utiliser les cinq valeurs précédentes pour toutes les années de prévision.

J'ai considéré les modèles suivants:

1. Bruit blanc. La prévision issue de ce modèle est égale à la moyenne observée:

$$\hat{A}_i = 189.9$$

2. Processus autorégressif d'ordre 1. La prévision dépend de l'apport de l'année précédente:

$$\hat{A}_i = 189.9 + 0.21(A_{i-1} - 189.9)$$

3. Moyenne mobile. La prévision est basée sur la moyenne de l'année précédente ou des cinq années précédentes:

$$\hat{A}_i = \frac{1}{k} \sum_{j=1}^k A_{i-j} \quad \text{avec } k=1 \text{ ou } k=5$$

4. Méthode des k voisins les plus proches. Pour la prévision de l'année i , on trouve les k années qui sont les plus similaires (voisin) à l'année $(i-1)$ puis on prend la moyenne des années qui suivent les k voisins. On considère k égal à 1 et à 5.
5. Modèle markov caché à deux états. Les paramètres du modèle sont d'abord estimés en utilisant toutes les observations. Pour faire la prévision de l'année i , on considère que seulement les données du passé sont connues. Il faut aussi déterminer la probabilité qu'on soit en état 1 ou en état 2 pendant l'année $i-1$. Il existe des techniques pour déterminer ces probabilités que nous appelons

respectivement α_1 et α_2 . La prévision est une fonction des probabilités α_1 et α_2 et les paramètres du modèle Markov caché:

$$\alpha_1(i-1) = \Pr\{S_{i-1} = 1 \mid x_1, x_2, \dots, x_{i-1}\}$$

$$\alpha_2(i-1) = \Pr\{S_{i-1} = 2 \mid x_1, x_2, \dots, x_{i-1}\}$$

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

$$\hat{A}_i = [\alpha_1(i-1)p_{11} + \alpha_2(i-1)p_{21}] \mu_1 + [\alpha_1(i-1)p_{12} + \alpha_2(i-1)p_{22}] \mu_2$$

Le résultat de l'analyse est donné dans le tableau suivant:

Modèle	REQM	Rang
Bruit blanc	19.8	3
AR(1) - normale	19.2	2
Valeur la plus récente	24.3	6
Cinq valeurs les plus récentes	21.0	4
Voisin le plus proche	28.2	7
Cinq voisins les plus proches	21.3	5
Markov caché à deux états	19.0	1

Selon le tableau, le modèle de Markov caché à deux états donne les meilleurs résultats, suivi par le modèle AR(1). Cependant la différence entre ces deux modèles est faible.

Conclusion de l'analyse préliminaire

Les modèles Markov caché et AR(1) seront retenus pour la prévision. On retient également le modèle bruit blanc afin de pouvoir comparer les résultats de plusieurs modèles.

Question 1: Est-ce que la moyenne historique (1943-2004) est le meilleur indicateur de la moyenne (la prévision) pour les cinq à dix prochaines années?

Pour la prévision, j'ai considéré les trois modèles retenus suite à l'analyse préliminaire. L'utilisation de plusieurs modèles permet de mieux apprécier l'impact du choix de modèle sur les conclusions.

Année	2005	2006	2007	2008	2009
Bruit blanc - normal	189.9	189.9	189.9	189.9	189.9
Bruit blanc - non param.	189.9	189.9	189.9	189.9	189.9
AR1	193.5	190.6	190.0	189.9	189.9
Markov caché*	187.9	188.6	188.7	189.0	189.1

* Obtenu par simulation

Sous l'hypothèse d'un processus de bruit blanc, l'écart-type de l'estimation de la moyenne est 2.5 TWh. La différence entre les modèles est donc généralement inférieure à l'incertitude sur l'estimation de la moyenne même.

De manière générale, on recommande l'utilisation d'une moyenne adaptative (régression non paramétrique, moyenne mobile, Markov caché) pour la prévision. Cependant, pour la période 2005-2009 il y a peu de différence entre les modèles. Il est donc acceptable d'utiliser la moyenne historique comme le meilleur indicateur de la moyenne pour les cinq prochaines années. Cette conclusion est également valable sur un horizon de 10 ans puisque l'estimation de la moyenne sur cet horizon, quelque soit le modèle privilégié parmi les quatre retenus, ne diffère pas significativement de la moyenne des apports énergétiques de la série historique (1943-2004).

Question 2: Quel est le déficit par rapport à la moyenne associé à une probabilité de dépassement égale à 2%?

Pour l'estimation des déficits associés à une probabilité de dépassement égale à 2%, j'ai considéré les mêmes modèles que dans la section précédente. Le premier tableau montre les résultats obtenus si on ne tient pas compte du passé. Une période de deux ans (par exemple) représente alors une période arbitraire du futur et non nécessairement 2005-06.

Période	1 an	2 ans	3 ans	4 ans
Bruit blanc –normale	-40.8	-57.6	-70.6	-81.5
Bruit blanc –non param. **	-40.5	-60.0	-74.4	-85.6
AR(1) – normale**	40.6	62.9	80.8	95.2
Markov caché **	38.6	57.5	73.3	87.6

*) Analytique

**) Obtenu par simulation

On peut observer que les valeurs actuelles utilisées par Hydro-Québec (-64TWh sur deux ans et -98TWh sur quatre ans) sont plutôt pessimistes. On peut également noter l'importance de l'autocorrélation lorsqu'il s'agit des déficits extrêmes. Les valeurs obtenues par le modèle AR(1) sont assez proches des valeurs utilisées actuellement et, parmi les modèles considérés ici, représentent une estimation conservatrice des déficits.

On souligne que les valeurs données dans le tableau ne sont pas associées à des années particulières. Il est possible d'établir des estimations de déficits plus précises pour les années 2005 à 2008 en prenant en considération l'état actuel du système. Pour les modèles de type bruit blanc, l'état actuel n'influence pas sur les valeurs. Pour les modèles AR(1) et Markov caché les résultats changent tel qu'indiqué dans le tableau suivant:

Période	2005	2005-06	2005-07	2005-08
AR(1) – normale*	39.9	59.7	75.4	89.6
Markov caché*	36.5	58.2	75.6	89.8

*) Obtenu par simulation

Le choix final des valeurs à utiliser demeure celui du décideur. S'il faut cependant se prononcer ici, on propose de considérer les résultats obtenus avec le modèle AR(1) qui sont conservateurs et plus proches des valeurs utilisées actuellement par Hydro-Québec.

Annexe

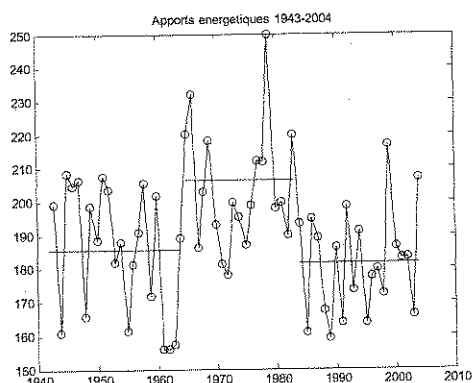


Figure 1. Série d'apports énergétiques

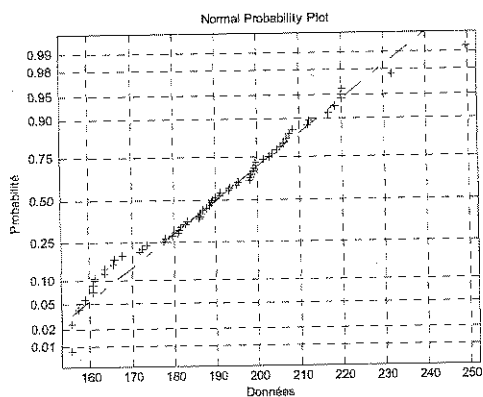


Figure 2. Distribution des apports énergétiques

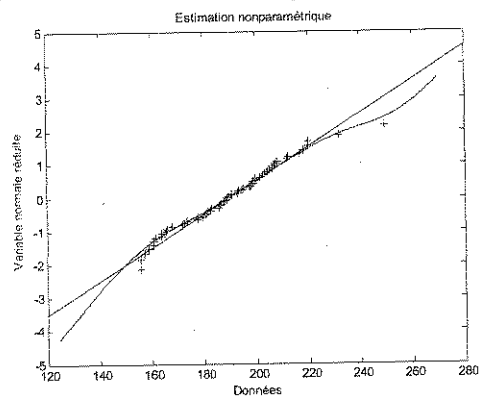
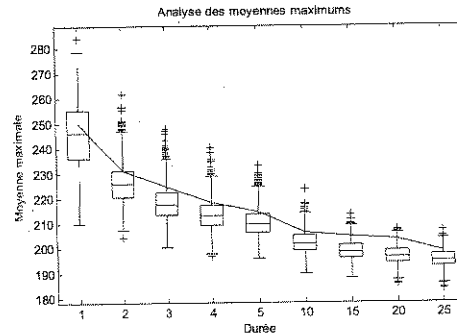
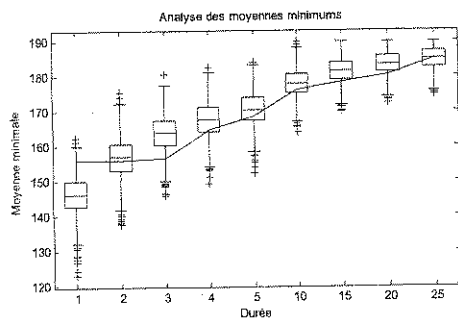


Figure 3. Distribution non paramétrique



Figures 4 et 5. Distribution des moyennes minimums et maximums sous l'hypothèse d'un bruit blanc.

Minimums – bruit blanc

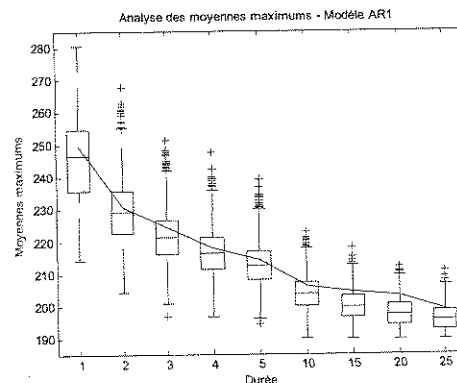
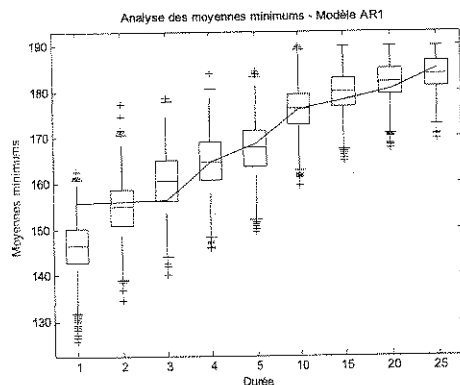
Durée	1	2	3	4	5	10	15	20	25
Valeur historique	155.8	155.9	156.4	164.5	168.5	176.1	178.1	180.3	184.9
Prob. de non-dépass.	97.7	43.2	10.0	27.6	34.8	32.8	19.5	20.4	50.4

Les probabilités inférieures à 10% sont considérées significatives.

Maximums – bruit blanc

Durée	1	2	3	4	5	10	15	20	25
Valeur historique	249.7	230.8	224.6	218.2	214.3	206.3	204.4	203.4	198.8
Prob. de non-dépass.	57.9	70.4	82.3	80.3	77.9	82.5	92.8	96.8	85.8

Les probabilités supérieures à 90% sont considérées significatives.



Figures 6 et 7. Distribution des moyennes minimums et maximums sous l'hypothèse d'un processus AR(1).

Minimums – AR(1)

Durée	1	2	3	4	5	10	15	20	25
Valeur historique	155.8	155.9	156.4	164.5	168.5	176.1	178.1	180.3	184.9
Prob. de non-dépass.	95.6	56.9	25.5	48.9	55.0	48.5	34.2	33.3	60.8

Les probabilités inférieures à 10% sont considérées significatives.

Maximums – AR(1)

Durée	1	2	3	4	5	10	15	20	25
Valeur historique	249.7	230.8	224.6	218.2	214.3	206.3	204.4	203.4	198.8
Prob. de non-dépass.	60.6	57.2	66.4	59.7	59.5	68.2	82.0	89.2	76.9

Les probabilités supérieures à 90% sont considérées significatives.

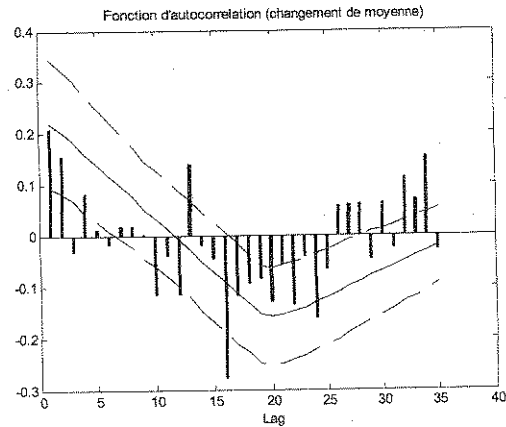
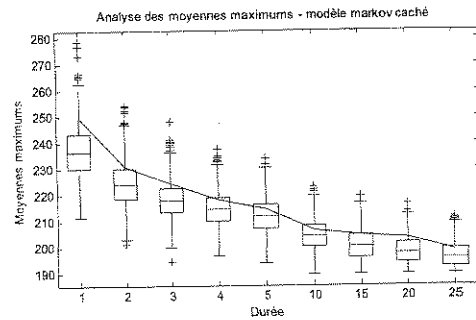
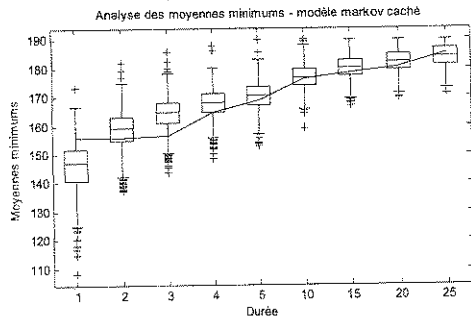


Figure 8. Fonction d'autocorrélation observée et simulée selon la procédure décrite dans le texte.



Figures 9 et 10. Distribution des moyennes minimums et maximums sous l'hypothèse d'un processus Markov caché.

Minimums – Markov caché

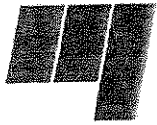
Durée	1	2	3	4	5	10	15	20	25
Valeur historique	155.8	155.9	156.4	164.5	168.5	176.1	178.1	180.3	184.9
Prob. de non-dépass.	89.5	31.3	8.9	23.7	35.9	46.2	34.2	34.1	62.5

Les probabilités inférieures à 10% sont considérées significatives.

Maximums – Markov caché

Durée	1	2	3	4	5	10	15	20	25
Valeur historique	249.7	230.8	224.6	218.2	214.3	206.3	204.4	203.4	198.8
Prob. de non-dépass.	89.5	78.5	81.3	72.1	65.2	64.5	75.4	84.7	72.2

Les probabilités supérieures à 90% sont considérées significatives.



Université du Québec

Institut national de la recherche scientifique

Eau, Terre et Environnement

**RAPPORT DE SYNTHÈSE
EXPERTISE
ÉVALUATION DE LA SÉRIE D'APPORTS
ÉNERGÉTIQUES 1943-2004 D'HYDRO-QUÉBEC**

Par

**Michel Slivitzky
Anne-Catherine Favre**

**Institut national de la recherche scientifique (INRS)
Centre Eau, Terre et Environnement
Université du Québec
490 de la Couronne, Québec (Québec) G1K 9A9**

Rapport d'expertise

2 novembre 2005

michel_slivitzky@ete.inrs.ca
anne-catherine_favre@ete.inrs.ca

TABLE DES MATIERES

1. Contexte du mandat	1
2. Analyse descriptive des données	1
3. Le changement de moyenne en 1985 est-il significatif ?	4
4. Que se passe-t-il en 1985 ?	5
5. Nos deux hypothèses	12
6. Deux modèles proposés	13
6.1 Mélange de deux lois normales	13
6.2 Loi normale après 1985	15
7. Réponses aux deux questions	16
7.1 Réponse à la question 1	16
7.2 Réponse à la question 2	16
8. Conclusions	17
9. Bibliographie	19

1. Contexte du mandat

En 1998 et en 2001, le comité d'experts mandaté par Hydro-Québec pour revoir les données historiques d'apports énergétiques a conclu que la moyenne historique des apports énergétiques demeurerait le meilleur indicateur à long terme de la prévision des apports énergétiques. Compte tenu de l'ajout des années 2001 à 2004 à la série de données disponible en 2000 et du fait que les méthodes d'analyse statistique ont continué de progresser, Hydro-Québec souhaite consulter des experts pour valider la série chronologique 1943-2004. Les deux questions spécifiques posées aux experts sont les suivantes :

1. Est-ce que, outre le changement de moyenne résultant de l'ajout des données de 2001 à 2004, la moyenne historique des apports énergétiques calculée sur la base des données historiques 1943-2004 reste toujours le meilleur indicateur de la moyenne à long terme (pour les 5-10 prochaines années) des apports énergétiques annuels ?
2. Quel est le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2% (écart actuellement utilisé -64 TWh sur deux ans et -98 TWh sur quatre ans) ?

2. Analyse descriptive des données

La figure 1 illustre la série temporelle des apports énergétiques annuels de 1943 à 2004.

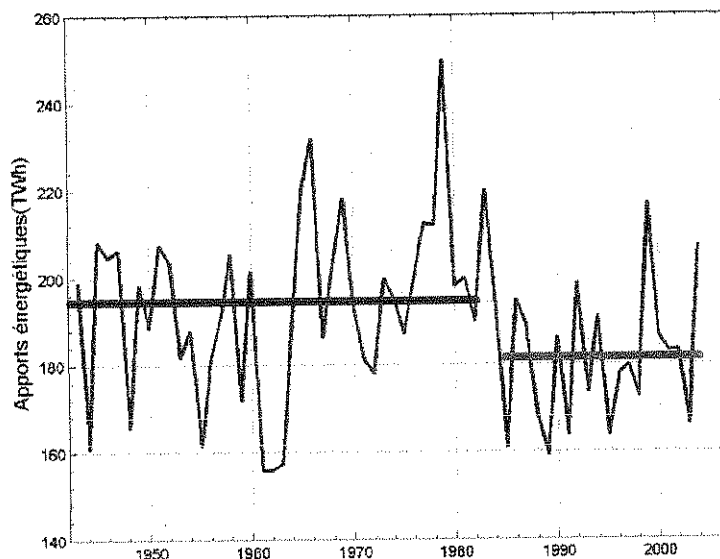


Figure 1 : Série temporelle des apports énergétiques annuels de 1943 à 2004. La moyenne de 1943 à 1984 est indiquée en rouge, alors que la moyenne de 1985 à 2004 est tracée en vert.

À première vue nous constatons une diminution à la fois de la moyenne et de la variabilité à partir de l'année 1985. Nous avons ainsi calculé quelques caractéristiques statistiques à la fois à partir de la série totale ainsi qu'en considérant les deux sous-séries 1943-1984 et 1985-2004. Le tableau 1 montre la moyenne, l'écart-type, le coefficient de variation (CV), le coefficient d'asymétrie et le coefficient d'aplatissement (kurtosis) pour les trois séries.

Tableau 1 : Caractéristiques statistiques (moyenne, écart-type, coefficient de variation (CV), coefficient d'asymétrie et coefficient d'aplatissement (kurtosis)) pour la série chronologique totale (1943-2004) ainsi que pour deux sous-séries (1943-1984) et (1985-2004).

série	moyenne	écart-type	CV	asymétrie	kurtosis
[1943,2004]	189.863	19.844	0.104	0.297	0.215
[1943,1984]	194.073	20.343	0.105	0.068	0.512
[1985,2004]	181.020	15.804	0.087	0.546	-0.155

Il est important, dans le but de modéliser les données, de tester leur normalité. Dans le tableau 1 nous constatons que les coefficients d'asymétrie et d'aplatissement sont proches de zéro, ce qui peut nous servir comme premier indicateur de normalité des données. Une inspection visuelle de la normalité a été effectuée par le biais de diagrammes quantiles-quantiles (« qqplot »). Les figures 2 à 4 montrent ces diagrammes basés sur la loi normale pour les trois séries. Dans les trois cas le centre des données est bien normalement distribué alors qu'une légère distorsion est visible dans la queue inférieure de la distribution. Plusieurs tests d'adéquation adaptés à la loi normale, tels Lilliefors, Shapiro-Wilks, Shapiro-Francia et d'Agostino ont été utilisés. Aucun de ces tests ne permet de rejeter l'hypothèse nulle de normalité.

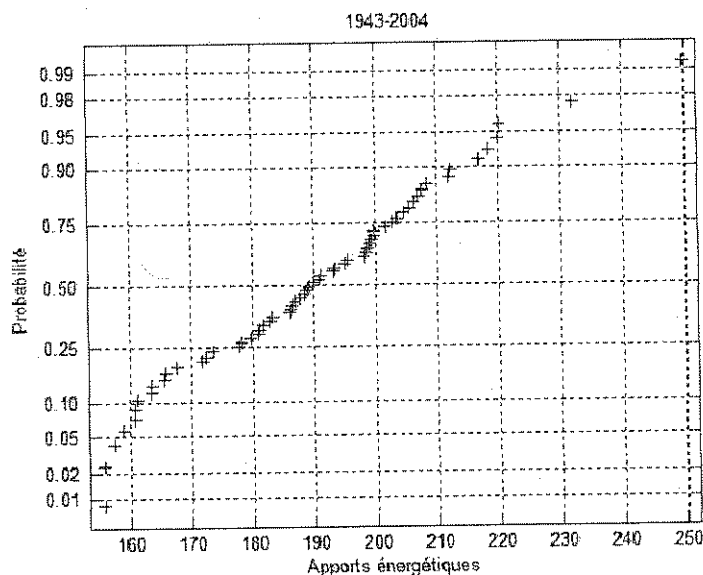


Figure 2 : Diagramme quantile-quantile (« qqplot ») basé sur la loi normale pour la série d'apports énergétiques 1943-2004.

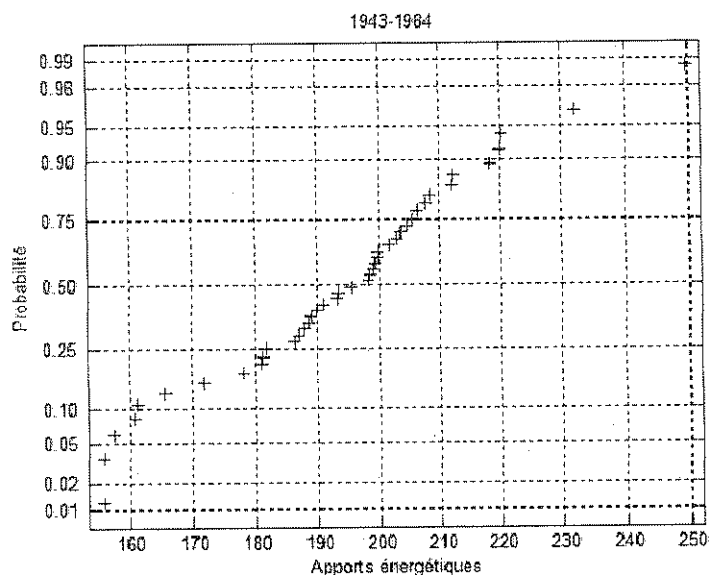


Figure 3 : Diagramme quantile-quantile (« qqplot ») basé sur la loi normale pour la série d'apports énergétiques 1943-1984.

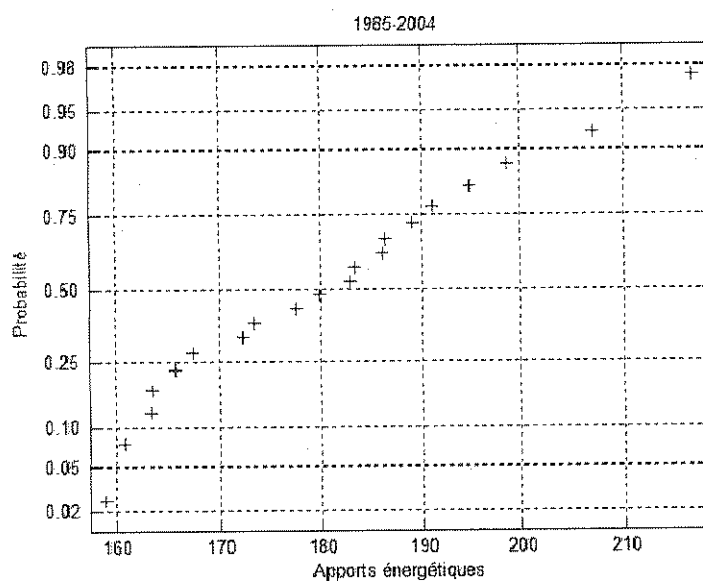


Figure 4 : Diagramme quantile-quantile (« qqplot ») basé sur la loi normale pour la série d'apports énergétiques 1985-2004.

3. Le changement de moyenne en 1985 est-il significatif ?

À la suite de cette analyse descriptive, il est légitime de se poser la question du caractère significatif du changement de moyenne en 1985. La figure 5 montre la moyenne mobile calculée sur 20 ans de la série des apports énergétiques.

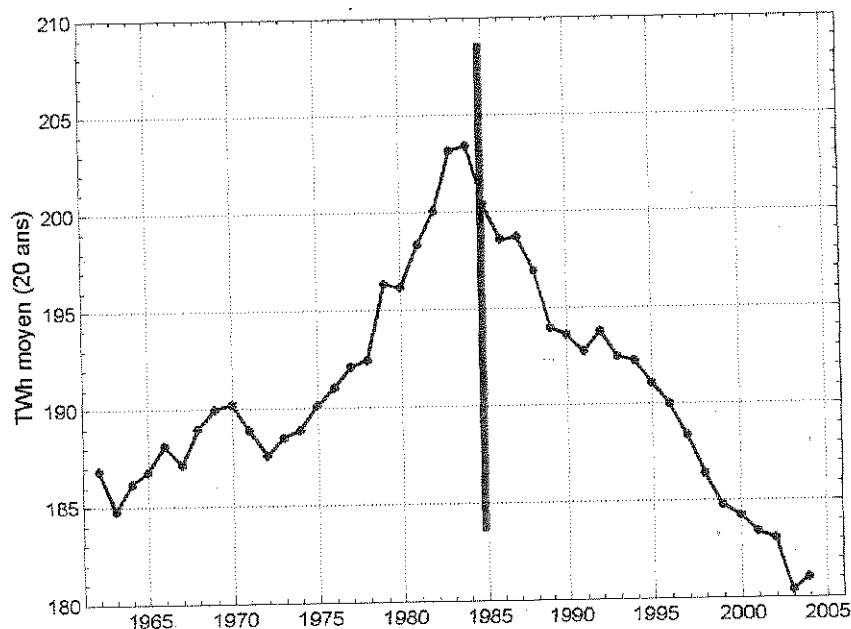


Figure 5 : Moyenne mobile sur 20 ans de la série d'apports énergétiques 1943-2004.

Il est frappant de constater que la moyenne mobile croît quasiment continûment jusqu'en 1984 puis elle décroît fortement. Nous déduisons du test du chi-carré que les variances avant et après 1985 ne sont pas significativement différentes. Nous avons déjà indiqué que les apports énergétiques sont normalement distribués, il est par conséquent possible d'utiliser la statistique t de Student afin de tester l'égalité des moyennes. Nous obtenons une p-valeur de 0.0071 avec le test unilatéral. Le test non paramétrique de Wilcoxon a également été appliqué. La p-valeur dans ce cas est de 0.0094. Nous concluons donc que la moyenne avant 1985 est significativement supérieure à la moyenne après cette date. La procédure de segmentation optimum de Hubert (2000) a été appliquée à ces données. Les résultats de cette approche sont résumés dans le tableau 2. Nous constatons qu'indépendamment du nombre de segments considérés, l'année 1985 constitue un pivot de transition.

Tableau 2 : Résultats de la procédure de segmentation optimum de Hubert en considérant une, deux, trois et quatre séparations.

Nombre de segments	Année début	Année fin
1	1943	2004
2	1943 1985	1985 2004
3	1943 1965 1985	1964 1984 2004
4	1943 1961 1964 1985	1960 1963 1984 2004

Le test bayésien de Lee et Heghinian (1977) a été ensuite appliqué en considérant la sous-période 1965-2004 (correspondant à la fusion des deux dernières sous-séries obtenues par la segmentation optimum en considérant trois séparations). La probabilité de changement obtenue est de 0.987 et l'intervalle de crédibilité à 95% pour la différence de moyenne est de [-34,-8.4]. La valeur zéro n'appartenant pas à cet intervalle, nous en déduisons qu'un changement de moyenne significatif a eu lieu durant cette période.

4. Que se passe-t-il en 1985 ?

Une littérature abondante, bien trop nombreuse pour être mentionnée ici, rapporte des changements importants, autour de l'année 1985, dans la circulation atmosphérique de l'hémisphère nord et de l'océan Atlantique nord. Ces modifications et changements pourraient être associés à la phase positive persistante de l'Oscillation Arctique (AO en anglais, Thompson et Wallace, 1998, 2001). La phase positive de l'AO (figure 6) se caractérise par un renforcement du vortex polaire de la surface à la basse stratosphère. Des vents froids soufflent sur l'Est Canadien alors que les perturbations de l'Atlantique apportent de la pluie avec des températures plus douces sur le Nord de l'Europe. Pendant la phase négative de l'AO (figure 7) de l'air froid continental pénètre dans le Midwest des États-Unis et l'on note des températures supérieures à la normale dans l'est du Canada.

Deux séries d'indices, couvrant les dernières années, peuvent être utilisées pour caractériser l'Oscillation Arctique pour la saison froide. La première (AO-1) est définie comme la moyenne des indices AO journaliers de janvier à mars (1950 – 2004) ; ces derniers étant calculés comme la première composante principale de l'anomalie des hauteurs géopotentiels de 1000 mb sur l'hémisphère nord (20°N - 90°N)¹. La deuxième série (AO-2) se définit comme la première composante principale des pressions mensuelles au niveau de la mer de décembre à mars (1899-

¹ Source : http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index

2004) sur l'hémisphère nord ($20^{\circ}\text{N} - 90^{\circ}\text{N}$)². La figure 8 montre ces deux indices sous forme de série temporelle pour la période 1950 à 2004. On remarque qu'ils sont très similaires et se déplacent vers des valeurs positives vers la fin des années 1980. On peut noter également que la série AO-1 présente des valeurs négatives relativement plus faibles.



Figure 6 : État de l'atmosphère lorsque l'indice AO est en phase positive³.

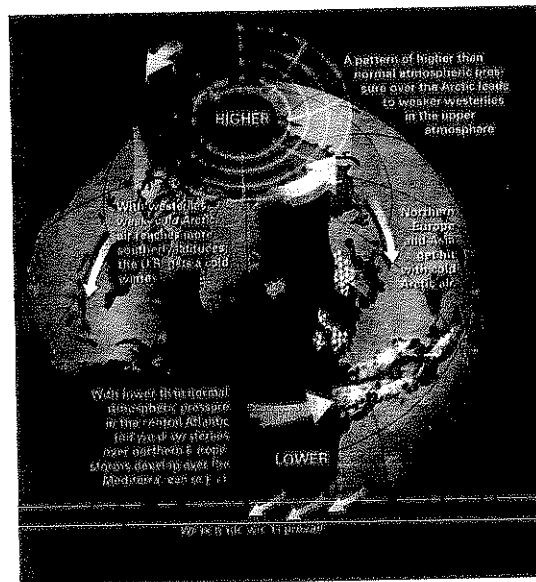


Figure 7 : État de l'atmosphère lorsque l'indice AO est en phase négative⁴.

² Source : <http://www.cgd.ucar.edu/cas/jhurrell/indices.info.html#nam>

³ Source : <http://jisao.washington.edu/data/aots/>

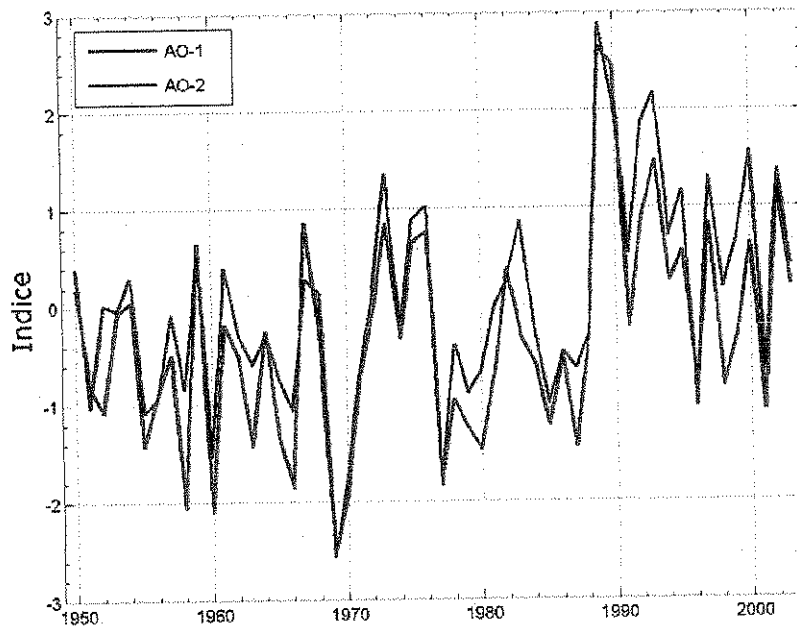


Figure 8 : Représentation des indices d'Oscillation Arctique AO-1 et AO-2 sous forme de série temporelle pour la période 1950 à 2004.

Thompson et Wallace (2001) ont analysé la fréquence des épisodes journaliers de l'indice AO normalisé pour les mois de janvier à mars de la période 1958 à 1997. Alors que pour toute la période, l'on note en moyenne 15 jours par hiver où l'indice était supérieur à 1.0 et 15 jours où il était inférieur à -1.0, pour la décennie 1958-1967, le nombre de jours où les indices étaient supérieur à +1.0 et inférieurs à -1.0 étaient respectivement 10.4 jours et 20.9 jours; pour la décennie 1988-1997 ces valeurs étaient respectivement 27.3 jours et 4.4 jours.

Les figures 9 et 10 illustrent des analyses sur les anomalies des précipitations et températures, pour les mois de novembre à avril couvrant la période 1950 à 1996, sur le continent nord américain pour des indices AO dans la phase positive. Alors que, dans l'est du continent, on remarque des anomalies moyennes de température de l'ordre de -1.0 °C à -1.5 °C (figure 10), pour les précipitations ces anomalies sont relativement beaucoup plus faibles de l'ordre de 0.5 à 0.8 cm/mois (figure 9).

⁴ Source : <http://jisao.washington.edu/data/aots/>

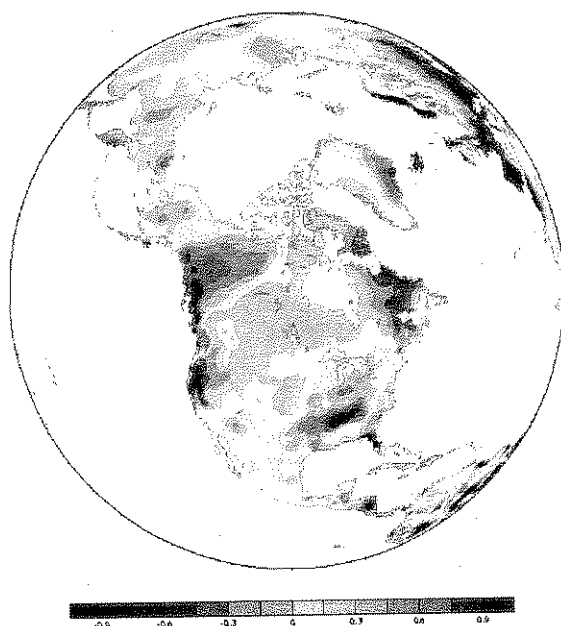


Figure 9 : Anomalie des précipitations (cm/mois) de novembre à avril pour la période de 1950 à 1996 lorsque l'indice AO est positif⁵.

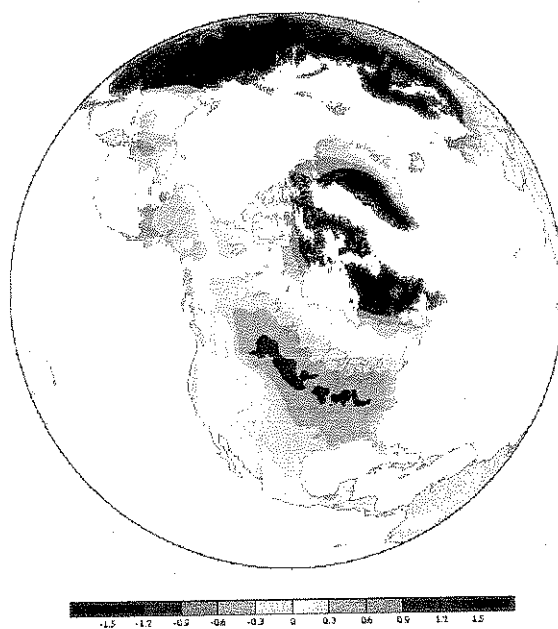


Figure 10 : Anomalie des températures (°C) de novembre à avril pour la période de 1950 à 1996 lorsque l'indice AO est positif⁶.

⁵ Source : <http://jisao.washington.edu/analyses0500/precipao.nam.gif>

⁶ Source : <http://jisao.washington.edu/analyses0500/sfctempao.nam.gif>

Peu d'analyses similaires ont été réalisées concernant les impacts dans l'est du Canada lorsque l'Oscillation Arctique est dans une phase négative ainsi que sur les mécanismes physiques correspondants. Thompson et Wallace (2001) montrent cependant que la climatologie du régime des pressions au niveau de la mer durant la phase négative est favorable à une circulation du nord-est avec une advection de masses d'air relativement chaudes et humides provenant de la mer du Labrador, alors qu'une phase positive est associée à une circulation provenant du nord-ouest avec advection d'air froid et sec. Déry et Wood (2004) s'appuient sur cet argument pour expliquer un lien entre l'Oscillation Arctique et les écoulements vers la baie d'Hudson et la baie d'Ungava. Dans une analyse des épisodes de blocage de la circulation sur l'Atlantique Nord, Huang et al. (2001) montrent que durant les épisodes négatifs de NAO l'on observe durant l'hiver 67% de plus de jours de blocage que durant les phases positives.

La figure 11 compare les écarts normalisés de l'indice AO-1 avec les écarts normalisés des apports énergétiques pour la période de 1950 à 2004. Nous remarquons un changement de régime vers 1985; en effet avant cette date les épisodes d'indice AO négatif correspondent à des apports énergétiques positifs alors qu'après cette date les épisodes d'AO positif correspondent à des apports énergétiques négatifs.

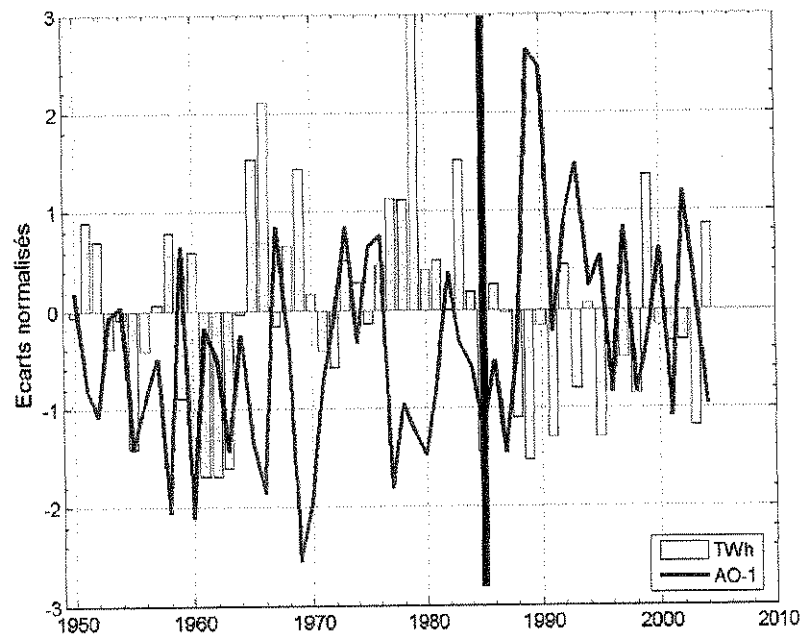


Figure 11 : Écarts normalisés de l'indice AO-1 et de la série énergétique pour la période 1950 à 2004.

Afin de préciser la nature des liens entre l'Oscillation Arctique et les apports énergétiques, nous avons utilisé l'approche des composites souvent utilisée dans le domaine des sciences atmosphériques. L'idée est de comparer la moyenne composite d'un certain nombre d'années

(non consécutives) correspondant aux valeurs extrêmes de l'indice avec la moyenne à long terme; un test statistique permet ensuite de vérifier dans quelle mesure la différence de ces deux valeurs est significative. Ainsi, par exemple, la différence de la moyenne annuelle composite des apports énergétiques correspondant aux six années extrêmes de la phase négative de l'indice AO-1 négatif avec la moyenne à long terme est de 21.2 TWh. Le test de Wilcoxon livre une p-valeur de 0.01; nous pouvons donc conclure que cette différence est significativement différente de zéro. Dans le tableau 3 nous présentons les différences des apports énergétiques moyens annuels pour les valeurs extrêmes des deux indices de AO, et pour un nombre d'années composites, variant de trois à dix, alors que dans le tableau 4 nous résumons les p-valeurs obtenues avec le test de Wilcoxon. Nous pouvons constater que les seules valeurs significativement différentes de zéro au seuil de 5% correspondent aux cinq à huit années composites avec les années extrêmes de l'indice AO-1 minimum; ces valeurs ont été surlignées en jaune dans ces deux tableaux.

Tableau 3 : Différence de production énergétique entre la moyenne composite et la moyenne à long terme pour les indices AO-1 et AO-2 maximum et minimum. Le nombre d'années composites considéré s'étend de trois à dix.

Nombre d'années composites								
	3	4	5	6	7	8	9	10
Indices AO maximum								
AO-1	-16.4	-13.7	-9.1	-8.1	-5.4	-5.9	-4.2	-5.5
AO-2	-16.4	-9.9	-8.5	-8.1	-5.4	-5.9	-8.1	-6.3
Indices AO minimum								
AO-1	19.2	15.4	20.8	21.2	19.4	16.9	11.5	7.6
AO-2	18.6	17.0	8.0	13.8	10.2	11.2	6.8	5.3

Tableau 4 : P-valeurs correspondant au test de Wilcoxon sur la différence de production énergétique entre la moyenne composite et la moyenne à long terme pour les indices AO-1 et AO-2 maximum et minimum. Le nombre d'années composites considéré s'étend de trois à dix.

Nombre d'années composites								
	3	4	5	6	7	8	9	10
Indices AO maximum								
AO-1	0.12	0.13	0.29	0.28	0.49	0.38	0.55	0.38
AO-2	0.12	0.32	0.33	0.29	0.50	0.39	0.23	0.36
Indices AO minimum								
AO-1	0.05	0.07	0.02	0.01	0.01	0.01	0.07	0.19
AO-2	0.09	0.06	0.26	0.10	0.23	0.13	0.34	0.47

Par la suite nous avons représenté les séries chronologiques des apports énergétiques et de l'indice d'Oscillation Arctique AO-1 en deux dimensions (échelle et temps) à l'aide d'ondelettes afin de représenter des structures (par exemple des périodicités locales intermittentes). Deux transformations ont été utilisées dans ce but : la transformation continue en ondelettes et la transformation croisée en ondelettes. La transformation continue s'obtient par convolution de la série temporelle considérée avec l'ondelette normalisée (dans notre cas nous avons utilisé l'ondelette de Morlet avec une fréquence adimensionnelle $\omega_0=6$) (Torrence et Compo, 1998). L'idée de la transformation croisée est de dégager des similarités entre deux séries temporelles. Elle s'obtient en multipliant la transformation continue de la première série temporelle par le complexe conjugué de la transformation continue de la deuxième série chronologique (Grinsted et al., 2004). Les figures 12 et 13 illustrent respectivement les transformations continues et croisées en ondelettes pour l'indice climatique AO-1 et la série énergétique pour la période 1950 à 2004. Dans ces figures, les contours en trait épais délimitent des zones significativement différentes d'un bruit rouge (au seuil de 5%) alors que les couleurs plus pâles indiquent le cône d'influence dans lequel les effets de bord sont susceptibles de déformer la figure. Dans la figure 12 nous constatons une structure significative autour de la période de trois ans en 1990 pour l'indice AO-1, alors que pour la série énergétique, une structure significative centrée autour de trois ans est visible en 1984. La figure 13 fait ressortir entre autres une similarité entre les deux séries autour de l'année 1985. Les relations de phase sont indiquées par des flèches.

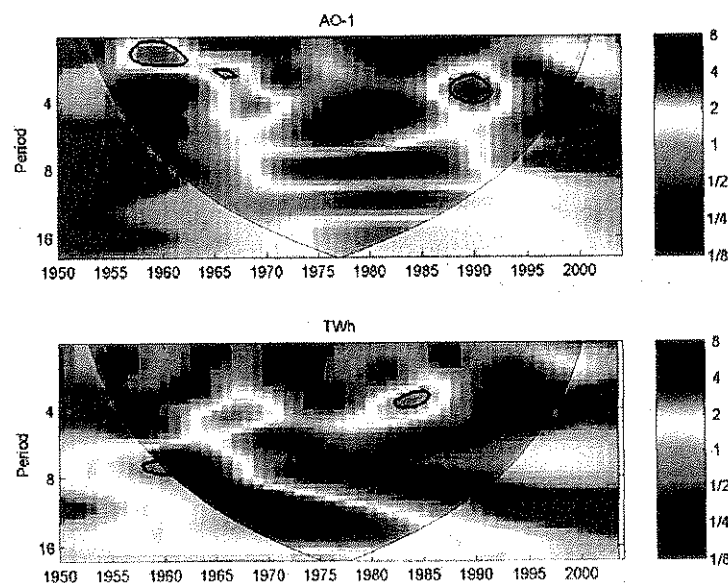


Figure 12 : Transformation continue en ondelettes de l'indice climatique AO-1 et de la série énergétique pour la période 1950 à 2004. Les couleurs indiquent la puissance de l'ondelette.

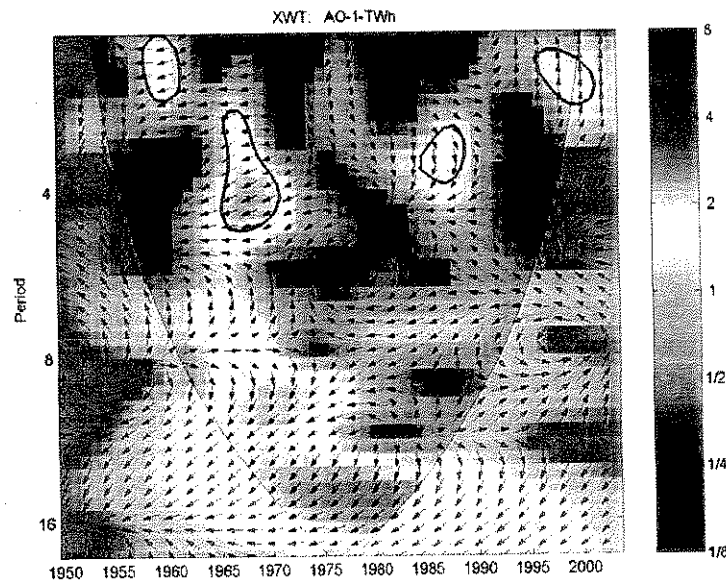


Figure 13 : Transformation croisée en ondelettes de l'indice climatique AO-1 et de la série énergétique pour la période 1950 à 2004. Les couleurs indiquent la puissance croisée.

5. Nos deux hypothèses

Nous formulons deux hypothèses, sur lesquelles sera basée la construction de nos modèles, en nous basant sur notre analyse approfondie de la série d'apports énergétiques. La première considère la non homogénéité de la moyenne qui a été prouvée à la section 3. La seconde formule l'absence de persistance significative. Les figures 14a et 14b montrent l'autocorrélogramme pour la série d'apports énergétiques pour la période totale (1943 à 2004) ainsi que pour la sous-période 1985 à 2004. Une seule valeur d'autocorrélation est significative correspondant au décalage 16 pour la série complète. Cet effet est inexplicable et est probablement induit par un effet d'échantillonnage. La figure 14 supporte donc notre deuxième hypothèse.

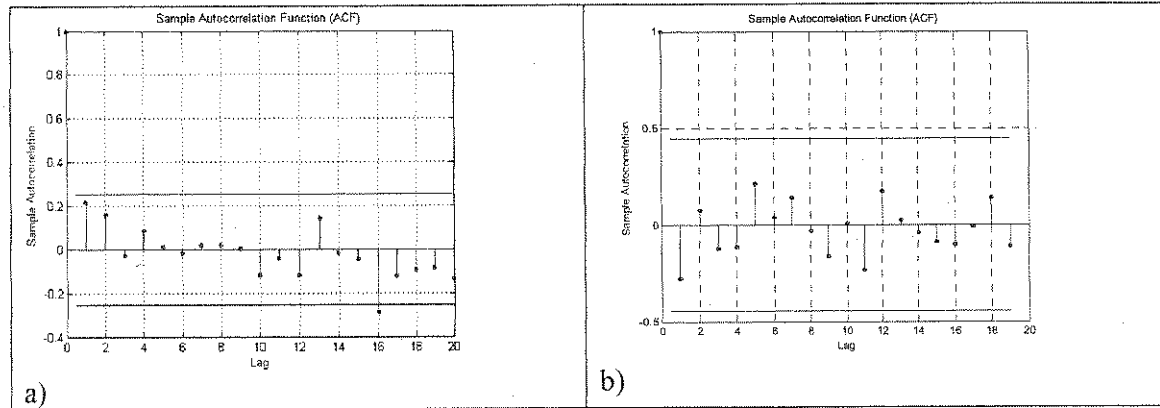


Figure 14 : Autocorrélogrammes de la série énergétique pour les périodes 1943 à 2004 (a) et 1985 à 2004 (b).

6. Deux modèles proposés

En nous basant sur les hypothèses de non homogénéité et d'absence de persistance énoncées à la section 5, deux modèles simples sont proposés pour la série d'apports énergétiques d'Hydro-Québec : un mélange de deux lois normales pour la période 1943-2004 et une seule loi normale après 1985.

6.1 Mélange de deux lois normales

On suppose que la série d'apports énergétiques X_t est distribuée selon un mélange de deux lois normales.

$$X_t \sim \pi N(\mu_1, \sigma_1^2) + (1 - \pi) N(\mu_2, \sigma_2^2) \quad (1)$$

On a alors que la série des déficits D_t suit également un mélange de distribution normale.

$$D_t \sim \pi N(\tilde{\mu}_1, \sigma_1^2) + (1 - \pi) N(\tilde{\mu}_2, \sigma_2^2) \quad (2)$$

où $\tilde{\mu}_1 = \mu_1 - \mu$, $\tilde{\mu}_2 = \mu_2 - \mu$ avec μ la moyenne à long terme des X_t .

La distribution du déficit énergétique sur deux ans peut être calculée facilement en considérant tous les cas possibles pour D_t et D_{t+1} d'appartenance à une des deux lois. Elle est égale à :

$$D_t + D_{t+1} \sim \pi^2 N(2\widetilde{\mu}_1, 2\sigma_1^2) + 2\pi(1-\pi)N(\widetilde{\mu}_1 + \widetilde{\mu}_2, \sigma_1^2 + \sigma_2^2) \\ + (1-\pi)^2 N(2\widetilde{\mu}_2, 2\sigma_2^2) \quad (3)$$

De la même manière que pour l'équation (3), la distribution du déficit énergétique sur quatre ans peut être calculée comme :

$$D_t + D_{t+1} + D_{t+2} + D_{t+3} \sim \pi^4 N(4\widetilde{\mu}_1, 4\sigma_1^2) + 4\pi^3(1-\pi)N(3\widetilde{\mu}_1 + \widetilde{\mu}_2, 3\sigma_1^2 + \sigma_2^2) \\ + 6\pi^2(1-\pi)^2 N(2\widetilde{\mu}_1 + 2\widetilde{\mu}_2, 2\sigma_1^2 + 2\sigma_2^2) \\ + 4\pi(1-\pi)^3 N(\widetilde{\mu}_1 + 3\widetilde{\mu}_2, \sigma_1^2 + 3\sigma_2^2) + (1-\pi)^4 N(4\widetilde{\mu}_2, 4\sigma_2^2)$$

(4)

L'estimation des paramètres du modèle est effectuée à l'aide de l'algorithme EM (Dempster, 1977) qui est usuellement utilisé dans le cas de mélanges de distributions. Cet algorithme comprend deux étapes : l'estimation des quantités inconnues indiquant à quelle distribution appartient l'observation i . Cette estimation est effectuée par le biais d'une espérance (étape E). L'estimation des paramètres par maximum de vraisemblance (étape M). Nous obtenons ainsi les estimateurs suivants par le biais de cet algorithme :

$$\begin{cases} \widehat{\pi} = 0.15 \\ \widehat{\mu}_1 = 161.0 \\ \widehat{\mu}_2 = 195.1 \\ \widehat{\sigma}_1 = 3.6 \\ \widehat{\sigma}_2 = 16.8 \end{cases} \quad (5)$$

Nous avons donc une probabilité de 15% que la moyenne soit égale à 161 TWh et une probabilité de 85% que la moyenne soit beaucoup plus élevée (égale à 195.1 TWh).

Notons que, suite à l'analyse exploratoire, nous avons considéré comme valeurs initiales de l'algorithme EM les moyennes avant et après 1985. Il est à remarquer que cet algorithme est sensible aux valeurs initiales, particulièrement aux moyennes et poids choisis lors de l'initialisation. La cause la plus probable est la faible taille de l'échantillon des apports énergétiques. Une étude de sensibilité a été effectuée avec diverses valeurs initiales pour les moyennes, les poids et les variances. L'algorithme converge parfois vers un résultat aberrant où l'une des variances est proche de zéro. Quand l'algorithme EM trouve un zéro, il reste « coincé » dans cette valeur, ce qui est une des caractéristiques de cette approche. Cependant quand l'algorithme converge vers une solution cohérente, celle-ci correspond aux valeurs que nous avons trouvées.

L'ajustement est vérifié en superposant l'histogramme normalisé et la densité du mélange avec les paramètres de l'équation (5). La figure 15 montre que l'ajustement obtenu est très satisfaisant.

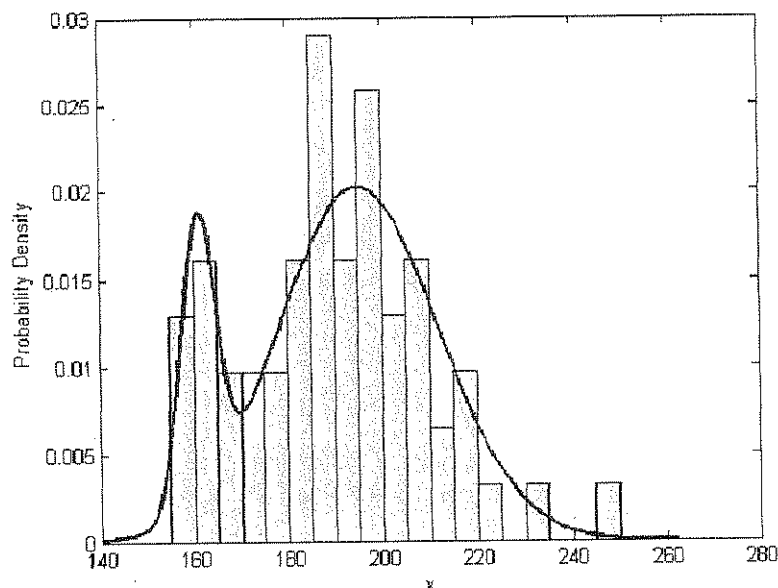


Figure 15 : Superposition de l'histogramme normalisé et de la densité du mélange de lois normales.

6.2 Loi normale après 1985

Nous avons montré dans les sections 3 et 4 que 1985 est vraiment une année de transition. Nous supposons que nous restons dans l'état actuel pour les dix prochaines années. De plus le caractère exceptionnel de la série énergétique est peut-être à rechercher dans le passé et non dans l'état actuel. Forts de ces constatations, nous proposons le modèle très simple suivant : la série d'apports énergétiques après 1985 suit une distribution normale et uniquement cette période sera prise en compte.

Nous avons donc dans ce cas que :

$$X_t \sim N(\mu_3, \sigma_3^2) \quad (6)$$

avec $\widehat{\mu}_3 = 181.0$, $\widehat{\sigma}_3 = 15.8$ (cf. tableau 1).

La distribution correspondante des déficits cumulés sur deux ans et sur quatre ans est donnée aux équations (7) et (8) par :

$$D_t + D_{t+1} \sim N\left(2\widetilde{\mu}_3, 2\sigma_3^2\right) \quad (7)$$

$$D_t + D_{t+1} + D_{t+2} + D_{t+3} \sim N\left(4\widetilde{\mu}_3, 4\sigma_3^2\right) \quad (8)$$

avec $\widetilde{\mu}_3 = \mu_3 - \mu$ où μ est la moyenne à long terme des X_t .

7. Réponses aux deux questions

Les développements théoriques énoncés dans la section 6 nous permettent de répondre aux questions qui constituent le cœur de ce mandat et qui ont été rappelées à la section 1.

7.1 Réponse à la question 1

En nous basant sur l'équation (1) et la définition d'un mélange de lois, le modèle 1 stipule que pour un horizon de dix ans, la série énergétique en 2014 peut s'écrire comme

$$X_{2014} = \begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \text{ avec prob. } \pi \\ X_2 \sim N(\mu_2, \sigma_2^2) \text{ avec prob. } (1-\pi) \end{cases} \quad (9)$$

L'espérance de la série énergétique se calcule alors simplement comme suit :

$$E(X_{2014}) = \pi\mu_1 + (1-\pi)\mu_2 = \mu = 189.9 \text{ TWh}$$

En utilisant le modèle 2 et un horizon de dix ans, le total énergétique s'exprime, en utilisant l'équation (6), comme suit :

$$X_{2014} \sim N(\mu_3, \sigma_3^2)$$

Dans ce cas $E(X_{2014}) = \mu_3 = 181.0 \text{ TWh}$

7.2 Réponse à la question 2

Nous avons déterminé la distribution des déficits cumulés calculés sur deux ans et quatre ans à l'aide du modèle 1 (équations 3 et 4). La question 2 concerne le calcul des quantiles à 2% de ces déficits qui s'obtiennent en inversant la distribution. Nous avons également calculé un intervalle

de confiance à 95 % correspondant à ces quantiles. Étant donné la faible taille des échantillons, ce calcul a été effectué à l'aide d'un bootstrap paramétrique (Davison et Hinkley, 1997). Les résultats obtenus à l'aide du modèle 1 sont exprimés dans l'équation 10.

$$\begin{aligned} q(D_t + D_{t+1})(2\%) &= -57.0 \text{ TWh} \quad IC = [-79.2 \text{ TWh}, -50.0 \text{ TWh}] \\ q(D_t + D_{t+1} + D_{t+2} + D_{t+3})(2\%) &= -79.4 \text{ TWh} \quad IC = [-103.4 \text{ TWh}, -56.7 \text{ TWh}] \end{aligned} \quad (10)$$

Les équations (7) et (8) expriment les distributions des déficits sur deux ans et quatre ans obtenues à partir du modèle 2. À l'aide de ces équations, nous obtenons les résultats suivants :

$$\begin{aligned} q(D_t + D_{t+1})(2\%) &= -63.6 \text{ TWh} \quad IC = [-66.7 \text{ TWh}, -40.7 \text{ TWh}] \\ q(D_t + D_{t+1} + D_{t+2} + D_{t+3})(2\%) &= -100.3 \text{ TWh} \quad IC = [-122.4 \text{ TWh}, -81.2 \text{ TWh}] \end{aligned} \quad (11)$$

8. Conclusions

Nous avons effectué une analyse approfondie de la série d'apports énergétiques 1943-2004 d'Hydro-Québec qui nous a conduit à postuler deux hypothèses pour la construction de modèles : la non homogénéité de la série énergétique et l'absence de persistance. Nous avons développé deux modèles simples cohérents avec ces hypothèses (section 6). Le tableau 5 résume les réponses obtenues pour la deuxième question avec les deux modèles.

Tableau 5 : Quantile à 2% du déficit énergétique (TWh) par rapport à la moyenne à long terme de 189.9.

Modèle	2 années		4 années	
Mélange de lois normales 1943-2004	-66.7	-40.6	-103.4	-56.7
Une seule loi normale 1985-2004	-79.2	-50.0	-122.4	-81.2

Nous ne privilégions pas un modèle par rapport à l'autre. Ce choix revient au décideur et est lié à un risque.

Notre analyse exploratoire a montré une rupture statistiquement significative en 1985. Il est à noter que l'estimation des paramètres du mélange de lois normales conduit à des valeurs très éloignées des moyennes et des variances obtenues en considérant les deux sous-échantillons avant et après rupture. Ce résultat n'est cependant pas du tout surprenant car notre modèle fréquentiel ne prend pas en compte ni le temps, ni des variables explicatives. Une extension naturelle au modèle, permettant de prendre en compte la persistance, serait de considérer une approche par chaînes de Markov cachées (Perreault et Fortin, 2003).

Cette étude a mis en évidence l'importance d'effectuer des études complémentaires afin de comprendre les liens non linéaires entre l'écoulement et l'Oscillation Arctique au Québec. Nous pouvons également nous poser la question quant au caractère exceptionnel de certaines périodes du point de vue de la production énergétique. Où le caractère hors du commun se situe-t-il ? Avant ou après l'année de transition 1985 ? Les périodes de « vaches grasses » ou de « vaches maigres » d'un point de vue énergétique constituent-elles l'ordinaire ? Cette question, bien qu'elle nous interpelle, reste ouverte car évidemment nous n'avons pas une série suffisamment longue pour y répondre de manière rigoureuse.

9. Bibliographie

- Davison, A.C. et Hinkley, D.V. (1997). Bootstrap methods and their application, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, pp. 582.
- Dempster, A.P., Laird, N.M. et Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society, Series B*, 39, 1-38.
- Déry, S.J. et Wood, E.F. (2004). Teleconnection between the Arctic Oscillation and Hudson Bay river discharge, *Geophysical Research Letters*, 31, L18205.
- Grinsted, A., Moore, J.C. et Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlinear Processes in Geophysics*, 11, 561-566.
- Huang, J., Shabbar, A. et Higuchi, K. (2001). The relationship between the wintertime North Atlantic Oscillation and blocking episodes in the North Atlantic, *International Journal of Climatology*, 21, 355-369.
- Hubert, P. (2000). The segmentation procedure as a tool for discrete modeling of hydrometeorological regimes, *Stochastic Environmental Research and Risk Assessment*, 14, 297-304.
- Lee, A.S.F. et Heghinian, S.M. (1977). A Shift of the Mean Level in a Sequence of Independent Normal Random Variables - A Bayesian Approach., *Technometrics*, 19, 503-506.
- Perreault, L. et Fortin, V. (2003). Valorisation de l'historique pour la prévision énergétique à longue échéance. Institut de recherche d'Hydro-Québec (IREQ), rapport de recherche No IREQ-2003-159C.
- Thompson, D.W.J. et Wallace, J.M. (1998). The Arctic Oscillation signature in the wintertime geopotential height and temperature fields, *Geophysical Research Letters*, 25, 1297-1300.
- Thompson, D.W.J. et Wallace, J.M. (2001). Regional Climate Impacts of the Northern Hemisphere Annular Mode, *Science*, 293, 85-89.
- Torrence, C. et Compo, G.P. (1998). A Practical Guide to Wavelet Analysis, *Bulletin of American Meteorological Society*, 79, 61-78.

Résumé du rapport technique sur l'évaluation de la série d'apports énergétiques 1943-2004 d'Hydro-Québec

Jean-François Angers, P.Stat

31 octobre 2005

1 Introduction

Dans ce rapport, nous essayons d'apporter une réponse aux deux questions suivantes :

Question 1 : Est-ce que, outre le changement de moyenne résultant de l'ajout des données de 2001 à 2004, la moyenne historique des apports énergétiques calculée sur la base des données historiques reste toujours le meilleur indicateur de la moyenne à long terme des apports énergétiques annuels ?

Question 2 : Quel est le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2% ?

Afin de répondre à ces questions, nous devons évaluer la probabilité

$$\mathbb{P} \left(\sum_{l=1}^k (Y_{n+l} - \bar{y}) \leq E_0 \mid y_1, \dots, y_n \right), \quad (1)$$

où $Y_{n+1}, \dots, Y_{n+k}$ représentent les apports futurs, $y_1, y_2, \dots, y_n$, les données historiques, $\bar{y} = 189,8626$ (la moyenne des apports énergétiques de 1943 à 2004) et E_0 est la valeur du déficit énergétique en supposant que le passé est garant de l'avenir pour un horizon de deux années ($k = 2$) et de quatre années ($k = 4$). Nous cherchons aussi la valeur de E_0 telle que l'équation (1) soit égale à 0,02. Nous allons utiliser une méthodologie bayésienne pour évaluer l'équation (1).

Dans ce résumé, nous commençons par décrire les modèles utilisés (sections 2 et 3 dans le rapport technique). À la section 3, nous répondons aux questions 1 et 2 énoncées à l'introduction (sections 2.4, 3.4 et 4 du rapport technique).

2 Modèles

Deux modèles ont été considérés pour répondre aux questions 1 et 2. Le premier modèle consiste à un modèle autorégressif. Ce modèle a été utilisé afin de tenir compte des apports

passés pour prédire les apports futurs. Il est présenté brièvement à la sous-section 2.1 de ce résumé et, avec plus de détails, à la section 2 du rapport technique.

Le choix du deuxième modèle est justifié par les données car lorsque nous faisons l'histogramme des apports (voir figure 1), nous pouvons voir apparaître deux modes, correspondant à des années humides et des années sèches. Ce modèle est décrit à la sous-section 2.2 de ce résumé et à la section 3 du rapport technique.

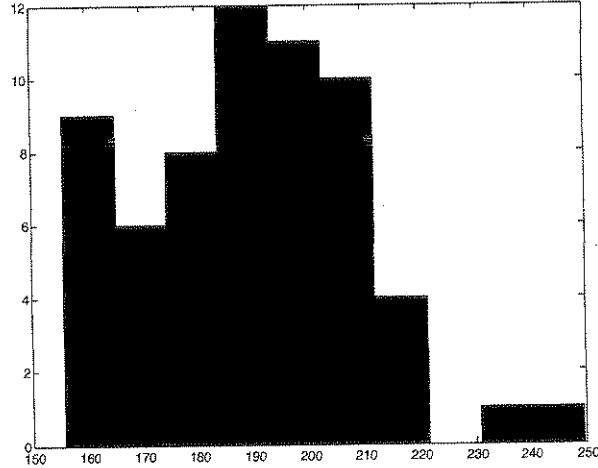


FIG. 1 – Histogramme des apports de 1943 à 2004

2.1 Modèle 1

Lors de la réunion du 22 juin 2005, il fut mentionné que les apports passés devraient servir à prédire les apports futurs. Pour vérifier ceci nous avons calculé les autocorrélations entre les apports (voir la figure 1 du rapport en annexe). Cependant, nous avons trouvé que les autocorrélations ne sont pas significatives. Par contre, comme l'autocorrélation d'ordre 1 est assez importante nous allons quand même utiliser un modèle AR(1), c'est-à-dire

$$y_t = \mu + \rho(y_{t-1} - \mu) + \varepsilon_t,$$

pour $t = 2, 3, \dots, n$, et où $\varepsilon_2, \varepsilon_3, \dots, \varepsilon_n$ sont n erreurs indépendantes, normales de moyenne 0 et de variance σ^2 .

Comme modèle *a priori*, nous utilisons un modèle conjugué (mais vague) sur μ et σ^2 et un modèle non informatif sur ρ . Ainsi nous avons

$$\begin{aligned} \mu \mid \sigma^2 &\sim N(\mu_0; \sigma^2/n_0); \\ \sigma^2 &\sim \text{IF}(\alpha/2; \beta/2); \\ \rho &\sim U_{(-1;1)}, \end{aligned}$$

où $n_0 \leq 1$, α et β sont « près » de 0.

En utilisant l'ensemble des données historiques, nous pouvons obtenir un intervalle de crédibilité de niveau 95% pour ρ , c'est-à-dire des valeurs de ρ_{inf} , ρ_{sup} telles que

$$\mathbb{P}(\rho_{inf} < \rho < \rho_{sup} \mid y_1, \dots, y_n) = 0,95$$

et que $\rho_{sup} - \rho_{inf}$ soit minimum. Lorsque $n_0 = 0,1$, $\alpha = 1$ et $\theta = 1$, l'intervalle de crédibilité de niveau 95% est $[-0,0175; 0,4938]$. Comme la valeur $\rho = 0$ appartient à l'intervalle de crédibilité, il est possible que les apports soient indépendants du passé.

2.2 Modèle 2

Afin de tenir compte de la présence de périodes sèches et humides, nous allons utiliser un mélange de modèles pour représenter la densité des apports. Ainsi, considérons la densité

$$f(y_t \mid p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) = p f_0(y_t \mid \theta_1, \sigma_1^2) + (1 - p) f_0(y_t \mid \theta_2, \sigma_2^2), \quad (2)$$

où y_t est l'apport au temps t , $p \in (0; 1)$ et $f_0(\cdot \mid \theta, \sigma^2)$ représente la densité normale de moyenne θ et de variance σ^2 .

Tout comme pour le premier modèle, nous utilisons des densités conjuguées (mais vagues) afin de simplifier les calculs. Nous avons donc :

$$\begin{aligned} p &\sim \text{be}(a; b), \\ \theta_i \mid \sigma_i^2 &\sim \text{N}(\theta_{0i}; \sigma_i^2/n_0), \\ \sigma_i^2 &\sim \text{II}(\alpha/2; \beta/2), \end{aligned}$$

où $n_0 \leq 1$, α et β sont « près » de 0. Les paramètres a et b sont choisis de façon à rendre la densité *a priori* la plus non informative possible.

3 Résultats

Dans cette section, nous allons répondre aux deux questions énoncées dans l'introduction. Plus de détails sont disponibles dans le rapport technique.

3.1 Question 1

La première question du mandat consistait à vérifier si la base des données historiques restait le meilleur indicateur de la moyenne à long terme. Pour répondre à cette question, nous avons utilisé les deux modèles de la section précédente mais en faisant changer l'historique de 1943-2004 à 1944-2004 jusqu'à 2000-2004. (Le déficit est quand même calculé à partir de la moyenne de l'ensemble des données historiques, c'est-à-dire de 1943 à 2004.) Les résultats sont données aux tableaux 1 et 4 du rapport technique. À la figure 2, nous retrouvons la valeur du déficit énergétique (courbe noire) en fonction des données historiques considérées pour les

TAB. 1 – Valeurs de E_0 et son intervalle de crédibilité à 95% pour un horizon de 2 ans basées sur l'historique 1943-2004

Modèle	Hyperparamètres	Déficit	Intervalle de crédibilité
1	$n_0 = 0,1; \alpha = \beta = 1$	-61,8	(-76,0; -51,0)
2	$n_0 = 0,1, \alpha = \beta = 1; a = b = 0,5$	-58,2	(-68,2; -47,0)

TAB. 2 – Valeurs de E_0 et son intervalle de crédibilité à 95% pour un horizon de 4 ans basées sur l'historique 1943-2004

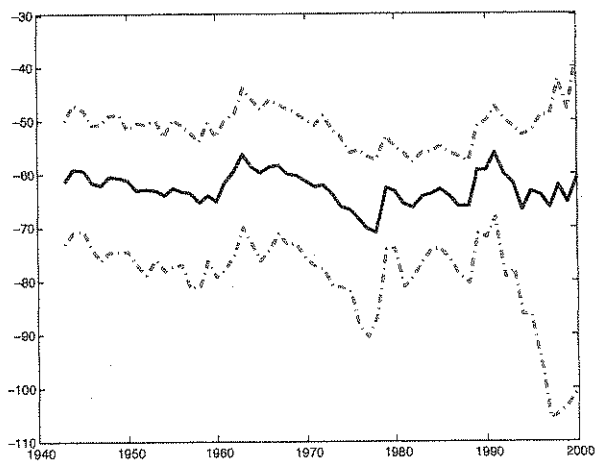
Modèle	Hyperparamètres	Déficit	Intervalle de crédibilité
1	$n_0 = 0,1; \alpha = \beta = 1$	-98,8	(-127,0; -79,5)
2	$n_0 = 0,1, \alpha = \beta = 1; a = b = 0,5$	-82,7	(-105,0; -62,8)

deux modèles et les horizons de deux et quatre ans. L'intervalle de crédibilité de niveau 95% (point à point) est aussi représenté par une courbe pointillée en vert. En examinant cette figure, nous voyons bien que, quel que soit l'historique considéré, la valeur du déficit est relativement constante. De plus, les intervalles de confiance contiennent les valeurs utilisées présentement par Hydro-Québec, soit -64TWh sur deux ans et -98TWh pour quatre ans. Les déficits pour un horizon de quatre ans avec le modèle deux sont inférieurs à ceux prédits par le modèle 1. Encore une fois, nous pouvons par contre constater que la courbe des déficits estimés par un modèle se trouve incluse dans l'intervalle de confiance obtenu par l'autre modèle.

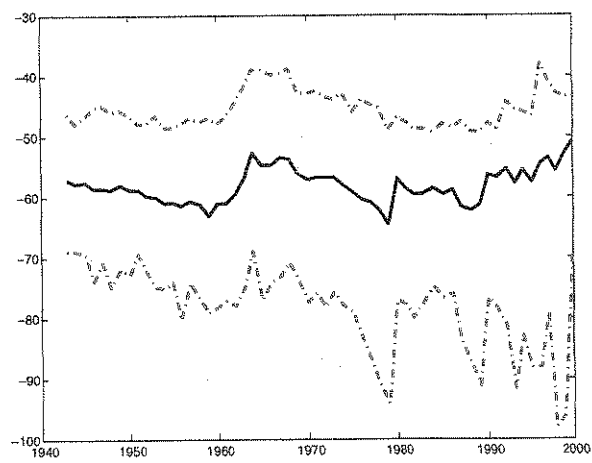
3.2 Question 2

La deuxième question du mandat était de prédire le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2%. Les valeurs utilisées actuellement sont de -64 TWh sur deux ans et de -98 TWh sur quatre ans.

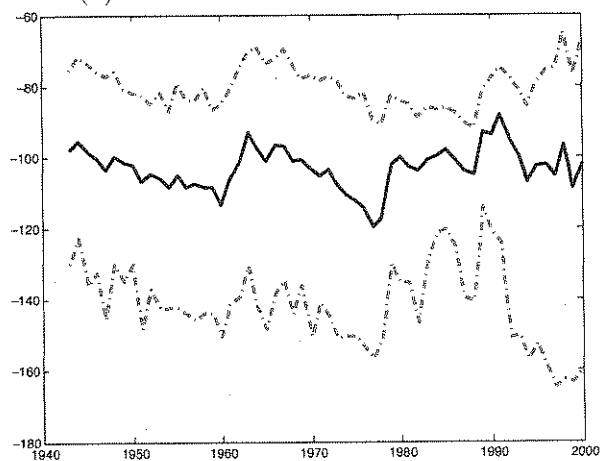
Nous avons calculé l'équation (1) en utilisant l'ensemble des données historiques, c'est-à-dire de 1943 à 2004 et ce pour les deux modèles présentés à la section 2 (voir les sections 2 et 3 du rapport en annexe pour plus de détails). En se basant sur les résultats des deux modèles (voir les tableau 1 et 2), nous pouvons conclure que ces valeurs sont encore actuelles car elles sont comprises dans les intervalles de crédibilité.



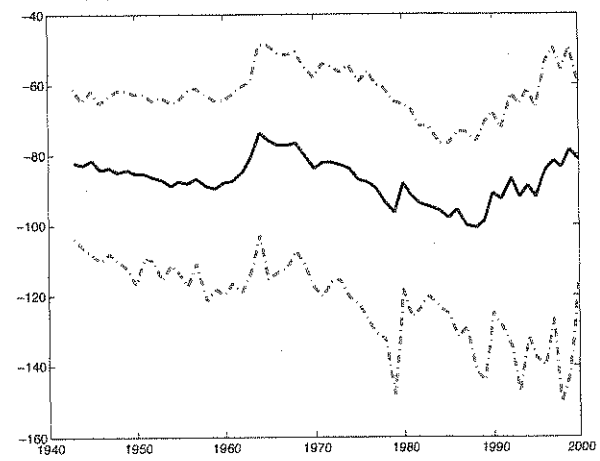
(a) Modèle 1 avec horizon de 2 ans



(b) Modèle 2 avec horizon de 2 ans



(c) Modèle 1 avec horizon de 4 ans



(d) Modèle 2 avec horizon de 4 ans

FIG. 2 – Influence de la période historique sur l'estimation du déficit

Rapport sur l'évaluation de la série d'apports énergétiques 1943-2004 d'Hydro-Québec

Jean-François Angers, P.Stat

31 octobre 2005

1 Introduction

Dans ce rapport, nous essayons d'apporter une réponse aux deux questions suivantes :

Question 1 : Est-ce que, outre le changement de moyenne résultant de l'ajout des données de 2001 à 2004, la moyenne historique des apports énergétiques calculée sur la base des données historiques reste toujours le meilleur indicateur de la moyenne à long terme des apports énergétiques annuels ?

Question 2 : Quel est le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2% ?

Pour ce faire, nous devons évaluer la probabilité

$$\mathbb{P} \left(\sum_{l=1}^k (Y_{n+l} - \bar{y}) \leq E_0 \mid y_1, \dots, y_n \right), \quad (1)$$

où $\bar{y} = 189,8626$ (la moyenne des apports énergétiques de 1943 à 2004) et E_0 est la valeur du déficit énergétique en supposant que le passé est garant de l'avenir pour un horizon de deux années ($k = 2$) et de quatre années ($k = 4$). Nous cherchons aussi la valeur de E_0 telle que l'équation (1) soit égale à 0,02. Nous allons utiliser une méthodologie bayésienne pour évaluer l'équation (1).

2 Modèle 1

Lors de la réunion du 22 juin 2005, il fut mentionné que les apports passés devraient servir à prédire les apports futurs. Pour vérifier ceci nous avons calculé les autocorrélations entre les apports. Celles-ci sont données à la figure 1.

Comme il avait été mentionné par Luc Perreault, le délai d'ordre 1 est le plus important mais il n'est pas significatif. (Le délai d'ordre 16 est significatif, mais il semble peu réaliste

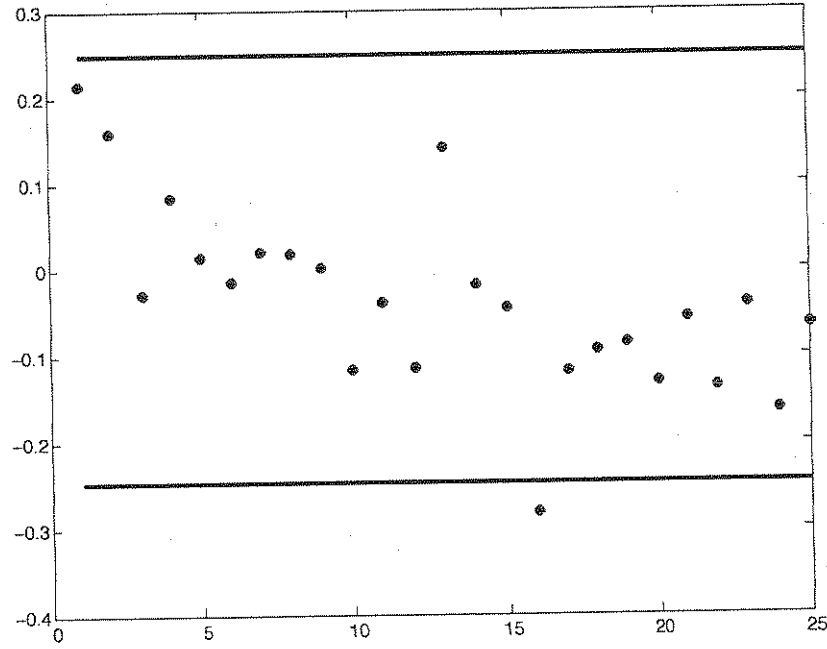


FIG. 1 – Autocorrélations entre les apports

de supposer que l'apport d'il y a 16 années puisse influencer celui de l'année en cours.) Nous avons donc décidé de modéliser le vecteur des apports de la façon suivante

$$y_t = \mu + \rho(y_{t-1} - \mu) + \varepsilon_t,$$

pour $t = 2, 3, \dots, n$, et où $\varepsilon_2, \varepsilon_3, \dots, \varepsilon_n$ sont n erreurs indépendantes, normales de moyenne 0 et de variance σ^2 . De plus, étant dans un cadre bayésien, $\mu \in \mathbb{R}$ et $\rho \in (-1; 1)$ sont aussi des variables aléatoires dont les densités seront spécifiées plus tard.

Posons $x_t = y_t - \rho y_{t-1}$ pour $t = 2, 3, \dots, n$ et $m = n - 1$. Ainsi, l'équation (2) peut s'écrire

$$x_t = (1 - \rho)\mu + \varepsilon_t.$$

La fonction de vraisemblance est donc

$$\begin{aligned} f(\underline{x} \mid \mu, \sigma^2, \rho) &= \prod_{i=2}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x_t - (1 - \rho)\mu)^2 \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=2}^n ([x_t - \bar{x}] + [\bar{x} - (1 - \rho)\mu])^2 \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp \left\{ -\frac{1}{2\sigma^2} [S + m(\bar{x} - \theta)^2] \right\} \end{aligned} \quad (2)$$

où $\bar{x} = m^{-1} \sum_{t=2}^n x_t$, $S = \sum_{t=2}^n (x_t - \bar{x})^2$ et $\theta = (1 - \rho)\mu$.

2.1 Modèles *a priori* et *a posteriori*

Comme modèle *a priori*, nous utilisons un modèle conjugué (mais vague) sur μ et σ^2 et un modèle non informatif sur ρ . Ainsi nous avons

$$\begin{aligned}\mu \mid \sigma^2 &\sim N(\mu_0; \sigma^2/n_0); \\ \sigma^2 &\sim \text{IG}(\alpha/2; \beta/2); \\ \rho &\sim U_{(-1,1)},\end{aligned}$$

où $n_0 \leq 1$, α et β sont « près » de 0 et $\text{IG}(a; b)$ représente la densité inverse gamma dont la densité est donnée par

$$f(x \mid a, b) = \frac{b^a}{\Gamma(a)} \frac{1}{x^{a+1}} \exp \left\{ -\frac{b}{x} \right\} \mathbb{I}_{\mathbb{R}_+}(x).$$

Ainsi, en conditionnant sur ρ and σ^2 , la densité *a priori* sur θ est

$$\theta \mid \sigma^2 \sim N \left((1 - \rho)\mu_0; \frac{(1 - \rho)^2 \sigma^2}{n_0} \right). \quad (3)$$

Les densités *a posteriori*, conditionnelles par rapport à ρ , sont les suivantes

$$\begin{aligned}\theta \mid \sigma^2, \rho, \underline{x} &\sim N \left(\frac{n_0(1 - \rho)\mu_0 + m(1 - \rho)^2 \bar{x}}{n_0 + m(1 - \rho)^2}; \frac{(1 - \rho)^2 \sigma^2}{n_0 + m(1 - \rho)^2} \right); \\ \mu \mid \sigma^2, \rho, \underline{x} &\sim N \left(\frac{n_0\mu_0 + m(1 - \rho)\bar{x}}{n_0 + m(1 - \rho)^2}; \frac{\sigma^2}{n_0 + m(1 - \rho)^2} \right); \\ \sigma^2 \mid \rho, \underline{x} &\sim \text{IG} \left(\frac{m + \alpha}{2}; \frac{1}{2} \left[\beta + S + \frac{n_0 m}{n_0 + m(1 - \rho)^2} (\bar{x} - (1 - \rho)\mu_0)^2 \right] \right).\end{aligned}$$

2.2 Calcul de la densité prédictive

Le but de ce projet est d'évaluer l'équation (1). En utilisant l'approche bayésienne, nous pouvons montrer que la densité prédictive de $(Y_{n+1}, \dots, Y_{n+k})$ (sachant $\underline{y}$) est donnée par

$$\begin{aligned}m(y_{n+1}, \dots, y_{n+k} \mid \underline{y}) &= \frac{\text{densité conjointe de } (y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k})}{\text{densité marginale de } (y_1, \dots, y_n)} \\ &= \frac{\int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k} \mid \mu, \sigma^2, \rho) \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho}{\int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n \mid \mu, \sigma^2, \rho) \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho} \\ &= \int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(y_{n+1}, \dots, y_{n+k} \mid y_1, \dots, y_n, \mu, \sigma^2, \rho) f(y_1, \dots, y_n \mid \mu, \sigma^2, \rho) \\ &\quad \times \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho \\ &\quad \Bigg/ \int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n \mid \mu, \sigma^2, \rho) \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho\end{aligned}$$

$$\begin{aligned}
&= \int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(x_{n+1}, \dots, x_{n+k} \mid y_1, \dots, y_n, \mu, \sigma^2, \rho) f(y_1, \dots, y_n \mid \mu, \sigma^2, \rho) \\
&\quad \times \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho \\
&\quad \Bigg/ \int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n \mid \mu, \sigma^2, \rho) \pi_1(\mu \mid \sigma^2) \pi_2(\sigma^2) \pi_3(\rho) d\mu d\sigma^2 d\rho \\
&= \int_{-1}^1 \int_0^\infty \int_{-\infty}^\infty f(x_{n+1}, \dots, x_{n+k} \mid y_1, \dots, y_n, \mu, \sigma^2, \rho) \\
&\quad \times \pi_1(\mu \mid \sigma^2 \rho, y_1, \dots, y_n) \pi_2(\sigma^2 \mid \rho, y_1, \dots, y_n) \pi_3(\rho) m(y_1, \dots, y_n \mid \rho) d\mu d\sigma^2 d\rho \\
&\quad \Bigg/ \int_{-1}^1 \pi_3(\rho) m(y_1, \dots, y_n \mid \rho) d\rho \\
&= \frac{\int_{-1}^1 \pi_3(\rho) m(y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k} \mid \rho) d\rho}{\int_{-1}^1 \pi_3(\rho) m(y_1, \dots, y_n \mid \rho) d\rho}, \tag{4}
\end{aligned}$$

où

$$\begin{aligned}
m(\underline{z} \mid \rho) &\propto \sqrt{\frac{n_0}{n_0 + m(1 - \rho)^2}} \frac{\Gamma([m + \alpha]/2)}{\Gamma(\alpha/2)} \frac{\beta^{\alpha/2}}{\left[\beta + S_z + \frac{n_0 m (\tilde{z} - (1 - \rho)\mu_0)^2}{n_0 + m(1 - \rho)^2} \right]^{(m + \alpha)/2}}, \\
\tilde{z} &= \frac{1}{m} \sum_{t=2}^{m+1} (z_t - \rho z_{t-1}), \\
S_z &= \sum_{t=2}^{m+1} (z_t - \rho z_{t-1} - \tilde{z})^2,
\end{aligned}$$

et $m + 1$ représente la longueur du $\underline{z}$. Il est à noter que l'intégrale de l'équation (4) doit se faire de façon numérique. Un graphique de l'équation (4) est donné pour $k = 1$ à la figure 2 (lorsque $n_0 = 0,1$, $\alpha = \beta = 1$).

2.3 Calcul de l'équation (1)

Pour évaluer l'équation (1), nous procédons comme suit :

Étape 1 : Générez une valeur de ρ selon

$$\pi_3(\rho \mid y_1, \dots, y_n) \propto m(y_1, \dots, y_n \mid \rho) \pi_3(\rho).$$

(Un graphique de $\pi_3(\rho \mid y_1, \dots, y_n)$ pour $n_0 = 0,1$, $\alpha = \beta = 1$ est donné à la figure 3.)

Étape 2 : Pour chaque valeur de ρ , générez des valeurs de $(y_{n+1}, \dots, y_{n+k})$, notée $(y_{i,n+1}, \dots, y_{i,n+k})$, $i = 1, 2, \dots, M$, selon

$$m(y_{n+1}, \dots, y_{n+k} \mid y_1, \dots, y_n, \rho) = \frac{m(y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k} \mid \rho)}{m(y_1, \dots, y_n \mid \rho)}.$$

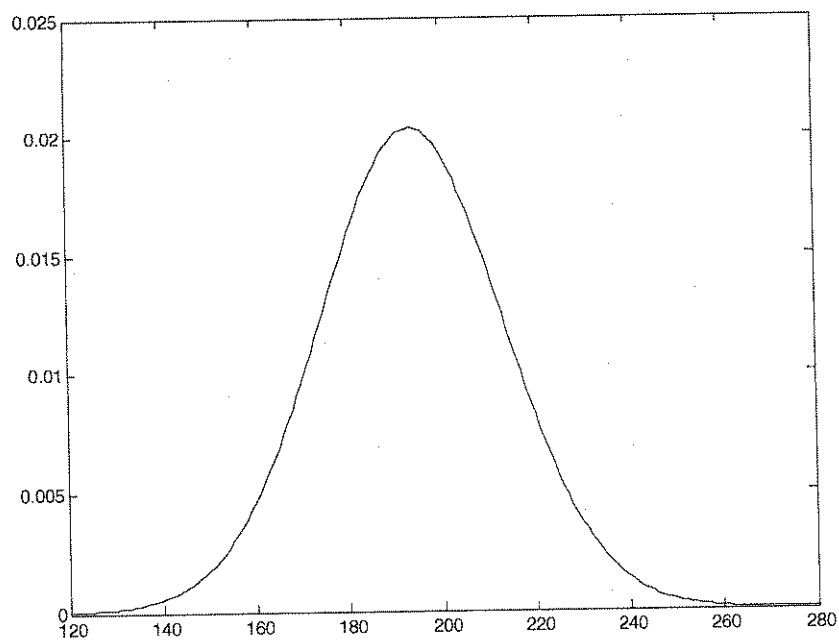


FIG. 2 – Densité marginale de $y_{n+1} \mid \rho = 0,24$

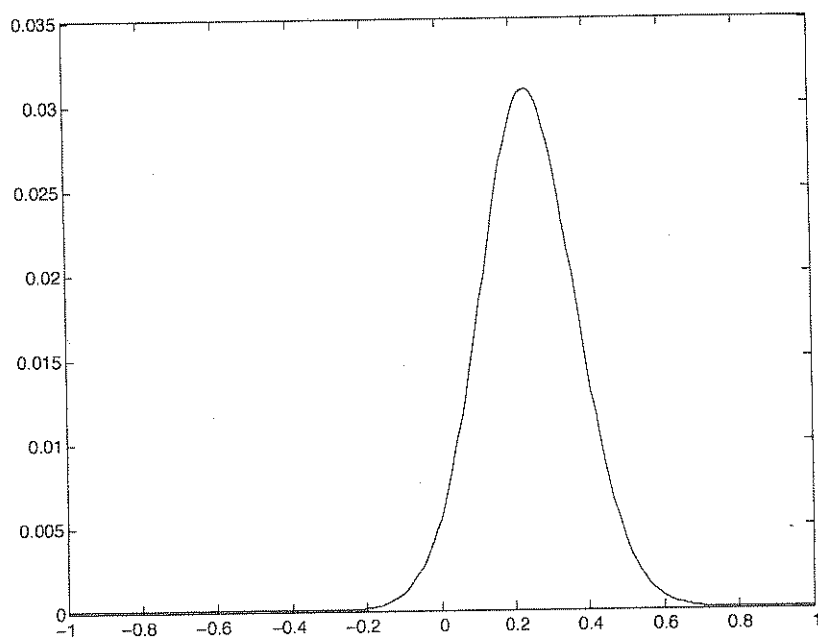


FIG. 3 – Densité *a posteriori* de ρ

Nous estimons l'équation (1) par

$$\hat{\mathbb{P}}_\rho = \frac{\#\{i \mid \sum_{l=1}^k (y_{i,n+l} - \bar{y}) \leq E_0\}}{M+1},$$

où $\#A$ correspond au cardinal de l'ensemble A .

Nous répétons les étapes 1 et 2 un grand nombre de fois et nous estimons l'équation (1) par

$$\hat{\mathbb{P}}\left(\sum_{l=1}^k (Y_{n+l} - \bar{y}) \leq E_0 \mid y_1, \dots, y_n\right) = \frac{\sum_{j=1}^M \hat{\mathbb{P}}_{\rho_j} \pi_3(\rho_j \mid y_1, \dots, y_n)}{\sum_{j=1}^M \pi_3(\rho_j \mid y_1, \dots, y_n)}. \quad (5)$$

2.4 Résultats

À l'aide de la figure 3, nous pouvons obtenir un intervalle de crédibilité de niveau 95% pour ρ , c'est-à-dire des valeurs de ρ_{inf} , ρ_{sup} telles que

$$\mathbb{P}(\rho_{inf} < \rho < \rho_{sup} \mid y_1, \dots, y_n) = 0,95$$

et que $\rho_{sup} - \rho_{inf}$ soit minimum. Lorsque $n_0 = 0,1$, $\alpha = 1$ et $\theta = 1$, l'intervalle de crédibilité est $[-0,0175; 0,4938]$. Nous ne pouvons donc pas conclure que l'autocorrélation d'ordre 1 soit significative au niveau 5%.

La première question du mandat consistait à vérifier si la base des données historiques restait le meilleur indicateur de la moyenne à long terme. Pour répondre à cette question, nous avons repris l'analyse avec le premier modèle mais en faisant changer la première année de l'historique de 1943, 1945 à ... par cinq ans. (Le déficit est toujours calculé à partir de la moyenne de toutes les observations.) Les résultats sont données au tableau 1 et à la figure 4 (déficit estimé et son intervalle de crédibilité de niveau 95% pour un horizon de deux et quatre ans). En examinant le tableau 1, nous voyons bien que, quelle que soit l'année de départ, les intervalles de confiance ne sont pas disjoints. Donc, quel que soit l'historique considérée, les valeurs des déficits sont sensiblement les mêmes.

Nous avons simulé $M = 250$ valeurs de ρ et nous avons évalué l'équation (5). Cette densité cumulative, basée sur l'ensemble des données historiques, est donnée aux figures 5 et 6 en plus de l'enveloppe convexe (minimum et maximum) pour le cas $(n_0, \alpha, \beta) = (0,1, 1, 1)$

Nous avons aussi considéré deux autres scénarios *a priori* possibles pour les valeurs de n_0 , α et β . Dans le tableau 2, nous avons calculé la valeur de E_0 ainsi qu'un intervalle de crédibilité de niveau 95%. Les trois scénarios considérés sont $(n_0, \alpha, \beta) = (0,1, 1, 1)$ (cas 1) $(0,01, 0,1, 0,1)$ (cas 2) et $(0,001, 0,01, 0,01)$ (cas 3).

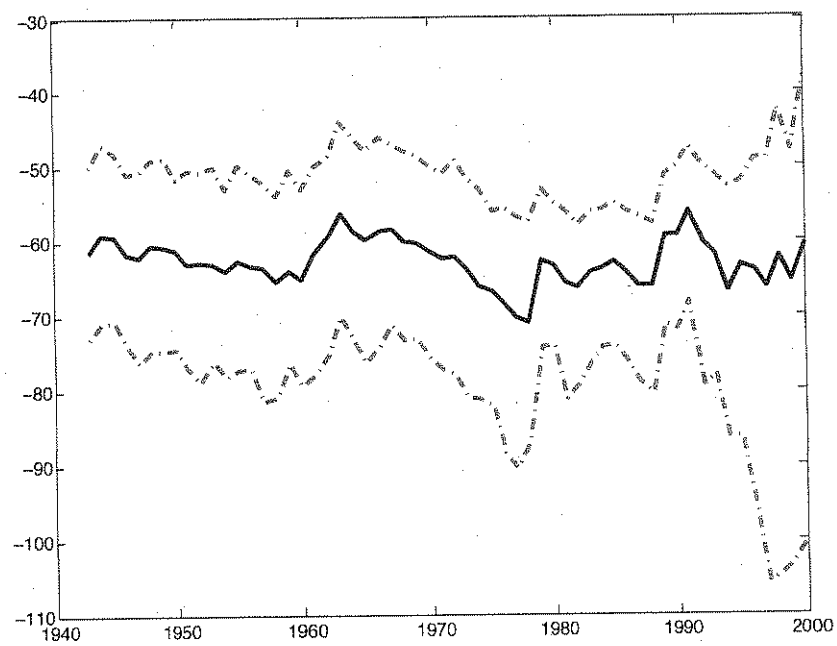
Nous voyons bien que l'information *a priori* (modélisée par n_0 , α et β) n'influence pas réellement les résultats de l'équation (5). Les valeurs de cette équation pour différentes valeurs de E_0 sont tabulées au tableau 3.

TAB. 1 – Valeurs de E_0 pour le modèle 1 et différentes années de début

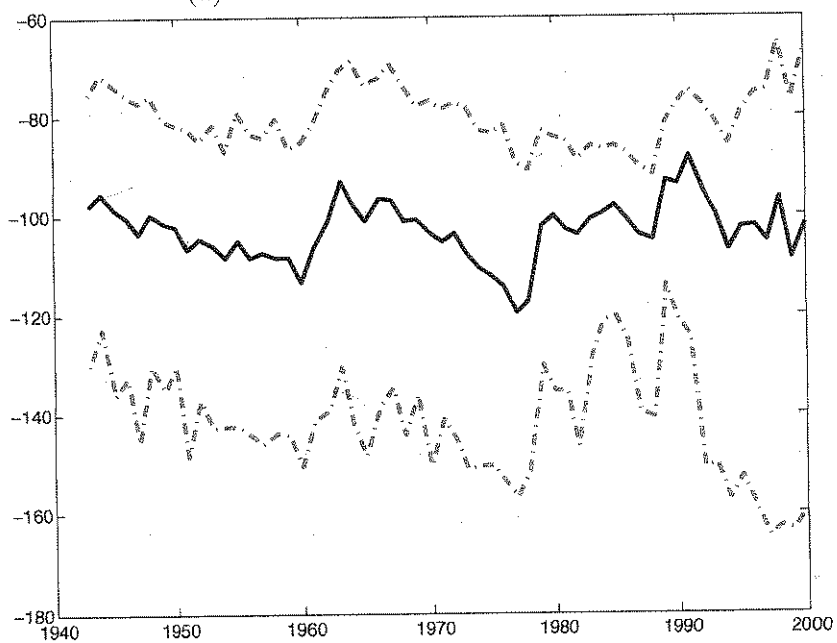
Année de départ	Horizon 2 ans		Horizon 4 ans	
	Déficit	Intervalle	Déficit	Intervalle
1943	-61,4335	(-72,9759 ; -49,9891)	-97,9856	(-130,2259 ; -75,7259)
1945	-59,2263	(-70,7238 ; -47,9891)	-98,4915	(-136,1001 ; -73,7292)
1950	-61,1633	(-74,3919 ; -51,4771)	-102,1475	(-130,2259 ; -81,9759)
1955	-62,4796	(-77,2259 ; -49,7391)	-105,0833	(-142,4858 ; -78,9887)
1960	-65,1271	(-79,2259 ; -52,9759)	-113,4032	(-150,2767 ; -85,2285)
1965	-59,7973	(-76,2259 ; -47,7292)	-101,0376	(-148,2259 ; -73,9891)
1970	-61,3581	(-75,7238 ; -49,9759)	-103,5788	(-150,2866 ; -76,9252)
1975	-66,6992	(-82,2259 ; -55,9759)	-112,4643	(-150,2767 ; -83,6546)
1980	-63,1247	(-73,9759 ; -54,7259)	-100,1103	(-135,2259 ; -84,4738)
1985	-62,6672	(-73,9759 ; -54,9390)	-98,0151	(-120,2238 ; -85,9920)
1990	-59,2247	(-71,9759 ; -49,9887)	-93,6722	(-120,9792 ; -77,7302)
1995	-63,2239	(-86,2259 ; -51,9891)	-102,4300	(-152,4666 ; -79,4738)
2000	-60,4258	(-100,2259 ; -37,9901)	-101,7855	(-159,6014 ; -66,9251)

TAB. 2 – Valeurs de E_0 pour le modèle 1 et les trois scénarios

Scénario	Horizon 2 ans		Horizon 4 ans	
	Moyenne	Intervalle	Moyenne	Intervalle
1	-61,8	(-76,0 ; -51,0)	-98,8	(-127,0 ; -79,5)
2	-62,2	(-75,2 ; -50,6)	-100,0	(-131,0 ; -78,7)
3	-61,9	(-75,0 ; -50,7)	-98,2	(-128,0 ; -76,0)



(a) Modèle 1 avec horizon de 2 ans



(b) Modèle 1 avec horizon de 4 ans

FIG. 4 – Influence de la période historique sur l'estimation du déficit à partir du modèle 1

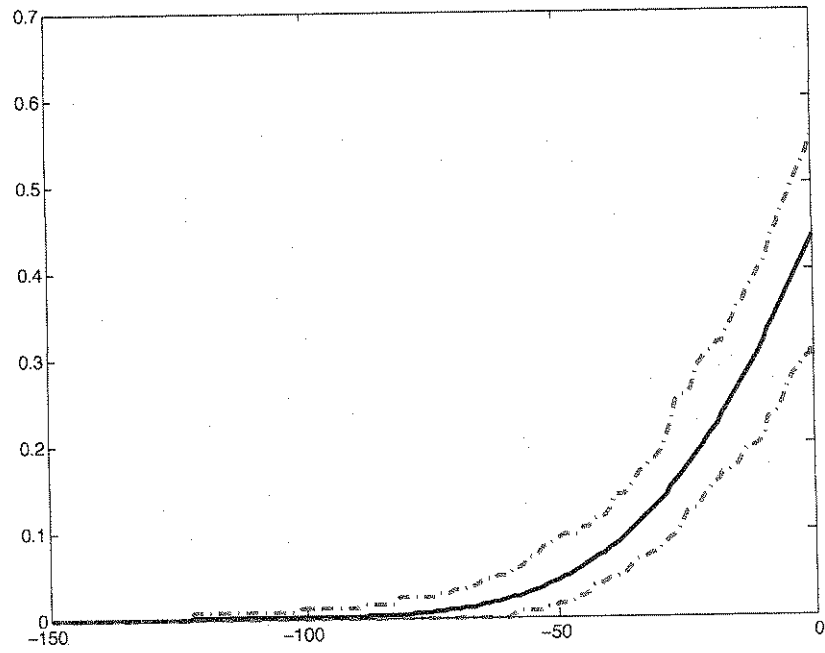


FIG. 5 – Équation (1) pour le modèle 1 avec $k = 2$

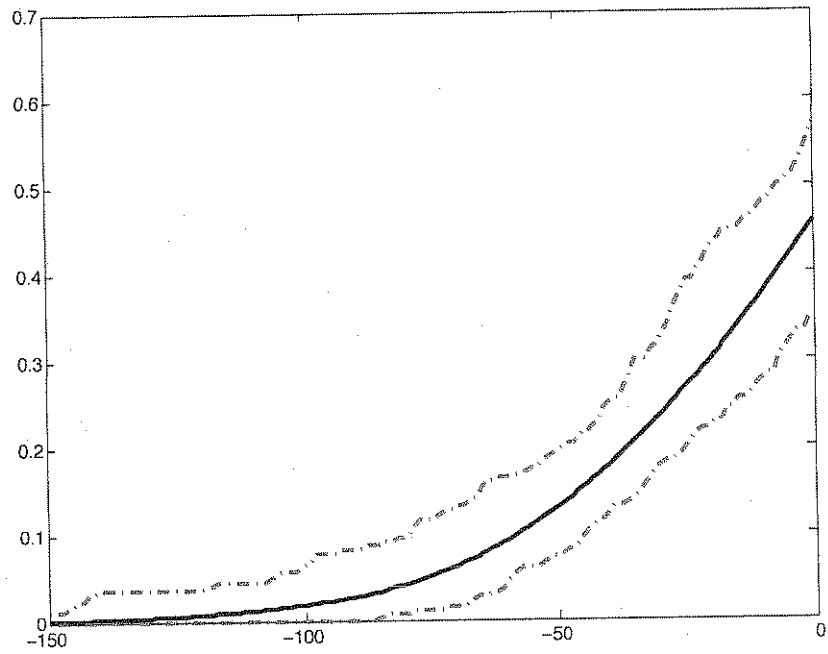


FIG. 6 – Équation (1) pour le modèle 1 avec $k = 4$

TAB. 3 – Valeurs de l'équation (1) pour le modèle 1 et les trois scénarios

Déficit	Horizon 2 ans			Horizon 4 ans		
	Cas 1	Cas 2	Cas 3	Cas 1	Cas 2	Cas 3
-115	0,0002	0,0002	0,0001	0,0090	0,0096	0,0092
-110	0,0003	0,0003	0,0002	0,0117	0,0116	0,0119
-105	0,0004	0,0004	0,0003	0,0147	0,0150	0,0149
-100	0,0006	0,0006	0,0007	0,0185	0,0189	0,0189
-95	0,0011	0,0010	0,0012	0,0232	0,0235	0,0237
-90	0,0018	0,0017	0,0019	0,0283	0,0295	0,0298
-85	0,0030	0,0028	0,0029	0,0352	0,0360	0,0364
-80	0,0046	0,0040	0,0046	0,0435	0,0438	0,0444
-75	0,0069	0,0063	0,0067	0,0526	0,0541	0,0544
-70	0,0100	0,0100	0,0106	0,0642	0,0653	0,0655
-65	0,0147	0,0146	0,0153	0,0771	0,0787	0,0782
-60	0,0210	0,0209	0,0223	0,0921	0,0941	0,0940
-55	0,0294	0,0301	0,0308	0,1092	0,1116	0,1119
-50	0,0413	0,0424	0,0430	0,1290	0,1316	0,1320
-45	0,0579	0,0580	0,0591	0,1522	0,1535	0,1549
-40	0,0785	0,0786	0,0784	0,1775	0,1785	0,1800
-35	0,1039	0,1030	0,1039	0,2045	0,2062	0,2080
-30	0,1330	0,1334	0,1359	0,2356	0,2360	0,2379
-25	0,1687	0,1701	0,1736	0,2676	0,2681	0,2715
-20	0,2113	0,2145	0,2176	0,3012	0,3029	0,3051
-15	0,2580	0,2636	0,2655	0,3376	0,3383	0,3419

3 Modèle 2

Afin de tenir compte de la présence de périodes sèches et humides, nous allons utiliser, comme deuxième modèle, un mélange de modèles pour représenter la densité des apports. Nous proposons donc d'utiliser une densité de la forme

$$f(y_t | p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) = p f_0(y_t | \theta_1, \sigma_1^2) + (1 - p) f_0(y_t | \theta_2, \sigma_2^2), \quad (6)$$

où y_t est l'apport au temps t , $p \in (0; 1)$ et $f_0(\cdot | \theta, \sigma^2)$ représente la densité normale de moyenne θ et de variance σ^2 . (La forme de $f(y_t | p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$ est donnée à la figure 7 lorsque les différents paramètres sont remplacés par leur moyenne *a posteriori*.)

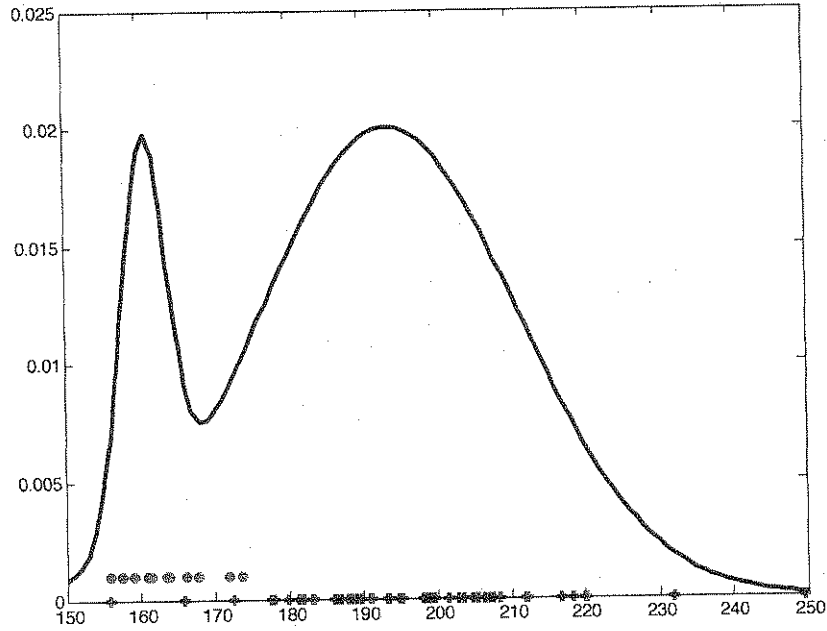


FIG. 7 – Densité utilisée pour le modèle (6) (Les points représentent les observations associées au premier élément du mélange et les + au deuxième.)

Afin de simplifier l'écriture de ce modèle, nous introduisons la variable latente z_t de densité Bernouilli de paramètre p . Ainsi, nous avons que

$$Y_t | z_t, \theta_{2-z_t}, \sigma_{2-z_t}^2 \sim N(\theta_{2-z_t}; \sigma_{2-z_t}^2),$$

et nous utiliserons l'algorithme EM pour évaluer les différentes expressions.

3.1 Modèles *a priori* et *a posteriori*

Tout comme pour le premier modèle, nous utilisons des densités conjuguées afin de simplifier les calculs. Nous avons donc :

$$\begin{aligned} p &\sim \beta e(a; b), \\ \theta_i \mid \sigma_i^2 &\sim N(\theta_{0i}; \sigma_i^2/n_0), \\ \sigma_i^2 &\sim \text{IF}(\alpha/2; \beta/2). \end{aligned}$$

En conditionnant sur $z_1, z_2, \dots, z_n$, nous pouvons montrer facilement que les densités *a posteriori* sont données par :

$$\begin{aligned} p \mid \underline{z} &\sim \beta e(a + n_1; b + n_2), \\ \theta_i \mid \underline{y}, \underline{z}, \sigma_i^2 &\sim N\left(\frac{n_0\theta_{0i} + n_i\bar{y}_{(i)}}{n_0 + n_i}; \frac{\sigma_i^2}{n_0 + n_i}\right), \\ \sigma_i^2 \mid \underline{y}, \underline{z} &\sim \text{IF}\left(\frac{\alpha + n_i}{2}; \frac{1}{2}\left[\beta + S_i + \frac{n_0n_i}{n_0 + n_i}(\bar{y}_{(i)} - \theta_{0i})^2\right]\right), \end{aligned} \tag{7}$$

où

$$\begin{aligned} n_1 &= \sum_{j=1}^n z_j, & n_2 &= \sum_{j=1}^n (1 - z_j) = n - n_1, \\ \bar{y}_{(1)} &= \frac{\sum_{j=1}^n z_j y_j}{\sum_{j=1}^n z_j}, & S_1 &= \sum_{j=1}^n z_j (y_j - \bar{y}_{(1)})^2, \\ \bar{y}_{(2)} &= \frac{\sum_{j=1}^n (1 - z_j) y_j}{\sum_{j=1}^n (1 - z_j)}, & S_2 &= \sum_{j=1}^n (1 - z_j) (y_j - \bar{y}_{(1)})^2. \end{aligned}$$

3.2 Densité prédictive

Comme à la sous-section 2.2, la densité prédictive de $y_{n+1}, \dots, y_{n+k}$ sachant $\underline{y}, \underline{z}, z_{n+1}, \dots, z_{n+k}$ est donnée par

$$\begin{aligned} m(y_{n+1}, \dots, y_{n+k} \mid \underline{y}, \underline{z}, z_{n+1}, \dots, z_{n+k}) \\ &= \frac{\text{densité conjointe de } (y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k}) \text{ sachant } \underline{z}, z_{n+1}, \dots, z_{n+k}}{\text{densité marginale de } (y_1, \dots, y_n) \text{ sachant } \underline{z}} \\ &= \frac{\int_0^1 \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n, y_{n+1}, \dots, y_{n+k} \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) \pi_1(\theta_1 \mid \sigma_1^2) \\ &\quad \times \pi_2(\sigma_1^2) \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) \pi_3(p) d\theta_1 d\theta_2 d\sigma_1^2 d\sigma_2^2 dp}{\int_0^1 \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(y_1, \dots, y_n \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) \pi_1(\theta_1 \mid \sigma_1^2) \pi_2(\sigma_1^2) \\ &\quad \times \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) \pi_3(p) d\theta_1 d\theta_2 d\sigma_1^2 d\sigma_2^2 dp} \end{aligned}$$

$$\begin{aligned}
& \times \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) \pi_3(p) d\theta_1 d\theta_2 d\sigma_1^2 d\sigma_2^2 dp \\
& = \int_0^1 \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \prod_{j=1}^{n+k} \left[p^{z_j} (1-p)^{z_j} f_0^{z_j}(y_j \mid \theta_{2-z_j}, \sigma_{2-z_j}^2) f_0^{1-z_j}(y_j \mid \theta_{2-z_j}, \sigma_{2-z_j}^2) \right] \\
& \quad \times \pi_1(\theta_1 \mid \sigma_1^2) \pi_2(\sigma_1^2) \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) \pi_3(p) d\theta_1 d\theta_2 d\sigma_1^2 d\sigma_2^2 dp \\
& \quad / \int_0^1 \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \prod_{j=1}^n \left[p^{z_j} (1-p)^{z_j} f_0^{z_j}(y_j \mid \theta_{2-z_j}, \sigma_{2-z_j}^2) f_0^{1-z_j}(y_j \mid \theta_{2-z_j}, \sigma_{2-z_j}^2) \right] \\
& \quad \times \pi_1(\theta_1 \mid \sigma_1^2) \pi_2(\sigma_1^2) \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) \pi_3(p) d\theta_1 d\theta_2 d\sigma_1^2 d\sigma_2^2 dp \\
& = \int_0^1 p^{a+n_1+\sum_{i=1}^k z_{n+i}} (1-p)^{b+n_2+\sum_{i=1}^k (1-z_{n+i})} dp \\
& \quad \times \int_0^\infty \int_{-\infty}^\infty \left[\prod_{\{j|z_j=1\}} f_0(y_j \mid \theta_1, \sigma_1^2) \right] \pi_1(\theta_1 \mid \sigma_1^2) \pi_2(\sigma_1^2) d\theta_1 d\sigma_1^2 \\
& \quad \times \int_0^\infty \int_{-\infty}^\infty \left[\prod_{\{j|z_j=0\}} f_0(y_j \mid \theta_2, \sigma_2^2) \right] \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) d\theta_2 d\sigma_2^2 \\
& \quad / \int_0^1 p^{a+n_1} (1-p)^{b+n_2} dp \int_0^\infty \int_{-\infty}^\infty \left[\prod_{\{j|z_j=1\}} f_0(y_j \mid \theta_1, \sigma_1^2) \right] \pi_1(\theta_1 \mid \sigma_1^2) \pi_2(\sigma_1^2) d\theta_1 d\sigma_1^2 \\
& \quad \times \int_0^\infty \int_{-\infty}^\infty \left[\prod_{\{j|z_j=0\}} f_0(y_j \mid \theta_2, \sigma_2^2) \right] \pi_1(\theta_2 \mid \sigma_2^2) \pi_2(\sigma_2^2) d\theta_2 d\sigma_2^2 \\
& = m_3(\underline{z}, z_{n+1}, \dots, z_{n+k}) m_1(\underline{y}, y_{n+1}, \dots, y_{n+k} \mid \underline{z}, z_{n+1}, \dots, z_{n+k}) \\
& \quad \times m_2(\underline{y}, y_{n+1}, \dots, y_{n+k} \mid \underline{z}, z_{n+1}, \dots, z_{n+k}) / m_3(\underline{z}) m_1(\underline{y} \mid \underline{z}) m_2(\underline{y} \mid \underline{z}),
\end{aligned}$$

où

$$\begin{aligned}
m_3(\underline{q}) & \propto \frac{\Gamma(m_1) \Gamma(m_2)}{\Gamma(m_1 + m_2)}, \\
m_i(\underline{x} \mid \underline{q}) & \propto \sqrt{\frac{n_0}{n_0 + n_i}} \left[\beta + S_i + \frac{n_0 n_i}{n_0 + n_i} (\bar{x}_{(i)} - \theta_{-i})^2 \right]^{-(n_i + \alpha)/2},
\end{aligned}$$

où

$$\begin{aligned}
m_1 & = \sum_j q_j, & m_2 & = \sum_j (1 - q_j), \\
\bar{x}_{(1)} & = \frac{\sum_j q_j x_j}{\sum_j q_j}, & S_1 & = \sum_j q_j (x_j - \bar{x}_{(1)})^2,
\end{aligned}$$

$$\bar{x}_{(2)} = \frac{\sum_j (1 - q_j) x_j}{\sum_j q_j}, \quad S_2 = \sum_j (1 - q_j) (x_j - \bar{x}_{(1)})^2.$$

Il est à noter que toutes les équations sont conditionnelles à $\underline{z}$. Pour obtenir les expressions non conditionnelles, il suffit de prendre l'espérance par rapport à $\underline{Z}$. Nous pouvons montrer que $Z_j \mid p, y, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2$ suit une densité de Bernoulli de paramètre

$$\frac{p f_0(y_j \mid \theta_1, \sigma_1^2)}{p f_0(y_j \mid \theta_1, \sigma_1^2) + (1 - p) f_0(y_j \mid \theta_2, \sigma_2^2)}. \quad (8)$$

3.3 Calcul de l'équation (1)

Comme à la section 2.3, l'évaluation de l'équation (1) se fait de façon similaire. Cependant, au lieu de générer des valeurs de ρ , nous générons les valeurs de $z_{n+1}, \dots, z_{n+k}$ selon la densité

$$m_3(z_{n+1}, \dots, z_{n+k} \mid \underline{z}) = \frac{m_3(z_1, \dots, z_n, z_{n+1}, \dots, z_{n+k})}{m_3(z_1, \dots, z_n)}.$$

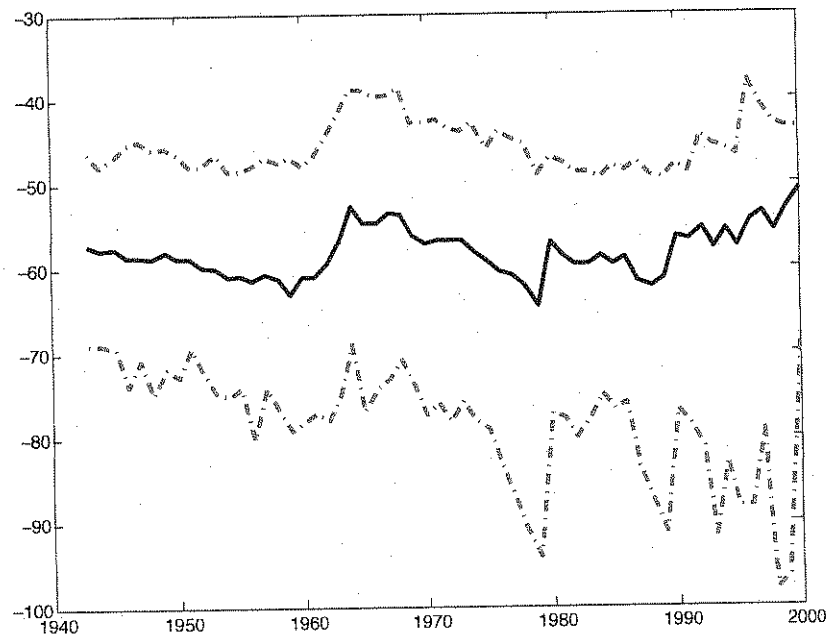
3.4 Résultats

Pour répondre à la première question, nous référons au tableau 4 et à la figure 8 (déficit estimé et son intervalle de crédibilité de niveau 95% pour un horizon de deux et quatre ans).

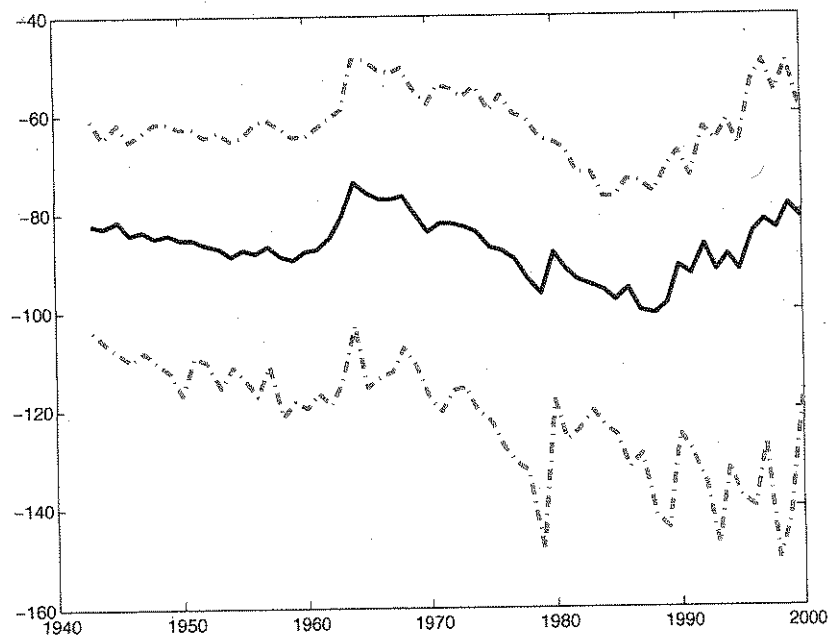
TAB. 4 – Valeurs de E_0 pour le modèle 2 et différentes années de début

Année de départ	Horizon 2 ans		Horizon 4 ans	
	Déficit	Intervalle	Déficit	Intervalle
1943	-57.3001	(-69.0940; -46.4848)	-82.0188	(-104.0000; -61.2348)
1945	-57.5862	(-69.5000; -46.7500)	-81.5119	(-108.2500; -62.0000)
1950	-58.9055	(-73.0000; -46.5940)	-85.1866	(-116.7348; -62.7500)
1955	-60.9415	(-74.2500; -48.7500)	-87.3011	(-113.5000; -64.7500)
1960	-61.0719	(-78.2500; -48.0000)	-87.8002	(-119.5000; -64.6980)
1965	-54.8828	(-77.0000; -39.0000)	-75.9539	(-115.3440; -48.7500)
1970	-57.2473	(-77.5000; -42.9480)	-83.7836	(-117.5000; -58.0000)
1975	-59.3659	(-79.0000; -46.0000)	-86.9583	(-122.2500; -59.2500)
1980	-57.0555	(-77.2500; -46.9480)	-87.9788	(-118.1980; -65.7500)
1985	-59.6007	(-76.7500; -48.0000)	-97.8080	(-125.1081; -77.0000)
1990	-56.3574	(-76.8581; -47.8440)	-90.8899	(-124.7500; -67.6980)
1995	-57.5878	(-88.2348; -46.7927)	-92.0825	(-137.7500; -66.9116)
2000	-50.9964	(-67.8440; -43.6061)	-81.6185	(-116.2500; -59.8694)

Les résultats obtenus avec le modèle 2 sont similaires à ceux obtenus avec le premier modèle. Cependant, pour un horizon de quatre ans, les valeurs estimées sont inférieures à celles du tableau 1. Cependant ces dernières se retrouvent à l'intérieur des intervalles de crédibilité du deuxième modèle et vice-versa.



(a) Modèle 2 avec horizon de 2 ans



(b) Modèle 2 avec horizon de 4 ans

FIG. 8 – Influence de la période historique sur l'estimation du déficit à partir du modèle 2

TAB. 5 – Estimés *a posteriori*

Paramètre	Cas 1	Cas 2	Cas 3
p	0,125	0,066	0,074
θ_1	160,68	161,04	161,11
θ_2	198,88	191,48	191,85
σ_1^2	9,08	1,33	4,62
σ_2^2	303,73	364,09	353,85

Pour répondre à la deuxième question, nous avons considéré, encore une fois, trois scénarios :

Cas 1 : $n_0 = 0,1$, $\alpha = \beta = 1$, $a = b = 0,5$ (densité de Jeffreys pour une proportion) ;

Cas 2 : $n_0 = 0,01$, $\alpha = \beta = 0,1$, $a = b = 1$ (densité uniforme sur l'intervalle $(0; 1)$) ;

Cas 3 : $n_0 = 0,001$, $\alpha = \beta = 0,01$, $a = b = 1$

Les estimés *a posteriori* des différents paramètres sont donnés au tableau 5 et les figures 9 et 10 donnent la probabilité du déficit énergétique sur un horizon de deux ou quatre années.

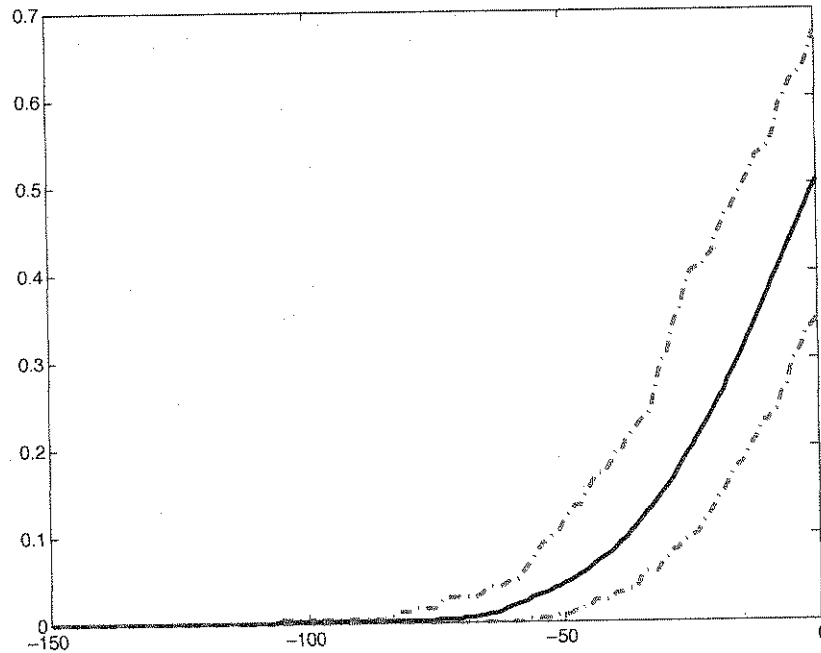


FIG. 9 – Équation (1) pour le modèle 2 et $k = 2$

Par cette table, nous pouvons voir que le scénario considéré a plus d'influence que pour le premier modèle sur l'estimation des paramètres. Cependant, l'impact final sur l'évaluation

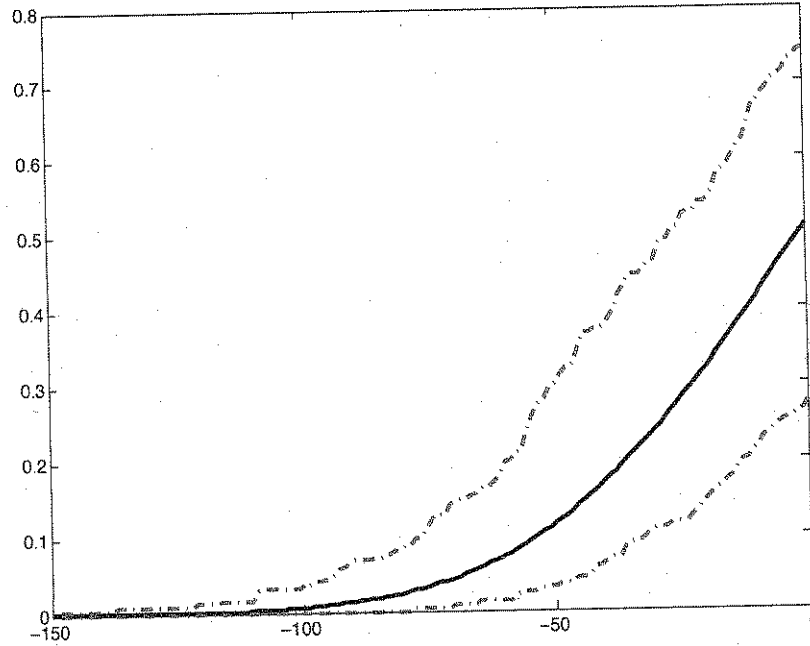


FIG. 10 – Équation (1) pour le modèle 2 $k = 4$

de l'équation (1) est négligeable (voir les tableaux 6 et 7). Au tableau 7, nous retrouvons

TAB. 6 – Valeurs de E_0 pour le modèle 2 et les trois scénarios

Scénario	Horizon 2 ans		Horizon 4 ans	
	Moyenne	Intervalle	Moyenne	Intervalle
1	-58,2	(-68,2 ; -47,0)	-82,7	(-105,0 ; -62,8)
2	-59,5	(-75,2 ; -45,8)	-85,5	(-112,0 ; -59,0)
3	-58,5	(-75,2 ; -46,0)	-83,9	(-112,0 ; -62,5)

les valeurs de l'équation (1) pour les différents scénarios. Nous pouvons constater que les probabilités sont similaires d'un cas à l'autre.

4 Conclusion

Dans ce rapport nous avons utilisé deux approches pour essayer de trouver la valeur du déficit énergétique de probabilité 0,02 pour un horizon de deux ou quatre années. Les deux modèles utilisés sont relativement insensibles aux choix des hyperparamètres et les résultats obtenus, sous ces deux modèles, sont relativement similaires. Pour un horizon de deux ans,

TAB. 7 – Valeurs de l'équation (1) pour le modèle 2 et les trois scénarios

Écart	Horizon 2 ans			Horizon 4 ans		
	Cas 1	Cas 2	Cas 3	Cas 1	Cas 2	Cas 3
-115	0,0000	0,0001	0,0000	0,0015	0,0025	0,0027
-110	0,0000	0,0001	0,0001	0,0025	0,0036	0,0039
-105	0,0000	0,0002	0,0001	0,0041	0,0056	0,0057
-100	0,0000	0,0003	0,0002	0,0063	0,0083	0,0076
-95	0,0001	0,0006	0,0004	0,0094	0,0115	0,0105
-90	0,0002	0,0009	0,0006	0,0135	0,0162	0,0146
-85	0,0004	0,0016	0,0009	0,0183	0,0221	0,0200
-80	0,0009	0,0026	0,0019	0,0247	0,0296	0,0266
-75	0,0017	0,0042	0,0033	0,0329	0,0396	0,0351
-70	0,0035	0,0068	0,0060	0,0441	0,0512	0,0458
-65	0,0076	0,0110	0,0102	0,0570	0,0649	0,0594
-60	0,0164	0,0182	0,0175	0,0733	0,0814	0,0756
-55	0,0296	0,0309	0,0288	0,0924	0,1013	0,0948
-50	0,0432	0,0448	0,0430	0,1143	0,1247	0,1180
-45	0,0611	0,0633	0,0628	0,1407	0,1507	0,1450
-40	0,0838	0,0886	0,0869	0,1711	0,1806	0,1746
-35	0,1157	0,1198	0,1175	0,2046	0,2136	0,2083
-30	0,1537	0,1593	0,1564	0,2393	0,2499	0,2445
-25	0,1998	0,2036	0,2028	0,2790	0,2895	0,2847
-20	0,2514	0,2558	0,2568	0,3216	0,3323	0,3272
-15	0,3094	0,3140	0,3148	0,3663	0,3766	0,3744

les estimés sont similaires et les intervalles de confiance ont une grande région commune. Cependant, pour un horizon de quatre années, les estimés sont légèrement différents. Le modèle 2 sous-estime systématiquement par rapport au modèle 1 mais les intervalles de confiance sont tous d'intersection non nulle.

La première question du mandat consistait à vérifier si la base des données historiques restait le meilleur indicateur de la moyenne à long terme. En examinant les tableaux 1 et 4 ainsi que les figures 4 et 8, nous voyons bien que, quelle que soit l'année de départ, les intervalles de confiance contiennent les valeurs utilisées présentement par Hydro-Québec, soit -64TWh sur deux ans et -98TWh pour quatre ans. Les déficits pour un horizon de quatre ans avec le modèle deux sont inférieures à ceux prédits par le modèle 1. Encore une fois, nous pouvons par contre constater que la courbe des déficits estimés par un modèle se trouve incluse dans l'intervalle de confiance obtenu par l'autre modèle.

La deuxième question du mandat était de prédire le « déficit » par rapport à la moyenne historique des apports énergétiques associé à une probabilité de dépassement égale à 2%. Les valeurs utilisées actuellement sont de -64 TWh sur deux ans et de -98 TWh sur quatre ans. En se basant sur les résultats des deux modèles (voir les tableaux 2 et 6), nous pouvons conclure que ces valeurs sont encore actuelles car elles sont comprises dans les intervalles de confiance. Ces valeurs sont en plus très près des valeurs obtenues par le premier modèle.