

Article

« Les progrès dans les prévisions : météorologie et économique »

John W. Galbraith

L'Actualité économique, vol. 81, n° 4, 2005, p. 559-593.

Pour citer cet article, utiliser l'adresse suivante :

<http://id.erudit.org/iderudit/014910ar>

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <http://www.erudit.org/apropos/utilisation.html>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : erudit@umontreal.ca

LES PROGRÈS DANS LES PRÉVISIONS : MÉTÉOROLOGIE ET ÉCONOMIQUE*

John W. GALBRAITH

*Département des sciences économiques
Université McGill*

RÉSUMÉ – La modélisation et la prévision en météorologie et en économique présentent un certain nombre de caractéristiques communes qui laissent à penser qu'il pourrait être intéressant de comparer leurs récents progrès en matière de prévision. Nous portons notre attention sur deux aspects de la prévision. Premièrement, nous étudions les mesures de la valeur ajoutée des prévisions et l'évolution de ces mêmes mesures au cours des 20 à 30 dernières années; deuxièmement, l'estimation et la représentation de l'incertitude de ces prévisions sont examinées. Nous considérons certaines variables quantitatives particulières comme étant représentatives de différents types de prévisions : température, croissance du PIB et volatilité des marchés financiers (variables continues); probabilité de précipitations et probabilité de récession (prévisions de probabilité); prévision de type 0/1 de précipitations ou de récession (prévisions binaires); tornades et krachs boursiers (événements rares). Nous effectuons un survol de l'information disponible à ce jour sur l'évolution de la qualité des prévisions et décrivons les méthodes en développement dans le but de comprendre et de représenter l'incertitude dans ces prévisions.

ABSTRACT – Meteorological and economic modelling and forecasting have a number of common features, which suggest that it may be interesting to compare their recent progress in forecasting. We concentrate on two aspects of forecasting: first, measures of the value added of forecasts, and the evolution of these measures over approximately the last twenty to thirty years; second, the estimation and representation of uncertainty of forecasts. We follow several particular quantities as representative of different types of forecasts: temperature, GDP growth and financial market volatility (continuous variables); probability of precipitation and probability of recession; 0/1 predictions of precipitation or recession (binary forecasts); tornadoes and market crashes (rare events). We describe the available information on the evolution of forecast quality, and describe also the evolving methods for understanding and representing uncertainty in the forecasts.

* Je suis particulièrement reconnaissant envers Gilbert Brunet et Richard Verret d'Environnement Canada ainsi qu'envers leurs collègues de Montréal et d'Edmonton, Jacques Derome de McGill et Simon Van Norden de HEC Montréal pour leur aide précieuse et pour toute l'information et les conversations que ceux-ci ont partagées avec moi. Je remercie aussi le Fonds québécois de la recherche sur la société et la culture (FQRSC), le Conseil de recherches en sciences humaines du Canada (CRSH) et le Centre interuniversitaire de recherche en analyse des organisations (CIRANO) pour leur appui général dans mes recherches, ainsi que les participants au 45^e congrès de la Société canadienne de science économique (2005) pour leurs commentaires pertinents.

INTRODUCTION

La météorologie et l'économique, en particulier la macroéconomie, ont beaucoup en commun. Chacune de ces sciences traite de grands systèmes simultanés dont la compréhension et la prédictibilité sont incomplètes, et ce même dans la mesure où les principes de l'évolution de ces systèmes sont adéquatement compris. En général, l'expérimentation à grande échelle est impossible. Les données à partir desquelles nous apprenons sur ces systèmes ne sont généralement pas d'origine expérimentale. Dans chaque cas, nous n'avons que des mesures imprécises sur le système actuel, de telle sorte que non seulement les prévisions, mais aussi la découverte de l'état actuel du système requièrent une estimation. Des prévisions peuvent être effectuées au sein des deux disciplines, à partir d'observations statistiques de l'évolution de ces systèmes ou de modèles structurels (la théorie) représentant ces systèmes, ou encore à partir d'une combinaison des deux. Les prévisionnistes en météorologie et en économique s'intéressent à des estimations ponctuelles ainsi qu'aux mesures de l'incertitude autour de ces prévisions et doivent donc composer avec les problèmes liés à la communication de cette incertitude au public. Des liens historiques existent aussi entre les deux disciplines : quelques unes des toutes premières prévisions météorologiques furent effectuées par l'économiste Stanley Jevons lors de son séjour en Australie¹.

Cette situation diffère d'avec celle qui entoure les prévisions dans d'autres sciences naturelles, où les similitudes avec les problèmes économiques sont moindres. Par exemple, les prévisions d'éruptions volcaniques, dont les progrès ont été rapides au cours des dernières années, traitent d'un phénomène circonscrit dans un lieu géographique précis pouvant être étudié à un degré de précision qui croît avec l'addition d'instruments de mesure sur le site du volcan. Les prévisions sismologiques quant à elles – bien qu'étant d'une complexité notoire – semblent tirer leur complexité de facteurs distincts de ceux applicables aux prévisions de grands systèmes simultanés.

De toute évidence, l'analogie entre l'économique et la météorologie a de sérieuses limites et les différences notables entre les deux disciplines ne sont pas difficiles à repérer. L'une d'elle repose sur le fait que les prévisions économiques de nature purement statistique (c'est-à-dire, celles basées sur des techniques de prévision en séries chronologiques, ne comportant essentiellement aucune structure provenant de la théorie économique) performant souvent aussi bien ou mieux que celles basées exclusivement sur la théorie économique. Par contre, les prévisions météorologiques actuelles de la plupart des variables quantitatives importantes surpassent indéniablement en précision les prévisions d'origine purement statistiques. Un autre contraste provient de l'existence potentielle d'anticipations engendrant elles-mêmes leur réalisation (*self-fulfilling prophecy*) en économie (par exemple, la prévision d'une récession peut causer une réduction de l'investissement et ainsi provoquer effectivement une récession), ce qui en général n'est

1. Je remercie Pierre Duguay et Gordon Fisher pour cette remarque.

pas problématique en météorologie (bien qu'il soit concevable que certaines prévisions météorologiques puissent générer une panique). Cette réalité ajoutée au fait que la valeur d'une prévision météorologique pour la société équivaut généralement à la somme de ses valeurs pour chaque individu – ce qui n'est pas le cas en économique – explique peut-être en partie les différences quant à l'origine publique ou privée de ces prévisions. Remarquez par exemple que les prévisions de tempêtes sévères proviennent d'organismes des gouvernements centraux dans tous les pays industrialisés alors que la prévision d'une récession n'est habituellement pas émise par les gouvernements².

Néanmoins, autant les objectifs que la direction du perfectionnement méthodologique de ces deux sciences partagent des similitudes substantielles. Il est donc naturel de considérer les progrès accomplis par nos collègues en météorologie comme autant de critères d'évaluation pour nos propres avancées. En gardant ceci à l'esprit, cet essai portera son attention sur les progrès accomplis en matière de prévision en économique et en météorologie au cours des dernières 20 à 30 années, en posant les questions suivantes :

- (i) Si nous avons une série chronologique des mesures précises de la valeur ajoutée de nos prévisions, y verrions-nous une tendance claire?
- (ii) Avons-nous accru de façon utile notre compréhension de l'incertitude dans nos prévisions?

Nous aurons besoin pour ce faire de considérer quelques exemples concrets de problèmes liés à la prévision dans chacune des deux disciplines, exemples illustrant jusqu'à un certain point le genre d'exercices de prévision qui sont entrepris. Nous allons donc nous référer à plusieurs types de problèmes de prévision survenant au sein de chaque discipline, avec des variables de prévision spécifiques provenant des catégories suivantes :

- Prévisions de variables continues : ex. température, pression atmosphérique, croissance du PIB, volatilité des rendements des actifs.
- Prévisions de probabilité : précipitations, récession.
- Prévisions binaires : précipitations, récession, oui ou non?
- Avertissement d'événements extrêmes (prévisions binaires d'événements rares) : tornade, krach boursier.

Dans au moins quelques-uns de ces cas, il est possible de caractériser de manière raisonnablement précise les progrès accomplis en prévision – d'où, en lien avec notre première question, la nécessité de mesures de la qualité et de la

2. L'existence de jeux à somme nulle dans la vie économique offre une justification, autre que le problème de la fourniture des biens publics, de la divergence possible entre la valeur sociale des prévisions et les dépenses individuelles pour celles-ci. Cette distinction peut toutefois s'effriter avec l'introduction de produits dérivés liés à la météo et donc d'occasions de profiter d'une information supérieure.

valeur d'une prévision. Nous tenterons aussi de traiter des contraintes à l'avancement du progrès dans ce domaine. Afin de répondre à la seconde question, nous devons examiner les développements dans la provision d'une caractérisation complète d'une distribution des résultats – par opposition aux seules prévisions ponctuelles – et autres améliorations dans notre compréhension de l'incertitude. Nous ne tenterons *pas* de fournir une étude des développements dans les techniques de prévision en tant que telles.

Le plan de cet essai se présente donc comme suit. Nous commencerons par considérer, dans la première section, une variété de mesures de précision ou de valeurs des prévisions applicables aux différents types de problèmes décrits précédemment. Dans la deuxième section, nous envisagerons l'évolution historique de ces mesures et de ces valeurs en météorologie, et nous effectuerons, dans la mesure du possible, des comparaisons avec l'économie. Aussi dans cette section, nous traiterons brièvement des contraintes limitant les avancées dans ce domaine. Dans la troisième section, nous aborderons les questions liées à la caractérisation de l'incertitude : caractérisation plus complète de la distribution, procédés permettant d'obtenir des mesures d'incertitude et méthodes de présentation de ces mesures au public. La dernière section offrira quelques conclusions générales ainsi que quelques suggestions de nature plus spéculatives.

1. MESURES DE LA QUALITÉ DES PRÉVISIONS

Afin de mesurer les tendances dans l'évolution de la précision des prévisions, nous devons tout d'abord clarifier ce qu'on entend exactement par précision. La réponse dépend du type de prévision à l'étude. Avant de débiter, nous remarquons que les mesures devant être décrites ci-dessous sont des quantités calculables à partir d'échantillons (telle une fréquence relative). Celles-ci, moyennant certaines conditions incluant la stationnarité stricte du processus, convergent vers les quantités correspondantes de la population (telle une probabilité ou une probabilité conditionnelle).

Un concept clé dans la plupart de ces mesures est celui de l'« aptitude » de la prévision, qui mesure la réduction proportionnelle dans une fonction de perte par rapport à une prévision de référence. Les scores d'aptitude ont été utilisés au moins depuis Brier (1950); voir Murphy et Winkler (1987), entres autres, pour un survol de la littérature pertinente. Une fonction de perte typique est l'erreur quadratique moyenne (EQM), et la prévision de référence habituelle en météorologie est la « climatologie », c'est-à-dire, les caractéristiques de la distribution non conditionnelle d'une variable mesurée sur une longue période, ex. la température moyenne à une date particulière estimée à partir de données sur une période de 100 ans. Plus précisément,

$$\text{Aptitude} \sim 1 - \frac{M_{\hat{f}}}{M_{ref}} \quad (1)$$

avec $M \equiv$ mesure de perte; $\hat{f} \equiv$ prévision; $ref \equiv$ référence (valeur climatologique; moyenne non conditionnelle). Cette valeur est inférieure à un, sauf dans le cas limite d'une prévision exacte, et normalement supérieure à zéro; une valeur négative indiquerait une prévision pire que celle obtenue en utilisant simplement la valeur de référence (une moyenne par exemple). Galbraith (2003) est arrivé à une définition similaire du contenu (*content*) d'une prévision. Il donne une expression analytique du contenu ou de l'aptitude de processus autorégressifs pour lesquels l'estimation des paramètres se fonde sur la perte de l'erreur quadratique moyenne³.

Clairement, une diminution de la perte de la prévision implique une aptitude augmentant vers l'unité.

Ces mesures peuvent non seulement être utilisées avec la première catégorie de prévisions – celle des variables continues telles la température ou la croissance du PIB – mais aussi avec les prévisions de probabilité. Pour ce dernier cas, Brier donne la définition

$$M_{\hat{y}}^B \equiv N^{-1} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad y_i = 0 / 1 \tag{2}$$

où \hat{y}_i est la probabilité estimée que l'événement y_i se produise ($y = 1$). Cela donne un équivalent à l'EQM pour les prévisions de probabilité que l'on peut utiliser dans la formule (1). Le critère de référence dans ce cas – M_{ref} – est obtenu en utilisant la probabilité non conditionnelle de y à la place de la probabilité estimée \hat{y}_i :

$$M_{ref}^B \equiv N^{-1} \sum_{i=1}^N (\bar{y} - y_i)^2 .$$

Remarquez qu'aussi bien pour les prévisions de probabilité que pour celles de niveau, l'aptitude aura tendance à diminuer lorsque l'on considère les prévisions de conditions dans un futur plus éloigné. D'ailleurs, on peut avantageusement mesurer l'horizon le plus lointain pour lequel une aptitude positive sera accessible à partir de nos méthodes de prévisions (cet horizon est appelé « *content horizon* » par Galbraith en 2003).

Une décomposition utile de (2) pour les prévisions de probabilité a été proposée par Murphy (1973). Supposons qu'un nombre L de catégories de probabilités différentes font l'objet d'une prévision, indexé par $\ell = 1, \dots, L$, ex. 0, 0,1, 0,2, ... 1; la probabilité non conditionnelle d'une réalisation étant donné que la prévision était ℓ est $\bar{y}_\ell = p(y = 1 | \hat{y}_\ell)$ et la décomposition de Murphy devient :

$$M_{\hat{y}}^B = N^{-1} \sum_{\ell=1}^L N_\ell (\hat{y}_\ell - \bar{y}_\ell)^2 - N^{-1} \sum_{\ell=1}^L N_\ell (\bar{y}_\ell - \bar{y})^2 + \bar{y}(1 - \bar{y}) \tag{3}$$

3. À l'opposé, la célèbre coefficient U de Theil divise la racine carrée de l'EQM par la racine carrée de la variation totale, le but étant d'extraire l'effet d'échelle plutôt que d'effectuer une comparaison directe avec la prévision de référence.

laquelle est interprétée comme fiabilité – résolution + « incertitude » (ce dernier terme étant la variance de la binomiale avec probabilité binomiale égale à \bar{y} , de telle sorte que sa valeur se situe dans l'intervalle (0, 0,25)).

Les prévisions binaires – où \hat{y} et y sont des variables de type 0/1 – peuvent aussi être évaluées en utilisant ces formules; il s'agit alors du cas particulier dans lequel $L = 2$ catégories de prévision de probabilité. Cependant, il existe d'autres mesures, bien adaptées à la prévision d'événements binaires rares, et ciblant spécifiquement les prévisions de type dichotomique. Afin de comprendre pourquoi le cas d'un événement rare est mal décrit par l'utilisation de mesures telles (2), imaginons l'exemple fictif de prévisions quotidiennes de tornades au centre-ville de Montréal. Chaque matin, on pourrait simplement annoncer « pas de tornade ». Puisque les tornades au centre-ville de Montréal sont extrêmement rares, cette prévision serait correcte presque à chaque jour, et la perte selon (2) serait près de zéro (bien sûr, puisque les prévisions climatologiques seraient très similaires, l'aptitude établie à partir de cette mesure, par (1), serait aussi près de zéro). Ces prévisions ne sont ainsi d'aucune utilité pour quiconque sachant que la probabilité non conditionnelle d'une tornade au centre-ville de Montréal est très faible. De plus, si l'on compare deux séquences de prévisions ne comprenant chacune qu'une très rare prévision d'une tornade, ces mesures seront très proches pour chacune de ces séquences.

On arrive au même résultat lorsque l'on utilise à la place de (2) une mesure telle que la proportion de prévisions exactes : la mesure semble très favorable (presque toutes les prévisions sont correctes), mais celle-ci est incapable de signaler les limitations des prévisions et est incapable de distinguer de façon significative des prévisions qui ne diffèrent entre elles qu'en de rares occasions.

Un certain nombre d'autres mesures accordant plus d'importance à différents aspects intéressants des résultats de prévisions binaires ont donc été proposées. Considérez le tableau de contingence 2×2 contenant les prévisions et les résultats pour un événement binaire :

		Événement observé	
		Oui	Non
Événement prévu	Oui	$n^{O/O}$	$n^{O/N}$
	Non	$n^{N/O}$	$n^{N/N}$

ou le premier exposant indique la prévision et le second l'état actuel du monde, ex. $n^{O/N}$ (oui | non) indique une prévision qu'il y aurait un événement observé dans le cas où, en fait, il n'y en a eu aucun, etc. Le nombre total de prévisions correspond à la somme de toutes les entrées du tableau. Nous pouvons alors calculer des mesures telles que (les termes usuels anglais sont entre parenthèses) :

Probabilité de détection (*probability of detection, hit rate*) : $\frac{n^{O/O}}{n^{O/O} + n^{N/O}}$;

Probabilité de fausse détection (*probability of false detection*) : $\frac{n^{O/N}}{n^{O/N} + n^{N/N}}$;

Taux de fausses alarmes (*false alarm rate*) : $\frac{n^{O/N}}{n^{O/O} + n^{O/N}}$;

Indice de menace (*threat score, critical success index*) : $\frac{n^{O/O}}{n^{O/O} + n^{O/N} + n^{N/O}}$.

La probabilité de détection considère la proportion des prévisions exactes sur le nombre de réalisations effectives de l'événement, (ex. sur 100 tornades, 30 ont été prédites). La probabilité de fausse détection donne quant à elle la proportion de prévisions inexactes de l'événement sur le nombre de fois où l'événement ne s'est pas produit (ex. sur 1 000 jours sans tornades, 20 jours de tornades ont été prédites). Remarquez que cette mesure diffère du taux de fausses alarmes, que l'on mesure par la proportion de prévisions inexactes d'un événement sur son nombre de prévisions (ex. sur 100 tornades prédites, il y a eu 35 cas où aucune n'est survenue).

L'indice de menace est peut-être moins transparent que les trois mesures précédentes. Afin de saisir son utilité, il nous faut revenir au problème qui a précédemment attiré notre attention : il est facile d'obtenir une importante proportion de prévisions exactes lorsqu'il s'agit d'événements rares, puisque la plupart des prévisions négatives s'avèreront correctes. C'est-à-dire, $n^{N/N}$ est un grand nombre, mais qui n'est pas tellement indicatif de l'aptitude ou de l'utilité d'une prévision. L'indice de menace élimine ces prévisions « faciles » : il donne la proportion de prévisions exactes d'un événement par rapport à toutes les prévisions sauf celles-ci. Il s'agit donc de la proportion de prévisions réussies par rapport au nombre de cas où l'événement a été prédit, s'est produit ou encore les deux.

Une dernière mesure que nous considérerons consiste en un ensemble de nombres exprimé à travers un graphique, le diagramme ROC (*Receiver Operating Characteristics*); voir le graphique 8 en dessous pour des exemples. Ce graphique trace simplement la probabilité de détection d'un signal pour une probabilité de fausse détection donnée, quelle qu'elle soit. Dans un contexte d'inférence statistique, l'utilisation d'un test de niveau élevé correspondra en général à une puissance élevée. De même, dans le cas présent, si nous acceptons une probabilité de fausse détection plus élevée, nous serons généralement en mesure d'obtenir une probabilité de détection correcte plus élevée. Une courbe située en haut et à gauche d'une autre implique un système d'alarme plus désirable (c'est-à-dire une meilleure prévision binaire).

D'autres mesures sont disponibles, notamment l'indice de Heidke (*Heidke Skill Score*) et l'indice de Hanssen-Kuipers (*Hanssen-Kuipers discriminant*); nous mettrons cependant l'emphase sur les mesures données plus haut à cause de leur disponibilité pour des séries chronologiques ayant été recensées en météorologie, dans les cas qui nous intéressent.

Nous avons jusqu'ici décrit certaines mesures de succès pour chacun des types de prévision mentionnés précédemment (continue, probabilité et binaire, incluant les prévisions binaires d'événements rares). En considérant l'évolution dans le temps de ces mesures, nous disposons maintenant d'un outil pouvant nous aider à répondre à la question posée dans l'introduction. Dans le cas des prévisions météorologiques, un certain nombre d'études de cette évolution ont été produites; pour ce qui est des variables économiques, l'information est plus difficile à dénicher.

2. ÉVOLUTION HISTORIQUE DES MESURES

La littérature météorologique contient un certain nombre d'études sur l'évolution temporelle de la précision des prévisions, mesurée comme nous l'avons tout juste décrit. Bien que ces études n'abordent pas d'ordinaire la signification statistique des améliorations dans les mesures, gardons à l'esprit que la fréquence de plusieurs prévisions météorologiques est au moins quotidienne, ce qui rend disponible pour évaluation un très grand nombre de points d'échantillon.

Nous considérerons ici les progrès accomplis dans le contexte d'exemples particuliers, mentionnés plus haut, en débutant par les prévisions de variables continues telles la température et la pression atmosphérique.

2.1 Variables météorologiques continues

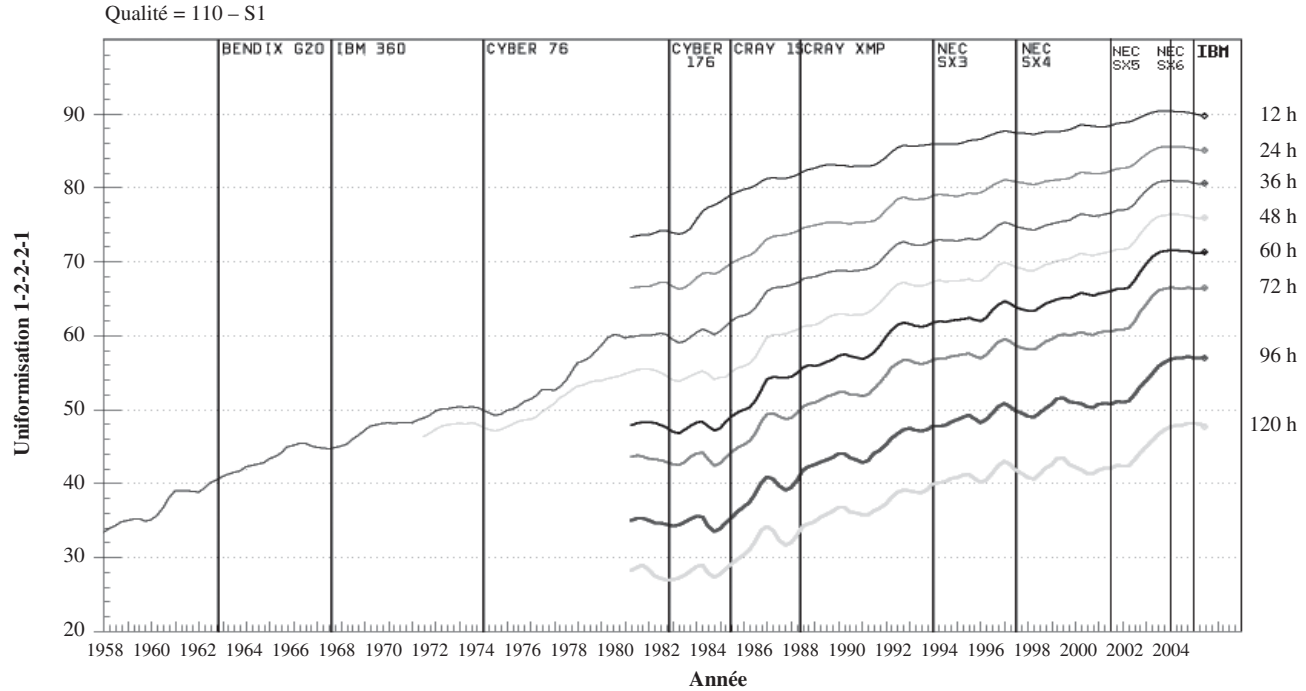
Une des séquences les plus explicites de semblables mesures est calculée par le Centre météorologique canadien à Dorval⁴. Les graphiques 1 et 2 représentent graphiquement des séquences de mesures d'un score S_1 ⁵, rapporté comme $110 - S_1$, de telle sorte que des nombres élevés indiquent une amélioration. Le graphique 1 comprend une séquence d'horizons de prévision de 12 heures jusqu'à 120 heures (5 jours), et le graphique 2 inclut les horizons de 5 à 10 jours.

4. Je suis très reconnaissant envers Richard Verret et ces collègues pour avoir produit pour les fins de cette présentation une mise à jour jusqu'au début 2005 de ce que j'ai appelé les graphiques 1 et 2.

5. Le score S_1 est défini sur un champ de points et mesure les erreurs dans les gradients (taux de variation) entre les points du quadrillage, comparant les gradients dans la prévision avec ceux dans les champs observés. Voir par exemple Wilks (1995) pour une formule.

GRAPHIQUE 1

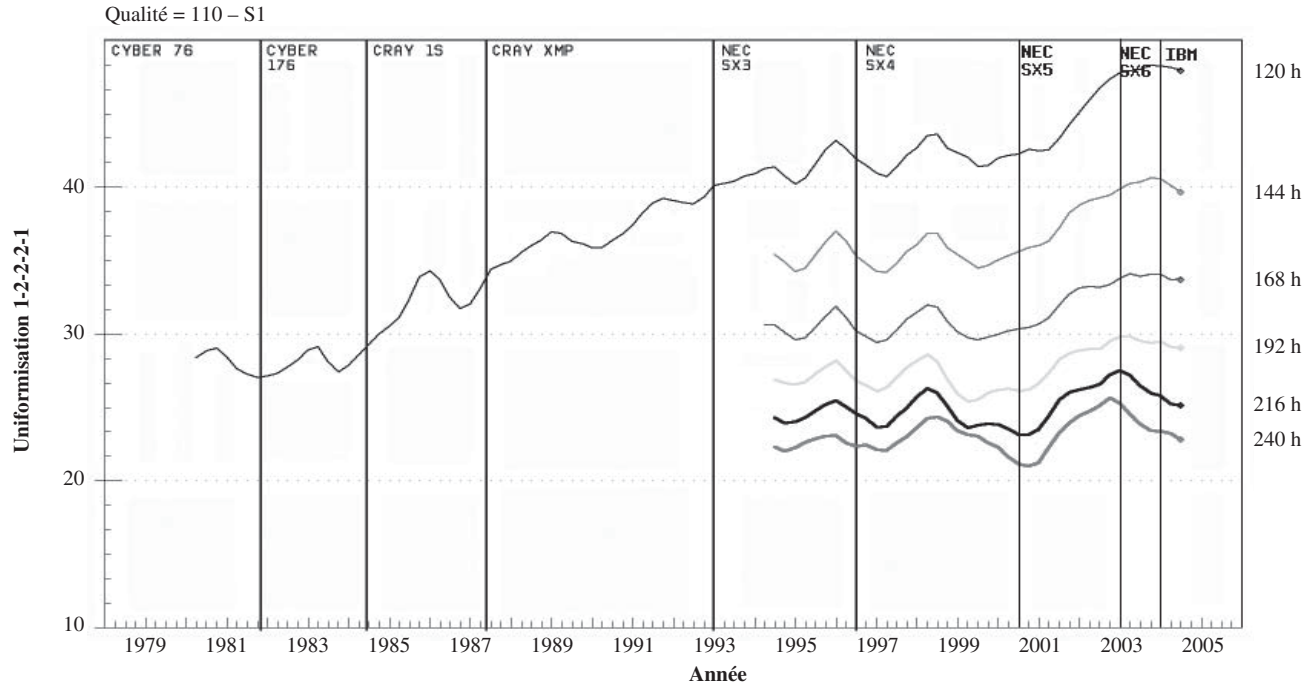
PRÉVISIONS DE LA PRESSION ATMOSPHÉRIQUE AU NIVEAU DE LA MER QUALITÉ DE LA PRÉVISION – VALEURS TRIMESTRIELLES



SOURCE : Centre météorologique canadien de Dorval

GRAPHIQUE 2

PRÉVISIONS DE LA PRESSION ATMOSPHÉRIQUE AU NIVEAU DE LA MER
 QUALITÉ DE LA PRÉVISION – VALEURS TRIMESTRIELLES



SOURCE : Centre météorologique canadien de Dorval

Ces séquences évaluent des prévisions de la valeur moyenne de la pression atmosphérique au niveau de la mer, une variable quantitative étroitement liée à la température et fréquemment étudiée. L'évaluation de prévisions 36 et 48 heures date respectivement de 1958 et 1971; les horizons jusqu'à 5 jours sont enregistrés à partir de 1981 et l'évaluation de prévisions d'horizons plus lointains dans le graphique 2 date de 1994. Remarquez que, évidemment, la mesure de précision décroît de manière monotone lorsque que l'on se déplace vers des horizons plus longs.

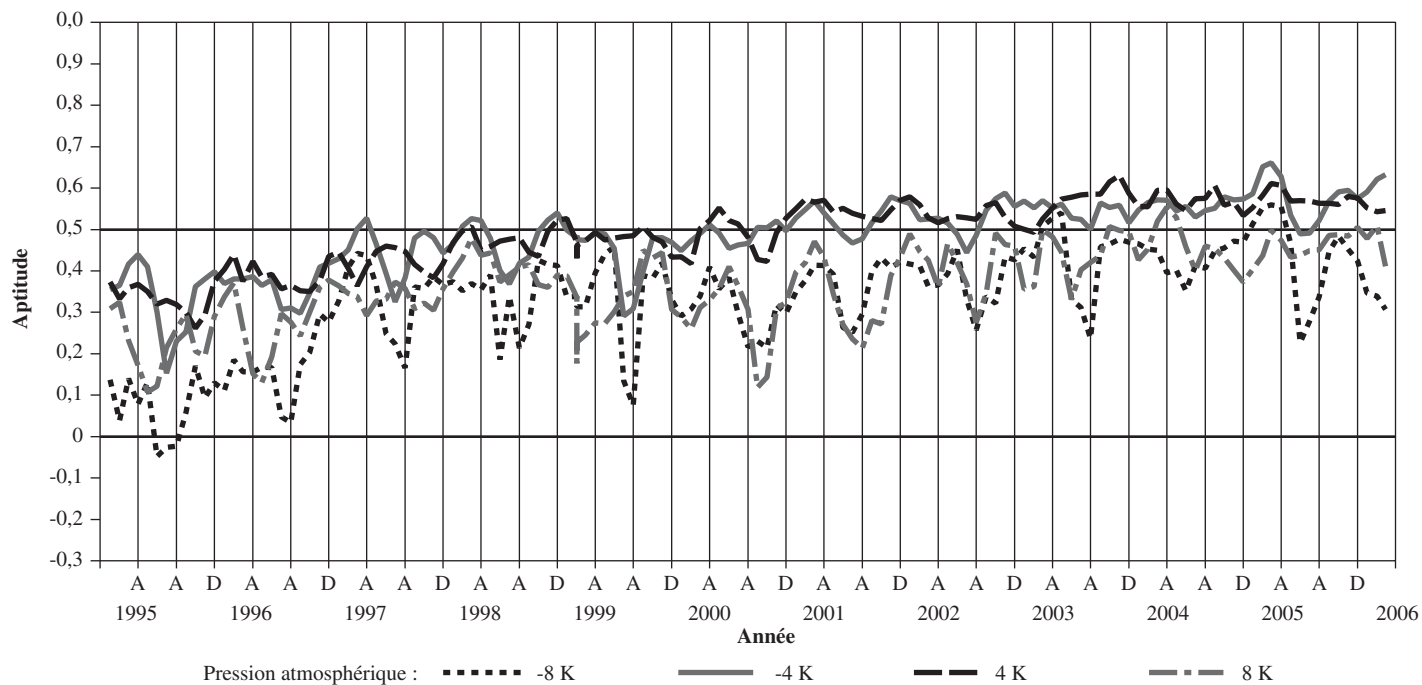
Dans cette mesure de la précision, une tendance à la hausse émerge clairement sur les prévisions jusqu'à cinq jours. Sur les prévisions 6 à 10 jours, pour lesquelles un intervalle de temps plus court est disponible pour évaluation, les signes d'une augmentation de la précision sont moins clairs, mais les estimations enregistrent certainement une amélioration des résultats depuis 2000 pour des horizons de 6 à 7 jours. Les scores S_1 ont généralement diminué de 20-25 points depuis 1980 (c'est-à-dire $110 - S_1$ a augmenté de cette quantité) pour les prévisions 1 à 5 jours.

Le prochain exemple provient du European Centre for Medium-Range Weather Forecasting (ECMWF) situé à Reading en Angleterre (voir Ghelli, 2004). Le graphique 3 rapporte les scores d'aptitude pour les prévisions de la température (à une pression atmosphérique donnée) 96 heures (4 jours) et 144 heures (8 jours). Bien que ces séries chronologiques ne couvrent qu'une période d'environ 10 ans, on perçoit clairement une tendance à l'amélioration. Remarquez que l'aptitude, de manière évidente, est non négligeable même pour l'horizon de huit jours (ce qui veut dire que même les prévisions huit jours montrent clairement une certaine valeur par rapport à la simple connaissance de la température moyenne à une date donnée).

Comme dernier exemple, le graphique 4 (graphique 5, p. 407, dans Carter *et al.*, 1989) indique l'erreur absolue moyenne de prévisions de température maximale 24 et 48 heures. Quoique celles-ci ne soient pas des scores d'aptitude – nous serions intéressés à comparer ces EAM avec celles provenant de prévisions basées sur la moyenne à long terme – elles indiquent encore une fois une tendance claire d'amélioration de la précision des prévisions.

GRAPHIQUE 3

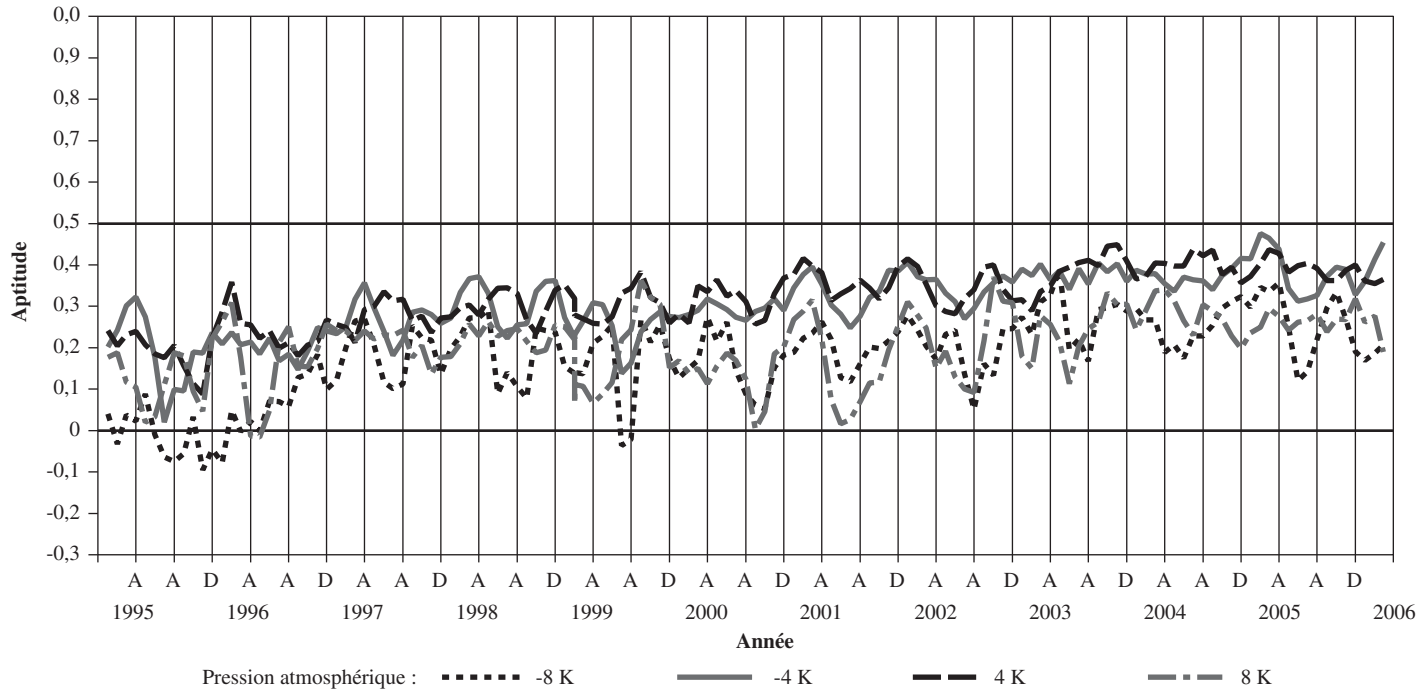
APTITUDE, MESURE DE BRIER
PRÉVISIONS DE TEMPÉRATURE POUR L'EUROPE, 96 HEURES À L'AVANCE



Source : Ghelli (2004)

GRAPHIQUE 3 (suite)

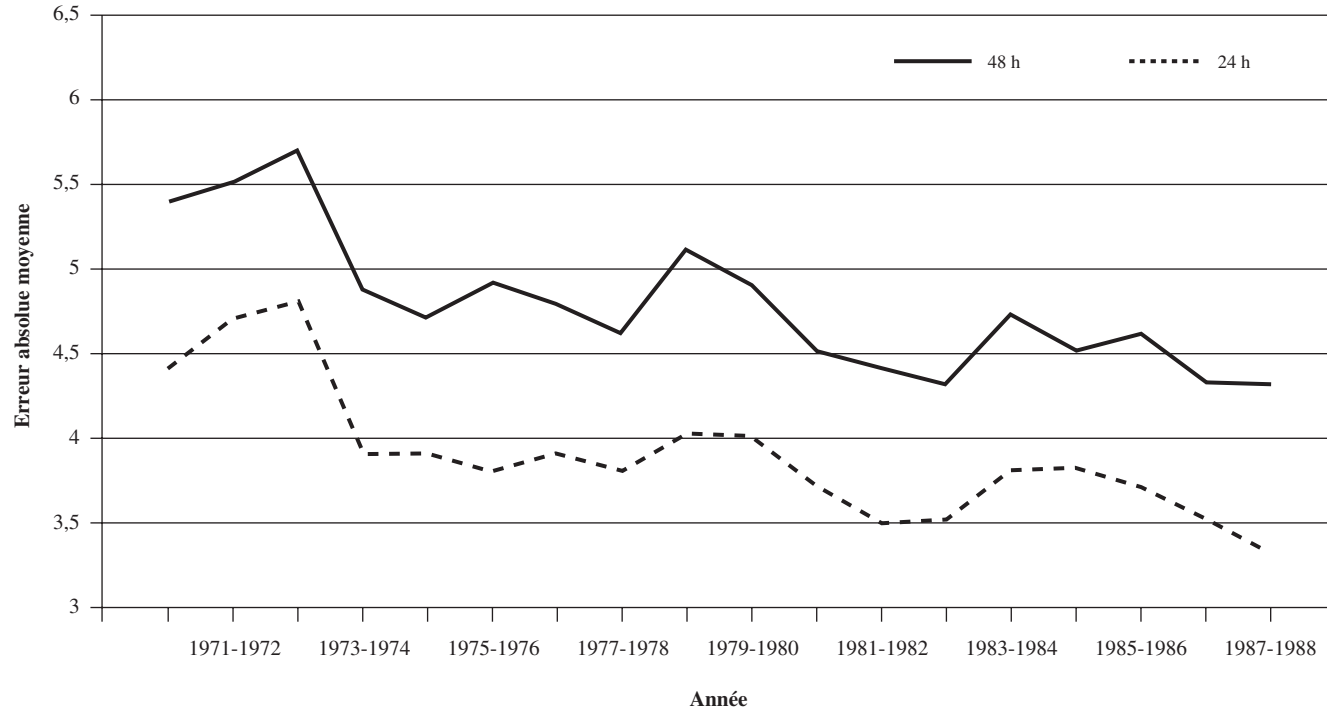
APTITUDE, MESURE DE BRIER
PRÉVISIONS DE TEMPÉRATURE POUR L'EUROPE, 144 HEURES À L'AVANCE



SOURCE : Ghelli (2004)

GRAPHIQUE 4

ERREUR ABSOLUE MOYENNE – PRÉCISIONS DE TEMPÉRATURE MAXIMALE

SOURCE : Carter *et al.* (1989), graphique 5

2.2 Prévisions de précipitations, binaires et de probabilité

On remarque aussi un progrès dans les prévisions de probabilité de précipitations. Le graphique 5 (encore Ghelli, 2004) rapporte les scores d'aptitude pour les probabilités de précipitations excédant un seuil donné (1 mm à 20 mm). Tous ces scores d'aptitude montrent une amélioration marquée dans les mesures sur une période de 10 ans; on remarquera toutefois que l'aptitude mesurée était près de 0 ou même négative durant certaines périodes dans le passé (rappelons ici qu'une mesure d'aptitude négative indique que des prévisions basées simplement sur la probabilité de précipitations non conditionnelle pour une date donnée auraient été meilleures que les prévisions effectives).

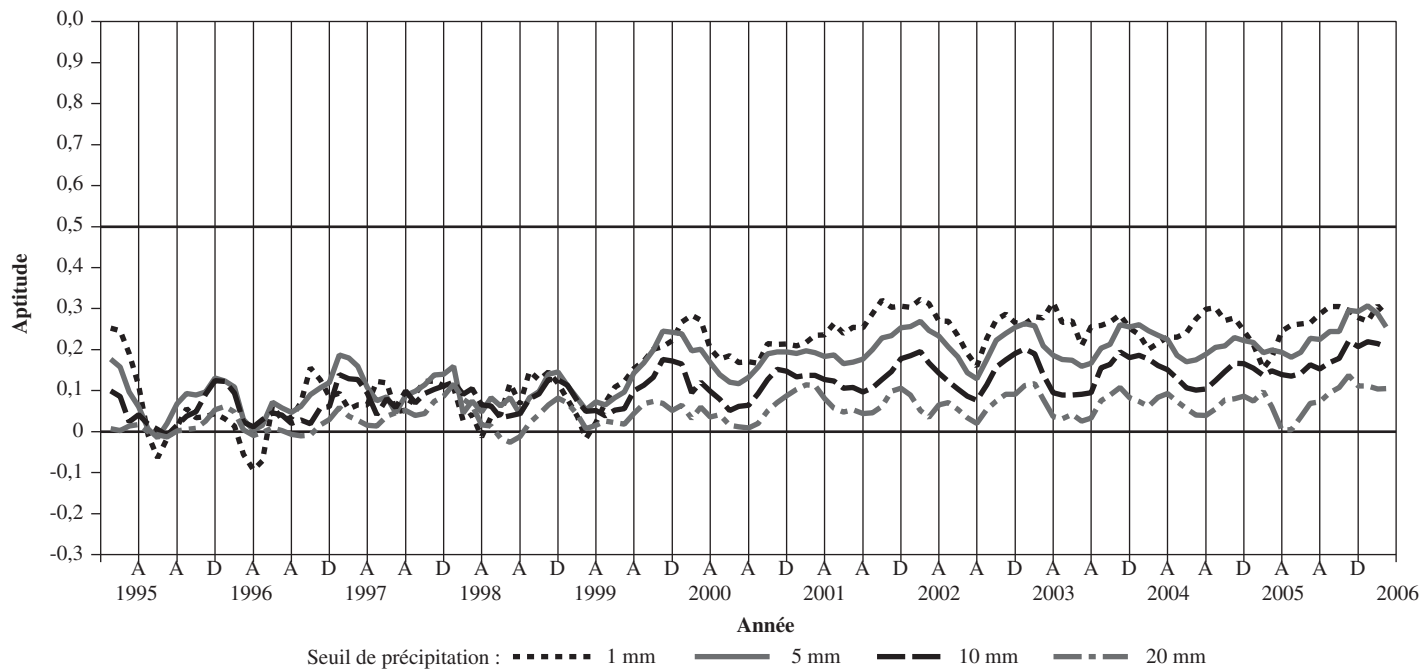
Le graphique 6 (Bosart, 2003, graphique 1, p. 521) montre (entre autres) les indices de menaces pour des prévisions binaires de précipitations cumulatives sur 24 heures, excédant un seuil donné à des horizons de 0-1 et 1-2 jours, aux États-Unis (48 états), de 1961 jusqu'à 2001. Bien qu'encore une fois, l'inférence statistique sur la pente estimée ne soit pas fournie, une tendance ascensionnelle claire est discernable; l'indice de menace pour un horizon de 1 jour, par exemple, augmente à partir de valeurs typiques aux environs de 0,18 durant les années soixante, jusqu'à 0,25 pour les données les plus récentes.

2.3 Avertissements de tornades

Le travail de pionnier de Finley (1884) a été étudié par nombre d'auteurs, y compris Murphy (1996), et la valeur d'un certain nombre de mesures pour prévisions binaires ont été calculées à partir du tableau de contingences 2×2 des avertissements de Finley. On peut donc attirer l'attention, pour des fins de comparaison avec des résultats récents, sur le fait que dans l'étude de Finley, l'indice de menace était de 0,228, la probabilité de détection de 0,549 et le taux de fausses alarmes de 0,720.

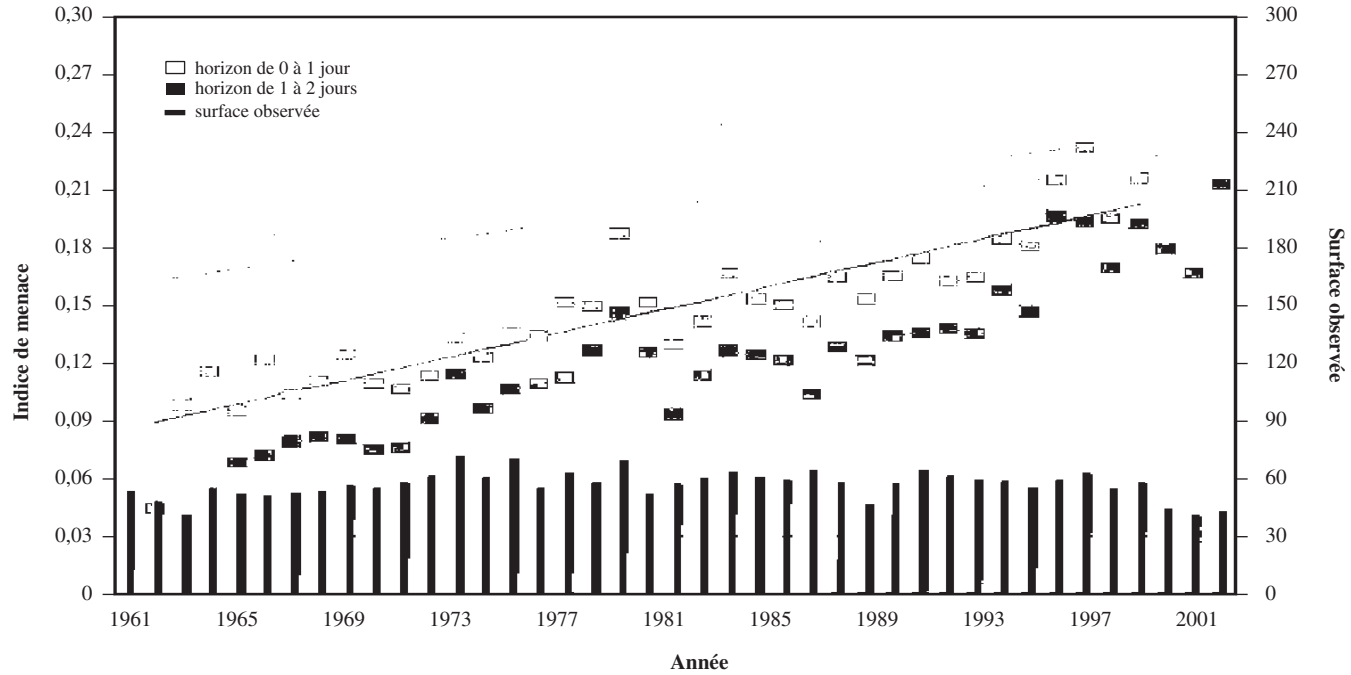
Le graphique 7a/b (graphiques 7 et 8 de Waldstreicher, 2005, p. 198) montre la probabilité de détection et le taux de fausses alarmes pour les avertissements de tornades dans l'est des États-Unis (de même que des données séparées pour la région de Raleigh, NC). Bien sûr, ces données ne sont pas exactement comparables avec celles de Finley, dont la couverture régionale ne correspondait pas précisément à cette définition de l'est des États-Unis. Néanmoins, il est intéressant de constater que, bien qu'il existe certaines preuves d'un progrès dans la probabilité de détection pendant l'intervalle de temps couvert ici, la valeur de celle-ci reste très similaire à celle de 0,55 à laquelle est arrivé Finley. Le taux de fausses alarmes, à peu près stable durant cette période est également semblable à celui trouvé par Finley (en fait légèrement plus élevé). Le graphique 7b rapporte les délais entre la prévision d'une tornade et sa réalisation (*lead times*); une mesure pour laquelle, sur l'ensemble de l'est des États-Unis, les progrès apparaissent limités.

GRAPHIQUE 5
APTITUDE, MESURE DE BRIER
PRÉVISIONS DE PRÉCIPITATION 4 JOURS À L'AVANCE



SOURCE : Ghelli (2004)

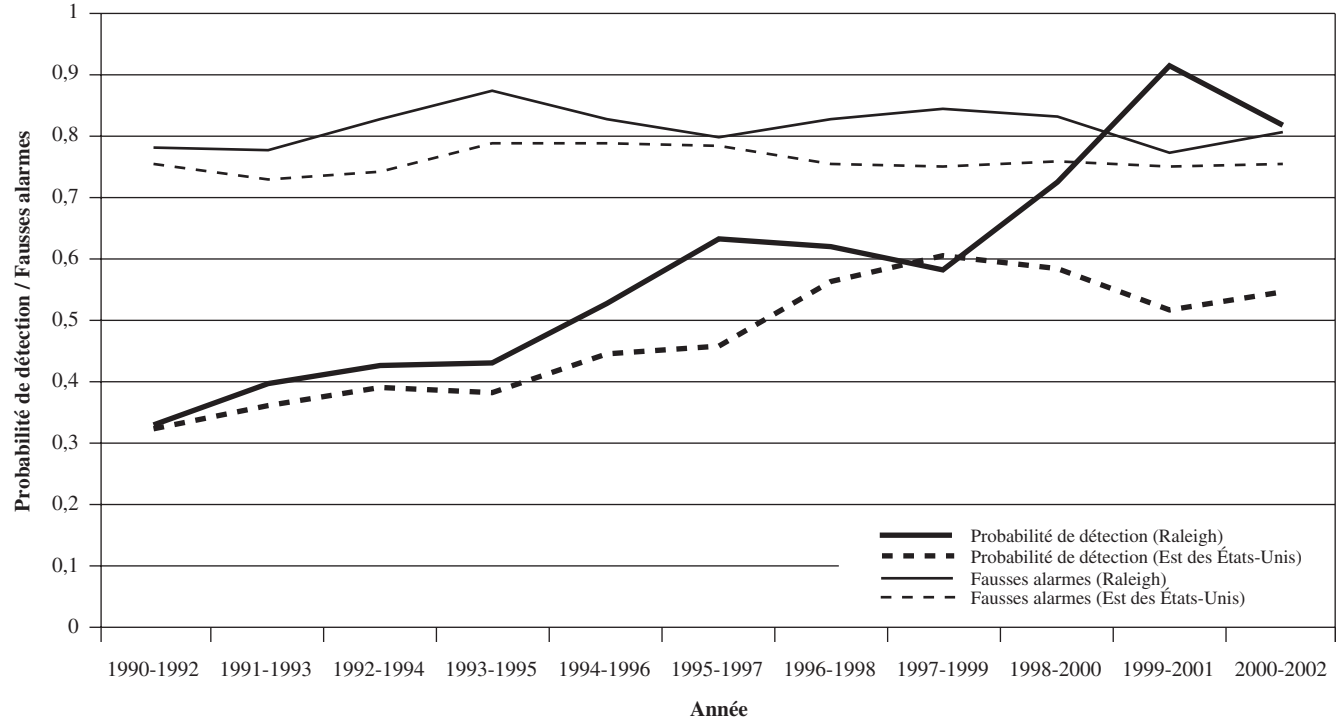
GRAPHIQUE 6
 INDICE DE MENACE
 PRÉCIPITATION CUMULATIVE SUR 24 HEURES, ÉTATS-UNIS (HORIZON DIFFÉRENT)



SOURCE : Bosart (2003), graphique 1

GRAPHIQUE 7A

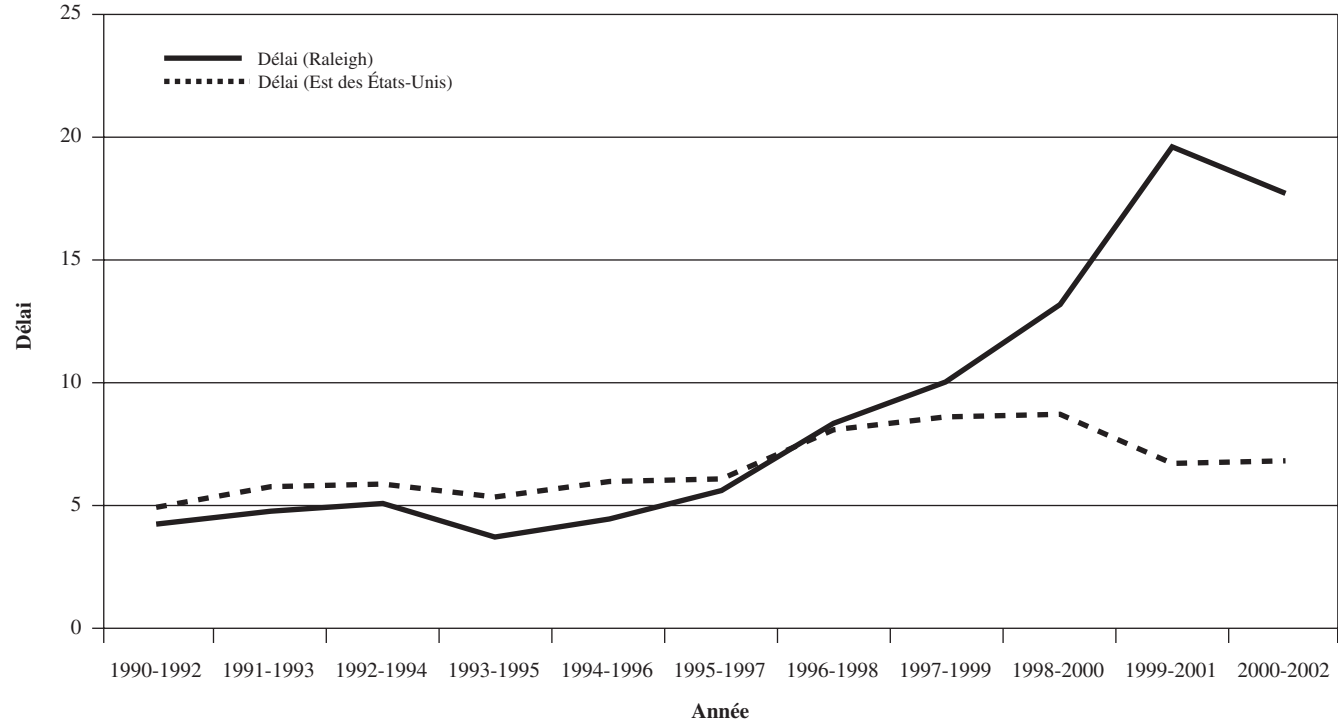
PROBABILITÉS DE DÉTECTION ET TAUX DE FAUSSE ALARME, TORNADES



SOURCE : Waldstreicher (2005), graphique 7

GRAPHIQUE 7B

DÉLAI ENTRE PRÉVISION ET RÉALISATION, TORNADES, RALEIGH, CAROLINE DU NORD ET RÉGIONS DE L'EST DES ÉTATS-UNIS

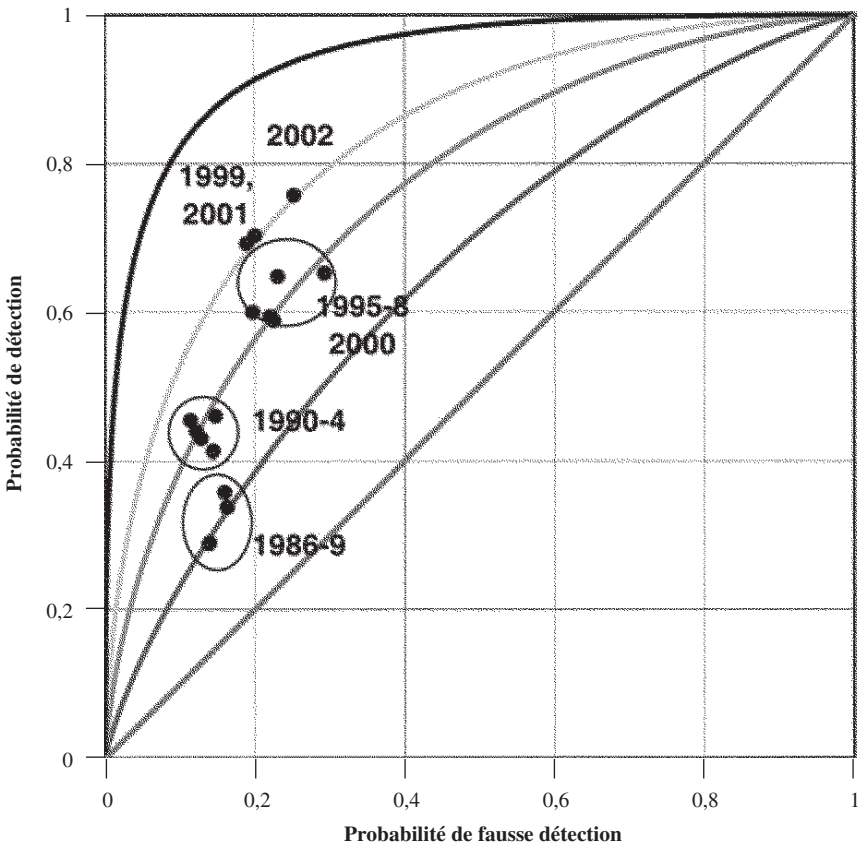


SOURCE : Waldstreicher (2005), graphique 8

Le graphique 8 (graphique 5, p. 842, dans Brooks, 2004) montre une séquence de points sur un intervalle historique similaire, représentée graphiquement avec une courbe ROC hypothétique passant près de chaque regroupement d'observations. Ces observations proviennent du National Weather Service des États-Unis (NWS). Elles décrivent les performances pour l'ensemble des États-Unis et elles tendent à suggérer une amélioration dans les ROCs – gardant à l'esprit que chaque point représente un groupe de plusieurs prévisions.

Néanmoins, les valeurs trouvées par Finley, antérieures à 1884, correspondraient ici à une des courbes intermédiaires. Globalement, cet examen n'évoque pas d'immenses progrès sur un long intervalle de temps. Il suggère plutôt que la prévision de tornades est une activité ou l'accomplissement d'un certain progrès

GRAPHIQUE 8
COURBES ROC



SOURCE : Brooks (2004), graphique 5

est assez aisé, mais où il est difficile de raffiner les prévisions jusqu'à l'obtention d'un haut niveau de précision. Cependant, les récents développements dans le radar Doppler laissent croire que des progrès appréciables pourraient être réalisés prochainement; nous y reviendrons plus loin.

Dans l'ensemble, la documentation sur l'évaluation météorologique présente deux caractéristiques dignes d'attention : d'une part une évaluation régulière et routinière de la qualité des prévisions et, d'autre part, des signes d'un progrès constant vers de meilleures prévisions. Les causes de ce constant progrès sont essentiellement liées au raffinement continu des modèles théoriques, à la disponibilité d'une puissance de calcul toujours plus grande, nécessaire à l'exploitation des nouveaux développements et à la densité croissante des réseaux de mesures et d'information.

Ce genre de documentation ou d'information n'est pas facilement accessible dans la littérature économique. Dans la prochaine sous-section, nous tenterons une modeste reconstitution de ce qui précède, afin de laisser entrevoir à quoi pourraient ressembler de telles séries, si elles avaient été calculées à partir de prévisions économiques en temps réel.

2.4 Prévisions économiques

L'évaluation des prévisions en économie, de la manière dont nous l'avons caractérisée ci-haut, fait face à plusieurs obstacles, dont le plus important est probablement la fréquence relativement faible des observations à prévoir. Pour les données financières, cela ne pose pas de problème, mais le fait qu'une des variables quantitatives les plus importantes pour ce genre de prévisions, la volatilité quotidienne, ne soit pas directement observable *ex post* ajoute un autre élément de difficulté.

La première variable économique que nous considérerons dans ce contexte est la croissance de la production réelle. Ces observations sont publiées sur une base trimestrielle, de telle sorte que les économistes disposent de 4 prévisions à évaluer annuellement contrairement, par exemple, aux 365 prévisions 24 heures de la température ou des précipitations. Il est donc impossible d'espérer maintenir un niveau de précision adéquat en évaluant des séries annuelles de la qualité des prévisions et toute tentative d'estimer ce à quoi une telle série pourrait ressembler sera inévitablement hautement approximative.

Nous tenterons cependant d'y parvenir. Pour ce faire, nous commencerons par considérer l'aptitude, ou le contenu, de prévisions de la croissance du PIB un trimestre à l'avance obtenues à partir des techniques actuelles les plus performantes⁶. Un petit nombre d'études, incluant Galbraith (2003) et Öller (1985), ont produit des estimés de l'aptitude dans le cadre de leur examen du futur le plus éloigné

6. Ces résultats sont basés sur des données canadiennes et sont similaires à ceux obtenus avec des données américaines.

jusqu'à ce que de telles prévisions aient de la valeur. En actualisant l'ensemble des données à partir d'une valeur de départ et en simulant une séquence de prévisions en temps réel, nous pouvons estimer l'aptitude ou le contenu des techniques actuelles. Il s'agit de la dernière observation enregistrée dans le graphique 9. Dans le but d'estimer l'aptitude accessible aux prévisionnistes à des dates antérieures, nous nous limitons volontairement aux techniques accessibles lors des époques et périodes étudiées; par exemple, nous n'admettons pas l'indice de diffusion de Stock et Watson (2002a, b) avant 1998 et nous estimons la perte que cela implique⁷. Nous estimons aussi l'effet de l'accumulation de points d'échantillon (à travers la réduction de l'incertitude d'un paramètre) – un exercice hautement impressionniste étant donné que l'hypothèse de stabilité des dynamiques exploitées par le modèle de prévision est elle-même une approximation grossière et que l'accumulation de l'information par augmentation de la taille d'un échantillon a moins de valeur lorsque les structures sont changeantes.

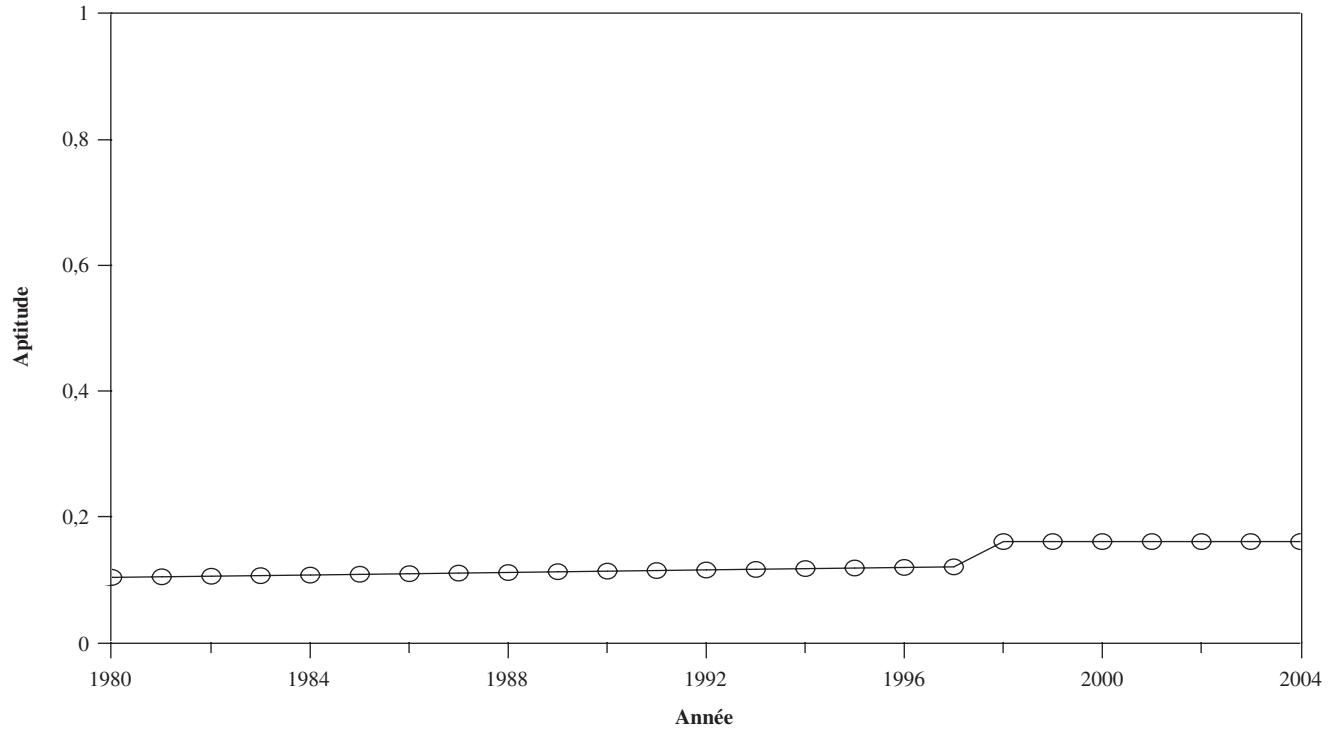
Le graphique qui en résulte, (graphique 9), suggère un progrès très lent dans la prévision de cette variable quantitative. L'aptitude de prévisions actuelles reste basse et ce faible taux d'accroissement reflète un phénomène largement observé : les techniques de pointe en prévision, quoique donnant des résultats supérieurs aux prévisions purement autorégressives disponibles depuis plusieurs années, ne sont que bien peu supérieures. Remarquez qu'en produisant le graphique 9, nous avons utilisé la version finale révisée des données; on peut s'attendre à ce que les résultats obtenus à partir de l'estimation initiale des données, accessibles à la date des prévisions, soient encore plus faibles.

Un portrait plus encourageant se dessine lorsque l'on considère les prévisions de volatilité, un des plus grands succès de l'économétrie financière. Le graphique 10 offre une reconstitution historique pour cette variable quantitative, obtenue à partir du même enchaînement d'étapes hautement approximatives décrites plus haut. Encore une fois, le dernier point de la séquence correspond à une estimation de l'aptitude de prévisions actuelles. Comme nous l'avons déjà fait remarquer, le tout est compliqué par le fait que la véritable volatilité quotidienne n'est jamais observée, mais peut être estimée par diverses méthodes utilisant des données intrajournalières à plus hautes fréquences; voir en particulier Zhang *et al.* (2005). Galbraith et Kisinbay (2005) ont estimé l'aptitude (contenu) de prévisions de la volatilité des taux de change *Yen/USD* et *DM/USD*; ces valeurs sont utilisées ici. Comme pour le graphique 9, nous effectuons une approximation de ce que le contenu correspondant aurait été si nous nous étions restreints aux méthodes de prévisions et aux données disponibles à une date historique donnée. Les sauts dans cette séquence de valeurs correspondent au développement des modèles ARCH (Engle, 1982), des modèles GARCH (Bollerslev, 1986) et de prévisions basées directement sur la variance réalisée (Anderson *et al.*, 2003, par exemple).

7. Nous utilisons la date d'apparition du document de travail NBER de Stock-Watson, plutôt que la date des articles publiés.

GRAPHIQUE 9

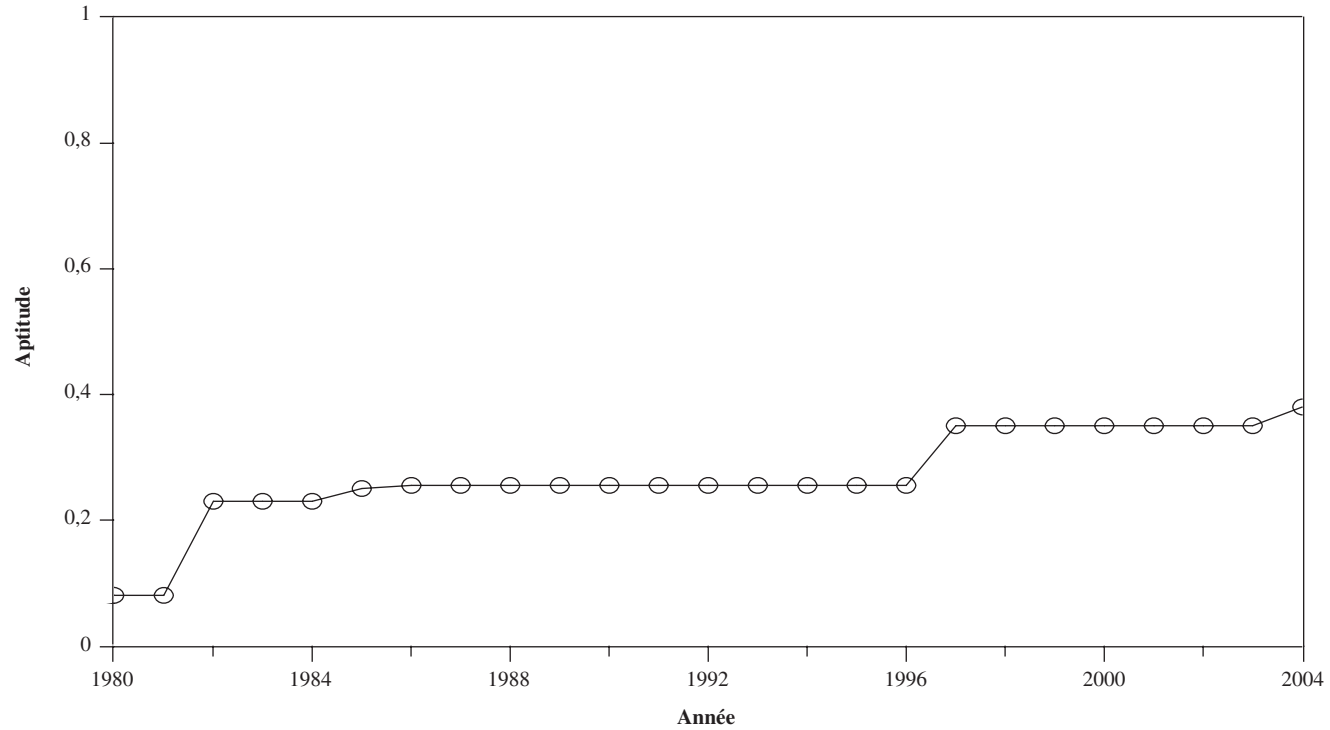
APTITUDE, PRÉVISIONS DE TAUX DE CROISSANCE, PNB, 3 MOIS – RECONSTRUCTION HISTORIQUE



SOURCE : Auteur

GRAPHIQUE 10

APTITUDE, PRÉVISIONS DE VARIANCE CONDITIONNELLE, 1 JOUR, YEN/\$US – RECONSTRUCTION HISTORIQUE



SOURCE : Auteur

Comme nous l'avions suggéré, sous cet angle la situation apparaît beaucoup plus encourageante. Nous sommes en présence d'un progrès appréciable et d'un niveau absolu substantiel d'aptitude de prévision accessible à partir des prévisions actuelles. La disponibilité de données à fréquence beaucoup plus haute que celles survenant dans des séries à applications macroéconomiques est un élément clé de ce progrès; pour la plupart des variables macroéconomiques, le taux de progression décrit dans le graphique 9 semble susceptible d'être une meilleure caractérisation.

Finalement, remarquez que pour les situations analogues à des prévisions de précipitations (prévisions de probabilités de récessions, par exemple) et des prévisions de tornades (prévision d'un krach boursier), le présent auteur n'a trouvé aucune évaluation économique régulière apte à une comparaison avec des résultats météorologiques.

3. CARACTÉRISATION DE L'INCERTITUDE

La météorologie et l'économique ont chacune effectué des avancées considérables dans la caractérisation et la communication de l'incertitude des prévisions. Cependant, ces sciences sont arrivées à ces nouveaux résultats en empruntant des chemins plutôt différents, ce qui reflète l'état de la science prévisionnelle dans celles-ci et, plus fondamentalement, ce qui reflète encore une fois les différences dans la disponibilité des données.

Une avancée essentielle à la compréhension de l'étendue des résultats possibles pour les prévisions météorologiques est venue de la prévision d'ensemble, qui s'apparente aux idées statistiques traditionnelles de combinaisons de prévisions. Un ensemble de prévisions consiste simplement en un groupe de prévisions de la même variable quantitative au même moment, effectué à l'aide de différentes procédures, modèles ou conditions initiales. Quoique ces ensembles soient utilisés à de multiples fins, la représentation de l'étendue des résultats possibles est un des principaux objectifs de la prévision d'ensemble.

En plus de l'incertitude sur le véritable processus régissant les conditions atmosphériques, l'incertitude entourant la mesure des conditions initiales pour une prévision est un élément crucial de l'incertitude de la prévision elle-même. Cette incertitude peut être représentée en sélectionnant un ensemble de conditions initiales possibles, et en produisant des prévisions basées sur chaque élément de cet ensemble. L'ensemble des prévisions qui en résulte illustre l'incertitude attribuable aux conditions initiales. Évidemment, d'autres sources d'incertitude, telle que l'incertitude sur le modèle, ne sont pas prises en compte en procédant ainsi. Néanmoins, pour ces prévisions basées sur des modèles où les conditions initiales sont un élément critique, l'ensemble défini de cette manière fournit une caractérisation utile d'une étendue de différents résultats.

Bien que pour des horizons courts l'erreur de prévision puisse être approximativement linéaire, pour des horizons moyens ou longs, l'erreur de prévision de même que la dépendance de ces prévisions aux conditions initiales est en général

non linéaire. En variant les conditions initiales, on génère un ensemble de prévisions qui divergeront l'une de l'autre à la longue, jusqu'à un degré qui ne sera pas d'ordinaire proportionnel à la différence dans les conditions initiales. Considérez par exemple le graphique 11 (Sivillo *et al.*, 1997, graphique 3, p. 814-815), représentant 17 ensembles de prévisions, donnant chacun une courbe de niveau pour le champ de cote à 500 hPa; les tableaux subséquents représentent des prévisions aux délais plus courts, de façon à ce que les parcours implicites des divers éléments de l'ensemble aient moins le temps de diverger. Remarquez que les valeurs des différents éléments de la prévision d'ensemble diffèrent l'une de l'autre par une quantité différente à des points distincts et que la largeur de la bande entre – par exemple – la bande la plus au nord et celle la plus au sud varie non seulement à travers le temps, mais aussi géographiquement. Un simple écart-type autour d'une bande centrale ne peut pas être une bonne approximation de cet ensemble de prévisions possibles à n'importe quel horizon temporel, bien que pour des horizons courts, la linéarité et donc la proportionnalité de l'erreur de prévision avec l'erreur dans les conditions initiales soit une approximation raisonnable (Sivillo *et al.*, 1997).

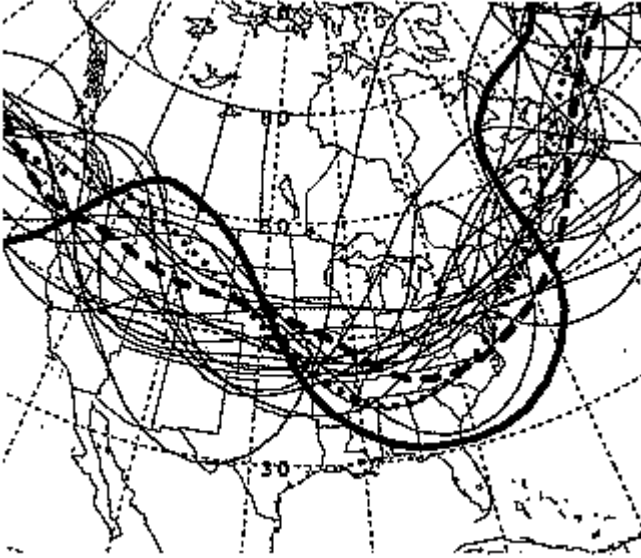
A contrario, comme première approximation il est probablement raisonnable de dire que la plupart des variables quantitatives *économiques* peuvent être prévues à l'aide de méthodes essentiellement statistiques au moins aussi bien que par les modèles économiques intégrés actuels, alors que les meilleures prévisions statistiques de la plupart des variables météorologiques sont clairement dominées par des prévisions basées sur des modèles structurels.

Considérez maintenant quelques représentations économiques de l'incertitude de prévision. Le graphique 12 (Britton *et al.*, 1998, graphique 9, p. 35) présente une surface tridimensionnelle représentant l'incertitude dans une prévision d'inflation à différents horizons dans le futur; il s'agit essentiellement d'un ensemble de fonctions de densité pour différents horizons de prévision. L'incertitude croissante à des horizons plus longs est transmise par « l'étalement » de la masse de probabilité lorsque l'horizon augmente. Une information similaire est communiquée de manière quelque peu différente par le diagramme « en éventail » du graphique 13. (Britton *et al.*, 1998, graphique 10, p. 36), décrivant l'incertitude d'une prévision de PIB. Dans ce cas, la gradation des couleurs plutôt que l'aire sous la courbe indique la masse de probabilité; l'ensemble d'intervalles délimités par chaque niveau de couleurs progressivement plus claires représentent des proportions progressivement plus larges de la masse de probabilité totale de la prévision (ex. 10 % de la probabilité repose dans l'intervalle le plus foncé, 20 % de la probabilité repose dans l'intervalle étendu prenant la prochaine gradation de couleur, *etc.*). Encore une fois, le graphique représente essentiellement un ensemble de fonctions de densité pour des prévisions à différents horizons dans le futur.

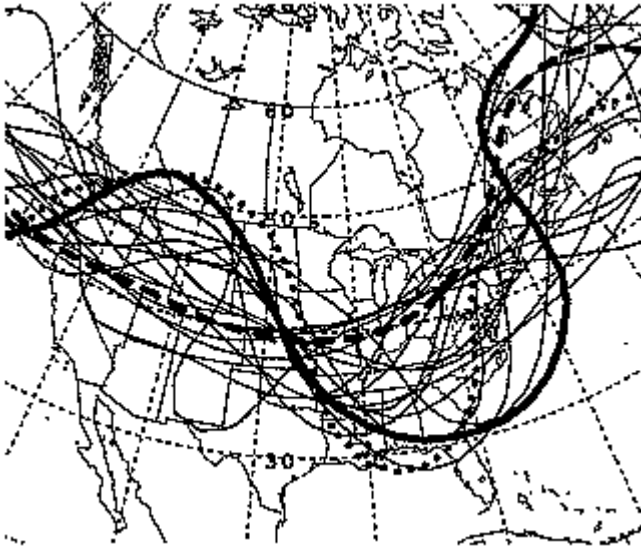
GRAPHIQUE 11

ENSEMBLE DES PRÉVISIONS

A – 9,5 jours à l'avance

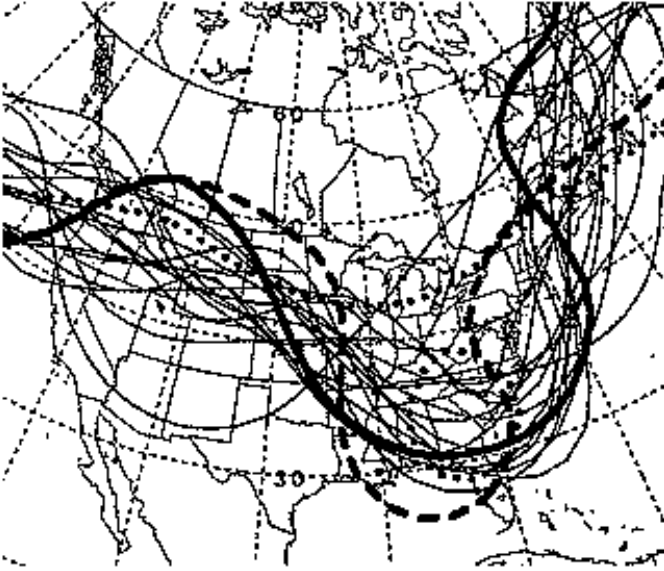


B – 8,5 jours à l'avance

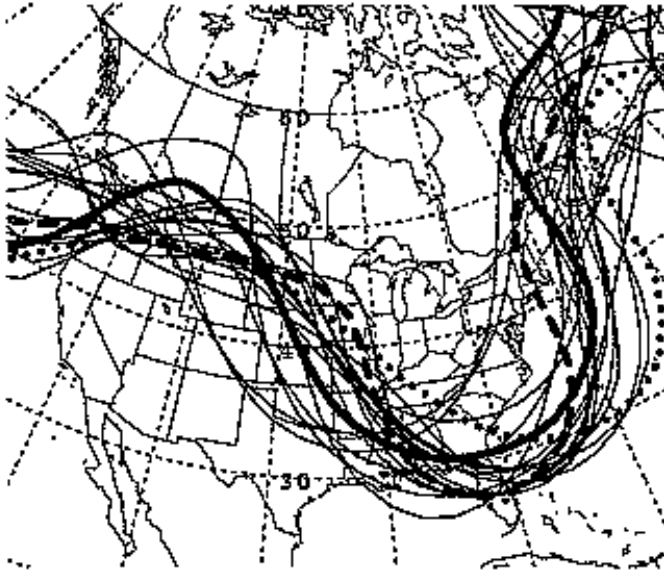


GRAPHIQUE 11 (suite)

C – 7,5 jours à l'avance

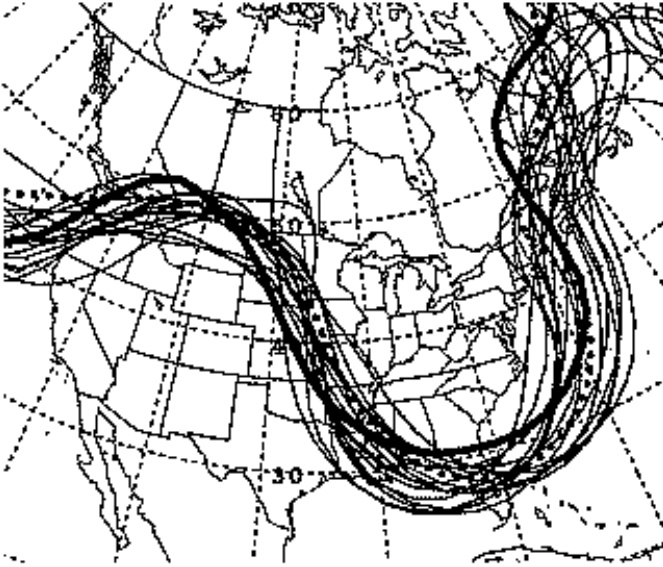


D – 6,5 jours à l'avance

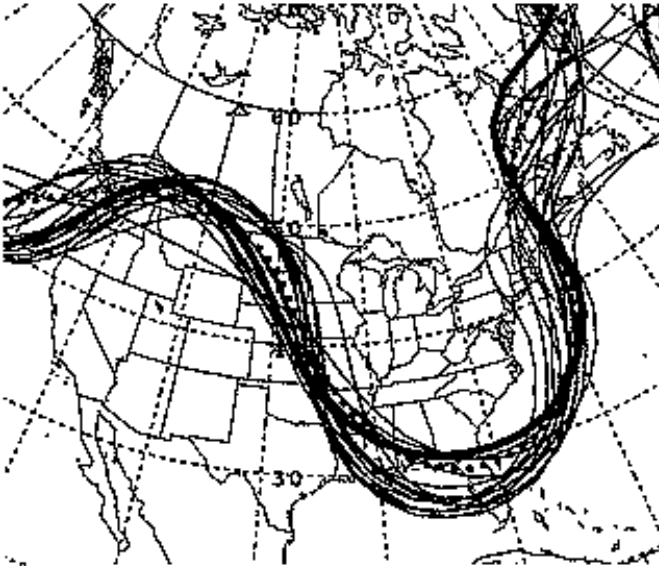


GRAPHIQUE 11 (suite)

E - 5,5 JOURS À L'AVANCE

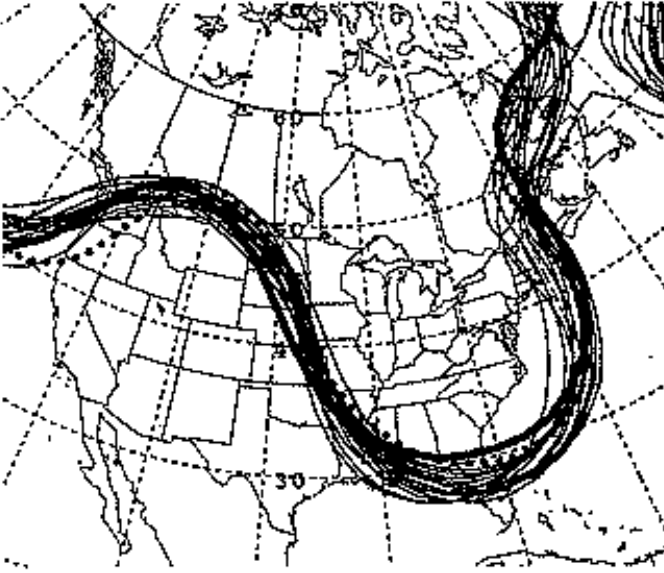


F - 4,5 jours à l'avance

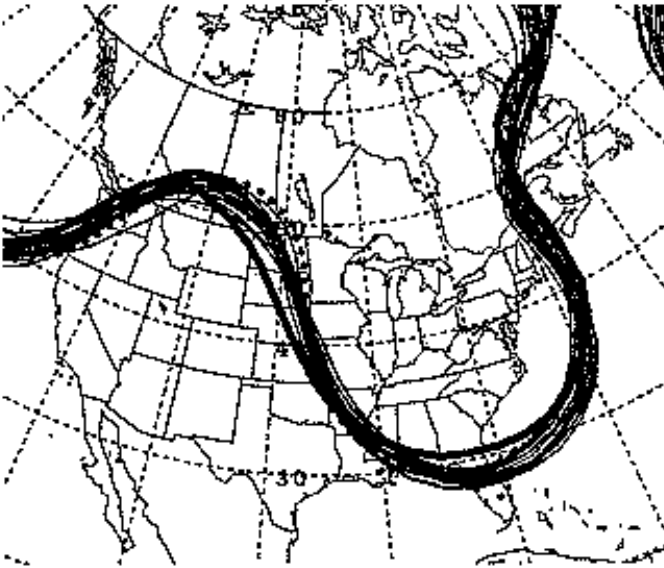


GRAPHIQUE 11B (suite)

G – 3,5 jours à l'avance

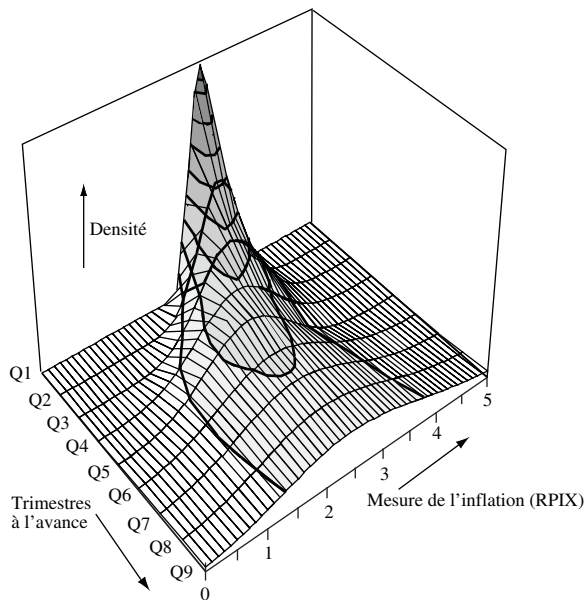


H – 2,5 jours à l'avance



GRAPHIQUE 12

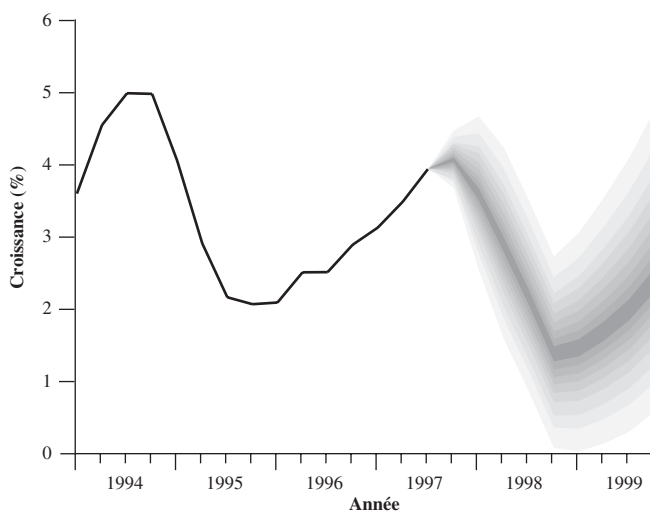
DENSITÉ DE PRÉVISIONS DE L'INFLATION



SOURCE : Britton *et al.* (1998), graphique 9

GRAPHIQUE 13

DISPERSION DE PRÉVISION DE CROISSANCE DU PIB



SOURCE : Britton *et al.* (1998), graphique 10

À première vue, les graphiques 12 et 13 peuvent sembler transmettre une quantité d'informations similaire à celle communiquée par le graphique 10. Mais il y a une importante différence. Dans le graphique 13, par exemple, la dispersion autour de la prévision centrale résulte d'un écart-type et d'un paramètre d'asymétrie, avec lesquels la densité à cet horizon est caractérisée. Dans la mesure où l'on connaît la forme fonctionnelle de la fonction, le graphique 13 peut être exprimé à n'importe quel horizon à l'aide de 3 nombres.

Dans l'un ou l'autre de ces cas, la représentation de l'incertitude des prévisions d'une manière qui puisse être comprise par un utilisateur non professionnel semble représenter un progrès considérable. Le degré d'incertitude des prévisions et le fait que cette incertitude puisse être décrite d'une manière qui a une signification précise ne sont pas bien compris par le public non professionnel, qui est l'audience ciblée par les prévisions économiques et météorologiques (c'est donc dire que l'un de nos buts est de fournir une information prédictive utile au grand public). L'utilisation de telles méthodes pour fournir plus d'informations qu'une simple valeur de prévision centrale est, à coup sûr, une initiative qui gagnerait à être plus répandue.

Néanmoins, nous devons mentionner que les caractérisations météorologiques de l'incertitude, quoique d'une portée limitée, apparaissent effectivement contenir davantage d'informations, un reflet de la quantité plus importante d'informations à haute fréquence à la disposition des météorologues pour l'incorporation dans des modèles de prévisions. Ce qui nous amène à l'un des thèmes de mes remarques finales.

CONCLUSION

L'un des motifs de notre examen des progrès accomplis par nos collègues chercheurs en météorologie est la prise en considération de ce qu'ils pourraient nous apprendre – si de telles leçons existent – dans le but d'obtenir plus de succès. Nous pourrions très bien trouver des idées intéressantes mais non applicables à l'économie en raison de contraintes au niveau des données, mais il serait étonnant qu'il n'y ait aucune idée en météorologie qui puisse être exploitée pour les fins de nos travaux.

Je ferai brièvement quelques suggestions, qui je pense pourraient être applicables, ou encore pourraient être appliquées de manière plus complète, aux données économiques.

Premièrement, il serait profitable de maintenir davantage de séries chronologiques de prévisions en sciences économiques. Bien que de nombreuses prévisions soient effectuées, ces prévisions ne sont souvent pas conservées en vue d'une évaluation future de la performance (il y a, bien sûr, plusieurs exceptions, tel que le « *Survey of Professional Forecasters* », maintenu par le « *Federal Reserve Bank of Philadelphia* »). Cela semble être une activité simple à laquelle les météorologistes s'adonnent ordinairement de manière plus consciencieuse (parmi les prévi-

sionnistes en économie, cependant, le « Federal Reserve Board » donne le bon exemple en conservant en archives des modèles de prévisions, des ensembles de données et des prévisions; voir Tetlow et Ironside, 2005 pour une analyse de cette information). L'évaluation de la performance présente et passée est, bien sûr, un élément clé d'un apprentissage visant à faire mieux.

Deuxièmement, il n'y a que peu de doutes sur la nécessité d'accroître la quantité d'informations que nous utilisons dans les prévisions économiques. Nos mesures du PIB paraissent une fois à chaque trimestre, mais assurément, d'autres informations (trafic routier?, trafic portuaire?) peuvent être exploitées lesquelles sont disponibles plus souvent que les mesures officielles du PIB⁸. Une alternative prometteuse réside dans les données enregistrées électroniquement à partir des terminaux de cartes de débits, alternative sur laquelle enquête actuellement le présent auteur, avec Greg Tkacz de la Banque du Canada. Une autre source d'information potentielle est l'utilisation de marchés de prévisions afin de résumer les vues et donc l'information, d'un grand nombre de participants au marché. Ces marchés peuvent, par exemple, être conçus pour donner un prix qui est une approximation de la probabilité d'un événement. (Voir Wolfers et Zitzewitz, 2005 au sujet de ces marchés et de l'interprétation des prix générés par ceux-ci.) Remarquez qu'il s'agit d'une source d'information qui n'est pas aussi valable pour les météorologistes; il est improbable que des participants au marché pariant sur le temps aient des sources d'informations supérieures à celles utilisées par les prévisionnistes professionnels, en partie à cause de l'intensité capitalistique de bonnes prévisions météorologiques. Lorsqu'il est question de phénomènes économiques cependant, il est entièrement concevable que des individus possèdent de l'information qui n'est pas reflété dans nos données officielles éparses.

La prévision est une activité importante, non seulement parce que les prévisions elles-mêmes peuvent constituer des guides utiles à la planification, mais aussi parce qu'il est naturel de tester nos connaissances par leur capacité prédictive. Nous sommes en état de mieux juger de nos progrès lorsque nous enregistrons soigneusement nos prévisions passées, et nous avons la possibilité de progresser si nous nous efforçons sans cesse d'exploiter plus d'informations pouvant avoir un pouvoir prédictif.

BIBLIOGRAPHIE

- ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD et P. LABYS (2003), « Modeling and Forecasting Realized Volatility », *Econometrica*, 71 : 579-625.
- ANDERSON, J.L. et W.G. STERN (1996), « Evaluating the Potential Predictive Utility of Ensemble Forecasts », *Journal of Climate*, 9 : 260-269.
- ATGER, F. (1999), « The Skill of Ensemble Prediction Systems », *Monthly Weather Review*, 127 : 1 941-1 953.

8. Je remercie Michel Truchon pour cette dernière suggestion.

- BOLLERSLEV, T. (1986), « Generalized Autoregressive Conditional Heteroskedasticity », *Journal of Econometrics*, 31 : 307-327.
- BOSART, L.F. (2003), « Whither the Weather Analysis and Forecasting Process? », *Weather and Forecasting*, 18 : 520-529.
- BRIER, G.W. (1950), « Verification of Forecasts Expressed in Terms of Probabilities », *Monthly Weather Review*, 78 : 1-3.
- BRITTON, E., P. FISHER et J. WHITELEY (1998), « Inflation Report Projections: Understanding the Fan Chart », Bank of England, *Quarterly Bulletin*, février : 30-37.
- BROOKS, H.E. (2004), « Tornado-warning Performance in the Past and Future: A Perspective from Signal Detection Theory », *Bulletin of the American Meteorological Society*, 85 : 837-843.
- BUIZZA, R. (1997), « Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction Systems », *Monthly Weather Review*, 125 : 99-119.
- CARTER, G.M., J.P. DALLAVALLE et H.R. GLAHN (1989), « Statistical Forecasts Based on the National Meteorological Center's Numerical Weather Prediction System », *Weather and Forecasting*, 3 : 401-412.
- DIEBOLD, F.X., T.A. GUNTHER et A.S. TAY (1998), « Evaluating Density Forecasts with Applications to Financial Risk Management », *International Economic Review*, 39 : 863-883.
- ENGLER, R.F. (1982), « Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation », *Econometrica*, 50 : 987-1008.
- FINLEY, J.P. (1884), « Tornado Predictions », *American Meteorological Journal*, 1 : 85-88.
- GALBRAITH, J.W. (2003), « Content Horizons for Univariate Time Series Forecasts », *International Journal of Forecasting*, 19 : 43-55.
- GALBRAITH, J.W. et T. KISINBAY (2005), « Content Horizons for Conditional Variance Forecasts », *International Journal of Forecasting*, 21 : 249-260.
- GHELLI, A. (2004), « Forecast Verification: Probabilistic Aspects. » in *Use and Interpretation of ECMWF Products*, European Centre for Medium-Range Weather Forecasting, Meteorological Training Course Lecture Programme.
- MARZBAN, C. (1998), « Scalar Measures of Performance in Rare-event Situations », *Weather and Forecasting*, 13 : 753-763.
- MURPHY, A.H. (1973), « A New Vector Partition of the Probability Score », *Journal of Applied Meteorology*, 12 : 595-600.
- MURPHY, A.H. (1996), « The Finley Affair: A Signal Event in the History of Forecast Verification », *Weather and Forecasting*, 11 : 3-20
- MURPHY, A.H. et R.L. WINKLER (1987), « A General Framework for Forecast Verification », *Monthly Weather Review*, 115 : 1330-1338.
- MURPHY, A.H. et D.S. WILKS (1998), « A Case Study of the Use of Statistical Models in Forecast Verification: Precipitation Probability Forecasts », *Weather and Forecasting*, 13 : 795-810.

- ÖLLER, L.-E. (1985), « How Far can Changes in Business Activity be Forecast? » *International Journal of Forecasting*, 1 : 135-141.
- ÖLLER, L.-E. et B. BAROT (2000), « The Accuracy of European Growth and Inflation Forecasts », *International Journal of Forecasting*, 16 : 293-315.
- PALTRIDGE, G.W. (1985), « The Value of Climate Forecasting », *Geophysical Surveys*, 7 : 273-290.
- ROEBBER, P.J. et L.F. BOSART (1996), « Complex Relationship between Forecast Skill and Forecast Value: A Real-world Analysis », *Weather and Forecasting*, 11 : 544-559.
- SIVILLO, J.K., J.E. AHLQUIST et Z. TOTH (1997), « An Ensemble Forecasting Primer », *Weather and Forecasting*, 12 : 809-818.
- STOCK, J.H. et M.W. WATSON (2002a), « Forecasting Using Principal Components from a Large Number of Predictors », *Journal of the American Statistical Association*, 97 : 1167-1179.
- STOCK, J.H. et M.W. WATSON (2002b), « Macroeconomic Forecasting using Diffusion Indexes », *Journal of Business and Economic Statistics*, 20 : 147-162.
- TETLOW, R.J. et B. IRONSIDE (2005), « Real-time Model Uncertainty in the United States: The Fed from 1996-2003 », Document de travail, Federal Reserve Board, Washington DC.
- TOTH, Z. et E. KALNAY (1997), « Ensemble Forecasting at NCEP and the Breeding Method », *Monthly Weather Review*, 125 : 3297-3319.
- VASIOFF, S.V. (2001), « Improving Tornado Warnings with the Federal Aviation Administration's Terminal Doppler Weather Radar », *Bulletin of the American Meteorological Society*, 82 : 861-874.
- WALDSTREICHER, J.S. (2005), « Assessing the Impact of Collaborative Research Projects on NWS Warning Performance », *Bulletin of the American Meteorological Society*, 86 : 193-203.
- WILKS, D.S. (1995), *Statistical Methods in the Atmospheric Sciences*, Academic Press.
- WILKS, D.S. (1996), « Statistical Significance of Long-range Optimal Climate Normal Temperature and Precipitation Forecasts », *Journal of Climate*, 9 : 827-839.
- WOLFERS, J. et E. ZITZEWITZ (2005), « Interpreting Prediction Market Prices as Probabilities. » Document de travail, University of Pennsylvania.
- ZHANG, L., P.A. MYKLAND et Y. AIT-SAHALIA (2005), « A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-frequency Data », *Journal of the American Statistical Association*, 100 : 1394-1411.