Rapport Technique

TR/CMGC/03/88

QUELQUES RESULTATS DES PREVISIONS COUPLEES DEMETER EFFECTUEES AU CERFACS

CERFACS Working Note

QUELQUES RESULTATS DES PREVISIONS COUPLEES DEMETER EFFECTUEES AU CERFACS

P ROGEL, E MAISONNAVE
Septembre 2003

1) Introduction

Cette note présente quelques résultats du système de prévision saisonnière du climat du CERFACS, et les compare à d'autres systèmes équivalents. Ces résultats ont été obtenus dans le cadre du projet européen DEMETER (Palmer et al., 2003, voir aussi http://www.ecmwf.int/research/demeter/).
Les résultats présentés ci-dessous ont été obtenus en effectuant des expériences systématiques de « prévisions » rétrospectives pour toutes les années entre 1987 et 2001. Le modèle utilisé couplé océan-atmosphère utilisé est composé d'ARPEGE-Climat pour l'atmosphère (ainsi que le sol), ORCA (version dite 2°) pour l'océan, et le coupleur OASIS. Les prévisions sont à six mois d'échéance, et démarrent au premier jour des mois de février, mai, août et novembre. Afin de prendre en compte l'incertitude liée à notre connaissance imparfaite de l'état réel du système climatique au moment de démarrer la prévision, ces prévisions sont répétées neuf fois. Chaque membre de ces ensembles démarre d'une condition initiale de l'océan légèrement différente des autres, cette différence reflétant en particulier notre méconnaissance de la quantité de chaleur stockée dans l'océan superficiel. La condition initiale pour l'atmosphère est l'état ré-analysé par le système du CEPMMT (ERA40, voir http://www.ecmwf.int/research/era/). La partie de ce système de prévision saisonnière qui touche à l'initialisation de l'océan est destinée à être adaptée pour tirer partie du système d'assimilation variationnelle « ORCAVAR » développé au CERFACS.
Dans la mesure où la prévision du climat aux échelles saisonnière doit non seulement prendre en compte l'incertitude des conditions initiales, mais aussi celle liée aux approximations faites dans les modèles, le but de DEMETER était de mettre au point un « multi-modèle », qui par combinaison des résultats de différents systèmes de prévision, permet une meilleure exploration de l'espace des possibles évolutions du système climatique à partir d'une date donnée. Cette approche, abondamment illustrée dans Palmer et al. (2003), permet également une intercomparaison assez rigoureuse de tous les modèles contribuant au projet. C'est dans cet esprit que cette note a été rédigée. Parmis les modèles participant à DEMETER, nous comparerons particulièrement nos résultats à ceux du modèle du CNRM, avec lequel nous partageons rigoureusement le même modèle d'atmosphère (ARPEGE), mais duquel diffèrent le modèle d'océan (deux versions d'OPA, avec des grilles et des physiques différentes, et un terme de rappel aux hautes latitudes) et le mode d'initialisation (en particulier, une introduction plus douce des perturbations en température dans le cas du CERFACS). Les résultats du CEPMMT, qui est le principal système opérationnel en Europe, constituent également une référence. Ce système est basé sur le modèle couplé HOPE-IFS, et l'océan est initialisé par interpolation optimale des observations in-situ de température.
Dans la suite, nous présenterons donc certains scores du CERFACS en les comparant à ceux de la plupart des modèles participant au multi-modèle DEMETER, d'abord dans les tropiques (et en particulier dans le Pacifique tropical) qui sont la principale source de prévisibilité à ces échelles de temps, et ensuite sur l'Hémisphère nord (avec un focus sur l'Europe) qui est une

zone d'intérêt pour le CERFACS. Dans une brève discussion, nous indiquerons quelques points permettant d'éclairer certaines différences de comportement entre modèles.

2) Performances tropicales

Un point crucial pour tout système de prévision saisonnière est sa bonne capacité à prédire les anomalies tropicales, et plus particulièrement les évènements « El Nino ». Nous montrons figure 1 les scores classiques de prévision d'anomalies de température dans la région du Pacifique tropical la plus représentative de cet évènement (appelée « Nino3 »), située dans la partie est de la bande équatoriale. Si au premier mois de la prévision, les anomalies restent naturellement très proches des observations, après 6 mois de prévision, les scores de corrélation d'anomalies restent très élevés (de l'ordre de 0.8), et l'écart quadratique moyen est

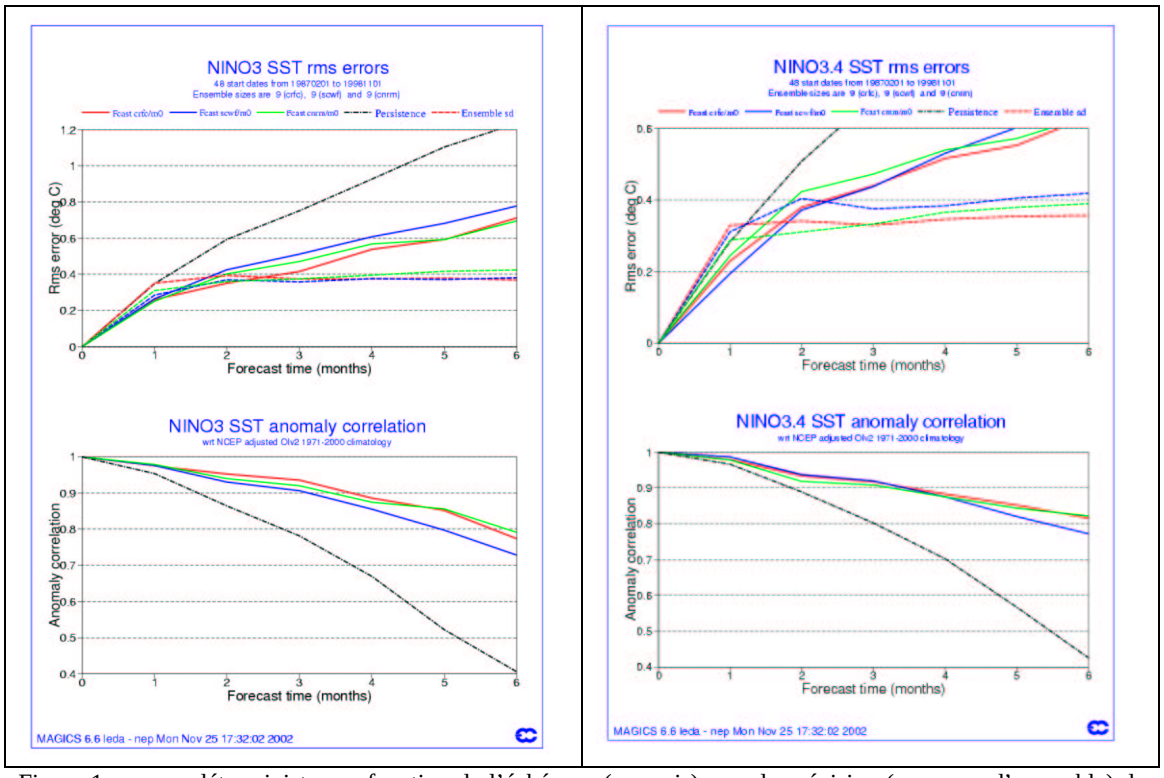

Figure 1 : scores déterministes en fonction de l'échéance (en mois) pour la prévision (moyenne d'ensemble) des anomalies de température dans la boite « Nino3 » (à gauche) et « Nino3.4 » (à droite). En haut, erreur quadratique moyenne, en bas, corrélation d'anomalies. Les modèles représentés sont le CERFACS (rouge), le CNRM (vert) et le CEPMMT (bleu), la persistance est en noir. Les pointillés (en haut) représentent la dispersion moyenne de chaque ensemble.

d'environ 0,7 °C. Ces scores montrent que le modèle du CERFACS est légèrement meilleur que tous les autres modèles en termes de corrélation, et significativement meilleur en termes d'écart quadratique (ceci est également vrai pour la comparaison avec les modèles qui ont des conditions initiales océaniques assimilées). Ceci signifie que de manière générale, les modèles savent prédire le signe des anomalies du Pacifique, mais que l'amplitude prédite est plus proche des observations dans le cas du CERFACS. Ceci se vérifie dans le cas de la prévision de l'évènement de 1997 (figure 2), où l'on voit que dans la phase de réchauffement

(de février à novembre), les anomalies prédites par le CERFACS sont systématiquement plus chaudes que les autres, et en particulier pendant le pic. Elles ont en revanche tendance à être un peu trop froide dans la phase de basculement vers « La Nina ». Cette bonne performance est également constatée sur les scores dans la partie est du bassin (boite « Nino4 », non montrée ici), où le modèle du CERFACS a des performances proches de celles du CNRM qui est le meilleur, et dans la zone centrale (boite « Nino3.4 », voir figure 1), dans laquelle le CERFACS a des performances très proches de celles du meilleur modèle (CEPMMT) pour les mois 1 à 3 de la prévision, et de celles du meilleur modèle (CNRM) pour les mois 4 à 6 de la prévision.
Associés à ces relativement bons scores, le modèle du CERFACS prédit également bien les anomalies de circulation atmosphérique (« Southern Oscillation Index ») ainsi que les précipitations dans la région du Pacifique tropical. En ce qui concerne le bassin Atlantique tropical, les performances du CERFACS sont plutôt moins bonnes que le CEPMMT, mais restent meilleures que le CNRM.

3) Performances sur l'Hémisphère nord et l'Europe

Plusieurs études (par exemple Cassou, 2001) ont montré que, si on prend l'océan comme source de prévisibilité, le potentiel de prévision aux échelles saisonnières est largement inférieur aux moyennes latitudes que dans les tropiques, du fait de la plus grande variabilité naturelle intrinsèque de l'atmosphère à ces latitudes. La prévision climatique saisonnière dans ces régions est par conséquent essentiellement un problème probabiliste, et consiste à étudier les fonctions de densité de probabilité (pdf) des variables du modèle, diagnostiquées à partir des ensembles de prévisions. Si celles-ci diffèrent notablement de la pdf des observations pour une prévision particulière, alors le système de prévision a détecté un signal. C'est en faisant des vérifications rétrospectives sur le plus grand nombre de cas possible que l'on peut établir les caractéristiques probabilistes du système de prévision. Nous le faisons dans la suite, pour les modèles du CERFACS et de DEMETER, en nous concentrant sur les variables de température et de précipitations, pour les prévisions des saisons d'hiver et d'été un mois en avance (moyenne des mois 2 à 4 de la prévision). Nous utilisons les résultats agrégés sur les régions de l'hémisphère nord (au nord de 20°N) et de l'Europe, et n'avons gardé pour illustration que ces dernières.

Afin de consolider les conclusions que nous pourrons tirer de ces résultats, plusieurs scores probabilistes sont utilisés dans la suite, en plus des corrélations d'anomalies (ACC, figure 3) :

- Le Ranked Probability Skill Score (RPSS, figure 4) :

Pour une variable donnée, ce score mesure l'écart entre les fonctions de densité de probabilité prévues et les vérifications. Pour la température, par exemple :

$$RPS = \sum \int_T (prob(T'<T) - v(T))^2 dp$$

prob(T'<T) est la probabilité prévue pour que l'anomalie de température soit inférieure à une valeur donnée T, diagnostiquée directement depuis le modèle ;
v(T) est 0 si la température effectivement observée est supérieure à T, 1 sinon.
Comme il s'agit d'un « skill score », le RPSS[1] mesure le gain par rapport au RPS de la prévision systématique climatologique : prob(T'<T) = pclim(T)[2] (c'est l'équivalent probabiliste de la persistence). Si les valeurs du RPSS sont positives, cela signifie que les

[1] $RPSS = 1 - RPS/RPS_{clim}$
[2] pclim est la fonction de distribution de probabilité climatologique ; par exemple : pclim(0) = 50% ; pour le tercile le plus chaud : pclim(0.46σ)= 33%.

prévisions apportent plus d’information que la climatologie, et une prévision parfaite aurait un score de 100.

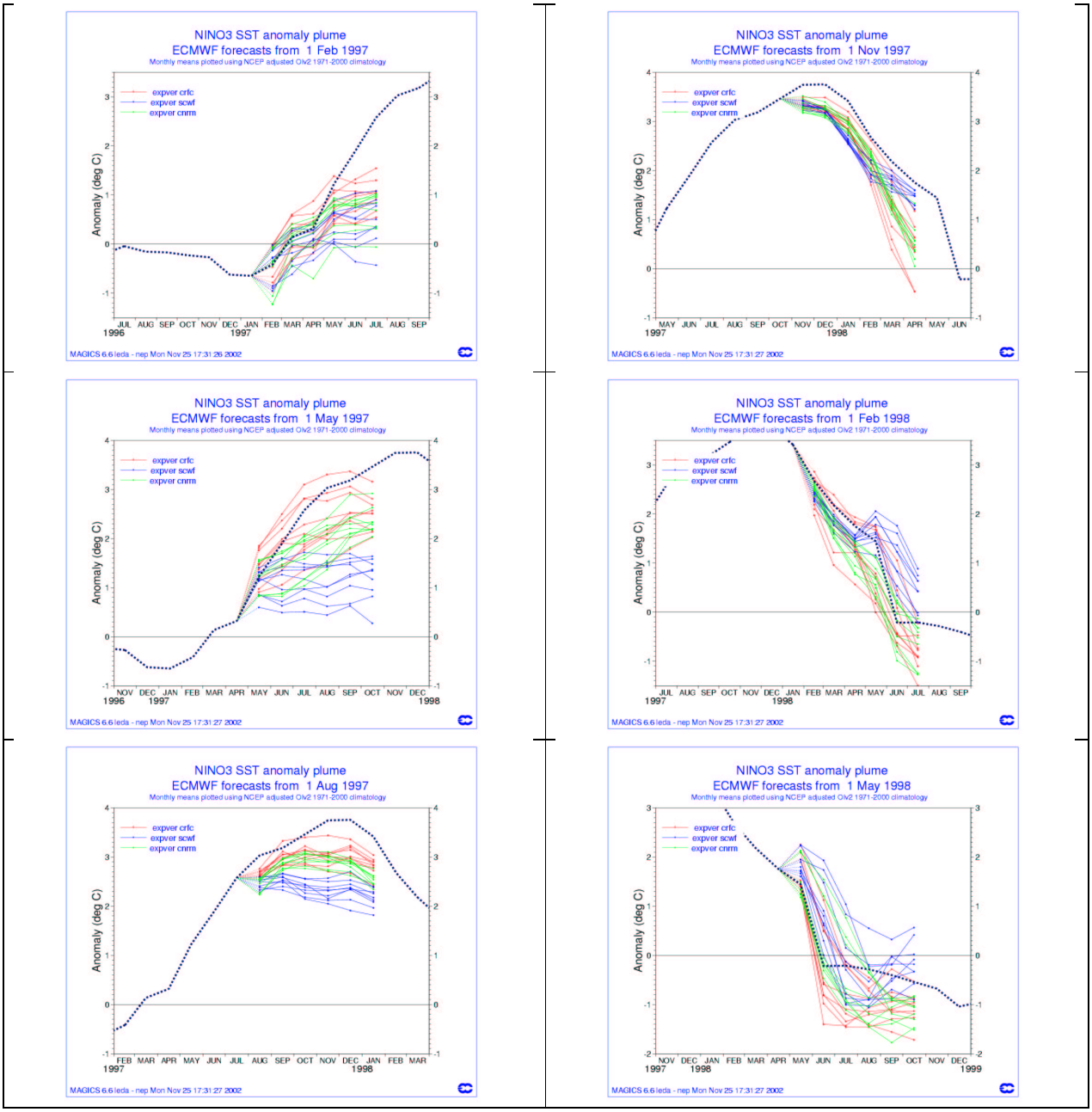

Figure 2: Séquence des prévisions d'anomalies de températures dans la boite « Nino3 »pour l'événement ENSO de 1997. Les modèles représentés sont le CERFACS (rouge), le CNRM (vert) et le CEPMMT (bleu). Le trait pointillé noir est l'anomalie observée.

- Le Relative Operating Characteristics (ROC, figure 5) :

Pour un événement donné, un utilisateur de prévisions probabilistes doit décider pour son propre usage à partir de quel seuil de probabilité prévue il considère que l'événement en cause va arriver. Une information utile vient de la variation relative de deux paramètres essentiels (le taux de fausses alarmes, et le taux de « coups au but », ou bonnes prévisions) que la prise en compte de ce seuil implique[3]. Le ROC mesure, pour un événement donné,

[3] Par exemple, si l'utilisateur ne veut pas prendre de risques et considère que ce seuil doit être le plus bas possible (par exemple prob(T'<0)=0), alors il est sûr que le taux de coups au but sera maximum, mais par contre, il aura

l'intégrale des variations relatives des taux de coups au but et de fausses alarmes, lorsque le seuil de décision varie de zéro à un. Ce score montre que le système de prévision surpasse la prévision climatologique s'il est supérieur à 0, et il serait de 0.5 pour une prévision parfaite.

- La valeur économique (figure 6) :

Nous nous situons toujours dans le cadre d'un utilisateur qui s'intéresse à la prévision d'un évènement qui peut lui infliger une perte économique (L) s'il se produit. L'utilisateur peut également se prémunir de cette perte, au coût C d'une action préventive. La valeur économique de la prévision (qui est en fait une évaluation objective de son utilité sur le long terme), tenant compte du coût moyen des fausses alarmes et de l'économie moyenne des coups au but, est calculée à l'aide des mêmes caractéristiques statistiques que le ROC, pour chaque valeur du seuil de décision et pour chaque valeur du rapport C/L. Cette courbe permet à l'utilisateur de choisir objectivement le seuil le plus approprié pour son cas. La courbe maximisant cette valeur est appelée valeur économique de la prévision, elle est exprimée en pourcentage du gain occasionné par une prévision parfaite.

Les résultats convergent pour indiquer que les prévisions du CERFACS pour l'hiver (aussi bien en température qu'en précipitations) sont parmi les meilleures du panel DEMETER. En effet, tant pour les scores déterministes (ACC) que pour les RPSS, qui sont exigeants, le CERFACS se place en tête. Pour ce qui est des RPSS, rappelons que la taille de l'ensemble a son importance, et que l'on ne peut ainsi directement comparer les scores issus d'un modèle individuel (9 membres) et issus du multi-modèle (63 membres). C'est pourquoi la majorité des modèles individuels a un RPSS négatif. Dans le cas du CERFACS, celui-ci est (légèrement) positif pour les prévisions de température d'hiver, et le moins négatif pour les précipitations (figure 4). Pour ces deux variables, ces relatifs bonnes performances ont pour conséquence que les scores liés à l'« utilité » des prévisions (ROC et valeur, figure 5 et 6) sont très hauts. Par exemple, les ROC sont les meilleurs, y compris par rapport au multi-modèle, pour quasiment tous les évènements, et la valeur est supérieure pour tous les utilisateurs potentiels pour les deux évènements présentés. Ces bons scores en hiver, tant d'un point de vue déterministe que probabiliste, sont également vérifiés lorsque l'on prend non plus l'Europe pour agréger les résultats, mais l'ensemble des latitudes extratropicales de l'hémisphère nord (non montré ici).

A l'inverse, les résultats convergent également pour indiquer que les prévisions saisonnières d'été du CERFACS sont plutôt dans la moyenne, voire inférieures aux autre. Par exemple, le RPSS en température est parmi les plus faibles (figure 4). Notons quand même que les prévisions de précipitation restent du même ordre de qualité que celles du CNRM (qui a le même modèle d'atmosphère).

4) Brève discussion

Une analyse exhaustive des différences entre modèles ferait l'objet d'une étude à part entière. Contentons nous ici d'avancer quelques hypothèses pour expliquer les différences entre les modèles du CERFACS et du CNRM, qui partagent rigoureusement le même modèle

---

comme contrepartie que beaucoup de fausses alarmes seront déclenchées. A l'inverse, s'il veut que la probabilité de prévision soit maximale (prob(T'<0)=1), alors il n'aura aucune fausse alarme, mais par contre, le nombre de coups au buts sera considérablement réduit, car un modèle raisonnablement calibré laissera presque toujours une possibilité que l'événement n'ait pas lieu. Entre ces deux extrêmes, un modèle qui ferait varier le taux de fausses alarmes et coups au but de façon proportionnelle n'aurait pas grande utilité. Il devient utile s'il existe des valeurs du seuil qui produisent plus de coups au but que de fausses alarmes.

d'atmosphère, mais diffèrent par le modèle d'océan et par le mode d'initialisation. Il est possible que les bons scores du CERFACS dans le Pacifique tropical sont liés à un meilleur traitement des couche de surface dans la génération des ensembles de conditions initiales. Cette procédure, qui consiste dans le cas du CNRM à extrapoler brutalement les perturbations de SST dans les premiers niveaux du modèle, affecte 8 des 9 membres dans le cas du CNRM, et est susceptible d'introduire des déséquilibres entre le champ de masse et la dynamique, y compris le schéma de mélange vertical. Elle peut donc avoir un effet important sur la moyenne d'ensemble.

Les différences de scores dans les moyennes latitudes, elles, peuvent provenir des erreurs systématiques très différentes (voir figure 7), elles-mêmes provoquées par des physiques océaniques différentes, mais aussi et surtout par un état océanique plus réaliste aux hautes latitudes dans le cas du CERFACS, suite au terme de rappel en température en profondeur.

Ces différences justifient en tout cas, de manière assez convaincante, le soin qui doit être apporté à la fabrication de conditions initiales océaniques.

Référence :

Cassou, C., Rôle de l'océan dans la variabilité basse fréquence de l'atmosphère sur la région Nord Atlantique-Europe, thèse de l'Université Paul Sabatier, 280 pp, 2001.

Palmer, T.N. , A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson, Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER), submitted to BAMS, 2003.

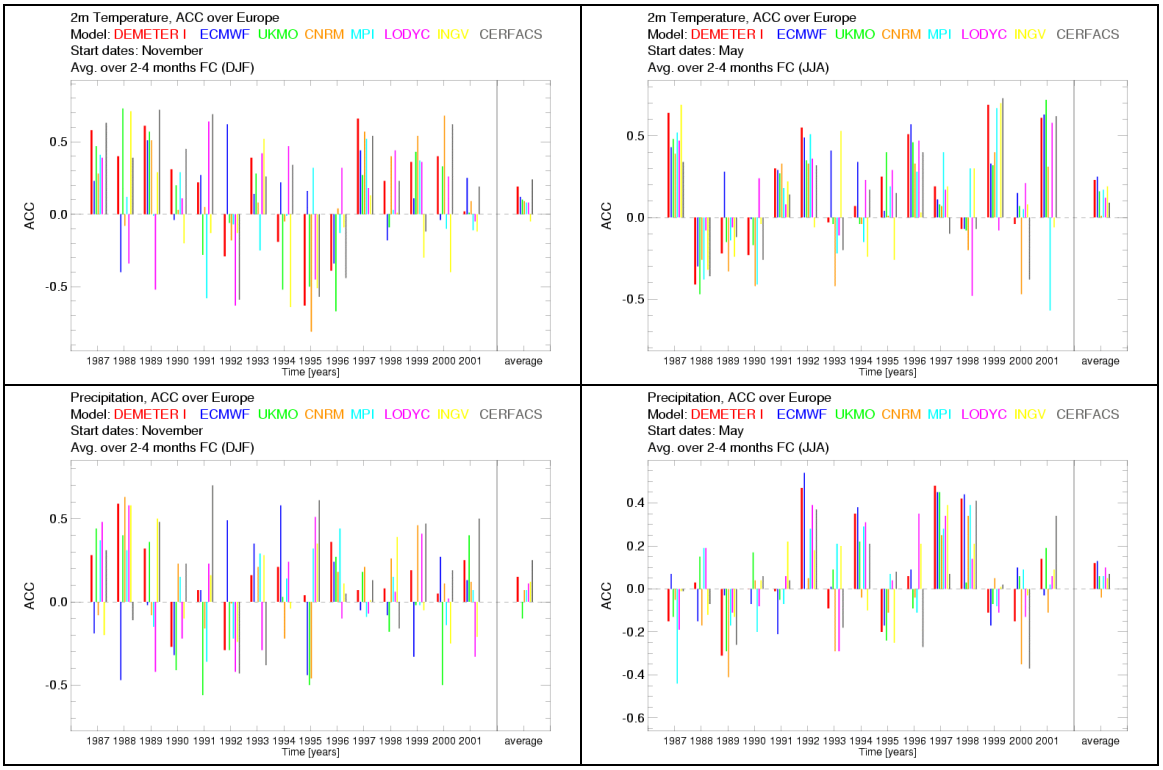

Figure 3: Corrélation d'anomalies pour les prévisions de température (en haut) et de précipitation (en bas), saisons d'hiver (DJF, à gauche) et d'été (JJA, à droite) sur la région Europe. L'échéance est de 1 mois (moyenne des mois 2-4 de la prévision). Pour chaque année, un score a été calculé par agrégation spatiale, et un score global par agrégation spatiale et temporelle. Les 7 modèles DEMETER sont représentés, ainsi que le « multi-modèle », le CERFACS est en noir.

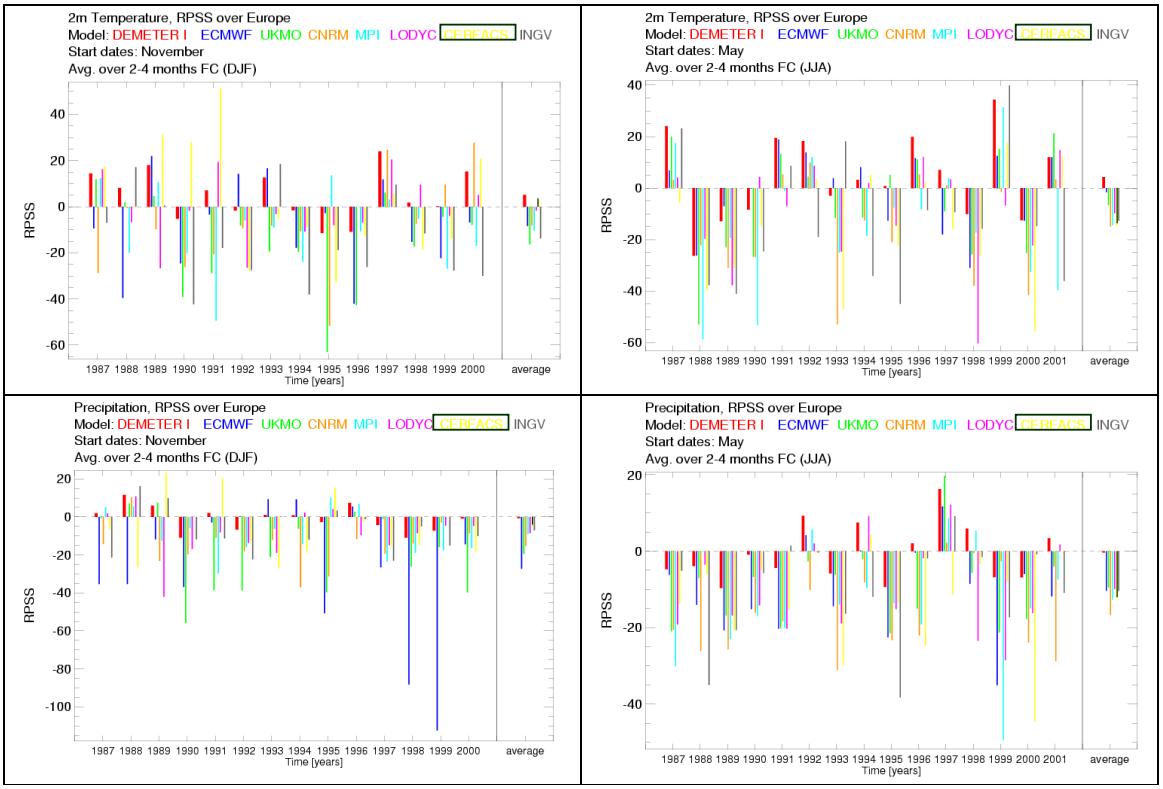

Figure 4: Même chose que la figure 3, pour les RPSS. Le CERFACS est en vert foncé.

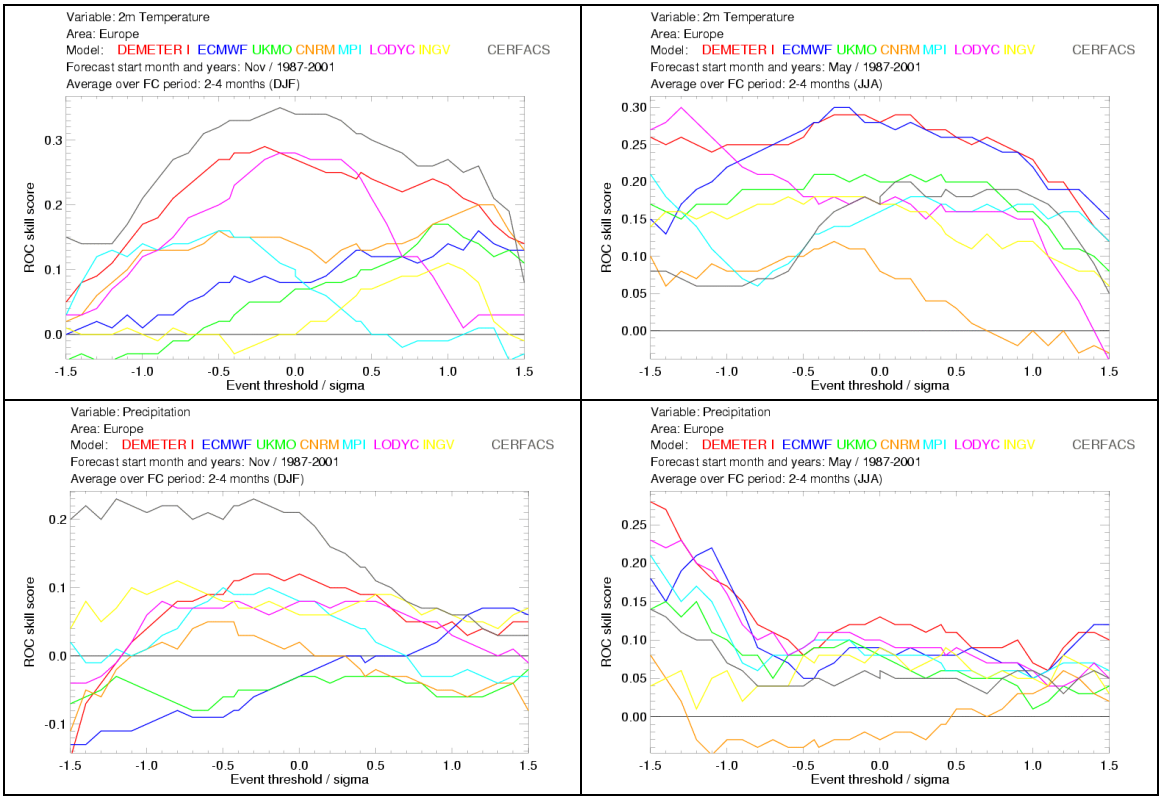

Figure 5: Même chose que pour la figure 3, pour les ROC. Les scores sont présentés pour tous les événements « anomalie de température (resp précipitation) supérieure à **x** écart-types », **x** est en abscisse. Le CERFACS est en noir.

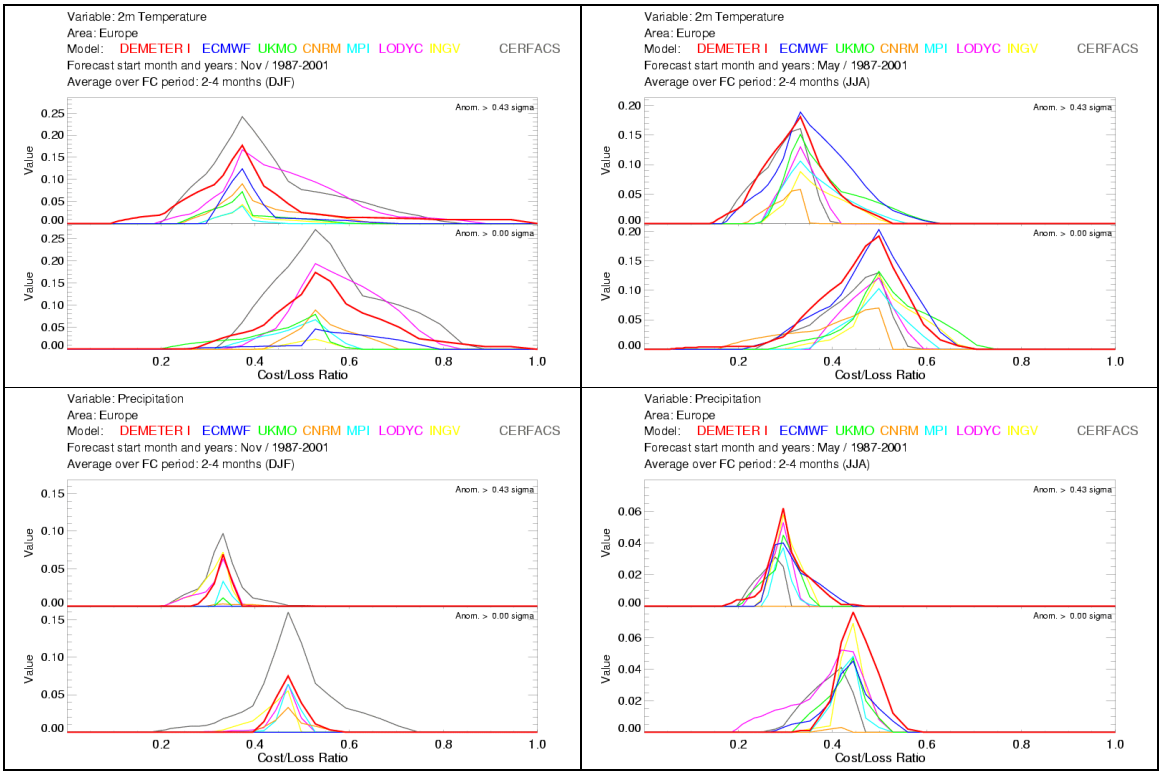

Figure 6: Même chose que pour la figure 3, pour la valeur économique. L'abscisse représente les utilisateurs potentiels (rapport C/L) et les scores ont été calculés pour 2 événements, « supérieur à la climatologie », et « dans le plus haut tercile ». Le CERFACS est en noir.

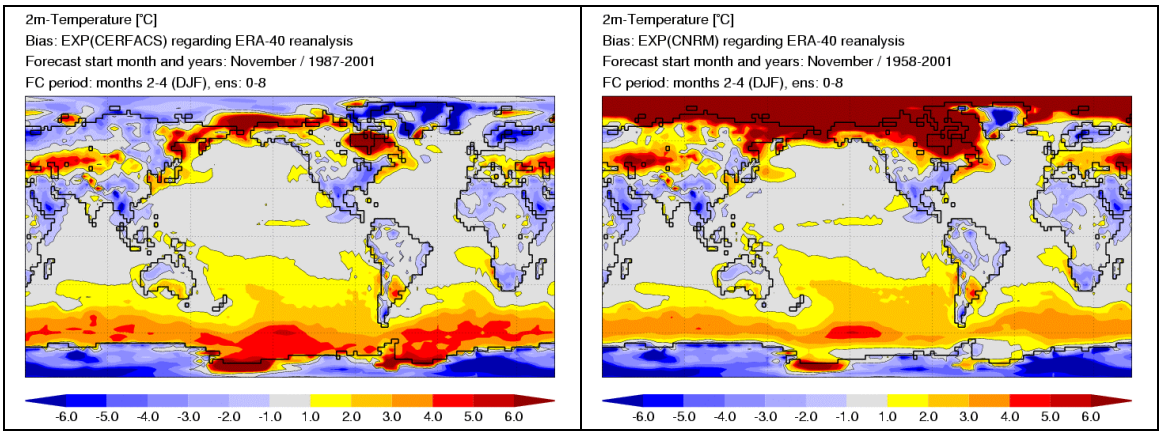

Figure 7: Erreur systématique en température pour la prévision d'hiver à 1 mois d'échéance, pour le CERFACS (à gauche) et le CNRM (à droite).