

PRÉVISIBILITÉ DU CYCLE DU BLÉ : MODÉLISATION PAR APPROCHE GLOBALE ET ASSIMILATION DE DONNÉES

Sylvain Mangiarotti^{1,2}, Malika Chassan², Laurent Drapeau¹

1 : Centre d'Etudes Spatiales de la Biosphère – Toulouse, France
sylvain.mangiarotti@ird.fr

2 : Joint International Laboratory Télédétection et Ressources en Eau en Méditerranée Semi-Aride (TREMA), Maroc

Résumé

Il a été récemment montré que des modèles dynamiques pouvaient être directement obtenus du signal d'agriculture pluvial observé depuis l'espace. L'analyse de ce modèle a permis de révéler un régime chaotique, c'est-à-dire à la fois déterministe et très sensible aux conditions initiales. De tels modèles ne nécessitent pas de forçage et peuvent donc être envisagés en mode prévisionnel. En contexte chaotique, une telle application nécessite le développement de méthodes adaptées afin d'ajuster au cours de l'intégration les états visités par le modèle, en cohérence avec les observations. Dans ce but, plusieurs schémas d'assimilation ont donc été développés : (a) une simple réinitialisation, (b) un filtre de Kalman étendu, (c) un filtre de Kalman d'ensemble et (d) un nudging direct et rétrograde. La prévisibilité effective a été estimée pour chacune de ces approches en se basant sur la croissance d'erreur en fonction de l'horizon de prévision. Les résultats mettent en évidence une bonne capacité prédictive du modèle lorsqu'utilisé avec les schémas d'assimilation de données les plus sophistiqués et contribue à valider le modèle.

Mots clés : Agriculture pluviale, assimilation de données, modélisation globale, chaos, prévisibilité.

Abstract

It was recently shown that dynamical models could be directly captured from the signal of rainfed crops observed from space. The regime of this dynamics was found to be chaotic that is both deterministic and highly sensitive to the initial conditions. Since requiring no external forcing, such model may be useful in a forecasting mode. However, in a chaotic context, such approach requires using adapted tools to update the state of the model in coherency with the observations. Several data assimilation schemes are developed and tested for the present context : (a) a direct re-initialisation scheme, (b) an Extended Kalman Filter, (c) an Ensemble Kalman Filter and (d) a Back and Forth Nudging. The effective predictability is estimated for each of these four approaches by considering the error increase as a function of the horizon of prediction. The results show the good predictive ability of the model when used with sophisticated data assimilation approaches, this results also contribute to validate the model.

Keywords : Rainfed agriculture, data assimilation, global modelling, chaos, predictability.

1. Introduction : prévisibilité du couvert végétal en zone semi-aride

Les limites de prévisibilité de l'agriculture pluviale ne résultent pas seulement de la complexité des processus en jeu à l'échelle locale, mais également de la dynamique à l'échelle régionale, impliquant la circulation atmosphérique et océanique. L'influence de l'Oscillation Nord Atlantique (NAO) a été suggérée dans plusieurs études (e.g., Knippertz et al. 2003). L'influence de la mer Méditerranée a également été montrée comme une influence forte sur le Nord du Maroc et comme faible en son centre et dans le sud. L'analyse des températures de surface de l'océan a également suggéré un effet indirect de l'océan Atlantique au niveau de l'équateur sur les rendements céréaliers du Maroc (Jarlan et al., 2012).

La plupart des modèles dédiés au couvert végétal sont

des modèles biophysiques semi-empiriques, nécessitant de disposer d'un forçage atmosphérique, et donc mal adapté à la prévision saisonnière. Les modèles autonomes basés sur des approches statistiques ont également été utilisés (e.g., Balaghi et al., 2007 ; Jarlan et al., 2012). En zone semi-aride, ces modèles statistiques simples, construits à partir d'un petit nombre de variables, ont montré de meilleures performances que les approches semi-empiriques, illustrant la nécessité de simplifier les modèles - quitte à devenir schématique - plutôt que de les complexifier trop, les modèles trop complexes conduisant souvent à des problèmes de calibration et d'optimisation de paramètre difficiles à résoudre correctement.

La théorie des systèmes dynamiques permet un autre point de vue, complémentaire. Appliqués à l'Afrique de l'Ouest, des travaux récents ont suggéré que 3 ou 4 variables pourraient suffire pour la modélisation

du couvert végétal de ces régions (Mangiarotti et al., 2010). Les comportements peuvent dépendre de l'échelle considérée et, en contexte prévisionnel, le nombre de variable peut aussi dépendre de l'horizon de prévision visé et du niveau de précision escompté (Mangiarotti et al., 2012a). Issue de la même théorie, la modélisation par approche globale consiste à modéliser une dynamique sous la forme d'un système canonique, et à partir d'une unique variable observée (Gouesbet et Letellier, 1994). Cette approche peut permettre d'obtenir des modèles non linéaires très parcimonieux (peu de paramètres) permettant une description à la fois concise et complexe de la dynamique globale du système. Si l'applicabilité de l'approche a pu être illustrée par de nombreux cas théoriques et expérimentaux (Letellier et al., 2009), son efficacité n'a pu être montrée qu'à un nombre très réduit de cas environnementaux (Maquet et al., 2007). Appliquée au cycle du blé dans le nord du Maroc, l'approche globale s'est présentée comme une approche puissante, montrant à la fois qu'une dynamique déterministe était sous-jacente au système et que la faible prévisibilité résultait de son régime chaotique (Mangiarotti et al., 2012b).

Un tel modèle pourrait permettre de réunir plusieurs avantages. En effet, de par sa formulation autonome il ne nécessite pas de forçage externe ; d'autre part il peut permettre un certain degré de non linéarité (et donc une certaine complexité) sans requérir d'hypothèse forte quant à sa formulation mathématique. Le principal objet de la présente étude est de tester les capacités d'un tel modèle en mode prévisionnel. En contexte chaotique, des outils spécifiques sont toutefois requis afin de compenser les effets de perturbations et les instabilités. En effet, quoique déterministes, les dynamiques chaotiques sont caractérisées par une forte sensibilité aux conditions initiales. En d'autres termes, de petites perturbations sur l'état courant du système vont être amplifiées exponentiellement, nécessitant donc un réajustement régulier des états visités par le modèle. L'assimilation de données a fait l'objet d'importants développements au cours des dernières décennies, plusieurs de ces méthodes seront ici testées.

Après avoir présenté les données satellite utilisées pour ce travail et le prétraitement qui leur a été appliqué, une présentation de l'approche de modélisation globale sera présentée ainsi que son utilisation en mode prévisionnel par l'adjonction de schémas d'assimilation de données. Quatre approches seront ici présentées et utilisées. Les résultats obtenus avec ces approches seront ensuite présentés et discutés pour en tirer conclusions et perspectives.

2. Données satellitaires

Les indices de végétation NDVI (Normalized Difference Vegetation Index) sont utilisés ici sur la période 1982-2008. Ils sont produits par le Global Inventory

Modelling and Mapping Study (GIMMS) du Global Land Cover Facility (GLCF) (Tucker, 2005). Ces produits sont issus de capteurs AVHRR (Advanced Very High Resolution Radiometer) situés à bord des satellites NOAA (National Oceanic Atmospheric Administration). Les indices NDVI permettent un suivi efficace des couverts herbacés et des cultures céréalières en zones semi-arides (Tucker, 1979). Le produit utilisé ici présente un échantillonnage à 10 jours et une résolution de $8 \times 8 \text{ km}^2$; il est corrigé des effets de dérive des capteurs et tient compte également des aérosols atmosphériques résultant des éruptions d'El Chichon et du Mont Pinatubo survenues respectivement en avril 1982 et juin 1991.

La zone étudiée dans ce travail se situe dans le nord du Maroc dans une fenêtre allant de $6,2^\circ\text{W}$ à $5,4^\circ\text{W}$ en longitude et de $34,6^\circ\text{N}$ à $35,4^\circ\text{N}$ en latitude. Bien que cette zone contienne une part de cultures irriguées, la production pluviale reste néanmoins dominante. Une sélection basée sur des cartes de couverture des sols (Mayaux et al., 2004) a été appliquée afin d'obtenir un signal plus caractéristique d'un type de culture particulier. Une fois la sélection appliquée, le signal a été agrégé spatialement à l'échelle de la zone étudiée en moyennant les mesures, permettant de réduire le niveau de bruit du signal. Un filtrage basé sur la méthode de Savitzky-Golay (1964) permettant aussi de calculer les dérivées successives du signal a été appliqué en prenant une fenêtre de filtrage d'environ 5 mois permettant de ré-échantillonner le signal et ses dérivées à 3 jours. Le signal avant et après filtrage est présenté en Figure 1.

3. Modélisation globale - assimilation de données

3.1. Modélisation par approche globale

L'approche de modélisation globale (Gouesbet & Letellier 1994) prend appui sur la théorie des systèmes dynamiques. Disposant d'un ensemble de variables permettant de décrire l'état d'un système à des instants successifs, cette théorie nous permet d'associer à cette succession d'états, une trajectoire caractéristique de la dynamique du système. Cette trajectoire peut être reconstruite dans un espace appelé l'espace des phases. En pratique, on ne dispose souvent que d'une seule variable. En s'appuyant sur le théorème de Takens (1981), on peut, partant d'une variable unique, remonter à une trajectoire difféomorphe à celle du système original (*i.e.*, équivalente à la trajectoire qui serait issue de l'ensemble des variables). Une trajectoire ainsi reconstruite à partir d'une seule série temporelle va nous servir de substitut à la trajectoire originale et pouvoir être utilisée comme source d'information pour construire notre modèle prédictif. L'approche revient à considérer la dynamique du milieu dans sa globalité, sans chercher à détailler les interactions entre variables. Selon un point de vue purement théorique, le modèle ainsi obtenu inclurait donc

non seulement la dynamique de la végétation, mais également la dynamique atmosphérique elle-même. En pratique, l'obtention d'un modèle aussi global va dépendre de l'observabilité du système étudié vu par la variable considérée (Letellier et al., 2005). Si l'approche ne peut nous permettre de distinguer les processus en jeu dans le comportement étudié, elle peut en revanche nous permettre d'en caractériser sa dynamique globale, aux échelles d'espace et de temps qui la caractérise. L'approche est très empirique car directement basée sur les observations, sans nécessiter d'hypothèse forte quant à la formulation de sa dynamique. Autrement dit, l'approche va nous permettre une modélisation de la dynamique sous-jacente au système, non contrainte par une formulation a priori.

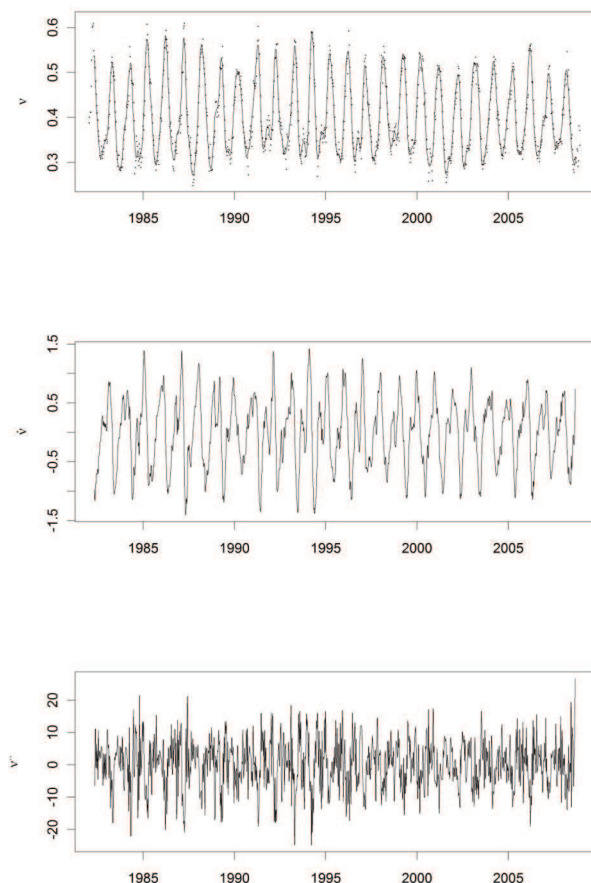


FIGURE 1 : Signal d'indice de végétation agrégé $v(t)$ avant filtrage et après filtrage.

La modélisation globale nécessite la mise en oeuvre d'une série d'outils d'analyse et de modélisation complexes. L'ensemble de ces outils a été développé au Centre d'Etudes Spatiales de la Biosphère et assemblé en deux packages PoMoS & GloMo codés sous R (Man-

giarotti et al., 2012b). Le premier de ces deux packages PoMoS (Polynomial Model Search) vise à rechercher la formulation algébrique optimale d'un problème à N variables d'état sous la forme d'un système d'équations aux dérivées ordinaires. Sa formulation générique lui permet également de rechercher des modèles globaux dans leur formulation polynomiale. Dans un cas tridimensionnel, partant de la série temporelle $x_1(t) = v(t)$ d'indice de la végétation, le modèle peut être formulé comme suit :

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dots \\ \dot{x}_3 = P(x_1, x_2, \dots, x_N) \end{cases} \quad (1)$$

où $P(x_1, x_2, \dots, x_N)$ est un polynôme dont les termes et le paramétrage sont à identifier.

Le second package GloMo (Global Modelling) permet plus spécifiquement d'identifier le paramétrage d'un modèle sélectionné, et d'en tester l'intégrabilité. Pour des raisons théoriques et pratiques, on privilégie généralement dans l'approche globale les modèles de petites dimensions qui garantissent une meilleure proximité à la dynamique originale et offrent une meilleure robustesse (Letellier et al., 2009). De façon pragmatique, ce choix pourrait être revu en contexte prévisionnel.

Appliquée à la dynamique au nord du Maroc, l'approche a permis d'obtenir un modèle tri-dimensionnel de 14 paramètres présentant une très bonne robustesse (Figure 2). L'obtention d'un modèle de type global peut être considérée comme un argument fort de présence d'un comportement déterministe (Letellier et al., 2009), sans exclure la présence de perturbations 'stochastiques' combinées (pouvant prendre la forme d'un bruit multiplicatif notamment). L'obtention de ce modèle nous fournit ici un argument puissant à l'hypothèse déterministe et nous montre par la même occasion la parfaite viabilité d'une approche autonome.

3.2. Différents schémas d'assimilation de données

Effectuer des prévisions en contexte chaotique nécessite assurément de disposer d'approches adaptées visant à compenser les effets de perturbation pouvant résulter de processus secondaires lesquels vont se propager au cours du temps et s'amplifier. Cette divergence de trajectoire - caractéristique des dynamiques chaotiques - fait des modèles chaotiques des bancs d'essais privilégiés pour tester les méthodes d'assimilation de données (e.g., (Miller et al., 1999)). Très peu de modèles ayant pu être directement obtenus d'observations réelles, l'application pratique d'un modèle chaotique en mode prédictif apparaît donc comme une première, de même que de l'utiliser pour tester et comparer différents schémas d'assimilation de données. Quatre schémas sont ici comparés présentant une complexité croissante. Chacun des schémas présente en alternance une phase d'intégration du modèle et une étape dite d'analyse. La phase

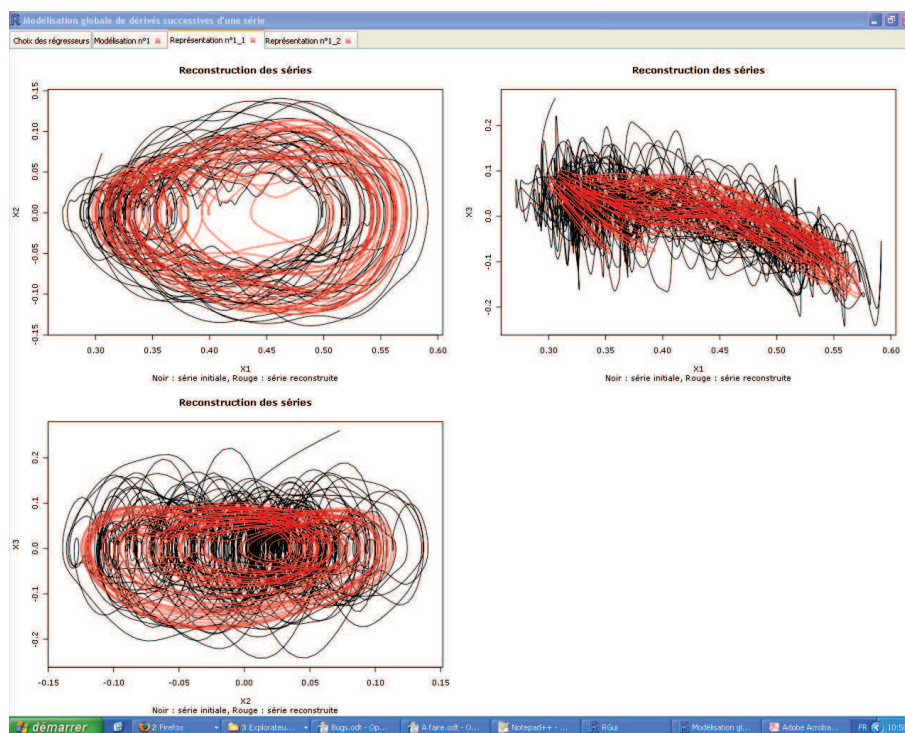


FIGURE 2 : Portrait de phase issu des données (en noir) et issu du modèle (en rouge).

d'intégration du modèle (Equation 1) est basée sur un algorithme d'intégration de type Runge-Kutta à pas variable. A pas de temps régulier (ici, tous les 10 jours), la phase d'intégration est interrompue par une étape d'analyse visant à ajuster l'état du modèle. Pour estimer les capacités prévisionnelles des différents schémas, ceux-ci sont tous utilisés ici en mode séquentiel (les observations n'étant rendues disponibles que sur une séquence d'assimilation incomplète).

Quatre méthodes sont considérées : (1) La réinitialisation (Réin) est la méthode la plus simple et consiste à réinitialiser brutalement le modèle en le forçant aux valeurs observées. Cette méthode s'apparente à la méthode de Cressmann (1959), mais appliquée ici à un système non spatialisé. L'approche ne prend en compte ni l'erreur du modèle, ni l'erreur des données. Elle revient à supposer les données parfaites et n'est donc en rien optimale. Cette méthode va jouer un rôle témoin afin de mieux distinguer l'apport des approches plus complexes. (2) Le filtre de Kalman étendu (ExKF) est une approche efficace permettant de prendre en compte les erreurs du modèle et des données. L'approche fait l'hypothèse d'une distribution Gaussienne des erreurs, peu adaptée au contexte chaotique. (3) Le filtre de Kalman d'Ensemble (EnKF ; Evensen 1994) ne nécessite pas d'hypothèse forte quant-à la distribution des erreurs. (4) Le Nudging direct et rétrograde (BFN) proposé par Auroux et Blum (2008) ne nécessite pas non plus d'hypothèse forte quant-à la distribution des erreurs et doit aussi permettre de mieux conserver le déterminisme des équations.

Pour disposer d'une estimation statistiquement significative, il est nécessaire de tester les capacités prédictives des approches sur un échantillon significatif. Une méthode consiste à séparer les observations en deux séquences (Mangiarotti et al., 2012a). Une première séquence de données destinées à être assimilée par le modèle pour un ajustement optimal de la trajectoire ; une seconde séquence, non destinée à être assimilée, conservée uniquement pour estimer les erreurs de prévision. En pratique une fenêtre d'assimilation de 5 années a été utilisée ici, contre une fenêtre prévisionnelle d'un an. Cette fenêtre de 5+1 années a été déplacée tout au long des 26 années de données disponibles avec un pas de temps de 10 jours permettant de disposer de l'erreur de prévision, au cours du temps, en fonction de l'horizon de prévision visé (voir Figure 3), constituant un ensemble d'estimation tout à fait significatif. La méthode a été appliquée à chacun des schémas d'assimilation de données.

4. Résultats et discussion

Les erreurs de prévision présentent de nombreux points communs quels que soient les schémas utilisés. On y distingue en particulier un cycle annuel assez marqué, avec des erreurs faibles dans l'entre saison, et des erreurs fortes avant la période de croissance qui tendent par fois à se réduire ensuite. Les erreurs pendant la phase de croissance sont plus souvent négatives, pouvant évoquer un biais du modèle. Toutefois ce biais

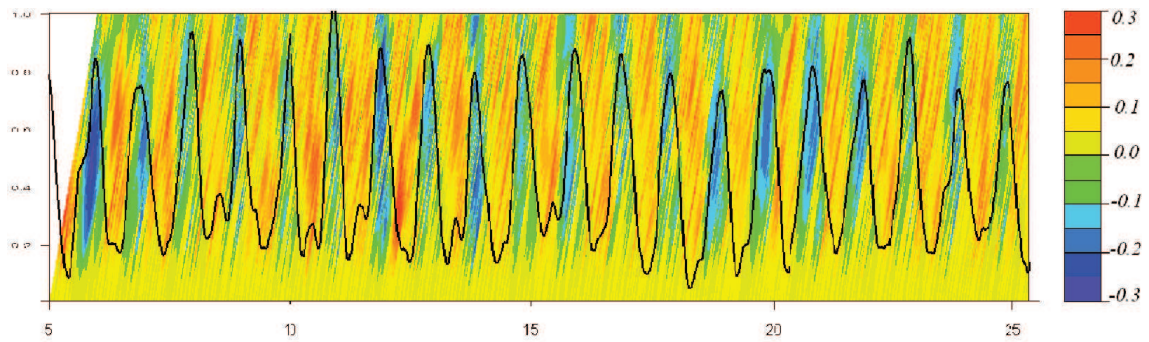


FIGURE 3 : Evolution des erreurs de prévision (erreur de NDVI sans unité) au cours du temps (abscisse en année) et en fonction de l'horizon de prévision visé (ordonnée également en année). Cas du filtre de Kalman étendu.

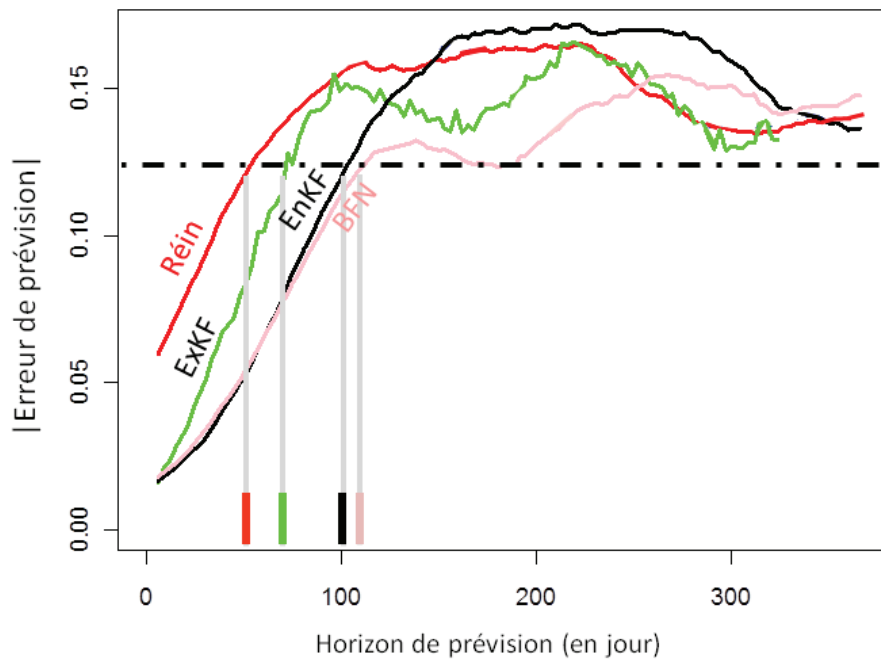


FIGURE 4 : 3ème quartile de la distribution d'erreur de prévision NDVI (sans unité) en fonction de l'horizon de prévision (en jour). Les résultats des quatre méthodes d'assimilation sont présentés : réinitialisation (Réin), filtre de Kalman étendu (ExKF), filtre de Kalman d'Ensemble (EnKF) et nudging direct et rétrograde (BFN).

ne semble pas systématique pouvant donc aussi refléter la propagation non-linéaire des erreurs conduisant à une perte de Gaussiennité. Plus subtilement, on y constate également l'influence des conditions initiales qui se traduisent ici par des traits obliques. On y distingue encore la croissance de l'erreur en fonction de l'horizon de prédiction : plus on cherche à prévoir à une échéance longue, plus l'erreur croît. Pour confronter les différentes approches proposées ici, on peut tracer l'évolution de cette erreur en fonction de l'horizon de prévision. Pour être représentatif de l'erreur maximum caractéristique, les courbes de la figure 4 présentent le 3ème quartile de l'erreur de prévision (en valeur absolue). Autrement dit, lors des prévisions, 75% des erreurs rencontrées sont inférieures aux valeurs présentées. Ces courbes présentent une convergence vers des valeurs toujours supérieures à 0,12. Cette valeur sera donc utilisée pour estimer des horizons de prévisibilités associés à chacune de ces méthodes.

Ces résultats illustrent bien la croissance rapide de l'erreur lorsque l'horizon de prévision croît et permet de classer les approches en termes de qualité de prévision. L'approche par réinitialisation ne permet pas de bonne prévision, même à court horizon, avec un horizon de prévisibilité proche de 50 jours. Par sa prise en compte des erreurs, le filtre de Kalman étendu permet de minimiser très significativement l'erreur à très court horizon mais son inaptitude à prendre en compte les non-linéarités de la dynamique conduisent à une croissance d'erreur très rapide. Les approches EnKF et BFN montrent bien leur capacité à prendre en compte les non-linéarités. Les deux approches présentent des résultats très similaires avec des horizons de prévisibilités proches de 100 jours, c'est-à-dire 3 mois ; l'approche BFN montre un léger avantage.

5. Conclusions

Combiné à des schémas d'assimilation de données, l'utilisation en mode prévisionnel de modèles canoniques obtenus par approche globale s'avère tout à fait envisageable. En contexte chaotique, les résultats illustrent bien la nécessité de choisir des schémas d'assimilation de données capables de prendre en compte tant les erreurs (des observations et du modèle) que les non linéarités du système. Parmi les quatre schémas d'assimilation comparés, les résultats obtenus avec les schémas les plus sophistiqués - le filtre de Kalman d'Ensemble et le Nudging direct et rétrogrades - fournissent ainsi des résultats très supérieurs aux autres approches. Par leur cohérence et leur qualité, ces résultats montrent donc l'applicabilité de modèles chaotiques de petite dimension à des contextes réels. Ces résultats contribuent également à valider le modèle.

Remerciements

Ce travail a été financé par le Centre National d'Etudes Spatiales et par le programme Les Enveloppes Fluides et l'Environnement (LEFE) de l'Institut National des Sciences de l'Univers (INSU). Les interactions avec le groupe de recherche DYnamique et COntôle des Ensembles Complexes (DYCOEC) y ont été très fructueuses.

Références

- Auroux, D. and Blum, J., 2008. A nudging-based data assimilation method for oceanographic problems : the back and forth nudging (bfn) algorithm. *Nonlinear Processes in Geophysics*, 15, 305-319.
- Balaghi, R., Tychon, B., Eerens, H. and Jlibene, M., 2007. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation* 10(4) :438-452.
- Cressmann, G.P., 1959. An operational objective analysis system. *Monthly Weather Review* 87(10) :367-374.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99(C5) :10143-10162.
- Gouesbet, G. and Letellier, C., 1994. Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets. *Physical Review E* 49(6) :4955-4972.
- Jarlan, L., Abaoui, J., Duchemin, B., Tourre, Y., Ouldabba, A., Mangiarotti, S., Kharrou, H., Khabba, S., Ballaghi, R., Le Page, M., Mokssit, A. and Chehbouni, A., 2012. Déterminants climatiques de la variabilité interannuelle des rendements en céréales et prévision précoce. Application à la province de Settat (Maroc). In : 9th International Conference African Association of Remote Sensing of the Environment, El Jadida, Morocco, Oct.-Nov. 2012.
- Knippertz, P., Fink, A.H., Reiner, A. and Speth, P., 2003. Three Late Summer/Early Autumn Cases of Tropical-Extratropical Interactions Causing Precipitation in Northwest Africa, *Monthly Weather Review* 131 :116-135.
- Letellier C., Aguirre L.A. and Maquet J., 2005. Relation between observability and differential embeddings for nonlinear dynamics. *Physical Review E* 71(2) :066213.1-066213.8.
- Letellier, C., Aguirre, L.A. and Freitas, U.S., 2009. Frequently asked questions about global modeling. *Chaos* 19.
- Mangiarotti, S., Coudret, R., Drapreau, L. and Jarlan, L., 2012b. Polynomial Model Search and Global Modelling - two new algorithms for global modelling of chaos. *Physical Review E* 86, 046205.
- Mangiarotti, S., Mazzega, P., Hiernaux, P. and Mougin, E. 2010. The vegetation cycle in West Africa from

AVHRR-NDVI data : Horizons of predictability versus spatial scales. *Remote Sensing of Environment* 114(9) :2036–2047.

Mangiarotti, S., Mazzega, P., Mougin, E.. and Hiernaux, P., 2012a. Predictability of vegetation cycles over the semi-arid region of Gourma (Mali) from forecasts of AVHRR-NDVI signals. *Remote Sensing of Environment* 123 :246–257.

Maquet, J., Letellier, C.. and Aguirre, L.A., 2007. Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems. *Mathematical Biology*, 55, pp. 21-39.

Mayaux, P., Bartholomé, E., Fritz, S.. and Belward, A., 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31, pp. 861-877.

Miller, R.N., Carter Jr., E.F. and Blue, S.T., 1999. Data assimilation into nonlinear stochastic models', *Tellus*, vol. 51A, pp. 167-194.

Savitzky, A. and Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Analytical Chemistry* 36 (8) :1627-1639.

Takens F., 1981. Detecting strange attractor in turbulence, in *Lecture Notes in Mathematics*, 898, Rand D. and L.S. Young eds., Springer Verlag, Berlin, pp. 366-381.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, pp. 127-150.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F. and Saleous, N.E., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing* 26(20) :4485-4498.