

Prévisions à court terme du niveau des aquifères : le cas de la nappe de Beauce

*Liliane BONNAL,
Pascal FAVARD*

**Groundwater level
short-term forecasts:
the case of the Beauce
water table**

Key-words:

groundwater, sustainable
management of an aquifer,
groundwater level forecast,
time series

Summary – *The objective of this article is to establish short term forecasts (one year), on the level of groundwater using conventional time series models. The interest of this work is to give policy makers reliable information on the evolution of groundwater levels. The forecasts which are essentially based on previous levels of groundwater are quite satisfactory. Indeed, the maximum forecast error is about half percent.*

**Prévisions
à court terme
du niveau des
aquifères : le cas de
la nappe de Beauce**

Mots-clés:

nappe phréatique, gestion
durable des aquifères,
prévisions du niveau
de la nappe,
séries temporelles.

Résumé – L'objectif de cette étude est de prévoir à court terme (c'est-à-dire pour une année), à partir des modélisations classiques appliquées aux séries temporelles, le niveau d'une nappe phréatique. L'intérêt de ce travail est, dans le cadre d'une gestion durable des aquifères, de fournir au décideur public des informations précises sur l'évolution des réserves en eau. Les valeurs prédites obtenues, essentiellement à partir des seuls niveaux de nappe passés, sont relativement satisfaisantes. En effet, l'erreur maximale de prévisions est de l'ordre d'un demi pour cent.

* *Université de Toulouse I, GREMAQ et LEERNA, 21 allée de Brienne, 31000 Toulouse.*

e-mail: bonnal@cict.fr

** *Université de La Rochelle, MSHS, 1, parvis Fernand Braudel, 17042 La Rochelle cedex.*

e-mail: pascal.favard@univ-lr.fr

Nous remercions Gilbert Colletaz et Brigitte Desaignes pour leurs commentaires. Nous sommes extrêmement reconnaissants à Christophe Bontemps et Ludovic d'Estampes pour leur aide.

BIEN que la France ait des dotations en eau relativement importantes, certaines régions vivent des situations tendues sur le plan de la disponibilité de la ressource. Ce problème a conduit les pouvoirs publics à prendre des mesures de rationnement pénalisant fortement les usagers ⁽¹⁾. Ces mesures sont toutefois jugées nécessaires, car des prélèvements importants en période de sécheresse peuvent avoir des effets externes (salinisation, pollution) sur le milieu aquatique et donc perturber l'équilibre naturel. En effet, dans certains cas, lorsque le niveau d'une nappe est insuffisant, la fonction réservoir de la nappe est irrémédiablement perdue. Pour éviter les problèmes d'irréversibilité et/ou de rationnements brutaux, il est indispensable d'avoir une bonne compréhension du fonctionnement d'une nappe phréatique. L'évaluation des disponibilités physiques en eau (abusivement appelée l'offre d'eau) va donc permettre d'améliorer la gestion de l'eau ⁽²⁾. Cette évaluation peut être faite pour différents horizons : à un an, afin de permettre aux agriculteurs de déterminer leur politique d'arrosage des cultures, à trois ans, afin qu'ils puissent mettre en place des politiques d'assolements, et enfin à très long terme, pour tenir compte des besoins futurs, en d'autres termes pour arbitrer de façon « optimale » entre générations.

Que peut faire le décideur public lorsque localement la demande globale en eau est supérieure à l'offre ⁽³⁾? Deux types de méthodes sont utilisées pour rétablir l'équilibre.

Le premier type de méthode, très appliqué par le passé, consiste à augmenter les disponibilités en eau. Pour ce faire, on peut par exemple construire des barrages, des canaux, des lacs collinaires... Il est évident que localement les quantités d'eau disponibles vont augmenter, mais, les problèmes liés à une affectation correcte de la ressource ne sont pas pour autant résolus. Le second type de méthode, proposé par l'économiste, consiste à maximiser la fonction de bien-être social intertemporel. Cette méthode permet de résoudre ce problème d'allocation, soit en déterminant un système de prix, soit en fixant un système de quotas. Afin de pouvoir mettre en place l'une de ces politiques économiques, il est indispensable d'évaluer en premier lieu les disponibilités de la ressource présentes et futures. Ces disponibilités dépendent essentiellement des infiltrations, des écoulements et des prélèvements. Les infiltrations (observées en saison hivernale) viennent alimenter les eaux souterraines de façon plus ou moins rapide selon la nature du sol (De Marsily, 1995).

⁽¹⁾ En particulier les agriculteurs dont la demande d'irrigation représente 65 % des consommations nettes d'eau.

⁽²⁾ La gestion « durable » des aquifères nécessite la connaissance des disponibilités physiques mais aussi des différentes demandes pour la période considérée et les périodes futures.

⁽³⁾ L'offre en eau représente ici la quantité d'eau pouvant être consommée sans mettre en péril l'équilibre naturel.

Mais l'évaluation de la quantité d'eau infiltrée dans la nappe est difficile ⁽⁴⁾. Les écoulements ne sont généralement pas mesurables et il semblerait même que ceux-ci ne soient pas toujours expliqués. Enfin, les prélèvements dépendent des demandes en eau. Il existe trois types d'agents économiques ayant chacun une fonction de demande ⁽⁵⁾ en eau spécifique: la demande domestique, la demande industrielle et la demande agricole. Toutefois, par manque d'information, l'évaluation de ces fonctions de demande est difficile à réaliser. De plus, comme nous ne disposons pas de travaux hydrogéologiques appliqués, l'étude de l'évolution du niveau des disponibilités en eau doit être réalisée à partir de modèles purement statistiques. Dans cette étude, nous allons mesurer les disponibilités d'une nappe d'eau ⁽⁶⁾ et plus particulièrement celles de la nappe de Beauce ⁽⁷⁾. Sachant que durant l'automne les exploitants décident des assolements qu'ils vont mettre en place l'année suivante, les prévisions seront toujours réalisées à partir du mois d'octobre d'une année donnée jusqu'au mois de septembre de l'année suivante.

Nous avons procédé en trois étapes. La première partie donne une brève description de la nappe de Beauce. La deuxième partie présente les modélisations considérées tandis que les résultats et les prévisions obtenus font l'objet de la troisième partie.

LA NAPPE DE BEAUCE : PRINCIPALES CARACTÉRISTIQUES

La Beauce est une région de grandes cultures céréalières (blé, maïs, orge), maraîchères et betteravière. Par la taille moyenne de ses exploitations (100 hectares), la Beauce se place parmi les régions céréalières les plus spécialisées d'Europe. Cette région est située au-dessus d'une réserve d'eau, la nappe de Beauce, qui s'étend sur environ 9 000 km² et sur six départements: le Loiret, l'Eure-et-Loir, le Loir-et-Cher, la Seine-et-Marne, les Yvelines et l'Essonne. Elle est donc localisée sur deux régions administratives: le Centre et l'Île-de-France. Elle appartient à deux bassins hydrographiques qui relèvent de deux agences: Loire-Bretagne et

⁽⁴⁾ Ce travail est réalisé par des hydro-géologues.

⁽⁵⁾ Pour plus de détails sur la valorisation des usages de l'eau, nous renvoyons à Amigues *et al.*, 1995.

⁽⁶⁾ Une nappe souterraine est une masse d'eau remplissant les interstices d'un terrain poreux et perméable (aquifère) de telle sorte qu'il y a toujours des liaisons par l'eau entre les pores. Toute nappe se forme par accumulation des eaux d'infiltrations au-dessus d'un terrain imperméable stoppant leur progression. Ce terrain est appelé *substratum* ou mur.

⁽⁷⁾ Le mur de la nappe de Beauce a grossièrement la forme d'une cuvette. L'épaisseur du réservoir est variable: elle peut être nulle par endroit et atteindre environ 150 mètres au centre de la cuvette. La nappe de Beauce est telle que les eaux ne sont pas immobiles dans les couches perméables (pour une présentation précise de la nappe de Beauce, cf. Diego et Fourcade, 1997).

Seine-Normandie. Le volume de cette nappe est estimé à 3 milliards de m³. Il existe peu de cours d'eau permanents sur ce territoire et la nappe est la source principale d'alimentation, en eau potable et en eau d'irrigation. Elle présente donc un intérêt économique très important.

Des phénomènes climatiques (tels que, par exemple, la sécheresse du début des années 1990) et des modifications du contexte économique agricole telles que la réforme de la Politique agricole commune ⁽⁸⁾ (PAC) de 1992 ont provoqué une baisse du niveau de la nappe, qui, à plusieurs reprises, s'est « dangereusement » rapproché du seuil critique pour finalement l'atteindre en 1993. Ce seuil est le niveau à partir duquel le risque de détérioration de la nappe est important. Les pouvoirs publics sont alors contraints de prendre des mesures de rationnement, en particulier en ce qui concerne les agriculteurs irrigués.

Le niveau de la nappe (figure 1), mesuré en mètres, est disponible mois par mois de janvier ⁽⁹⁾ 1975 à mai 1998. C'est une moyenne pondérée des 9 piézomètres répartis sur la région. Pour l'ensemble de la période observée, la hauteur moyenne de la nappe est de 107,39 mètres; son maximum, 110,82 mètres, a été atteint en juillet 1983 et son minimum, 103,86 mètres, en septembre 1993. Ce niveau est inférieur au seuil critique fixé à 103,9 mètres, limite au-dessous de laquelle il y a risque d'assèchement des rivières.

Figure 1.
Evolution du niveau
moyen de la nappe de
Beauce (en mètres)



⁽⁸⁾ La région a connu des changements de modes et de types de cultures induits par la réforme de la Politique agricole commune (PAC) de 1992. En effet, suite à la mise en place de cette réforme (dont la principale mesure est d'accorder une prime à l'hectare progressivement croissante de 1992 à 1996 aux cultures de blé, de maïs, d'orge et de pois sur surfaces irriguées, ces dernières devant être plafonnées), les cultures ont été modifiées et la part des surfaces irriguées est progressivement passée de un tiers à deux tiers (en 1995) des superficies cultivées (cf. Favard et Foncel, 1998). La proportion de surface irriguée est maintenant de l'ordre de 60 % pour le blé, 80 % pour le maïs et 25 % pour l'orge.

⁽⁹⁾ On dispose aussi de relevés hebdomadaires mais, pour des raisons techniques, seules les données mensuelles ont été exploitées.

MODÉLISATION

Lors de l'étude d'une série temporelle, il est nécessaire d'identifier la tendance et la saisonnalité qui peuvent éventuellement être observées. Cette tendance et cette saisonnalité peuvent être soit déterministes, soit stochastiques. Si elles sont déterministes, leur correction s'obtient en introduisant dans le modèle une variable associée au temps (*trend*) et des variables dichotomiques associées aux différentes périodes d'observations des données (mois lorsque les données sont annuelles, trimestres pour des données trimestrielles...). Si elles sont stochastiques, il sera alors nécessaire de différencier la série. Dans ce dernier cas, la modélisation la plus généralement utilisée est de type SARIMA. Un SARIMA $(p,d,q)(P,D,Q)_s$ est défini par la relation :

$$\begin{aligned} & (1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}) (1 - B^S)^D (1 - \phi_1 B - \dots - \phi_p B^p) (1 - B)^d Y_t \\ & = (1 + \theta_1 B + \dots + \theta_q B^q) (1 + \Psi_1 B^S + \dots + \Psi_Q B^{QS}) \varepsilon_t \end{aligned} \quad (1)$$

où $Y_t (t = 1, \dots, T)$ est la série brute, $\varepsilon_t (t = 1, \dots, T)$ le terme d'erreur associé, B l'opérateur retard tel que $B^S Y_t = Y_t - Y_{t-S}$ et S la période (égale à 12 dans l'étude car les données sont mensuelles). Les Φ , ϕ , θ et Ψ sont les paramètres à estimer. Les lettres minuscules (resp. majuscules) sont associées à la partie non saisonnière (resp. saisonnière) du modèle. d (resp. D) détermine le degré de différentiation de la série et permet de corriger la stationnarité (resp. de différentiation saisonnière et permet de corriger la saisonnalité), p (resp. P) caractérise l'ordre du polynôme AR (resp. AR saisonnier) et q (resp. Q) l'ordre du polynôme MA (resp. MA saisonnier).

Avant de déterminer les degrés (p, q, P, Q) du modèle, il est nécessaire de connaître les éventuels degrés de différentiation d et D . Accepter de différencier la série à l'ordre d pour la tendance et D pour la saisonnalité implique que ces dernières sont stochastiques. Deux méthodes peuvent être utilisées pour déterminer d et D .

La première permet de déterminer ces degrés à partir du calcul des variances de la série différenciée $\text{Var} [(1 - B)^\ell (1 - B^S)^j Y_t]$; avec ⁽¹⁰⁾ $\ell = 0, \dots, 5$; $j = 0, \dots, 3$, d et D sont alors tels que :

$$(d, D) = \underset{\ell, j}{\text{argmin}} (\text{Var} [(1 - B)^\ell (1 - B^{12})^j Y_t]).$$

La seconde consiste à réaliser des tests de racines unitaires avec saisonnalité (les principaux tests sont présentés dans l'encadré).

⁽¹⁰⁾ Les valeurs des variances associées aux séries différenciées sont, en général, décroissantes puis croissantes. Généralement, la valeur maximale prise par ℓ est 5 et celle prise par j est 3. Ces dernières seront plus élevées si la variance de la série différenciée pour $\ell > 5$ et/ou $j > 3$ continue de décroître (cf. Box et Jenkins, 1976).

Tests de racines unités saisonnières

Il existe plusieurs méthodes pour tester si la série étudiée a une tendance et une saisonnalité stochastique, c'est-à-dire s'il est nécessaire de différencier la série à l'ordre 1 et à l'ordre $S = 12$ (ou 4 si les données sont trimestrielles). Trois de ces méthodes sont proposées dans cet encadré.

- **Hasza et Fuller (1982)** considèrent le modèle

$$y_t = \phi_1 y_{t-1} + \phi_S y_{t-S} + \phi_{S+1} y_{t-(S+1)} + u_t$$

où u_t est un bruit blanc. On teste alors, selon une méthode classique, l'hypothèse nulle définie par $\phi_1 = \phi_S = -\phi_{s+1} = 1$, l'hypothèse alternative est alors aucune racine unitaire.

- **Osborn, Chui, Smith et Birchenhall (OCSB, 1988)** ont étendu cette approche à l'aide de la régression suivante :

$$\Theta_{p-S-1}(B)(1-B)(1-B^S)y_t = \sum_{s=1}^S \delta_s D_{s,t} + \pi_1 (1-B^S)y_{t-1} + \pi_2 (1-B)y_{t-S} + \varepsilon_t$$

où $\Theta_{p-S-1}(B)$ est un polynôme en B à l'ordre $p-S-1$ ou $p-S-1$ est déterminé tel que les résidus estimés suivent un bruit blanc. Si l'on accepte l'hypothèse $\pi_2 = 0$, la différence à l'ordre S est appropriée et si $\pi_1 = \pi_2 = 0$ le filtre $(1-B)(1-B^S)$ est nécessaire. Bien que la racine de $(1-B)$ soit une des racines de $(1-B^S)$, on ne peut rien conclure à partir du paramètre π_1 . Les valeurs des statistiques associées à ces deux tests sont données par exemple dans Franses (1998). Si l'hypothèse de test $\pi_1 = \pi_2 = 0$ est rejetée, on cherche alors les racines unitaires à l'aide d'un test standard de type Dickey et Fuller.

- **Hylleberg, Engle, Granger et Yoo (HEGY, 1990)** ont proposé un test qui est une extension de l'approche de Dickey, Hasza et Fuller (1984). Le test qui va être présenté ici est celui associé à une différenciation saisonnière mensuelle ⁽¹¹⁾ de type $Y_t - Y_{t-12}$. L'équation associée au test de HEGY est définie par :

$$\Theta(B)y_{8,t} = \mu_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-2} + \pi_5 y_{4,t-1} + \pi_6 y_{4,t-2} + \pi_7 y_{5,t-1} + \pi_8 y_{5,t-2} + \pi_9 y_{6,t-1} + \pi_{10} y_{6,t-2} + \pi_{11} y_{7,t-1} + \pi_{12} y_{7,t-2}$$

où $\Theta(B)$ est un polynôme en B tel que les résidus estimés suivent un bruit blanc,

$$\begin{aligned} y_{1,t} &= (1+B)(1+B^2)(1+B^4+B^8)y_t, \\ y_{2,t} &= -(1-B)(1+B^2)(1+B^4+B^8)y_t, \\ y_{3,t} &= -(1-B^2)(1+B^4+B^8)y_t, \\ y_{4,t} &= -(1-B^4)(1-B\sqrt{3+B^2})(1+B^2+B^4)y_t, \\ y_{5,t} &= -(1-B^4)(1+B\sqrt{3+B^2})(1+B^2+B^4)y_t, \\ y_{6,t} &= -(1-B^4)(1-B^2+B^4)(1-B+B^2)y_t, \\ y_{7,t} &= -(1-B^4)(1-B^2+B^4)(1+B+B^2)y_t, \\ y_{8,t} &= -(1-B^{12})y_t, \end{aligned}$$

Les variables $y_{h,t}$, $h = 1, \dots, 8$, sont associées aux différentes racines du polynôme. Pour les séries mensuelles, on peut considérer les statistiques du «*t*» de Student pour π_1 et π_2 et les statistiques de «*F*» de Fischer pour $\{\pi_3, \pi_4\}$, $\{\pi_5, \pi_6\}$, $\{\pi_7, \pi_8\}$, $\{\pi_9, \pi_{10}\}$, $\{\pi_{11}, \pi_{12}\}$. Les valeurs critiques de ces tests sont données dans Beaulieu et Miron (1993) ou Franses (1998).

⁽¹¹⁾ Le test associé à une différenciation saisonnière trimestrielle est présenté par exemple dans Franses (1998).

Le calcul des variances des séries différenciées est donné dans le tableau 1. Quatre séries, notées $Z_{h,t}$ $h = 1, \dots, 4$, pourraient être envisagées :

$$\begin{aligned} Z_{1t} &= (1 - B) Y_t, \\ Z_{2t} &= (1 - B) (1 - B^{12}) Y_t, \\ Z_{3t} &= (1 - B)^2 Y_t, \\ Z_{4t} &= (1 - B)^2 (1 - B^{12}) Y_t. \end{aligned} \tag{2}$$

Tableau 1.
Variance des séries différenciées

		<i>l</i>					
		0	1	2	3	4	5
<i>j</i>	0	3,8192	0,0198	0,0183	0,0469	0,1498	0,5155
	1	0,7304	0,0287	0,0335	0,0972	0,3256	1,1458
	2	1,0420	0,0851	0,1030	0,3062	1,0361	3,6502
	3	2,6718	0,2791	0,3514	1,0607	3,6047	12,693

Parmi les tests de racines unitaires saisonniers, deux ont été réalisés : le test de HEGY et celui de OCSB. Les modèles permettant de réaliser ces tests incluent comme variables auxiliaires la constante, des variables dichotomiques mensuelles et une tendance linéaire déterministe. La statistique associée au test OCSB est égale à 0,085. La valeur critique de ce test à 5 % étant de 35,87 ; on accepte donc une double différenciation à l'ordre 1. En ce qui concerne le test de HEGY, afin d'obtenir des résidus non corrélés, les retards 2 et 4 ont été introduits. Les statistiques du « *t* » de Student associées à π_1 et π_2 sont respectivement égales à 1,54 et - 0,67. Les statistiques de « *F* » de Fischer sont telles que $F(\pi_3, \pi_4) = 39,6$; $F(\pi_5, \pi_6) = 352,3$; $F(\pi_7, \pi_8) = 283$; $F(\pi_9, \pi_{10}) = 223,5$ et $F(\pi_{11}, \pi_{12}) = 89,7$. Si l'on compare ces résultats aux valeurs critiques (cf. Franses, 1998 par exemple), il apparaît que, à partir du test de HEGY, la différenciation la plus appropriée à la série est à l'ordre 1 pour la tendance et à l'ordre 2 pour la saisonnalité (soit $(1 - B) (1 - B^2)$).

RÉSULTATS EMPIRIQUES

Choix et validation des modèles retenus

L'objectif de cette étude est de déterminer les prévisions à court terme (pour 12 mois) de la série du niveau de la nappe. Cette série est observée de janvier 1975 à mai 1998. Une fois les degrés de différenciation (*d* et *D*) définis, il est nécessaire de déterminer l'ordre des différents polynômes retards (équation (1)). Les différentes séries considérées sont : la série brute corrigée d'une tendance et d'une saisonnalité déterministe, les quatre séries Z_{ht} $h = 1, \dots, 4$, (définies en (2)) suggérées par le calcul des variances (qui inclut la série proposée par le test de OCSB) et la série proposée par le test de HEGY. Pour chaque série, le modèle retenu est celui qui minimise le critère d'information d'Akaike (AIC). Ce critère permet de sélectionner

entre plusieurs modèles et d'évaluer les estimations selon le nombre de paramètres. Il est défini par $AIC(k) = n \text{Ln } \hat{\sigma}_{ML}^2 + 2k$, où n est le nombre d'observations utilisées pour estimer les paramètres, k est le nombre de paramètres à estimer pour le modèle ARMA et $\hat{\sigma}_{ML}^2$ est la variance empirique.

Afin de tester la validité et la robustesse des différents modèles, les estimations ont été réalisées sur plusieurs années en faisant, pour chacune d'elles, varier la date de fin d'observation. En pratique, il apparaît que les modèles qui semblent être bien adaptés aux données observées ne fournissent pas forcément les meilleures prévisions. Il est alors possible, après avoir sélectionné les modèles selon le critère AIC, de les évaluer à partir des prévisions en utilisant l'indicateur de moyenne des carrés des erreurs de prévisions ⁽¹²⁾ défini comme suit :

$$MSPE = \frac{1}{m} \left[\sum_{k=1}^m (\hat{y}_{n+k} - y_{n+k})^2 \right].$$

Les MSPE (multipliées par 100) des modèles sélectionnés par AIC sont données dans le tableau 2. Les valeurs des moyennes ont été multipliées par 100. Elles sont calculées selon le modèle et l'année de fin d'observation.

Tableau 2. Performance prévisionnelle des modèles retenus

Variable(*)	Modèle(**)(***)	1990	1991	1992	1993	1994	1995	1996	1997	
Transformations Box et Jenkins										
1, Y_t	AR(1-4)	<i>c, M, t</i>	11,4	2,3	8	66	60,6	19,2	1,1	8,2
2, Z_{1t}	ARMA(1,2)	<i>c, M</i>	25,1	3,2	2,9	72,7	31,4	43,2	3,3	3,3
3, Z_{2t}	AR(1-4)	<i>c, M</i>	3,1	2,9	2,1	29,6	2,0	105,7	0,8	30,4
4, Z_{3t}	SARIMA(1,1,2) (0,1,1-2) ₁₂		8	1	2,2	39,7	17,1	53,4	0,9	5,8
5, Z_{4t}	SARIMA(2,2,0) (0,1,1-2) ₁₂		0,9	2,6	2,8	33	5,1	90,6	1,3	0,2
Transformation indiquée par le test de HEGY										
6. (1-B) (1-B ²) Y_t	SARIMA(0,1,1-3) (0,1,1) ₂		0,5	11,8	3,4	90,9	24,1	68,5	2,6	17,3

(*) Cette colonne caractérise la série brute Y_t et ses transformations. Les prévisions sont évaluées pour la série brute et mesurées en mètres. Les erreurs de prévisions sont elles aussi exprimées en mètres.

(**) Les modèles incluent les retards notés dans la parenthèse. « 1-4 » signifie que tous les retards de 1 à 4 de la série considérée sont inclus dans le modèle, « 2 » signifie que seul le retard d'ordre 2 est inclus dans le modèle.

(***) Le modèle peut inclure la constante (*c*), des variables dichotomiques saisonnières (*M*) et une variable temporelle (*t*).

⁽¹²⁾ Pour tester si les critères de MSPE sont significativement différents entre deux modèles A et B, il est possible d'utiliser la statistique de test notée *S* et définie par

$$S = \left(\frac{m}{4} \right)^{-\frac{1}{2}} \left[\sum_{k=1}^m d_k - \frac{m}{2} \right] \sim N(0,1),$$

où par exemple d_k ($k = 1, \dots, m$) vaut 1 si $MSPE_A$ est supérieure à $MSPE_B$ et 0 sinon. Ce critère peut être utilisé si le nombre de prévisions est suffisamment grand (cf. Diebold et Mariano, 1995). D'autres critères peuvent être utilisés tels que le critère MAPE.

A la vue de ce tableau, les modèles 4 et 5 sont ceux qui ont, sur l'ensemble des années, les plus petites MSPE. Le modèle 4 est associé à la série différenciée à l'ordre 1 pour la tendance et à l'ordre 1 pour la saisonnalité ($d=1$ et $D=1$, cette différenciation était acceptée par le test de OCSB et suggérée par le critère de la variance minimale). Le modèle 5 est quant à lui associé à la série différenciée à l'ordre 2 pour la tendance et à l'ordre 1 pour la saisonnalité ($d=2$ et $D=1$, cette série était suggérée par le critère de la variance minimale). La tendance et la saisonnalité semblent donc être stochastiques. Dans la suite de l'étude, nous allons utiliser ces deux transformations.

Des tests associés à la variance et à la normalité des résidus ont été réalisés. Seuls les tests portant sur la variance sont présentés. Jusqu'à présent, nous avons supposé que la variance des erreurs de mesure (et donc la variance du niveau de la nappe) était homoscédastique, c'est-à-dire constante au cours de la période d'observation (1975-1998). Or, la variance pourrait dépendre de la période d'observation. Les modèles qui permettent de corriger ce problème sont les modèles autorégressifs conditionnellement hétéroscédastiques (les modèles ARCH, cf. Franses (1998) pour plus de détails sur ces modèles). Pour analyser la présence d'hétéroscédasticité, plusieurs tests ont été proposés. Le test utilisé ici est un test du multiplicateur de Lagrange. L'objectif de ce test est de détecter l'hétéroscédasticité des résidus d'un modèle linéaire défini de la façon suivante: $y = X' \beta + u$ où $y = (y_1, \dots, y_T)'$, X est une matrice de dimension (T, k) , β est le vecteur de paramètres associés, $u = (u_1, \dots, u_T)'$, avec $u_t \sim N(0, \sigma_t^2)$ et $\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}^2 = h(z_t' \alpha)$. Le principe est de tester la nullité des $(\alpha_i)_{i=1}^p$, ce qui rend la variance constante.

Les tests réalisés sont à l'ordre deux⁽¹³⁾ ($p=2$), c'est-à-dire que l'on suppose que $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2$. Les valeurs des statistiques empiriques du test sont toutes supérieures à 10 pour le modèle 4, et à 18 pour le modèle 5. On accepte donc l'hypothèse de la présence d'hétéroscédasticité. Les modèles considérés sont maintenant des SARIMA avec résidus ARCH. Les prévisions associées ayant subi peu de modifications, les MSPE ne seront pas présentées (les conclusions restant inchangées). De plus, on peut noter que, d'après le test de Bera et Jarque, on accepte la normalité des résidus.

Les prévisions

Bien que les coefficients associés aux différents modèles retenus soient stables, pour les différentes années considérées, les erreurs de prévisions sont quant à elles, selon les années, quelque peu différentes. Ces

⁽¹³⁾ Des tests à l'ordre supérieur ont été réalisés et se sont avérés non significatifs.

erreurs de prévisions calculées pour les douze premiers mois (d'octobre de la dernière année considérée à septembre de l'année suivante) mesurent l'écart entre la valeur observée et la valeur prédite c'est-à-dire $Y_t - \hat{Y}_t$. Excepté pour les années 1993 et 1995, les erreurs maximales de prévision observées sont inférieures à 35 centimètres. Toutefois, pour la période 1993-1994, la nappe est, dès février, très fortement sous-estimée (l'erreur observée étant de 1 mètre pour le mois de septembre aussi bien pour le modèle 4 que pour le modèle 5), tandis que pour la période 1995-1996, la nappe est à l'inverse très fortement surestimée (l'erreur observée étant en septembre proche de 1,20 mètre pour le modèle 4 et de 1,80 mètre pour le modèle 5). Peut-on expliquer ces mauvaises prévisions ?

Lorsque l'on s'intéresse à l'évolution du niveau de la nappe (figure 1), on constate que celle-ci connaît une décroissance à partir de 1989 jusqu'au milieu de l'année 1993. Cette décroissance semble, en grande partie, liée à la période de sécheresse observée dans la Beauce entre 1989 et 1992 (cf. Bonnal et Favard, 1998 et 2000). Etant donné la trajectoire passée de la nappe, le modèle considéré (de par sa structure) maintient la tendance à la baisse du niveau de la nappe enregistrée depuis plus de trois ans et n'arrive pas à prévoir la rupture de tendance qui a lieu à la fin de l'année 1993. En effet, la modélisation considérée ne peut pas prévoir la fin d'un choc exogène à notre modèle (fin de la période de sécheresse), lorsque celui-ci a été durable. Ce phénomène peut donc expliquer la sous-estimation de la nappe observée pour la période 1993-1994.

On constate que la réforme de la Politique agricole commune de 1992 a eu pour conséquence de développer les cultures sur sols irrigués. A la suite de cette réforme, la part des surfaces irriguées, sur la région étudiée, est progressivement passée d'un tiers des surfaces cultivées à deux tiers en 1995 (cf. Favard et Foncel, 1998). Ces changements ont dû, a priori, conduire les agriculteurs à augmenter leur demande en eau. Ceci a pour conséquence, toutes choses restant égales par ailleurs, de diminuer le niveau de la nappe. Une fois encore, ne prenant pas en compte de chocs exogènes, la modélisation considérée ne permet pas d'anticiper la baisse du niveau de la nappe à partir des seules observations des niveaux de nappe passés. Ceci explique alors la surestimation du niveau de la nappe pour la période 1995-1996.

Nous allons donc utiliser une modélisation plus générale dans laquelle il est possible d'introduire des variables d'intervention (variables exogènes), c'est-à-dire considérer une modélisation SARIMAX. Ces variables exogènes notées X vont nous permettre de prendre en compte des ruptures de tendance considérées comme des chocs exogènes sur la série. La première variable explicative considérée caractérise la période de sécheresse observée de juin 1989 à août 1992. La deuxième variable explicative est associée à la réforme de la PAC de 1992. Il semblerait que les effets attendus de cette politique aient été pleinement observés à partir de 1995 (cf. Favard et Foncel, 1998). Pour un exploitant, irriguer im-

plique des investissements notamment en matériel de pompage et de distribution d'eau. Il semble alors assez intuitif que les agriculteurs ne puissent pas faire ces investissements de manière instantanée. Par conséquent, ils étalent dans le temps cette augmentation de capital fixe. Le modèle avec variables d'intervention s'écrit alors :

$$\begin{aligned} & (1 - \Phi_1 B^S - \dots - \Phi_p B^{pS}) (1 - B^S)^D (1 - \phi_1 B - \dots - \phi_p B^p) (1 - B)^d Y_t \\ & = \sum_{j=1}^V \frac{P_{X_j}(B)}{\Theta_{X_j}} X_{jt} + (1 + \theta_1 B + \dots + \theta_q B^q) (1 + \Psi_1 B^S + \dots + \Psi_Q B^{QS}) \varepsilon_t \end{aligned}$$

où X_{jt} est la $j^{ième}$ variable d'intervention et $P_{X_j}(B)$; $\Theta_{X_j}(B)$ sont des polynômes retardés associés à X_{jt}

Les estimations ont été reprises pour chacune des années. Toutefois, seules les variables d'intervention considérées sur la période 1993-1995 se sont avérées significatives. De plus, on ne considère qu'un choc exogène par année ($v=1$). Pour l'estimation réalisée en 1993, la variable d'intervention, X_{1t} (nommée Sec_t) caractérise le phénomène de sécheresse observé sur la période 1989-1992. Cette variable prend la valeur 1 durant la période de sécheresse et la valeur 0 sinon. Pour les estimations réalisées en 1994 et 1995, la variable d'intervention considérée, X_{2t} (nommée Pac_t) est associée à la réforme de la PAC. Cette variable prend la valeur 1 pour les années 1994, 1995 et 1996, le choc exogène lié à cette politique pouvant avoir eu encore des répercussions en 1996. Après cette date, on suppose que la modélisation retenue incorpore directement cet impact. Pour chacune de ces variables d'intervention le dénominateur $\Theta_{X_j}(B)$ est toujours égal à 1. Le numérateur $P_{X_j}(B)$ inclut les variables Sec ou Pac selon les années avec des retards éventuels.

Un test du rapport de maximum de vraisemblance permet de mesurer la pertinence de la prise en compte de ces variables exogènes. La statistique associée à ce test suit un Chi-deux à deux degrés de liberté. Les valeurs empiriques du test sont comprises entre 3 et 4 selon les années pour le modèle 4 et supérieures à 4 pour le modèle 5. Pour ce dernier, l'estimation avec variables d'intervention est rejetée pour l'année 1994 (la valeur du test étant de 4,2). En revanche, la prise en compte, d'une part, de la variable associée à la sécheresse pour l'année 1993 est acceptée avec une erreur de 10 % (la valeur du test étant égale à 5), et, d'autre part, de la variable associée à la PAC pour l'année 1995 est acceptée avec une erreur de 5 % (la valeur du test étant égale à 6,8).

Si l'on s'intéresse aux MSPE associées à ces différentes années, il apparaît que pour le modèle 4 sans variable d'intervention ⁽¹⁴⁾, (resp. modèle 5 avec variables d'intervention), ces valeurs sont de 40,6 (5,6) pour

⁽¹⁴⁾ Les valeurs des MSPE ne correspondent pas à celles du tableau 1 car les modèles considérés incluent la correction de l'hétéroscédasticité (SARIMA-ARCH).

l'année 1993, 12,3 (9,0) pour l'année 1994, 51,5 (8,3) pour l'année 1995. A la vue de ces valeurs, nous pouvons considérer que le modèle permettant de prévoir le plus précisément le niveau de la nappe est le modèle 5, c'est-à-dire celui pour lequel nous avons différencié à l'ordre 2 pour la tendance et à l'ordre 1 pour la saison. Notons toutefois que les tests de racines unités saisonnières effectués précédemment avaient rejeté la différenciation à l'ordre 2⁽¹⁵⁾. Ces différenciations supposent qu'il existe une relation fixe entre les trois derniers niveaux de la nappe. Or, les pluies efficaces mettent en moyenne trois mois pour pénétrer à l'intérieur de la nappe⁽¹⁶⁾, ceci est dû au fait que la nappe de Beauce est peu profonde et parfois même affleurante, ce phénomène étant saisonnier (d'où la différenciation à l'ordre 12).

Les erreurs de prévisions du niveau de la nappe pour le modèle 5 avec variables d'intervention sont présentées dans le tableau 3. Ces variables d'intervention permettent d'améliorer les prévisions associées à la période 1993-1995. En effet, pour l'année 1993, l'erreur maximale de prévision observée pour le mois de septembre était proche de 1 mètre, elle n'est plus que de 40 centimètres après la prise en compte de la période de sécheresse. La réforme de la PAC de 1992 a un effet très fort sur les prévisions du niveau de la nappe réalisées pour l'année 1995, l'erreur maximale passant de 1,70 mètres à 55 centimètres (le niveau étant encore surestimé). Les graphiques des prévisions, associés au modèle 5 pour différentes années, sont donnés dans les figures 2, 3 et 4.

Tableau 3. Erreurs de prévisions (en mètres) pour le modèle SARIMAX(2,2,0)(0,1,2)₁₂ avec bruit ARCH(2)

Mois	Année							
	1990-1991	1991-1992	1992-1993	1993-1994	1994-1995	1995-1996	1996-1997	1997-1998
Octobre	0,03	-0,04	0,00	0,03	-0,06	0,00	-0,02	-0,02
Novembre	0,07	-0,02	0,01	0,06	-0,13	0,06	0,00	-0,01
Décembre	0,09	0,02	0,00	0,10	-0,17	0,08	0,00	-0,02
Janvier	0,11	0,08	0,01	0,22	-0,32	0,15	-0,05	0,05
Février	0,05	0,07	0,01	0,34	-0,30	0,06	-0,12	0,06
Mars	-0,04	0,20	-0,06	0,40	-0,35	-0,01	-0,09	0,04
Avril	-0,09	0,29	-0,07	0,37	-0,31	-0,16	-0,21	0,12
Mai	-0,17	0,28	0,03	0,28	-0,25	-0,40	-0,17	0,17
Juin	-0,14	0,24	0,18	0,24	-0,21	-0,53	-0,11	
Juillet	-0,11	0,20	0,27	0,11	-0,32	-0,51	0,03	
Août	-0,08	0,12	0,27	0,12	-0,46	-0,44	-0,02	
Septembre	-0,07	0,10	0,34	0,17	-0,45	-0,21	0,02	

Pour les années 1990 à 1992, ainsi que pour 1996 et 1997, les erreurs de prévisions ont été obtenues à partir du modèle sans variable d'intervention avec bruit ARCH (modèle 5).

⁽¹⁵⁾ Cette différenciation est en général rejetée pour la plupart des séries temporelles.

⁽¹⁶⁾ Cf. Diego et Fourcade (1997), Bonnal et Favard (2000).

Figure 2.
Prévisions du niveau
de la nappe pour la
période 1993-1994
(en mètres)

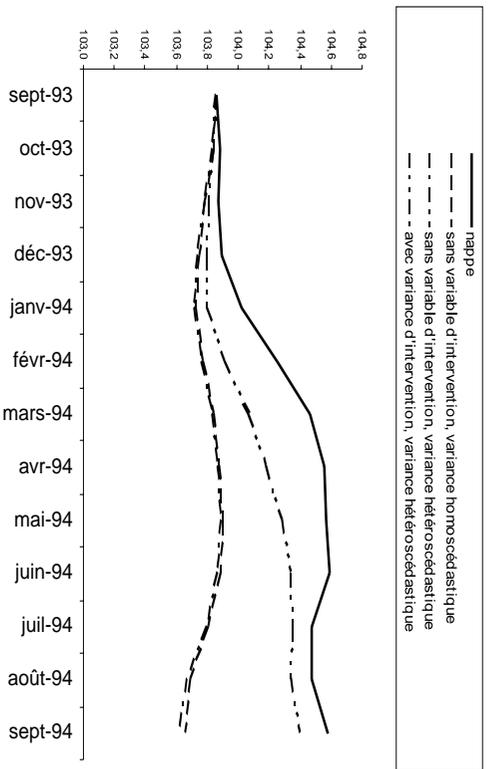


Figure 3.
Prévisions du niveau
de la nappe pour la
période 1995-1996
(en mètres)

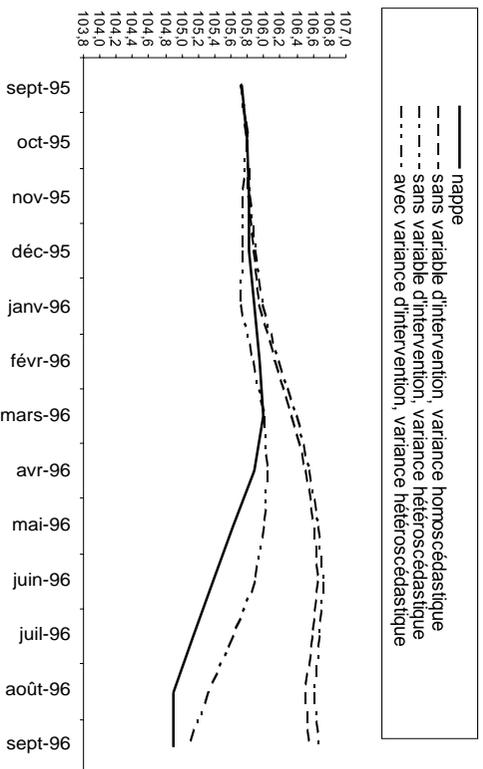
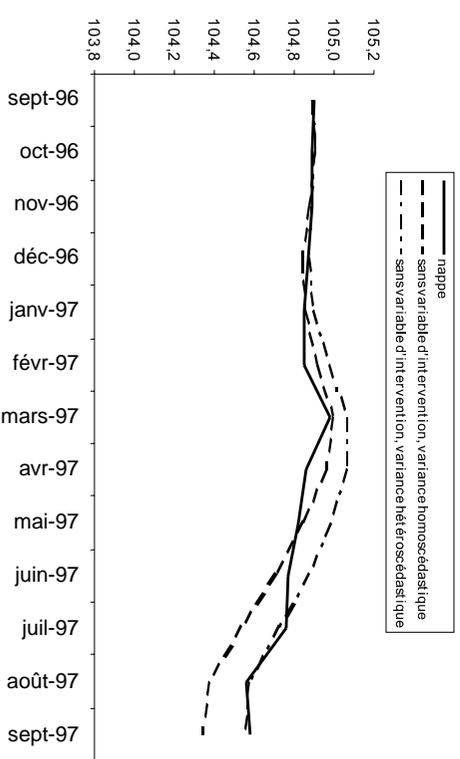


Figure 4.
Prévisions du niveau
de la nappe pour la
période 1996-1997
(en mètres)



CONCLUSION

L'objectif de ce travail était de réaliser des prévisions du niveau de la nappe de Beauce. A court terme, les prévisions obtenues sont satisfaisantes et, de plus, ne nécessitent pas de conjectures sur les paramètres climatiques. Malgré la durée relativement courte de la période de prévision, ce travail permet au décideur public de mettre en place une politique de court terme. Il peut, par exemple à l'aide de quotas, faire en sorte que le niveau de la nappe ne soit pas inférieur au seuil critique qu'il s'est fixé. De plus, grâce à une assez bonne connaissance de l'évolution probable du niveau de la nappe du mois d'octobre de l'année considérée jusqu'au mois de septembre de l'année suivante, les agriculteurs peuvent dès cette date mettre en place une politique d'arrosage des cultures adaptée aux disponibilités anticipées en eau et/ou aux quotas.

Bien que les prévisions de court terme soient satisfaisantes et utiles du point de vue de la politique agricole, de bonnes prévisions à moyen et long terme (deux ou trois années) permettraient, d'une part, au décideur public de faire des arbitrages intertemporels, en annonçant par exemple des quotas sur trois années, et de se servir alors de la nappe comme d'un réservoir, par exemple en diminuant les consommations présentes pour augmenter les consommations futures, et, d'autre part, aux agriculteurs d'adapter leurs assolements en conséquence.

Or, la limite principale de cette étude est l'absence de prévisions de plus long terme (sur deux ou trois années) satisfaisantes. Ce résultat pourrait s'expliquer par le fait que le processus générateur des niveaux de la nappe n'est pas, comme le laisse supposer la modélisation considérée, un processus à mémoire courte mais plutôt à mémoire longue (Hurst, Black et Simaika, 1965). L'effet de Hurst (qui caractérise les processus à mémoire longue) semble être souvent observé lorsque l'on étudie des séries géophysiques et, plus particulièrement, des séries hydrologiques. Il faut toutefois noter que cet effet est, pour certaines séries utilisées par Hurst, contesté (Mesa et Poveda, 1993).

Une autre modélisation possible serait par exemple de considérer des modèles non linéaires avec effet de seuil. Ce travail d'amélioration des prévisions de long terme sera réalisé dans une recherche future, les enjeux économiques de ce type de travail étant, dans le contexte actuel, évidents.

BIBLIOGRAPHIE

- AMIGUES (J.-P.), BONNIEUX (F.), LE GOFFE (P.), POINT (P.), 1995 — *Valorisation des usages de l'eau*, Paris, INRA-Economica.
- BEAULIEU (J.-J.), MIRON (J. A.), 1993 — Seasonal unit roots in aggregate US data, *Journal of Econometrics*, 55, pp. 305-328.
- BONNAL (L.), FAVARD (P.), 1998 — Préviation du niveau de la nappe phréatique de Beauce, Rapport technique ERNA n° 98.03.003, Université de Toulouse I.
- BONNAL (L.), FAVARD (P.), 2000 — Préviation du niveau de la nappe phréatique de Beauce à court, moyen et long terme, Rapport technique LEERNA n° 00.01.005, Université de Toulouse I.
- BOX (G.E.P.), JENKINS (G.), 1976 — *Time series analysis forecasting and control*, San Diego, Holden Day.
- DE MARSILY (G.), 1995, — *L'eau*, Paris, Dominos.
- DICKEY (D. A.) HASZA (D. P.), FULLER (W. A.), 1984 — Testing for unit roots in seasonal time series, *Journal of the American Statistical Association*, 79, pp. 355-367.
- DIEBOLD (F. X.), MARIANO (R. S.), 1995 — Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, pp. 253-263.
- DIEGO (V.), FOURCADE (F.), 1997 — Irrigation et gestion de la nappe de Beauce, mémoire de DEA, Université de Toulouse I.
- FAVARD (P.), FONCEL (J.), 1998 — Estimation de la demande en eau d'irrigation: le cas de la Beauce, Rapport technique ERNA n° 98.02.002, Université de Toulouse I.
- FRANSES (P. H.), 1998 — *Times series models for business and economic forecasting*, Cambridge, Cambridge University Press.
- HASZA (D. P.), FULLER (W. A.), 1982 — Testing for nonstationary parameter specifications in seasonal time series models, *Annals of Statistics*, 10, pp. 1209-1216.
- HYLLEBERG (S.), ENGLE (R. F.), GRANGER (C. W. J.), YOO (B. S.), 1990 — Seasonal integration and cointegration, *Journal of Econometrics*, 44, pp. 215-238.

- HURST (H. E.), BLACK (R. P.), SIMAIKA (V. M.), 1965 — *Long-term storage: an experimental study*, London, Constable.
- MESA (O. J.), POVEDA (G.), 1993 — The Hurst effect: the scale of fluctuation approach, *Water Resources Research*, 29(12), pp. 3995-4002.
- OXBORN (D. R.), CHUI (A. P. L.), SMITH (J. P.), BIRCHENHALL (C. R.), 1988 — Seasonality and the order of integration for consumption, *Oxford Bulletin of Economics and Statistics*, 50, pp. 361-377.