

Une étude comparative de logiciels de prévision automatique de séries chronologiques

Valentina Stan § Gilbert Saporta

Conservatoire National des Arts et Métiers

292 Rue Saint Martin

F 75141 Paris Cedex 03

valentina_titu@yahoo.fr saporta@cnam.fr

Résumé : Cette étude a pour objectif de comparer 5 logiciels de prévision automatique de séries chronologiques, parmi les plus utilisés. On analyse leurs performances sur 50 séries, selon 6 critères à l'aide de plusieurs analyses en composantes principales.

Mots clés: prévision, logiciels de prévision automatique, séries chronologiques

Abstract: The objective of this study is to compare 5 software, amongst the most used, which automatically estimate time series. For this software are analyzed the performances by using 50 time series, according to 6 criteria, with several principal components analyses.

Keywords: forecasting, software for automatically forecasting, time series.

Introduction

La prévision de séries chronologiques demande en général une expertise importante de la part de l'utilisateur. Depuis plusieurs années on assiste à la diffusion de logiciels de prévision entièrement automatiques. Si il existe de nombreuses comparaisons de méthodes, il n'y avait pas à notre connaissance de comparaisons de ces outils logiciels. Nous présentons ici une comparaison systématique de 5 logiciels très répandus.

1. Logiciels testés

1.1 Forecast Pro 4.2 [2]

Le logiciel propose 15 méthodes, les principales catégories sont : lissage exponentiel, Box–Jenkins, moyenne mobile, modèles pour données discrètes ou intermittentes, ajustement de courbes, régression dynamique, Census X11, multi–niveaux. Le système expert du logiciel ajuste automatiquement les paramètres de chacun des modèles en compétition. Le critère utilisé par défaut pour choisir la meilleure méthode de prévision est MAD.

1.2 KTS – 304 [3]

Kxen Time Series (KTS) est un des composants analytiques de la suite KXEN Analytic Framework. KTS construit ses modèles en extrayant les quatre composantes d'une série chronologique : la tendance, les périodicités (ou les parties récurrentes dans le signal), les fluctuations qui représentent les perturbations affectant une série chronologique et qui dépendent de son état passé (c'est le "phénomène de mémoire"), le résidu.

La prévision donnée par KTS est la somme des trois premières composantes, et utilise des techniques proches des réseaux de neurones.

Pour choisir le meilleur modèle, les critères utilisés par KTS sont :

- la confiance sur l'horizon le plus grand possible par rapport à l'horizon de prévision choisi;
- l'erreur minimum cumulée sur l'horizon maximum de confiance trouvé précédemment.

1.3 SAS 8.2 [4]

SAS propose plus de 15 méthodes, dont : tendance linéaire, tendance linéaire avec erreurs autorégressives, tendance linéaire avec termes saisonniers, composante saisonnière avec variables indicatrices, les méthodes de lissage, marche aléatoire avec dérive, méthodologie Box–Jenkins, la moyenne. Pour choisir la meilleure méthode de prévision, le critère utilisé par défaut est le RMSE, mais l'utilisateur a la possibilité de le changer.

1.4 SPSS Decision Time 1.1 [5]

SPSS Decision Time 1.1 est le premier logiciel de prévision réalisé par SPSS Inc. A l'aide de «Forecast Wizard» l'utilisateur est guidé tout au long du processus de prévision. 10 méthodes sont intégrées, depuis les classiques lissages exponentiels jusqu'aux modèles ARIMA. La meilleure méthode de prévision est choisie par défaut à l'aide du critère MAPE, mais l'utilisateur peut choisir lui-même d'autres critères.

1.5 Statgraphics 5.1 [6]

Le logiciel propose 14 méthodes : marche aléatoire, moyenne, tendance linéaire, tendance quadratique, tendance exponentielle, courbe en S, moyenne mobile, lissage exponentiel simple, lissage exponentiel de Brown, lissage exponentiel de Holt, lissage exponentiel quadratique, lissage exponentiel de Winter, ARMA, SARMA. Le critère d'information d'Akaike est utilisé par défaut pour déterminer la meilleure méthode, mais il peut être changé par utilisateur.

Les principales différences entre les logiciels:

- Statgraphics est le seul qui ne réalise pas automatiquement la transformation de la série, c'est l'utilisateur qui doit l'indiquer. Les autres logiciels cherchent automatiquement la transformation à appliquer à la série parmi les transformations usuelles (logarithme, racine carrée ...).
- Pour Forecast Pro et Statgraphics l'utilisateur doit spécifier si la série est ou non saisonnière, les autres déterminent automatiquement la saisonnalité.
- Parmi les 5 logiciels, Statgraphics est le seul qui présente une comparaison des modèles.
- SPSS Decision Time et KTS offrent la possibilité d'introduire dans le processus de prévision des événements qui ont «affecté» la série et qui sont considérés importants par l'utilisateur.

2. Séries utilisées

Pour réaliser cette étude nous avons utilisé 50 séries chronologiques de plusieurs types :

Périodicité des données	Type de séries						Total
	Macro	Industrie	Finance	Démographie	Environnement	Autres	
Annuelle	1	7	0	3	2	0	13
Trimestrielle	2	4	1	0	0	0	7
Mensuelle	4	11	2	2	1	0	20
Journalière	0	0	3	1	1	1	6
Autres	0	0	0	0	0	4	4
Total	7	22	6	6	4	5	50

La principale source pour sélectionner ces séries est la M3-Competition, l'étude la plus connue, coordonnée par Makridakis et Hibon (2000) [1] qui compare des méthodes de prévision pour 3003 séries.

Quelques exemples :

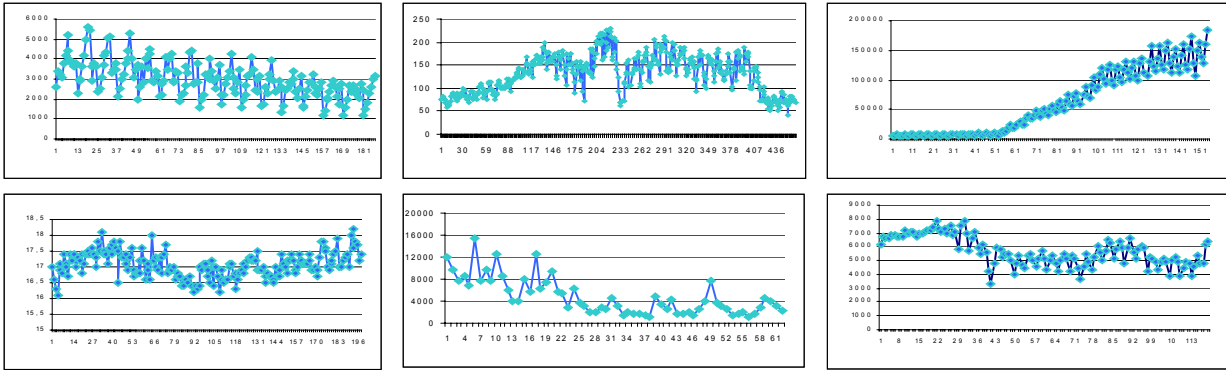


Figure n°1 : Des exemples de séries utilisées

Pour comparer les résultats obtenus en utilisant chaque logiciel on a procédé de la manière suivante:

- on supprime les dernières observations :
 - 6 pour les séries annuelles longues et 2 pour les séries très courtes;
 - 8 pour les séries trimestrielles;
 - 12 pour les séries mensuelles et 2 pour les séries très courtes;
 - 30 pour les séries journalières;
 - 18 pour les autres types des séries, sauf la série 50 pour laquelle 10 observations sont supprimées.
- chaque logiciel prévoit pour chaque série les observations supprimées;
- les valeurs prévues sont comparées avec les valeurs réelles.

3. Critères et normalisations

En notant e_i les erreurs de prévision, les critères suivants ont été utilisés pour évaluer les performances :

1) l'écart type : $SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2}$; 2) l'écart absolu moyen : $MAD = \frac{1}{n} \sum_{i=1}^n |e_i - \bar{e}|$;

3) l'erreur absolue moyenne : $MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$; 4) l'écart absolu moyen en pourcentage : $MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$;

5) l'erreur quadratique moyenne : $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$; 6) l'erreur maximale = $\max |e_i|$.

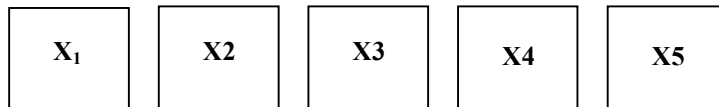
Sauf MAPE, dont on verra plus loin qu'il se comporte différemment des autres, ces critères sont sensibles à l'échelle de mesure et au niveau des séries et ne peuvent donc être comparés d'une série à l'autre. Deux normalisations ont été étudiées : en fonction de la moyenne de la série et en fonction de la moyenne des valeurs du critère obtenues par les différents logiciels pour la même série. Ensuite, en utilisant les valeurs normalisées on a considéré pour chaque normalisation deux types d'analyses.

4. Les deux types d'analyse

Pour chaque logiciel on dispose d'un tableau de dimension : $(43 \times 6)^1$ ou (50×6) , d'où 5 ou 4 tableaux, selon que l'on intègre KTS ou non dans l'étude. Pour analyser simultanément ces tableaux on peut les assembler de deux manières différentes :

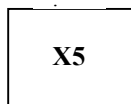
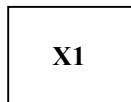
¹ KTS ne réalise pas de prévisions pour les séries très courtes et pour la série 38 qui a beaucoup de valeurs nulles.

4.1 Juxtaposition horizontale :



Objectif : étudier dans le cadre de chaque logiciel les corrélations entre critères.

4.2 Empilement vertical :



Objectif : étudier les corrélations entre les 6 critères. Il y a donc 6 colonnes et 215 individus (43 séries x 5 logiciels). A chaque série correspond donc 5 points. Les logiciels sont alors les modalités d'une variable nominale illustrative supplémentaire ce qui permettra de réaliser un classement pour les logiciels.

5. Résultats²

5.1 Juxtaposition horizontale

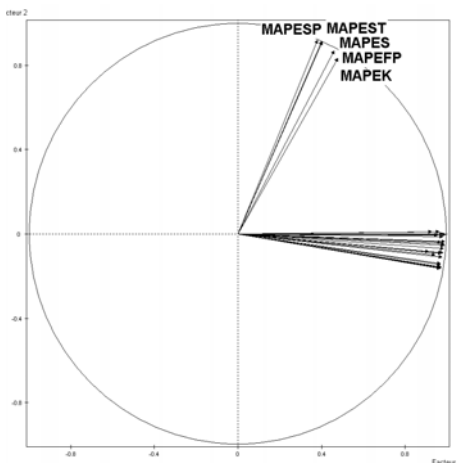


Figure n°2 : Cercle de corrélation

Les deux premiers axes factoriels extraient 95,41% de l'inertie du nuage de points.

Le cercle de corrélation montre que le critère MAPE a un comportement indépendant des autres et que tous les autres critères sont très semblables entre eux quel que soit le logiciel. Si MAPE est mis en variable supplémentaire le cercle de corrélation montre que tous les autres critères sont très semblables entre eux quel que soit le logiciel

Dans le plan principal on remarque que les séries sont près du centre du graphique : les valeurs des critères sont globalement proches des valeurs moyennes.

On voit que les séries 4 et 45 ont un comportement complètement différent des autres et contribuent fortement à l'inertie portée par chaque axe. La série qui contribue le plus au premier axe est S4 (56,8% de l'inertie de cet axe est due à cette série); S45 contribue le plus au deuxième axe (avec 52,3%). La série 4 («La valeur de marché du stock total de vente sur les échanges enregistrés entre 1947-1987») est une série annuelle sans saisonnalité et qui a une évolution constante jusqu'à une croissance brusque. La série 45 («Le PIB réel aux Etats Unis - exprimé en \$ 1982 - entre le premier trimestre 1947 et le troisième trimestre 1991») est une série trimestrielle, sans saisonnalité mais qui présente des grandes fluctuations.

Ces deux séries pourraient avoir une grande influence pour la détermination de ces axes, cependant si on considère les deux séries comme individus illustratifs, le cercle de corrélation reste le même.

² Les analyses ont été effectuées avec le logiciel SPAD 5.5.

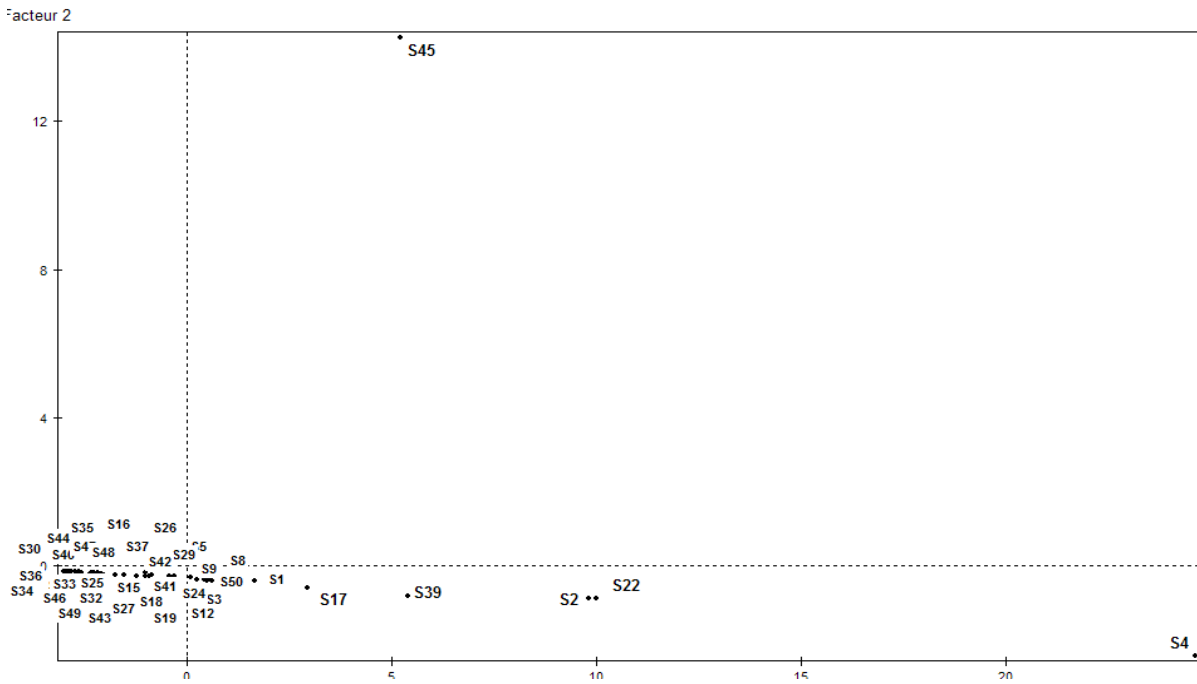


Figure n°3 : Plan principal (séries)

5.2 Empilement vertical

On trouve à nouveau que MAPE est différent des autres critères. La figure 5 montre chaque logiciel comme centre de gravité de 43 séries.

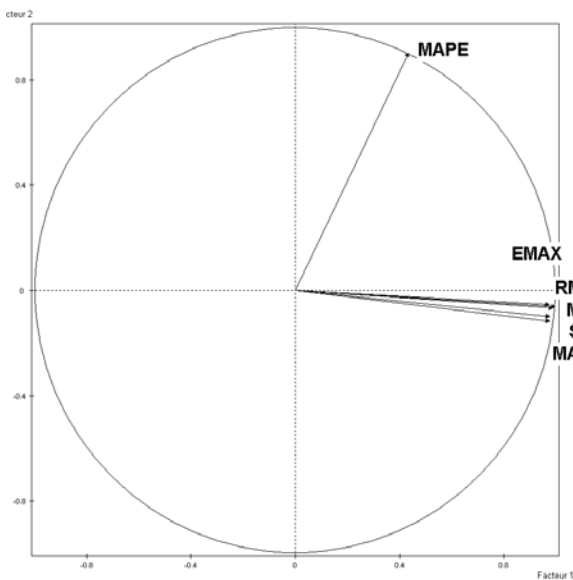


Figure n°4 : Cercle de corrélation

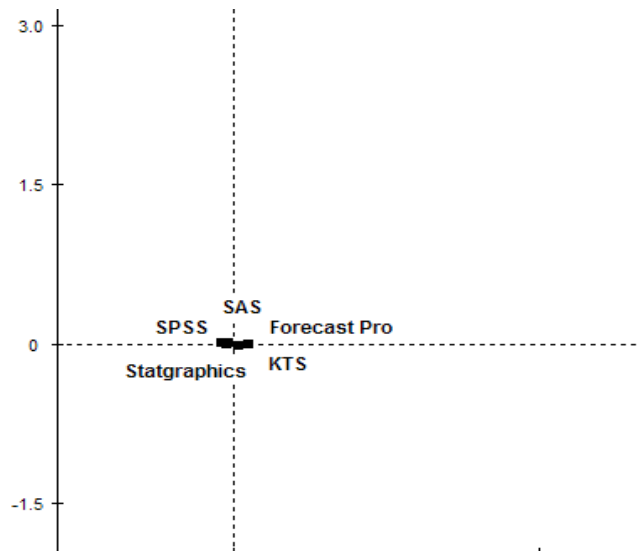


Figure n°5 : Graphique des logiciels analysés

Tous les logiciels se confondent pratiquement avec le centre du graphique; donc, on ne peut pas faire une distinction entre eux, concernant les critères de performance utilisés. Cette affirmation est confirmée par le calcul des zones de confiance qui se recouvrent complètement: on ne peut donc conclure à la supériorité d'un logiciel sur les autres.

Remarque :

- Lorsqu'on utilise les 50 séries et 4 logiciels (sans KTS), la conclusion est la même : les 4 logiciels se «confondent» avec le centre du graphique, en conséquence on ne peut pas les départager selon les critères considérés.

Conclusions

- La principale conclusion (obtenue après avoir réalisé 8 ACP pour les 2 normalisations) est qu'on ne peut pas réaliser un classement des 5 logiciels testés, leurs performances sont pratiquement identiques du point de vue des critères considérés, quelle que soit la normalisation utilisée, ce qui est rassurant pour l'utilisateur.
- Il y a des situations dans lesquelles les modèles les plus complexes ne donnent pas de meilleurs résultats que des modèles simples : par exemple, pour la série 37, les méthodes de lissage exponentiel utilisées par SPSS et SAS donnent des résultats meilleurs que les modèles ARIMA utilisés par Forecast Pro et Statgraphics.
- On ne peut bien sûr pas dire si un logiciel choisit le « vrai » modèle. Il aurait fallu pour cela procéder à des simulations, mais nous avons choisi dans ce travail de n'utiliser que des séries réelles, ce qui fait qu'en toute rigueur nos conclusions ne sont valables que pour les 50 séries utilisées et pour les critères de performances considérés.

Bibliographie

- [1] Makridakis, S. et Hibon, M. (2000) *The M3-Competition: results, conclusion and implication*, International Journal of Forecasting, vol. 16
- [2] <http://www.forecastpro.com>
- [3] <http://www.kxen.com>
- [4] <http://www.sas.com>
- [5] www.spss.com.fr
- [6] www.statgraphics.com