méthodes probabilistes de prévision

par Mme Marie-Claude Viano, de l'Université de Lille 1

Remerciements

Je remercie Georges Oppenheim qui a fourni les graphiques qui illustrent le texte.

Marie-Claude Viano

Je vais parler de prévision, et plus exactement des méthodes de prévision qui utilisent les probabilités.

Commençons par un exemple, celui de la consommation électrique.

Prévoir la consommation d'électricité est très important pour la bonne raison que l'électricité ne se stocke pas. Il ne faut arriver à l'instant *t* ni en ayant produit insuffisamment parce qu'en ce cas on risque des coupures de courant, ni en ayant trop produit puisqu'on ne peut stocker. Il est donc primordial d'arriver à prévoir le mieux possible les besoins de demain, en s'appuyant sur ce que l'on a consommé aujourd'hui, hier, avant-hier, ou même dans les six derniers mois.

La figure 1 représente cinq jours de la consommation d'électricité dans toute la France. Les mesures ont été faites toutes les demi-heures. En réalité elles ont été faites pendant cinq semaines entre juin et juillet 1991. Ce sont les cinq derniers jours de cette période de cinq semaines qui ont été représentés dans la figure 1.

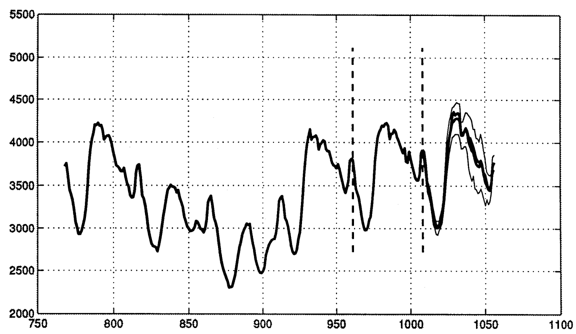
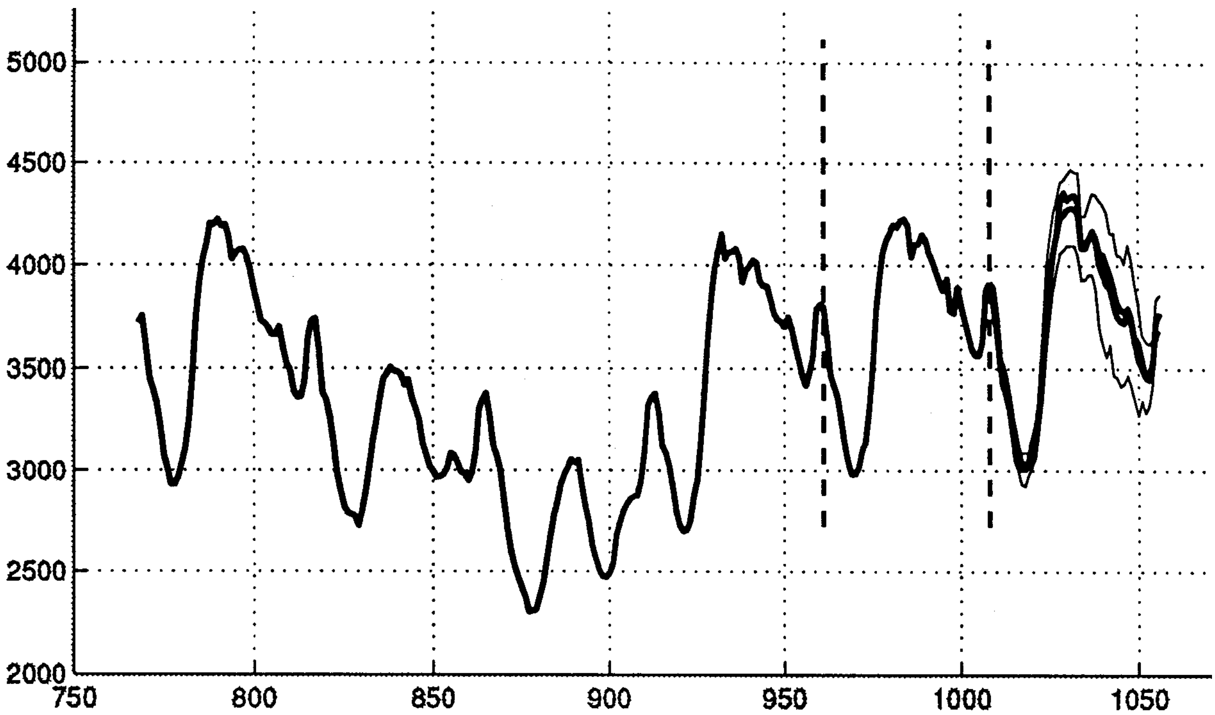


Figure 1.— Données et prévisions. Consommation d'électricité. Cinq jours de consommation, demi-heure par demi-heure. On prévoit une journée entière.

Cette succession de données a servi à tester un certain nombre de méthodes de prévision. L'extrémité droite de la courbe en trait plein représente la consommation effective au cours du dernier mardi et du début du dernier mercredi de la période observée.

En trait plus léger, en haut et en bas, on a tracé un couloir de prévision qui signifie la



(Figure 1.)

chose suivante : on prévoit qu'à l'instant *t* la consommation d'électricité va être située entre la valeur basse et la valeur haute.

C'est ainsi que se présentent souvent les résultats des prévisions. En comparant à la courbe en trait plein vous pouvez comprendre que cette prévision est de bonne qualité. Je reviendrai sur cet exemple en fin d'exposé.

Quelques idées, à propos d'exemples simples

Je vais commencer par dégager les principes de base qui sont derrière la plupart des méthodes probabilistes de prévision.

Le pile ou face

On a joué à pile ou face cinquante fois successives et on a reporté le résultat sur la figure 2 en représentant pile par 1 et face par 2. A la question :

Comment prévoir le résultat
du cinquante-et-unième coup ?

tout le monde répondra, avec raison, qu'aucune méthode ne peut donner une probabilité d'erreur inférieure à 1/2, et que le fait de connaître le résultat des 50 coups précédents n'est d'aucune aide!

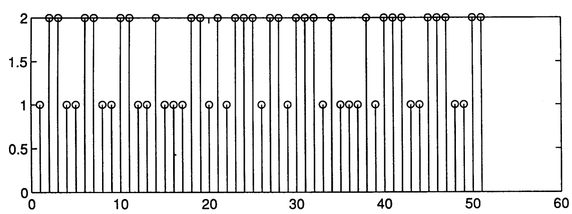


Figure 2.— Cinquante coups de pile ou face.

Le pile ou face avec mémoire

Imaginez maintenant une pièce de monnaie “truquée”, qui ait deux chances sur trois de reproduire au coup $n + 1$ ce qu’elle a donné au coup n.

Evidemment une telle pièce n’existe pas mais un jeu fictif est très facile à réaliser. Il suffit d’une pièce ordinaire et d’un dé à six faces.

On amorce le jeu en lançant la pièce (admettons qu’elle tombe sur pile). Puis on lance le dé. S’il donne un chiffre entre 1 et 4 on décide que le deuxième résultat du jeu fictif va être le même que celui du lancer de la pièce (ce sera donc pile, dans l’exemple choisi) et qu’il sera différent dans le cas contraire. Puis on lance à nouveau le dé. Si il donne un chiffre entre 1 et 4 on décide que le troisième résultat du jeu fictif est le même que le deuxième et qu’il est différent dans le cas contraire. Et on continue.

Un calcul de probabilités facile montre que pour ce jeu fictif la probabilité d’obtenir pile est toujours 1/2 et que la probabilité que les résultats de deux coups successifs soient les mêmes est bien 2/3.

On a réalisé cinquante coups de ce jeu fictif et on a reporté le résultat sur la figure 3.

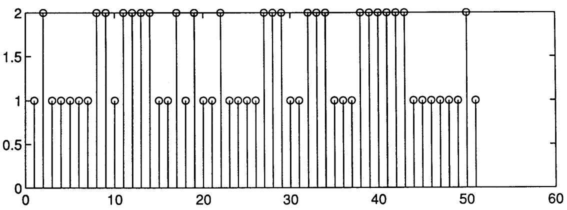


Figure 3.— Cinquante coups de pile ou face.

Comment prévoir le résultat du cinquante-et-unième coup? On remarque que le dernier coup a donné un pile. Comme on sait que la pièce a deux chances sur trois de se répéter il paraît raisonnable de prévoir pile.

La règle adoptée consiste à prendre pour prévision du coup $n + 1$ le résultat du coup n. La probabilité d’erreur est 1/3.

Cette règle est meilleure, par exemple, que celle qui consisterait à prévoir systématiquement pile (pour celle-ci la probabilité d’erreur est 1/2). Si on avait utilisé une vraie pièce, la probabilité d’erreur aurait été 1/2 pour chacune des deux règles précédentes.

Le pile ou face avec mémoire imparfaitement connue

Compliquons un peu les choses en reprenant l’exemple de la pièce truquée. Vous disposez des résultats des 50 coups, vous savez que la pièce a une probabilité p de se répéter. Mais vous ignorez la valeur de p. C’est ce qui se passe dans la réalité. On sait que le phénomène a une certaine mémoire, on s’est fait une idée de la forme de cette mémoire (ici on sait que la pièce se souvient seulement du résultat immédiatement antérieur), mais on ignore son intensité.

La première chose à faire est d’essayer de reconstituer l’information manquante, donc, dans le cas qui nous occupe, d’obtenir ce qu’on appelle une estimation de p.

Regardons la figure 3. Sur 50 coups, donc sur 49 passages d’un coup au suivant, il y a 30 répétitions et 19 changements. La proportion des répétitions est 30/49, donc un peu supérieure à 0,6.

Une méthode d’estimation de p pourrait consister précisément à choisir comme valeur de p la proportion de répétitions avant le dernier coup joué (elle donnerait ici comme estimation la valeur 30/49). En fait c’est une très bonne méthode, mais les arguments probabilistes qui justifient ce jugement sont en dehors des limites de cet exposé. Disons seulement qu’une des qualités de cette méthode est que plus le nombre de coups joués est grand plus la probabilité qu’elle fournisse une valeur très différente de 1/3 est petite. Faute de connaître la vraie valeur de p on va lui attribuer la valeur 30/49 et agir comme dans le cas d’une pièce qui aurait la probabilité 30/49 de se répéter. On va donc prévoir pile pour le coup cinquante et un.

Les idées de base

Quatre idées se dégagent des exemples précédents :

Première idée :

> Pas de prévision intéressante sans une certaine dépendance entre les données (avec la pièce de monnaie ordinaire on ne peut rien faire).

Deuxième idée :

> Tout travail de prévision s'appuie sur des hypothèses sur la forme de cette dépendance (la pièce truquée se souvient du coup précédent seulement).

Troisième idée :

> Cette dépendance ne change pas au cours du temps (le paramètre p reste constant).

Quatrième idée :

> Préalablement à la prévision il faut se livrer au travail de statisticien qui consiste à utiliser les données dont on dispose pour évaluer les paramètres qui caractérisent la dépendance (estimer p, dans l'exemple de la pièce avec mémoire imparfaitement connue).

Quelques modèles d'usage courant

Les séries d'observations ne sont que très rarement des successions de 1 et de 2 et il a bien fallu imaginer des dépendances un peu plus complexes que celle de notre pièce de monnaie fictive.

La situation habituelle est la suivante : on a observé des grandeurs, par exemple des consommations d'électricité, depuis l'instant 1 jusqu'à l'instant n.

Appelons-les $x_1, x_2, \ldots, x_n$. On cherche à prévoir x_{n+1}, et dans ce but on commence par étudier les dépendances entre les grandeurs observées.

Le plus simple

Un lien de dépendance très simple pourrait être $x_2 = a\, x_1, \ldots, x_n = a\, x_{n-1}$ (bref $x_{j+1} = \mathrm{a}\, x_j$ pour les j entre 1 et n–1). Il aurait l'avantage de conduire sans hésiter à prévoir x_{n+1} par $a\, x_n$. Mais bien évidemment il est parfaitement irréaliste ! Aussi on le modifie un peu en $x_{j+1} = a\, x_j + \varepsilon_j$.

Ici, la succession des ε_j est ce qu'on appelle un bruit blanc. Vous pouvez facilement en fabriquer un en appuyant n–1 fois successivement sur la touche "`random`" de votre calculatrice. Disons brièvement que c'est une succession de nombres au hasard sans lien les uns avec les autres tous situés entre – 1 et 1 (avec votre calculatrice vous trouvez des nombres entre 0 et 1, multipliez-les par 2 et enlevez 1 au résultat …).

Supposons un moment que les données évoluent vraiment selon la relation

$$x_{j+1} = a\, x_j + \varepsilon_j \,.$$

On a deux questions à résoudre :

si a est connu, comment prévoir x_{n+1}, et :

si a est inconnu, comment utiliser les données pour en donner une bonne valeur approchée ?

La réponse à la première question est simple.

Comme les ε_j n'ont aucune mémoire, toute la dépendance entre les données est contenue dans le paramètre a. De ce fait la meilleure règle consiste à prévoir x_{n+1} par $a\ x_n$ (préciser en quoi cette règle est meilleure que les autres sort du cadre de cet exposé).

On a représenté dans les figures 4 et 6 deux séries de cent données (les points sont reliés par des segments de droite par souci de lisibilité) qui évoluent selon les relations $x_{j+1} = a\ x_j + \varepsilon_j$ avec $a = 0{,}8$ pour la figure 4 et $a = -0{,}8$ pour la figure 6. On va prévoir x_{101} par $0{,}8\ x_{100}$ dans le premier cas et par $-0{,}8\ x_{100}$ dans le second.

La réponse à la deuxième question est plus compliquée.

Dessinons dans le plan le nuage des points de coordonnées (x_j, x_{j+1}). Des données liées par les relations $x_{j+1} = a\ x_j + \varepsilon_j$ fourniront un nuage allongé le long de la droite d'équation $y = a\ x$. La figure 5 est le nuage construit à partir des données de la figure 4, la figure 7 à partir de la figure 6.

La « pente » du nuage donne donc une indication sur la valeur de a. Pour être plus précis disons qu'on sait trouver la droite qui approche le mieux le nuage. Vous la connaissez peut être déjà, on l'appelle « droite des moindres carrés ».

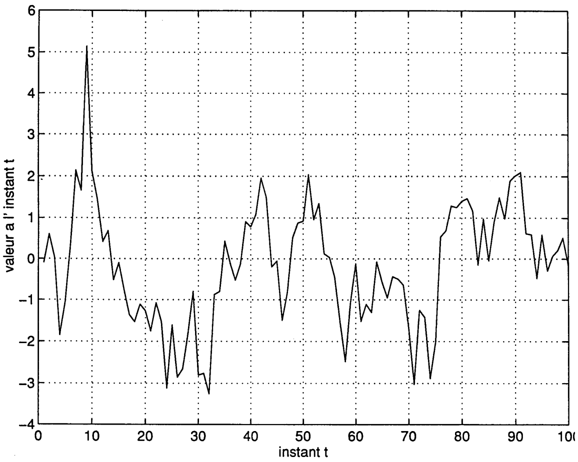


Figure 4.— $x_{j+1} = 0{,}8\ x_j + \varepsilon_j$.

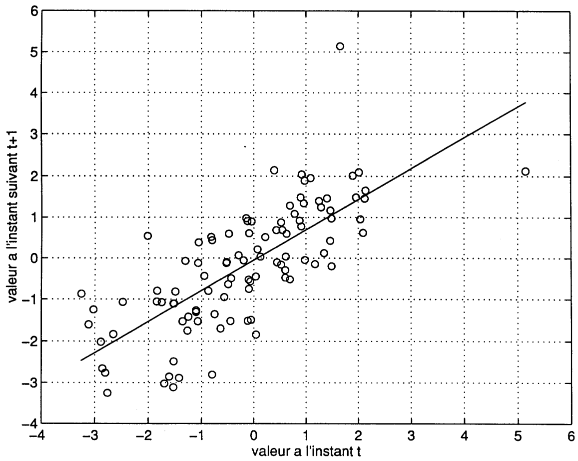


Figure 5.— Droite des moindres carrés.

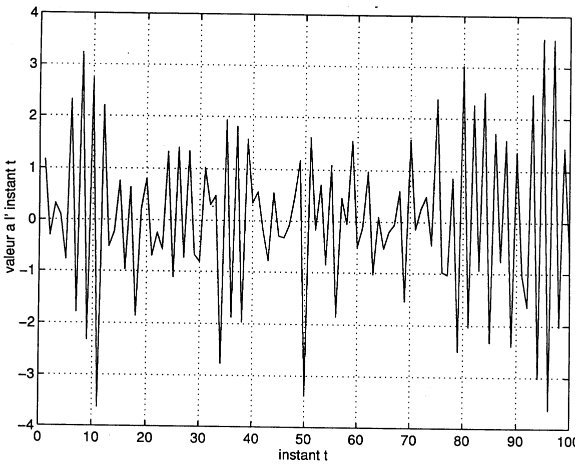


Figure 6.— $x_{j+1} = -0.8\ x_j + \varepsilon_j$.

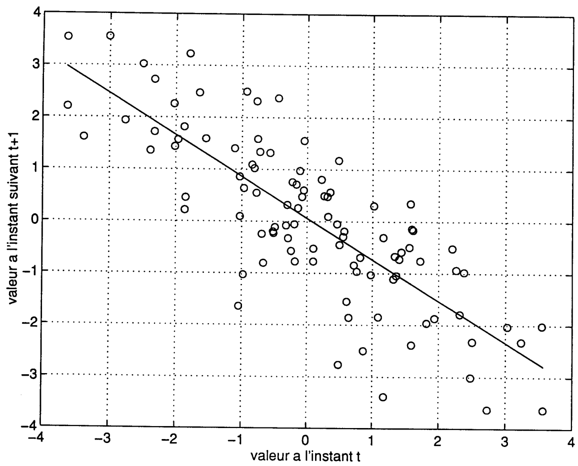


Figure 7.— Droite des moindres carrés.

La droite des moindres carrés a pour pente

$$\frac{\sum_{j=1}^{n-1} x_j\, x_{j+1}}{\sum_{j=1}^{n} x_j^2}$$

Cette droite est représentée sur les figures 5 et 7. Elle a pour pente 0,7422 dans le premier cas et – 0,8039 dans le second. Cette méthode d'estimation du paramètre a est très bonne. Des considérations probabilistes montrent que plus il y a de données, moins elle a de chances de fournir une valeur éloignée de la valeur de a qui a servi à construire les données.

Revenons enfin au problème de la prévision dans les figures 4 et 6.

Si on sait que les données sont liées par les relations $x_{j+1} = a\, x_j + \varepsilon_j$ mais si on ignore la valeur de a, il est tout à fait recommandé de lui attribuer la valeur

$$\frac{\sum_{j=1}^{n-1} x_j\, x_{j+1}}{\sum_{j=1}^{n} x_j^2}$$

et d'effectuer la prévision comme si ce qu'on a trouvé était la valeur de a. Dans les deux exemples cela conduit à prévoir x_{101} par $0{,}7422/x_{100}$ dans le premier cas et par $-0{,}8039/x_{100}$ dans le second.

Compliquons un peu

A ce stade on devrait se poser la question du bien fondé du modèle $x_{j+1} = a\ x_j + \varepsilon_j$. Comment voit-on que ce modèle de dépendance est adapté aux données sur lesquelles on travaille ? Que faire s'il est inadapté ?

En réalité on dispose de tout un stock de modèles du même type, par exemple

$$x_{j+1} = a_1\, x_j + a_2\, x_{j-1} + \varepsilon_j$$

ou

$$x_{j+1} = a_1\, x_j + a_2\, x_{j-1} + \varepsilon_j + b\, \varepsilon_{j-1}.$$

On les appelle « modèles linéaires ». On dispose aussi de moyens informatiques puissants qui permettent, pour chaque jeu de données, de choisir dans le stock des modèles linéaires le modèle le mieux adapté, et enfin de faire le calcul de la prévision … C'est le développement des moyens informatiques qui a permis l'essor des techniques de prévision.

Un exemple

Pour terminer voici les données, en nombre de voyageurs, du transport aérien depuis janvier 1949 jusqu'à décembre 1960. Il y a 132 données (une par mois). Elles sont représentées sur la figure 8.

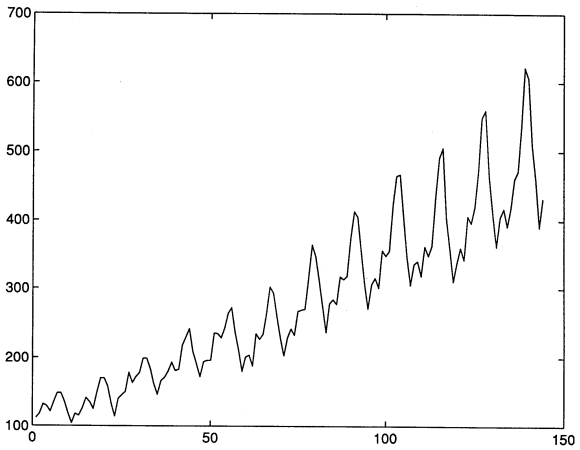


Figure 8.— Trafic aérien de Janvier 1949 à Décembre 1960. En abscisse : le temps en mois. En ordonnée le trafic.

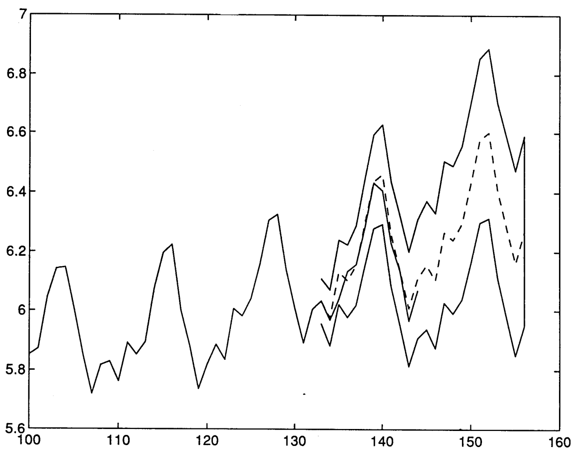


Figure 9.— Les prévisions et l'intervalle de prévision.

Sur la figure 9 on a matérialisé :

1- la prévision pour les 24 mois suivants assortie du couloir de prévision dont il a été dit un mot dans l'introduction : ce sont les trois courbes en trait léger,

2- les vraie données du 132-ième mois : c'est la courbe en trait gras.

On voit que la prévision et la réalité sont très proches.

Comment ont été obtenues ces prévisions ?

Manifestement ces données n'ont pas du tout l'allure de celles des figures 5 et 6. On y voit une tendance moyenne à l'augmentation et une succession de motifs de même allure qui vont en s'allongeant vers le haut. La tendance traduit l'augmentation globale du trafic aérien et chaque motif correspond au déroulement d'une année, avec ses périodes de creux et de pointe.

Le premier soin est de se débarrasser de ces effets qui ne doivent rien au hasard.

On prend le log des données pour gommer un peu les pics. Puis on travaille sur la suite des $\log(x_j) = y_j$.

Pour faire disparaître la tendance à l'augmentation on change les y_j en $y_j - y_{j-1} = z_j$.

Pour faire disparaître les motifs on change les z_j en $z_j - z_{j-12} = t_j$.

Alors, alors seulement, on obtient une série de données pour la quelle on choisit un modèle linéaire, dont on estime les paramètres pour prévoir (enfin!) les valeurs $t_{132}, \dots, t_{156}$.

Puis on n'oublie pas de faire les trois transformations inverses qui conduisent aux prévisions de $x_{132}, \dots, x_{156}$ représentées sur la courbe médiane en trait léger de la figure 9.

Conclusion

Vous venez de faire connaissance, de façon très partielle, avec la méthode de prévision la plus courante : celle qui consiste à chercher un modèle dans le stock des modèles linéaires. Elle figure dans tous les logiciels de prévision et est utilisée quotidiennement dans de nombreux domaines comme la météorologie, l'économétrie et dans de nombreux secteurs de l'industrie. On peut dire que pratiquement tous les problèmes mathématiques qu'elle a pu poser sont résolus et qu'elle n'est plus du domaine de la recherche.

Par contre cette méthode est loin d'être adaptée à toutes les situations pratiques. En gros pour qu'elle donne de bons résultats il faut que les données présentent une certaine régularité.

Ce n'est bien sûr pas toujours le cas et d'autres stocks de modèles sont étudiés, comme par exemple ceux du type

$$x_{j+1} = f(x_j) + \varepsilon_j$$

où f est une fonction a priori inconnue.

Avec ce genre de modèle les choses deviennent du point de vue mathématique beaucoup plus intéressantes, et on arrive à faire des prévisions dans des situations plus difficiles que celles que vous venez de voir. Les prévisions de la consommation d'électricité de la figure 1 ont été obtenues de cette façon.

Il me semble utile de conclure en remarquant que dans certaines situations comme la survenue d'un événement exceptionnel toutes les méthodes de prévision donnent de mauvais résultats.