

LA MSP

► Les situations non standard (4^e partie)

► Par Daniel LAFAYE de MICHEAUX, directeur R&D de GPC System, et Alain PALSKEY.

Dans les trois articles précédents, nous avons successivement abordé les cas de mesures non gaussiennes, de mesures issues d'équipements parallèles et de grandeurs régulées couramment rencontrées dans la pratique et pour lesquelles l'utilisation des cartes de contrôle classiques n'est pas adaptée. Nous avons terminé par les mesures sensibles à un phénomène d'usure ou d'encrassement qui présentent des dérives linéaires inévitables. Nous allons à présent généraliser les méthodes appliquées aux dérives linéaires à des mesures dont le graphe "serpente" (cf. figure 1) ; celles-ci présentent naturellement une moyenne qui n'est pas constante dans le temps.

Traitement habituel des mesures qui "serpentent"

La situation de la figure 1 est très fréquente : elle se présente presque toujours lorsqu'on mesure une température de four, un niveau dans une cuve. Ce graphe est caractéristique d'une mesure transgressant à la fois les trois hypothèses de la situation standard :

1. Constance dans le temps de la cible et de la variabilité de la grandeur mesurée.
2. Indépendance statistique des mesures successives.

3. Distribution gaussienne de la mesure. Une mesure sensible à un phénomène d'usure, évoquée dans l'article précédent⁽¹⁾, est un cas particulier où le "serpent" prend une forme de dents de scie.

Comment surveille-t-on habituellement de telles mesures ? On construit une carte avec des limites de Shewhart calculées à partir de l'écart type évalué sur l'ensemble de l'historique (écart type long terme) : il y aura alerte si le serpent se met à louvoyer un peu trop loin de sa route habituelle (cf. limites orange sur la

figure 1). On applique donc la technique classique alors que le comportement des données ne vérifie pas les hypothèses la justifiant ! Cette pratique contestable ne permet de détecter que les dysfonctionnements les plus manifestes, laissant passer inaperçues au niveau du processus des anomalies lourdes de conséquences pour la qualité du produit.

Si l'on veut une détection plus fine des dysfonctionnements, il convient de perfectionner le modèle de données.

Modèles de mesures qui "serpentent"

Sur la figure 1, l'observation x_t de la mesure à l'instant t (ou pour le numéro t de produit) n'est pas indépendante de la mesure précédente x_{t-1} , puisqu'il est clair que x_t est très généralement plus proche de x_{t-1} que de x_{t-5} ou de x_{t-10} . Prenons le cas d'une mesure de niveau dans un réservoir ; la différence ε_t entre débits entrant et sortant est une valeur aléatoire de moyenne constante nulle, on a :

$$x_t = x_{t-1} + \varepsilon_t \quad (1)$$

Pour supprimer l'effet de serpent, il faut non plus surveiller x_t mais :

$$d_t = x_t - x_{t-1} = \varepsilon_t \quad (2)$$

L'équation (1) peut être généralisée en posant que x_t est fonction linéaire de x_{t-1} : $x_t = a_0 + a_1 \cdot x_{t-1} + \varepsilon_t$ (3)

ou, encore plus général : x_t est fonction linéaire de x_{t-1} et x_{t-2} :

$$x_t = a_0 + a_1 \cdot x_{t-1} + a_2 \cdot x_{t-2} + \varepsilon_t \quad (4)$$

Les relations (3) et (4) définissent des modèles appelés autorégressifs (AR), respectivement d'ordre 1 et 2.

Dans ces équations, ε_t est appelé l'innovation à l'instant t et on peut souvent supposer que ε_t satisfait aux trois hypothèses de la situation standard.

Une anomalie apparaissant à l'instant t ne peut perturber x_t qu'en perturbant l'innovation ε_t . On pourra donc la détecter en surveillant ε_t par une carte de contrôle standard.

Mais pour identifier ε_t , il faudrait disposer des paramètres a_0 , a_1 , a_2 , qui sont inconnus. Nous allons donc devoir

(1) Qualité références, octobre 2003, pp. 71-75.

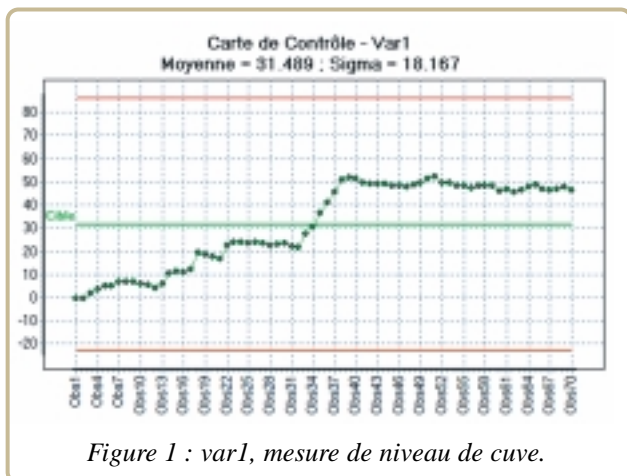


Figure 1 : var1, mesure de niveau de cuve.

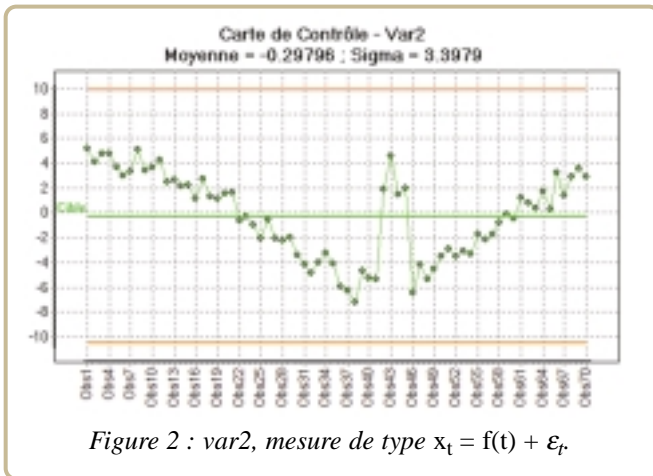
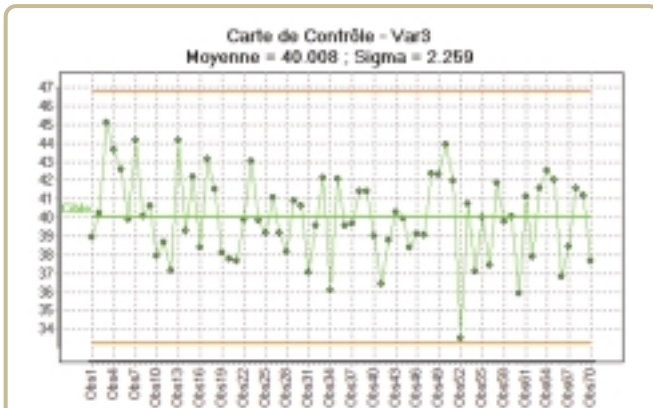


Figure 2 : var2, mesure de type $x_t = f(t) + \varepsilon_t$



Autocorrélations estimées pour Var3

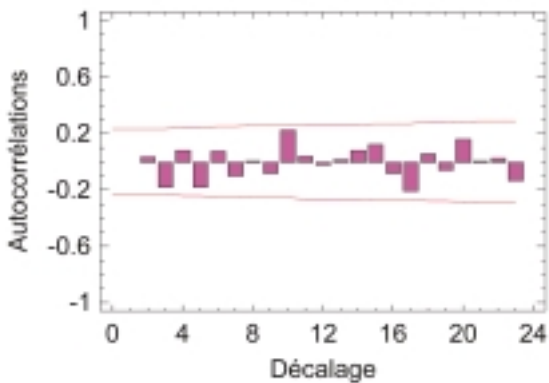


Figure 3 : var3, valeurs brutes et fonction d'autocorrélation.

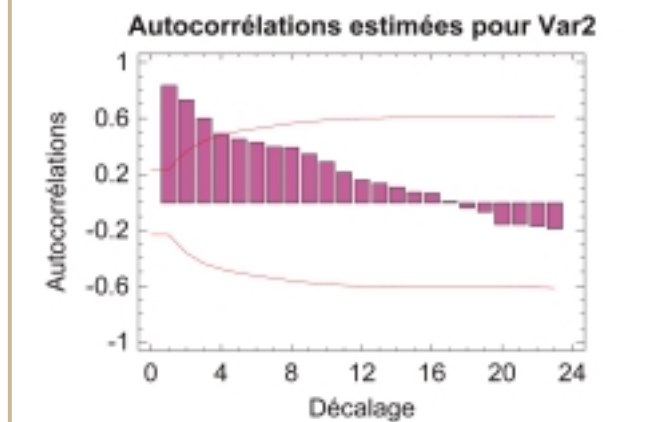
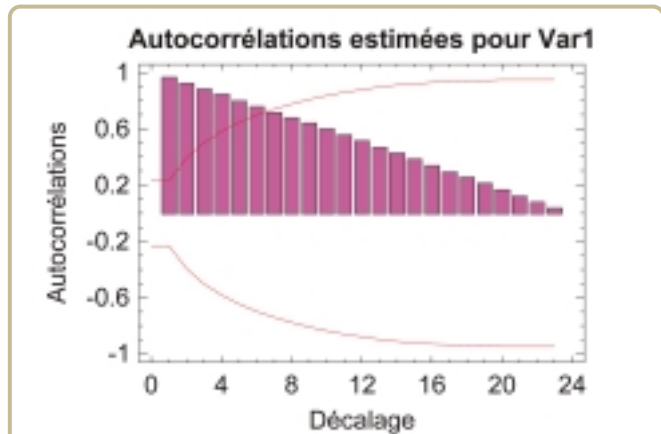


Figure 4 : corrélogrammes de var1 et var2.

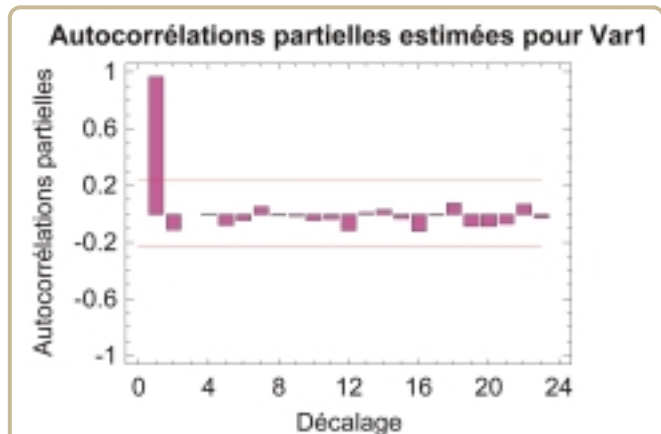


Figure 5 : fonction d'autocorrélation partielle de var1.

les estimer à partir des observations disponibles dans l'historique.

Remarquons que le modèle de variable d'usure : $x_t = x_{t_0} + a \cdot (t-t_0) + \varepsilon'_t$ (5) est un cas particulier du modèle de l'équation (3) en prenant $a0 = a, a1 = 1$ et $\varepsilon_t = \varepsilon'_t - \varepsilon'_{t-1}$. Ce modèle (5) peut être par ailleurs généralisé sous la forme $x_t = f(t) + \varepsilon_t$, f désignant une fonction du temps. La figure 2 montre le graphe d'une mesure relevant de ce modèle.

Lorsque la fonction f est inconnue, on ne peut distinguer $f(t)$ de ε_t qu'en supposant une variation lente de f .

On pourra détecter les anomalies en surveillant par une carte de contrôle standard la variable calculée $y_t = x_t - f(t)$. Cela n'est possible que si f est connue (cas très rare) ; sinon, on pourra utiliser avec profit le modèle autorégressif précédent.

Quand doit-on renoncer à une carte de contrôle sur données brutes ?

On doit y renoncer à chaque fois qu'on prend conscience que l'hypothèse 2 sur l'indépendance des mesures successives n'est pas vérifiée. Pour ce faire, plusieurs tests statistiques sont disponibles ; pour notre part, nous proposons l'examen visuel de la fonc-

tion d'autocorrélation, qui peut être conforté par le "test du portemanteau" ou "test de Box-Pierce".

Considérons la variable X formée par la série des mesures x_t et la variable X_h obtenue en retardant de h les mesures de la variable X :

$$X_h(t) = X(t-h) = x_{t-h}$$

La fonction d'autocorrélation de la variable X est définie pour chaque valeur de décalage h par le coefficient de corrélation $\rho(h)$ entre la variable X et la variable X_h .

Bien évidemment, $\rho(0) = 1$ et la valeur théorique de $\rho(h)$ est nulle lorsque x_t vérifie l'hypothèse 2 dans la situation standard.

Un exemple de données relevant de la situation standard avec sa fonction d'autocorrélation est donné figure 4 : les coefficients expérimentaux $\hat{\rho}(h)$, estimés à partir des données, sont proches de 0 (pour $h > 0$) et ils sont à l'intérieur d'un intervalle de confiance délimité par les limites rouges sur le corrélogramme (graphique de la fonction d'autocorrélation). En revanche, les modèles non standard décrits dans le paragraphe précédent donnent des valeurs $\hat{\rho}(h)$ significativement différentes de 0, ce qui caractérise le non-respect de l'hypothèse 2. On peut le constater sur les corrélogrammes associés aux données des figures 1 et 2 montrés figure 5.

Le test du portemanteau consiste à examiner si les K premiers $\hat{\rho}(h)$ sont significativement différents de 0. On compare la valeur :

$$Q' = N(N+2) \sum_{h=1}^K \frac{(\hat{\rho}(h))^2}{N-h} \quad (6)$$

aux valeurs prises par une loi du χ^2 à K degrés de liberté : on refuse l'hypothèse d'indépendance des x_t si $Q' > \chi^2_{1-\alpha}(K)$. Dans la formule (6), N désigne le nombre d'observations dans l'historique.

Dans le cas des données de la figure 1 où $N = 70$, en choisissant $K = 23$ et $\alpha = 5\%$, on obtient $Q' = 542$, ce qui est nettement supérieur à la valeur $\chi^2_{1-\alpha}(23) = 35,2$ que l'on peut trouver dans une table statistique du χ^2 .

Le même test appliqué sur les données de la figure 4 donne une valeur qui conduit à accepter l'hypothèse d'indépendance des mesures successives.

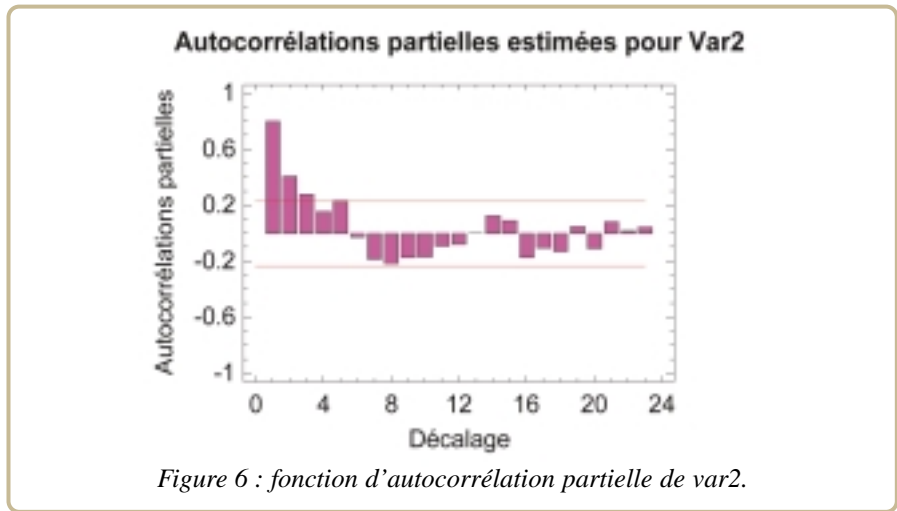


Figure 6 : fonction d'autocorrélation partielle de var2.

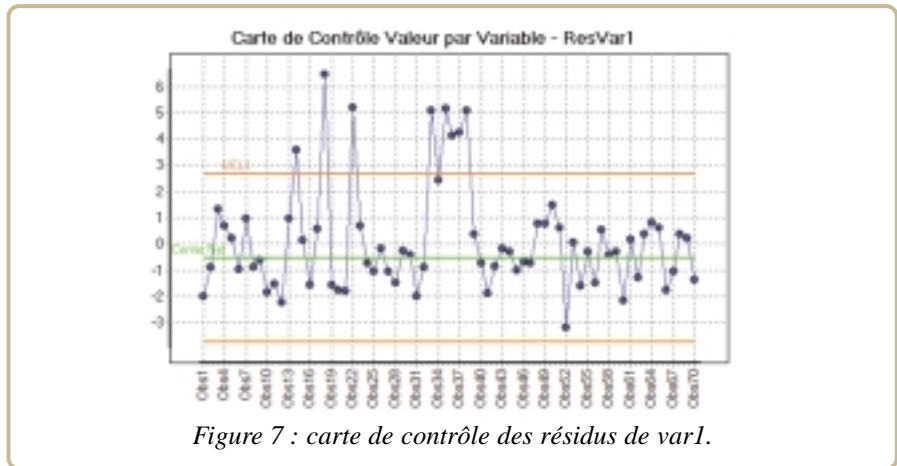


Figure 7 : carte de contrôle des résidus de var1.

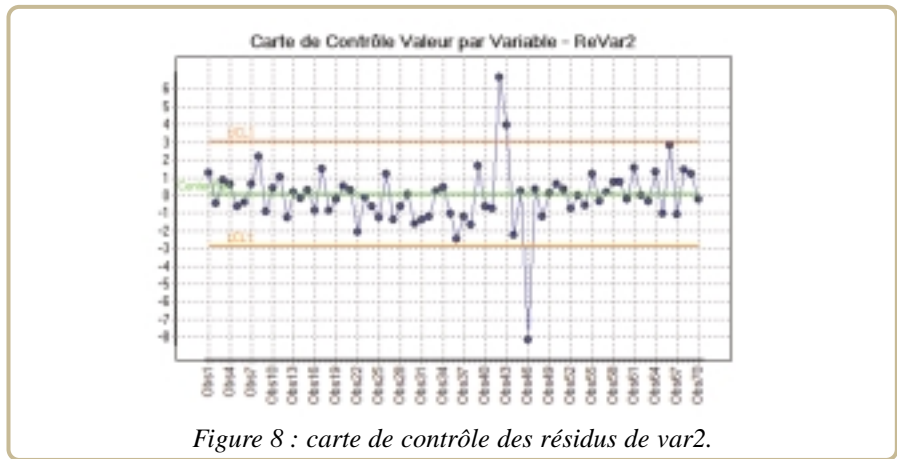


Figure 8 : carte de contrôle des résidus de var2.

Choix du modèle autorégressif

Pour déterminer l'ordre du modèle AR, on considère la valeur prise, pour les différents décalages h , par la fonction d'autocorrélation partielle $\tau(h)$.

Les valeurs $\tau(h)$ peuvent être obtenues en utilisant un logiciel statistique classique traitant des séries chronologiques ; leur mode de calcul déborde le sujet de cet article.

La figure 7 montre la fonction d'autocorrélation partielle associée aux données de la figure 1 : seule la valeur $\tau(1)$ est significativement différente de 0 et donc le modèle AR à considérer est d'ordre 1 : c'est celui de la formule (3).

La figure 8 montre la fonction d'autocorrélation partielle associée aux données de la figure 2 : les valeurs $\tau(1)$ et $\tau(2)$ sont significativement différentes

de 0 et c'est donc le modèle AR d'ordre 2 qui convient : celui de la formule (4). Il est possible de considérer des modèles autorégressifs d'ordre supérieur à 2, mais cela est rarement nécessaire dans la pratique.

Estimation du modèle autorégressif, calcul de la variable à contrôler

Si c'est un modèle d'ordre 2 qui a été choisi, il faut déterminer, à partir de l'historique observé, les estimations $\hat{\mathbf{a}}_0$, $\hat{\mathbf{a}}_1$ et $\hat{\mathbf{a}}_2$ des coefficients \mathbf{a}_0 , \mathbf{a}_1 et \mathbf{a}_2 du modèle. Cette estimation est encore donnée par les logiciels de statistiques : il s'agit des coefficients de la régression linéaire de la variable X sur les variables décalées X_1 et X_2 .

La série des résidus de cette régression $\hat{\varepsilon}_t = \mathbf{x}_t - (\hat{\mathbf{a}}_0 + \hat{\mathbf{a}}_1 \cdot \mathbf{x}_{t-1} + \hat{\mathbf{a}}_2 \cdot \mathbf{x}_{t-2})$ est la variable qu'il convient de suivre avec une carte de contrôle classique pour mieux détecter les anomalies.

Les graphes des variables transformées $\hat{\varepsilon}_t$ issues des données des figures 1 et 2, montrent des anomalies qui n'apparaissent pas sur les cartes de contrôle classiques appliquées aux données brutes.

Extension de la méthode

La démarche présentée consiste donc à suivre non pas les mesures brutes, mais les différences $\mathbf{d}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ ou encore les écarts entre la mesure \mathbf{x}_t et une valeur attendue calculée à partir d'un modèle supposé.

Cette démarche est très générale et peut être appliquée pour des modèles plus complexes de séries chronologiques permettant de prédire une mesure à partir de la connaissance de mesures antérieures : modèles MA (moving average), Arima, Sarima. On trouvera des exposés sur ces modèles dans la littérature et de nombreux outils logiciels permettent de calculer les résidus à surveiller.

Cette démarche peut aussi être étendue à la surveillance de l'écart entre une mesure et sa prédiction obtenue à l'aide d'autres mesures. Le modèle de prédiction peut alors être donné par des relations théoriques entre variables physico-chimiques ou bien par un modèle de relations simples (linéaires, polynomiales...) dont les coefficients auront été estimés par l'étude statistique d'un historique.

En conclusion

Au travers de ces quatre derniers articles traitant des situations non standard, nous avons voulu montrer qu'il est utile de s'échapper de la routine des cartes classiques si l'on veut traiter avec plus de rigueur et d'efficacité les données collectées dans les ateliers.

Des solutions existent pour obtenir, à l'aide de transformations bien adaptées, des cartes plus sensibles permettant de mieux détecter et identifier les anomalies véritables.