

C'EST BON A SAVOIR

EXPLORER LE TEMPS AVEC SAS: quelques astuces en visualisation préliminaire

Jean-Paul VALOIS
ELF EXPLORATION PRODUCTION, 64018 PAU CEDEX
Tél : 05 59 83 44 89 , Email : jean-paul.valois@elf-p.fr

PRESENTATION

Le système SAS dispose de différentes procédures pour étudier les séries chronologiques (module SAS/ETS). Ce module est en outre accompagné d'un interface graphique. Bien qu'il constitue le coeur de l'offre SAS en la matière, il ne fait pas l'objet du présent article, orienté analyse préliminaire.

Le choix d'un modèle (termes autorégressifs, moyennes mobiles...) peut en effet être facilité si l'on opère une analyse exploratoire, particulièrement si plusieurs variables sont en jeu¹(1). Le module SAS/INSIGHT ne permet pas lui-même d'effectuer les traitements chronologiques, mais peut néanmoins rendre de précieux services, en préalable à la modélisation: 'scatter plots', reconnaissance des tendances, des changements de régimes de la série, constatation du lien de ces changements avec certaines variables, etc. Des graphiques générés par SAS/GRAPH peuvent également permettre des constatations utiles.

Plusieurs astuces vont faciliter une telle exploration préliminaire des données temporelles. Nous nous intéressons ici à l'affichage des dates dans SAS/INSIGHT et SAS/GRAPH.

1. Afficher les dates en clair sous INSIGHT

Les dates SAS sont gérées en nombre de jours (depuis 01/01/60). Sans précaution particulière, les graphiques de INSIGHT vont utiliser pour les graduations ces nombres impénétrables du type 8245.... 9560..

1.1. Une solution interactive.

Sous INSIGHT, Edit --> Format --> Other --> permet de choisir le format approprié (par exemple year2. pour '96') ; désigner la variable à représenter par ce format (par ex: la variable DATE de la table).

Si l'on sauve alors la table ainsi modifiée (File --> Save --> Data), le format d'affichage choisi sous INSIGHT sera reproduit lors des autres appels d'INSIGHT sur cette table ainsi que pour

¹ J.P. VALOIS , F. PETIT, G. HALLE, B. TRUONG VAN (1998) Analyse exploratoire et régression sur série chronologique, application à l'affaissement d'un champ pétrolier, XXXè journées ASU, Rennes (en préparation).

toute visualisation de la table par les classiques fenêtres SAS d'affichage plein écran (FSVIEW). On peut bien sûr modifier à nouveau par la suite le format, et prendre par exemple le format DDMMYY8 ('31/12/96').

Comme les tableurs, INSIGHT offre donc un moyen simple et interactif de choisir le format pour afficher une date. Des essais interactifs sont possibles et peuvent réconcilier certains utilisateurs avec l'aspect hirsute de certains noms de format sous SAS.

1.2. Une solution par programmation préalable.

Il peut s'avérer intéressant d'utiliser une étape DATA préliminaire pour générer une nouvelle variable. On choisit ci-après un format qui reflète les années, mais qui contienne le reste de l'information sous forme décimale par exemple 75,5 pour 30/06/75 ou 76,999 pour 31/12/76. La lecture est alors plus aisée que celle en nombre de jours, et l'on conserve un format numérique, intéressant dans divers traitements (alors que la traduction 31/12/76 est une chaîne de caractères).

La macro suivante utilise la date SAS (en nombre de jours depuis 01/01/60), et extrait successivement la partie année (par ex: 75) puis le quantième du jour, enfin convertit ce quantième en pourcentage.

```
%MACRO YEARDEC (DATAIN = _last_ , NBJOURS = NBJOURS ,
                OUT = myres , YDEC = _YDEC );

%PUT -- MACRO YEARDEC----- nbjours --> date decimale ex 60.999 pour 31/12/60 -- ;

DATA &OUT (DROP = __TIM2 - __TIM9) ;
    LENGTH __TIM2 $4. __TIM3 $2. __TIM7 $7 __TIM8 $5 ;

    SET &DATAIN ;
    IF &NBJOURS NE . THEN DO ;

/*----- EXTRACTION DE DECENNIE ----- */

    __TIM2 = put (&NBJOURS , year4. );
    __TIM3 = SUBSTR (__TIM2 , 3 2) ;

/*----- EXTRACTION DU QUANTIEME JOURS - CONV DECIMALE --- */

    __TIM4 = put (&nbjours , julday4. );
    __TIM5 = ( __TIM4 / 366 ) - 0.001 ; /* la soustraction évite un 100 pour 31/12 */

    __TIM7 = PUT ( __TIM5 , 7 5) ;
    __TIM8 = SUBSTR (__TIM7,2,5) ;

/*----- CONCATENATION ----- */

    __TIM9 = __TIM3 || __TIM8 ;
    &YDEC = INPUT (__TIM9 , 7 4) ;
    END ;
    ELSE &YDEC = . ;
run ;

%PUT -- FIN MACRO YEARDEC----- ;
%MEND ;
```

Utilisation sur la table: mytable avec la variable à transformer : mydate (=nb de jours SAS):

```
%YEARDEC (DATAIN = mytable , NBJOURS = mydate.
           OUT = myres , YDEC = _YDEC );
```

on obtient la table myres avec la nouvelle variable _YDEC.

Remarque:

Avec ce format 75.2 = 14 mars 75 (20 % de l'année) et non février 75!

2. Appel de INSIGHT par la PROC INSIGHT

La méthode précédente ne garantit pas que les axes seront gradués de façon agréable: l'automatisme d' INSIGHT peut très bien faire démarrer par défaut un axe à la mi-année par exemple.

On peut certes sous INSIGHT régler ensuite les axes de façon interactive. Mais il peut s'avérer astucieux de commander directement des axes adaptés aux besoins. C'est possible en appelant INSIGHT par la procédure appropriée au lieu de l'habituel menu déroulant.

Cette méthode peu connue s'utilise par une instruction programme du Display Manager (fenêtre bleue):

```
PROC INSIGHT DATA = mytable ;
```

Dans les lignes suivantes, on veut déclencher SAS/INSIGHT sur la table mytable, générer automatiquement un LINE PLOT des variables: «écart» en ordonnée, «andec» en abscisse. On souhaite en outre un axe horizontal s'étalant entre les années 60 et 90, avec annotations tous les 5 ans, et graduations tous les ans - soit 4 graduations par période de 5 ans - (figure 1):

```
PROC INSIGHT ;
OPEN mytable </NODISPLAY >;
LINE écart * andec
/XAXIS = 60 90 5 4 60 90 ;          RUN ;
```

(en minuscule, les données prises pour exemple, en majuscule les noms codés SAS obligatoires).

Remarques:

1/ L'option / NODISPLAY (facultative) supprime l'affichage en mode tableur des données elles-mêmes. Ce peut être intéressant dans quelques cas, par exemple si l'on programme une application SAS/AF et que l'on souhaite mettre l'utilisateur en contact immédiat avec le graphique souhaité.

2/ Cet appel d'INSIGHT par procédure fait gérer INSIGHT par le Display Manager SAS , elle est plus économe en ressources que l'appel par Menu interactif (et fournit donc de meilleures performances).

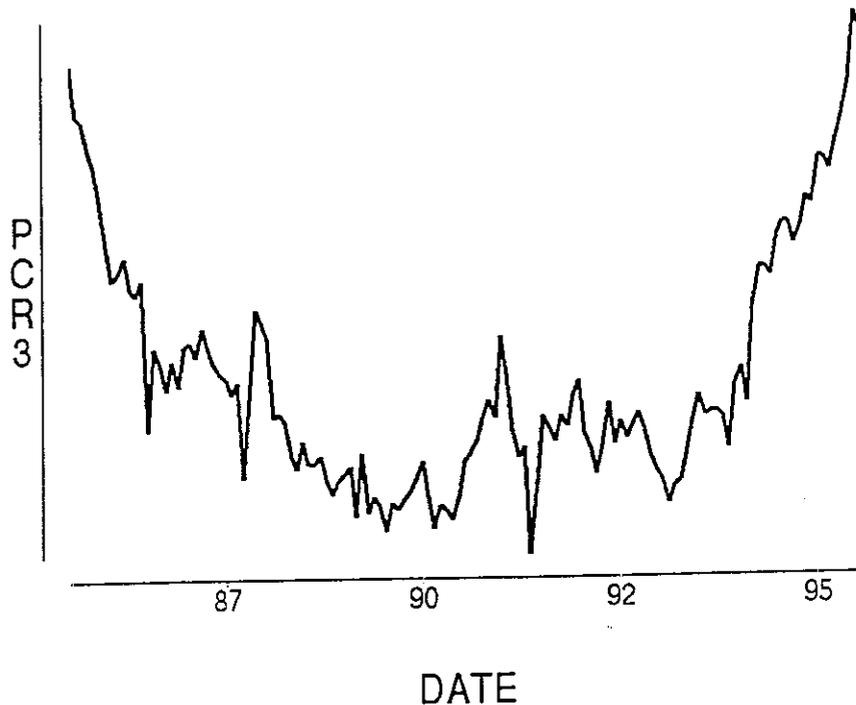


Figure 1 : Graphique produit par SAS/INSIGHT. Les graduations peuvent être contrôlées soit de façon interactive (Edit / Formats, cf. paragraphe 1), soit en appelant le module par PROC INSIGHT et en contrôlant les axes par des options lors de cet appel (cf. paragraphe 2).

3. Les dates sur les axes de SAS/GRAPH

On utilise ici une date SAS (nombre de jours depuis 01/01/60).

Par défaut, les axes produits sous SAS/GRAPH (par la proc GPLOT par exemple), vont eux aussi être gradués en nombre de jours. Or il est plus sympathique de repérer événements ou tendances d'après nos habitudes calendaires.

On peut provoquer l'affichage en jours par une instruction format dans la PROC GPLOT:

```
PROC GPLOT DATA = mytable ;
      FORMAT date DDMMYY8 ;
      PLOT   écart * date ;           RUN ;
```

(en minuscule, les données prises pour exemple, en majuscule les noms codés SAS obligatoires).

Mais les graduations sont fournies à intervalles réguliers, et l'alternance de mois de longueur variable (30, 31, 28 jours) va produire un affichage peu sympathique selon l'exemple suivant:

26/11/81 14/06/82 31/12/82 19/07/83 04/02/84

ces intervalles sont réguliers en nombre de jours, mais pas dans nos repères calendaires habituels!

Solution proposée:

Imposer les limites de l'axe en nombre de jours par l'instruction :

```
AXISn ORDER =( n1 TO n2) ;
```

Pour obtenir des graduations régulières correspondant à nos repères, l'option BY est disponible. Si l'on choisit 1461 jours, on couvre exactement 4 ans (en incluant une année bissextile).

Syntaxe complète :

```
AXIS1 ORDER = (8000 TO 9461 BY 1461 ) MINOR = (nb = 3) ;
PROC GPLOT DATA = mytable ;
  FORMAT date DDMMYY8 ;
  PLOT   écart * date / HAXIS = AXIS1 ;          RUN;
```

L'axe va alors comporter des périodes entières de 4 ans. Le deuxième terme de ORDER (après le TO) doit englober la fin de la période voulue, il n'est pas nécessaire de calculer exactement un multiple de 1461 (dans notre exemple 9600 fait l'affaire également).

Le premier terme de ORDER contrôle le jour calendaire pris pour le début de l'axe. Dans notre exemple, on tombe sur le 26 NOV! Il est souvent plus commode de commencer au 1er janvier. Comment trouver la date SAS, en nombre de jours, correspondant par exemple au 01/01/85 ?

Nos découvertes précédentes avec INSIGHT vont ici nous être utiles.

On ouvre INSIGHT en option «NEW» (pas de table SAS appelée dans la fenêtre tableur listant les données). On affecte le format DDMMYY8. à l'une des colonnes (par exemple la colonne A). On saisit un chiffre, par exemple 8000, dès que l'on fait RETURN la conversion en date calendrier s'affiche: 26/11/1981.

En quelques itérations, on trouve que 8036 est le chiffre à fournir pour obtenir le 01/01/82.

On peut alors programmer la proc GPLOT :

```
AXIS1 ORDER = (8036 TO 11000 BY 1461 ) MINOR = (nb = 3) ;
PROC GPLOT DATA = mytable ;
  FORMAT date DDMMYY8. ;
  PLOT   écart * date / HAXIS = AXIS1 ;          run ;
```

et l'on obtient comme graduations (figure 2) ;

et l'on obtient comme graduations (figure 2) :

```
01/01/82      01/01/86      01/01/90
```

Remarques:

1/ On peut simplifier les graduations de l'axe, en optant par exemple pour FORMAT date YEAR2.

2/ On peut aussi utiliser pour SAS/GRAPH la conversion en date décimale proposée en \$1.2.

Bonne exploration !

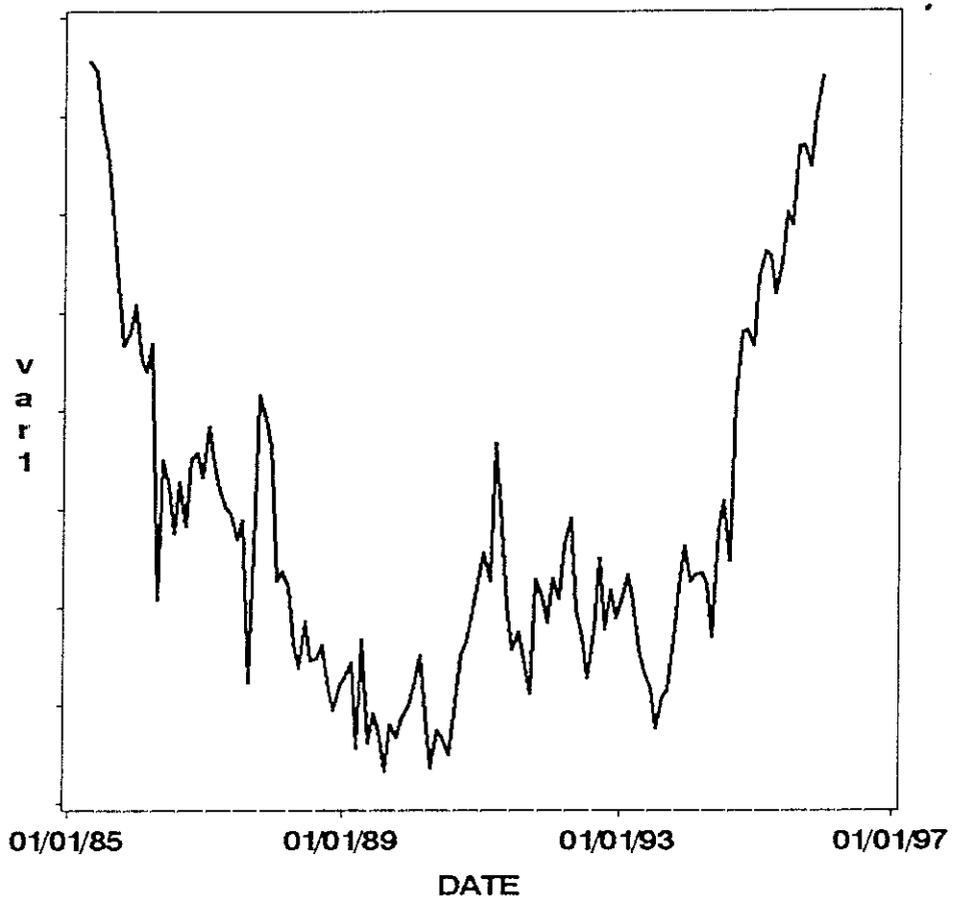


Figure 2 : Graphique produit en utilisant SAS/GRAPH. Une instruction `AXISn`, jointe à une option `FORMAT` dans la procédure `GPLOT`, permet de fournir des graduations correspondant aux repères calendaires (cf. paragraphe 3).