

# CurveModels

CurveModels : Lowess.....	2
CurveModels : Lowess_Discrete.....	4
CurveModels : Polynomials.....	6
CurveModels : Numerical.....	8

## CurveModels : Lowess



Module graphique: tracé de courbes modélisées par régression Lowess (LOcally WEighted regression Scatter plot Smoothing de Cleveland<sup>1</sup>, détails dans Chambers & Coll<sup>2</sup> p. 95-123).



Cette option permet de tracer des courbes modélisant des séries de valeurs par régression Lowess.



La fenêtre de dialogue principale présente neuf paramètres:

Parameter	Default/Value
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	Rh.cnta 39 15
Number of points (Min=4) ?	6
Weight file (optional)	
Variable label file (or #)	Code_VarK
Draw model (1=yes, 2=no)	
Draw observed points (1=yes, 2=no)	
Draw residual sticks (1=yes, 2=no)	

1- Nom du fichier X (fichier binaire contenant les abscisses des points des courbes). Par défaut, des abscisses égales à la suite 1, 2, 3, ... sont utilisées.

2- Le numéro de la colonne à utiliser dans le fichier X. Toutes les courbes ont les mêmes abscisses, définies par les valeurs contenues dans cette colonne. Valeur par défaut=1.

3- Nom du fichier Y (fichier binaire contenant les ordonnées des points des courbes). Il n'y a pas de valeur par défaut: **ce paramètre est obligatoire**.

4- Nombre de points utilisés pour la régression Lowess. Le nombre minimal est de 4. Les valeurs inférieures sont refusées. Valeur par défaut=4.

5- Fichier de pondération. Une pondération non uniforme permet d'accorder plus d'importance à certains points. Par défaut, la pondération est uniforme.

6- Labels par variables. Cette option permet de rajouter un label sur chaque graphique correspondant à une colonne du fichier Y. Par défaut les graphiques n'ont pas de label. La réponse # permet d'affecter un label constitué par le numéro d'ordre de la variable dans les colonnes de Y.

7- Tracé des modèles (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un trait reliant les points de la courbe modélisée.

8- Tracé des points observés (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un petit carré (3 pixels de côté) pour chaque point observé.

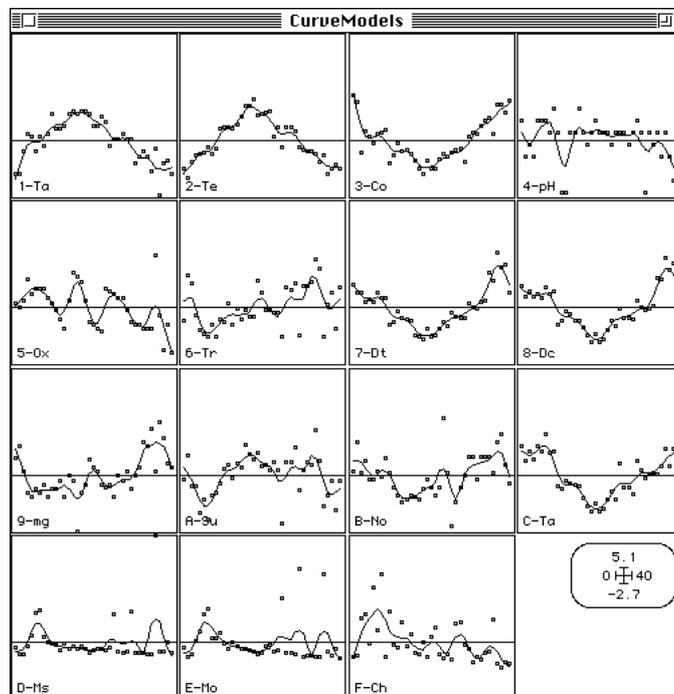
9- Tracé des résidus (oui=1, non=2, valeur par défaut=non). La valeur oui conduit au tracé d'un trait reliant chaque point observé à la courbe modélisée.



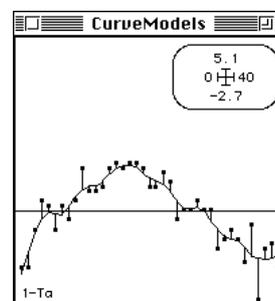
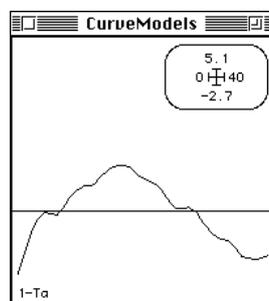
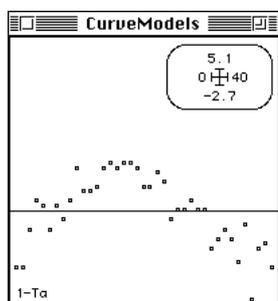
La sélection des variables du fichier Y s'effectue classiquement avec la fenêtre de sélection des lignes et des colonnes. Il n'est par contre pas possible de sélectionner des groupes de lignes et d'effectuer des régressions sur ces groupes.



Utiliser la carte Rhône de la pile ADE-4•Data pour obtenir le fichier Rh.txt (39-15). Le transformer en binaire (fichier Rh) et effectuer l'ACP normée pour obtenir le fichier Rh.cnta qui contient les variables normées. Représenter la régression Lowes sur ces 15 variables physico-chimiques en utilisant une fenêtre de lissage de 6 points et avec comme étiquettes des graphiques le fichier de labels Code\_varX (paramètre numéro 5) :



Sélectionner uniquement la première variable et modifier les paramètres 7, 8 et 9 pour obtenir différentes présentations:



Pour choisir le nombre de voisins utiliser UniVarReg : Lowess -> Error



<sup>1</sup>Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* : 74, 829-836.

<sup>2</sup> Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983) *Graphical methods for data analysis*. Duxbury Press, Boston. 1-395.

## CurveModels : Lowess\_Discrete



Module graphique: Tracé de courbes de régression locale.



Cette option permet de tracer des courbes modélisant des séries de valeurs par régression Lowess. Dans l'option CurveModels : Lowess, un point estimé est calculé pour chaque point observé. Dans la présente option l'estimation locale est calculée en  $n$  points régulièrement réparti sur l'intervalle des valeurs prises en abscisse.



La fenêtre de dialogue principale (File selection) présente neuf paramètres:

Paramètre	Valeur par défaut
X file (default = 1, 2, 3, ..., n)	H12 48 2
X file column number (default = 1)	
Y file (no default)	H12 48 2
Number of points (Min=4) ?	15
Weight file (optional)	
Variable label file (or #)	
Draw model (1=yes, 2=no)	
Draw observed points (1=yes, 2=no)	
Point number (default = 20)	

1- Nom du fichier X (fichier binaire contenant les abscisses des points des courbes). Par défaut, des abscisses égales à la suite 1, 2, 3, ... sont utilisées.

2- Le numéro de la colonne à utiliser dans le fichier X. Toutes les courbes ont les mêmes abscisses, définies par les valeurs contenues dans cette colonne. Valeur par défaut=1.

3- Nom du fichier Y (fichier binaire contenant les ordonnées des points des courbes). Il n'y a pas de valeur par défaut: **ce paramètre est obligatoire**.

4- Nombre de points utilisés pour la régression Lowess. Le nombre minimal est de 4. Les valeurs inférieures sont refusées. Valeur par défaut=4.

5- Fichier de pondération. Une pondération non uniforme permet d'accorder plus d'importance à certains points. Par défaut, la pondération est uniforme.

6- Labels par variables. Cette option permet de rajouter un label sur chaque graphique correspondant à une colonne du fichier Y. Par défaut les graphiques n'ont pas de label. La réponse # permet d'affecter un label constitué par le numéro d'ordre de la variable dans les colonnes de Y.

7- Tracé des modèles (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un trait reliant les points de la courbe modélisée.

8- Tracé des points observés (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un petit carré (3 pixels de côté) pour chaque point observé.

9- Nombre de points de la courbe modèle (20 par défaut).

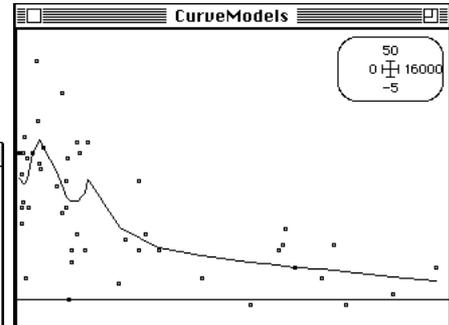


La sélection des variables dans le fichier de données (et dans le fichier des modèles) s'effectue classiquement avec la fenêtre de sélection des lignes et des colonnes. Il n'est par contre pas possible de sélectionner des groupes de lignes.



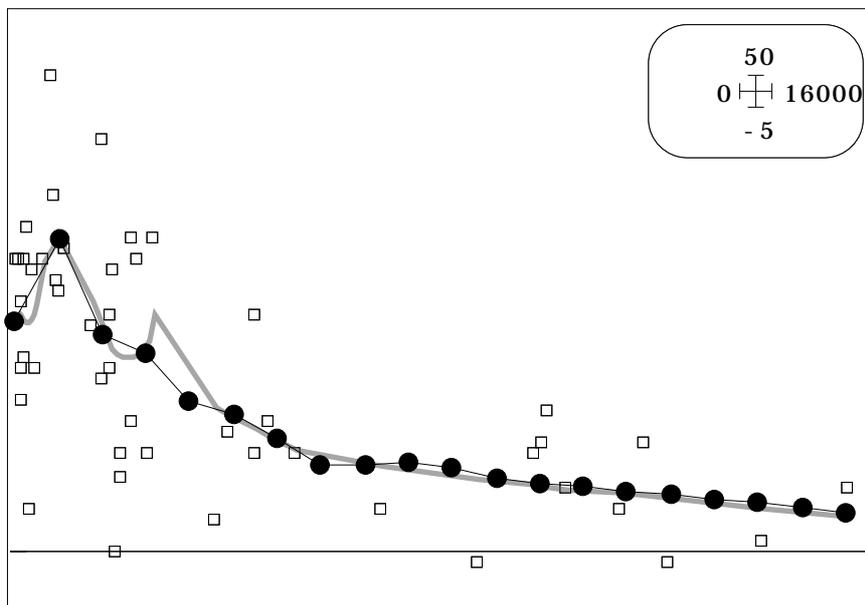
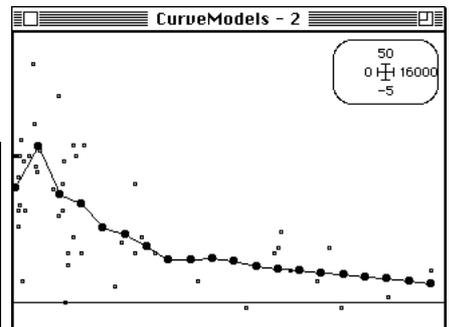
Utiliser la carte Monde et le fichier X12. Tracer la courbe de régression locale de la variable 2 (taux de croissance de la population) en fonction de la variable 1 (Produit Intérieur Brut) :

Lowess			
H file (default = 1, 2, 3, ..., n)	<input type="text" value="H12"/>	48	2
H file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="H12"/>	48	2
Number of points (Min=4) ?	<input type="text" value="15"/>		



Tracer la même courbe en 20 points régulièrement répartis :

Lowess_Discrete			
H file (default = 1, 2, 3, ..., n)	<input type="text" value="H12"/>	48	2
H file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="H12"/>	48	2
Number of points (Min=4) ?	<input type="text" value="15"/>		



## CurveModels : Polynomials



Module graphique: tracé de courbes modélisées par régression polynomiale.



Cette option permet de tracer des courbes modélisant des séries de valeurs par régression polynomiale.



La fenêtre de dialogue principale (File selection) présente neuf paramètres:

Label	Value	Column
X file (default = 1, 2, 3, ..., n)	39	1
X file column number (default = 1)	1	
Y file (no default)	Rh.cnta	15
Order of polynomial (default = 1) ?	2	
Weight file (optional)		
Variable label file (optional)	Code_VarK	
Draw model (1=yes, 2=no)		
Draw observed points (1=yes, 2=no)		
Draw residual sticks (1=yes, 2=no)		

1- Nom du fichier X (fichier binaire contenant les abscisses des points des courbes). Par défaut, des abscisses égales à la suite 1, 2, 3, ... sont utilisées.

2- Le numéro de la colonne à utiliser dans le fichier X. Toutes les courbes ont les mêmes abscisses, définies par les valeurs contenues dans cette colonne. Valeur par défaut=1.

3- Nom du fichier Y (fichier binaire contenant les ordonnées des points des courbes). Il n'y a pas de valeur par défaut: **ce paramètre est obligatoire**.

4- Degré du polynôme. Valeur par défaut=1. Le valeur maximale de ce paramètre est égale au nombre total de points moins 1.

5- Fichier de pondération. Une pondération non uniforme permet d'accorder plus d'importance à certains points. Par défaut, la pondération est uniforme.

6- Labels par variables. Cette option permet de rajouter un label sur chaque graphique correspondant à une colonne du fichier Y. Par défaut les graphiques n'ont pas de label. La réponse # permet d'affecter un label constitué par le numéro d'ordre de la variable dans les colonnes de Y.

7- Tracé des modèles (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un trait reliant les points de la courbe modélisée.

8- Tracé des points observés (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un petit carré (3 pixels de côté) pour chaque point observé.

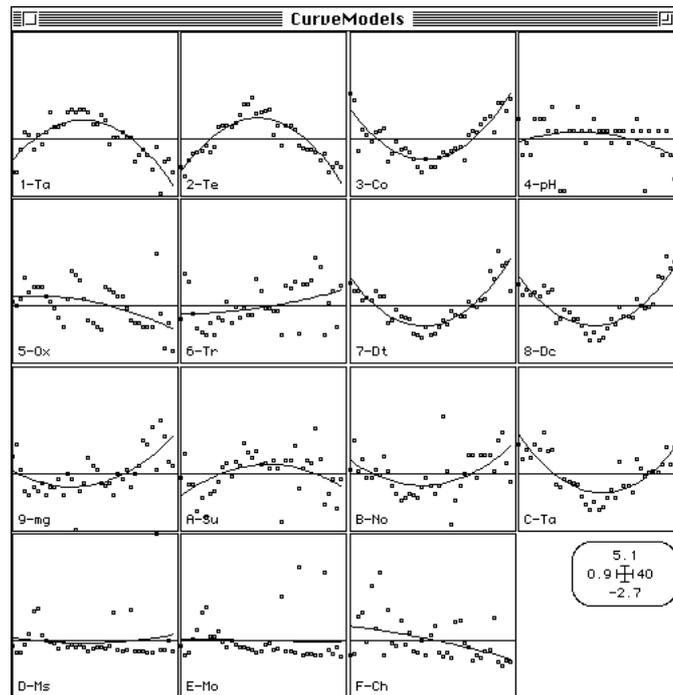
9- Tracé des résidus (oui=1, non=2, valeur par défaut=non). La valeur oui conduit au tracé d'un trait reliant chaque point observé à la courbe modélisée.



La sélection des variables du fichier Y s'effectue classiquement avec la fenêtre de sélection des lignes et des colonnes. Il n'est par contre pas possible de sélectionner des groupes de lignes et d'effectuer des régressions sur ces groupes.



Procéder comme pour CurveModels : Lowess. Avec un polynôme de degré deux sur les 15 variables on obtient la figure suivante :



Remarquer que la régression sur un polynôme de degré un fourni la droite de regression linéaire, sans préjuger toutefois de la qualité de cette régression, et sans en fournir de description numérique.



Pour choisir le degré du polynôme utiliser UniVarReg : Polynomial -> Error.

## CurveModels : Numerical



Module graphique: superposition de courbes et de séries de points.



Cette option permet de superposer des courbes (représentant un modèle quelconque) et des séries de points (correspondant à des valeurs observées).



La fenêtre de dialogue principale (File selection) présente huit paramètres:

Label	Value
# file (default = 1, 2, 3, ..., n)	
# file column number (default = 1)	
Model values file (no default)	Rh5Outd^2 39 5
Data values file (no default)	Rh5 39 5
Variable label file (or #)	Code_Var#5
Draw model (1=yes, 2=no)	
Draw observed points (1=yes, 2=no)	
Draw residual sticks (1=yes, 2=no)	1

1- Nom du fichier X (fichier binaire contenant les abscisses des points et des courbes). Par défaut, des abscisses égales à la suite 1, 2, 3, ... sont utilisées.

2- Le numéro de la colonne à utiliser dans le fichier X. Toutes les courbes ont les mêmes abscisses, définies par les valeurs contenues dans cette colonne. Valeur par défaut=1.

3- Nom du fichier des modèles (fichier binaire contenant les ordonnées des courbes). Ce fichier doit avoir le même nombre de lignes et de colonnes que le fichier des données. Il n'y a pas de valeur par défaut: **ce paramètre est obligatoire**

4- Nom du fichier des données (fichier binaire contenant les ordonnées des séries de points). Ce fichier doit avoir le même nombre de lignes et de colonnes que le fichier des modèles. Il n'y a pas de valeur par défaut: **ce paramètre est obligatoire**

5- Labels par variables. Cette option permet de rajouter un label sur chaque graphique correspondant à une colonne du fichier Y. Par défaut les graphiques n'ont pas de label. La réponse # permet d'affecter un label constitué par le numéro d'ordre de la variable dans les colonnes de Y.

6- Tracé des modèles (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un trait reliant les points de la courbe modélisée.

7- Tracé des points observés (oui=1, non=2, valeur par défaut=oui). La valeur oui conduit au tracé d'un petit carré (3 pixels de côté) pour chaque point observé.

8- Tracé des résidus (oui=1, non=2, valeur par défaut=non). La valeur oui conduit au tracé d'un trait reliant chaque point observé à la courbe modélisée.



La sélection des variables dans le fichier de données (et dans le fichier des modèles) s'effectue classiquement avec la fenêtre de sélection des lignes et des colonnes. Il n'est par contre pas possible de sélectionner des groupes de lignes.



Utiliser l'exemple Rhône. On désire superposer le polynôme de degré 2 obtenu par régression sur la variable numéro 3 (conductivité) avec les points des cinq variables suivantes: 3, 7, 8, 9 et 12 (conductivité, dureté totale, dureté calcique, magnésium, TAC, cf figure ci-dessus). Commencer par extraire la variable numéro 3 du fichier Rh.cnta avec FilesUtil : Row-Col Selection (fichier de sortie Rh3, 39-1):

Row-Col Selection

Input file  39 15

Selection of rows (default = all)

Selection of columns (default = all)

Output file

Quit Ok

Extraire de même les variables 3, 7, 8, 9 et 12 dans le fichier Rh5:

Row-Col Selection

Input file  39 15

Selection of rows (default = all)

Selection of columns (default = all)

Output file

Quit Ok

Utiliser UniVarReg : Initialize pour préparer la régression polynomiale:

Initialize

Explanatory variable

Selected column (default = 1)

Y file: dependent variables  39 1

Option: row weighting

Output file name

Quit Ok

Effectuer cette régression sur un polynôme de degré deux:

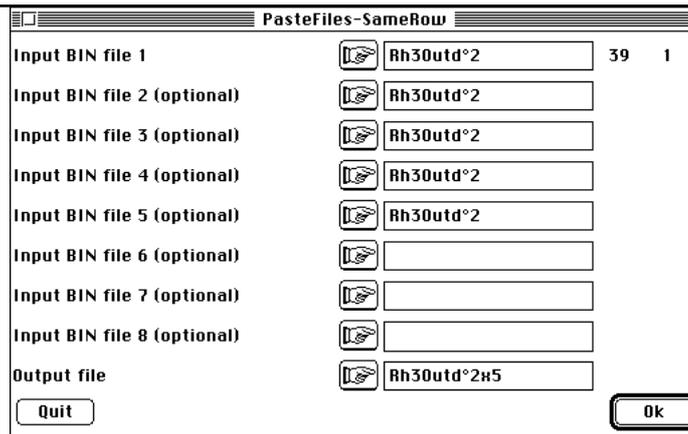
Polynomial -> Model

Input file

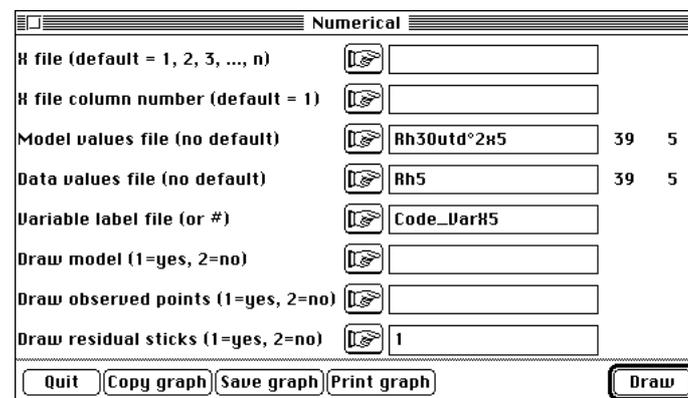
Order of polynomial (default = 2) ?

Quit Ok

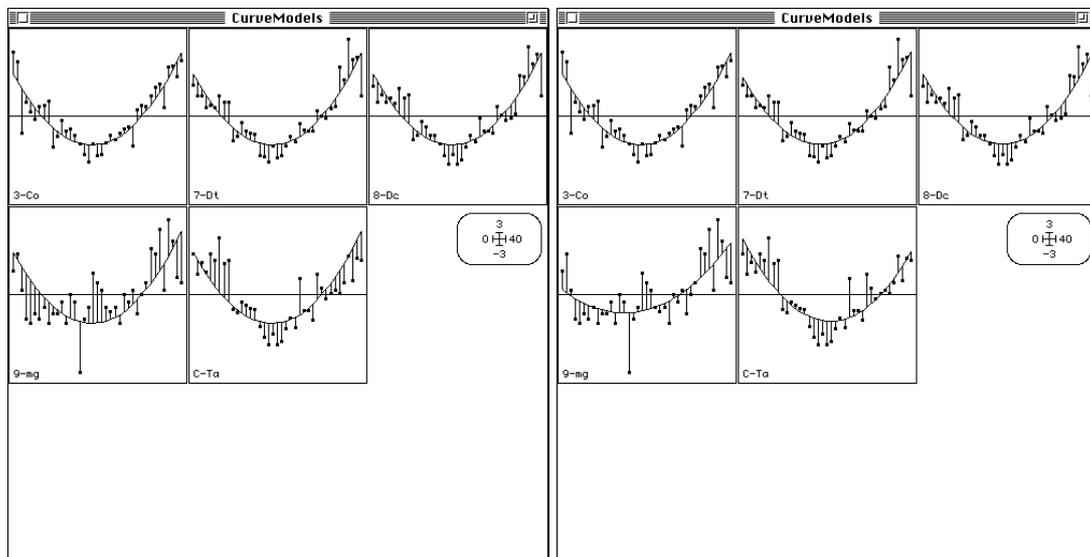
Le fichier obtenu s'appelle Rh3Outd<sup>o2</sup>. Il faut le recopier cinq fois avec FilesUtil : PasteFiles-SameRow dans un fichier qu'on appellera Rh3Outd<sup>o2x5</sup> :



Superposer enfin ces cinq courbes sur les cinq variables du fichier Rh5 avec CurveModels : Numerical.



On obtient ainsi la figure suivante (à gauche), qui diffère peu de celle obtenue en effectuant les 5 régressions polynomiales séparées (à droite):



On trouvera de nombreux autres exemple d'utilisation de cette option dans le module UnivarReg.