

UniVarReg

UniVarReg : Initialize.....	2
UniVarReg : Lowess -> Curves.....	4
UniVarReg : Lowess -> Error.....	7
UniVarReg : Lowess -> Model.....	9
UniVarReg : Lowess -> New Data.....	12
UniVarReg : Polynomial -> Curves.....	13
UniVarReg : Polynomial -> Error.....	14
UniVarReg : Polynomial -> Model.....	15
UniVarReg : Polynomial -> New Data.....	16

UniVarReg : Initialize



Utilitaire de contrôle de données.



L'option s'emploie pour définir un jeu de paramètres qui sera utilisé par toutes les autres options de ce module. On s'intéresse ici au cas d'une **variable explicative unique**, de k variables à expliquer et de modèles de prédiction univariés indépendants pour chacune des variables à prédire.



L'option utilise une seule fenêtre de dialogue :

Explanatory variable	<input type="text" value="date"/>	19	1
Selected column (default = 1)	<input type="text"/>		
Y file: dependent variables	<input type="text" value="Gre"/>	19	14
Option: row weighting	<input type="text"/>		
Output file name	<input type="text" value="DG"/>		

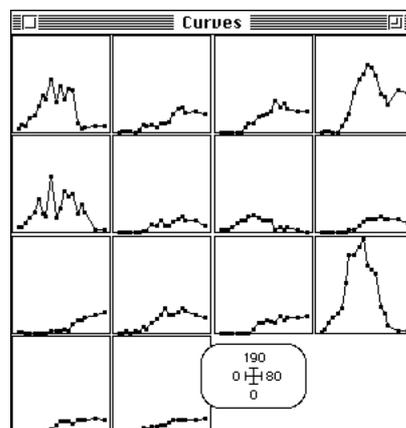
Quit Ok

- Nom du fichier binaire contenant la variable explicative.
- Numéro de la colonne utilisée dans le fichier binaire contenant la variable explicative (par défaut, c'est la première).
- Nom du fichier binaire contenant les variables à prédire.
- Fichier de pondération des lignes (par défaut, on utilise la pondération uniforme).
- Nom générique des fichiers de sortie.



Utiliser la carte Grèbes de la pile ADE-4•Data pour obtenir les fichiers date (19-1) et Gre (19-14). Représenter (Curves : Lines) les variables de Gre en fonction de la variable unique de date :

X file (default = 1, 2, 3, ..., n)	<input type="text" value="date"/>	19	1
X file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="Gre"/>	19	14



Enregistrer l'association des deux fichiers par la présente option:

```
-----  
New TEXT file DG.uni contains the parameters:  
----> Explanatory variables: date [19][1]  
----> Selected variable: 1  
----> Dependent variable file: Gre [19][14]  
----> Row weight file: Uniform_weight  
-----
```

On entre dans les autres options du module par le fichier ---.uni ainsi créé.



Le fichier de la variable explicative peut être omis :

Dans ce cas, la variable explicative prend les valeurs 1, 2, ..., n (le nombre de lignes du fichier des variables à prédire) et est stockée dans un fichier au nom explicite :

```

-----
New TEXT file DGbis.uni contains the parameters:
----> Explanatory variables: DGbis.12...n [19][1]
----> Selected variable: 1
----> Dependent variable file: Gre [19][14]
----> Row weight file: Uniform_weight
-----

```

UniVarReg : Lowess -> Curves



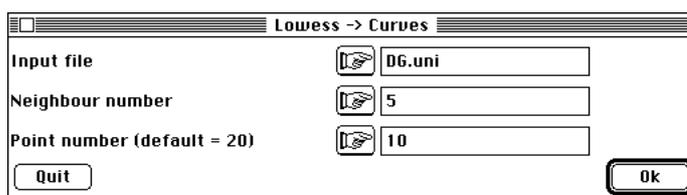
Méthode exploratoire de prédiction univariée par régression locale.



Lissage de courbes par la méthode **Lowess** (**l**ocally **w**eighted regression scatter plot **s**moothing) de Cleveland¹, décrite en détails dans ². On choisit p points régulièrement répartis dans l'intervalle de variation de la variable explicative. En chacun de ses points on estime la variable à prédire par le résultat d'une régression simple pondérée (les poids utilisés dérivent d'une fonction tricube). Cette régression utilise les v plus proches voisins de chaque point : v est le paramètre de lissage. La courbe lissée dépend complètement de ce choix.



L'option utilise une seule fenêtre de dialogue :



 Fichier des paramètres créé par UniVarReg : Initialize.

 Nombre de voisins v utilisés par la régression locale.

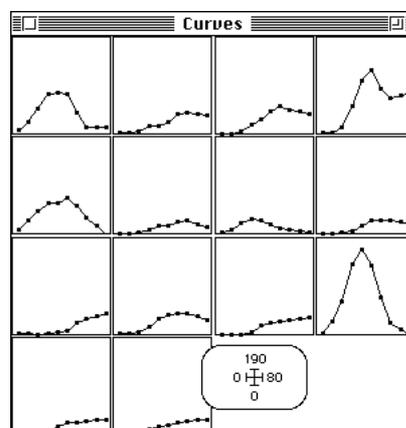
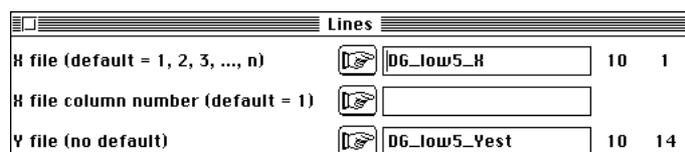
 Nombre de points p utilisé pour la prédiction par régression locale.



Utiliser l'exemple introduit dans UniVarReg : Initialize :

```
Lowess curves
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Number of neighbours used: 5
-----
Output file: DG_low5_X
--- Number of rows: 10, column: 1
This file contains the values (x) used for lowess estimation
Output file: DG_low5_Yest
--- Number of rows: 10, columns: 14
This file contains the lowess model for each variable (columns),
-----
```

Représenter le modèle par Curves : Lines :



On peut, évidemment superposer données et modèles par deux utilisations de cette option.



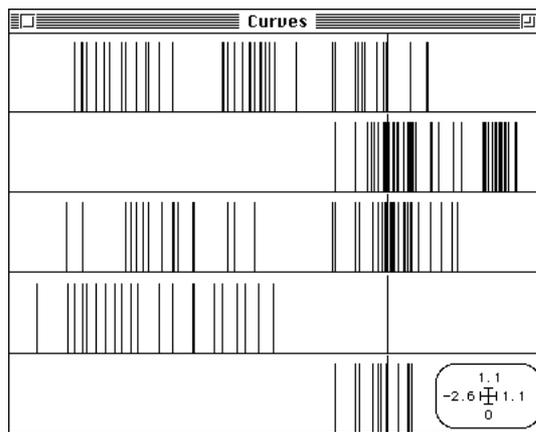
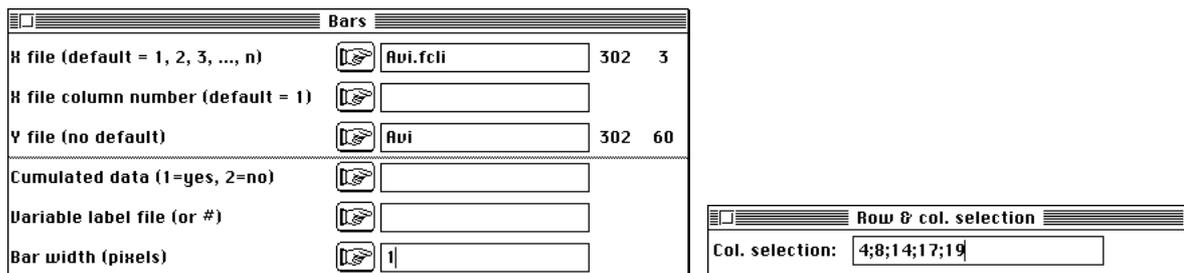
Pour choisir le paramètre de lissage v , utiliser UniVarReg : Lowess -> Error.

L'option fonctionne comme UniVarReg : Lowess -> Model, mais fait l'estimation en des points régulièrement répartis au lieu de les faire pour les valeurs observées de la variable prédictrice. Dans UniVarReg : Lowess -> New data, on fait l'estimation en des points définis par des valeurs supplémentaires de la variable prédictrice.

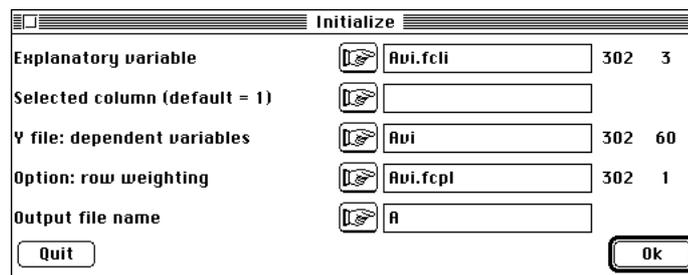
L'option est donc intéressante quand le nombre d'observations est grand et que la prédiction en chaque point observé n'est pas souhaitable. Utiliser par exemple la carte Provence_Corse de la pile ADE-4•Data. Faire l'AFC du tableau Avi (302-60).



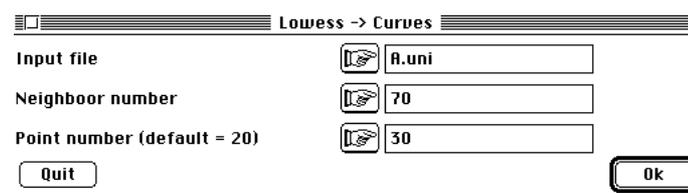
Représenter la présence de 5 des taxons sur l'ordination des 302 relevés faites par le premier facteur de cette analyse :



Pour modéliser ces séries de valeurs 0-1 (0 absence de l'espèce, 1 présence de l'espèce), associer les fichiers :



Prédire ces valeurs en 30 points de l'intervalle sur 70 voisins :



Représenter les modèles en conservant le même paramétrage :

Lines			
X file (default = 1, 2, 3, ..., n)	<input type="text" value="A_low70_H"/>	30	1
X file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="A_low70_Yest"/>	30	60

Représenter de plus la position des 302 relevés, en ajoutant un fichier par MatAlg :

Row, Col → [Rij] with Rij = 1	
Row number	<input type="text" value="302"/>
Column number	<input type="text" value="1"/>
Output file	<input type="text" value="Un302"/>
<input type="button" value="Quit"/>	<input type="button" value="Ok"/>

Bars			
X file (default = 1, 2, 3, ..., n)	<input type="text" value="Aui.fcli"/>	302	3
X file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="Un302"/>	302	1

Assembler les résultats obtenus :



La représentation des données (présence) est enrichie par la représentation des modèles (estimation des fréquences) qui tient compte de la répartition initiale des relevés.



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

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

UniVarReg : Lowess -> Error



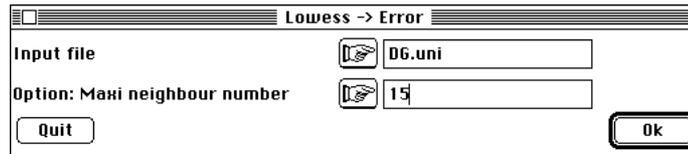
Méthode exploratoire de prédiction univariée par régression locale : utilitaire pour le choix du nombre de voisins.



Dans la régression locale on prédit la valeur prise en un point par la variable à prédire en utilisant une régression pondérée sur les ν plus proche valeurs. On peut donc calculer, en fonction de ν , l'erreur moyenne ¹ commise sur tous les points où sont connues variable prédictrice et variable prédite. On peut tracer l'évolution de cette erreur en fonction du nombre de voisins utilisés et faire ainsi un choix motivé.



L'option utilise une seule fenêtre de dialogue :



Fichier des paramètres créé par UniVarReg : Initialize.

Maximum du nombre de voisins ν utilisés par la régression locale. On ne dépasse pas, en général, 70% du nombre total de points. Limiter ce maximum pour les grands effectifs afin de limiter les calculs nécessaires.

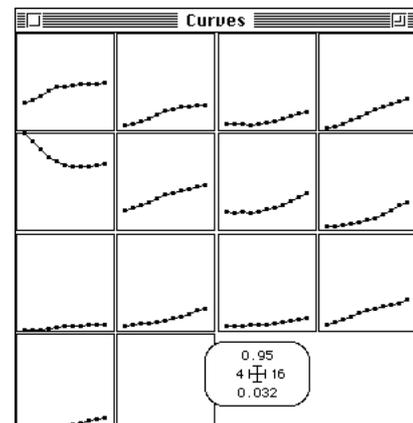
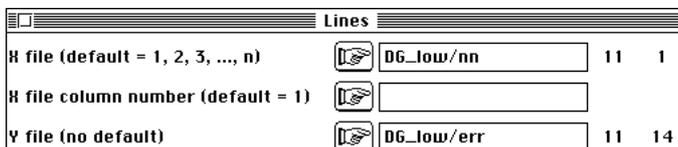


Utiliser l'exemple introduit dans la fiche de UniVarReg : Initialize :

```
lowesserror: total error of 1D-locally weighted regression
X coordinates file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Data file: Gre
--- Number of rows: 19, columns: 14
Maximal neighbour number: 15
Uniform weight
Output file: DG_low/err
--- Number of rows: 11, columns: 14
This file contains, for each variable (columns),
and for each number of neighbors (rows) the mean
estimation error for all sampling points
Output file: DG_low/nn
--- Number of rows: 11
This file contains the number of neighbours used in each locally weighted
regression
+2.94E-01 +7.88E-02 +8.56E-02 +5.52E-02 +9.48E-01 +2.19E-01 +1.99E-01
+6.21E-02 +3.31E-02 +7.71E-02 +7.55E-02 +9.26E-02 +4.20E-02 +3.72E-02
+3.15E-01 +9.31E-02 +8.51E-02 +6.72E-02 +8.64E-01 +2.39E-01 +1.94E-01
...

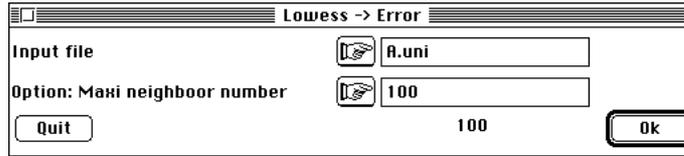
```

L'option édite ces erreurs pour une lecture directe et permet de tracer les courbes des erreurs par Curves : Lines :





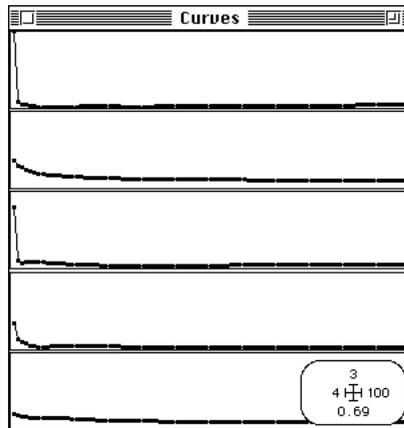
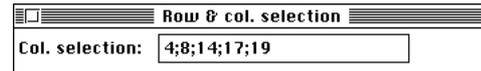
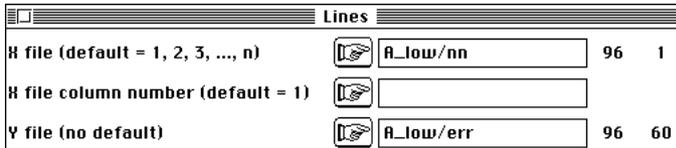
Utiliser l'exemple introduit dans la fiche de UniVarReg : Lowess -> Curves :



```

lowesserror: total error of 1D-locally weighted regression
X coordinates file: Avi.fcli
--- Number of rows: 302, columns: 3
--- Selected column: 1
Data file: Avi
--- Number of rows: 302, columns: 60
Maximal neighbour number: 100
Weight file: Avi.fcpl
--- Number of rows: 302, columns: 1
Output file: A_low/err
--- Number of rows: 96, columns: 60
This file contains, for each variable (columns),
and for each number of neighbors (rows) the mean
estimation error for all sampling points
Output file: A_low/nn
--- Number of rows: 96
This file contains the number of neighbours used in each local regression

```



Dans le premier cas, la croissance immédiate des courbes impose un petit nombre de voisins. Dans le second, en augmentant le nombre de voisins, on ne modifie pas sensiblement l'erreur, tout en augmentant le lissage du modèle.



N.B. Le point de mesure lui-même n'est pas utilisé dans la prédiction. Il s'agit de prévoir quel nombre de voisins utiliser pour des points qui ne sont pas des points de mesures. Le nombre de voisins est celui qui est effectivement utilisé. Cette technique est employée dans des régressions locales à deux dimensions dans ².



¹ Cleveland, W.S. & Devlin, S.J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* : 83, 596-610.

² Thioulouse , J., Chessel, D., Champely, S. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics* , 2, 1-14.

UniVarReg : Lowess -> Model



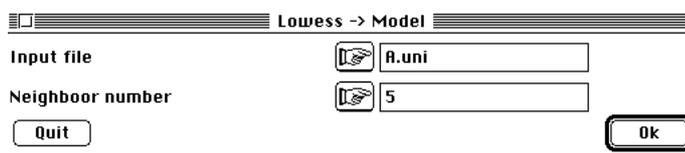
Méthode exploratoire de prédiction univariée par régression locale.



Lissage de courbes par la méthode **Lowess** (**l**ocally **w**eighted regression scatter plot smoothing, cf. UniVarReg : Lowess -> Curves). L'option conserve sur disque les valeurs représentées directement dans CurveModels : Lowess.



L'option utilise une seule fenêtre de dialogue :



 Fichier des paramètres créé par UniVarReg : Initialize.

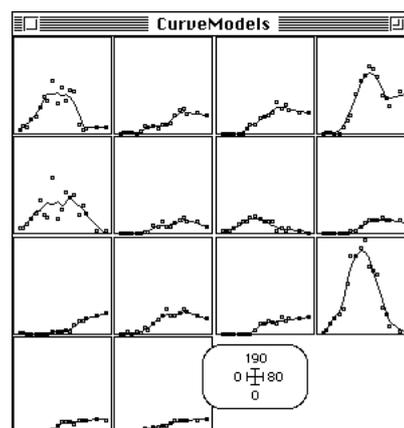
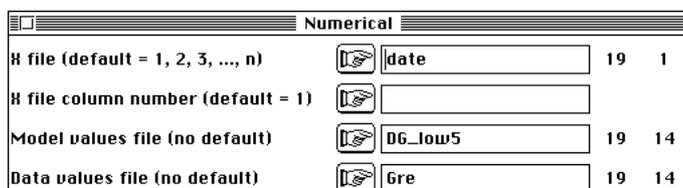
 Nombre de voisins ν utilisés par la régression locale.



Utiliser l'exemple introduit dans la fiche de UniVarReg : Initialize :

```
Lowess model
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Number of neighbours: 5
Weights = 0.052632
Output file: DG_low5
--- Number of rows: 19, columns: 14
This file contains the lowess model for each variable (columns),
Variable 1 | Err/Var  0.153
Variable 2 | Err/Var  0.035
Variable 3 | Err/Var  0.038
Variable 4 | Err/Var  0.022
Variable 5 | Err/Var  0.434
...
Variable 13 | Err/Var  0.015
Variable 14 | Err/Var  0.020
```

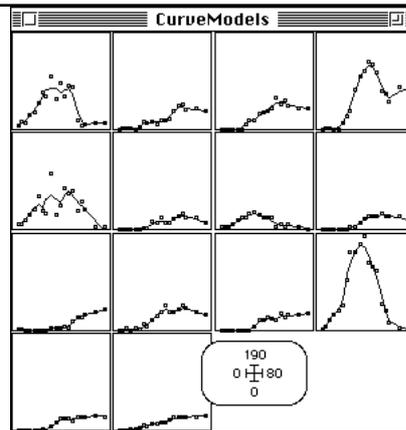
Représenter données et modèles par CurveModels : Numerical :



Le listing contient, outre la possibilité de faire cette figure, les valeurs relatives de l'erreur rapportée à la variance initiale.

La figure peut être obtenue directement par CurveModels : Lowess :

Lowess	
X file (default = 1, 2, 3, ..., n)	date 19 1
X file column number (default = 1)	
Y file (no default)	Gre 19 14
Number of points for regression ?	6 19 14



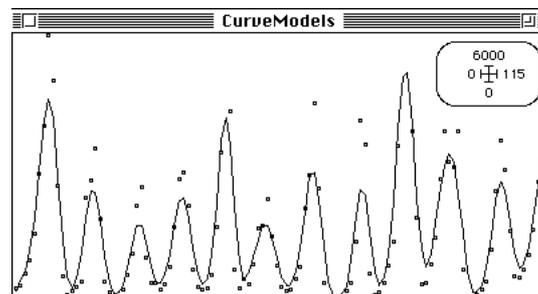
Il faut utiliser un point de plus dans le module graphique, pour obtenir le même résultat. La procédure du module graphique, en effet, compte un point voisin supplémentaire qui n'est en fait pas utilisé (le dernier des voisins a un poids nul).

Pour une exploration purement graphique l'option CurveModels : Lowess suffit. Pour préciser l'erreur associée au nombre de voisins, on utilisera le présent module qui permet en outre des estimations soit en un nombre limité de points (UniVarReg : Lowess -> Curves) soit pour des points supplémentaires (UniVarReg : Lowess -> New Data)

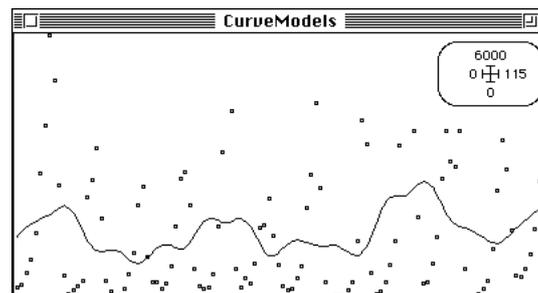


La régression locale utilise les ν plus proches voisins de chaque point : ν est le paramètre de lissage. La courbe lissée dépend complètement de ce choix. Ci-dessous lissage de la chronique du nombre de lynx capturés dans le district de la rivière Mackenzie, au Nord-Ouest du Canada de 1821 à 1934 (141 points : données de Elton et Nicholson 1942¹, reprises dans Ord 1979², disponibles dans la carte Lynx de la pile ADE-4•Data) avec les valeurs 5, 20, 50 du paramètre de lissage.

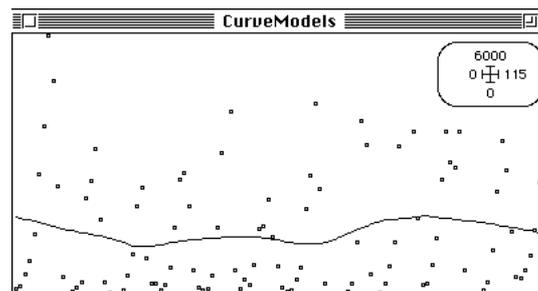
Lowess	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	Lynx
Number of points for regression ?	6



Lowess	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	Lynx
Number of points for regression ?	21



Lowess	
X file (default = 1, 2, 3, ..., n)	
X file column number (default = 1)	
Y file (no default)	Lynx
Number of points for regression ?	51



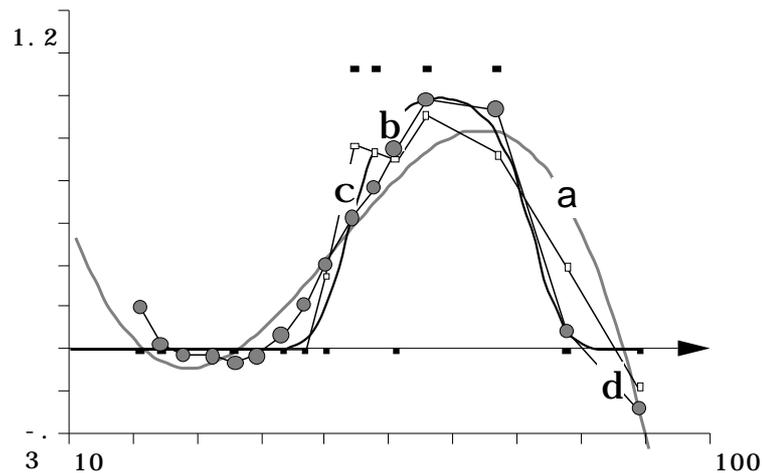
Voir un autre exemple dans la fiche 1 du fascicule 5 de la documentation thématique.

Comme exercice, utiliser le jeu de données :

X	20	23	26	30	33	36	40	43	46	50	53	56	60	70	80	90
Y	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0

pour comparer divers ajustements obtenus à partir du jeu de données artificielles³.

- a — Régression polynomiale de degré 3 (UniVarReg : Polynomial -> Model).
- b — Régression logistique Gaussienne.
- c — Régression linéaire sur les 3 premiers vecteurs propres de l'opérateur de voisinage (OrthoVar : Modelling).
- d — Régression locale pondérée. (extrait de ⁴).



¹ Elton, C. & Nicholson, M. (1942) The ten year cycle in numbers of lynx in Canada. *Journal of Animal Ecology* : 11, 215-244.

² Ord, J.K. (1979) Time-series and spatial patterns in ecology. In : *Spatial and temporal analysis in ecology*. Cormack, R.M. & Ord, J.K. (Eds.) International Co-operative Publishing House, Fairland, Maryland. 1-94.

³ Ter Braak, C.J.F. & Looman, C.W.N. (1986b) Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* : 65, 3-11.

⁴ Méot, A., Chessel, D. & Sabatier, R. (1993) Opérateurs de voisinage et analyse des données spatio-temporelles. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 45-72.

UniVarReg : Lowess -> New Data



Méthode exploratoire de prédiction univariée par régression locale.



Lissage de courbes par la méthode **Lowess** (**l**ocally **w**eighted regression scatter plot smoothing, cf. [UniVarReg : Lowess -> Curves](#) et [UniVarReg : Lowess -> Model](#)). L'option permet l'extension de la prédiction à des valeurs supplémentaires de la variable prédictive.



L'option utilise une seule fenêtre de dialogue :

Input file		DG.uni	
Neighbour number		5	
# values: supplementary data		date	19 1
Selected column (default = 1)			15

Quit Ok

Fichier des paramètres créé par [UniVarReg : Initialize](#).

Nombre de voisins v utilisés par la régression locale.

Fichier des valeurs supplémentaires de la variable prédictive.

Numéro de la colonne utilisée (par défaut, c'est la première).



Utiliser l'exemple introduit dans la fiche de [UniVarReg : Initialize](#) :

```
Lowess model
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Neighbour number used: 5
X supplementary coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
-----
Output file: DG_low5_ls
--- Number of rows: 19, columns: 14
This file contains the lowess model for each variable (columns),
-----
```

Vu le choix du fichier des valeurs supplémentaires, le contenu du fichier `DG_low5_ls` est évidemment identique au contenu du fichier `DG_low5` créé par [UniVarReg : Lowess -> Model](#).

UniVarReg : Polynomial -> Curves



Méthode exploratoire de prédiction univariée par régression polynomiale.



Lissage de courbes par des modèles polynomiaux. On choisit p points régulièrement répartis dans l'intervalle de variation de la variable explicative. En chacun de ses points on estime la variable à prédire par le résultat d'une régression polynomiale de degré fixé.



L'option utilise une seule fenêtre de dialogue :

Polynomial -> Curves

Input file

Order of polynomial (default = 2) ?

Point number (default = 20)

Quit 15 Ok

Fichier des paramètres créé par UniVarReg : Initialize.

Degré utilisé la régression polynomiale (par défaut, c'est 2).

Nombre de points p utilisé pour la prédiction (par défaut, c'est 20).



Utiliser l'exemple introduit dans UniVarReg : Initialize :

Polynomial model

•••

Order of polynomial: 3

Output file: DGd°3_X

--- Number of rows: 15, column: 1

This file contains the x values used for polynomial estimation

Output file: DGd°3_Yest

--- Number of rows: 15, columns: 14

This file contains the polynomial model for each variable (columns),

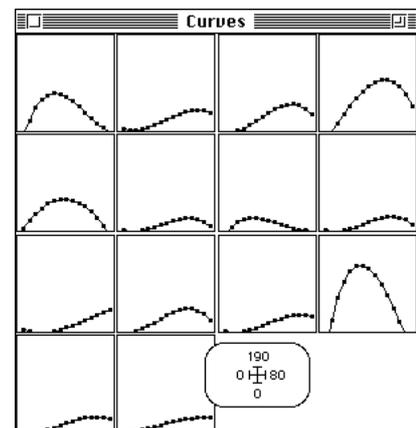
Représenter le modèle par Curves : Lines :

Lines

X file (default = 1, 2, 3, ..., n) 15 1

X file column number (default = 1)

Y file (no default) 15 14



On peut, évidemment, superposer données et modèles par deux utilisations de cette option.



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

L'option fonctionne comme UniVarReg : Polynomial -> Model, mais fait l'estimation en des points régulièrement répartis au lieu de les faire pour les valeurs observées de la variable prédictrice. Dans UniVarReg : Polynomial -> New data, on fait l'estimation en des points définis par des valeurs supplémentaires de la variable prédictrice.

UniVarReg : Polynomial -> Error



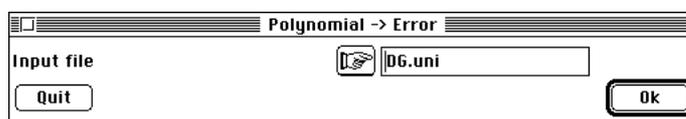
Méthode exploratoire de prédiction univariée par régression polynomiale : utilitaire pour le choix du degré du polynôme.



L'option édite les erreurs et les paramètres des régressions polynomiales indépendantes de chacune des variables à expliquer sur la variable explicative.



L'option utilise une seule fenêtre de dialogue :



 Fichier des paramètres créé par UniVarReg : Initialize.



Utiliser l'exemple introduit dans la fiche de UniVarReg : Initialize :

```
Polynomial model
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
-----
Predicted variable n° 1
-----
d°|Error/Vari|  a0  |  a1  |  a2  |  a3  |  a4  |  a5  |  a6  |
-----
1| 9.979e-01|4.311e+01|-1.455e+00|
2| 4.243e-01|6.780e+01| 3.232e+00|-2.469e+01|
3| 3.423e-01|7.116e+01|-1.525e+01|-2.990e+01| 9.724e+00|
4| 2.050e-01|8.369e+01|-1.928e+00|-6.983e+01|-1.138e+00| 1.471e+01|
5| 1.999e-01|8.504e+01|-8.789e+00|-7.573e+01| 9.793e+00| 1.774e+01|-3.300e+00|
6| 1.863e-01|8.850e+01| 1.896e-01|-9.924e+01|-8.419e+00| 4.206e+01| 3.550e+00|-6.259e+00|
-----
...
Predicted variable n° 14
-----
d°|Error/Vari|  a0  |  a1  |  a2  |  a3  |  a4  |  a5  |  a6  |
-----
1| 5.201e-02|1.289e+01| 9.617e+00|
2| 4.741e-02|1.358e+01| 9.747e+00|-6.812e-01|
3| 2.891e-02|1.308e+01| 1.245e+01| 8.177e-02|-1.423e+00|
4| 2.738e-02|1.349e+01| 1.289e+01|-1.217e+00|-1.777e+00| 4.785e-01|
5| 2.738e-02|1.348e+01| 1.294e+01|-1.171e+00|-1.861e+00| 4.551e-01| 2.546e-02|
6| 2.469e-02|1.301e+01| 1.171e+01| 2.044e+00| 6.309e-01|-2.872e+00|-9.117e-01| 8.564e-01|
-----
```

L'option édite pour chaque variable à prédire et chaque degré de 1 à 6 l'erreur relative et les paramètres estimés du modèle polynomial (du type $a_0+a_1x+a_2x^2+\dots+a_kx^k$). Le degré du polynôme à utiliser varie dans le cas présent d'une chronique à l'autre.

UniVarReg : Polynomial -> Model



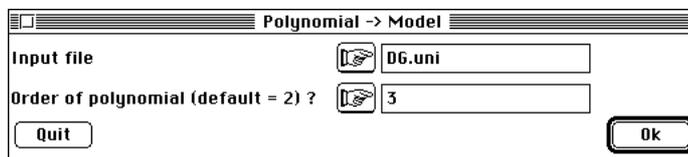
Méthode exploratoire de prédiction univariée par régression polynomiale.



L'option conserve sur disque les valeurs représentées directement dans CurveModels : Polynomial.



L'option utilise une seule fenêtre de dialogue :



 Fichier des paramètres créé par UniVarReg : Initialize.

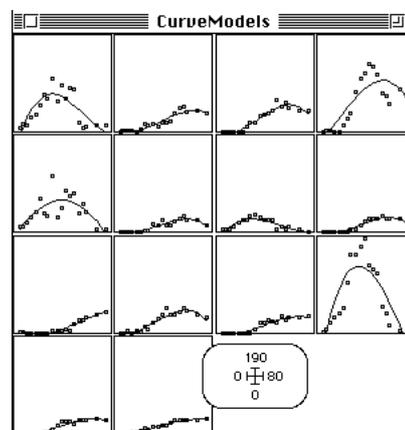
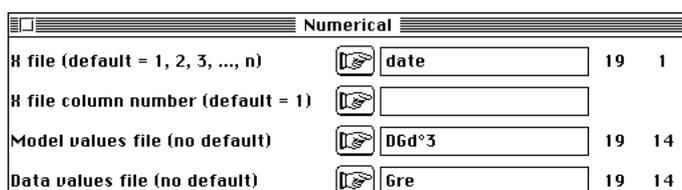
 Degré du polynôme utilisé (par défaut, c'est 2).



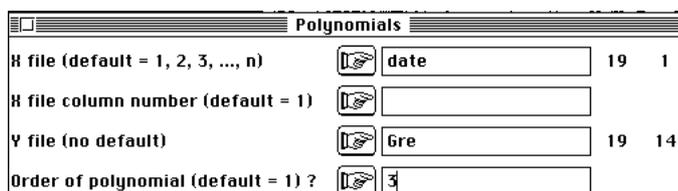
Utiliser l'exemple introduit dans la fiche de UniVarReg : Initialize :

```
Polynomial model
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Order of polynomial: 3
Output file: DGd°3
--- Number of rows: 19, columns: 14
This file contains the polynomial model for each variable (columns),
```

Représenter données et modèles par CurveModels : Numerical :



La figure peut être obtenue directement par CurveModels : Polynomial :



Pour une exploration purement graphique l'option CurveModels : Polynomial suffit. Pour préciser l'erreur associée au nombre de voisins, on utilisera le présent module qui permet en outre des estimations soit en un nombre limité de points (UniVarReg : Polynomial -> Curves) soit pour des points supplémentaires (UniVarReg : Polynomial -> New Data).

UniVarReg : Polynomial -> New Data



Méthode exploratoire de prédiction univariée par régression polynomiale.



Lissage de courbes par régression polynomiale, cf. UniVarReg : Polynomial -> Curves et UniVarReg : Polynomial -> Model). L'option permet l'extension de la prédiction à des valeurs supplémentaires de la variable prédictive.



L'option utilise une seule fenêtre de dialogue :

Polynomial -> New Data

Input file D6.uni

Order of polynomial (default = 2) ? 3

values: supplementary data date 19 1

Selected column (default = 1)

Quit Ok

Fichier des paramètres créé par UniVarReg : Initialize.

Degré du polynôme utilisé (par défaut, c'est 2).

Fichier des valeurs supplémentaires de la variable prédictive.

Numéro de la colonne utilisée (par défaut, c'est la première).



Utiliser l'exemple introduit dans la fiche de UniVarReg : Initialize :

```
Polynomial model
Y file: Gre
--- Number of rows: 19, columns: 14
X coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
Order of polynomial: 3
X supplementary coordinate file: date
--- Number of rows: 19, columns: 1
--- Selected column: 1
-----
Output file: dated°3_ls
--- Number of rows: 19, columns: 14
This file contains the polynomial model for each variable (columns),
-----
```

Vu le choix du fichier des valeurs supplémentaires, le contenu du fichier DGd°3_ls est évidemment identique au contenu du fichier DGd°3 créé par UniVarReg : Polynomial -> Model.