

Détermination du polynôme des moindres carrés par une méthode algébrique

FLORIAN PLATEL
www.metgen.org

1. Détermination du polynôme

Soient n un entier strictement supérieur à 1 et $F = \{(x_i, y_i) \in \mathbb{R}^2, i \in [1 \dots n]\}$, un ensemble de n observations et p un entier strictement inférieur à n .

L'on cherche le polynôme P à coefficients réels noté :

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_px^p,$$

pour lequel $\sum_{i=1}^n (P(x_i) - y_i)^2$ est minimum.

En posant

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & \dots & x_n^p \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_p \end{pmatrix},$$

le problème posé revient alors à trouver un vecteur β minimisant la norme de $(X \times \beta - Y)$. On appelle $\hat{\beta}$ le vecteur minimisant cette norme (cf. figure 1).

Soit $D = \{X \times \beta, \beta \in \mathbb{R}^p\}$. De façon triviale, D est un sous espace vectoriel de \mathbb{R}^n . D'après le théorème de Pythagore, $\hat{\beta}$ est tel que le vecteur $(Y - X \times \hat{\beta})$ appartient à l'orthogonal de D .

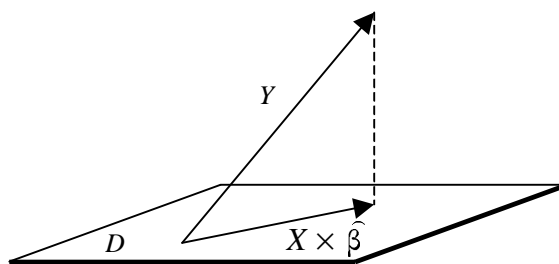


Fig. 1. – Projection du vecteur des observations sur l'espace vectoriel D .

Soit β un élément de \mathbb{R}^p . On note $(|)$ le produit scalaire euclidien. D'après ce qui précède

$$(X \times \beta | (Y - X \times \hat{\beta})) = 0,$$

et l'on a successivement :

$${}^t(X \times \beta) \times (Y - X \times \hat{\beta}) = 0,$$

$$\text{puis : } {}^t\hat{\beta} \times ({}^tX \times Y - {}^tX \times X \times \hat{\beta}) = 0.$$

Ceci étant vrai pour tout vecteur β , on obtient :

$${}^tX \times Y - {}^tX \times X \times \hat{\beta} = 0.$$

Si ${}^tX \times X$ est inversible, alors :

$$\hat{\beta} = ({}^tX \times X)^{-1} \times {}^tX \times Y. \quad (1)$$

2. Pondération des données

Les données étant souvent entachées d'incertitudes (a priori différentes) il est alors tentant d'accorder plus d'importance aux données ayant de faibles incertitudes, dont la connaissance est de ce fait plus assurée, qu'aux valeurs présentant de mauvaises incertitudes. Cela pourra conduire par exemple à minimiser l'influence de valeurs issues de conditions extrêmes, proches des limites des instruments de mesure.

Soit $\{w_i, i \in [1..n]\}$, un ensemble de n réels positifs ou nuls tels que pour $i \in [1..n]$, w_i représente le degré de confiance que l'on a pour le doublet (x_i, y_i) . Pour prendre en compte l'influence de la pondération dans l'approximation il suffit de multiplier chacune des composante de X et de Y par la racine carrée de la pondération qui lui est associée. Pratiquement, cela revient à construire la matrice diagonale $W_{1/2}$ avec les racines carrées des w_i sur la diagonale et des 0 ailleurs, puis à remplacer X par $W_{1/2} \times X$ et Y par $W_{1/2} \times Y$. La simplicité de cette prise en compte de la pondération résulte de la linéarité du formalisme matriciel employé.

Les calculs se simplifient et la formule (1) avec prise en compte de la pondération devient :

$$\hat{\beta} = ({}^t X \times W \times X)^{-1} \times {}^t X \times W \times Y, \quad (2)$$

avec :

$$W = \begin{pmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix}.$$

La difficulté est maintenant de choisir une pondération. Afin de tenir compte des incertitudes, on pourra prendre par exemple comme valeurs de pondération pour le doublet d'observations (x_i, y_i) avec $i \in \{1, \dots, n\}$:

$$w_i = \frac{1}{U_i^2}, \quad (3)$$

$$\sum_{j=1}^n \frac{1}{U_j^2}$$

U_k étant l'incertitude élargie sur y_k ($k \in \{1, \dots, n\}$).

Il convient néanmoins de garder à l'esprit que la pondération est ici présentée dans un cadre purement déterministe et non probabiliste.

3. Choix du degré du polynôme

Soit $n \geq 2$ le nombre de doublets numériques (x, y) . Le degré k du polynôme calculé pourra être compris entre 0 et $n - 1$. S'il est aisé d'éliminer les cas triviaux $k = 0$ et

$k = 1$, il est moins évident de déterminer à partir de quelle valeur de k l'approximation est optimale. Il n'existe en effet pas de méthode permettant de déterminer le degré du polynôme « optimum » dans le cas général.

On peut néanmoins remarquer que lorsque l'on augmente le degré k du polynôme, les résidus

$$R_i = |y_i - P_k(x_i)| \quad (i \in \{0, \dots, n\}),$$

sont de plus en plus petits. Toutefois, lorsque k augmente, la proximité du graphe de P aux points (x_i, y_i) s'accompagne souvent d'oscillations en dehors des points d'interpolation qui sont le plus souvent peu crédibles sur le plan physique. Cette remarque souligne l'importance de systématiquement utiliser un outil de représentation graphique tel que **PolyReg**.

Dans le cas où le modèle physique qui lie y_i à x_i est un polynôme (cas des relations entre la résistance d'une sonde de platine et la température par exemple) dont on connaît le degré on peut tracer

$$R(k) = \sum_{i=1}^n (y_i - P_k(x_i))^2$$

en fonction de k . Lorsque le degré « optimum » du polynôme sera atteint on observera alors une décroissance rapide de la courbe comme indiqué sur la figure 2.

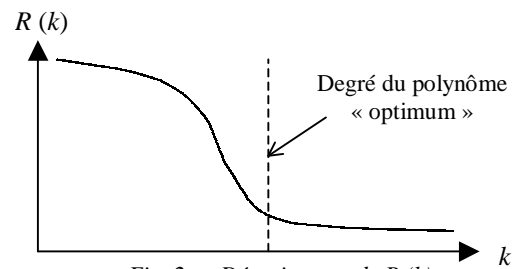


Fig. 2. - Décroissance de $R(k)$

4. Justesse « algébrique » du polynôme

La justesse du polynôme au sens algébrique dépend de la « proximité » entre $W_{1/2} \times X \times \hat{\beta}$ et $W_{1/2} \times Y$.

Cette proximité peut être caractérisée par le cosinus de l'angle formé entre $W_{1/2} \times X \times \hat{\beta}$ et $W_{1/2} \times Y$.

Celui-ci se calcule très simplement par la formule :

$$\cos(\theta) = \frac{(W_{1/2} \times Y \mid W_{1/2} \times X \times \hat{\beta})}{\|W_{1/2} \times Y\| \cdot \|W_{1/2} \times X \times \hat{\beta}\|},$$

ce qui s'écrit :

$$\cos(\theta) = \frac{({}^t Y \times W \times X \times \hat{\beta})}{\sqrt{{}^t Y \times W \times Y} \cdot \sqrt{{}^t \hat{\beta} \times {}^t X \times W \times X \times \hat{\beta}}}.$$

Le polynôme sera d'autant plus juste que la valeur de $|\cos \theta|$ sera proche de 1.

5. Cas du polynôme du premier degré

En conservant les notations précédentes, ce cas de figure revient à prendre $p = 1$. Dès lors les données pour le calcul du polynôme deviennent :

$$X = \begin{pmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}.$$

Le calcul de a_0 et a_1 est effectué dans un premier temps pour des données non pondérées à l'aide de la formule (1). On a alors successivement :

$${}^t X \times X = \begin{pmatrix} \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ n & \sum_{i=1}^n x_i \end{pmatrix},$$

puis, si cette matrice est inversible

$$({}^t X \times X)^{-1} = \frac{1}{\left(\sum_{i=1}^n x_i\right)^2 - n \cdot \sum_{i=1}^n x_i^2} \times \begin{pmatrix} \sum_{i=1}^n x_i & -\sum_{i=1}^n x_i^2 \\ -n & \sum_{i=1}^n x_i \end{pmatrix},$$

et d'autre part

$${}^t X \times Y = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix},$$

ce qui donne finalement :

$$({}^t X \times X)^{-1} \times {}^t X \times Y = \frac{1}{\left(\sum_{i=1}^n x_i\right)^2 - n \cdot \sum_{i=1}^n x_i^2} \times \begin{pmatrix} \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i y_i \end{pmatrix},$$

soit :

$$a_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2},$$

$$a_1 = \frac{n \cdot \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}.$$

On retrouve donc bien les formules habituelles de la droite des moindres carrés que l'on peut trouver par exemple dans la référence [1].

De même le calcul pour des données pondérées donne successivement :

$$({}^t X \times W \times X)^{-1} = \frac{1}{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i\right)^2} \times \begin{pmatrix} \sum_{i=1}^n w_i x_i^2 & -\sum_{i=1}^n w_i x_i \\ -\sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i \end{pmatrix},$$

$${}^t X \times W \times Y = \begin{pmatrix} \sum_{i=1}^n w_i y_i \\ \sum_{i=1}^n w_i x_i y_i \end{pmatrix},$$

$$({}^t X \times W \times X)^{-1} \times {}^t X \times W \times Y =$$

$$\frac{1}{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i\right)^2} \times \begin{pmatrix} \sum_{i=1}^n w_i y_i \sum_{i=1}^n w_i x_i^2 - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i x_i y_i \\ \sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i y_i \end{pmatrix}$$

soit :

$$a_0 = \frac{\sum_{i=1}^n w_i x_i^2 \sum_{i=1}^n w_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i\right)^2},$$

$$a_1 = \frac{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i \right)^2}$$

Ce qui correspond également aux formules usuelles de la droite des moindres carrés pondérée [1].

6. Application

L'étalonnage d'une chaîne de température dotée d'un thermocouple de type K a donné la table d'étalonnage reproduite dans le tableau 1.

Tableau 1
Table d'étalonnage de la chaîne de température.

Valeur Etalon (°C)	Valeur Appareil (°C)	Erreur (°C)	Incertitude (°C)
0,00	-0,5	-0,5	0,5
149,47	150,2	0,7	1,0
298,8	297,5	-1,3	1,5
500,4	499,9	-0,5	2,0
700,0	698,9	-1,1	2,0
899,8	897,3	-2,5	3,0
1 099,5	1 095,0	-4,5	4,0

Quel est le meilleur polynôme donnant l'erreur en fonction de la valeur appareil sur la plage allant de -0,5 °C à 1 095,0 °C ?

En appliquant la formule (3) on obtient les pondérations w_i reproduites dans le tableau 2.

Tableau 2
Pondérations calculées selon la formule (3).

Valeur Appareil (°C)	Erreur (°C)	Incertitude (°C)	Pondération w_i
-0,5	-0,5	0,5	0,65
150,2	0,7	1,0	0,16
297,5	-1,3	1,5	0,07
499,9	-0,5	2,0	0,04
698,9	-1,1	2,0	0,04
897,3	-2,5	3,0	0,02
1 095,0	-4,5	4,0	0,01

En utilisant les données du tableau 2 avec le logiciel **PolyReg**, on obtient les polynômes donnés dans le tableau 3.

Remarque : les polynômes ont été recopiés avec toute la « résolution informatique » même si cela n'a pas de signification physique.

De prime abord, l'on pourrait être tenté de penser que plus le degré est élevé, meilleur est le polynôme. En fait c'est vrai mathématiquement mais ça ne l'est pas physiquement : le polynôme est meilleur aux points des observations, mais la courbe obtenue semble s'éloigner du modèle physique en présentant des oscillations a priori non nécessaires.

Le tableau 4 donne quant à lui la valeur de $|\cos \theta|$ ainsi que le résidu absolu maximum et la somme des carrés des résidus. Les valeurs de $|\cos \theta|$ confirment que plus le degré est élevé, plus le vecteur des observations Y est proche de sa projection. Les valeurs des sommes des carrés des résidus qui ont été représentées graphiquement sur la figure 3 montrent que si les données suivent une fonction polynomiale, on peut alors penser à l'incertitude de mesure près qu'il s'agit avec une assez bonne approximation d'un polynôme de degré 2.

Tableau 3
Polynômes des moindres carrés.

Degré	Polynôme
1	P [X] = -0,260445881348811 -0,00166170000512021 * X
2	P [X] = -0,418661079050611 +0,00295009775245173 * X -6,04266173927926E-6 * X^2
3	P [X] = -0,438367096842552 +0,00450383261036768 * X -1,14142338731453E-5 * X^2 +3,96284284705524E9 * X^3
4	P [X] = -0,469274577204221 +0,013984577806196 * X -7,33445930543909E-5 * X^2 +1,12159881037947E-7 * X^3 -5,50253420454774E-11 * X^4
5	P [X] = -0,478223698412436 +0,0344480627341301 * X -0,000269648771736598 * X^2 +6,911565376685E-7 * X^3 -7,20572032919988E-10 * X^4 +2,60560884204014E-13 * X^5
6	P [X] = -0,466793289903802 +0,0660773058431544 * X -0,000671003117527796 * X^2 +2,41594058233986E-6 * X^3 -3,99212673255854E-9 * X^4 +3,08134594997113E-12 * X^5 -9,02228889724659E-16 * X^6

Tableau 4
Cosinus et résidus.

Degré (k)	$ \cos \theta $	Résidu (Max abs)	$R(k)$
1	0,694	2,420	8,689
2	0,837	1,224	2,233
3	0,843	1,296	2,664
4	0,902	1,022	2,772
5	0,972	0,665	0,969
6	1,000	0,000	0,000

L'utilisation d'un outil graphique tel que *PolyReg* révèle ici tout son intérêt. En effet, sur les figures 4 à 9 en annexe, l'on a représenté les courbes données par *PolyReg* et l'on « voit bien » que pour les degrés supérieurs à 4 les courbes sont « peu crédibles ».

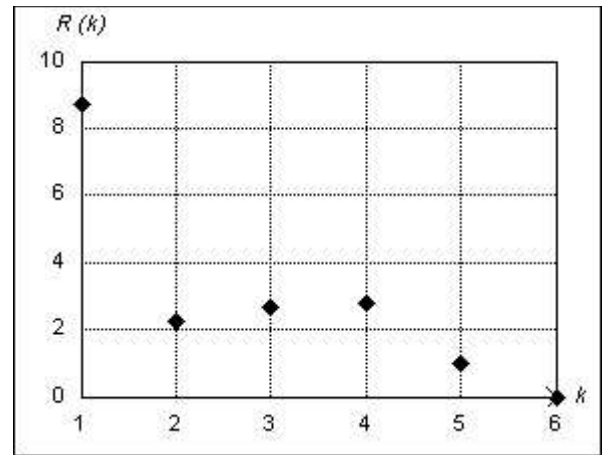


Fig. 3. – Somme des carrés des résidus en fonction du degré du polynôme.

Référence

- [1] TAYLOR J., "An introduction to Error Analysis", 2^e édition, *University Science Books*, 1997.

ANNEXE
Représentation graphique des polynômes

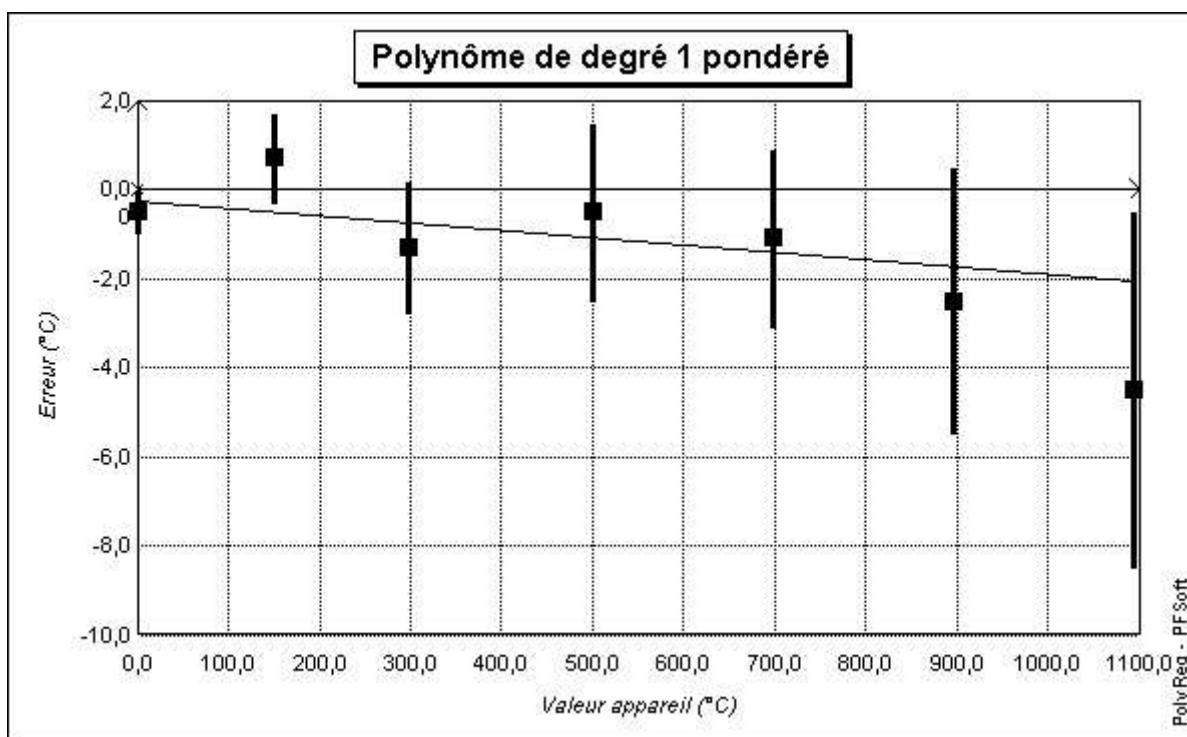


Fig. 4. – Modélisation par un polynôme de degré 1.

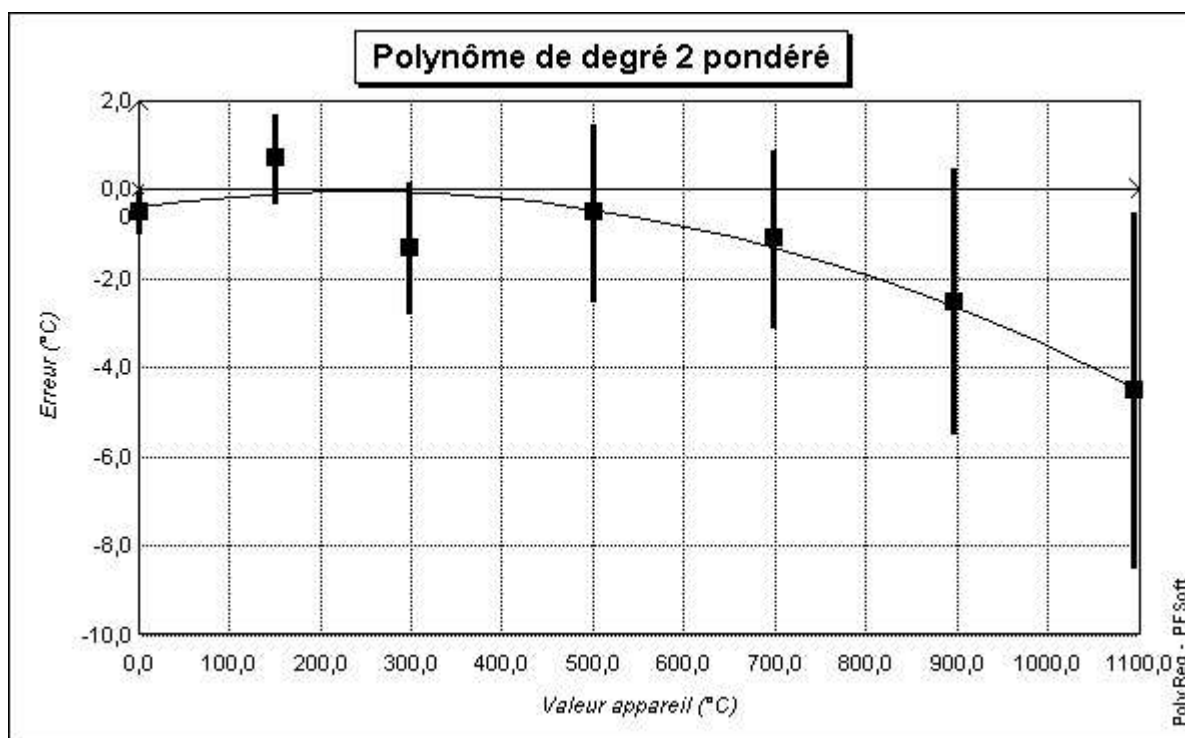


Fig. 5. – Modélisation par un polynôme de degré 2.

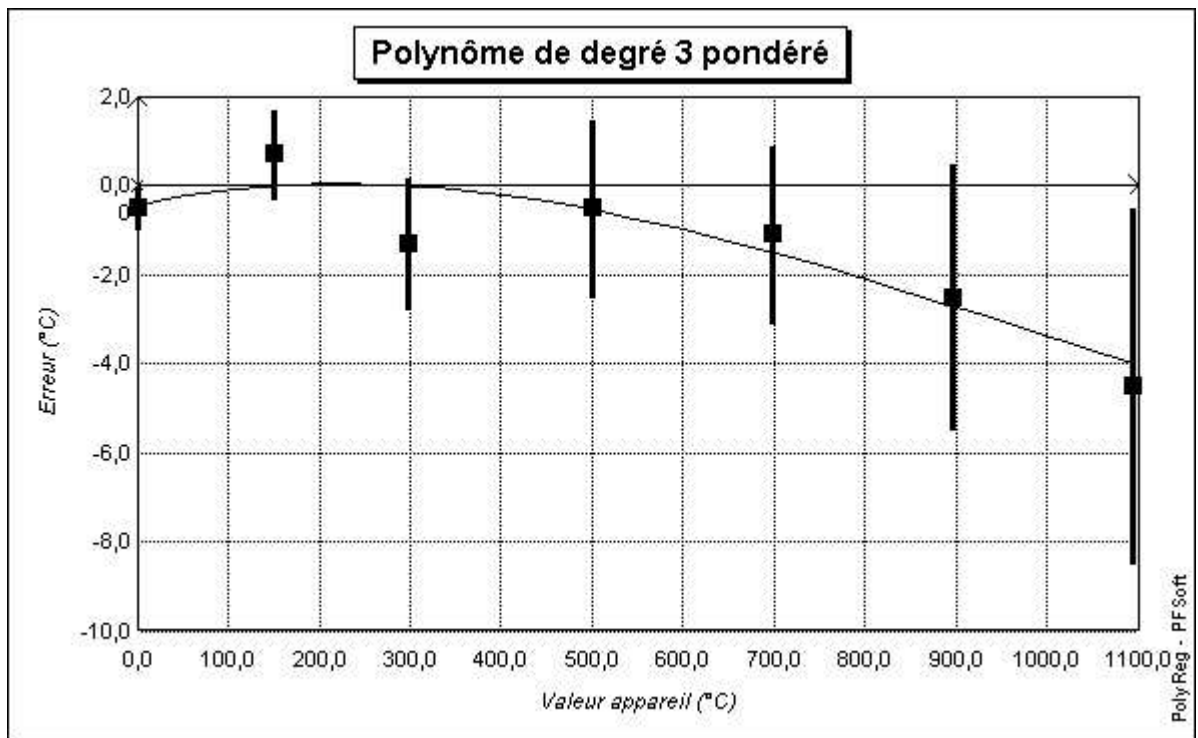


Fig. 6. – Modélisation par un polynôme de degré 3.

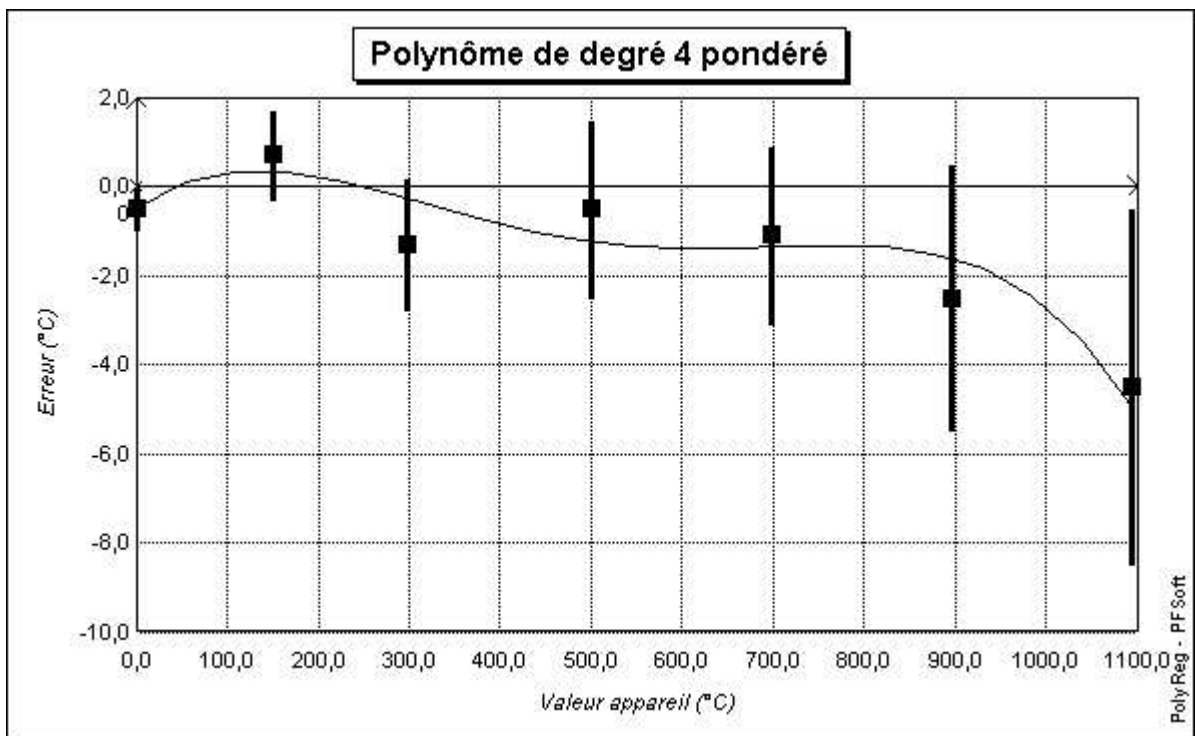


Fig. 7. – Modélisation par un polynôme de degré 4.

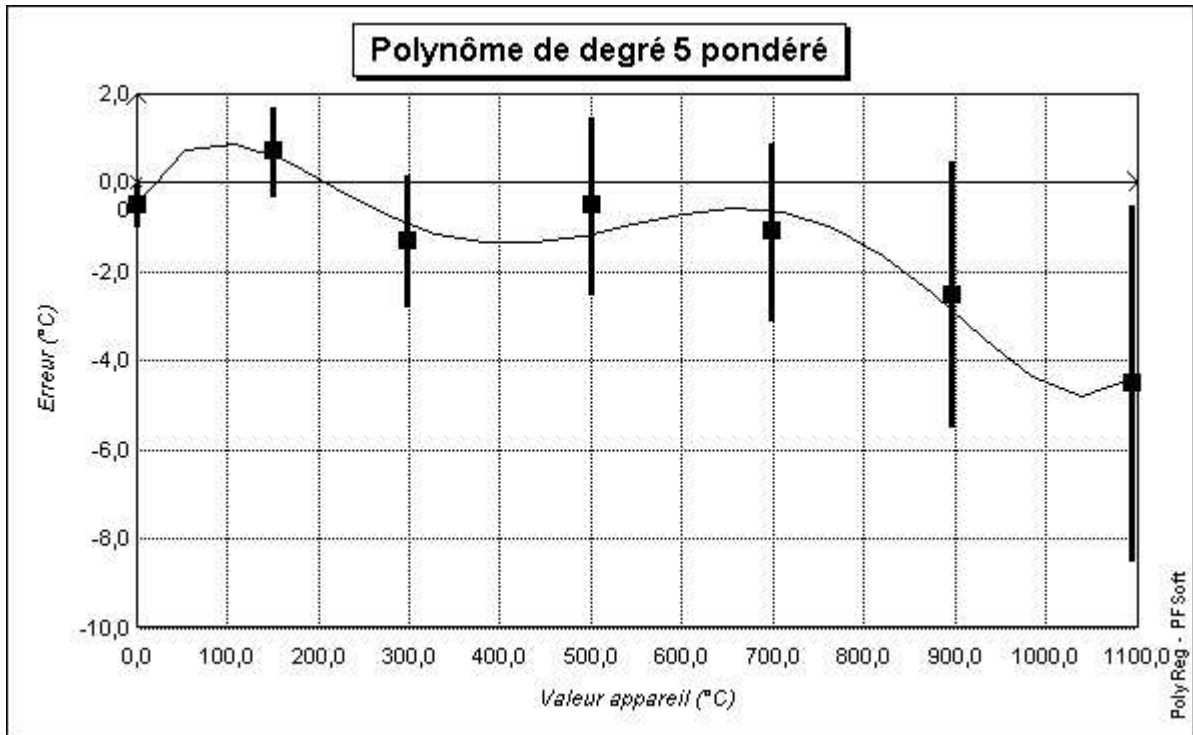


Fig. 8. – Modélisation par un polynôme de degré 5.

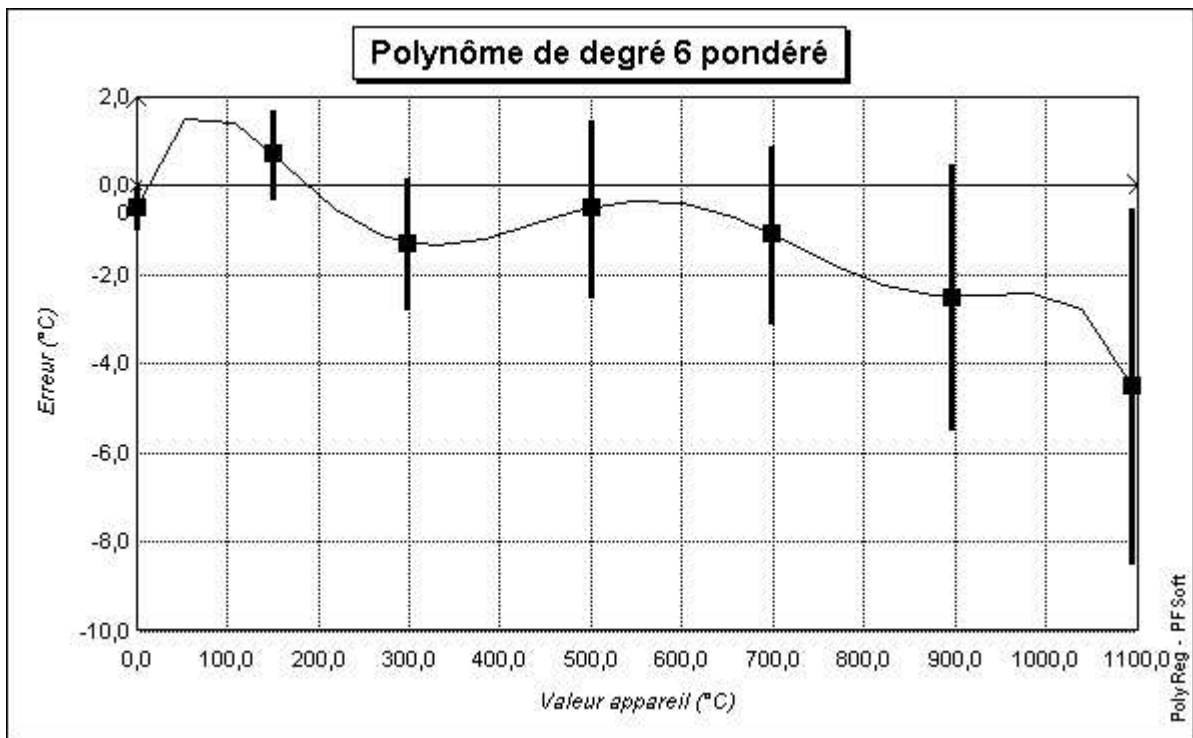


Fig. 9. – Modélisation par un polynôme de degré 6.