

REGPOLY : Un logiciel pour estimer la fonction d'étalonnage et calculer un résultat de mesure à partir d'une indication

Catherine Yardin^a

LNE, Service Mathématiques et Statistiques, 75724 Paris cedex 15, France

Résumé. Regpoly estime une fonction d'étalonnage polynômiale avec des méthodes de régression pondérée (GLS). Elles prennent en compte à des degrés divers l'hétéroscédasticité et la corrélation des variables. Regpoly calcule avec la fonction estimée une nouvelle valeur de y en mode direct et une nouvelle valeur de x en mode inverse. Il propage à cette valeur prévue, l'incertitude des variables en entrée et le cas échéant, celle du prédicteur. Regpoly est une interface graphique utilisable aisément. Les données d'entrée sont soigneusement vérifiées et notamment la structure des incertitudes. Regpoly comprend des outils statistiques et graphiques d'aide à l'interprétation des résultats. Le logiciel est diffusé avec un recueil d'exemples issus de cas concrets analysés au laboratoire.

Abstract. Regpoly estimates a polynomial calibration function with weighted regression methods (generalized least squares). They take differently into account heteroscedasticity and correlations in variables. Regpoly uses the estimated function to evaluate a new value for y (direct evaluation) and a new value for x (inverse evaluation). It propagates uncertainties associated with input variables and sometimes with the predictor to the new value. Regpoly is a graphical interface simple to use. The input data are scrupulously checked and especially the uncertainties structure. Regpoly incorporates statistical and graphical tools to assist with the analysis of the results. The software is distributed with a collection of examples issued from real cases analyzed in LNE.

1 Introduction

Le VIM [1] définit l'étalonnage comme « une opération qui, dans des conditions spécifiées, établit en une première étape une relation entre les valeurs et les incertitudes de mesure associées qui sont fournies par des étalons et les indications correspondantes avec les incertitudes associées, puis utilise en une seconde étape cette information pour établir une relation permettant d'obtenir un résultat de mesure à partir d'une indication ».

Ainsi, l'étalonnage consiste à déterminer le modèle de mesure (mathématique) qui convertit une indication en un résultat de mesure. Dans le cas d'un instrument de mesure utilisé sur une plage de mesure, le modèle d'étalonnage comporte une fonction mathématique du type $y = f(x)$ où la fonction f diffère de la fonction constante. Cette fonction dépend du modèle utilisé (modèle du résultat de mesure, modèle de la correction)

et du choix des variables x et y (par exemple, x est l'indication et y le résultat de mesure ou inversement).

La fonction d'étalonnage est mentionnée dans le VIM sans être définie ; dans les publications, elle correspond souvent à la fonction f reliant l'indication y à la variable x qui représente avec ambiguïté la valeur étalon et le résultat de mesure. Dans cet article, la notion de fonction d'étalonnage est étendue à toute fonction f associée à un modèle d'étalonnage quel qu'il soit. Cette fonction est estimée en utilisant l'information (les valeurs des étalons et des indications) et la relation mentionnée à l'étape 1 de l'étalonnage. Cette fonction est utilisée dans le sens de l'estimation ou dans le sens inverse pour évaluer le résultat ciblé.

Le laboratoire d'étalonnage ou le métrologue doit donc effectuer deux types de calcul :

- estimer la fonction f
- utiliser la fonction f pour faire des prévisions

^a Corresponding author: catherine.yardin@lne.fr

Il est clair que ces calculs ne peuvent pas être menés sans avoir recours à un logiciel.

) qui décrit son contenu. Les rubriques sont présentées dans les paragraphes 3 et 4.

2 Un logiciel pour l'étalonnage

2.1 Le besoin

Sur le plan statistique, l'étalonnage présente les particularités suivantes :

- la fonction d'étalonnage est estimée avec des données qui comportent de l'incertitude sur les deux variables x et y ,
- la fonction d'étalonnage peut être utilisée dans le sens d'estimation ou dans le sens inverse,
- l'incertitude associée à un résultat prévu avec la fonction intègre les incertitudes de la fonction et du prédicteur

Il est rare de disposer de fonctionnalités adaptées dans les logiciels de calcul actuellement disponibles y compris les logiciels de statistiques.

Dans Regpoly, la fonction d'étalonnage est estimée avec des méthodes de régression pondérée qui intègrent la variation des incertitudes (hétéroscédasticité) dans le domaine d'étalonnage. Elle est utilisée, en mode direct ou en mode inverse, pour prévoir une nouvelle valeur. Regpoly propage à cette valeur, l'incertitude des variables en entrée.

Ce cœur du logiciel est complété par des outils de validation et d'aide à l'interprétation des résultats. Il s'agit de tests statistiques, de calculs complémentaires ou de représentations graphiques. Ces fonctionnalités sont nécessaires pour pallier au manque d'expérience sur le sujet.

L'estimation de la fonction est traitée au paragraphe 3 et son utilisation au paragraphe 4. Auparavant, la structure générale du logiciel est décrite.

2.2 Organisation générale de Regpoly

Regpoly est un programme exécutable développé avec le logiciel Matlab et compilé pour être utilisable hors de l'environnement Matlab. Il se compose de 2 interfaces graphiques : un écran d'accueil (Figure 1) et un écran de calcul (Figure 2). La navigation dans le logiciel s'effectue principalement en cliquant sur des boutons ou en sélectionnant une option dans une liste déroulante. Aucune compétence en programmation n'est requise.

L'écran de calcul regroupe l'ensemble des opérations liées à l'estimation et à l'utilisation de la fonction. Ces opérations sont organisées séquentiellement dans des rubriques numérotées : Données d'entrée (I), Degré du polynôme (II) , etc...

Ces rubriques sont activées au fur et à mesure de la progression dans le processus de calcul. Chaque rubrique comprend un bouton d'information (représenté par l'objet

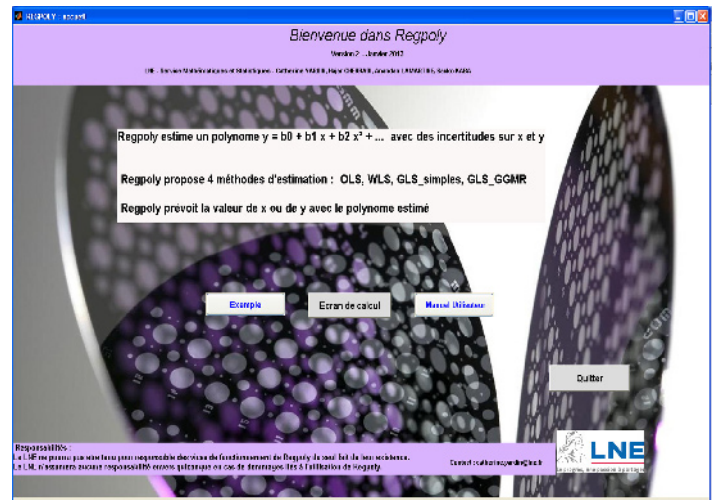


Figure 1. Ecran d'accueil de Regpoly

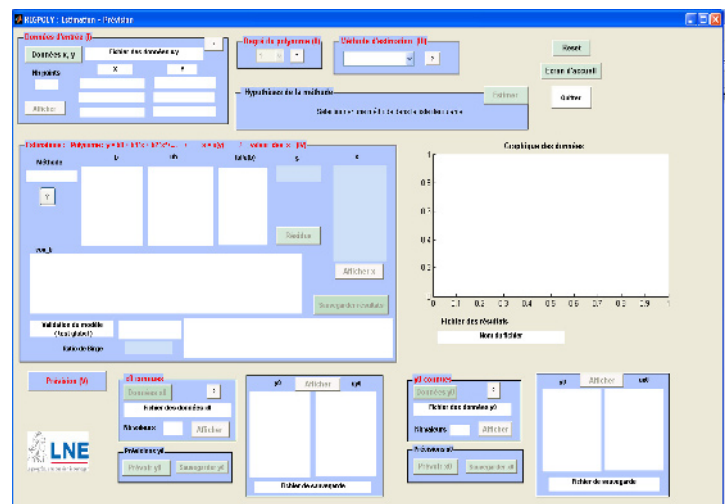


Figure 2. Ecran de calcul de Regpoly

Les données en entrée et en sortie du logiciel sont stockées dans des fichiers Excel (version Excel 2000). A chaque étape, Regpoly teste la validité des données d'entrée (valeurs importées des variables ou cohérence des choix effectués par l'utilisateur). En cas de non conformité, un message erreur, assez explicite, avertit l'utilisateur.

Regpoly est diffusé avec un manuel utilisateur et une recueil d'exemples. Ces exemples sont issus de [2] et de cas concrets analysés au laboratoire. Ils concernent divers domaines de la métrologie, des fonctions d'étalonnage reliant des grandeurs de nature identique ou différente, des calculs de prévision en mode direct ou en mode inverse, des études sur l'influence des incertitudes et notamment des covariances qui sont progressivement introduites dans le système.

3 Estimation de la fonction d'étalonnage

Regpoly estime la fonction $y = f(x)$ sans tenir compte du sens d'utilisation de la fonction. Cette opération concerne les rubriques (I) à (IV) du logiciel.

3.1 Données d'entrée (I)

Il s'agit des valeurs avec les incertitudes associées des variables x et y . Elles proviennent de l'étape 1 de l'étalonnage. Par exemple, les valeurs des étalons et les indications de l'instrument étalonné, y compris leurs incertitudes. Et le cas échéant, les covariances enregistrées dans une matrice de variances-covariances.

En pratique, cette étape nécessite une analyse préalable et parfois longue des incertitudes de manière à déterminer clairement l'incertitude associée à chaque variable. L'évaluation des covariances peut se révéler difficile.

Cette étape n'est pas prise en charge par le logiciel. Cela n'exclut pas la possibilité de tester différentes hypothèses de corrélation.

Les données d'entrée sont vérifiées et parfois corrigées. Une variable peut être de 2 types : avec incertitude, sans incertitude. Les vérifications sur les incertitudes concernent :

- le caractère non nul du vecteur des incertitudes associé à une variable (autrement la variable est du type "sans incertitude"),
- la symétrie et l'inversibilité de la matrice de variances-covariances,
- le niveau des covariances avec le cas échéant, un message informant l'utilisateur lorsque les valeurs sont trop corrélées.

Regpoly ne traite pas le cas des répétitions.

3.2 Degré du polynôme (II)

Dans Regpoly, la fonction d'étalonnage est un polynôme de degré k d'équation :

$$y = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k \quad (1)$$

L'utilisateur choisit le degré k et Regpoly crée les termes x^2, \dots, x^k . Le polynôme (1) est toujours estimé avec le terme constant.

Pour calculer les résidus (écarts entre le polynôme estimé et les données réelles) et réaliser les tests statistiques de validation, la valeur de k ne doit pas excéder $(n - 2)$ où n est le nombre de couples (x,y) des données d'entrée ; le maximum étant fixé à $k = 6$. Cette valeur est jugée suffisante pour représenter un grand nombre de fonctions instrumentales.

Il n'est pas possible d'inverser le rôle des variables x et y , à moins de modifier le fichier des données d'entrée.

3.3 Méthodes d'estimation (III)

Regpoly propose 3 méthodes d'estimation type moindres carrés pondérés : WLS, GLS_simples et GLS_GGMR. Ces méthodes associent aux données des poids qui dépendent des incertitudes des variables x et y .

Les WLS (weighted least squares) traitent du cas où x est connue sans incertitude et y est connue avec une incertitude qui varie dans le domaine d'étalonnage. Les valeurs de y ne sont pas corrélées.

Les GLS_GGMR (generalised least squares_Generalised Gauss Markov regression) considèrent que les deux variables sont connues avec une incertitude variable. Les valeurs de x et de y peuvent être corrélées.

Les GLS_simples (generalised least squares) concernent les deux cas : x avec incertitude, x sans incertitude. Dans le premier cas, les GLS_simples approximent de manière simple l'estimateur des GLS_GGMR. Dans le deuxième cas, ils généralisent les WLS en tenant compte des corrélations des valeurs de y .

Compte tenu des incertitudes associées aux données de l'étalonnage, la méthode GLS_GGMR est la plus adaptée. Elle est symétrique au sens où elle estime le même polynôme quel que soit le sens d'écriture du modèle.

Les WLS peuvent être utilisés dans le cas où l'incertitude sur une des variables est jugée négligeable. Lorsqu'il existe des corrélations, ils sont remplacés par les GLS_simples.

Regpoly comprend également la méthode OLS (ordinary least squares) dans laquelle seule la variable y a une incertitude constante et inconnue. Les OLS ne sont absolument pas adaptés au contexte de l'étalonnage. Mais ils sont à l'origine des méthodes décrites précédemment et ils apportent une aide non négligeable à l'interprétation des résultats. Les estimations OLS sont souvent comparées à celles obtenues avec les autres méthodes.

Regpoly détermine les coefficients du polynôme avec leur incertitude associée. Ce polynôme estimé permet de déterminer les valeurs $y_{modélisées}$ avec une incertitude. Lorsque la méthode GLS_GGMR est sélectionnée, Regpoly détermine également les valeurs de la variable x . A priori, Regpoly est un des seuls logiciels à fournir ce type d'information.

Les calculs évoqués précédemment sont réalisés avec les outils matriciels. Dans le cas où la variable x est sans incertitude, l'estimateur est défini analytiquement. Lorsque x et y sont connues avec une incertitude, l'estimateur des paramètres inconnus (coefficients du polynôme et/ou valeurs de la variable x) est défini par un système d'équations non linéaires résolu par calcul numérique. Dans la méthode GLS_GGMR, Regpoly utilise la méthode de Gauss-Newton et l'algorithme proposé dans [2,4].

3.4 Estimations (IV)

Cette rubrique traite des estimations des paramètres inconnus du modèle et des résultats liés à la validation de la fonction.

Regpoly restitue la plupart des résultats usuellement associés aux méthodes de régression : valeur des coefficients du polynôme, incertitude –type et matrice de variances-covariances associées, valeur et incertitude-type des x (GLS_GGMR), écart-type des résidus (OLS), statistique du Fisher (OLS), statistique du Chi-deux et ratio de Birge (WLS, GLS_simples et GLS_GGMR), statistique de Student, .. Le polynôme estimé et son intervalle de confiance sont représentés. Par ailleurs, il est possible de visualiser les résidus ou les résidus pondérés.

Regpoly comprend des outils de validation globale et partielle. Regpoly valide globalement le polynôme estimé en réalisant un test de Fisher (méthode OLS) ou un test du Chi-deux (autres méthodes). La statistique du Chi-deux est le rapport entre les écarts au polynôme et les incertitudes associées aux variables. En cas de non validation, Regpoly indique des facteurs susceptibles d'expliquer ce résultat. Par exemple, "incertitude en entrée insuffisante" dans le cas du Chi-deux. Regpoly calcule avec le Chi-deux, le ratio de Birge dont la valeur cible est 1 quel que soit le modèle. L'utilisateur peut également vérifier que les résidus ne sont pas trop élevés et qu'il n'existe pas de valeur aberrante.

La validation partielle est réalisée avec la statistique de Student calculée sur chaque coefficient du modèle :

$$\frac{|b_k|}{u(b_k)} \quad (2)$$

Lorsque ce rapport est trop faible (inférieur à la valeur critique), le coefficient n'est pas significatif et le terme x_k correspondant doit être retiré du modèle. Parfois, un degré trop élevé du polynôme conduit à rendre non significatif l'ensemble des coefficients et diminuer le degré corrige ce défaut.

Il est conseillé d'estimer la fonction en augmentant au fur et à mesure le degré du polynôme.

4 Utilisation de la fonction d'étalonnage

La fonction est utilisée pour prévoir une nouvelle valeur de x ou de y . On désigne par le terme "prédicteur", la valeur utilisée en entrée du calcul et par le terme "prévision", la valeur déduite du calcul. L'utilisation de la

fonction d'étalonnage concerne la rubrique (V) du logiciel, avec deux sous-rubriques : "Prévisions y_0 " et "Prévisions x_0 ".

4.1 Prédicteurs (x_0 connues, y_0 connues)

Le prédicteur est indiqué avec une incertitude associée, qui peut être nulle. Par simplicité, Regpoly effectue le calcul de prévisions en groupant plusieurs prédicteurs supposés non corrélés (groupe de x_0 connues ou de y_0 connues). Il vérifie que ces valeurs appartiennent au domaine d'étalonnage avec une tolérance de $\pm 10\%$. Ainsi, il est tenu compte de l'écart possible entre les plages de mesure théorique et pratique de l'étalonnage.

4.2 Prévisions

Une prévision y_0 est évaluée en utilisant (1) à la valeur x_0 :

$$y_0 = b_0 + b_1 x_0 + b_2 x_0^2 + \dots + b_k x_0^k \quad (3)$$

Regpoly vérifie que cette valeur correspond au domaine d'étalonnage avec une tolérance de $\pm 30\%$. En cas de dépassement, l'utilisateur en est informé.

L'incertitude associée aux coefficients estimés est propagée à la prévision y_0 . Le calcul est effectué en utilisant la loi de propagation des incertitudes.

$$u_f^2(y_0) = u_b^2(b_0 + b_1 x_0 + b_2 x_0^2 + \dots + b_k x_0^k) \quad (4)$$

Une prévision x_0 est évaluée en mode inverse. Ce calcul est souvent appelé "calibration". Il consiste à résoudre le système suivant :

$$\text{déterminer } x_0 \text{ tel que } y_0 = b_0 + b_1 x_0 + b_2 x_0^2 + \dots + b_k x_0^k \quad (5)$$

Le nombre de racines x_0 vérifiant le système dépend du degré du polynôme. Ces racines sont testées avant d'être validées. Regpoly élimine :

- les racines complexes
- les racines extérieures au domaine d'étalonnage avec une tolérance de $\pm 10\%$

Le calcul d'incertitude est également plus complexe que dans le calcul direct. Il repose sur l'utilisation des matrices jacobiniennes du polynôme par rapport aux coefficients (J_b) et aux termes en x (J_x) :

$$u_f^2(x_0) = J_x^{-1} J_b U_b J_b' J_x^{-1} \quad (6)$$

Regpoly affiche la liste des racines validées avec leur incertitude.

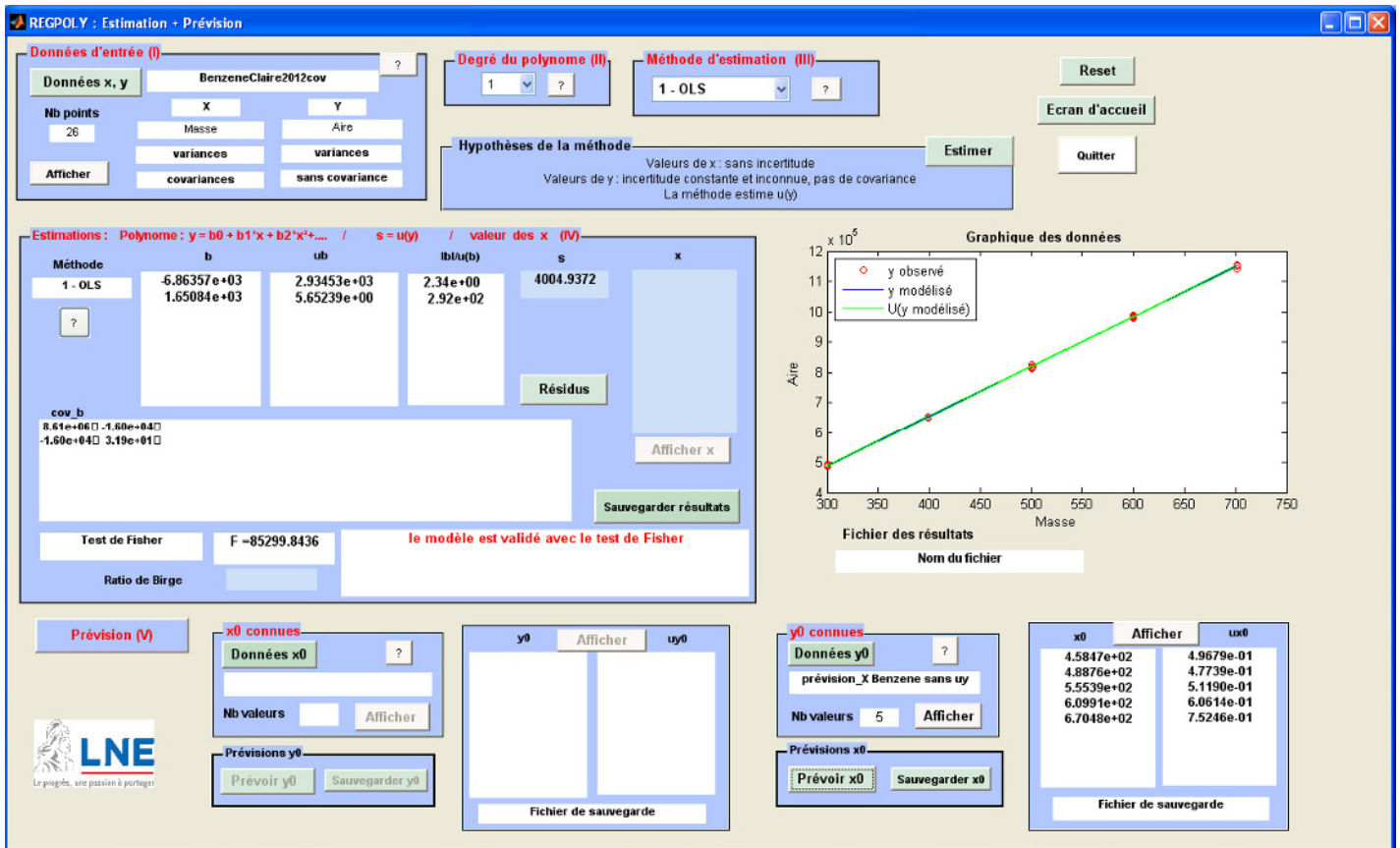


Figure 3. Estimations OLS

4.3 Incertitude totale

L'incertitude évaluée avec l'équation (4) ou l'équation (6) est l'incertitude due à la fonction estimée. En termes statistiques, ce calcul correspond à une prévision moyenne [3].

Le prédicteur peut comporter une incertitude non nulle. Par exemple, l'incertitude associée à l'indication évaluée par l'utilisateur de l'instrument. Regpoly propage à la prévision, l'incertitude du prédicteur.

Dans le cas d'une prévision y_0 , l'incertitude de x_0 est propagée en appliquant la formule de propagation des incertitudes à l'équation (3). Dans ce calcul, les différents termes en x_0 sont supposés non corrélés.

$$u_f^2(y_0) = u_{x_0}^2(b_0 + b_1 x_0 + b_2 x_0^2 + \dots + b_k x_0^k) \quad (7)$$

Dans le cas d'une prévision x_0 , l'incertitude de y_0 est propagée en utilisant la matrice jacobienne du polynôme en x_0 :

$$u_{y_0}(x_0) = \frac{u(y_0)}{J_{x_0}} \quad (8)$$

Les deux composantes d'incertitude – polynôme et prédicteur - sont supposées indépendantes et sommées quadratiquement.

5 Exemple

Il s'agit du dosage de la masse de Benzène par chromatographie en phase gazeuse réalisé au LNE (exemple 2 de [3]).

Les valeurs étalons des masses de Benzène sont très corrélées avec un coefficient de corrélation proche de 0,98. Les indications sont des aires de pic supposées non corrélées.

La gamme de mesure s'étend de 300 ng à 2000 ng et 26 tubes étalons sont analysés. La fonction d'étalonnage est une droite :

$$\text{Masse} = b_0 + b_1 \text{ aire de pic} \quad (9)$$

La droite est estimée avec les méthodes OLS et GLS_GGMR. Puis elle est utilisée pour déduire la valeur de masses de Benzène dans le domaine d'étalonnage. Les résultats sont présentés dans les figures 3 et 4. Les coefficients du polynôme et les valeurs prévues des masses diffèrent peu d'une méthode à l'autre. Par contre, les incertitudes des OLS sont très faibles.

6 Conclusions et perspectives

En estimant un polynôme jusqu'au degré 6, Regpoly permet d'évaluer la fonction d'étalonnage de divers instruments de mesure.

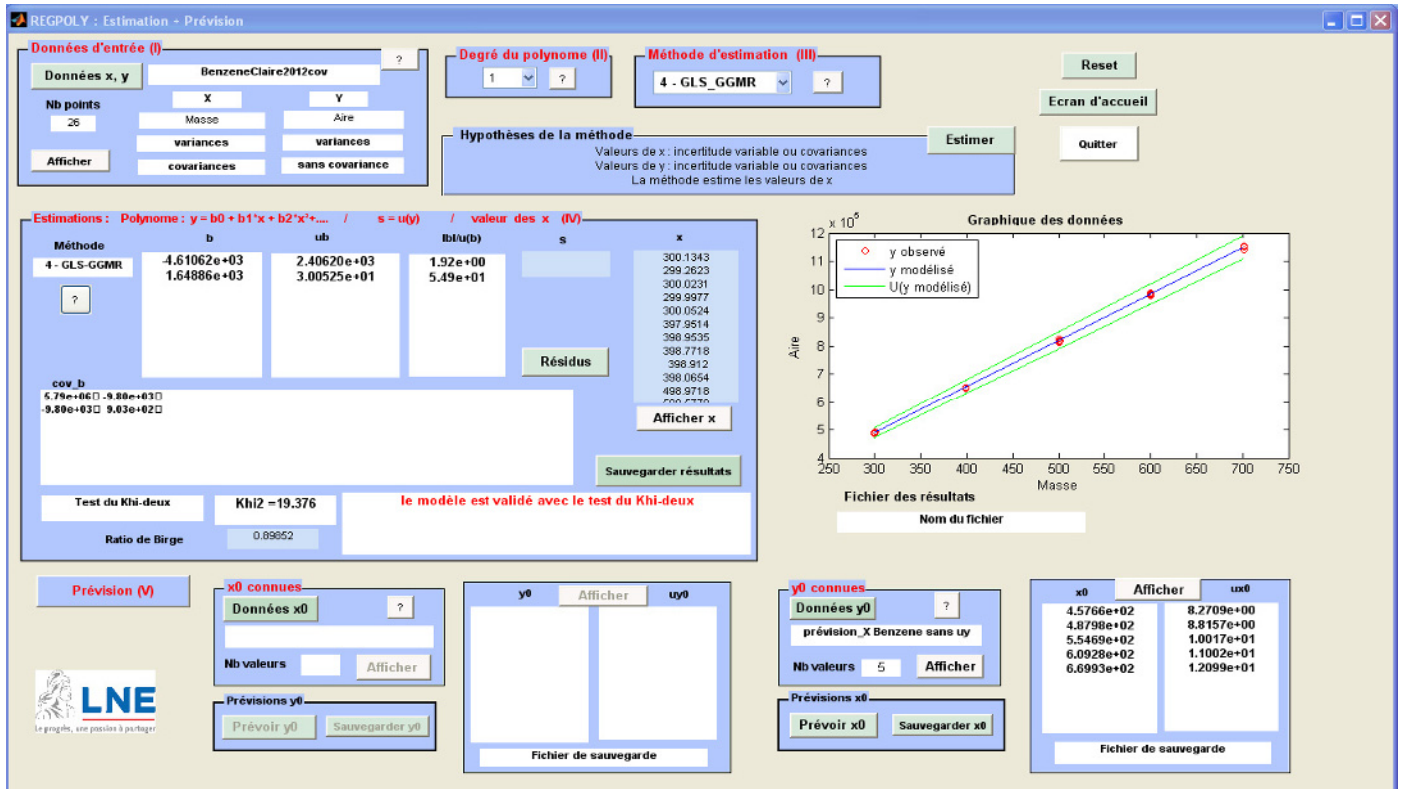


Figure 4. Estimations GLS_GGMR

- l'utilisation partielle des incertitudes des données d'entrée
- l'inversion des variables x et y dans le logiciel

Il couvre ainsi de nombreux domaines de la métrologie (ex : débitmétrie des gaz, analyse chimique, thermométrie, ..).

En utilisant la fonction estimée dans les deux sens et en propageant les incertitudes de la fonction et du prédicteur, Regpoly s'adresse aux laboratoires d'étalonnage et à leurs clients, utilisateurs finals des instruments.

Regpoly guide l'utilisateur dans l'interprétation et la validation de la fonction estimée. Le recueil d'exemples, empruntés au monde concret du laboratoire, confirme cette qualité.

Par ailleurs, Regpoly est utilisable en dehors du cadre de l'étalonnage. Il permet d'estimer une fonction avec des incertitudes sur les deux variables.

Pour l'avenir, des améliorations sont envisagées. Sur le plan technique, elles porteront principalement sur l'aide à l'interprétation des résultats avec la restitution des valeurs y modélisées et l'ajout de graphiques. Des modules spéciaux - Validation ou Interprétation des résultats, qui regrouperaient des fonctionnalités existantes, pourraient être créés.

Sur le plan ergonomique, elles faciliteront ou permettront :

- la saisie des covariances ou des corrélations
- la saisie ou la modification des données d'entrée directement dans le logiciel

Enfin, Regpoly pourrait intégrer un module de comparaison des résultats obtenus avec différentes méthodes comme les simulations de Monte-Carlo ou les techniques Bayésiennes.

7 Remerciements

L'auteur remercie Hajar Cherradi, Anandan Lamartine et Sacko Kaba, stagiaires au Service Mathématiques et Statistiques du LNE, pour leur participation à la réflexion sur les méthodes de régression pondérée et l'élaboration de Regpoly.

References

- 1 ISO/CEI GUIDE 99, Vocabulaire international de métrologie, (2007)
- 2 ISO/TS 28037, Determination and use of straight-line calibration functions, (2010)
- 3 C. Yardin, Estimer la droite d'étalonnage avec les moindres carrés généralisés et évaluer le résultat de mesure, Revue Française de Métrologie, **Volume 2012 – 3**, n°31, pp. 21 – 39, (2013)

4 M. Milton, P. Harris, I. Smith, A. Brown and B. Goody,
Implementation of a generalized least-squares method for
determining calibration curves from data with general
uncertainty structures, Metrologia, 43, S291-S298, (2006)