

# Génération d'observations pour la validation ou la comparaison de logiciels d'ajustement de mesures par moindres carrés

■ Stéphane DURAND

*Dans un précédent article, paru en 2012 (cf. [9]) dans la revue XYZ n°132, nous avons réalisé une validation du logiciel CoMeT par comparaison à d'autres logiciels d'ajustement de mesures topographiques faisant usage de la méthode des moindres carrés. Pour réaliser ces comparaisons, nous avons simulé différents jeux d'observations dans des réseaux de tailles variables, et comparé les résultats obtenus en termes de coordonnées ajustées et d'estimateur du facteur unitaire de variance. Dans le présent article, nous allons rappeler et discuter quelques méthodes classiques de génération d'observations et surtout, présenter plus en détails la méthode que nous avons exploitée à l'époque, qui possède quelques propriétés particulières.*

■ MOTS-CLÉS

Moindres carrés, génération d'observations, comparaison/ validation de logiciels

sentant le facteur unitaire de variance et  $Q$  la matrice de covariance, aussi appelée matrice des cofacteurs, sur les observations (i.e. les précisions a priori sur les mesures et les covariances entre les mesures).

Cette hypothèse de comportement normal des erreurs accidentelles confère des propriétés statistiques intéressantes à la solution des moindres carrés, permettant de construire différents tests d'hypothèse pour vérifier le bon déroulement du traitement, mais également de déterminer des régions de fiabilité et de confiance. De nombreux ouvrages traitent de ces notions, et le lecteur intéressé pourra se reporter par exemple à [6] ou [14].

Pour juger de la qualité d'un traitement par moindres carrés, deux quantités sont particulièrement utilisées. Si  $\hat{X}$  désigne la solution des moindres carrés, le vecteur des résidus de mesures correspond à une estimation des valeurs des erreurs accidentelles sur les mesures, et se définit par :

$$\hat{V} = B - A(\hat{X} - X_0) \quad (\text{éq. 3})$$

Le second indicateur de la qualité d'un traitement par moindres carrés est l'estimateur du facteur unitaire de variance, défini par :

$$\hat{s}_0^2 = \frac{\hat{V}^T P \hat{V}}{n - m}$$

Comme indiqué par exemple dans [6], il est courant de réaliser, à l'issue d'un traitement par moindres carrés, un test global de validation du traitement comme première étape dans un processus de recherche de fautes ou d'erreurs de pondération éventuelles

## Introduction

Dans les sciences de la mesure, et plus particulièrement dans les sciences associées au métier du topographe, la méthode d'ajustement la plus largement employée est la méthode des moindres carrés.

De manière synthétique, il s'agit de déterminer un ensemble de  $m$  paramètres inconnus, à partir d'un ensemble de  $n$  observations surabondantes. Formellement, on peut regrouper les  $m$  inconnues du problème dans un vecteur  $X$  et les  $n$  observations dans un vecteur  $L$  :

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, L = \begin{bmatrix} l_1 \\ \vdots \\ l_n \end{bmatrix}, m < n$$

Dans la méthode classique des moindres carrés, chaque mesure  $l_i$  est liée au vecteur  $X$  des inconnues par la relation :

$$l_i = f_i(X) + e_i$$

Expression dans laquelle  $f_i(X)$  représente le modèle fonctionnel, liant

mesure et inconnues, aussi appelé l'équation d'observation, et  $e_i$  représente la valeur de l'erreur associée à la mesure, inhérente au processus de mesure utilisé. En regroupant l'ensemble des erreurs de mesures dans un vecteur  $E$ , on obtient la relation matricielle suivante :

$$L = f(X) + E \quad (\text{éq. 1})$$

Étant donnée une valeur approchée  $X_0$  des paramètres, l'équation 1 est linéarisée au premier ordre sous la forme :

$$B = A(X - X_0) + E \quad (\text{éq. 2})$$

Expression dans laquelle  $B$  correspond à la différence entre les observations  $L$  et leurs valeurs approchées  $f(X_0)$ , calculées à partir des valeurs approchées de  $X$  et du modèle fonctionnel utilisé, et  $A$  correspond à la jacobienne en  $X_0$  de la fonction  $f$ .

Le modèle stochastique classiquement utilisé consiste à considérer que le vecteur  $E$  des erreurs sur les mesures est uniquement constitué d'erreurs accidentelles et suit une loi normale centrée, de matrice de covariance  $s_0^2 Q$ ,  $s_0^2$  repré-



dans les mesures. Ce test est basé sur la quantité :

$$s = (n - m) \frac{s_0^2}{s^2} \quad (\text{éq 3b})$$

En faisant l'hypothèse que le vecteur des erreurs accidentelles de mesure suit une loi normale centrée de covariance  $s_0^2 Q$ , désignée par  $N(0, s_0^2 Q)$ , la statistique suit une loi du chi-deux à  $n-m$  degrés de libertés. Dans le cas d'un test bilatéral, on fixe un seuil de confiance  $p$ , en général compris entre 80 et 99 %, et on détermine, à partir de la loi du chi-deux à  $n-m$  degrés de liberté, la région d'acceptation  $[\gamma_1, \gamma_2]$  du test. On vérifie alors que la quantité  $s$  appartient à l'intervalle  $[\gamma_1, \gamma_2]$ . Si cela est le cas, le test du chi-deux est validé au seuil  $p$ , autrement dit le vecteur des résidus semble globalement bien suivre la loi  $N(0, s_0^2 Q)$ . Si cela n'est pas le cas, le test du chi-deux échoue, autrement dit il existe des éléments du vecteur des résidus pour lesquels des fautes sont présentes ou une mauvaise pondération a été utilisée.

Un second test statistique couramment employé pour affiner la détection de fautes ou d'erreurs de pondération sur les mesures dans le cas d'erreurs normalement distribuées est celui dit de Baarda (cf. [2] ou [6]). Ce test fait l'hypothèse que le facteur unitaire de variance  $s_0^2$  est connu et est basé sur la notion de résidu normé  $w_i$  défini par :

$$\forall i = 1, \dots, n, w_i = \frac{\hat{v}_i}{s_0 \sqrt{q_{v_i v_i}}} \quad (\text{éq. 3c})$$

Dans cette expression,  $\hat{v}_i$  désigne l'élément  $i$  du vecteur des résidus  $\hat{V}$  et  $q_{v_i v_i}$  l'élément  $(i,i)$  de la matrice des cofacteurs  $Q_{\hat{V}}$  sur les résidus :

$$Q_{\hat{V}} = Q - A (A^T Q^{-1} A)^{-1} A^T$$

Pour des erreurs accidentelles suivant une loi  $N(0, s_0^2 Q)$ , chaque résidu normé est censé suivre la loi normale centrée réduite  $N(0, 1)$ . Pour un seuil de confiance  $p$  donné, il est possible de calculer un intervalle  $[-\gamma, \gamma]$  tel que l'on considérera qu'aucune faute et aucun problème de pondération n'affecte la mesure  $i$  si la valeur du résidu normé associé  $w_i$  se trouve dans cet intervalle. Par exemple, pour un seuil de confiance de 95 %, la valeur de rejet est  $\gamma = 1.96$ .

## Vérifier / comparer / tester des logiciels de traitements par moindres carrés

Dès lors que l'on cherche à évaluer un logiciel de traitement d'observations faisant usage de la méthode des moindres carrés, ou à comparer plusieurs logiciels entre eux, la première idée qui vient naturellement à l'esprit est d'utiliser un ou plusieurs jeux d'observations acquises spécifiquement pour ces tests, ou issues de campagnes anciennes. Après traitement de ces observations, on étudiera en particulier les valeurs ajustées des paramètres, mais également les valeurs des résidus et de l'estimateur du facteur unitaire de variance.

Utiliser des observations acquises sur le terrain a l'avantage d'assurer que les tests réalisés le sont avec des observations reflétant bien la réalité du terrain. Néanmoins, il n'est pas toujours possible de maîtriser entièrement les mesures réalisées sur le terrain, surtout sur des données anciennes : erreurs de mise en œuvre non documentées ou manque d'information sur les conditions de mesures, manque de précision sur les mesures de hauteur de voyant ou d'appareil, correction de réfraction pas assez fine... etc. De plus, il n'est pas toujours possible de réaliser sur le terrain l'acquisition du jeu de mesures idéal pour tester un logiciel : matériel non disponible, impossibilité pratique de mesure, manque de temps.

Pour tester ou comparer des logiciels d'ajustement par moindres carrés, il est possible de recourir à la génération de jeux d'observations. Un exemple relativement connu dans le monde de la topographie est celui du logiciel Bernese, de l'Institut Astronomique de l'Université de Bern (cf. [8]) qui propose un module GPSSIM assez complet pour la génération d'observations GNSS. En reprenant l'équation (1), il suffit pour cela de définir le modèle fonctionnel  $f$  à utiliser, de fixer les valeurs des paramètres  $X$  ainsi que les valeurs associées aux erreurs sur les mesures  $E$ . On peut alors calculer un ou plusieurs jeux de valeurs pour les observations  $L$ . Certes, les mesures ainsi obtenues ne rempla-

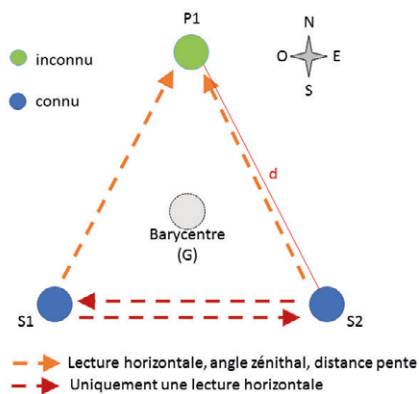
ceront pas des valeurs réelles acquises sur le terrain. Mais la génération d'observations permet de disposer de jeux de mesures pour des campagnes d'observations difficiles à réaliser en pratique. Elle autorise également la construction de jeux de données où l'on peut faire varier un paramètre en particulier, et uniquement ce paramètre.

## Génération d'observations

Dans le cas de l'article proposé en 2012 (cf. [9]), l'objectif était par exemple de comparer, à des fins de validation, le logiciel CoMeT (cf. [10]) avec d'autres logiciels d'ajustement de mesures tachéométriques. Un des points étudiés était la manière dont les différents logiciels tiennent compte de la sphéricité de la Terre et son influence sur les résultats obtenus (coordonnées, résidus).

Pour réaliser cette étude, nous avons généré des observations sur différents réseaux de points possédant une géométrie commune, constituée de deux points connus en coordonnées 3D, notés S1 et S2, et d'un point inconnu P1, répartis sur les 3 sommets d'un triangle équilatéral de barycentre G. La Figure 1 illustre la forme des réseaux et précise les types d'observations considérés dans chaque réseau. Différents jeux de mesures avaient été générés, en faisant varier la longueur du côté du triangle,  $d$ , de 25 m à 50 km.

Afin d'illustrer la présente contribution, nous allons reprendre cette idée d'un réseau de trois points formant un triangle équilatéral. Nous allons considérer comme barycentre les coordonnées RGF93 de la station permanente MAN2. Partant d'une taille  $d$  de réseau, nous pouvons alors définir les coordonnées théoriques des points du réseau, dans le repère local associé au barycentre G par rapport à l'ellipsoïde IAG-GRS80. Les coordonnées planes locales des points se déduisent de la forme du triangle et nous avons fixé les coordonnées verticales locales des points comme suit : 10 m pour S1, -10 m pour S2 et 0 m pour P1. Ce faisant, nous pouvons disposer de valeurs théoriques pour l'ensemble des paramètres  $X_{th}^d$  d'un réseau de taille  $d$ , en particulier les coordonnées cartésiennes



**Figure 1.** Forme générale des réseaux utilisés pour générer les jeux de mesures. Tous les réseaux ont en commun le barycentre G.

géocentriques RGF93 des points. Il est alors possible de calculer les valeurs théoriques des observations  $f(X_{th}^d)$ , en utilisant le modèle fonctionnel 3D Géodésique de CoMeT, en négligeant la réfraction sur les angles zénithaux, et en considérant une déviation de la verticale nulle en chaque point du réseau par rapport à l'ellipsoïde IAG-GRS80.

Afin de générer un jeu d'observations pour notre réseau de taille  $d$ , il nous reste, en suivant l'équation (1), à construire un vecteur d'erreurs sur les mesures  $E$ .

## Mesures générées sans erreur

Une première manière de construire le vecteur  $E$  est de le prendre nul. Dans ce cas, si on ajuste nos observations avec un logiciel utilisant le même modèle fonctionnel que celui ayant permis la génération des mesures, les paramètres ajustés correspondent exactement, au critère de convergence choisi près, aux valeurs théoriques de ces mêmes paramètres utilisés pour générer les observations. De plus, le vecteur des résidus, qui est une valeur estimée du vecteur des erreurs sur les mesures, correspond exactement à la valeur de  $E$  utilisée pour générer les mesures et est donc nul. Par conséquent, la valeur de l'estimateur du facteur unitaire de variance est également nulle.

Regardons de plus près les conséquences sur les tests statistiques classiquement utilisés à l'issue d'un

traitement par moindres carrés pour la recherche de fautes. Au niveau du test sur les résidus normés, la valeur du résidu étant nulle, le résidu normé est également nul et sa valeur est exactement au milieu de l'intervalle  $[-\gamma, \gamma]$ . Le test sur le résidu normé est donc toujours validé, quel que soit le seuil de confiance utilisé.

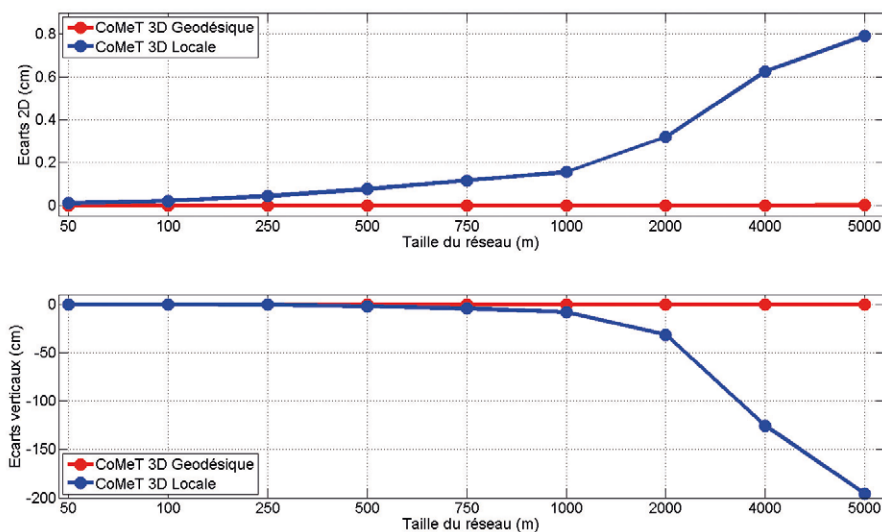
Concernant le test du chi-deux, les résidus étant tous nuls, l'estimateur du facteur unitaire de variance vaut zéro, et la quantité  $s$  de l'équation 3b sera obligatoirement nulle. Si on regarde la forme de la densité de probabilité de la loi du chi-deux, on remarque que, quel que soit le degré de liberté utilisé, les valeurs de la densité de probabilité sont supérieures ou égales à zéro. Ainsi, pour un seuil de confiance inférieur à 100 %, la valeur nulle de la quantité  $s$  sera toujours en dehors de la région d'acceptation  $[\gamma_1, \gamma_2]$  et notre test de validation globale échouera. Comme les mesures générées ne sont pas affectées d'erreurs (ni du coup de fautes), c'est donc un problème de pondération sur les mesures. En effet, puisque nous n'avons pas ajouté d'erreurs à nos mesures, nous devrions considérer une précision nulle sur nos observations... Ce qui n'est pas réaliste.

Ainsi, ne pas ajouter d'erreurs à nos mesures générées n'est pas optimal lorsque le même modèle fonctionnel est utilisé pour générer et ajuster les mesures.

Si on réalise un ajustement avec un logiciel utilisant un modèle fonctionnel différent, les résidus sur les observations seront représentatifs des différences entre les modèles fonctionnels. De même, les coordonnées ajustées obtenues seront différentes des coordonnées théoriques.

A titre d'illustration, nous avons généré avec le logiciel CoMeT, en utilisant son modèle fonctionnel 3D Géodésique, des observations sans erreur pour des réseaux de taille variant de 50 m à 5 km. Nous avons ajusté ces mesures d'une part en utilisant le même modèle fonctionnel et d'autre part en utilisant le modèle 3D Locale de CoMeT. Dans ce dernier, l'ajustement des mesures se fait dans le repère local associé au barycentre commun des réseaux, le point G, par rapport à l'ellipsoïde IAG-GRS80, en considérant une déviation de verticale nulle en G, et sans aucune prise en compte de la sphéricité de la Terre. Comme le montrent les résultats de l'article de 2012, nous devons donc nous attendre à des différences sur les résultats entre les deux calculs qui augmentent avec la taille du réseau, et surtout au niveau des écarts sur les coordonnées verticales des points.

La figure 2 montre les écarts en coordonnées planes et verticales par rapport aux coordonnées théoriques des points, obtenus lors de l'ajustement des jeux d'observations sans erreurs en utilisant les deux modèles



**Figure 2.** Écarts entre les coordonnées ajustées et les coordonnées théoriques en planimétrie et en vertical, en cm, pour différentes tailles de réseaux.

► fonctionnels. On constate comme attendu qu'en utilisant le même modèle fonctionnel pour l'ajustement et la génération des mesures, les coordonnées ajustées sont identiques aux coordonnées théoriques. Lorsque l'on utilise pour l'ajustement des mesures le modèle fonctionnel en 3D Locale de CoMeT, les écarts aux coordonnées théoriques augmentent avec la taille du réseau, résultat cohérent avec les différences de modèles fonctionnels : on obtient un écart sur les coordonnées verticales nul pour des réseaux de taille inférieure à 100 m, de 2 cm pour un réseau à 250 m et de presque 2 mètres à 5 km.

### Mesures générées avec un vecteur d'erreur suivant la loi normale

Certes, utiliser un vecteur d'erreurs nul correspond à la manière la plus simple de générer des observations. Cela conduit malheureusement à un échec systématique du test du chi-deux. Une seconde méthode consisterait donc à générer de manière aléatoire un vecteur  $E$  suivant la loi normale centrée en tenant compte des précisions a priori sur les observations.

De nombreuses méthodes peuvent être utilisées pour générer des nombres aléatoires suivant la loi normale centrée réduite, comme par exemple la méthode de Box-Muller (cf. [3]), qui réalise cette génération à partir de nombres aléatoires suivant une loi uniforme, sachant que tous les langages de programmation actuels proposent au minimum une fonctionnalité de génération de nombres aléatoires suivant la loi uniforme dans l'intervalle  $]0,1[$ . Il est ensuite possible de passer de la loi normale centrée réduite à n'importe quelle loi normale centrée d'écart-type donné.

Plutôt que de générer un seul vecteur d'erreurs et d'étudier les résultats de l'ajustement des mesures associées, il est courant de générer un lot de  $k$  vecteurs d'erreurs, et d'étudier statistiquement les résultats du traitement les  $k$  jeux de mesures associés, la valeur de  $k$  étant choisie de manière que les résultats aient une réelle signification

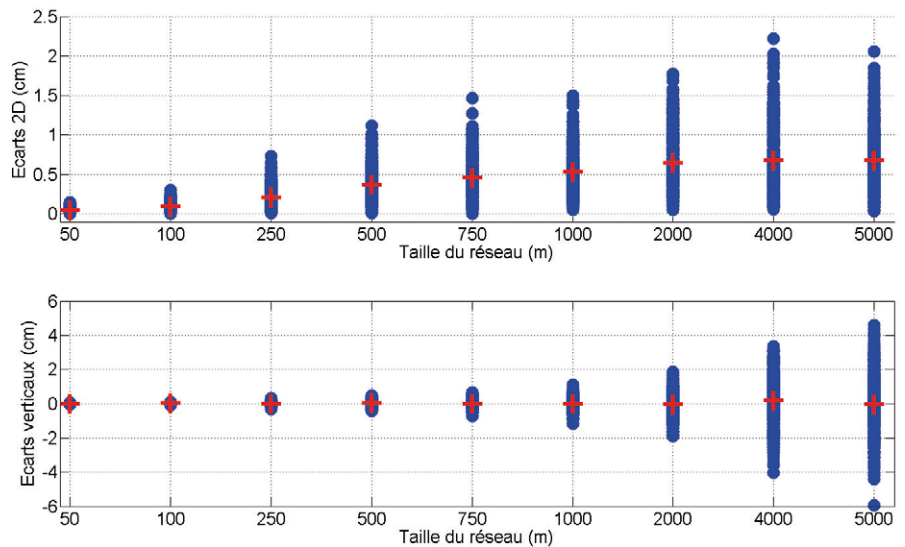


Figure 3. En bleu, les écarts entre coordonnées théoriques et coordonnées ajustées en planimétrie (en haut) et en vertical (en bas). En rouge, les écarts moyens sur les 100 jeux ajustés pour chaque taille de réseau.

statistique. On parle alors d'analyse par la méthode de la propagation des distributions (ou simulation Monte Carlo).

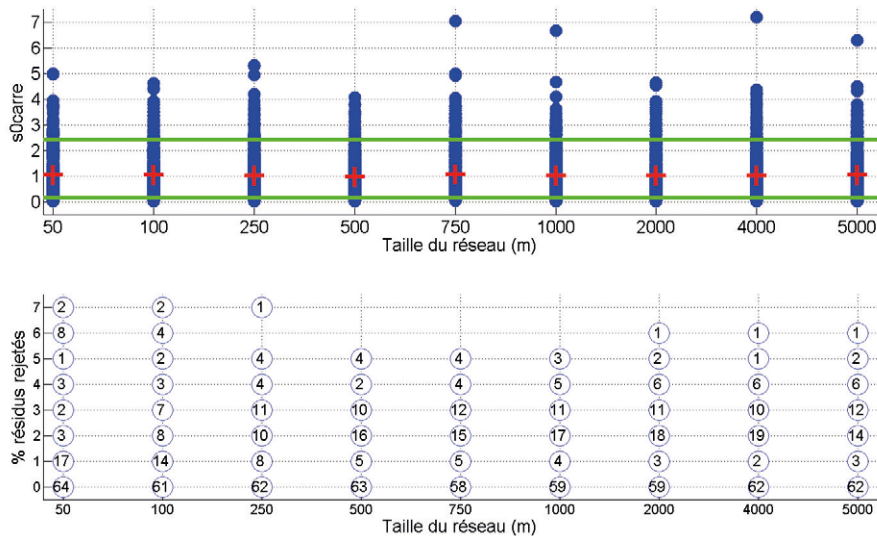
Afin d'illustrer nos propos, nous allons considérer une précision a priori sur les mesures angulaires de 0.3 mgon et sur les mesures de distances de 5 mm et générer, pour chaque taille de réseau, 500 jeux d'observations avec des vecteurs d'erreurs constitués de valeurs aléatoires issues de la loi normale centrée. Ce faisant, nous considérons donc connu le facteur unitaire de variance avec une valeur  $s_0^2=1$ .

La figure 3 représente les écarts en coordonnées planes et verticales, en centimètres, entre les coordonnées ajustées en utilisant le même modèle fonctionnel que celui utilisé pour la génération des mesures et les coordonnées théoriques du point P1, en fonction de la taille du réseau. Pour chaque taille de réseau, on précise également en rouge la valeur moyenne sur l'ensemble des 500 jeux générés.

Sur les écarts en coordonnées planes, on constate une évolution de l'étendue des valeurs qui augmente fortement avec la taille du réseau. L'étendue est d'environ 1.4 mm pour le réseau de taille 50 m puis de 14 mm pour celui de taille 750 m pour atteindre 20 mm pour le réseau de 5 km. On constate

également que la valeur moyenne de l'écart planimétrique augmente avec la taille du réseau pour atteindre 0.6 cm pour le réseau de taille 5 km. Sur les écarts en coordonnées verticales, on constate également une augmentation de l'étendue des valeurs en fonction de la taille du réseau, plus prononcée encore que sur les coordonnées planes (1 mm à 50 m, 13 mm à 750 m et 110 mm à 5 km). On remarque également que la valeur moyenne sur les 500 jeux de mesures générées par réseau augmente avec la taille du réseau (nulle à 50 m, autour de 0.5 mm à 5 km).

La figure 4 (en haut) représente les valeurs obtenues pour l'estimateur du facteur unitaire de variance sur chaque jeu de mesures générés. En rouge apparaît la valeur moyenne de l'estimateur du facteur unitaire de variance sur les 500 jeux générés, naturellement proche de l'unité. Sur cette figure sont également représentées en vert les bornes de la région d'acceptation du test du chi-deux, en considérant un degré de liberté de 3 et un seuil de confiance de 87 %. Sur cette figure, on constate une étendue des valeurs de l'estimateur du facteur de variance importante, allant de 5 pour le réseau de taille 50 m à 6.3 pour le réseau de 5 km. On constate que pour un certain nombre de jeux de mesures, entre 5 et 10 %, le test du chi-deux n'est pas validé au seuil de 87 %.



**Figure 4.** En haut, valeurs de l'estimateur du facteur unitaire de variance en bleu et bornes de la région d'acceptation du test du chi-deux en vert (seuil de 87 %). En bas, nombre de jeux de données possédant un certain nombre de résidus rejetés (test de Baarda) au seuil de 87 %.

Avec ce même seuil de confiance, la *figure 4* (en bas) représente pour chaque taille de réseau, le nombre de résidus ne validant pas le test de Baarda. On constate que pour environ 60 % des jeux de mesures générées, quelle que soit la taille du réseau, tous les résidus valident le test de Baarda. Il reste donc toujours environ 40 % des jeux de mesures où au moins un et au maximum 6 résidus sur les 8 possibles ne valident pas le test.

Ainsi, en générant des erreurs de manière aléatoire en suivant la loi normale centrée, et en tenant compte des précisions a priori sur les observations, il peut arriver que les erreurs générées influent fortement sur la solution ajustée. Le problème apparaît surtout lors de la génération aléatoire d'un vecteur d'erreurs de taille peu importante ( $n < 50$ ), comme cela est le cas dans nos réseaux exemples. En effet, même si les valeurs d'erreurs générées aléatoirement proviennent d'une loi normale centrée, la taille du vecteur fait que lorsqu'une valeur générée possède une valeur absolue un peu forte, elle n'est pas compensée par la présence en nombre important d'autres valeurs d'erreurs plus faibles, et a donc une forte influence sur la solution des moindres carrés. Cela peut conduire à un échec du test du chi-deux et au rejet de certains résidus normés.

### Mesures générées avec contraintes particulières

Dans l'article de 2012, notre objectif principal était de comparer, à des fins d'inter-validation, les résultats de différents logiciels sur des jeux de données générés issus de réseaux de taille variable, de 25 m à 50 km. Afin de faciliter l'analyse des résultats, tant en termes d'écart aux coordonnées théoriques que de valeurs de l'estimateur du facteur unitaire de variance, nous avons généré plusieurs jeux de mesures par taille de réseau, suivant la loi normale centrée, en tenant compte des précisions a priori sur les mesures et en utilisant le modèle fonctionnel 3D Géodésique de CoMeT. Pour chaque taille de réseau, nous avons ensuite sélectionné un seul jeu de mesures, pour lequel, lors de l'ajustement par le logiciel CoMeT avec le même modèle fonctionnel, la valeur de l'estimateur du facteur unitaire de variance était proche de 1, et les écarts entre coordonnées théoriques et ajustées étaient très faibles. Ce choix de jeux de mesures particuliers nous permettait de juger rapidement des différences entre les résultats obtenus par les différents logiciels. Ainsi, un logiciel possédant un modèle fonctionnel proche de celui utilisé par CoMeT pour générer les mesures devait produire des écarts faibles entre coordonnées ajustées et théoriques, et une valeur

de l'estimateur du facteur unitaire de variance proche de 1.

Entre 2014 et 2015, nous avons repris cette idée et développé de manière plus précise une méthodologie de génération d'erreurs telle que, lorsque les mesures générées sont ajustées avec le même modèle fonctionnel que celui utilisé pour les générer, elles répondent aux critères suivants :

(a) la solution ajustée correspond exactement aux paramètres théoriques utilisés pour générer les mesures  
 (b) quel que soit le seuil de confiance choisi, le test du chi-deux est validé  
 (c) étant donné un seuil de confiance  $p$ , tous les résidus normés valident le test de Baarda.

Pour répondre à la condition (a), rappelons que lors de la génération des observations, comme indiqué dans l'équation 1, on part de valeurs théoriques  $X_{th}$ , pour les paramètres et on forme un vecteur d'observations  $L$  par :

$$L = f(X_{th}) + E$$

Lors de l'ajustement des mesures, on utilise l'équation linéarisée (équation 2) et on procède par itération sur la valeur approchée des paramètres. Considérons pour simplifier que la valeur approchée  $X_0$  pour les paramètres corresponde aux paramètres théoriques  $X_{th}$  utilisés pour générer les observations. L'équation 2 devient alors :

$$\hat{X} - X_{th} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} E$$

Si l'on souhaite que les paramètres ajustés correspondent à leurs valeurs théoriques, il faut choisir le vecteur  $E$  non nul tel que :

$$(A^T Q^{-1} A)^{-1} A^T Q^{-1} E = 0 \quad (\text{éq. 4})$$

En effet, nous avons déjà discuté auparavant du cas  $E$  nul, qui n'était pas optimal pour notre besoin. En pratique, l'équation 4 exprime le fait que le vecteur  $E$  appartient au noyau, noté  $ker(K_1)$ , de l'application linéaire :

$$K_1 = (A^T Q^{-1} A)^{-1} A^T Q^{-1}$$

La matrice  $Q$  étant par construction symétrique définie positive, on peut trouver, par décomposition de Cholesky, une matrice  $R$  telle que  $Q^{-1} = R^T R$ . Posons alors  $K$  l'application linéaire définie par :

$$K = (A^T Q^{-1} A)^{-1} A^T R^T$$





Si le vecteur  $E$  appartient au noyau de  $K_1$  alors le vecteur  $y = RE$  appartient au noyau de  $K$ . Comme la matrice  $K$  est de taille  $n \times m$  et de rang  $m$  par construction, la dimension du noyau de  $K$  est  $n - m$ . En considérant  $\beta$  une base orthogonale du noyau de  $K$ , pour tout vecteur  $y$  de dimension  $n - m$ , la quantité  $\beta y$  appartient au noyau de  $K$ . Nous pouvons donc construire facilement un vecteur d'erreurs  $E$  répondant à l'équation 4 en choisissant un vecteur réel  $y$  de taille  $n-m$  et en formant :

$$E = R^{-1}\beta y \quad (\text{éq. 5})$$

Pour répondre à la condition (b), rappelons que le test du chi-deux est basé sur la quantité  $s$  de l'équation 3b qui suit une loi du chi-deux à  $n-m$  degrés de liberté si le vecteur des erreurs de mesures suit la loi  $N(0, s_0^2 Q)$ . En considérant la condition (a) réalisée, i.e.

$\hat{X} = X_{th}$ , le vecteur des résidus obtenu correspond, d'après l'équation 3, au vecteur :

$$\hat{V} = E \quad (\text{éq. 6})$$

La valeur centrale pour la loi du chi-deux étant son degré de liberté, on peut assurer que le test du chi-deux sera validé quel que soit le seuil de confiance choisi dès lors que l'on peut trouver  $E$  tel que :

$$E^T Q^{-1} E = (n - m) s_0^2$$

La condition (a) étant supposée satisfaite, on peut utiliser l'équation 5 pour exprimer la condition (b) sous la forme :

$$y^T y = (n - m) s_0^2 \quad (\text{éq. 7})$$

Pour répondre à la condition (c), il faut fixer un seuil de probabilité  $p$ , et en déduire les bornes de la région d'acceptation du test de Baarda  $[-\gamma, \gamma]$  par lecture dans la table inverse de la fonction de répartition de la loi normale centrée réduite. D'après la définition du résidu normé de l'équation 3c, la condition (c) peut s'écrire :

$$-\gamma s_0 q_{vv} \leq \hat{V} \leq \gamma s_0 q_{vv}$$

Expression dans laquelle  $q_{vv}$  est un vecteur réel de taille  $n$  contenant les racines carrées des éléments diagonaux de la matrice des cofacteurs  $Q_{\hat{V}}$

sur les résidus, autrement dit les quantités de l'équation 3c. En supposant les conditions (a) et (b) satisfaites, et en utilisant les équations 5 et 6, il vient :

$$-\gamma s_0 q_{vv} \leq R^{-1}\beta y \leq \gamma s_0 q_{vv} \quad (\text{éq. 8})$$

Notre recherche d'un vecteur d'erreur  $E$  satisfaisant aux conditions (a), (b) et (c), est ainsi équivalente à trouver un vecteur réel  $y$  de taille  $n-m$ , satisfaisant aux équations 7 et 8.

Une possibilité est alors de considérer la solution  $\hat{z}$  du problème de programmation quadratique (QP) :

$$\max_{z \in \mathbb{R}^{n-m}} \Phi(z) = z^T z \quad (\text{éq. 9})$$

s.c.  $z \in U$

Dans ce problème, l'ensemble des solutions admissibles correspond au polyèdre :

$$U = \{u \in \mathbb{R}^{n-m} \text{ tq } -\gamma s_0 q_{vv} \leq R^{-1}\beta u \leq \gamma s_0 q_{vv}\} \quad (\text{éq. 10})$$

Considérons également la sphère de rayon  $r$  en dimension  $n-m$  définie par :

$$S_r = \{u \in \mathbb{R}^{n-m} \text{ tq } u^T u = r^2\}$$

Par définition, toute solution du problème de l'équation 9 remplit les conditions (a) et (c). Si cette solution est telle que  $\Phi(\hat{z}) \geq (n - m) s_0^2$ , alors il suffit de projeter la solution  $\hat{z}$  sur la sphère  $S_{s_0\sqrt{n-m}}$ , orthogonalement à la sphère, pour trouver une valeur  $\hat{y}$  satisfaisant à la condition (b). Certes,  $\hat{y}$  n'est plus la solution du problème de l'équation 9, mais comme il se trouve à l'intérieur du polyèdre  $U$ , il remplit les conditions (a) à (c). Si  $\Phi(\hat{z}) < (n - m) s_0^2$  alors il n'est pas possible de trouver un vecteur  $y$  satisfaisant les conditions (a) à (c) et la solution  $\hat{z}$  est celle qui remplit au plus près la condition (b). Il peut être alors intéressant de remplacer la condition (b) par la condition (b') suivante : (b') Au seuil de confiance  $p$  choisi, le test du chi-deux est validé.

La condition (b') permet ainsi de considérer la région d'acceptation  $[\gamma_1, \gamma_2]$  du test du chi-deux au seuil  $p$ , et de chercher la projection de  $\hat{z}$  sur la sphère plutôt  $S_{s_0\sqrt{\gamma_1}}$  que sa projection sur la sphère  $S_{s_0\sqrt{n-m}}$ .

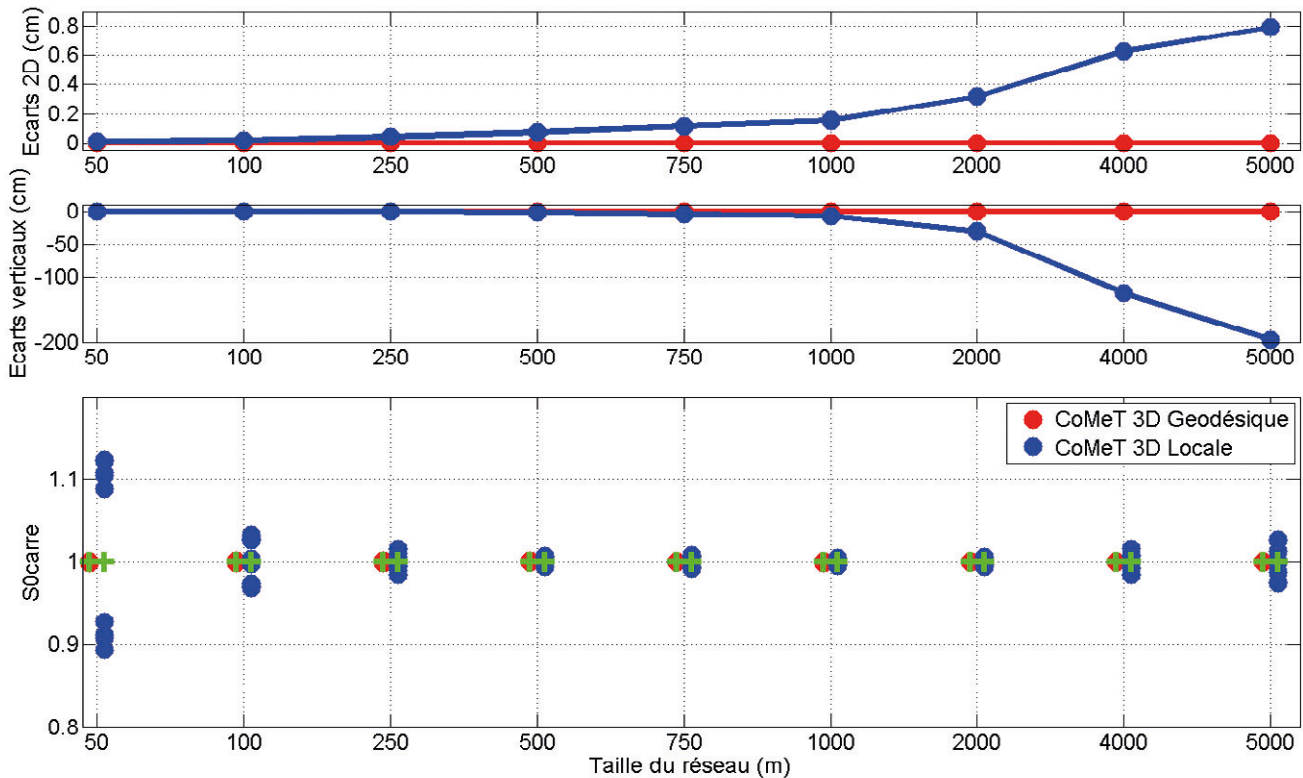
Étant donné que la fonction quadratique présente dans la fonction objectif est la matrice identité, donc symétrique

définie positive, ce problème est de type NP-Difficile (cf. [12]) et est appelé un problème d'optimisation quadratique non convexe. La difficulté dans la recherche d'une solution de ce problème est qu'un maximum local ne correspond pas forcément à un maximum global. Dans [13], plusieurs méthodes pour résoudre un problème QP dans le cas non convexe sont décrites. Les méthodes les plus récentes utilisent des techniques de séparation et évaluation (*Branch and Bound* - B&B), où la partie séparation est basée sur les conditions du premier ordre de Karush-Kuhn-Tucker (cf. [11]) et où des relaxations polyédrales semi-définies sont résolues à chaque nœud du processus B&B (cf. [5] et [7]).

Une autre méthode, plus simple et plus adaptée à notre cas, est d'utiliser une méthode d'énumération de sommets telles que celles décrites dans [1] et [4]. L'objectif des méthodes d'énumération est de fournir l'ensemble des sommets d'un polyèdre. Sachant que la solution du problème de l'équation 9 se trouve forcément sur un sommet du polyèdre  $U$ , on peut alors pour chaque sommet calculer la valeur de la fonction objectif et déduire la solution du problème de l'équation 9.

Le point de départ des méthodes d'énumération est de partir du fait que pour un polyèdre tel que défini dans l'équation 10, un point  $u$  est un sommet du polyèdre  $U$ , si et seulement si, il est l'unique solution d'un sous-ensemble de  $n-m$  inéquations indépendantes résolues comme des équations. Les avantages de telles méthodes sont d'une part leur simplicité de mise en œuvre par rapport aux méthodes de résolution d'un problème QP non convexe, et d'autre part, le fait que, pour notre besoin, elles peuvent être stoppées dès lors que l'on trouve un sommet  $u \in U$  tel que  $u^T u \geq (n - m) s_0^2$ .

Afin d'illustrer la méthode, nous avons généré, avec un seuil de confiance de 87 %, pour chacun des réseaux, 10 jeux d'observations, en prenant soin à chaque fois d'utiliser un sommet différent du polyèdre  $U$ , afin d'obtenir des vecteurs d'erreurs différents, même si, du fait de la forme du polyèdre, les coordonnées des différents sommets



**Figure 5.** En haut, écarts entre coordonnées ajustées et théoriques sur les composantes planimétriques pour chaque taille de réseau et chacun des 10 jeux de mesures taille de réseau. Au milieu, écarts sur les coordonnées verticales. En bas, valeurs de l'estimateur du facteur unitaire de variance pour chacun des 10 jeux simulés pour chaque taille de réseau suivant les deux modèles fonctionnels utilisés.

peuvent parfois être très semblables. Ces jeux d'observations générés ont été ajustés en utilisant le modèle fonctionnel 3D Géodésique ainsi que le modèle 3D Locale de CoMeT. La *figure 5* présente les résultats obtenus en termes d'écarts aux coordonnées théoriques planes et verticales mais également de valeurs de l'estimateur du facteur unitaire de variance.

Le premier constat que l'on peut faire sur cette *figure 5* est que pour le modèle 3D Géodésique qui a servi à générer les observations, les écarts obtenus entre coordonnées ajustées et théoriques sont nuls. Les mesures générées respectent bien la condition (a). On remarque également avec ce modèle fonctionnel que les valeurs de l'estimateur du facteur unitaire de variance sont toutes identiques et égales à 1. Les mesures générées respectent donc bien la condition (b).

Le second constat est que lors de l'ajustement des mesures avec le modèle fonctionnel 3D Locale, les résultats obtenus pour chaque réseau sont identiques pour chacun des 10 jeux de mesures. Ces résultats sont identiques

à ceux de la *Figure 2*, où l'on considèrerait un vecteur d'erreurs nul. Ainsi, les écarts constatés sont uniquement liés à la différence de modèle fonctionnel.

La *Figure 5* (bas) présente les valeurs de l'estimateur du facteur unitaire de variance pour chaque jeu de mesures ajusté, pour chacun des modèles fonctionnels, avec en vert les valeurs moyennes sur les 10 jeux par taille de réseau. On constate que pour le modèle fonctionnel 3D Géodésique (courbe rouge), la valeur de l'estimateur est toujours égale à 1, conformément au critère (b). Pour le modèle fonctionnel 3D locale, les valeurs obtenues sont en moyenne égales à 1, mais avec étendues pouvant atteindre 0.2 pour le réseau à 50 m. Cela nous permet là encore de juger la présence de modèles fonctionnels différents.

## Conclusion

Cet article fait suite à celui publié dans la revue en 2012, sur la validation du logiciel CoMeT par comparaison à d'autres logiciels d'ajustement, sur la base d'observations générées sur

des réseaux de géométrie particulière. Après avoir présenté et illustré quelques méthodes classiques permettant de générer un vecteur d'erreurs à ajouter aux mesures théoriques (issues d'un modèle fonctionnel particulier), nous avons détaillé notre méthodologie particulière de génération.

Cette méthode, dont l'idée était déjà présente dans l'article de 2012, s'est surtout développée d'un point de vue théorique entre 2013 et 2015. Elle permet d'assurer que le vecteur d'erreurs  $E$  ajouté aux mesures théoriques conduira, lors de l'ajustement avec le même modèle fonctionnel, à valider les deux tests classiques réalisés à l'issue d'un traitement par moindres carrés : test du chi-deux et test de Baarda. Notre méthodologie assure également dans ce cas que la solution ajustée correspond exactement à la solution théorique pour les paramètres. Ces caractéristiques permettent de faciliter l'analyse des résultats de comparaisons sur des jeux de données où l'on fait varier un ou plusieurs paramètres (taille du réseau, hauteur de point...).





Cette méthode de génération du vecteur d'erreurs  $E$  est intégrée dans le logiciel CoMeT, et nous avons montré, sur un exemple pratique, son fonctionnement effectif. ●

## Contact

**Stéphane DURAND**

École supérieure des géomètres et topographes  
Laboratoire géomatique et foncier  
stephane.durand@lecnam.net

## Références

- [1] D. Avis and K. Fukuda, *A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra*, Discrete Comput. Geom. 8 (1992), 295-313.
- [2] W. Baarda, *A Testing Procedure for use in Geodetic Networks*, Delft, Computing Centre of the Delft Geodetic Institute, 97 pages, 1964.
- [3] George E. P. Box, Mervin E. Muller, « *A Note on the Generation of Random Normal Deviates* », The Annals of Mathematical Statistics Vol. 29, No. 2 (Jun., 1958), pp. 610-611
- [4] D. Bremner, K. Fukuda, and A. Marzetta. *Primal-dual methods for vertex and facet enumeration*. Discrete Comput. Geom., 20 (1998), 333-357.

- [5] Samuel Burer and Dieter Vandembussche, *A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations*, Math. Program., Ser. A 113 (2008), 259-282.
- [6] W.F. Caspary *Concepts of network and deformation analysis*. Edited by. J. M. Rüeger. Third (corrected) Impression, 2000.
- [7] Jieqiu Chen, Samuel Burer, *Globally solving nonconvex quadratic programming problems via completely positive programming*. Mathematical Programming Computation 4, Issue 1 (2012), pp 33-52.
- [8] Dach R., Lutz S., Walser P. and Fridez P., *Bernese GNSS software version 5.2*, Astronomical Institute, University of Bern, 2015
- [9] Durand S. et Guérin C., *Validation du logiciel CoMeT d'ajustement de mesures topographiques*, Revue XYZ n°132, pp32-40, 3<sup>e</sup> trimestre 2012
- [10] Durand S., *CoMeT-Compensation de Mesures Topographiques*, Manuel utilisateur, Laboratoire Géomatique et Foncier, Cnam, 2017 (<http://comet.esgt.cnam.fr>)
- [11] Kuhn, H. W.; Tucker, A. W. (1951). « *Nonlinear programming* ». Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. pp. 481-492. MR 0047303.
- [12] P. M. Pardalos, S. A. Vavasis, *Quadratic*

*programming with one negative eigenvalue is NP-hard*, J. Global Optim., 1(1), 15-22, 1991.

[13] P. Pardalos, *Global optimization algorithms for linearly constrained indefinite quadratic problems*, Comput. Math. Appl., 21, 87-97, 1991.

[14] Sillard, Patrick, *Estimation par moindres carrés*, Collection ENSG-IGN, Hermès, Lavoisier, 2001

## ABSTRACT

*In a previous paper, published in 2012 in this journal, we conducted comparisons between the CoMeT software and others least squares network adjustment applications in order to validate the CoMeT application. To perform such comparisons, we generated several sets of measurements in networks sharing the same geometry but different side lengths, and compared the obtained results in terms of adjusted coordinates and estimated variance factor. The purpose of this contribution is to review and discuss some common methods used for generating observations and to present in more detail our method, the one used in the previous article, which present some specific properties.*

### Olivier Reis

*Ingénieur géomètre-topographe*  
ENSAI Strasbourg - Diplômé de l'Institut  
de traducteurs et d'interprètes (ITI) de Strasbourg  
9, rue des Champs  
F-57200 SARREGUEMINES  
Téléphone / télécopie : 03 87 98 57 04  
Courriel : o.reis@infonie.fr

**Pour toutes vos traductions d'allemand  
et d'anglais en français en topographie  
– géodésie –  
géomatique – GNSS**

### Reinhard Stölzel

*Ingénieur géomètre-topographe*  
Interprète diplômé de la  
Chambre de commerce et d'industrie de Berlin  
Heinrich-Heine-Straße 17, D-10179 BERLIN  
Téléphone : 00 49 30 97 00 52 60  
Télécopie : 00 49 30 97 00 52 61  
Courriel : stoelzelr@aol.com

**Pour toutes vos traductions de français  
et d'anglais en allemand en géomatique  
– GNSS –  
infrastructures de transport**

**Des topographes traducteurs d'expérience à votre service**