

UTILISATION DES INCERTITUDES ASSOCIEES AUX MESURES ALTIMETRIQUES SATELLITAIRES DE PLAN D'EAU

JC Bader, G Belaud, UMR G-Eau, Montpellier, 14/06/2010

RESUME

On cherche comment utiliser les incertitudes I associées aux mesures altimétriques satellitaires de plan d'eau Y , pour quantifier l'imprécision des régressions établies entre de telles mesures et des cotes X mesurées au sol. Le but consiste à connaître les erreurs types résiduelles de régression qu'on obtiendrait si chaque valeur Y_i de la série de base était remplacée par une distribution normale de moyenne Y_i et d'écart type I_i . Par une méthode numérique basée sur l'utilisation de nombreux jeux de données différents, on montre les résultats suivants de portée générale, dont certains sont justifiés de façon théorique :

Pour la régression exprimant Y en fonction de X :

- Les droites de régression sont identiques sur la série de base et la série étendue
- L'erreur type résiduelle de régression $E1re$ sur la série étendue est fonction de l'erreur type résiduelle de régression $E1rb$ sur la série de base et de la moyenne quadratique $MQIb$ des incertitudes I sur la série de base :

$$E1re = (E1rb^2 + MQIb^2)^{0.5} \quad (0)$$

Pour la régression exprimant X en fonction de Y :

- Les droites de régression calculées sur la série de base et la série étendue diffèrent. Leurs coefficients directeurs respectifs $A2b$ et $A2e$ sont tels que le rapport $A2e/A2b$, toujours inférieur ou égal à 1, décroît en fonction du rapport $MQIb/ETYb$, où $ETYb$ désigne l'écart type de Y sur la série de base.
- Sur la série étendue, les écarts de X par rapport à la droite de régression calée sur la série de base ont une moyenne nulle. La moyenne quadratique $E3re$ de ces écarts est fonction de $MQIb$ et de l'erreur type résiduelle de régression $E2rb$ calculée sur la série de base :

$$E3re = (E2rb^2 + (A2b \times MQIb)^2)^{0.5} \quad (00)$$

NOTATIONS

Dans les notations, le suffixe 0 apposé à un nom de paramètre signifie que la valeur de celui-ci est arbitrairement choisie. Les suffixes b et e apposés à un nom de paramètre signifient que celui-ci est calculé sur la série de base ou sur la série étendue.

$A0$	pente de relation linéaire moyenne entre X et Y
$A1$	pente de régression linéaire entre X et Y
$A2$	pente de régression linéaire entre Y et X
$B0$	ordonnée à l'origine de la relation linéaire moyenne entre X et Y

$B1$	ordonnée à l'origine de la régression linéaire entre X et Y
$B2$	ordonnée à l'origine de la régression linéaire entre Y et X
D_i	résidu de rang i (écart entre Y_i et $A0 \times X_i + B0$)
$E1r$	écart type résiduel de la régression linéaire entre X et Y
$E2r$	écart type résiduel de la régression linéaire entre Y et X
$E3re$	moyenne quadratique des écarts entre X et la régression de Y calée sur série de base
ETI	écart type des incertitudes I sur Y
ETD	écart type des résidus D
ETX	écart type des valeurs de X
ETY	écart type des valeurs de Y
$f(a,b,c)$	loi normale inversée de probabilité a , moyenne b et écart type c
I_i	incertitude sur la valeur de Y_i
KI	coefficient de régression double associé à $MQIb$, reliant $E1re$ à $E1rb$ et $MQIb$
KR	coefficient de régression double associé à $E1rb$, reliant $E1re$ à $E1rb$ et $MQIb$
$M3e$	moyenne des écarts entre X et la régression de Y calée sur la série de base
MI	moyenne des valeurs d'incertitude I sur Y
MD	moyenne des valeurs de D
MQI	moyenne quadratique des valeurs de I
MX	moyenne des valeurs de X
MY	moyenne des valeurs de Y
N	nombre de couples X_i, Y_i
R^2	coefficient de détermination ou de Nash de la régression linéaire entre X et Y
Rr^2	coefficient de détermination de la régression double reliant $E1re$ à $E1rb$ et $MQIb$
X_i	valeur de rang i de la variable X
Y_i	valeur de rang i de la variable Y

OBJECTIF

On considère deux séries X et Y de dimension $N0$ représentant des niveaux de plan d'eau, susceptibles d'être reliées par régression linéaire après éventuel décalage de temps, constant ou variable. Chaque valeur de X est une cote observée au sol, alors que chaque valeur de Y est une cote mesurée par altimétrie satellitaire, associée à une valeur d'incertitude I .

En première approximation, on considère que X et Y ont des distributions normales et qu'en outre, chaque valeur Y_i correspond à la moyenne d'une distribution normale d'écart type I_i .

Etant donné l'écart type résiduel $E1rb$ sur Y de la régression linéaire entre X et Y et la moyenne quadratique $MQIb$ des incertitudes I , on désire connaître l'écart type résiduel $E1re$ de la régression entre X et Y qu'on obtiendrait en remplaçant chaque valeur Y_i par sa distribution normale d'écart type I_i . On cherche donc à établir une relation de portée générale entre $E1re$, $E1rb$ et $MQIb$.

On se pose la même question pour la régression linéaire entre Y et X .

METHODE

Les relations recherchées sont établies par une démonstration numérique, basée sur de nombreux jeux de données aléatoires vérifiant les conditions de normalité énoncées plus haut. Les résultats obtenus pour la régression entre X et Y sont également justifiés de façon théorique, grâce aux hypothèses de normalité.

Constitution d'un jeu de données

Un jeu de données consiste en une série de base de NO triplets X, Y, I et en une série étendue associée de $100NO$ couples X, Y , élaborée à partir de la série de base.

La série de base est définie à partir de l'octuplet $NO, MX0, ETX0, A0, B0, ETD0, MIO, ETIO$, dont les valeurs sont choisies de façon arbitraire.

Les valeurs croissantes de X_i sont définies ainsi pour chaque rang i entre 1 et NO :

$$X_i = f((i-0.5) / NO ; MX0 ; ETX0) \quad (1)$$

On définit ensuite une série de résidus ordonnés D_i et d'incertitudes ordonnées I_i de la façon suivante, pour chaque rang i entre 1 et NO :

$$D_i = f((i-0.5) / NO ; 0 ; ETD0) \quad (2)$$

$$I_i = f((i-0.5) / NO ; 0 ; ETIO) \quad (3)$$

Après une redistribution en ordre aléatoire des valeurs de chaque série D_i et I_i , les valeurs de la série Y_i sont alors calculées de la façon suivante :

$$Y_i = A0 \times X_i + B0 + D_i \quad (4)$$

La série étendue est obtenue en remplaçant chacun des NO triplets X_i, Y_i, I_i de la série de base par 100 couples X_{ij}, Y_{ij} définis ainsi pour chaque rang j entre 1 et 100 :

$$X_{ij} = X_i \quad (5)$$

$$Y_{ij} = Y_i + f((j-0.5)/100 ; 0 ; I_i) \quad (6)$$

Conclusion : chaque jeu de données, constitué d'une série de base et d'une série étendue associée, est défini par les valeurs arbitraires de l'octuplet $NO, MX0, ETX0, A0, B0, ETD0, MIO, ETIO$, un classement aléatoire des résidus D_i et un classement aléatoire des incertitudes I_i .

Statistiques calculées pour chaque jeu de valeurs

La série de base présente par construction les valeurs moyennes suivantes pour X, Y, I et D :

$$MX_b = MX0 \quad (7)$$

$$MY_b = A0 \times MX0 + B0 \quad (8)$$

$$MI_b = MIO \quad (9)$$

$$MD_b = 0 \quad (10)$$

Les écarts types de X, I et D sur la série de base présentent l'égalité suivante :

$$ETX_b/ETX0 = ETI_b/ETIO = ETD_b/ETD0 = k \quad (11)$$

avec $k < 1$ et tendant vers 1 quand NO augmente (par ex : $k=0.975$ pour $NO = 25$)

Les régressions entre X et Y sont calculées dans les deux sens à la fois sur la série de base et sur la série étendue, donnant les paramètres $A1$ et $A2$ (pentes de régression), $B1$ et $B2$

(ordonnées à l'origine), $E1r$ et $E2r$ (écarts types résiduels) et R^2 (coefficient de détermination). La moyenne quadratique MQI des incertitudes I sur Y est calculée sur la série de base. Enfin, on calcule sur la série étendue la moyenne $M3e$ et la moyenne quadratique $E3re$ des écarts entre X et la droite de régression de paramètres $A2b$ et $B2b$ établie sur la série de base.

RESULTATS

Régressions entre X et Y

Au total, 180 jeux de données différents ont été testés, correspondant à 12 octuplets de paramètres arbitraires différents, avec pour chacun d'eux 15 redistributions aléatoires des valeurs de résidus D et d'incertitudes I . Les résultats moyens obtenus pour chaque octuplet de paramètres sont donnés dans le tableau 1.

Pour chaque jeu de données, les coefficients de régression entre X et Y sont strictement identiques pour la série de base et la série étendue ($A1b = A1e$ et $B1b = B1e$), comme le montre la figure 1 pour un jeu de données particulier.

Les paramètres $A1$, $B1$, $E1r$ et R^2 de la régression entre X et Y varient évidemment beaucoup en fonction des valeurs imposées à l'octuplet de paramètres arbitraires, mais assez peu (hormis $B1$), en fonction des redistributions aléatoires de D et de I . Ces variations sont cependant suffisantes pour pouvoir établir pour chaque octuplet de paramètres les régressions linéaires doubles contraintes suivantes, basées sur 15 triplets $E1re$, $E1rb$, $MQIb$:

$$E1re = KR \times E1rb + KI \times MQIb \quad (12)$$

Le coefficient de détermination de cette régression est strictement égal à 1 pour chacun des douze octuplets de paramètres testés. On remarque par ailleurs (fig.2) que les coefficients de régression KR et KI sont très bien reliés à la valeur moyenne du rapport $MQIb/E1rb$:

$$KR = (1 / (1 + (MQIb / E1rb)^2))^{0.5} \quad (13)$$

$$KI = ((MQIb / E1rb)^2 / (1 + (MQIb / E1rb)^2))^{0.5} \quad (14)$$

Les relations 12 à 14 donnent finalement :

$$E1re = (E1rb^2 + MQIb^2)^{0.5} \quad (15)$$

Cette relation est confirmée par la figure 3 qui regroupe les résultats obtenus avec chacun des 180 jeux de données testés.

Régressions entre Y et X

Au total, 195 jeux de données différents ont été testés, correspondant à 13 octuplets de paramètres arbitraires différents, avec pour chacun d'eux 15 redistributions aléatoires des valeurs de résidus D et d'incertitudes I . Les résultats moyens obtenus pour chaque octuplet de paramètres sont donnés dans le tableau 2.

Contrairement aux régressions entre X et Y , les régressions entre Y et X peuvent être très différentes lorsqu'elles sont calculées sur la série de base ou sur la série étendue (fig. 4). Or dans le cas pratique qui nous intéresse, on dispose seulement d'une série de base de données

sur laquelle il est possible d'établir la régression entre Y et X , d'équation :

$$X = A2b \times Y + B2b \quad (16)$$

C'est donc bien par rapport à cette droite de régression qu'il convient d'évaluer les écarts de X sur la série étendue, et en particulier leur moyenne $M3e$ et leur moyenne quadratique $E3re$. On constate que $M3e$ est parfaitement nul pour chacun des 195 jeux de données testés. Les résultats obtenus permettent par ailleurs d'établir la relation suivante (fig. 5) :

$$E3re = (E2rb^2 + (A2b \times MQIb)^2)^{0.5} \quad (17)$$

Dans tous les cas, les rapports $E2re/E3re$ et $A2e/A2b$ sont inférieurs ou égaux à 1. Ils décroissent globalement en fonction du rapport MQI/ETY (fig.6), mais pourraient aussi dépendre d'autres facteurs.

Justification théorique de la relation 15

On note Ya la valeur de Y estimée de manière empirique. On a aussi l'incertitude empirique MQI . On suppose que $(Y-Ya)$ suit une loi gaussienne $LG(0,I)$, MQI étant une estimation empirique de I .

On a fait la régression entre Ya et X , avec les coefficients A et B , erreur type Erb . On suppose que $(Ya-(A.X+B))$ suit une loi gaussienne $LG(0,Erb)$.

La somme des deux variables aléatoires gaussiennes $(Y-Ya)$ et $(Ya-(A.X+B))$ reste une variable aléatoire gaussienne de moyenne nulle et de variance $Ere^2=I^2+Erb^2$.

<i>NO</i>	100	100	100	100	15	75	50	50	50	50	100	100
<i>MXO</i>	1000	1000	1000	1000	500							
<i>ETXO</i>	700	700	700	700	1000							
<i>A0</i>	3.0	3.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	10.0	10.0	0.5
<i>B0</i>	100	100	500	500	500							
<i>ETD0</i>	900	900	900	90	90	150	500	1000	1000	1000	330	480
<i>MIO</i>	550	550	55	55	55	55	120	1100	2500	1800	1800	1800
<i>ETIO</i>	200	20	20	20	20	20	30	30	300	300	600	600
<i>MXb</i>	1000	1000	1000	1000	500							
<i>ETXb</i>	696	696	696	696	671	694	691	691	691	691	696	994
<i>MYb</i>	3100	3100	3100	3100	3100	2100	2100	2100	2100	10500	10500	750
<i>ETYb</i>	2306	2234	2269	2089	2009	1401	1453	1641	1676	7043	6958	691
<i>MQIb</i>	585	550	58	58	58	58	124	1100	2517	1824	1896	1896
<i>Mlb</i>	550	550	55	55	55	55	120	1100	2500	1800	1800	1800
<i>ETIb</i>	198.7	19.9	19.9	19.9	19.2	19.8	29.6	29.6	296.2	296.2	596.2	596.2
<i>min(Ii)</i>	34.8	498.5	3.5	3.5	18.3	5.5	50.2	1030.2	1802.1	1102.1	254.5	254.5
<i>MDb</i>	0.000	0.000	0.000	0.000	0.000							
<i>ETDb</i>	894.3	894.3	894.3	89.4	86.3	148.7	493.7	987.4	987.4	987.4	327.9	476.9
<i>A1b</i>	3.06	2.95	3.00	3.00	2.99	2.01	1.98	1.90	1.97	10.09	9.99	0.50
<i>B1b</i>	42.6	152.8	100.5	98.7	107.8	92.2	120.0	197.9	134.7	406.5	507.9	497.7
<i>E1rb</i>	890.5	888.0	891.4	88.7	82.2	147.3	488.1	980.7	978.3	967.7	326.7	474.9
<i>Rb²</i>	0.850	0.841	0.845	0.998	0.998	0.989	0.887	0.641	0.655	0.981	0.998	0.525
<i>A1e</i>	3.06	2.95	3.00	3.00	2.99	2.01	1.98	1.90	1.97	10.09	9.99	0.50
<i>B1e</i>	42.6	152.8	100.5	98.7	107.8	92.2	120.0	197.9	134.7	406.5	507.9	497.7
<i>E1re</i>	1063	1043	893	106	101	158	503	1469	2686	2055	1912	1943
<i>ETYe</i>	2378	2300	2270	2090	2010	1403	1458	1972	3012	7273	7208	2007
<i>Re²</i>	0.800	0.794	0.845	0.997	0.997	0.987	0.880	0.444	0.204	0.920	0.930	0.063
<i>MQIb / Erb</i>	0.657	0.620	0.066	0.659	0.708	0.397	0.253	1.122	2.573	1.885	5.804	3.993
<i>KR</i>	0.837	0.851	0.998	0.836	0.805	0.928	0.969	0.666	0.361	0.459	0.171	0.243
<i>KI</i>	0.544	0.523	0.065	0.545	0.591	0.370	0.246	0.741	0.927	0.883	0.979	0.964
<i>Rr²</i>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tableau 1 : résultats obtenus pour les régressions entre X et Y avec 180 jeux de données. Chaque colonne correspond à 15 jeux de données obtenus à partir des paramètres des 8 premières lignes, par redistribution aléatoire des valeurs de résidus D et d'incertitudes I . Les valeurs en gras sont strictement identiques pour les 15 jeux de données associés à un octuplet de paramètres donné. Les autres valeurs correspondent à la moyenne obtenue sur les 15 jeux de données.

<i>N0</i>	100	100	100	100	100	100	100	100	100	100	100	100	100
<i>MX0</i>	500	500	500	500	500	500	500	500	500	500	100	500	100
<i>ETX0</i>	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	2000	1000	500
<i>A0</i>	0.5	0.2	0.2	0.2	10.0	10.0	10.0	5.0	1.0	3.0	3.0	10.0	0.6
<i>B0</i>	500	500	500	500	500	500	500	500	500	500	100	500	100
<i>ETD0</i>	480	1000	5	125	125	500	500	500	500	500	500	500	50
<i>M0</i>	1800	50	2000	500	50	20	1000	1000	850	35	120	1000	2000
<i>ET0</i>	600	10	500	120	10	5	5	300	300	10	10	5	500
<i>MXb</i>	500	500	500	500	500	500	500	500	500	500	100	500	100
<i>ETXb</i>	994	994	994	994	994	994	994	994	994	994	1987	994	497
<i>MYb</i>	750	600	600	600	5500	5500	5500	3000	1000	2000	400	5500	160
<i>ETYb</i>	691	1013	199	235	9939	9941	9950	5003	1093	3044	5971	9942	301
<i>MQIb</i>	1896	51	2061	514	51	21	1000	1043	901	36	120	1000	2061
<i>Mlb</i>	1800	50	2000	500	50	20	1000	1000	850	35	120	1000	2000
<i>ETIb</i>	596.2	9.9	496.8	119.2	9.9	5.0	5.0	298.1	298.1	9.9	9.9	5.0	496.8
<i>min(I)</i>	254.5	24.2	712.1	190.9	24.2	7.1	987.1	227.3	77.3	9.2	94.2	987.1	712.1
<i>MDb</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>ETDb</i>	476.9	993.6	5.0	124.2	124.2	496.8	496.8	496.8	496.8	496.8	496.8	496.8	49.7
<i>A1b</i>	0.504	0.201	0.200	0.202	10.002	9.992	10.001	5.010	0.980	3.023	2.995	9.993	0.597
<i>B1b</i>	498.1	499.6	500.0	499.2	499.1	503.8	499.5	494.8	509.9	488.7	100.5	503.5	100.3
<i>E1rb</i>	475.1	989.3	4.9	123.4	123.7	495.1	495.2	495.8	494.7	493.0	495.1	495.2	49.1
<i>Rb²</i>	0.524	0.046	0.999	0.723	1.000	0.998	0.998	0.990	0.794	0.974	0.993	0.998	0.973
<i>A1e</i>	0.504	0.201	0.200	0.202	10.002	9.992	10.001	5.010	0.980	3.023	2.995	9.993	0.597
<i>B1e</i>	498	500	500	499	499	504	500	495	510	489	101	503	100
<i>E1re</i>	1943	991	2048	525	134	496	1110	1149	1023	494	509	1110	2048
<i>ETYe</i>	2007	1015	2057	563	9939	9941	9999	5110	1413	3044	5973	9991	2070
<i>Re²</i>	0.063	0.046	0.009	0.127	1.000	0.998	0.988	0.949	0.475	0.974	0.993	0.988	0.021
<i>A2b</i>	1.041	0.190	4.995	3.594	0.100	0.100	0.100	0.198	0.811	0.322	0.332	0.100	1.632
<i>B2b</i>	-280.6	386.0	-2496.7	-1656.5	-49.8	-49.1	-48.6	-92.9	-311.3	-144.5	-32.7	-49.0	-161.1
<i>E2rb</i>	684.7	970.4	24.7	522.1	12.4	49.5	49.5	98.5	450.4	161.0	164.8	49.5	81.3
<i>A2e</i>	0.123	0.190	0.047	0.629	0.100	0.100	0.099	0.189	0.485	0.322	0.332	0.099	0.034
<i>B2e</i>	407.4	386.3	472.0	122.8	-49.8	-49.1	-43.2	-68.5	15.2	-144.4	-32.6	-43.6	94.5
<i>E2re</i>	962.0	970.5	989.0	928.2	13.4	49.5	110.3	223.5	719.7	161.4	169.5	110.4	491.7
<i>E3re</i>	2077.2	970.5	10227.1	1908.6	13.4	49.5	110.8	227.3	854.5	161.4	169.5	110.9	3342.2
<i>M3e</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tableau 2 : résultats obtenus pour les régressions entre Y et X avec 195 jeux de données. Chaque colonne correspond à 15 jeux de données obtenus à partir des paramètres des 8 premières lignes, par redistribution aléatoire des valeurs de résidus D et d'incertitudes I . Les valeurs en gras sont strictement identiques pour les 15 jeux de données associés à un octuplet de paramètres donné. Les autres valeurs correspondent à la moyenne obtenue sur les 15 jeux de données.

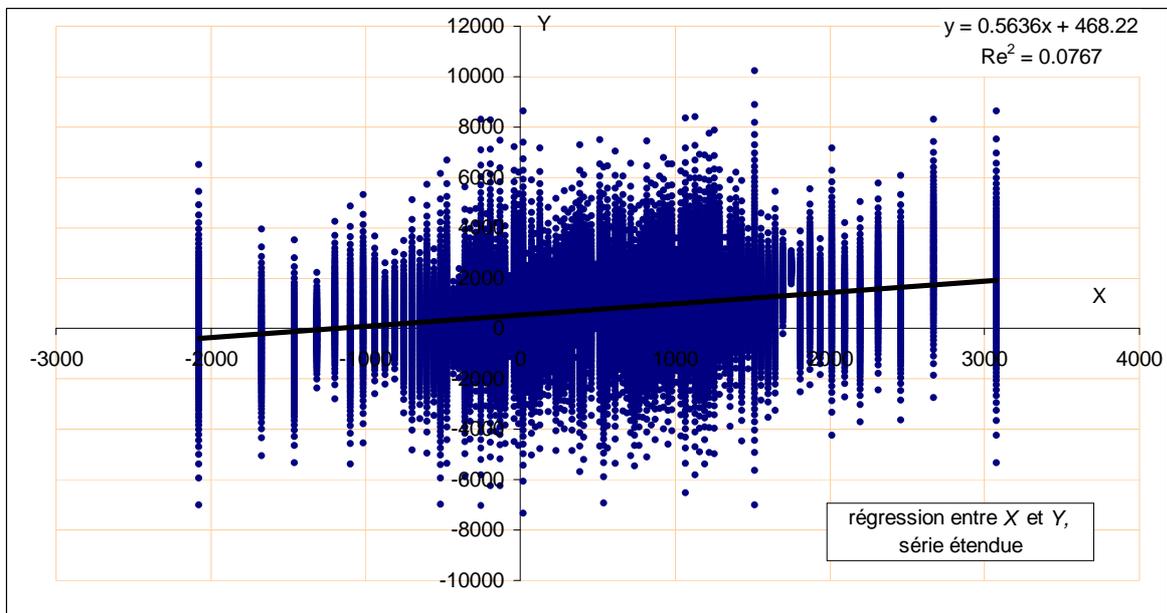
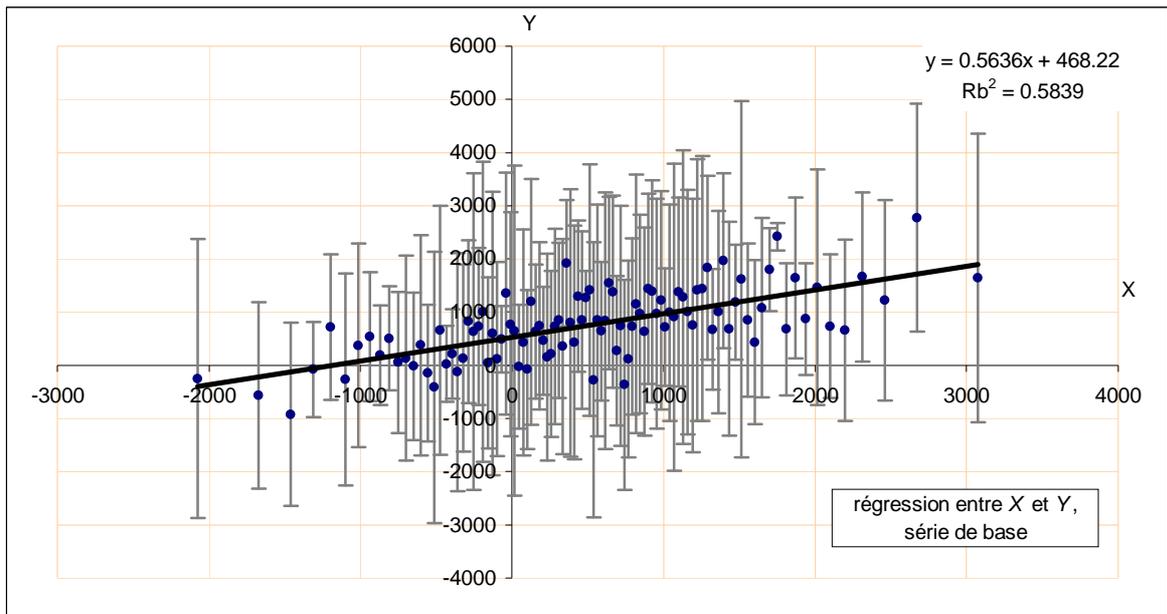


Figure 1 : similarité de la régression linéaire entre X et Y, établie sur la série de base ou la série étendue associée. Exemple pour un jeu de données caractérisé par : $N0=100$; $MX0=500$; $ETX0=1000$; $A0=0.5$; $B0=500$; $ETD0=480$; $MIO=1800$; $ETIO=600$.

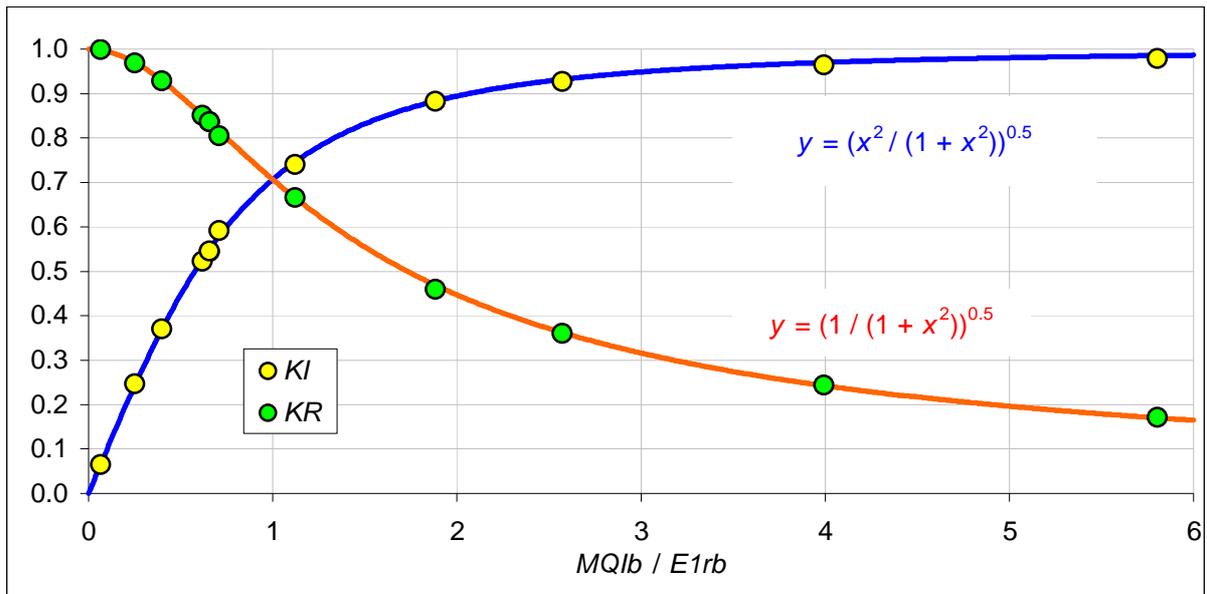


Figure 2 : variations des coefficients de régression KI et KR en fonction du rapport $MQIb/E1rb$

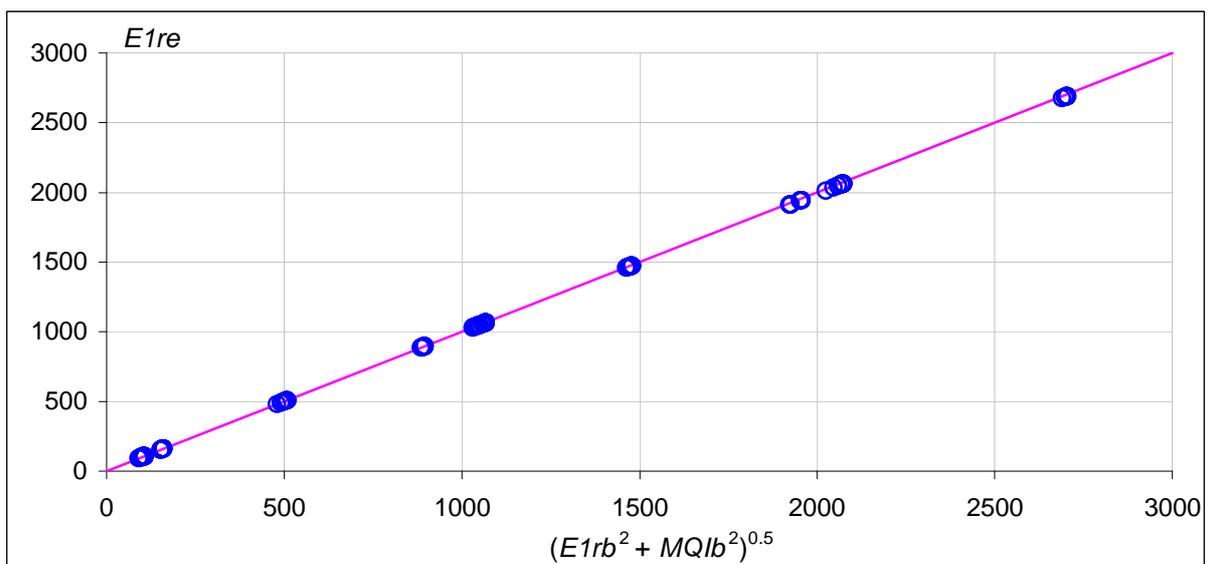


Figure 3 : relation observée sur les 180 jeux de données testés, entre l'erreur type résiduelle de régression $E1re$ de la série étendue, l'erreur type résiduelle de régression $E1rb$ de la série de base et la moyenne quadratique $MQIb$ des incertitudes I sur Y de la série de base

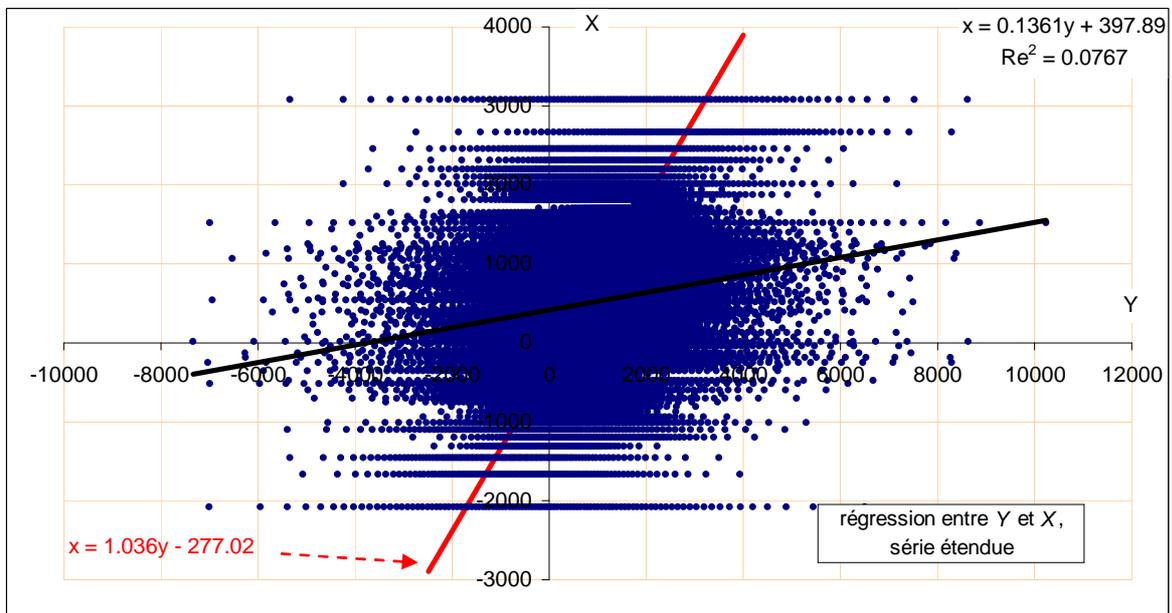
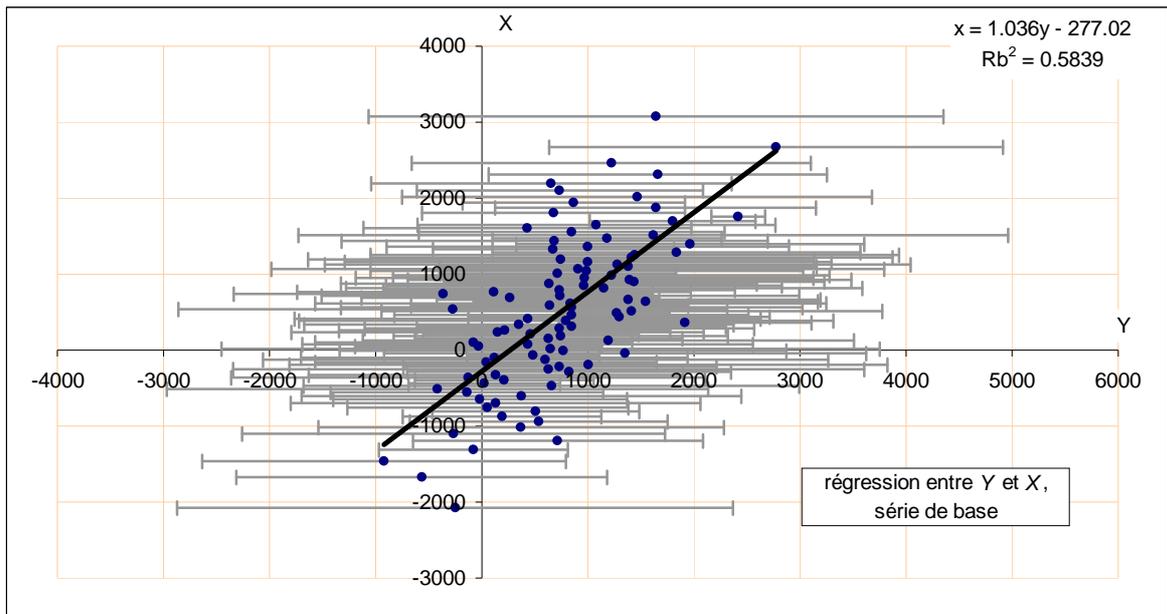


Figure 4 : exemple de régressions différentes entre Y et X, obtenues sur une série de base et sur la série étendue associée (jeu de données identique à celui de la figure 1)

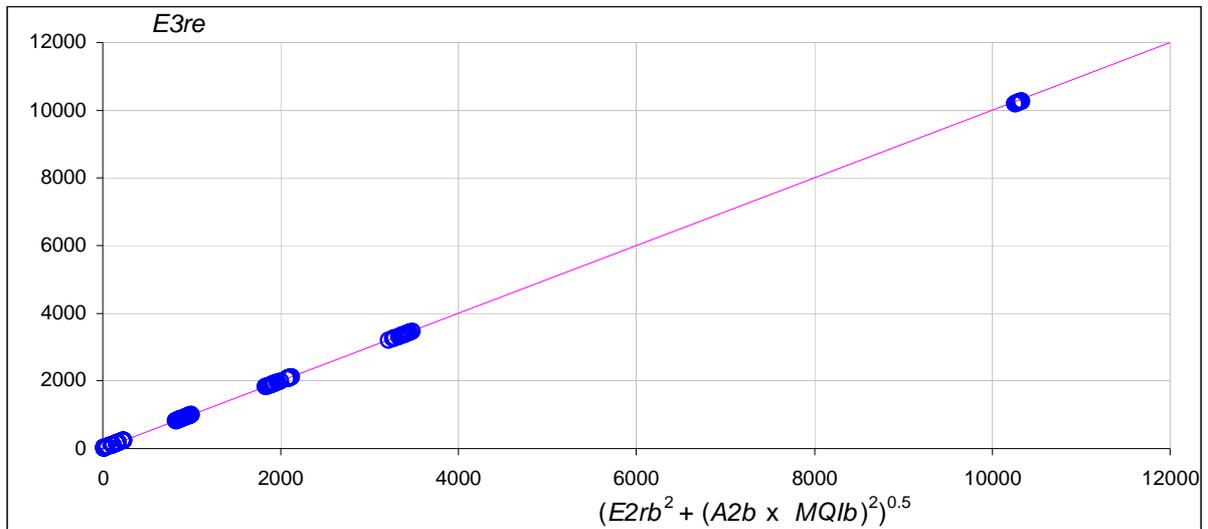


Figure 5 : relation observée sur les 195 jeux de données testés, entre la moyenne quadratique $E3re$, sur la série étendue, des écarts de X à la droite $X = A2b \times Y + B2b$, d'une part, et les paramètres suivants calculés sur la série de base, d'autre part : $E2rb$ (erreur type résiduelle de régression), $MQIb$ (moyenne quadratique des incertitudes I sur Y) et $A2$ (coefficient de régression).

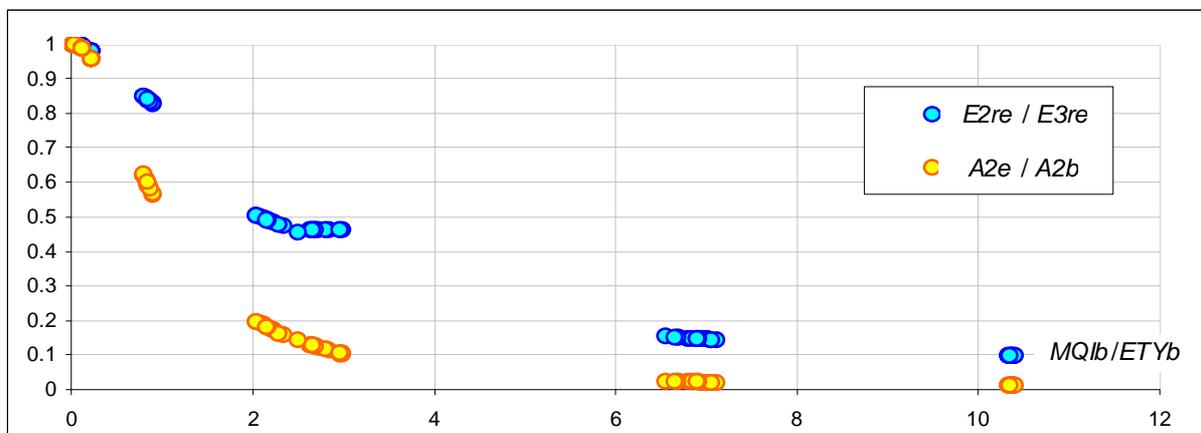


Figure 6 : rapports $E2re/E3re$ et $A2e/A2b$ en fonction du rapport $MQIb/ETYb$