

École Polytechnique de Montréal
Département de mathématiques et de génie industriel et département C.G.M.

MTH2302C – Probabilités et statistique

Examen final – Hiver 2010, Mercredi 28 avril de 9h30 à 12h00

QUESTION 1 (14 points)

La pluie annuelle maximale (X_i en mm) mesurée sur une période de 6 heures a été relevée à la station météorologique du Mont-Tremblant pour les 50 dernières années. On considère que les variables aléatoires X_i sont indépendantes.

La moyenne des 50 valeurs observées est $\bar{x} = 20$ mm et la variance expérimentale est $s^2 = 9$ mm². La distribution obtenue est :

Intervalle (mm)	$0 \leq x < 17.475$	$17.475 \leq x < 19.24$	$19.24 \leq x < 20.76$	$20.76 \leq x < 22.525$	$22.525 \leq x$
Fréquence observée	7	12	11	8	12

Les classes précédentes sont équiprobables pour une $N(20, 9)$.

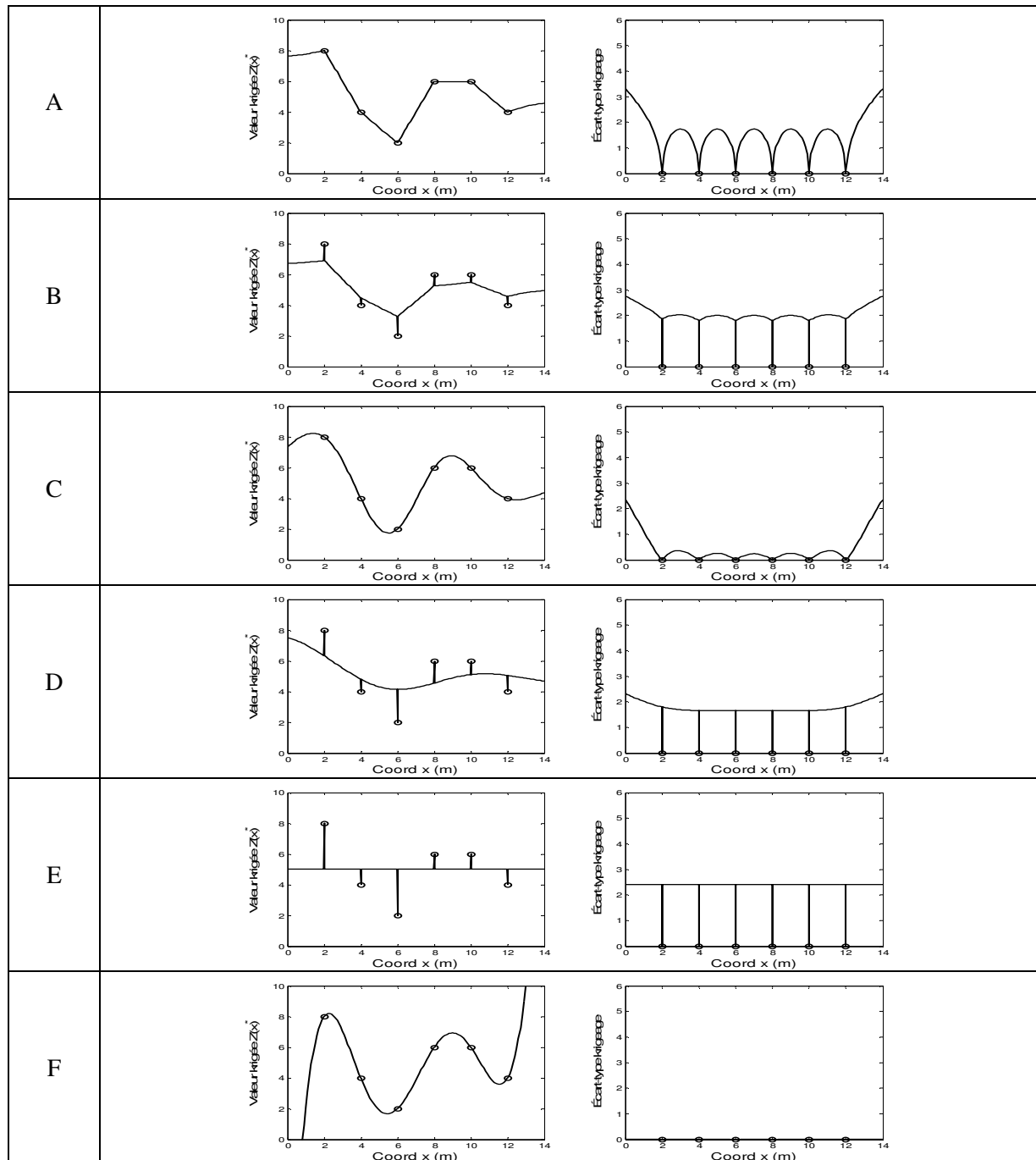
- a) *Faites le test d'ajustement permettant de vérifier, au niveau $\alpha = 5\%$, si la distribution est compatible avec une loi normale.*
- b) *Peut-on affirmer que la moyenne de la population μ_x est significativement supérieure à 19 mm? Faites le test requis au niveau $\alpha = 5\%$. Poser clairement les hypothèses H_0 et H_1 du test et la règle de décision.*

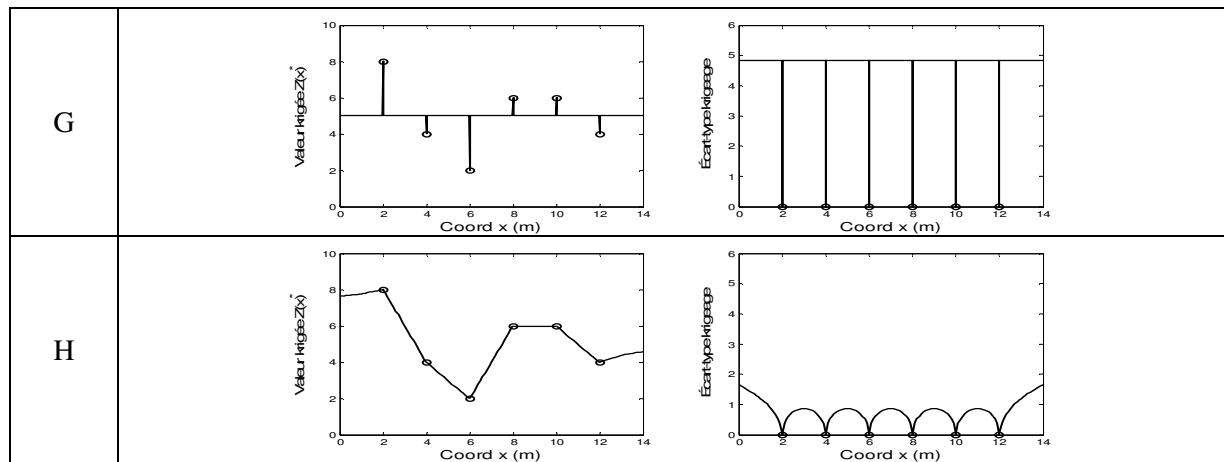
Réponses

Question 1 (suite)

QUESTION 2 (12 points)

Chacune des 8 figures suivantes (A à H, sur 2 pages) montre, selon la coordonnée x (m), 6 valeurs observées Z_i ($i=1\dots 6$) (o), l'interpolation obtenue par krigeage $Z(x)$ (colonne de gauche) et l'écart-type de krigeage $\sigma_k(x)$ (colonne de droite) obtenus avec un certain modèle de variogramme.





Le tableau suivant présente différents modèles de variogrammes. Associez à chaque modèle du tableau une seule figure (indiquez la lettre correspondante dans la dernière colonne).

Réponses

	Modèle	Figure
i.	Effet de pépite pur, $C_0 = 20$	
ii.	Effet de pépite pur, $C_0 = 5$	
iii.	Sphérique, portée=10 m, $C_0=0$, $C=20$	
iv.	Sphérique, portée=10 m, $C_0=0$, $C=5$	
v.	Sphérique, portée=10 m, $C_0=2$, $C=8$	
vi.	Gaussien, portée effective = 5 m, $C_0=0$, $C=10$	
vii.	Gaussien, portée effective = 20 m, $C_0=0$, $C=10$	
viii.	Gaussien, portée effective = 10 m, $C_0=2$, $C=8$	

QUESTION 3 (20 points)

Pour 20 sables différents, on effectue en laboratoire la mesure de la porosité (x , en fraction) et du \log_{10} (conductivité hydraulique (m/s)) (y). On effectue ensuite la régression linéaire de y sur x . On obtient les résultats suivants :

$\sum_{i=1}^{20} x_i$	$\sum_{i=1}^{20} y_i$	$\sum_{i=1}^{20} x_i^2$	$\sum_{i=1}^{20} y_i^2$	$\sum_{i=1}^{20} x_i y_i$
5.266	-89.633	1.449	404.95	-23.189

$\sum_{i=1}^{20} (x_i - \bar{x})$	$\sum_{i=1}^{20} (y_i - \bar{y})$	$\sum_{i=1}^{20} (x_i - \bar{x})^2$	$\sum_{i=1}^{20} (y_i - \bar{y})^2$	$\sum_{i=1}^{20} (x_i - \bar{x})(y_i - \bar{y})$
0	0	0.06255	3.2511	0.41155

De plus, le coefficient de détermination $R^2 = 0.83$

- Estimez les coefficients β_0 et β_1 par moindres carrés.
- Calculez SS_E et SS_R .
- Faites le test pour vérifier le caractère significatif de la régression (coefficient β_1) au niveau $\alpha = 5\%$.
- Un sable présente une porosité $x=0.3$. Construisez l'intervalle de confiance bilatéral de niveau 95% pour $E[Y | x=0.3]$.
- Quelle serait l'équation de prédiction (donnez les valeurs des coefficients) si l'on voulait plutôt prédire la porosité x à partir de la mesure de $y = \log_{10}$ (conductivité hydraulique) ?

Réponses

QUESTION 3 (suite)

QUESTION 4 (18 points)

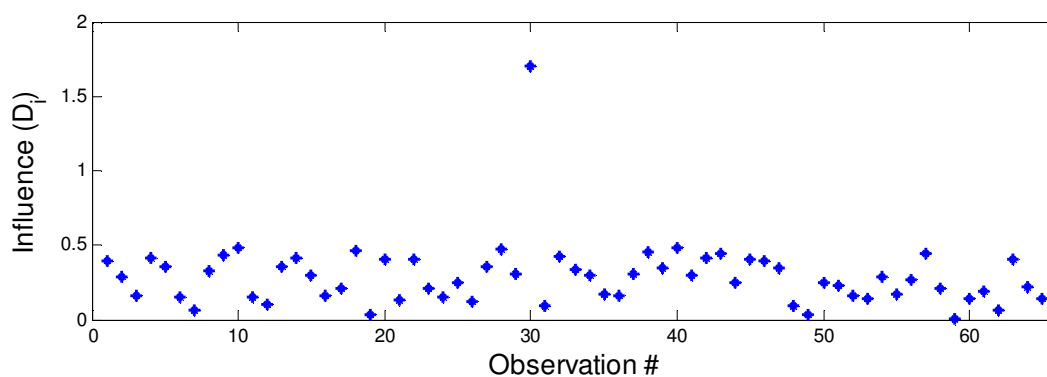
On a mesuré la gravité (en mgal) en 65 points d'une zone donnée. On désire estimer en tout point $E[Z(x,y)|x,y]$ à l'aide d'un modèle polynomial de régression selon les coordonnées x et y de la zone. On considère plusieurs modèles polynomiaux d'ordre croissant.

Le tableau suivant présente les résultats pour des ordres croissants de polynômes :

Ordre du polynôme	Nombre total de coefficients dans le modèle (incluant la constante)	R^2	SS_E (mgal ²)
1	3	0.4	1800
2	6	0.5	1500
3	10	0.54	1380
4	15	0.55	1350
5	21	0.56	1320

On a aussi $S_{yy} = 3000$ (mgal²)

- Le polynôme d'ordre 1 est-il significatif au niveau $\alpha = 5\%$? Faites le test requis.
- Partant du polynôme d'ordre 1 et augmentant successivement l'ordre du polynôme, déterminer l'ordre à retenir si le critère est de garder le modèle d'ordre le plus élevé pour lequel l'ajout des variables est significatif au niveau $\alpha = 5\%$.
- La figure suivante montre l'influence des observations dans la régression. Commentez le résultat.



- Le modèle de régression suppose des résidus non-corrélés. Décrivez le problème potentiel avec cette hypothèse dans le contexte particulier de l'étude. Quel outil géostatistique vous permettrait de vérifier cette hypothèse (expliquer)?

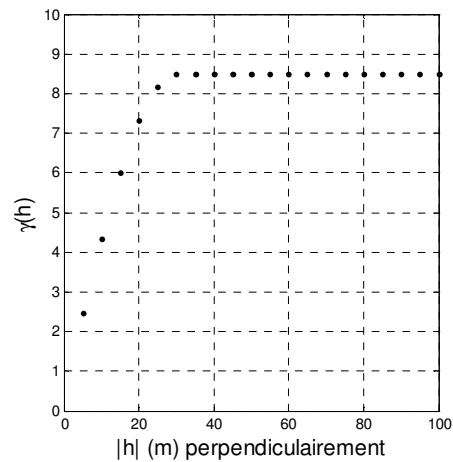
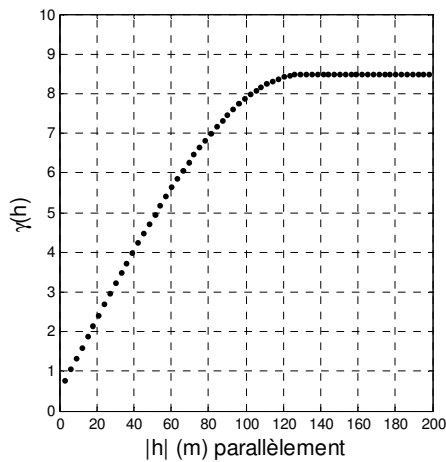
Question 4 (suite)

Réponses

QUESTION 4 (suite)

QUESTION 5 (12 points)

Un relevé bathymétrique d'un chenal maritime est effectué. On mesure l'élévation (m) du fonds du chenal. On calcule le variogramme expérimental de l'élévation ($Z(x)$), dans les directions parallèle et perpendiculaire à l'allongement du chenal, et l'on obtient :

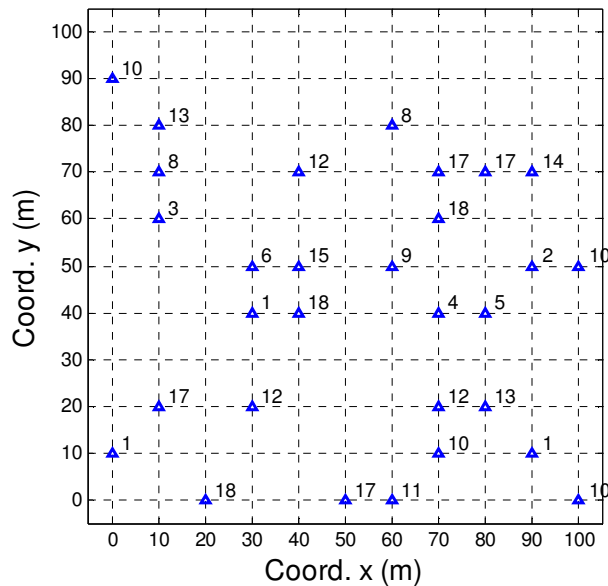


Suggérez un modèle de variogramme 2D s'ajustant à ces données. (Note : votre réponse doit être complète et précise et permettre de calculer la valeur théorique du variogramme pour toute paire de points peu importe la direction et la distance définies par la paire de points; indiquez les unités des paramètres)

Réponse

QUESTION 6 (12 points)

La figure suivante montre le résultat d'une campagne d'échantillonnage de sols contaminés aux hydrocarbures (en %).

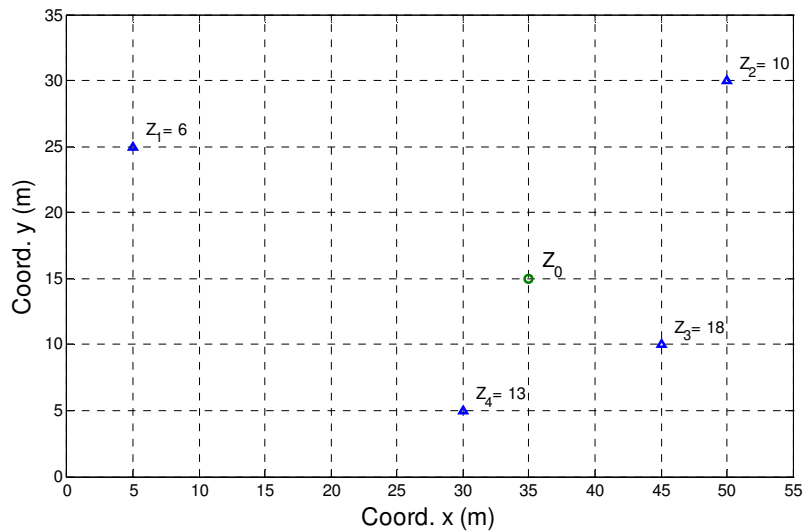


Calculez le variogramme de la concentration en hydrocarbures dans la direction NE-SO (azimut 45°) pour la distance $\|h\| = 10\sqrt{2}$ m. Indiquez clairement les paires utilisées dans le calcul en donnant les valeurs de chaque point de la paire. Indiquez les unités du résultat.

Réponse

QUESTION 7 (12 points)

Les données suivantes donnent la quantité de particules fines (en %) observées en quatre points d'un dépôt de sable (Z_1 à Z_4). On désire estimer par krigeage ordinaire la quantité de fines au point x_0 de coordonnée (35,15). Le modèle de variogramme est sphérique, isotrope, avec $C_0=20\%$, $C=80\%$ et $a=30$ m.



Pour éviter des calculs, le tableau suivant donne les valeurs de la covariance sphérique ayant $C=1$ pour différentes distances $\|h\|$ utiles.

$\ h\ $	$\sqrt{25+100} = 11.18$	$\sqrt{25+225} = 15.811$	$\sqrt{25+400} = 20.616$	$\sqrt{225+225} = 21.213$
Cov($\ h\ $; avec $C=1$)	0.4669	0.2826	0.131	0.11612

Construisez le système de krigeage ordinaire (ne pas résoudre)

QUESTION 7 (suite)

Corrigé

Q1-

a) H_0 La distribution est normale, H_1 , la distribution n'est pas normale.

$$\sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} = 1/10 (3^2 + 2^2 + 1^2 + 2^2 + 2^2) = 2.2 < \chi_{5-1-2, \alpha=5\%}^2 = 5.99, \text{ donc on accepte } H_0$$

b) $H_0 \mu_X = 19\text{mm}$ vs $H_1 \mu_X > 19\text{mm}$

$$t_0 = \frac{(20-19)\sqrt{50}}{\sqrt{9}} = 2.36 > t_{49, \alpha=5\%} = 1.677. \text{ Donc on rejette } H_0.$$

Q2-

	Modèle	Figure
i.	Effet de pépite pur, $C_0 = 20$	G
ii.	Effet de pépite pur, $C_0 = 5$	E
iii.	Sphérique, portée=10 m, $C_0=0$, $C=20$	A
iv.	Sphérique, portée=10 m, $C_0=0$, $C=5$	H
v.	Sphérique, portée=10 m, $C_0=2$, $C=8$	B
vi.	Gaussien, portée effective = 5 m, $C_0=0$, $C=10$	C
vii.	Gaussien, portée effective = 20 m, $C_0=0$, $C=10$	F
viii.	Gaussien, portée effective = 10 m, $C_0=2$, $C=8$	D

Q3-

a) $b_1 = S_{xy}/S_{xx} = 0.41155/0.06255 = 6.5795$ $b_0 = -89.633/20 - b_1(5.266/20) = -6.214$

b) $SSR = S_{yy} * R^2 = 3.2511 * 0.83 = 2.6984$ $SSE = 3.2511 - 2.6984 = 0.5527$

c) $H_0 \beta_1 = 0$ $H_1 \beta_1 \neq 0$

$$\frac{2.6984/1}{0.5527/(20-2)} = 87.88 \gg F_{1,18, \alpha=5\%} = 4.41$$

d) $\hat{y} = -6.214 + 6.5795 * 0.3 = -4.24$

$$\text{Intervalle : } -4.24 \pm t_{18, \alpha = 2.5\%} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} = -4.24 \pm 2.10 \sqrt{0.5527(0.05 + \frac{(0.3 - 0.2633)^2}{.0625})}$$
$$= -4.24 \pm 0.42 = [-4.66, -3.82]$$

$$e) \hat{x} = \alpha_0 + \alpha_1 y \quad a_1 = S_{xy}/S_{yy} = 0.41155/3.2511 = 0.1266 \quad a_0 = 5.266/20 - 0.1266(-89.633/20) = 0.8307$$

4- a)

H₀ La régression est significative

H₁ La régression n'est pas significative

$$\text{SSR} = 3000 - 1800 = 1200$$

$$\frac{1200/2}{1800/(65-3)} = 20.67 > F_{2,62, \alpha = 5\%} = 3.15 \Rightarrow \text{La polynôme d'ordre 1 est significatif}$$

b)

$$\frac{300/3}{1500/(65-6)} = 7.87 > F_{3,59, \alpha = 5\%} = 2.76 \Rightarrow \text{Le polynôme d'ordre 2 améliore significativement la régression}$$

$$\frac{120/4}{1380/(65-10)} = 1.20 < F_{4,55, \alpha = 5\%} = 2.54 \Rightarrow \text{Le polynôme d'ordre 3 n'améliore pas significativement la régression}$$

⇒ on retient le polynôme d'ordre 2.

c) L'observation #30 a une influence anormalement grande ($D_i > 1$). Il faudrait enlever cette observation et reprendre la régression.

d) Il est douteux dans un contexte spatial que les résidus de la régression soient non-corrélés spatialement. On devrait calculer le variogramme des résidus et vérifier si l'on a bien un effet de pépite pur.

Q5- Un modèle sphérique présentant une anisotropie géométrique avec effet de pépite. La direction de plus grande continuité est parallèle à l'allongement du chenal, la direction de plus petite continuité est perpendiculaire. Les portées correspondantes sont $a_{\text{paral}} = 130$ m, $a_{\text{perp}} = 30$ m. L'effet de pépite est 0.5 m^2 et la composante sphérique a $C = 8 \text{ m}^2$

Q6- On trouve 7 paires :

(1-17), (11-10), (10-13), (1-15), (5-2), (9-18), (18-17).

$$\gamma_e(10\sqrt{2}; \text{azimut } 45) = 1/14(16^2 + 1^2 + 3^2 + 14^2 + 3^2 + 9^2 + 1^2) = 39.5 \text{ \%}^2$$

Q7-

Les covariances pour les h fournis sont 80 x les valeurs données, soit :

h	$\sqrt{25+100} = 11.18$	$\sqrt{25+225} = 15.811$	$\sqrt{25+400} = 20.616$	$\sqrt{225+225} = 21.213$
Cov(h _i h _j ; avec C=80)	37.35	22.61	10.48	9.29

$$\begin{bmatrix} 100 & 0 & 0 & 0 & 1 \\ 0 & 100 & 10.518 & 0 & 1 \\ 0 & 10.518 & 100 & 22.61 & 1 \\ 0 & 0 & 22.61 & 100 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 9.289 \\ 37.349 \\ 37.349 \\ 1 \end{bmatrix}$$