Université de Technologie de Compiègne

Chapitre 2 / Modélisation stochastique en mécanique

Apprentissage statistique pour la mécanique

Identifier une réponse mécanique et contrôler une prédiction

Master Structures et Systèmes Mécaniques Complexes (SMC)

Mokhtar Z. Alaya et Salim Bouzebda

Version 2026

Apprendre et valider la réponse de la poutre

La poutre du chapitre 1 sert de banc d'essai : sa réponse exacte est connue.

Nous la cachons pendant l'apprentissage, puis nous l'utilisons pour valider.

Les essais doivent identifier la réponse et permettre de contrôler ses erreurs.

Identifier une réponse mécanique

On apprend une relation entre entrées physiques et réponse observée.

1 · Charger

Y_1, Y_2 en N

Géométrie et appuis fixés

→

2 · Mesurer

M_1 observé en N mm

Capteur étalonné

→

3 · Prédire

Une nouvelle charge

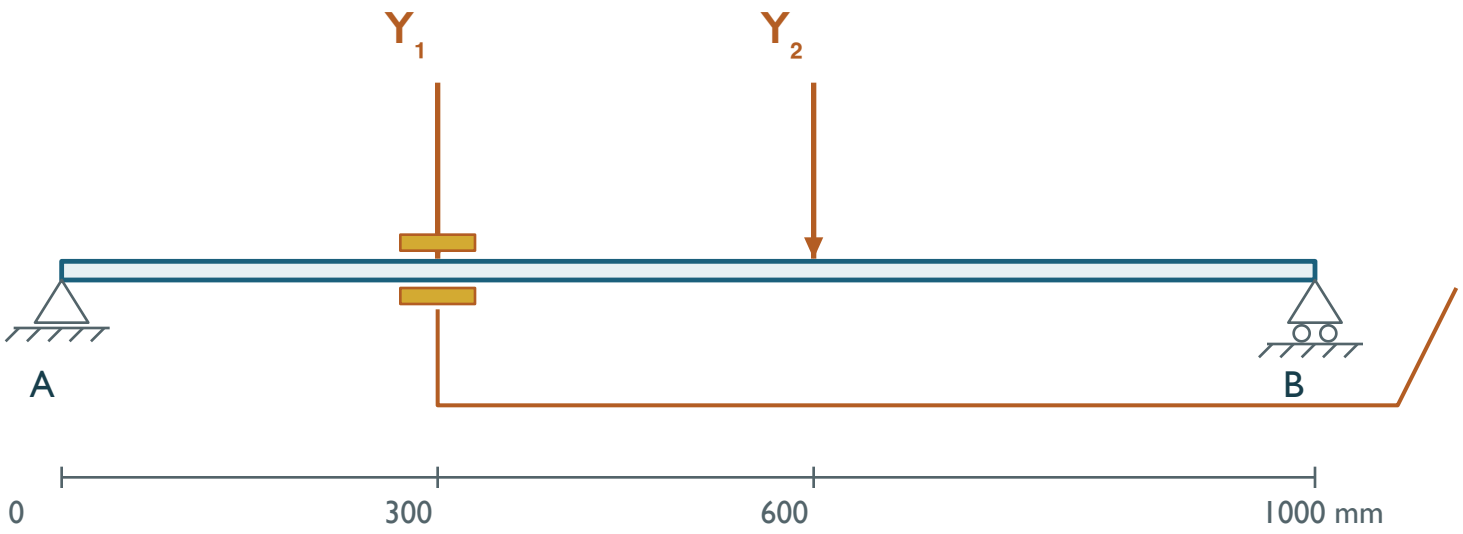
Une réponse estimée

Données = (Y_1, Y_2, M_1 observé) pour chaque essai

Le plan d'essais et le protocole de validation font partie du modèle appris.

Une traverse instrumentée

Les jauges étalonnées donnent accès au moment à la section x = 300 mm.



De la déformation au moment

Jauge → signal électrique

Étalonnage → M_1 observé

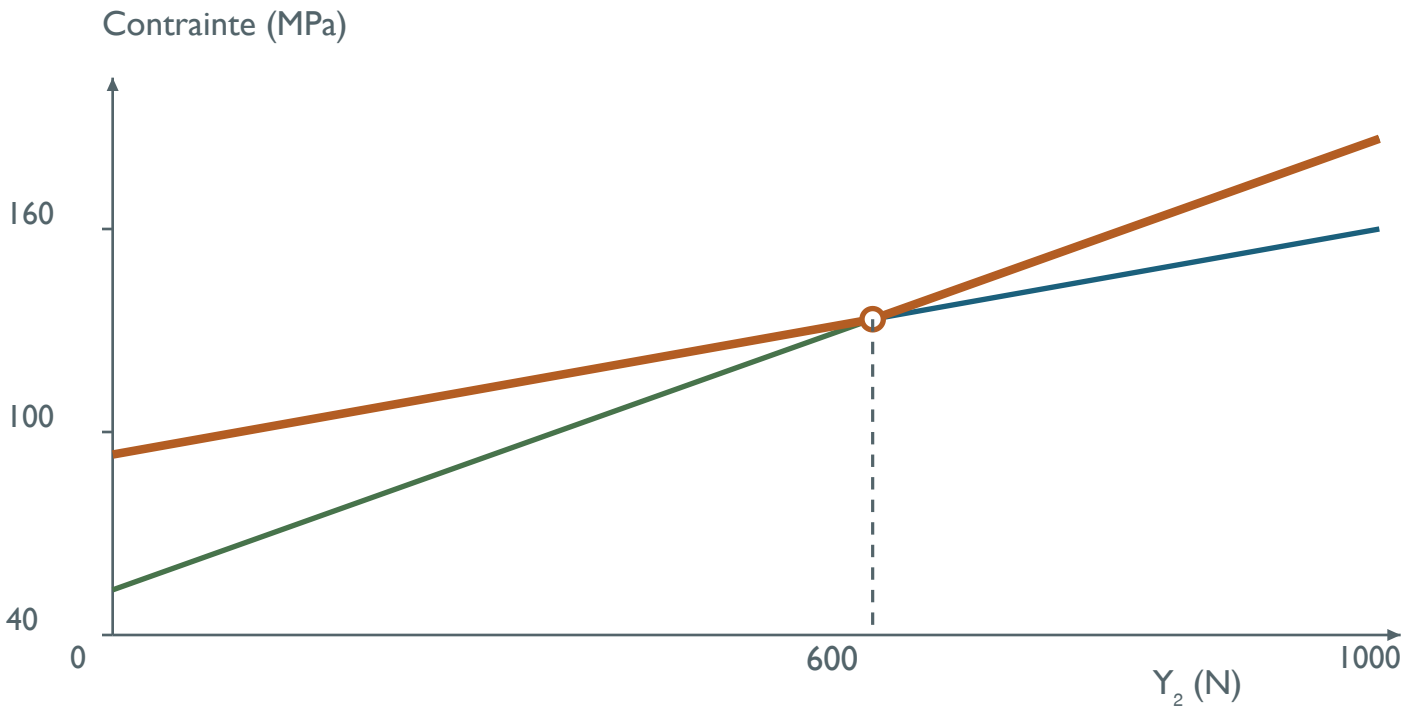
$M_1 = 210Y_1 + 120Y_2$

Unités : Y en N ; M_1 en N mm

Mesurer une réponse exige de distinguer la grandeur physique et le signal du capteur.

Deux cibles, deux difficultés

La contrainte maximale est l'enveloppe de deux réponses de section.



Exemple calculé

Y_1 = 800 N fixé

W = 1800 mm³

— M_1/W

— M_2/W

— $\sigma max = max(M_1,M_2)/W$

La section critique change : σmax reste continue, mais sa pente peut changer.

Définition 2.1 · Modèle d'apprentissage supervisé

Un jeu d'apprentissage est une collection de couples $\mathcal{D}_N = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$, où $\boldsymbol{x}_i \in \mathbb{R}^p$ est une entrée et $y_i \in \mathbb{R}$ une réponse observée. Dans le modèle de régression $y_i = f(\boldsymbol{x}_i) + \varepsilon_i$, la fonction f représente la réponse mécanique à apprendre et ε_i l'erreur d'observation. Un prédicteur $\widehat{f}$ est une fonction construite à partir de $\mathcal{D}_N$ pour estimer la réponse à une nouvelle entrée.

Une entrée fixe le chargement et une sortie mesure la réponse de la structure. Apprendre sert à prévoir cette réponse pour un nouveau chargement.

Le modèle d'observation

Chaque couple associe une entrée physique et une réponse observée

Le bruit décrit l'imprécision de cette réponse, à charges fixées

Le modèle suppose ici des erreurs conditionnellement centrées

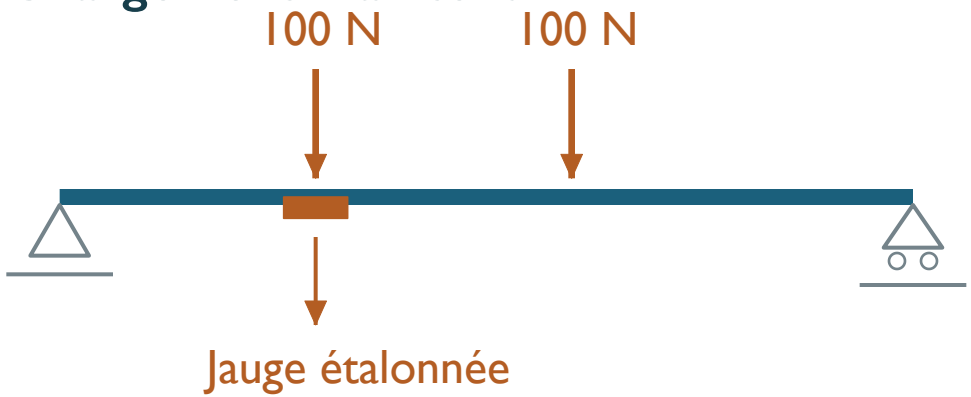
$$y_i = f(x_i) + \varepsilon_i$$

$$\mathbb{E}[\varepsilon_i \mid x_i] = 0$$

Répéter un essai à charges fixées

À chargement identique, la dispersion des mesures renseigne le bruit d'observation.

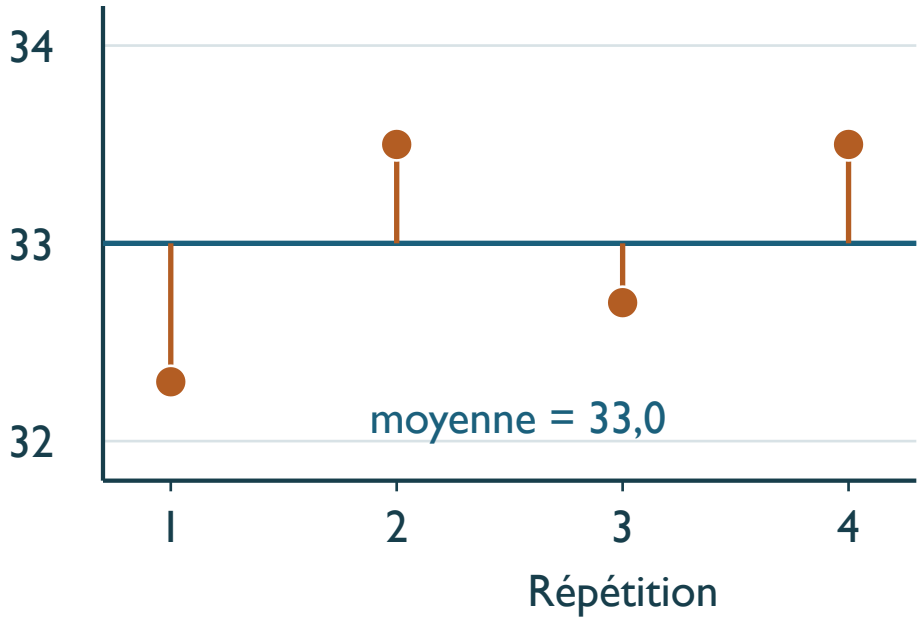
Chargement maintenu



Moment exact : 33 N m

$$\mathrm{Var}(\overline{M}) = s_\varepsilon^2 / N$$

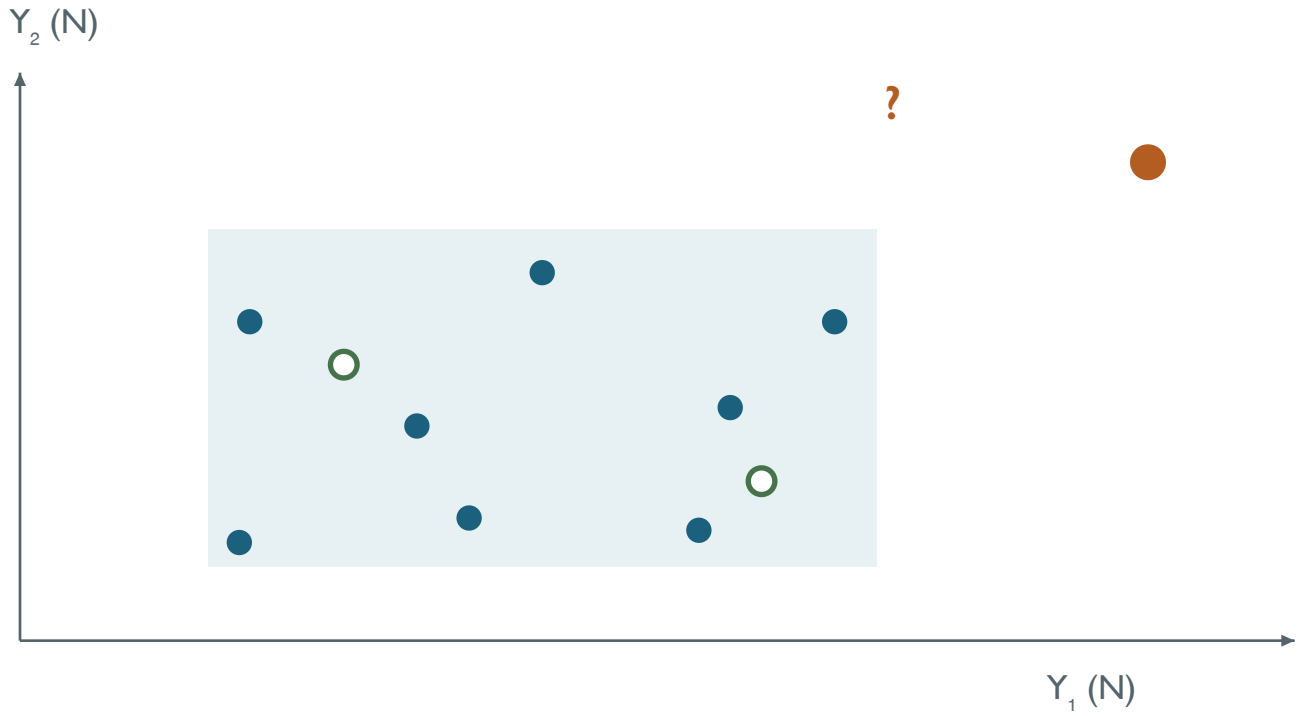
Mesures répétées (N m)



Des répétitions indépendantes précisent la moyenne ; elles ne diversifient pas les charges.

Domaine d’apprentissage et domaine d’usage

Les essais doivent explorer les directions de chargement utiles en service.



Trois statuts à distinguer

- Essais d’apprentissage
- Test tenu à l’écart
- Nouveau domaine

Même unité, usage différent.

Tester hors du domaine appris est une évaluation nouvelle, pas une garantie d’extrapolation.

Proposition 2.2 · Prédicteur optimal pour l'erreur quadratique

Soit $(\boldsymbol{X}, Y)$ un couple aléatoire tel que $\mathbb{E}[Y^2] < \infty$. Notons $m(\boldsymbol{x}) = \mathbb{E}[Y \mid \boldsymbol{X} = \boldsymbol{x}]$. Pour toute fonction mesurable g telle que $\mathbb{E}[g(\boldsymbol{X})^2] < \infty$, on a, pour presque toute entrée,

$$\begin{aligned}\mathbb{E}[(Y - g(\boldsymbol{x}))^2 \mid \boldsymbol{X} = \boldsymbol{x}] \\ = \mathrm{Var}(Y \mid \boldsymbol{X} = \boldsymbol{x}) + (m(\boldsymbol{x}) - g(\boldsymbol{x}))^2.\end{aligned}$$

La moyenne conditionnelle minimise donc le risque $\mathbb{E}[(Y - g(\boldsymbol{X}))^2]$ parmi ces prédicteurs ; le minimiseur est unique à égalité presque sûre près.

À charges fixées, les mesures restent dispersées par le bruit. Pour une erreur quadratique, leur moyenne conditionnelle est la cible que le modèle doit apprendre.

Définition 2.3 · Matrice de régression et moindres carrés

Avec p caractéristiques et une constante, on écrit

$$\widehat{f}(\boldsymbol{x}) = \theta_0 + \sum_{j=1}^{p} \theta_j x_j, \qquad D = \begin{pmatrix} 1 & \boldsymbol{x}_1^\mathsf{T} \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_N^\mathsf{T} \end{pmatrix}.$$

La matrice $D \in \mathbb{R}^{N\times(p+1)}$ est la matrice de régression. L'estimateur des moindres carrés est un minimiseur du critère $J(\boldsymbol{\theta}) = \|\boldsymbol{y} - D\boldsymbol{\theta}\|^2$ sur $\mathbb{R}^{p+1}$.

Matrice de régression de la poutre

Une ligne de D contient la constante et les charges d'un essai

L'intercept s'exprime en N mm, les deux pentes en mm

Les moindres carrés minimisent la somme des résidus au carré

$$\widehat{y} = \theta_0 + \theta_1 Y_1 + \theta_2 Y_2$$

$$y \simeq D\theta, \qquad D_{i,:} = (1, Y_{1i}, Y_{2i})$$

Lemme 2.4 · Rang et matrice de Gram

Pour toute matrice réelle D, $D^\mathsf{T} D$ est symétrique positive semi-définie et $\ker(D^\mathsf{T} D) = \ker D$. Elle est définie positive, donc inversible, si et seulement si les colonnes de D sont indépendantes.

Le rang indique si le plan d'essais sépare les influences des charges. Des sollicitations redondantes peuvent laisser plusieurs jeux de coefficients indiscernables.

Proposition 2.5 · Équations normales et identifiabilité

Pour tout $D \in \mathbb{R}^{N\times(p+1)}$ et tout $\boldsymbol{y} \in \mathbb{R}^N$, le problème des moindres carrés admet au moins une solution. Ses solutions sont exactement les vecteurs vérifiant

$$D^\mathsf{T}(D\widehat{\boldsymbol{\theta}} - \boldsymbol{y}) = 0.$$

La solution est unique si et seulement si D est de rang colonne plein ; dans ce cas, $\widehat{\boldsymbol{\theta}} = (D^\mathsf{T}D)^{-1}D^\mathsf{T}\boldsymbol{y}$. Même lorsque les coefficients ne sont pas uniques, le vecteur ajusté $D\widehat{\boldsymbol{\theta}}$ est unique : c'est la projection orthogonale de $\boldsymbol{y}$ sur l'espace engendré par les colonnes de D.

Les équations normales traduisent le meilleur compromis sur les résidus des essais. Leur unicité dépend de la diversité des chargements, pas du seul nombre de mesures.

Quatre essais pour séparer les influences

Charges en N : (100,0), (0,100), (100,100), (200,100)

Moments en N mm : 21000, 12000, 33000, 54000

Le rang de D permet de séparer les deux influences et la constante

$$\widehat{\theta} = \begin{pmatrix} 0 \\ 210 \\ 120 \end{pmatrix}$$

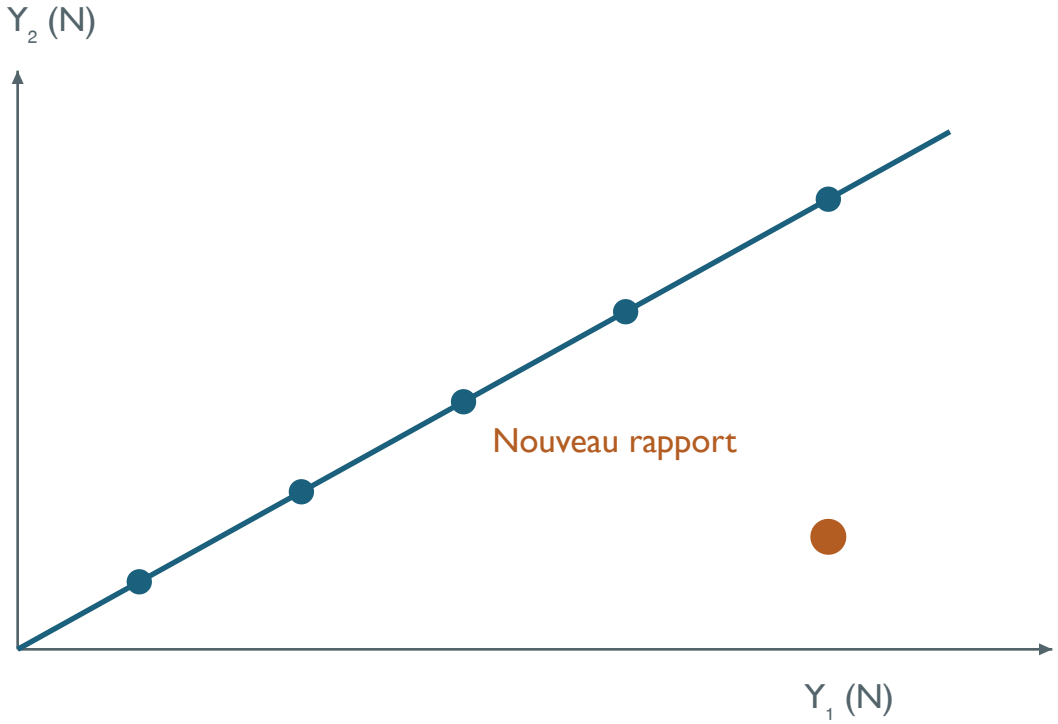
Remarque 2.6 · Identification et protocole de chargement

Des données nombreuses ne suffisent pas à garantir ce rang. Si toutes les charges expérimentales vérifient $Y_2 = 2Y_1$, répéter mille fois la même relation ne permet pas de séparer l'influence des deux charges. Il faut un plan d'expériences qui excite plusieurs directions du problème mécanique. Des prédictions précises sur ce protocole ne prouvent donc pas que chaque coefficient mécanique est identifié.

Répéter les mêmes proportions de charge ne permet pas de séparer deux influences. Il faut modifier les directions de sollicitation pour identifier les coefficients.

Charges proportionnelles : identifiabilité

Répéter un même rapport de charges ne sépare pas les deux influences.



Deux modèles indiscernables

Sur $Y_2 = 2Y_1$:

$210Y_1 + 120Y_2 = 450Y_1$

$150Y_1 + 150Y_2 = 450Y_1$

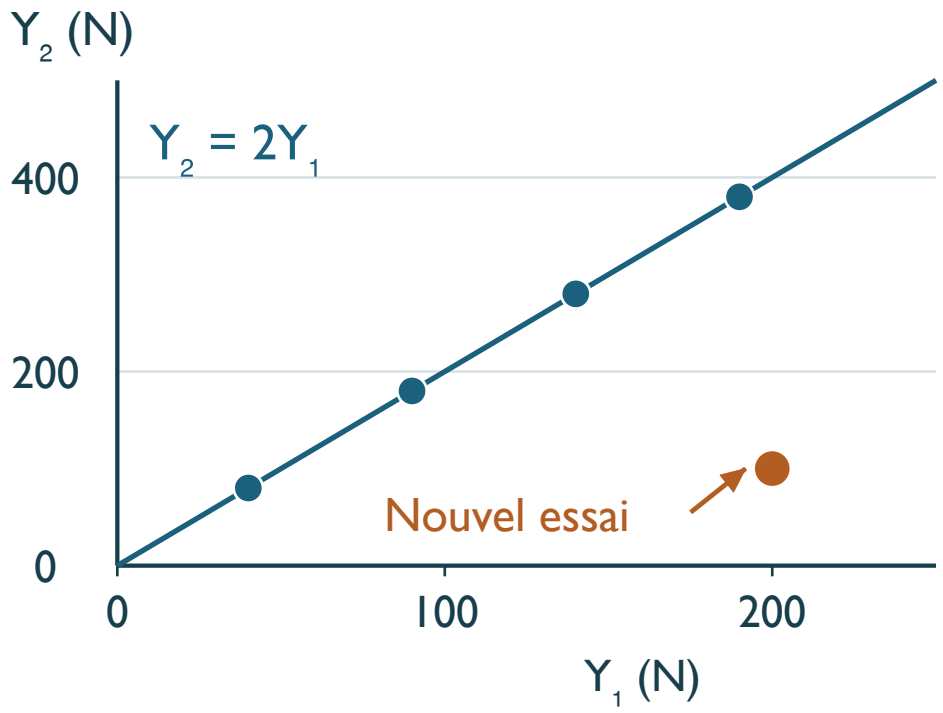
À (200,100) N :

54 000 ≠ 45 000 N mm

Il faut un plan de rang suffisant ; l'indépendance statistique des charges n'est pas requise.

Une direction de charge manquante

Sur une même droite de chargement, des influences différentes restent indiscernables.



Deux modèles compatibles

$$\widehat{M}_A = 210Y_1 + 120Y_2$$

$$\widehat{M}_B = 150Y_1 + 150Y_2$$

Sur la droite : $M = 450Y_1$

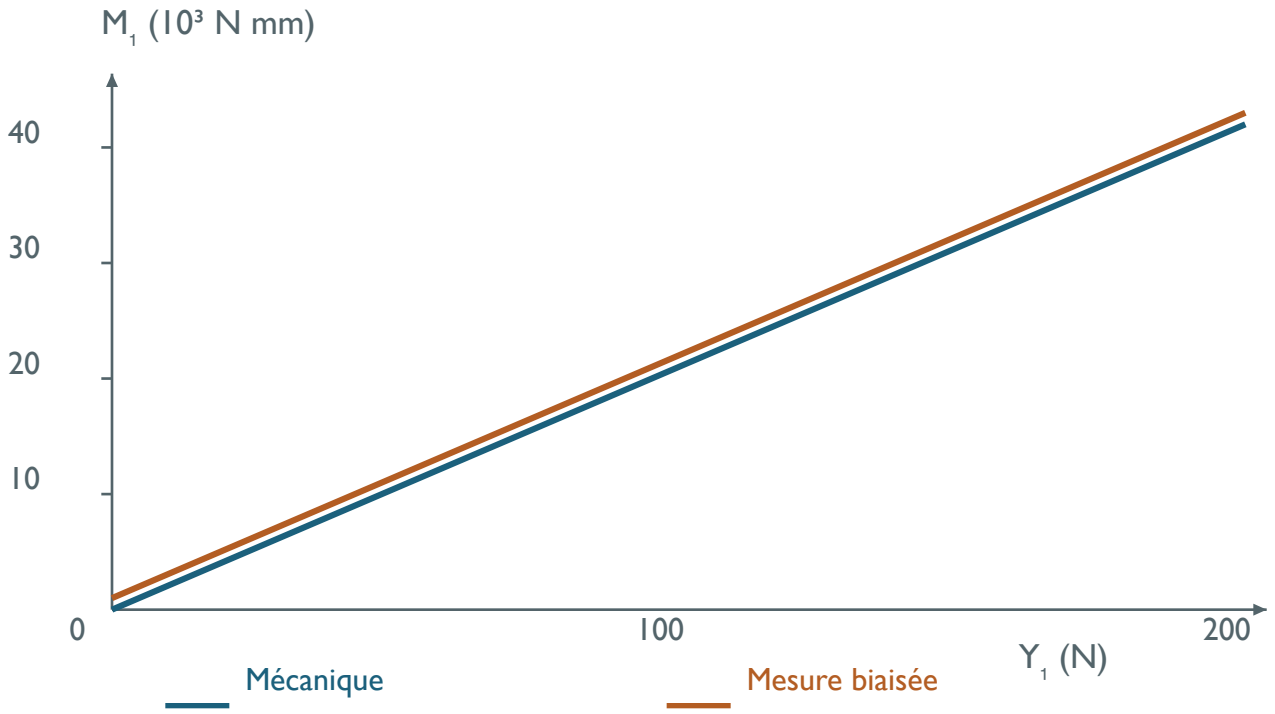
Au nouvel essai (200,100) N :

54 000 et 45 000 N mm

Changer les proportions des charges permet de distinguer les deux modèles.

Un décalage de capteur

Un biais constant modifie l'intercept, pas la réponse mécanique réelle.



Même pente, origine décalée

M_1 observé = M_1 + 1000 N mm

θ_0 = 1000 N mm

θ_1 = 210 mm

θ_2 = 120 mm

Conserver un intercept permet d'identifier un décalage constant du capteur.

Proposition 2.7 · Espérance et covariance des coefficients

Supposons le modèle affine exact $\boldsymbol{y} = D\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, D de rang colonne plein, et des erreurs de moments d'ordre deux finis telles que $\mathbb{E}[\boldsymbol{\varepsilon} \mid D] = 0$ et $\mathrm{Cov}(\boldsymbol{\varepsilon} \mid D) = s_\varepsilon^2 I$. Alors les moindres carrés sont sans biais conditionnellement au plan d'essais : $\mathbb{E}[\widehat{\boldsymbol{\theta}} \mid D] = \boldsymbol{\theta}$.

De plus,

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (D^\mathsf{T} D)^{-1} D^\mathsf{T} \boldsymbol{\varepsilon},$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}} \mid D) = s_\varepsilon^2 (D^\mathsf{T} D)^{-1}.$$

La normalité des erreurs n'est pas requise pour ces identités.

Corollaire 2.8 · Réponse moyenne et nouvelle mesure

Sous les hypothèses de la proposition 2.7, pour une entrée fixée et $\boldsymbol{d}_* = (1, \boldsymbol{x}_*^\mathsf{T})^\mathsf{T}$, on a

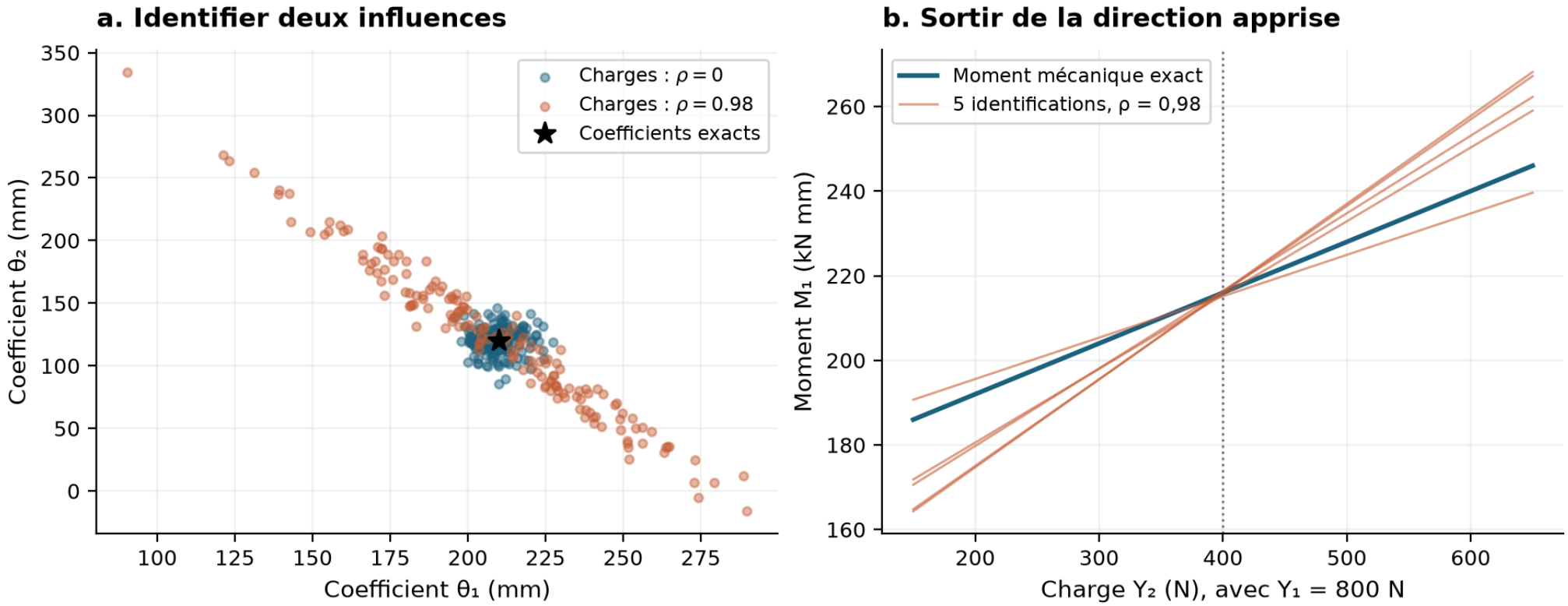
$$\mathrm{Var}(\boldsymbol{d}_*^\mathsf{T}\widehat{\boldsymbol{\theta}} \mid D) = s_\varepsilon^2 \boldsymbol{d}_*^\mathsf{T}(D^\mathsf{T}D)^{-1}\boldsymbol{d}_*.$$

Si $y_* = \boldsymbol{d}_*^\mathsf{T}\boldsymbol{\theta} + \varepsilon_*$ est une nouvelle mesure, avec $\mathbb{E}[\varepsilon_* \mid D] = 0$, $\mathrm{Var}(\varepsilon_* \mid D) = s_\varepsilon^2$ et une erreur indépendante des erreurs d'apprentissage conditionnellement à D, alors

$$\mathrm{Var}\big(y_* - \boldsymbol{d}_*^\mathsf{T}\widehat{\boldsymbol{\theta}} \mid D\big) = s_\varepsilon^2\left[1 + \boldsymbol{d}_*^\mathsf{T}(D^\mathsf{T}D)^{-1}\boldsymbol{d}_*\right].$$

Prévoir la réponse moyenne et prévoir une nouvelle mesure sont deux tâches distinctes.
La nouvelle mesure ajoute son propre bruit à l'incertitude de l'identification.

Précision des coefficients et prédiction



Des charges presque colinéaires rendent les coefficients sensibles au bruit

Des prédictions proches des essais peuvent pourtant rester précises

Proposition 2.9 · Décomposition biais–variance de l'erreur

À une entrée fixée $\boldsymbol{x}_*$, supposons que $\widehat{f}(\boldsymbol{x}_*)$ possède un moment d'ordre deux conditionnellement au plan D. Répétons les campagnes et notons $b_* = \mathbb{E}[\widehat{f}(\boldsymbol{x}_*) \mid D] - f(\boldsymbol{x}_*)$ le biais. Pour une nouvelle observation $y_* = f(\boldsymbol{x}_*) + \varepsilon_*$, dont le bruit est, conditionnellement à D, indépendant de l'apprentissage, centré et de variance s_ε^2,

$$\mathbb{E}[(y_* - \widehat{f}(\boldsymbol{x}_*))^2 \mid D] = s_\varepsilon^2 + b_*^2 + \mathrm{Var}(\widehat{f}(\boldsymbol{x}_*) \mid D).$$

Un modèle trop rigide peut manquer la réponse et un modèle trop sensible peut suivre le bruit. La validation mesure l'effet de ce compromis sur de nouveaux essais.

Définition 2.10 · Régression ridge

Pour un paramètre $\lambda \geq 0$, la régression ridge minimise

$$\|\boldsymbol{y} - D\boldsymbol{\theta}\|^2 + \lambda \sum_{j=1}^{p} \theta_j^2.$$

L'intercept θ_0 n'est pas pénalisé ; la matrice de pénalisation est $P = \mathrm{diag}(0, 1, \ldots, 1)$.

La pénalisation freine les coefficients très sensibles aux données. Elle stabilise l'apprentissage en acceptant un biais, dont l'intérêt doit être validé.

Régularisation et contrôle par validation

La pénalité porte sur les pentes, avec P=diag(0,1,1)

Pour λ>0, elle donne une solution unique même si le rang manque

Elle introduit un biais et son intérêt se vérifie par validation

$$\min_{\theta} \left\{ \|y - D\theta\|_2^2 + \lambda \|P\theta\|_2^2 \right\}$$

$$\left(D^{\mathsf{T}} D + \lambda P^{\mathsf{T}} P\right) \widehat{\theta} = D^{\mathsf{T}} y$$

Proposition 2.11 · Solution ridge et unicité

Supposons $N \geq 1$, une première colonne de D égale à $\mathbf{1}$ et $\lambda > 0$. Alors $D^\mathsf{T} D + \lambda P$ est définie positive et la solution ridge est unique, même si D n'est pas de rang colonne plein :

$$(D^\mathsf{T} D + \lambda P)\widehat{\boldsymbol{\theta}}_\lambda = D^\mathsf{T} \boldsymbol{y}.$$

La régularisation permet un calcul unique même lorsque le plan manque de directions. Elle ne crée toutefois pas l'information mécanique absente des essais.

Validation croisée

Un jeu de test indépendant reste réservé

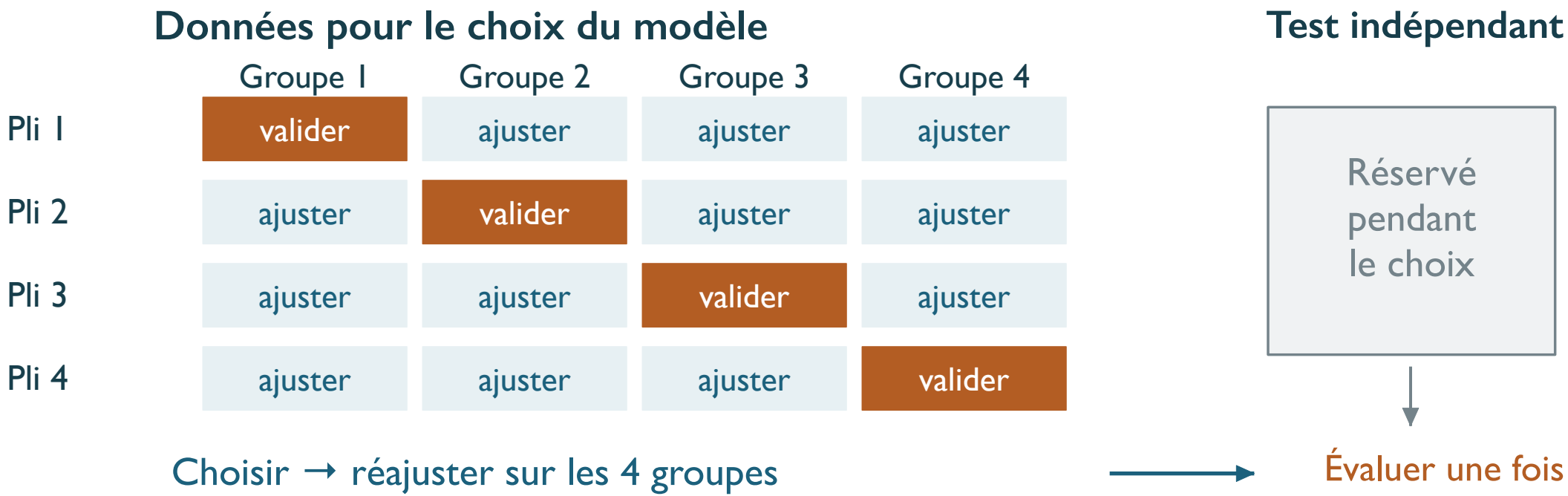
Dans chaque pli, les $K-1$ groupes d'apprentissage servent au centrage, à la réduction et à l'ajustement

L'erreur sur le groupe restant, agrégée sur les K plis, sert à choisir les hyperparamètres

Transformations et modèle sont réajustés sur tout l'apprentissage avant le test final

Le test reste à l'écart du choix

Chaque ligne représente un ajustement complet, transformations comprises.



Le centrage et la réduction sont appris uniquement sur les groupes utilisés pour ajuster.

Mesurer une erreur mécanique

La RMSE s'exprime dans l'unité de la cible

Les résidus et l'erreur maximale localisent les zones difficiles

Le voisinage du seuil mérite un contrôle spécifique

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)^2}$$

Pour une contrainte, la RMSE s'exprime en MPa.

Exemple 2.12 · Une pente rétrécie par ridge

Pour isoler le mécanisme de ridge, considérons une entrée adimensionnée $x = (-1, 0, 1)$ et des réponses $y = (-2, 1, 4)$. La réponse moyenne est $\bar{y} = 1$. Les moindres carrés retrouvent la relation $y = 1 + 3x$, puisque

$$\widehat{\theta}_1 = \frac{\Sigma_i \, x_i (y_i - \bar{y})}{\Sigma_i \, x_i^2} = \frac{6}{2} = 3.$$

Avec une pénalisation $\lambda\theta_1^2$ et $\lambda = 2$, la pente devient $6/(2+2) = 1{,}5$, tandis que l'intercept reste un. L'ajustement des trois points est moins précis. Cet exemple sans bruit montre que régulariser n'améliore pas automatiquement un modèle : la régularisation introduit un biais pour limiter la sensibilité aux données, et son utilité doit être évaluée sur la tâche visée.

Proposition 2.13 · Filtrage spectral de ridge

Après centrage des entrées et de la cible $\boldsymbol{y}_c$, soit $X = \Sigma_{j=1}^{r} s_j \boldsymbol{u}_j \boldsymbol{v}_j^\top$ la décomposition en valeurs singulières de la matrice des entrées, avec $s_j > 0$. Pour $\lambda > 0$, les pentes ridge sont

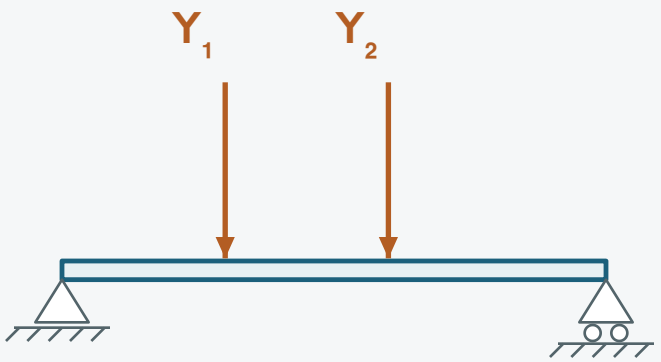
$$\widehat{\boldsymbol{\theta}}_{\lambda,1:p} = \sum_{j=1}^{r} \frac{s_j}{s_j^2 + \lambda} (\boldsymbol{u}_j^\top \boldsymbol{y}_c) \boldsymbol{v}_j.$$

La solution des moindres carrés de norme minimale remplace chaque facteur $s_j/(s_j^2 + \lambda)$ par $1/s_j$. Les petites valeurs singulières sont donc les plus fortement amorties relativement aux moindres carrés.

Un substitut de la réponse mécanique

Un métamodèle remplace une évaluation coûteuse, sous contrôle de sa précision.

Référence mécanique



Calcul ou essai de référence

Jeu de réponses

Entrées : (Y_1, Y_2)

Sortie : σmax

Domaine documenté

Substitut

Prédiction rapide

Incertitude conditionnelle

Validation indépendante

Le substitut accélère la prédiction ; il ne dispense pas de validation mécanique.

Une réponse inconnue au-delà du modèle linéaire

Les moindres carrés ont appris des coefficients d'influence.

Une réponse plus complexe demande d'apprendre une fonction entière.

Le processus gaussien décrit cette fonction encore imparfaitement connue.

Définition 2.14 · Processus gaussien

Une famille de variables aléatoires $(f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$ est un processus gaussien si, pour tout entier $n \geq 1$ et toute collection $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, le vecteur $(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n))$ est conjointement gaussien, éventuellement dégénéré. On note $f \sim \mathcal{GP}(m, k)$, où

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \qquad k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(f(\boldsymbol{x}), f(\boldsymbol{x}')).$$

Ces deux fonctions déterminent toutes les lois de dimension finie du processus.

À charges fixées, la réponse du solveur est déterministe mais encore mal connue. Le GP représente cette connaissance incomplète avant et après les observations.

Définition 2.15 · Noyau de covariance admissible

Une fonction $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un noyau de covariance admissible si elle est symétrique et si, pour tout n, tout choix d'entrées et tout $\boldsymbol{a} \in \mathbb{R}^n$,

$$\sum_{i,j=1}^{n} a_i a_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0.$$

Ainsi toute matrice $K = (k(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j}$ est positive semi-définie. Cette propriété n'impose pas que chacun de ses coefficients soit positif, ni que K soit inversible.

Un noyau relie les réponses possibles à des chargements voisins. Son admissibilité garantit que ces relations forment une covariance cohérente pour tout ensemble d'essais.

Noyau et échelles de variation

L'amplitude fixe l'échelle des réponses possibles dans l'a priori

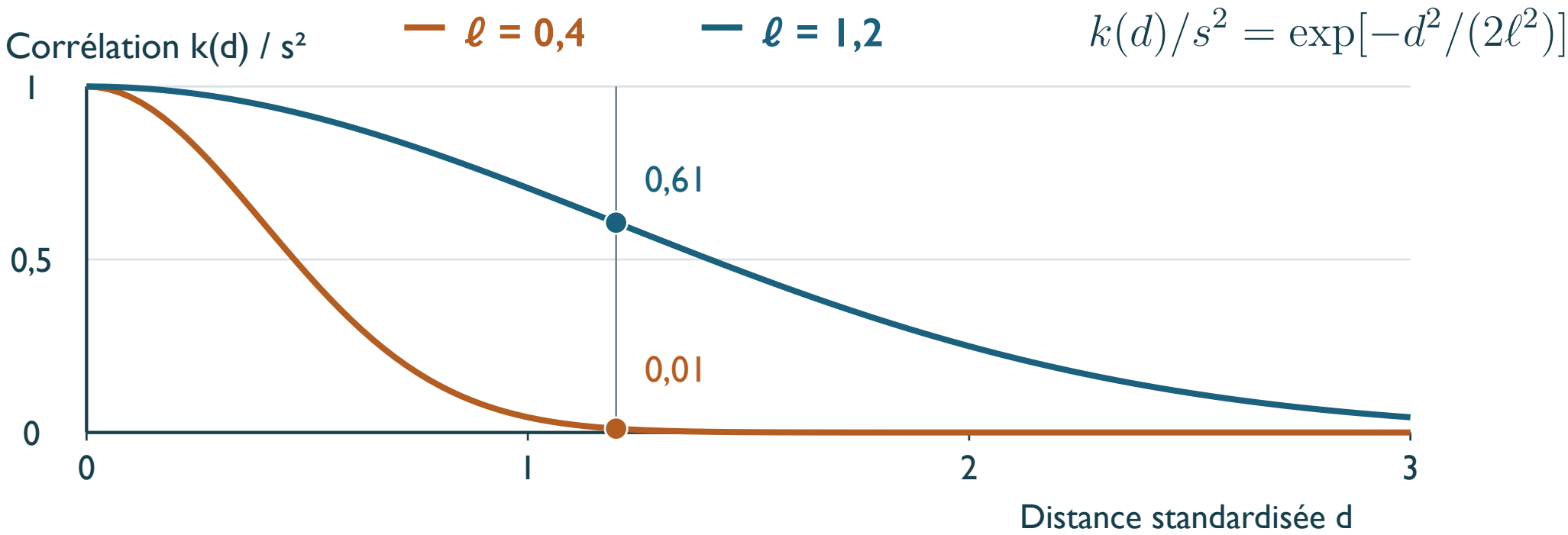
Les longueurs portent les unités des entrées

Une longueur élevée impose une variation lente selon cette direction

$$k(x, x') = s^2 \exp\left[-\frac{1}{2}\sum_j \frac{(x_j - x'_j)^2}{\ell_j^2}\right]$$

La portée d'une observation

La longueur du noyau fixe jusqu'où deux conditions de charge se ressemblent dans l'a priori.



Une grande longueur transmet l'information plus loin ; elle doit rester adaptée à la mécanique.

Régularité et changement de branche

Le maximum de deux réponses lisses peut présenter un angle

Le noyau de Matérn 3/2 impose moins de régularité que le noyau exponentiel quadratique

La précision près de la cassure reste à vérifier

$$r^2 = \sum_j \frac{(x_j - x'_j)^2}{\ell_j^2}, \qquad r \geq 0$$

$$k(x, x') = s^2(1 + \sqrt{3}\, r) \exp(-\sqrt{3}\, r)$$

Remarque 2.16 · Incertitude sur une fonction de réponse

Pour une géométrie et des charges imposées, le solveur de poutre fournit une réponse déterministe. Le GP est ici une représentation de notre connaissance incomplète de cette fonction. Avant les observations, plusieurs courbes sont possibles dans ce modèle ; après les observations, les courbes compatibles avec elles reçoivent davantage de poids. Répéter le calcul exact à une charge fixée ne fait donc pas varier physiquement la contrainte selon le GP.

Lemme 2.17 · Décorrélation de blocs gaussiens

Si les vecteurs $\boldsymbol{U}$ et $\boldsymbol{V}$ sont conjointement gaussiens, alors ils sont indépendants si et seulement si $\mathrm{Cov}(\boldsymbol{U}, \boldsymbol{V}) = 0$. La seule normalité de chaque composante séparément ne suffit pas.

Théorème 2.18 · Conditionnement d'un processus gaussien (1/2)

Soit $f \sim \mathcal{GP}(0, k)$ et des observations à entrées fixées $y_i = f(\boldsymbol{x}_i) + \varepsilon_i$, où les ε_i sont indépendants, de loi $\mathcal{N}(0, s_\varepsilon^2)$ et indépendants du processus f. Les hyperparamètres sont fixés. Posons $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{k}_* = (k(\boldsymbol{x}_i, \boldsymbol{x}_*))_i$ et $k_{**} = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$.

Supposons $K_y = K + s_\varepsilon^2 I$ inversible, ce qui est assuré si $s_\varepsilon^2 > 0$. Alors $f_* = f(\boldsymbol{x}_*)$ vérifie

$$\begin{pmatrix} \boldsymbol{y} \\ f_* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K + s_\varepsilon^2 I & \boldsymbol{k}_* \\ \boldsymbol{k}_*^\mathsf{T} & k_{**} \end{pmatrix}\right).$$

Théorème 2.18 · Conditionnement d'un processus gaussien (2/2)

La loi conditionnelle de la *fonction latente* est $f_* \mid \boldsymbol{y} \sim \mathcal{N}(m_*, s_*^2)$, avec

$$m_* = \boldsymbol{k}_*^\mathsf{T}(K + s_\varepsilon^2 I)^{-1}\boldsymbol{y},$$
$$s_*^2 = k_{**} - \boldsymbol{k}_*^\mathsf{T}(K + s_\varepsilon^2 I)^{-1}\boldsymbol{k}_*.$$

En particulier, $0 \leq s_*^2 \leq k_{**}$.

Corollaire 2.19 · Prédiction d'une nouvelle observation

Sous les hypothèses du théorème 2.18, soit $y_* = f_* + \varepsilon_*$, où $\varepsilon_* \sim \mathcal{N}(0, s^2_{\varepsilon,*})$ est indépendant de f et des erreurs d'apprentissage. Alors

$$y_* \mid \boldsymbol{y} \sim \mathcal{N}\big(m_*, s^2_* + s^2_{\varepsilon,*}\big).$$

La variance d'une nouvelle mesure comprend donc le bruit futur, tandis que s^2_* décrit la fonction latente conditionnellement aux observations.

Une bande sur la fonction ne décrit pas toute la dispersion d'un futur capteur. Pour prédire une nouvelle mesure, il faut ajouter la variance de son bruit indépendant.

Corollaire 2.20 · Interpolation sans bruit

Si $s_\varepsilon^2 = 0$ et K est inversible, le GP retrouve exactement chaque observation : au point $\boldsymbol{x}_* = \boldsymbol{x}_i$, $m_* = y_i$ et $s_*^2 = 0$.

Un calcul mécanique exact doit être retrouvé aux points déjà observés. Cette propriété fournit un contrôle simple du modèle dans le cas sans bruit.

Calcul de la moyenne et de la variance

La factorisation de Cholesky évite l'inverse explicite

Le jitter stabilise un calcul et reste distinct du bruit physique

Une variance nettement négative révèle une anomalie

$$K_y = LL^\mathsf{T}, \qquad Lz = y, \qquad L^\mathsf{T}\alpha = z$$

$$Lv = k_*, \qquad m_* = k_*^\mathsf{T}\alpha, \qquad s_*^2 = k_{**} - v^\mathsf{T}v$$

Remarque 2.21 · Résolution numérique et bruit physique

Calculer les prédictions par factorisation de Cholesky, sans inverser explicitement K_y. Un petit terme diagonal numérique, ou *jitter*, peut stabiliser la factorisation ; il ne constitue pas à lui seul un modèle de bruit physique. S'il est ajouté, les calculs portent sur la matrice perturbée et l'interpolation n'est plus exactement celle du corollaire 2.20.

Une observation exacte : calcul explicite

L'entrée et la réponse sont adimensionnées, avec un a priori centré

Interpolation au point observé ; retour à l'a priori loin de l'observation

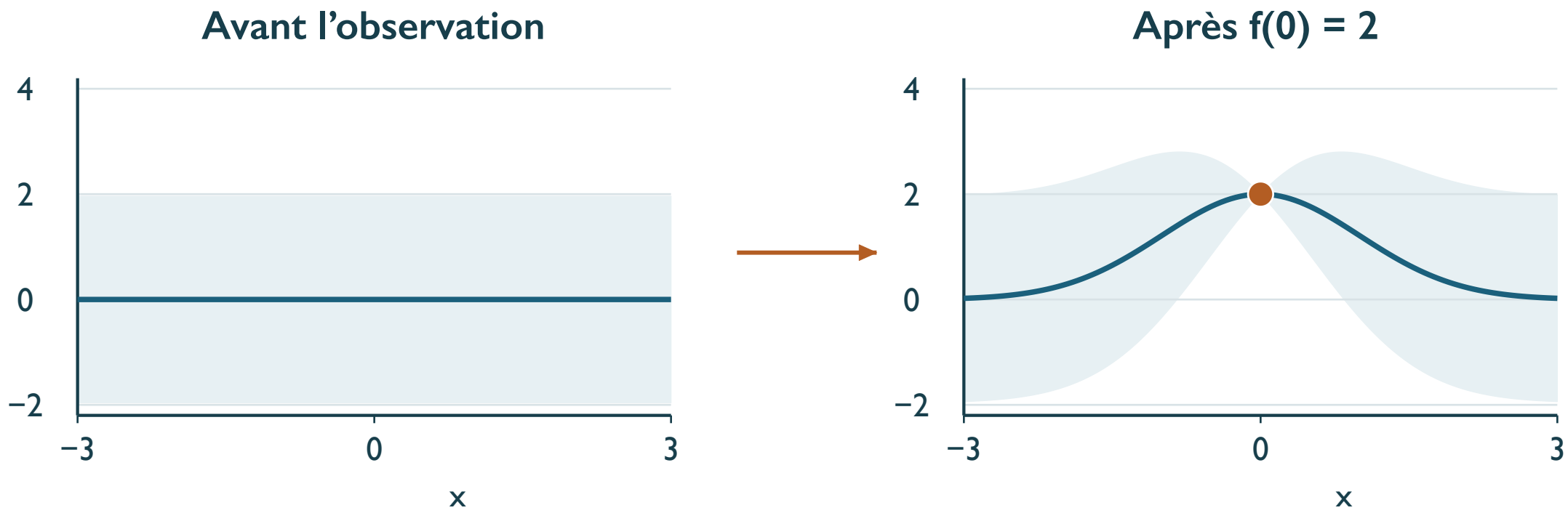
$$f(0) = 2, \qquad k(x, x') = \exp\left[-\frac{(x - x')^2}{2}\right]$$

$$m(x) = 2\exp\left(-\frac{x^2}{2}\right)$$

$$s^2(x) = 1 - \exp(-x^2)$$

Une observation transforme le GP

Adimensionné, noyau $\exp[-(x - x')^2/2]$; courbe moyenne et rubans ponctuels à 95 %.



Au point observé sans bruit, la variance s'annule ; elle revient loin des données.

Exemple 2.22 · Deux observations sans bruit (1/2)

Prenons une entrée et une réponse adimensionnées, un noyau $k(x, x') = \exp(-(x - x')^2/2)$ et deux observations sans bruit $f(0) = 0$, $f(1) = 1$. Notons $c = e^{-1/2}$ et $a = e^{-1/8}$. Au milieu de l'intervalle, $x_* = 1/2$,

$$K = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}, \qquad \boldsymbol{k}_* = \begin{pmatrix} a \\ a \end{pmatrix},$$

$$K^{-1} = \frac{1}{1 - c^2} \begin{pmatrix} 1 & -c \\ -c & 1 \end{pmatrix}.$$

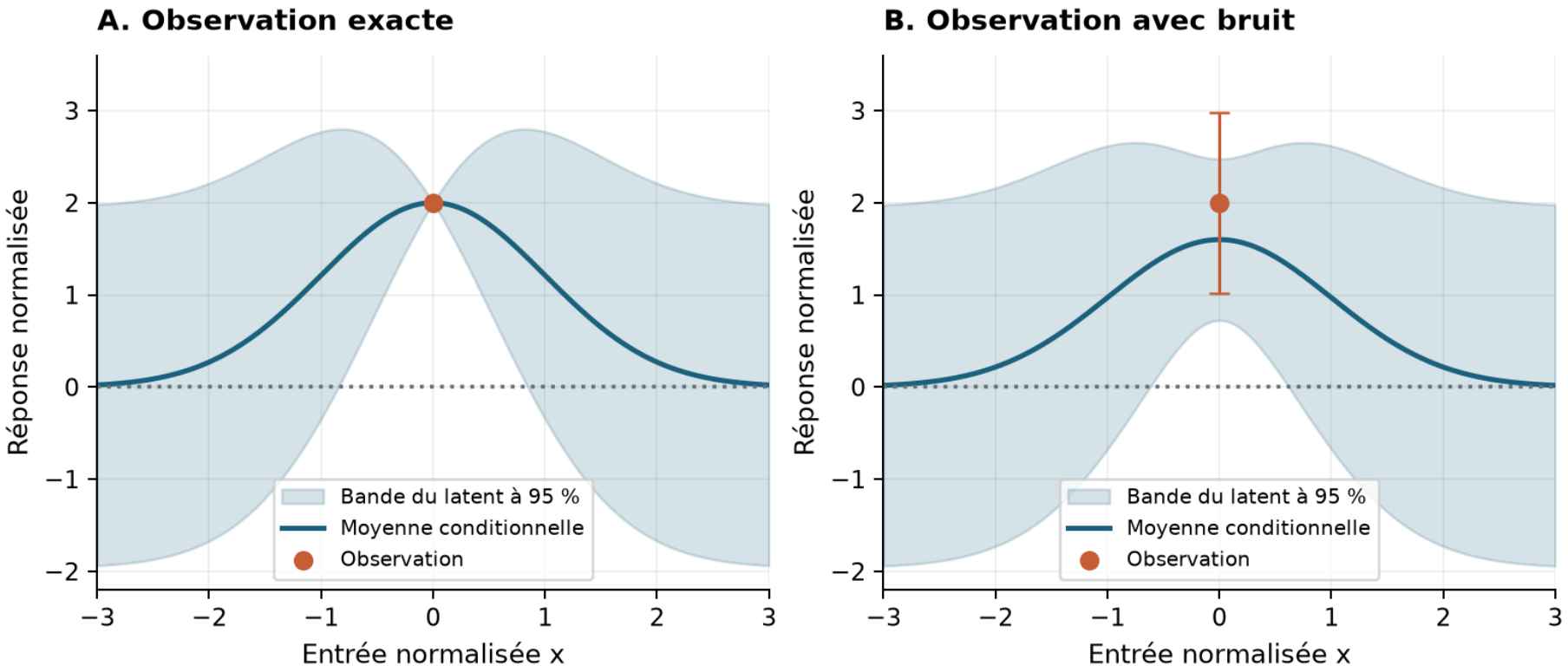
On obtient

$$m_* = \frac{a}{1 + c} \simeq 0{,}5493, \qquad s_*^2 = 1 - \frac{2a^2}{1 + c} \simeq 0{,}0305.$$

Exemple 2.22 · Deux observations sans bruit (2/2)

Le résultat n'est pas simplement l'interpolation affine $1/2$: il dépend du noyau et de la moyenne a priori nulle. Aux points observés, la moyenne retrouve exactement les observations et la variance est nulle, dans ce calcul sans bruit. Cet exemple permet de vérifier une implémentation avant d'utiliser un optimiseur.

Réponse latente et nouvelle mesure



La bande décrit l'incertitude sur la fonction latente

Pour une nouvelle mesure, le bruit indépendant ajoute sa variance

Proposition 2.23 · Prédiction conjointe en plusieurs points

Sous les hypothèses du théorème 2.18, soit $\boldsymbol{f}_*$ le vecteur des réponses latentes en un ensemble fini de points tests. Notons $K_{X*} = \mathrm{Cov}(\boldsymbol{y}, \boldsymbol{f}_*)$ et $K_{**} = \mathrm{Cov}(\boldsymbol{f}_*)$. La loi conditionnelle de $\boldsymbol{f}_*$ est gaussienne de moyenne $K_{X*}^{\mathsf{T}} K_y^{-1} \boldsymbol{y}$ et de covariance

$$\mathrm{Cov}(\boldsymbol{f}_* \mid \boldsymbol{y}) = K_{**} - K_{X*}^{\mathsf{T}} K_y^{-1} K_{X*}.$$

Une courbe prédite doit rester cohérente entre plusieurs chargements. La covariance conditionnelle conserve ces liens, au lieu de traiter chaque point comme une fonction indépendante.

Complément · une cible Monte-Carlo

La variance d'une moyenne dépend du nombre de trajectoires indépendantes

Chaque entrée peut avoir une précision d'estimation différente

Cette variance d'estimation fixe la variance du bruit d'observation du GP

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$$

$$\text{Variance d'estimation} \simeq \frac{s_i^2}{N_i}$$

Charges aléatoires et incertitude du GP

À charges fixées, le GP décrit l'incertitude sur la réponse

La variation des charges disperse aussi sa moyenne prédite

Le bruit indépendant s'ajoute pour une nouvelle mesure

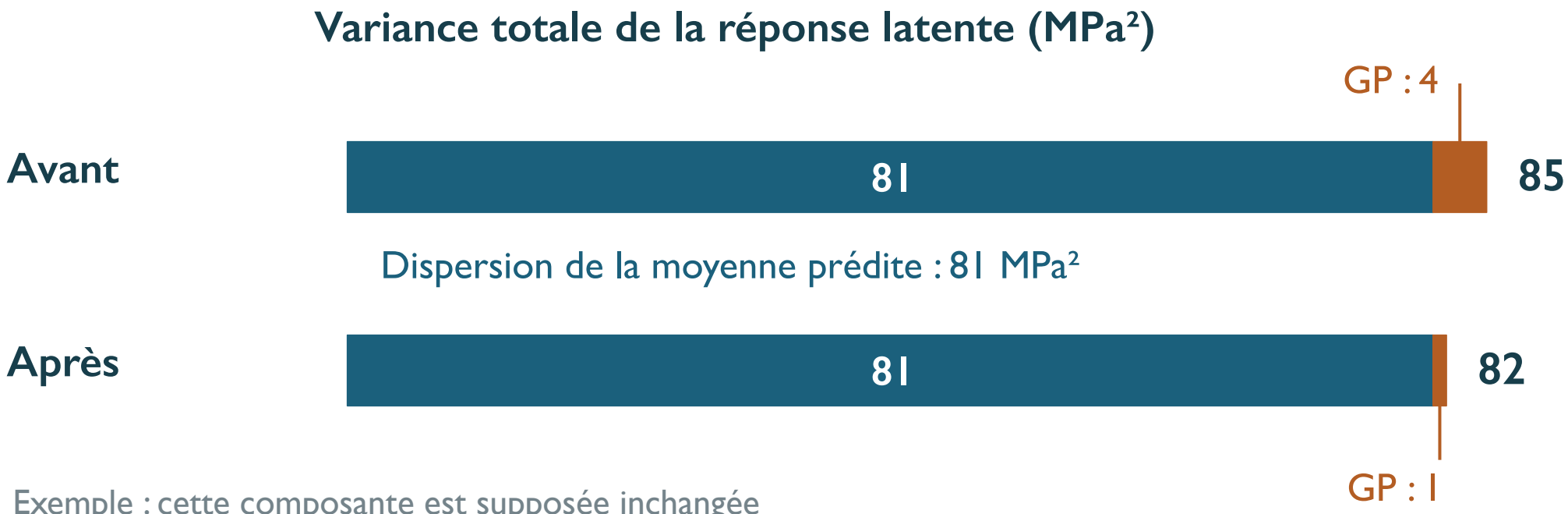
$$\mathrm{Var}(F(\mathbf{Y}) \mid \mathcal{D})$$

$$= \mathbb{E}_{\mathbf{Y}|\mathcal{D}}[s^2_{\mathcal{D}}(\mathbf{Y})] + \mathrm{Var}_{\mathbf{Y}|\mathcal{D}}[m_{\mathcal{D}}(\mathbf{Y})]$$

Charges et GP indépendants conditionnellement aux données ; moments d'ordre deux finis

Plus d'essais, mêmes charges

Illustration : l'incertitude du GP diminue, la loi des sollicitations reste la même.



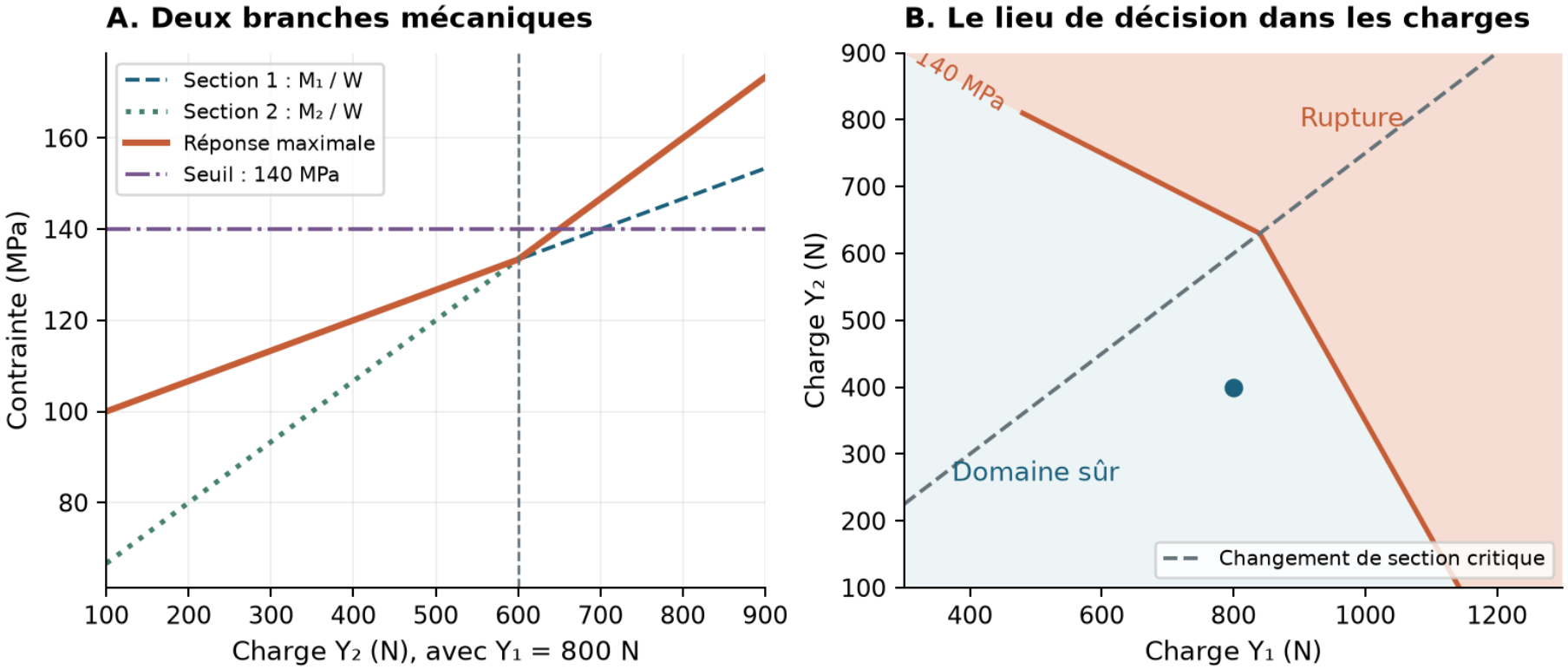
Mieux connaître la réponse ne réduit pas, à lui seul, la variabilité des efforts.

Remarque 2.24 · Portée d'un intervalle crédible

Pour $s_* > 0$, l'intervalle $m_* \pm 1{,}96 s_*$ a une probabilité conditionnelle d'environ $95\,\%$ pour la réponse latente au point fixé. C'est un intervalle crédible *ponctuel* sous le modèle retenu et ses hyperparamètres fixés. Ce n'est ni une bande simultanée sur toutes les entrées, ni une garantie de couverture de la réponse mécanique vraie. Vérifier empiriquement les erreurs, les zones d'extrapolation et les désaccords de classification autour du seuil admissible.

Le ruban exprime une incertitude ponctuelle sous le modèle choisi. Une validation mécanique reste nécessaire, surtout près du seuil de rupture et hors du domaine appris.

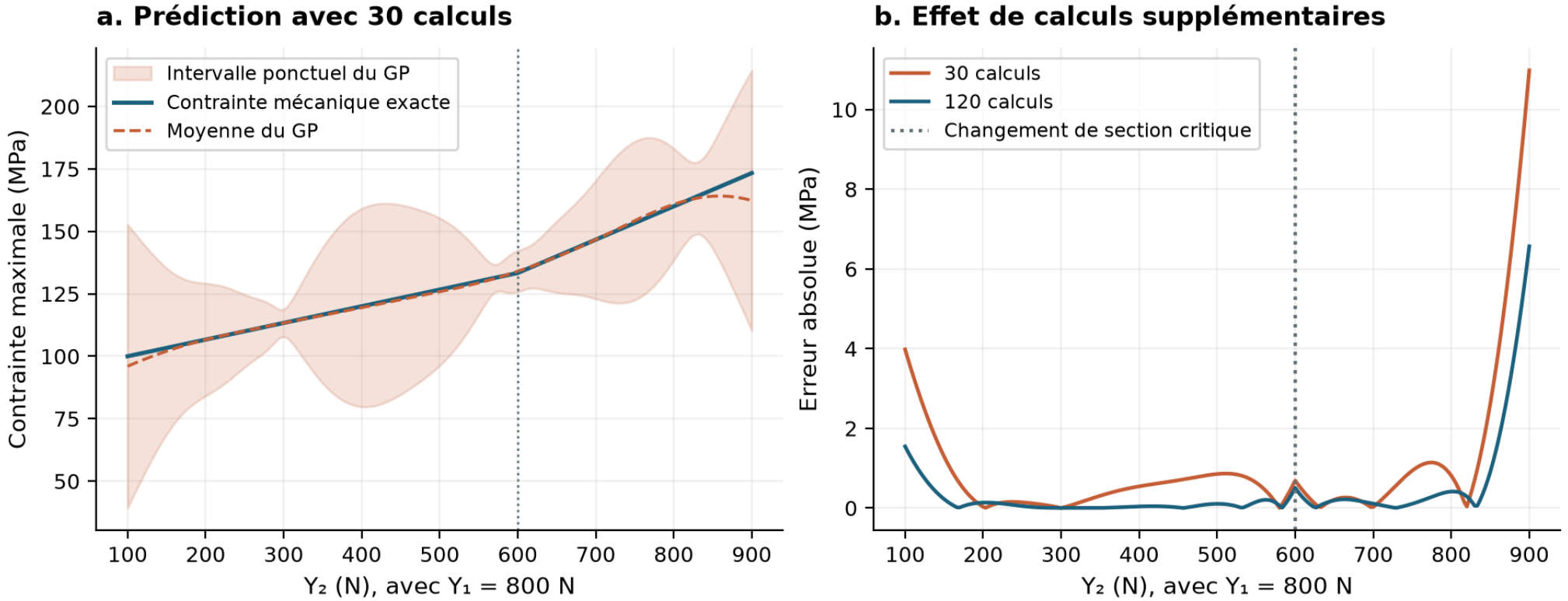
Section critique et seuil de rupture



Pour des charges positives, les branches se croisent sur $Y_2 = 0{,}75 Y_1$

À $Y_1 = 800$ N, la cassure est à 600 N et le seuil de 140 MPa à 650 N

Budget d'apprentissage et précision



À noyau fixé, 30 puis 120 calculs abaissent ici la RMSE de 1,99 à 1,02 MPa

L'amélioration varie selon la position

Un test indépendant de 400 charges

Ici, les hyperparamètres sont ajustés sur 120 couples ; le test en utilise 400 autres

Le taux observé de 87 % reste inférieur au niveau nominal de 95 %

$$\mathrm{RMSE}_{\mathrm{test}} \simeq 0{,}365\,\mathrm{MPa}$$

$$\text{Couverture ponctuelle observée : } \frac{348}{400} = 87\,\%$$

Rupture · deux classifications comparées

5000 charges communes : 479 ruptures exactes et 487 prédites

Les 8 désaccords sont des faux positifs dans cet échantillon

Ce résultat ne garantit pas l'absence de défaillance manquée en exploitation

Erreur de réponse et décision de rupture

Une erreur d'au plus ε conserve la décision à plus de ε du seuil

Le désaccord dépend aussi des charges proches du seuil

$$I = \mathbf{1}_{\{f(\mathbf{Y}) \geq \sigma_a\}}, \quad \widehat{I} = \mathbf{1}_{\{\widehat{f}(\mathbf{Y}) \geq \sigma_a\}}$$

$$|P_f - P_f^{\mathrm{app}}| \leq \mathbb{P}(I \neq \widehat{I}), \qquad \varepsilon > 0$$

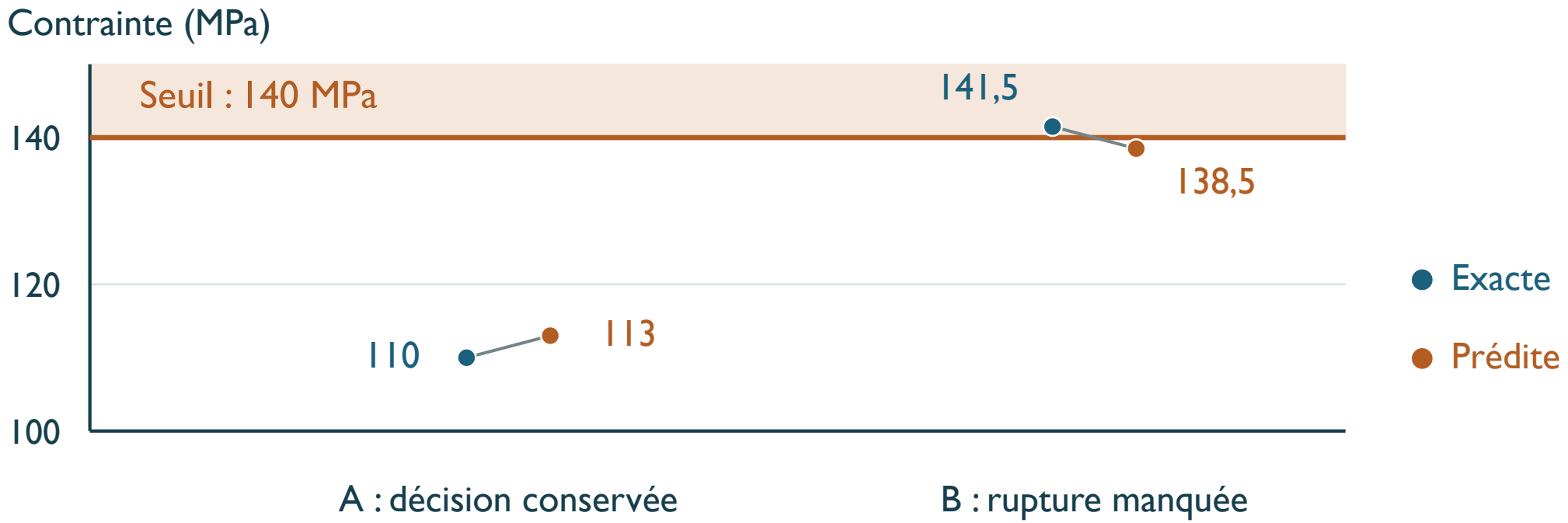
$$\mathbb{P}(I \neq \widehat{I}) \leq \mathbb{P}(|f(\mathbf{Y}) - \sigma_a| \leq \varepsilon)$$

$$+\mathbb{P}(|\widehat{f}(\mathbf{Y}) - f(\mathbf{Y})| > \varepsilon)$$

Même loi physique des charges et même convention de rupture

Une erreur, deux conséquences

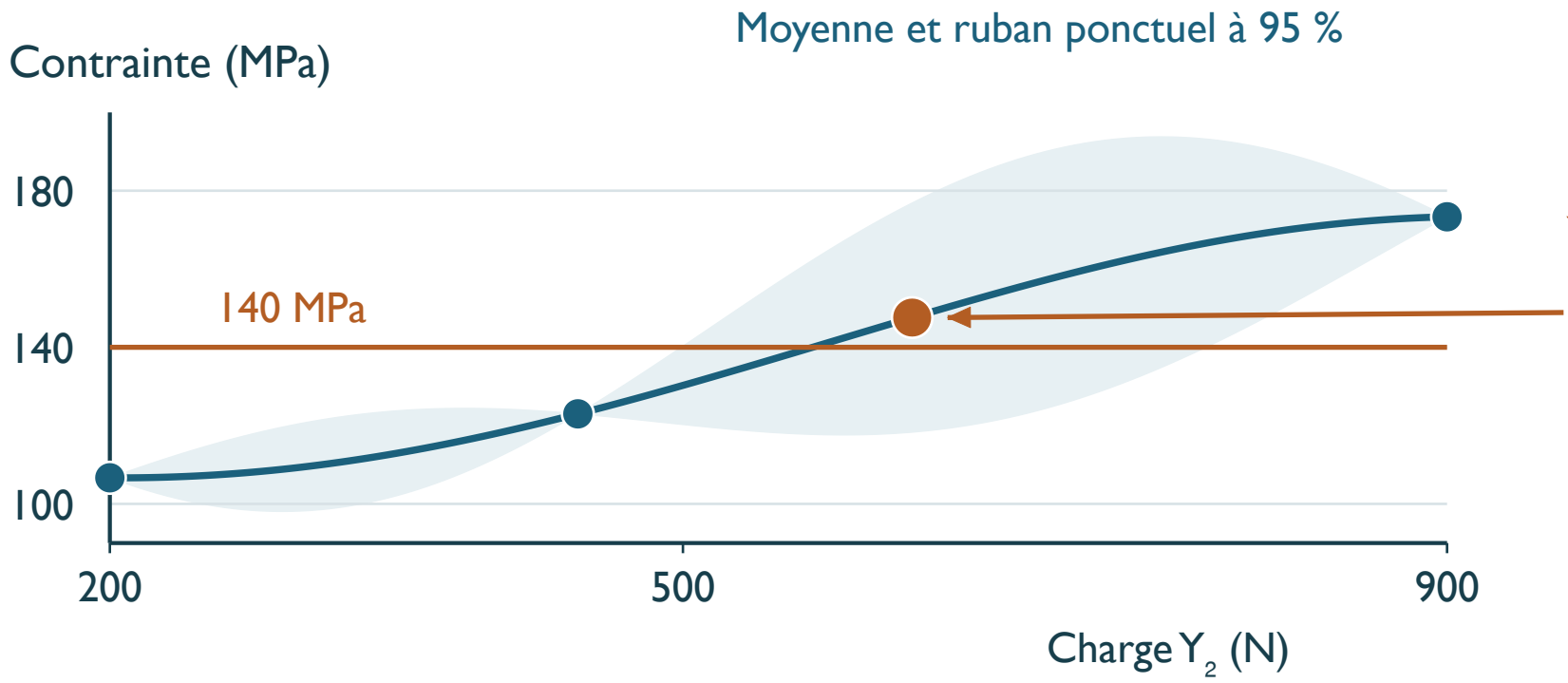
Les deux prédictions ont une erreur absolue de 3 MPa.



Près du seuil, une faible erreur peut transformer une rupture en annonce de sécurité.

Où ajouter un calcul mécanique ?

Y_1 = 800 N fixé ; trois réponses connues, un budget pour une quatrième.



Point proposé

Y_2 = 620 N

Priorité illustrative

$$\frac{|m - \sigma_a|}{s} \quad (s > 0)$$

Un calcul près du seuil précise la décision ; le test indépendant reste nécessaire.

Chapitre 3 · Le chargement devient un profil spatial

Ici, l'entrée était un couple de charges : le GP décrivait une réponse inconnue.

Au chapitre 3, une réalisation sera un profil de charge le long de la poutre.

La covariance reste centrale, mais son indice et son sens physique changent.

Références et repères du cours

James et al., An Introduction to Statistical Learning with Applications in Python, 2023

Rasmussen et Williams, Gaussian Processes for Machine Learning, 2006

Polycopié révisé : chapitre 2, résultats 2.1–2.24