

Quantification et mitigation de la fatigue d'alarmes causée par les systèmes de détection des défauts

Abdoul R. Diallo¹, Lazhar Homri¹, Jean-Yves Dantan¹, Frédéric Bonnet², Thomas Boeuf²

¹ *Arts et Metiers Institute of Technology, Université de Lorraine, LCFC, Metz, France*

² *ArcelorMittal Research SA, Voie romaine, Maizières-lès-Metz, France*

22/05/2025

Sûreté / Surveillance / Supervision (S3)

Bagages abandonnés



360 objets oubliés par semaine dans les gares et les trains



2022

12 millions de voyageurs
35500 trains impactés

Procédure en cas de bagages abandonnés

1

Annonce sonore

2

Périmètre de sécurité

3

Equipe de démineurs

<https://www.groupe-sncf.com/fr/groupe/coulisses/circulation-trains/bagage-abandonne>

Pictogrammes générés par une IA générative

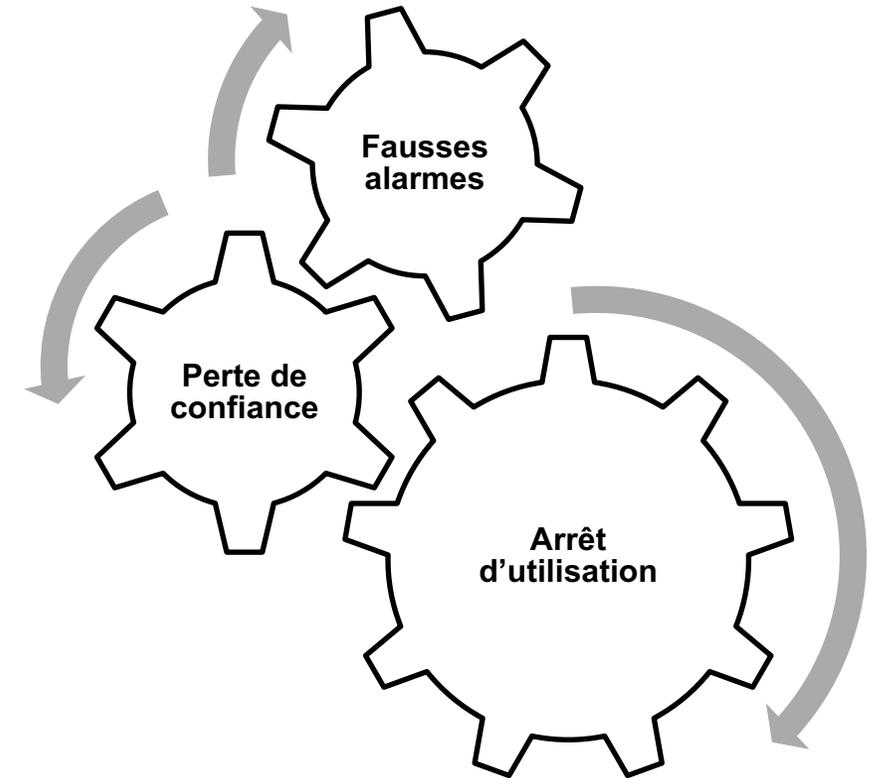
Fatigue d'alarmes

13 février 2023



Il ne voulait pas arriver en retard à Paris : le passager d'un TGV ouvre lui-même un « colis suspect » en gare de Reims

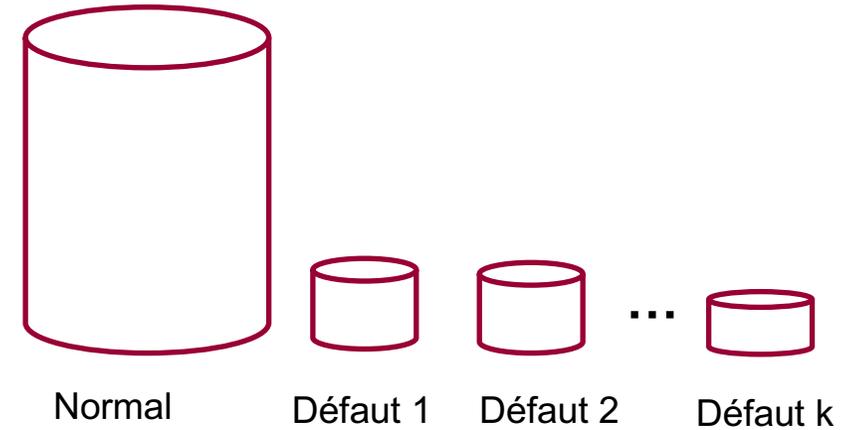
Alors que le train s'apprêtait à être évacué, un passager a lui-même ouvert un sac à main oublié qui ne contenait que des chewing-gums et des post-it.



Quid des modèles de détection des défauts dans l'industrie ?

Questions de recherche

1. Quels sont les indicateurs utilisés pour l'évaluation des modèles?
 1. Ces indicateurs mesurent-ils la fatigue d'alarmes ?
 2. Comment mesurer dans le cas contraire la fatigue d'alarmes ?
2. Comment prévenir ou à minima réduire la fatigue d'alarmes ?



Déséquilibre des données

- Cartes de contrôles
- Méthodes statistiques multivariées (Analyse en composantes principales (ACP), Moindres carrés partiels, analyse en composantes indépendantes...)
- Autoencodeur et ses variantes
- Classifieurs à une classe (Isolation Forest, One-Class Support Vector Machine)

Abid, A., Khan, M. T., & Iqbal, J. (2021). A review on fault detection and diagnosis techniques: basics and beyond. *Artificial Intelligence Review*, 54(5), 3639–3664.

Li, W., Gu, S., Zhang, X., & Chen, T. (2020). Transfer learning for process fault diagnosis : Knowledge transfer from simulation to physical processes. *Computers and Chemical Engineering*, 139, 106904.

Evaluation des modèles de détection

Matrice de confusion



Taux de fausses alarmes

$$FAR = \frac{FP}{FP + TN}$$

True negative rate

$$TNR = \frac{TN}{TN + FP} = 1 - FAR$$

Taux de détection des défauts

Recall $TPR = \frac{TP}{TP + FN}$ Sensitivity

True positive rate

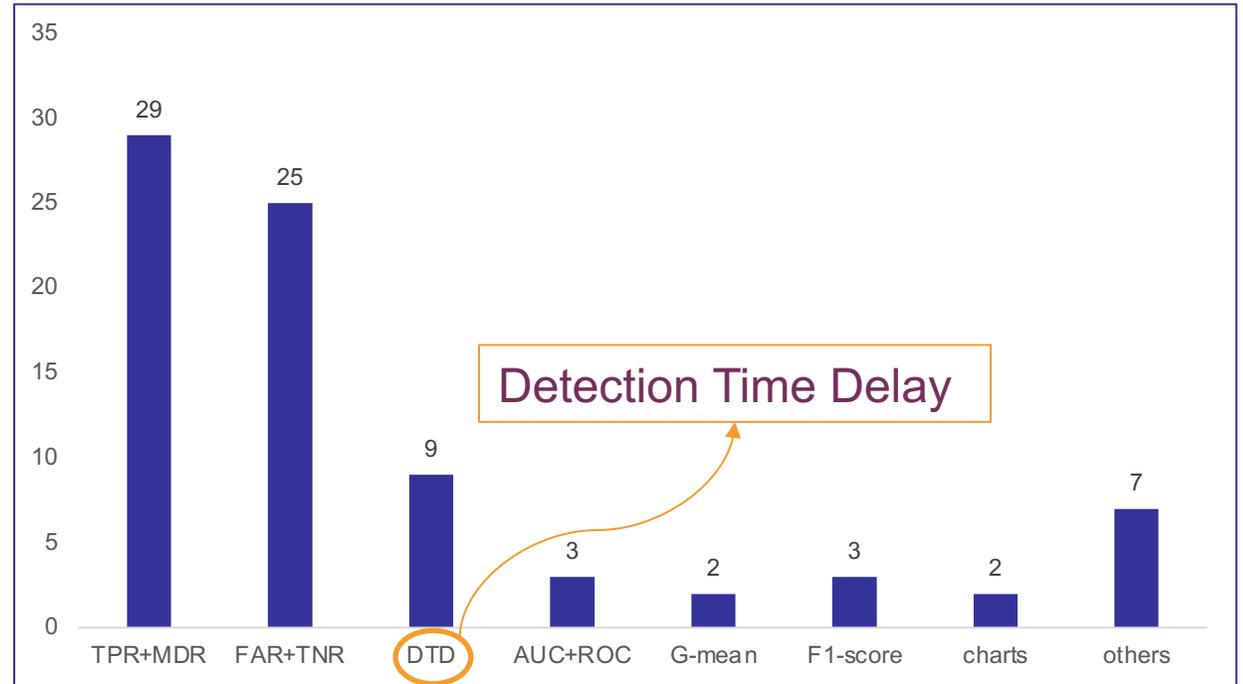
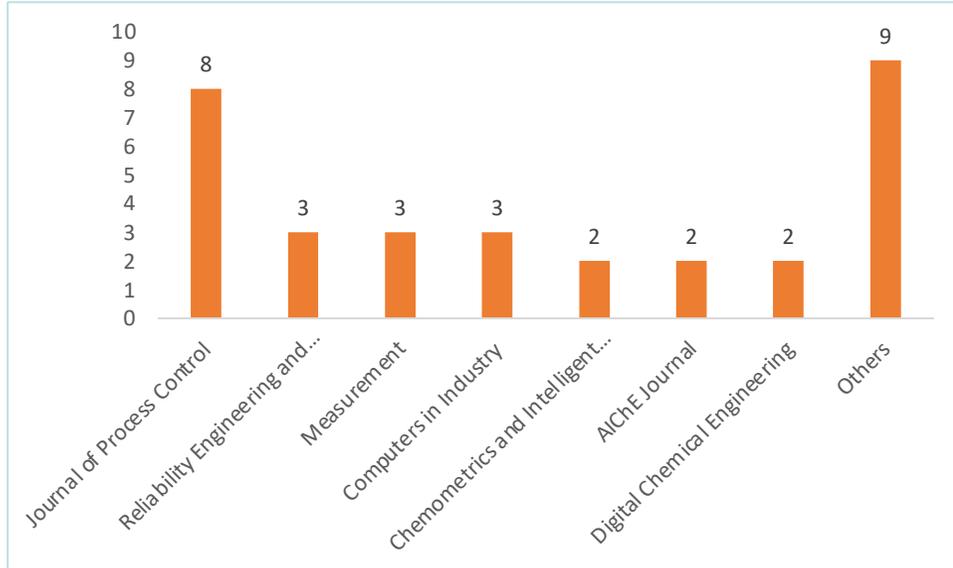
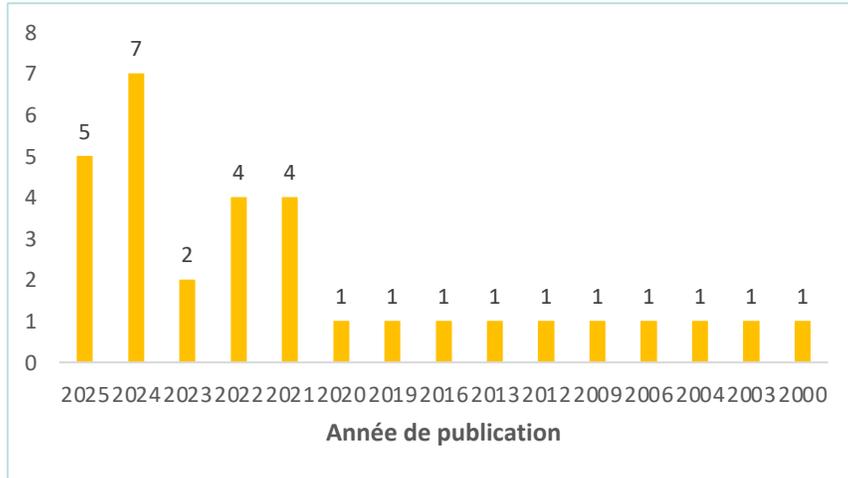
Missed detection rate

$$MDR = \frac{FN}{FN + TP} = 1 - TPR$$

Non-detection rate

Missed alarm rate

Evaluation des modèles de détection



Le taux de détection des défauts (TPR+MDR) et le taux de fausses alarmes (FAR+TNR) sont les indicateurs les plus utilisés pour l'évaluation des modèles.

Ces indicateurs mesurent-ils la fatigue d'alarmes ?

Mesure de la fatigue d'alarmes

Notations

NOC : Le process fonctionne normalement

A : Le modèle déclenche une alarme

- $\mathbb{P}(A \setminus \text{NOC})$: taux de fausses alarmes
- $\mathbb{P}(A \setminus \overline{\text{NOC}})$: taux de détection des défauts
- $\mathbb{P}(\text{NOC} \setminus A)$: fatigue d'alarmes

$$\mathbb{P}(\text{NOC} \setminus A) = \frac{\mathbb{P}(A \setminus \text{NOC}) \cdot \mathbb{P}(\text{NOC})}{\mathbb{P}(A)}$$

Hypothèses

- $\mathbb{P}(\text{NOC}) \gg \mathbb{P}(\overline{\text{NOC}})$
- Système de détection « parfait » :
 - Taux de détection des défauts = 100%
 - Taux de fausses alarmes = α

Mesure de la fatigue d'alarmes

Notations

NOC : Le process fonctionne normalement

A : Le modèle déclenche une alarme

- $\mathbb{P}(A \setminus \text{NOC})$: taux de fausses alarmes
- $\mathbb{P}(A \setminus \overline{\text{NOC}})$: taux de détection des défauts
- $\mathbb{P}(\text{NOC} \setminus A)$: fatigue d'alarmes

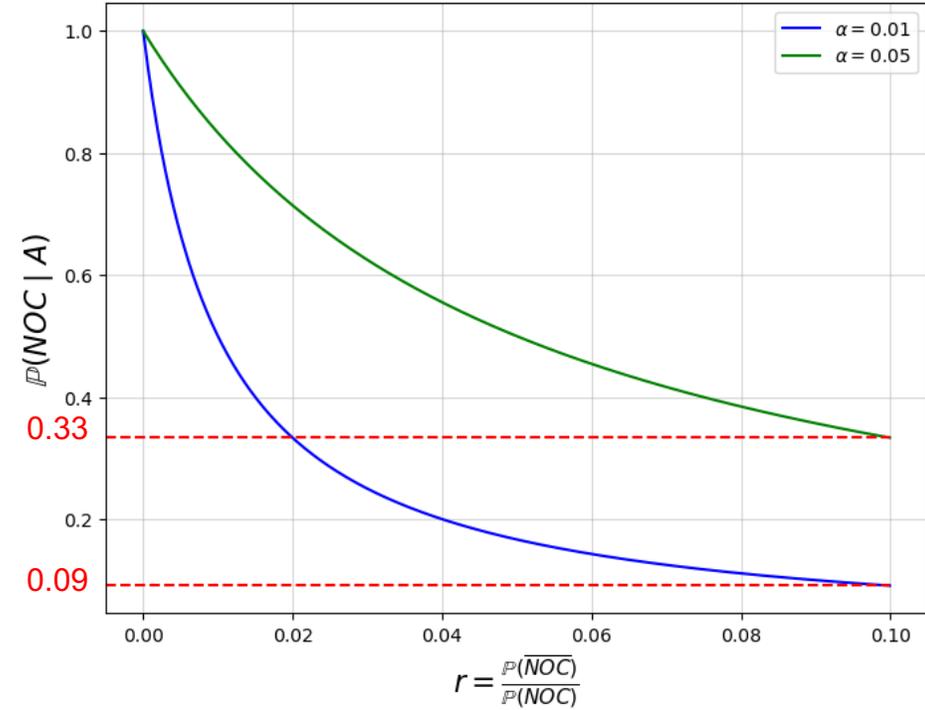
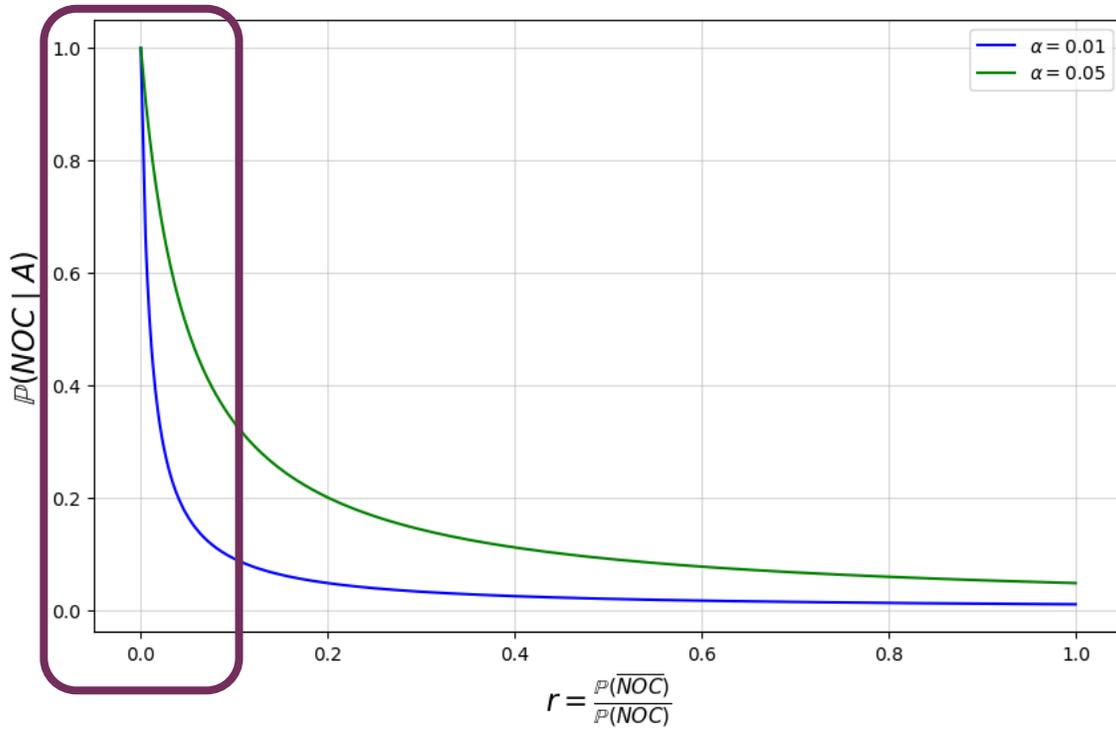
Hypothèses

- $\mathbb{P}(\text{NOC}) \gg \mathbb{P}(\overline{\text{NOC}})$
- Système de détection « parfait » :
 - Taux de détection des défauts = 100%
 - Taux de fausses alarmes = α

$$\mathbb{P}(\text{NOC} \setminus A) = \frac{\mathbb{P}(A \setminus \text{NOC}) \cdot \mathbb{P}(\text{NOC})}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \setminus \text{NOC}) \cdot \mathbb{P}(\text{NOC})}{\mathbb{P}(A \setminus \text{NOC}) \cdot \mathbb{P}(\text{NOC}) + \mathbb{P}(A \setminus \overline{\text{NOC}}) \cdot \mathbb{P}(\overline{\text{NOC}})} = \frac{\mathbb{P}(A \setminus \text{NOC})}{\mathbb{P}(A \setminus \text{NOC}) + \mathbb{P}(A \setminus \overline{\text{NOC}}) \frac{\mathbb{P}(\overline{\text{NOC}})}{\mathbb{P}(\text{NOC})}}$$

\parallel
 r

Mesure de la fatigue d'alarmes pour un système « parfait »



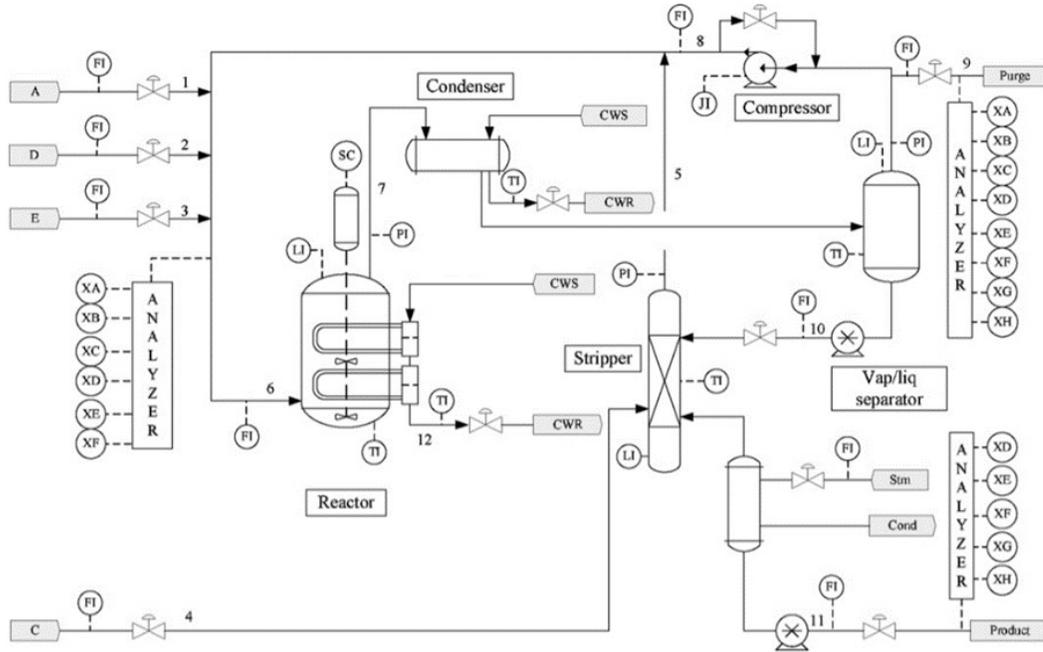
- Le déséquilibre a un grand effet sur $\mathbb{P}(NOC \setminus A)$

- Il existe un grand écart entre $\mathbb{P}(NOC \setminus A)$ et le FAR

$\mathbb{P}(NOC \setminus A)$ = Taux de fausses découvertes (FDR)

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Illustration avec le Tennessee Eastman Process (TEP)



Evaluation du modèle

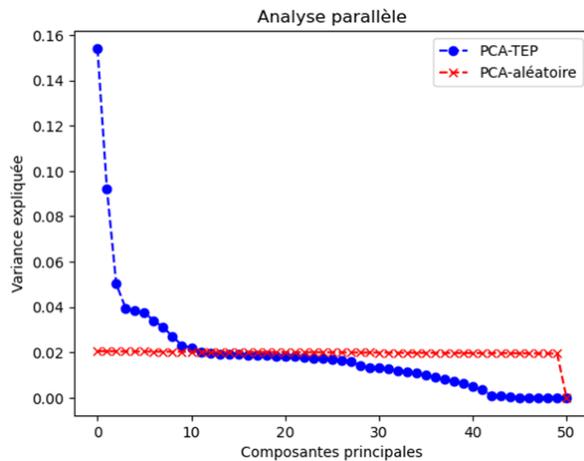
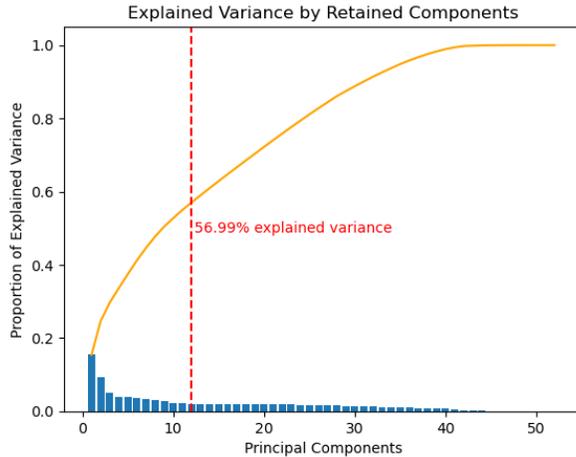
- 500 x 960 échantillons de fonctionnement normal
- 800 échantillons pour chaque défaut

- Fréquence d'échantillonnage : 3 minutes
- Durée d'une simulation
 - Entraînement : 25 heures ➡ 500 échantillons
 - Evaluation : 48 heures ➡ 960 échantillons
- Nombre total de simulations : 500
- Nombre de classe :
 - Une classe de fonctionnement normal
 - 20 classes de défauts
- Taille des données pour chaque classe:
 - Entraînement : 500 x 500
 - Evaluation : 500 x 960

Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 17(3), 245–255. [https://doi.org/10.1016/0098-1354\(93\)80018-](https://doi.org/10.1016/0098-1354(93)80018-)

Rieth, Cory A.; Amsel, Ben D.; Tran, Randy; Cook, Maia B., 2017, "Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation", <https://doi.org/10.7910/DVN/6C3JR1>, Harvard Dataverse, V1

Résultat sur le TEP



$$Q_\alpha = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}}$$

Traditionnelle

$\alpha = 1\%$

$FAR = 0,94\%$

$TPR = 61,81\%$

$FDR = 31,39\%$

Prédiction conforme (Marginale)

$FAR = 1,02\%$

$TPR = 62,09\%$

$FDR = 32,93\%$

$\alpha = 5\%$

$FAR = 4,96\%$

$TPR = 67,03\%$

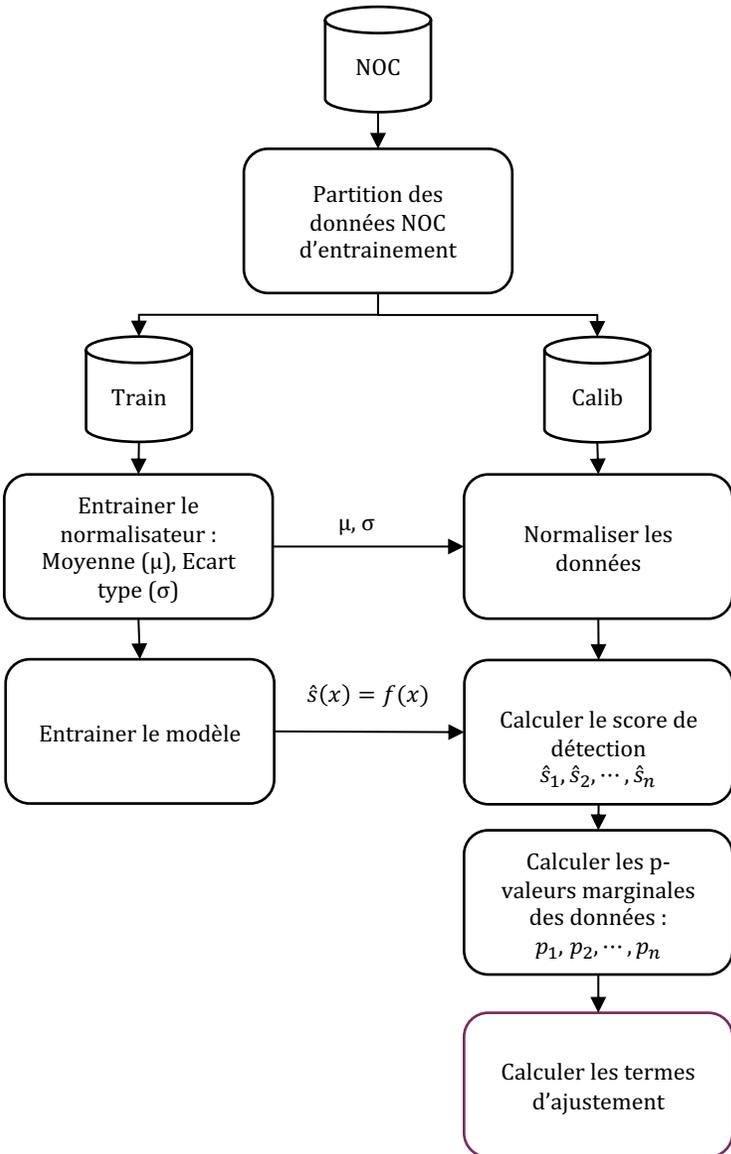
$FDR = 68,96\%$

Jackson, J. E., & Mudholkar, G. S. (1979). Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, 21(3), 341. <https://doi.org/10.2307/1267757>

Taux de détection (Traditionnelle)

Fault	$\alpha = 1\%$	$\alpha = 5\%$
Fault 1	99.62	99.88
Fault 2	98.75	98.88
Fault 3	1.25	6.12
Fault 4	98.50	99.62
Fault 5	11.12	19.12
Fault 6	100	100
Fault 7	99.88	100
Fault 8	97.38	98.12
Fault 9	1.38	5.88
Fault 10	23.62	43.00
Fault 11	70.25	79.38
Fault 12	88.25	92.50
Fault 13	93.12	94.00
Fault 14	100	100
Fault 15	1.25	6.00
Fault 16	9.25	24.00
Fault 17	89.00	93.00
Fault 18	93.50	93.62
Fault 19	15.38	34.00
Fault 20	44.75	53.50

Réduction de la fatigue d'alarmes



P-valeur marginale

$$p(x) = \frac{1 + |\{X_i \in X_{calib} : \hat{s}(X_i) \geq \hat{s}(x)\}|}{n + 1}$$

Validité marginale

$$\mathbb{P}[p(x) \leq \alpha] \leq \alpha$$

P-valeur Conditionnelle

$$p'(x) = b_{[(n+1)p(x)]}, \quad p(x) \in [0,1]$$

Validité conditionnelle

$$\mathbb{P}[\mathbb{P}[p'(x) \leq \alpha] \leq \alpha] \geq 1 - \delta$$

$$b_0 = 0, \quad b_{n+1} = 1 \quad 0 \leq b_1 \leq b_2 \leq \dots \leq b_n \leq 1$$

Méthode DKW

$$b_i = \min \left\{ \frac{i}{n} + \sqrt{\ln \frac{C}{\delta} \frac{1}{2n}}, 1 \right\}$$

Méthode de Simes

$$b_{n+1-i} = 1 - \delta^{1/k} \left(\frac{i \cdots (i - k + 1)}{n \cdots (n - k + 1)} \right)^{1/k}$$

Méthode Asymptotique

$$b_i = \min \left\{ \frac{i}{n} + c_n(\delta) \frac{\sqrt{i(n-i)}}{n\sqrt{n}}, 1 \right\}$$

Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1956). Asymptotic Minimax Character of the Sample Distribution Function and of the Classical. *The Annals of Mathematical Statistics*, 27(3), 642–669. <https://www.jstor.org/stable/2237374>

Sarkar, S. K. (2008). Generalizing Simes' test and Hochberg's stepup procedure. *The Annals of Statistics*, 36(1), 337–363. <https://doi.org/10.1214/009053607000000550>

Bates, S., Candès, E., Lei, L., Romano, Y., & Sesia, M. (2023). Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1). <https://doi.org/10.1214/22-AOS2244>

Réduction de la fatigue d'alarmes

On dispose de m p-valeurs :

$$p_1, p_2, \dots, p_m$$

Ranger les p-valeurs par ordre croissant :

$$p_{(1)} < p_{(2)} \leq \dots \leq p_{(m)}$$

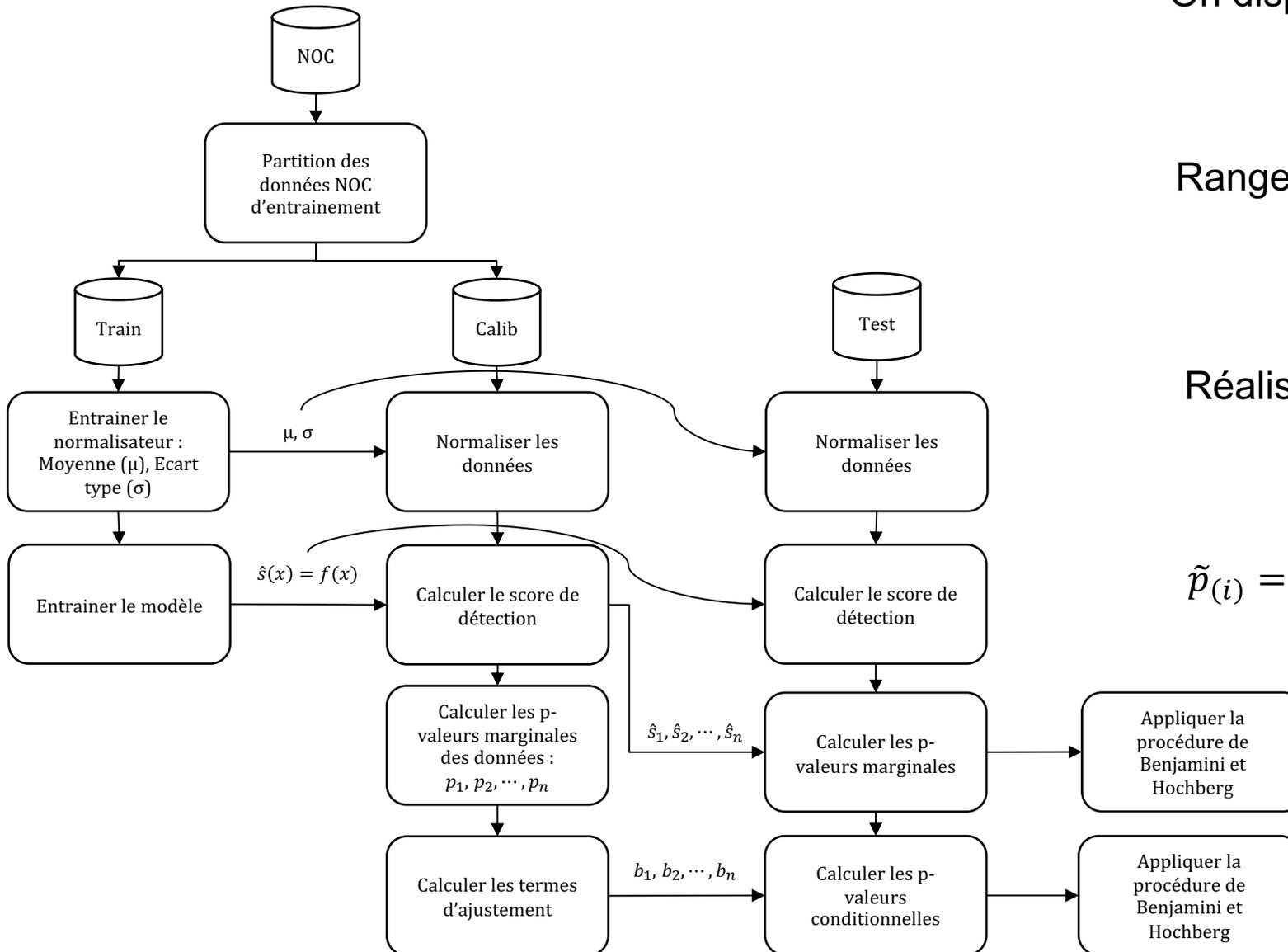
Réaliser la transformation suivante :

$$\tilde{p}_{(m)} = p_{(m)}$$

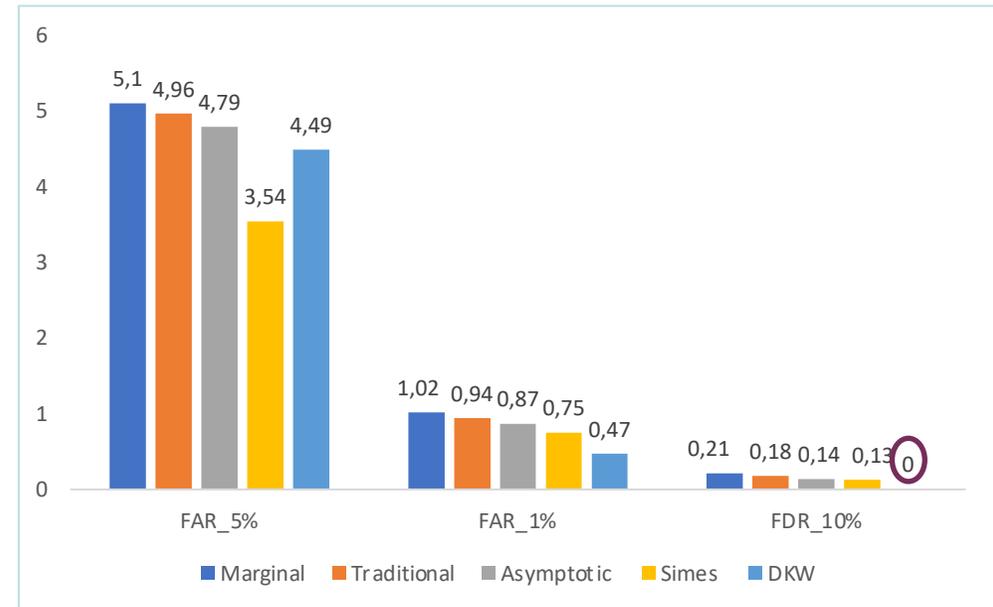
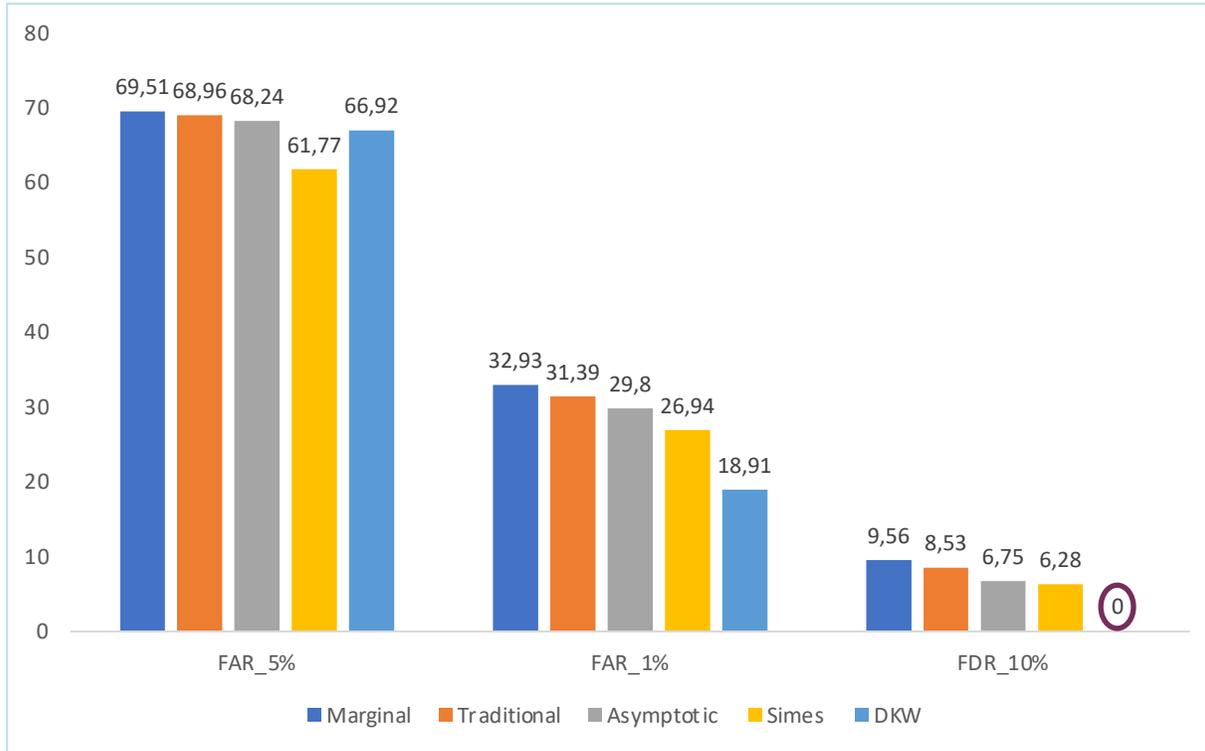
$$\tilde{p}_{(i)} = \min \left\{ \frac{m}{i} p_{(i)}, \tilde{p}_{(i+1)} \right\}, i = 1, \dots, m - 1$$

Une alarme est déclenchée si

$$\tilde{p}_{(i)} \leq q$$



Réduction de la fatigue d'alarme - TEP



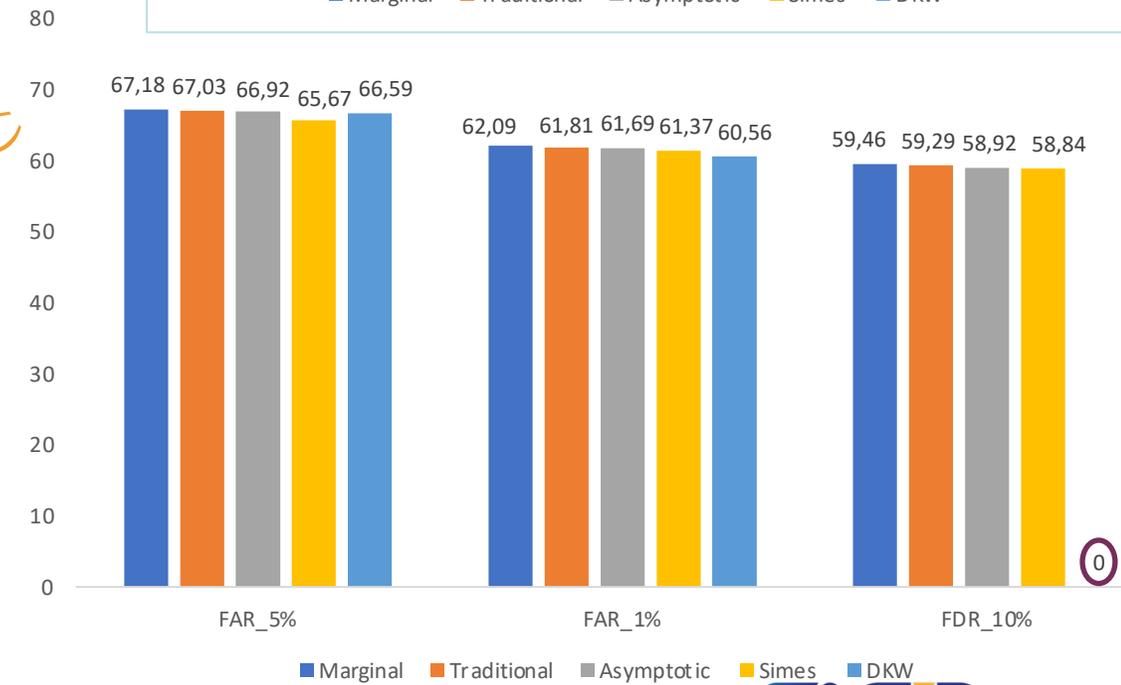
FAR (%)

TPR (%)

Evolution du FDR (%)

Pourcentage de réduction des différents indicateurs

Approach	1%			5%		
	FAR	FDR	Mean TPR	FAR	FDR	Mean TPR
Traditional	80,85	72,83	4,08	96,37	87,63	11,55
Marginal	79,41	70,97	4,24	95,88	86,25	11,49
Simes ($\delta=0.1$)	82,67	76,69	4,12	96,33	89,83	10,4
Asymptotic ($\delta=0.1$)	83,91	77,35	4,49	97,08	90,11	11,95
DKW ($\delta=0.1$)	100	100	100	100	100	100



Conclusions et perspectives

- Ce travail permet de mesurer et de diminuer la fatigue d'alarmes
- Il a été aussi appliqué sur une ligne de galvanisation
- Le contrôle du FDR en temps réel ?
- Quel risque de fatigue d'alarmes peut – on tolérer ?
- Dans certaines applications, le coût d'un défaut non détecté est énorme

Les patients et les professionnels de la santé seraient prêts à accepter 2 250 faux positifs supplémentaires pour une vraie détection de cancer de plus

Halligan, S., Altman, D. G., & Mallett, S. (2015). Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *European Radiology*, 25(4), 932. <https://doi.org/10.1007/S00330-014-3487-0>

Quantification et mitigation de la fatigue d'alarmes causée par les systèmes de détection des défauts

Abdoul R. Diallo¹, Lazhar Homri¹, Jean-Yves Dantan¹, Frédéric Bonnet², Thomas Boeuf²

¹ Arts et Metiers Institute of Technology, Université de Lorraine, LCFC, Metz, France

² ArcelorMittal Research SA, Voie romaine, Maizières-lès-Metz, France

22/05/2025

Merci pour votre attention

