



JOURNÉES MAS 2022 :
SESSION PARALLÈLE MATHS - ENTREPRISES

CALIBRATION DE MODÈLES DE SYSTÈMES COMPLEXES POUR LA
CONSTRUCTION DU JUMEAU NUMÉRIQUE DU VÉHICULE AUTONOME

SOMMAIRE

01 CONTEXTE ET PROBLÉMATIQUE

02 DÉMARCHE GÉNÉRALE
INFÉRENCE BAYÉSIENNE ET MODÈLE DE SUBSTITUTION

03 PREMIERS RÉSULTATS OBTENUS
POUR LE MODÈLE DE SUBSTITUTION





01
CONTEXTE ET PROBLÉMATIQUE

01 – CONTEXTE ET PROBLÉMATIQUE

- **Contexte et problématique : validation et certification du véhicule autonome**

- Réalisation de tests réels sur piste qui sont coûteux en argent et en temps
- Équipe responsable des simulateurs : intégrer des essais simulés dans le processus de certification et d'homologation du véhicule

Les simulations numériques sont-elles suffisamment corrélées aux tests réels pour pouvoir être utilisées légalement ?

- **Objectif : calibration du simulateur**

- Développer une méthodologie qui permet de jauger la qualité des simulations puis de les réajuster
- Prouver qu'il est possible de compléter voir de remplacer les tests sur pistes par ceux sur simulateur

01 – CONTEXTE ET PROBLÉMATIQUE

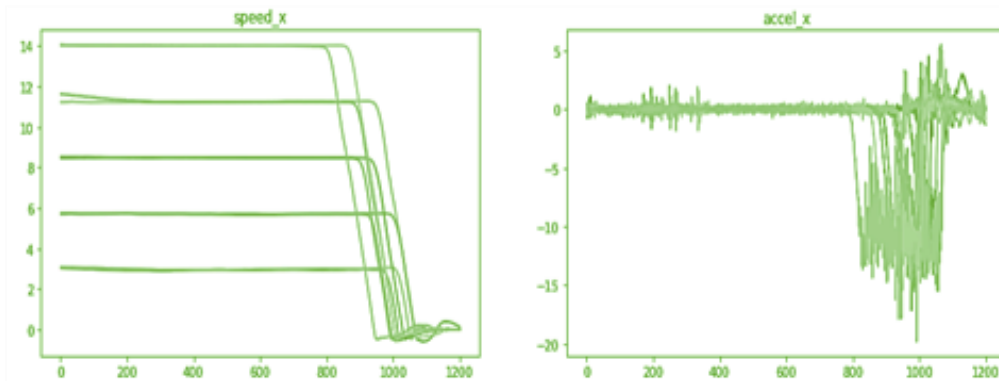
■ Données et outils à disposition :

- **Données réelles** : peu nombreuses, réalisées sur piste par l'équipe TVC au centre technique d'Aubevoye
- **Données simulées** : générées avec la **plateforme ADCAE** qui permet d'accéder au logiciel de simulation SCANeR

■ Les données réelles

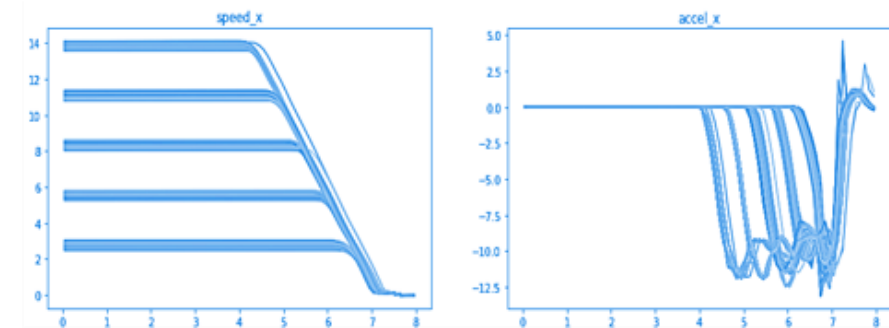
- Entrées : valeurs des paramètres incertains
- Sorties : séries temporelles plus bruitées

→ *uniquement quelques expériences :
environ 3 essais pour chaque configuration*

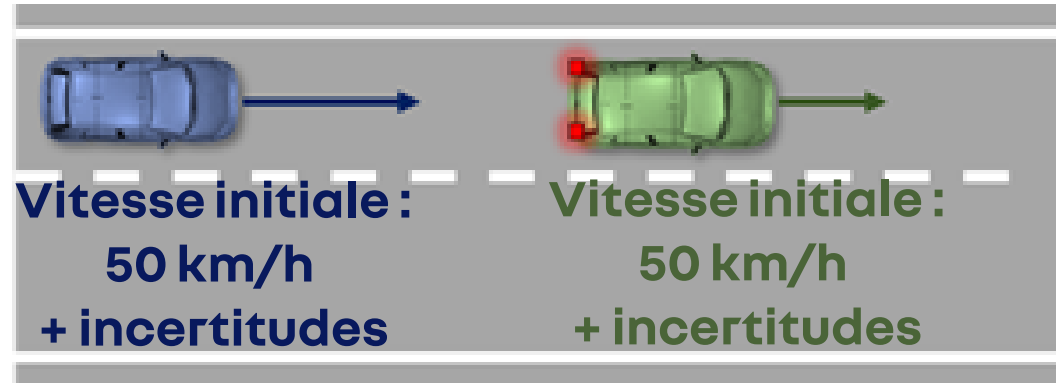


■ Les données simulées

- Entrées : paramètres de scénario, MADA et ENV (efficacité de freinage, vitesse initiale, ...)
- Sorties : séries temporelles décrivant le comportement des véhicules (vitesse, accélération, ...)



01 – CONTEXTE ET PROBLÉMATIQUE



- **Pour chaque test réel : valeurs nominales**

- Renseignées par le TVC (Aubevoye)
- Entachées d'incertitudes : erreurs de capteur + une certaine tolérance

→ *Si le TVC dit que la vitesse initiale est de 50 km/h, il est possible qu'en réalité la voiture ait démarré à 50.5 km/h*

- **Calibrer le simulateur :**

- Tester des valeurs autour des valeurs nominales
- Déterminer quelles valeurs permettent d'obtenir la simulation la plus proche des tests réels

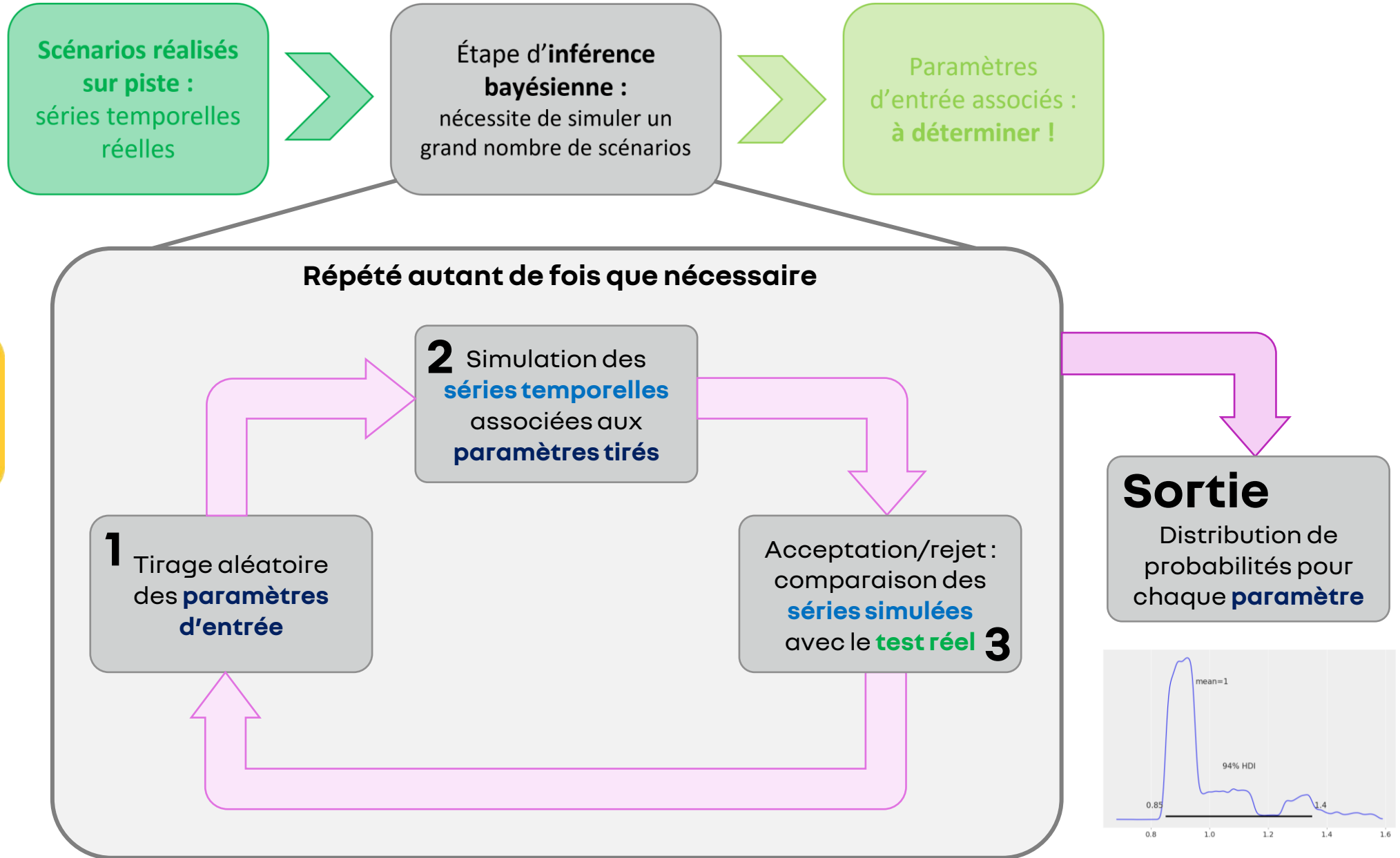
→ *Utilisation d'inférence bayésienne* ←

02

DÉMARCHE GÉNÉRALE



02 - DÉMARCHE GÉNÉRALE

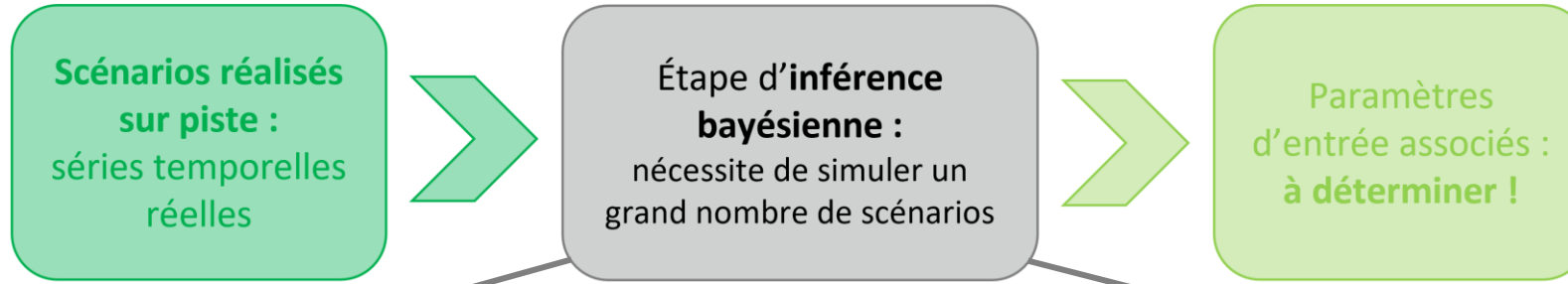


2 : avec **ADCAE**

Plateforme **ADCAE**



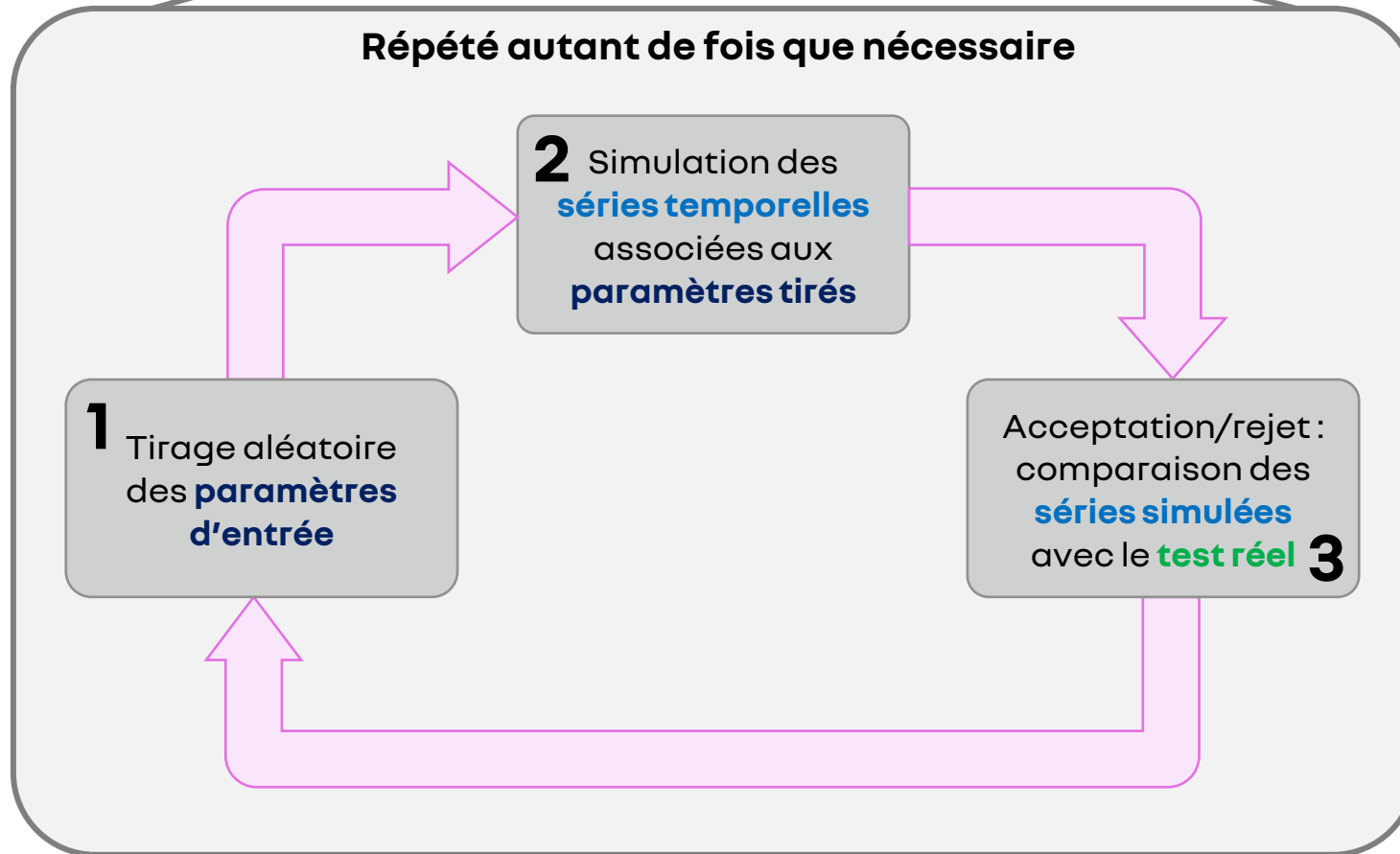
02 - DÉMARCHE GÉNÉRALE



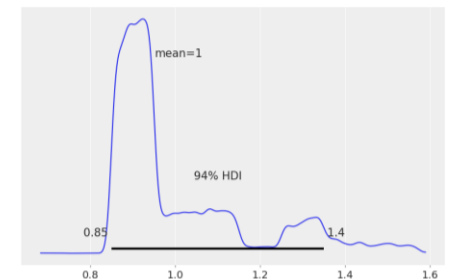
2 : avec **ADCAE**



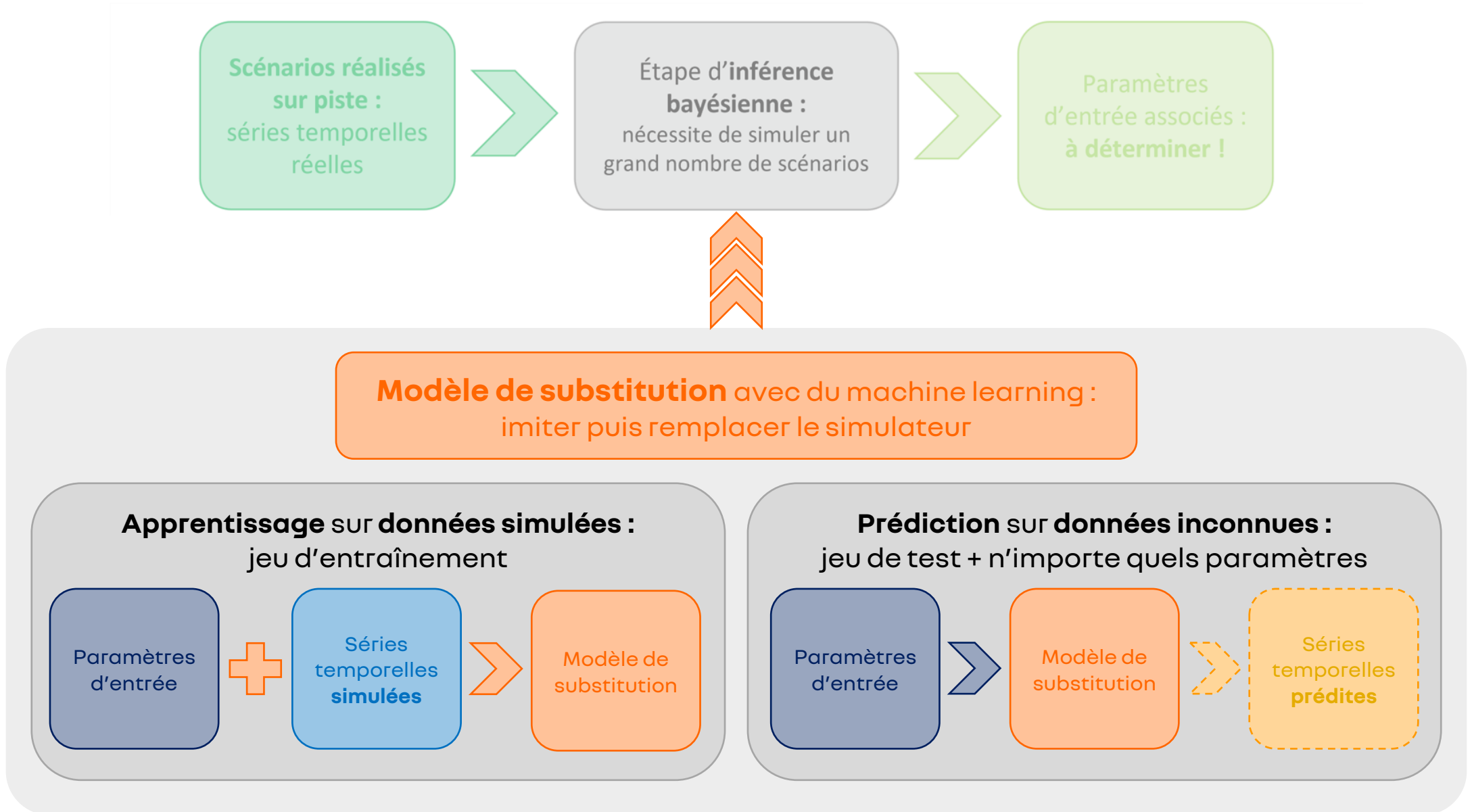
Intelligence artificielle



Sortie
Distribution de probabilités pour chaque **paramètre**

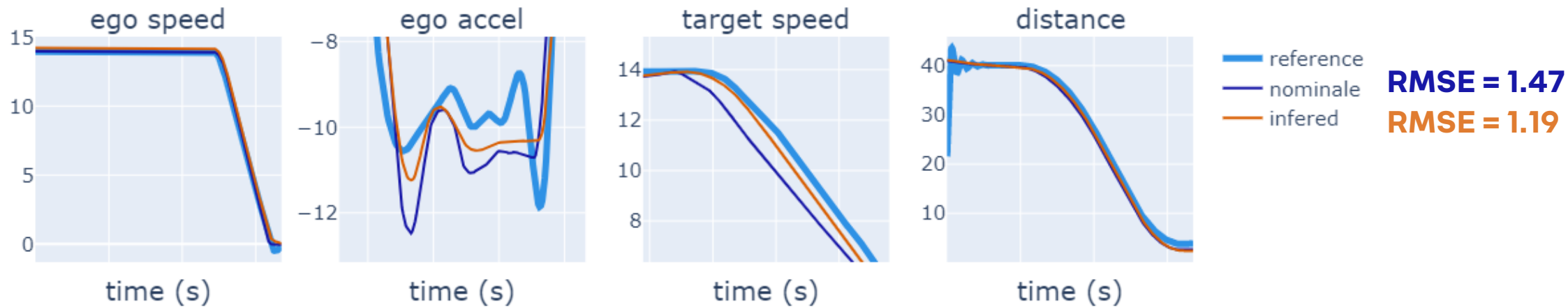


02 - DÉMARCHE GÉNÉRALE

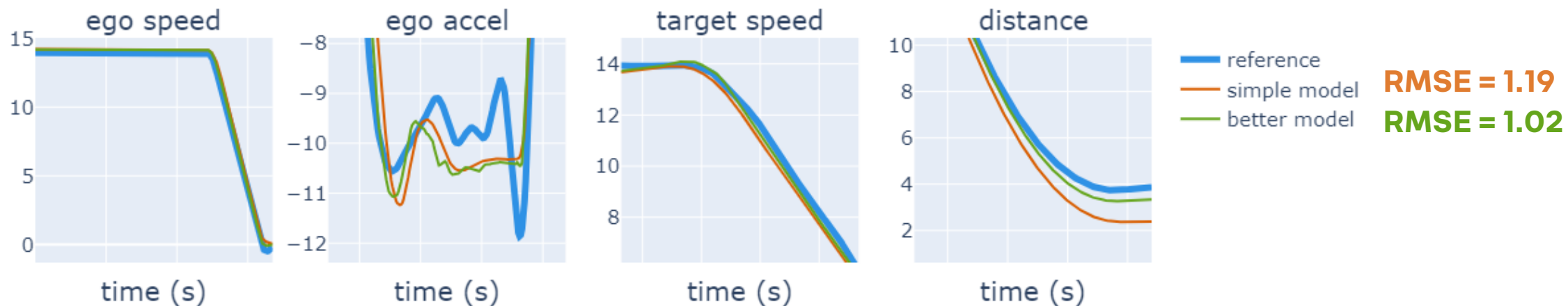


02 - DÉMARCHE GÉNÉRALE

■ De meilleurs résultats en recalibrant le simulateur

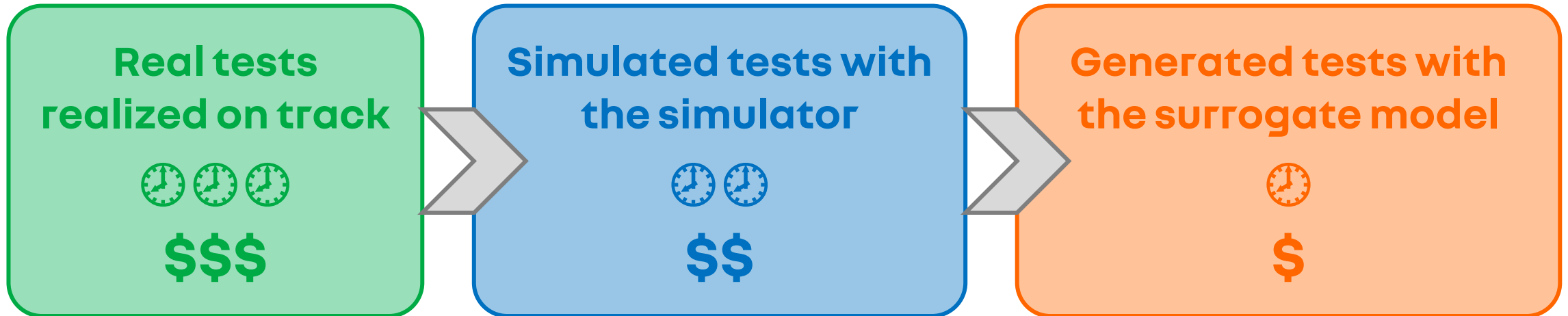


■ Plus le modèle est précis, plus les résultats sont bons



02 - DÉMARCHE GÉNÉRALE

■ Pour résumer :



03

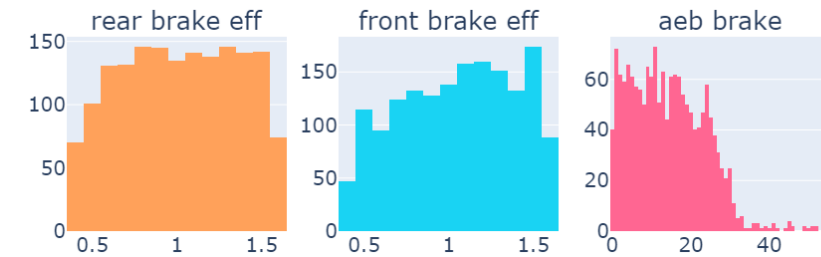
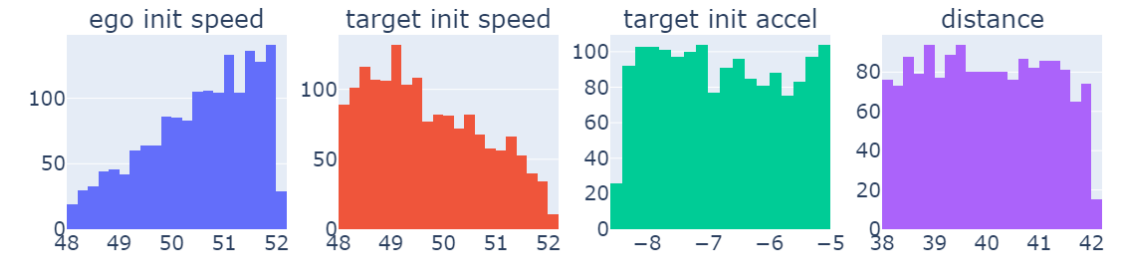
PREMIERS RÉSULTATS OBTENUS POUR LE MODÈLE DE SUBSTITUTION
SUR DONNÉES DE RENAULT

03 – PREMIERS RÉSULTATS OBTENUS

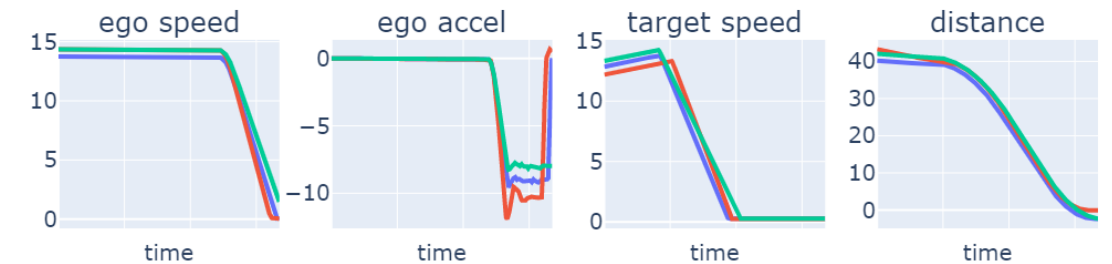
- **Application au scénario :**
 - Deux véhicules en mouvement : test de l'AEB du second véhicule (ego) suite à un freinage brusque du premier (target)
- **Format des données :**
 - Entrée : vecteurs de taille 7
 - Sortie : vecteurs de taille 2684 (*4 séries temporelles*)
- **Taille totale jeux de données :**
 - 1642 scénarios
- **Métrique pour déterminer la précision du modèle :**
 - L'erreur quadratique moyenne :

$$\text{RMSE}(u, v) = \sqrt{\frac{\sum_{i=1}^n (u_i - v_i)^2}{n}}$$

- **Paramètres d'entrée :**



- **Exemples de scénarios pour trois jeux de paramètres :**



03 – PREMIERS RÉSULTATS OBTENUS

■ Sélection de la méthode la plus prometteuse

		k -NN	KRR	CNN	DF	RF	PCA-RF
RMSE ($\times 10^{-2}$)	train	9.09	0.04	7.09	4.22	1.24	1.95
	test	29.2	5.89	7.49	7.17	7.36	12.78
training time		0.07 sec	0.32 sec	25 min	13 min	43 sec	9.47 sec
prediction time		0.01 sec	0.02 sec	0.12 sec	0.90 sec	0.12 sec	0.90 sec



TEST : RMSE ($\times 10^{-2}$)					
	ego speed	ego accel	target speed	distance	mean
KRR	2.13	7.45	1.35	12.65	5.90
RF	0.12	1.39	2.35	10.05	3.48

k -NN : k nearest neighbors ; KRR : Kernel Ridge Regression ; CNN : Convolutional Neural Networks ;
 DF : Deep Forest ; RF : Random Forests ; PCA-RF : Random Forests with PCA

03 – PREMIERS RÉSULTATS OBTENUS

■ Comparaison des temps de calculs

1 modèle global

- Entraînement :
43 sec
- Prédire 165 scénarios :
< 0.2 seconde

RMSE test = 7.36

4 modèles spécifiques

- Entraînement :
52 sec
- Prédire 165 scénarios :
< 0.3 seconde

RMSE test = 3.48

■ RMSE moyenne pour chaque pas de temps sur données tests



CONCLUSION

CONCLUSION – ET POUR LA SUITE ?

■ Pour conclure :

- La méthode des forêts aléatoires permet d'obtenir nos prédictions les plus précises
- Le modèle de substitution est plus précis et permet ainsi d'avoir des résultats meilleurs et plus fiables
- La méthodologie utilisée permet de recalibrer les paramètres d'entrée du simulateur et ainsi avoir une simulation plus réaliste

■ Pour la suite :

- Complexifier le modèle de substitution en utilisant différentes méthodes pour chaque série temporelle
- Travailler sur la partie inférence bayésienne en :
- Développement de mon propre package pour réduire les temps de calculs
 - Spécifier la démarche pour qu'elle s'adapte parfaitement à notre cadre
 - Jauger l'influence de la qualité du modèle sur cette étape

MERCI

The background features a series of parallel diagonal lines in various shades of gray and black, creating a sense of depth and movement. On the right side, there are several bright cyan-colored geometric shapes, including triangles and parallelograms, which stand out against the darker background.

ANNEXE 1 – DESCRIPTION DE L'INFÉRENCE

■ Formalisme inférence bayésienne :

- Déterminer les paramètres X à partir des données réelles $Z = z_0$ connues
→ Cela revient à déterminer la densité de probabilité :

$$\mathbb{P}(X|Z = z) \propto \mathbb{P}(Z = z|X)\pi(X)$$

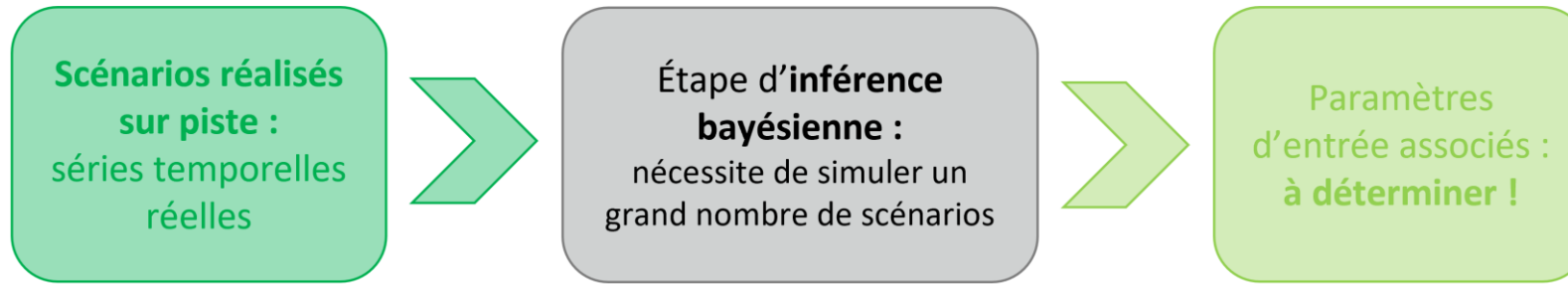
- $\pi(X)$: prior sur les paramètres, loi uniforme
- $\mathbb{P}(Z = z|X)$: vraisemblance du modèle, déterminée à l'aide du modèle de substitution

➤ Méthode ABC : transforme itérativement le prior en posterior en propageant les paramètres échantillonnés à travers une série de distributions

■ Algorithme : pour des données z_0 connues

1. Échantillonner un paramètre x^* selon le prior $\pi(X)$
2. Simuler une base de données z^* à l'aide d'une fonction qui associe à x des données de la même dimension que les données z_0
→ *étape réalisée à l'aide du modèle de substitution pour remplacer SCANeR*
3. Comparer les données créées z^* aux données connues z_0 en utilisant une distance d et un seuil de tolérance ε

ANNEXE 2 - DÉMARCHE GÉNÉRALE



- **Pour réaliser l'inférence bayésienne, on répète ces étapes autant de fois que souhaité/nécessaire :**

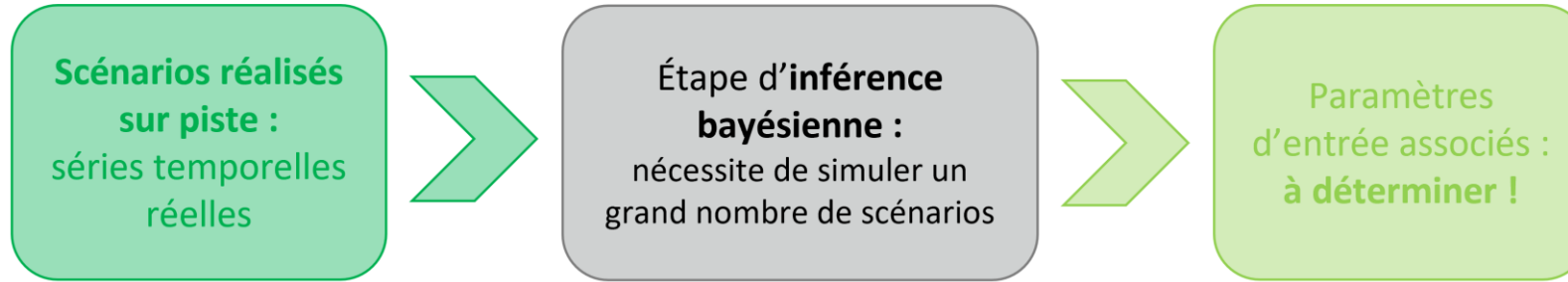
1. Sélection aléatoire de valeurs pour les paramètres d'entrée
→ Prior : lois uniformes sur leur intervalle de définition

- **Choix des intervalles :**

- Valeurs nominales : fournies par le TVC pour chaque scénario, elles sont entachées d'incertitudes dues à des erreurs de capteur et une certaine tolérance
- Idée : tester des valeurs autour de ces valeurs nominales et déterminer lesquelles permettent d'obtenir la simulation la plus proche du scénario réel

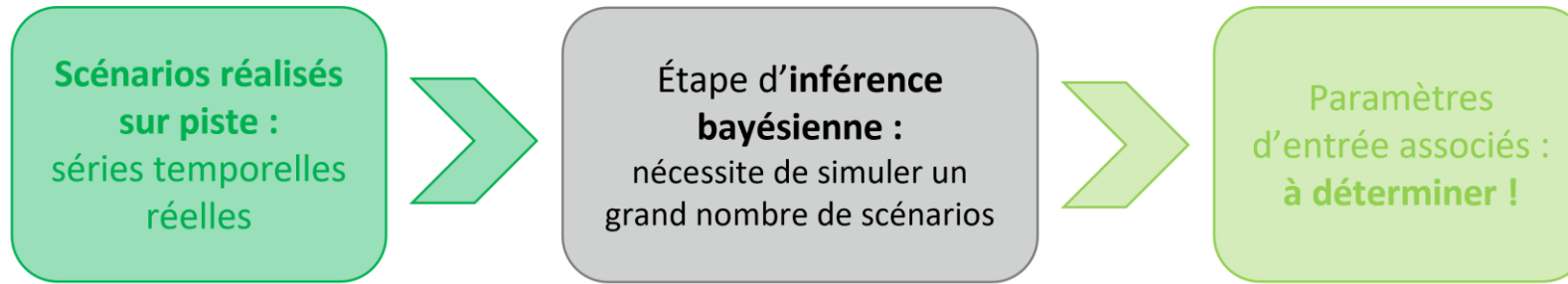


ANNEXE 2 - DÉMARCHE GÉNÉRALE



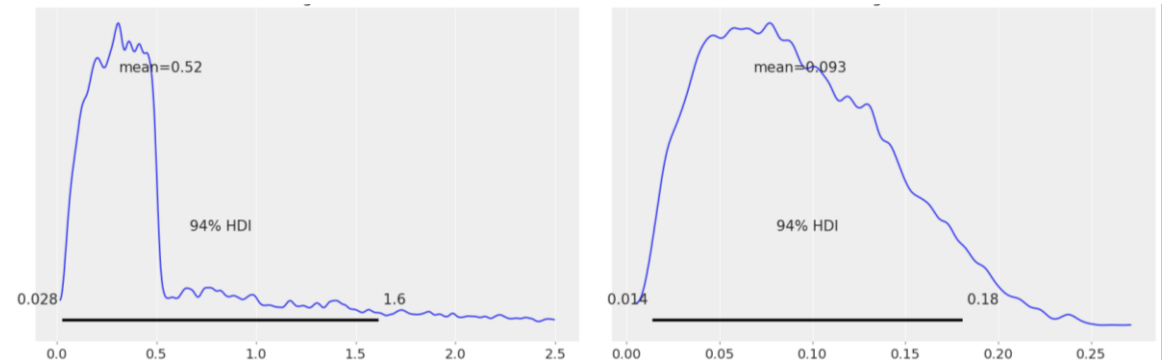
- **Pour réaliser l'inférence bayésienne, on répète ces étapes autant de fois que souhaité/nécessaire :**
 1. Sélection aléatoire de valeurs pour les paramètres d'entrée
→ Prior : lois uniformes sur leur intervalle de définition
 2. Pour des paramètres choisis à l'étape précédente : prédiction des séries temporelles associées à l'aide d'**ADCAE**
 3. Comparaison des séries prédites avec le(s) **test(s) réel(s)** :
 - Si les prédictions sont suffisamment proches des tests réels, les paramètres proposés sont conservés.
 - Sinon, ils sont rejetés.

ANNEXE 2 - DÉMARCHE GÉNÉRALE

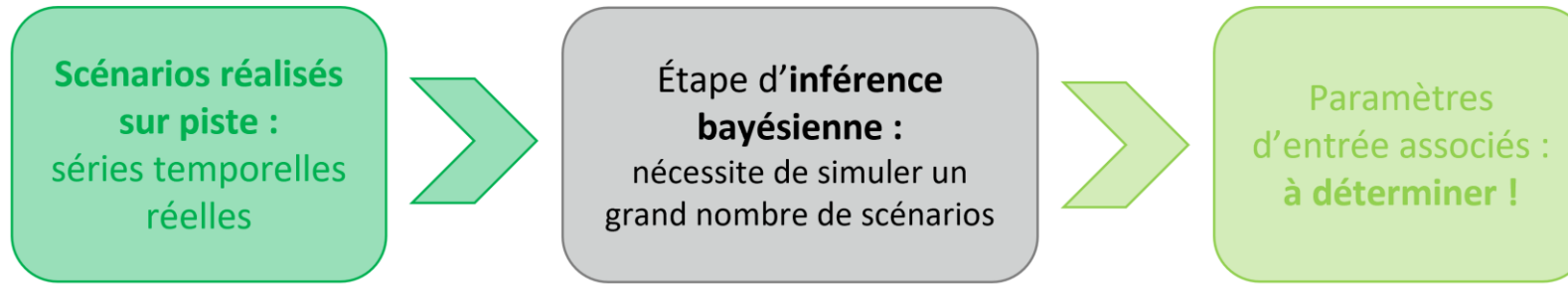


■ **Pour réaliser l'inférence bayésienne, on répète ces étapes autant de fois que souhaité/nécessaire :**

1. Sélection aléatoire de valeurs pour les paramètres d'entrée
→ Prior : lois uniformes sur leur intervalle de définition
2. Pour des paramètres choisis à l'étape précédente : prédiction des séries temporelles associées à l'aide d'**ADCAE**
3. Comparaison des séries prédites avec le(s) **test(s) réel(s)** :
 - Si les prédictions sont suffisamment proches des tests réels, les paramètres proposés sont conservés.
 - Sinon, ils sont rejetés.
4. Finalement, en sortie : on obtient des distributions de probabilités pour chaque paramètre



ANNEXE 2 - DÉMARCHE GÉNÉRALE



■ **Pour réaliser l'inférence bayésienne, on répète ces étapes autant de fois que souhaité/nécessaire :**

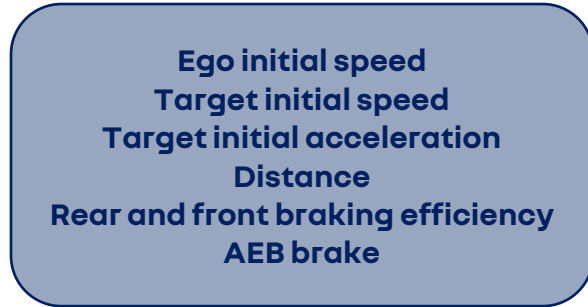
1. Sélection aléatoire de valeurs pour les paramètres d'entrée
→ Prior : lois uniformes sur leur intervalle de définition
2. Pour des paramètres choisis à l'étape précédente : prédiction des séries temporelles associées à l'aide d'**ADCAE** → remplacer par le **modèle de substitution**



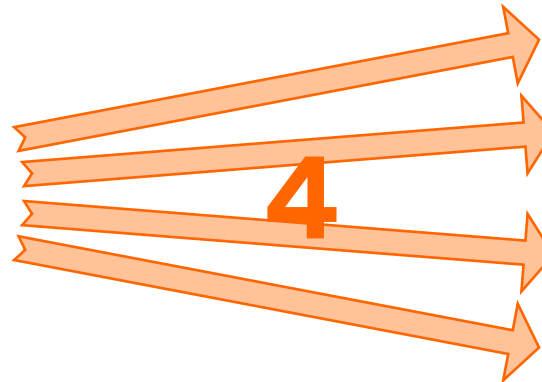
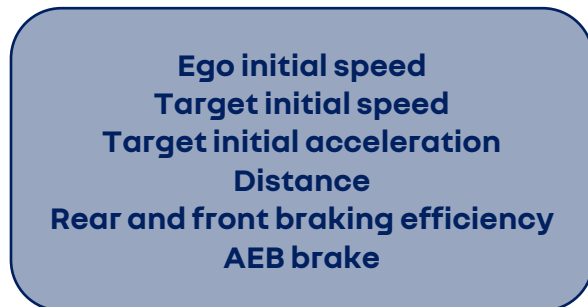
3. Comparaison des séries prédites avec le(s) **test(s) réel(s)** :
 - Si les prédictions sont suffisamment proches des tests réels, les paramètres proposés sont conservés.
 - Sinon, ils sont rejetés.
4. Finalement, en sortie : on obtient des distributions de probabilités pour chaque paramètre

ANNEXE 3 - DÉVELOPPEMENT DU MODÈLE DE SUBSTITUTION

■ Ancienne approche : 1 modèle général

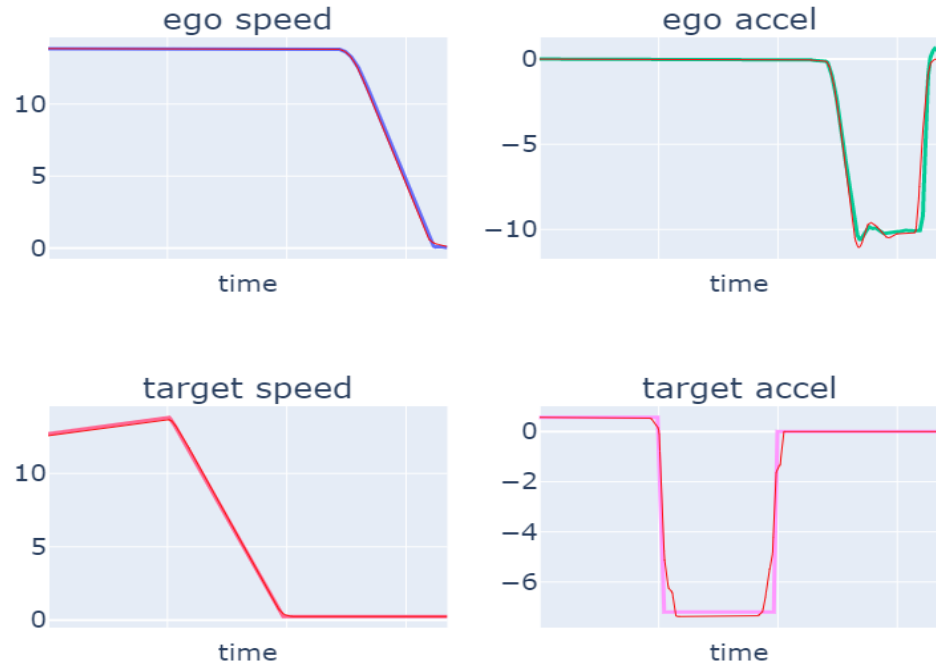


■ Nouvelle approche : 4 modèles spécifiques

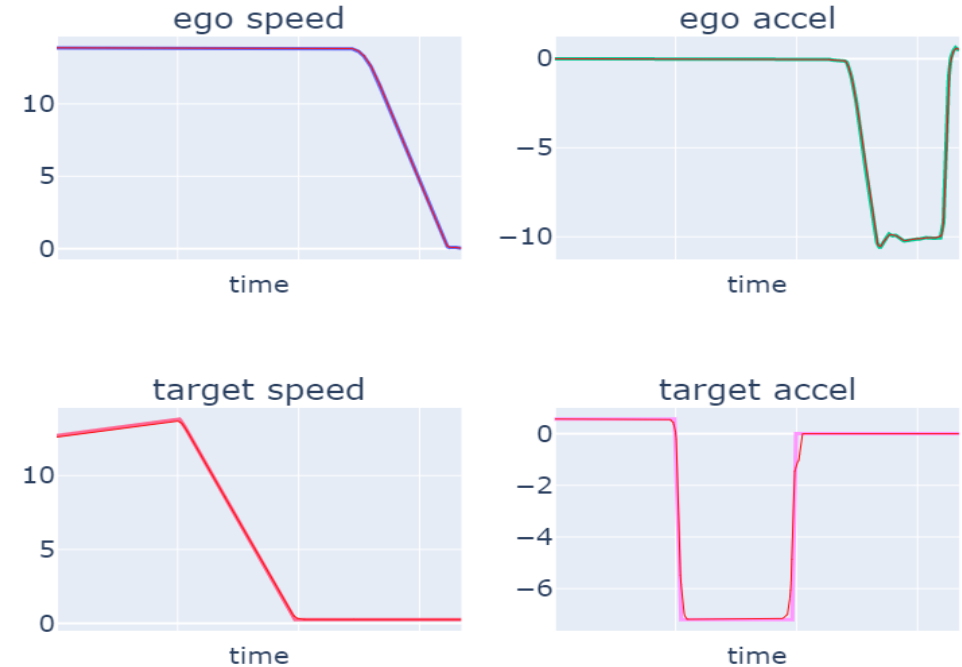


ANNEXE 3 - DÉVELOPPEMENT DU MODÈLE DE SUBSTITUTION

■ 1 modèle global

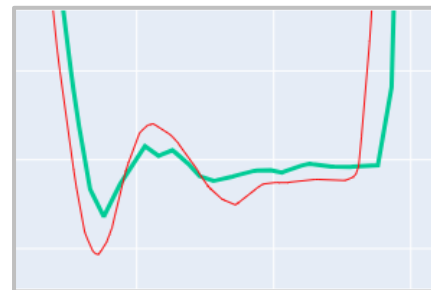


■ 4 modèles spécifiques

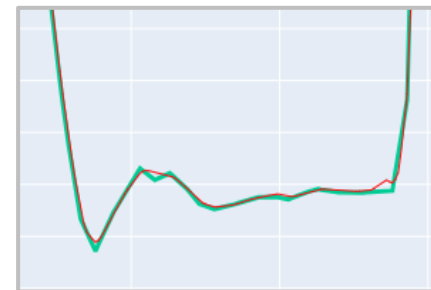


■ Zoom of ego accel

- Temps de calcul :**
- Entraînement: **43 sec**
 - Prédire 165 scénarios: **< 0.2 seconde**



1 modèle



5 modèles

Temps de calcul :

- Entraînement: **52 sec**
- Prédire 165 scénarios: **< 0.3 seconde**

ANNEXE 3 – DÉVELOPPEMENT DU MODÈLE DE SUBSTITUTION

■ RMSE de chaque série temporelle

TRAIN : RMSE ($\times 10^{-2}$)					
	ego speed	ego accel	target speed	distance	mean
1 model	0.21	1.53	0.83	2.39	1.24
5 models	0.03	0.22	0.43	2.10	0.70
TEST : RMSE ($\times 10^{-2}$)					
	ego speed	ego accel	target speed	distance	mean
1 model	1.35	10.27	5.1	12.71	7.36
5 models	0.12	1.39	2.35	10.05	3.48

■ Scaled-RMSE de chaque série temporelle

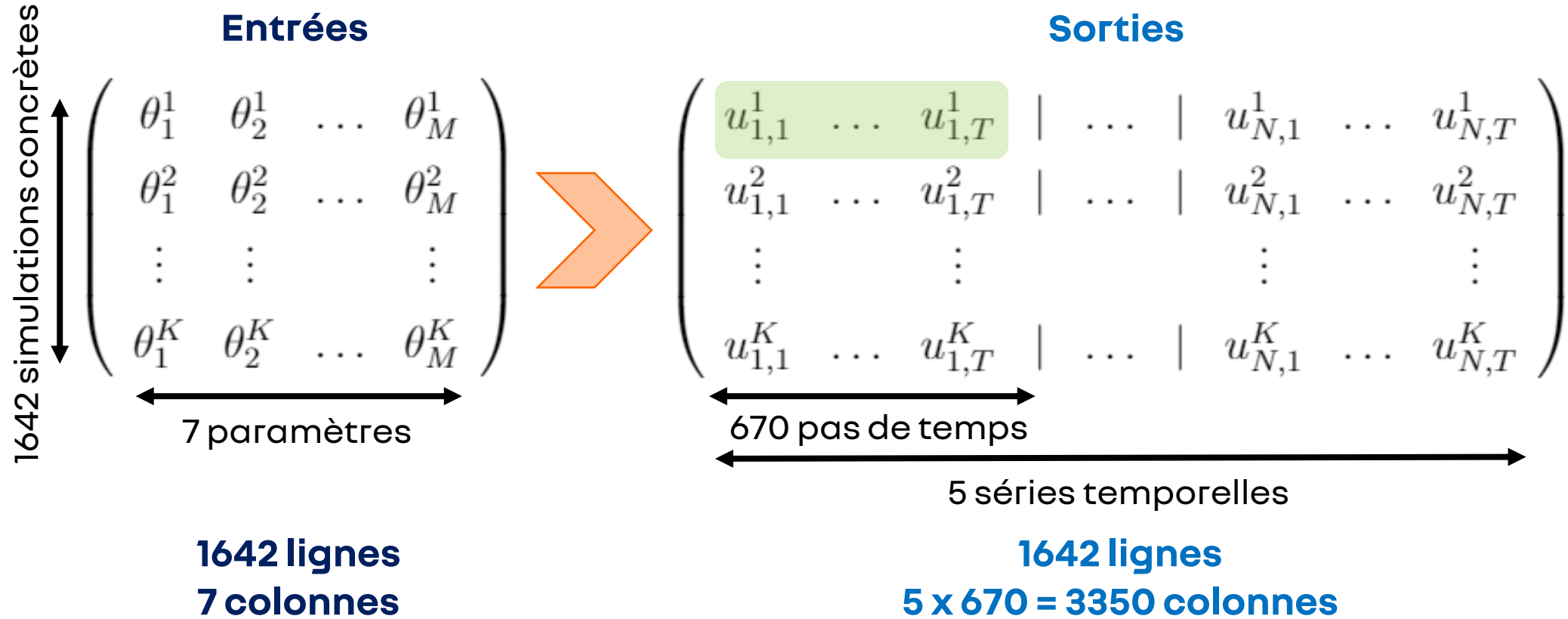
TRAIN : s-RMSE ($\times 10^{-4}$)					
	ego speed	ego accel	target speed	distance	mean
1 model	0.11	12.76	0.64	0.35	3.46
5 models	0.02	12.87	0.30	0.46	3.41
TEST : s-RMSE ($\times 10^{-4}$)					
	ego speed	ego accel	target speed	distance	mean
1 model	0.83	9.38	4.07	2.66	4.24
5 models	0.07	1.26	2.16	2.58	1.52

■ RMSE moyenne pour chaque pas de temps



ANNEXE 4 - DIFFÉRENTS FORMATS DE DONNÉES

Format 1:



ANNEXE 4 – DIFFÉRENTS FORMATS DE DONNÉES

■ **Format 2 :**

Entrées

$$\begin{pmatrix} \theta_1^1 & \theta_2^1 & \dots & \theta_M^1 & 1 \\ " & " & \dots & " & 2 \\ \vdots & \vdots & & \vdots & \vdots \\ " & " & \dots & " & T \end{pmatrix}$$

$$\dots$$

$$\begin{pmatrix} \theta_1^K & \theta_2^K & \dots & \theta_M^K & 1 \\ " & " & \dots & " & 2 \\ \vdots & \vdots & & \vdots & \vdots \\ " & " & \dots & " & T \end{pmatrix}$$

8 paramètres

670 x 1642 = 1.1M lignes
8 colonnes



Sorties

$$\begin{pmatrix} u_{1,1}^1 & u_{2,1}^1 & \dots & u_{N,1}^1 \\ u_{1,2}^1 & u_{2,2}^1 & \dots & u_{N,2}^1 \\ \vdots & \vdots & & \vdots \\ u_{1,T}^1 & u_{2,T}^1 & \dots & u_{N,T}^1 \end{pmatrix}$$

$$\dots$$

$$\begin{pmatrix} u_{1,1}^K & u_{2,1}^K & \dots & u_{N,1}^K \\ u_{1,2}^K & u_{2,2}^K & \dots & u_{N,2}^K \\ \vdots & \vdots & & \vdots \\ u_{1,T}^K & u_{2,T}^K & \dots & u_{N,T}^K \end{pmatrix}$$

5 séries temporelles

670 x 1642 lignes
5 colonnes

670 pas de temps

1642 simulations concrètes