

Modélisation Substitutive de la Croissance des Cultures par Apprentissage Profond Guidé par la Physique

Julian Agudelo^{1, 2}, Vincent Guigue¹, Cristina Manfredotti¹, Evelyne Lutton¹, et Hadrien Piot²

¹UMR MIA Paris Saclay, AgroParisTech and INRAE

²Agrial, Agricultural Service

Résumé

Une modélisation précise de la croissance des cultures est essentielle au progrès de la recherche agronomique et à l'amélioration des pratiques agricoles. Cependant, les simulateurs fondés sur les processus sont souvent limités par des exigences importantes de paramétrisation, tandis que les modèles d'apprentissage automatique se restreignent généralement à la prévision d'états de développement végétatif ponctuels sans fournir de simulations complètes. Ces limites motivent le développement de modèles substituts efficaces, physiquement cohérents et robustes dans des contextes avec données limitées ou hétérogènes.

Dans cet article, nous proposons un modèle substitut séquence-à-séquence (seq2seq) guidé par la physique pour le modèle de croissance des cultures WOFOST (World Food Studies). Notre modèle est conçu pour reproduire la dynamique du développement des cultures tout en limitant les besoins en paramétrisation et en données d'entrée. La cohérence physique est assurée grâce à l'utilisation de contraintes architecturales et de pénalités physiques dans la fonction de perte, intégrant des biais observationnels, inductifs et d'apprentissage appropriés. Nous présentons des résultats empiriques montrant que l'ajout de contraintes architecturales et des pénalités fondées sur la physique améliore la précision prédictive, oriente l'apprentissage vers des hypothèses plus généralisables et renforce la robustesse dans des conditions hors distribution (OOD). En réduisant les besoins en entrées, l'approche proposée offre un cadre de modélisation de la croissance des cultures plus accessible pour l'étude de questions agronomiques dans des conditions réelles.

Mots-clés: Réseaux de neurones guidés par la physique, modélisation substitutive, modélisation de la croissance des cultures, séries temporelles.

1 Introduction

La modélisation de la croissance des cultures est essentielle pour aider à la conception des systèmes agricoles et pour orienter les pratiques et politiques rurales (Zhang et al., 2023). Les approches de modélisation se divisent généralement en deux catégories : les modèles fondés sur les processus (Process-Based Models, PBM), qui simulent la croissance des cultures à partir des mécanismes biophysiques sous-jacents (Fischer, 2015), et les modèles fondés sur les données, qui apprennent directement des relations complexes à partir des données (Zhang et al., 2023).

Avec l'adoption croissante de l'apprentissage automatique, les modèles fondés sur les données ont montré de bonnes performances prédictives pour des tâches liées à la modélisation des cultures. Cependant, ils se limitent souvent à la prévision d'états ponctuels ou d'indicateurs agrégés, tels que l'indice de surface foliaire (Zhang et al., 2021; Hussain et al., 2025), la concentration en nutriments (Trontelj ml. and Chambers, 2021; Jain et al., 2024) et, le plus souvent, le rendement des cultures (Van Klompenburg et al., 2020; Shahhosseini et al., 2021; Sharma et al., 2023; Yenikar et al., 2025). Par conséquent, ils offrent une représentation limitée de la dynamique de croissance des plantes et sont moins adaptés à la modélisation du phénomène dans son ensemble (Maestrini et al., 2022). De plus, ils peuvent présenter

une faible capacité de généralisation et manquent souvent d’interprétabilité agronomique. Les PBM fournissent une modélisation de la croissance des cultures interprétable, en offrant des trajectoires temporelles des états des cultures. Néanmoins, leur déploiement pratique est souvent contraint par des exigences élevées en matière de paramétrage et par la nécessité de disposer de données d’entrée détaillées, ce qui peut restreindre leur usage opérationnel (Fischer, 2015; Shahhosseini et al., 2021).

Pour pallier les limites de ces deux approches, la recherche récente s’est de plus en plus orientée vers les modèles hybrides (He et al., 2023; Cao et al., 2026). Ces modèles combinent la connaissance mécanistique des PBM pour garantir des prédictions physiologiquement cohérentes avec des composants fondés sur les données, capables de capturer des relations complexes qui sont propres aux systèmes agricoles, mais non incluses dans la formulation mécanistique (Maestrini et al., 2022).

Dans cette perspective de modélisation hybride, notre approche s’appuie sur le cadre général de l’«apprentissage informé par la physique» (Physics-informed learning), formalisé par Karniadakis et al. (2021). Nous explorons en particulier trois mécanismes complémentaires pour intégrer la connaissance physique dans le modèle et introduire les biais appropriés :

1. **Biais observationnel** : en entraînant le modèle sur des jeux de données physiquement plausibles.
2. **Biais inductif** : en incluant, par construction, des modules imposant la monotonie dans la génération des séries temporelles concernées.
3. **Biais d’apprentissage** : en intégrant des pénalités physiques dans la fonction de perte.

WOFOST est un modèle mécanistique de croissance des cultures qui simule la croissance en s’appuyant sur des représentations explicites des processus physiologiques et de leurs interactions avec les conditions environnementales et les pratiques de gestion (De Wit et al., 2019). Dans la continuité de ces travaux, nous développons un modèle substitut seq2seq guidé par la physique pour approximer WOFOST.

Contributions

- Nous proposons un modèle substitut d’ordre réduit de WOFOST, utilisant un nombre restreint de variables d’entrée afin d’en améliorer l’utilisation opérationnelle.
- Nous intégrons des connaissances a priori dans le substitut grâce à des biais observationnels, inductifs et d’apprentissage appropriés, renforçant la cohérence physique.
- Nous fournissons des preuves empiriques en appui des affirmations suivantes : (1) la prédiction de séquences complètes de développement offre des performances supérieures à celles de la prédiction ponctuelle du rendement ; (2) les contraintes architecturales améliorent la cohérence physique et la précision ; (3) les pénalités physiques dans la fonction de perte renforcent la cohérence physique des sorties ; (4) les contraintes physiques orientent le processus d’apprentissage vers des hypothèses plus généralisables, améliorant la robustesse en OOD.

2 État de l’art

L’apprentissage guidé par la physique constitue une famille d’approches en apprentissage automatique dans lesquelles la connaissance du domaine, exprimée par des lois physiques, des PBMs ou des contraintes mécanistiques, est explicitement intégrée dans le processus d’apprentissage. Du point de vue de l’apprentissage profond, ce paradigme remonte aux années 1990, lorsque les réseaux neuronaux ont été proposés comme solveurs alternatifs pour les équations différentielles (Lee and Kang, 1990; Lagaris et al., 1998). Cependant, limitées par la maturité technique et les capacités de calcul de l’époque, ces approches sont restées largement en sommeil jusqu’à la publication des travaux fondamentaux introduisant les réseaux de neurones informés par la physique (Physics-Informed Neural Networks, PINNs) (Raissi et al., 2019) et les réseaux de neurones guidés par la physique (Physics-Guided Neural Networks, PGNNs) (Jia et al., 2020). Ces travaux, avec d’autres recherches récentes, constituent un

ensemble de techniques centrées sur l’usage de contraintes physiques pour régulariser l’entraînement des réseaux neuronaux.

Plusieurs études ont montré l’efficacité de l’apprentissage profond guidé par la physique pour la construction de modèles de substitution de systèmes physiques, en particulier dans le contexte de solveurs à grande dimension (Xian and Wang, 2024) et de systèmes fondés sur des équations différentielles (Haghighat et al., 2021; Donnelly et al., 2024). Parallèlement, les approches purement fondées sur les données ont été appliquées avec succès pour substituer des modèles mécanistiques à grande échelle dans différents domaines (Garzón et al., 2022; Luo et al., 2023; Gherman et al., 2023). Cependant, l’utilisation de l’apprentissage profond guidé par la physique pour remplacer des simulateurs caractérisés par des composants hétérogènes, des processus discrets et une différentiabilité limitée, reste encore relativement peu explorée.

Dans le contexte de la simulation de la croissance des cultures, l’apprentissage guidé par la physique est de plus en plus utilisé pour améliorer les systèmes de prédiction du rendement, en couplant les PBMs avec des réseaux neuronaux. Ce couplage prend plusieurs formes : les états internes du PBM sont fournis en entrée ou comme cibles auxiliaires pour les réseaux neuronaux (Shahhosseini et al., 2021), les sorties du PBM servent à constituer des données de préentraînement (Fang et al., 2026) ou à régulariser les fonctions de perte (He et al., 2023). Cependant, malgré ces avancées, l’application de l’apprentissage guidé par la physique à la construction de modèles substitués pour les modèles de croissance des cultures fondés sur les processus reste rare. La plupart des substitués existants visent à prédire les états finaux (Corrales et al., 2022; von Bloh et al., 2024), plutôt que l’ensemble du processus de croissance.

À notre connaissance, ce travail constitue la première tentative de développement d’un modèle substitut pour un modèle de croissance des cultures fondé sur des processus à l’aide d’une approche d’apprentissage guidé par la physique.

3 Formulation du problème

Dans WOFOST, la croissance des cultures est simulée à partir d’un ensemble de paramètres d’entrée :

$$\mathbf{X}_{\text{wof}} = (\mathbf{c}, \mathbf{s}, \mathbf{g}, \{\mathbf{p}_t\}_{t=0}^{\mathcal{T}}, \{\mathbf{m}_t\}_{t=0}^{\mathcal{T}}) \quad (1)$$

où \mathbf{c} , \mathbf{s} et \mathbf{g} sont des vecteurs regroupant respectivement les caractéristiques propres à la culture, les propriétés du sol et les conditions initiales du site analysé. Les séries temporelles $\{\mathbf{p}_t\}_{t=0}^{\mathcal{T}}$ et $\{\mathbf{m}_t\}_{t=0}^{\mathcal{T}}$ sont collectées à une résolution quotidienne et apportent des informations complémentaires : $\{\mathbf{p}_t\}_{t=0}^{\mathcal{T}}$ correspond aux pratiques agronomiques, tandis que $\{\mathbf{m}_t\}_{t=0}^{\mathcal{T}}$ regroupe les observations météorologiques. \mathcal{T} désigne le nombre total de jours simulés.

À partir de ces entrées, WOFOST génère des séries temporelles multivariées quotidiennes correspondant à plusieurs variables indiquant l’état de développement de la culture $\{\mathbf{z}_t\}_{t=T}^{\mathcal{T}}$:

$$\text{WOFOST} : \mathbf{X}_{\text{wof}} \in \mathbb{R}^{D_{\text{wof}}} \rightarrow \{\mathbf{z}_t\}_{t=T}^{\mathcal{T}} \in \mathbb{R}^{d \times (\mathcal{T}-T)} \quad (2)$$

où T correspond à la date de semis.

L’objectif de notre étude est de développer un modèle substitut \mathcal{S} capable de prédire un sous-ensemble de ces variables, soit les séries temporelles multivariées $\{\mathbf{y}_t\}_{t=T}^{\mathcal{T}}$, avec $\mathbf{y}_t \subset \mathbf{z}_t$. Les prédictions reposent sur un ensemble de variables d’entrée défini par $\mathbf{X} = \mathbf{X}'_{\text{wof}} \cup \mathbf{X}_{\text{ind}}$, où $\mathbf{X}'_{\text{wof}} \subset \mathbf{X}_{\text{wof}}$ représente le sous-ensemble des entrées WOFOST facilement disponibles, et \mathbf{X}_{ind} regroupe des variables dites indirectes :

$$\mathcal{S} : \mathbf{X} \in \mathbb{R}^{D_x} \rightarrow \{\mathbf{y}_t\}_{t=T}^{\mathcal{T}} \in \mathbb{R}^{b \times (\mathcal{T}-T)} \quad (3)$$

avec $D_{\text{wof}} > D_x$ et $d > b$. Le vecteur \mathbf{X} comprend 9 variables statiques et 7 variables dynamiques, tandis que \mathbf{X}_{wof} contient généralement entre 80 et 90 variables statiques et entre 7 et 9 variables dynamiques.

Bien que les entrées indirectes ne soient pas représentées explicitement dans les équations du PBM, nous supposons qu’elles capturent des aspects pertinents de la dynamique des cultures. Par exemple, l’identifiant d’un exploitant agricole ne décrit pas directement le développement de ses cultures, mais permet au modèle de prendre en compte des motifs liés à des facteurs non observés, tels que les pratiques agricoles ou les conditions spécifiques du sol.

Les variables de développement des cultures retenues dans $\{\mathbf{y}_t\}_{t=T}^{\mathcal{T}}$ sont listées dans le Tableau 1.

Variable	Description	Unité
DVS	Stade de développement de la culture	–
LAI	Indice de surface foliaire	–
TAGP	Biomasse totale hors sol	kg ha ⁻¹
TWSO	Biomasse des organes de réserve	kg ha ⁻¹
TWLV	Biomasse des feuilles	kg ha ⁻¹
TWST	Biomasse des tiges	kg ha ⁻¹
TWRT	Biomasse des racines	kg ha ⁻¹
DMI	Augmentation quotidienne de matière sèche	kg ha ⁻¹ j ⁻¹
ASRC	Nutriments nets disponibles	kg _{CH₂O} ha ⁻¹ j ⁻¹
GASS	Taux d’assimilation (stress hydrique)	kg _{CH₂O} ha ⁻¹ j ⁻¹
MRES	Taux de respiration de maintenance	kg _{CH₂O} ha ⁻¹ j ⁻¹

TABLE 1 – Variables de développement des cultures incluses dans \mathbf{y}_t .

4 Jeu de données

Nous utilisons des données réelles fournies par la coopérative agricole Agrial, située dans le nord-ouest de la France. Ces données couvrent 200 086 parcelles agricoles enregistrées entre 2019 et 2024, incluant des informations sur les espèces cultivées, les dates de semis et les identifiants des exploitants. Les données incluent huit espèces : orge, blé, maïs, colza, fève, tournesol, millet et sorgho. Les enregistrements de parcelles sont croisés avec les données météorologiques de la coopérative et interpolés au niveau des parcelles à l’aide de la méthode de pondération par distance inverse. En raison de la rareté des informations sur les pratiques agricoles et de la flexibilité limitée des entrées WOFOST, nous avons soit supprimé la majorité des pratiques agricoles, soit transformées en variables statiques.

Ces données sont utilisées pour exécuter le modèle WOFOST et générer des séries temporelles de développement des cultures. Lorsque certaines valeurs de paramètres sont indisponibles, les paramètres par défaut fournis par WOFOST (De Wit and Boogaard, 2021) sont adoptés : (1) les paramètres régionaux de culture pour les traits phénologiques manquants, (2) les paramètres du sol cohérents avec la classification de la texture du sol (issues de nos données), et (3) les valeurs initiales standard pour la concentration en CO₂ et la WAV (quantité initiale d’eau dans la zone racinaire), conformément aux pratiques établies (De Wit et al., 2019). Ainsi, le jeu de données se compose d’un sous-ensemble des entrées WOFOST et de leurs sorties correspondantes.

Enfin, ce jeu de données est partitionné de deux manières selon le scénario d’évaluation hors distribution (OOD) :

- La première partition, notée $\mathcal{D}_{\text{OOD-G}}$, est conçue pour évaluer l’OOD uniquement selon l’axe géographique. Les observations situées dans la zone grisée de la Figure 1 constituent un sous-ensemble OOD de 21 491 échantillons, tandis que les 178 595 observations restantes sont considérées comme des données *in-distribution* et réparties en ensembles d’entraînement, de validation et de test selon un ratio 80-10-10.

- La deuxième partition, notée $\mathcal{D}_{\text{OOD-GP}}$, cible l'évaluation OOD sur les axes géographique et phénologique. Le sous-ensemble OOD contient uniquement 2864 échantillons d'orge provenant de la zone grisée de la Figure 1. Les 161 124 observations restantes (hors orge) sont réparties en ensembles d'entraînement, de validation et de test selon un ratio 80-10-10. Toutes les observations correspondant à l'orge ou à la zone géographique spécifiée sont exclues des ensembles d'entraînement, de validation et de test.

Les principaux résultats présentés dans cet article sont obtenus avec la partition $\mathcal{D}_{\text{OOD-GP}}$, car son évaluation permet d'estimer la capacité de généralisation du modèle dans des conditions proches d'un déploiement réel. La partition $\mathcal{D}_{\text{OOD-G}}$ est utilisée uniquement lorsque cela permet une analyse ciblée.

Les données sont anonymisées et accessibles publiquement sur Hugging Face¹.

Prétraitement des données Étant donné que le jeu de données provient de différents logiciels de gestion agricole ainsi que des déclarations des exploitants pour la Politique Agricole Commune (PAC) européenne, un prétraitement est nécessaire afin d'éviter les données aberrantes. À cette fin, nous avons filtré les données pour garantir la plausibilité agronomique des parcelles. Les variables catégorielles statiques ont été codées sous forme d'indices entiers, suivant les pratiques standard pour l'apprentissage de représentations par embeddings. Les dates de semis ont été représentées en utilisant un codage cyclique sinus-cosinus afin de préserver leur structure périodique. Enfin, toutes les variables numériques statiques et les séries temporelles d'entrée ont été standardisées par normalisation centrée-réduite.

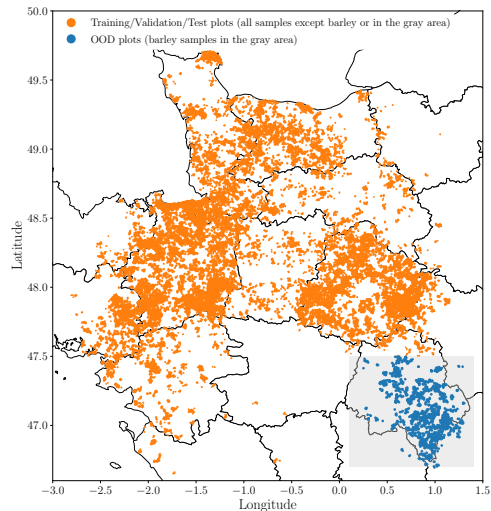


FIGURE 1 – Localisation des échantillons hors distribution (OOD) (en bleu).

5 Un Modèle Séquence-à-Séquence Guidé par la Physique

Le modèle proposé utilise une architecture séquence-à-séquence (seq2seq) (Sutskever et al., 2014) dans le cadre du paradigme des réseaux de neurones guidés par la physique (PGNN) (Jia et al., 2020). Les architectures seq2seq sont particulièrement adaptées à la modélisation substitutive de systèmes dynamiques, car elles transforment des séries temporelles d'entrée en trajectoires de sortie en conservant une représentation latente de l'historique (Sutskever et al., 2014). Cela permet de capturer les dépendances à long terme, essentielles pour des processus tels que la croissance des cultures et le développement phénologique. De plus, les unités récurrentes se sont montrées compatibles avec l'intégration de contraintes physiques (Zheng et al., 2023; Tao et al., 2025). Enfin, la séparation explicite entre encodeur et décodeur facilite l'application de contraintes physiques lors de la génération des trajectoires. La Figure 2 illustre l'architecture proposée.

L'encodeur traite l'entrée \mathbf{X} en séparant les variables statiques spécifiques au site \mathbf{x}_c des mesures météorologiques dynamiques jusqu'à la date de semis, $\{\mathbf{m}_t\}_{t=0}^T$. Les variables statiques sont encodées via un Perceptron Multi-Couche (MLP), tandis que la série temporelle est traitée par une Gated Recurrent Unit (GRU) (Cho et al., 2014).

Le décodeur, également basé sur une GRU, transforme ensuite la représentation latente $\mathbf{h}^{(E)}$ dans b séries temporelles $\{\mathbf{y}_t\}_{t=T}^{T+n}$ générées de manière autorégressive. n représente l'horizon de prédiction maximal, tel que $T + n = \mathcal{T}$.

1. https://huggingface.co/datasets/JulAgu/PG_S2S-Crop_development/tree/main

À chaque étape de décodage $t > T$, les mesures météorologiques du jour courant \mathbf{m}_t sont concaténées à l'état caché \mathbf{h}_t , produisant l'état augmenté $\tilde{\mathbf{h}}_t = \mathbf{h}_t \oplus \mathbf{m}_t$. Cet état augmenté est ensuite passé à travers une pile de b MLPs, qui prédisent les valeurs rassemblées dans le vecteur $\hat{\mathbf{y}}_t$.

Parallèlement, un classificateur binaire prend $\tilde{\mathbf{h}}_t$ en entrée et estime la probabilité \hat{p}_t que la culture ait atteint le stade de récolte, servant de critère d'arrêt pour le décodage.

Comme aucune sortie précédente n'est disponible à la première étape du décodage T , l'entrée initiale du décodeur $\hat{\mathbf{y}}_{T-1}$ est générée en passant les mesures météorologiques \mathbf{m}_{T-1} dans un MLP.

5.1 Canaux Monotones par Construction

Parmi les variables $\{\mathbf{y}_t\}_{t=T}^T$, la production cumulée de biomasse et le stade de développement sont décrits par cinq variables monotones. Sur cette base, nous introduisons un module implémentant une architecture résiduelle, dans laquelle un MLP couplé à une activation ReLU en sortie garantit que les résidus restent positifs.

Pour une variable monotone i et un état caché augmenté du décodeur $\tilde{\mathbf{h}}_t$, le module proposé génère :

$$\hat{y}_t^i = \max(\mathcal{M}(\tilde{\mathbf{h}}_t), 0) + \hat{y}_{t-1}^i \quad (4)$$

où \mathcal{M} est un MLP conventionnel.

On peut visualiser intuitivement l'effet de ces canaux comme un biais inductif, contraignant l'espace d'hypothèses à des fonctions monotones par rapport aux variables concernées.

5.2 Pénalités Physiques

L'apprentissage des paramètres du modèle est guidé par une fonction de perte composite incluant l'objectif de prédiction supervisée et trois termes de régularisation physique :

$$\mathcal{L} = \alpha \mathcal{L}_d + \beta \mathcal{L}_{s1\text{-phy}} + \gamma \mathcal{L}_{s2\text{-phy}} + \delta \mathcal{L}_{d\text{-phy}} \quad (5)$$

où $\alpha := 1 - (\beta + \gamma + \delta)$, $\mathcal{L}_{d\text{-phy}}$ favorise la cohérence physique dans le temps, alors que $\mathcal{L}_{s1\text{-phy}}$ et $\mathcal{L}_{s2\text{-phy}}$ correspondent à des pénalités appliquées à chaque pas de temps.

La perte ℓ_2 , \mathcal{L}_d , quantifie l'erreur de prédiction sur toutes les sorties :

$$\mathcal{L}_d = \text{MSE}(\{\mathbf{y}_t\}_{t=T}^T, \{\hat{\mathbf{y}}_t\}_{t=T}^T) \quad (6)$$

$\mathcal{L}_{s1\text{-phy}}$ garantit la cohérence interne de la biomasse :

$$\mathcal{L}_{s1\text{-phy}} = ((\hat{y}_t^{\text{TWSO}} + \hat{y}_t^{\text{TWLV}} + \hat{y}_t^{\text{TWST}}) - \hat{y}_t^{\text{TAGP}})^2 \quad (7)$$

$\mathcal{L}_{s2\text{-phy}}$ impose la validité de la relation entre les nutriments disponibles et la respiration de maintenance :

$$\mathcal{L}_{s2\text{-phy}} = ((\hat{y}_t^{\text{GASS}} - \hat{y}_t^{\text{MRES}}) - \hat{y}_t^{\text{ASRC}})^2 \quad (8)$$

$\mathcal{L}_{d\text{-phy}}$ contraint la matière sèche totale disponible pour la partition le jour $t - 1$ à être cohérente avec l'augmentation totale de biomasse entre les jours $t - 1$ et t :

$$\mathcal{L}_{d\text{-phy}} = (\hat{y}_{t-1}^{\text{DMI}} - (\hat{y}_t^{\text{TWRT}} - \hat{y}_{t-1}^{\text{TWRT}}) - (\hat{y}_t^{\text{TAGP}} - \hat{y}_{t-1}^{\text{TAGP}}))^2 \quad (9)$$

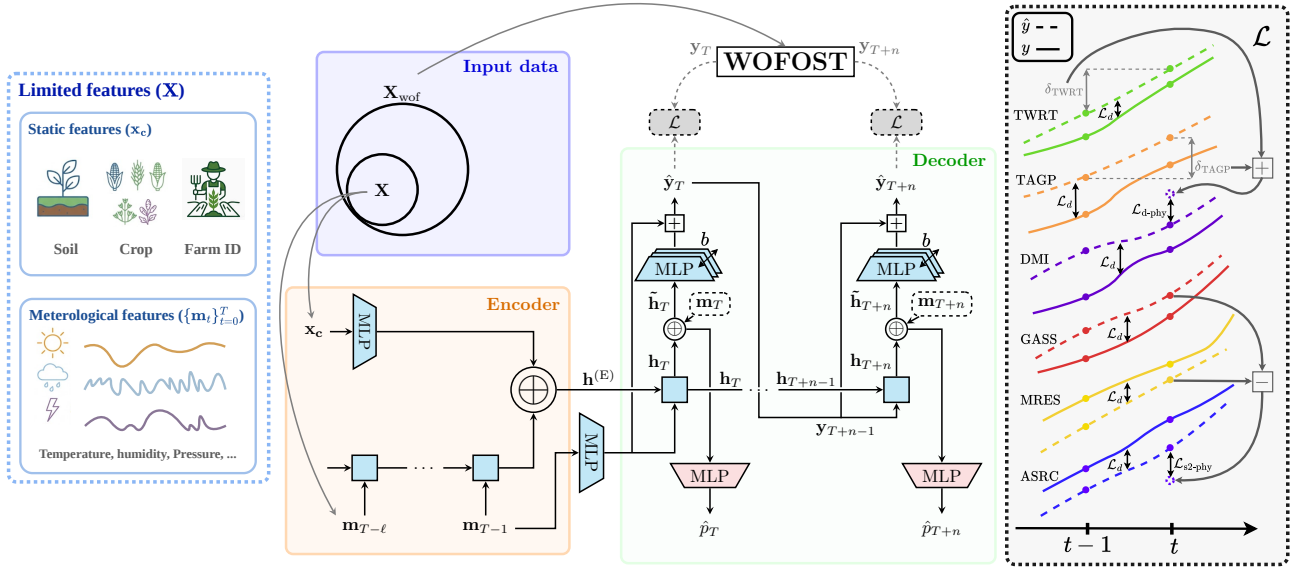


FIGURE 2 – Vue d’ensemble de l’architecture proposée. \oplus désigne la concaténation et $+$ la sommation. Suivant une approche similaire à WOFOST, le décodeur intègre progressivement les entrées météorologiques, tandis que la monotonie est imposée au sein des b MLPs du décodeur. La boîte de gauche détaille l’ensemble restreint d’entrées fournies au modèle de substitution, tandis que celle de droite illustre symboliquement la perte \mathcal{L}_d et les pertes physiques $\mathcal{L}_{d\text{-phy}}$ et $\mathcal{L}_{s2\text{-phy}}$.

6 Cadre expérimental

Les expériences ont été conçues pour étudier l’intégration progressive des connaissances biophysiques, en partant d’une prédiction ponctuelle du rendement purement basée sur les données, jusqu’à un modèle séquence-à-séquence entièrement guidé par la physique. Nous suivons trois étapes successives :

1. Intégration des connaissances observationnelles en prédisant des séquences temporelles : d’abord uniquement les biomasse (S-S2S), puis l’ensemble des variables de développement des cultures incluses dans \mathbf{y}_t (S2S).
2. Imposition d’un biais inductif pour garantir un comportement physiquement plausible en assurant la monotonie des variables concernées via une contrainte structurelle (S2S $_m$).
3. Ajout de lois physiques comme biais d’apprentissage en introduisant des termes de pénalité dans la fonction de perte (S2S $_{msd}$).

Des expériences supplémentaires ont été conduites afin d’évaluer systématiquement les interactions entre les différentes stratégies d’intégration des connaissances, en considérant toutes les combinaisons possibles des trois étapes décrites précédemment (S-S2S $_m$, S2S $_{ms}$, S2S $_{md}$, S2S $_{sd}$, S2S $_s$ et S2S $_d$). Le code nécessaire pour reproduire les expériences est disponible sur GitHub².

Les expériences sont réalisées sur une machine équipée d’un processeur Intel(R) Core(TM) i7-14700 et d’un GPU NVIDIA RTX A6000. L’implémentation repose principalement sur PyTorch. L’entraînement est effectué avec l’optimiseur AdamW (Loshchilov and Hutter, 2017) et le scheduler de taux d’apprentissage *one-cycle* (Smith, 2018). Tous les cycles d’entraînement utilisent 60 *epochs*, une taille de batch fixée à 64 et un taux d’apprentissage initial de $1e-4$. Pendant l’entraînement, du *teacher forcing* est utilisé, permettant au décodeur de traiter l’ensemble de la séquence en parallèle et accélérant considérablement l’apprentissage. Lors de l’inférence, le modèle passe en décodage autoregressif.

2. https://github.com/JulAgu/pg-sequence_learning

L’optimisation des hyperparamètres se limite aux coefficients des termes de la fonction de perte. Les hyperparamètres liés à l’architecture, tels que la profondeur du réseau ou le nombre de couches, sont maintenus constants, et tous les poids du modèle sont initialisés par des valeurs aléatoires uniformes dans l’intervalle $[-1/\sqrt{m}, 1/\sqrt{m}]$, où m est le nombre de variables dans \mathbf{X} .

Baselines Nous incluons des modèles de référence de prédiction ponctuelle ignorant la structure temporelle; à savoir, des modèles substitués qui prédisent $y_{t=T}^{\text{TWSO}}$ (le rendement) à partir de \mathbf{X} . Plus précisément, nous considérons : (1) un MLP opérant sur une représentation aplatie de \mathbf{X} , combinant les variables statiques et l’ensemble de la séquence météorologique, similaire au modèle substitut DSSAT proposé par de Freitas Cunha et al. (2023); et (2) un MLP appliqué à la représentation latente produite par un MLP-GRU traitant la séquence météorologique complète et les variables statiques ($X \cup \{\mathbf{m}_t\}_T^T$), c’est-à-dire la même architecture que celle utilisée pour l’encodeur du modèle proposé (Figure 2), mais avec un MLP comme décodeur.

6.1 Métriques d’évaluation

La performance des modèles est évaluée à l’aide de l’erreur absolue moyenne normalisée (MASE) (Hyndman and Koehler, 2006), qui fournit une mesure invariante à l’échelle de la précision de prédiction sur des séries temporelles multivariées. En normalisant la MAE par rapport à celle d’une prévision naïve à un pas. La MASE permet la comparaison entre différentes séries temporelles et fournit une métrique moyenne agrégée :

$$\overline{\text{MASE}} = \frac{1}{I+1} \sum_{i=0}^I \frac{\frac{1}{n+1} \sum_{t=T}^{\mathcal{T}} |y_t^i - \hat{y}_t^i|}{\frac{1}{n} \sum_{t=T+1}^{\mathcal{T}} |y_t^i - y_{t-1}^i|} \quad (10)$$

où $I = 10$ et $i = 0, 1, \dots, 10$ correspond à DVS, LAI, ..., MRES (voir Table 1).

De plus, compte tenu de l’intérêt pratique pour la prédiction du rendement, nous évaluons les substitués sur $y_{\mathcal{T}}^{\text{TWSO}}$ en utilisant la MAE, même si cet objectif n’a pas été explicitement renforcé pendant l’entraînement.

Pour mesurer la cohérence physique, nous introduisons des métriques spécifiques fondées sur les termes de régularisation introduits en 5. \mathbf{V}_{mon} mesure la fraction moyenne de la longueur de la série temporelle où la monotonie est violée pour les variables concernées.

$$\mathbf{V}_{mon} = \frac{1}{n-1} \sum_{t=T}^{\mathcal{T}-2} \sum_{i=0}^J \max(0, \hat{y}_t^i - \hat{y}_{t+1}^i) \quad (11)$$

où $J = 5$ et $j = 0, 1, \dots, 5$ correspond à DVS, TAGP, TWSO, TWLV, TWST, TWRT.

\mathbf{V}_{biom} correspond à la moyenne de la différence absolue entre la somme des biomasses des organes aériennes et la biomasse totale :

$$\mathbf{V}_{biom} = \frac{1}{n+1} \sum_{t=T}^{\mathcal{T}} |(\hat{y}_t^{\text{TWSO}} + \hat{y}_t^{\text{TWLV}} + \hat{y}_t^{\text{TWST}}) - \hat{y}_t^{\text{TAGP}}| \quad (12)$$

\mathbf{V}_{ass} correspond à la moyenne de la différence absolue entre les nutriments disponibles et les nutriments nets :

$$\mathbf{V}_{ass} = \frac{1}{n+1} \sum_{t=T}^{\mathcal{T}} |(y_t^{\text{GASS}} - \hat{y}_t^{\text{MRES}}) - \hat{y}_t^{\text{ASRC}}| \quad (13)$$

Enfin, \mathbf{V}_{mat} est définie comme la moyenne de la différence absolue entre la partition totale disponible et l’augmentation journalière de biomasse :

$$\mathbf{V}_{mat} = \frac{1}{n+1} \sum_{t=T}^{\mathcal{T}} |\hat{y}_{t-1}^{\text{DMI}}(\hat{y}_t^{\text{TWRT}} - \hat{y}_{t-1}^{\text{TWRT}}) - (\hat{y}_t^{\text{TAGP}} - \hat{y}_{t-1}^{\text{TAGP}})| \quad (14)$$

Nous utilisons la différence absolue plutôt que la différence au carré (comme dans les Équations (7), (8), (9)) afin de conserver les unités des variables.³

Bien que ces métriques capturent des aspects variés de la cohérence physique, leurs échelles différentes rendent la comparaison difficile. Pour pallier ce problème, nous introduisons le *PhysicScore*, une mesure agrégée des violations normalisées min-max entre modèles.

$$\text{PhysicScore} = \frac{1}{4} \sum_k^4 \frac{\mathbf{V}_k - \min(\mathbf{V}_k)}{\max(\mathbf{V}_k) - \min(\mathbf{V}_k)} \quad (15)$$

où $k = 1, 2, 3, 4$ correspond respectivement à *mon, ass, biom, mat*. $\min(\mathbf{V}_k)$ et $\max(\mathbf{V}_k)$ sont calculés sur l'ensemble des modèles.

Enfin, pour évaluer la performance hors distribution (OOD), nous définissons un indicateur de dégradation de la MASE :

$$\Delta \overline{\text{MASE}}_{\text{OOD}} = \frac{(\overline{\text{MASE}}_{\text{OOD}} - \overline{\text{MASE}})}{\overline{\text{MASE}}} \quad (16)$$

7 Résultats expérimentaux

Dans cette section, l'indice m désigne la contrainte architecturale de monotonie, tandis que s et d indiquent que la fonction de perte inclut respectivement les termes de pénalité statiques ($\mathcal{L}_{s1\text{-phy}}$, $\mathcal{L}_{s2\text{-phy}}$) et dynamiques ($\mathcal{L}_{d\text{-phy}}$). Le Tableau 2 résume les principaux résultats pour les trois familles de modèles : les modèles de référence à prédiction ponctuelle décrits en Section 6 (MLP et MLP-GRU), les modèles seq2seq entraînés uniquement sur des séquences de sortie liées à la biomasse (S-S2S, S-S2S $_m$), et finalement les modèles seq2seq qui intègrent d'autres séries temporelles et différentes combinaisons de contraintes physiques.

Modèle	$\overline{\text{MASE}}$	$\text{MAE}_{\text{yield}}$	$\Delta \overline{\text{MASE}}_{\text{OOD}}$	$\overline{\mathbf{V}}_{mon}$	\mathbf{V}_{biom}	\mathbf{V}_{ass}	\mathbf{V}_{mat}
MLP	-	2079.236	-	-	-	-	-
MLP-GRU	-	1935.846	-	-	-	-	-
S-S2S	0.287	1445.836	0.380	0.091	-	-	-
S-S2S $_m$	0.266	1134.727	0.564	0	-	-	-
S2S	0.143	671.954	7.881	0.182	10.254	2.252	9.738
S2S $_s$	0.161	722.312	7.633	0.141	4.519	0.660	12.830
S2S $_d$	0.140	573.741	7.264	0.192	12.756	2.890	2.060
S2S $_{sd}$	0.139	480.627	8.043	0.180	1.356	0.510	2.210
S2S $_m$	0.131	549.320	9.763	0	24.401	2.069	1.114
S2S $_{ms}$	0.141	495.730	8.255	0	134.119	0.890	1.509
S2S $_{md}$	0.212	2183.760	5.514	0	1297.219	2.581	5.470
S2S $_{msd}$	0.150	474.619	6.833	0	69.199	0.702	4.003

TABLE 2 – Résultats agrégés sur $\mathcal{D}_{\text{OOD-GP}}$. Les configurations Pareto-optimales sont en orange.

Étant donnée la nature multi-objectif de la tâche, l'analyse fondée uniquement sur les métriques individuelles peut s'avérer trompeuse pour évaluer les compromis entre précision du substitut, cohérence physique et robustesse hors distribution. Afin de fournir une évaluation plus claire de ces objectifs

3. Il convient de rappeler que $\mathcal{T} = T + n$

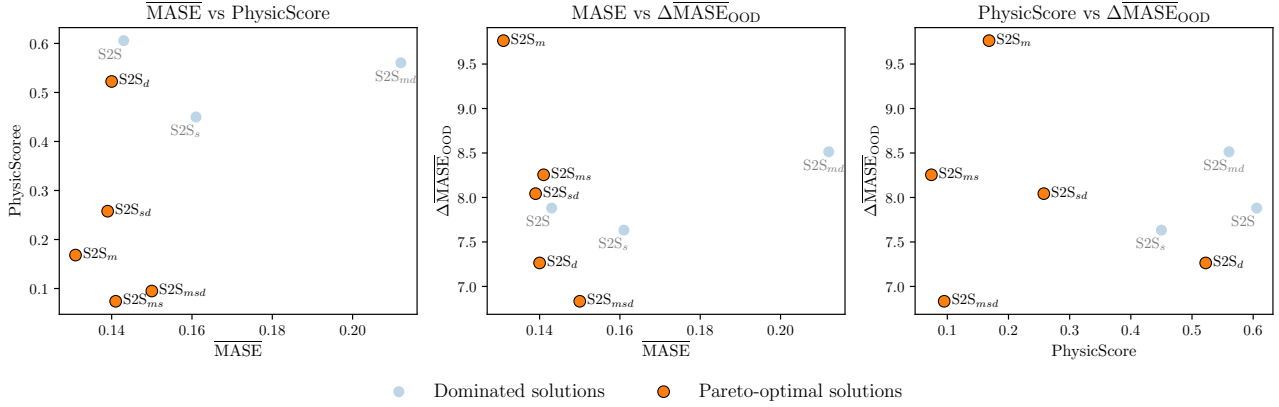


FIGURE 3 – Projections 2D du front de Pareto tridimensionnel (les valeurs les plus faibles sont les meilleures). La dominance est calculée en 3D selon PhysicScore, MASE et $\Delta\overline{\text{MASE}}_{\text{OOD}}$. Les solutions Pareto-optimales sont mises en évidence en orange ; leur position illustre les compromis optimaux entre les trois objectifs d’évaluation. Par exemple, lorsque la performance hors distribution (OOD) est prioritaire, le modèle S2S_{msd} apparaît comme le plus favorable.

concurrents, nous calculons un front de Pareto tridimensionnel permettant d’identifier l’ensemble des modèles non dominés (Sener and Koltun, 2019). Dans notre configuration expérimentale, les modèles Pareto-optimaux sont S2S_d , S2S_{sd} , S2S_m , S2S_{ms} et S2S_{msd} . La Figure 3 présente des projections bidimensionnelles de cet espace tridimensionnel, permettant de distinguer les solutions dominées des solutions non dominées.

Bien que la vitesse d’exécution ne constitue pas notre objectif principal, notre substitut est nettement plus rapide que WOFOST. Dans la configuration actuelle, le modèle exploite l’accélération GPU et atteint 160 ms par batch (128 échantillons dans nos expériences), contre 300 ms par échantillon pour WOFOST exécuté sur CPU.

L’apprentissage séquentiel surpasse la prédiction ponctuelle Les modèles seq2seq entraînés sur les trajectoires WOFOST surpassent les approches de prédiction ponctuelle, même lorsque ces dernières sont spécifiquement optimisées pour la prédiction du rendement. Le Tableau 2 montre que la prédiction de séquences complètes améliore systématiquement la précision du rendement. De plus, l’ajout de signaux auxiliaires apporte des gains moyens supplémentaires de 55 % sur la $\overline{\text{MASE}}$ et de 61 % sur la $\text{MAE}_{\text{yield}}$. Ces résultats suggèrent que les trajectoires générées par WOFOST agissent comme des connaissances observationnelles a priori : la prédiction de variables supplémentaires améliore de manière systématique les performances globales et par variable (voir l’Annexe B pour les résultats détaillés).

Effet des contraintes physiques Les modèles intégrant des contraintes physiques présentent différents avantages par rapport au substitut purement guidé par les données S2S. Certaines configurations améliorent la précision prédictive globale, tandis que d’autres renforcent la cohérence physique. Comme indiqué dans le Tableau 2, l’utilisation exclusive de contraintes souples, sous forme de termes de pénalisation physique, induit des comportements distincts. Le modèle S2S_d améliore simultanément la précision du rendement et la cohérence dynamique (\mathbf{V}_{mat}). Le modèle S2S_s réduit significativement les violations statiques (\mathbf{V}_{biom} et \mathbf{V}_{ass}), au prix d’une dégradation de la $\overline{\text{MASE}}$. Enfin, la combinaison des deux pénalités (S2S_{sd}) offre le meilleur compromis entre précision et cohérence physique, atteignant la plus faible $\text{MAE}_{\text{yield}}$ tout en réduisant fortement les violations statiques et dynamiques, bien qu’avec une légère perte de robustesse hors distribution.

De manière similaire, les configurations imposant une contrainte monotone stricte éliminent entièrement les violations de monotonie ($\mathbf{V}_{mon} = 0$) par construction, garantissant une dynamique d’accumulation de biomasse physiquement plausible. Toutefois, ce biais inductif peut entrer en conflit avec d’autres contraintes physiques, comme le montre l’augmentation systématique des violations du bilan de biomasse. Les effets d’interaction entre contraintes rendent cette incompatibilité encore plus visible, notamment pour $S2S_{md}$, qui présente une forte dégradation de la précision prédictive accompagnée de violations sévères de biomasse. À l’inverse, le modèle entièrement contraint $S2S_{msd}$ atténue cet effet et obtient la plus faible dégradation hors distribution tout en conservant une précision compétitive. Les conflits entre objectifs concurrents lors de l’entraînement des réseaux neuronaux sont bien connus dans la littérature. Des travaux antérieurs ont notamment montré que l’introduction de régularisations physiques peut entraîner des vitesses de convergence déséquilibrées entre composantes de la fonction de perte (Wang et al., 2022), et que la combinaison de contraintes par construction (dures) et souples peut accentuer ce phénomène, conduisant à une dégradation de la précision prédictive (Beucler et al., 2021).

Dans l’ensemble, comme illustré en Figure 3, l’intégration progressive de connaissances biophysiques ne conduit pas à un seul modèle dominant. Elle produit plutôt une famille de substituts Pareto-optimaux surpassant systématiquement l’approche purement guidée par les données en termes de cohérence physique, tout en maintenant une précision prédictive compétitive. Du point de vue de la modélisation substitutive, ces modèles constituent des solutions préférables, car ils respectent davantage les contraintes biophysiques connues sans sacrifier les performances globales. En outre, la capacité à produire des prédictions physiquement cohérentes tout au long du processus de croissance renforce la confiance des utilisateurs, en leur permettant de suivre l’évolution de la plante et de vérifier la plausibilité de son comportement. Les modèles proposés héritent ainsi de l’un des principaux atouts des modèles biophysiques, tout en nécessitant un nombre plus réduit de variables d’entrée.

Robustesse hors distribution (OOD) et capacité d’apprentissage en *few-shot* Un avantage majeur de l’apprentissage automatique guidé par la physique réside dans sa capacité à maintenir une cohérence au-delà du domaine d’entraînement (Bonfanti et al., 2024). Des travaux antérieurs ont notamment montré que les PINNs peuvent produire des extrapolations fiables et surpasser les modèles purement guidés par les données (Guo et al., 2024). Comme indiqué dans le Tableau 2, le modèle intégrant toutes les contraintes physiques $S2S_{msd}$ présente la plus faible dégradation de performance en conditions hors distribution. Toutefois, sa précision reste inférieure à celle d’un prédicteur naïf ($\overline{\text{MASE}} > 1$), ce qui empêche son utilisation directe. Cette limitation peut être levée grâce à des stratégies d’adaptation telles que le *few-shot learning*, permettant d’obtenir des résultats satisfaisants dans de nouveaux contextes.

La figure 4 présente les performances de plusieurs modèles dans différentes configurations hors distribution (OOD). Dans l’ensemble, les modèles guidés par la physique montrent une plus grande robustesse face à la dégradation des performances. Plus précisément, pour le scénario OOD le long de l’axe géographique uniquement ($\mathcal{D}_{\text{OOD-G}}$), l’intégration des connaissances physiques permet d’utiliser directement le modèle. Dans le scénario OOD le plus exigeant ($\mathcal{D}_{\text{OOD-GP}}$), les performances sont fortement dégradées ($\overline{\text{MASE}} > 1$). Cependant, il est possible de restaurer efficacement les performances grâce à un apprentissage *few-shot* : un ajustement fin du modèle pendant 10 *epochs* sur 10 parcelles sélectionnées aléatoirement dans le domaine cible OOD suffit à obtenir des performances utilisables en pratique.

8 Conclusions, limites et perspectives

Nous développons un modèle substitut seq2seq guidé par la physique pour la simulation de la croissance des cultures. Ce modèle approxime WOFOST en réduisant significativement les variables d’entrée

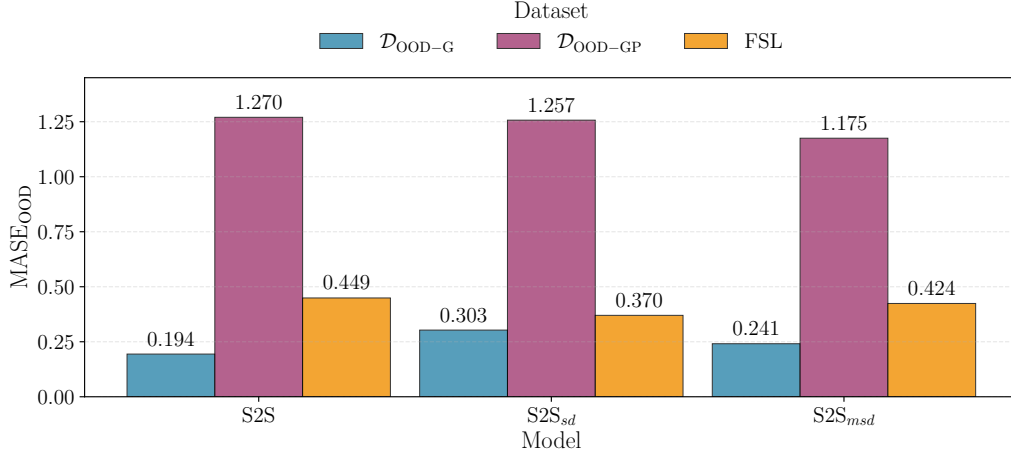


FIGURE 4 – Précision de certains modèles selon les trois modes de partitionnement. La valeur $\overline{\text{MASE}}$ est systématiquement indiquée pour le sous-ensemble OOD. FSL désigne les résultats obtenus en appliquant un apprentissage par peu d'exemples (*few-shot*) aux modèles $\mathcal{D}_{\text{OOD-GP}}$.

et en simplifiant son utilisation opérationnelle. Notre approche intègre systématiquement des connaissances biophysiques en introduisant des biais d'observation, inductifs et d'apprentissage appropriés via des contraintes architecturales et des pénalités physiques dans la fonction de perte. Nos résultats suggèrent que certaines combinaisons de ces contraintes améliorent la précision des prédictions, renforcent la cohérence physique et offrent une plus grande robustesse dans des scénarios hors-distribution par rapport aux modèles purement fondés sur les données. Nous proposons un ensemble d'options viables pour l'intégration des connaissances dans un contexte applicatif spécifique, offrant des compromis entre précision et plausibilité physique, ce qui en fait un outil prometteur pour la recherche agronomique.

Le modèle substitut est entraîné sur des simulations générées par WOFOST, et hérite donc de ses hypothèses, paramétrages et biais sous-jacents. Par conséquent, dans son état actuel, il ne peut pas dépasser les performances du modèle original et nécessite une calibration pour s'adapter à des conditions agronomiques particulières. De la même manière, la version spécifique du substitut présentée dans cet article est entraînée sur un ensemble non exhaustif de scénarios climatiques et de pratiques agricoles, limité à notre zone d'étude, ce qui peut restreindre sa capacité à extrapoler de manière fiable vers des conditions extrêmes ou inédites.

L'intégration de la quantification de l'incertitude et le développement des mécanismes de calibration seront au cœur des prochaines étapes afin d'améliorer l'applicabilité du modèle dans le monde réel. Par ailleurs, bien que nos résultats fournissent des intuitions sur la compatibilité entre différentes contraintes physiques, une étude systématique des mécanismes sous-jacents responsables de ces incompatibilités permettra d'introduire progressivement des contraintes supplémentaires et d'élargir le domaine de validité.

Références

- Ahmad, S., Abbas, G., Fatima, Z., Hussain, S., Khan, M. A., Ahmed, M., Wilkerson, C. J., and Hoogenboom, G. (2025). Modeling the impact of climate warming on tomato phenology. *Agricultural and Forest Meteorology*, 374 :110825.
- Bucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., and Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.*, 126 :098302.
- Bonfanti, A., Santana, R., Ellero, M., and Gholami, B. (2024). On the generalization of PINNs outside

- the training domain and the hyperparameters influencing it. *Neural Computing and Applications*, 36(36) :22677–22696.
- Cao, B., Yu, L., Zhong, L., Qiao, S., Tan, S., Huang, X., and Wang, H. (2026). Physics-guided deep learning for crop yield estimation. *European Journal of Agronomy*, 172 :127850.
- Cho, K., Merriënboer, B. v., Bahdanau, D., and Bengio, Y. (2014). On the Properties of Neural Machine Translation : Encoder-Decoder Approaches. arXiv :1409.1259 [cs].
- Corrales, D. C., Schoving, C., Raynal, H., Debaeke, P., Journet, E.-P., and Constantin, J. (2022). A surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern france. *Computers and Electronics in Agriculture*, 192 :106578.
- de Freitas Cunha, R. L., Silva, B., and Avegliano, P. B. (2023). A comprehensive modeling approach for crop yield forecasts using ai-based methods and crop simulation models.
- De Wit, A. and Boogaard (2021). A gentle introduction to WOFOST. *Wageningen Environmental Research*, 168 :1–50.
- De Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., Van Kraalingen, D., Supit, I., Van Der Wijngaart, R., and Van Diepen, K. (2019). 25 years of the WOFOST cropping systems model. *Agricultural Systems*, 168 :154–167.
- Donnelly, J., Daneshkhah, A., and Abolfathi, S. (2024). Physics-informed neural networks as surrogate models of hydrodynamic simulators. *Science of The Total Environment*, 912 :168814.
- Fang, R., Yu, L., Xie, N., Wang, K., Rao, K., Zhong, L., Chen, J., and Wu, Z. (2026). A knowledge-informed fusion network for process-aware crop yield prediction. *International Journal of Applied Earth Observation and Geoinformation*, 146 :105154.
- Fischer, R. (2015). Definitions and determination of crop yield, yield gaps, and of rates of change. *Field Crops Research*, 182. Publisher : Elsevier BV.
- Garzón, A., Kapelan, Z., Langeveld, J., and Taormina, R. (2022). Machine learning-based surrogate modeling for urban water networks : Review and future research directions. *Water Resources Research*, 58(5) :e2021WR031808. e2021WR031808 2021WR031808.
- Gherman, I. M., Abdallah, Z. S., Pang, W., Goroehowski, T. E., Grierson, C. S., and Marucci, L. (2023). Bridging the gap between mechanistic biological models and machine learning surrogates. *PLOS Computational Biology*, 19(4) :1–16.
- Guo, F., Li, A., Yue, B., Xiao, Z., Xiao, F., Yan, R., Li, A., Lv, Y., and Su, B. (2024). Improving the out-of-sample generalization ability of data-driven chiller performance models using physics-guided neural network. *Applied Energy*, 354 :122190.
- Haghighat, E., Raissi, M., Moure, A., Gomez, H., and Juanes, R. (2021). A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379 :113741.
- He, E., Xie, Y., Liu, L., Chen, W., Jin, Z., and Jia, X. (2023). Physics guided neural networks for time-aware fairness : An application in crop yield prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12) :14223–14231.
- Hussain, S., Teshome, F. T., Tulu, B. B., Awoke, G. W., Hailegnaw, N. S., and Bayabil, H. K. (2025). Leaf area index (lai) prediction using machine learning and uav based vegetation indices. *European Journal of Agronomy*, 168 :127557.

- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4) :679–688.
- Jain, S., Sethia, D., and Tiwari, K. C. (2024). A critical systematic review on spectral-based soil nutrient prediction using machine learning. *Environmental Monitoring and Assessment*, 196(8) :699.
- Jia, X., Willard, J., Karpatne, A., Read, J. S., Zwart, J. A., Steinbach, M., and Kumar, V. (2020). Physics-Guided Machine Learning for Scientific Discovery : An Application in Simulating Lake Temperature Profiles. arXiv :2001.11086 [cs].
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6) :422–440.
- Khordadi, M. J., Olesen, J. E., Alizadeh, A., Nassiri Mahallati, M., Ansari, H., and Sanaeinejad, H. (2019). Climate change impacts and adaptation for crop management of winter wheat and maize in the semi-arid region of iran. *Irrigation and Drainage*, 68(5) :841–856.
- Lagaris, I., Likas, A., and Fotiadis, D. (1998). Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5) :987–1000.
- Lee, H. and Kang, I. S. (1990). Neural algorithm for solving differential equations. *Journal of Computational Physics*, 91(1) :110–131.
- Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. In *International Conference on Learning Representations*.
- Luo, J., Ma, X., Ji, Y., Li, X., Song, Z., and Lu, W. (2023). Review of machine learning-based surrogate models of groundwater contaminant modeling. *Environmental Research*, 238 :117268.
- Maestrini, B., Mimić, G., Van Oort, P. A., Jindo, K., Brdar, S., Athanasiadis, I. N., and Van Evert, F. K. (2022). Mixing process-based and data-driven approaches in yield prediction. *European Journal of Agronomy*, 139 :126569.
- Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P. (2011). The rcp greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, 109(1) :213.
- Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks : A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378 :686–707.
- Sener, O. and Koltun, V. (2019). Multi-task learning as multi-objective optimization.
- Shahhosseini, M., Hu, G., Huber, I., and Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11(1). Publisher : Springer Science and Business Media LLC.
- Sharma, P., Dadheech, P., Aneja, N., and Aneja, S. (2023). Predicting agriculture yields based on machine learning using regression and deep learning. *IEEE Access*, 11 :111255–111264.
- Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters : Part 1 - learning rate, batch size, momentum, and weight decay. *ArXiv*, abs/1803.09820.
- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’14, page 3104–3112, Cambridge, MA, USA. MIT Press.

- Tao, S., Shen, S., Li, Y., Wang, Q., Gao, P., and Mugume, I. (2016). Projected crop production under regional climate change using scenario data and modeling : Sensitivity to chosen sowing date and cultivar. *Sustainability*, 8(3).
- Tao, Y., Chen, W.-H., Li, Z.-B., and He, W.-Y. (2025). Dynamic displacement reconstruction of bridge based on physics-informed recurrent neural network. *Advances in Bridge Engineering*, 6(1) :12.
- Trontelj ml., J. and Chambers, O. (2021). Machine learning strategy for soil nutrients prediction using spectroscopic method. *Sensors*, 21(12).
- Van Klompenburg, T., Kassahun, A., and Catal, C. (2020). Crop yield prediction using machine learning : A systematic literature review. *Computers and Electronics in Agriculture*, 177 :105709.
- von Bloh, M., Lobell, D. B., and Asseng, S. (2024). Knowledge informed hybrid machine learning in agricultural yield prediction. *Comput. Electron. Agric.*, 227 :109606.
- Wang, S., Yu, X., and Perdikaris, P. (2022). When and why pinns fail to train : A neural tangent kernel perspective. *Journal of Computational Physics*, 449 :110768.
- Xian, J. and Wang, Z. (2024). A physics and data co-driven surrogate modeling method for high-dimensional rare event simulation. *Journal of Computational Physics*, 510 :113069.
- Yenkikar, A., Mishra, V. P., Bali, M., and Ara, T. (2025). An explainable ai-based hybrid machine learning model for interpretability and enhanced crop yield prediction. *MethodsX*, 15 :103442.
- Zhang, N., Zhou, X., Kang, M., Hu, B.-G., Heuvelink, E., and Marcelis, L. F. M. (2023). Machine learning versus crop growth models : an ally, not a rival. *AoB PLANTS*, 15(2). Publisher : Oxford University Press (OUP).
- Zhang, Z., Xin, Q., and Li, W. (2021). Machine learning-based modeling of vegetation leaf area index and gross primary productivity across north america and comparison with a process-based model. *Journal of Advances in Modeling Earth Systems*, 13(10).
- Zheng, Y., Hu, C., Wang, X., and Wu, Z. (2023). Physics-informed recurrent neural network modeling for predictive control of nonlinear processes. *Journal of Process Control*, 128 :103005.

A De l'exploration de scénarios à l'analyse agronomique inverse

Les modèles de cultures fondés sur les processus sont largement utilisés pour explorer des scénarios agronomiques et évaluer des stratégies d'adaptation. Cependant, leur utilisation à des fins d'optimisation, comme la recherche de la combinaison d'entrées permettant d'atteindre des objectifs agronomiques précis, reste complexe. Ces modèles sont souvent non linéaires, discontinus et définis implicitement via des simulations numériques itératives. Résoudre des problèmes inverses dans ce contexte nécessite généralement des méthodes de recherche globale très coûteuses en calcul. Le modèle substitut séquence-à-séquence guidé par la physique que nous proposons pour WOFOST présente un avantage majeur, puisqu'il permet de résoudre des problèmes inverses par optimisation fondée sur le gradient, facilitant ainsi l'identification de configurations d'intrants qui produisent les résultats souhaités.

Pour les agronomes, cela revêt un intérêt particulier, car les essais sur le terrain sont coûteux, longs et soumis à la variabilité environnementale. Ainsi, notre substitut neural accélère non seulement l'analyse agronomique, mais élargit également le type de questions pouvant être abordées.

Pour illustrer cette capacité, nous étudions une question de recherche reprise dans plusieurs travaux de la littérature agronomique (Tao et al., 2016; Khordadi et al., 2019; Ahmad et al., 2025) : « Comment ajuster les dates de semis comme stratégie d'adaptation aux scénarios de changement climatique ?

». Pour y répondre, nous sélectionnons aléatoirement une parcelle de blé dans le jeu de données et remplaçons la série temporelle météorologique par les données issues des scénarios d’émissions +2.5, +4.5 et +8.5 des trajectoires de concentration représentatives (RCP) pour l’année 2035 (Meinshausen et al., 2011). Nous choisissons le modèle $S2S_{msd}$ et figeons ses paramètres, à l’exception des neurones représentant la date de semis dans la couche d’entrée. Ensuite, nous appliquons une descente de gradient stochastique avec une fonction de perte définie comme la valeur négative de \hat{y}_T^{TWSO} sur 10 *epochs*, obtenant ainsi la date optimale selon le substitut. Enfin, nous évaluons le résultat en le comparant à une recherche aléatoire basée sur un échantillonnage uniforme de 100 dates simulées avec WOFOST sur une période de quatre semaines centrée sur la date identifiée.

RCP	Date $S2S_{msd}$	Rendement $S2S_{msd}$	Date WOF	Rendement WOF
2.6	17/11	8300	10/11	7930
4.5	21/11	9210	23/11	8812
8.5	19/11	6825	02/11	7000

TABLE 3 – Rendements simulés avec WOFOST pour différents scénarios RCP et dates de semis. Pour une comparaison équitable, les rendements (colonnes 3 et 5) sont calculés avec WOFOST.

Pour l’expérience présentée dans le Tableau 3, notre modèle substitut identifie généralement des dates de semis qui conduisent à des rendements supérieurs ou comparables à ceux obtenus par une recherche aléatoire classique avec WOFOST. Pour les scénarios RCP 2.5 et 4.5, les dates prédites par $S2S_{msd}$ donnent des rendements légèrement supérieurs à ceux de WOFOST, ce qui suggère que le modèle peut optimiser efficacement les dates de semis dans des conditions de changement climatique modérées. En revanche, dans le scénario plus extrême RCP 8.5, la date prédite par $S2S_{msd}$ conduit à un rendement inférieur à celui obtenu avec la date WOFOST, ce qui indique que, malgré son guidage physique, la nature data-driven de notre substitut limite ses performances en dehors de son domaine d’entraînement.

La différentiabilité du modèle substitut ouvre des perspectives très prometteuses pour les problèmes inverses. L’efficacité de cette approche en fait un outil puissant pour les agronomes, offrant des capacités de simulation largement supérieures à celles des méthodes classiques basées sur l’échantillonnage. Bien que l’optimisation basée sur le gradient permette de résoudre plus facilement des problèmes inverses, si des contraintes supplémentaires ne sont pas imposées, les solutions obtenues peuvent correspondre à des pratiques agronomiquement irréalistes ou opérationnellement inapplicables.

Par ailleurs, étant donné que le modèle proposé est un substitut de type réduit, nous sacrifions un certain niveau de détail dans les entrées. Par exemple, alors que WOFOST permettrait de faire varier des dizaines de variables d’entrée pour trouver un profil phénologique adapté à un scénario climatique, notre substitut fournit une représentation latente de ce profil cible qui ne peut actuellement pas être décodé en variables explicites.

Modèle	Variable	MAE	RMSE	MASE	V_{mon}
S-S2S	DVS	0.121	0.262	0.257	0.044
	LAI	0.573	1.219	0.304	-
	TAGP	1037.848	2052.792	0.200	0.044
	TWSO	626.664	1552.684	0.389	0.187
S-S2S _m	DVS	0.101	0.219	0.215	0.000
	LAI	0.526	0.976	0.279	-
	TAGP	1094.163	2123.050	0.211	0.000
	TWSO	583.454	1530.782	0.362	0.000
S2S	DVS	0.026	0.063	0.055	0.000
	LAI	0.327	0.699	0.171	-
	TAGP	783.862	1900.356	0.148	0.057
	TWSO	381.649	1186.900	0.234	0.264
	TWLV	139.638	286.787	0.110	0.245
	TWST	319.820	761.757	0.125	0.239
	TWRT	70.802	182.185	0.098	0.289
	DMI	16.305	48.613	0.151	-
	ASRC	24.310	72.550	0.151	-
	GASS	32.205	91.413	0.155	-
MRES	9.444	30.197	0.170	-	

TABLE 4 – Résultats par variable sur \mathcal{D}_{OOD-GP} : modèles S-S2S et modèle S2S sans contraintes physiques

B Résultats Détaillés

Modèle	Variable	MAE	RMSE	MASE	V_{mon}
S2S _s	DVS	0.044	0.088	0.092	0.006
	LAI	0.353	0.732	0.185	-
	TAGP	839.382	1899.940	0.158	0.025
	TWSO	389.404	1168.657	0.239	0.208
	TWLV	184.278	349.274	0.145	0.233
	TWST	369.941	789.119	0.144	0.176
	TWRT	108.920	253.762	0.150	0.201
	DMI	17.484	47.546	0.162	-
	ASRC	26.061	71.010	0.162	-
	GASS	33.931	88.754	0.164	-
MRES	9.606	28.800	0.173	-	
S2S _d	DVS	0.026	0.067	0.055	0.000
	LAI	0.329	0.716	0.172	-
	TAGP	729.736	1879.193	0.138	0.088
	TWSO	372.187	1178.814	0.228	0.201
	TWLV	140.060	305.237	0.110	0.258
	TWST	293.194	747.280	0.114	0.327
	TWRT	80.262	193.256	0.111	0.277
	DMI	15.602	48.399	0.144	-
	ASRC	23.204	72.161	0.144	-
	GASS	31.116	91.249	0.150	-
MRES	9.516	30.657	0.171	-	
S2S _{sd}	DVS	0.019	0.054	0.040	0.000
	LAI	0.321	0.711	0.168	-
	TAGP	759.595	1914.497	0.143	0.044
	TWSO	388.732	1212.058	0.239	0.220
	TWLV	134.060	281.939	0.105	0.233
	TWST	293.744	737.226	0.115	0.245
	TWRT	72.200	175.336	0.099	0.340
	DMI	16.074	48.969	0.149	-
	ASRC	24.065	73.076	0.149	-
	GASS	32.166	92.199	0.155	-
MRES	9.457	30.351	0.170	-	
S2S _m	DVS	0.016	0.040	0.033	0.000
	LAI	0.312	0.695	0.163	-
	TAGP	703.791	1866.177	0.133	0.000
	TWSO	362.283	1175.186	0.222	0.000
	TWLV	125.057	277.962	0.098	0.000
	TWST	273.549	730.686	0.107	0.000
	TWRT	70.154	160.560	0.097	0.000
	DMI	15.029	48.557	0.139	-
	ASRC	22.365	72.372	0.139	-
	GASS	29.993	91.251	0.145	-
MRES	8.960	30.135	0.161	-	

TABLE 5 – Résultats par variable sur \mathcal{D}_{OOD-GP} : S2S et variantes simples

Modèle	Variable	MAE	RMSE	MASE	V_{mon}
S2S _{ms}	DVS	0.015	0.039	0.032	0.000
	LAI	0.339	0.711	0.178	-
	TAGP	752.585	1877.201	0.142	0.000
	TWSO	371.354	1179.302	0.228	0.000
	TWLV	142.921	282.821	0.112	0.000
	TWST	302.048	734.440	0.118	0.000
	TWRT	80.026	163.459	0.110	0.000
	DMI	16.126	48.340	0.149	-
	ASRC	24.142	72.057	0.150	-
	GASS	32.538	91.164	0.157	-
	MRES	9.680	30.352	0.174	-
S2S _{md}	DVS	0.020	0.049	0.041	0.000
	LAI	0.468	0.986	0.245	-
	TAGP	763.471	1837.925	0.144	0.000
	TWSO	1328.996	3231.431	0.815	0.000
	TWLV	134.081	280.946	0.105	0.000
	TWST	288.919	738.920	0.113	0.000
	TWRT	88.609	176.566	0.122	0.000
	DMI	18.831	51.857	0.174	-
	ASRC	28.138	77.387	0.175	-
	GASS	37.995	98.793	0.183	-
	MRES	12.295	33.150	0.222	-
S2S _{msd}	DVS	0.023	0.056	0.049	0.000
	LAI	0.350	0.730	0.184	-
	TAGP	826.579	1817.856	0.156	0.000
	TWSO	367.588	1125.963	0.226	0.000
	TWLV	165.122	304.896	0.130	0.000
	TWST	317.018	701.262	0.124	0.000
	TWRT	99.919	204.754	0.138	0.000
	DMI	16.727	48.328	0.155	-
	ASRC	25.001	72.154	0.155	-
	GASS	33.309	90.643	0.161	-
	MRES	9.895	29.830	0.178	-

TABLE 6 – Résultats par variable sur \mathcal{D}_{OOD-GP} : modèles S2S combinant différentes contraintes.

C Optimisation des coefficients

Pour explorer de manière systématique l’effet des coefficients de la fonction de perte sur les performances du modèle, nous avons réalisé une recherche par grille limitée aux coefficients de la fonction de perte guidée par la physique, α , β , γ et δ , qui contrôlent la contribution relative de chaque composante de la perte. Le tableau 7 présente les différentes configurations testées, chaque ligne correspondant à une combinaison spécifique de valeurs de coefficients utilisée lors de la recherche.

Model	α	β	γ	δ
S2S _s	0.8	0.1	0.1	-
	0.9	0	0.1	-
	0.9	0.1	0	-
	0.33	0.33	0.33	-
	0.5	0	0.5	-
	0.5	0.5	0	-
S2S _d	0.9	-	-	0.1
	0.8	-	-	0.2
	0.5	-	-	0.5
S2S _{sd}	0.7	0.1	0.1	0.1
	0.7	0	0.15	0.15
	0.7	0.15	0	0.15
	0.25	0.25	0.25	0.25
	0.8	0.1	0	0.1
	0.8	0	0.1	0.1
	0.33	0	0.33	0.33
	0.33	0.33	0	0.33
S2S _{ms}	0.8	0.1	0.1	-
	0.9	0	0.1	-
	0.9	0.1	0	-
	0.33	0.33	0.33	-
	0.5	0	0.5	-
	0.5	0.5	0	-
S2S _{md}	0.9	-	-	0.1
	0.8	-	-	0.2
	0.5	-	-	0.5
S2S _{msd}	0.7	0.1	0.1	0.1
	0.7	0	0.15	0.15
	0.7	0.15	0	0.15
	0.25	0.25	0.25	0.25
	0.8	0.1	0	0.1
	0.8	0	0.1	0.1
	0.33	0	0.33	0.33
	0.33	0.33	0	0.33

TABLE 7 – Valeurs des coefficients pour les différentes configurations d’optimisation S2S selon les quatre types de paramètres : α , β , γ et δ . Un tiret (-) indique qu’aucun coefficient n’a été attribué à ce paramètre dans la configuration correspondante.

Le tableau 8 présente les hyperparamètres retenus pour les expériences principales. Ces configurations ont été sélectionnées à partir de la recherche par grille précédente en fonction de leurs performances sur l’ensemble de validation.

Model	α	β	γ	δ
S2S _s	0.8	0.1	0.1	-
S2S _d	0.9	-	-	0.1
S2S _{sd}	0.7	0.1	0.1	0.1
S2S _{ms}	0.9	0	0.1	-
S2S _{md}	0.9	-	-	0.1
S2S _{msd}	0.7	0.1	0.1	0.1

TABLE 8 – Hyperparamètres retenus pour les expériences principales avec les modèles S2S. Chaque ligne indique les coefficients sélectionnés (α , β , γ , δ) utilisés dans la fonction de perte pour la configuration correspondante. Un tiret (-) signifie qu'aucun coefficient n'a été attribué pour ce paramètre.