

# Méthodes probabilistes d'apprentissage profond avec a priori physiques de représentations interprétables

Yoël ZERAH, Silvia VALERO, Jordi INGLADA

CESBIO-UMR 5126, 18 avenue Edouard Belin, 31401 Toulouse CEDEX 9, France

yoel.zerah@cesbio.cnes.fr, silvia.valero@cesbio.cnes.fr, jordi.inglada@cesbio.cnes.fr

**Résumé** – Nous proposons une méthodologie d'apprentissage profond, intégrant des contraintes physiques, de représentation latentes probabilistes et interprétables, sémantiquement liées aux paramètres de modèles physiques. La méthode est appliquée à des images satellitaires optiques pour estimer des métriques de phénologie de végétation.

**Abstract** – We propose a physics-guided deep learning methodology for discovering probabilistic latent spaces to be semantically grounded to the parameters of physical models. The method is applied on optical satellite images to estimate phenological vegetation metrics.

## 1 Introduction

Réaliser l'apprentissage automatique de représentations de données permet d'en extraire l'information pertinente pour une application postérieure : cela permet de réduire la dimension de données d'entrée, de découvrir des motifs, de généraliser des comportements. Les méthodes d'apprentissage de représentations sont usuellement non supervisées, les "bonnes" représentations d'une donnée ne sont pas nécessairement connues et doivent être découvertes par le système apprenant. Cela affranchit également l'apprentissage de jeux de données étiquetés fastidieux ou coûteux à produire.

Cependant les représentations produites par apprentissage font souvent face à des problèmes d'interprétabilité. De plus, elles ne sont pas forcément compréhensibles par les humains, et ne sont pas justifiables. A ce titre, l'approche de la *désintrication*<sup>1</sup>[8] impose des structures et *a priori* aux représentations apprises pour les rendre plus interprétables. Néanmoins, une signification de ces représentations doit souvent être retrouvée et vérifiée *a posteriori*.

Dans le cas particulier des mesures physiques issues de capteurs, le comportement de la donnée observée fait typiquement l'objet de modélisations réalisées par des experts. Nous considérons que les modèles physiques sont de "bonnes" représentations de ces données particulières, et proposons de contraindre l'apprentissage de représentations à être lié à ces modèles physiques, avec un modèle génératif profond.

Nous appliquons ici notre méthodologie à des séries temporelles issues d'imagerie satellite d'observation de la Terre, en particulier celle de la constellation Sentinel-2 (S2). Nous considérons des modèles paramétriques *phénologiques*<sup>2</sup> [10]

décrivant l'évolution saisonnière des indices spectraux de pixels d'images S2 correspondant à de la végétation.

Nous proposons ici une méthode d'apprentissage intégrant des modèles physiques dans une structure d'auto-encodeurs, et l'appliquons à la prédiction de *paramètres phénologiques* comme représentation de séries temporelles d'indices spectraux issues d'images S2.

## 2 Apprentissage guidé par un modèle

### 2.1 Auto-encodeur à décodeur-simulateur

L'apprentissage de représentations basé sur les modèles génératifs est classiquement fondée sur des architectures de type auto-encodeur (AE). L'encodeur réduit la donnée d'entrée en un *code* et le décodeur est un processus générant une reconstruction de la donnée d'entrée à partir de ce code. Les AE sont optimisés pour maximiser la qualité de la reconstruction, par exemple en minimisant une erreur quadratique moyenne. Le code inféré est une représentation de la donnée maximisant la compression, et n'est pas forcément interprétable tel quel.

Le décodeur est un modèle génératif prenant en entrée le code, et appris lors de l'entraînement. Pour que le code soit interprétable, nous proposons que le modèle contenu dans le décodeur ne soit pas appris, mais imposé. Pour cela, nous remplaçons le réseau de neurones du décodeur par un simulateur (différentiable) de données  $\Omega_\phi$ , qui implémente un modèle physique. Les paramètres de ce modèle sont alors les composantes du code  $\phi$ , interprétables par définition (voir figure 1). Dans notre approche, seul l'encodeur est entraîné. Une fois optimisé, il réalise une inversion du modèle du décodeur, et produit une estimation des paramètres de ce modèle en tant que représentation. L'approche d'auto-encodeur à décodeur-simulateur (AEDS) a notamment été utilisée dans le

1. **Disentanglement.**

2. L'étude des variations des phénomènes périodiques, annuels en l'occurrence, de la végétation en fonction du climat.

cas d'images [2], et étendue aux auto-encodeurs variationnels (VAE) dans [7].

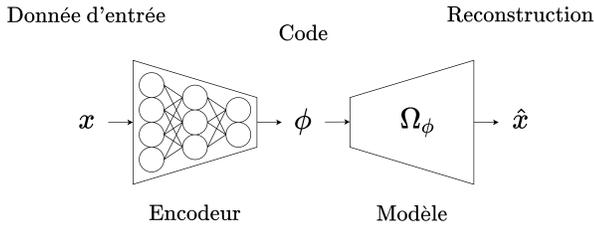


FIGURE 1 – Auto-encodeur à décodeur-simulateur

## 2.2 Représentations probabilistes

### 2.2.1 Inférence de distributions

Les VAE [6] sont une classe d'AE dont l'encodeur ne produit pas directement un code, mais infère les paramètres d'une distribution de probabilité préalablement choisie parmi une famille de distributions. Ce sont les réalisations de cette distribution (constituant l'*espace latent*), obtenues de manières différentiables<sup>3</sup>, qui sont décodées.

Encoder l'information sous la forme d'une distribution fournit une information plus riche qu'une estimation ponctuelle, avec notamment une description de l'incertitude.

À l'instar du VAE, nous réalisons l'inférence de représentations comme les paramètres de distributions de probabilité - mais dont les tirages sont les paramètres du modèle employé dans le décodeur. Nous proposons ainsi une version probabiliste de l'AEDS.

### 2.2.2 Fonction de coût

Pour entraîner un modèle d'AEDS probabiliste, nous utilisons comme fonction objectif la NLL<sup>4</sup> gaussienne de la donnée d'entrée  $x = (x_1, \dots, x_n)$  par rapport à la distribution des reconstructions (obtenue par composition de la distribution latente avec le générateur du décodeur) :

$$\mathcal{L}(x, \mu_{\hat{x}}, \Sigma_{\hat{x}}) = \frac{1}{2} (x - \mu_{\hat{x}})^\top \Sigma_{\hat{x}}^{-1} (x - \mu_{\hat{x}}) + \ln \left( |\Sigma_{\hat{x}}|^{1/2} \right) \quad (1)$$

avec  $\mu_{\hat{x}} = (\mu_{\hat{x},1}, \dots, \mu_{\hat{x},n})$  le vecteur des moyennes par composante de la distribution des reconstructions  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$  et  $\Sigma_{\hat{x}}$  la matrice de covariance des composantes de  $\hat{x}$ . Minimiser cette fonction objectif revient à minimiser l'erreur moyenne de reconstruction, et ajuster la variance des reconstructions - faible si l'erreur<sup>5</sup> est faible, élevée si l'erreur est élevée. La NLL comme fonction objectif permet d'éviter l'effondrement des distributions inférées vers des distributions ponctuelles [1].

Cette fonction objectif est équivalente au terme de reconstruction d'un VAE classique à décodeur gaussien — nous ne

faisons pas l'hypothèse d'un a priori  $p(\phi)$  particulier, et ne pouvons alors pas inclure le terme de Kullback-Leibler.

Lors de l'entraînement, pour chaque donnée d'entrée encodée, nous effectuons le tirage de plusieurs échantillons des distributions latentes (contrairement au VAE où l'on en tire un seul). Le vecteur des moyennes  $\mu_{\hat{x}}$  et la matrice de covariance  $\Sigma_{\hat{x}}$  utilisés dans la fonction objectif sont calculées empiriquement sur la distribution de reconstructions issues du décodage de chaque échantillon latent.

## 2.3 Contraintes sur l'espace latent

Les paramètres du modèle qui doivent être prédits par l'AEDS probabiliste peuvent correspondre à des paramètres physiques, soumis à des contraintes, telles que des relations d'ordre, ou l'appartenance à des intervalles fermés  $[\alpha, \beta]$ . L'intervalle d'appartenance de ces variables est une information a priori complémentaire au modèle du décodeur que l'on peut régler.

Pour effectuer le tirages de variables  $\phi$  sur  $[\alpha, \beta]$ , nous effectuons un tirage  $z$  selon une distribution à support dans  $[0, 1]$ , puis réalisons une transformation affine :  $\phi = (\beta - \alpha)z + \alpha$ . La sélection de cette distribution est aussi un choix de conception.

Nous utilisons ici la distribution de Kumaraswamy [3], dépendant de deux paramètres  $a$  et  $b$ , notée  $\mathcal{K}(a, b)$  de support borné  $[0, 1]$ . C'est une loi flexible, dont l'expression analytique de la fonction de répartition inverse est connue :  $F_{\mathcal{K}}^{-1}(x, a, b) = (1 - (1 - x)^{1/b})^{1/a}$

Afin de réaliser des tirages de cette loi de manière différentiable par rapport à  $a$  et  $b$ , nous utilisons la *méthode de la transformée inverse* :

$$u \sim \mathcal{U}(0, 1) \Rightarrow F_{\mathcal{K}}^{-1}(u) \sim \mathcal{K}(a, b)$$

Pour réaliser le tirage de variables ordonnées  $\phi_1 < \phi_2$ , nous tirons préalablement  $\phi_1$ , puis calculons la somme :  $\phi_2 = \phi_1 + \Delta_{1,2}$  avec  $\Delta_{1,2} > 0$ , variable appartenant à un intervalle borné, tirée selon la procédure décrite ci-dessus.

## 3 Le modèle phénologique

Les séries temporelles d'images satellites optiques permettent le calcul d'indices de végétations. Ces indices permettent d'estimer des métriques caractérisant les stades phénologiques de la végétation, par inversion de modèles physiques qui les simulent.

### 3.1 L'indice NDVI

L'indice NDVI (*Normalized Difference Vegetation Index*), qui reflète l'activité photosynthétique au sol, est une mesure permettant d'identifier les phases de la phénologie végétale. Cet indice est exprimé à partir des réflectances dans le rouge  $\rho_R$  et le proche infrarouge  $\rho_{NIR}$  issues d'images multi-spectrales : 
$$\text{NDVI} = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \in [-1, 1].$$

3. Selon la procédure de *reparamétrisation*.

4. Negative Log Likelihood

5. L'erreur de reconstruction peut être due à des bruits aléatoires de la donnée d'entrée (*incertitude aléatoire*), ou à l'incapacité du modèle à décrire la donnée (*incertitude épistémique*).

### 3.2 Le modèle paramétrique double sigmoïde

L'évolution temporelle de l'indice NDVI de certaines végétations et cultures agricoles a une forme classiquement modélisable par une fonction double-sigmoïde [9] :

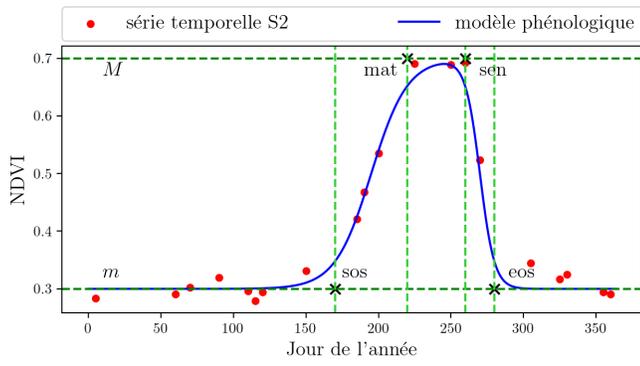
$$\Omega_\phi(t) = (M - m) (S_1(t) - S_2(t)) + m \quad (2)$$

avec les fonctions sigmoïdes :

$$S_1(t) = \left( 1 + \exp \left( 2 \frac{\text{sos} + \text{mat} - 2t}{\text{mat} - \text{sos}} \right) \right)^{-1} \text{ et}$$

$$S_2(t) = \left( 1 + \exp \left( 2 \frac{\text{sen} + \text{eos} - 2t}{\text{eos} - \text{sen}} \right) \right)^{-1}.$$

Il s'agit d'un modèle paramétrique dépendant d'un jeu  $\phi$  de 6 variables phénologiques, décrites en figure 2.



Variable	Définition
$M$	maximum de la double sigmoïde
$m$	minimum de la double sigmoïde
sos	date de début de croissance de la double sigmoïde (Start Of Season)
mat	date de fin de croissance de la double sigmoïde (Maturity)
sen	date de début de décroissance de la double sigmoïde (Senescence)
eos	date de fin de décroissance (End Of Season)

FIGURE 2 – Modèle phénologique double-sigmoïde

Les variables phénologiques sont ordonnées et bornées, nous intégrons ces contraintes grâce aux méthodes décrites en 2.3.

Le processus d'encodage de séries temporelles de NDVI en variables phénologiques et de reconstruction est illustré en figure 3.

## 4 Résultats expérimentaux

### 4.1 Données

Nous disposons d'un jeu de données de séries temporelles d'indices spectraux issues d'imagerie S2 acquise sur une année, avec leur classe d'occupation du sol, et à partir desquelles on calcule les séries temporelles de NDVI. La constellation S2 visite chaque scène une fois tous les 5 jours<sup>6</sup>. Mais en raison de la couverture nuageuse, chaque pixel de la série

6. Tous les 2-3 jours pour les zones à l'intersection des fauchées des deux satellites.

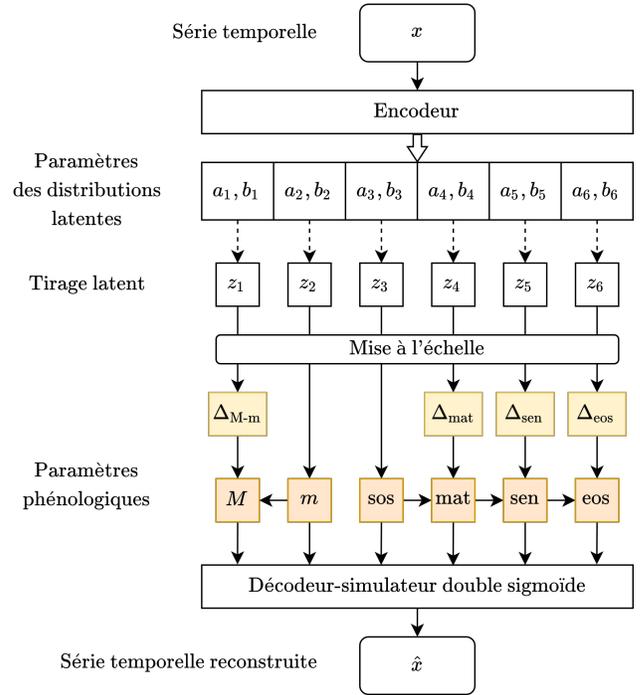


FIGURE 3 – Encodage d'une série temporelle avec un AEDS probabiliste à distributions de Kumaraswamy et inférence de variables phénologiques

d'image comporte en moyenne 25 acquisitions. Les séries temporelles des pixels sur une année sont donc linéairement interpolées et ré-échantillonnées tous les 5 jours

Ce jeu de données contient des classes de cultures agricoles, pour lesquelles le modèle phénologique est approprié, mais aussi des classes pour lesquelles il n'est pas adapté.

Pour calculer diverses métriques de validation, nous générons un jeu de données de séries temporelles synthétiques, en utilisant le modèle phénologique. Nous bruitons les séries temporelles obtenues, en choisissant aléatoirement 25 dates parmi une grille temporelle échantillonnée à 5 jours, masquons les autres, et appliquons un bruit gaussien de variance aléatoirement choisie entre 0 et 0.1. Ce jeu de données permet aussi de réaliser des apprentissages supervisés des variables phénologiques (régression, voir Table 1), que nous comparons à notre méthode.

### 4.2 Configuration expérimentale

Nous réalisons plusieurs expériences d'inférence de variables phénologiques à partir de séries temporelles de NDVI par plusieurs méthodes, détaillées en Table 1.

Les réseaux de neurones utilisés dans les encodeurs des AEDS probabilistes et pour la régression supervisée<sup>7</sup> ont la même architecture, un perceptron à 4 couches cachées (à 400, 500, 300 et 100 neurones).

7. régression supervisée de distributions de Kumaraswamy, avec comme fonction objectif la NLL de Kumaraswamy sur les paramètres inférés

TABLE 1 – Liste des expériences d’estimation de paramètres phénologiques à partir de séries temporelles de NDVI

Exp.	Modèle d’inférence	Jeu de données d’entraînement	Méthode supervisée	Estimateur ponctuel
1	MCMC [5]	-	Non	Médiane
2	Algorithme à Région de Confiance [4]	-	Non	-
3	Régression	Simulé	Oui	Mode
4	AEDS probabiliste	Simulé	Non	Mode
5	AEDS probabiliste	Réel	Non	Mode

### 4.3 Performances d’inférence

Nous évaluons chaque méthode sur le jeu de données simulées, l’erreur moyenne absolue (Table 2) par rapport au mode des distributions inférées, ainsi que le taux d’appartenance à l’intervalle 5–95 centiles de la véritable valeur de chaque variable phénologique à l’intervalle inféré (Table 3).

TABLE 2 – Erreur absolue moyenne d’estimation ponctuelle

Variable Phénologique	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5
$M$	<b>0.05</b>	0,24	<b>0,05</b>	0,14	0,11
$m$	<b>0.02</b>	<b>0,02</b>	<b>0,02</b>	0,04	0,05
sos	8.79	12,05	<b>8,68</b>	12,69	15,60
mat	10.72	20,14	<b>9,29</b>	25,74	20,49
sen	<b>11.70</b>	21,16	13,35	20,48	20,23
eos	<b>11.71</b>	13,05	13,36	14,27	17,55

TABLE 3 – Taux d’appartenance à l’intervalle 5–95 centiles (en gras : les taux les plus proches de 90%)

Variable Phénologique	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5
$M$	86.32	33,58	98,75	<b>92,13</b>	92,87
$m$	<b>83.44</b>	2,81	99,98	99,99	99,97
sos	86.53	86,45	95,41	<b>88,37</b>	83,18
mat	87.51	<b>87,71</b>	98,90	69,02	79,32
sen	<b>85.84</b>	85,40	99,08	51,51	56,51
eos	84.86	84,39	99,34	<b>89,71</b>	85,30

Les résultats montrent que les erreurs d’estimation ponctuelles des AEDS probabilistes sont moins bonnes que celle du MCMC et de la régression supervisée.

Pour l’estimation des intervalles de confiance à 90% le MCMC offre les performances les plus consistantes. L’estimation par algorithme à région de confiance n’est pas fiable pour chaque variable, et ne converge pas systématiquement par ailleurs. La régression supervisée surestime systématiquement les intervalles de confiance, qui ne sont pas informatifs. Ceux produits par les AEDS probabilistes sont plus fiables que les cas 2 et 3, mais ont une qualité qui dépend de la variable considérée. Les AEDS ont des performances similaires, et le modèle entraîné sur les données réelles est globalement un peu plus fiable, nonobstant la présence d’échantillons ne correspondant pas au modèle phénologique dans le jeu d’entraînement.

## 5 Conclusion

Nous avons développé une méthode permettant l’inférence de représentations interprétables et probabilistes, et l’avons évaluée sur le modèle phénologique de séries temporelles. Si les performances d’inférence n’atteignent pas encore le niveau d’autres méthodes plus répandues — le choix d’un encodeur simpliste limitant les capacités d’inférence — notre méthode ne requiert pas de jeu de données étiqueté pour l’entraînement, et est extrêmement rapide à l’inférence. La méthode est flexible, s’adaptant à tout modèle différentiable, et peut employer d’autres distributions paramétriques.

## Références

- [1] Haleh AKRAMI et al. “Addressing Variance Shrinkage in Variational Autoencoders using Quantile Regression”. In : *CoRR* abs/2010.09042 (2020).
- [2] Miguel A. ARAGON-CALVO. “Self-supervised learning with physics-aware neural networks – I. Galaxy model fitting”. In : *Monthly Notices of the Royal Astronomical Society* 498 (2020), p. 3713-3719.
- [3] Fábio Mariano BAYER, Débora Missio BAYER et Guilherme PUMI. “Kumaraswamy autoregressive moving average models for double bounded environmental data”. In : *Journal of Hydrology* 555 (2017), p. 385-396. ISSN : 0022-1694.
- [4] Thomas F. COLEMAN et Yuying LI. “An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds”. In : *SIAM Journal on Optimization* 6.2 (1996), p. 418-445.
- [5] Xiaojie GAO, Josh M. GRAY et Brian J. REICH. “Long-term, medium spatial resolution annual land surface phenology with a Bayesian hierarchical model”. In : *Remote Sensing of Environment* 261 (2021), p. 112484.
- [6] Diederik P KINGMA et Max WELING. *Auto-Encoding Variational Bayes*. 2014. arXiv : 1312 . 6114 [stat.ML].
- [7] Naoya TAKEISHI et Alexandros KALOUSIS. “Physics-Integrated Variational Autoencoders for Robust and Interpretable Generative Modeling”. In : *CoRR* (2021).
- [8] Michael TSCHANNEN, Olivier BACHEM et Mario LUCIC. “Recent Advances in Autoencoder-Based Representation Learning”. In : *ArXiv* abs/1812.05069 (2018).
- [9] Xi YANG et al. “Regional-scale phenology modeling based on meteorological records and remote sensing observations”. In : *Journal of Geophysical Research* 117 (2012).
- [10] Xiaoyang ZHANG et al. “Monitoring vegetation phenology using MODIS”. In : *Remote Sensing of Environment* 84.3 (2003), p. 471-475.