



HAL
open science

Learning Sentinel-2 Spectral Dynamics for Long-run Predictions using Residual Neural Networks

Joaquim Estopinan, Guillaume Tochon, Lucas Drumetz

► **To cite this version:**

Joaquim Estopinan, Guillaume Tochon, Lucas Drumetz. Learning Sentinel-2 Spectral Dynamics for Long-run Predictions using Residual Neural Networks. EUSIPCO 2021: 29th European Signal Processing Conference, Aug 2021, Dublin (virtual), Ireland. 10.23919/EUSIPCO54536.2021.9616304 . hal-03265010

HAL Id: hal-03265010

<https://imt-atlantique.hal.science/hal-03265010>

Submitted on 18 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning Sentinel-2 Spectral Dynamics for Long-run Predictions using Residual Neural Networks

Joaquim Estopinan
EPITA Research and Development
Laboratory (LRDE)
Le Kremlin-Bicêtre, France
joaquim.estopinan@inria.fr

Guillaume Tochon
EPITA Research and Development
Laboratory (LRDE)
Le Kremlin-Bicêtre, France
guillaume.tochon@lrde.epita.fr

Lucas Drumetz
Lab-STICC, UMR CNRS 6285
IMT Atlantique
Brest, France
lucas.drumetz@imt-atlantique.fr

Abstract—Making the most of multispectral image time-series is a promising but still relatively under-explored research direction because of the complexity of jointly analyzing spatial, spectral and temporal information. Capturing and characterizing temporal dynamics is one of the important and challenging issues. Our new method paves the way to capture real data dynamics and should eventually benefit applications like unmixing or classification. Dealing with time-series dynamics classically requires the knowledge of a dynamical model and an observation model. The former may be incorrect or computationally hard to handle, thus motivating data-driven strategies aiming at learning dynamics directly from data. In this paper, we adapt neural network architectures to learn periodic dynamics of both simulated and real multispectral time-series. We emphasize the necessity of choosing the right state variable to capture periodic dynamics and show that our models can reproduce the average seasonal dynamics of vegetation using only one year of training data.

Index Terms—remote sensing, multispectral images, time-series, spectral dynamics, recurrent neural networks

I. INTRODUCTION

The Sentinel-2 mission is part of the European Earth observation project Copernicus, a program jointly led by the European Space Agency (ESA) and the European Commission. It intends to provide authorities and interested actors with open multispectral data reflecting Earth surface changes, with proper natural resources management as end goal. The mission consists in two polar-orbiting satellites synchronized on the same sun orbit, and diametrically opposite to one another. As each Sentinel-2 satellite has a temporal revisit of ten days, the common temporal revisit of a given location on earth under the same viewing conditions goes down to five days [1].

The Sentinel-2 mission (or the soon to be launched pending hyperspectral EnMAP mission [2]) gives access to spectral time-series with a high temporal revisit. Such multidimensional data require a particular care to extract their rich information, but proved to greatly increase carried out task results when processed successfully [3]–[5]. The automatic learning of spectral dynamics would provide a new spectral

insight on the observed areas. It could benefit a large panel of applications such as spectral unmixing, anomaly detection, data assimilation [6] and predictions on vulnerable species conservation status [7]. It was proven in [8] that knowing spectral dynamics improves an endmember estimation task on synthetic data.

In the observed scenes, pure materials of interest are varying through time in their spectral signature or/and their spatial extent because of seasonal changes for instance. Reference [9] showed that a state-space model is convenient to model these evolutions. For a given material described by a n -dimensional state variable \mathbf{X}_t , the *state equation* of the model is given by an ordinary differential equation (ODE) when only temporal variations of the spectra are considered:

$$\frac{d\mathbf{X}_t}{dt} = \mathcal{F}(\mathbf{X}_t) + \eta_t \quad (1)$$

where \mathcal{F} is the dynamical operator and η_t an error accounting for both noise and model approximation. Our objective is thereby to directly learn from data these dynamics with specific neural network (NN) architectures [8].

The contribution of this article is threefold: (i) a particular state variable describing spectral signatures is shown efficient to learn periodic dynamics with residual neural networks (ResNets), (ii) long-run predictions on synthetic data show that the developed architecture is clearly more adapted for such task than a long short-term memory (LSTM) [10] architecture, (iii) first results on real data demonstrate the potential of our method for applications like classification or unmixing.

In the following, Section II focuses on methods and NNs to learn spectral dynamics. Then, Sections III and IV introduce conducted experiments and results on synthetic and real Sentinel-2 data, respectively. Finally, Section V gathers our conclusions on this work and its implications.

II. DYNAMICS LEARNING

A. ResNet architecture

Recurrent neural networks (RNNs) architectures, such as LSTM networks [10], are natural candidates when it comes

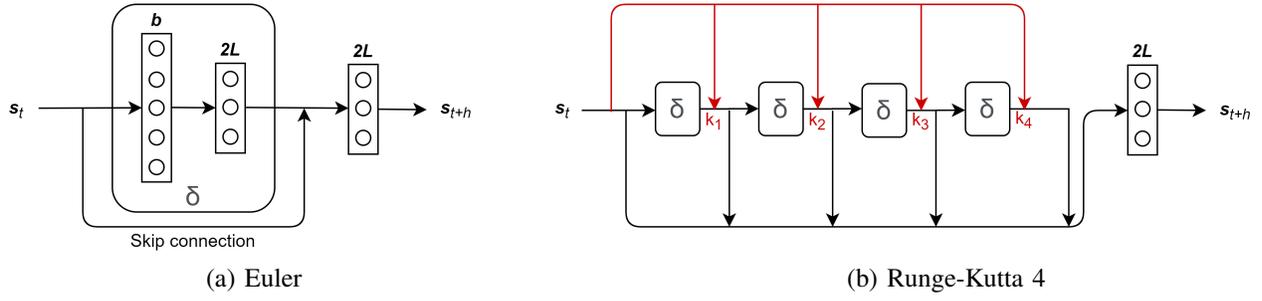


Fig. 1. ResNets architectures implementing (a) Euler and (b) Runge-Kutta integration schemes.

to learning and processing time-series data. Among RNNs architectures, ResNets with shared weights across layers have recently been proven to be particularly adapted to estimate dynamical operators [11]. Using them allows to reformulate the problem of learning an input/output relationship to the learning of a deviation from the identity [11], [12], making them suited to process data generated by differential equations. Mathematically, we assume that spectral signatures can be obtained by the following equation inherited from the integration of the ODE (1):

$$\mathbf{s}_{t+h} = \mathbf{s}_t + h\delta(\mathbf{s}_t), \quad (2)$$

\mathbf{s}_t being the pixel L -dimensional reflectance at time-step (t), L is the number of exploited spectral bands, δ is the infinitesimal operator of the integrated ODE and h is the integration step (hereafter set to 1 without loss of generality). In the ResNet typical architecture, a hard-cabled *skip connection* constrains the learning of the difference between \mathbf{s}_{t+h} and \mathbf{s}_t , see Fig. 1 (a). The problem is thus shifted to the learning of the infinitesimal operator δ . To achieve this, two numerical integration procedures are classically considered: the first-order Euler method and the more refined fourth-order Runge-Kutta 4 (RK4) method. These numerical integration schemes are hard-coded in the architectures. The ResNets are then learning the ODE through the weights attributed to the infinitesimal operator δ .

B. Euler and Runge-Kutta 4 integration schemes

The classical first-order Euler method consists in approximating δ by the temporal derivative of the targeted function:

$$\mathbf{s}_{t+1} = \mathbf{s}_t + \mathbf{s}'_t \quad (3)$$

An approximation of \mathbf{s}'_t is therefore learned through the shared weights of δ , see Fig. 1 (a) (the rationale for the dimension $2L$ is further explained in Section II-C). The two layers encoding δ are distributed through the whole time-series. The local error, *i.e.* the error per step, is on the order $O(h^2)$ [13].

The architecture of the RK4 integration scheme is displayed by Fig. 1 (b). Here, the approximation of δ is given by a weighted average of four slopes taken at the beginning, the middle and the end of the integration step (their respective weighting coefficients being k_1, k_2, k_3 and k_4), and the four

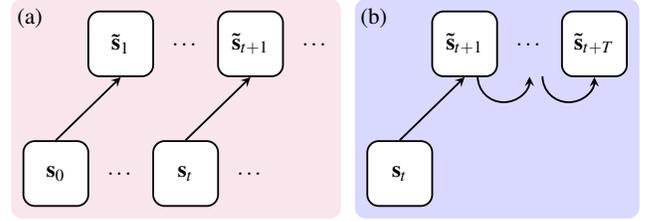


Fig. 2. (a) Training on short-term predictions and (b) test on long-term predictions.

δ blocks are put in series. δ is still learned by the weights of two consecutive fully connected layers. b , the number of connections between them, is of particular interest as it regulates the complexity given to the encoding of δ , and thus its expressiveness degree. Finally, the ResNets have to be provided with enough input information to accurately learn the dynamical operator's nature.

C. State augmentation with derivatives

By modifying the initial state of the input data, *i.e.* more concretely adding or removing features to the n -dimensional vectors in entry, the RNNs can lead to very different predictions. Initially, L -dimensional reflectance vectors \mathbf{s}_t were used as state variable. However, with this configuration, the tested RNNs were not able to capture the periodicity of simulated spectral dynamics. This was because the complete state of the 2^{nd} order system was never observed as the derivatives were missing [14]. Thus, the state variable was augmented with reflectance left derivatives \mathbf{s}'_t :

$$\mathbf{X}_t \leftarrow \begin{bmatrix} \mathbf{s}_t \\ \mathbf{s}'_t \end{bmatrix} \in \mathbb{R}^{2L}, \quad (4)$$

It explains the layers $2L$ output dimensions in Fig. 1. Choosing the right state variable allowed to learn dynamics periodicity on simulated data during the training, see Section III. If necessary, one can go one step further by adding higher-order derivatives to the state variable.

D. Training and testing strategies

RNNs are trained on successive one time-step prediction tasks, as depicted by Fig. 2 (a). To predict $\tilde{\mathbf{s}}_t$, the RNNs are fed with the ground truth \mathbf{s}_{t-1} at the previous date.

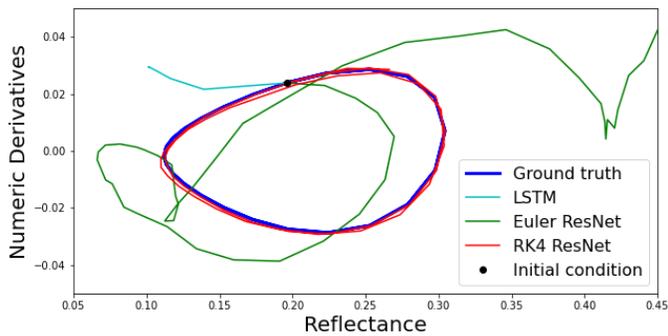


Fig. 3. Phase diagrams of LSTM, Euler ResNet and RK4 ResNet compared to the ground-truth trajectory. Represented long-run predictions are of 50 time-steps, 100^{th} spectral band.

Repeated along the training set, this process eventually enables to reach an interesting estimation of δ . The weights of the two fully connected layers encoding the infinitesimal operator are updated for each short-term prediction done in the training set. Such one time-step predictions on reflectance values, introduced in [8], naturally flow from the the ResNets architecture implementing an integration scheme according to (2).

Now, the objective is to predict the reflectance of a pure pixel at time $(t + T)$ knowing its trajectory until date (t) only. To do so, successive one time-step predictions are done from date (t) up to date $(t + T)$. While the first prediction is indeed done from known data at time-step (t) , the RNN input is then fed with the prediction at the previous time-step until the date $(t + T)$ is reached, as shown by Fig. 2 (b). It is a significantly harder task than the short-term predictions done during training since there is a higher chance that the overall error accumulates through the successive predictions, resulting in a trajectory drifting away from its expected behavior. Nevertheless, long-term predictions are useful to forecast dynamics ahead of time, and can be used to check that the learned dynamical systems has desirable topological properties (e.g. stability, convergence or even chaotic properties depending on the dynamics of the data [14]).

III. VALIDATION ON SYNTHETIC DATA

Simulated data were used to develop and test the different recurrent architectures [8]. Time-series were constructed with hyperspectral ($L = 144$) albedos derived from the DFC 2013 Houston dataset [15], a daily-periodic cosine varying illumination angle and simplified Hapke model from [16]. We obtained a toy dataset following nonlinear dynamics. To learn these synthetic dynamics, the training was typically done on a twenty-day windows with a thirty-minute sample-step, 2500 epochs and *Adam* optimizer. Fig. 3 gathers phase diagrams of long-run prediction from LSTM, Euler ResNet and RK4 ResNet strategies compared to the ground truth (100^{th} spectral band of the grass spectral signature). RK4 ResNet clearly appears as the only network able to provide correct *long-run* predictions (all architectures perform fine for 1 step-ahead predictions). LSTM and Euler ResNet do not manage to learn the

TABLE I
RMSE AND SAE FOR LSTM AND RESNETS LONG-RUN PREDICTIONS

t +	RMSE $\times 10^{-3}$			SAE $\times 10^{-3}$		
	LSTM	Euler	RK 4	LSTM	Euler	RK 4
12	27.7	31.4	0.45	4.30	18.0	0.01
24	48.5	32.9	3.01	4.30	13.0	0.01
36	27.5	92.5	2.33	4.30	10.9	0.03
120	49.8	2367	33.6	4.30	23.6	0.67

slightly nonlinear daily-periodic variations, with predictions respectively collapsing and diverging after some oscillations. Table I presents the root-mean-square errors (RMSE) and spectral angle errors (SAE) for $(t + T)$ long-run predictions. RK4 ResNet predictions result in very low errors for first 120 time-steps (2.5 days). Then errors eventually begin to grow, which is reasonable given the iterative construction of long-run predictions (see Fig. 2 (b)). Finally, it is worth mentioning that the synthetic dynamics are here the same for all spectral band, though with different albedo inputs for each band. However, this strong modeling hypothesis is likely not to be adapted to real data.

IV. EXPERIMENTS AND RESULTS ON REAL DATA

In this Section, we apply the developed spectral dynamic learning methodology on real Sentinel-2 time-series data.

A. Data description

Images produced by Sentinel-2 satellite imaging sensors are composed of 4 bands with 10 m spatial resolution (covering visible and near infrared domains), 6 bands with 20 m spatial resolution (in near-infrared and short-wave infrared domains) and 3 bands with 60 m spatial resolution. Here, we only retained 10 m and 20 m bands ($L = 10$), and upsampled those latter bands with bi-cubic interpolation to reach a common 10 m spatial resolution. The chosen area, located in South Puerto Rico, is the same as the one presented in [17], for which auxiliary data (such as rain records) have been collected at the close *Ensenada* station. The overall time-series runs from December 19th 2018 to July 31st 2020, resulting in a total of 118 individual images. Cloudy (thus unexploitable) images were replaced by Cressman interpolations [18] conducted thanks to (timewise) nearby available data, with a Gaussian radius of interpolation set to $R = 10$ days. Finally, the complete time-series was globally normalized, *i.e.* divided by the maximum reflectance value among all pixels, dates and spectral bands (the difference between 10 m and 20 m cells sensibility was also taken into account). RGB images extracted from the series are displayed in Fig. 5. We trained our model with forest pixels localized within the red rectangle. Forest indeed has a quite high intra-class spectral variability but the proposed approach is expected to learn an average spectral behavior. In the following, we only use the RK4 network since it proved to be the only one suited to long term predictions.

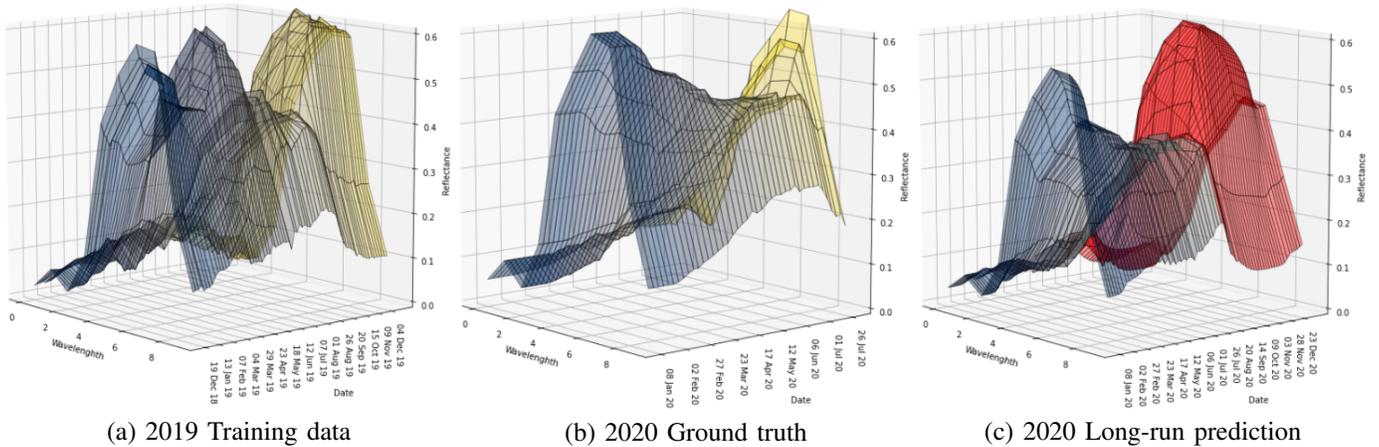


Fig. 4. Temporal evolution of a forest spectrum (a) in 2019 (whole year), (b) in 2020 (half year) and (c) 2020 long-run prediction (whole year) for the same pixel.

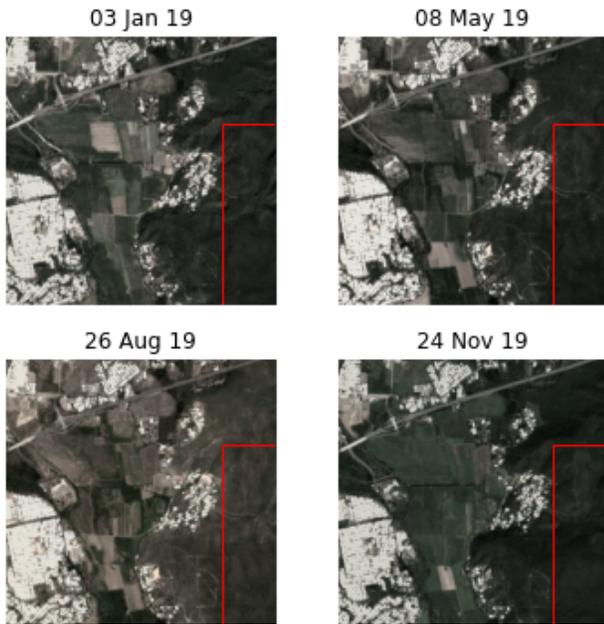


Fig. 5. Four RGB images extracted from the time-series.

B. Obtained results

The training was done on 74 images ranging from December 19th 2018 to December 19th 2019. Fig. 4 (a) displays the spectral evolution of a given forest pixel along those dates. Data included in the training set for a targeted pixel comprised its four direct neighbors, leading to a batch size equal to 5. 60000 epochs were needed with *Adam* optimizer and *MSE* loss function to reach convergence and realistic results. The value given to *b* turned out to be a major success criterion, the architecture being fixed otherwise. We found out that around 150 connections were optimal. Fig. 4 (b) represents the evolution of the same pixel as in Fig. 4 (a) between January 8th 2020 and July 31st 2020, serving as ground truth. Finally, Fig. 4 (c) represents the long-run prediction over

73 dates (from January 8th 2020 to January 12th 2021) for this pixel. The red portion accounts for predictions after the 31st of July 2020, to easily compare prior predictions with the ground truth. Predictions match the spectral decreasing tendency observable on the ground truth and peak prediction occurs slightly later than in reality (just after the 31st of July). RK4 ResNet nonetheless achieves to learn specific and periodic patterns. The quality of the predictions decreases with time, but remains realistic. RMSE on January 18th is of 5.0×10^{-2} and of 7.3×10^{-2} on April 17th. Because of the peak late prediction, it finally reaches 1.3×10^{-1} on July 26th.

V. CONCLUSION

In this work, we aimed at learning spectral dynamics patterns and periodicity directly from data, without prior knowledge. We showed that, using the right state variable, the developed ResNet implementing RK4 integration scheme succeeded in that task on simulated and real time-series multispectral data. Long-run predictions on synthetic data were proven significantly better than with LSTM or Euler ResNet, and realistic enough on a real multispectral time-series to capture seasonal variations in vegetation. Future research avenues include optimizing hyperparameters and adding an energy conservation constraint to further improve the results. A prior classification step on time-series to automatically integrate in training set spectrally consistent zones is also a promising direction, as is the stochastic modeling of the spectral variability of vegetation (prediction of a probability density instead of point estimates). This work could eventually benefit applications for space time interpolation of multispectral data, scene unmixing and forecasting problems among many others.

REFERENCES

- [1] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sensing of Environment*, vol. 120, pp. 25 – 36, 2012.
- [2] L. Guanter, H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner, C. Chlebek, et al., "The EnMAP spaceborne imaging spectroscopy mission for earth observation," *Remote Sensing*, vol. 7, no. 7, pp. 8830–8857, 2015.
- [3] W. He and N. Yokoya, "Multi-temporal Sentinel-1 and -2 data fusion for optical image simulation," *ISPRS International Journal of Geo-Information*, vol. 7, no. 10, pp. 389, 2018.
- [4] D. Tapete and F. Cigna, "Appraisal of opportunities and perspectives for the systematic condition assessment of heritage sites with Copernicus Sentinel-2 high-resolution multispectral imagery," *Remote Sensing*, vol. 10, no. 4, pp. 561–583, 2018.
- [5] N. Yokoya, X. X. Zhu, and A. Plaza, "Multisensor coupled spectral unmixing for time-series analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2842–2857, 2017.
- [6] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer Science & Business Media, 2009.
- [7] B. Cazorla, J. Cabello, J. Peñas, P. Garcillán, D. A. Reyes, and D. Alcaraz-Segura, "Incorporating ecosystem functional diversity into geographic conservation priorities using remotely sensed ecosystem functional types," *Ecosystems*, pp. 1–17, 2020.
- [8] L. Drumetz, M. Dalla Mura, G. Tochon, and R. Fablet, "Learning endmember dynamics in multitemporal hyperspectral data using a state-space model formulation," *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2483–2487, 2020.
- [9] S. Henrot, J. Chanussot, and C. Jutten, "Dynamical spectral unmixing of multitemporal hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3219–3232, 2016.
- [10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 1997.
- [11] R. Fablet, S. Ouala, and C. Herzet, "Bilinear residual neural network for the identification and forecasting of dynamical systems," *EUSIPCO 2018 : European Signal Processing Conference*, pp. 1–5, 2018.
- [12] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. 2018, vol. 31, pp. 6571–6583, Curran Associates, Inc.
- [13] E. Hairer, C. Lubich, and M. Roche, *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, vol. 1409, Springer, 2006.
- [14] S. Ouala, D. Nguyen, L. Drumetz, B. Chapron, A. Pascual, F. Collard, L. Gaultier, and R. Fablet, "Learning latent dynamics for partially-observed chaotic systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2020.
- [15] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, 2014.
- [16] L. Drumetz, J. Chanussot, and C. Jutten, "Spectral unmixing: A derivation of the extended linear mixing model from the Hapke model," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019.
- [17] M. A. Goenaga, M. C. Torres-Madronero, M. Velez-Reyes, S. J. Van Bloem, and J. D. China, "Unmixing analysis of a time series of hyperion images over the Guánica dry forest in Puerto Rico," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 329–338, 2013.
- [18] G. P. Cressman, "An operational objective analysis system," *Monthly Weather Review*, vol. 87, no. 10, pp. 367–374, 1959.