

---

# Learning Granger Causal Feature Representations

---

Gherardo Varando<sup>\* 1</sup> Miguel-Ángel Fernández-Torres<sup>\* 1</sup> Gustau Camps-Valls<sup>\* 1</sup>

## Abstract

Tackling climate change needs to understand the complex phenomena occurring on the Planet. Discovering teleconnection patterns is an essential part of the endeavor. Events like El Niño Southern Oscillation (ENSO) impact essential climate variables at large distances, and influence the underlying Earth system dynamics. However, their automatic identification from the wealth of observational data is still unresolved. Nonlinearities, nonstationarities and the (ab)use of correlation analyses hamper the discovery of true causal patterns. We here introduce a deep learning methodology that extracts nonlinear latent functions from spatio-temporal Earth data and that are Granger causal with the index altogether. We illustrate its use to study the impact of ENSO on vegetation, which allows for a more rigorous study of impacts on ecosystems globally.

## 1. Introduction

The Earth is a complex, dynamic and networked system. Land, atmosphere and climate interact constantly, and at different spatial and temporal scales (Dijkstra, 2013). The study of the intertwined connections between all spheres is a subject of active research. A paradigmatic case in climate sciences are teleconnection patterns; atmospheric changes in one place may largely impact other processes in distant regions. Phenomena like El Niño-Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO), change regimes and patterns of key essential variables like moisture, greenness and precipitation. Discovering spatio-temporal patterns is one of the main goals of the climate community, and the key to *analyze* and *understand* the underlying physical dynamics and processes driving the Earth system (Guttman, 1989; Larson, 2010).

Summarizing such relations have typically relied on extracting principal components from spatio-temporal data cubes

and analyzing their (linear) correlation with the ‘summarizing index’ (e.g. ENSO or NAO). A plethora of both linear and nonlinear dimensionality reduction have been used: extensions of Empirical Orthogonal Functions (EOF) (Bauer-Marschallinger et al., 2013; Volkov, 2014; Forootan et al., 2016) that seek for rotated (Lian & Chen, 2012), oblique (Hendrickson & White, 1964), and Varimax (Kaiser, 1958) transformations, as well as more powerful nonlinear methods based on neural networks (Hsieh et al., 2006), Gaussian processes (Lutinen & Ilin, 2009), low-rank tensor representations (Yu et al., 2015) and kernel methods (Bueso et al., 2020). Then, a simple lag-correlation analysis between the modes and the index of interest is performed. The approach is effective but two main -and critical- problems emerge: (1) the feature extraction is generally disconnected and learned independently from the index, which leads to a quite arbitrary selection of the modes of variability explaining the phenomena, and importantly (2) the (ab)use of correlation analysis typically leads to spurious detections, which do not tell anything about true causal effects. In this paper we introduce a methodology to address both problems jointly.

Learning representations is now best done with neural networks, like autoencoders, normalizing flows or unsupervised convolutional networks (Hinton & Salakhutdinov, 2006; Goodfellow et al., 2016; Romero et al., 2016; Reichstein et al., 2019; Camps-Valls et al., 2021). Neural networks cope well with highly structured (spatio-temporal) data, are computationally efficient, and can be engineered to minimize meaningful criteria. In our case, we aim to learn expressive representations that should be causal too. The goal can be thus framed within the recent field of *causal representation learning* (Schölkopf et al., 2021), which aims to discover high-level causal variables from low-level observations. Discovering teleconnections fits nicely in this framework as fundamental observable variables are often spatio-temporal, while high level relationships between aggregated variables or indices are believed to hold.

We propose learning Granger causal feature representations with neural networks. We introduce an autoencoder neural network (Hinton & Salakhutdinov, 2006) trained not only to optimize the reconstruction error but also the Granger causality criterion (Granger, 1980) between the latent representation and the index under study, e.g. ENSO or NAO. The architecture allows us to learn latent feature representations

---

<sup>\*</sup>Equal contribution <sup>1</sup>Image Processing Laboratory (IPL). Correspondence to: Gherardo Varando <gherardo.varando@uv.es>, Miguel-Ángel Fernández-Torres <miguel.a.fernandez@uv.es>.

that are Granger causal with the target index. We illustrate its use in assessing the impact of ENSO on vegetation.

## 2. Granger penalized autoencoder

A plethora of unsupervised neural nets are available for learning representations from spatio-temporal data. Autoencoders (Hinton & Salakhutdinov, 2006) aim at reproducing the inputs at the output layer by using the high-abstraction features learned in the intermediate layers. The use of autoencoders in remote sensing and geoscience data is widespread: from feature extraction and image classification (Zabalza et al., 2016; Othman et al., 2016), to spectral unmixing (Guo et al., 2015; Su et al., 2019), and the decomposition of spatio-temporal Earth data (Tibau et al., 2018).

While the standard autoencoder may retrieve latent features containing sufficient information for data reconstruction, performing post-extraction analysis on such latent representation could be challenging. For example, the obtained representation is not assured to be *causal*, and not even meaningful for causal discovery methods. To address this issue, we here propose to train an autoencoder with an additional loss term that enforces Granger causality between one of the elements of the latent representation and a target signal. The obtained latent representation is thus rewarded to be (at least partially) a causal effect of the target external signal.

### 2.1. Model overview

Let us define a general autoencoder as  $\xi = \psi \circ \phi$  composed of an encoding function  $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^D$  and a decoding function from the  $D$ -dimensional latent representation to the reconstructed space,  $\psi : \mathbb{R}^D \rightarrow \mathbb{R}^N$ , see Fig. 1. Now, given a sequence of input data points  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbb{R}^N$  sampled uniformly in time and a corresponding target signal  $y_1, \dots, y_T$  (e.g. ENSO), we consider the following optimization problem:

$$\begin{aligned} \{\phi^*, \psi^*\} = \arg \min & \left\{ \frac{1}{T} \sum_{t=1}^T \|\mathbf{x}_t - \psi(\phi(\mathbf{x}_t))\|_2^2 \right. \\ & \left. - \beta \text{GCI}_m(\phi_1(\mathbf{x}_{1:T}), y_{1:T} | \phi(\mathbf{x}_{1:T})) \right\} \end{aligned} \quad (1)$$

where  $\text{GCI}_m(\phi_1(\mathbf{x}_{1:T}), y_{1:T} | \phi(\mathbf{x}_{1:T}))$  is the Granger causality index between the target signal  $y_{1:T}$  and the first component of the latent representation  $\phi_1(\mathbf{x}_{1:T})$ . Specifically, for a fixed maximum number of lags  $L \in \mathbb{Z}^+$ , it is defined as

$$\text{GCI}_m(\phi_1(\mathbf{x}_{1:T}), y_{1:T} | \phi(\mathbf{x}_{1:T})) = \log \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} \right),$$

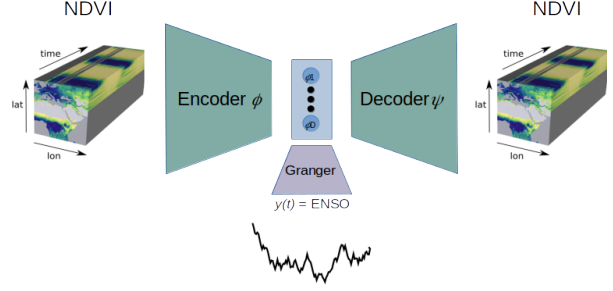


Figure 1. Learning causal representations with the neural Granger autoencoder. The architecture is trained to jointly minimize the reconstruction error of NDVI and the Granger causality index between  $\phi_1(\mathbf{x}(t))$  and ENSO  $y(t - \tau)$ ,  $\tau > 0$ .

where  $\hat{\sigma}_0^2$  and  $\hat{\sigma}_1^2$  are the estimated error variances of the following vector autoregressive models:

$$\begin{aligned} \phi_1(\mathbf{x}_t) &= \sum_{d=1}^D \sum_{l=1}^L b_{d,l} \phi_d(\mathbf{x}_{t-l}) + \varepsilon_0, \\ \phi_1(\mathbf{x}_t) &= \sum_{d=1}^D \sum_{l=1}^L b'_{d,l} \phi_d(\mathbf{x}_{t-l}) + \sum_{l=1}^L a_l y_{t-l} + \varepsilon_1. \end{aligned}$$

It should be noted that the autoencoder itself is time-agnostic but the Granger causality component of the optimized loss imposes time-dependent structure to the latent space. In addition, the optimized loss is equal to the mean squared error (MSE) for  $\beta = 0$ .

### 2.2. Explaining representations

An important shortcoming of learning nonlinear representations with autoencoders is their visualization and interpretation. Actually, interpretability has been identified as a potential weakness of deep neural networks, in particular for the geosciences (Lary et al., 2016; Gil et al., 2018; Blair et al., 2019; Reichstein et al., 2019). EXplainable AI (XAI) has emerged as an important field in machine learning (Samek & Müller, 2019). In this paper we visualize the spatio-temporal activations of the causal feature representation with a technique called *Neuron Integrated Gradients* (NIG) (Sundararajan et al., 2017; Shrikumar et al., 2018), which allows to attribute the latent representations  $\phi_d(\mathbf{x})$  at the output of the encoder  $\phi$  to each input spatial location  $x_n$  by integrating the gradient along the straight-line path in the input space from the baseline  $\tilde{x}_n$  to the input  $x_n$ , and then defines feature-wise scores:

$$R_d(x_n) = (x_n - \tilde{x}_n) \cdot \int_{\alpha=0}^1 [\nabla \phi_d(\tilde{\mathbf{x}} + \alpha \cdot (\mathbf{x} - \tilde{\mathbf{x}}))]_n d\alpha \quad (2)$$

Note that if  $\phi_d(\tilde{\mathbf{x}}) \approx 0$ , the scores fulfill  $\phi_d(\mathbf{x}) = \sum_n R_d(x_n)$ , so the attribution values deploy a complete explanation.

### 3. Experimental results

In order to assess our approach, we analyze the relationship between ENSO and vegetation greenness represented via the normalized difference vegetation index (NDVI) (Rouse et al., 1974). The index has been used a multitude of studies as proxy of vegetation status and health, as well as to quantify impacts of drought and plant stress.

#### 3.1. Data collection

The NDVI was directly computed from MODIS reflectance data derived from the MCD43A4.006 BRDF- Adjusted Reflectance 16-Day L3 Global 500m product (Schaaf & Wang, 2015; Schaaf et al., 2002), which are disseminated from the Land Processes Distributed Active Archive Center (LP DAAC) also available at Google Earth Engine (GEE). The MCD43A2 MODIS product, which contains ancillary quality information for the corresponding MCD43A4 product, was also used for avoiding low-quality BRDF estimates. We computed the NDVI at 16 d temporal and  $0.5^\circ$  spatial scales over 2007-2017 (11 years). Missing values in the computed NDVI were filled with linear interpolation and as customary we removed the yearly seasonality from NDVI by subtracting the means across the selected years.

We used time series of the ENSO34 climate index from the Royal Netherlands Meteorological Institute (KNMI), which is calculated daily based on Sea Surface Temperature (SST) anomalies averaged across the central equatorial Pacific Ocean (5N-5S, 170W-120W). ENSO34 time series were resampled to the weekly resolution by simple binning.

#### 3.2. Training and implementation details

The autoencoder is implemented as a dense feedforward neural network with encoder layers of sizes 256, 128, 64, 32 and 16 and symmetrical decoder layers, being  $D = 4$  the dimension chosen for the intermediate latent space and  $L = 20$  the lag order. We use Leaky ReLU (0.1) as the activation function for all hidden layers. The network is trained on a standard laptop (CPU 1.80 GHz  $\times$  8 cores and 16 GiB of RAM) during 500 epochs, using Adam optimizer with a learning rate of 0.001. The code has been developed in Python using PyTorch library, and is publicly available on GitHub<sup>1</sup>.

#### 3.3. Experimental setup

First, we consider the time period consisting of the years 2014-2017 to train the autoencoder described above. The selected study period includes the strong 2015-2016 El Niño event which has been shown to have an impact on global photosynthesis (Luo et al., 2018). Then, we train a model

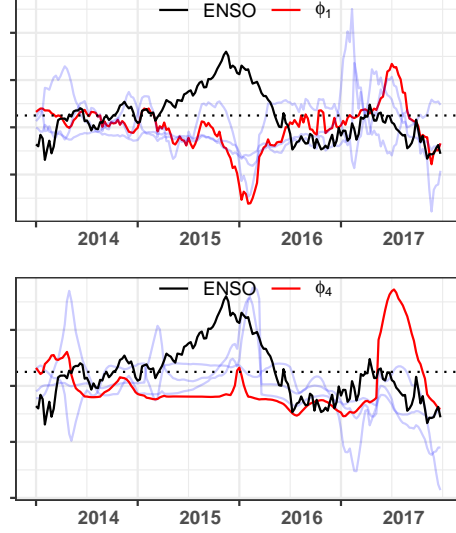


Figure 2. ENSO index and standardized latent features with and without the Granger penalization.

with  $\beta = 0.01$  and compare it with the same autoencoder architecture trained without the Granger causal loss ( $\beta = 0$ ). Models achieved comparable reconstruction errors for both choices of  $\beta$ .

We perform the standard Granger causality test (with Bonferroni correction for multiple testing) between ENSO and each component of the learned latent space  $\phi_d$ . As expected, and shown in Fig. 2, the first component ( $\phi_1$ ) of the model trained with  $\beta = 0.01$  (top) obtains a significant corrected  $p$ -value  $< 10^{-20}$  in the Granger test with respect to ENSO being a cause. The model trained with  $\beta = 0$  (bottom) also achieves one latent component ( $\phi_4$ ) for which the Granger test obtains a significant level of the test statistic but, in contrast, with a much higher  $p$ -value  $< 0.01$ .

#### 3.4. Causal footprints of ENSO on vegetation

The 2015/16 El Niño Event officially started in March 2015, long lived and was one of the strongest on the long term record. Several studies have reported severe impacts worldwide, from central America and Caribbean to south-east Asia, and mainly in (east and south-east) Africa, which was impacted by important vegetation degradation (Kalisa et al., 2019).

Figure 3 shows clear differences between the causal patterns of ENSO on vegetation before (2014) and during (2015) the event. Northwest African Sahara was critically impacted (Kogan & Guo, 2017) in 2015, which is clearly identified by the method. Ethiopia was hit by a severe drought on both the first and second growing seasons, and pastoral livelihoods largely affected, as previously reported in (Philippon et al., 2014) with simple linear correlation. Similarly, Eritrea was impacted with great reductions on ce-

<sup>1</sup><https://github.com/IPL-UV/LatentGranger>

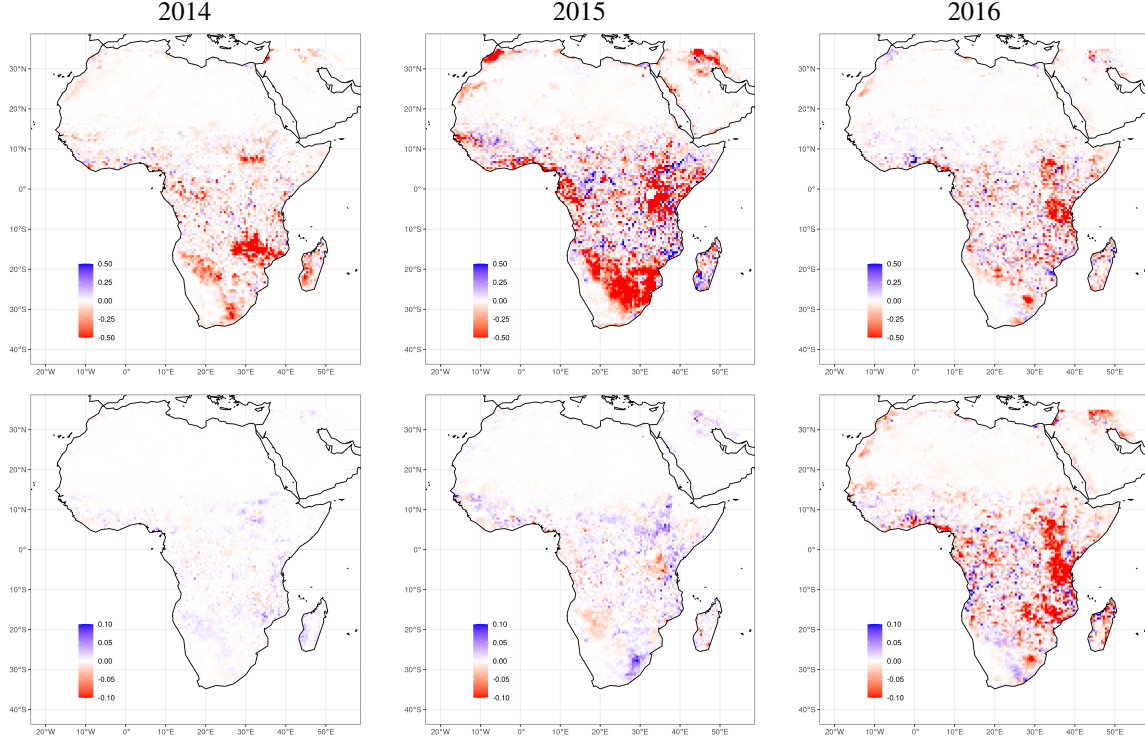


Figure 3. Neural integrated gradients for  $\beta = 0.01$  (top) and  $\beta = 0$  (bottom) during second to last observation of each year (columns).

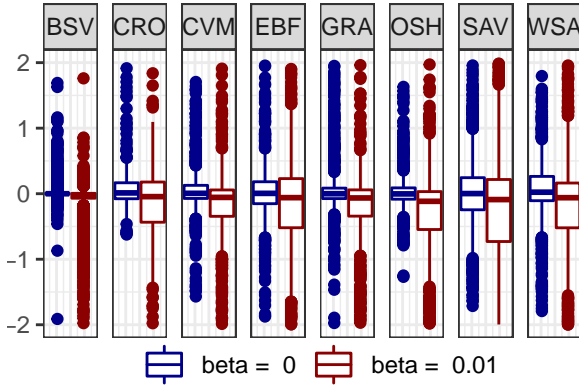


Figure 4. Box-plots for values of the integrated gradients maps per main Africa biomes in the year 2015. Values outside the range  $[-2, 2]$  are not shown for a better visualization.

real production in 2016. Another clear pattern stands out in Southern Africa, where El Niño led to drier than the average conditions, which impacted cereal production in subsequent years. In this region, vegetation photosynthetic activity tends to be dampened during the warm phases of ENSO and to increase during the cold ones; such negative gradients appear as early as December quite remarkably (note that for  $\beta = 0$  even a misleading faint positive activation appears).

These results are confirmed when looking at the NIG per biome in 2015, cf. Fig. 4, where more negative activations

(causal reduction in vegetation health status) are observed when the causal term is included,  $\beta \neq 0$ , especially significant for the most affected biomes; savannas and croplands as previously reported (Kogan & Guo, 2017).

#### 4. Conclusions and future work

We introduced a neural network methodology to learn Granger causal feature representations from spatio-temporal Earth data. We illustrated its performance in the study of the impact of ENSO on vegetation. Spatially explicit maps were derived via integrated gradients of the final model, which allowed us identifying well-known teleconnection patterns as well as unveiling other unreported patterns in the literature.

The proposed methodology is generic and modular, and allows many instantiations. One could think of replacing the Granger causality index by others, like the Geweke or conditional independence tests. While we used a fully connected dense network for simplicity, convolutional nets or other architectures could be used as well. In the future, we aim to study generalization and robustness of the causal representations. The methodology paves the way to gain insights on physical processes from Earth data, and leave mere (and potentially spurious) correlation patterns behind.

## References

- Bauer-Marschallinger, B., Dorigo, W. A., Wagner, W., and Dijk, A. I. J. M. V. How oceanic oscillation drives soil moisture variations over mainland australia: An analysis of 32 years of satellite observations. *Journal of Climate*, 26(24), 2013.
- Blair, G., Henrys, P., Leeson, A., Watkins, J., Eastoe, E., Jarvis, S., and Young, P. Data science of the natural environment: a research roadmap. *Frontiers in Environmental Science*, 7:121, 2019.
- Bueso, D., Piles, M., and Camps-Valls, G. Nonlinear PCA for Spatio-Temporal Analysis of Earth Observation Data. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8), 2020.
- Camps-Valls, G., Tuia, D., Zhu, X., and Reichstein, M. E. *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences*. Wiley & Sons, 2021. ISBN 978-1-119-64614-3.
- Dijkstra, H. A. *Nonlinear Climate Dynamics*. Cambridge University Press, 2013.
- Forootan, E., Khandu, Awange, J., Schumacher, M., Anyah, R., Dijk, A. V., and Kusche, J. Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over australia. *Remote Sensing of Environment*, 172, 2016.
- Gil, Y., Pierce, S. A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M., et al. Intelligent systems for geosciences: an essential research agenda. *Communications of the ACM*, 62(1):76–84, 2018.
- Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.
- Granger, C. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352, 1980. ISSN 0165-1889. doi: [https://doi.org/10.1016/0165-1889\(80\)90069-X](https://doi.org/10.1016/0165-1889(80)90069-X).
- Guo, R., Wang, W., and Qi, H. Hyperspectral image unmixing using autoencoder cascade. In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4. IEEE, 2015.
- Guttman, N. B. Statistical descriptors of climate. *Bulletin of the American Meteorological Society*, 70(6):602–607, 1989.
- Hendrickson, A. E. and White, P. O. Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17(1):65–70, 1964.
- Hinton, G. E. and Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006.
- Hsieh, W. W., Wu, A., and Shabbar, A. Nonlinear atmospheric teleconnections. *Geophysical research letters*, 33(7), 2006.
- Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.
- Kalisa, W., Igbawua, T., Henchiri, M., Ali, S., Zhang, S., Bai, Y., and Zhang, J. Assessment of climate impact on vegetation dynamics over east africa from 1982 to 2015. *Scientific reports*, 9(1):1–20, 2019.
- Kogan, F. and Guo, W. Strong 2015–2016 el niño and implication to global ecosystems from space data. *International Journal of Remote Sensing*, 38(1):161–178, 2017.
- Larson, J. W. Can we define climate using information theory? *IOP Conference Series: Earth and Environmental Science*, 11(1):012028, 2010.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10, 2016.
- Lian, T. and Chen, D. An evaluation of rotated EOF analysis and its application to tropical pacific SST variability. *Journal of Climate*, 25(15), 2012.
- Luo, X., Keenan, T. F., Fisher, J. B., Jiménez-Muñoz, J.-C., Chen, J. M., Jiang, C., Ju, W., Perakalapudi, N.-V., Ryu, Y., and Tadić, J. M. The impact of the 2015/2016 el niño on global photosynthesis using satellite remote sensing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1760):20170409, 2018. doi: 10.1098/rstb.2017.0409.
- Luttinen, J. and Ilin, A. Variational gaussian-process factor analysis for modeling spatio-temporal data. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A. (eds.), *Advances in Neural Information Processing Systems* 22, pp. 1177–1185. Curran Associates, Inc., 2009.
- Othman, E., Bazi, Y., Alajlan, N., Alhichri, H., and Melgani, F. Using convolutional features and a sparse autoencoder for land-use scene classification. *International Journal of Remote Sensing*, 37(10):2149–2167, 2016.
- Philippon, N., Martiny, N., Camberlin, P., Hoffman, M., and Gond, V. Timing and patterns of the enso signal in africa over the last 30 years: Insights from normalized difference vegetation index data. *Journal of Climate*, 27(7):2509–2532, 2014.



- Reichstein, M., Camps-Valls, G., Stevens, B., Denzler, J., Carvalhais, N., Jung, M., and Prabhat. Deep learning and process understanding for data-driven Earth System Science. *Nature*, 566:195–204, Feb 2019.
- Romero, A., Gatta, C., and Camps-Valls, G. Unsupervised deep feature extraction for remote sensing image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 54(3):1349–1362, 2016. ISSN 0196-2892. doi: <http://dx.doi.org/10.1109/TGRS.2015.2478379>.
- Rouse, J., Haas, R. H., Schell, J. A., Deering, D. W., et al. Monitoring vegetation systems in the great plains with ERTS. *NASA special publication*, 351(1974):309, 1974.
- Samek, W. and Müller, K.-R. Towards explainable artificial intelligence. In *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 5–22. Springer, 2019.
- Schaaf, C. and Wang, Z. MCD43A4 MODISTerra+Aqua BRDFAlbedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006, 2015. Data set.
- Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., et al. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote sensing of Environment*, 83(1-2):135–148, 2002.
- Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., and Bengio, Y. Toward causal representation learning. *Proceedings of the IEEE*, 109(5): 612–634, 2021. doi: 10.1109/JPROC.2021.3058954.
- Shrikumar, A., Su, J., and Kundaje, A. Computationally efficient measures of internal neuron importance. *arXiv preprint arXiv:1807.09946*, 2018.
- Su, Y., Li, J., Plaza, A., Marinoni, A., Gamba, P., and Chakravorty, S. Daen: Deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4309–4321, 2019.
- Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pp. 3319–3328. PMLR, 2017.
- Tibau, X.-A., Requena-Mesa, C., Reimers, C., Denzler, J., Eyring, V., Reichstein, M., and Runge, J. Supernovae: Vae based kernel pca for analysis of spatio-temporal earth data. 2018.
- Volkov, D. L. Do the north atlantic winds drive the nonseasonal variability of the arctic ocean sea level? *Geophysical Research Letters*, 41(6), 2014.
- Yu, R., Cheng, D., and Liu, Y. Accelerated online low-rank tensor learning for multivariate spatio-temporal streams. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, pp. 238–247, 2015.
- Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., Du, P., and Marshall, S. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing*, 185:1–10, 2016.