Advances in Machine Learning for Earth Observation

Gustau Camps-Valls
Image Processing Lab (IPL)
Universitat de València

http://isp.uv.es  @isp_uv_es  gustau.camps@uv.es
Earth observation
“Earth observation (EO) is the gathering of information about planet Earth’s physical, chemical and biological systems via remote sensing technologies supplemented by earth surveying techniques, encompassing the collection, analysis and presentation of data.”
Detecting life in plants...
Detecting life in plants ...
Detecting life in plants ...
Detecting life in plants ...
Detecting life in plants ...
Detecting life in plants ...
Earth observation applications

- Identify objects, classify the land cover and detecting changes
- Estimate the content of bio-geo-physical and bio-chemical parameters
Earth observation and friends
Earth observation meets machine learning
Earth observation - Machine learning

- One soil map: https://map.onesoil.ai
- Global wealth maps: http://penny.digitalglobe.com
- Global data refinery: http://www.descarteslabs.com
- Disease mapping: https://www.healthmap.org/en
- Water risk atlas: https://www.wri.org/applications/maps/aqueduct-atlas
- Flood analyzer: http://floods.wri.org
AI promises to transform scientific discovery ...

“Unlike earlier attempts ... [AI systems] can see patterns and spot anomalies in data sets far larger and messier than human beings can cope with.”

July 7 2017 Issue
... yet only when some things happen!

- Strong spatial and temporal correlations
- Big data accessible
- Cheap computing resources available
- Fast scalable ML models available
- No expert knowledge needed
- High prediction accuracy is enough
Challenges in Earth system science
Big data challenges

1. Data size now exceeds 100 petabytes, and is growing quasi-exponentially.
2. The speed of change exceeds 5 petabytes a year, and acquisition frequencies of 10 Hz or more;
3. Reprocessing and versioning are common challenges.
4. Data sources can be multi-dimensional, spatially integrated, from the organ level (such as leaves) to the global level.
5. Earth has diverse observational systems, from remote sensing to in situ observations.
6. The uncertainty of data can stem from observational errors or conceptual inconsistencies.
Statistical challenges

1. High dimensional data: multi-temporal, multi-angular and multi-source
2. Non-linear and non-Gaussian feature relations
3. Data misalignments and distortions
4. Irrelevant features and biased sampling strategies
5. Uneven sampling, skewed distributions and anomalies in the wild
6. Few supervised information is available
Philosophical challenges

- **Consistency issue:** ML models do not respect Physics
- **Learning issue:** ML are excellent approximators, yet no fundamental laws are learned
- **Interpretability issue:** Big data is good to estimate correlations, what about causations?
Outline

1. Advances in spatio-temporal data processing
   ◦ Classification
   ◦ Regression
   ◦ Dimensionality reduction
2. Big data in the Google cloud
3. ML models should be consistent with Physics
4. Understanding is more important than predicting
Spatio-temporal data classification
Neural networks for spatio-temporal classification

- Convolutional neural nets (CNN): hierarchical structure exploits spatial relations
- Long short-term memory (LSTM): recurrent network that accounts for memory/dynamics

“A Deep Network Approach to Multitemporal Cloud Detection”
Probabilistic and scalable classifiers

● Gaussian processes as an alternative to neural nets
● GPs allow a probabilistic treatment, confidence intervals, feature ranking, deep too!
● Gaussian processes start to be scalable ...

“Remote Sensing Image Classification With Large-Scale Variational Gaussian Processes,”
Multitask learning

- Multiple inter-related outputs? Data from multiple sources?
- Learn to fuse heterogeneous information
Anomalous change detection

● Pervasive and anomaly changes in the wild

May13, 14.0°  Aug13, 43.6°  Nov13, 29.3°

Regression, fitting, parameter retrieval
Spatio-temporal variable prediction

- STA is common place in climate informatics, neuroscience, video processing, NLP, ...
- Current approaches: CNN + LSTM, space-time Gaussian processes
- Novel approaches: distribution regression and variational deep GPs

\[ \mathbb{P} \mapsto \mu_k(\mathbb{P}) \mapsto \mathbb{P} \mapsto [\mathbb{E}\phi_1(X), \ldots, \mathbb{E}\phi_s(X)] \in \mathbb{R}^s \]

\[ \langle \mu_k(\mathbb{P}); \mu_k(\mathbb{Q}) \rangle_{\mathcal{H}_k} = \mathbb{E}_{X \sim \mathbb{P}, Y \sim \mathbb{Q}} k(X, Y) \]

"A Survey on Gaussian Processes for Earth Observation Data Analysis"  

"Deep Gaussian Processes for Retrieval of bio-geo-physical parameters",  

"Nonlinear Distribution Regression for Remote Sensing Applications"  
Adsuara, Perez,Muñoz, Mateo, Piles, Camps-Valls, IEEE TGARS 2019
Multioutput and transfer learning

- Multioutput regression: compactness & speed
- Transfer learning

Multioutput regression and gap filling

● Transfer learning across time, sensors and space

● LAI and FAPAR across time and space

● Soil moisture and sensor fusion

“Gap filling of biophysical parameters with multi-output GPs”

“Latent force GP models for EO time series prediction”
Luengo, Muñoz, Piles, Camps-Valls, IEEE IGARS, 2019
Dimensionality reduction and modes of variability
Sparse coding in unsupervised deep nets

- CNN trained to extract sparse features → features + linear classifier suffice!

\[ \sum_i T_{ij} = 1 \forall i \in [1...N] \]

\[ \sum_i T_{ij} = \frac{N_i}{N_s} \forall j \in [1...N_s] \]

Minimize \( E = \|H - T\|^2 \)

**PS:** for each sample only one output must be active

**LS:** even distribution of outputs, no dead outputs

“Unsupervised Deep Feature Extraction for Remote Sensing Image Classification”
Kernel multivariate data analysis

- Transform data to max var/corr/covar

- Transform data to max SNR

- Transform data to max information


Spatio-temporal analysis of the Earth cubes

- PCA/EOF is popular, yet cannot cope with nonlinear spatio-temporal relations
- ROCK PCA
  - copes with nonlinearities
  - extracts spatial and temporal components
  - very fast

“Rotated Complex Kernel PCA for spatio-temporal data decomposition”
Bueso, Piles, Camps-Valls, IEEE TGARS, 2018
Spatio-temporal analysis of the Earth cubes

- SM decomposition
  - Meaningful compression
  - Climate-specific modes of variability
  - Boreal and Equatorial modes of SM variability dominate
  - Seasonal and ENSO related temporal modes

“Rotated Complex Kernel PCA for spatio-temporal data decomposition”
Bueso, Piles, Camps-Valls, IEEE TGARS, 2018
Spatio-temporal analysis of the Earth cubes

- PC3 highly correlates with ENSO + new spatial patterns uncovered

- Nonlinear cross-correlation uncovers unreported SM-ENSO lags

“Rotated Complex Kernel PCA for spatio-temporal data decomposition”
Bueso, Piles, Camps-Valls, IEEE TGARS, 2018
Efficiency
Google Earth Engine (something that Europeans love)
Google Earth Engine: cloud detection in the cloud

- Exploit temporal information and change detection

“Multitemporal Cloud Masking in the Google Earth Engine”
Mateo, Gómez, Amorós, Muñoz, and Camps-Valls. Remote Sensing 7 (10) :1079, 2018

“Cloud masking and removal in remote sensing image time series”
Global maps of LAI, FAPAR, FVC, canopy water content by inverting PROSAIL with ML...
Google Earth Engine: spatialization of plant traits

- Global maps at 500 m resolution of specific leaf area, leaf dry matter content, leaf nitrogen and phosphorus content per dry mass, and leaf nitrogen/phosphorus ratio.

“A methodology to derive global maps of leaf traits using remote sensing and climate data”
Moreno, Camps-Valls, Kattge, Robinson, Reichstein, ... and Running.
Remote Sensing of Environment 218 (12):69-88, 2018
Physics-aware machine learning

\[ F(X, H(t)|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle ) = y \]
The truth is that...

“Models without data are fantasy. Data without models are chaos.”

Patrick Crill, Stockholm University, quoted in Science, 2014, in “Methane on the rise again”, vol 343, pp. 493-495
Physics-driven ML: constrained optimization

- ML that minimizes model violations and predictions are dependent of physical laws

\[
\text{PhysLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \| w \|_2^2 + \gamma \Omega(\hat{y}, \Phi)
\]

\[
\Omega(\hat{y}, \Phi) = \text{sum of physical violations of } \hat{y}
\]

\[
\text{FairLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \| w \|_2^2 + \gamma I(\hat{y}, s)
\]
Physics-driven ML: joint model-data ML

Let ML talk to physical models

\[
\text{JointLoss} = \text{Cost}(y, \hat{y}) + \lambda_1 \| w \|^2_2 + \gamma \Omega(\hat{y}, \Phi)
\]
\[
\Omega(\hat{y}, \Phi) = \text{Cost}_s(y_s, \hat{y}_s)
\]

**Setup**

- ERMES project: 3 rice sites, 85% European production
- Landsat 8 + in situ measurements + PROSAIL simulations
- In situ LAI measurements: \( r = 70-300 \) (3 countries, 2 years)
- Simulations: \( s = 2000 \) (Landsat 8 spectra and LAI)

**Filling the space ...**
Physics-driven ML: hybrid modeling framework

PERSPECTIVE

Deep learning and process understanding for data-driven Earth system science

Markus Reichstein¹,², Gustau Camps-Valls³, Bjorn Stevens⁴, Martin Jung¹, Joachim Denzler²,⁵, Nuno Carvalhais⁶,⁷ & Prabhat⁷

Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we argue that these contextual cues should be used as part of deep learning (an approach that is able to extract spatio-temporal features automatically) to gain further process understanding of Earth system science problems, improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales, for example. The next step will be a hybrid modelling approach, coupling physical process models with the versatility of data-driven machine learning.

Physics-driven ML: hybrid modeling framework

- ML that learns laws of physics (e.g. consistency model-data, convection, advection, mass and energy conservation)

A: “Physisizing” a deep learning architecture by adding one or several physical layers after the multilayer neural network

B: A motion field is learned with a convolutional-deconvolutional net, and the motion field is further processed with a physical model


Physics-driven ML: emulation of complex codes

- GP Emulation = Mathematical tractability + Global sensitivity analysis + Speed

**Figure**: Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis, Verrelst, Camps-Valls et al, Remote Sensing of Environment, 2016

**Figure**: Emulation as an accurate alternative to interpolation in sampling radiative transfer codes, Vicent and Camps-Valls, IEEE Journal Sel. Topics Rem. Sens, Appps. 2018
Physics-driven ML: encoding and learning ODE/PDEs

- Who needs Lorenz?

1. True Lorenz System
   \[
   \begin{align*}
   \dot{x} &= a(y - x) \\
   \dot{y} &= x(r - z) - y \\
   \dot{z} &= xy - bz.
   \end{align*}
   \]

2. Sparse Regression to Solve for Active Terms in the Dynamics

3. Identified System
   \[
   \begin{align*}
   \dot{x} &= \Theta(x^T) \chi_1 \\
   \dot{y} &= \Theta(x^T) \chi_2 \\
   \dot{z} &= \Theta(x^T) \chi_3.
   \end{align*}
   \]

"Discovering governing equations from data by sparse identification of nonlinear dynamical systems" Brunton, Proctor, Kutz, PNAS 2016

- Who needs Navier Stokes?

\[
\psi_t = -u \psi_x - v \psi_y + 0.01(\psi_{xx} + \psi_{yy})
\]


- Who needs Schrödinger?

\[
\psi_t = 0.5i \psi_{xx} + i|\psi|^2 \psi
\]

"Discovering governing equations from data by sparse identification of nonlinear dynamical systems" Brunton, Proctor, Kutz, PNAS 2016
Understanding is more important than fitting

\[ F(X, Y) = \text{Diagram} \]
Feature selection & ranking

- Filters & wrappers

Neuron and bases visualization

- What did the network learn?
- How do bases change in time, with real/simulations/together, under extremes?

“Processing of Extremely high resolution LiDAR and optical data”, Campos-Taberner, Camps-Valls et al, 2016
Graphical models and causal discovery

- Causality discovery learns cause and effects relations from data
- What for? Hypothesis testing, model-data comparison, causes of extreme impacts

A platform for causal discovery

- **CauseMe**: [http://causeme.uv.es](http://causeme.uv.es)
  - Download time series with ground truth
  - Run your causal discovery algorithm offline
  - Upload your causal graph
  - Get your results!

“Inferring causation from time series with perspectives in Earth system sciences”

“Causal Inference in Geoscience and Remote Sensing from Observational Data,”
Conclusions
Conclusions

- Machine learning in EO and climate
  - Many techniques ready to use
  - Huge community, exciting tools

- Solid mathematical framework to deal with
  - Multivariate data
  - Multisource data
  - Structured spatio-temporal relations
  - Nonlinear feature relations
  - Fitting and classification

- Risks & remedies
  - Understanding is more complex
  - Physics consistency a must
  - Physics-driven ML & Explainable AI
Thanks!

http://isp.uv.es
@isp_uv_es
gustau.camps@uv.es
Propaganda

---

ERC Consolidator Grant 2015-2020
Statistical Learning for Earth Observation Data Analysis
Gustau Camps-Valls, Universitat de València, Spain
✉️ gustau.camps@uv.es | 🌐 http://isp.uv.es

---

1 PhD + 1 PostDoc positions!

- Remote sensing and geosciences
- Machine learning
  - Regression, time series analysis
  - Graphical models
  - Causal inference

E-mail me: gustau.camps@uv.es