Classification of Landcover Using Satellite Image Time Series and Variational Gaussian Process

Traitement des Données Massives et Apprentissage: Applications en geophysique, écologie et SHS - Grenoble

---

Valentine Bellet, Mathieu Fauvel, Jordi Inglada

Mercredi 7 Juin 2023

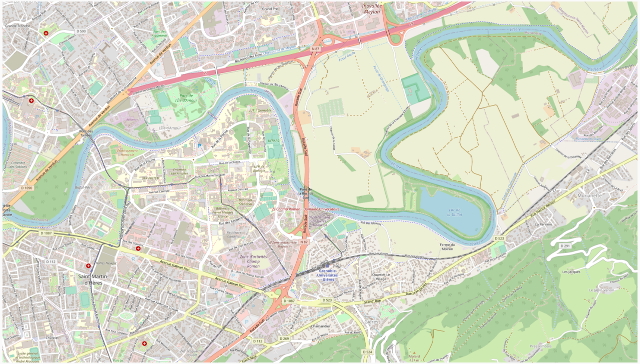
CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UPS, Toulouse, FRANCE

Outline

Introduction

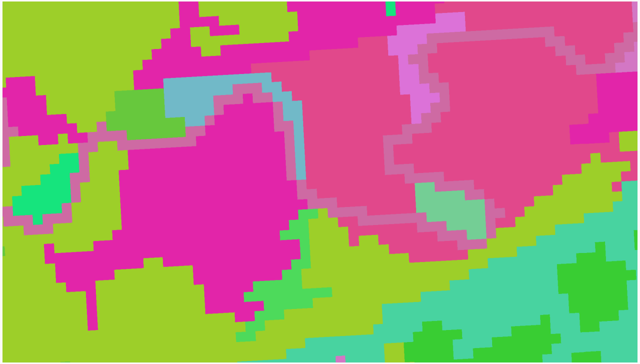
Land Cover Data

- Categorization of Earth's surface
- State and trend
- Used in many applications
  - ⋆ Geography
  - ⋆ Ecology
  - ⋆ Hydrology
- Remote sensing



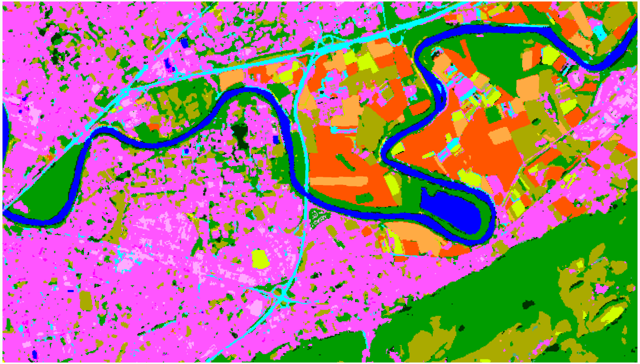
Land Cover Data

- Categorization of Earth's surface
- State and trend
- Used in many applications
  - ⋆ Geography
  - ⋆ Ecology
  - ⋆ Hydrology
- Remote sensing



Land Cover Data

- Categorization of Earth's surface
- State and trend
- Used in many applications
  - ⋆ Geography
  - ⋆ Ecology
  - ⋆ Hydrology
- Remote sensing



Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

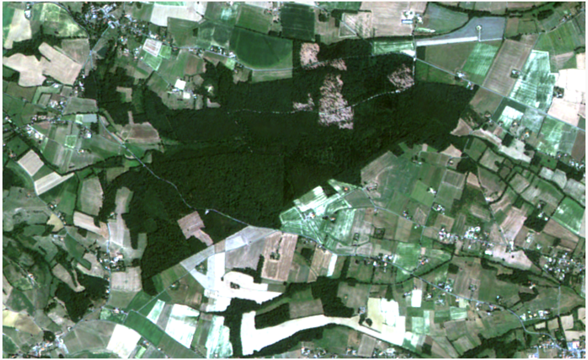
Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia



High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

High Spatial Resolution Satellite Image Time Series

Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

Sentinel 2

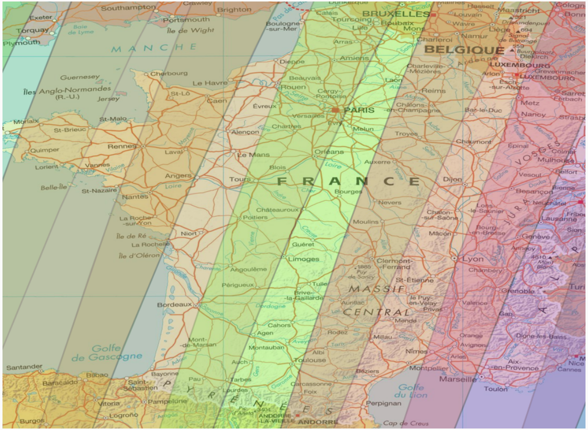
- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

Pôle Theia

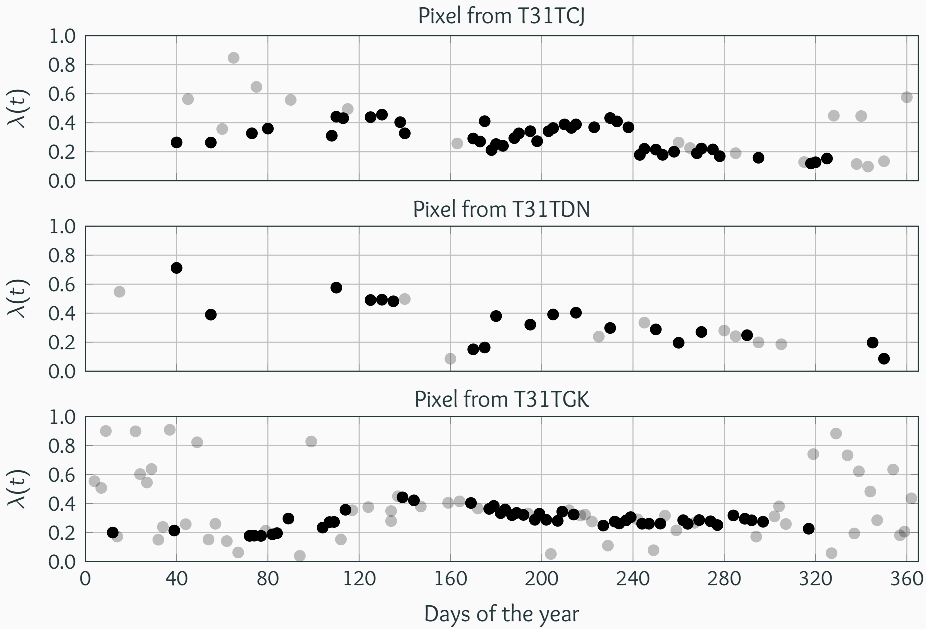
Sentinel 2

- Global coverage
- High spatial resolution (10m & 20m)
- 13 spectral bands
- Revisit every 5 days
- Free and open source data

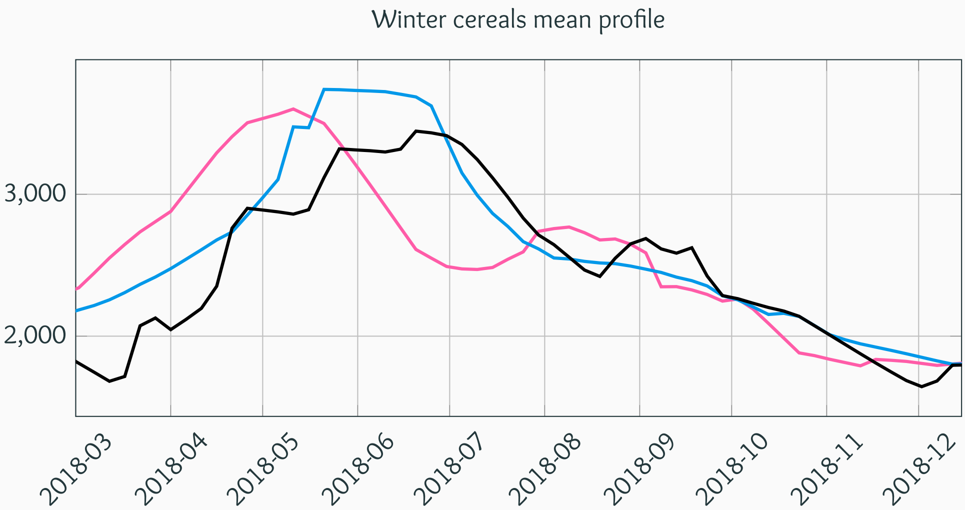
Pôle Theia



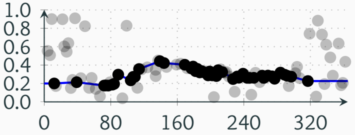
Irregularly sampled SITS



Spatial Variability



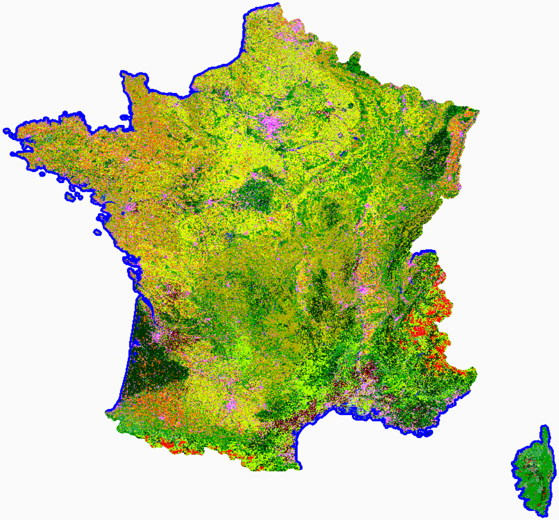
- Temporal resampling



- Input feature

$$\mathbf{x} = \big[\lambda_1(t_1), \ldots, \lambda_1(t_T), \ldots, \lambda_{10}(t_T), \\ \mathrm{si}_1(t_1), \ldots, \mathrm{si}_p(t_T) \ldots\big]$$

- Each sample is associated with class membership $Z \in \{1, \ldots, C\}$
- Classification algorithms + spatial stratification [Ing+17]



OSO products

Objectives & Contributions

Limitations of (some) current approaches

- Two-step approaches might not be optimal w.r.t. final objective:

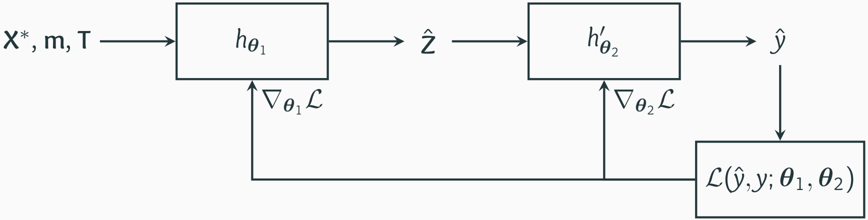
  Reconstruction error VS classification error

- Do not scale
- Do not exploit spectro-spatio-temporal structure of SITS

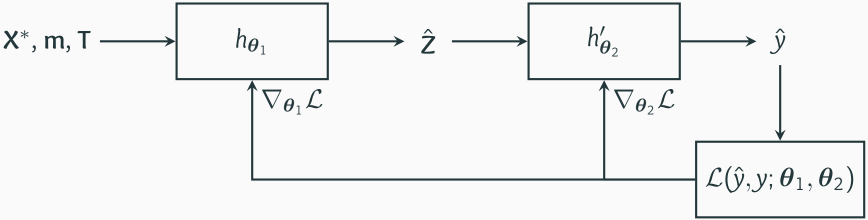
Contributions

- Classification of irregular and unaligned SITS
  - ⋆ Learn *latent smoother*
- Evaluation of Sparse Variational Gaussian process
  - ⋆ Covariance operator
  - ⋆ Large scale processing (variability & volume)

Workflow & Notations



Workflow & Notations



- $\mathcal{S} = \{\mathbf{x}^i, \mathbf{y}^i, \mathbf{T}^i\}_{i=1}^N, y \in \{0,1\}^C$
- $\mathbf{x}^i(t) \in \mathbb{R}^D$, observed at T^i times $\{t_1^i, \ldots, t_{T_j^i}^i\} = \mathbf{T}^i$
- $\mathbf{T} = \bigcup_{i=1}^n \mathbf{T}^i = \{t_1, \ldots, t_T\}$
- Mask time series $\mathbf{m}^i \in \{0,1\}^T$, such as $\mathbf{m}_j^i = 1$ if $t_j \in \mathbf{T}^i$ else 0, $\forall t_j \in \mathbf{T}$
- Augmented pixel time series $\mathbf{x}^{*i}$, such as $\mathbf{x}^{*i}(t_j) = \mathbf{x}^i(t_j)$ if $\mathbf{m}_j^i = 1$ else $\mathbf{0}$, $\forall t_j \in \mathbf{T}$
- Sample matrix

$$\mathbf{X}^{*i} = \left[\mathbf{x}^{*i}(t_1)| \ldots |\mathbf{x}^{*i}(t_T)\right] \in \mathbb{R}^{D \times T}$$

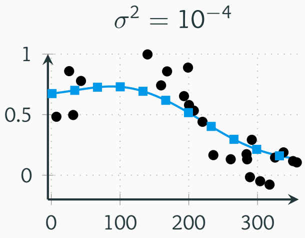
Latent Representation

This section is based on

V. Bellet, M. Fauvel, J. Inglada and J. Michel. End-to-end Learning for Land Cover Classification using Irregular and Unaligned SITS by Combining Attention-Based Interpolation with Sparse Variational Gaussian Processes. 2023. ⟨hal-04112115⟩ [Bel+23]

- **Problem:** find **Z**, regular and aligned on R times $\{r_1, \ldots, r_R\}$
- Nadaraya-Watson kernel smoother [HTF01, Chapter 6]:

$$\hat{x}_\ell(r) = \sum_{j=1}^{T} \frac{K(r, t_j) m_j}{\sum_{j'=1}^{T} K(r, t_{j'}) m_{j'}} x^*_\ell(t_j) \text{ with } K(r, t_j) = \exp\left\{-d(r, t_j)\right\} \text{ and } d(r, t_j) = \sigma^{-2}(r - t_j)^2$$



(Multi) Time Attention (mTAN)

- Use attention mechanism and temporal embedding [SM21]

$$d(r,t) = \frac{\phi(r)^\top \mathbf{W_q}^\top \mathbf{W_k} \phi(t)}{\sqrt{E}} \text{ with } \phi(t) = \begin{bmatrix} \omega_1 t + \alpha_1 \\ \sin(\omega_2 t + \alpha_2) \\ \vdots \\ \sin(\omega_E t + \alpha_E) \end{bmatrix}$$

- Time attention smoother

$$\begin{aligned} \hat{x}_\ell(r) &= \sum_{j=1}^{T} \frac{\exp\left(\frac{\phi(r)^\top \mathbf{W}_q^\top \mathbf{W}_k \phi(t_j)}{\sqrt{E}}\right) m_j}{\sum_{j'=1}^{T} \exp\left(\frac{\phi(r)^\top \mathbf{W}_q^\top \mathbf{W}_k \phi(t'_j)}{\sqrt{E}}\right) m_{j'}} x^*_\ell(t_j) \\ &= \text{softmax}\left\{\frac{(\Phi(\mathbf{T})^\top \mathbf{W}_k^\top \mathbf{W}_q \phi(r)) \odot \mathbf{m}}{\sqrt{E}}\right\}^\top \mathbf{x}^*_\ell \\ &= \gamma_r^\top \mathbf{x}^*_\ell \end{aligned}$$

Spatially Aware Spectro-Temporal Feature Reduction

- Spatial positional encoding [SL20]

$$\begin{aligned} \varphi : \mathbb{R}^2 &\to \mathbb{R}^{D\times T} \\ (\psi_1, \psi_2) &\mapsto \text{MLP}\Big(\sin(\psi_1\nu_1), \cos(\psi_1\nu_1), \ldots, \cos(\psi_2\nu_{F/4})\Big) \\ &= \mathbf{P} \end{aligned}$$

- Latent representation

$$\mathbf{Z} = \mathbf{B}\Big[\mathbf{X}^* + \mathbf{P}\Big]\mathbf{\Gamma}$$

where $\mathbf{Z} = [\mathbf{z}(r_1), \ldots, \mathbf{z}(r_R)] \in \mathbb{R}^{D'\times R}$, $\mathbf{B} \in \mathbb{R}^{D'\times D}$ and $\mathbf{\Gamma} = [\gamma_{r_1}, \ldots, \gamma_{r_R}] \in \mathbb{R}^{T\times R}$.

- Spatial positional encoding [SL20]

$$\begin{aligned}\varphi : \mathbb{R}^2 &\rightarrow \mathbb{R}^{D\times T}\\ (\psi_1, \psi_2) &\mapsto \text{MLP}\Big(\sin(\psi_1\nu_1), \cos(\psi_1\nu_1), \ldots, \cos(\psi_2\nu_{F/4})\Big)\\ &= \mathbf{P}\end{aligned}$$

- Latent representation

$$\mathbf{Z} = \mathbf{B}\Big[\mathbf{X}^* + \mathbf{P}\Big]\mathbf{\Gamma}$$

where $\mathbf{Z} = [\mathbf{z}(r_1), \ldots, \mathbf{z}(r_R)] \in \mathbb{R}^{D'\times R}$, $\mathbf{B} \in \mathbb{R}^{D'\times D}$ and $\mathbf{\Gamma} = [\gamma_{r_1}, \ldots, \gamma_{r_R}] \in \mathbb{R}^{T\times R}$.

- $||\mathbf{Z}^i - \mathbf{Z}^j||_F^2 = ||\mathbf{B}\big(\mathbf{X}^{i*}\mathbf{\Gamma}^i - \mathbf{X}^{j*}\mathbf{\Gamma}^j\big)||_F^2 + ||\mathbf{B}\big(\mathbf{P}^i\mathbf{\Gamma}^i - \mathbf{P}^j\mathbf{\Gamma}^j\big)||_F^2 + 2\Big\langle \mathbf{B}\big(\mathbf{X}^{i*}\mathbf{\Gamma}^i - \mathbf{X}^{j*}\mathbf{\Gamma}^j\big), \mathbf{B}\big(\mathbf{P}^i\mathbf{\Gamma}^i - \mathbf{P}^j\mathbf{\Gamma}^j\big)\Big\rangle_F$

Sparse Variational Gaussian Process

This section is based on

V. Bellet, M. Fauvel and J. Inglada, “Land Cover Classification With Gaussian Processes Using Spatio-Spectro-Temporal Features,” in IEEE Transactions on Geoscience and Remote Sensing, 2023, [BFI23]

Gaussian Process

- A GP f is defined by m and k: $f \sim \mathcal{GP}(m, k)$
- Any finite size observation $\left[f(\mathbf{Z}^1), \ldots, f(\mathbf{Z}^N)\right]^\top \sim \mathcal{N}_N(\boldsymbol{\mu}, \mathbf{K})$
- ψ *link function* between the GP and the target
- Univariate regression
  - $y = \psi\big(f(\mathbf{Z})\big) = f(\mathbf{Z}) + \epsilon$ with $\epsilon \sim \mathcal{N}_1(0, \sigma^2)$
  - Model fitting

$$\log p(\mathbf{y}|\mathbf{Z}, \boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2}\log\left(|\mathbf{K} + \sigma^2\mathbf{I}_n|\right) - \frac{n}{2}\log(2\pi)$$

  - Prediction

$$\hat{y}|\tilde{\mathbf{Z}} \sim \mathcal{N}_1\left(m(\tilde{\mathbf{Z}}) + \mathbf{k}_{\tilde{\mathbf{Z}}}^\top(\mathbf{K} + \sigma^2\mathbf{I}_n)^{-1}(\mathbf{y} - \boldsymbol{\mu}), k(\tilde{\mathbf{Z}}, \tilde{\mathbf{Z}}) - \mathbf{k}_{\tilde{\mathbf{Z}}}^\top(\mathbf{K} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{k}_{\tilde{\mathbf{Z}}} + \sigma^2\right)$$

- Several applications in remote sensing: see [Cam+16]

Gaussian Process for Multiclass Classification

- Multi-output / Multivariate Gaussian Process: $\mathbf{f} = [f_1, \ldots, f_C]$, $\mathfrak{m} \in \mathbb{R}^C$ and $\mathcal{K} \in \mathbb{R}^{C \times C}$

$$\mathfrak{m}(\mathsf{Z}) = \begin{bmatrix} m_1(\mathsf{Z}) & \ldots & m_C(\mathsf{Z}) \end{bmatrix}^\top,$$

$$\mathcal{K}(\mathsf{Z}, \mathsf{Z}) = \begin{bmatrix} k_{11}(\mathsf{Z}, \mathsf{Z}) & \ldots & k_{1C}(\mathsf{Z}, \mathsf{Z}) \\ \vdots & k_{pp'}(\mathsf{Z}, \mathsf{Z}) & \vdots \\ k_{C1}(\mathsf{Z}, \mathsf{Z}) & \ldots & k_C(\mathsf{Z}, \mathsf{Z}) \end{bmatrix},$$

- Any finite size observation $\left[\mathbf{f}(\mathsf{Z}^1), \ldots, \mathbf{f}(\mathsf{Z}^N)\right]^\top \sim \mathcal{N}_{NC}(\boldsymbol{\mu}, \mathbf{K})$
- Link function: $\mathbf{y} = \text{softmax}\big(\mathbf{f}(\mathsf{Z})\big)$

Gaussian Process for Multiclass Classification

- Multi-output / Multivariate Gaussian Process: $\mathbf{f} = [f_1, \ldots, f_C]$, $\mathfrak{m} \in \mathbb{R}^C$ and $\mathcal{K} \in \mathbb{R}^{C \times C}$

$$\mathfrak{m}(\mathbf{Z}) = \begin{bmatrix} m_1(\mathbf{Z}) & \ldots & m_C(\mathbf{Z}) \end{bmatrix}^\top,$$

$$\mathcal{K}(\mathbf{Z}, \mathbf{Z}) = \begin{bmatrix} k_{11}(\mathbf{Z}, \mathbf{Z}) & \ldots & k_{1C}(\mathbf{Z}, \mathbf{Z}) \\ \vdots & k_{pp'}(\mathbf{Z}, \mathbf{Z}) & \vdots \\ k_{C1}(\mathbf{Z}, \mathbf{Z}) & \ldots & k_C(\mathbf{Z}, \mathbf{Z}) \end{bmatrix},$$

- Any finite size observation $\left[\mathbf{f}(\mathbf{Z}^1), \ldots, \mathbf{f}(\mathbf{Z}^N)\right]^\top \sim \mathcal{N}_{NC}(\boldsymbol{\mu}, \mathbf{K})$
- Link function: $\mathbf{y} = \text{softmax}\big(\mathbf{f}(\mathbf{Z})\big)$
- Issues in large scale classification
  - ⋆ Computational complexity in $\mathcal{O}\big((CN)^3\big)$, memory footprint in $\mathcal{O}\big((CN)^2\big)$,
  - ⋆ ψ = softmax $\rightarrow$ no-analytic solution for *likelihood* and *posterior*

Linear Model of Co-Regionalization

- Linear combination of independant univariate GPs: $g_c \sim \mathcal{GP}(m_c, k_l)$ [GG97]

$$\mathbf{f} = \mathbf{A}\mathbf{g}$$

- Computational complexity reduce to $\mathcal{O}(CN^3)$
- $\mathbf{A}$ is learned as parameter of the model

Sparse Approximation

- Introduce latent variables / inducing points: $\mathbf{Z}_u$ with $u \in \{1, \ldots, M\}$

$$\mathbf{K}_{NN} \approx \mathbf{K}_{NM}\mathbf{K}_{MM}^{-1}\mathbf{K}_{MN}$$

- Computational complexity reduce to $\mathcal{O}(CNM^3)$

Linear Model of Co-Regionalization

- Linear combination of independant univariate GPs: $g_c \sim \mathcal{GP}(m_c, k_l)$ [GG97]

$$\mathbf{f} = \mathbf{A}\mathbf{g}$$

- Computational complexity reduce to $\mathcal{O}(CN^3)$
- $\mathbf{A}$ is learned as parameter of the model

Sparse Approximation

- Introduce latent variables / inducing points: $\mathbf{Z}_u$ with $u \in \{1, \ldots, M\}$

$$\mathbf{K}_{NN} \approx \mathbf{K}_{NM}\mathbf{K}_{MM}^{-1}\mathbf{K}_{MN}$$

- Computational complexity reduce to $\mathcal{O}(CNM^3)$
- How to choose/find the inducing points ?

Learn them as parameters but size grow as $CMD'R$

Variational Inference

- Optimize a lower bound of the log-likelihood (ELBO) [BKM17]
- For univariate SGP [HMG15]:

$$\mathcal{E}(q) = \sum_{i=1}^{N} \mathbb{E}_{q'(g(\mathbf{Z}^i)|\boldsymbol{\theta}^{\nu},\boldsymbol{\theta})}\left[\log p\left(y^i|g(\mathbf{Z}^i)\right)\right] - \mathrm{KL}\left[q\left(g(\mathbf{Z}_u)|\boldsymbol{\theta}^{\nu}\right)\|\, p\left(g(\mathbf{Z}_u)|\boldsymbol{\theta}\right)\right],$$

with

$$q\left(g(\mathbf{Z}_u)|\boldsymbol{\theta}^{\nu}\right) \sim \mathcal{N}_M\left(\mathbf{m}, \mathbf{S}\right)$$
$$q'\left(g(\mathbf{Z}^i)|\boldsymbol{\theta}^{\nu}, \boldsymbol{\theta}\right) \sim \mathcal{N}_1\left(g(\mathbf{Z}^i)|\ \mathbf{k}_{Mi}^{\top}\mathbf{K}_{MM}^{-1}\mathbf{m}, k(\mathbf{Z}^i, \mathbf{Z}^i) - \mathbf{k}_{Mi}^{\top}\mathbf{K}_{MM}^{-1}\left(\mathbf{K}_{MM} - \mathbf{S}\right)\mathbf{K}_{MM}^{-1}\mathbf{k}_{Mi}\right)$$

- Expectation approximate with MC sampling and reparametrisation trick

Variational Inference

- Optimize a lower bound of the log-likelihood (ELBO) [BKM17]
- For univariate SGP [HMG15]:

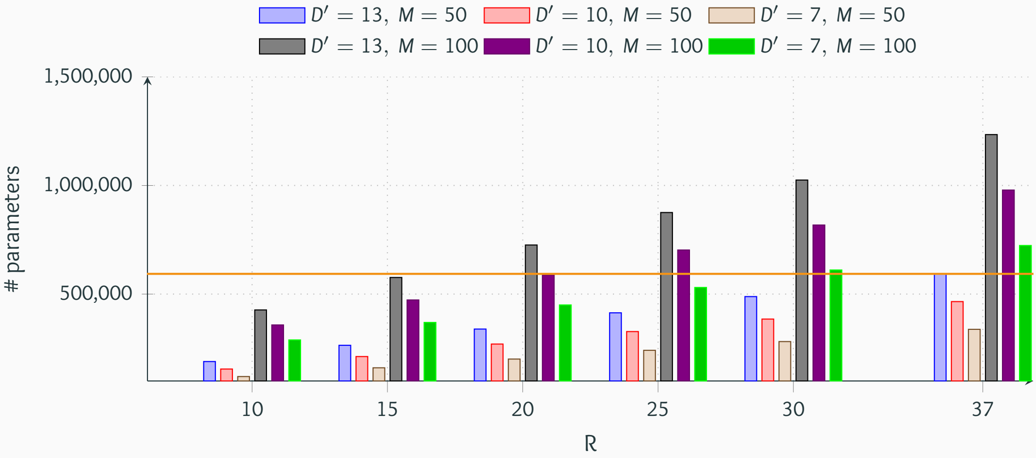
$$\mathcal{E}(q) = \sum_{i=1}^{N} \mathbb{E}_{q'(g(\mathbf{Z}^i)|\boldsymbol{\theta}^v,\boldsymbol{\theta})}\Big[\log p\big(y^i|g(\mathbf{Z}^i)\big)\Big] - \mathrm{KL}\Big[q\big(g(\mathbf{Z}_u)|\boldsymbol{\theta}^v\big) \,\|\, p\big(g(\mathbf{Z}_u)|\boldsymbol{\theta}\big)\Big],$$

with

$$q\big(g(\mathbf{Z}_u)|\boldsymbol{\theta}^v\big) \sim \mathcal{N}_M\big(\mathbf{m}, \mathbf{S}\big)$$
$$q'\big(g(\mathbf{Z}^i)|\boldsymbol{\theta}^v, \boldsymbol{\theta}\big) \sim \mathcal{N}_1\Big(g(\mathbf{Z}^i)|\ \mathbf{k}_{Mi}^\top \mathbf{K}_{MM}^{-1}\mathbf{m}, k(\mathbf{Z}^i, \mathbf{Z}^i) - \mathbf{k}_{Mi}^\top \mathbf{K}_{MM}^{-1}\big(\mathbf{K}_{MM} - \mathbf{S}\big)\mathbf{K}_{MM}^{-1}\mathbf{k}_{Mi}\Big)$$

- Expectation approximate with MC sampling and reparametrisation trick
- Extension to LMC straightforward ...

Model Complexity

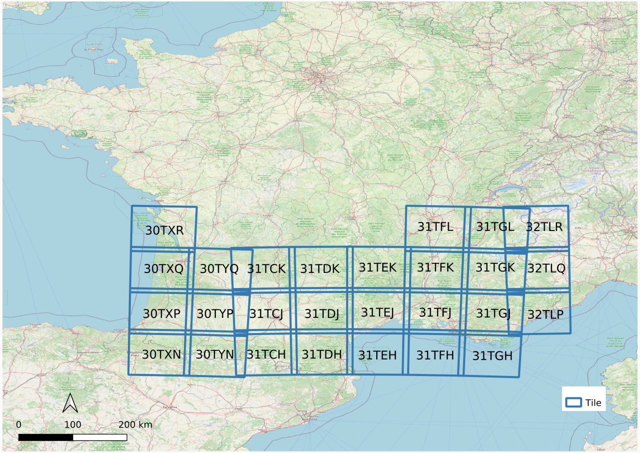


Experiments

Data set

- All S2 acquisitions between [01-2018, 12-2018]
- 10 bands + 3 spectral indices
- $T = 303$ & $D = 13$
- 23 land cover classes
  - ⋆ Training: 4000 pixels/class
  - ⋆ Validation: 1000 pixels/class
  - ⋆ Test: 10,000 pixels/class
  - ⋆ 9 random (train, val, test) sets

Training	*Validation*	*Test*
92 000	23 000	230 000



Data set

- All S2 acquisitions between [01-2018, 12-2018]
- 10 bands + 3 spectral indices
- $T = 303$ & $D = 13$
- 23 land cover classes
  - ⋆ Training: 4000 pixels/class
  - ⋆ Validation: 1000 pixels/class
  - ⋆ Test: 10,000 pixels/class
  - ⋆ 9 random (train, val, test) sets

Training	*Validation*	*Test*
92 000	23 000	230 000

Color	Code	Name
	CUF	Continuous urban fabric
	DUF	Discontinuous urban fabric
	ICU	Industrial and commercial units
	RSF	Road surfaces
	RAP	Rapeseed
	STC	Straw cereals
	PRO	Protein crops
	SOY	Soy
	SUN	Sunflower
	COR	Corn
	RIC	Rice
	TUB	Tubers / roots
	GRA	Grasslands
	ORC	Orchards and fruit growing
	VIN	Vineyards
	BLF	Broad-leaved forest
	COF	Coniferous forest
	NGL	Natural grasslands
	WOM	Woody moorlands
	NMS	Natural mineral surfaces
	BDS	Beaches, dunes and sand plains
	GPS	Glaciers and perpetual snows
	WAT	Water bodies

Classification Accuracy - Specto-Temporal Information Only

- Comparison with *prior* linear interpolation

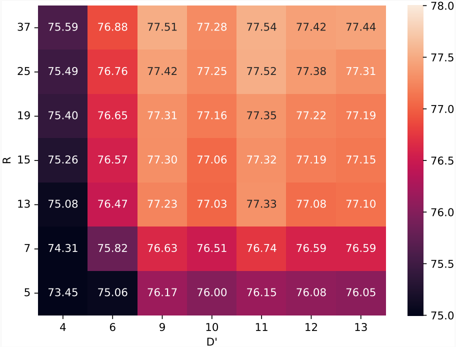
	mTAN-SVGP	linInter-SVGP	linInter-RF
OA (%)	77.4 (0.2)	67.3 (0.4)	65.4 (0.4)
Time (s)	1317.4	336.6	54.6

Classification Accuracy - Specto-Temporal Information Only

- Comparison with *prior* linear interpolation

	mTAN-SVGP	linInter-SVGP	linInter-RF
OA (%)	77.4 (0.2)	67.3 (0.4)	65.4 (0.4)
Time (s)	1317.4	336.6	54.6

- Influence of R and D'



OA (%)

Training Time (s)

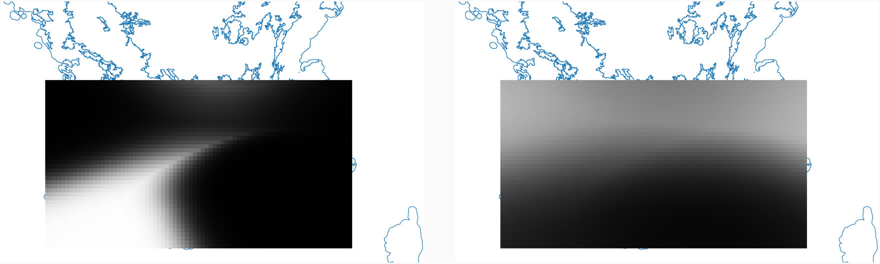
Classification Accuracy - Spatio-(Reduced)-Spectro-Temporal Information

- $R = 13$ & $D' = 9$ and spatial positional encoding

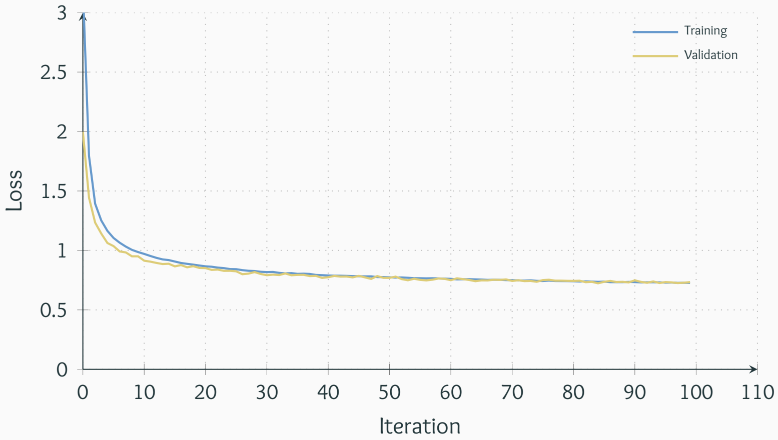
	With	Without
OA (%)	78.6 (0.2)	77.2 (0.2)
Time (s)	834	871

- $R = 13$ & $D' = 9$ and spatial positional encoding

	With	Without
OA (%)	78.6 (0.2)	77.2 (0.2)
Time (s)	834	871

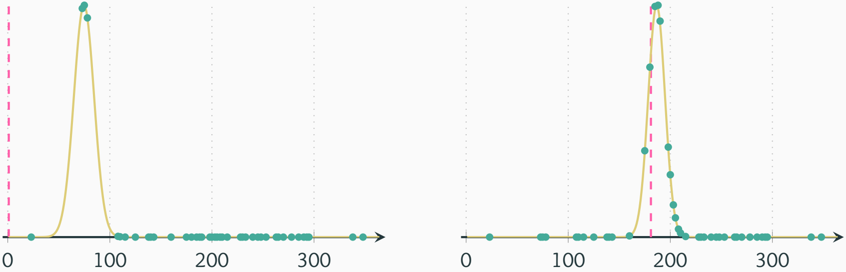


Convergence

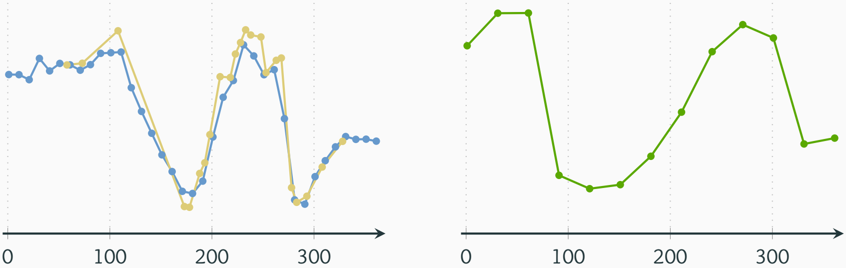


Parameters Analysis

- Similarity kernel - normalized attention



- Latent feature (for *corn* pixel)



Comparison With Others Methods

Competitive Methods

- mTAN + MLP
- mTAN + LTAE [SL20]
- (raw + mask) + LTAE

	mTAN-SVGP	mTAN-MLP	mTAN-LTAE	raw-LTAE
OA (%)	78.6	71.5	74.6	81.5
Time (s)	834	1207	840	1279

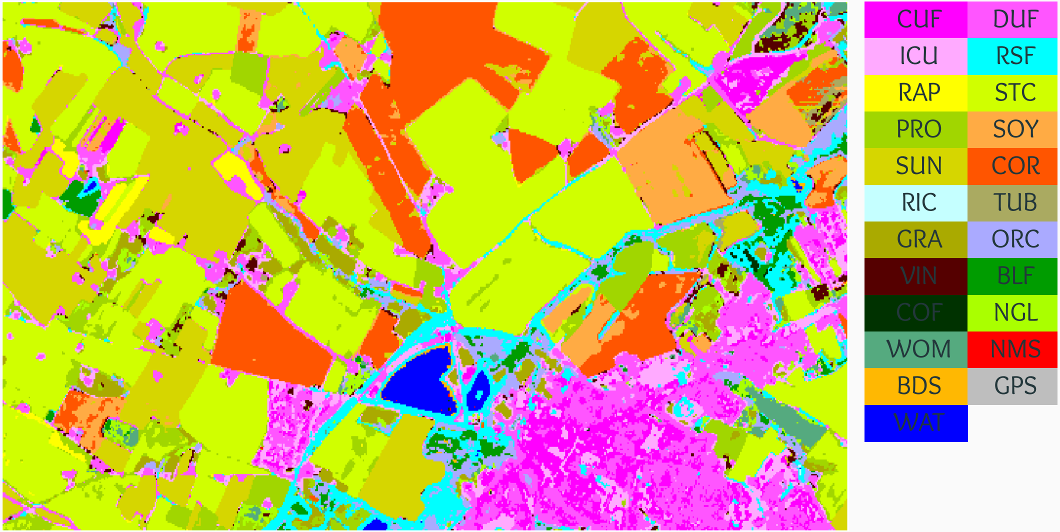
Temporal shift

- One tile for test: change t to $t + \delta$

	0	1	2	3	5
mTAN-SVGP	77.4 (0.6)	77.4 (0.8)	77.4 (0.8)	77.4 (0.8)	77.0 (0.6)
raw-LTAE	80.0 (0.6)	76.0 (2.0)	66.5 (4.4)	56.6 (9.2)	52.3 (8.2)

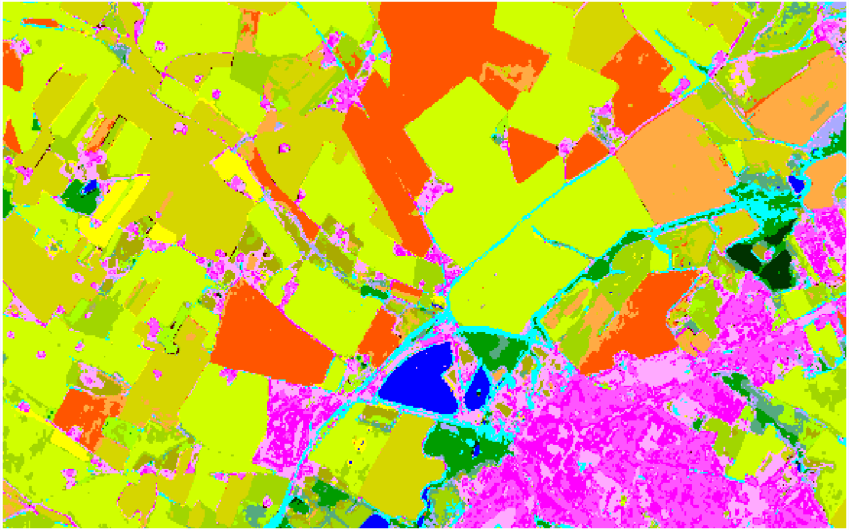
Classification Maps

Classification Maps



Classification Maps

g



CUF	DUF
ICU	RSF
RAP	STC
PRO	SOY
SUN	COR
RIC	TUB
GRA	ORC
VIN	BLF
COF	NGL
WOM	NMS
BDS	GPS
WAT	

Conclusions

Conclusions & Perspectives

Conclusions

- Classification of irregular and unaligned SITS
- Scale to large data set
- Versatility

Perspectives

- Regression
- Process multi-year data
- Spatio-temporal attention
- Extend smoother

References

[1] Valentine Bellet, Mathieu Fauvel, and Jordi Inglada. "Land Cover Classification With Gaussian Processes Using Spatio-Spectro-Temporal Features." In: *IEEE Transactions on Geoscience and Remote Sensing* (2023), pp. 1–1. ISSN: 1558-0644. DOI: 10.1109/tgrs.2023.3234527. URL: http://dx.doi.org/10.1109/TGRS.2023.3234527.

[2] Valentine Bellet et al. "End-to-end Learning for Land Cover Classification using Irregular and Unaligned SITS by Combining Attention-Based Interpolation with Sparse Variational Gaussian Processes." working paper or preprint. June 2023. URL: https://hal.science/hal-04112115.

[3] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians." In: *Journal of the American Statistical Association* 112.518 (2017), pp. 859–877.

[4] Gustau Camps-Valls et al. "A Survey on Gaussian Processes for Earth-Observation Data Analysis: A Comprehensive Investigation." In: *IEEE Geoscience and Remote Sensing Magazine* 4.2 (2016), pp. 58–78. DOI: 10.1109/MGRS.2015.2510084.

[5] P. Goovaerts and D.C.E.E.P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Applied geostatistics series. Oxford University Press, 1997. ISBN: 9780195115383.

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[7] James Hensman, Alex Matthews, and Zoubin Ghahramani. "Scalable Variational Gaussian Process Classification." In: *In Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*. 2015, pp. 351–360.

[8] Jordi Inglada et al. "Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series." In: *Remote Sensing* 9.1 (2017). ISSN: 2072-4292. DOI: 10.3390/rs9010095.

[9] Vivien Sainte Fare Garnot and Loic Landrieu. "Lightweight temporal self-Attention for classifying satellite images time series." In: *Workshop on Advanced Analytics and Learning on Temporal Data*. AALTD. Sept. 2020.

[10] Satya Narayan Shukla and Benjamin Marlin. "Multi-Time Attention Networks for Irregularly Sampled Time Series." In: *International Conference on Learning Representations*. 2021.

Thank you for your attention

https://gitlab.cesbio.omp.eu/fauvelm/tdma2023