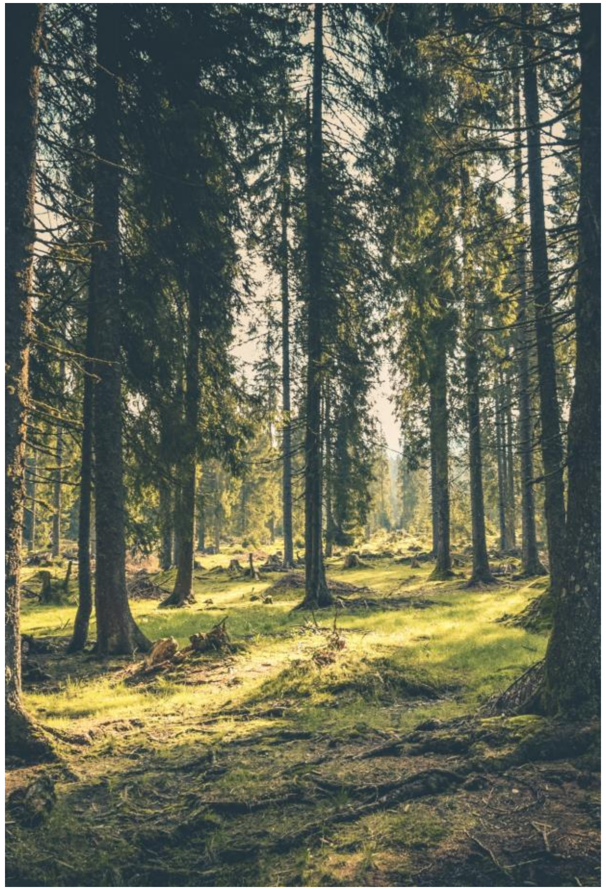
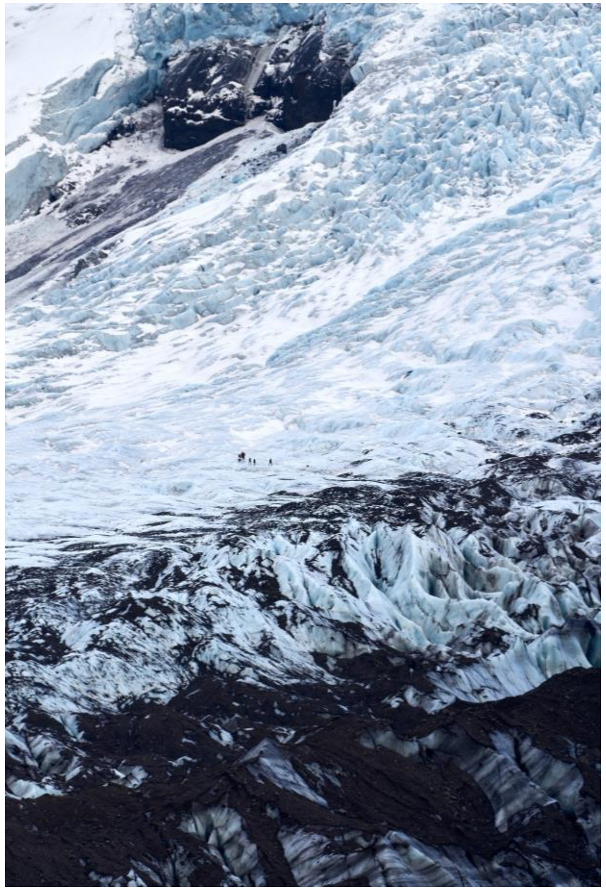
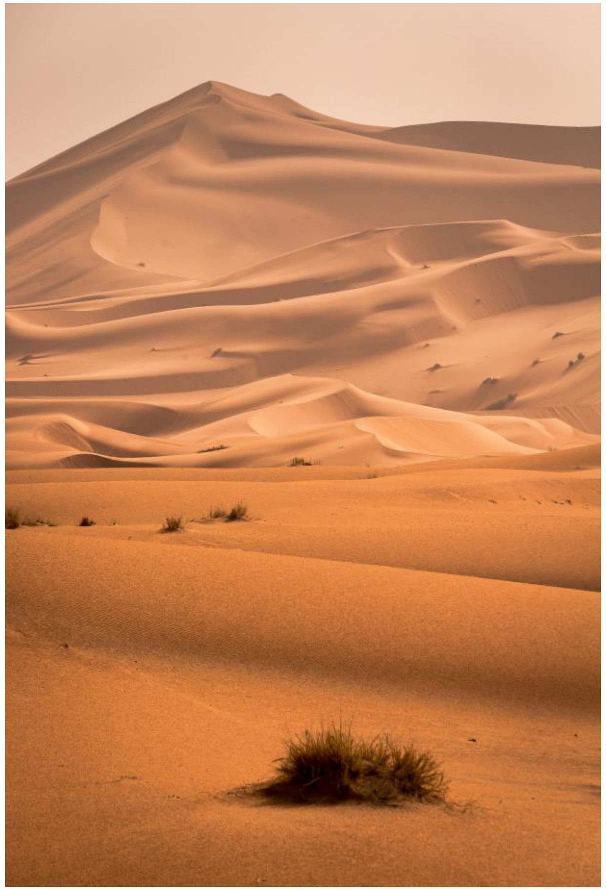





Global hydro-climatic bioms identified via MTL

Global hydro-climatic bioms identified via multitask learning

- Christina Papagiannopoulou,
  Diego G. Miralles, Matthias Demuzere,
  Niko E. C. Verhoest and Willem Waegeman
- Ghent University, Belgium

- Published: October 2018

<u>Abbreviations and clarifications:</u>

- Biome = ecological region containing all its plants, animals and nonliving things
- Hydo-climate biome = regions of coherent vegetation–climate behaviour
- MTL – Multitask learning
- NDVI - Normalized Difference Vegetation Index
- ASO - Alternative structure optimization

Why identifying biomes?

- Gain a better understanding of complex interactions among different environmental variables.
- Used as a diagnostic of climate change by exploring their shifting boundaries
- Predict future climatic zone distributions using climate projections (tipping- & turning points)
- Unravel anomalous relationships between climate and vegetation dynamics

- Aim: data-driven approach that aims to quantify the response of vegetation to local climate variables in a supervised setting at a global scale

Data

Papagiannopoulou et al. (2017a)

- Combination of 21 datasets (mixutre of satelite and in situ observations)
- 1° lat – 1° long. resolution (13072 locations)
- 360 months – 30 years (1981 – 2010)
- Most important features (of 3209 climate variables):
    - land surface temperature
    - near-surface air temperature,
    - longwave–shortwave surface radiative fluxes
    - precipitation
    - snow water equivalent
    - soil moisture

- NDVI - Normalized Difference Vegetation Index
    - detrended & multiyear average for sesonal expectation

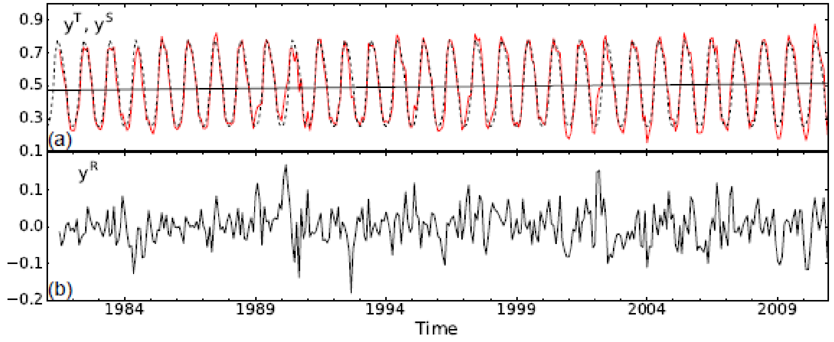
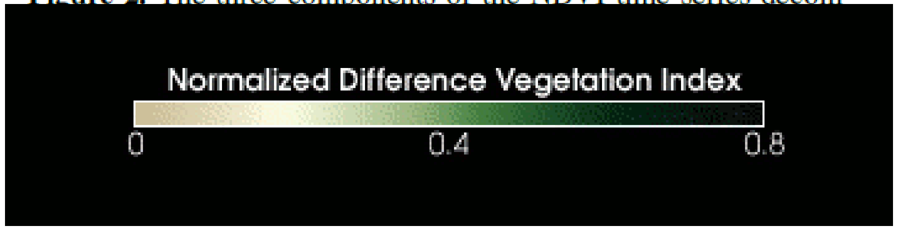
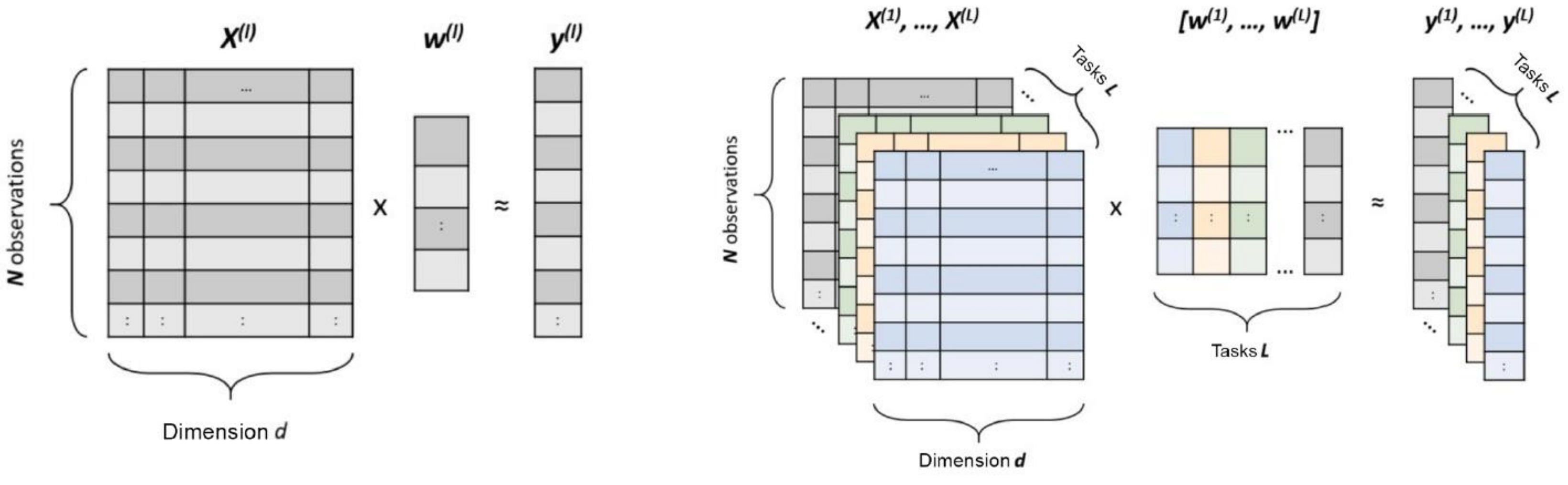


Figure 2. The three components of the NDVI time series decom-



Methods – Single task learning

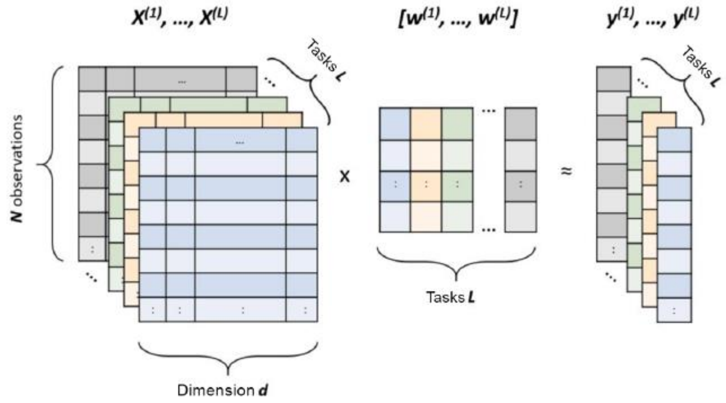
Pixel based apporach:



(a)
(b)

Minimzation problem: $$\min_{\boldsymbol{w}^{(l)}} \sum_{i=1}^{N} \mathcal{L}\left(\boldsymbol{w}^{(l)} \boldsymbol{x}_i^{(l)}, y_i^{(l)}\right) + \lambda ||\boldsymbol{w}^{(l)}||^2,$$ (Ridge Regression)

Methods - Multitask learning



- Making use of spatial relationship between taks
- Multitask minimization task

$$\min_{w^{(1)},\ldots,w^{(L)}} \sum_{l=1}^{L} \sum_{i=1}^{N} \mathcal{L}\left(\boldsymbol{w}^{(l)} \boldsymbol{x}_i^{(l)}, y_i^{(l)}\right) + \Omega\left(\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(L)}\right),$$

- With Ω as a factor of relatedness between taks

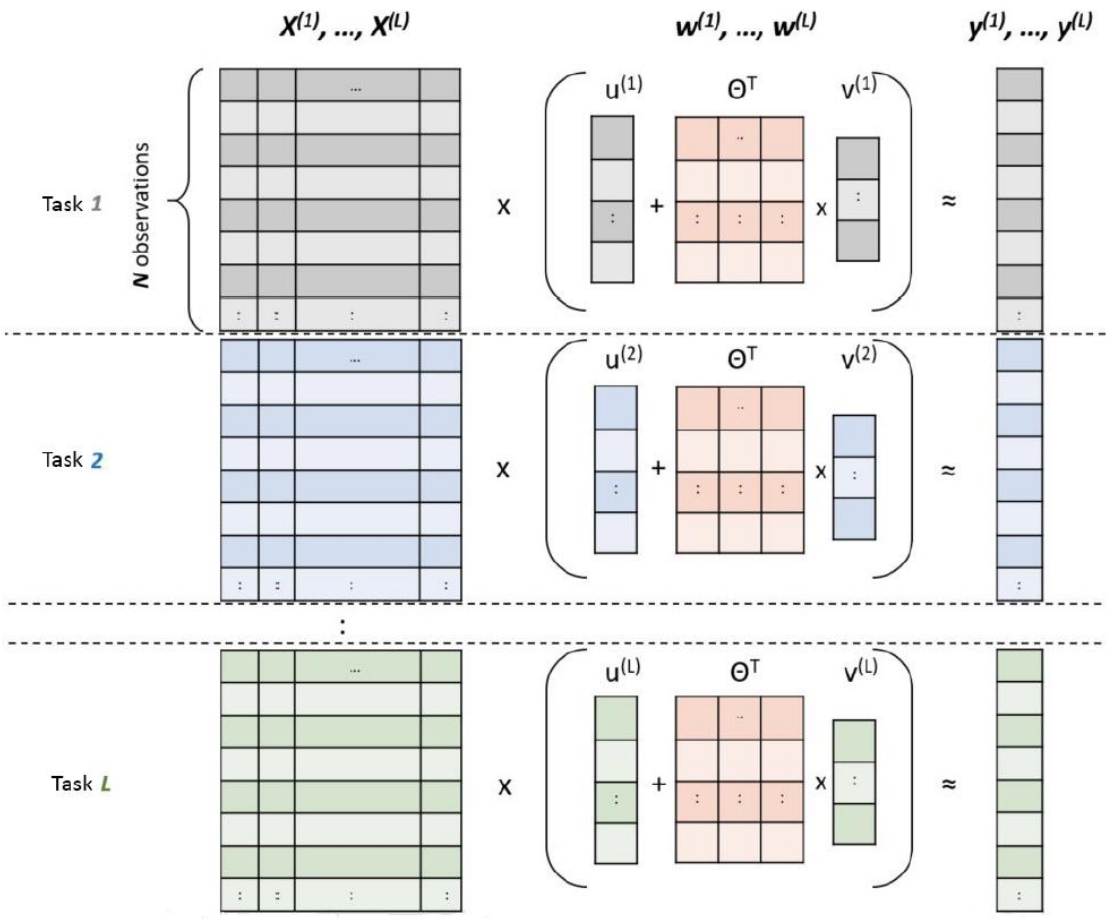
- SVD-based ASO for dimensionality reduction
- Compute low-dimensional feature map θ to look for similiarities within the tasks
- Spectral clustering – hierarchical agglomerative clustering

Alternative Structure Optimization (ASO)

- Step I: a model function for each individual task is learned
- Step 2: weight vector is decomposed by an SVD

$$f^{(l)}(\boldsymbol{x}) = \boldsymbol{w}^{(l)}\boldsymbol{x}_i^{(l)} = \boldsymbol{u}^{(l)}\boldsymbol{x}_i^{(l)} + \boldsymbol{v}^{(l)}\boldsymbol{\Theta}\boldsymbol{x}_i^{(l)},$$

- $w^{(l)} \in \mathbb{R}^d$ - *weight vector of location l*
- $u^{(l)} \in \mathbb{R}^d$ - *high dim. weights representation*
- $v^{(l)} \in \mathbb{R}^h$ - *low dim. weights representation*
- $\Theta^T \in \mathbb{R}^{dxh}$ - *low-dim. feature map*



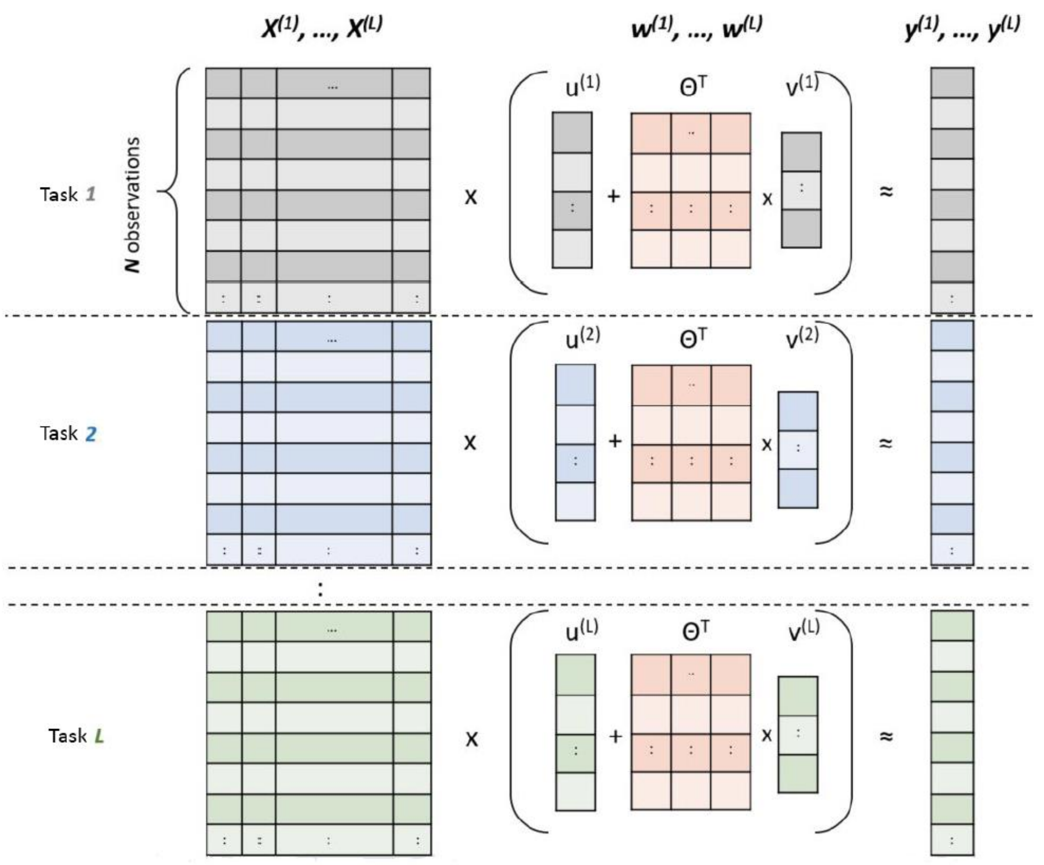
Alternative Structure Optimization (ASO)

- New minimization problem for ASO:

$$\min_{\{w^{(l)}, v^{(1)}\}, \Theta\Theta^T = \mathbf{I}} \sum_{l=1}^{L} \left(\sum_{i=1}^{N} \mathcal{L}(\boldsymbol{w}^{(l)} \boldsymbol{x}_i^{(l)}, y_i^{(l)}) + \lambda^{(l)} \|\boldsymbol{u}^{(l)}\|_2^2 \right),$$

- With $||u^{(l)}||_2^2$ being a regularization term that controls task relatedness among all tasks
  - Penalizes difference between high- and low-dim. space

$u^l = \omega^{(l)} - \Theta^T v^{(l)}$



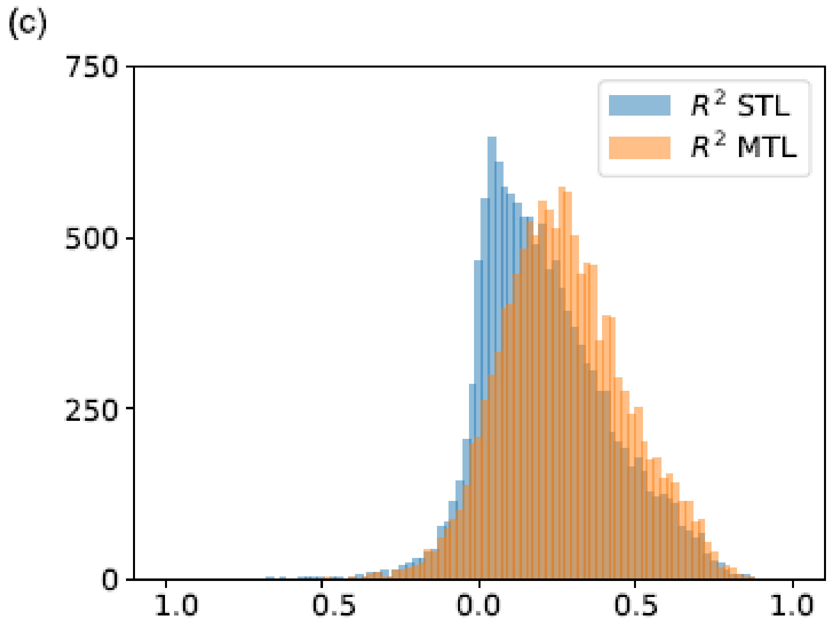
Alternative Structure Optimization (ASO)

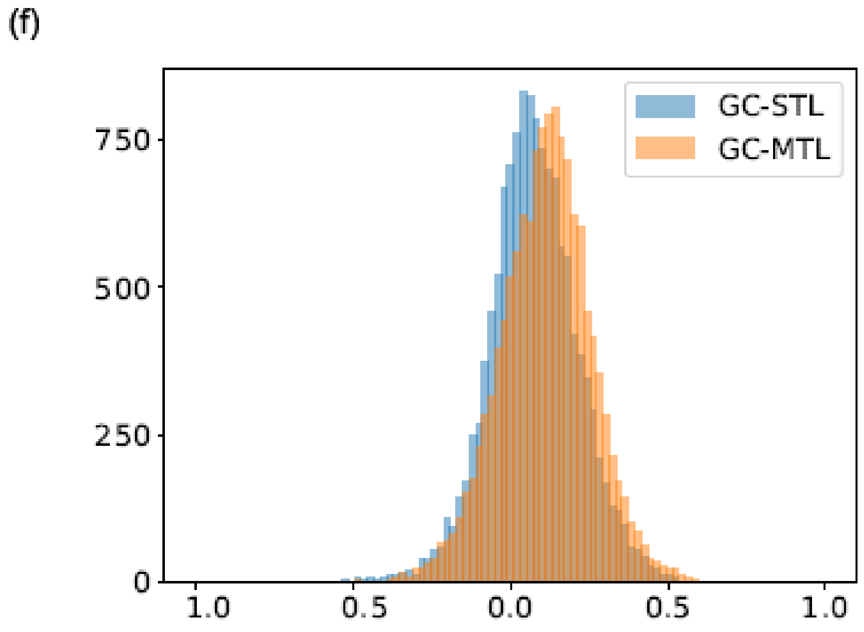
$u^l = \omega^{(l)} - \Theta^T v^{(l)}$

Algorithm 1 SVD-ASO

Input: training data $D^{(l)} = \left\{\left(\boldsymbol{x}_i^{(l)}, y_i^{(l)}\right)\right\}_{i=1,\ldots,N}$,
where $l = 1, \ldots, L$
Parameters: h and $\boldsymbol{\lambda} = \left\{\lambda^{(1)}, \ldots, \lambda^{(L)}\right\}$
Output: $\boldsymbol{\Theta} \in \mathbb{R}^{h \times d}$ and $\mathbf{V} = \left[\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}\right]^T \in \mathbb{R}^{L \times h}$
Initialize: $\boldsymbol{w}^{(l)} = 0,\ l = 1, \ldots, L,$ and $\boldsymbol{\Theta}$ to random
repeat
  for $l = 1$ **to** L **do**
    with fixed $\boldsymbol{\Theta}$ and $\boldsymbol{v}^{(l)} = \boldsymbol{\Theta}\boldsymbol{w}^{(l)}$, solve the optimization problem of Eq. (3) for $\boldsymbol{u}^{(l)}$:
    $\text{argmin}_{\boldsymbol{u}^{(l)}} \sum_{i=1}^{N} \mathcal{L}\left(\boldsymbol{u}^{(l)}\boldsymbol{x}_i^{(l)} + (\boldsymbol{v}^{(l)}\boldsymbol{\Theta})\boldsymbol{x}_i^{(l)}, y_i^{(l)}\right) + \lambda^{(l)}\|\boldsymbol{u}^{(l)}\|_2^2$
    $\boldsymbol{w}^{(l)} = \boldsymbol{u}^{(l)} + \boldsymbol{\Theta}^T\boldsymbol{v}^{(l)}$
  end for
  Apply an SVD decomposition on
  $\mathbf{W} = \left[\sqrt{\lambda^{(1)}}\boldsymbol{w}^{(1)}, \ldots, \sqrt{\lambda^{(L)}}\boldsymbol{w}^{(L)}\right]$:
  $\mathbf{W} = \mathbf{V_1}\mathbf{D}\mathbf{V_2}^T$ (with diagonals of $\mathbf{D}$ in descending order)
  $\boldsymbol{\Theta} = \mathbf{V_1}^T[:h,:]$ // update $\boldsymbol{\Theta}$ to the first h rows of $\mathbf{V_1}^T$
until convergence

Performance of MTL compared to STL

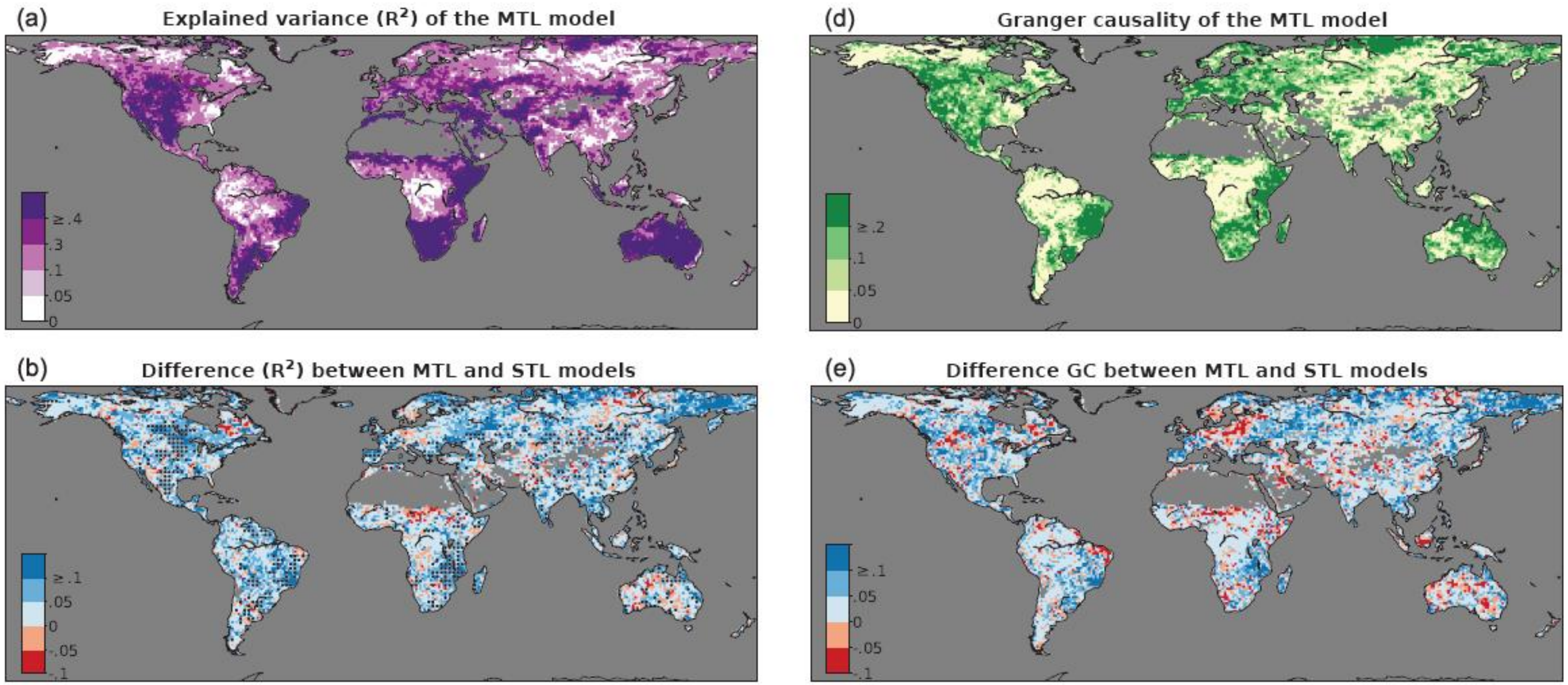




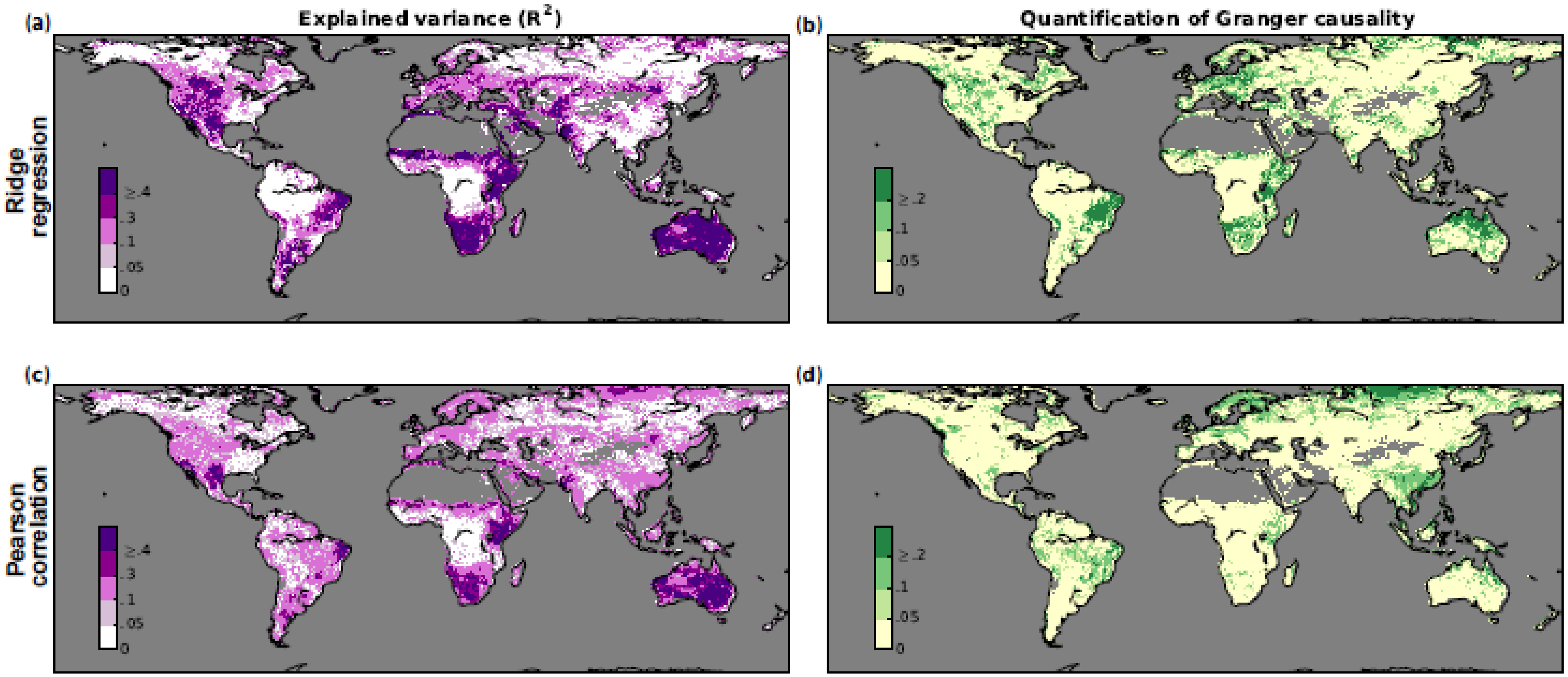
$$R^2(y, \hat{y}) = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\sum_{i=P+1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=P+1}^{N} (y_i - \bar{y})^2},$$

With $\hat{y}$: the predicted value
$\bar{y}$: the mean of the timeseries

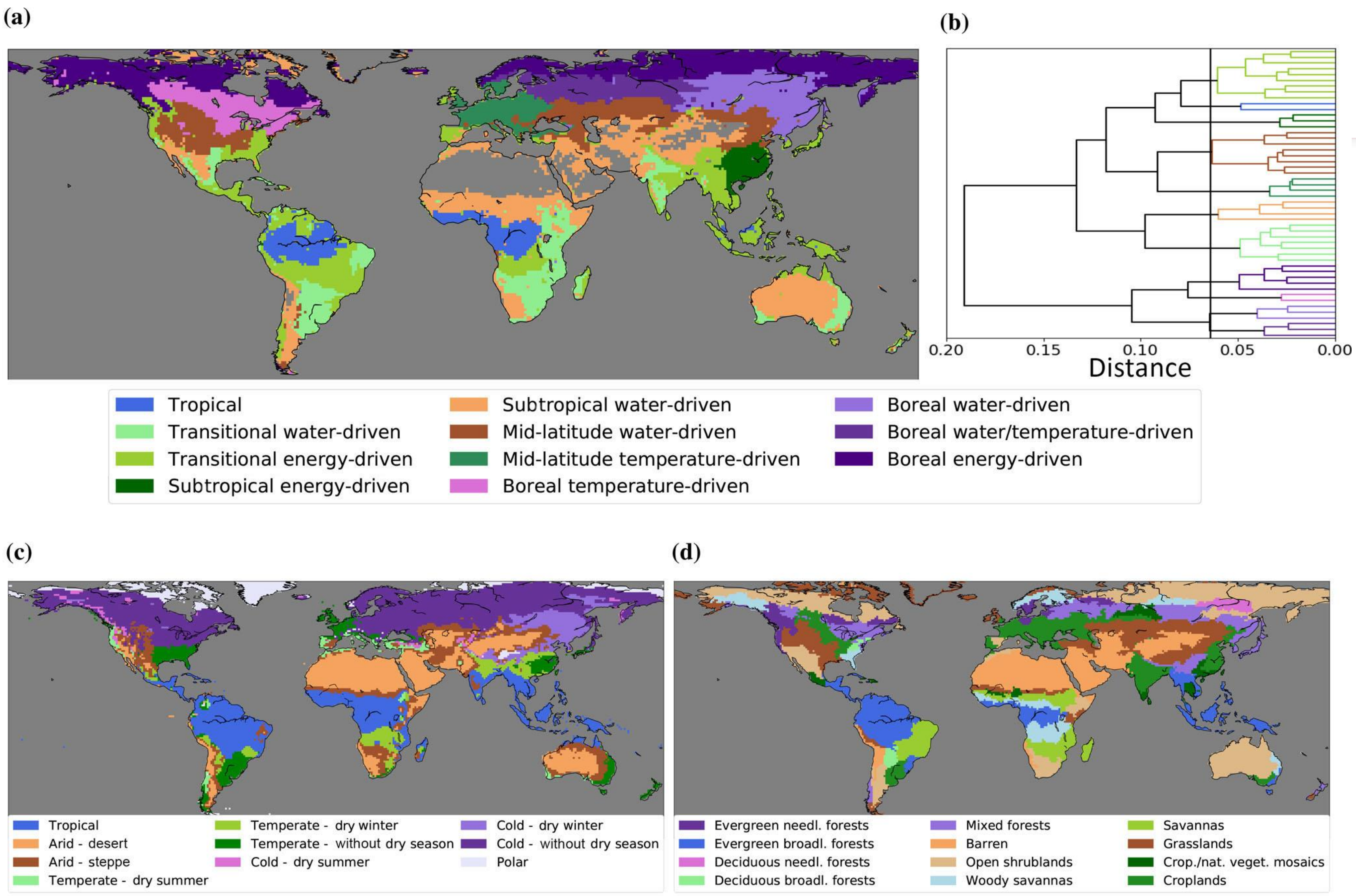
Comparison of Performance of MTL and STL



Variability and Granger Causality of the STL model



Classification of the MTL approach



Scatter plots of MTL, Köppen-Geiger and IGBP