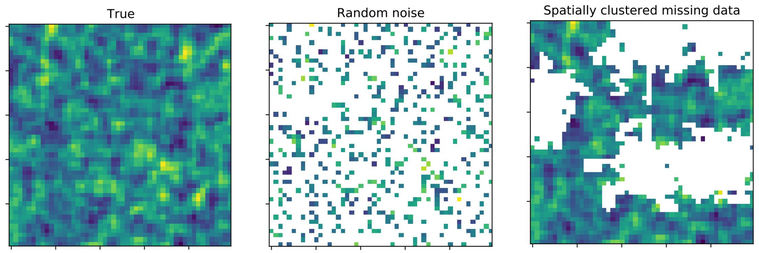
Value propagation-based spatial interpolation

Laurens Arp, Mitra Baratchi, Holger Hoos

Motivation

Problem: (spatial) interpolation

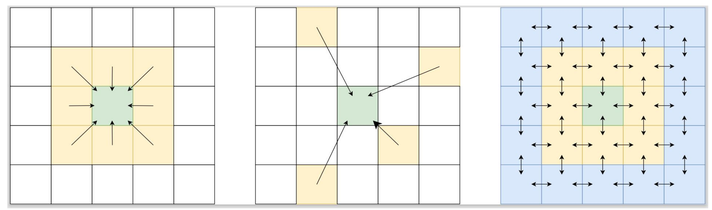
- How can **missing (spatial) data** best be filled in?
  - For example: cloud cover or sensor defects in remote sensing (RS) data
- Why should we care?
  - Real-world data cannot be assumed to have laboratory-level consistency and availability (task 7.1)
  - Maximise data exploitation
  - Important preprocessing step for downstream applications
  - Relevance to data-driven AutoAI preprocessing pipelines



Method and experiments

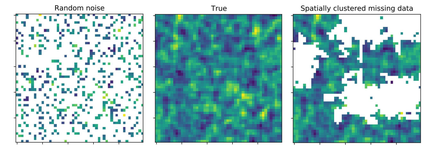
Proposed method

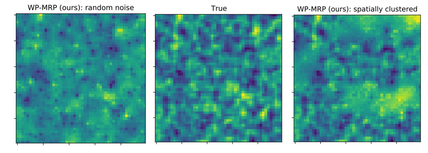
- Existing **local** methods (left)
  - Spatial autoregressive models (SAR, ARMA)
  - Moving average models (MA, ARMA)
  - Convolutional neural networks (CNN)
- Existing **distance-based** methods (middle)
  - Gaussian processes/Kriging
- Ours: **system-oriented** method (right)
  - Recursively propagate known values through system (task 7.2)
  - Predict/apply weight based on spatial features
  - Iterate until equilibrium (MRP)

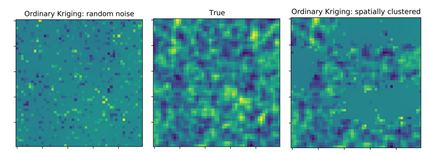


Results and discussion

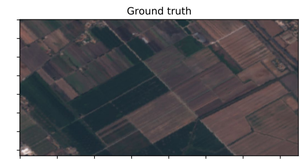
Synthetic data

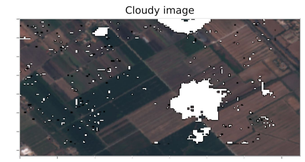






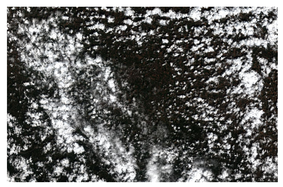
Cloud removal





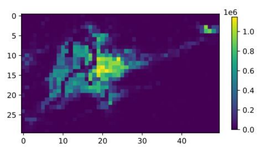
Work in progress: remote sensing cloud removal

- Target: cloudy Sentinel-2 MSI (optical) images
- Features
  - Cloud-free old MSI images
  - Sentinel-1 SAR (radar) images
- Role of AutoAI
  - Preprocessing of RS data not trivial
    - MSI: atmospheric correction, sunlight intensity, normalisation methods
    - SAR: calibration, multi-looking, speckle filtering, terrain correction
- Implications for AutoAI
  - Automatically fill in missing data as part of preprocessing pipeline

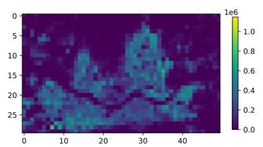


Experiments

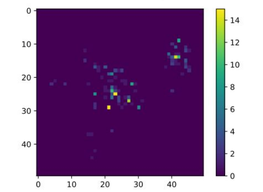
Baselines: Kriging, SAR, MA, ARMA, non-spatial regression, CNNs



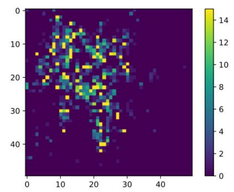
(a) GDP Taipei



(b) GDP Seoul

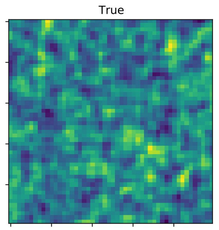


(c) COVID-19 Daegu



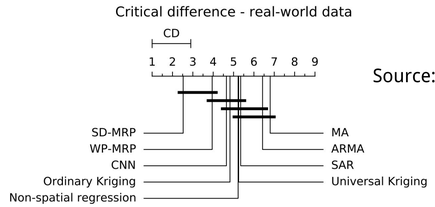
(d) COVID-19 Seoul

Left: real-world data
Bottom: synthetic data



Conclusions and future work

- Visually plausible results
- Quantitatively
  - Significantly better than baselines on real-world data (MAE)
  - Scaled better to large datasets than kriging, CNNs



Source: https://github.com/LaurensArp/VPint

- Future work
  - Incorporate into fully automated preprocessing pipeline for remote sensing data