PATCH-FILL

Multiscale Gap-Filling for Earth System Data Cubes

Charly Zimmer[1], Anja Neumann[1], Miguel D. Mahecha[2], Josefine Umlauft[1]

[1]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Leipzig University, Leipzig, Germany
[2]Institute for Earth System Science and Remote Sensing, Leipzig University, Leipzig, Germany

UNIVERSITÄT LEIPZIG

EARTH SYSTEM DATA LAB

ScaDS.AI DRESDEN LEIPZIG

esa

Preprint

Motivation

- Many applications in Earth system science require **gap-free** data sets
- Remote sensing data are plagued by gaps due to **cloud obstruction**, **incomplete satellite coverage**, and **low-quality flags**
- Gap-Filling in remote sensing data with the help of machine learning often requires model architectures that are tailored specifically to underlying dataset characteristics such as scale, resolution or range of values
- This limits the **transferability** to other gap-filling scenarios
- Training these models is hindered by the **lack of adequate training samples**, as they must be gathered from gap-afflicted data themselves

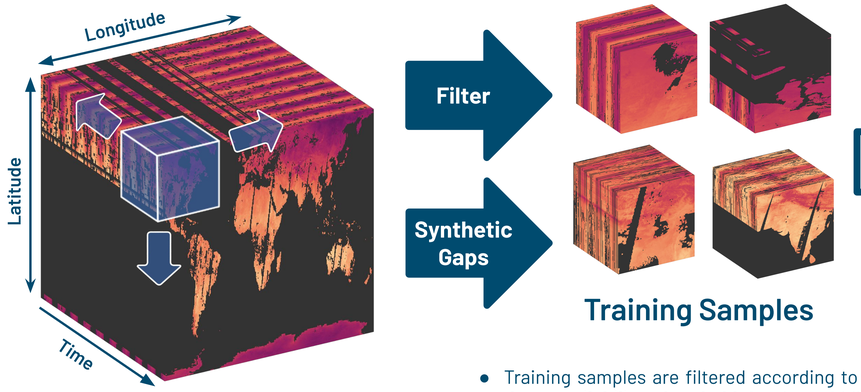
Cloud Obstruction

Incomplete Satellite Coverage

Our Approach

- We present a generic gap-filling method that is spatio-temporal, univariate and multiscale
- We employ a **patch-wise gap-filling framework**
- By introducing **masked loss functions** and providing effective methods for **synthetic gap generation**, we are able to leverage gap-afflicted datasets and gather large volumes of training samples from them
- We perform gap-filling on multiple climatic variables from Earth System Data Cubes (ESDC) using a 3D CNN architecture, making this the **first global-scale gap-filling solution on ESDC**
- We validate the generalization capabilities of the model by performing **Spatial Leave-One-Out (SLOO)** Cross Validation

① Dataset Sampling

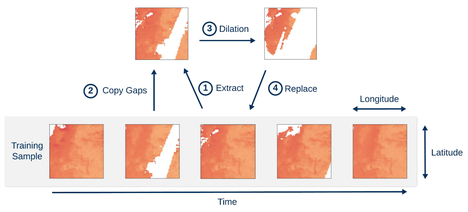


- The data cube is iterated along every dimension to extract training samples of a pre-defined shape (= "***kernel shape***"), containing patches of subsequent time steps
- For Land Surface Temperature:
  Time: 5 timesteps
  Longitude/Latitude: 64 grid points

- Training samples are filtered according to a **data availability threshold** in the middle patch of a training sample (= "***gap patch***")
- The gap patch is used to construct a target by introducing **synthetic gaps** so that the areas that were synthetically removed can act as target values
- By using a data availability threshold, we allow for target patches that contain gaps (= "***incomplete targets***")

Synthetic Gap Generation

- The system extracts the middle patch ① and imposes the gap areas from the previous patch on it ② . Gaps are enlarged using dilation ③ . The processed patch replaces the gap patch, while the original gap patch becomes the target ④.



② Model Training

U Net

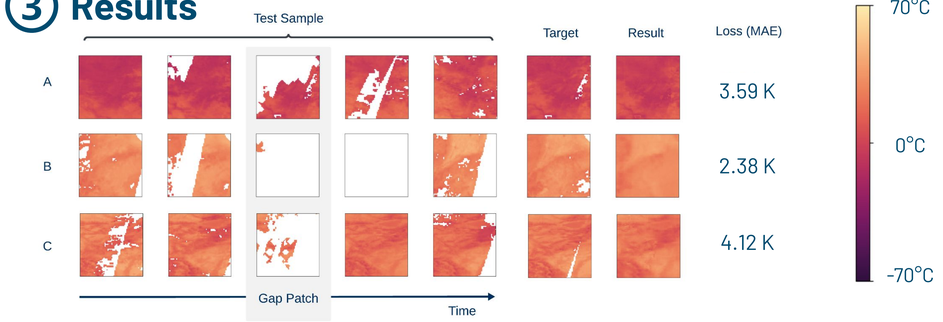
3D CNN
Strided Convolutions ⬇
Transposed Convolutions ⬆
Gap Encoding as Mean over Input

Masked Loss Function

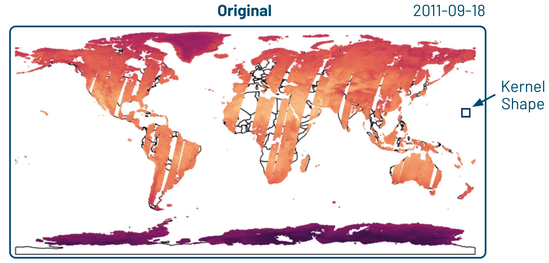
$$Masked\ MAE = \frac{\sum_{i=1}^{64}\sum_{j=1}^{64} |target_{i,j} - pred_{i,j}| \ \forall\ (i,j) \in \blacksquare}{|\blacksquare|}$$

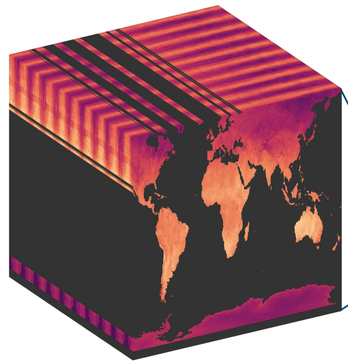
- Introduces the ability of ignoring selected areas in the loss computation
- Needed because of incomplete targets and measurements in the gap patches
- The mask is computed as the areas that are **not provided in the gap patch** but are **provided in the target patch**

③ Results

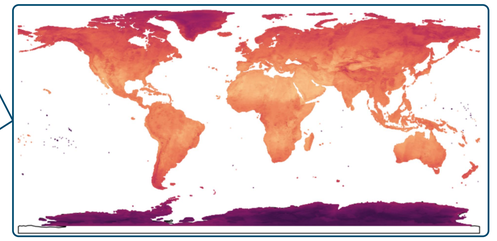


Variable:	Land Surface Temperature
Time Span:	2002-05-19 - 2011-12-23
Kernel Shape:	(5, 64, 64)
Data Threshold:	90%
Training Samples:	428,483

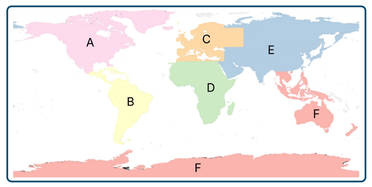




Gap-Filling Solution



SLOO Cross Validation



Masked MAE (K)

Test Set	AR6	Test Samples	Training	Validation	Test
A	0-6	93,613	3.28	3.32	3.86
B	7-15	16,582	3.02	3.08	2.69
C	16-19	39,282	3.39	3.47	3.58
D	20-27	40,814	3.46	3.51	2.89
E	28-37	162,466	3.19	3.31	3.77
F	38-45	75,675	3.07	3.11	2.79

Acknowledgments

We thank the European Space Agency (ESA) for funding the DeepESDL project and acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research 'Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig', project identification number: ScaDS.AI

This presentation participates in OSPP
EGU
Outstanding Student & PhD candidate Presentation contest

EGU Abstract

charly.zimmer
@uni-leipzig.de