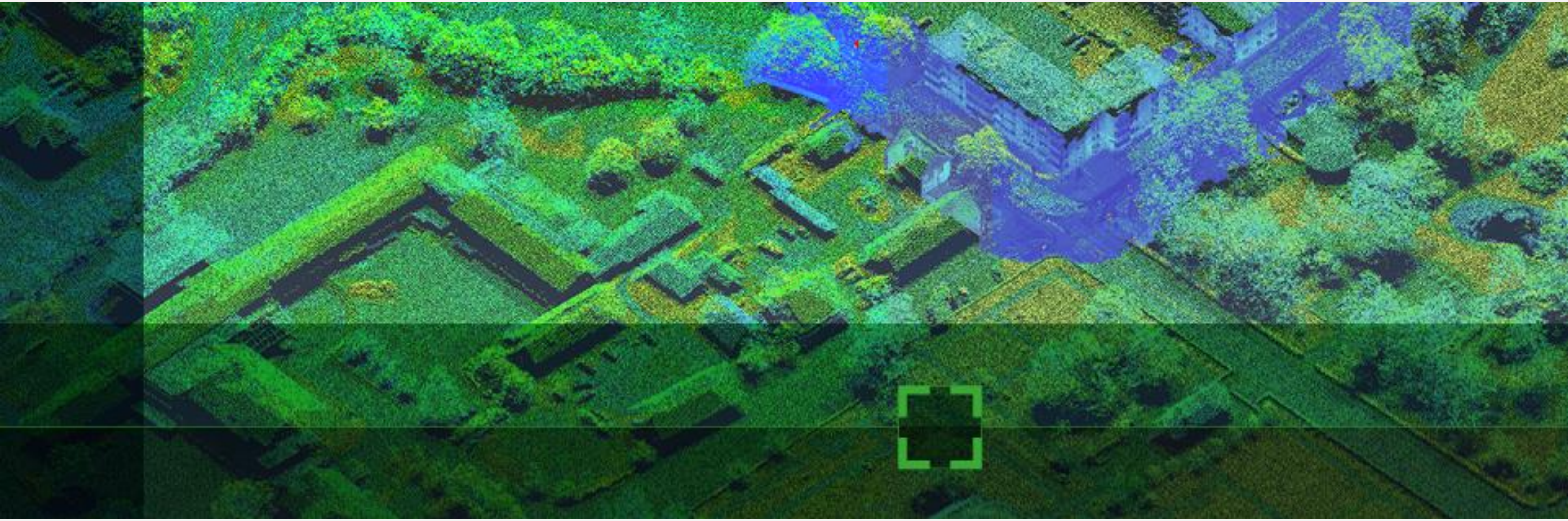
Automatically Find New Buildings - A Pilot Project Using Sentinel-1

Torgeir Ferdinand Klingenberg, Norwegian Mapping Authority

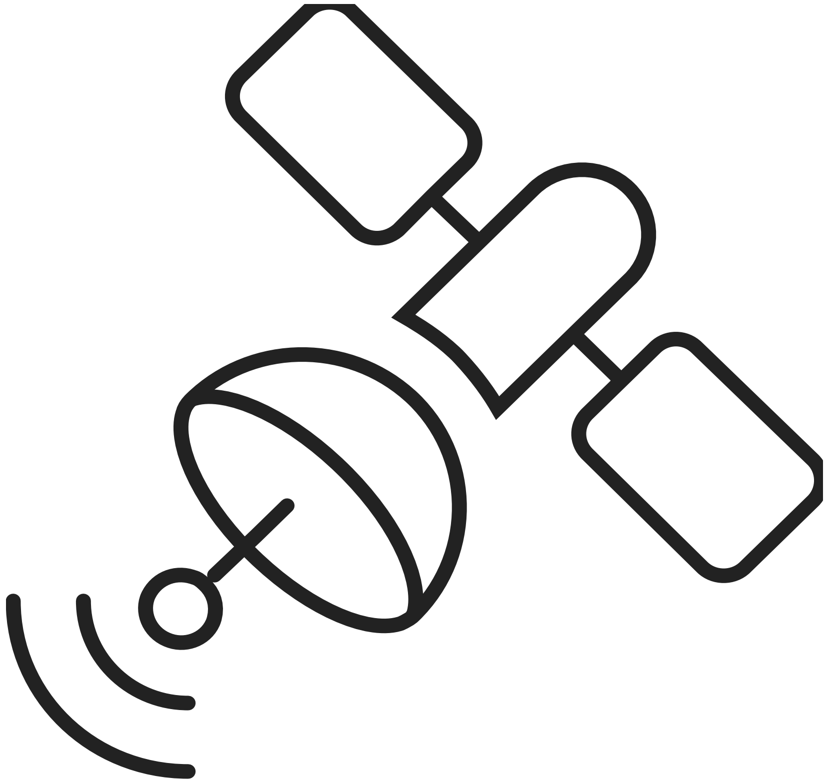


Kartverket

11.01.2024

Table of content

- Introduction
- Data
- Results
- On going work: Machine learning

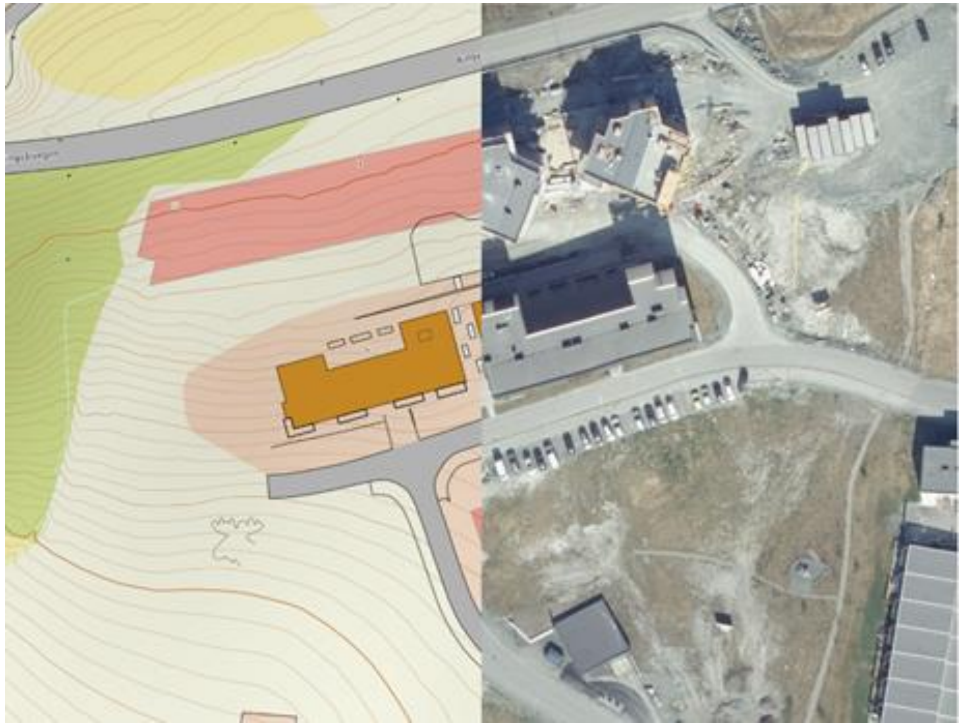


Kartverket

Introduction

Need and purpose

- Find new buildings/building changes
- More predictable cost estimates
- More updated database
- Useful for municipalities and others who want to detect buildings and building changes - e.g., illegal construction



Founding and partners

- In the framework of FPCUP, the following four types of action are supported:
  - National and multinational information/training events
  - To build an active dialogue
  - Promote national and multinational innovative initiatives
  - Development and piloting of downstream applications and services

Copernicus
Framework Partnership Agreement for Copernicus User Uptake

https://www.copernicus-user-uptake.eu/

Founding and partners

- Norwegian Mapping Authority (NMA) applied for EU/FPCUP funds through Norwegian Space Agency (NOSA), and was approved
  - A pilot project with the aim of investigating whether Sentinel-1 can be used to detect buildings and building changes.

Framework Partnership Agreement for Copernicus User Uptake

1. NOSA has its own contract with DLR for FPCUP
2. NMA contract with NRS
3. NMA contract with NORCE for the development

Norsk Romsenter
Norwegian Space Agency

Kartverket

NORCE

Object of the exercise

The aim is to discover the potential that can be achieved with Sentinel-1.

- Important questions/problems to be answered are:
  - How small buildings and building extensions can be detected?
  - How often is it possible to deliver the target result?

- Delivery
  - Knowledge and methodology necessary to achieve the target result must be transferred to the Norwegian Mapping Authority.
  - A description of the methodology/method and a description of the use of the method must be provided in the form of a manual.

Delivery

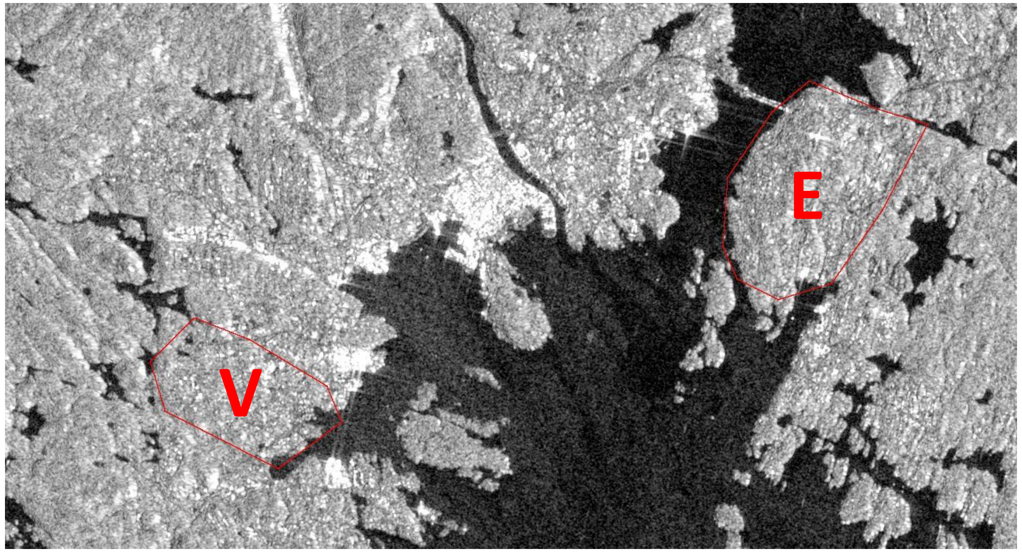
- The target result must be provided as surfaces with reliability targets in the Geography Markup Language (GML) data format.
- The target result must be delivered in the map projection EUREF89 UTM.
- The target result is to be achieved by using data from the Sentinel-1 radar satellites.
- The target result must be achieved by using a sufficient number of Sentinel-1 acquisitions (time series)
- The methodology must be designed so that the target result can at least deliver an annual update. But more frequent deliveries are wanted to be assessed against quality
- The target result must be achievable using open-source code. Necessary code must be published on Github, Gitlab or similar. Where the supplier has trade secrets that prevent the sharing of source code, it is desirable that this solution be used in addition to the open solution.

Data

Data

- Sentinel-1 backscatter 2016-2021 (NORCE)
- Property information (vector) (NMA/Kristiansand municipality)
  - Construction building date

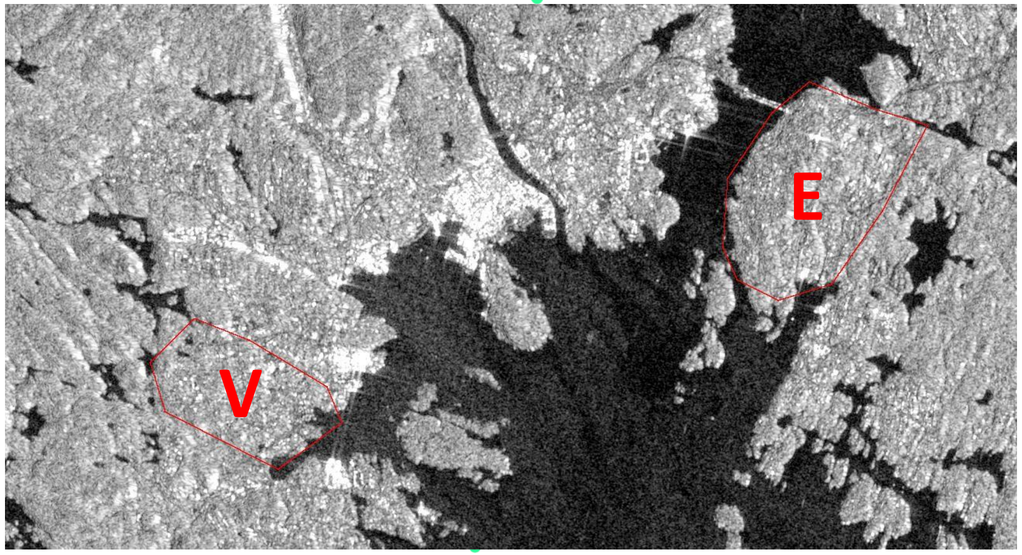
- Additional dataset:
  - Aerial images from country-wide program and Geovekst



Sentinel-1 overview

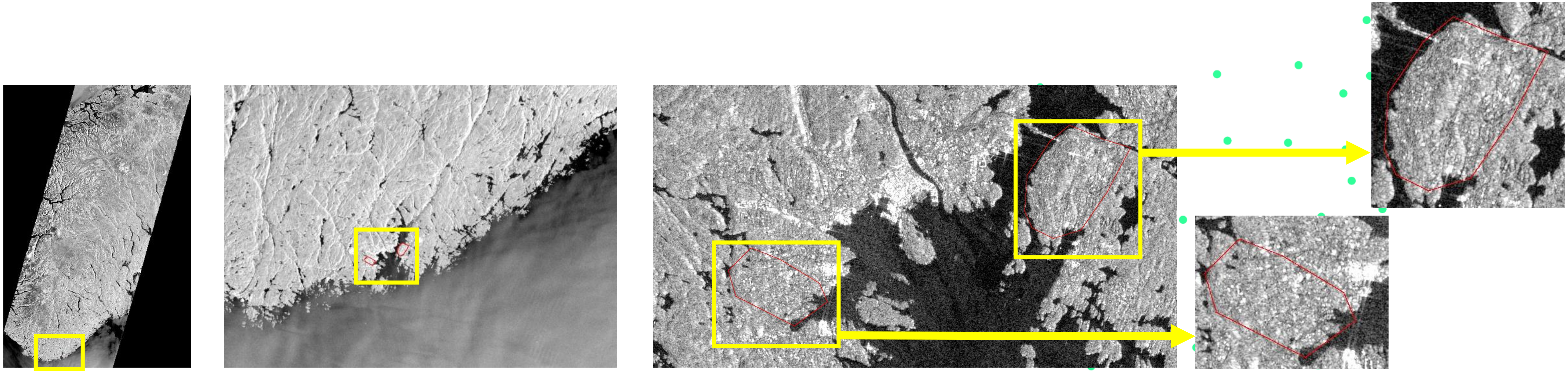
- Has geocoded data stack 1.1.2016-1.1.2022 (6 years) within AOI=[6 462 000,6 469 000, 82 000,95 000] (UTM WGS-84, z33) 10m pixel -> 1300x700 pixel
- Geometries 2016-2021
  - 015ASC covers western area only 223
  - 037DES OK 316
  - 044ASC OK 309
  - 110DES covers western area only 318
  - 117DES OK 182
  - 139DES OK 314
  - Totally 1672

Year	*Scenes/Year*
2016	136
2017	319
2018	295
2019	274
2020	318
2021	318



Processing

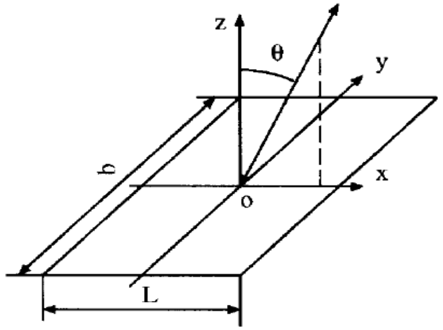
- Splits into 2 areas (smaller stacks, faster processing)
  - AOI_V= [83 880,86 330,6 463 000,6 465 000] -245x200 pixel
  - AOI_E= [91 220, 93850, 6 465 100,6 468 000] - 263x290 pixel
- Clipping:

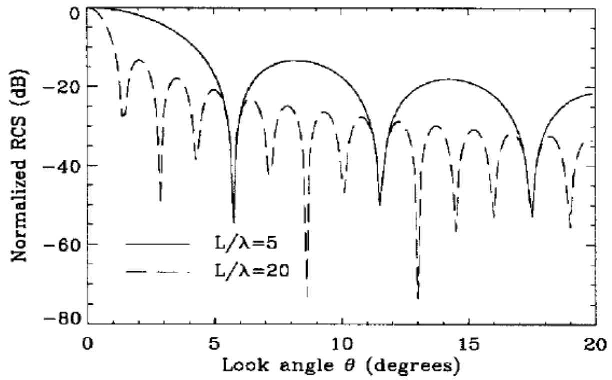


Radar backscatter (teory)

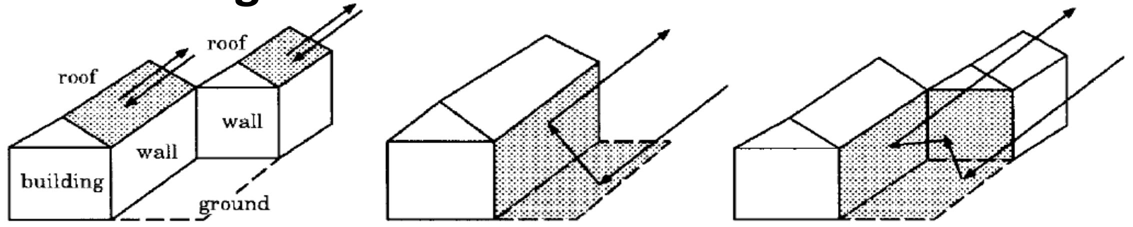
- Backscatter from various objects is generally difficult to model, and will depend heavily on the angle of incidence and direction, as well as the shape and orientation of the building
- Possibilities for single, double and triple bounce effects where co-pole and cross-pole can behave differently.

Flat plain





Building



(a) Single Bounce Scattering (b) Double Bounce Scattering (c) Triple Bounce Scattering

Dong, Y., Forster, B., & Ticehurst, C. (1997). Radar backscatter analysis for urban environments. *International journal of remote sensing*, *18*(6), 1351-1364.

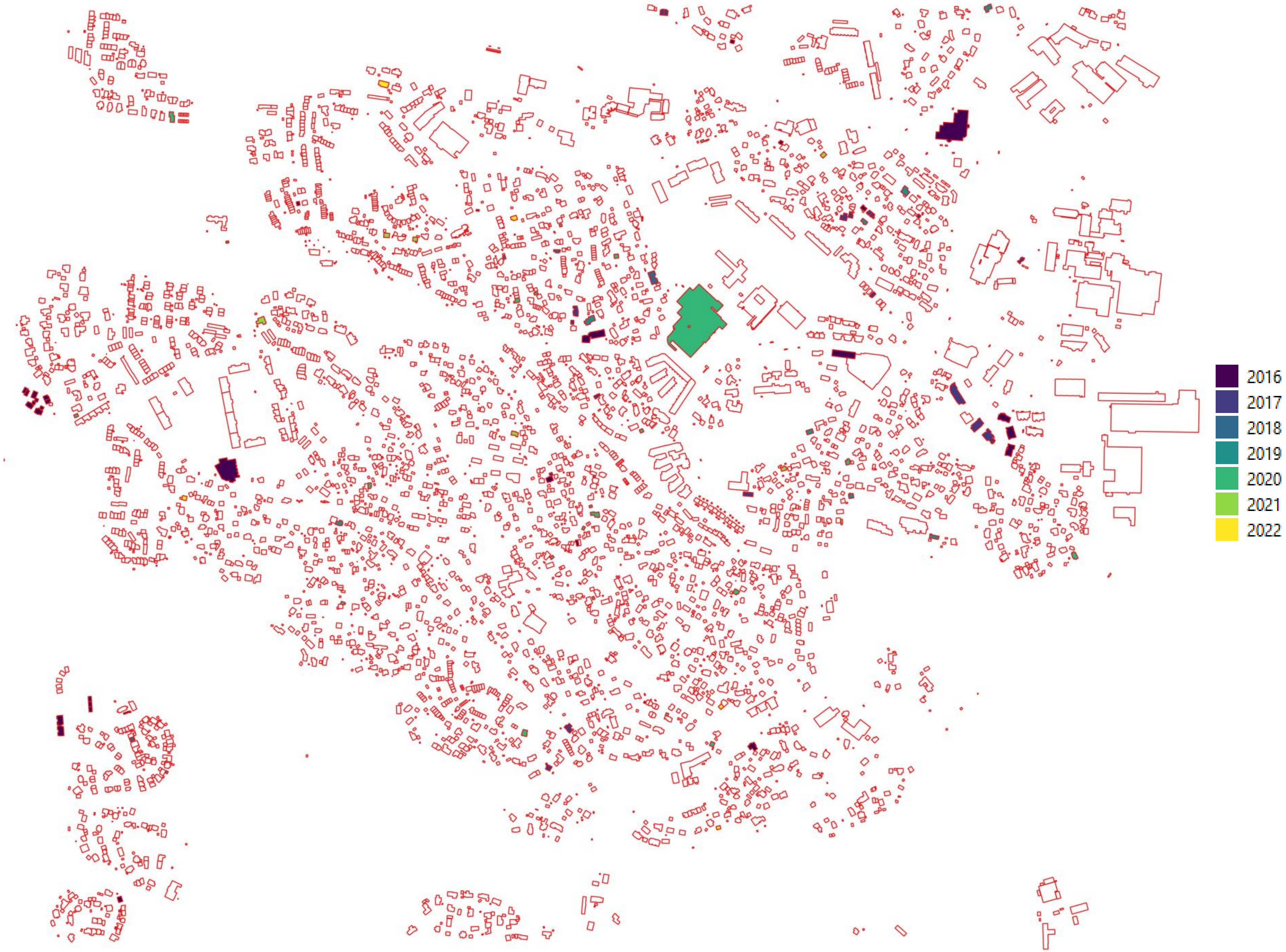
Original

East

Kartverket

Original

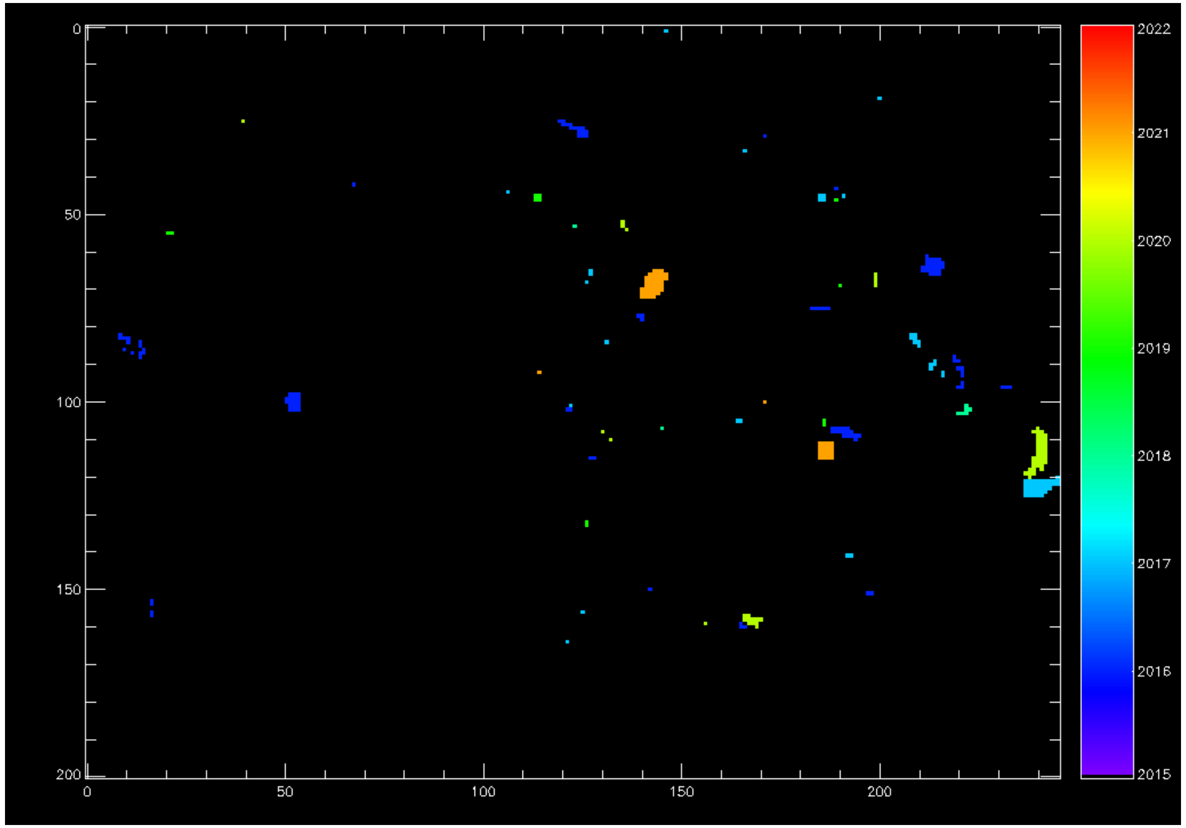
West



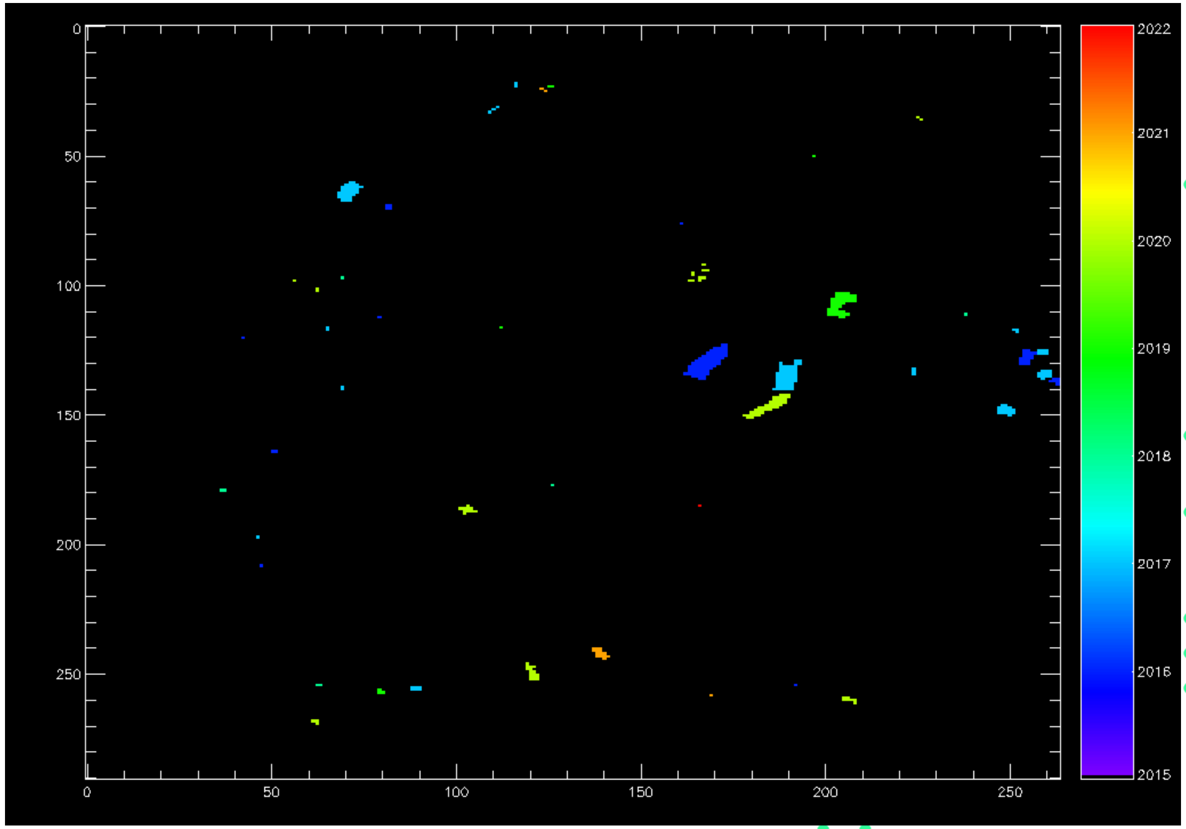
Kartverket

Ground truth – areas with visible change

West (354 pixels)



East (566 pixels)

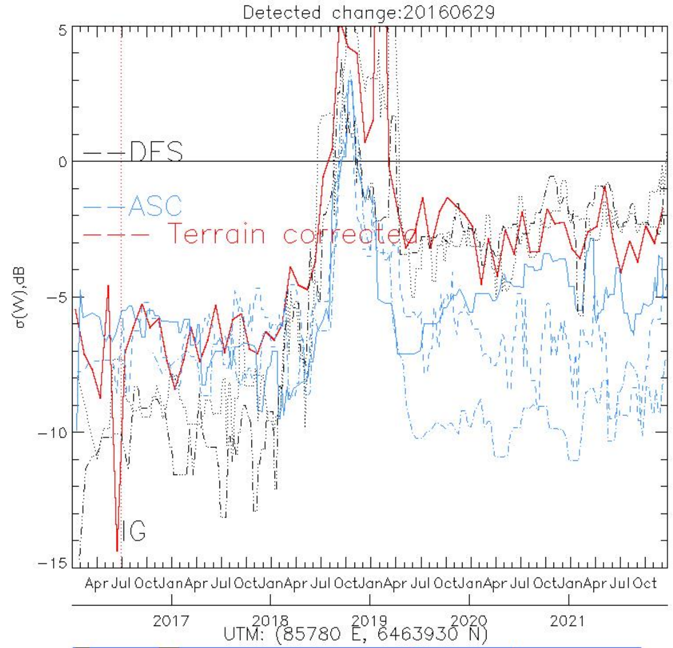


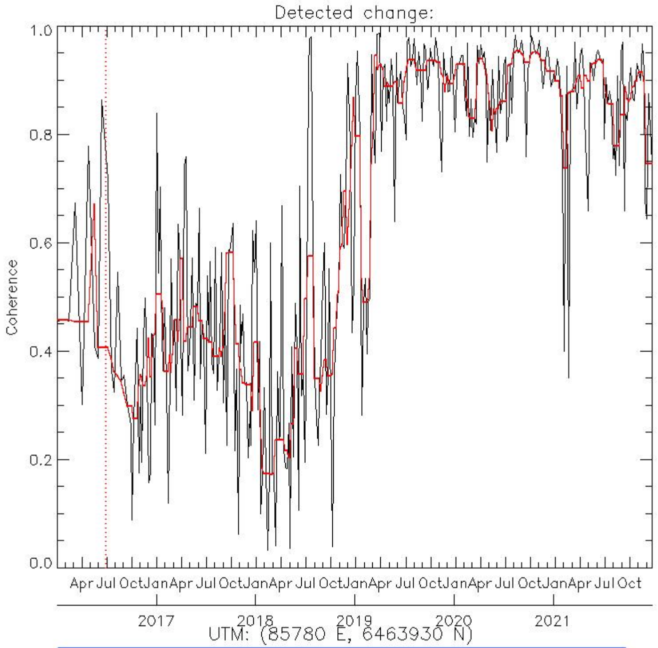
Kartverket

Results

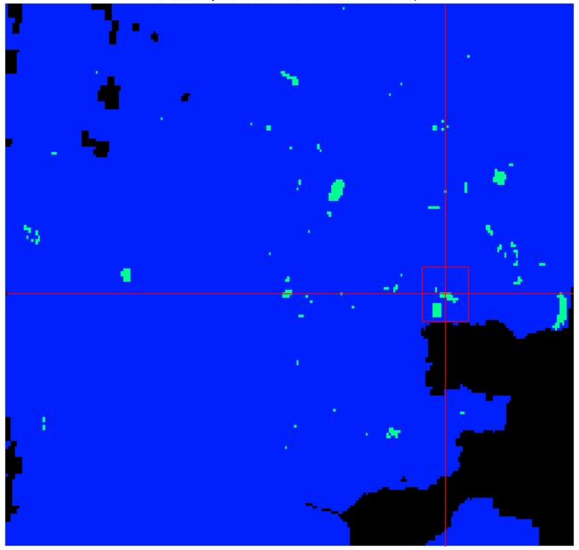
Kartverket

Demolished building and new building

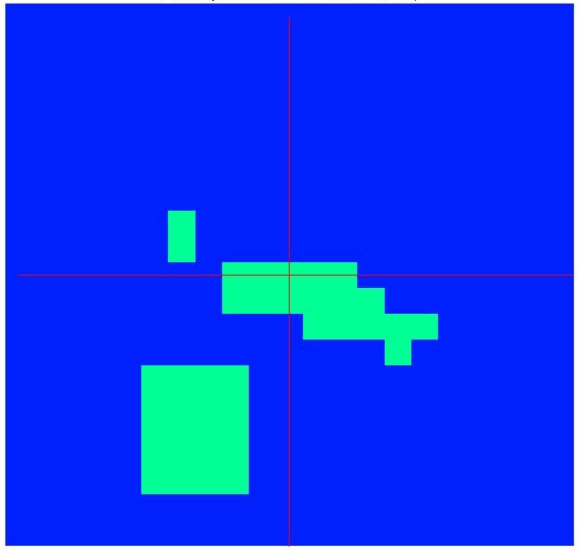




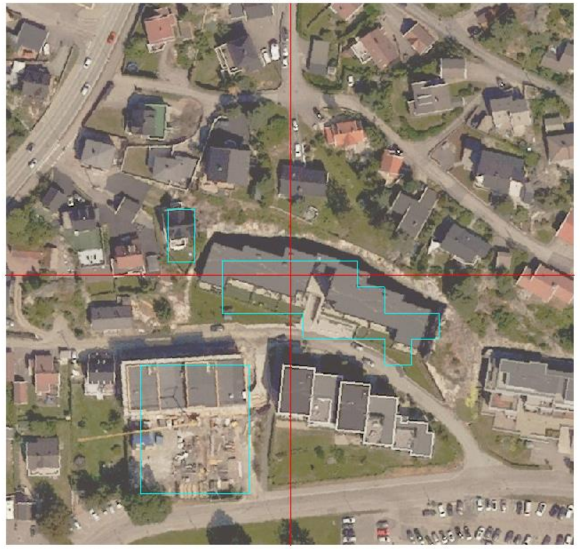
UTM: (85780 E, 6463930 N)



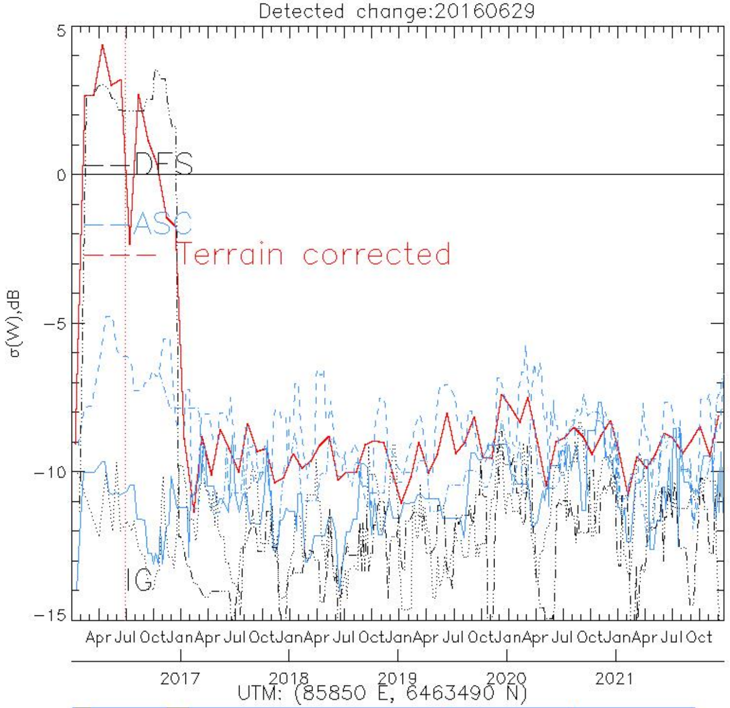
UTM: (85780 E, 6463930 N)

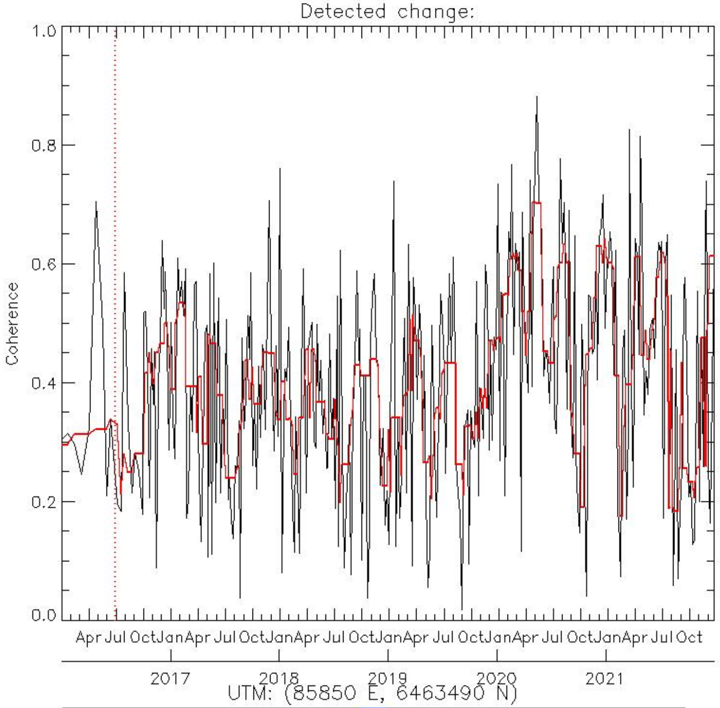


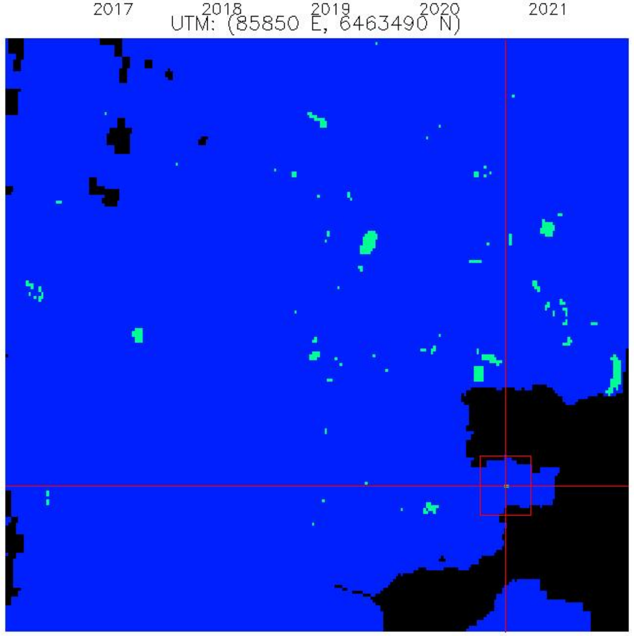
NiB, 2021

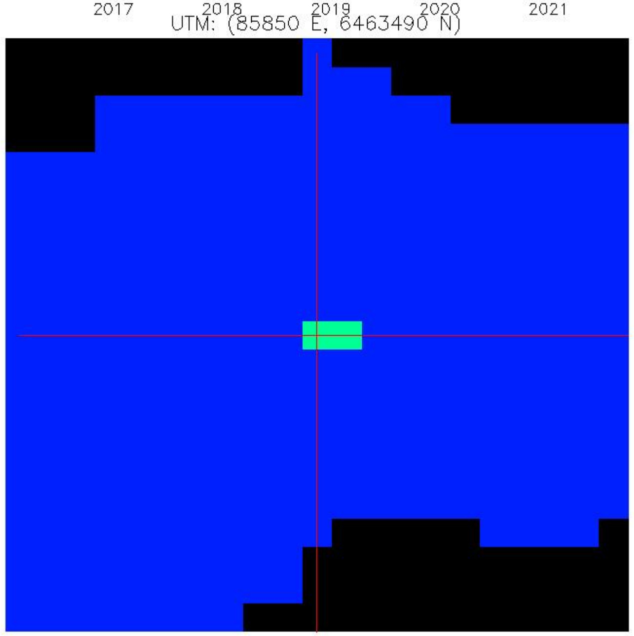


Demolished building

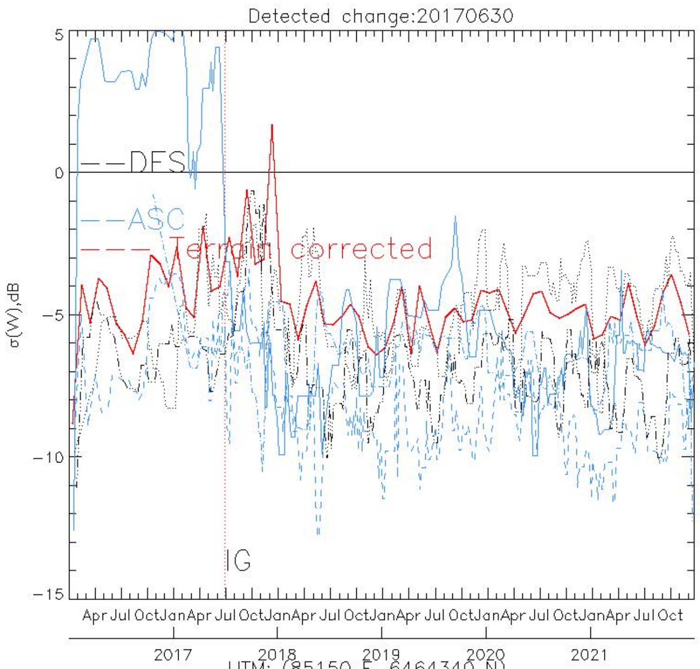


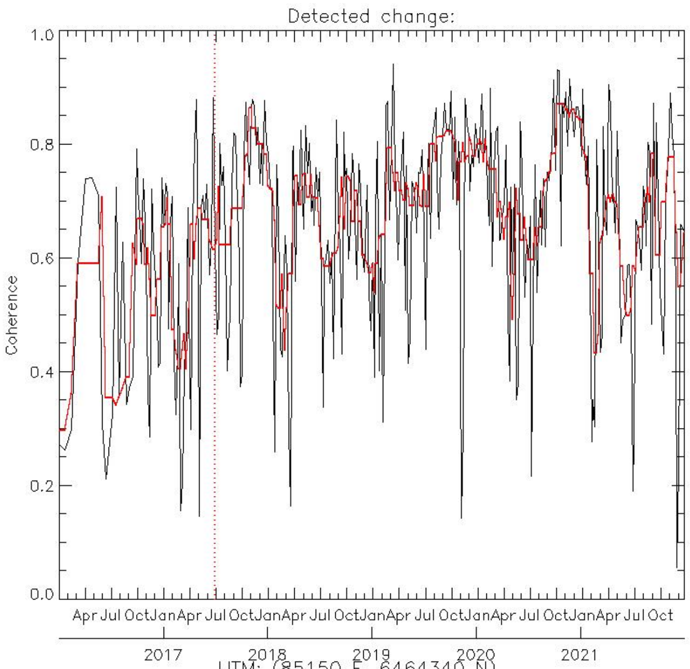


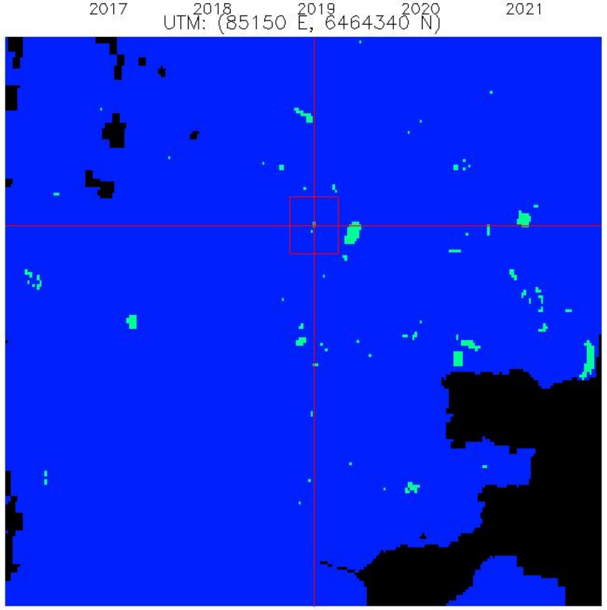


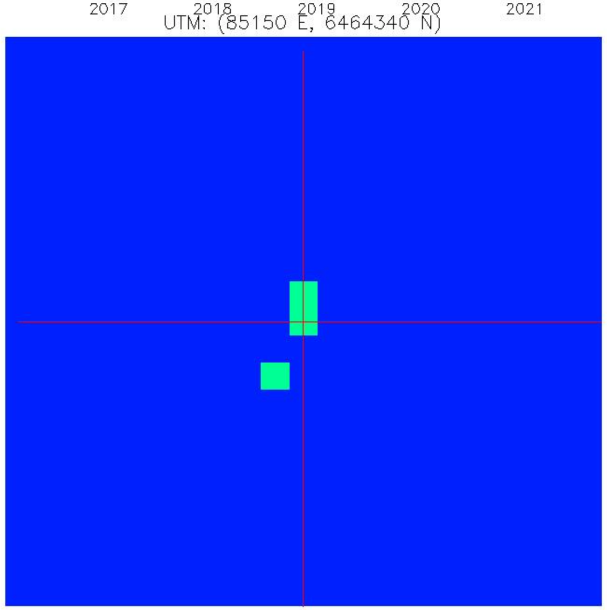


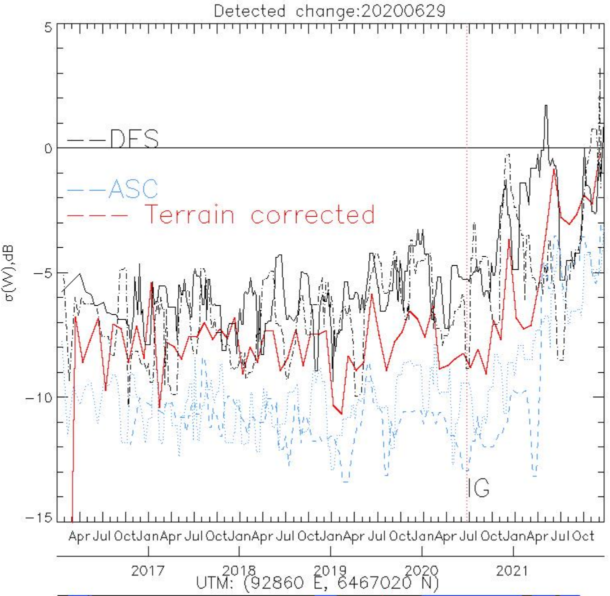
Building change

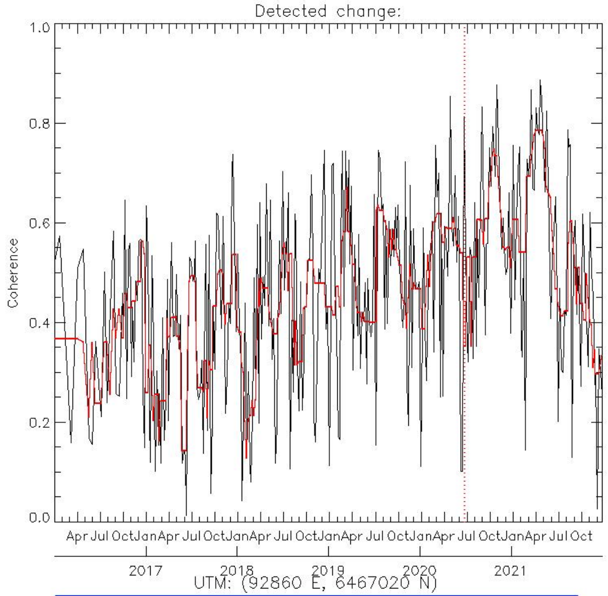


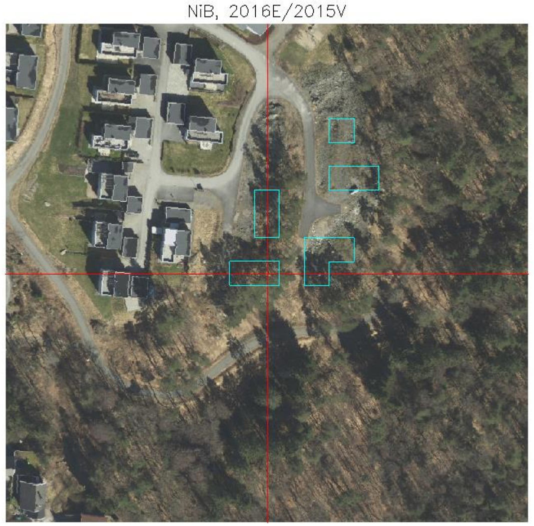


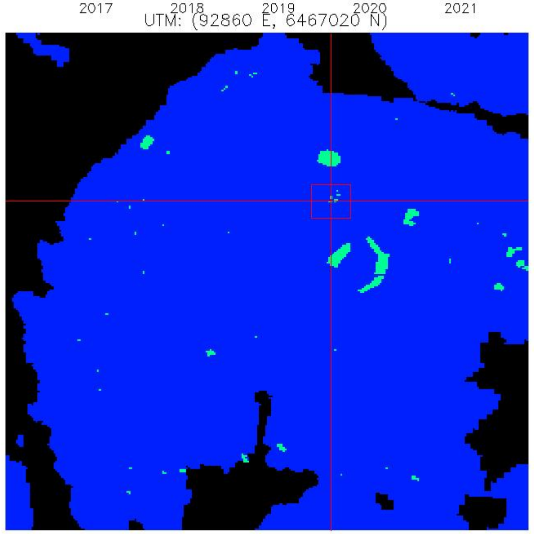


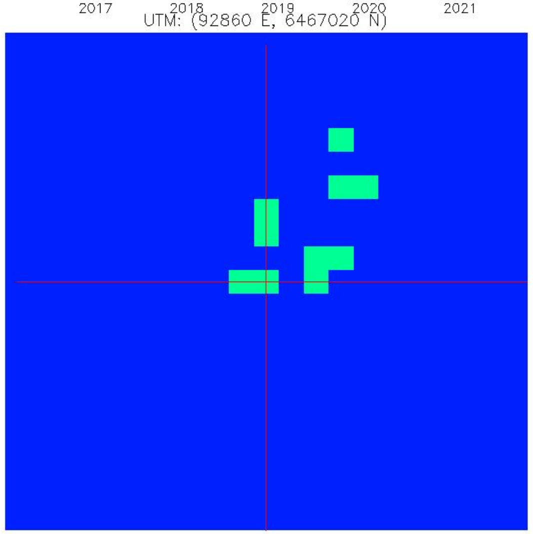



Detected change:20200629
--DES
--ASC
--- Terrain corrected
σ(V),dB
IG
Apr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul Oct
2017 2018 2019 2020 2021


Detected change:
Coherence
Apr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul Oct
2017 2018 2019 2020 2021

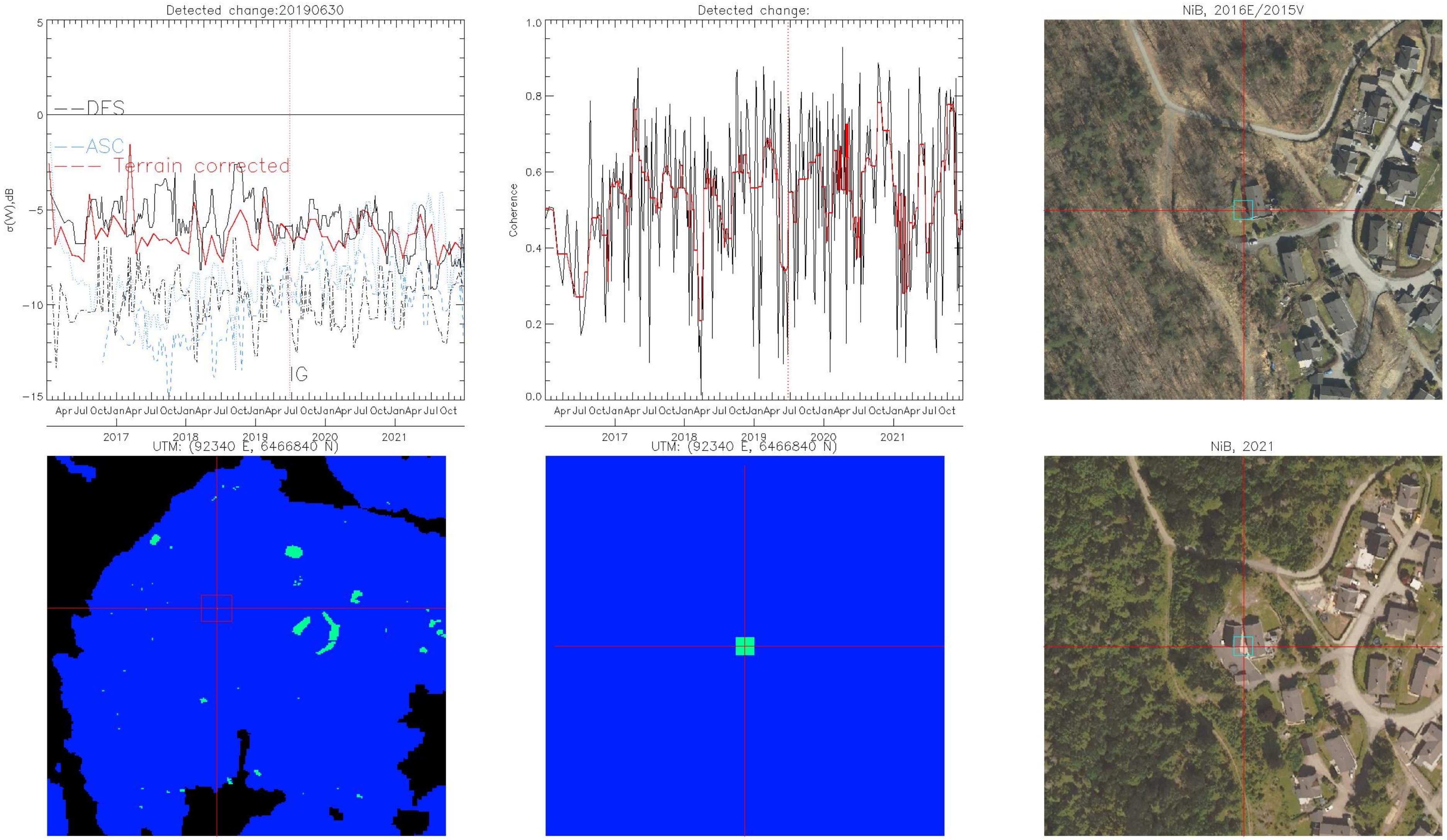

NiB, 2016E/2015V

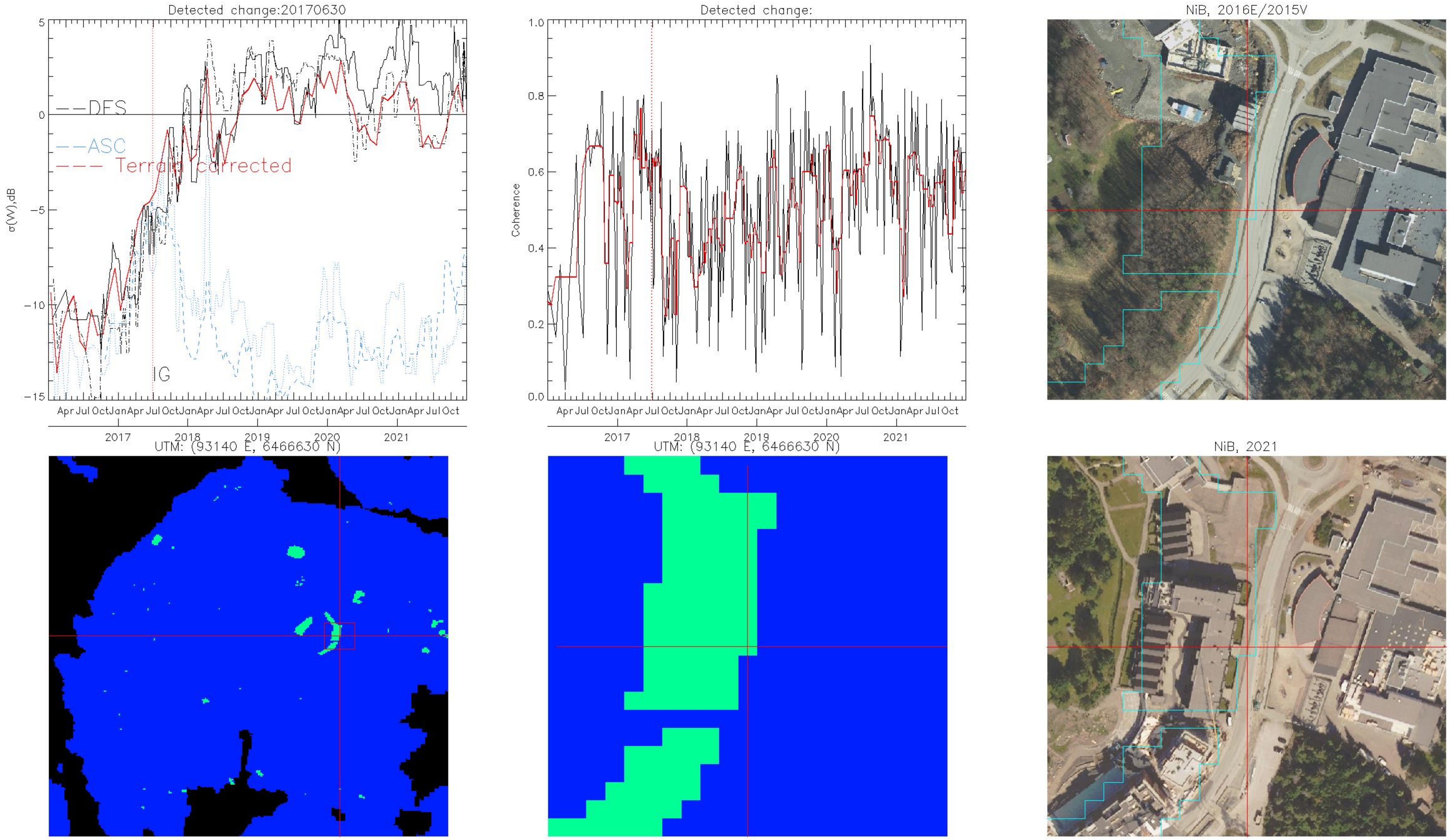

UTM: (92860 E, 6467020 N)


UTM: (92860 E, 6467020 N)


NiB, 2021

Difficult


Detected change:20190630
5
0
−5
−10
−15
DES
ASC
Terrain corrected
σ(W),dB
IG
Apr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul OctJanApr Jul Oct
2017
2018
2019
2020
2021
Detected change:
1.0
0.8
0.6
0.4
0.2
0.0
Coherence
NiB, 2016E/2015V
UTM: (92340 E, 6466840 N)
UTM: (92340 E, 6466840 N)
NiB, 2021



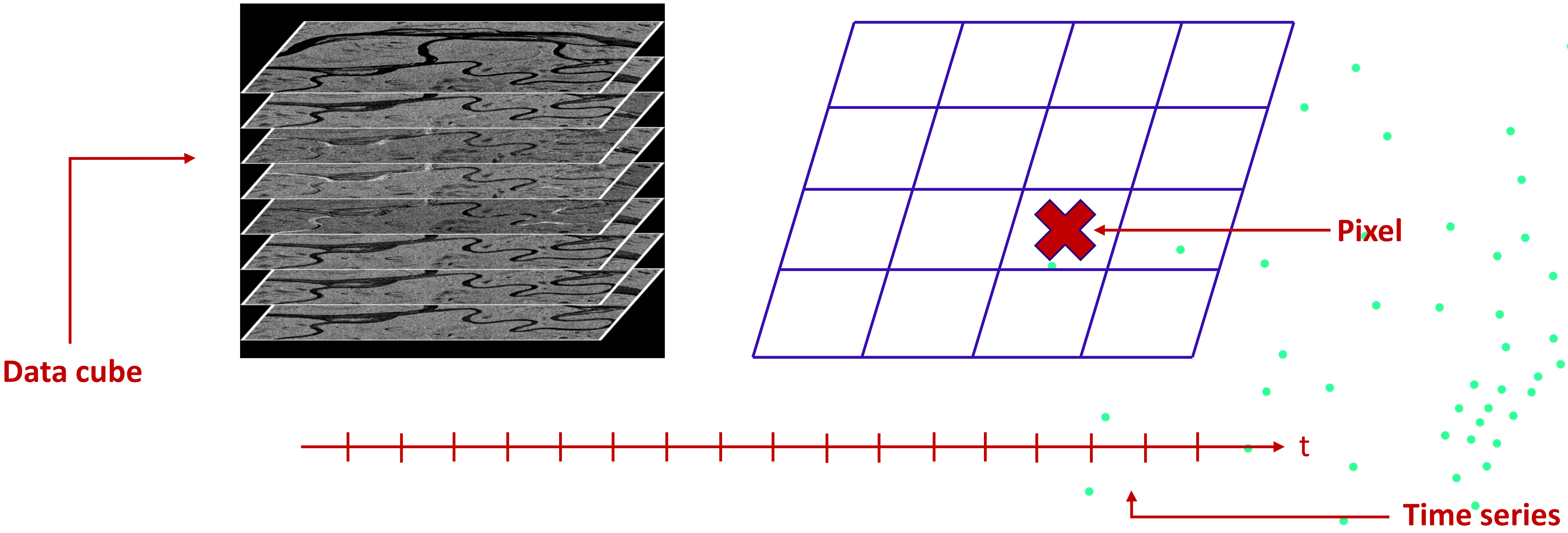
Detected change:20170630
DES
ASC
Terrain corrected
σ(W),dB
IG
Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct
2017 2018 2019 2020 2021
Detected change:
Coherence
Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct Jan Apr Jul Oct
2017 2018 2019 2020 2021
NiB, 2016E/2015V
UTM: (93140 E, 6466630 N)
UTM: (93140 E, 6466630 N)
NiB, 2021

Detection of building changes in SAR image data

A machine learning perspective

Data

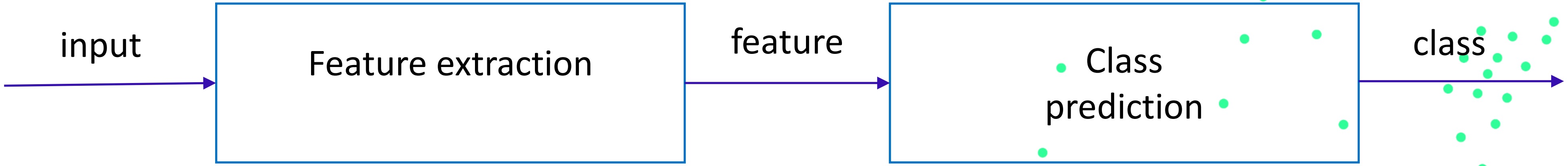
- Satellite images are resampled to a common grid
- Gives a data cube where each pixel is represented by a time series



Problem definition and terminology

Change detection = binary classification

- Two classes: «change» and «no change»
- Classification = feature selection + class prediction
  - Input: time series
  - Feature: numbers that characterise the object to be classified
  - Feature selection: transformation of time series (input data) into an alternative representation used as input to the class prediction
  - Time series classification requires specialised feature extraction!



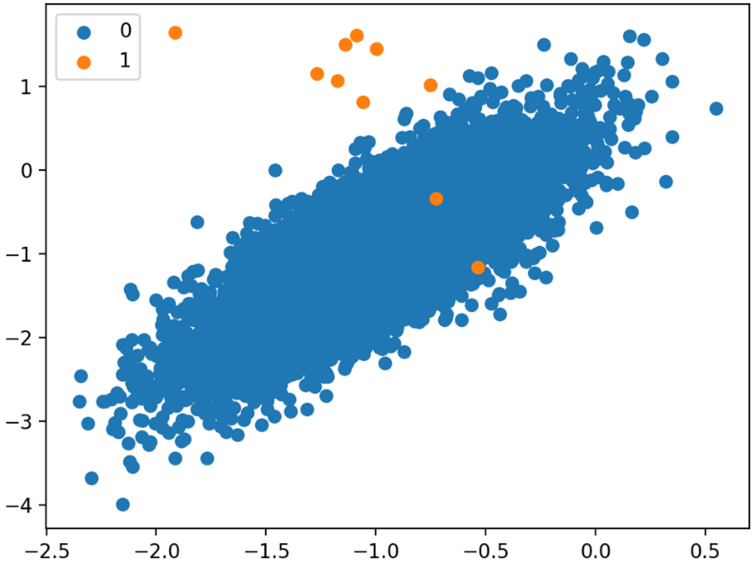
Challenges

NORCE

- Imbalanced classes
- Time series with different lengths
- Time series with irregular sampling
- Noisy class labels (uncertain location of change in time and space)
- Heterogeneous input data
- Use of contextual information

Imbalanced classes

- Few pixels with change, many without



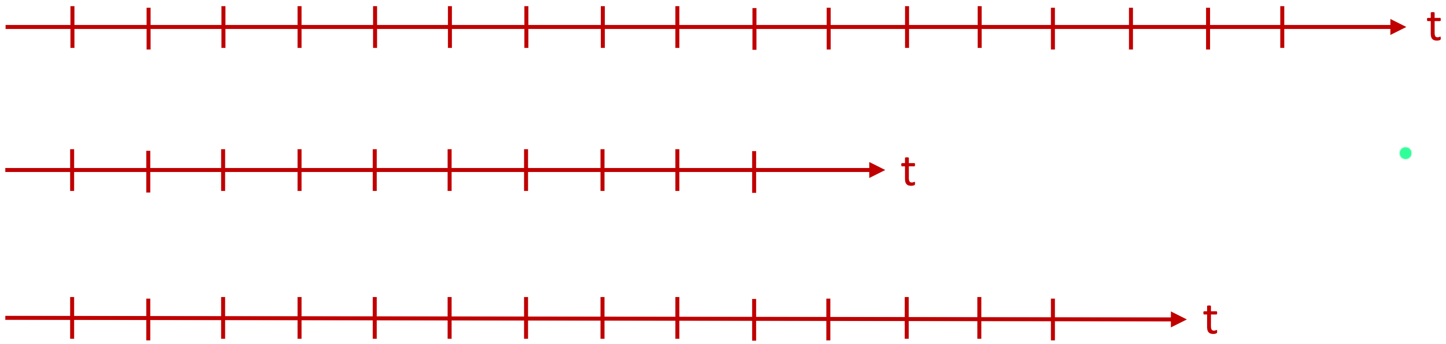
⇒ Changes are underrepresented in training dataset

⇒ Ineffective training of the machine learning (ML) model

- Can be handled by standard methods, such as class weighting
(balancing the proportion of change and unchanged pixels presented to the model during training by uneven sampling)

Time series with different lengths

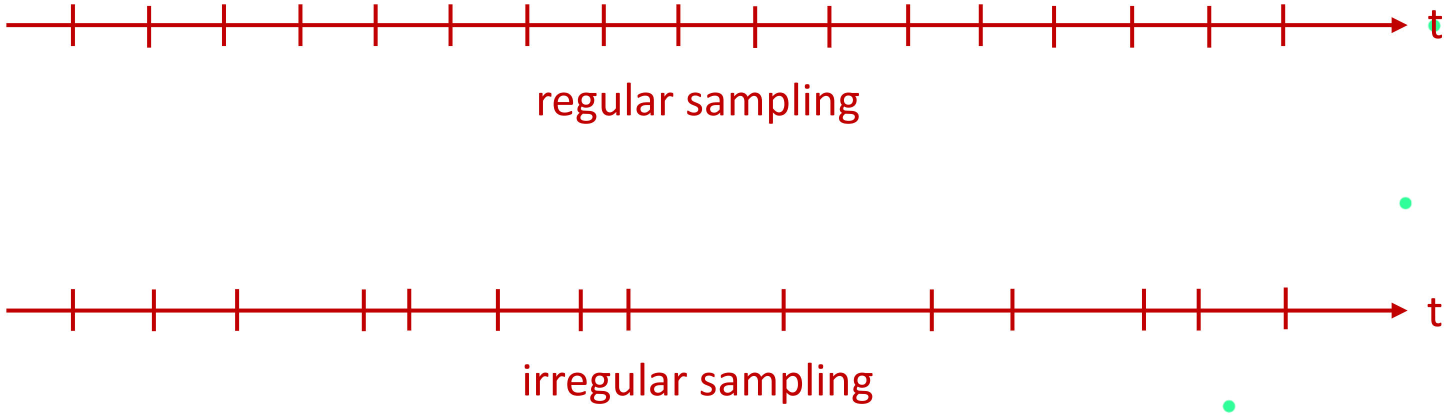
- Varying number of observations for different locations and years



- ML model must handle input time series with different lengths!

Time series with irregular sampling

- Available observations may be sampled at irregular intervals



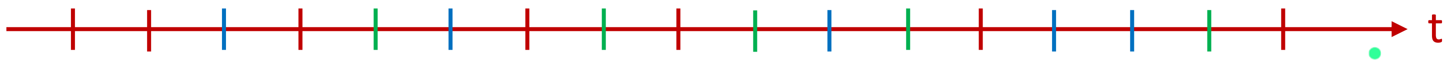
- ML model must handle irregular sampling

Noisy class labels

- Uncertainty in space
  - Location of changes in satellite data may not correspond with polygons of buildings that are reported as subject to change
- Uncertainty in time
  - Onset of visible changes in satellite data may not correspond with reported starting time of construction work
- Machine learning term: *noisy labels*
- The label noise is *not random*, which makes is *more difficult* to handle from a machine learning point of view

Heterogeneous input data

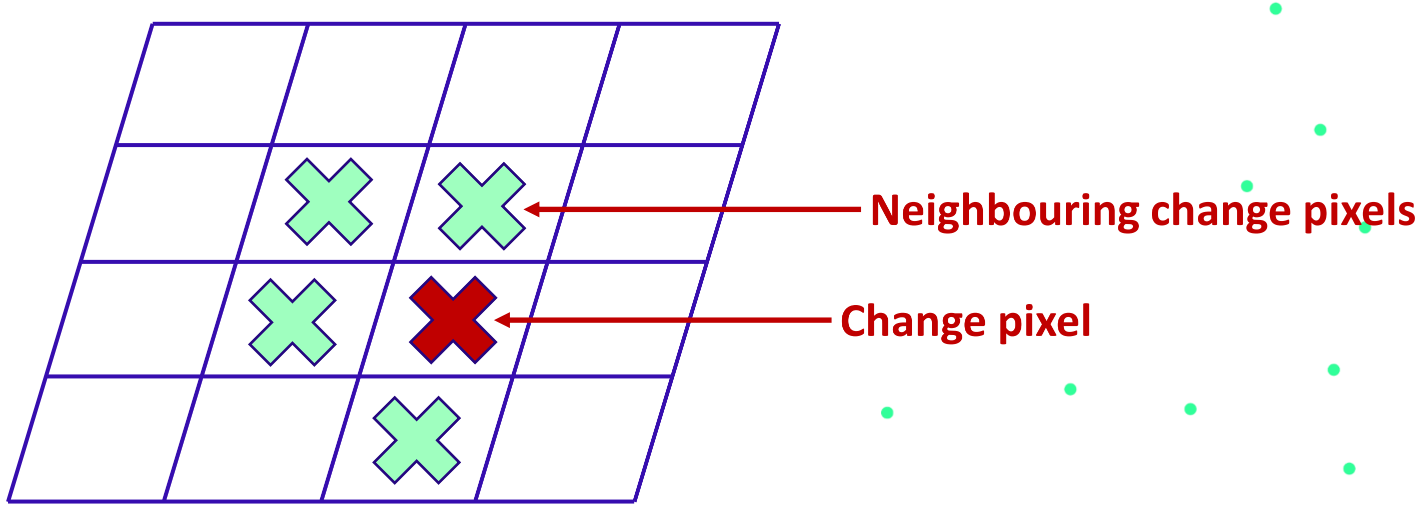
- Dataset contains observations of different variables:
  - Radar backscatter intensity
    - From different polarisations (property of electromagnetic wave)
    - From different geometries (radar pointing angles)
  - Coherence (correlation between radar measurements at two times)
- How to combine different data sources or detection results?



available time series of measurements from different variables

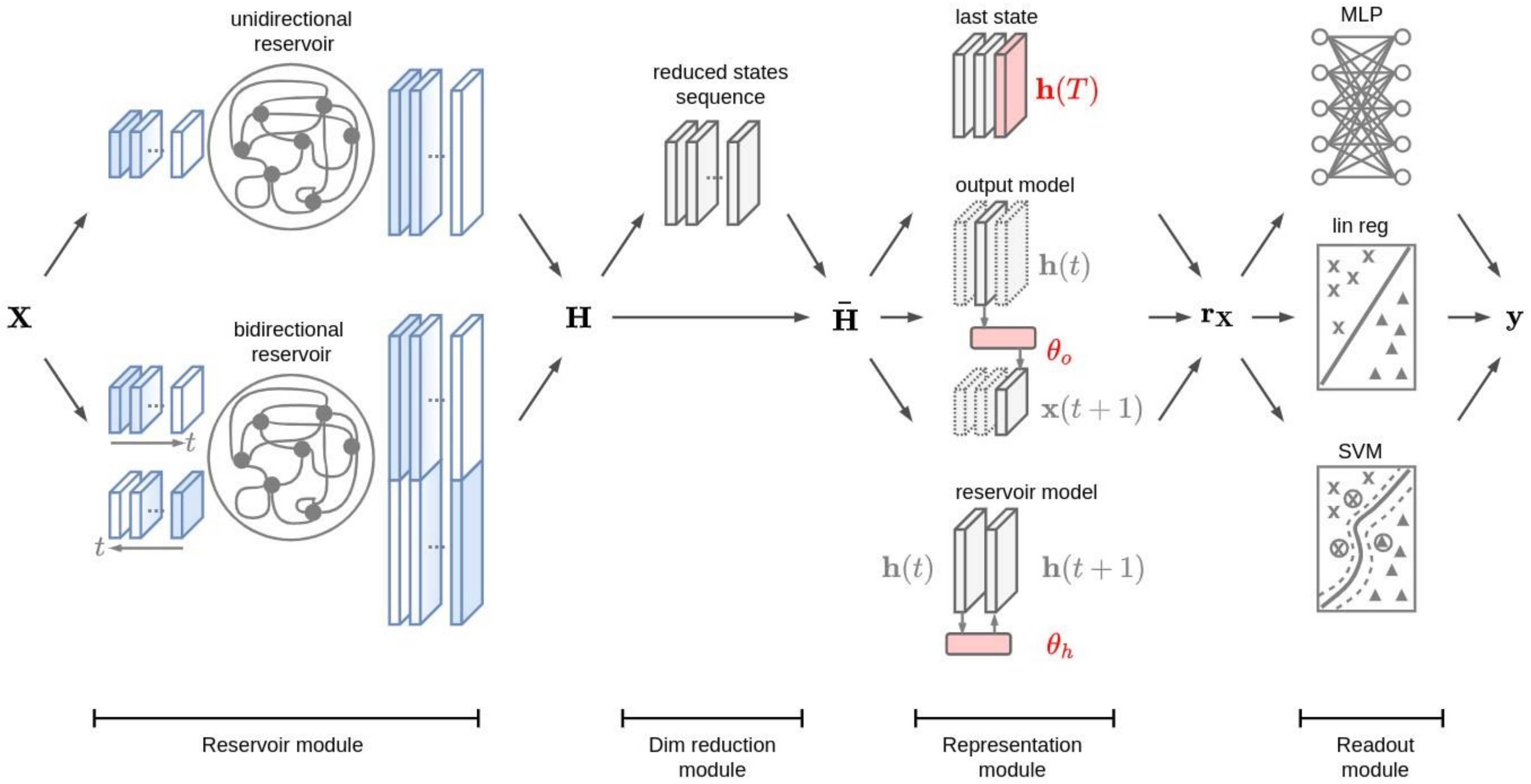
Use of contextual information

- A change pixel is often surrounded by other change pixels



- How to exploit the information from neighbouring pixels to improve the change detection?

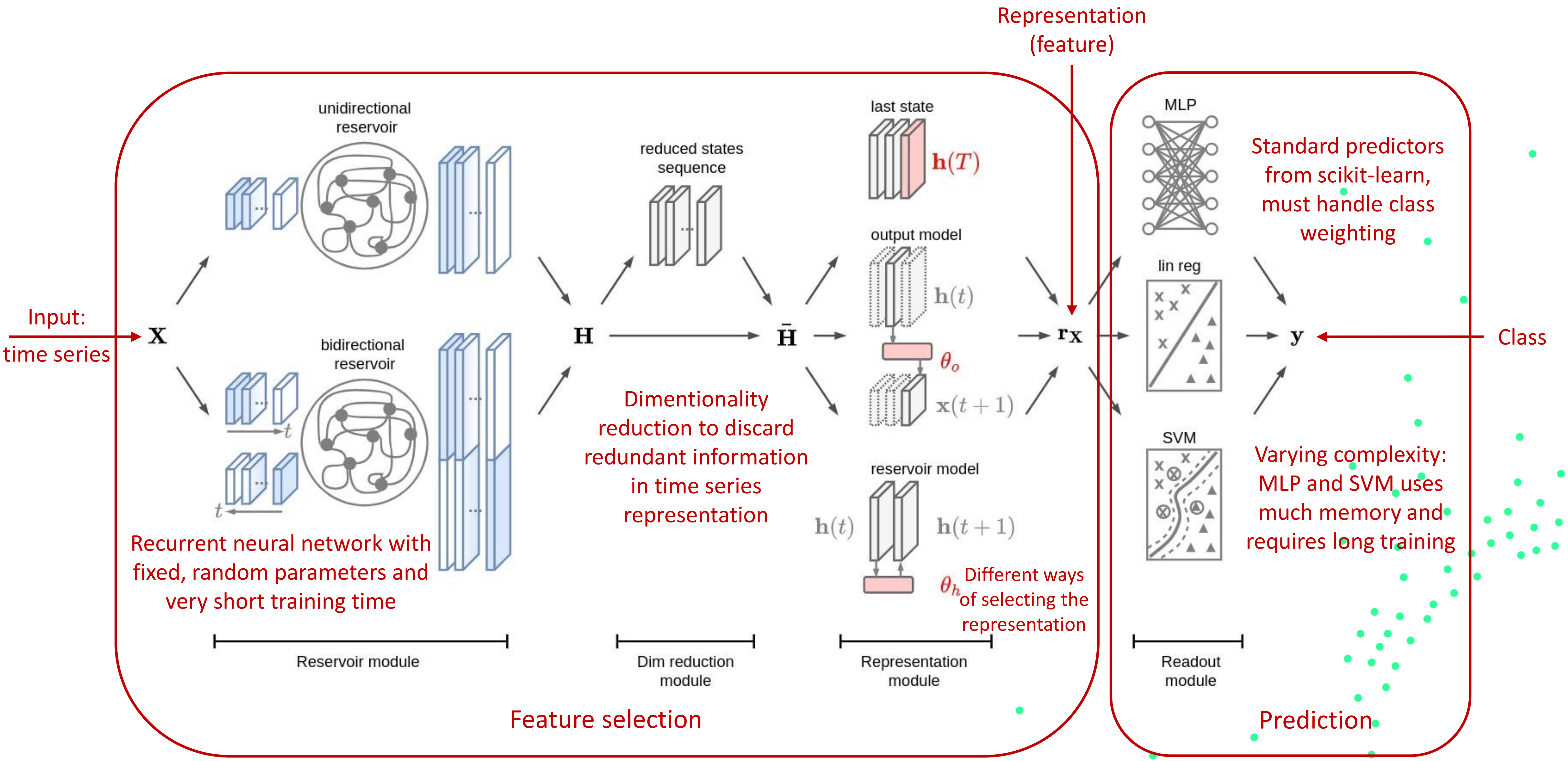
Machine learning model of choice



Machine learning model

- Bianchi et al. (2021): «Reservoir computing approaches for representation and classification of multivariate time series»
  - Uses *reservoir computing (RC)* for feature selection (1st stage):
    - Generates a *reservoir* consisting of a collection of *recurrent neural networks* specialised at extracting information and patterns from sequential data (such as time series)
    - The reservoir has fixed, random parameters that are *not trained*. Yet, the size of the reservoir ensures a rich and efficient *representation* of the time series
    - The length of the representation is independent of input time series length and sampling
    - The RC network requires much less training data and compute time than than alternative deep neural networks for time series processing
  - Can use standard classifiers for class prediction (2nd stage):
    - Linear classifier, support vector machine, multilayer perceptron

Machine learning model



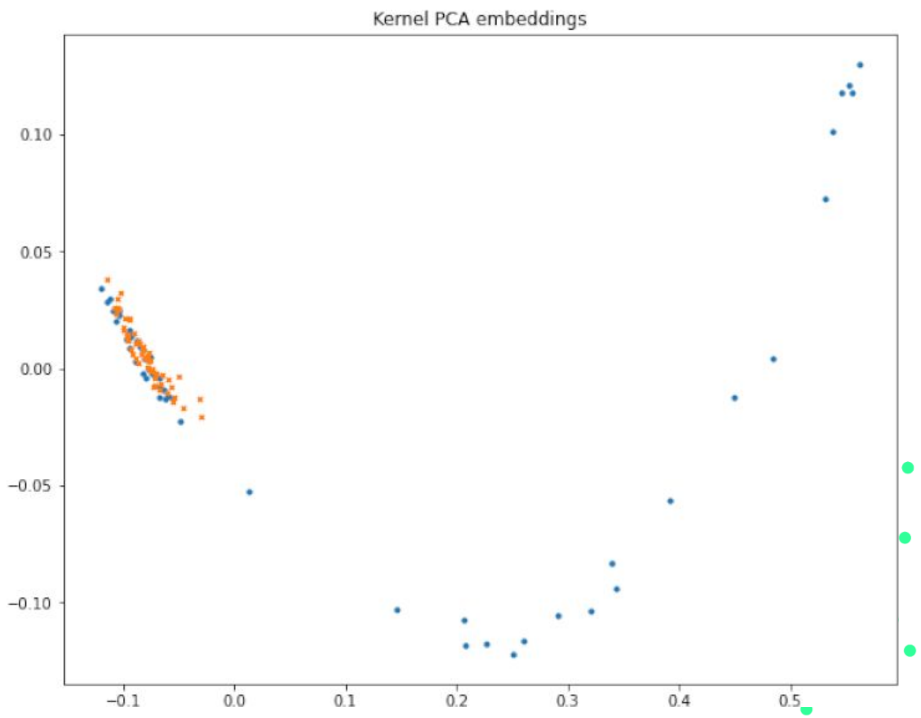
Challenges and solutions

- Imbalanced classes
  - Solved with standard method: class weighting to balance classes
- Time series with different lengths
- Time series with irregulær sampling
  - The RB network provides representations of the time series that are independent of length and sampling
- Noisy class labels
  - Key problem that requires solution
- Heterogene input data
  - Investigating alternatives: combine separate detections on homogeneous data eller detection with complete dataset as input
- Use of contextual information
  - Requires more advanced ML models
    Awaiting better class labels

Critical: Re-labelling of noisy training data

Alternative ways of handling:

- Manual re-labelling (laboreous!)
- Semi-automatic re-labelling
  - Manual re-labelling supported by automated analysis and customised graphical user interface
- Fully automated re-labelling
  - Example from project (figure at the right):
    - Clustering of pixel representations into: «change» and «no-change»
    - Change pixels that group together with no-change pixels can be defined as outliers and re-labelled by an algorithm
    - Our case requires adaption of existing algorithms for re-labelling and handling of noisy labels



Abstract representation of pixels in two dimentions to study clustering/grouping

Questions?

Contact info:

→ Torgeir Ferdinand Klingenberg

→ Torgeir.Ferdinand.Klingenberg@kartverket.no

Contact info NORCE :

→ Eirik Malnes

→ eima@norceresearch.no

Kartverket

Kartverket