

# Dataset information

---

## Contents

- `data.npy` is a `numpy.array` of shape `(samples, sequence length, features)` which represents the complete dataset. This dataset can then be split based on the first dimension, i.e., samples, into train/validation/test sets.  
**samples:** this dimension represents the number of time sequences, i.e. number of pixel-wise time series that have been produced from the stack of Sentinel-1 images.  
**sequence length:** this dimension represents the number of steps per sequence, i.e., the number of dates used to create the time series. The time difference between each step is 6 days.  
**features:** this dimension represents the number of channels, i.e., VV and VH Sentinel-1 polarization channels.
- `labels.py` is a `numpy.array` of shape `(samples,)` which represents the label/target value of each sequence. Each label can take two possible values:  
**0:** means that the sequence represents a non-flooded time series pattern at the last steps (i.e., absence of flood event)  
**1:** means that the sequence represents a flooded time series pattern at the last steps (i.e., presence of flood event).

## Reference

<https://doi.org/10.5281/zenodo.6510223>