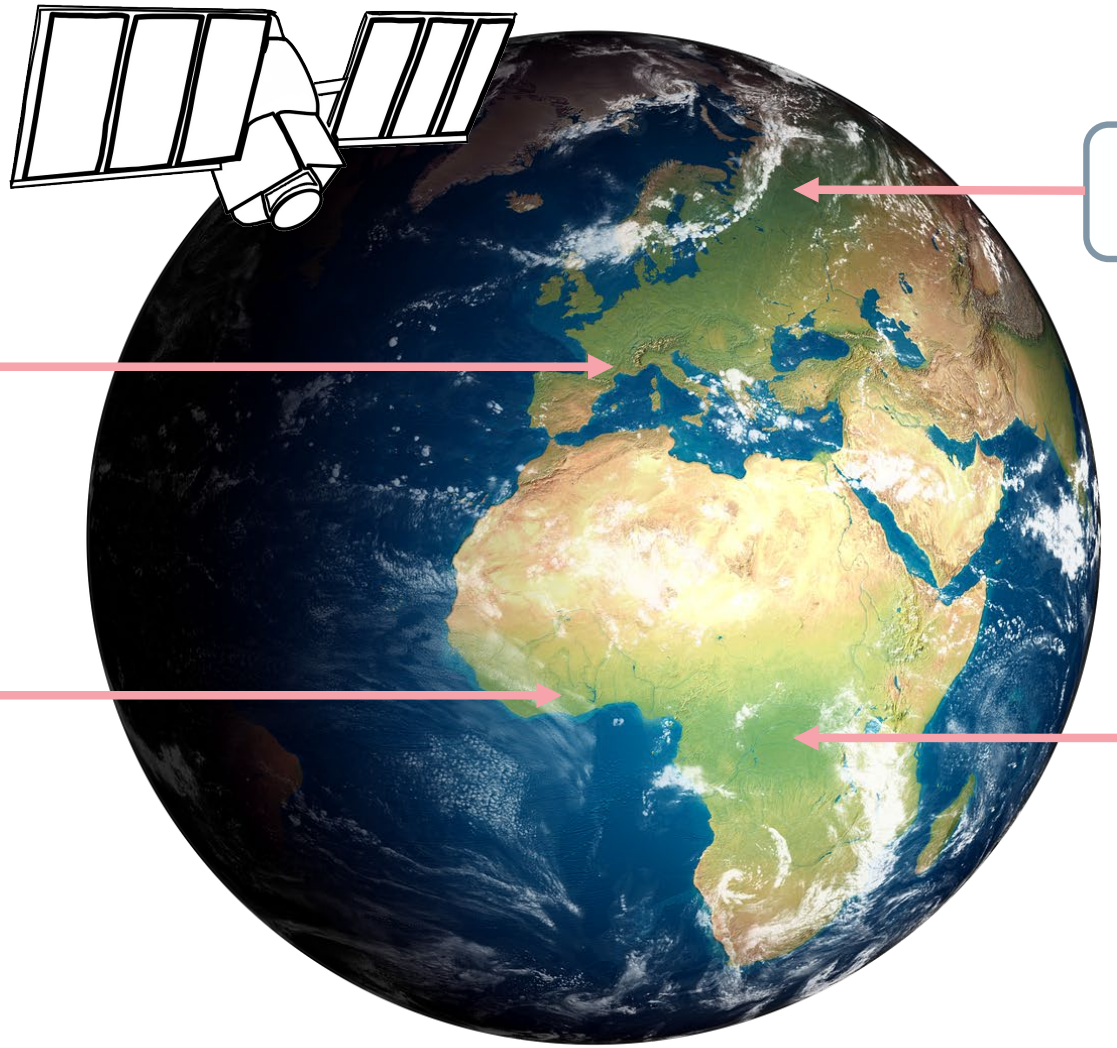
# Benefits and Opportunities of Explainable Machine Learning in the Environmental Sciences

Ribana Roscher

*AI4EO Future Lab, Technical University of Munich*

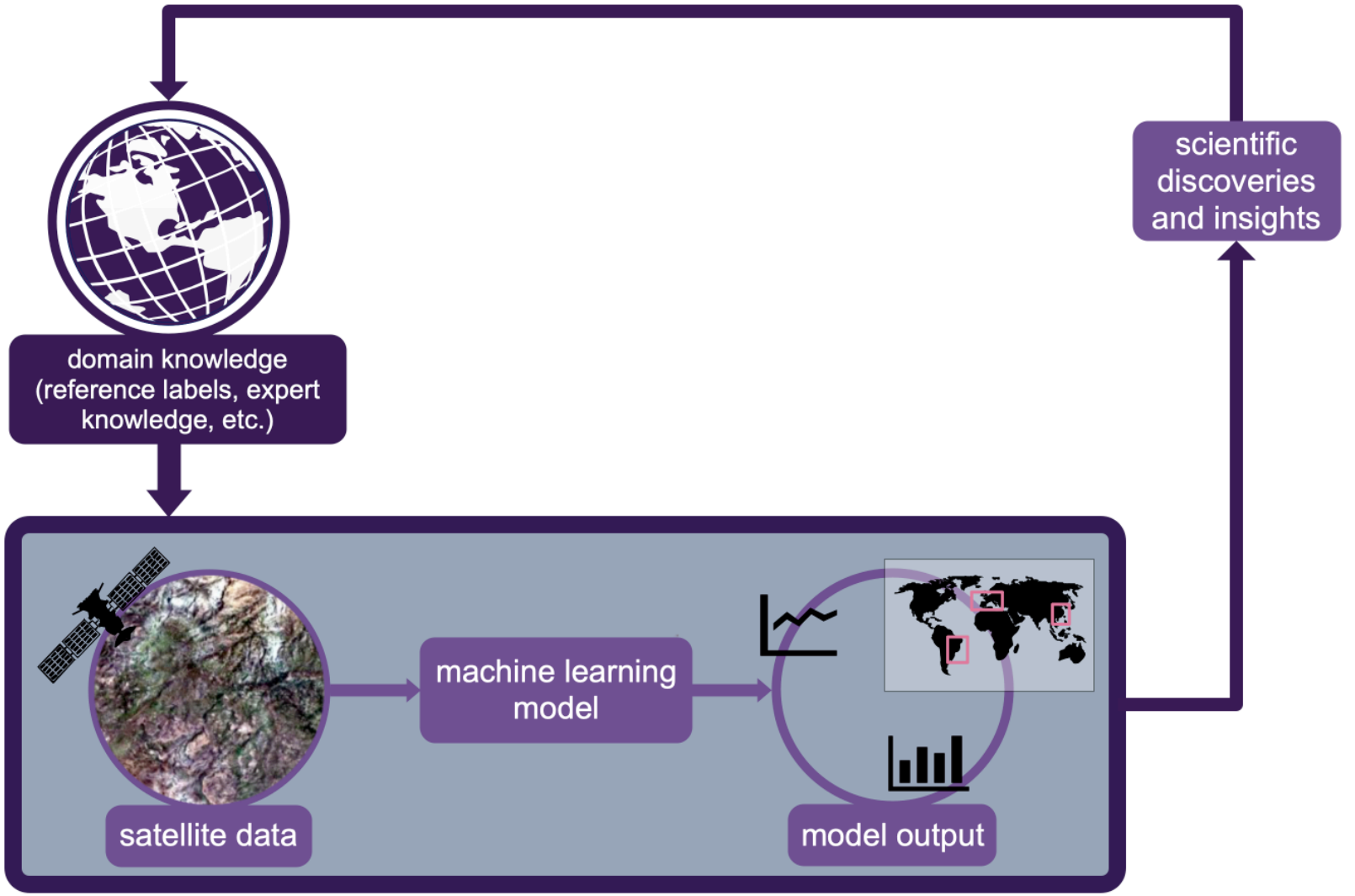


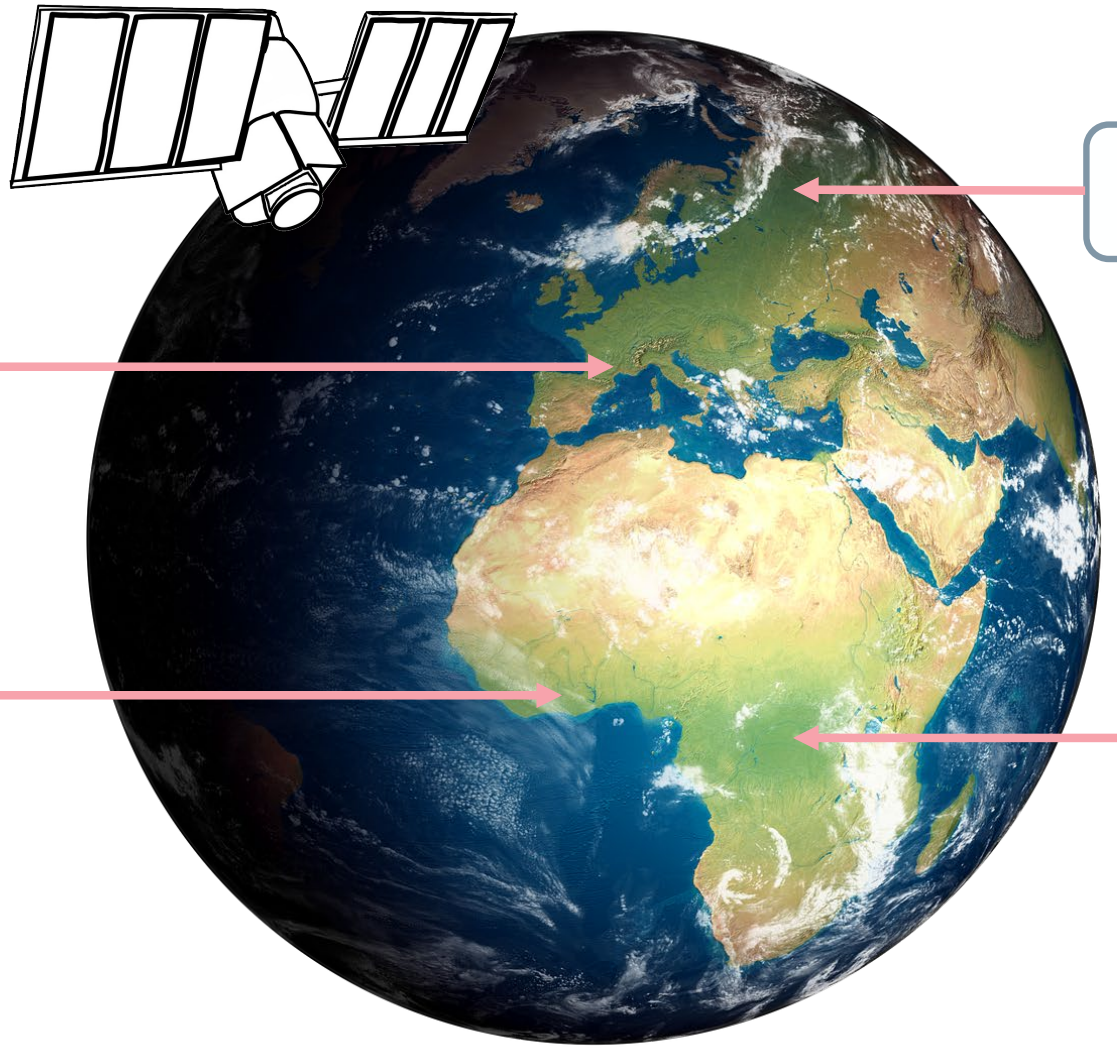
Has this area changed in recent months?

What is here?

What would it look like next month?

What is the biomass here?


scientific discoveries and insights
domain knowledge (reference labels, expert knowledge, etc.)
machine learning model
satellite data
model output



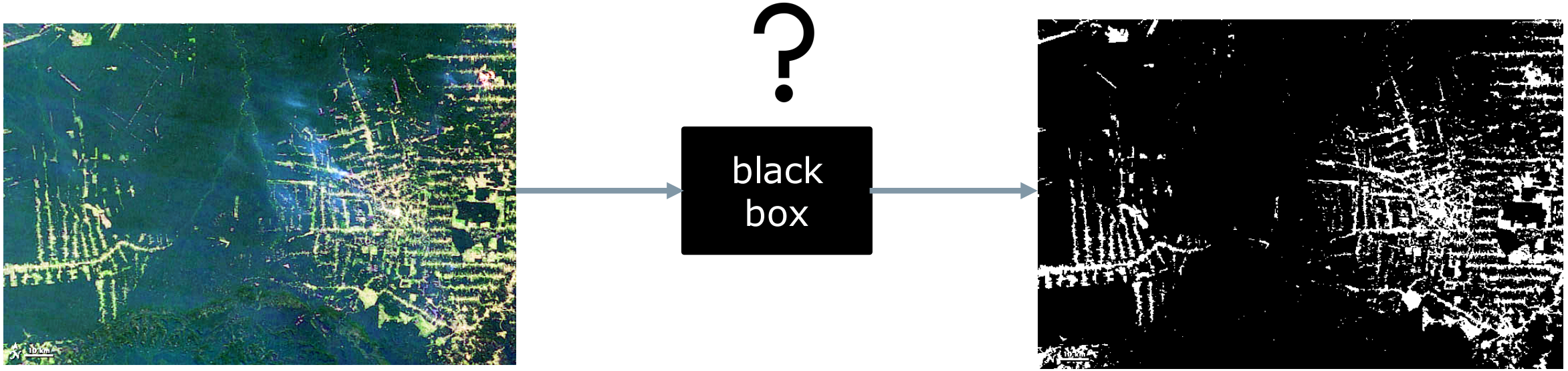
Why does it look like this?

Why is it classified as healthy vegetation?

Why does my algorithm perform poor here?

How does my algorithm come to the conclusion that this is water?

**Deep neural networks are the prime example for black box behavior.**

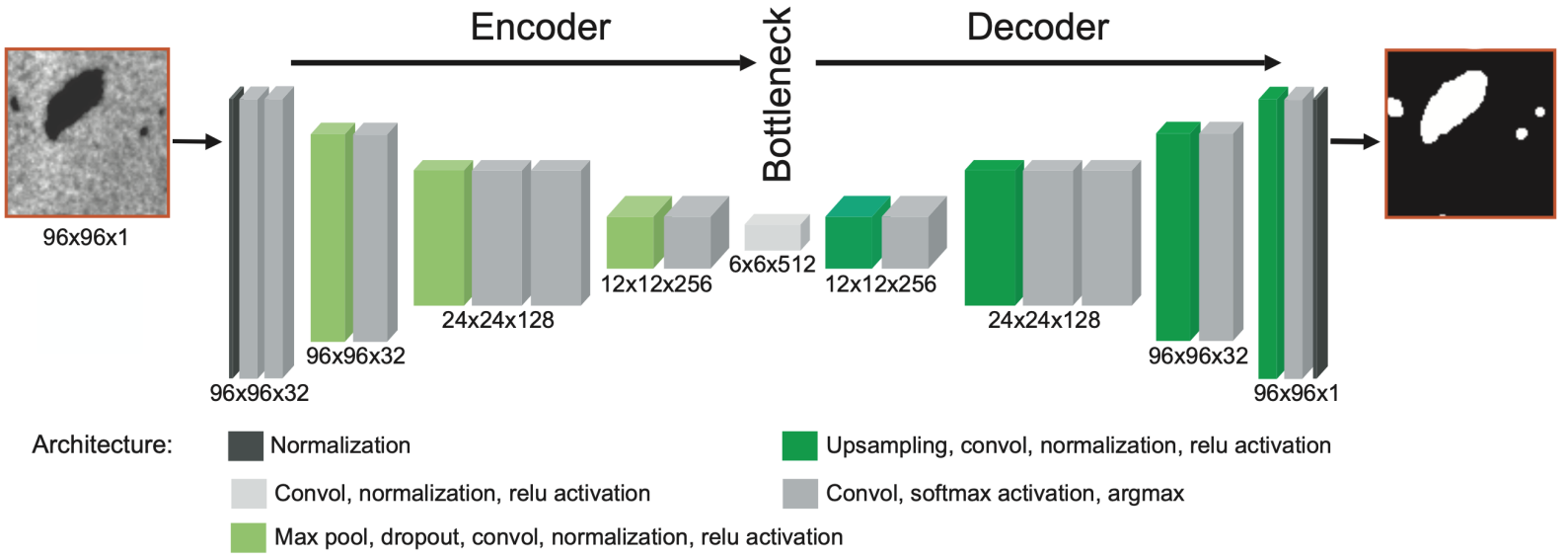



- Transparancy
- Interpretability
- Explainability

# Transparancy

Transparency of a machine learning approach concerns its different ingredients, including

- overall model structure
- individual model components
- learning algorithm
- how the specific solution is obtained by the algorithm



# Interpretability and explainability

## Interpretability

- Present some properties of a machine learning model (model structure, training data, learning procedure, …) in **understandable terms** to a human
- Can be obtained, for example, by visualizing relevant patterns or understandable proxy models

## Explainability

- Combine interpretable entities with **domain knowledge** (and analysis goal)
- Adadi & Berrada (2018) provide four reasons to seek explanations: to justify decisions, to (enhance) control, to improve models, and to discover new knowledge

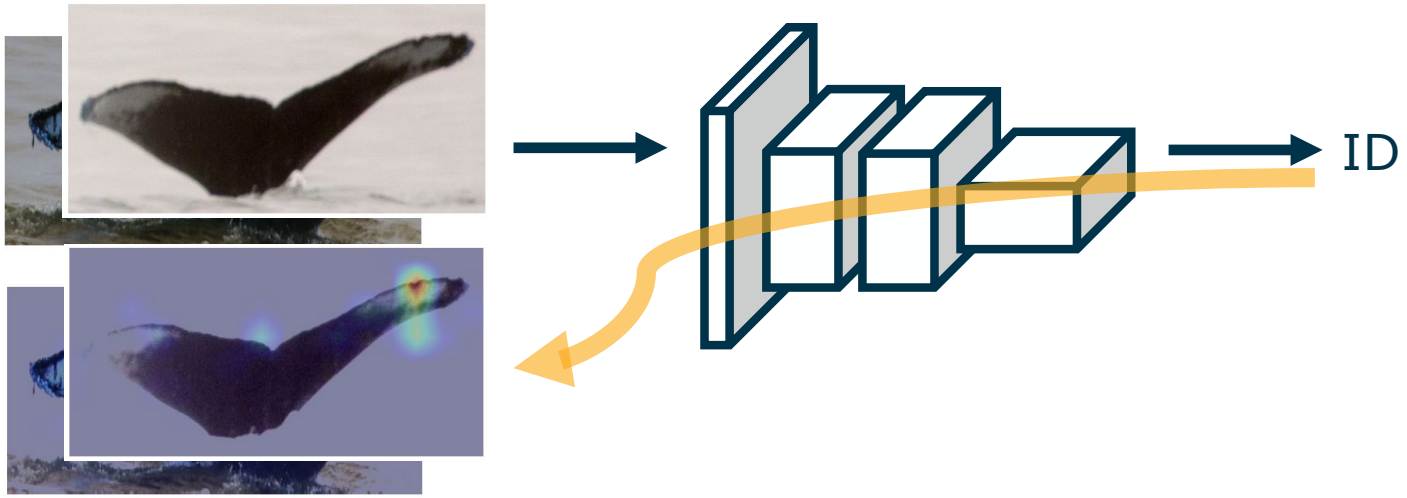
## Why do we distinguish?

➢ Explanation changes with application domain

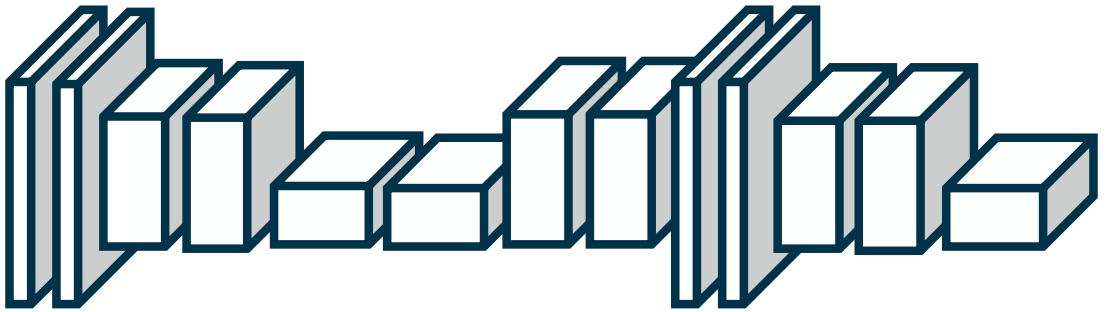
Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160.
Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access*, *8*, 42200-42216.

# Approaches

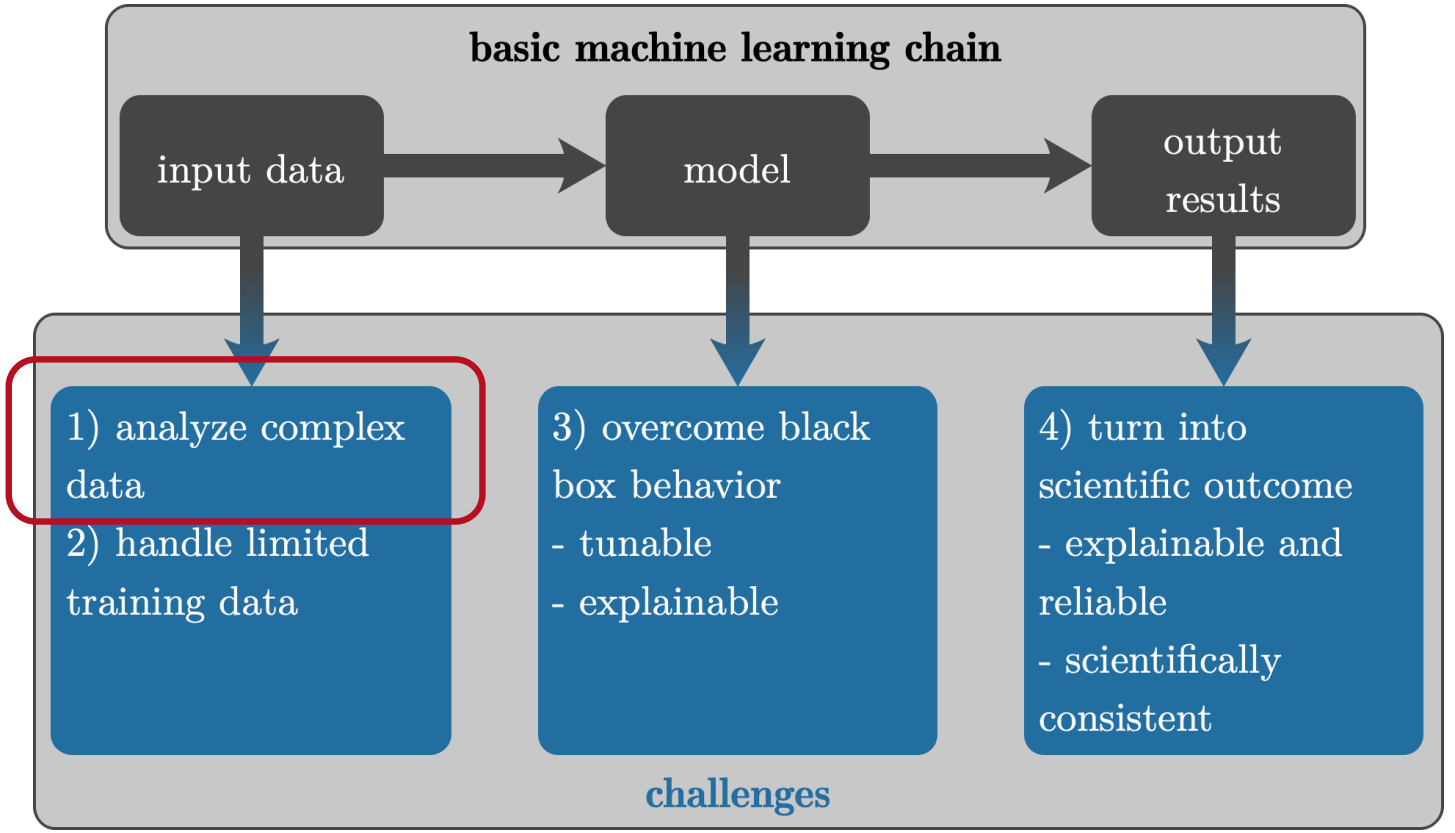
## Explaining output by the input




## Explaining the model (parts)

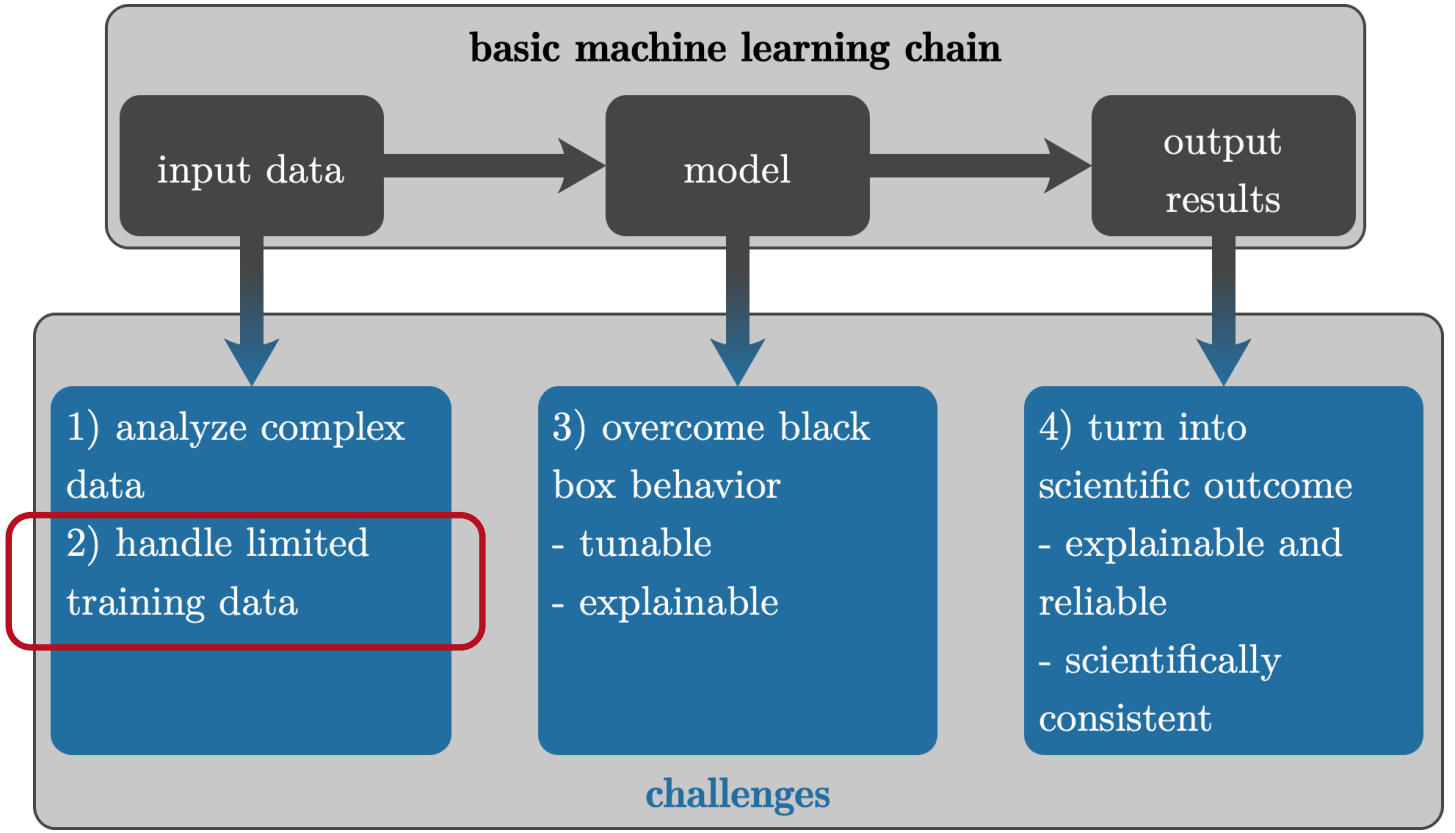


# Can explainable machine learning by useful in remote sensing?



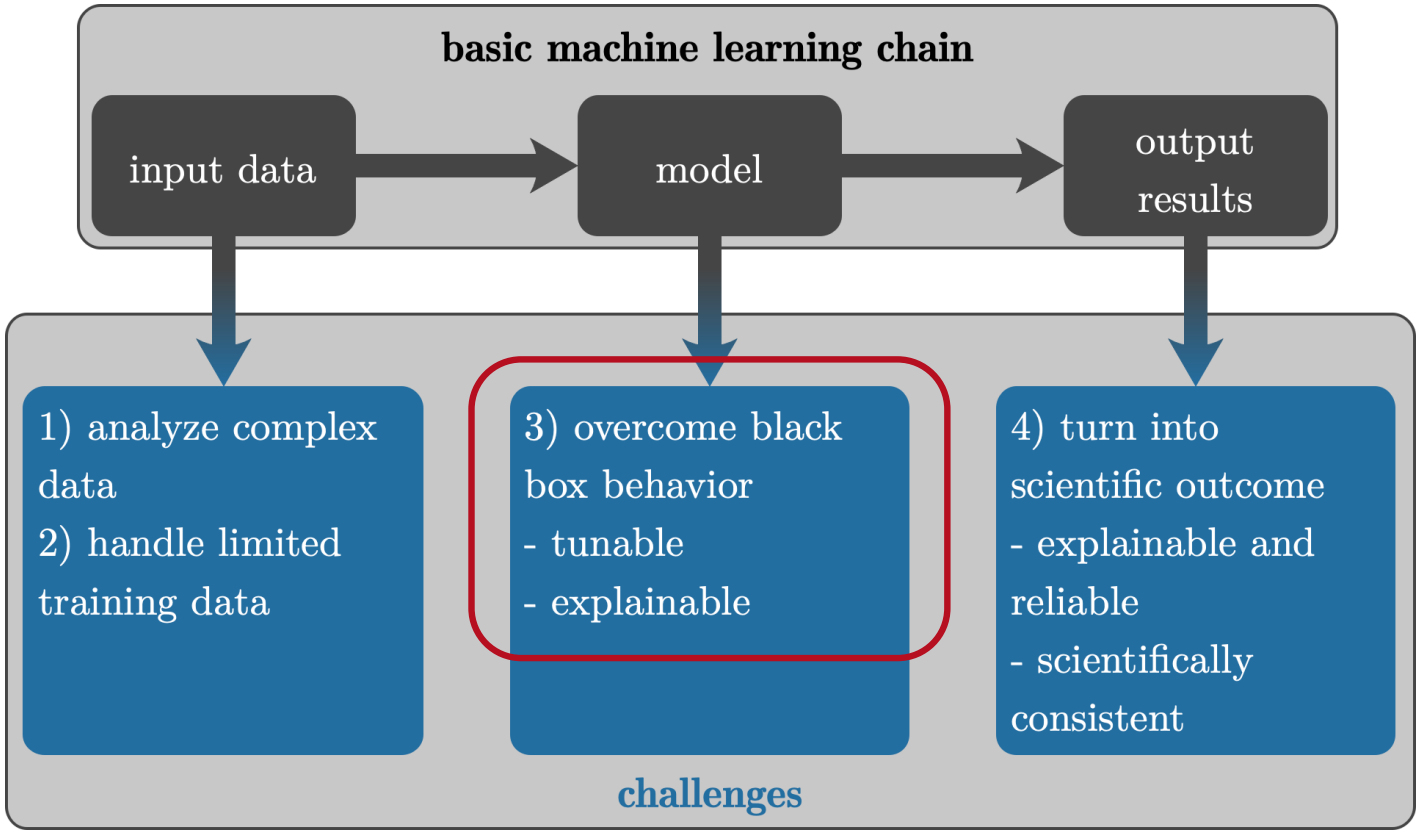

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explain it to Me-Facing Remote Sensing Challenges in the-and Geosciences with Explainable Machine Learning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*, 817-824.

# Can explainable machine learning by useful in remote sensing?



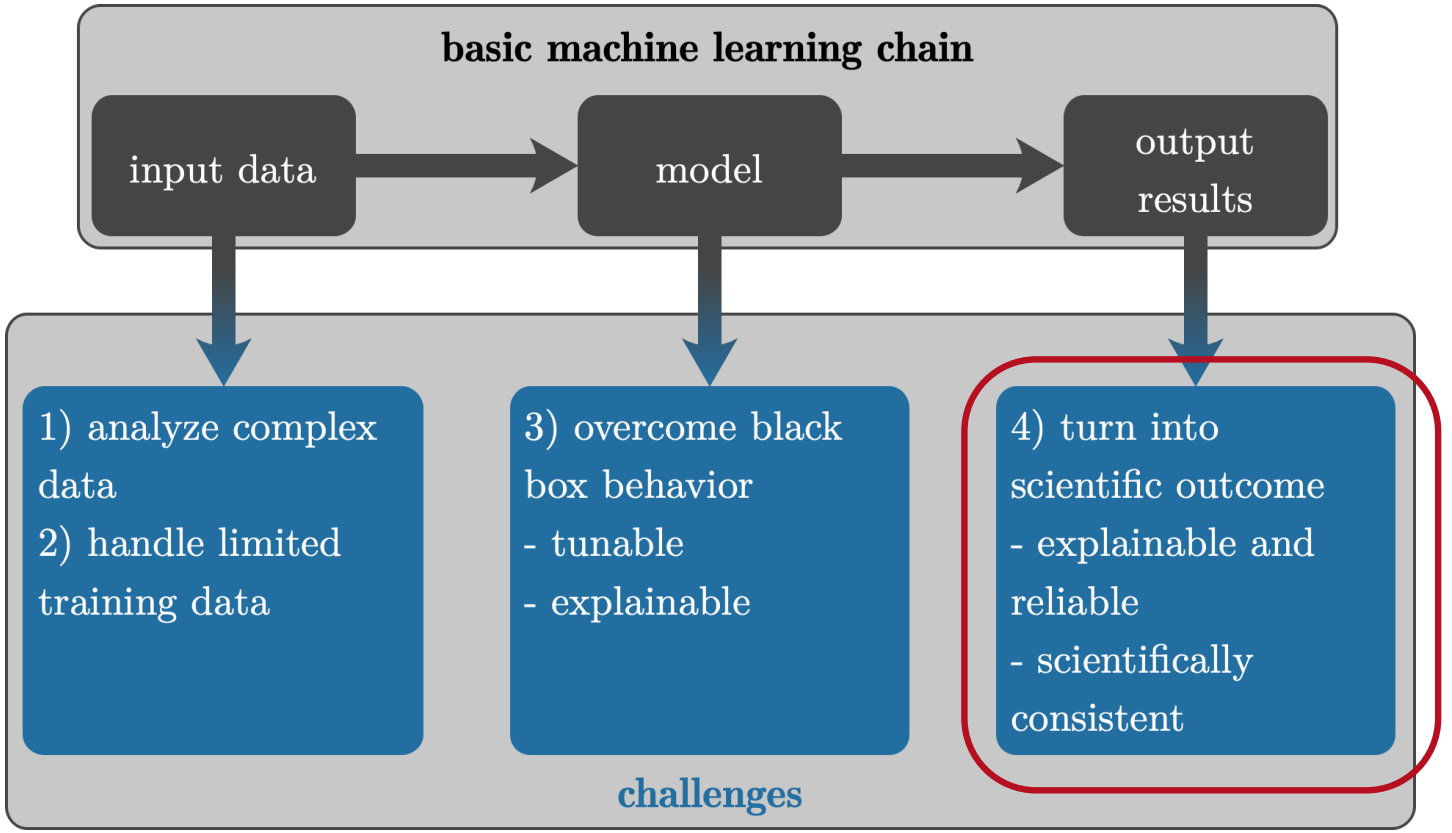

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explain it to Me-Facing Remote Sensing Challenges in the-and Geosciences with Explainable Machine Learning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*, 817-824.

# Can explainable machine learning by useful in remote sensing?



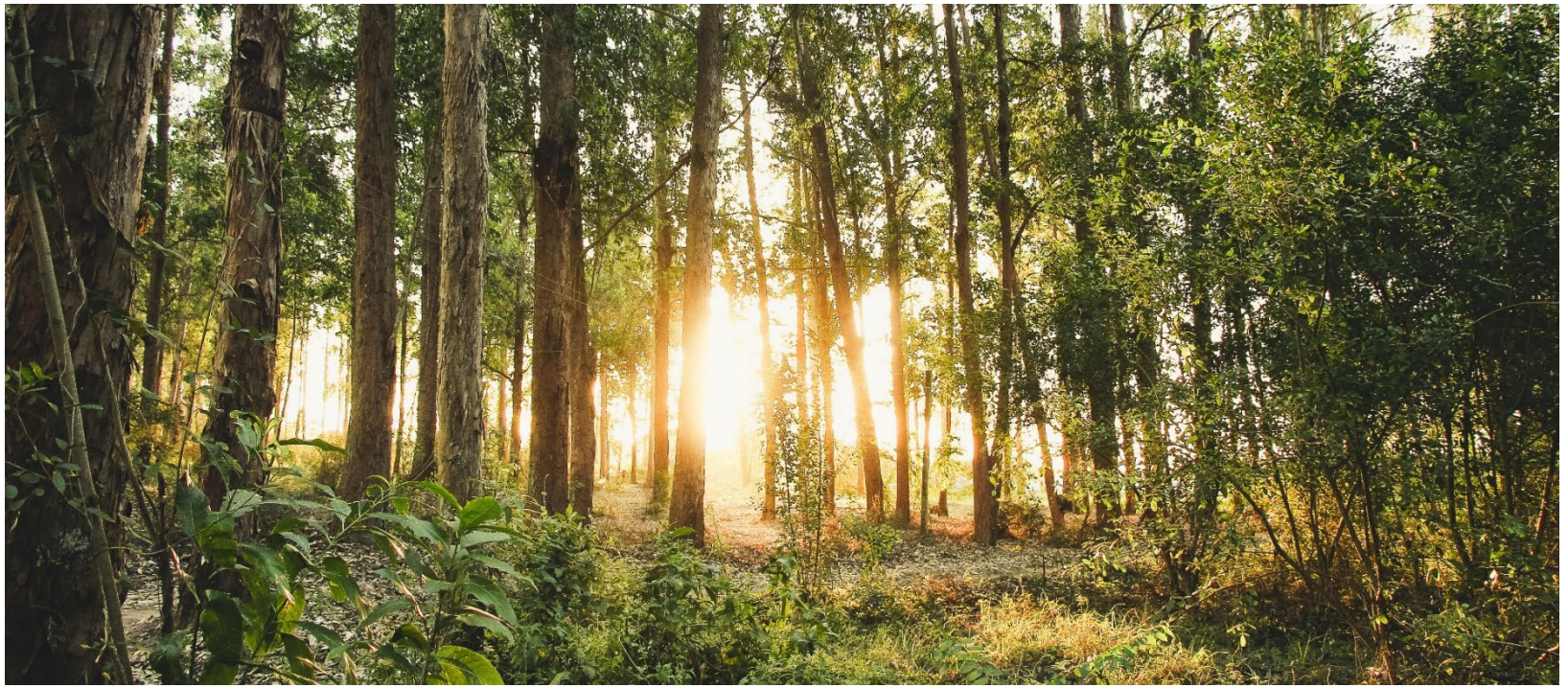

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explain it to Me-Facing Remote Sensing Challenges in the-and Geosciences with Explainable Machine Learning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*, 817-824.

# Can explainable machine learning by useful in remote sensing?




Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explain it to Me-Facing Remote Sensing Challenges in the-and Geosciences with Explainable Machine Learning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*, 817-824.

# What makes nature wild?

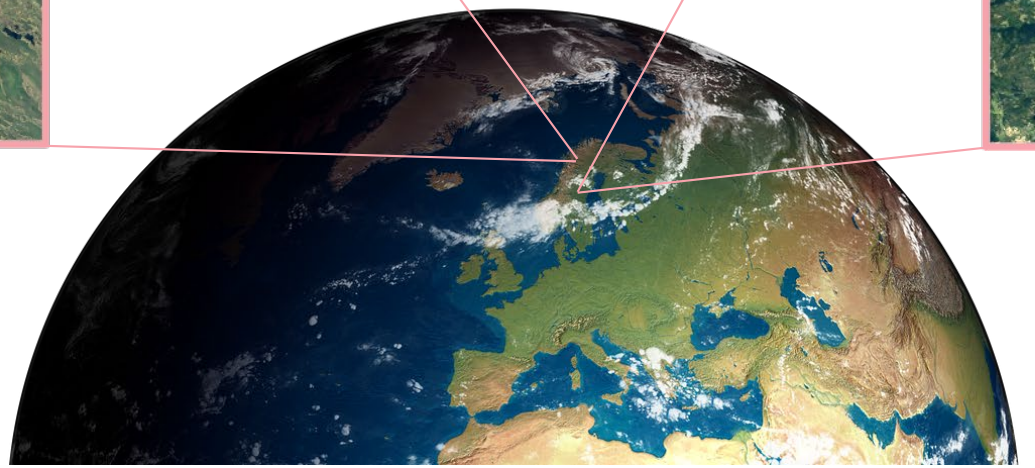


**No existing definition** which can be used for machine learning

➢Use explainable ML to **discover concepts** for wilderness and deepen our understanding about the land cover class so that it is useful for mapping

# What makes nature wild?

nature reserve Sjaunja, Sweden



Stockholm, Sweden

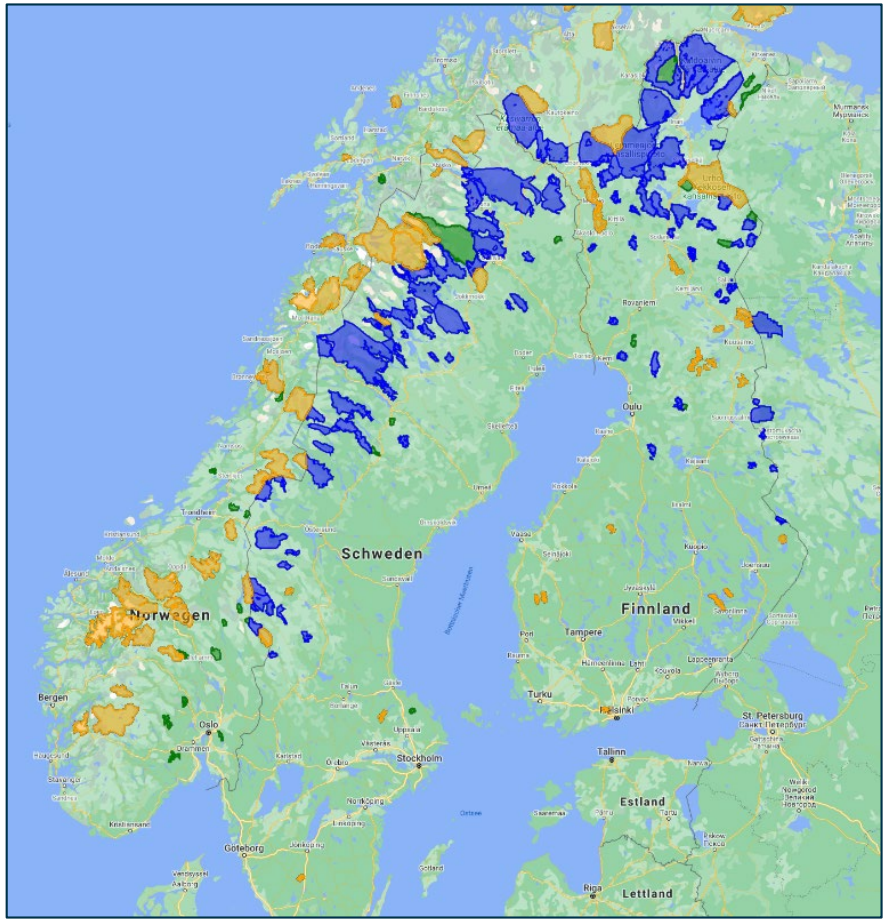
# What makes nature wild?

## World Database on Protected Areas (WDPA)

- category Ia: strict nature reserve
- category Ib: wilderness areas

„**Protected** areas that are **usually large** unmodified or **slightly modified** areas, retaining their **natural character** […], **without […] significant human habitation**, which are protected and managed so as to preserve their natural condition."

- category II: national park

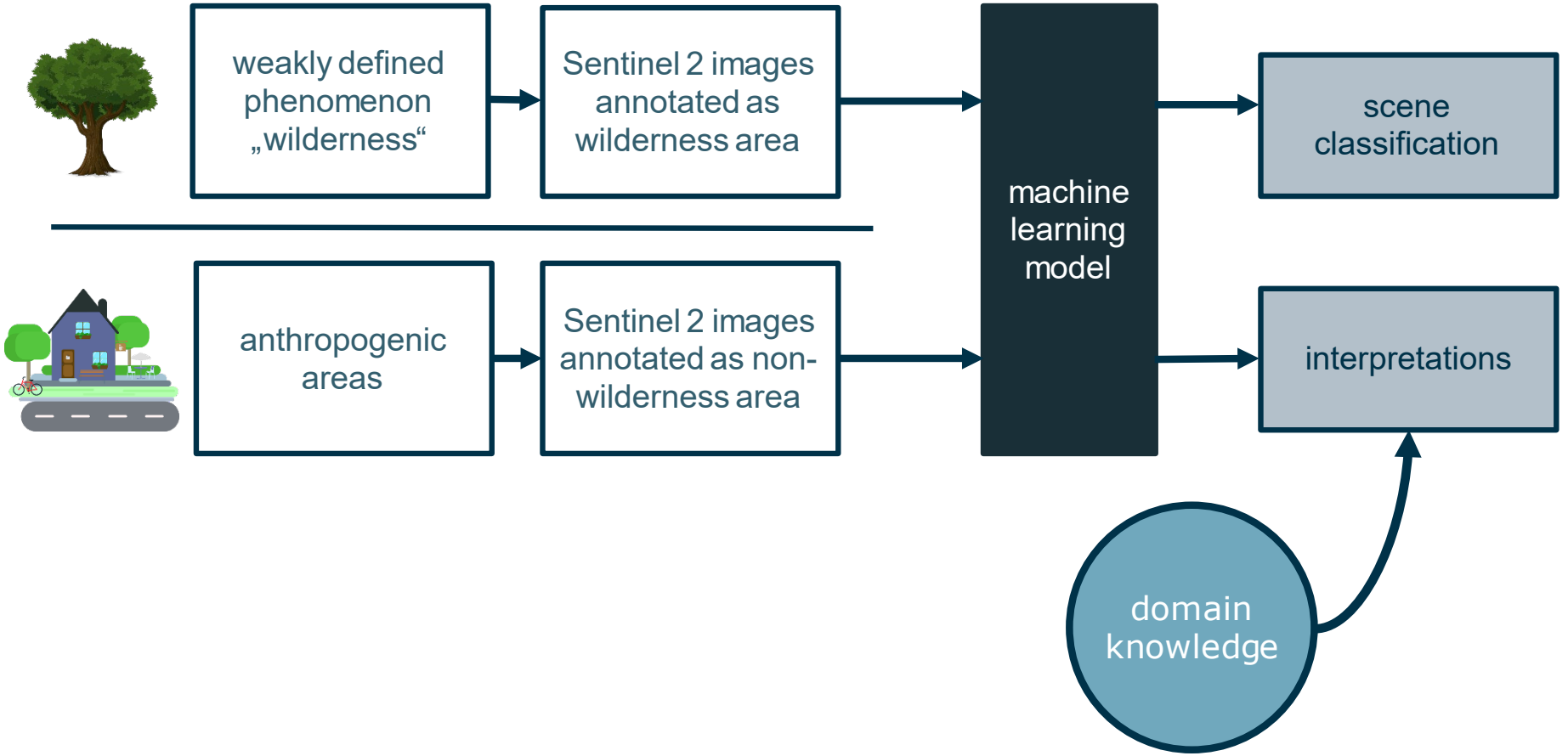



## Assumption

- protected areas (WDPA) can be associated with wilderness

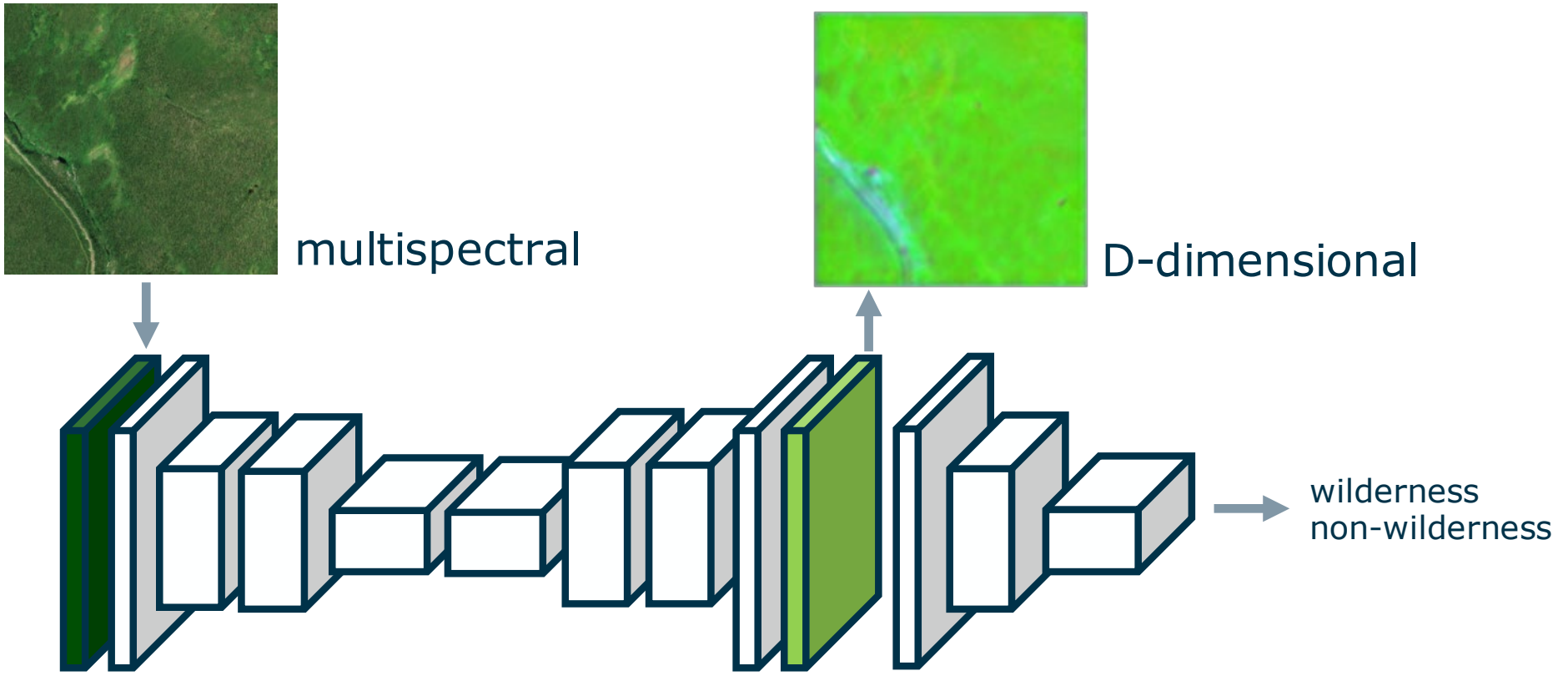
definition held by the International Union for Conservation of Nature and Natural Resources (IUCN)
https://www.iucn.org/theme/protected-areas/about/protected-area-categories/category-ib-wilderness-area

# Conceptual framework



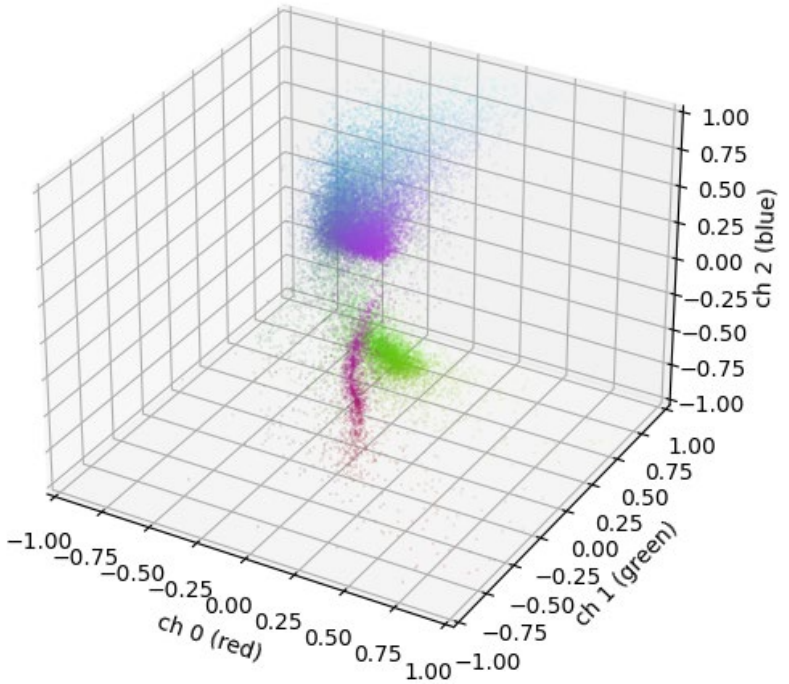

Stomberg, T., Weber, I., Schmitt, M., & Roscher, R. (2021). jUngle-Net: Using explainable machine learning to gain new insights into the appearance of wilderness in satellite imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *3*, 317-324.

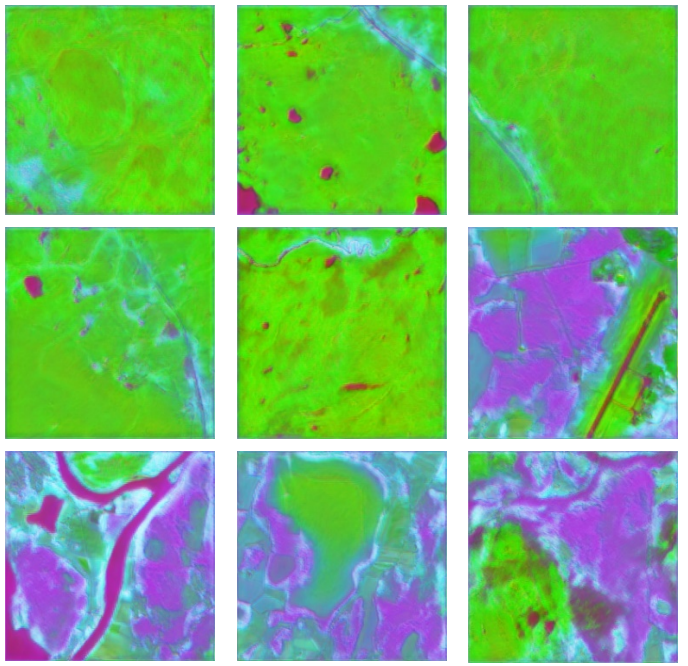
# jUngle-Net



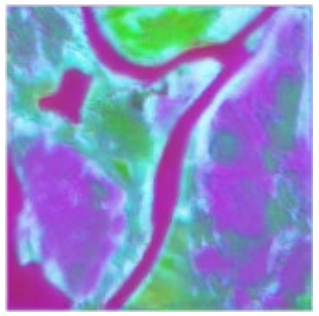

input ➔ U-net ➔ activation map ➔ CNN+FCN ➔ output

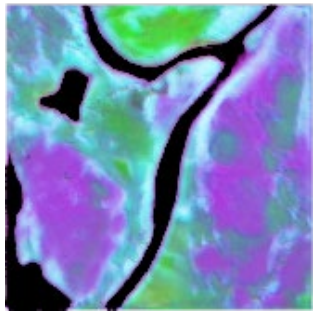
# Activation space

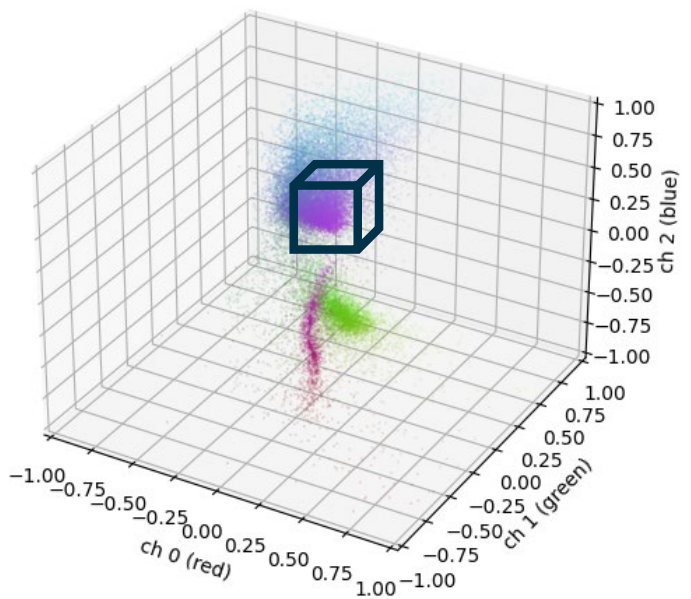


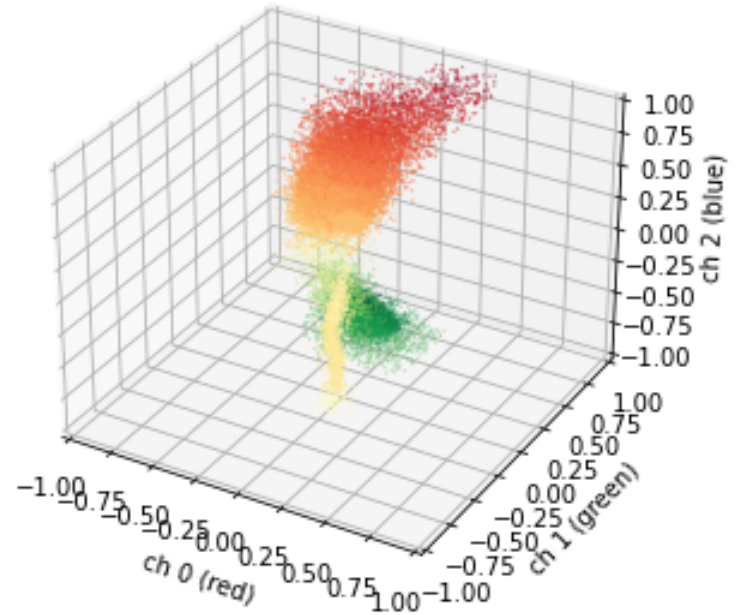



# Sensitivity analysis

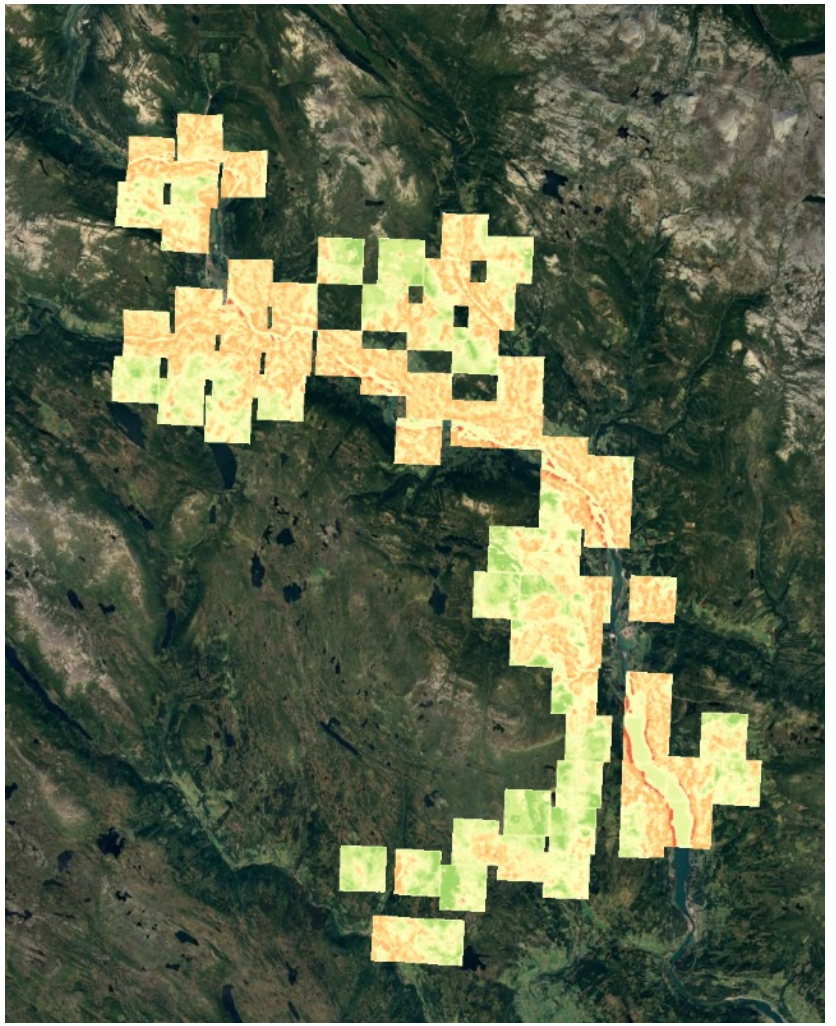




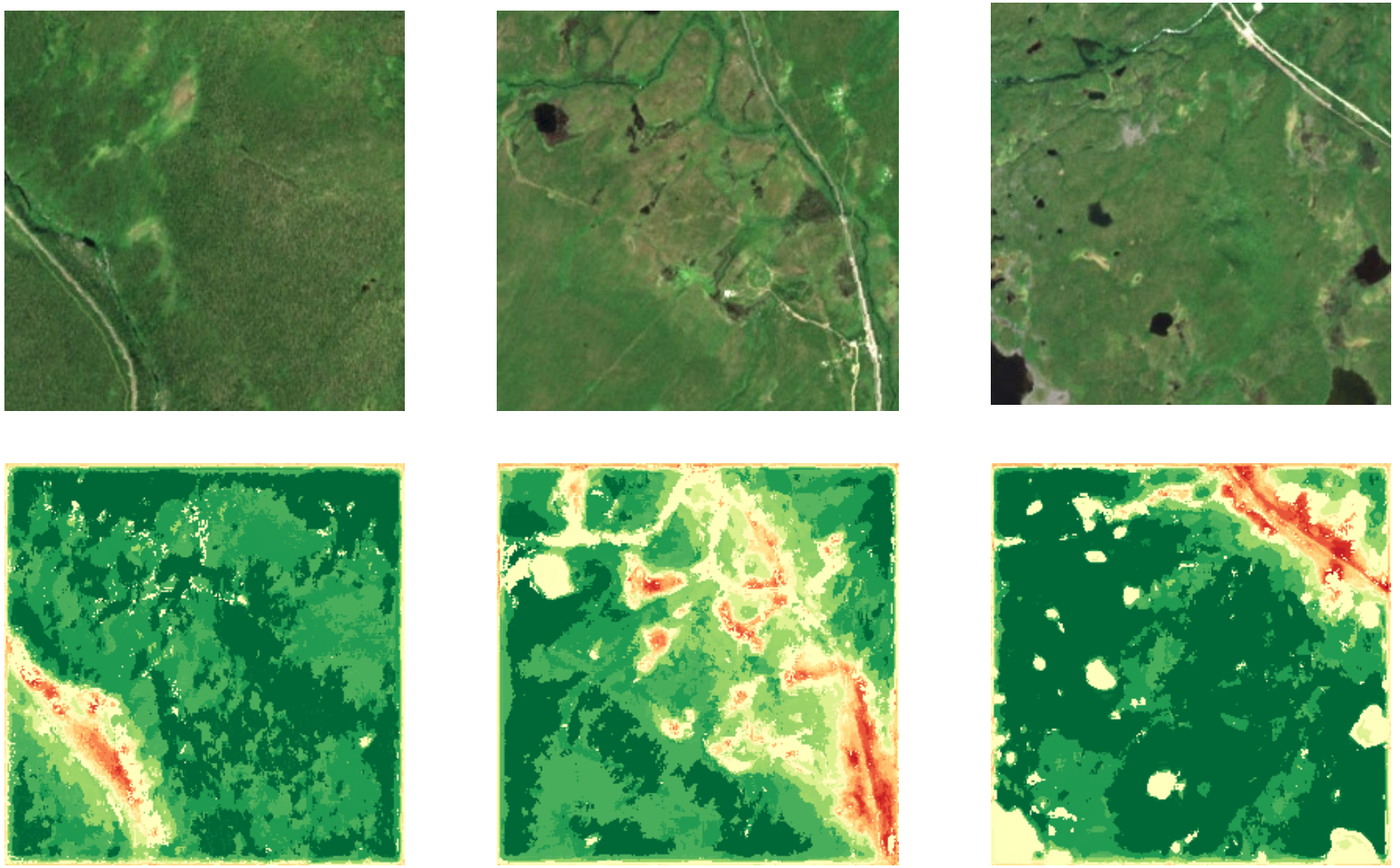


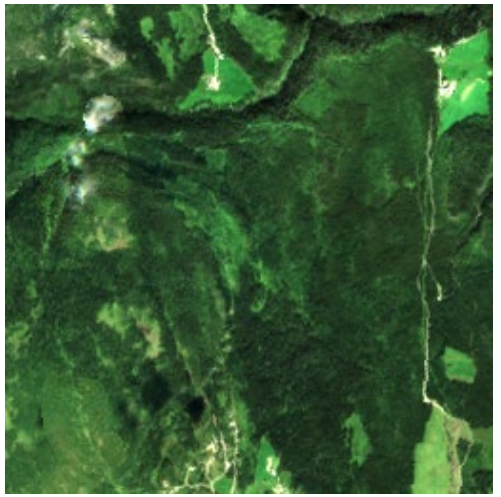
# Sensitivity analysis

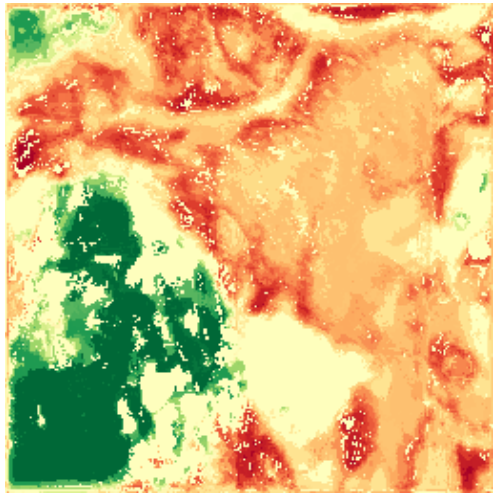


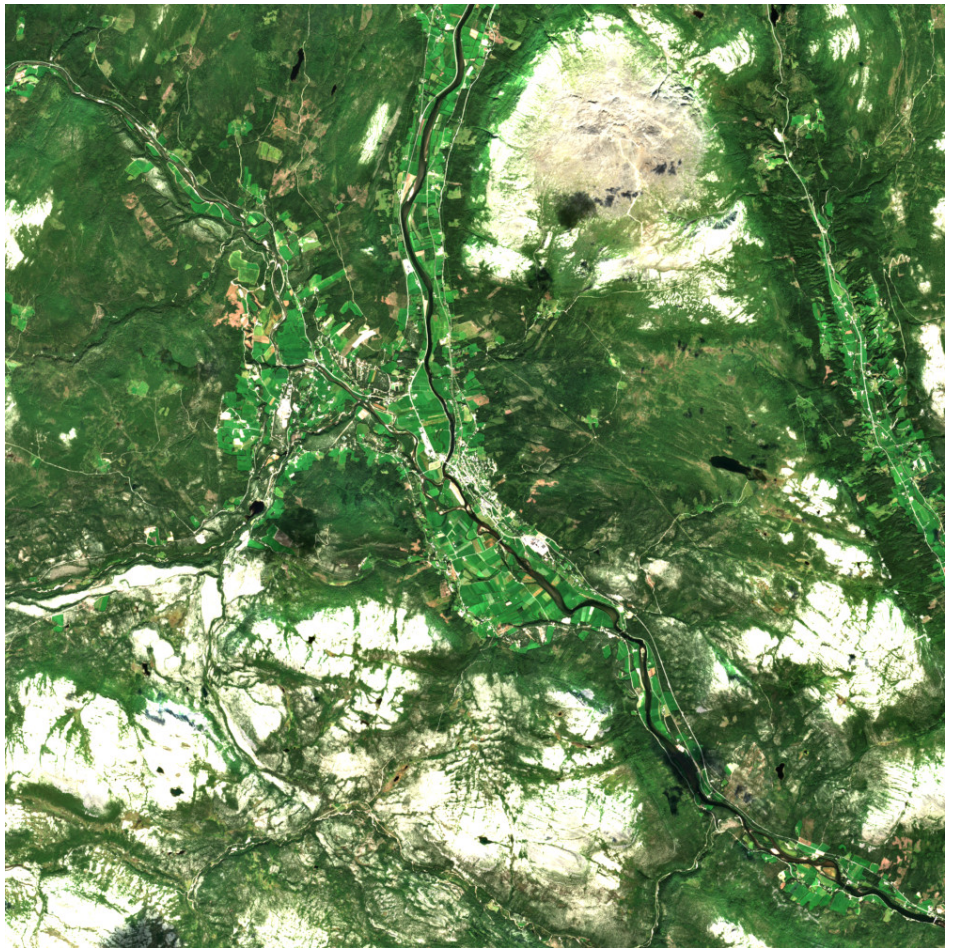
# Results: streets

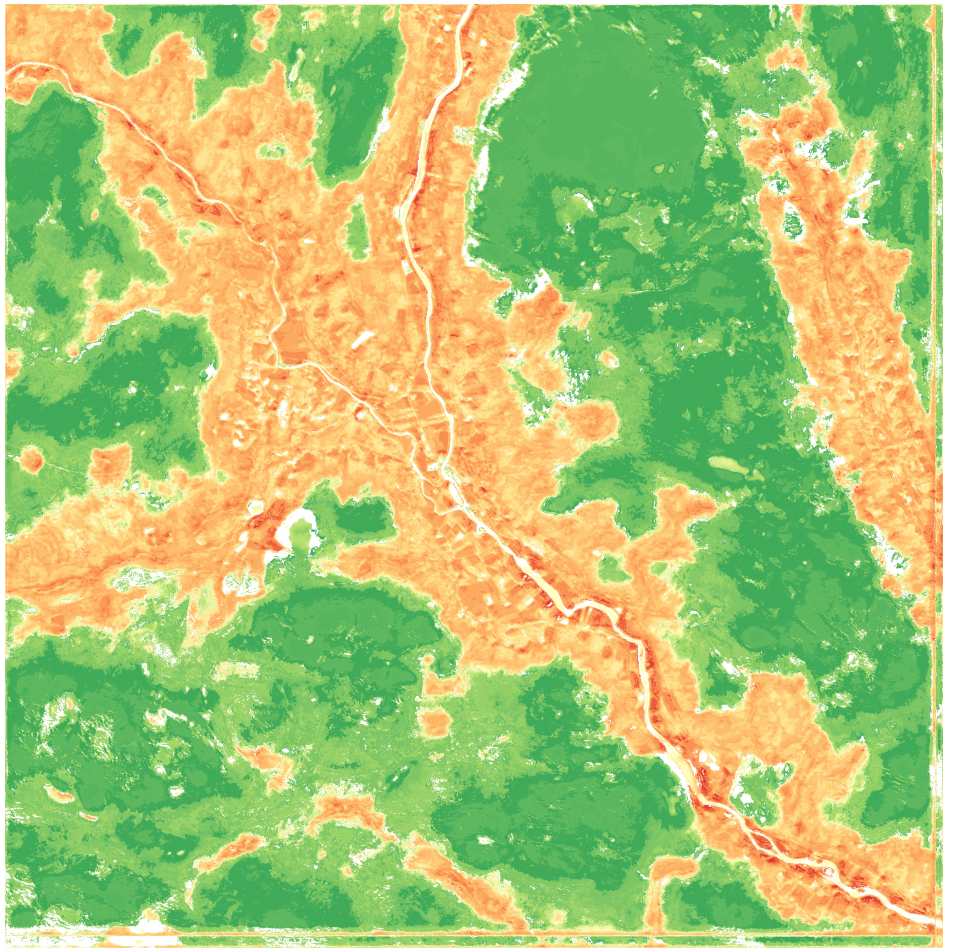


# Results: anthropogenic regions

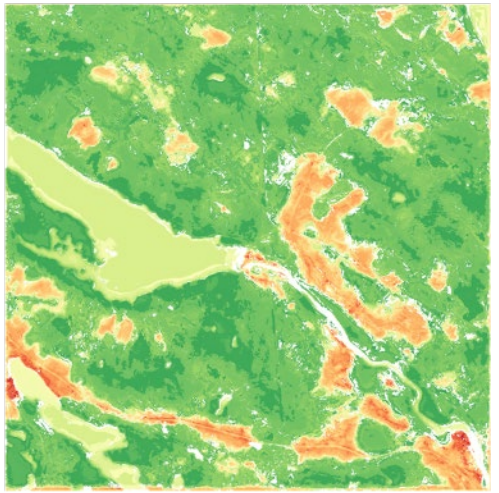


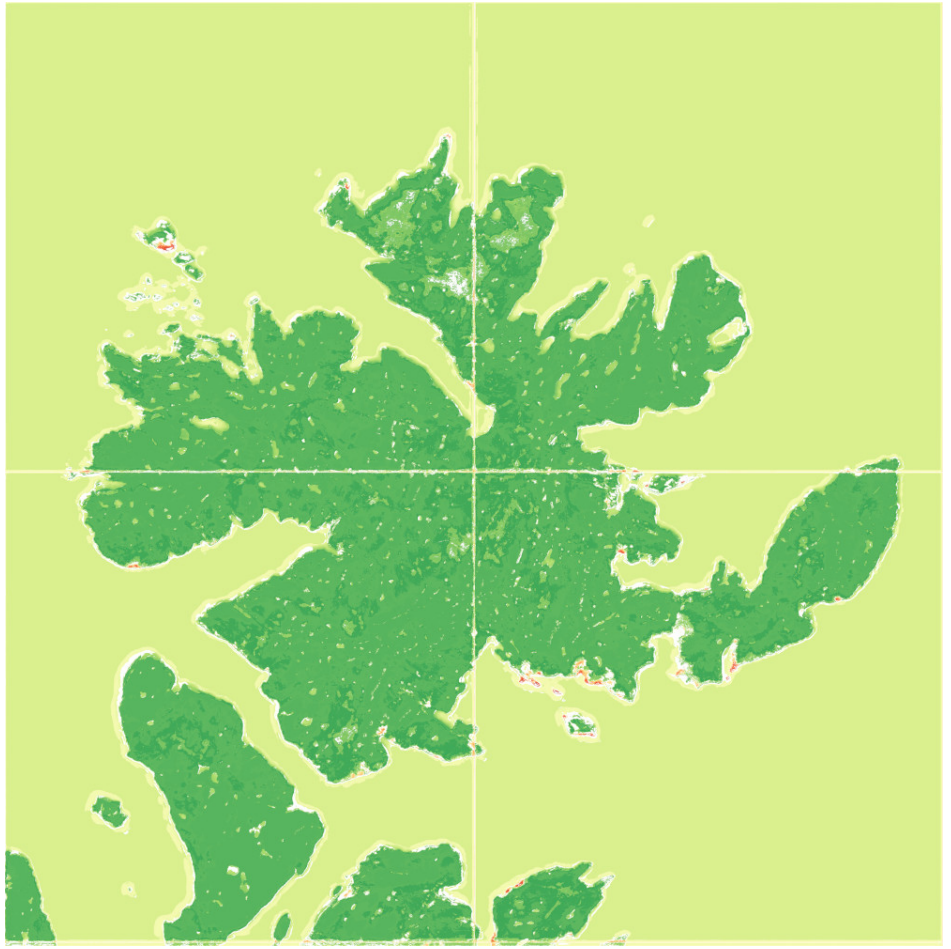






# Results: specific regions





# Findings

- jUngle-Net allows to find sensitive concepts and helps to better understand wilderness from a technical point of view
- Domain knowledge necessary (ongoing collaboration with Institute of Science and Ethics, University of Bonn)
- Iterative process necessary to guide the method to improve findings

# Conclusion

- Seeking interpretations and explanations is nothing new, it got more attention with the rise of deep neural networks
- Interpretations can lead to wrong or insufficient explanations – be aware of confirmation bias
- Explainable machine learning can tackle challenges in remote sensing by going **beyond accuracy maximization**
    - Helps to formulate hypotheses
    - Discover new knowledge and insights