CSC – IT CENTER FOR SCIENCE

HELMHOLTZ ASSOCIATION

Scientific workflows for different data analysis models

Shahbaz Memon[1,2], Helmut Neukirchen[1], Thomas Zwinger[3]

1) *University of Iceland, Iceland*
2) *Jülich Supercomputing Centre, Germany*
3) *CSC, Finland*

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES

JÜLICH FORSCHUNGSZENTRUM

Outline

Two case studies:

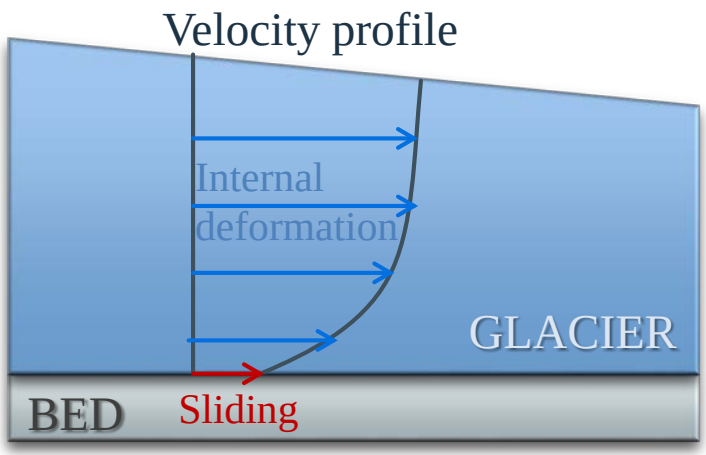
- Glacier Modeling
  - Already presented in detail last year
    - Based on application from Dorotheé Vallot
- Parallel classification of remotely sensed images

Ice flow: Movement of the ice

- Deformation of ice
  - Fracture (crevasses)
  - Internal deformation or creep
- Basal sliding

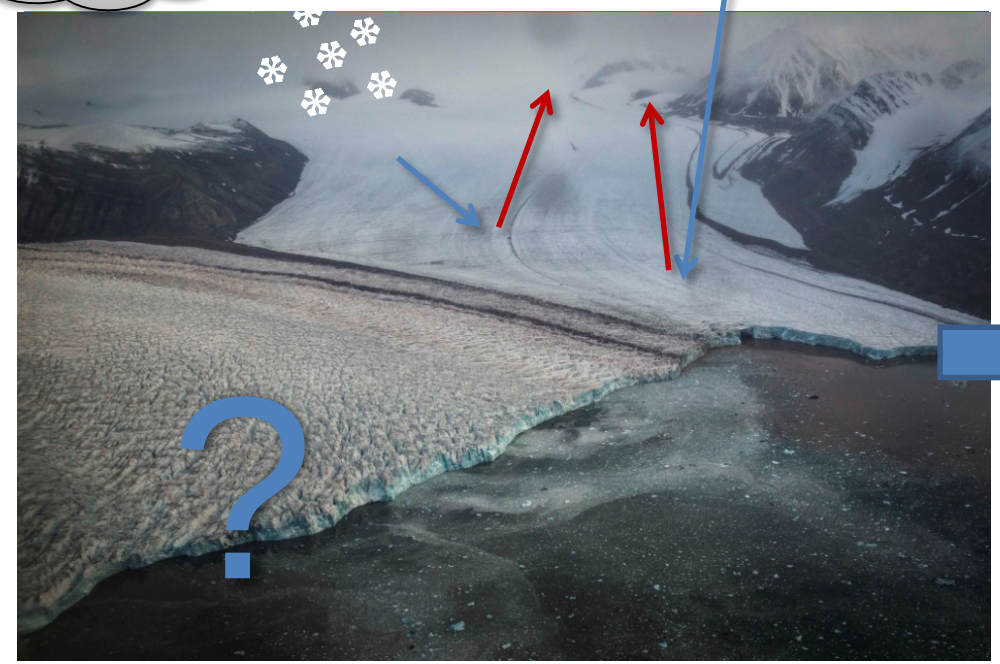
Photo: D. Vallot

➡ Can be modelled as a continuous process



Calving process

Accumulation

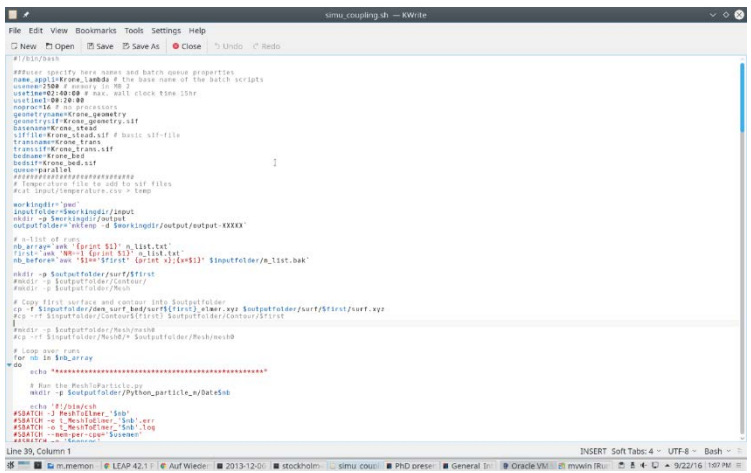


Ablation

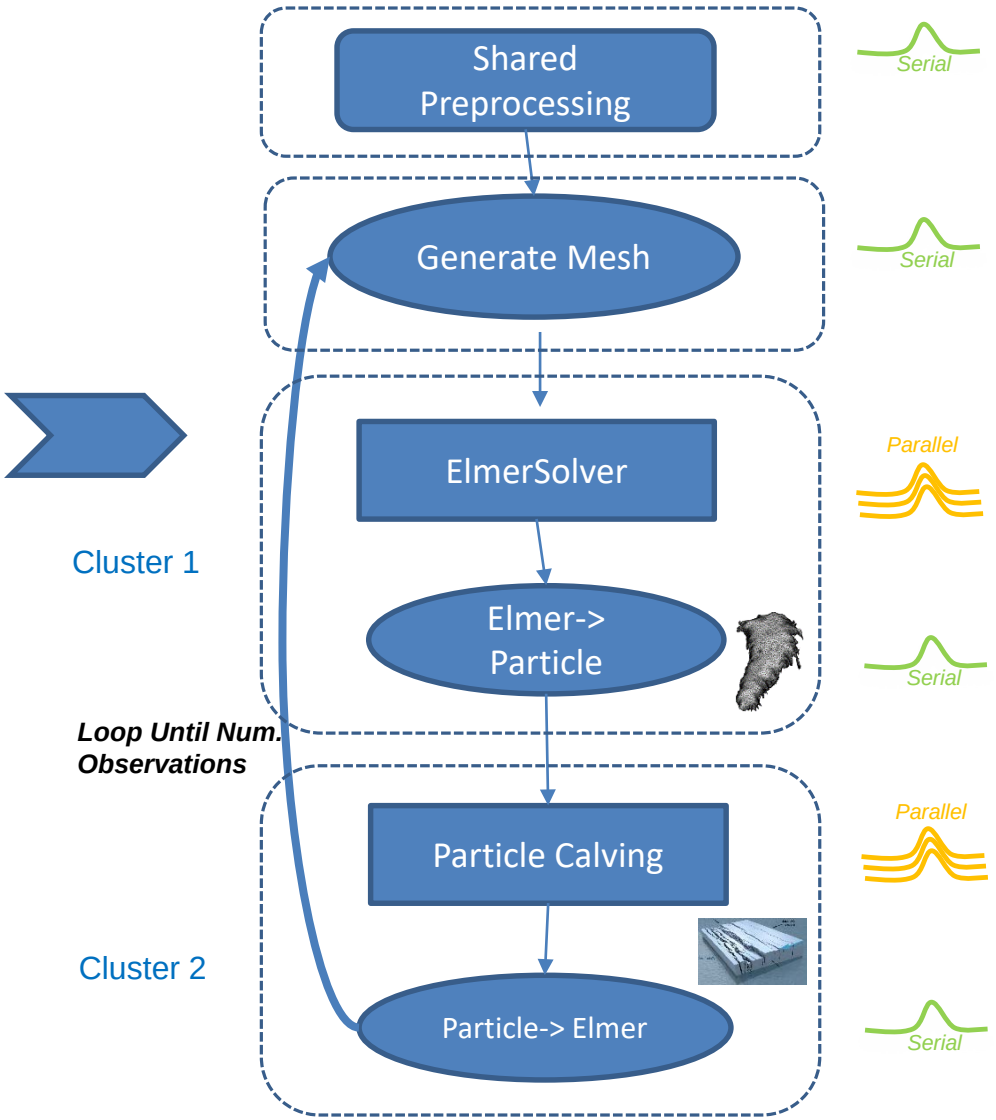
Calving occurs when tensile stresses are large enough to propagate fractures through the ice

Can be modelled as a discrete process

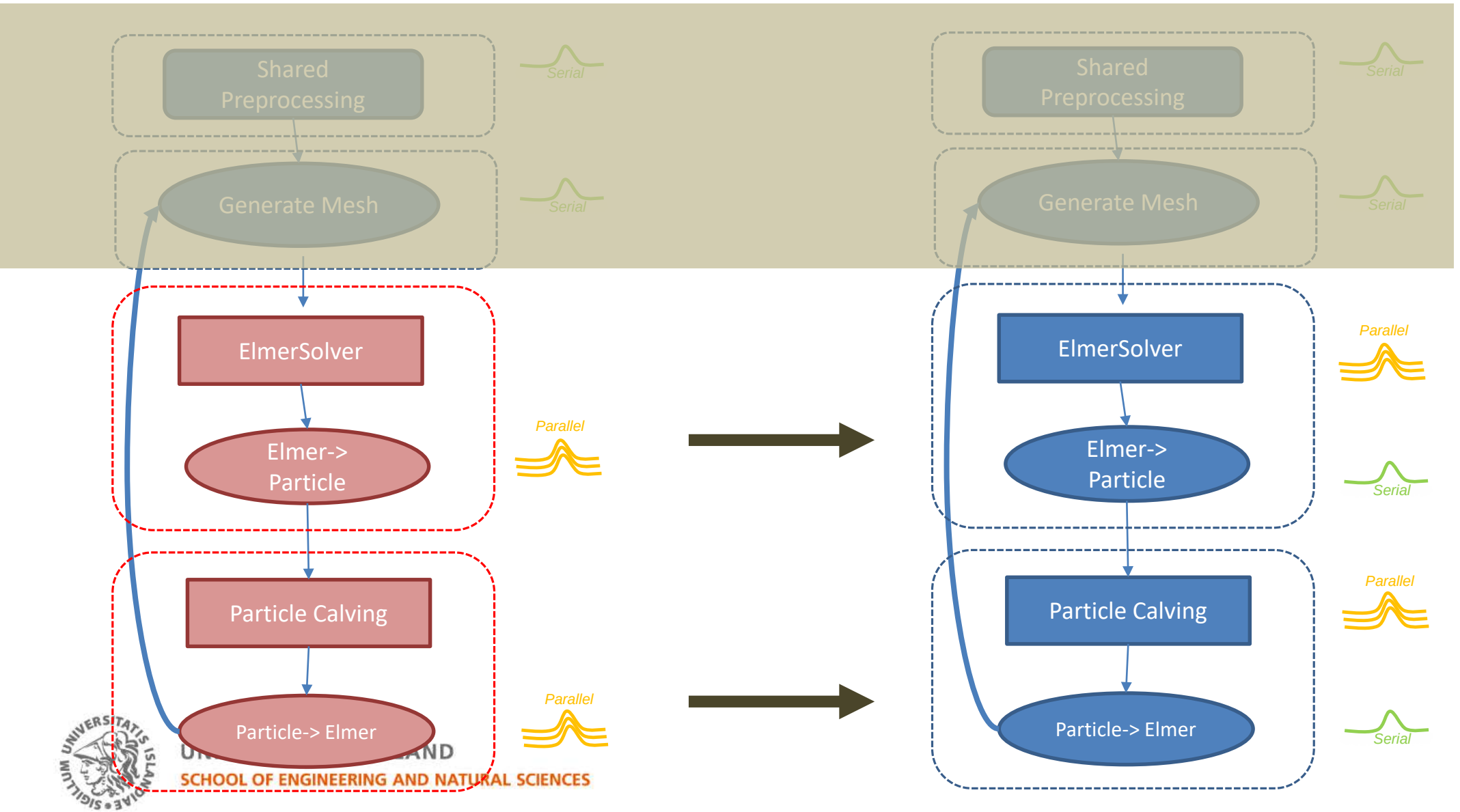
Glacier Modeling Workflow



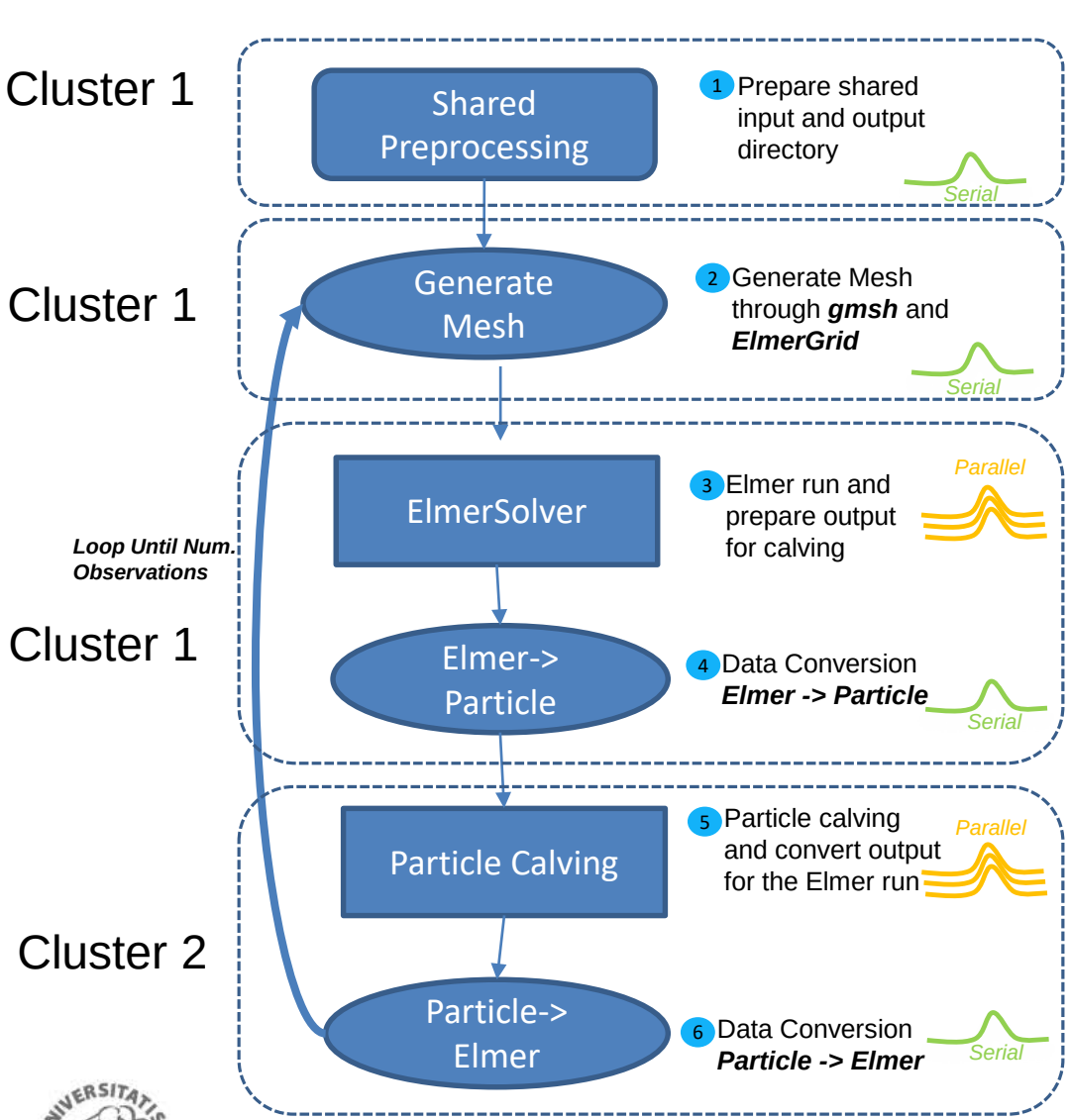
Main script

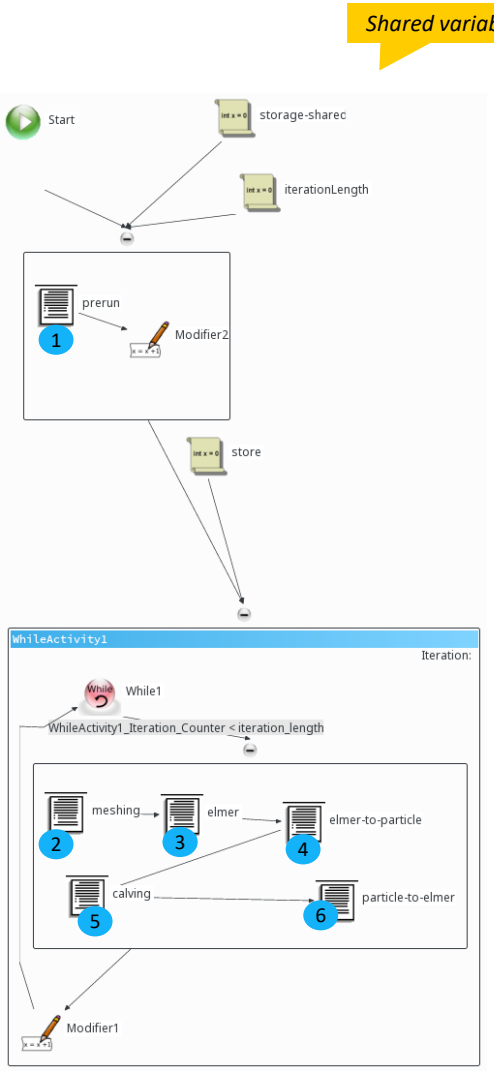


Workflow Enhancements



Workflow Realization in UNICORE





Summary: Glacier Modeling

- UNICORE deployed on CSC's Sisu and Taito Cluster
  - Thanks to CSC Admins and Thomas Zwinger
- Applications Involved: ElmerI
- Current workflow instance works with small data set
- Results:
  - EGU 2017: Coupling of a continuum ice sheet model and a discrete element calving model using a scientific workflow system
  - Journal/Conference: Complete use case with evaluation and usability analysis.

Case Study: Image Classification

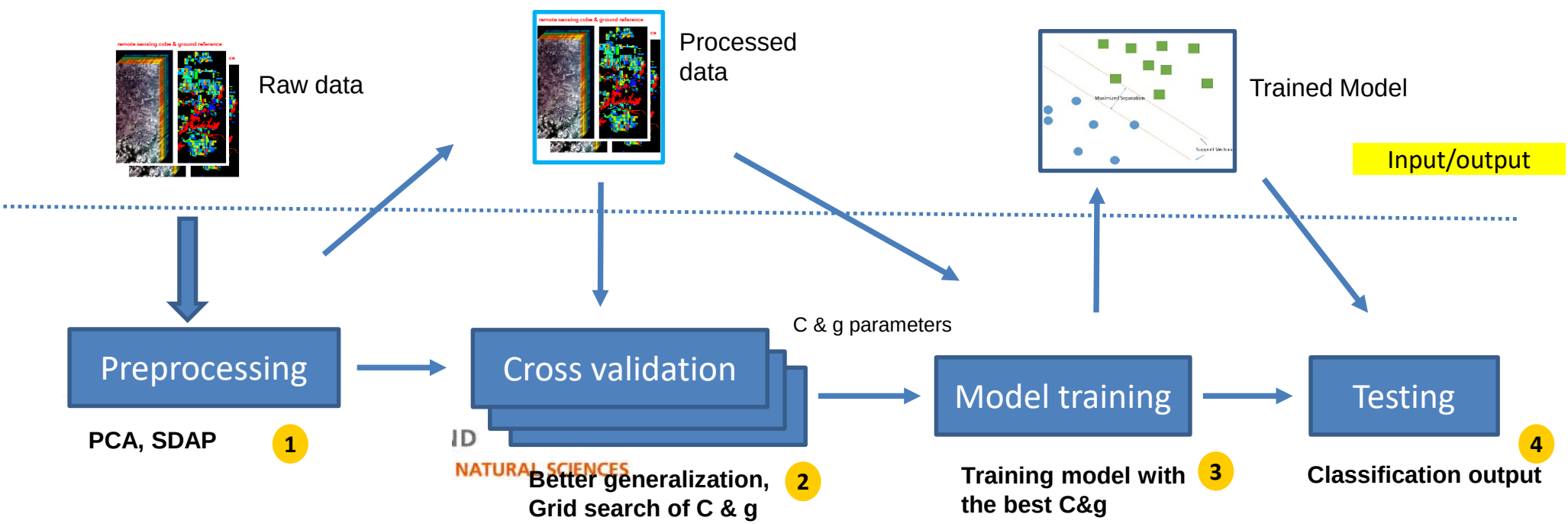
- Indian Pines dataset, multi-spectral dataset
- Acquired in 1992 through the AVIRIS sensor over an agricultural site composed of fields and regular geometry
- Land-cover classification problem consist of similar spectral classes and mixed pixels
- Each scene is preprocessed and generates 30 features with 1417x617, spatial resolution of 20m

Method: Support Vector Machines(SVM)

- SVM is a robust method to discover linear and non-linear decision boundaries with less amount of data
- Used in many remote sensing applications

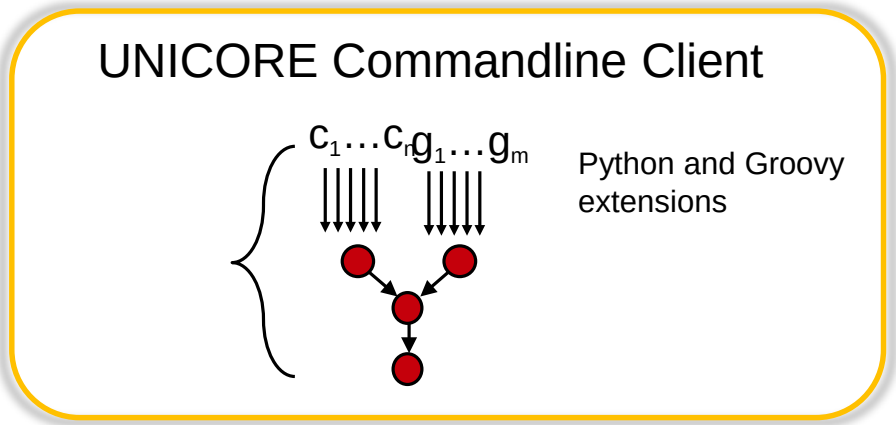
$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$$

Common Analysis Steps



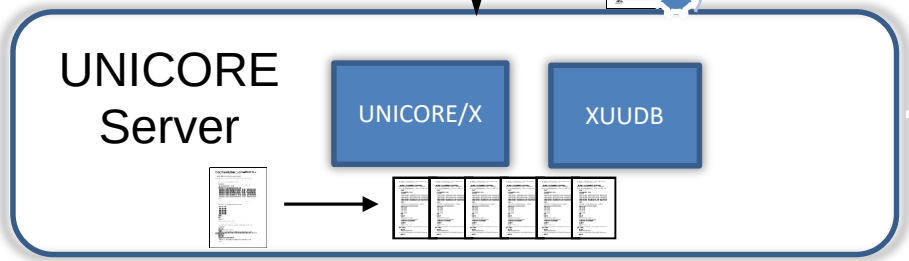
Analysis: Experiment setup

Job request creation and submission

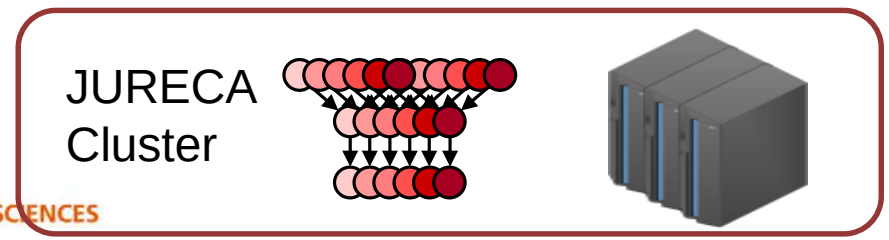


Single Job request over HTTP(s)

Request Validation and enactment



Execution on Target system



Summary: Multi-spectral Classification

- UNICORE deployment on Jülich's JURECA Cluster
- Command line client implementation available for cross validation and model selection
- Results and next steps:
  - IGARSS 17: Facilitating Efficient Data Analysis of Remotely Sensed Images Using Standards-based Parameter Sweep Models (Done)
  - Outlook: Automate the whole scenario including the data management, pre-processing, testing and accuracy tasks. (In Progress)