3rd Summer School on High-Performance and Disruptive Computing in Remote Sensing

HDCRS - Working Group of the IEEE GRSS Earth Science Informatics Technical Committee (ESI TC)

Prof. Gabriele Cavallaro (Forschungszentrum Jülich | University of Iceland)
Prof. Dora Blanco Heras (University of Santiago de Compostela)
Dr. Jin Sun (Nanjing University of Science and Technology)
SPIN-OFFS

CITIUS
Centro Singular de Investigación en Tecnologías Inteligentes

TEAM +140 PEOPLE
35 SENIOR RESEARCHERS

COORDINATED EUROPEAN PROJECTS

NANOVIR
Nanoscale Design using Virtual Reality

HYBRIDS
Interactive Natural Language for Explainable AI (Coordinator)

Multimodal Fusion of Sensor Information

R&D RESULTS 2019-21

ARTICLES

72% in Q1
36% in D1
22% in top 10% most cited

9,2 M €
16% income from European projects

OPPORTUNITIES FOR POSTDOCS

- Juan de la Cierva contracts
- Ramón y Cajal contracts
- Marie Skłodowska-Curie Postdoctoral Fellowships (MSCA-PF)
- ERC-Starting Grant
- Postdoc CITIUS

Welcome package with special benefits for each call. Contact us!

citiusus

citius.kmt@usc.es

Based in Santiago de Compostela, UNESCO World Heritage City. One of the 100 best places in the world by TIME Magazine 2021.

Based in Santiago de Compostela, UNESCO World Heritage City. One of the 100 best places in the world by TIME Magazine 2021.
CiTIUS: Scientific Areas

- High performance computing
- Electronic design of intelligent devices
- Devices and computing resources
- Remote Sensing
- Intelligent Technologies
  - Computer Vision
  - Virtual and augmented reality
  - Language technologies
  - Robotics
- Support to machine intelligence
- Social, economic, ethical and political framework
- Natural and artificially intelligence
  - Intelligent systems and environments
  - Trustworthy AI

Automatic learning and reasoning
The HPC and remote sensing group

- 20 years in HPC (15 years in Remote Sensing)
- 7 PhD in Remote Sensing and HPC finished
- 2 seniors and 5 researchers
- 5 PhD students and 10 master students
WHAT IS REMOTE SENSING?
Observing an object from a distance without physical contact
APPLICATIONS OF REMOTE SENSING

Observing objects and phenomena from a distance without physical contact allows for numerous applications

- Non-invasive method in contrast to in situ or on-site observation
- Efficient and continuous observation of the Earth and its changes
- Satellite platforms provide repetitive and consistent view
EARTH OBSERVATION AT DIFFERENT SCOPES

Solutions developed at three levels of increasing complexity (time response, amount of data, computational resources):

Local: a company, a group of users,…
Time: short

Regional or national: public institutions, companies,…
Time: medium

Global: governments, institutions,…
Time: long
CITIUS GROUP ON RS AND HPC

Solutions at the first and second level for extraction of information aimed at helping the final users in making decisions.

- Mainly from multi and hyperspectral sensor imaging. Integrating information from different sources.
- With strong focus on monitoring agriculture and natural ecosystems.
- Increasing the amount and quality of the information and providing it in a timely manner.

Watershed monitoring at regional level in Spain
CITIUS GROUP ON RS AND HPC

Solutions that require a variety of techniques and approaches

Analyzing, capturing

Preprocessing, storing

Solving: classification, change detection, anomaly detection, ...

Threats

ResBaGA
SATELLITES ORBITING THE EARTH IN 2022

8261 individual satellites orbiting the Earth at the end of January; an increase of 11.84% compared to April 2021

- To the end of January 2022, there have been 12293 objects launched into space
- Communications (3135), **Earth observation (1030)**, Technology development/demonstration (385)
- Navigation/positioning (154), Space science/observation (108), Earth science (22), Other purposes (18)

https://www.pixalytics.com/satellites-in-2022/
https://physicsworld.com/a/space-debris-threat-to-geosynchronous-satellites-has-been-drastically-underestimated/
ESA dedicated to observing Earth from space ever since the launch of its first Meteosat weather satellite back in 1977.

- ESA launched a range of different types of satellites over the last 40 years,
- The objective is to understand the complexities of our planet, particularly with respect to global change
- Applications: weather forecasting, Earth-science research, enhancing agriculture and maritime safety, aiding disaster response, etc.
The Copernicus Open Access Hub provides complete, free and open access to Sentinel data

Platforms: Sentinel-1, Sentinel-2, Sentinel-3 and Sentinel-5P user products

490,000 registered users with an average daily download volume of 203 TiB
EARTH OBSERVATION DATA
Information about the physical, chemical, and biological systems of the planet Earth

A market in constant expansion

**Upstream**
Space Infrastructure
Manufacturing of satellites, ground systems, launchers, ground operations

**Midstream**
Data Sales
Data acquisition, data processing, data archiving, data distribution and delivery

**Downstream**
Value-adding services Data
Processing, adaptation to users' need, integration with data from other sources

Digital Transformation Monitor, Big Data in Earth Observation, EASME/COSME/2017:

EUSPA EO and GNSS Market report 2022:
EO IS BEYOND BIG DATA
Volume, Variety, Veracity, ... Velocity and ...

High Dimensional
Data reflect complex relationships between natural and social phenomena

High Complexity
Data is incorporated in highly complex models

High Uncertainty
Data errors or incompleteness of data is unavoidable

MACHINE LEARNING AND DEEP LEARNING IN REMOTE SENSING

- Classical ML such as Support Vector Machine (SVM) and Random Forest (RF) since the ’90s
- DL unleashed advances in the last decade
- Figures may differ depending on the source, but the overall trend remains consistent
EMERGING COMPUTING PARADIGMS

- Supercomputing
- Specialized Hardware Computing
- Quantum Computing
- Edge Computing
- Cloud computing
- Blockchain
MOST BREAKTHROUGHS REQUIRE HEAVY COMPUTE POWER, USING MANY ACCELERATORS SIMULTANEOUSLY

- GPT-3: natural language generation, language understanding
- CLIP, DALL-E 2, Stable Diffusion: image understanding and image generation
- AlphaFold 2: protein structure prediction
- AlphaZero, MuZero: learning control in highly dimensional state-action spaces

https://openai.com/blog/chatgpt
Jenia Jitsev, Towards Scalable Deep Learning, Scalable Learning & Multi-Purpose AI Lab, Helmholtz AI, LAION @ JSC
CURRENT POPULARITY OF SUPERCOMPUTERS

- Big tech companies are announcing their AI supercomputers
- Supercomputing now goes far beyond traditional scientific computing, which was driven by large governments
- The field is currently propelled by major industries building highly specialized supercomputers


https://www.forbes.com/sites/jamesmorris/2022/10/06/teslas-biggest-news-at-ai-day-was-the-dojo-supercomputer-not-the-optimus-robot/?sh=22ba4ab780bd
https://ai.facebook.com/blog/ai-rsc/
https://www.thesun.co.uk/tech/5072741/google-nasnet-ai-child-reinforcement-learning/
Torsten Hoefler, “Efficient AI: From supercomputers to smartphones”, Scalable Parallel Computing Lab @ ETH Zurich, https://youtu.be/xxwT4SjG4o
PRE DEEP LEARNING ERA (1950-2010)

- Training compute doubled every 17 to 29 months
- This increase is roughly in line with Moore's Law


The trend accelerated along with the popularity of deep learning.

From around 2010, compute doubles every 4 to 9 months.
LARGE-SCALE ERA AROUND 2015

- In late 2015, a new trend of large-scale models emerged
- Computational capacity significantly higher than that of other models published in the same year (e.g., release of AlphaGo)
- This growth trend is slower than the overall DL trend, with a doubling time of roughly every 8 to 17 months
GPT-3 MODEL (175 BILLION WEIGHT PARAMETERS)

- AlexNet, winner ILSRVC 2012 – 60M
- ResNet-50, winner ILSRVC 2015 – 25M
- CLIP ViT L/14, multi-modal learning, 2021 - 600M

Jenia Jitsev, Towards Scalable Deep Learning, Scalable Learning & Multi-Purpose AI Lab, Helmholtz AI, LAION @ JSC
TIME REQUIRED FOR A FULL TRAINING OF GPT-3 MODEL

- ≈ 355 years with one Nvidia Volta 100 GPU
- ≈ 90 years with one Nvidia Ampere 100 GPU
- ≈ 16 days if scaled well with 2,000 A100 GPUs on JUWELS Booster

Jenia Jitsev, Towards Scalable Deep Learning, Scalable Learning & Multi-Purpose AI Lab, Helmholtz AI, LAION @ JSC
TRAINING OF MODELS REQUIRES ACCELERATORS

- GPUs (currently NVIDIA dominant), TPUs (Google)
- GPUs: generic deep learning hardware (parallelizing matrix/tensor operations via vectorization)
- Specialized hardware, e.g. in-memory computing chips, Graphcore IPU: Colossus MK2, Cerebras Wafer Scale Engine 2 (850k cores)

https://cloud.google.com/tpu
https://www.graphcore.ai/products/ipu
https://www.cerebras.net/product-chip/
MODERN HPC SYSTEMS ARE NOT ONLY MASSIVELY PARALLEL BUT ALSO HETEROGENEOUS

- Cost-effective scaling
- Effective resource-sharing
- Fit application diversity
  - Large-scale simulations
  - Data analytics
  - Machine- and Deep Learning
- Composability of heterogeneous resources

Modular Supercomputing Architecture (MSA)
The Forschungszentrum Jülich in Germany will house Europe's first exascale computer, named JUPITER.

This machine is set to surpass the threshold of one quintillion (a “1” followed by 18 zeros) calculations per second.

JUWELS 20 = JUPITER (when using the diameters of planets as a measurement scale).
JUPITER WILL BE A MODULAR SUPERCOMPUTER

- Its modules will be able to interactively and dynamically work together (it is not just a set of modules/machines)
- Target >20x application performance compared to JUWELS Booster
- Performance: >1 Exabyte storage (with different storage technologies), > 1 EF (GPU), 7,5 PB/s (x86 or ARM)

Thomas Lippert, "Jupiter Ascending, Driving the Forefront of Europe’s HPC Effort", EuroHPC Summit 2023, https://www.eurohpcsummit.eu
DESTINATION EARTH

- HPC is used intensively to run highly complex simulations of the Earth system at the very high resolution of the digital twins
- The flexibility of modular systems is suitable for DestinE
- It's important to quantify the uncertainty of the forecasts and to perform complex data fusion, data handling, and access operations.
GRSS TECHNICAL COMMITTEES

- Earth Science Informatics
  - Geoscience Spaceborne Imaging Spectroscopy
  - Instrumentation and Future Technologies
  - Remote sensing Environment, Analysis and Climate Technologies

- Frequency allocations in Remote Sensing
- Image Analysis and Data Fusion
- Modeling in Remote Sensing
- GRSS Standards for Earth Observations
EARTH SCIENCE INFORMATICS (ESI) TECHNICAL COMMITTEE

Objectives
Advance application of informatics to geoscience and remote sensing

https://www.grss-ieee.org/technical-committees/earth-science-informatics/

Chairs
Peter Baumann
Manil Maskey

TWO WORKING GROUPS

High-performance and Disruptive Computing in Remote Sensing (HDCRS)
Gabriele Cavallaro
Dora Blanco Heras
Jin Sun

Databases in Remote Sensing (DBRS)
Dai-Hai Ton That
Kesheng (John) Wu
Khalid Belhajjame
High Performance and Disruptive Computing in Remote Sensing Working Group

Main Objective:
Connect and support the community of interdisciplinary researchers in remote sensing who are specialized in emerging computing paradigms

The idea:
Innovative computing technologies applied to efficient computation of remote sensing problems

COME ABOARD

About Earth Science Informatics

MISSION

The Earth Science Informatics Technical Committee (ESI TC) provides a venue for informatics professionals to exchange ideas and share knowledge. It aims at advancing application of informatics to geosciences and remote sensing, assessing technology to support data stewardship and management, and promoting best practices and lessons learned.

The mission of the ESI TC is to bring together informatics experts and practitioners to share ideas and information to support open science and maximize the use of science data for research and applications.

https://www.grss-ieee.org/technical-committees/earth-science-informatics/
HDCRS Activities
HDCRS SUMMER SCHOOL - 2022

High Performance and Disruptive Computing in Remote Sensing
30 May - 2 June, 2022

Teaching material and videos available at
UPCOMING ACTIVITIES AT IGARSS 2023

Tutorial on End-to-End Machine Learning with Supercomputing and in the Cloud

Hands-on experience in Supercomputer Systems and Cloud Computing Services for Remote Sensing Applications

Community Contributed Sessions

(1) Scalable Parallel Computing for Remote Sensing
(2) Quantum computing next generation HPC
(3) Quantum Machine Learning algorithms for EO

https://2023.ieeeigarss.org/
UPCOMING ACTIVITIES AT M2GARSS 2024

2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium
15 - 17 April 2024 • Oran, Algeria • Hybrid: Online Platform

Tutorial on “Quantum Computing for Earth Observation”
https://2024.m2garss.org/tutorials.php
ABOUT THIS SCHOOL

Network with students and young professionals, as well as senior researcher and professors who work on interdisciplinary research in remote sensing.

- Supercomputing
- Distributed computing
- Specialized hardware computing
- Edge computing
- Blockchain
- Parallel Programming Models
- Quantum Computing
- Neuromorphic computing

Will be organized every year and cover different research topics.
# AGENDA – MONDAY MAY 29TH

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30 - 09:45</td>
<td>Welcome at the University of Iceland and Opening of the Summer School</td>
<td>Jón Atli Benediktsson</td>
</tr>
<tr>
<td>09:45 - 10:30</td>
<td>Work and Activities of the HDCRS Working Group</td>
<td>Dora Blanco Heras</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gabriele Cavallaro</td>
</tr>
<tr>
<td>10:30 - 11:00</td>
<td>Coffee break</td>
<td>-</td>
</tr>
<tr>
<td>11:00 - 12:00</td>
<td>Lowering the Barrier for Modern Cloud-based Geospatial Big Data Analysis by Combined Use of Innovative and Traditional Infrastructure</td>
<td>Serkan Girgin</td>
</tr>
<tr>
<td>12:00 – 13:30</td>
<td>Lunch break</td>
<td>-</td>
</tr>
<tr>
<td>13:30 - 15:00</td>
<td>Hyperspectral technology: inspiring ideas, challenges and opportunities</td>
<td>José Francisco López</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roberto Sarmiento</td>
</tr>
<tr>
<td>15:00 - 15:30</td>
<td>Coffee break</td>
<td>-</td>
</tr>
<tr>
<td>15:30 - 16:30</td>
<td>An Overview of the European HPC Strategy and Highlights from the Icelandic HPC Communities</td>
<td>Hemanadhan Myneni</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Morris Riedel</td>
</tr>
<tr>
<td>20:00 - 22:00</td>
<td>Social Dinner</td>
<td>-</td>
</tr>
</tbody>
</table>

*Greenwich Mean Time (GMT)
SOCIAL DINNER
20:00 – 22:00

- Höfnin Restaurant, Geirsgata 7, 101 Reykjavík
## AGENDA – TUESDAY MAY 30TH
Data Science at Scale: Harmonized Landsat Sentinel (HLS) Case Study

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 - 10:30 (GMT)</td>
<td>Data Science Overview, Large Scale Data Harmonization</td>
</tr>
<tr>
<td>10:30 - 11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00 - 12:00</td>
<td>Analysis and Exploration: Analytics Platform, Science Analysis</td>
</tr>
<tr>
<td>12:00 - 13:30</td>
<td>Lunch break</td>
</tr>
<tr>
<td>13:30 - 15:30</td>
<td>Application: Geospatial Foundation Model</td>
</tr>
<tr>
<td>15:30 - 16:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:00 - 17:00</td>
<td>Inference with Fine-Tuned Model and Conclusion</td>
</tr>
</tbody>
</table>

Speakers: Manil Maskey, Sean Harkins, Brian Freitag, Muthukumaran Ramasubramanian, Iksha Gurung, Linsong Chu, Paolo Fraccaro, Johannes Jakubik, Blair Edwards

*Greenwich Mean Time (GMT)*
## AGENDA – WEDNESDAY MAY 31TH
Remote Sensing Deployable Analysis Environment

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 - 10:30</td>
<td>HPC, RSDAT, and the EO Software Ecosystem</td>
</tr>
<tr>
<td>10:30 - 11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00 - 12:00</td>
<td>Deployment with RS-DAT and Data Retrieval</td>
</tr>
<tr>
<td>12:00 - 13:30</td>
<td>Lunch break</td>
</tr>
<tr>
<td>13:30 - 15:00</td>
<td>Scaling EO workflows with HPC</td>
</tr>
<tr>
<td>15:00 - 15:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>15:30 - 17:00</td>
<td>Hands-on session</td>
</tr>
</tbody>
</table>

Speakers: Francesco Nattino, Meiert Willem Grootes, Pranav Chandramouli

*Greenwich Mean Time (GMT)*
# AGENDA – THURSDAY JUNE 1\textsuperscript{TH}

Quantum Computing for Earth Observation

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker/Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 - 10:30</td>
<td>Introduction to Quantum Computing and its Ecosystem</td>
<td>Riccardo Mengoni</td>
</tr>
<tr>
<td>10:30 - 11:00</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>11:00 - 12:00</td>
<td>Introduction to Quantum Computing and its Ecosystem</td>
<td>Riccardo Mengoni</td>
</tr>
<tr>
<td>12:00 - 13:30</td>
<td>Lunch break</td>
<td></td>
</tr>
<tr>
<td>13:30 - 14:30</td>
<td>ESA Quantum Computing for Earth Observation (QC4EO): Current Activities and Perspectives</td>
<td>Bertrand Le Saux</td>
</tr>
<tr>
<td>14:30 - 15:30</td>
<td>Is Space ready for the Quantum Leap? A Thales Alenia Space Perspective on Quantum Technologies for Earth Observation</td>
<td>Mattia Verducci, Tommaso Catuogno</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>Closing</td>
<td></td>
</tr>
</tbody>
</table>

*Greenwich Mean Time (GMT)*
Thank you for your attention