Programme Seminar *Big Data Processing in Astronomy and Earth Observation*

10:00	Opening by **Maurice Bouwhuis (SURF)** Maurice Bouwhuis is innovation-programme manager at SURF and in charge of the program that deals with the relationship and development projects in which SURF partners with the different Research Communities.
10:15	**Measuring Air Quality and Greenhouse Gas Emission from Space using TROPOMI: Past, Presence and Future** The Tropospheric Monitoring Instrument (TROPOMI) aboard the ESA Sentinel 5 Precursor (S5P) satellite is the first of the ESA Sentinel missions of the Copernicus program to monitor air quality and climate change. Since 2018, the mission has provided continuous observations of the global distribution of a suite of atmospheric trace gases with a moderate spatial resolution of about 5×7 km^2. With a focus on NO_2, CO, and CH_4, we will demonstrate how TROPOMI has kept an eye on our efforts toward a clean, healthy, and sustainable environment for more than six years in space. To further explore the extensive experience in the Netherlands on atmospheric remote sensing using satellite observation, a Dutch consortium of SRON, KNMI, TNO, and ISISspace realizes a next-generation mission. The Twin Anthropogenic Greenhouse Gas Observers (TANGO) mission is a pioneering Cubsat satellite mission that comprises two satellites measuring CO_2, CH_4, and NO_2 with a spatial resolution of 300x300 m^2 for selected target areas. Tango will be launched in 2027 and envisages a unique European contribution to monitoring the emission of localized point sources of greenhouse gases. We will outline the Tango mission concept and the synergistic observation strategy with a global survey mission like TROPOMI, paving the way for future air quality and greenhouse gas emission monitoring from space. **Jochen Landgraf (SRON)** Jochen Landgraf is senior scientist at SRON Netherlands Institute for Space Research leading a group for atmospheric trace gas remote sensing. His research is focused on remote sensing from satellite measurements, particularly from TROPOMI, Sentinel-5, OCO-2, GOSAT-1 and GOSAT-2 observations. He is lead scientist of the TANGO-Carbon instrument that will measure CO2 and CH4 from space.
10:35	**Gaia: the two-billion-star surveyor** The European Space Agency's Gaia mission is observing well over two billion astronomical sources ranging from asteroids in the solar system to stars in the Milky Way to the very distant quasars. I will briefly summarize the mission aims and then describe the data processing infrastructure, the types of data produced, and how the results are stored and made available to the astronomical community. I will also touch on the use of HPC and AI in the data processing for Gaia. **Anthony Brown (UL)** Anthony Brown is associate professor at Leiden Observatory, Leiden University, and has been involved in the ESA Gaia mission since 1997. He currently chairs the Gaia Data Processing and Analysis Consortium, a team of about 450 European astronomers, IT, and project specialists who are in charge of turning the raw measurements from the Gaia spacecraft into a three-dimensional map of over two billion stars in our home galaxy, the Milky Way. Anthony is very broadly interested in the astronomical research that can be done with the aid of Gaia data, from studies of our own solar system to understanding the formation history of the Milky Way.
10:55	**Serving satellite observations of the Earth system** Clearly, climate change results in more extreme weather and associated ocean anomalies, hence calling for improved weather advisories in the future for loss reduction and more accurate climate projections for adaptation. A better understanding of Earth's dynamics may provide benefits in both areas. Winds determine weather, hurricanes, waves and surges, energy production, ocean forcing, heat and carbon budgets, climate change, andsoforth. Hence, satellite winds are extensively used by marine forecasters, in NWP, by oceanographers, wind engineers, off-shore industry, safety authorities and climate scientists alike. More dynamical information is becoming

	available due to new and extending international virtual constellations of satellite wind-sensing instruments. This new information is hence beneficial for above-mentioned applications, while broad user implementation depends on a chain of standards in calibration, data processing, quality control and quality monitoring on different levels. In the presentation focus will be put on the EUMETSAT Satellite Application Facilities (SAF), the EU Copernicus Services and on international collaboration in ESA Data Innovation Science Clusters (DISC). **Ad Stoffelen (KNMI)** Ad Stoffelen (fellow IEEE) is currently responsible for a team at KNMI working on active remote sensing with international involvement in ESA, EUMETSAT, NASA, Indian and Chinese missions. For EUMETSAT he manages operational (24/7) and R&D wind activities in the Ocean & Sea Ice SAF, arranging the servicing of ocean surface winds from a virtual constellation of scatterometer (7 scatterometers). The KNMI winds are further provided at a higher processing level in the Copernicus Marine Service Wind TAC for real-time and climate monitoring purposes. Ad's group is moreover closely involved and leading the geophysical processing of data from the ESA Aeolus and EarthCare missions, again in a broad international context. He leads the Ocean Surface Winds task group of the Coordination Group of Meteorological Satellites (CGMS), co-leads the International Ocean Vector Winds Science (IOVWST), is part of the ESA Aeolus Science Advisory Group (SAG), the EUMETSAT/ESA Scatterometer SAG, and part of the ESA Harmony mission SAG. He led and took part in international ESA, EUMETSAT, EU and NOAA projects on extreme winds, storm surges, wind energy, ocean waves and currents, and data assimilation for Numerical Weather Prediction (NWP). He is a professional trainer for weather nowcasting, NWP and climate applications.
11:15	Coffee break
11:45	**Real-time data processing for the BlackGEM Array of telescopes** The BlackGEM Array of telescopes was built to detect optical flashes from the mergers of double neutron stars. Gravitational wave detectors can detect the ripple in space-time that is caused by such a merger, but cannot precisely localise the event on the sky. BlackGEM is designed to quickly follow up on a gravitational wave detection by taking deep images of a large part of the sky to find the faint optical afterglow of the collision. The optical light originates from nuclear reactions that take place during the merger and are thought to be the main sites of heavy element production in the Universe. The BlackGEM telescopes have become operational in the Atacama desert in Chile in 2023. In this talk we discuss the cloud-based real-time data processing techniques and infrastructure, which makes science-ready data products available to the BlackGEM consortium within 20 minutes after the data was taken in Chile. We also discuss the BlackHoleFinder citizen science app, through which the general public helps the project to identify events that deserve further attention. **Steven Bloemen (RU)** Steven has a PhD in astrophysics (KU Leuven, 2013) and is now project manager and deputy director at the Radboud Radio Lab, the astrophysics instrumentation group at Radboud University. As the project manager of the BlackGEM Array he oversees the realisation of the telescopes in Chile, from the design phase to the operations phase.
12:05	**Ruisdael Observatory: (Atmospheric) Clouds of Data** Current challenges in atmospheric studies are to understand and to represent as explicit as possible chaotic phenomena such as cloud formation, its organization, intensification and resulting rain patterns, or the turbulent dispersion of atmospheric pollutants. To this end, one of the main aims of the Ruisdael Observatory is to integrate high-resolution numerical simulations with observations, namely a TestBed in which large amounts of observations are confronted and systematically evaluated with high thee-dimensional high resolution numerical (~20 m or below) performed at very large domains (> 500 km). This team effort leads to three challenges related to the treatment of large amounts of data: the systematic and comprehensive analysis of numerical results, the integration with observations, and the visualizations. All these challenges require novel storage methods that allow for a fast and flexible retrieval of these data. During my presentation, I will discuss current efforts in dealing with

	these three challenges applied to the Netherlands and two other regions sensitive to climate modifications: the Atlantic Oceanic trade winds or the Amazon Basin. **Jordi Vila-Guerau de Arellano (WUR)** Jordi Vilà-Guerau de Arellano is professor at the Meteorology and Air Quality Section in Wageningen University. His research deals with the turbulent phenomena of the clear and cloudy boundary layers. He has a special interest on research topics related to phenomena across scales, and their connection with nearby disciplines such as atmospheric chemistry or photosynthesis. To this end, he develops and uses methodologies based on high-resolution numerical simulations integrating in his analysis the use of advanced observational techniques.
12:25	**Data Intensive Radio Astronomy with LOFAR and SKA** Modern radio telescopes are amongst the most data intensive machines on the planet: our existing instruments are already generating tens of petabytes per year, while new and upgraded systems over the next year will move towards the exascale. Simply collecting that volume of data is already a challenge — securely archiving it, making it available to a broad community, and facilitating data analysis and the scientific process even more so. In this talk, I'll discuss the systems — both technical and organizational — that ASTRON and our colleagues in the International LOFAR Telescope and the SKA Regional Centre Network are putting in place to maximise the impact of these facilities. **John Swinbank (ASTRON)** John Swinbank is the Science Data Centre Programme Manager at ASTRON, the Netherlands Institute for Radio Astronomy, where he is developing techniques and services to make data from cutting-edge radio telescopes — LOFAR, Westerbork/Apertif, and ultimately the Square Kilometre Array — available to the widest possible community in the most scientifically productive way possible. Before joining ASTRON in 2020, he was the Deputy Project Manager for Data Management at the Vera C. Rubin Observatory (formerly LSST), with a particular interest in the software pipelines used to generate both annual data releases and rapid transient alerts, and was also a Research Professor at the University of Washington in Seattle. He has worked on the development of astronomical transient alert systems within the context of the International Virtual Observatory Alliance, and contributed to a wide range of open-source astronomical software, including participating in the governance of the Astropy Project.
12:45	Lunch
13:45	**Gravitational wave data challenges for LISA** The European Space Agency LISA gravitational wave mission will be sensitive to gravitational wave signals in the milli-Hz frequency range, where a large number of different sources with very different signals emit gravitational waves simultaneously. This poses significant challenges to the data processing. But even worse, the signals are buried in laser frequency noise that is many orders of magnitude stronger than the signals. The raw data thus first needs significant processing to subtract the noise. These challenges are tackled by an international consortium of partners from Europe united in the LISA distributed data processing center (DDPC) and the US. **Gijs Nelemans (RU)** Gijs Nelemans studied Physics and Astronomy at Utrecht University and did his PhD on binary systems at the University of Amsterdam. After a post-doc at the Institute of Astronomy of Cambridge University he came to Radboud University, where he is now professor of Gravitational Wave Astrophysics. Together with Elena Rossi he is chairing the LISA-NL consortium in which the Dutch contributions to the LISA mission are brought together.
14:05	**Remote Sensing - Deployable Analysis Environment - power your workflow with HPC** Contemporary research, both in earth observation and astronomy, is often characterized by large data volumes on the scale of tens to hundreds of terabytes and beyond, with these volumes growing incessantly. While enabling novel research, this growth also poses challenges to the exploitation of the data, as the computational and storage resources available on workstations often no longer suffice.

	Here we present the Remote Sensing - Deployable Analysis EnvironmenT (RS-DAT), a readily-deployable and user-friendly framework enabling researchers to scale up their (data intensive) analysis and inference workflows using high-performance/high-throughput computing (HPC/HTC) systems and associated storage resources, while making use of the tools and interfaces (Jupyter Notebooks, Python, etc.) they are already accustomed to. RS-DAT seamlessly integrates and extends the Python and PyData ecosystem, including Dask and Xarray, and provides tools for data storage and retrieval on mass storage systems (e.g., dCache at SURF). Although initially developed and tested on the Snellius, Spider and Surf Research cloud HPC/HTC systems in the context of Earth observation and remote sensing science, RS-DAT is domain agnostic at its core and capable of scaling analysis workflows across a wide range of fields. This session aims at introducing researchers to the capabilities of RS-DAT, while inviting engagement and input on its further development and specific adaptation to provide HPC/HTC powered workflows across different research domains. **Meiert Grootes (NLeSC)** Dr. Meiert Willem Grootes is a Senior Research Software Engineer at the Netherlands eScience Center. After obtaining a PhD in astrophysics from the University of Heidelberg he pursued research on galaxy evolution as an independent postdoctoral fellow at the Max-Planck Institute for Nuclear Physics and ESA, including work on data analysis pipelines for satellite images, machine learning based galaxy classification techniques, radiative transfer models, and statistical analysis with specific focus on the GAMA and KiDS surveys.At the eScience center he is involved in projects from the Environment & Sustainability and Natural Sciences & Engineering sections, with a focus on earth observation, data storage and access, and machine learning. He is further interested in HPC/HTC and hardware acceleration, and their application in research.
14:25	**The National Research IT Infrastructure** SURF is the collaborative ICT organization for Research and Education in the Netherlands. Together with scientists, we develop solutions to tackle complex research problems by adopting the strengths of the latest technologies. In this talk we discuss the various platforms and services offered by SURF for computing, data storage, dynamic scaling. We also discuss about emerging technologies and the interaction between different technological domains that play a crucial role for designing future infrastructures and applications. **Natalie Danezi (SURF)** Natalie Danezi is a senior technical consultant at SURF. Holder of BSc and MSc degrees in Computer Science, she is specialised in computational and software techniques in engineering and distributed systems technologies such as grid and cloud computing. The core of her work is assisting researchers from various scientific fields to build their own production platforms on top of powerful data processing facilities.
14:45	**Break-out sessions** moderated by **Daniela Huppenkothen (SRON)** Daniela Huppenkothen is a staff scientist and NWO WISE fellow at the SRON Netherlands Institute for Space Research, where she works at the interface of astronomy, statistics, and machine learning. She is particularly interested in how we can use modern statistical tools and machine learning methods about the universe. Her interdisciplinary research has made her curious about how researchers collaborate, and how to facilitate interdisciplinary collaborations. She has run numerous hackathons and unconferences and organizes a meta-unconference about participant-driven events called “Hack the Hackathon”. Previously, she was the Associate Director for the DIRAC Institute at the University of Washington, and a Moore-Sloan Data Science Fellow at New York University’s Center for Data Science. She holds a PhD in astronomy and astrophysics from the University of Amsterdam.
16:15	**Global Water Watch** Global Water Watch (GWW) provides free, globally accessible, near-real-time information on water. GWW does so by both an interactive website, as well as an API for interacting with the data

	programmatically. Behind GWW is an algorithm that requires a global Petabyte-sized dataset of the full history of multiple optical satellite missions. To overcome these data challenges, global water watch makes use of many state-of-the-art technologies and services. During this talk we describe the platform's architecture and the lessons learned so far. **Jaap Langemeijer (Deltares)** Jaap Langemeijer works at Deltares at the Deltares Software Centre. He works mostly on projects and products that are focused on Earth Observation and other large geospatial data, as well as automation. Before his career at Deltares, Jaap worked as a consultant "Big Data" at Capgemini, and studied geophysics at the University of Utrecht.
16:35	**The Euclid NL Data Centre** The Euclid satellite will deliver nearly ~2 Million 0.5 Square degree images resulting in a 3D map of 2 Billion Galaxies covering the evolution of the Universe in the last 10 Billion year. The formal NL contribution to the Euclid satellite is provided by a National data centre in Groningen, and support from Leiden on the Infrared pipeline. Apart from providing compute and storage resources to an international network of 9 data centres, we also build and delivered the distributed archive, storage and processing systems to the network and a mirrored the database to ESAC. The system is inspired by the AstroWISE system build for OmegaCAM at the VST. **Edwin Valentijn (RUG)** Edwin A. Valentijn is prof astronomical information Technology at the Kapteyn Astronomical Institute Groningen, Founder of AstroWISE (2003), Founder of Euclid (2008), National project manager for the Euclid satellite, chair of the Information Universe and author of "Powers of Two" (Springer 2022). His research focusses on the nature of Dark Matter.
16:55	**Closing & Drinks**