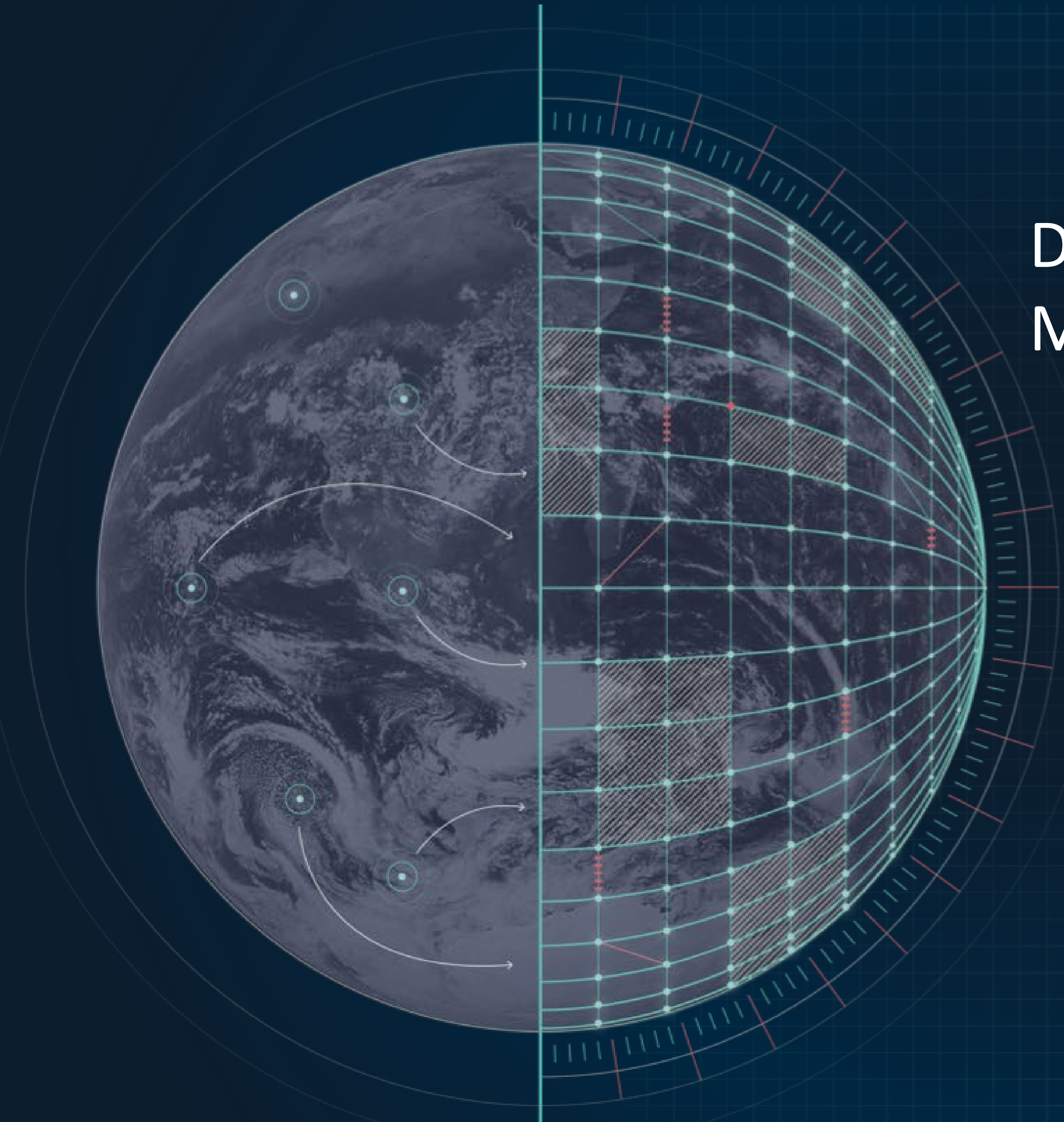


DESTINATION EARTH

DIGITAL TWINS – MODELLING-OBSERVATION FUSION

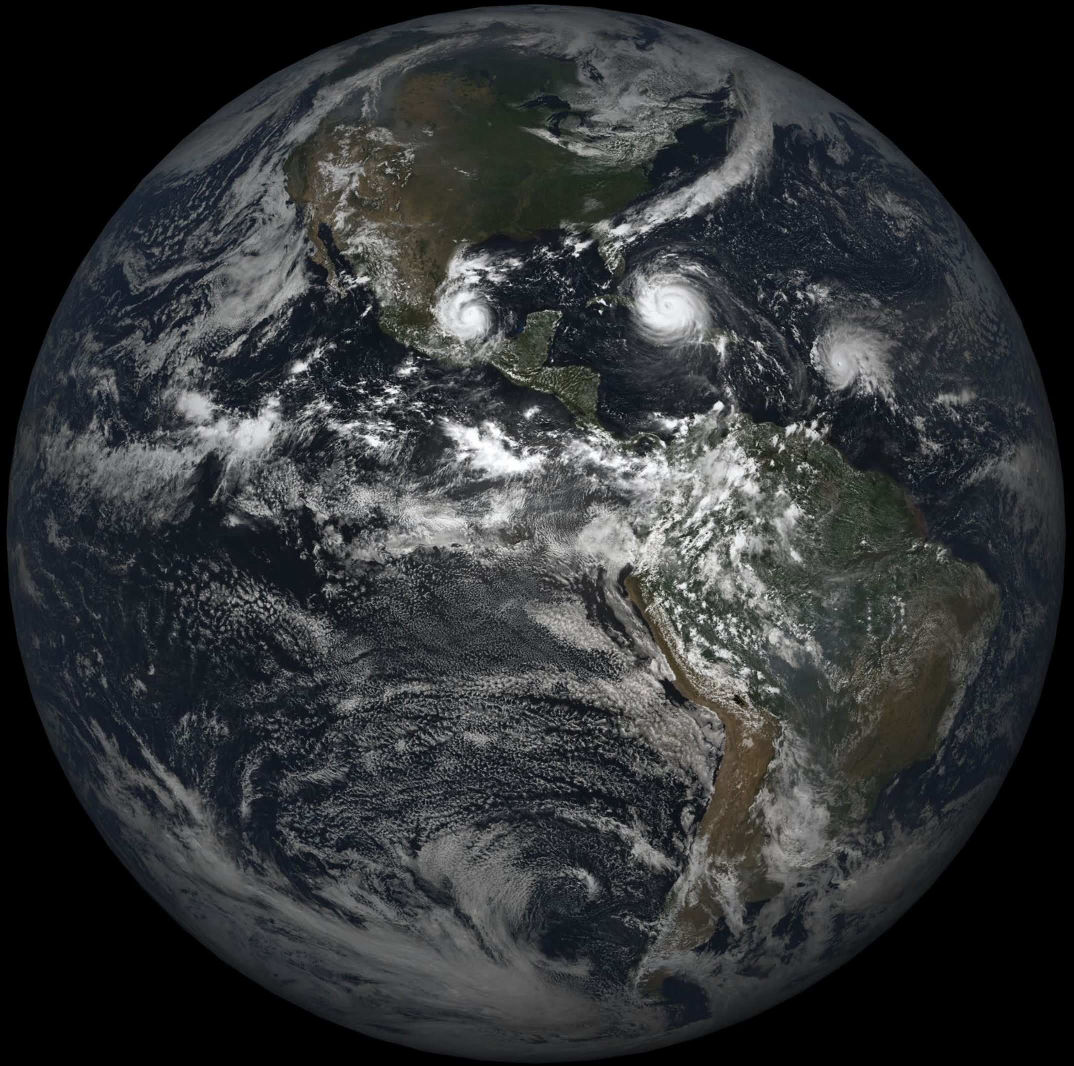
Nils Wedi & Peter Lean

Funded by the European Union

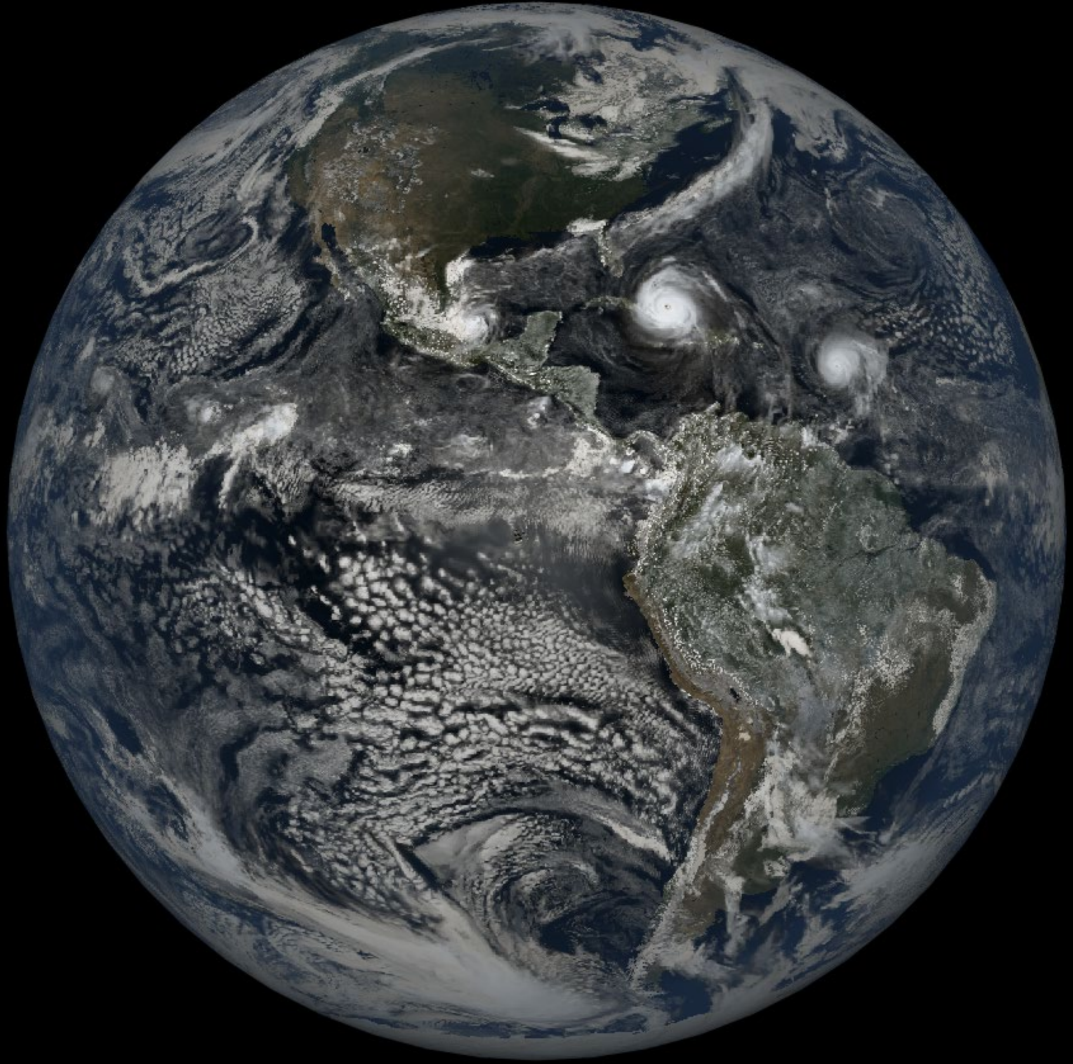
ECMWF

RTTOV-MFASIS: simulated imagery in the visible..

GOES16_ABI CH2_3_1 composite 20170908 1800 UTC



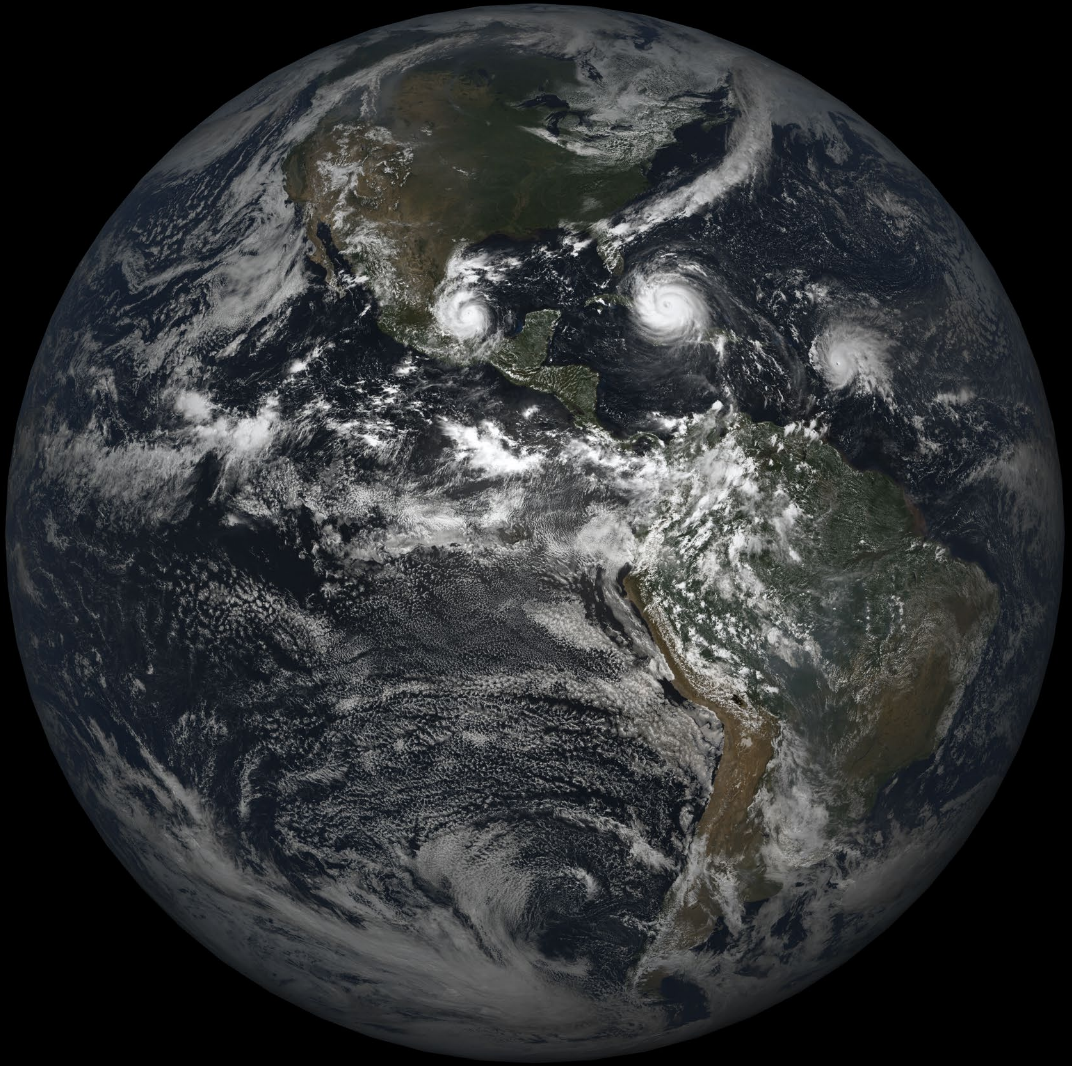
IFS FC+18h at 9 km (oper)



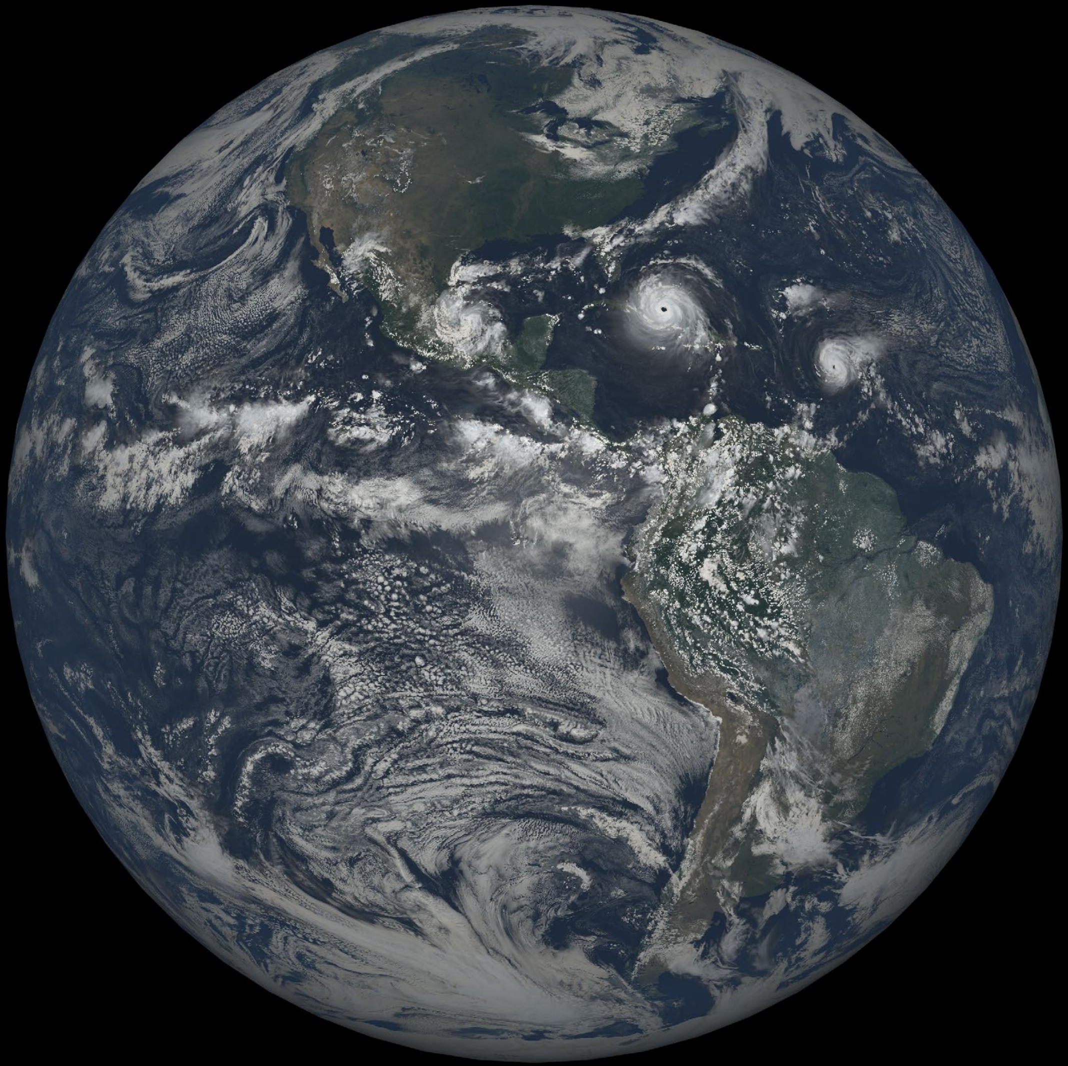
Philippe Lopez

RTTOV-MFASIS: simulated imagery in the visible..

GOES16_ABI CH2_3_1 composite 20170908 1800 UTC



IFS FC+18h at 2.5 km



Philippe Lopez

Observations: per 12-hour assimilation cycle

24 Billion

Incoming observations

500 Million

Pass pre-processing and ingested into ECMWF's IFS model

30 Million

Assimilated by IFS data assimilation

351 Billion

1km global IFS model simulation points

HIGH-LEVEL SCHEMATIC OF NUMERICAL WEATHER PREDICTION (NWP)

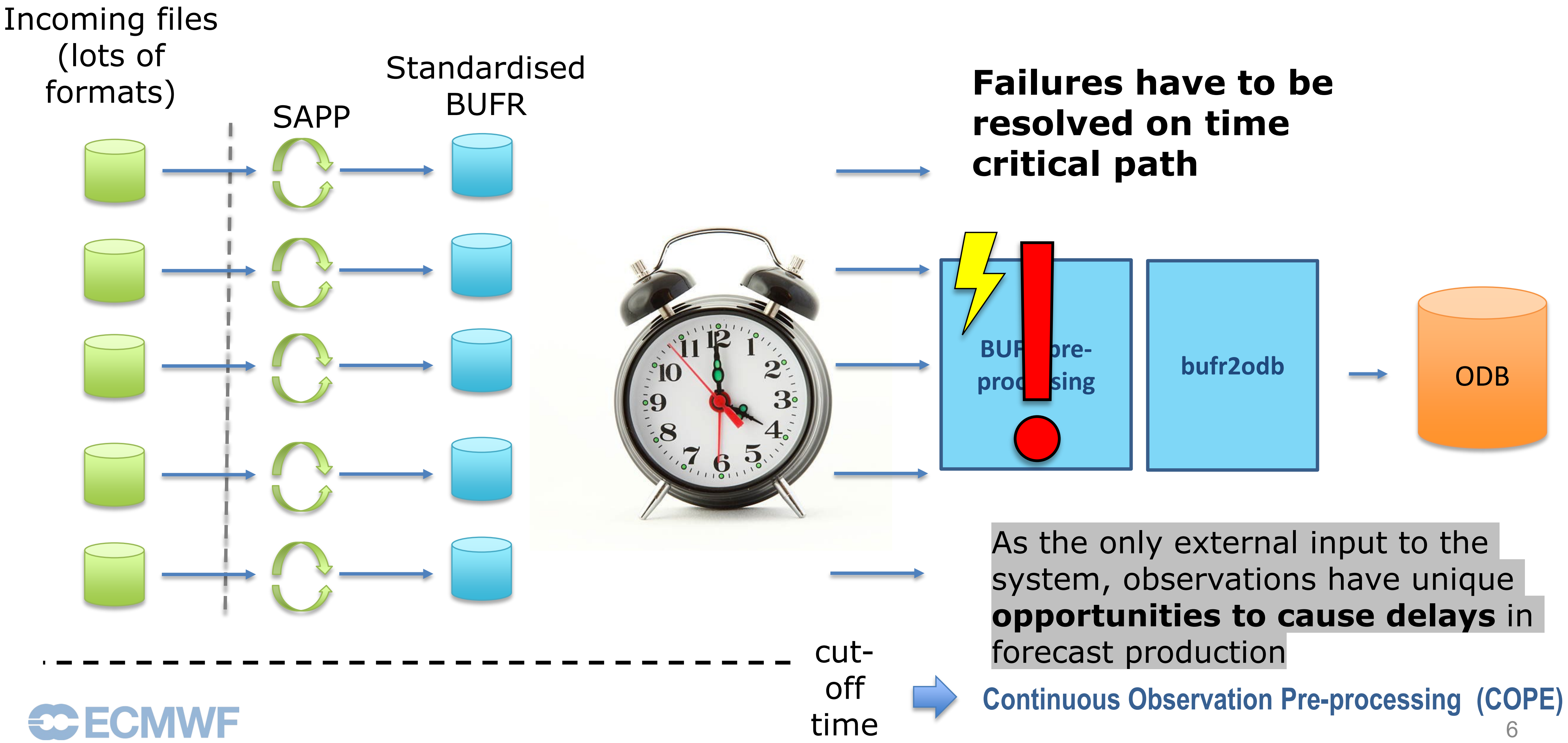
Input: observations

Output: forecasts

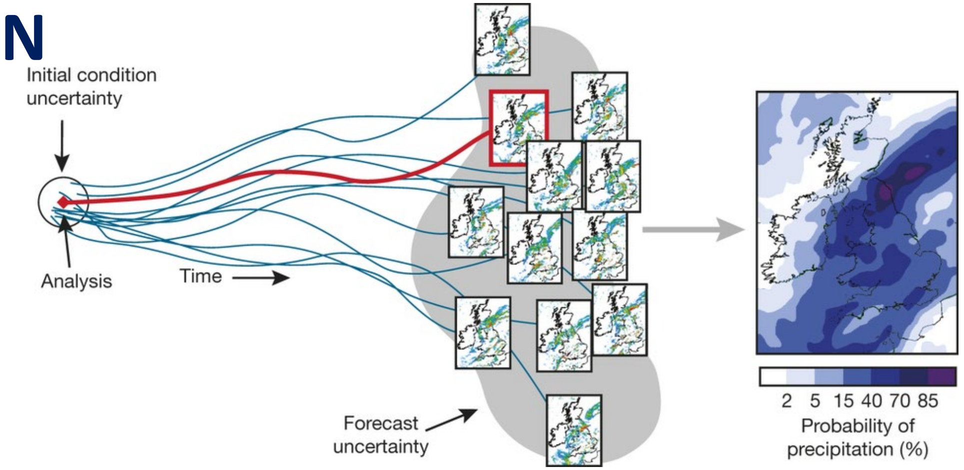
- *In an ideal world*, forecasts would be produced instantaneously and use all observations that have been made up until that time, **but**

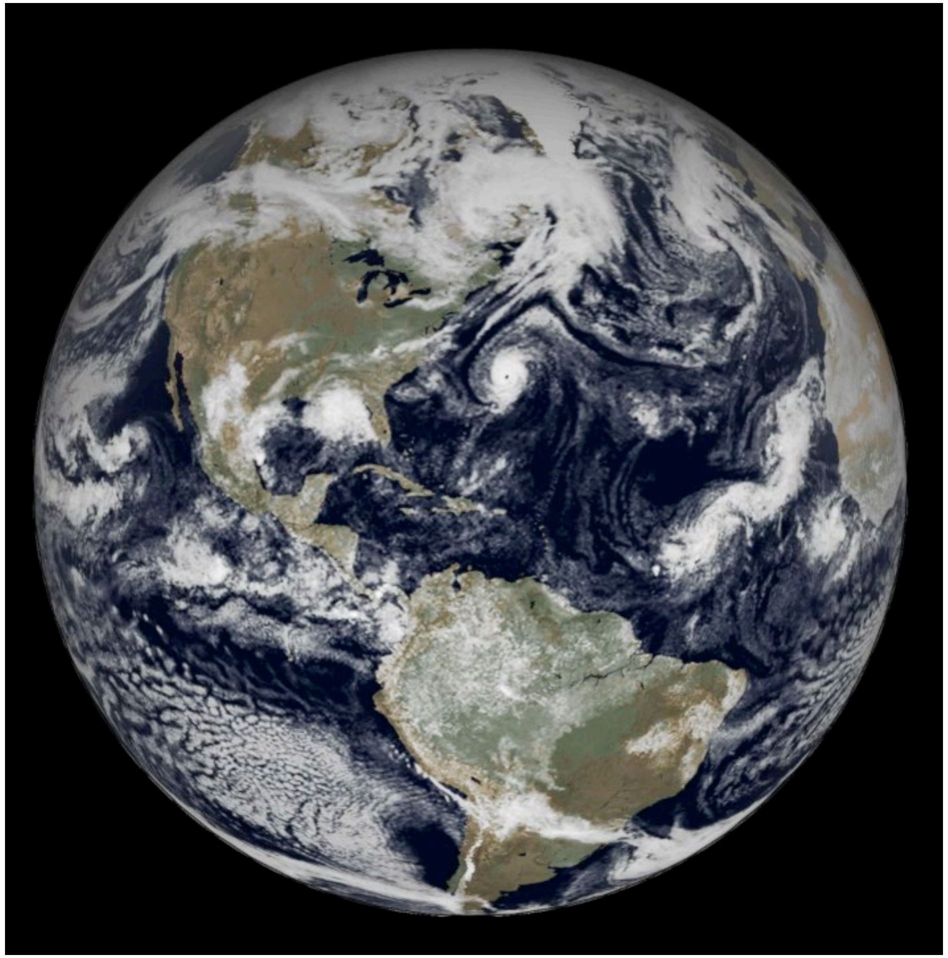
- **Processing time**; generating an analysis and forecast is computationally expensive
  - ➢ By the time the forecast is completed, the observations that went into producing it may be several hours old (which is significant)
  - ➢ Minimising the time spent on observation processing and data assimilation is crucial if we are to gain maximum value from the composition of all observations

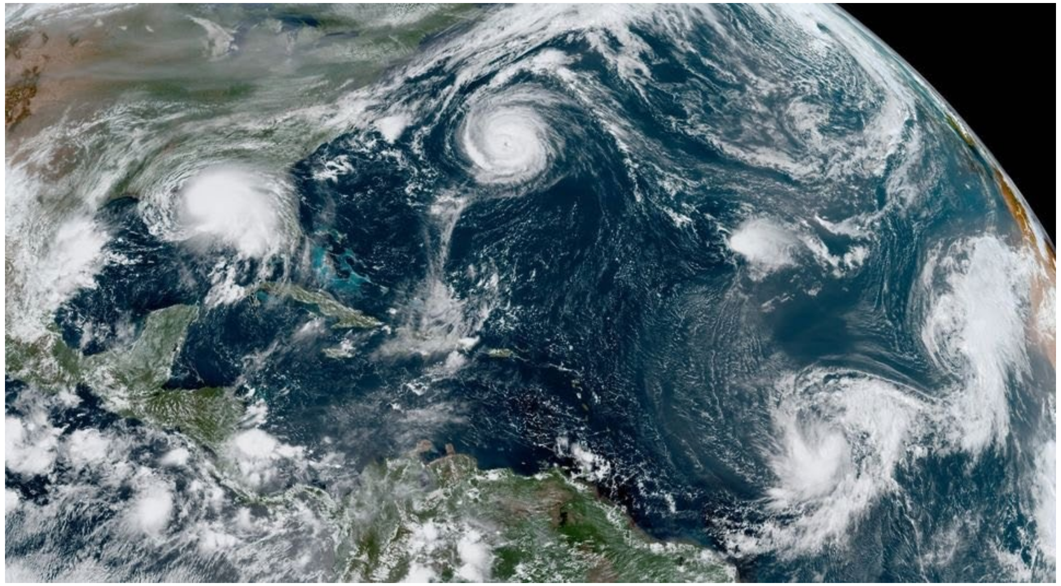
FAULT TOLERANCE OF OBSERVATION PRE-PROCESSING ENGINE



UNCERTAINTY QUANTIFICATION







NOAA

~9km global grid 51 Ensemble members 20200913 00 UTC + 41 h

Simon Lang & Irina Sandu

Data volume of the global ensemble state today
~ 900GB per time step
Hourly data in a 15day forecast -> ~300TB
Time-critical production in 1 hour real time.

ECMWF

EUROHPC: €8 BILLION PROGRAMME TO TAKE US TOWARDS EXASCALE

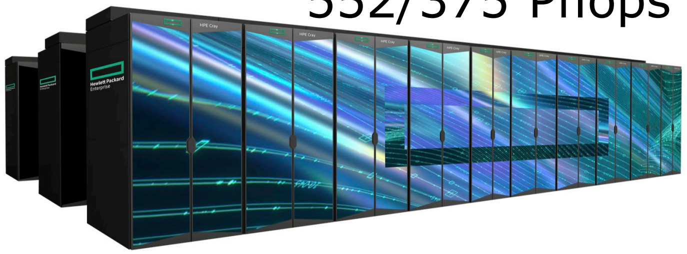
- 3 large (O(100PFlops)) supercomputers in Finland, Italy, Spain
- 5 smaller ones (size of Archer in UK) in Luxembourg, Slovenia, Portugal, Czech Republic, Bulgaria
- 1 high-end supercomputer (~1000 Pflops) by 2023-2024 tbd

Supercomputers

Currently six EuroHPC supercomputers are under construction across Europe:

LUMI

LUMI Finland,
552/375 Pflops

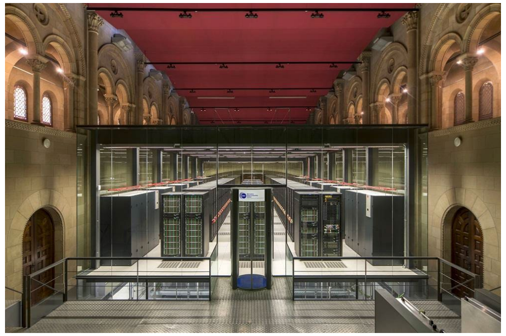


© HPE

The LUMI system will be a Cray EX supercomputer supplied by Hewlett Packard Enterprise (HPE) and located in Finland.

Sustained performance:	375 petaflops
Peak performance:	552 petaflops
Compute partitions:	GPU partition (LUMI-G), x86 CPU-partition (LUMI-C), data analytics partition (LUMI-D), container cloud partition (LUMI-K)
Central Processing Unit (CPU):	The LUMI-C partition will feature 64-core next-generation AMD EPYC™ CPUs
Graphics Processing Unit (GPU):	LUMI-G based on the future generation AMD Instinct™ GPU
Storage capacity:	LUMI's storage system will consist of three components. First, there will be a 7-petabyte partition of ultra-fast flash storage, combined with a more traditional 80-petabyte capacity storage, based on the Lustre parallel filesystem, as well as a data management service, based on Ceph and being 30 petabytes in volume. In total, LUMI will have a storage of 117 petabytes and a maximum I/O bandwidth of 2 terabytes per second
Applications:	AI, especially deep learning, and traditional large scale simulations combined with massive scale data analytics in solving one research problem
Other details:	LUMI takes over 150m2 of space, which is about the size of a tennis court. The weight of the system is nearly 150 000 kilograms (150 metric tons)

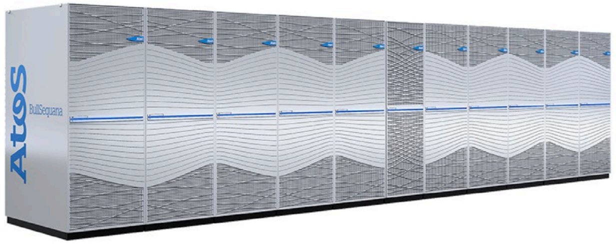
The LUMI consortium includes the Swiss CSCS, building Alps ~300-500 Pflops by 2023



MareNostrum 5
~200 Pflops

LEONARDO

Leonardo Italy:
322/249 PFlops



© Atos

LEONARDO will be supplied by ATOS, based on a BullSequana XH2000 supercomputer and located in Italy.

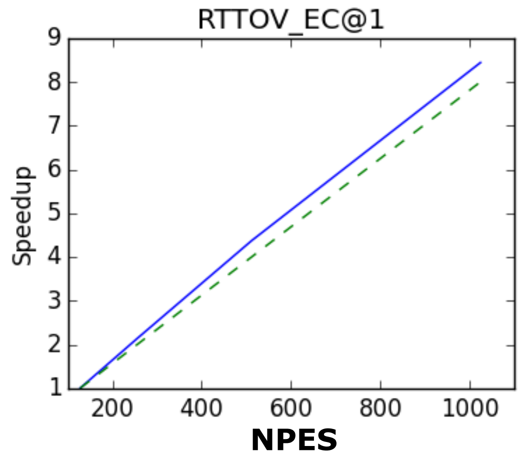
Sustained performance:	249.4 petaflops
ak rformance:	322.6 petaflops
mpute rtitions:	Booster, hybrid CPU-GPU module delivering 240 PFlops, Data-Centric, delivering 9 Pflops and featuring DDR5 Memory and local NVM for data analysis
ntral ocessing it (CPU):	Intel Ice-Lake (Booster), Intel Sapphire Rapids (data-centric)
aphics ocessing it (GPU):	NVIDIA Ampere architecture-based GPUs, delivering 10 exaflops of FP16 Tensor Flow AI performance
orage pacity :	Leonardo is equipped with over 100 petabytes of state-of-the-art storage capacity and 5PB of High Performance storage
plications:	The system targets: modular computing, scalable computing applications, data-analysis computing applications, visualization applications and interactive computing applications, urgent and cloud computing
her details:	Leonardo will be hosted in the premises of the Tecnopolo di Bologna. The area devoted to the EuroHPC Leonardo system includes 890 sqm of data hall, 350 sqm of data storage, electrical and cooling and ventilation systems, offices and ancillary spaces

The current proposal is to set aside a maximum of **10% of the Union's access time for strategic initiatives**.

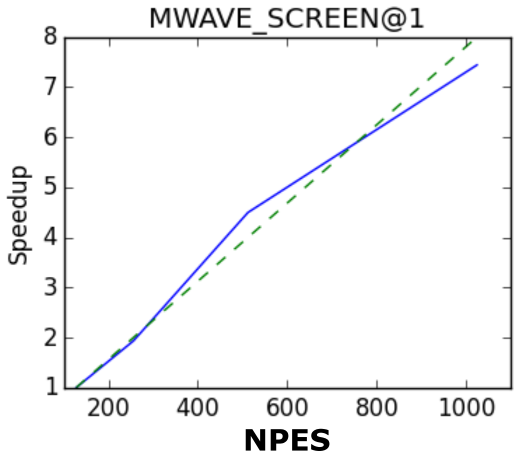
Courtesy P.L. Vidale

Fortunately, most current observation processing scales well on HPC

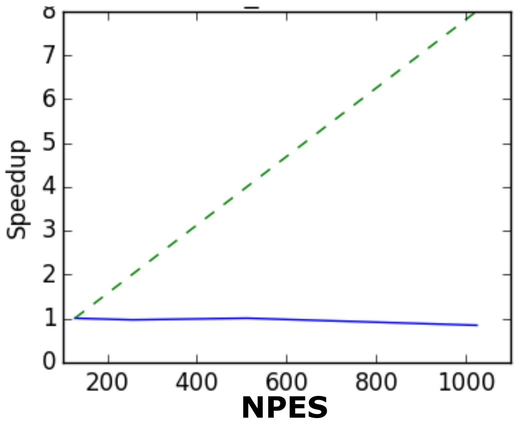
Radiative transfer;
(calculation of observation equivalents)



MW satellite quality control



Thinning



Bottleneck : legacy thinning algorithm involves a global sort of the data on a single MPI task

Most current **observation processing** is inherently an **embarrassingly parallel** problem and scales easily

However, there are a few notable **exceptions** e.g. old **thinning algorithm** which requires **a global view of the data.**

The **data assimilation algorithm(s)** scale similar to the model, but may need **careful adaptation to emerging HPC** architectures

ECMWF

DESTINATION EARTH

TECHNOLOGY

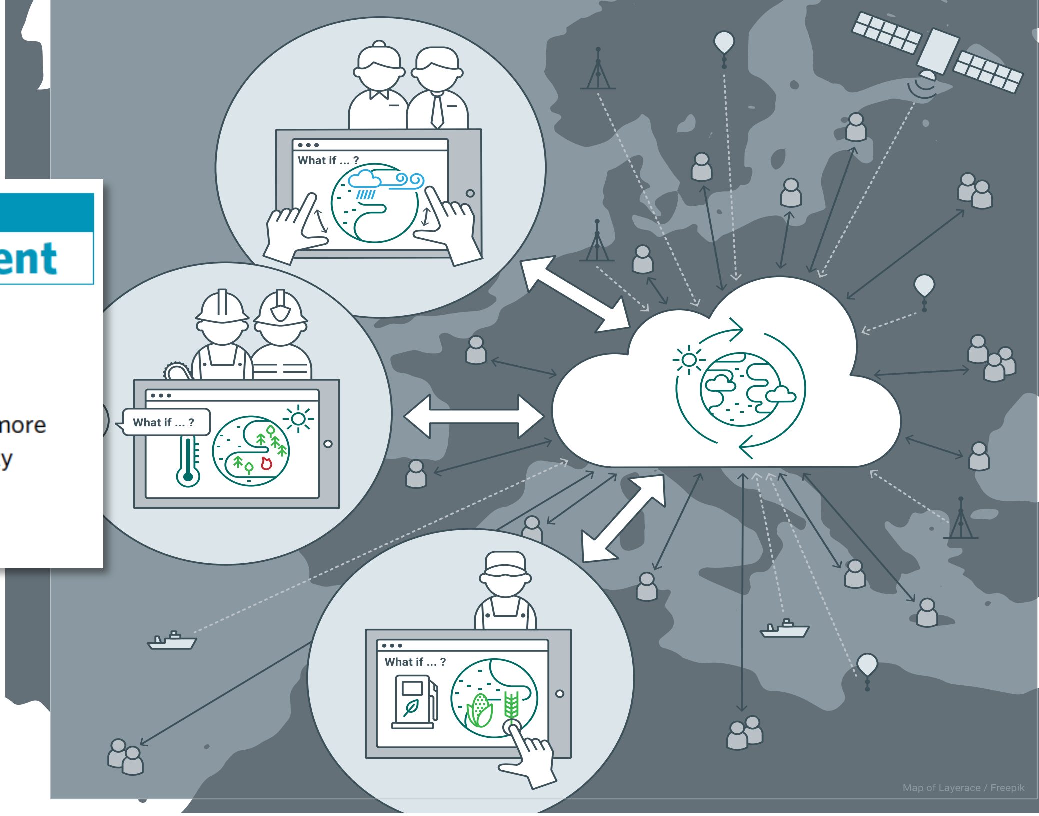
Check for updates

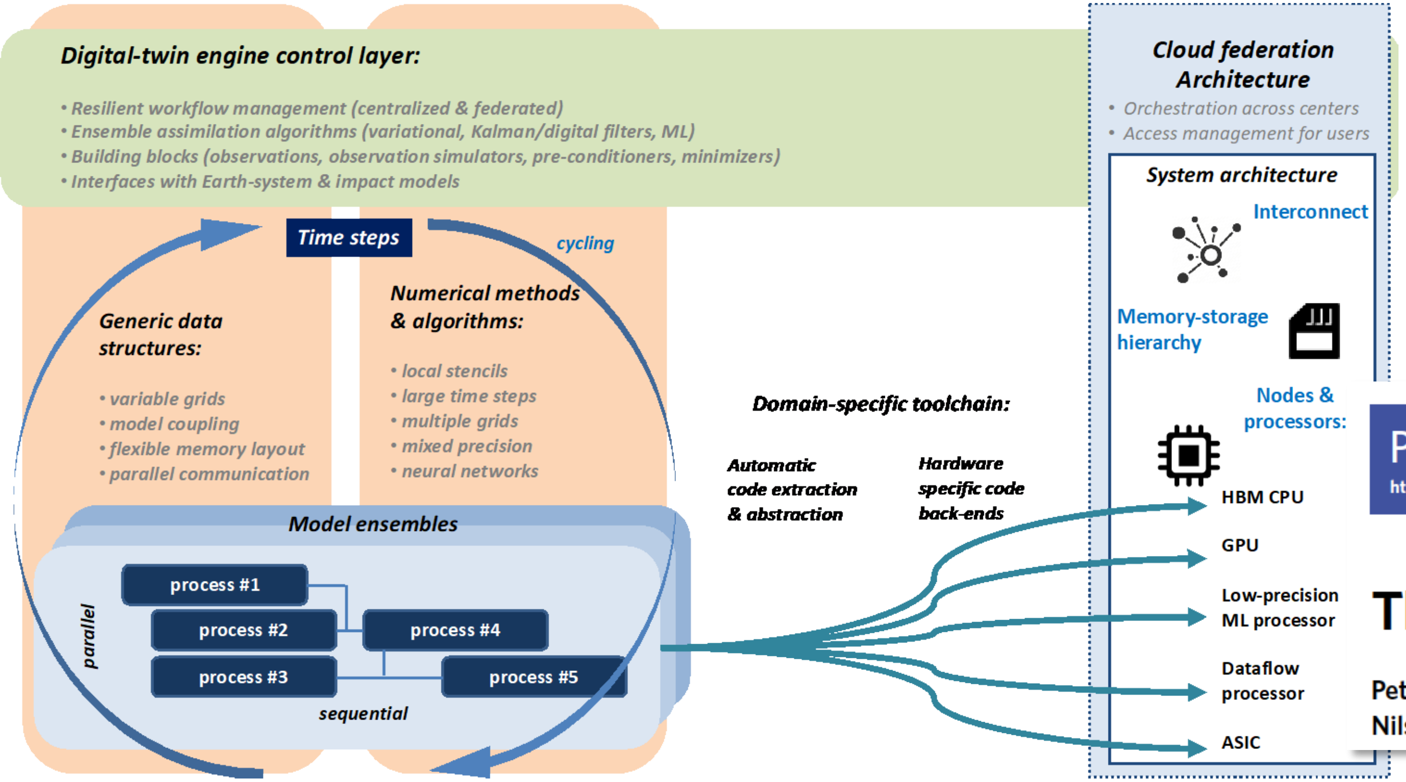
comment

A digital twin of Earth for the green transition

For its green transition, the EU plans to fund the development of digital twins of Earth. For these twins to be more than big data atlases, they must create a qualitatively new Earth system simulation and observation capability using a methodological framework responsible for exceptional advances in numerical weather prediction.

Peter Bauer, Bjorn Stevens and Wilco Hazeleger



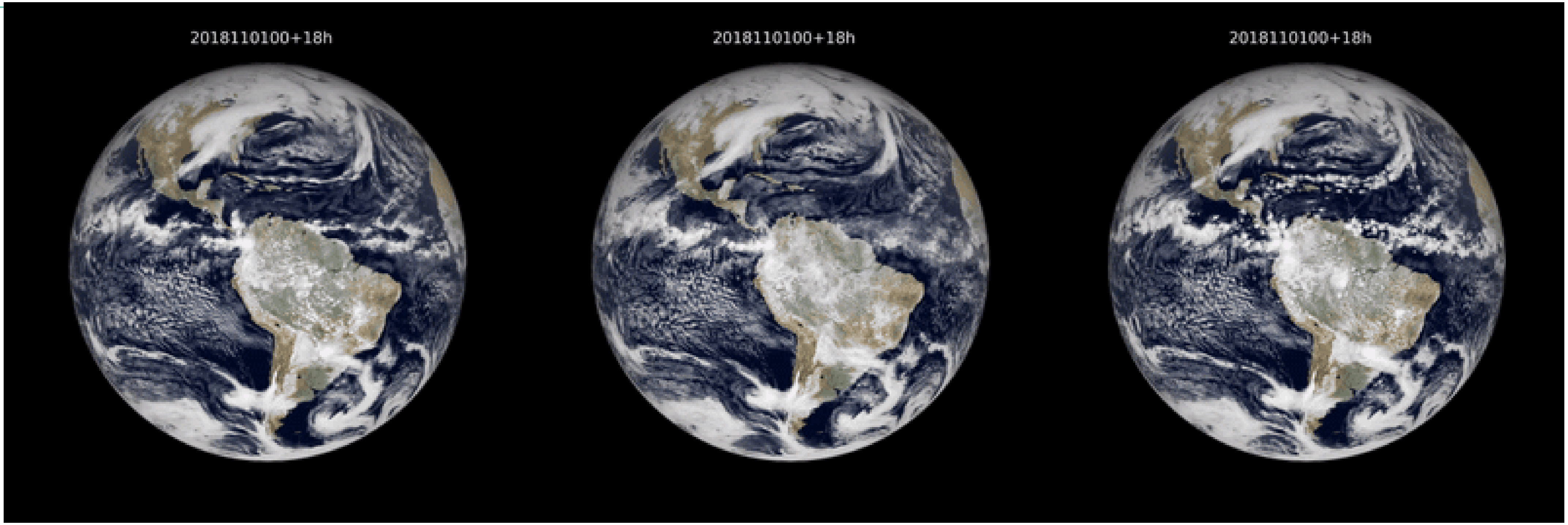


PERSPECTIVE
https://doi.org/10.1038/s43588-021-00023-0
nature computational science
Check for updates

The digital revolution of Earth-system science

Peter Bauer[1], Peter D. Dueben[1], Torsten Hoefler[2], Tiago Quintino[3], Thomas C. Schulthess[4] and Nils P. Wedi[1]

3-HOURLY ACCUMULATED RADIATIVE FLUXES AT THE TOP OF THE ATMOSPHERE



1.4 km **9 km** **9 km w/o deep**

Wedi et al, James 2020

Seasonal simulations at 1km grid-spacing, which can explicitly resolve storms, associated with deep precipitating clouds, the effects of the landscape on the atmosphere, the effects of ocean eddies on ocean heat transport and its interaction with ice-sheets. These may be used for **future observing system simulation experiments** (OSSE).

Global km-scale forecasts & projections

summit

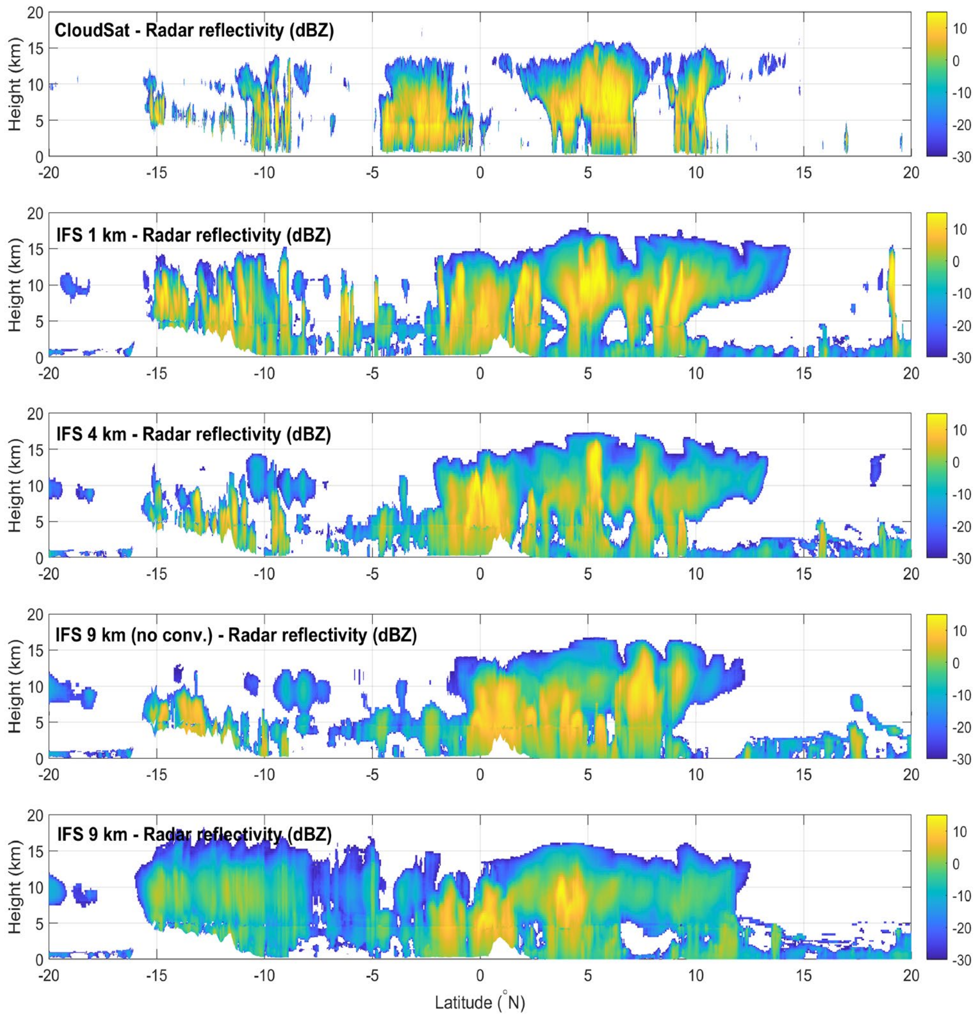
This research used resources of the ***Oak Ridge Leadership Computing Facility***, which is a DOE office of Science User Facility supported under contract DE-AC05-00OR22725.

ECMWF

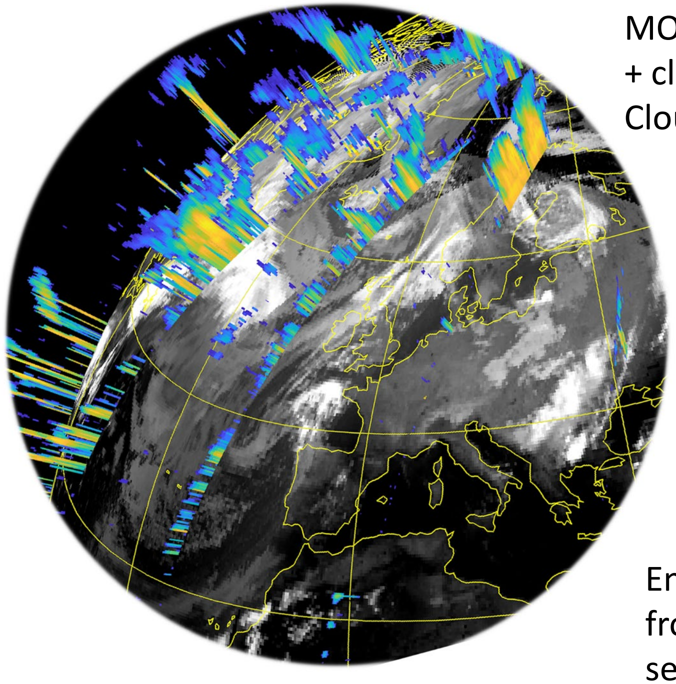
DESTINATION EARTH

DATA ASSIMILATION

Example: EarthCARE mission preparation

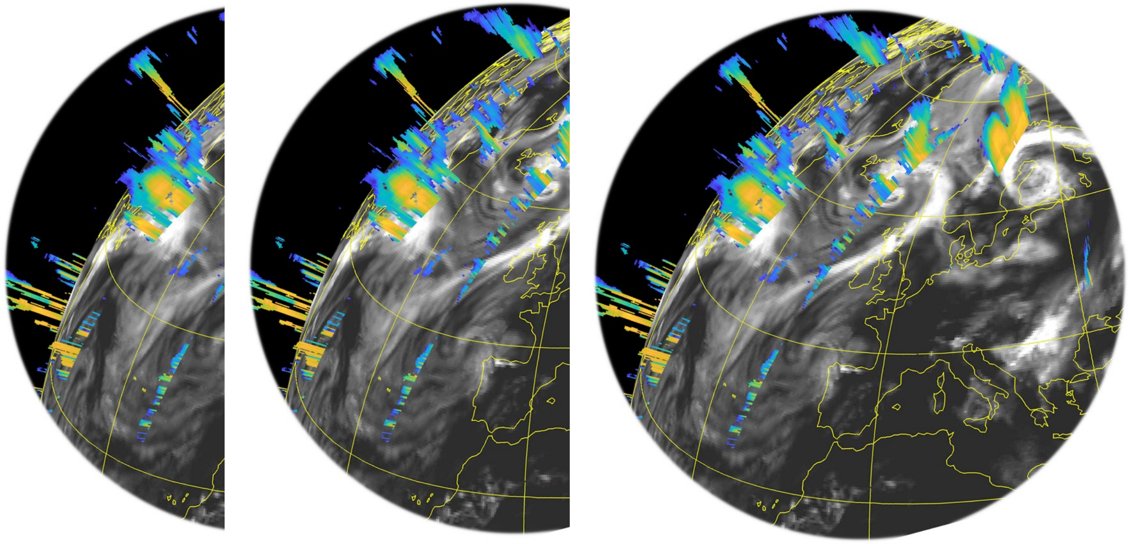


Direct comparison **in observation space**



MODIS aqua infrared channel
+ cloud radar cross sections from CloudSat

Ensemble of simulated satellite images from the IFS model with analysed cross-section cloud profiles



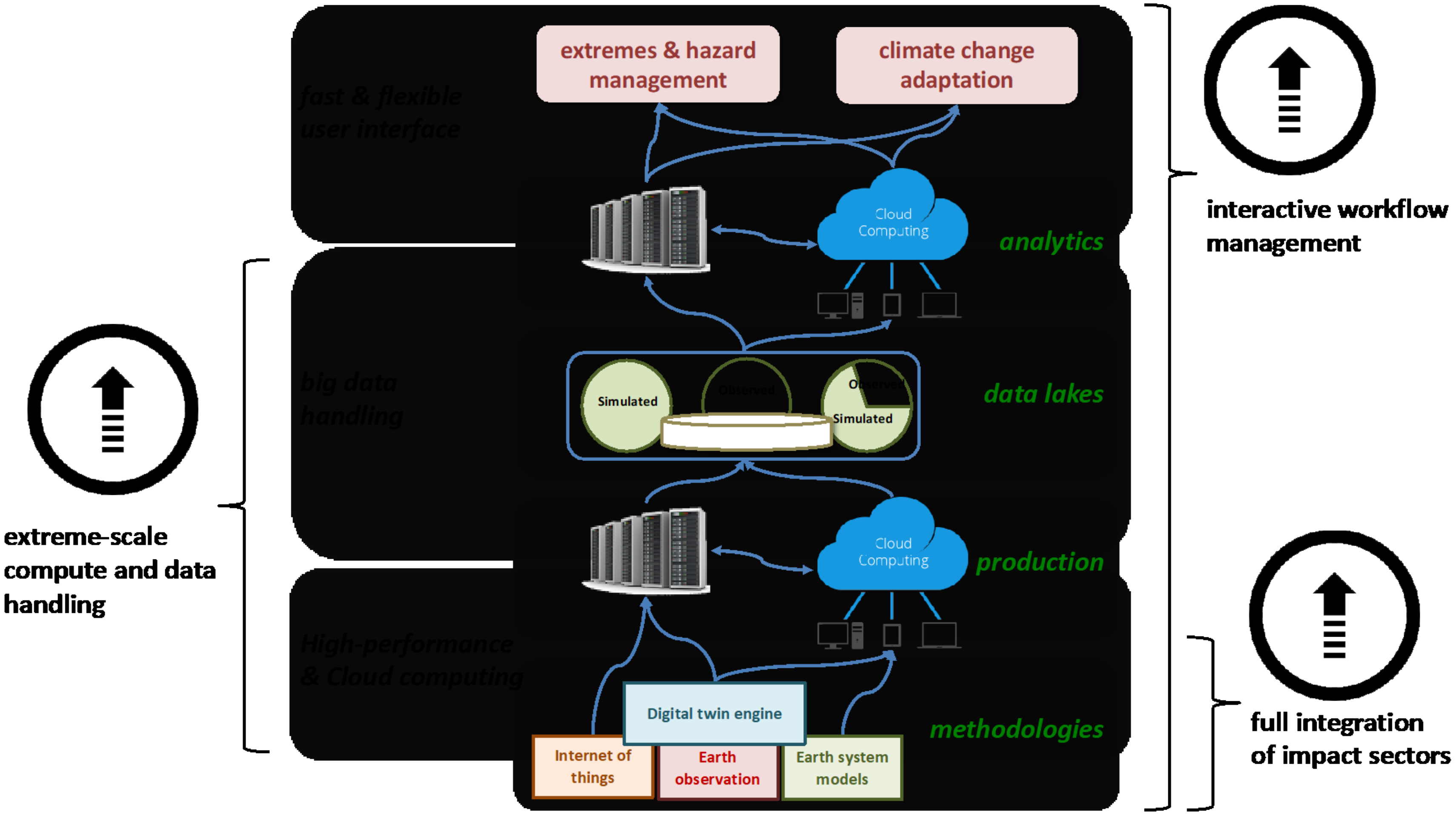
EarthCARE work was partially supported by the ***ESA funded*** project, Operational Assimilation of Space-borne Radar and Lidar observations for Numerical Weather Prediction (4000116891/16/NL/LvH) and the follow-on PEARL Cloud - Preparation for EarthCARE Assimilation of Radar and Lidar Cloud Observations - ESA ESTEC contract (4000128669/19/NL/CT).

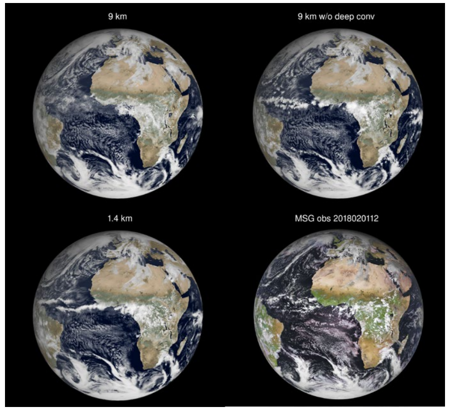
ECMWF

Mark Fielding and Marta Janiskova

WORKFLOWS

New horizons of machine learning, blurring the real and the physical(ly simulated) world





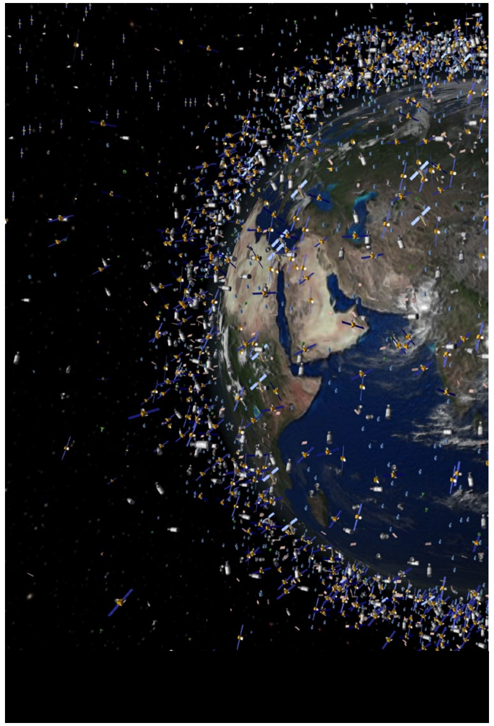
Digital Twins

Edge-computing
On-the-fly processing
Big ML/AI data pipelines
Hypercube data access
...

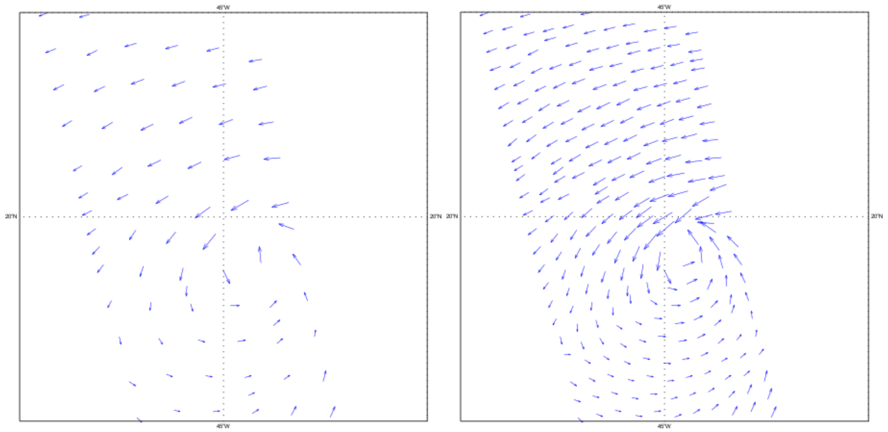
OPPORTUNITIES & CHALLENGES

- **New observing platforms** available for high-resolution Earth-System models
  - Constellations of small satellites
  - IoT observations
    - Large volume of lower quality data requiring advanced QC at scale
    - Manage inhomogeneous distribution with high temporal frequency
- **High resolution data assimilation**
  - Currently, we only assimilate around 5% of available observations with data heavily thinned to avoid spatially correlated errors between observations
    - Using a higher percentage of the incoming observations
    - Representing spatial error correlations will be more computationally demanding and likely involve non-local communication patterns implying reduced HPC scaling performance
    - Adapting to emerging HPC architectures is a challenge

MEOP
Courtesy: Phil Browne

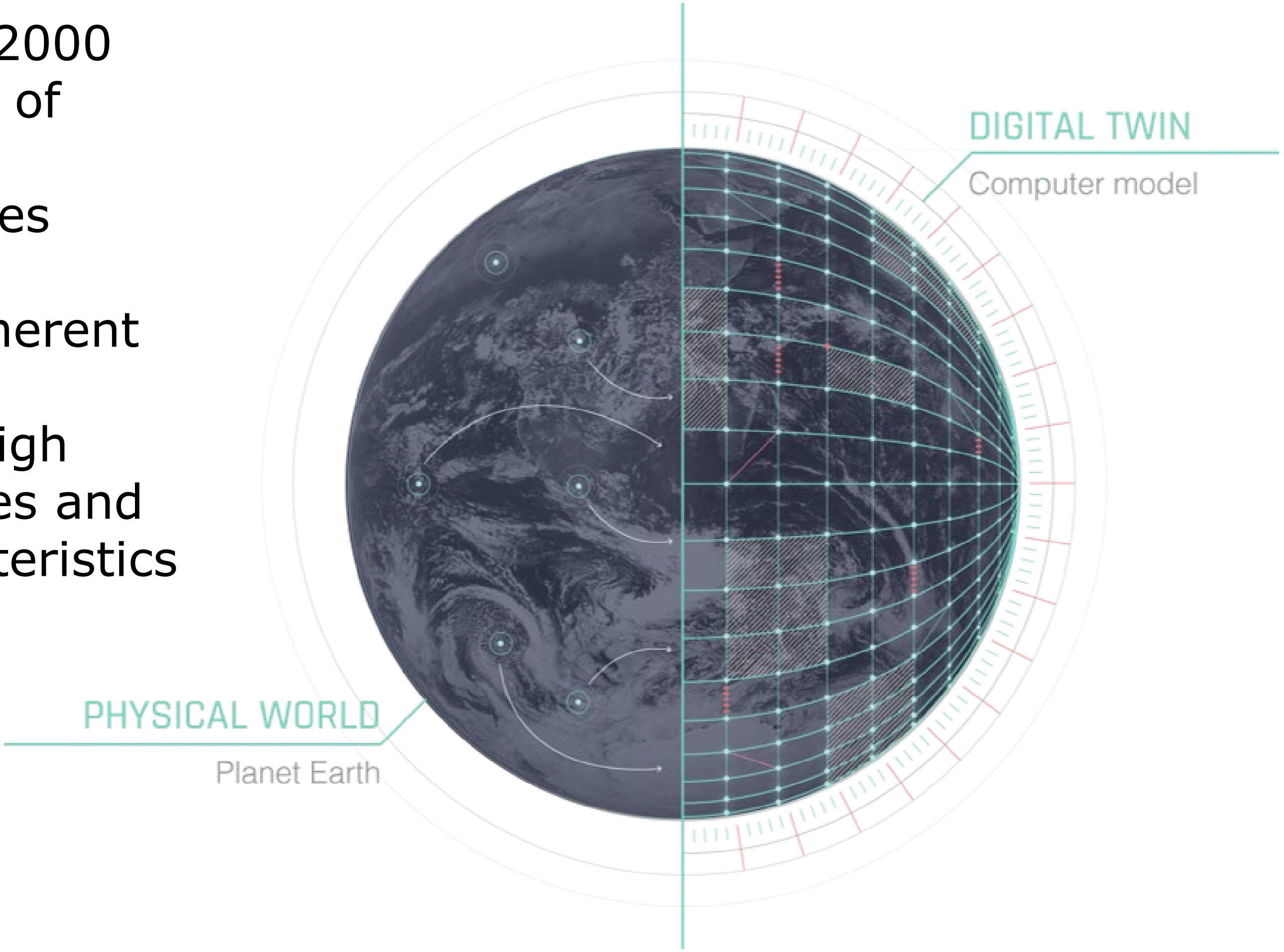


ESA



Assimilation of scatterometer data
Courtesy: Giovanna De Chiara

- Huge explosion in observation numbers since 2000 driven by satellite data, even bigger explosion of modelling (& ML training) data in DestinE
- Model adaptation to emerging HPC architectures challenging, observation processing has so far remained relatively affordable thanks to its inherent scalability
- Opportunities & challenges with advances in high resolution data assimilation, ML/AI technologies and new observing platforms with different characteristics (e.g. IoT)



CONTACT AND FURTHER INFORMATION

www.ecmwf.int/destine

Funded by the European Union
ECMWF
esa
EUMETSAT