AlpsB

A Geographically Distributed Infrastructure to Facilitate Large-Scale Training of Weather and Climate AI Models

Alex Upton, Jerome Tissieres, Maxime Martinasso **(CSCS)**

May 07, 2025

Table of Contents

1. The AI Revolution in Weather and Climate Modelling

2. Overview of AlpsB

3. Use Cases & Data

4. Challenges of deploying a geographically distributed system

# The AI Revolution in Weather and Climate Modelling

- ML is the '**third revolution**' in W&C Modelling

  - **Quiet revolution**: Better observations & more compute power

  - **Digital revolution**: Km-scale simulations on supercomputers

- Traditionally **physics-based** models used for weather forecasting

  - **High** computational requirements

  - **Sensitive** to approximations of physical laws

- ML models train algorithms to recognise patterns in data and make **predictions quickly**

  - Require **less computational power** and time

  - Consume **less energy** compared to traditional models

  - Already matching or surpassing **accuracy** of traditional NWP

The AI Revolution in Weather and Climate Modelling

- HRES is the **gold standard** today for global weather forecasting

- 2024 paper[1] compared the performance in forecasting near-surface temperature and wind speed extremes up to 10 days ahead in a semi-operational setting:

  - **IFS HRES** (ECMWF)

  - **GraphCast** (Google DeepMind)

  - **Pangu-Weather** (Huawei)

- Results suggest that data-driven models outperform IFS HRES in forecasting 2m temperature and 10m wind speed on average in **most regions**

[1] Olivetti & Messori https://doi.org/10.5194/gmd-17-7915-2024
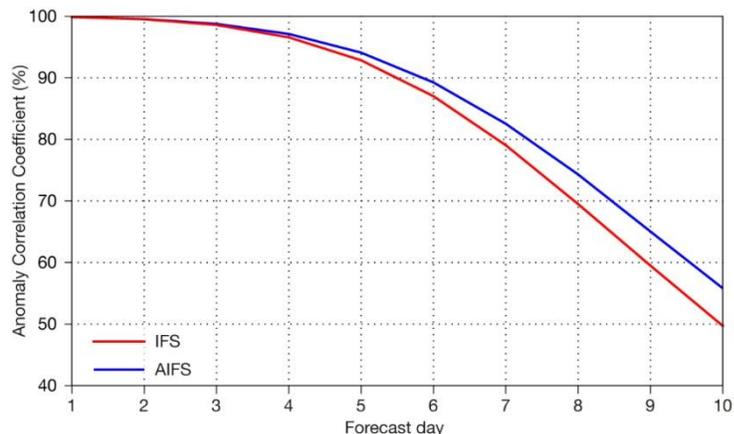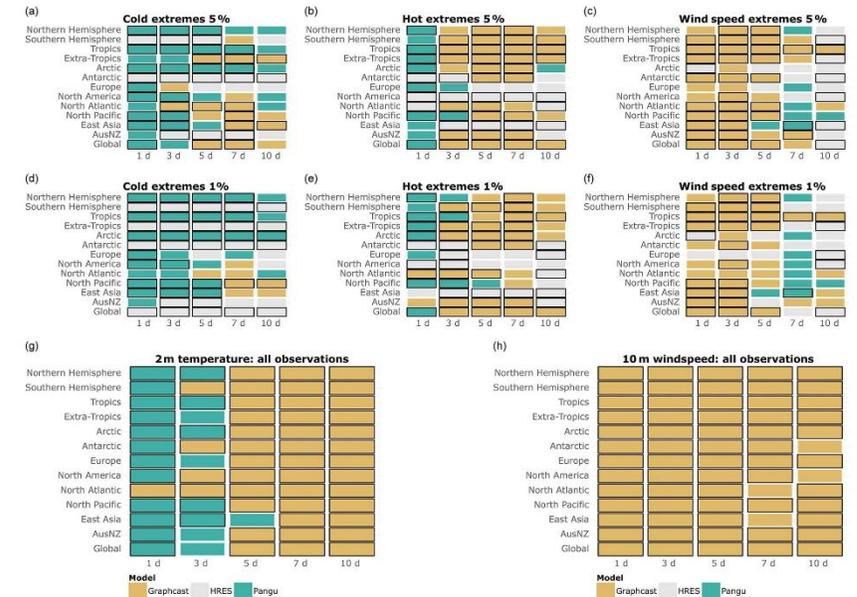
- ECMWF has implemented a data-driven forecast model, the **A**rtificial **I**ntelligence/Integrated **F**orecasting **S**ystem (AIFS)

- **AIFS outperforms IFS** for the northern hemisphere Anomaly Correlation Coefficient (ACC) for geopotential height at 500 for 2022 (higher values indicate better skill)

CSCS   SWISS|TWINS   ETH zürich

# The AI Revolution in Weather and Climate Modelling

- Training an AI model requires **two main elements:**

  - A **large dataset**
  - A **large-scale computing system** with thousands of GPUs

- The Main W&C Training Dataset: **ERA5**

  - Whilst different **MLWP** models employ different techniques, **all were trained using ERA5**
  - ERA5 is the fifth generation ECMWF **atmospheric reanalysis of the global climate**
  - Covers period from **1940 to present-day**
  - ERA5 is **5PB+** and growing…
  - **ERA6** is the next generation of reanalysis dataset currently in development, at least **25PB+…**

- Moving these huge datasets to the large-scale compute is becoming a **bottleneck**

- This leads us on to **AlpsB**, a **CSCS** system installed at **ECMWF** currently in development

850 hPa temperature and 500 hPa geopotential
ERA5 hourly - 00:00 on 1 January 2023

temperature (°C)
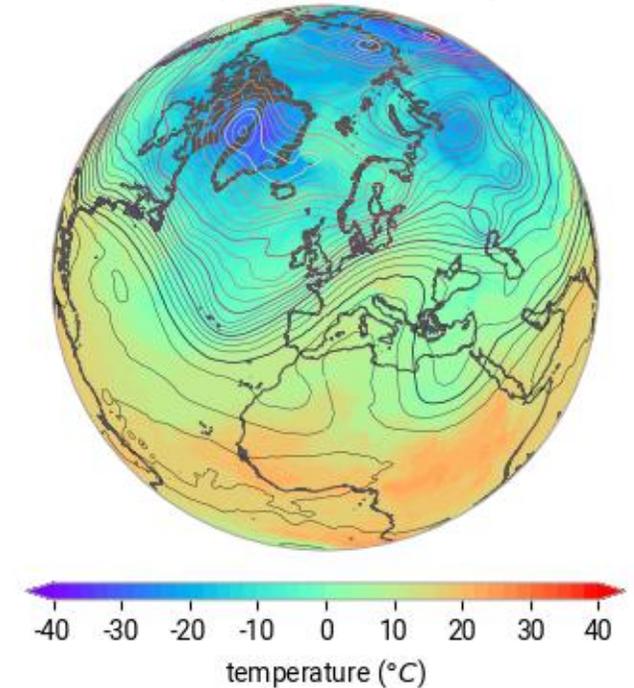
AlpsB

- AlpsB is a **storage-focused system** designed to operate as a Data Hypercube

- AlpsB is **physically installed** in the **ECMWF** datacentre in **Bologna, Italy**

  - ECMWF is the **European Centre for Medium-Range Weather Forecasts**
  - Produces data and data products for **member and cooperating states**

- **Directly connected** to Alps in Lugano and ECMWF Infrastructure

- Offers CSCS users **high-performance access** to ECMWF data

- Part of a **multi-site** distributed infrastructure:

  - **Lugano –** Alps
  - **Lausanne –** AlpsE
  - **Bologna –** AlpsB
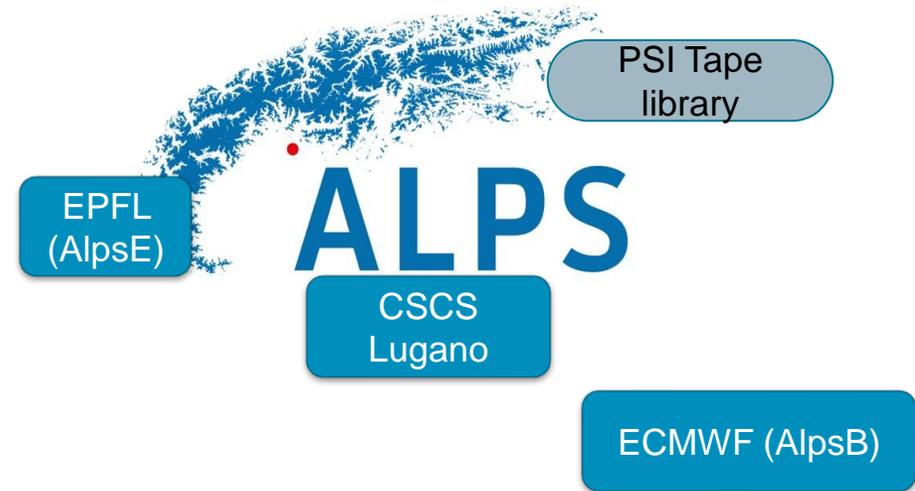
AlpsB infrastructure @ ECMWF Bologna

AlpsB infrastructure @ ECMWF Bologna

AlpsB infrastructure

- 3 **Compute nodes**
  - 2x AMD EPYC 7713 64c, 512GB RAM, 30TB local SSD
- 11 **Storage/Compute nodes**
  - 1x AMD EPYC 9534 64c, 384GB RAM, 61TB local SSD
- 1 **Clusterstor appliance (Lustre)**
  - HPE/Cray E1000, 2 PB usable space, ~20 GB/s
- 2 **Slingshot HSN switches**
  - HPE/Cray Slingshot-11, 200Gb, 1 HSN group
- 5 **mgmt. network switches**
  - HPE/Aruba switches, 100Gb, leafs/spine topology.

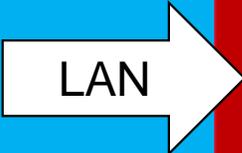## Network infrastructure

- One VXLAN/EVPN network fabric, and one out-of-band network
  - 6 **switches leafs/spines/border-leafs**
    - Cisco Nexus 9300-series, 400Gb, leafs/spine topology
  - 1 **low-speed switch for mgmt.**
    - Cisco Nexus 9300-series, 10/25/100Gb, copper
  - 1 **Out-of-band router/switch**
    - MikroTik CRS310-8G+2S+IN
  - 1 **Console server**
    - Vertiv Avocent ACS6016

AlpsB Connectivity
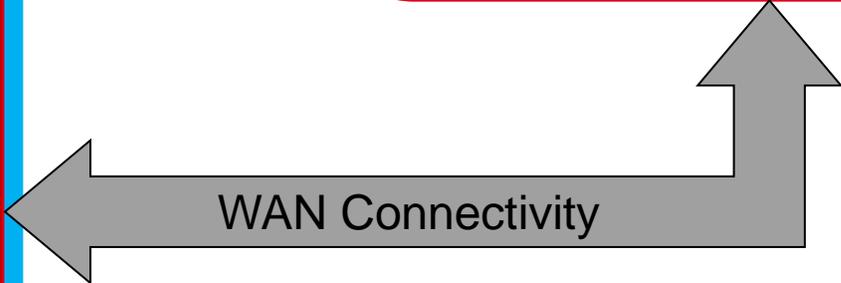

ECMWF Datacentre, Bologna

ECMWF Supercomputer

CSCS Datacentre, Lugano

Alps Research Infrastructure

LAN

LAN

AlpsB
Data Hypercube
Infrastructure

WAN Connectivity

Meteorological Archival and
Retrieval System (MARS)

AlpsB WAN Connection Bologna – Lugano

- WAN circuit between Alps and AlpsB shown below

  - Two **dedicated** and **redundant** 100Gbps layer-2 connections between CSCS network equipment in Bologna and Lugano

  - Circuits involves **SWITCH**, **GÉANT**, and **GARR**:

    - SWITCH provides a L2 service over a **redundant optical network**

    - GÉANT uses **two different fibre paths**, one Geneva - Milano, one Geneva - Marseilles - Milano

    - GARR uses **two different fibre paths** between two different locations in both Milano and Bologna

  - Observed speed **~ 92 Gbps** on each link, between test servers on both sides

  - On top of these two links, between the CSCS border-leafs, we apply **end-to-end encryption** of the data
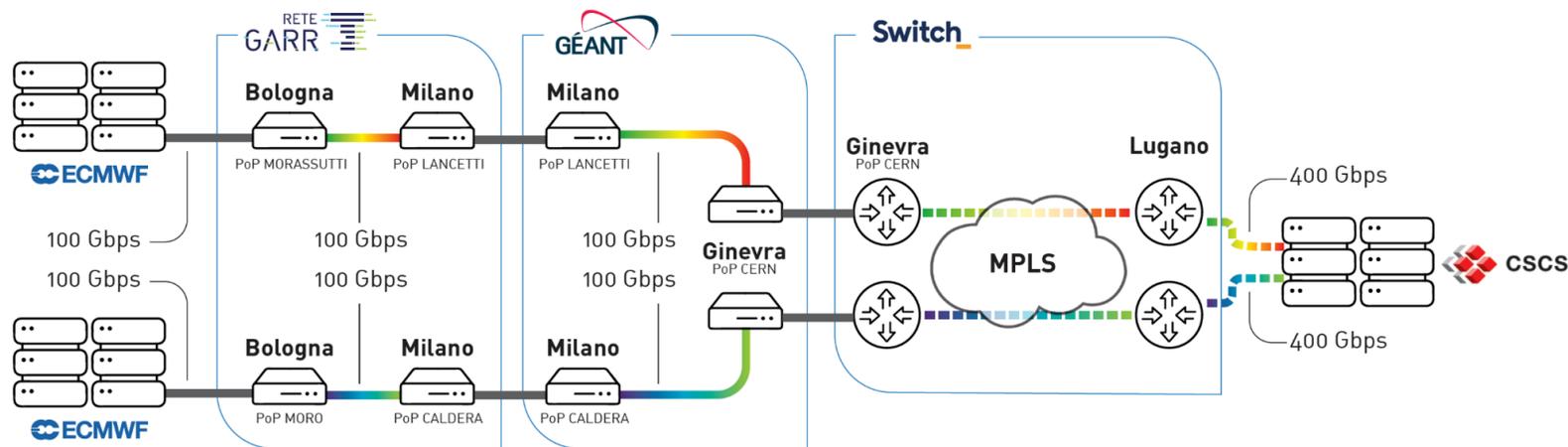
WAN Connectivity end-to-end optical circuits

Circuit 1

Circuit 2

CERN (CH)
CSCS (CH)
MILANO (IT)
ECMWF (IT)
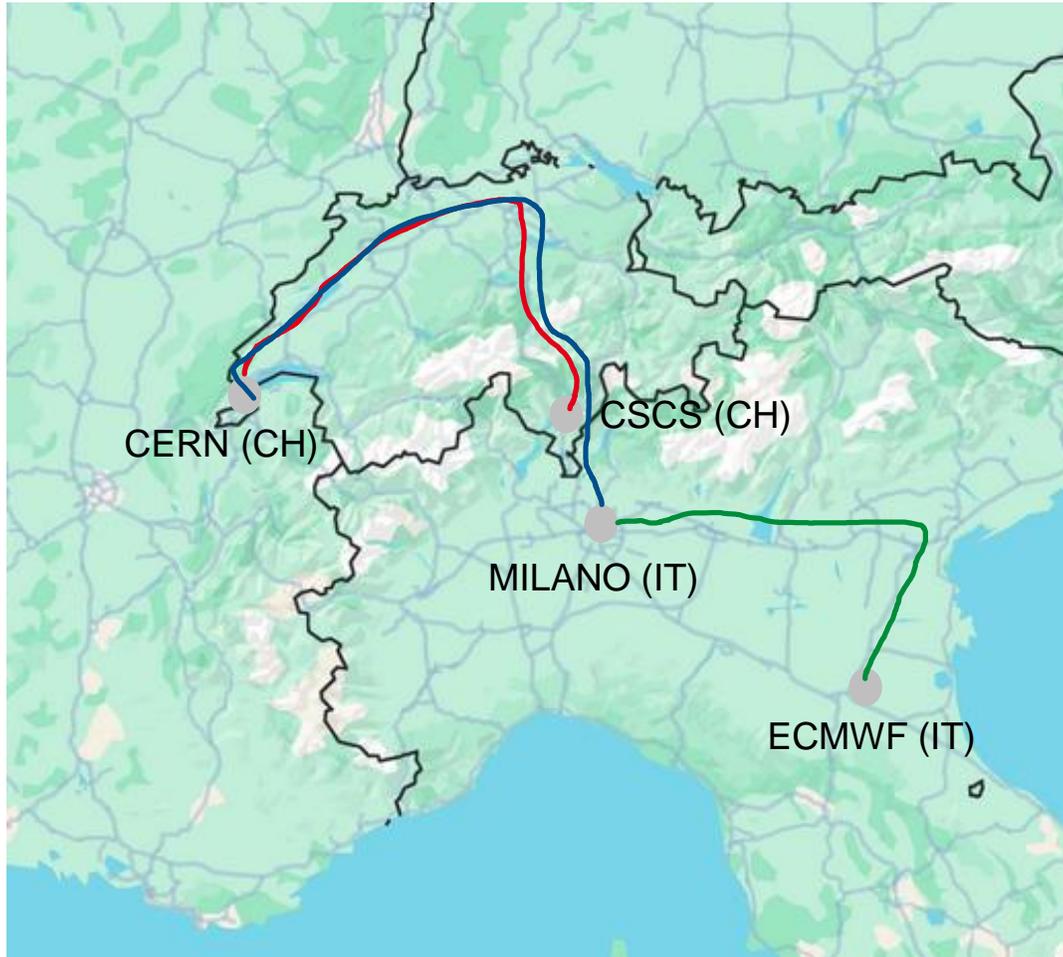
SWITCH EoMPLS Service
GÉANT Managed Wavelength Service
GARR T-service

AlpsB Use Cases and Data

- AlpsB offers **high performance access to ECMWF data** from the CSCS Alps system

  - This includes **operational forecasts**, **ERA5** and access to **MARS**

- Supports **existing use cases** as well as enabling new ones:

  - **MeteoSwiss** access to forecasts used for **boundary conditions** for national forecasting

  - Training of W&C **AI models**

  - **Climate** simulations

  - **Air pollution** source detection improvement

- ECMWF produces ~ **320TB of forecast data per day**, that will be **streamed directly** to AlpsB (4 forecasts a day, ~ **80TB** per forecast **every 6 hours**)

  - Dedicated **600 TB SSD storage for forecast data**, idea to store the last 4-5 forecasts (value of real-time operational forecasts decreases quickly)

- AlpsB will store this data locally in a **FDB** database:

  - FDB (Fields Data Base) is a **domain-specific object store** for storing, indexing and retrieving GRIB data....) developed by ECMWF
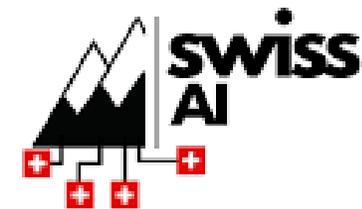
Polytope

- AlpsB will expose a **Polytope** web service that can be queried via API calls, to retrieve an extract of the data:

  - Polytope is a library **developed by ECMWF** for extracting complex data from datacubes

  - Provides an **API** for non-orthogonal access to data

  - Can be used to efficiently extract complex features such as **polygon regions** or **spatio-temporal paths**

  - Polytope will be deployed on AlpsB K8s vCluster, **co-managed** by CSCS & ECWMF

- Using the Polytope web service, it will also be possible to query ECMWF **Meteorological Archival and Retrieval System (MARS)**
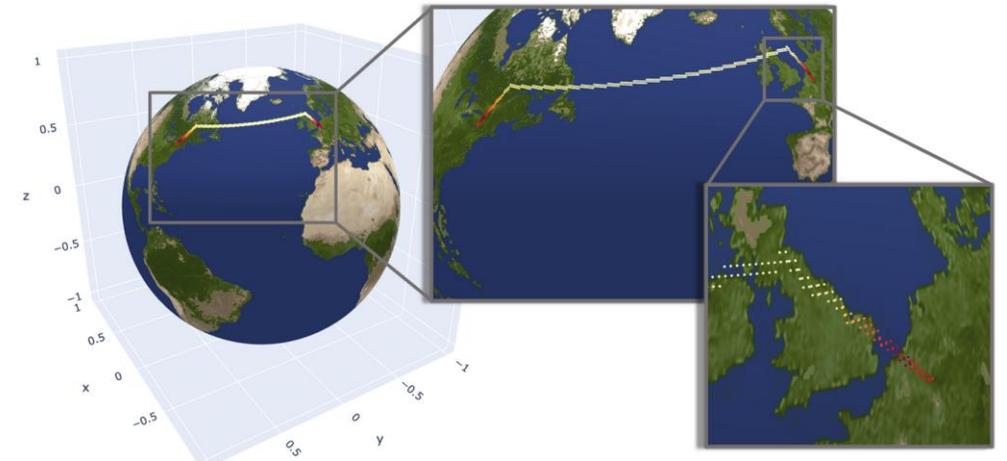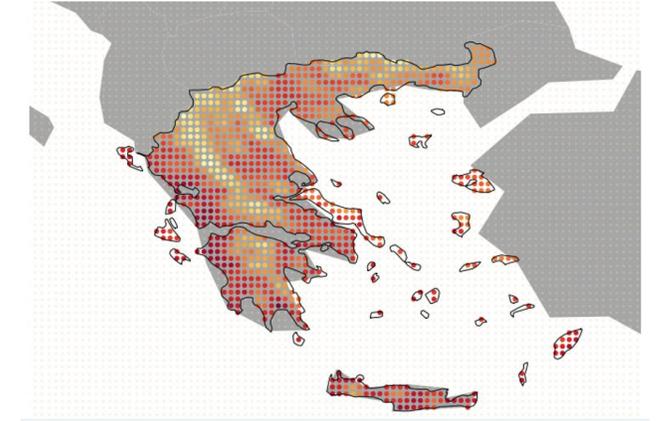
AlpsB Use Case – C2SM

- C2SM is the **Center for Climate Systems Modeling**

- Initial survey has identified **ERA5** as the main dataset of interest, with interest also in the **operational forecasts**

- In addition, **AIFS** also of interest, in particular use of this is projected to increase amongst the W&C community
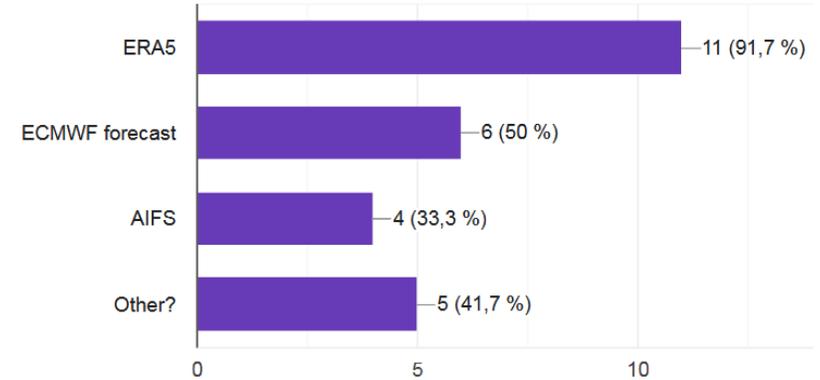
ECMWF's AI forecasts become operational

25 February 2025

AIFS OPERATIONAL

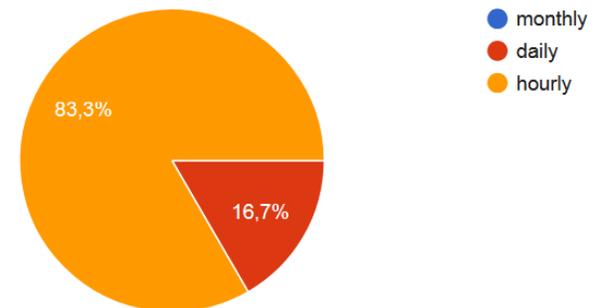What kind of data would you like to have immediate access to on AlpsB?

12 Antworten

ERA5 Data

An initial ERA5 dataset should have the temporal resolution of

12 Antworten

Comparison of existing ERA5 dataset access

Source	Dataset Summary	Advantages	Disadvantages
WeatherBench 2	• **Subset** of ERA5	• **Easily accessible** via GCP bucket in the cloud • Google cloud storage interface/cmd line access	• **Reduced selection** of years and variables so does not capture the whole dataset • GCP access required
Copernicus Climate Data Store	• **Complete** ERA5 dataset	• Complete ERA5 dataset • **Interface** that provides commands to download variables/years of interest	• **Queueing and prioritisation** of downloads often leading to slow downloads as a result, **significant bottleneck** considering potential size of dataset • ECMWF account required
German Climate Computing Centre (DKRZ)	• **Curated version** of ERA5	• **Generally quicker download than via CCDS**	• **DKRZ account** required • Delay in updating to the latest version of the ERA5 dataset
AlpsB	• **W&C community version** of ERA5	• Pre-fetched dataset (**no queueing**) • Curated based on feedback from W&C community • **Fastest download speed** due to dedicated connection between data source and compute • **Polytope Integration** for optimised data retrieval	• CSCS account & access required • CH W&C community focused

Challenges of a geodistributed infrastructure

- First CSCS system deployed **outside of Switzerland**

- **Technical challenges**:

  - Connecting infrastructure in **Italy** to Alps system in **Switzerland**

    - **Existing network connection** Swiss/Italian border **passes through a river**

    - Solution: Dedicated georedundant connection via **three network providers**

  - No easy **physical access**

    - Solution: Setup out-of-bound connectivity in case of **failure/misconfiguration of WAN circuit**, in order to allow backup connectivity as a failsafe

- **Operational challenges:**

  - **Support**

    - **Not physically on-site** with the HW

    - Solution: Operational workflow including **on-site engineering support** from ECMWF, as well as **support contract with AlpsB vendor** (HPE Italy)

  - **Indirect customer** for Milan – Bologna network circuit

    - Unable to have legal contract with Garr as CSCS/ETHZ is **not an Italian entity**

    - Solution: set up mechanism for **syncing and alerting** between CSCS and ECMWF **network teams**

Challenges of a geodistributed infrastructure

- In addition, a number of **administrative challenges**

- **Administrative challenges:**

  - Physical deployment of a system in a **foreign datacentre**
    - Solution: **housing agreement** with ECMWF covering power supply, cooling and support, as well as VAT and other aspects including insurance
  - **Procurement of hardware**
    - Swiss-based suppliers that we have existing agreements with were **unable to deliver** to Italy
    - Solution: procure HW directly from provider **based in country of machine installation**
    - Purchase from provider in CH and export/import to Italy complicated (this happened with network HW and required permission from **Italian Ministry of Foreign Affairs**)
  - **VAT rules** different between Italy and Switzerland
    - Different VAT rates in **Italy (22%)** and **Switzerland (8.1%)**
    - Solution: discussed with **VAT lawyers** in both Italy and Switzerland, also to clarify rules around **reverse-charge VAT**

CSCS SWISS|TWINS ETH zürich

Current Status of AlpsB and Future Timeline

- Status end of **April 2025**

  - **Legal and administrative** 'hurdles' cleared

  - Installation and configuration of AlpsB hardware in **Bologna**

  - Tests and validation of the **100Gb WAN links** between Lugano and Bologna

  - Interconnection of the **ECMWF datacentre network** to the CSCS AlpsB infrastructure network in Bologna

  - **vCluster** on AlpsB provisioned

  - **Security** assessment and hardening completed

- Next Steps

  - Definition and deployment of the **Polytope service**

    - **ECMWF installation of SW stack on AlpsB vCluster** – in progress

  - Definition and optimisation of **W&C datasets** on AlpsB

    - **Discussions with Swiss W&C Community** – in progress

  - Validation of AlpsB and network performances on **entire ensemble of forecasts and data slices**

  - AlpsB **in production** and accessible via **Alps Climate & Weather platform**
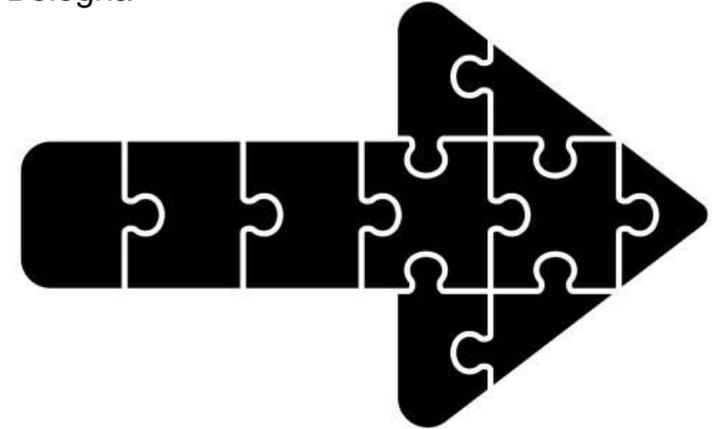
Acknowledgements

- SwissTwins Funding: **SERI**

- **ECMWF**

- **CSCS colleagues**
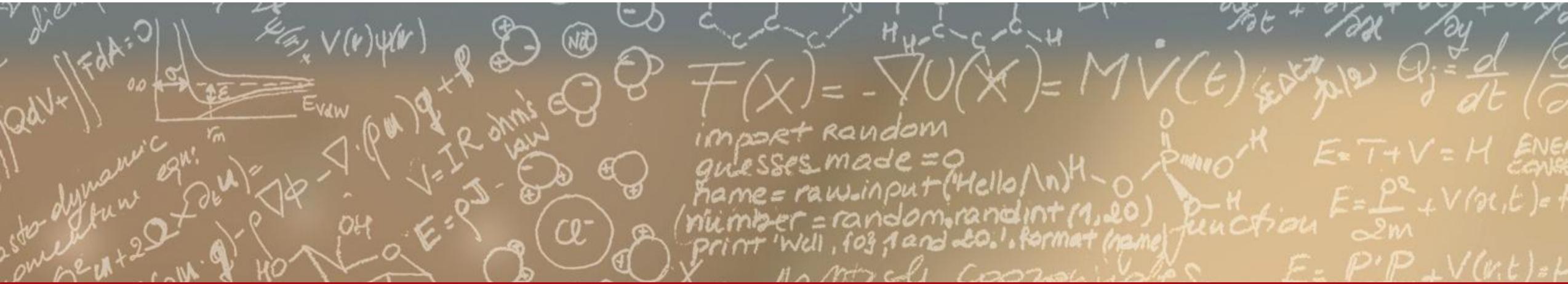
Thanks for your attention, questions…