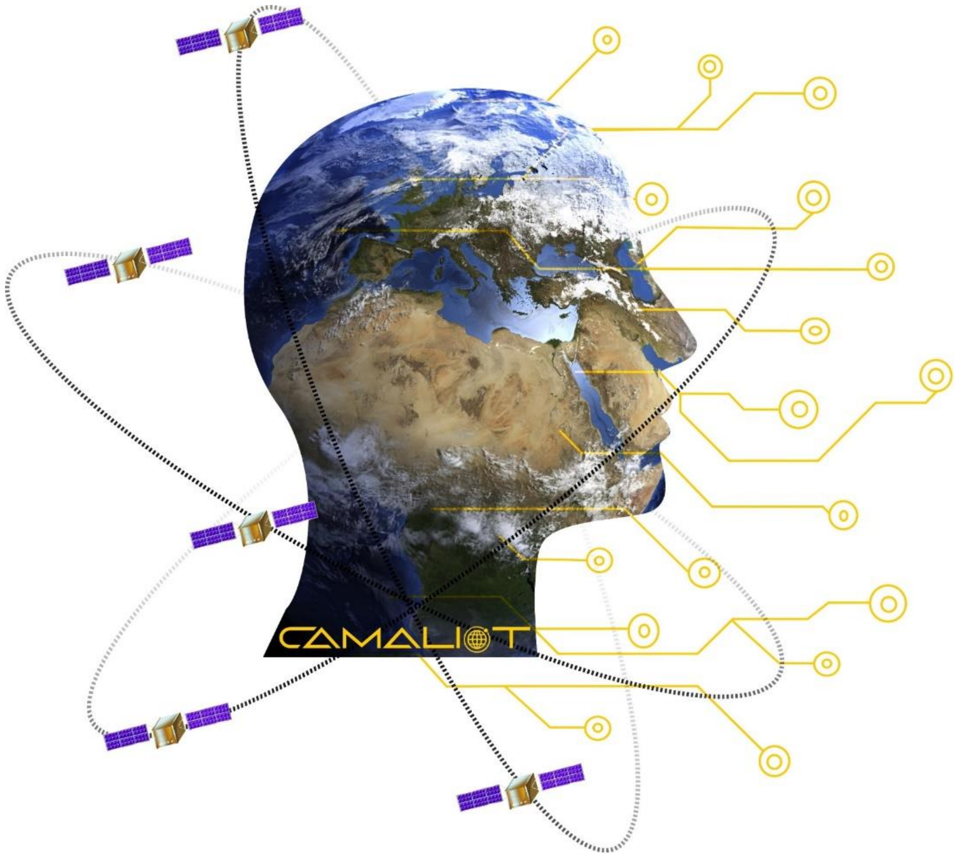
The CAMALIOT project

Linda See, B. Soja, G. Kłopotek, M. Awadaljeed, L. Crocetti, Y. Pan, M. Rothacher, R. Weinacker, T. Sturn, I. McCallum, Navarro, V.

May 2022

CAMALIOT

Appli**ca**tion of **ma**chine **l**earning technology for GNSS **IoT** data fusion

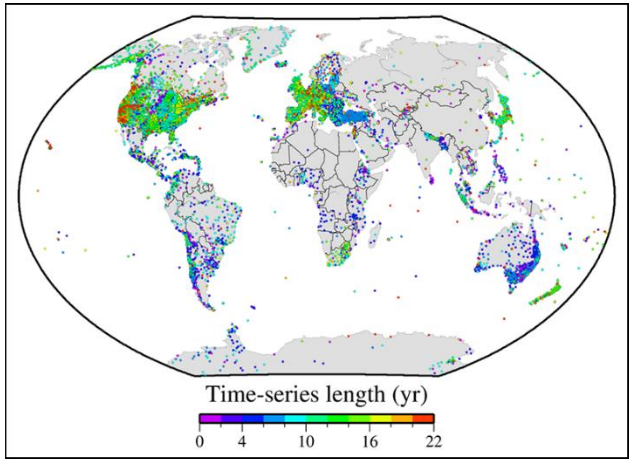


Motivation

- GNSS infrastructure has been growing significantly in recent years
  - Four global constellations, including the European Galileo system
  - Tens of thousands of geodetic stations
  - Millions of GNSS-capable devices (smartphones, drones,…)
  - Unprecedented spatio-temporal resolution

- GNSS data far from fully exploited for science and society:
  - Earth and space weather forecasts
  - Geo-hazard detection

- GNSS Science Support Centre (GSSC) by ESA
  - Integrate GNSS data and processing capabilities from various domains
  - Offer science exploitation services



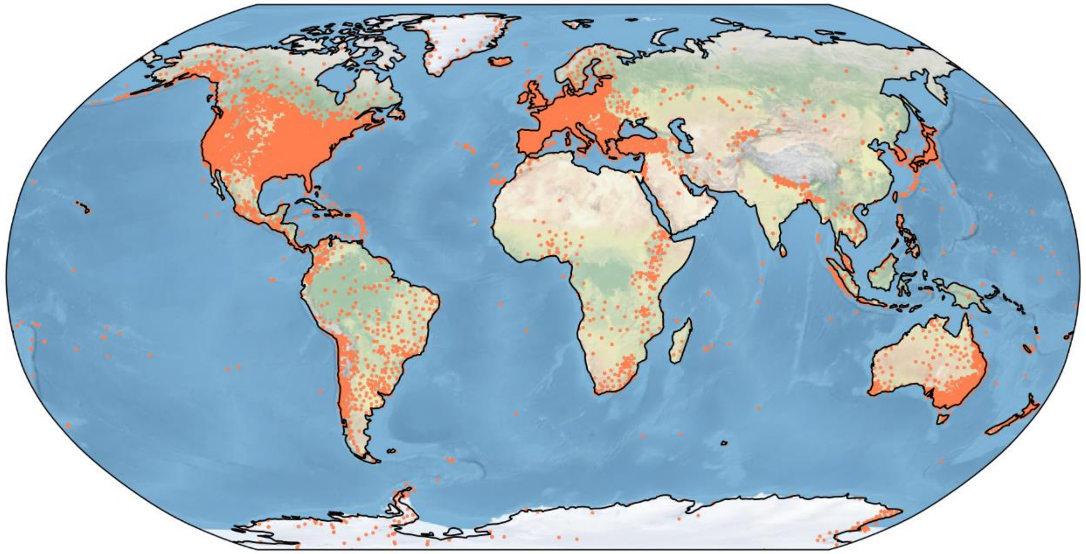
Blewitt et al., 2018

Motivation – GNSS IoT data

- Raw GNSS data available from Android smartphones
  - Android version 7.0+
  - Recent smartphone generations with dual-frequency receivers

- GNSS capabilities in various IoT applications
  - Smart cities, navigation, agriculture, robotics,…
  - Driven by the low-cost GNSS receivers

Geodetic GNSS stations



Motivation – big data GNSS processing / machine learning

- Increasing popularity of data science / machine learning approaches in several disciplines
  - Availability of big data
  - Increase in computational power
  - Development of machine learning algorithms and infrastructure

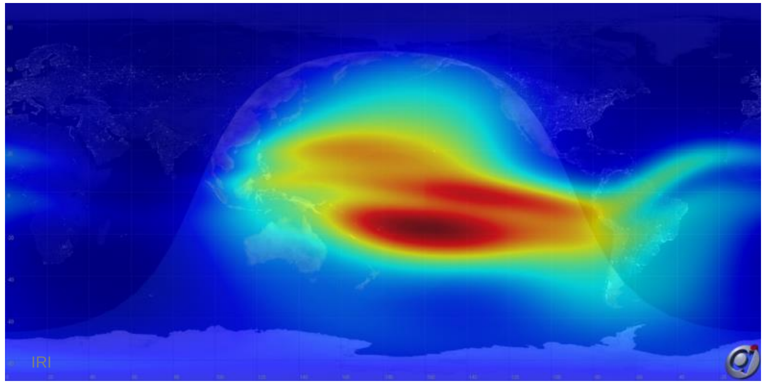
- Machine learning suitable to extract information from “challenging” datasets
  - Inhomogeneous, low-quality data (like GNSS IoT data)
  - Ability to assimilate data from very different domains

Motivation – ionosphere

- GNSS observations sensitive to ionospheric effects
  - VTEC determination by geometry-free linear combination (multi-frequency, code vs. carrier phase)
  - Decent also using single-frequency GNSS observations from low-cost devices
  - Derivation of global ionospheric maps (GIMs)

- Ionospheric models provide corrections for single-frequency devices (e.g., low-cost / IoT)

- Effect of solar activity and geomagnetic phenomena on ionospheric parameters



Motivation – troposphere

- GNSS allows estimation of tropospheric parameters
  - Zenith wet delay, gradients
  - Derived atmospheric quantities, e.g. precipitable water vapor (PWV)

- Tropospheric products to improve positioning performance

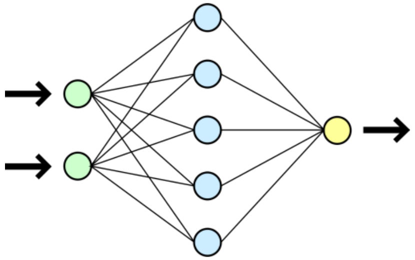
- Tropospheric parameters assimilated in numerical weather models to improve weather forecasts

Goals

- GNSS IoT data
  - Investigate alternative sources of GNSS IoT data
  - Crowdsourcing GNSS data

openclipart.org

- GNSS big data processing
  - Develop framework for automated and robust GNSS processing
  - Machine learning for anomaly detection and data selection

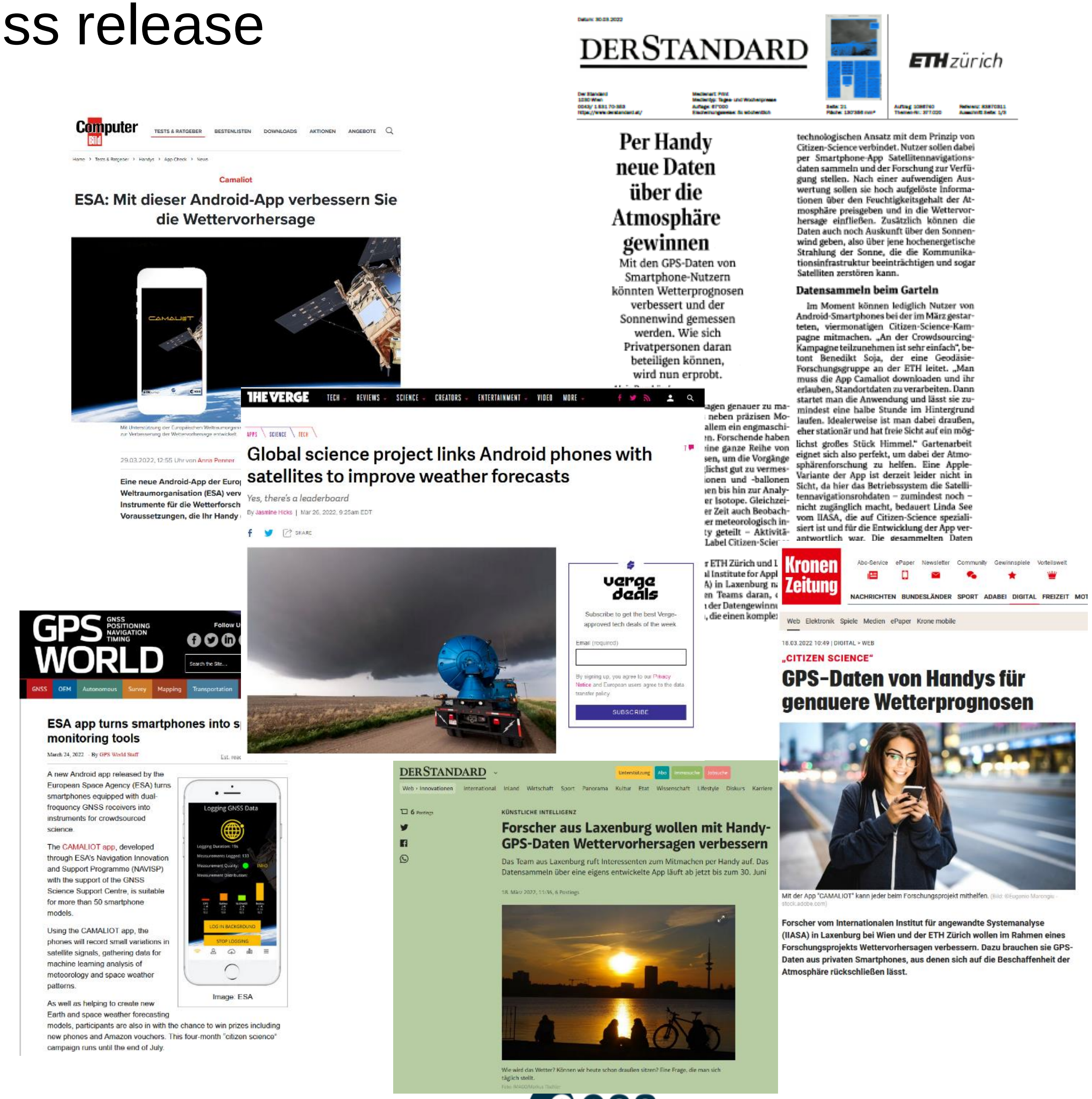


- Science use cases
  - Troposphere – Earth weather
  - Ionosphere – space weather

ETH zürich IIASA CAMALIOT esa

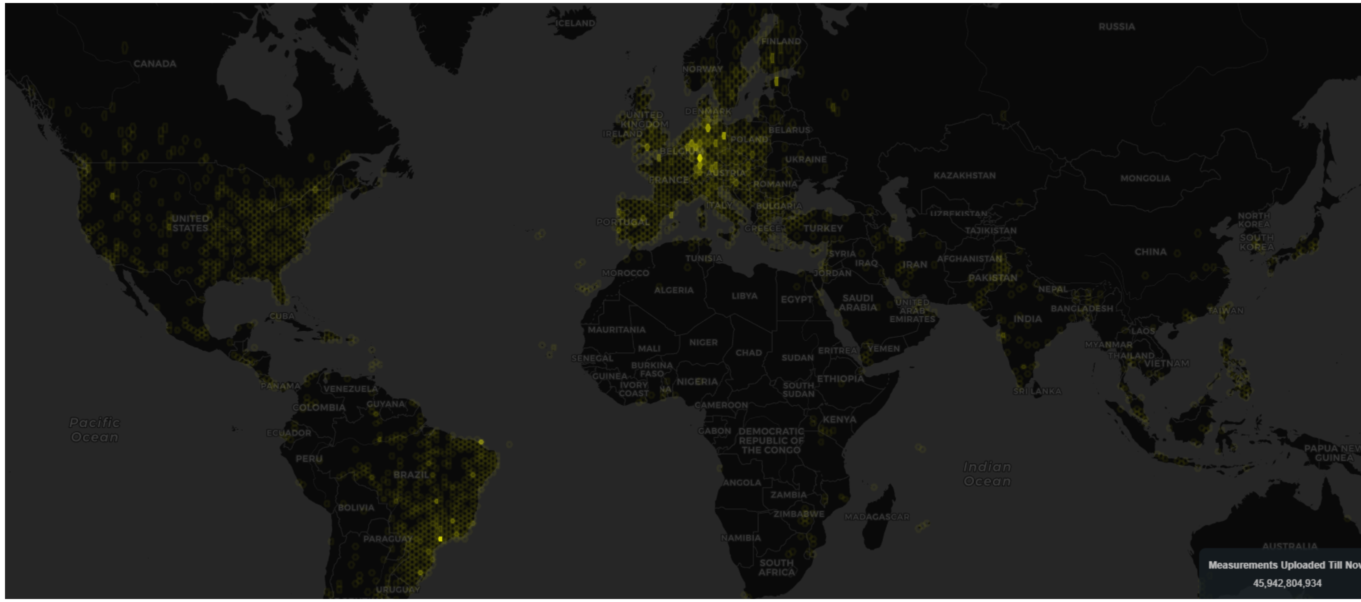
Responses to the CAMALIOT press release

- Online media
- Print media
- Interviews
  - Der Standard journalist
  - NY Times book editor
  - Popular Science journalist
- Contact established
  - ZAMG, MeteoSwiss, UBIMET
  - Jua, tomorrow.io, Miratlas



Crowdsourcing campaign

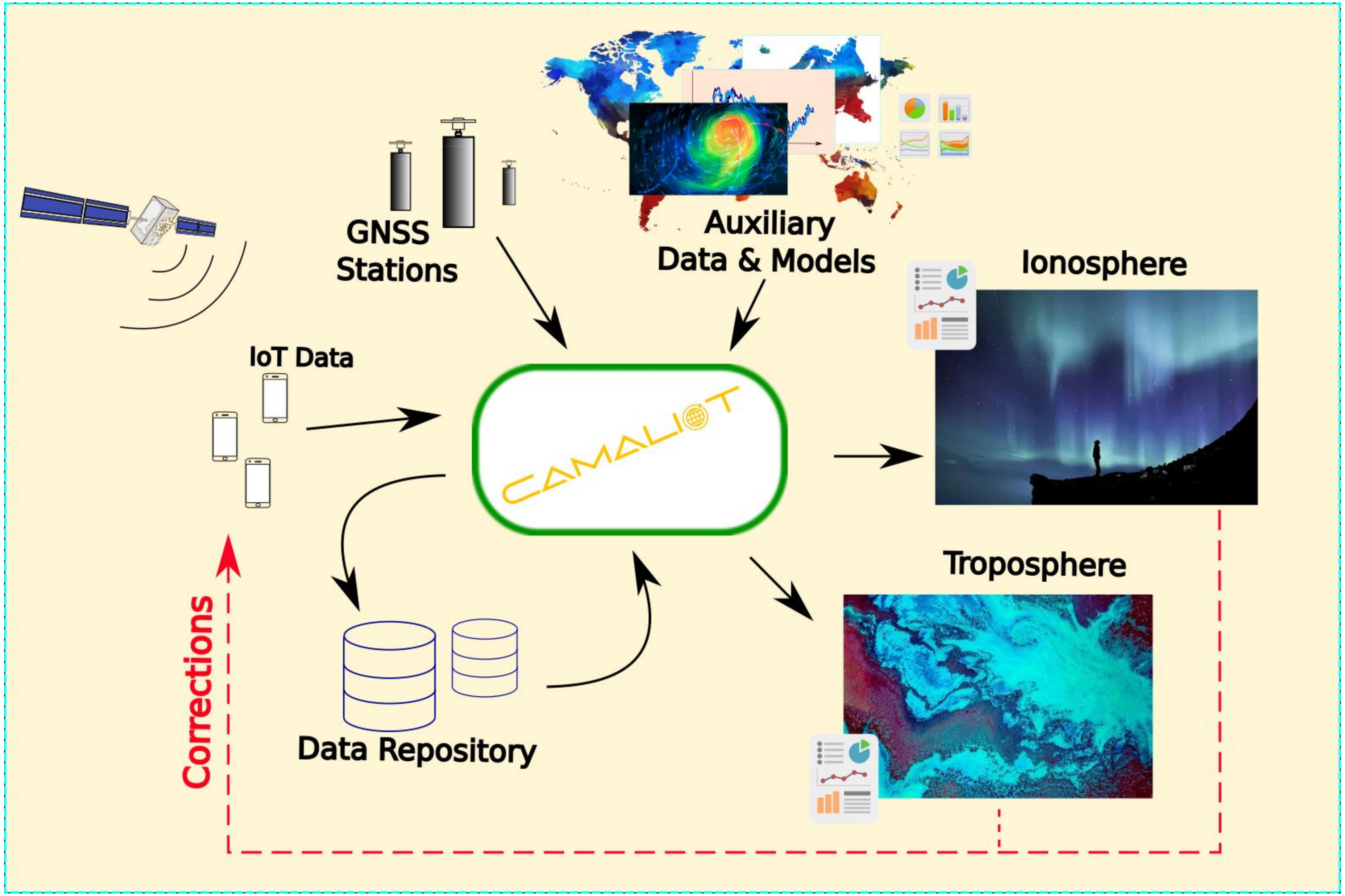
- Started March 17
- Android app with 35k+ installations
- Almost 11k registered users
- > 45 billion GPS observations collected so far



https://www.camaliot.org

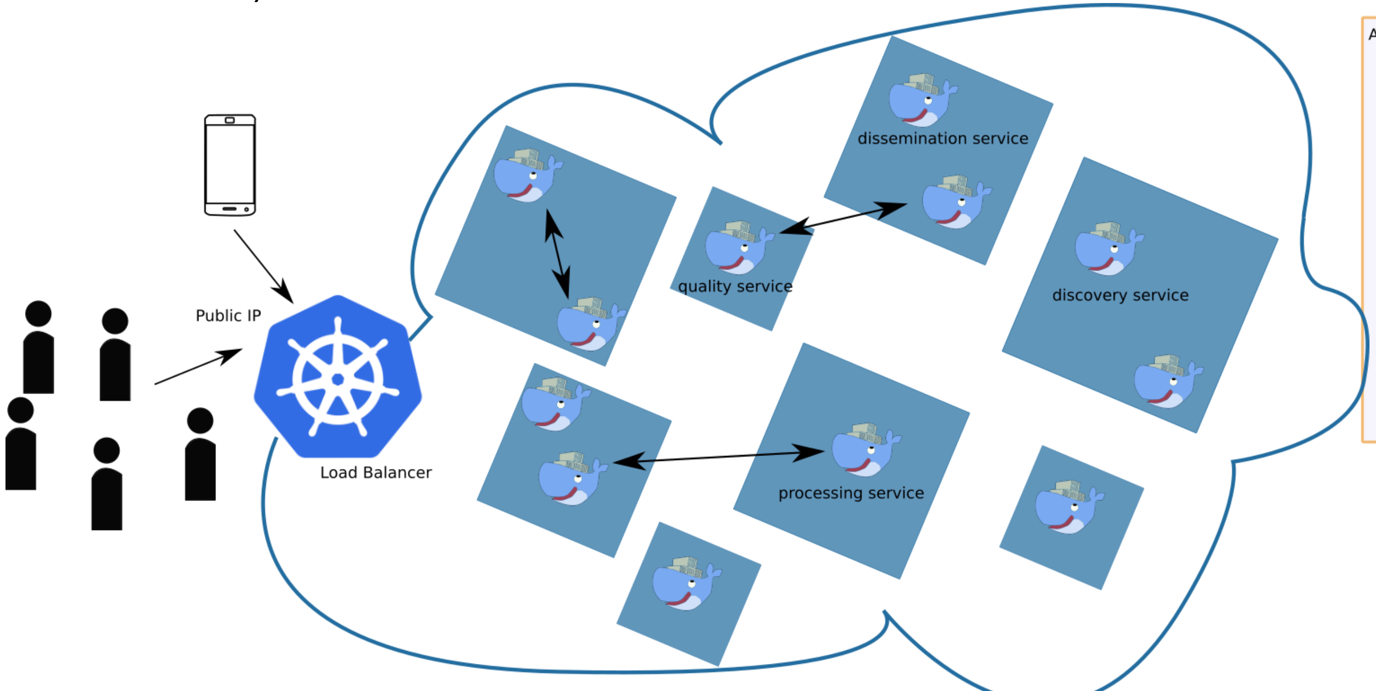
https://www.youtube.com/watch?v=rYJDW35R9jE

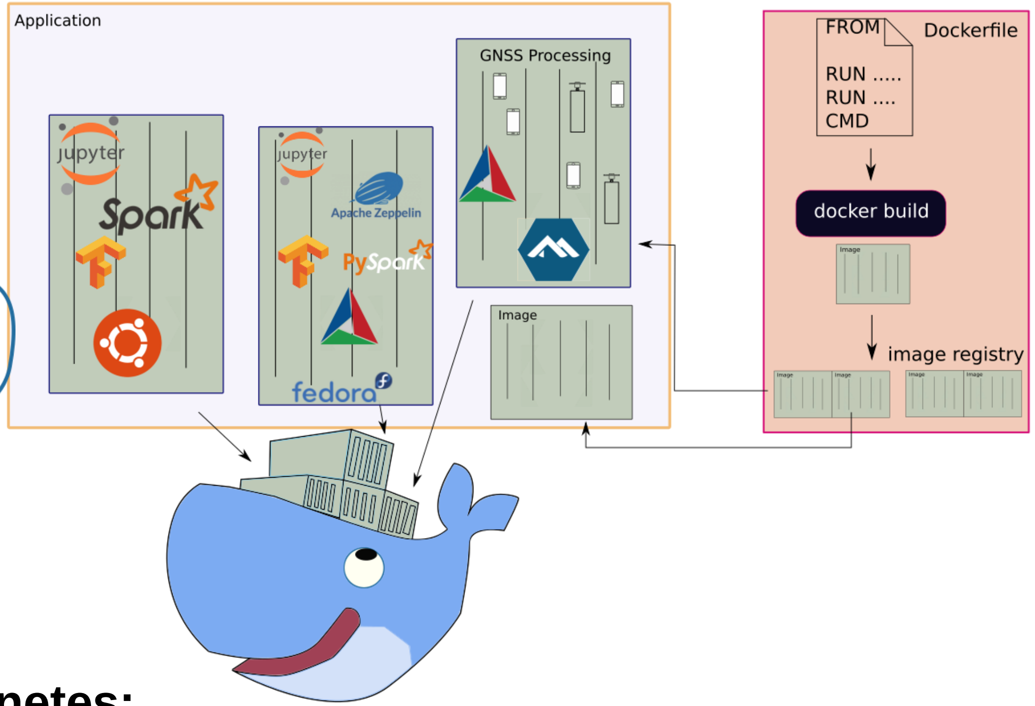
CAMALIOT high-level overview



ESA GNSS Science Support Centre (GSSC)

GSSC, microservices and Kubernetes cluster





Containerized applications (docker containers) on Kubernetes:

- packages up the code and the dependencies

- always run the same, regardless of the infrastructure

- micro-services: application broken into smaller, independent pieces, can be deployed and managed dynamically

ETH zürich CAMALIOT IIASA

CAMALIOT SW Architecture (1)

EXOSCALE

DigitalOcean

Exoscale and DigitalOcean:

- CAMALIOT SW running on resources provided by Exoscale
- Continuous Integration (CI) and Continuous Deployment with DigitalOcean

EXOSCALE

A solid European cloud hosting alternative.

Compute
- High performance SSD cloud servers
- Self-sustained zones for resilient deployments
- Snapshots and Custom Templates
- Instance Pools to manage groups of machines
- Anti affinity groups
- IAM and organizations management

Kubernetes
- Scalable Kubernetes Service
- Start Kubernetes clusters in 100 seconds
- Scale up and down Worker Nodes
- Full control plane lifecycle management
- CLI, API, portal, Terraform support
- Deep NLB integration

DBaaS
- Fully managed database as a service
- Wide range of open source databases
- Managed PostgreSQL, MySQL, Apache Kafka and Redis™ databases
- Start within minutes
- End-to-end GDPR compliance

Object Storage
- Highly available multi-redundancy object storage
- Low latency, high bandwidth public or private secure HTTP(s) access
- S3 compatible API for simple tooling integration

GPU Servers
- From 1 up to 8 NVIDIA GPU cards
- Direct passthrough access for maximum performance
- All the advantages of a regular Compute server

Networking
- Dual 25 Gbps secure private networking
- Security groups to manage firewall rules
- Network Load Balancer
- Elastic IP addresses
- IPv6 for instances

https://www.exoscale.com

Data Security, GDPR Compliance & Data Privacy Laws.

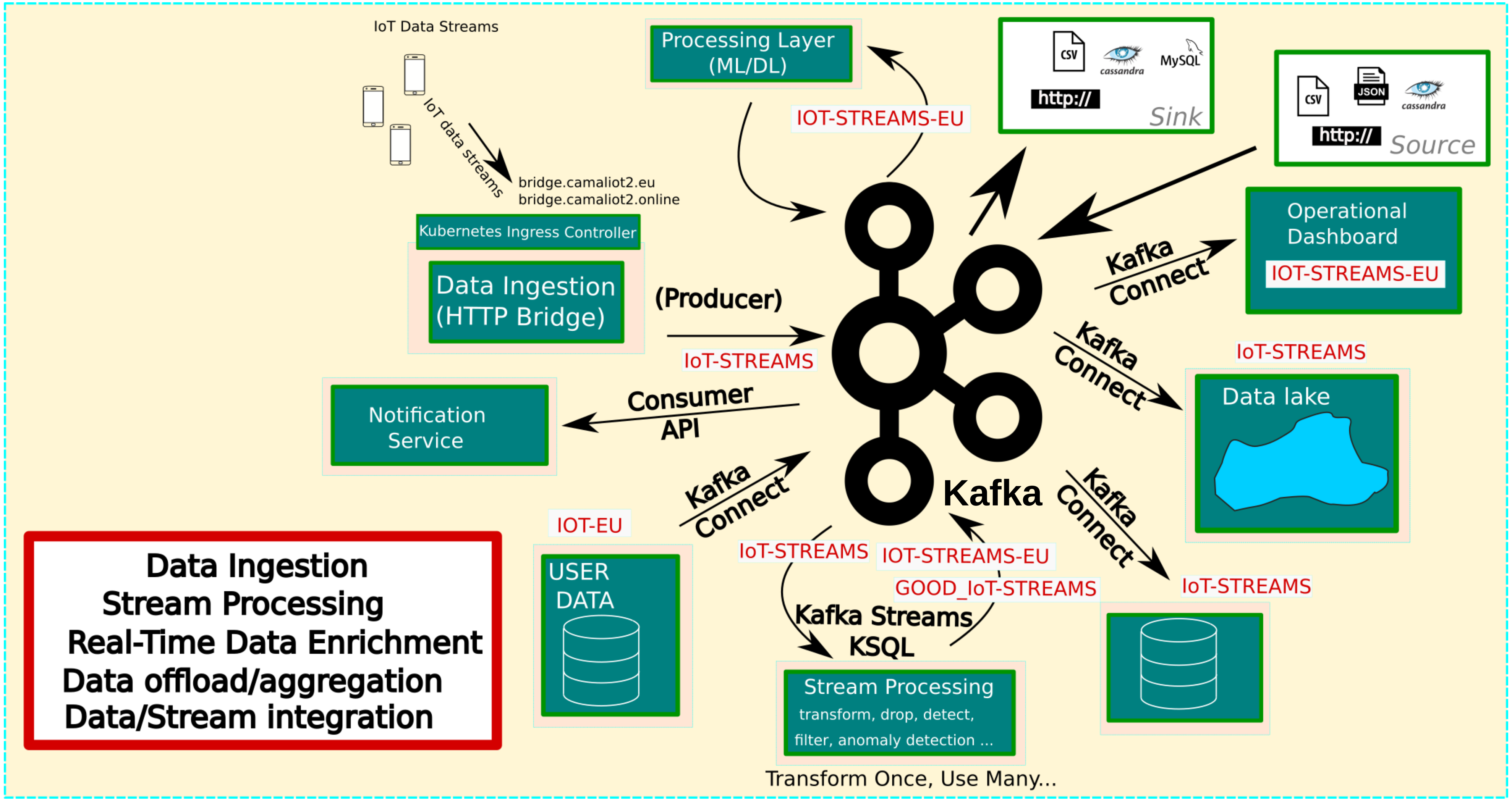
We take security seriously. Exoscale is the ideal partner for those looking to respect the strict Swiss data privacy laws. We also help engineers ensure that the workloads started in Austria, stay in Austria. Same goes for workloads started in Germany: we guarantee that they will remain in Germany. This means that your projects are GDPR-compliant when you work with Exoscale.

https://www.exoscale.com

CAMALIOT SW Architecture (2)

Crowdsourced GNSS observations: collection and processing as realized with the Kafka ecosystem

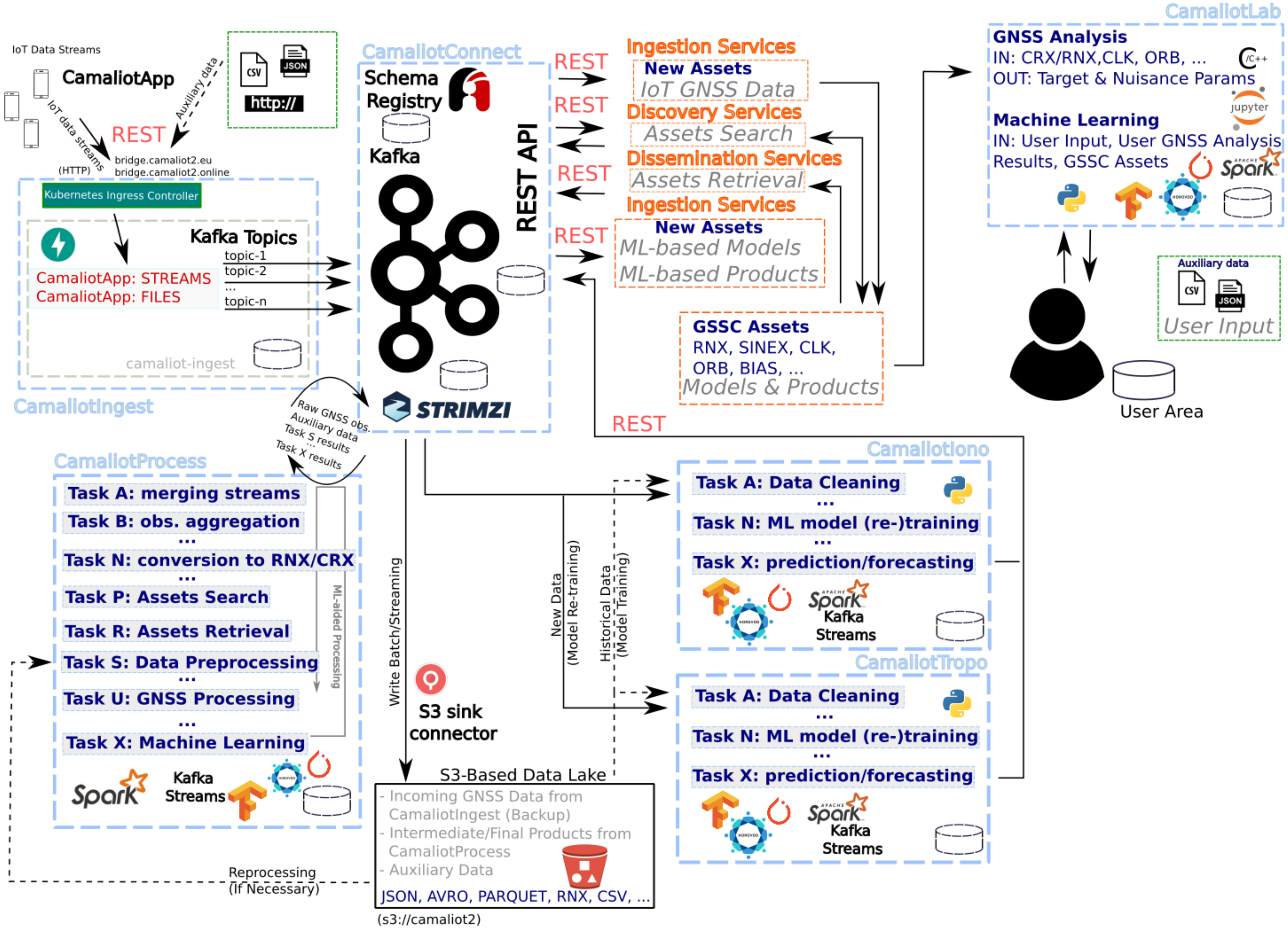
Event-driven Processing with History



CAMALIOT SW Architecture (3)

Crowdsourced GNSS observations: collection and processing as realized with the Kafka ecosystem

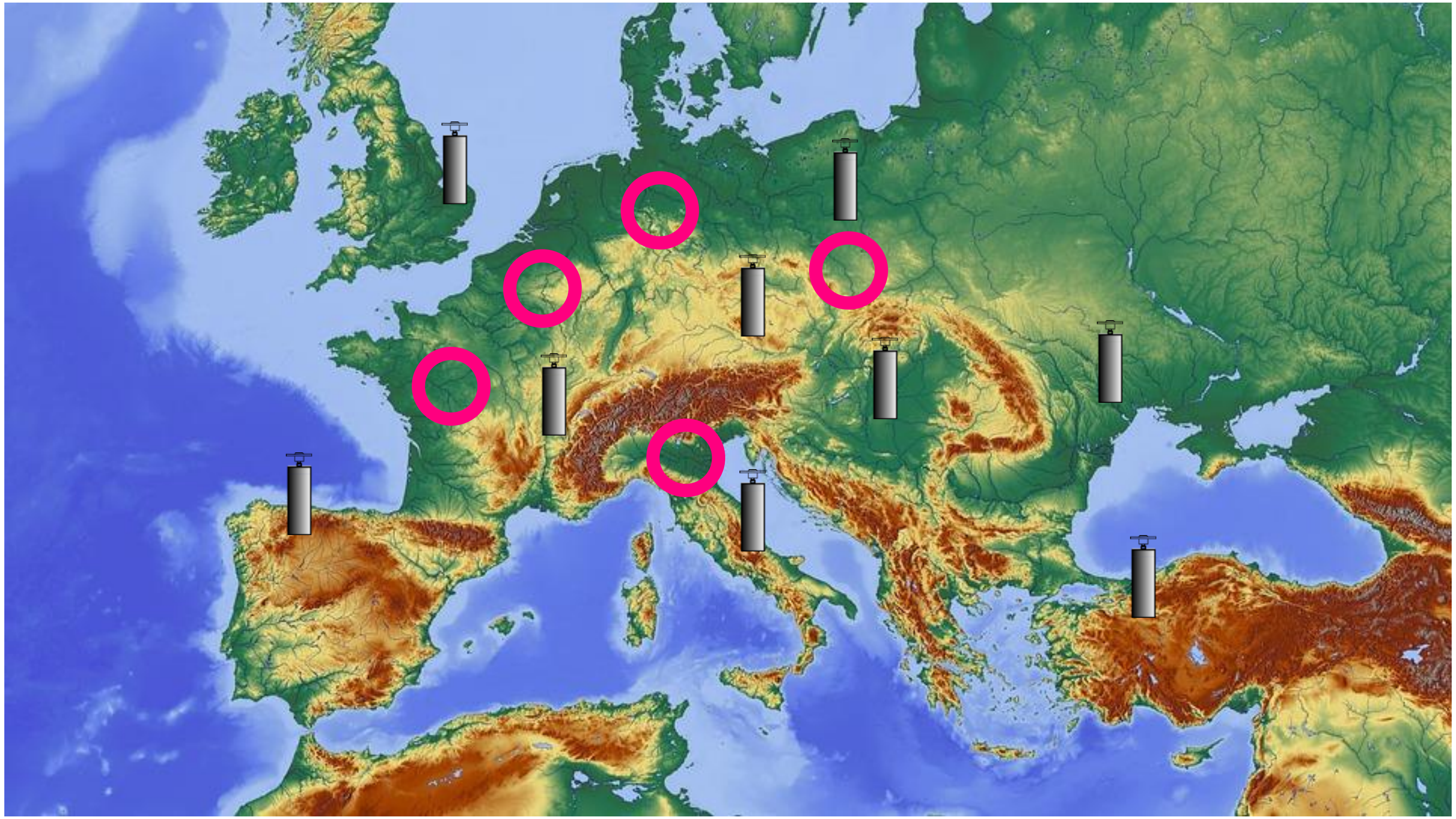
Event-driven Processing with History

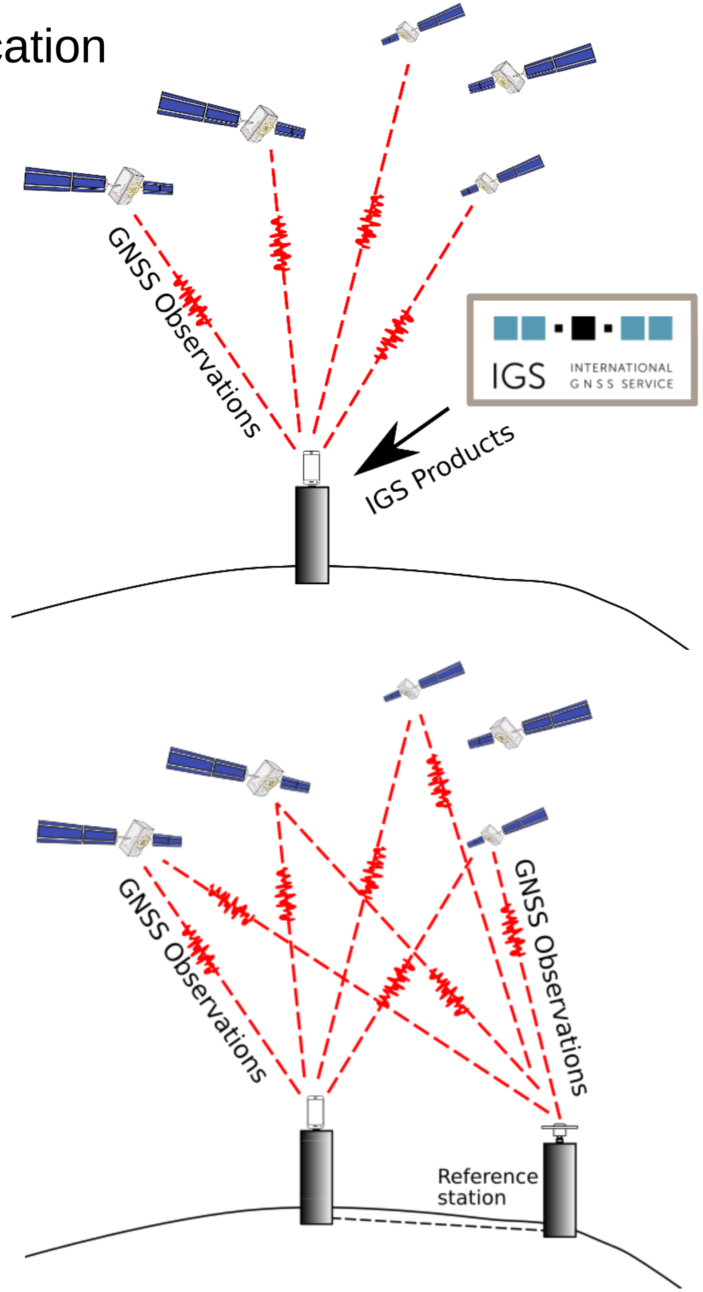


ETH zürich CAMALIOT IIASA

GNSS IoT Data Collection & Processing (1)

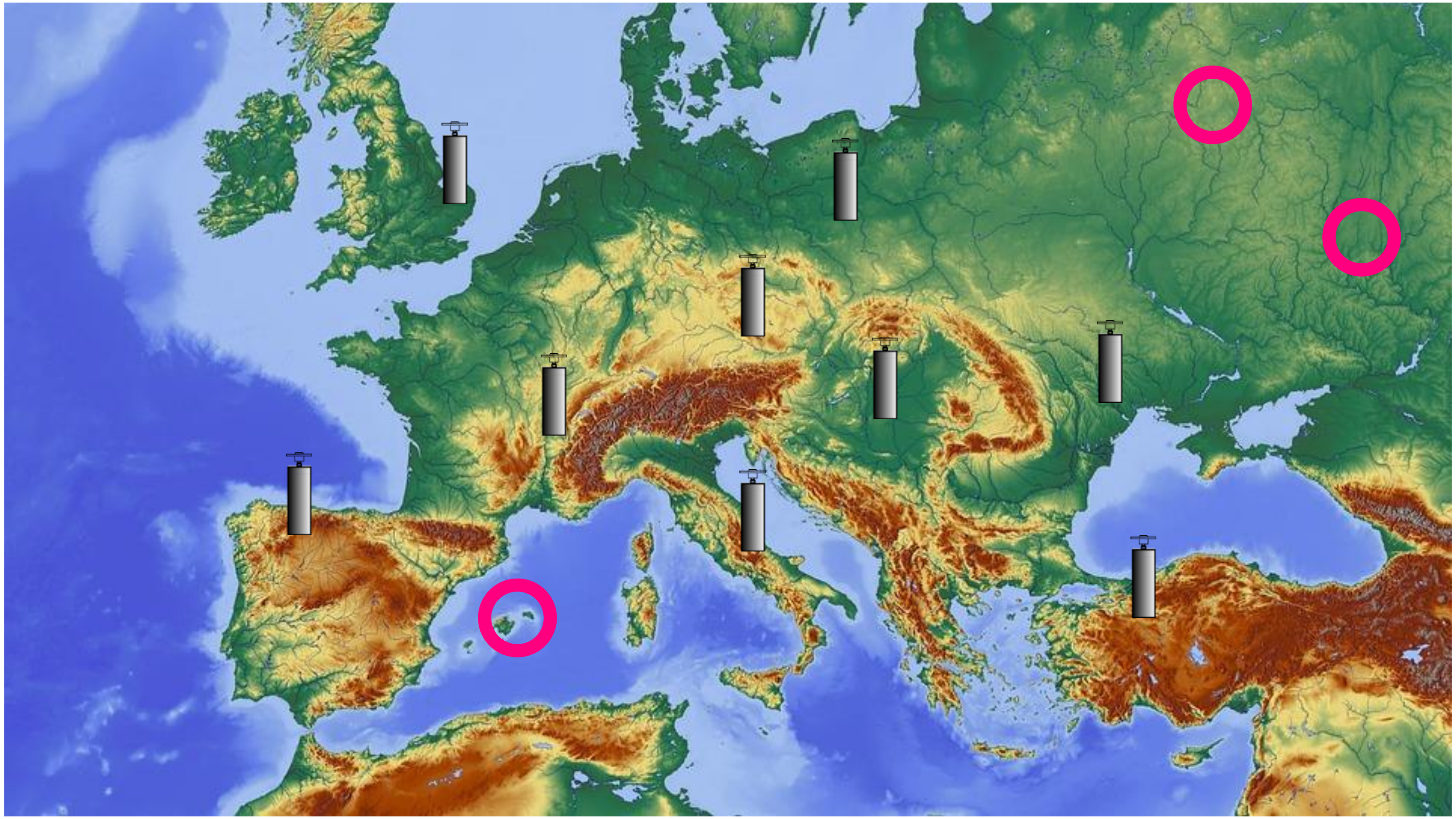
Case I – Crowdsourced GNSS Observations for Network Densification

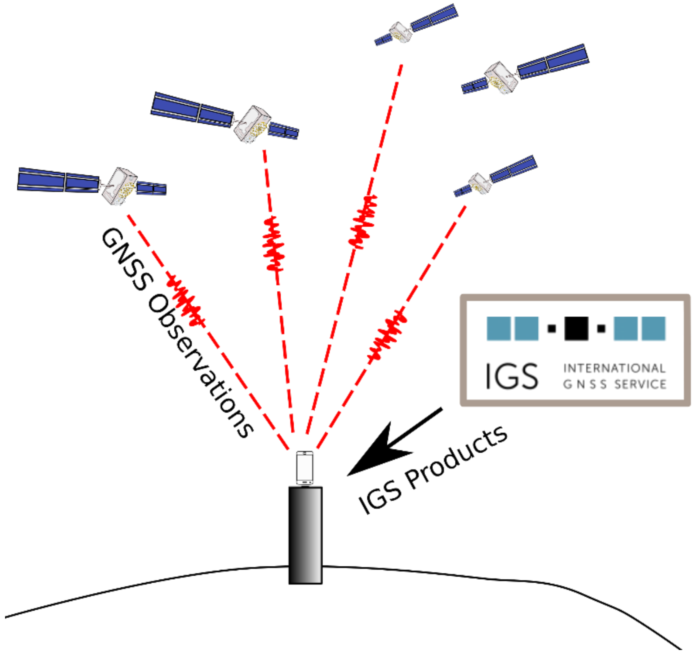




GNSS IoT Data Collection & Processing (2)

Case II – Crowdsourced GNSS Observations: Network Expansion

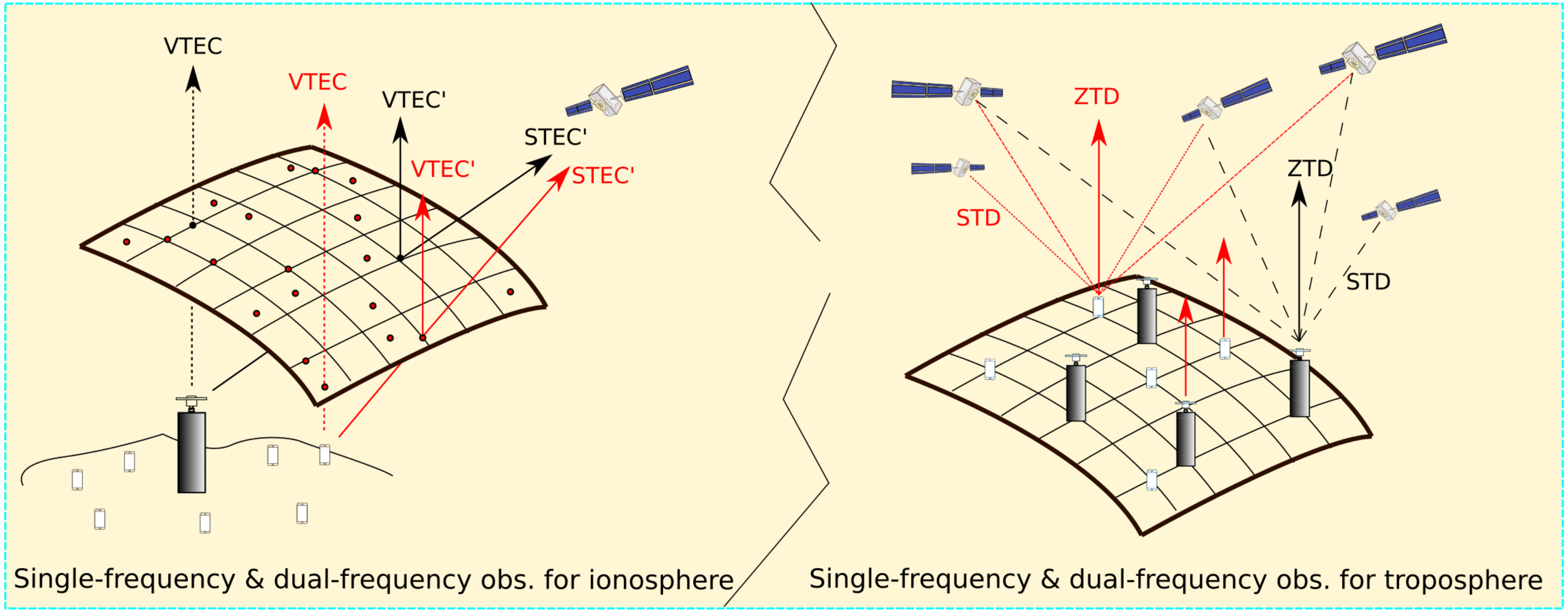




ETH zürich CAMALIOT IIASA

GNSS IoT Data Collection & Processing (3)

CamaliotGnss component for GNSS processing



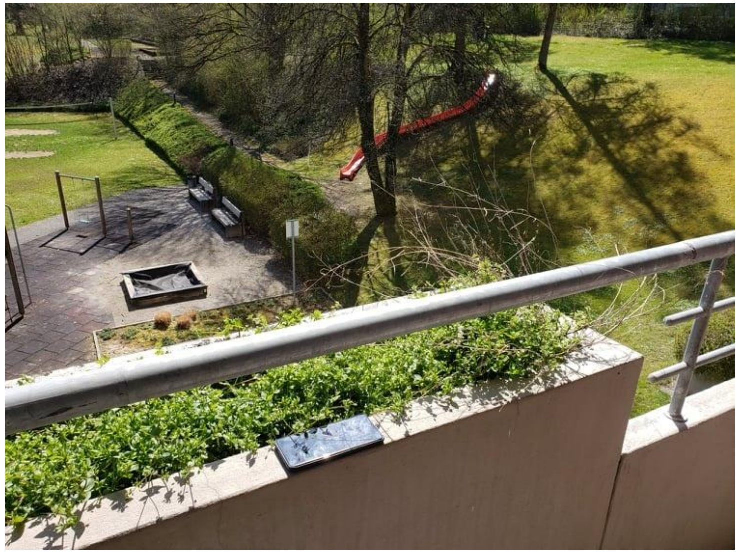
<u>CamaliotGnss as a part of CamaliotProcess:</u>

- Based on RTKLIB v2.4.3 b34
- Kalman Filter for parameter estimation: forward/backward/combined runs
- Support for RINEX2/RINEX3.0{2-4} with broadcast/precise orbit and clock products
- Support for baseline processing
- Support for ionosphere-free multi-constellation PPP

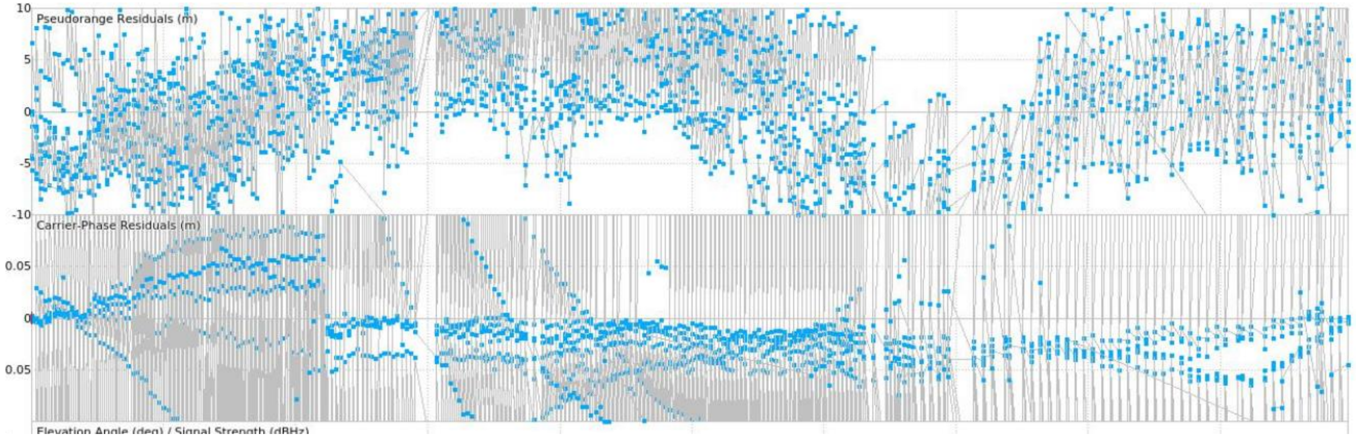
ETH zürich CAMALIOT IIASA

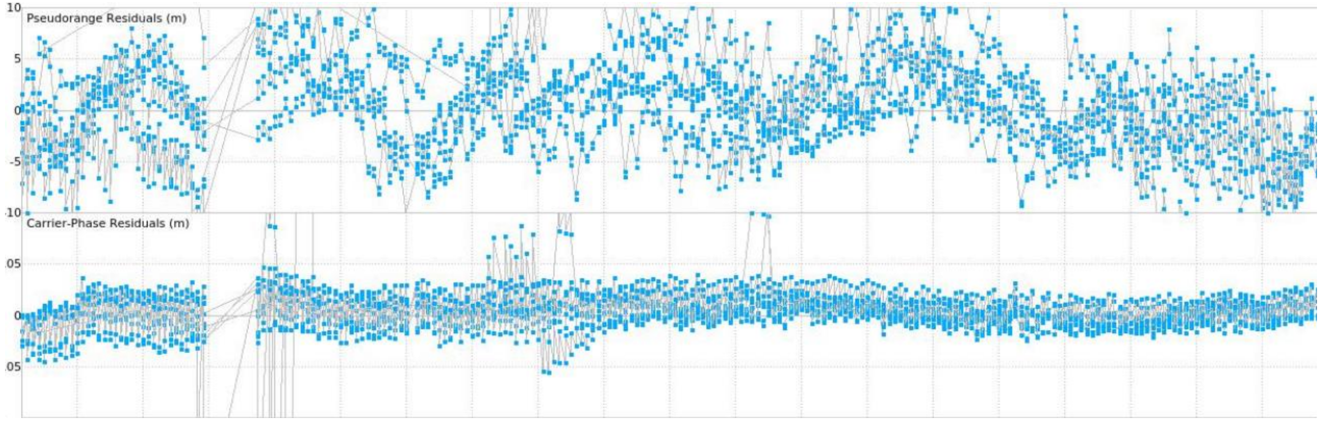
GNSS IoT Data Collection & Processing (4)

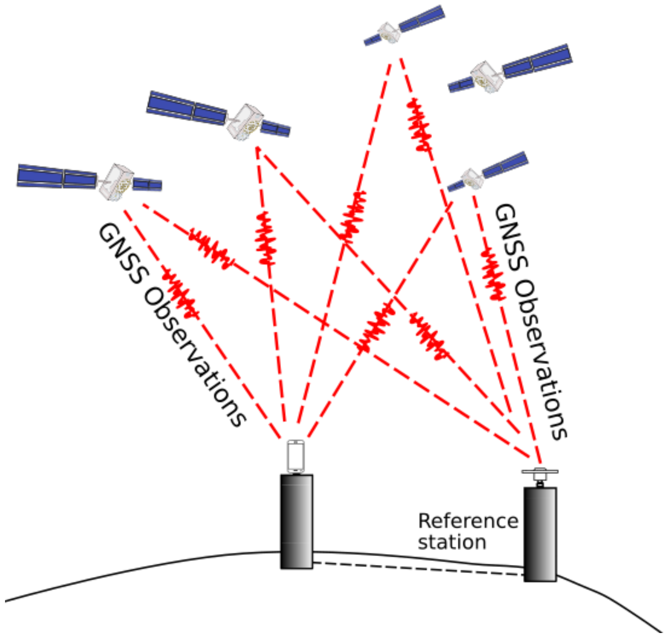
Samsung Galaxy S20 (SM-G981B) – Limited Visibility – 45 minutes – 0.8-km baseline



G+C L1: 3   5-minute batch: no pre-processing





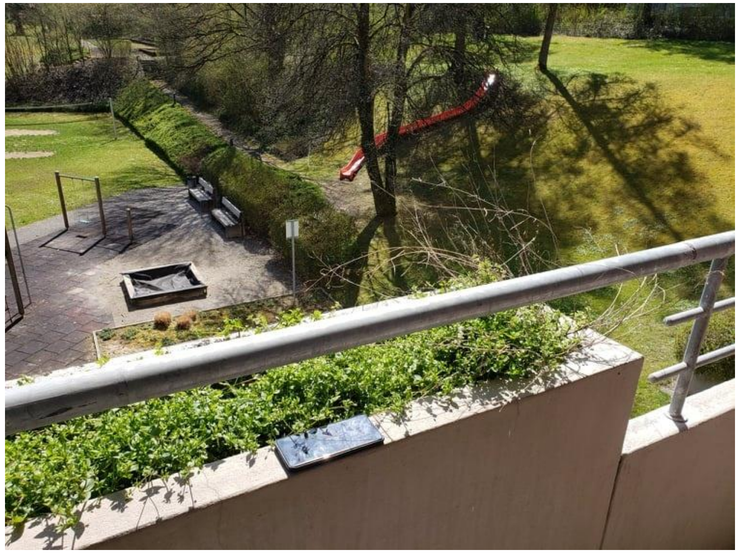


G+C L1: 1   5-minute batch: no pre-processing

ETH zürich CAMALIOT IIASA

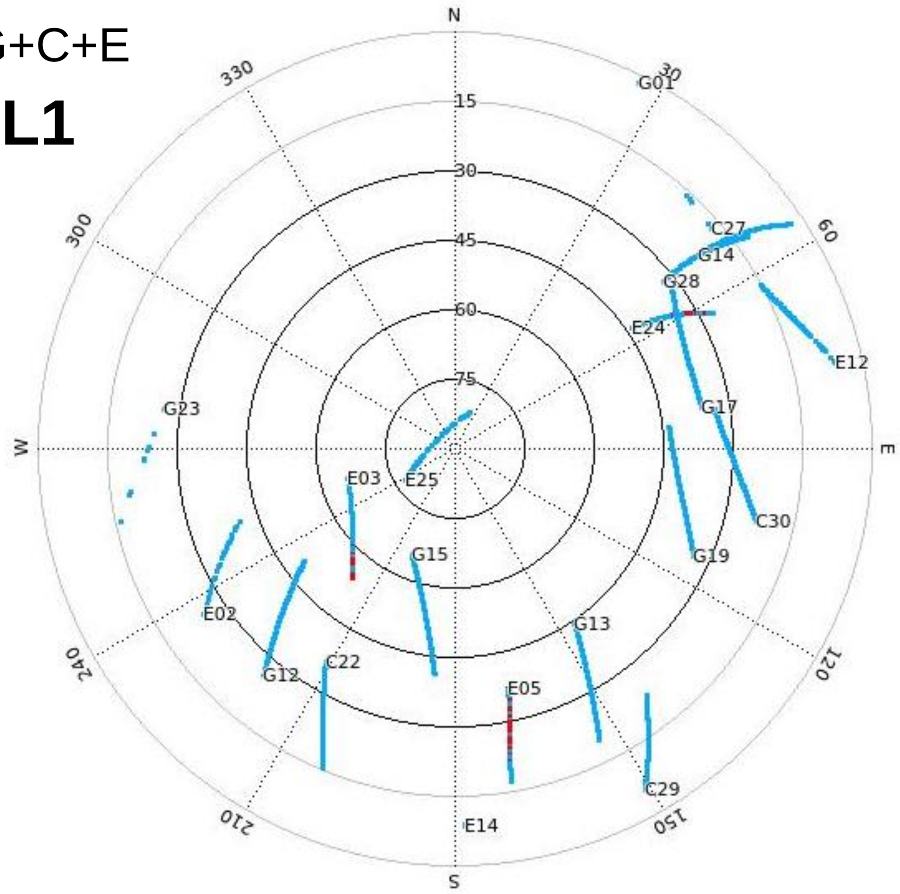
GNSS IoT Data Collection & Processing (5)

Samsung Galaxy S20 (SM-G981B) – Limited Visibility – 45 minutes – 0.8-km baseline

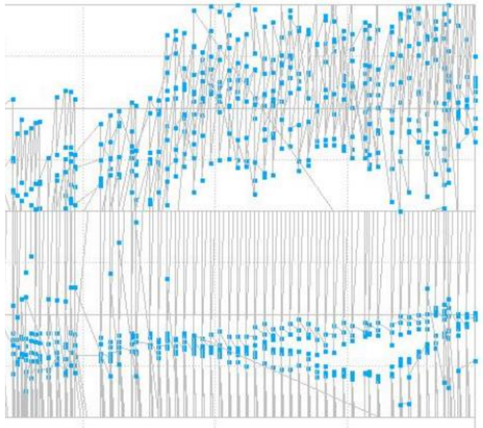


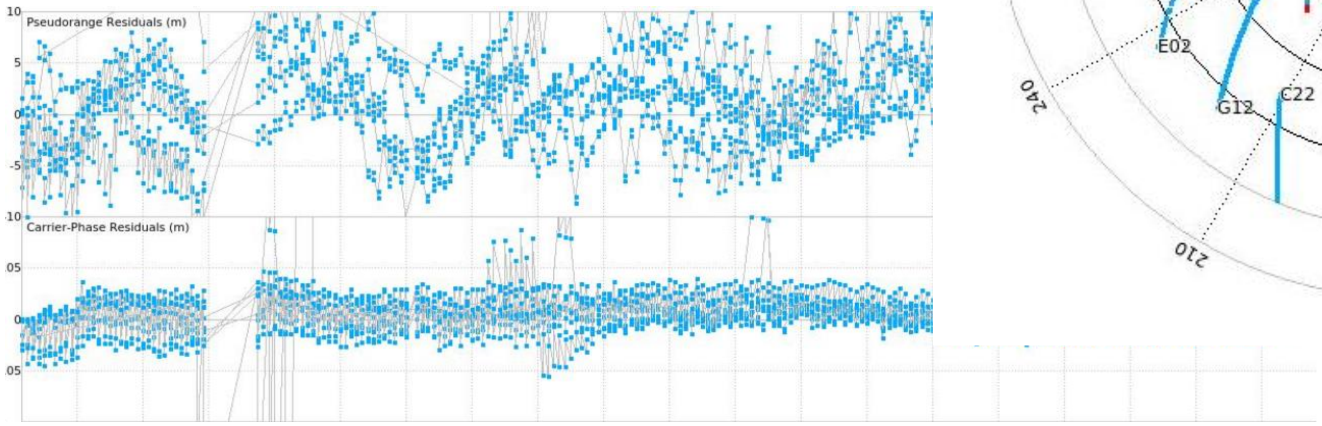
G+C+E

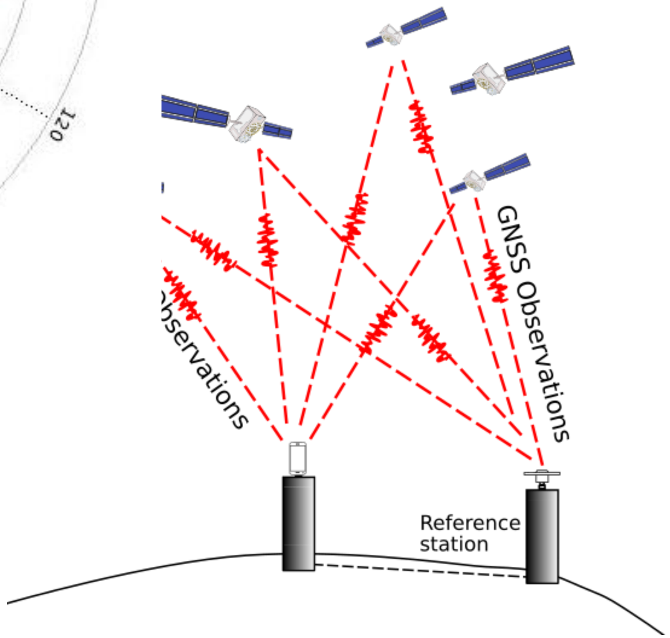
L1



no pre-processing



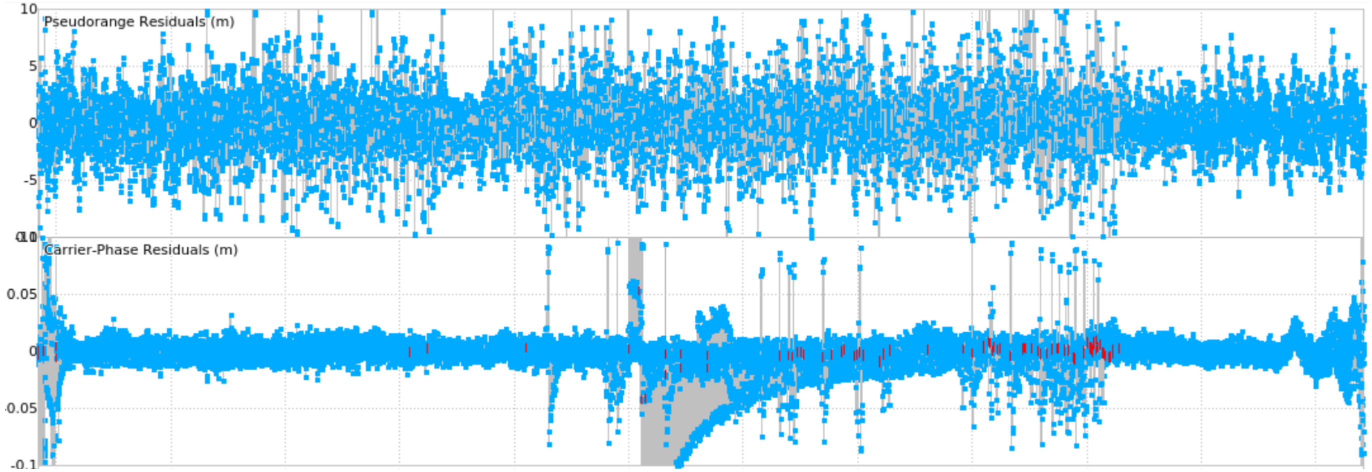


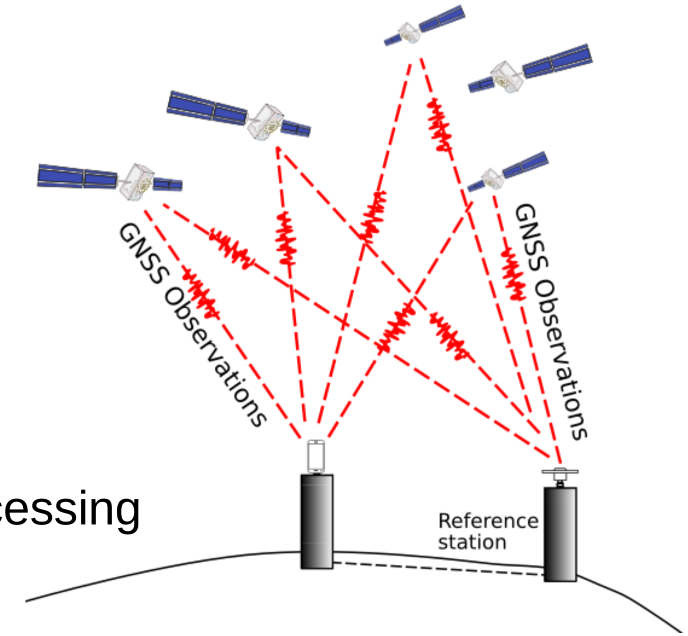


G+C L1: 1   5-minute batch: no pre-processing

GNSS IoT Data Collection & Processing (6)

Samsung Galaxy S20 (SM-G981B) – Open Field – 60 minutes – 5.6-km baseline



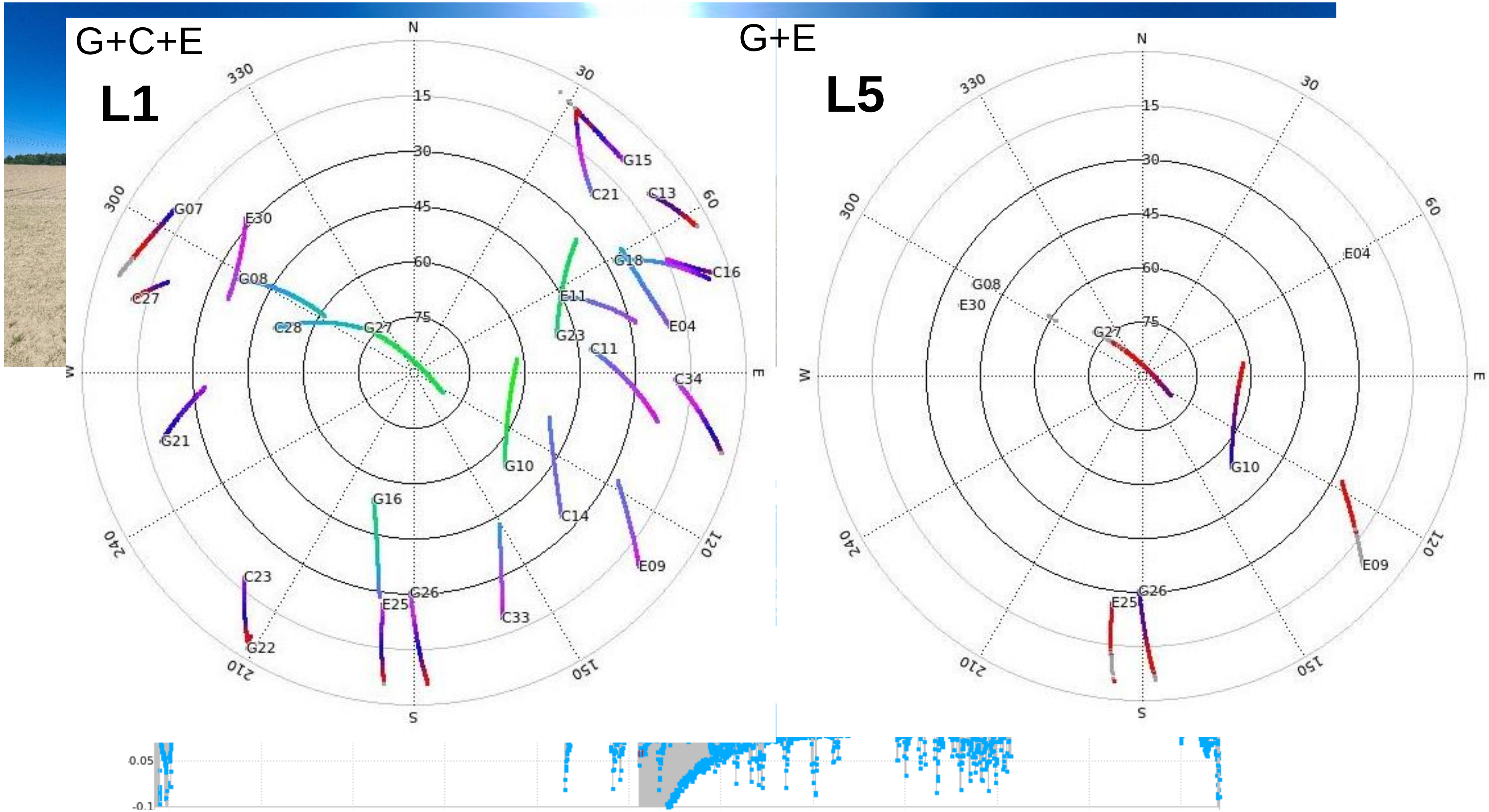


G+C L1: single 60-minute window: no pre-processing

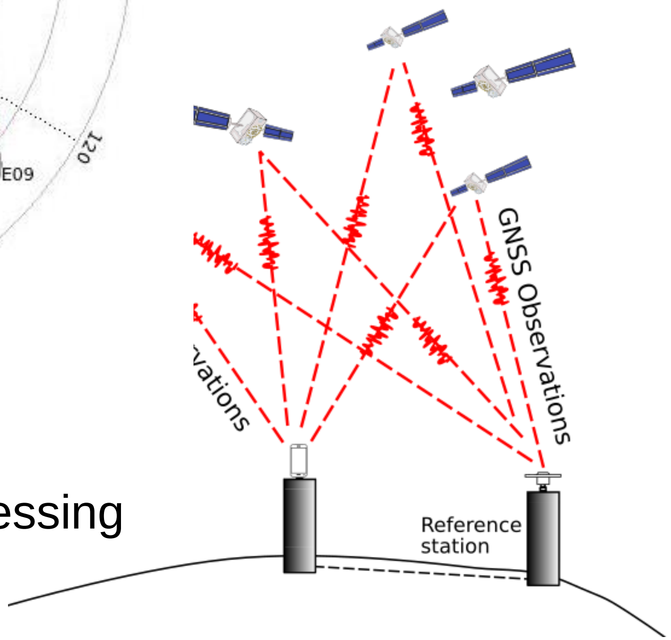
ETH zürich CAMALIOT IIASA

GNSS IoT Data Collection & Processing (6)

Samsung Galaxy S20 (SM-G981B) – Open Field – 60 minutes – 5.6-km baseline



G+C L1: single 60-minute window: no pre-processing



ETH zürich CAMALIOT IIASA

GNSS IoT Data Collection & Processing (7)

Smartphone-based TEC: First insights Based on Dual-frequency Carrier-Phase Observations

<u>Xiaomi Mi8</u>

<u>Initial investigation:</u>

- 3-hour observations on Nov 29, 2021

- Geometry-free (L1/L5) combination for STEC based on phase observations from RINEX3 (Geo++ RINEX Logger)

- Observations in the vicinity of the AGNES ETH2 station

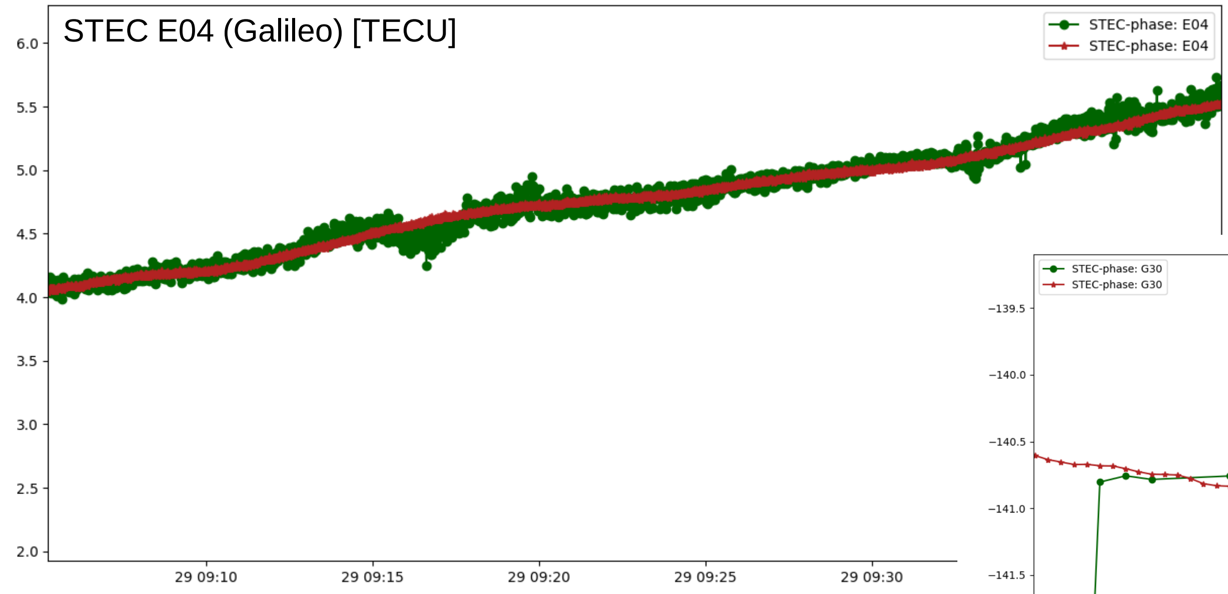
- Investigating “raw” satellite-specific STEC time series for both Xiaomi Mi8 and ETH2

<u>ETH2</u>

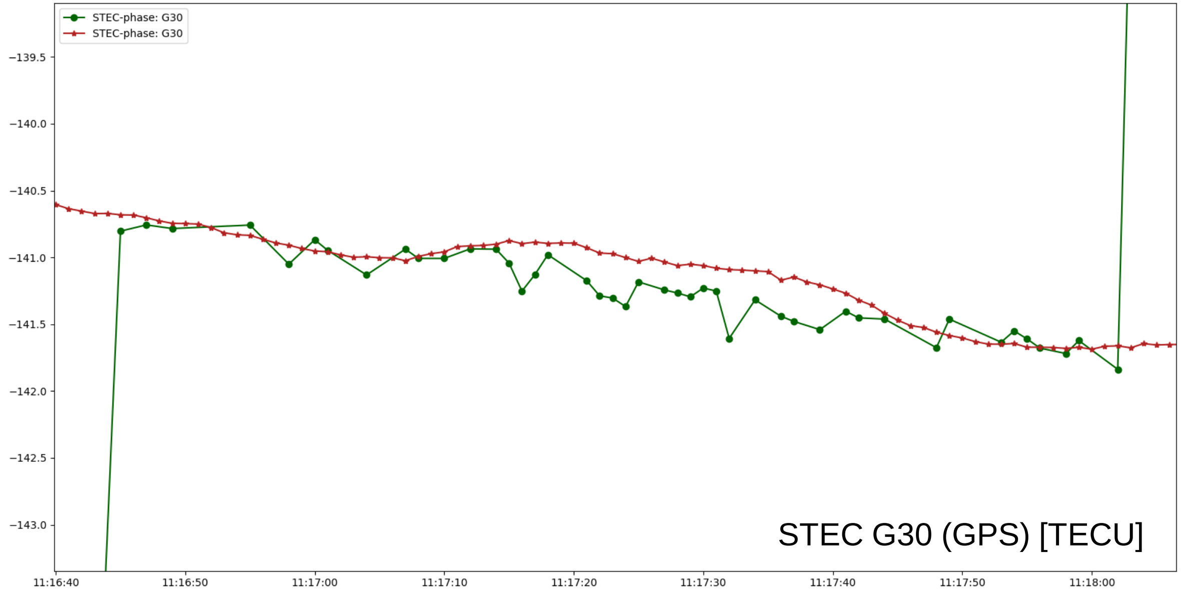
GNSS IoT Data Collection & Processing (9)

Smartphone-based TEC: First insights Based on Dual-frequency Carrier-Phase Observations

<u>Examples of satellite-specific STEC time series from Xiaomi Mi8 (green) and ETH2 (red)</u>



- STEC from Xiaomi Mi8 aligned (offset) manually to STEC from ETH2



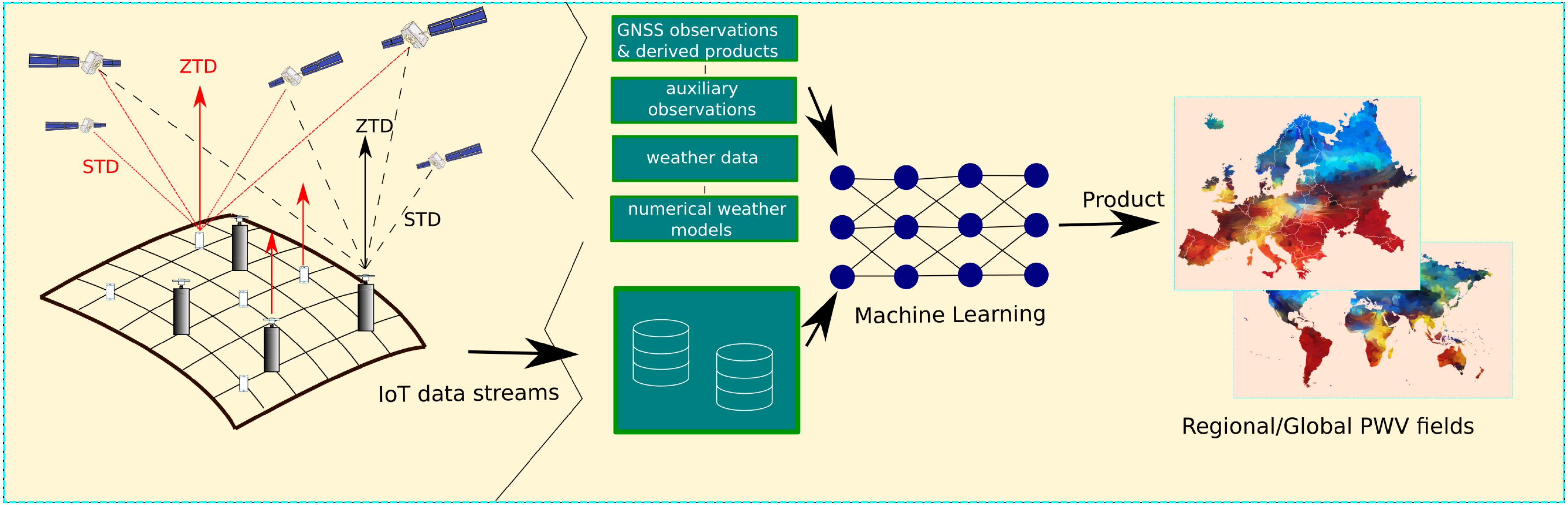
ETH zürich CAMALIOT IIASA

GNSS Science Use Cases

1. Troposphere modeling
2. Prediction of tropospheric parameters

3. Ionosphere modeling
4. Prediction of ionospheric parameters

Troposphere modeling



Troposphere modeling

- Prerequisites:
  - Tropospheric parameters from GNSS (IoT) data processing
    - Zenith wet delay (ZWD), gradients, precipitable water vapor (PWV)
  - Download external meteorological models (re-analysis and forecast) and data (meteo stations)
- Flow (automated):
  - Tropo data pre-processing / cleaning
  - Meteo data pre-processing / cleaning
  - Neural network (e.g., CNN, GNN) – prediction in space
    - Input: tropo parameters from IoT devices and GNSS stations, meteorological data
    - Target: tropo parameters at grid points
    - Training and validation based on geodetic tropo estimates and IGS products
- Output: improved 2-D field of ZWD, gradients, PWV

Troposphere / meteorological external data and models

- Geodetic observations of tropospheric parameters
  - **GNSS**, GNSS radio occultation, GNSS tomography
  - VLBI, DORIS,…
  - Water vapor radiometers
- Meteorological observations
  - Radiosonde observations, weather balloons, weather stations, weather buoys
  - Meteorological satellites
- Weather models / numerical weather prediction (NWP)
  - Global models
    - **ECMWF**, GFS, HRRR, NCEP,…
  - Regional models
    - e.g. COSMO in Switzerland
  - Forecasts and re-analysis

Prediction of ZWD in space (and time)

- Aim: predict ZWD in space (and time) using ML algorithms

- **Target: ZWD**
- **Features:**
  Latitude, Longitude, Time,
  + meteorological data (ERA5)

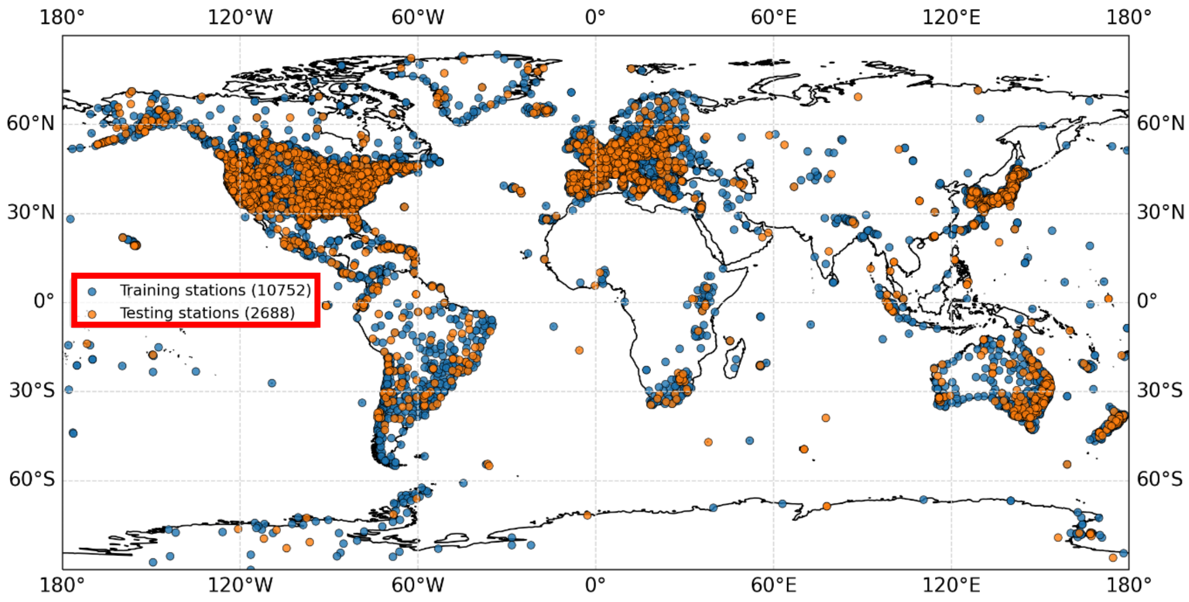
$\underbrace{\text{– u, v (wind speed) – t2m (temperature) – z (geopotential) – sp (surface pressure) – tp (total precipitation)}}_{\textbf{ground}}$

$\underbrace{\text{– z (geopotential) – r (relative humidity) – q (specific humidity) – t (temperature) – u, v (wind speed)}}_{\textbf{1000hPa}}$

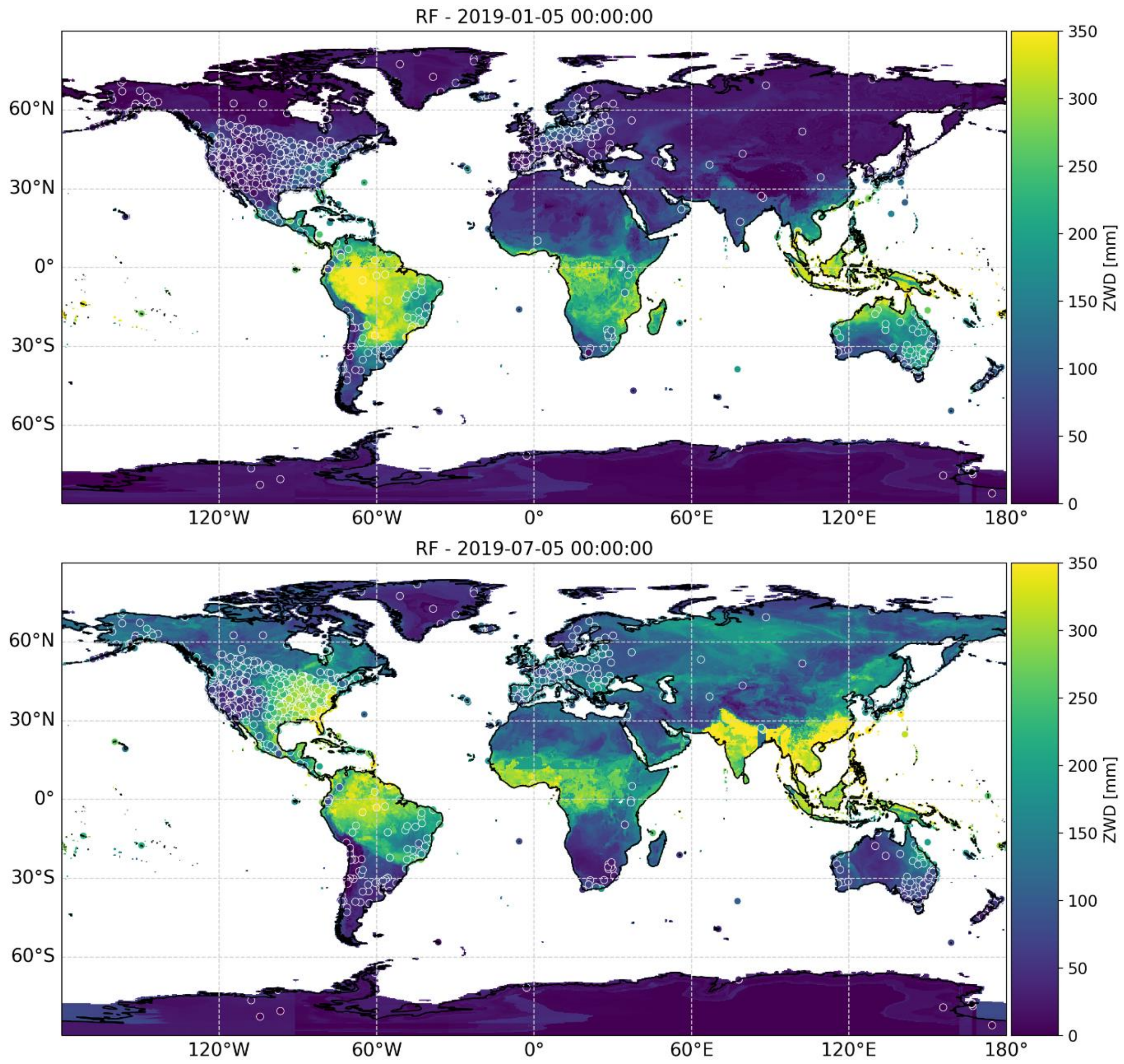
- Temporal resolution: hourly
- Spatial resolution: 0.25°
- Time span: year 2019

Global model

Distribution of training and test stations for all available stations (2019)



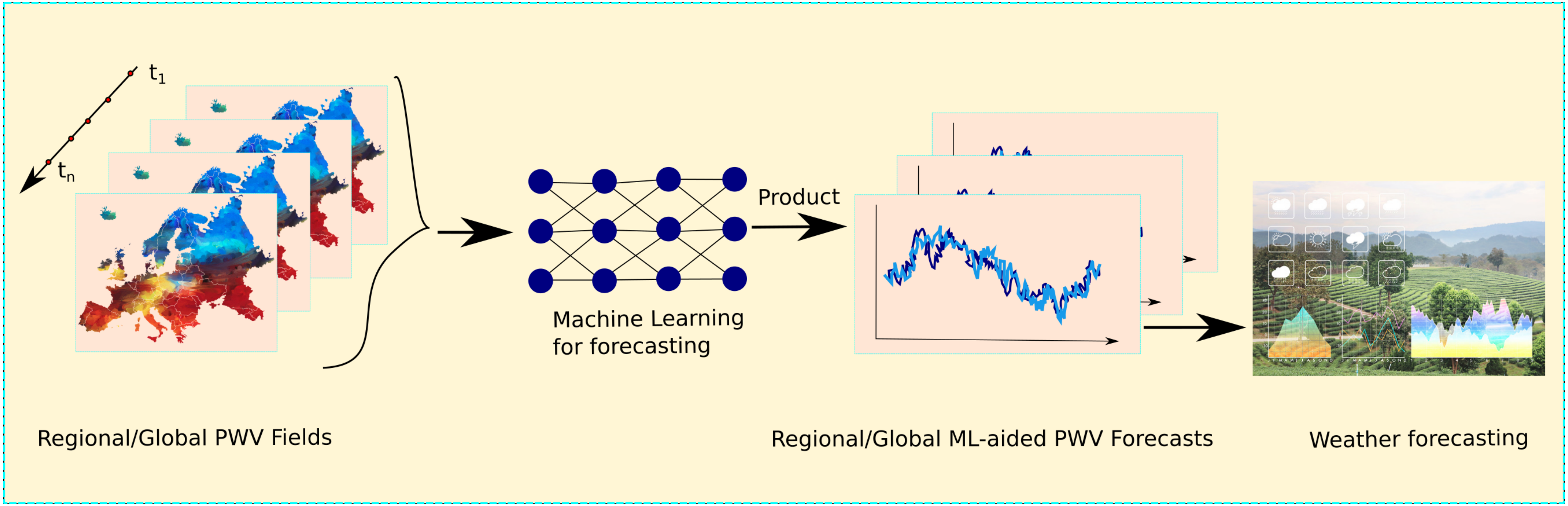
ZWD predictions



Overview of results – performance for testing stations
year 2019

	RMSE [mm]		R^2			
	Random Forest	XGBoost	Random Forest	XGBoost	Train + Test stations	#stations
Europe	14.96	13.75	0.903	0.918	2403 + 601	3004
North America	15.43	13.79	0.956	0.965	5284 + 1321	6605
WORLD	**18.07**	**18.68**	**0.947**	**0.944**	**10752 + 2688**	**13440**
Australia	18.17	15.23	0.917	0.942	532 + 133	665
Africa	27.27	25.67	0.853	0.870	241 + 61	302
South America	30.07	26.68	0.921	0.938	433 + 109	542

Prediction of tropospheric parameters

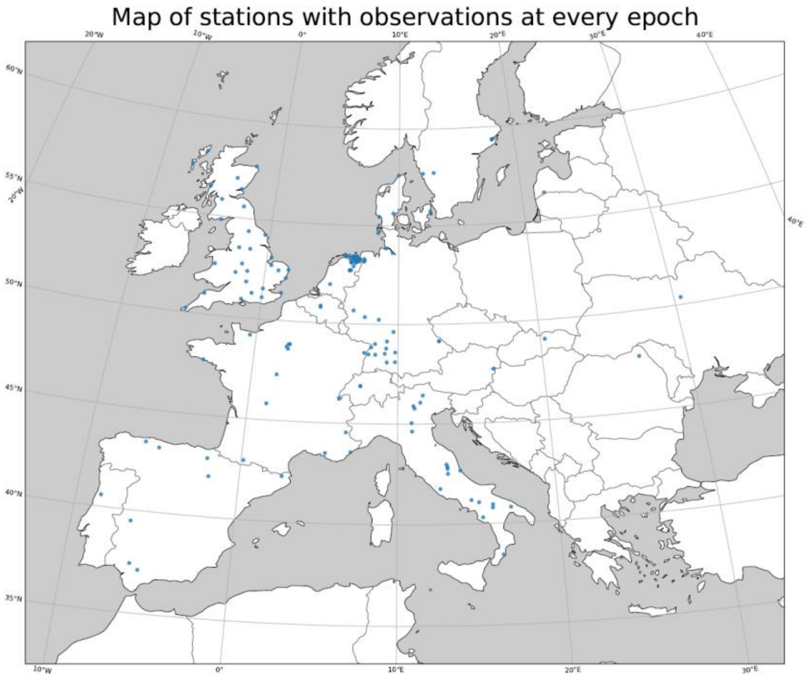


Prediction of tropospheric parameters

- Prerequisites:
  - 2-D field of ZWD, gradients, PWV from previous use case available for several epochs in the past
  - Meteo forecast models
- Flow (automated):
  - Tropo data pre-processing / cleaning
  - Meteo data pre-processing / cleaning
  - Recurrent neural network (e.g., LSTM) – prediction in time
    - Input: ZWD, gradients, PWV at grid points for several past epochs (e.g., 2 days), external data (past and future)
    - Target: ZWD, gradients, PWV at grid points for future epochs (e.g., 1 hour)
    - Training and validation based on tropo data from the past
- Output: 2-D field of ZWD, gradients, PWV predicted to future epochs

Troposphere modeling and prediction example

- Test case covering aspects of both tropospheric use cases
  - model GNSS tropospheric delays based on meteorological data
  - predict tropospheric delays into the future
- Data:
  - 145 GNSS station in Europe,
    - 5-min tropospheric zenith wet delays (ZWD)
  - Meteorological data from ECMWF ERA5 (1-h)
    - Pressure
    - Temperature
    - Relative humidity
    - Wind speed
    - Precipitation
    - Specific humidity
    - Geopotential



- Based on past 48 hours of meteorological data, predict ZWD 1 hour into the future

Machine learning approach

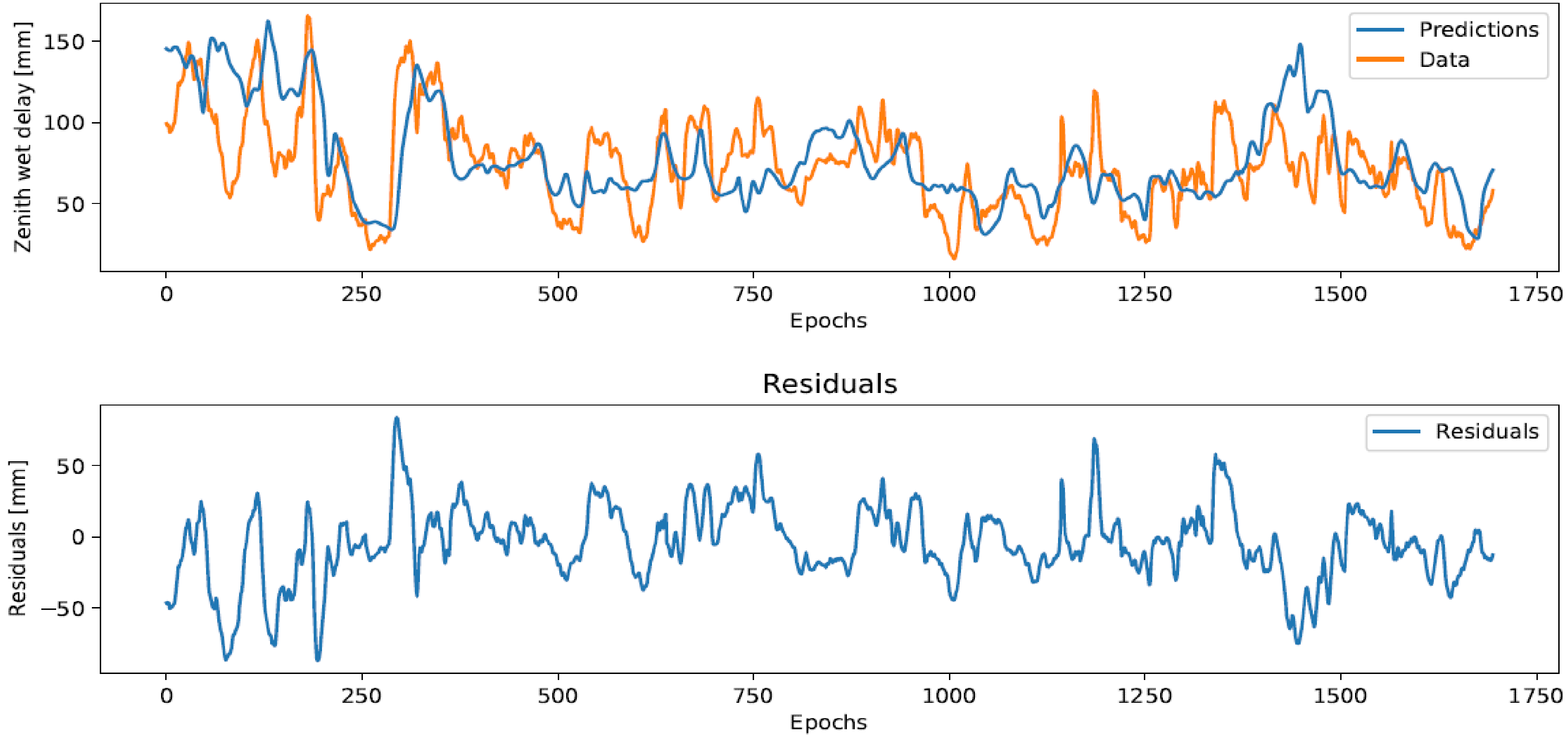
- Data pre-processing
- Models for individual stations
- Sliding window approach
- Split into training, validation, and test sets (70:10:20 ratio)
- LSTM neural network

Layer	Type	Nodes
1	LSTM	256
2	LSTM*	128
3	LSTM*	128
4	LSTM*	128
5	LSTM*	128
6	Dense	1

*…with dropout layer

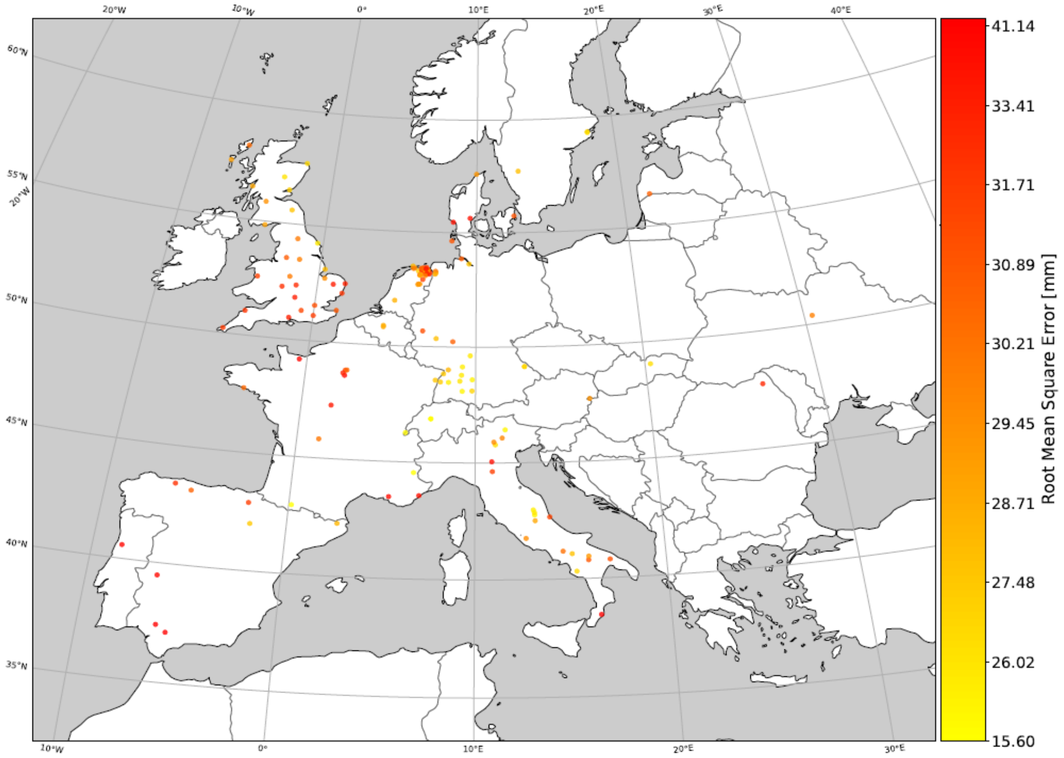
Results for station WTZR (Wettzell, Germany)

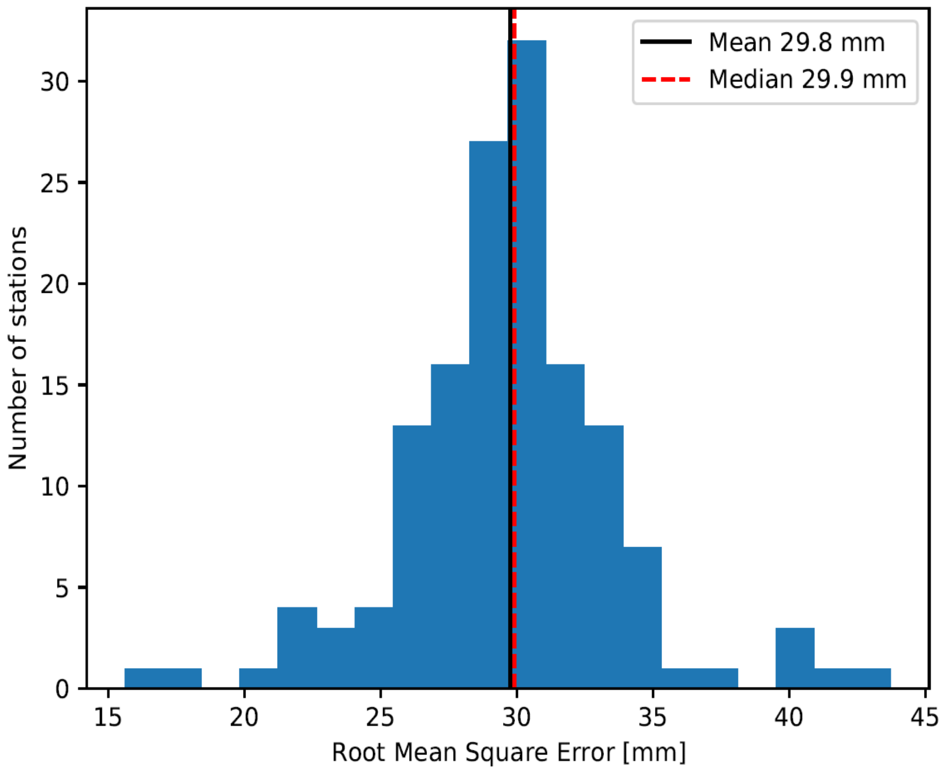
- RMSE of 26.6 mm for test set



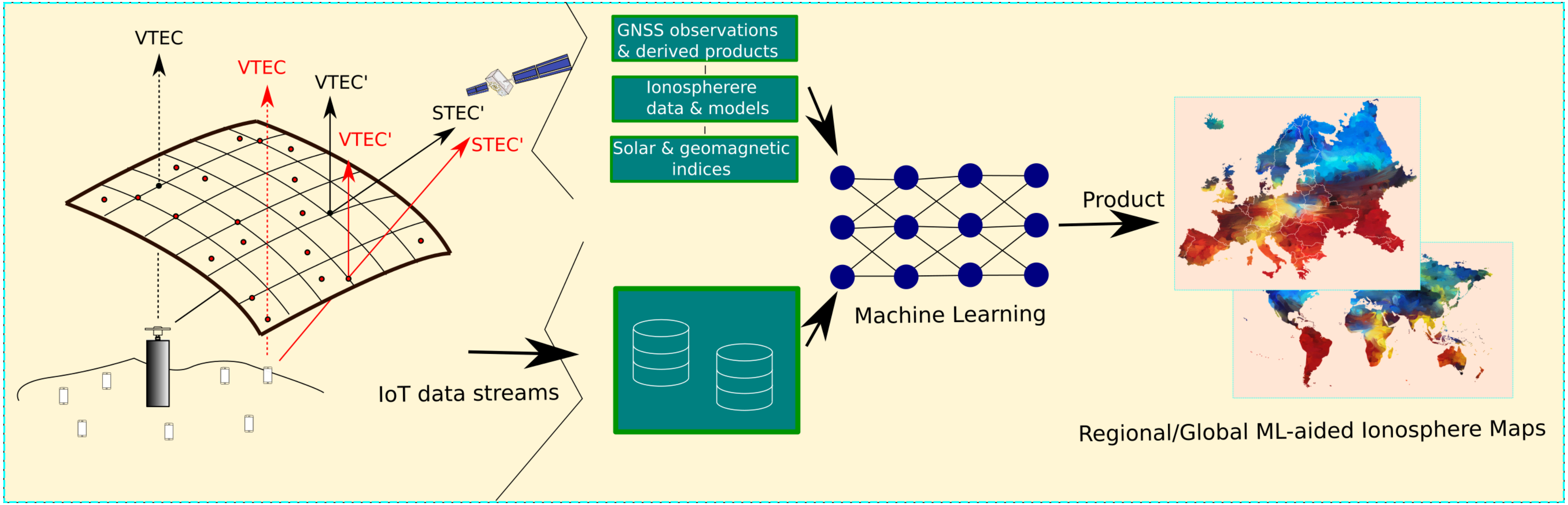
Results for 145 GNSS stations

- RMSE of 29.8 mm for test set





Ionosphere modeling



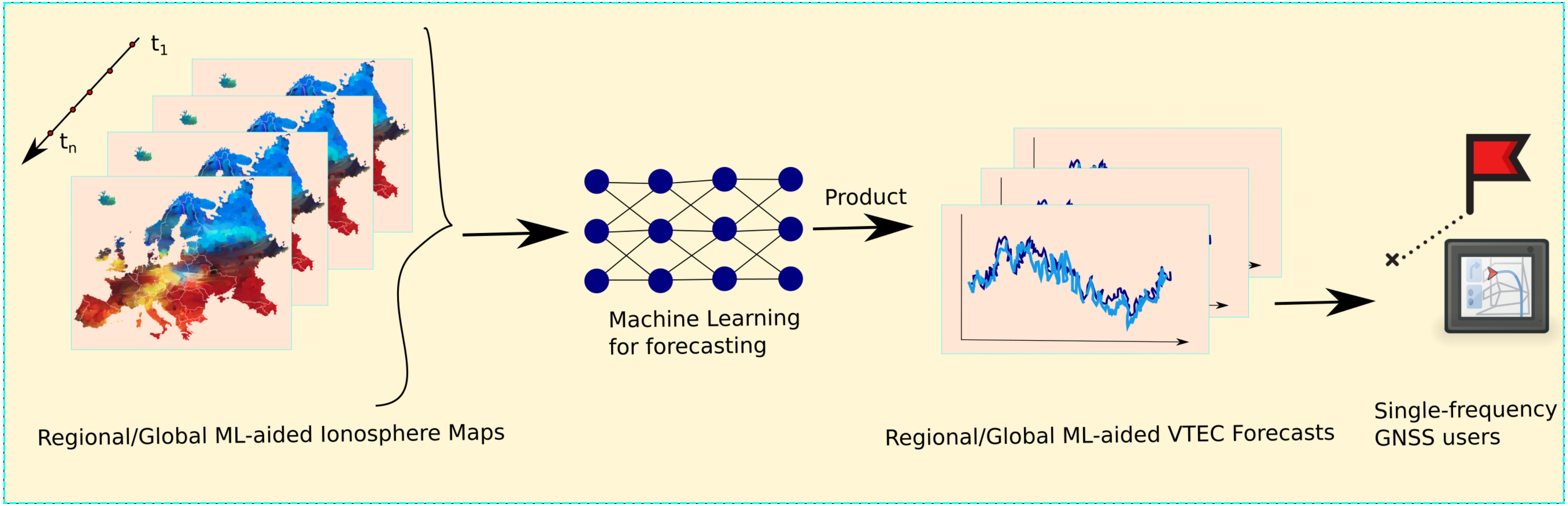
Ionosphere modeling

- Prerequisites:
  - VTEC from GNSS (IoT) data processing
    - geometry-free linear combination → slant TEC → VTEC
  - Download external models and data (ionosphere / space weather)
- Flow (automated):
  - VTEC pre-processing / cleaning
  - External data pre-processing / cleaning
  - Neural network (e.g., CNN, GNN) – prediction in space
    - Input: VTEC at IPPs related to IoT devices and GNSS stations, external data
    - Target: VTEC at grid points of global ionospheric map
    - Training and validation based on geodetic VTEC estimates, IGS products, or ionosphere models
- Output: improved VTEC grids similar to IGS GIM products

Ionosphere / space weather external data and models

- Geodetic observations of ionospheric parameters
  - **GNSS**, DORIS, altimetry, SWARM
- Other ionospheric measurements
  - Beacon measurements, ionosonde measurements, astronomic observations, e.g. LOFAR
- Solar observations
  - Coronagraphs (e.g. LASCO / STEREO) and other spacecraft observations
  - Global Ionosphere Flare detection network, Real Time Solar Wind observation network, solar flux measurements (**F10.7**), **Sunspot numbers**
- Products / services / models
  - Global ionospheric maps (**GIM from IGS combination** / analysis centers)
  - Ionospheric models (**IRI**, NeQuick), thermosphere models (GITM)
  - Products / indices / models for solar activity, solar wind, space weather
    - ESA Space Situational Awareness (SSA) / Space Weather (SWE)

Prediction of ionospheric parameters



Prediction of ionospheric parameters

- Prerequisites:
  - GIMs (VTEC grids) from previous use case available for several epochs in the past
  - External forecast data / models
- Flow (automated):
  - GIM pre-processing / cleaning
  - External data pre-processing / cleaning
  - Recurrent neural network (e.g., LSTM) – prediction in time
    - Input: VTEC at grid points for several past epochs (e.g., 2 days), external data (past and future)
    - Target: VTEC at grid points for future epochs (e.g., 1 hour)
    - Training and validation based on GIM data from the past
- Output: VTEC grid predicted to future epochs

Science Use Case – Ionosphere Temporal Prediction

Problem:

- ionospheric disturbances worsen the positioning accuracy of single-frequency GNSS-capable devices
- external information to correct for ionospheric delays in single frequency GNSS measurements

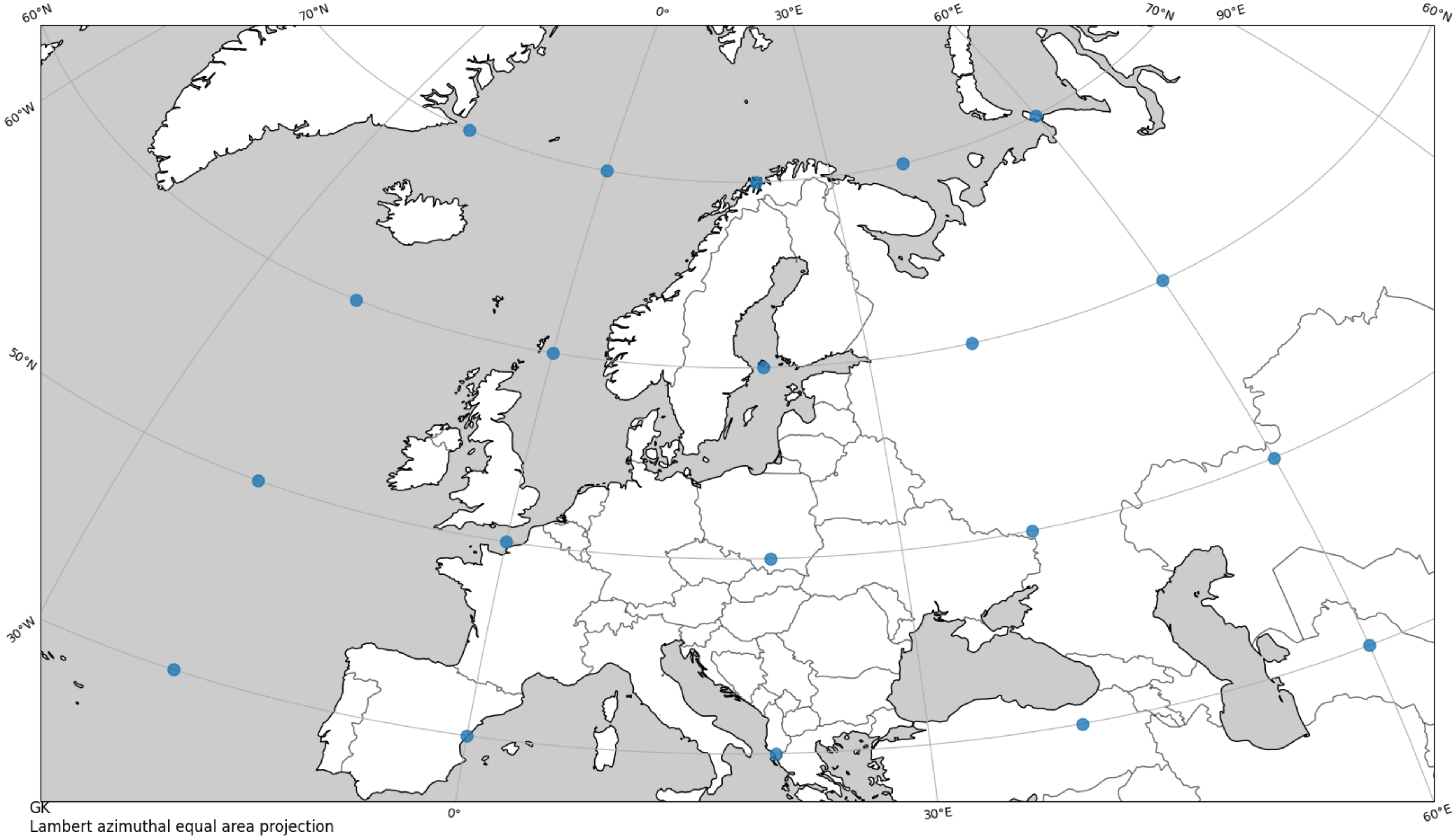
Solution:

- predict vertical total electron content (VTEC) in the future

Plans:

- predict VTEC for 24 hours using machine learning and based on VTEC time series extracted from global ionosphere maps (GIMs)
- incorporate additional models, indices, data sets of the ionosphere and space weather to improve the machine learning models

Grid Points Considered



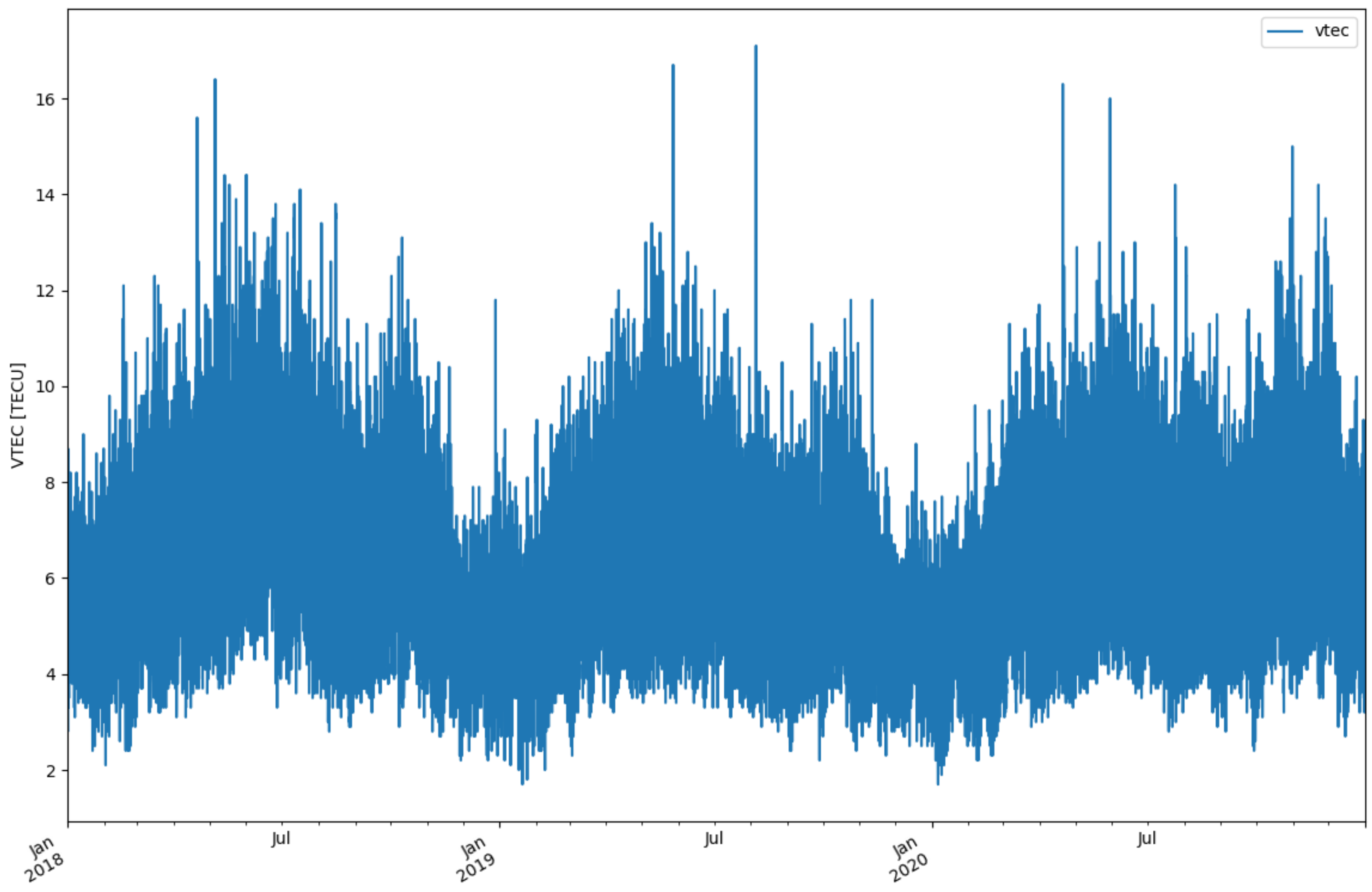
Data and Feature Engineering

- VTEC time series from IGS final solution with two-hour temporal resolution for year 2018 and 2019

- 5 input features: VTEC, and daily and yearly fluctuations as sin and cos functions

- will include space weather data in the future

VTEC Time Series (50 lat and 0 lon)



Multi-steps Prediction Models

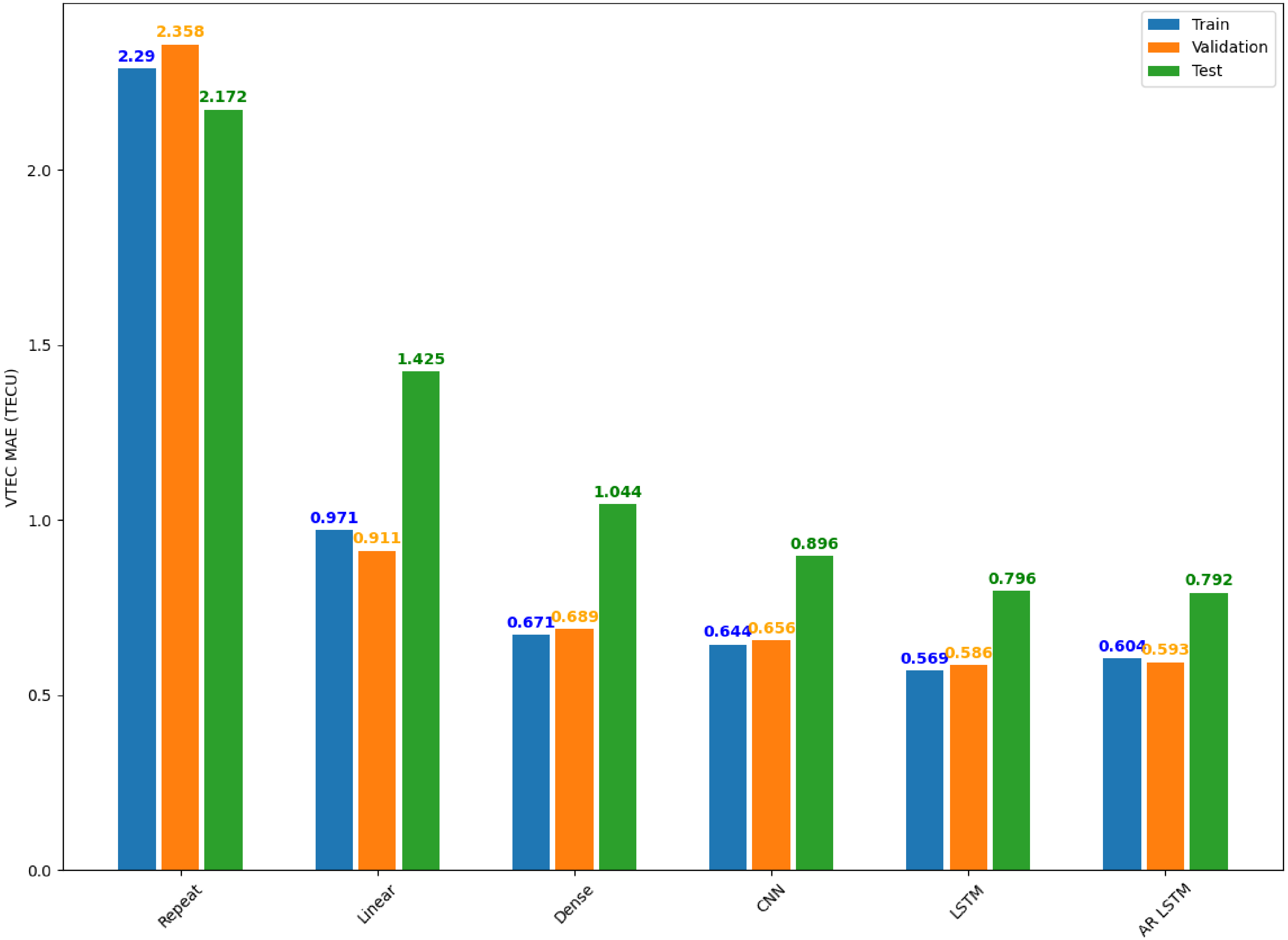
single-shot: make the predictions all at once

- Baseline Model
- Linear Model
- Dense Model
- CNN Model
- Long Short-Term Memory (LSTM) Model

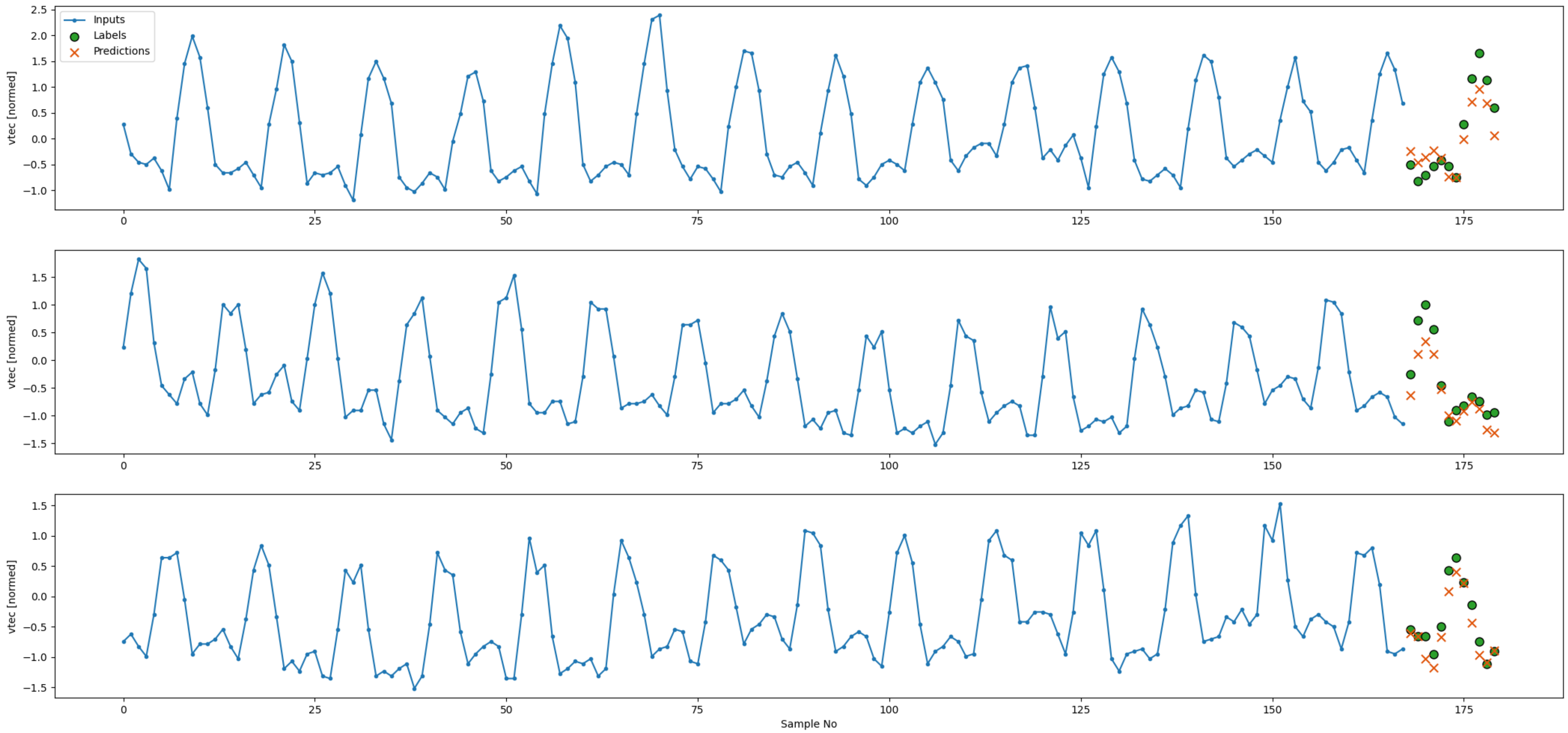
auto-regressive (AR): make one prediction at a time and feed the output back to the model

- AR-LSTM Model

Performance – Mean Absolute Error (MAE)



Prediction of LSTM Model



Thank you!
The CAMALIOT project

Linda See, B. Soja, G. Kłopotek, M. Awadaljeed, L. Crocetti, Y. Pan, M. Rothacher, R. Weinacker, T. Sturn, I. McCallum

May 2022