ClimAlytics AT

David Kalteis, Dominik Forsthuber, Michael Kerscher

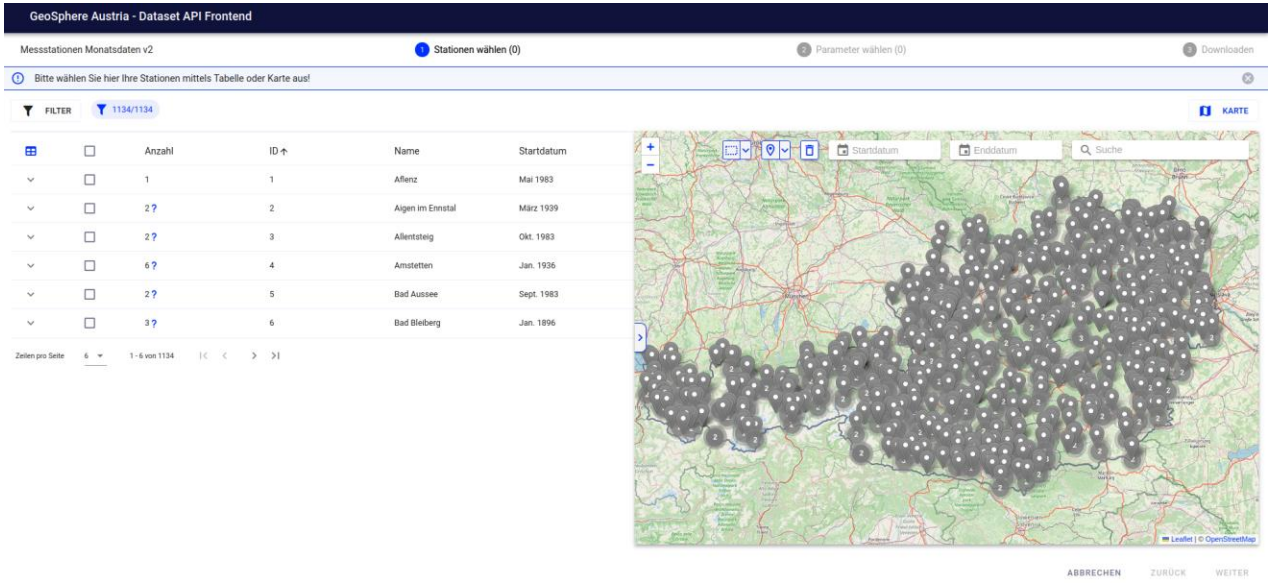
BIG Data - Project

HAGENBERG | LINZ | STEYR | WELS

FH OBERÖSTERREICH
UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Dataset

- Provider: GeoSphere Austria
- Agency: Bundesanstalt für Geologie, Geophysik, Klimatologie und Meteorologie



Technology Stack

Research Questions

- *How does the long-term trend in mean monthly temperature vary with elevation?*
- *Which geographic zones (valleys, plateaus, alpine corridors) show the largest shifts in “hot days” (≥ 30 °C) and “frost days” (≤ 0 °C) since 1970?*
- *How do station installation dates and validity periods create spatio-temporal gaps, and where are the largest “data deserts”?*
- *Which seasonal windows and locations optimize safety—combining sunshine hours, wind-gust flags, and frost/heat indicators?*

RQ1: Long-Term Temperature Trends

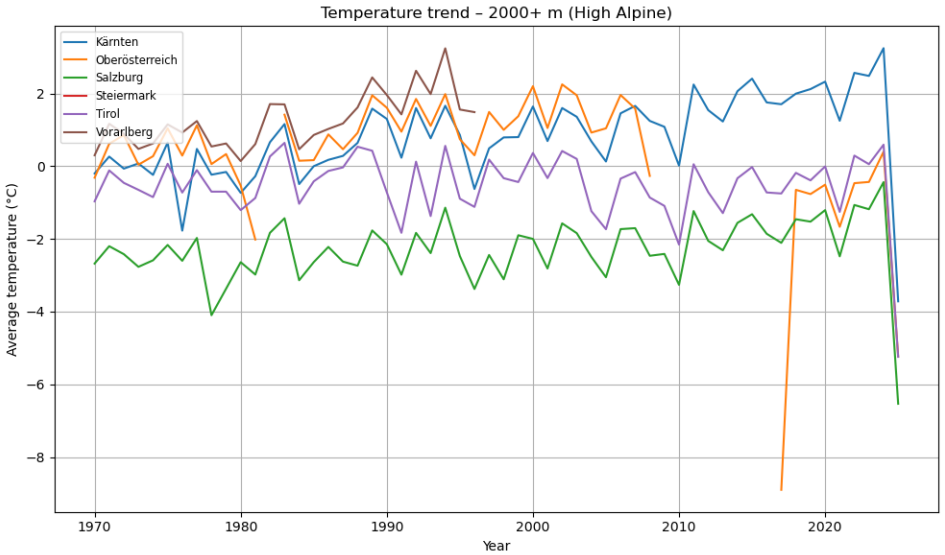
Objective:

- Long-term temperature trends differ by elevation

Methodolgy:

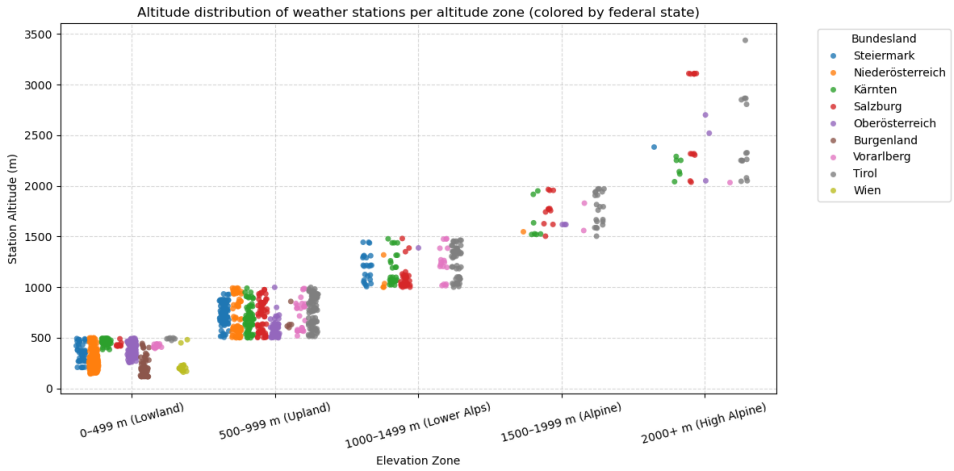
- Dataset of measurements with relevant columns: `stations_id, date, tl_mittel`
- Join stations metadata to classify into one of five elevation bands:
  - 0 – 499 m -> Lowland
  - 500 – 999 m -> Upland
  - 1000 – 1499 m -> Lower Alps
  - 1500 – 1999 m -> Alpine
  - 2000+ m -> High Alpine
- Aggregate by year, elevation zone and region

RQ1: Long-Term Temperature Trends



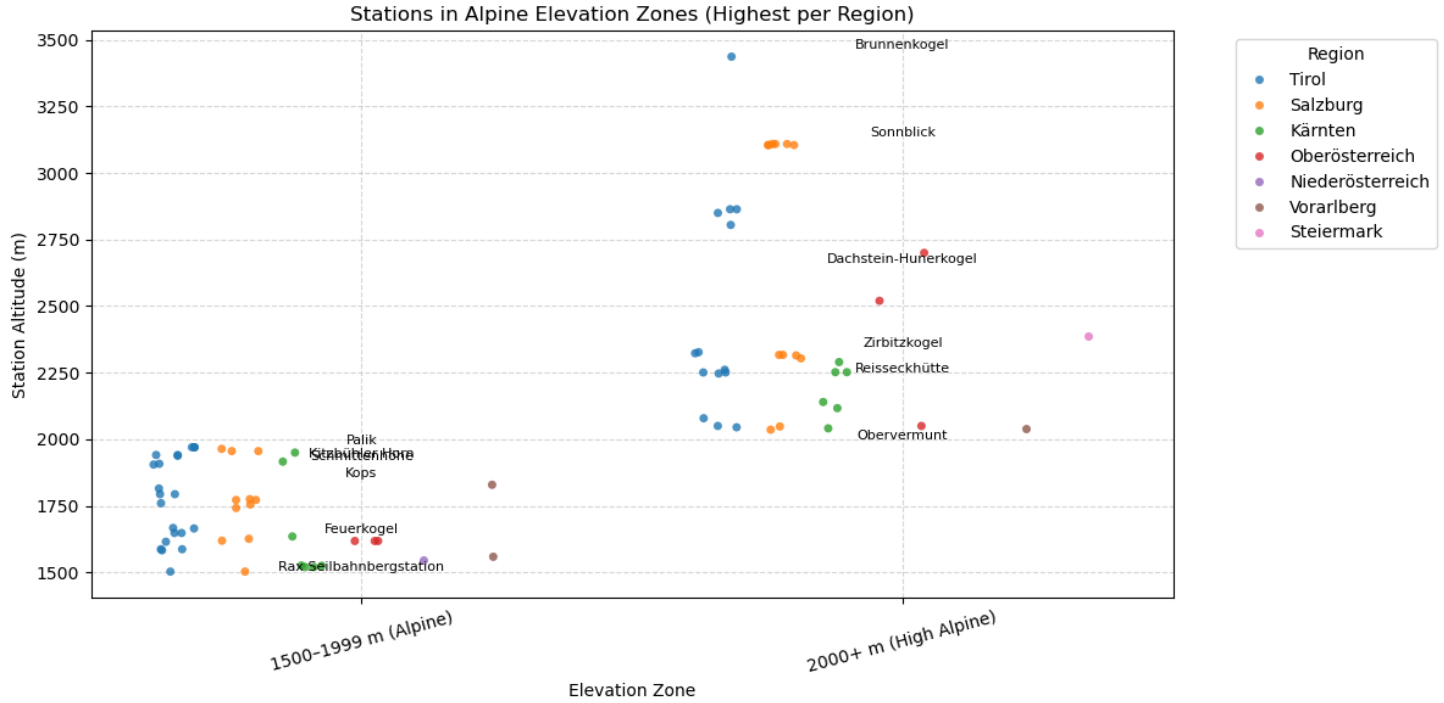
- Average temperature from 1970 – 2025
- All regions with High Alpine zone
- General increase
- Changes and missing values are visible
  - > Sensor malfunctions
  - > Missing or incomplete historical data
  - > Gaps due to station shutdowns

RQ1: Long-Term Temperature Trends



- Distribution of weather stations per elevation zone
- To validate
  - > the results of temperature trends
  - > Stations in High Alpine are located well above 2000+
- Regional diversity is given

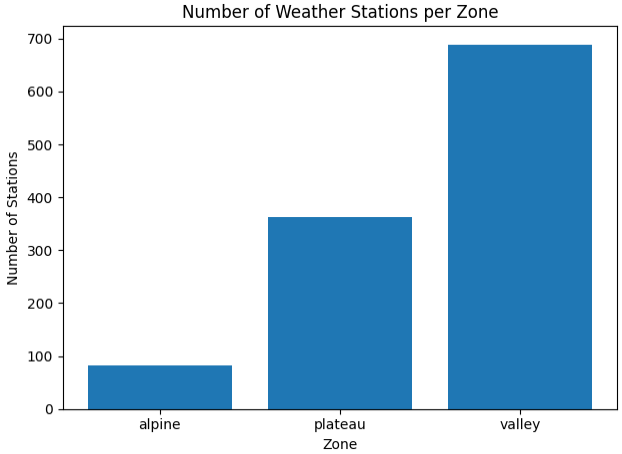
RQ1: Long-Term Temperature Trends



RQ2: Spatial Patterns of Extremes

- Trends in “hot days” (≥ 30 °C) and “frost days” (≤ 0 °C) since 1970
  - ***tage_tropen**,Tropentage,"Zahl der Tropentage, Tagesmaximum der Lufttemperatur in 2 m Höhe >=30.0°C",d*
  - ***tage_frost**,Frosttage,"Zahl der Frosttage, 24-Stunden-Minimalwert der Lufttemperatur in 2m Höhe < 0.0 °C",d*

- Zones defined by elevation:
  - **Valley**: ≤ 700 m
  - **Plateau**: 701–1500 m
  - **Alpine**: > 1500 m



RQ2: Full Time Series Analysis

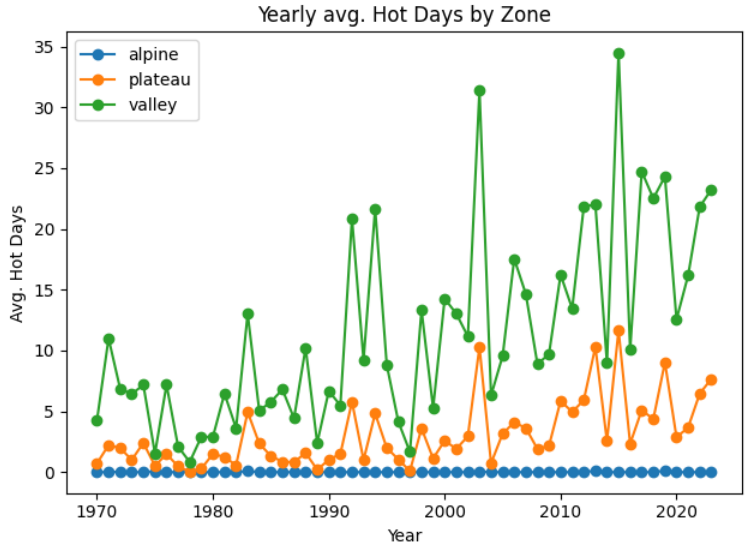
Data Processing

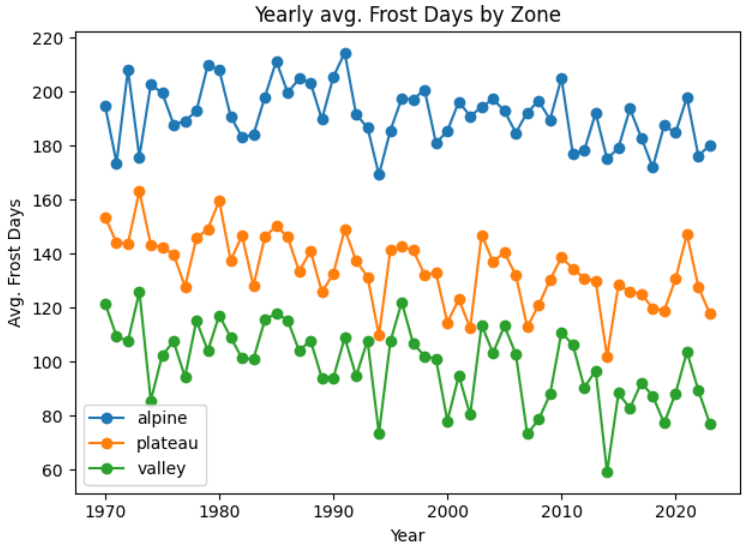
- Ingest raw CSV into Spark; parse date & extract year
- Join station metadata; attach elevation → assign zone
- Repartition by year & zone, write to Parquet for fast queries

Visualization

- Compute annual, per-station averages of hot & frost days
- Plot time series (1970–2023) for each zone

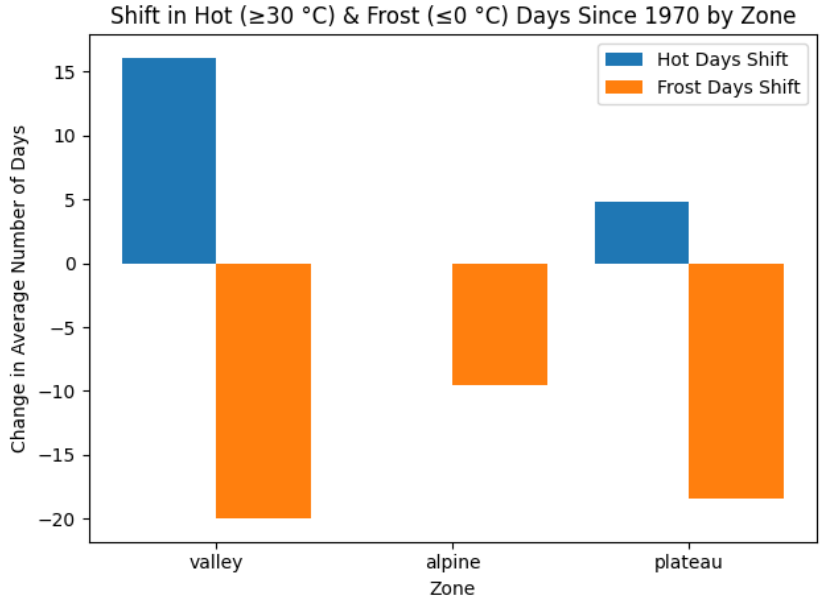
RQ2: Full Time Series Analysis





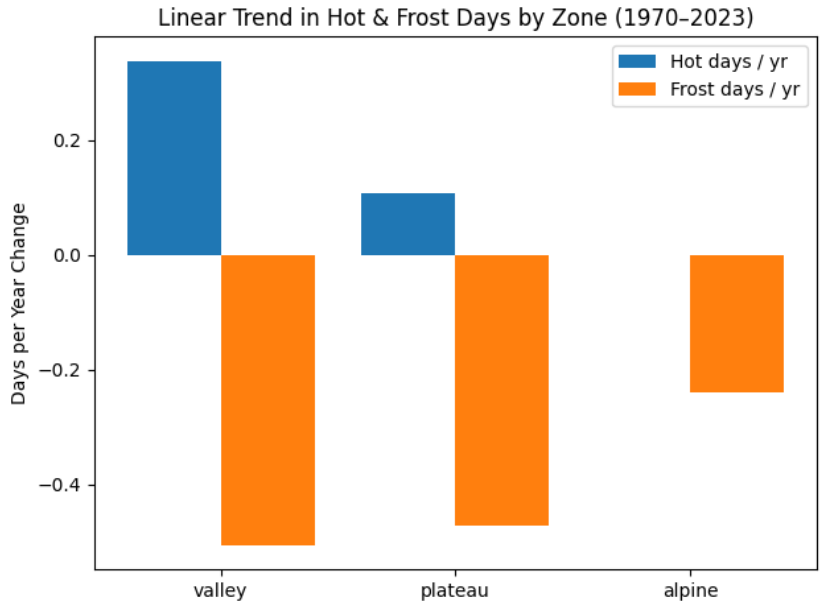
RQ2: End–Minus–Start Difference

- Calculate mean(2015–23) – mean(1970–79) by zone
- Highlights net shifts over 50 years
- Valleys: largest ↑ hot days & ↓ frost days; plateaus moderate; alpine minimal



RQ2: Linear Regression Trends

- Fit a straight trend line to each zone's annual hot-day and frost-day averages
- Slope = average change in days per year (↑ or ↓)
- Incorporates every year's data, making the trend resistant to outlier years



RQ3: Spatio-Temporal Data Coverage Pro.

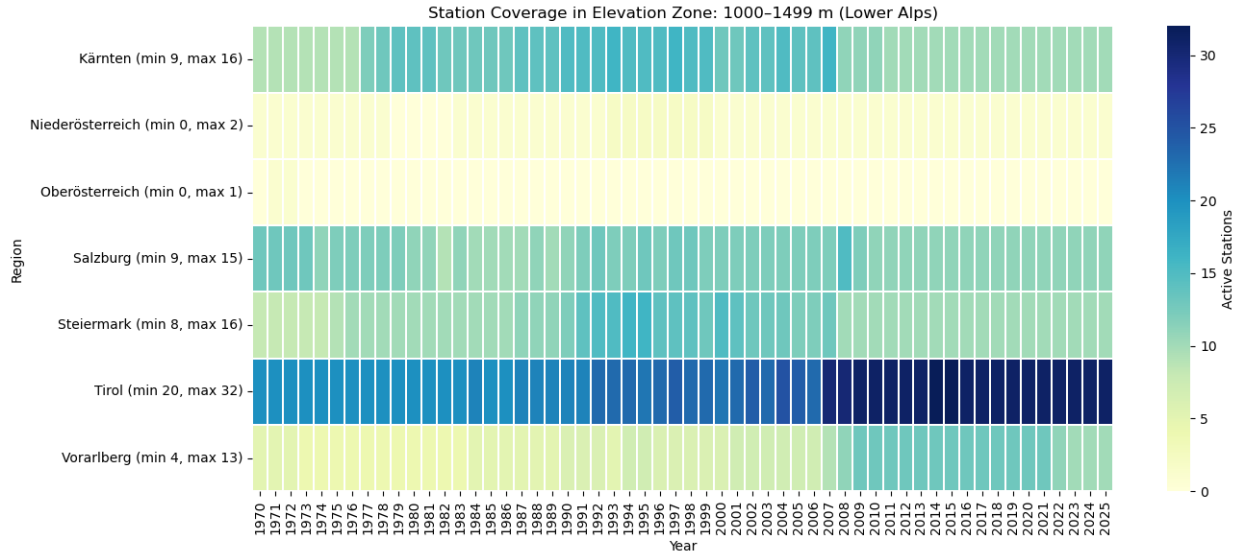
Objective:

- Identify spatial and temporal coverage gaps

Methodolgy:

- Two main steps:
  1. Metadata-Based Coverage
  2. Real Measurement-Based Coverage
     - To verify acutal data presence

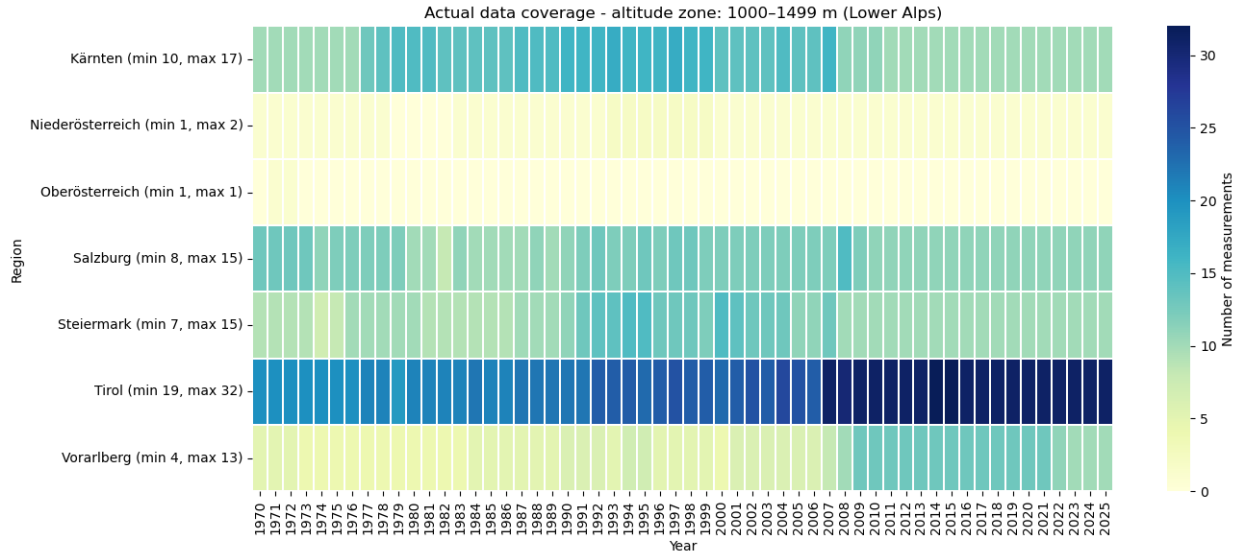
RQ3: Spatio-Temporal Data Coverage Pro.



Metadata-Based Coverage

- Using station metadata
- Filter by operational dates -> start / end
- Classify into elevation zones by altitude
- Count active stations per year x elevation zone x Bundesland

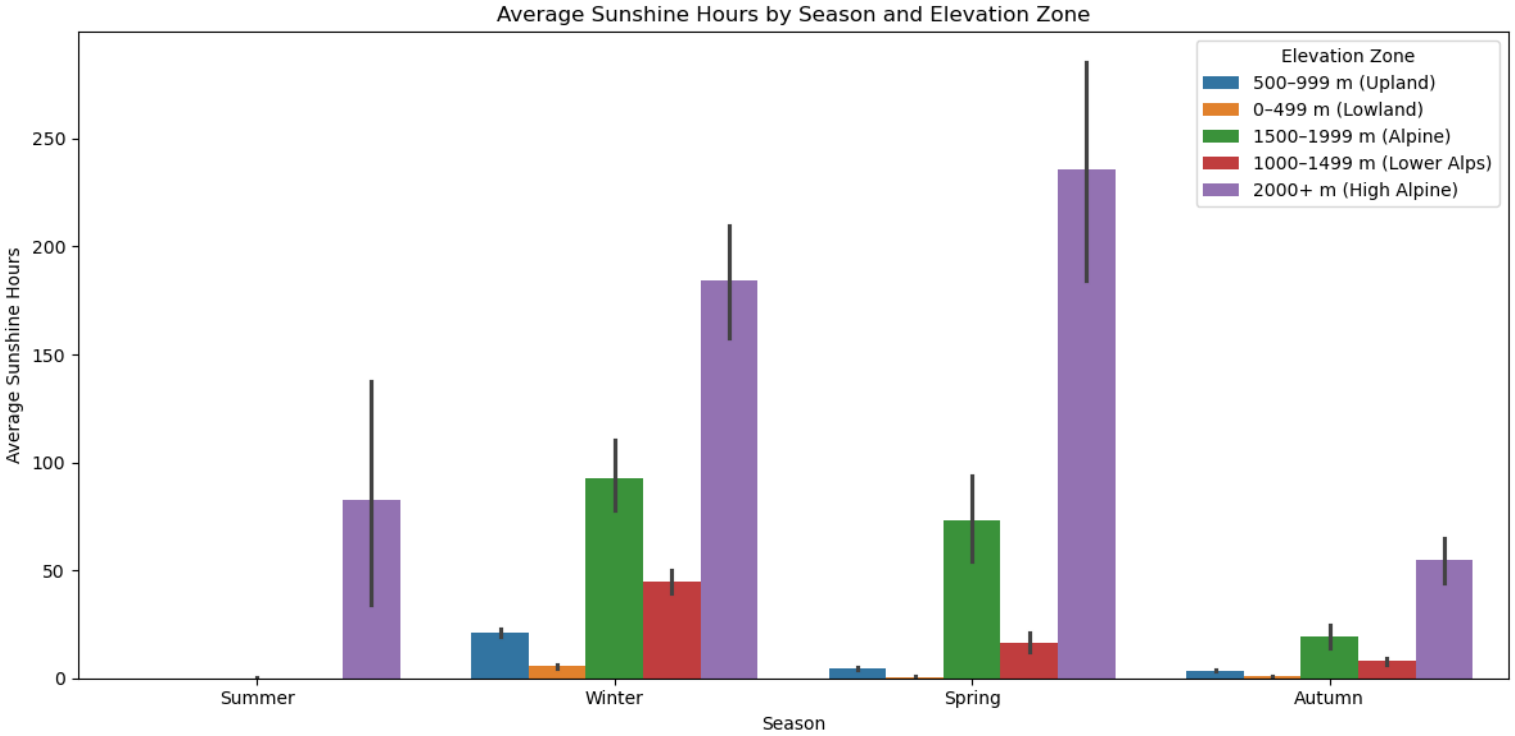
RQ3: Spatio-Temporal Data Coverage Pro.



Real Measurement-Based Coverage

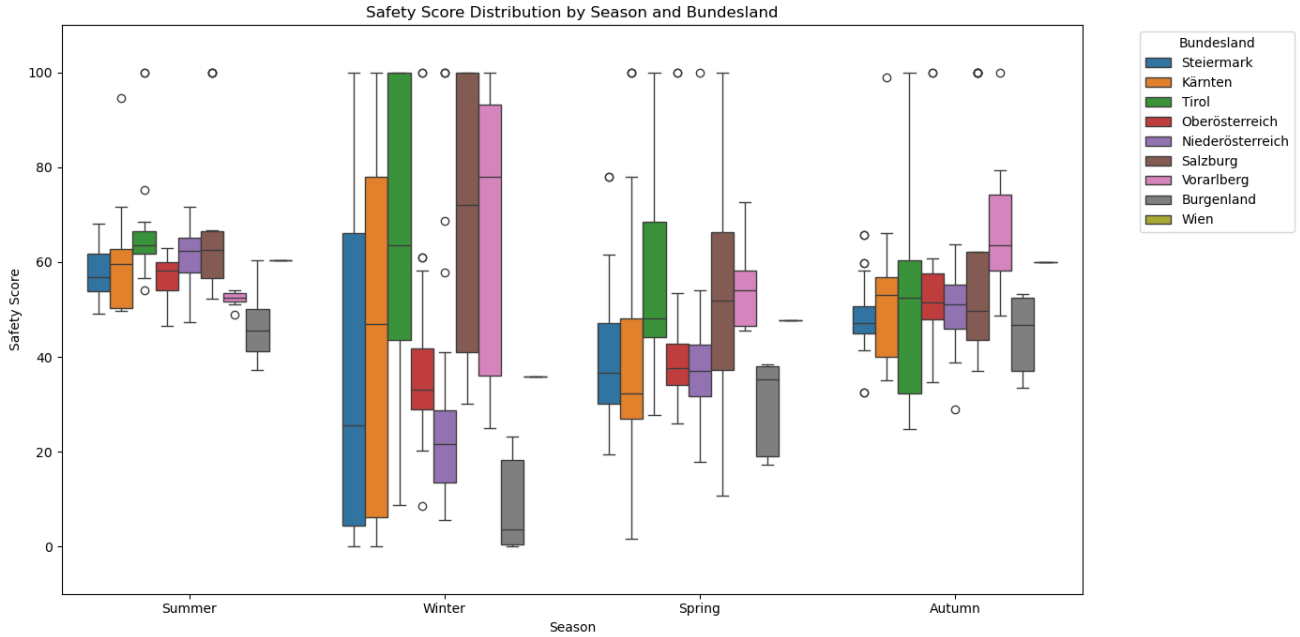
- Filter climate dataset for rows with non-null measurements
- Join with metadata and classify by elevation zone and Bundesland
- Count distinct active stations per year x zone x Bundesland

RQ4: Operational Window Optimization.

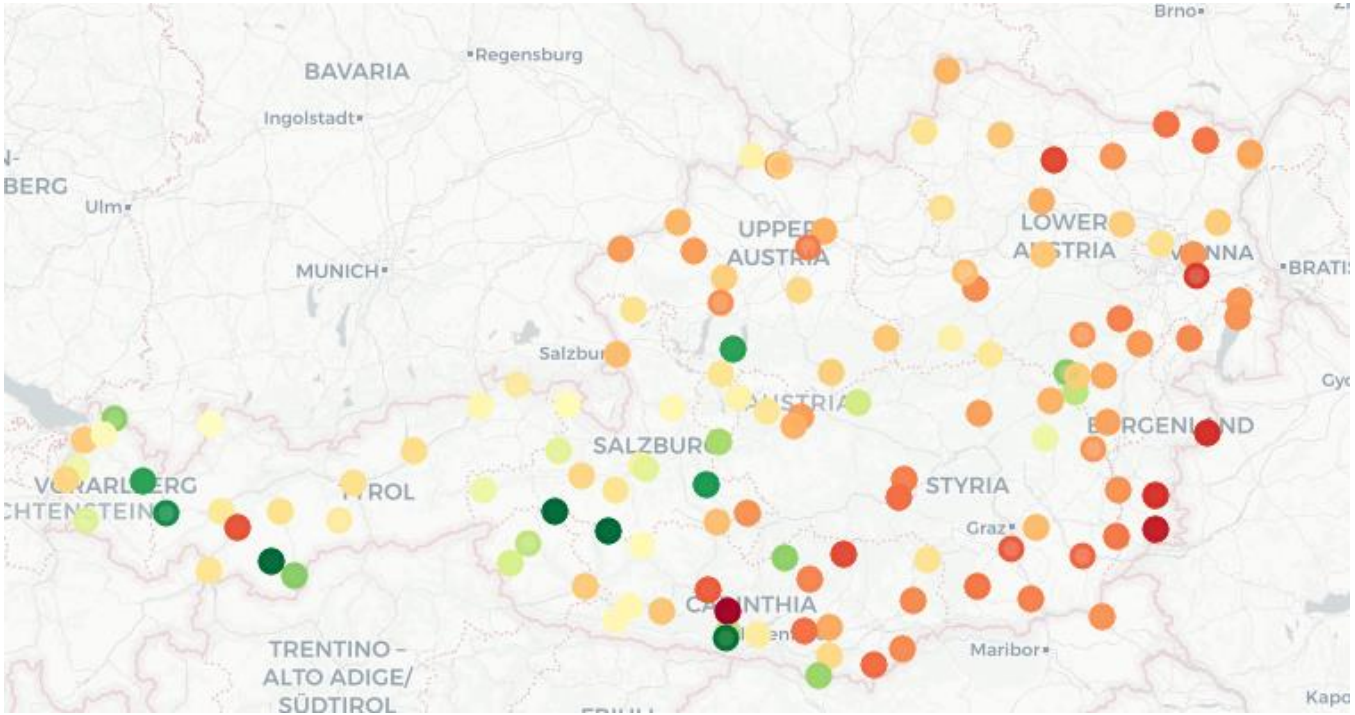


RQ4: Operational Window Optimization.

- Score = Sunshine - Wind Gust Freq - Frost Freq - Heat Freq



RQ4: Operational Window Optimization.



ClimAlytics AT

David Kalteis, Dominik Forsthuber, Michael Kerscher

BIG Data - Project

HAGENBERG | LINZ | STEYR | WELS

FH OBERÖSTERREICH
UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA