Data Quality Control

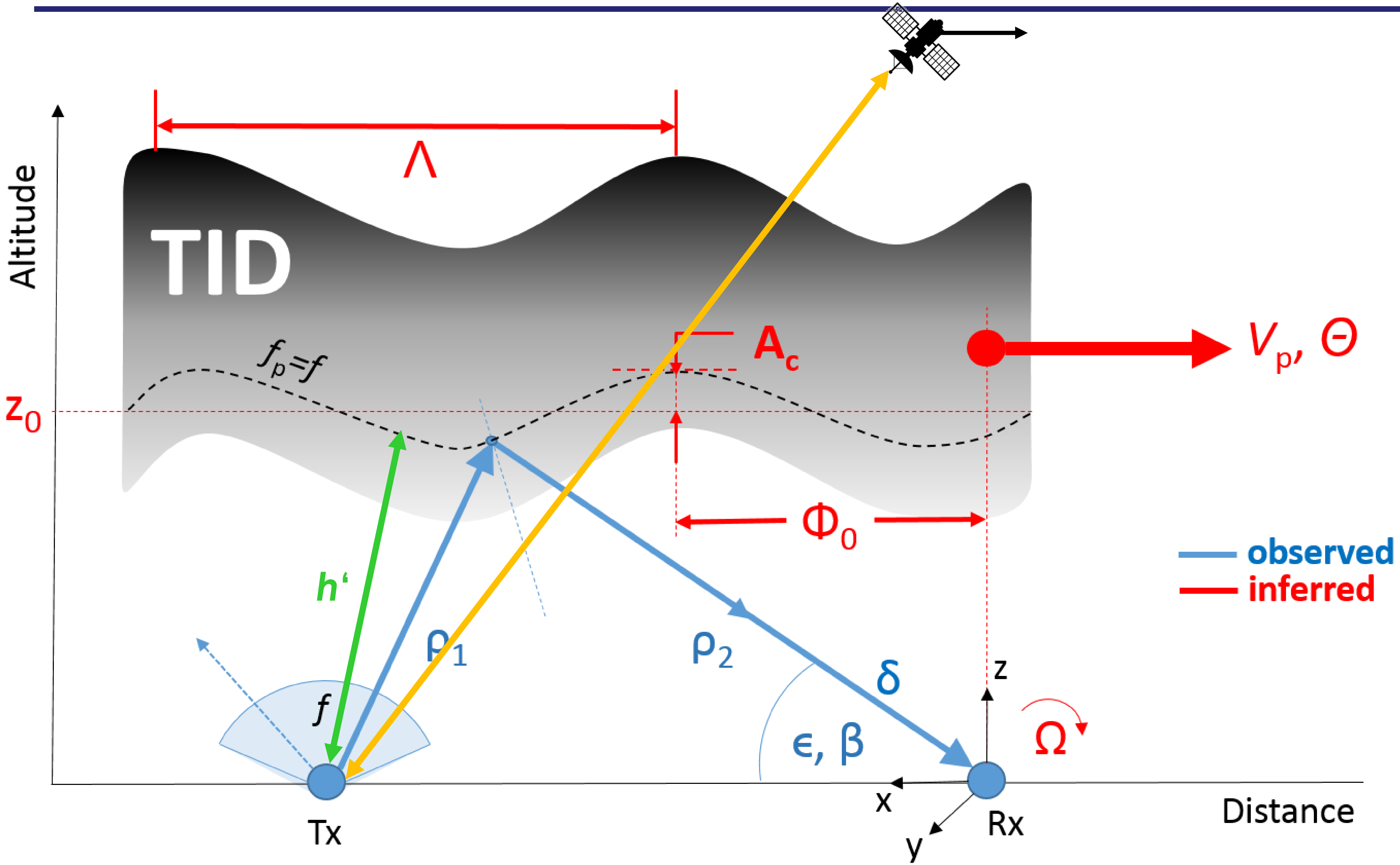
for robust TID detection

Ivan Galkin (BGD), on behalf of T-FORS and DISPEC teams

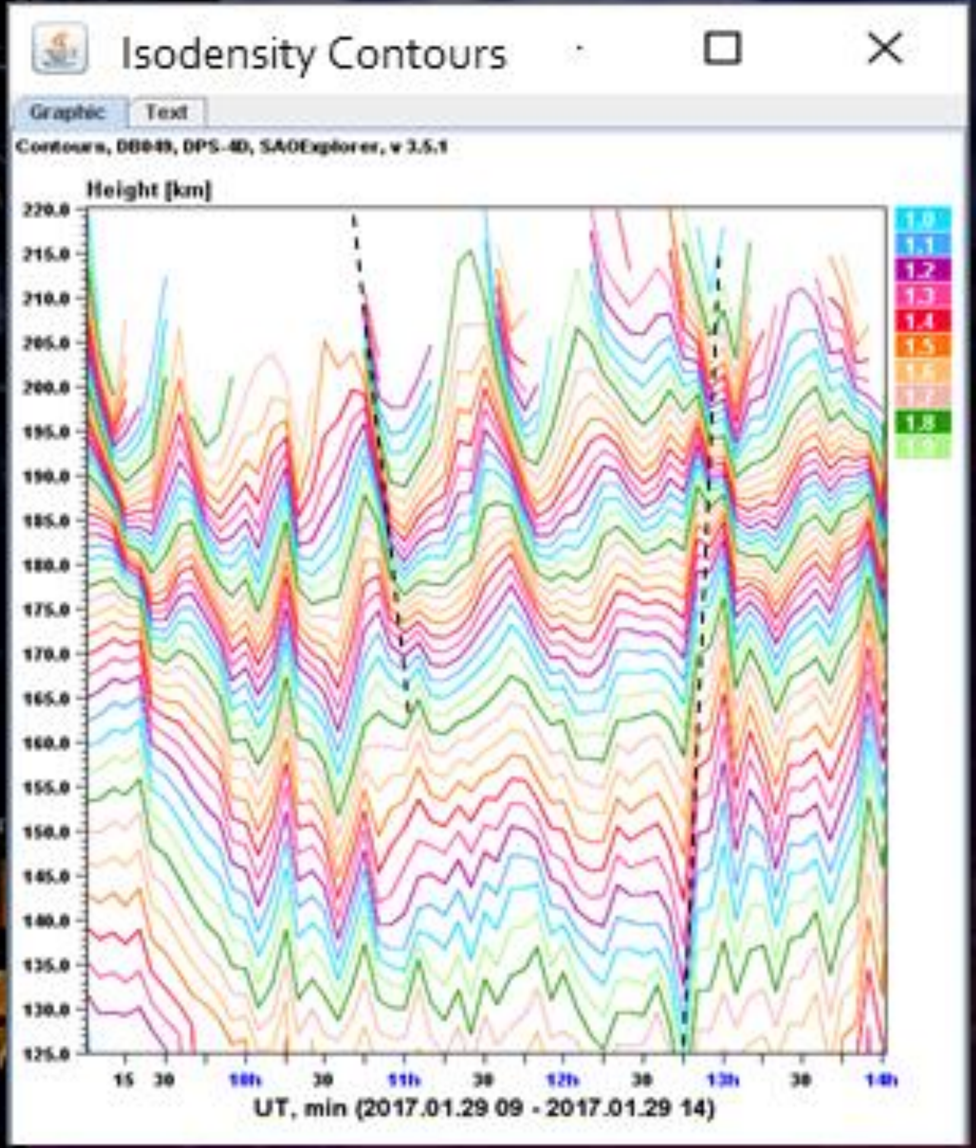
Outline

- Background: TID Detection
- TID Detection Resources and their QC systems
  - TID Catalogues (strictest QC)
  - Real-time ionosonde data for HF-INT technique (elaborate QC)
  - Real-time ionosonde data for HF-TID technique (failing QC requirements)
  - Real-time GNSS TID detection in VTEC time series and keograms (advanced QC)

TID is a density perturbation wave



HF versus other TID sensors

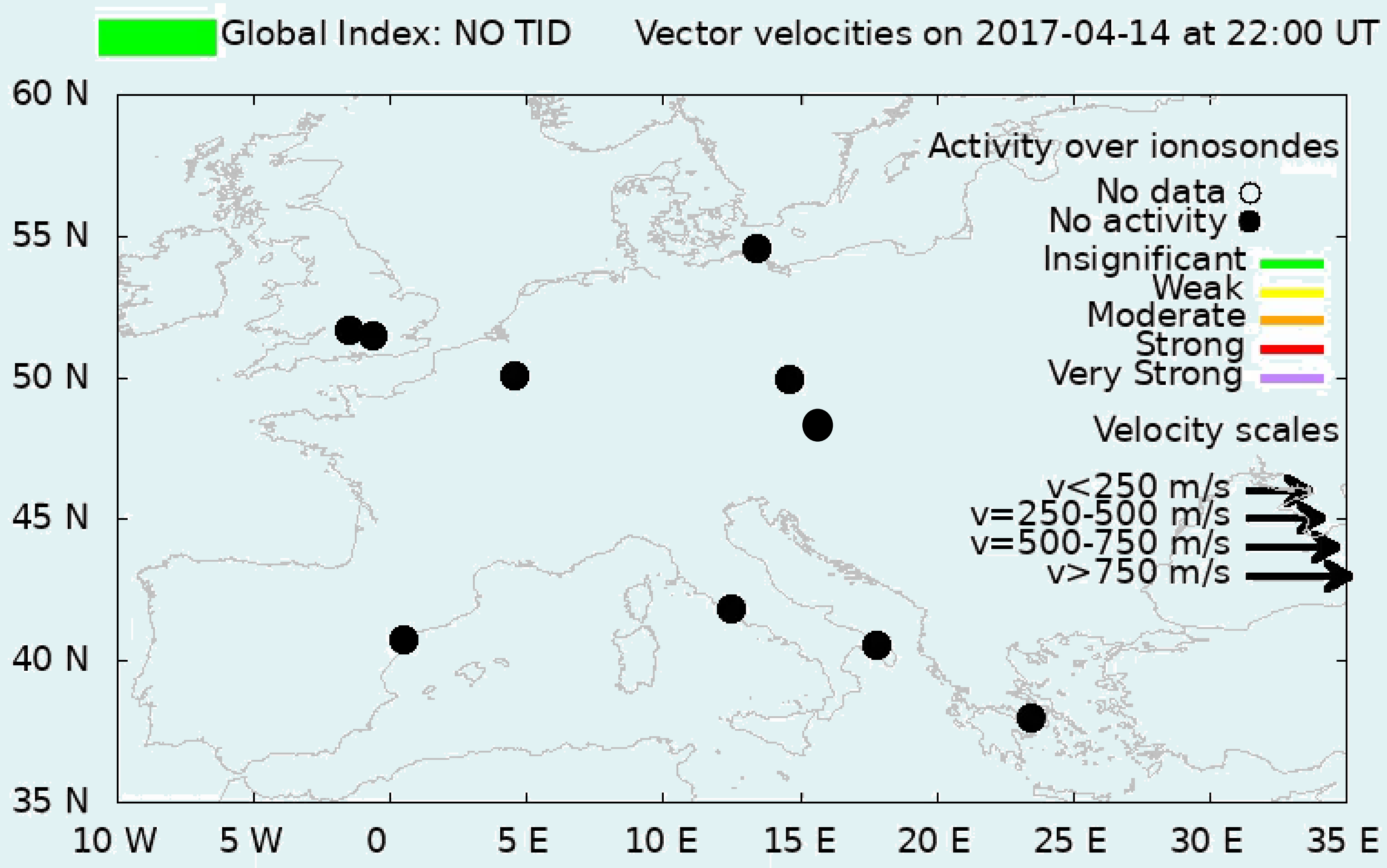


Data courtesy Tobias Verhulst, RMI

- 1D Altitude profile of TID
  - Detailed view of propagation along z-axis
  - Bottomside F-region only
- Sensitivity
  - Detection of a few % TID (and weaker)
  - “TID are always present” © Dima P.
- Direct measurement
  - Static platform
  - No slant-to-vertical transformation needed
- Automatic intelligent systems available
  - Replicate data human analysis

LSTID Catalogue

- **Visual inspection to determine LSTIDs events**
  - ➢ Looking for coherent velocity propagation
  - ➢ 1604 LSTIDs events detected and recorded above Europe between 01/2014 and 12/2022
- **Determination of onset time and duration**
  - ➢ Approximative
- **Average of the main characteristics of the TID for all stations and during the whole event.**



Single-site observer can detect TID, its amplitude, and period
But not propagation direction and velocity
Network of sensors is needed for triangulation

MSTID Catalogue

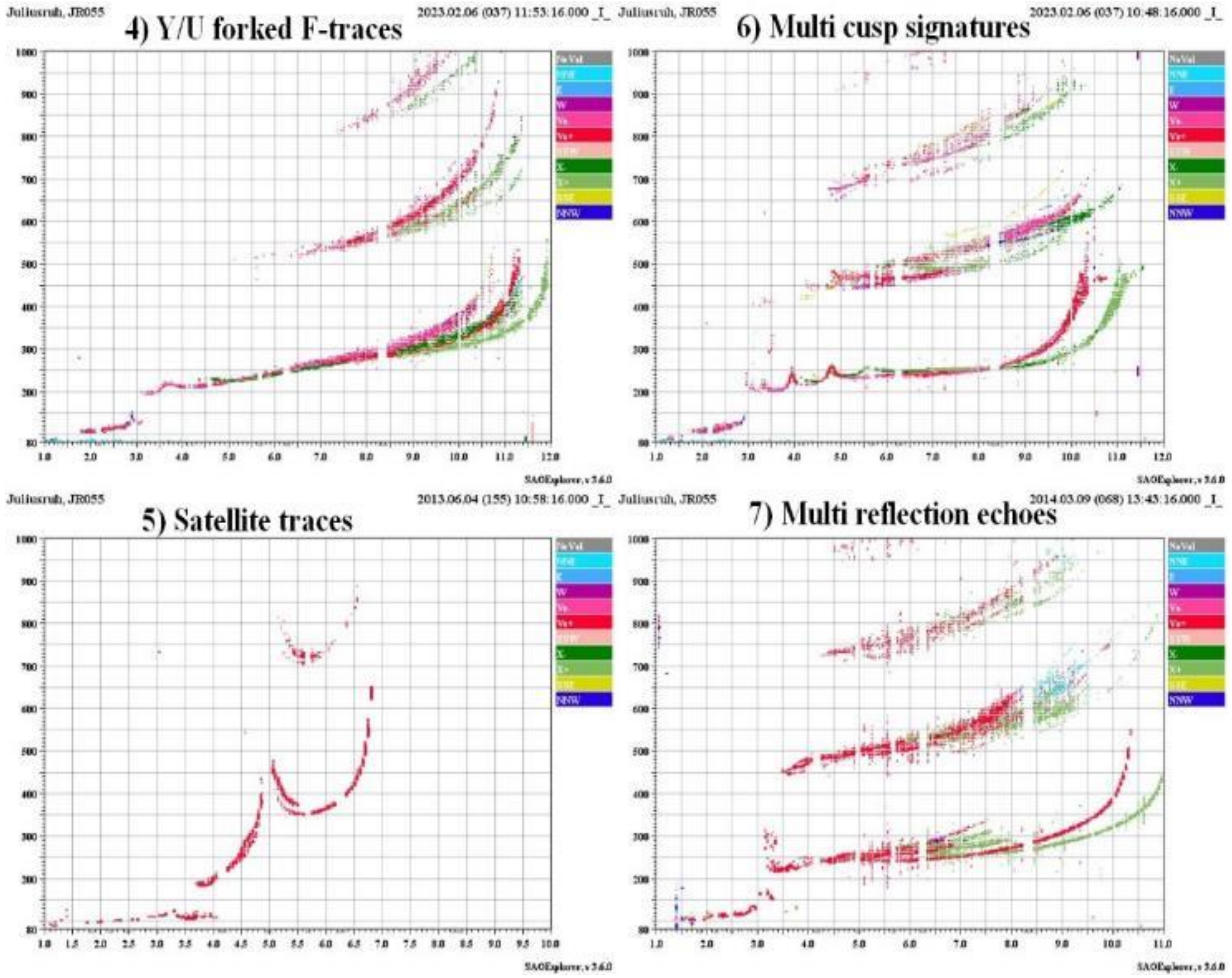
- **Visual inspection to determine MSTIDs events**

    - Looking for ionogram characteristics such as spread, multi-reflection echoes, satellite traces, Y/U fork structures, Multi-cusp signatures

    - We analyzed the data for the year 2014, 2016 and 2020 for the four seasons i.e. two equinoxes (March and September) and two solstices (June and December)

    - In total 24*366=8784 ionograms per station were analyzed manually. Total 10 stations were included.

    - Special high-cadence campaign in September 2024, data analysis in progress



- **Determination of onset time and duration**

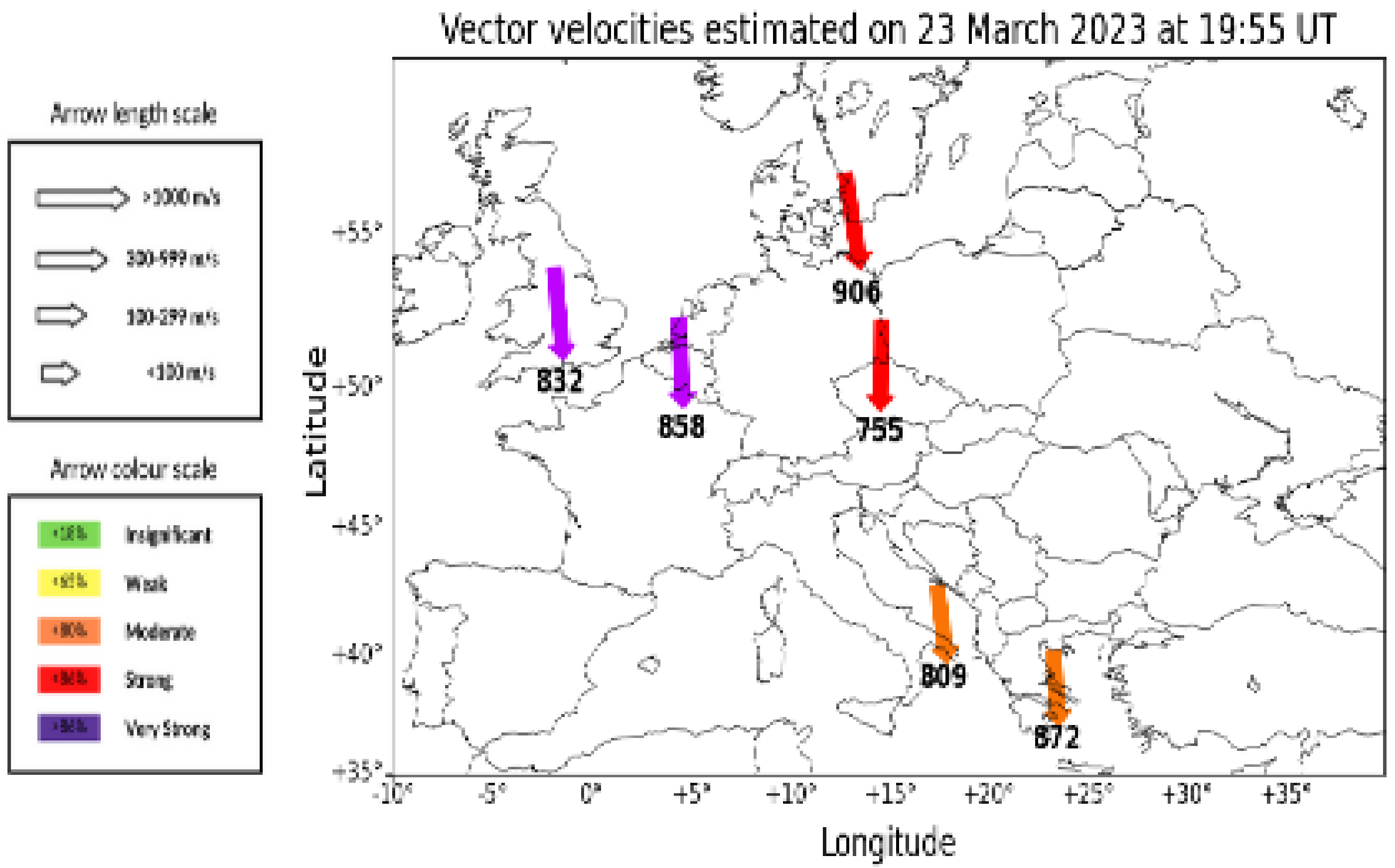
    - Approximative

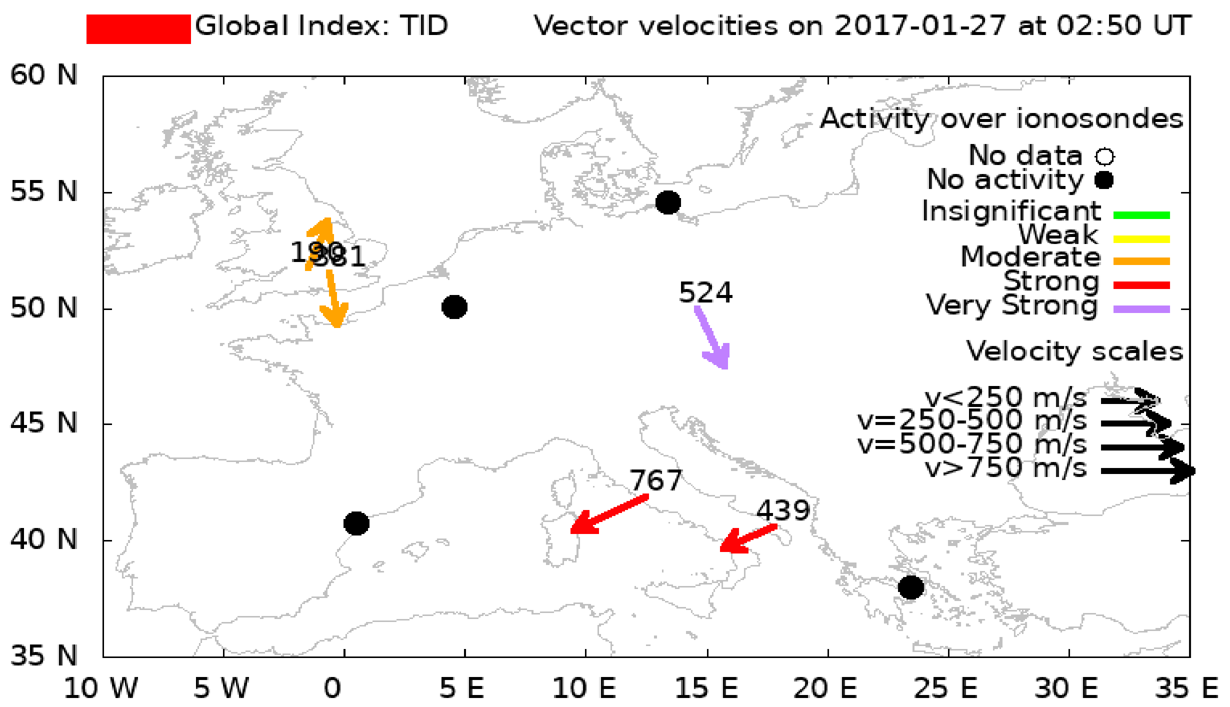
Real-time Data Quality Systems

1. **Ionosonde data for HF-INT technique (elaborate QC)**
2. **Ionosonde data for HF-TID technique (failing QC requirements)**
3. **GNSS TID detection in VTEC time series and keograms (advanced QC)**

1. HF-INT Quality Control System

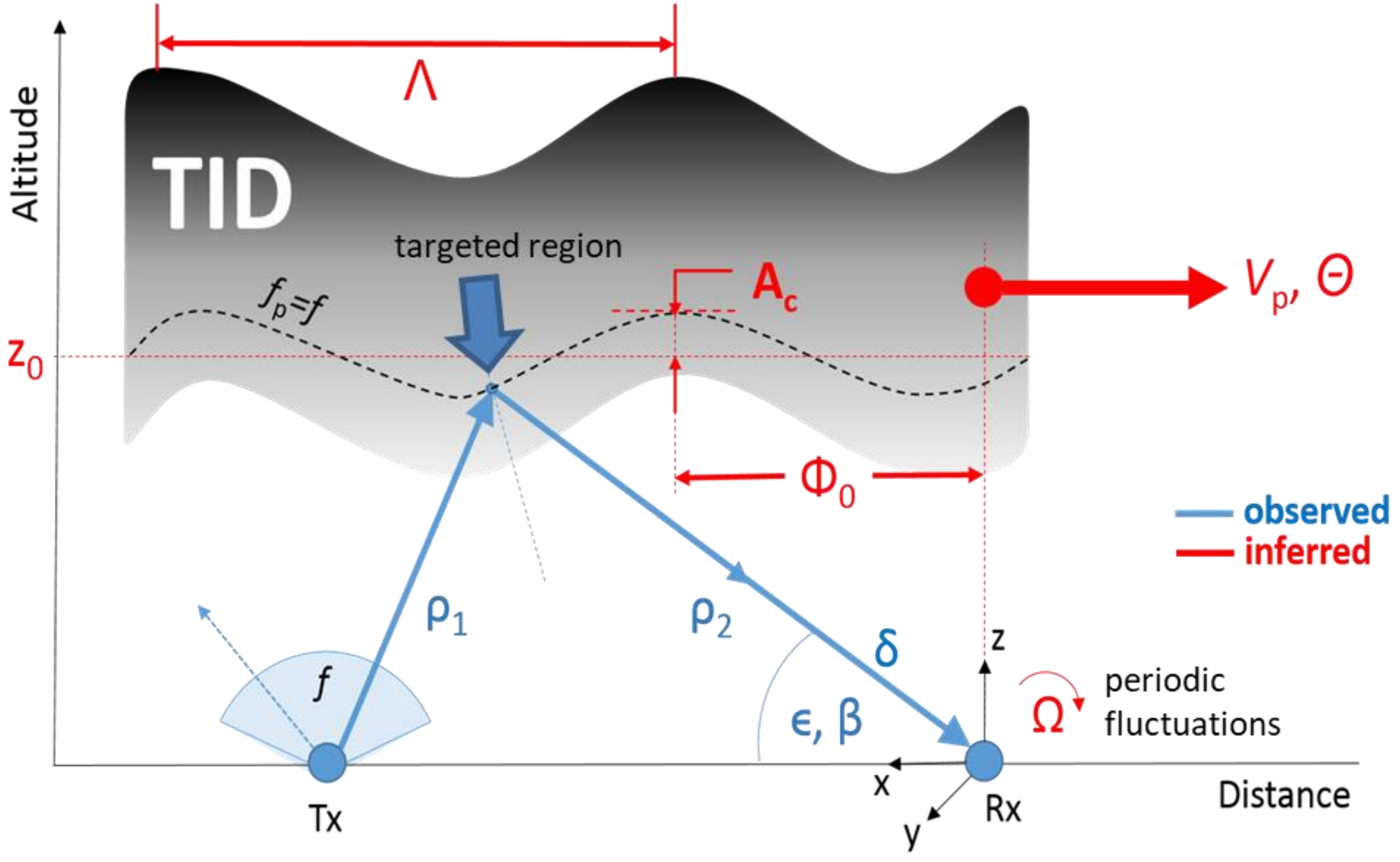
- Vertical sounding with triangulation
- Harmonization of multiple TID evaluations
  - At least 4 stations must agree
- Missing MUF(3000) values
  - Instrument issues
  - Es, D-region absorption
  - Low confidence
  - Poor connectivity to LGDC
- Boundary station effects
  - 180° uncertainty





2. HF-TID Technique

$$N(z_0,t;x,y) = N_{bg}(z_0,t;x,y)\left[1 + A_N \cos\left\{\Omega t - K\left[x\cos\Theta + y\sin\Theta\right] + \Phi_0\right\}\right]$$



$$\mathfrak{D}(\Omega_k) = \frac{2\pi}{N}\sum_{n=0}^{N-1}\delta_n \exp(-i\,\Omega_k n)$$ Signal's Doppler spectrum

$$\mathfrak{B}(\Omega_k) = \frac{2\pi}{N}\sum_{n=0}^{N-1}\beta_n \exp(i\Omega_k n)$$ Signal's azimuth spectrum

$$\mathfrak{E}(\Omega_k) = \frac{2\pi}{N}\sum_{n=0}^{N-1}\epsilon_n \exp(-i\Omega_k n)$$ Signal's elevation spectrum

$$A_N = Abs\left[\frac{i\lambda\,\mathfrak{D}(\Omega)}{2z_0\Omega sin\epsilon_0}\right] \qquad \Theta = Re[\mathfrak{H}]$$

$$\mathfrak{K} = \frac{\mathfrak{S}}{\mathfrak{D}(\Omega)}\frac{cos\epsilon_0}{z_0\lambda}$$

$$\mathfrak{H} = arctan(\mathfrak{C}_1, \mathfrak{C}_2)$$

$$\mathfrak{C}_1 = \frac{-i\lambda\,\mathfrak{D}(\Omega) + 2z_0\Omega\,\mathfrak{E}(\Omega)tan\epsilon_0}{\mathfrak{S}}$$

$$\mathfrak{S} = \sqrt{\mathfrak{T}_1 + \mathfrak{T}_2 + \mathfrak{T}_3}$$

$$\mathfrak{C}_2 = \frac{-2z_0\Omega\,\mathfrak{B}(\Omega)}{\mathfrak{S}}$$

$$\mathfrak{T}_1 = -\mathfrak{D}^2(\Omega)\lambda^2 sin^2\varepsilon_0$$
$$\mathfrak{T}_2 = -i4z_0\lambda\Omega sin\epsilon_0 tan\epsilon_0\,\mathfrak{D}(\Omega)\mathfrak{E}(\Omega)$$
$$\mathfrak{T}_3 = 4z_0^2\Omega^2\left[\mathfrak{E}^2(\Omega)tan^2\epsilon_0 + \mathfrak{B}^2(\Omega)\right]$$

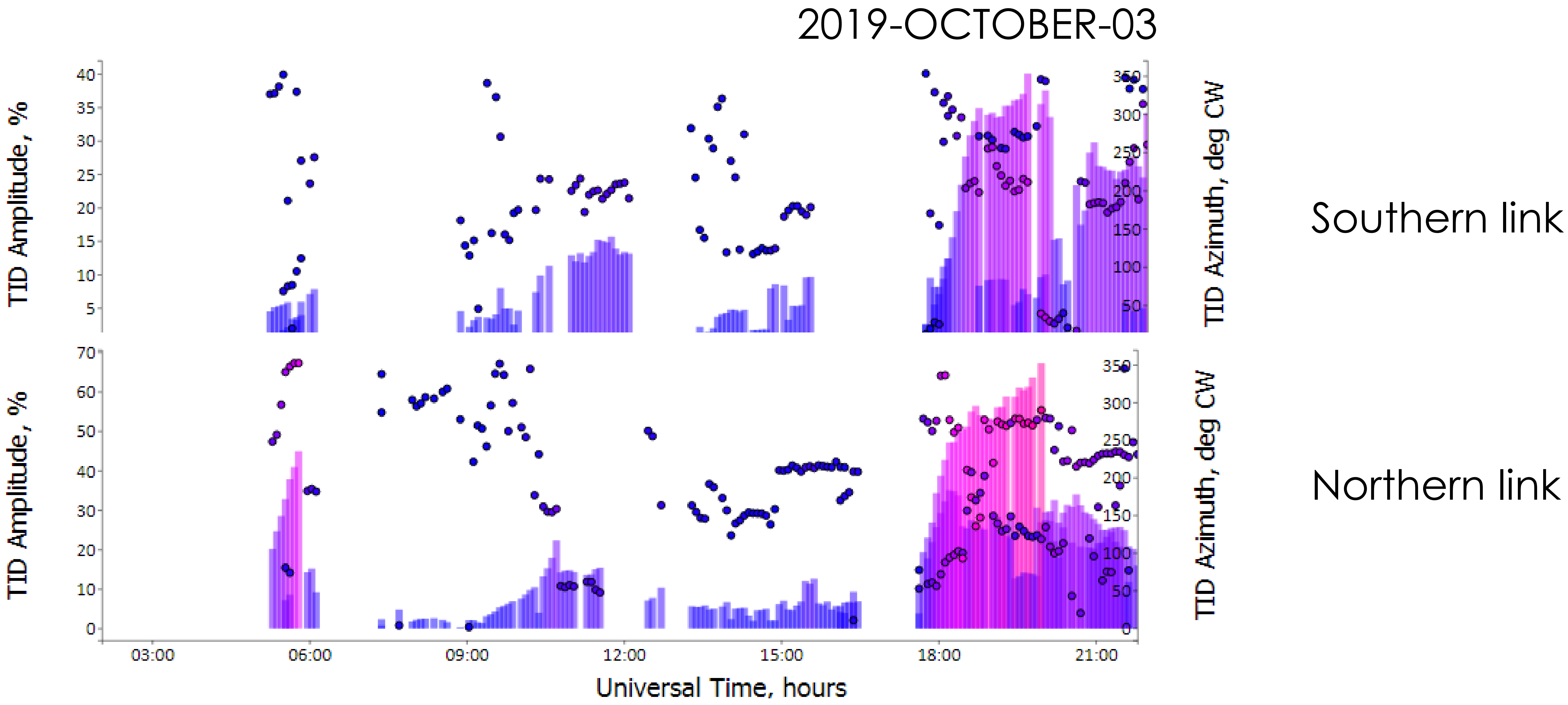
$$V_p = Abs\left[\frac{\Omega}{\mathfrak{K}}\right]$$

$$\mathfrak{D}(\Omega_k) = \frac{2\pi}{N}\sum_{n=0}^{N-1}\delta_n \exp(-i\,\Omega_k n)$$

$$\mathfrak{B}(\Omega_k) = \frac{2\pi}{N}\sum_{n=0}^{N-1}\beta_n \exp(i\Omega_k n)$$

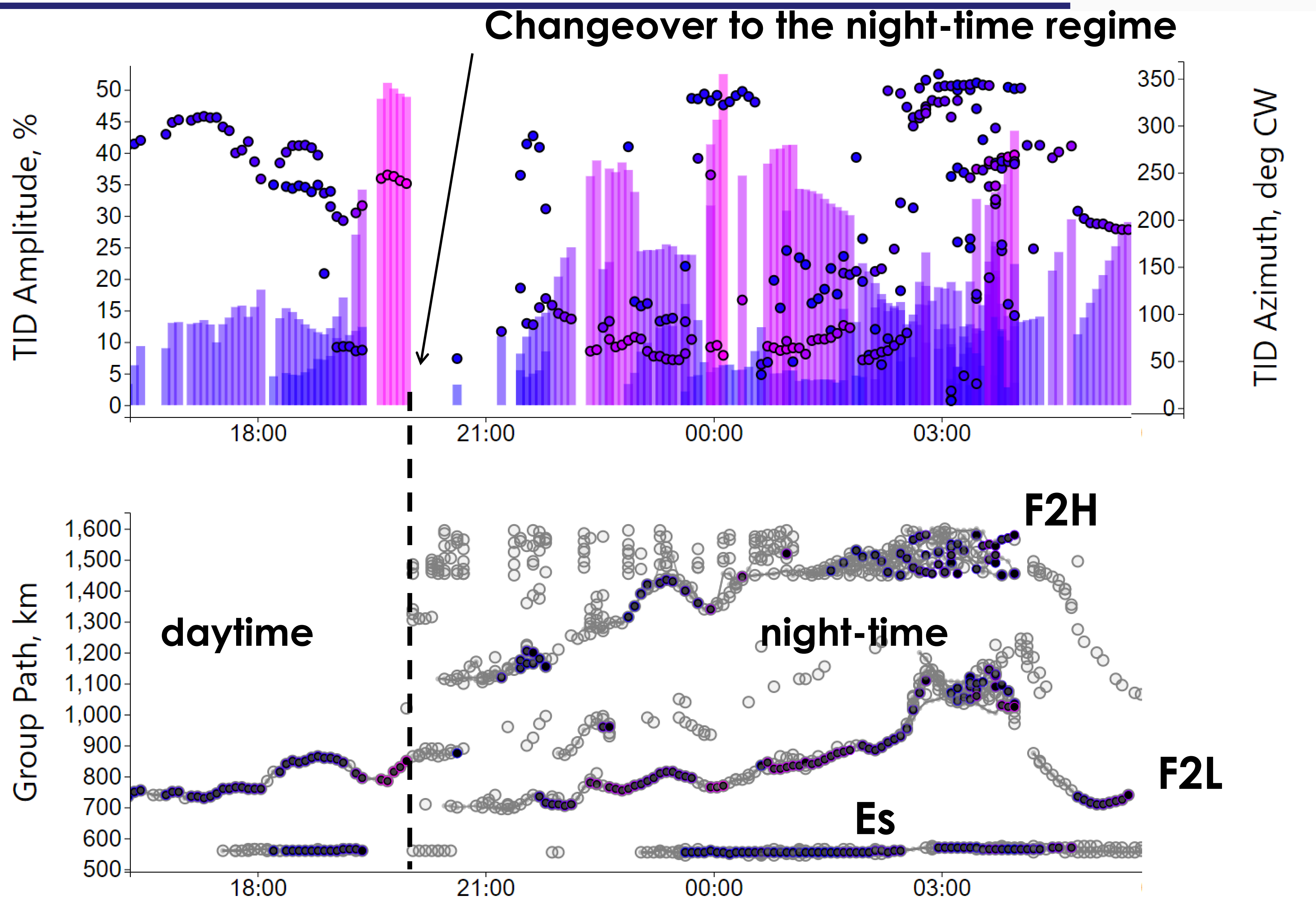
TID Detection is OK

TID characterization – not so much…



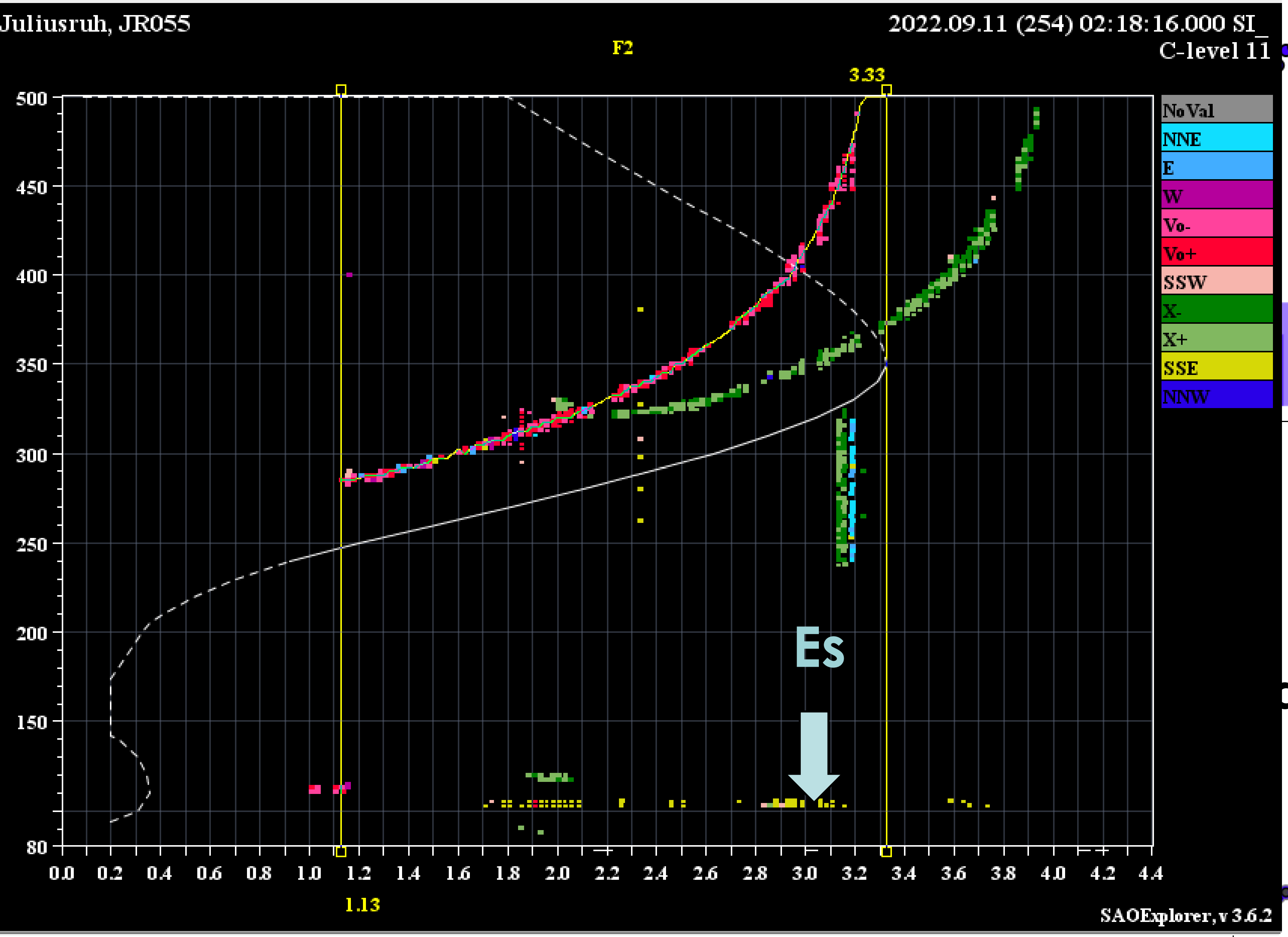
HF-TID Quality Issues (1)

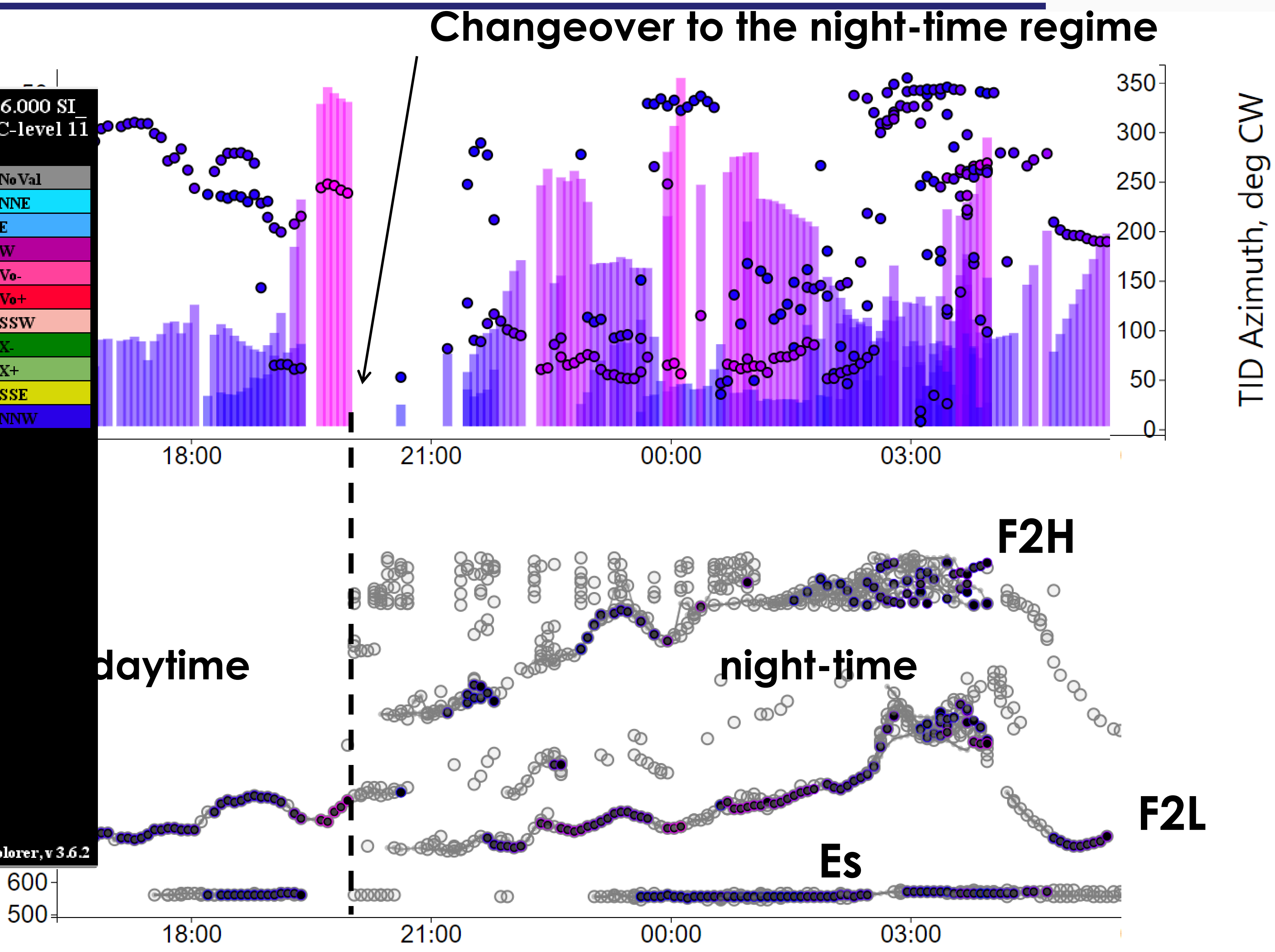
- Input signal tracks
  - ANNAE algorithm from ARTIST-5 used
  - Susceptible to signal jitter
  - Susceptible to multi-path propagation
  - More elaborate track identification is needed
    - Need F2L track only, closest to the mirror model
    - RayTRIX CQP is used to rule out E
- SW development is ongoing



Universal Time, September 11-12, 2022

HF-TID Quality Issues (1)





Universal Time, September 11-12, 2022

HF-TID Quality Issues (2)

- HF-TID method stiff requirements
  - Uses signal's amplitude and phase at Ω frequency (TID period)
    - Doppler frequency, zemith angle, and azimuth angle
      - Range is not used by this technique!
      - Total 6 values to derive all TID parameters
    - Azimuth variation is small, below 2°
    - High SNR is required, above 40 dB
      - Have to use high coherent integration time of 40+ sec
        - » Digisonde schedules are too busy already
  - Assumptions made:
    - Small perturbation model
    - Mirror reflection from ionosphere

GNSS Data Quality

- Good news: bias is not important
  - Detrending is done to high-pass TID variations
    - at the order of 300 seconds time scale for MSTIDs
  - But the TID-related variations are small
    - Amplitudes at the level of few tenths of TECU
    - Yet, one order of magnitude higher than the carrier phase measurement error
- Good news: the GNSS Cycle Slip is not an issue, either
  - recent progress in detecting and correcting these
  - should not be an error source into the processing pipeline

Thank you for your attention!

Funded by the European Union

The T-FORS project is funded by the European Union (GA-101081835). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Sprinting Review of D1.5 Ch#4

PART THREE. WHICH MEASURES WE DESIGNED FOR T-FORS

- Robustness of our models
  - accepts categorical features
  - handle missing values
  - provide automatic feature scaling
- Model Retraining (INGV)
  - not a simple retraining
  - handles covariate shift
    - adjust the decision function
    - look upstream at data quality
  - handles concept shift
    - trending…
- Digisonde data conditioning
  - Input features:
    - Autoscaling Confidence Score
    - HF-INT Quality Indicator
      - from TechTIDE project
      - weighted average of metrics describing data availability at sites
  - Output labels: DQF4 LSTID Catalog
- Performance monitoring (NOA)
  - drift detection and prevention
  - fine-tuning (retraining…)