SH23E-2613 - 16 Dec 2025, 14:15–17:45
Session: The Future of the International Sunspot Number

Reconstruction of the Sunspot Number Series : Gathering Data

Kalugodu Chandrashekhar[1], Laure Lefèvre[1], Benjamin Mampaey[1], Senthamizh Pavai Valliappan[1], Hisashi Hayakawa[2], Christian Ritter[3]
[1] Royal Observatory of Belgium (ROB), [2] Nagoya University, [3] Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract

The Sunspot Number (S_N) is one of the longest continuous records of solar activity. S_N V2.0 significantly improved the historical series through targeted recalibrations, but recent studies have shown that some scale changes—such as the 1849 transition—can only be fully understood by returning to the original raw observations. This motivates a complete reconstruction (S_N V3.0) in which uncertainties and calibration steps arise directly from the quality of each dataset rather than from global statistical corrections.

To enable this reconstruction, the historical data foundation has expanded dramatically in recent years. WDC–SILSO digitized the Mittheilungen (1610–1944) between 2017 and 2019, converting Wolf's original Zürich journals into machine-readable datasets with modern metadata and provenance. As part of FARSUN, the Zürich observation tables for 1945–1979 were digitized from 2023 to 2025, completing the recovery of the full Zürich legacy. In parallel, major community efforts (e.g., Hayakawa, Arlt, Vaquero, Carrasco) have located, transcribed, and published numerous additional early telescopic sunspot records since 2010.

Together, these datasets greatly expand the number of overlapping observations and allow independent cross-checks across observers and epochs. This richer observational base strengthens scale-transfer analyses, improves calibration robustness, and provides the foundation for an uncertainty-quantified reconstruction of the long-term sunspot series. These advances make a defensible S_N V3.0 possible and ensure that the Sunspot Number remains a reliable reference for solar activity studies.

Four-Axis Methodology

1. **Gather Sources**: track down dispersed observing logs (Wolf's *Mittheilungen*, Zürich tables, Gruithuisen manuscripts, Adams drawings) and capture provenance metadata at ingest.
2. **Process Data**: use handwriting/OCR engines (Transkribus models, custom scripts) alongside manual double-keying for fragile entries.
3. **Validate & Standardize**: reconcile aliases, observer IDs, and calendars; model group/spot counting behaviour to derive calibration curves.
4. **Disseminate**: publish machine-readable products, APIs, and VO-ready services with versioned uncertainty bundles for each release.

FARSuN Historical Sunspot Database

A sustainable, interoperable reference database ensuring that four centuries of solar observations remain accessible and scientifically reliable for the next generation of Sun-climate research.

- Legacy sources (Mitteilungen, Wolf Institute, Shreya, Stephen, etc.) have been integrated into the current schema for cross-validation and QC.
- Automated quality checks detect inconsistencies (e.g., NG > NS, duplicates, missing metadata).
- Metadata harmonization for observers, instruments, and observatories is underway.

Next Steps

- Finalize canonical schema and close remaining legacy mappings.
- Expand automated QC dashboards and uncertainty modeling.
- Scale up digitization and extraction (Adams, Gruithuisen, Stark, etc.) using HTR and citizen-science workflows.
- Expose the full dataset through the EPN-TAP service, ensuring FAIR data availability.
- Use this harmonized database as the foundation for S_N V3.0 calibration and long-term solar activity studies.

Methods

- Dual-mode transcription: tailored Transkribus handwriting models plus manual keying for complex tables.
- Data hygiene: observer alias reconciliation, calendar conversions, and spot/group recount audits.
- Uncertainty modelling: propagate observer calibration curves and note spotless-day confidence flags.
- Automation: reproducible Python pipelines with checkpoints for citizen-science ingestion and review.

Quality Gates

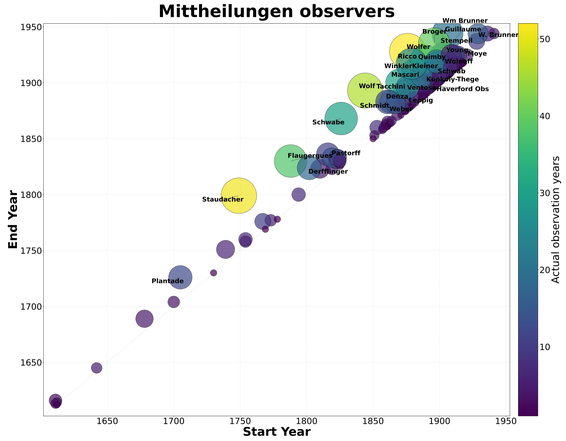
- Two-person verification for Adams and Gruithuisen counts before publishing derived indices.
- Cross-compare FARSuN transcriptions with Wolf's *Mittheilungen* and modern SILSO indices for drift detection.
- Metadata harmonisation ensures unique observer IDs, instrument notes, and weather context survive downstream.
- Git-tracked pipelines and notebooks capture every transformation for audit and reuse.

Key Datasets

Mittheilungen Dataset (1610–1945)

The Mittheilungen journals, compiled by Rudolf Wolf and successors at the Zürich Observatory, form a central source of historical sunspot observations. They document daily group and spot counts from 1848 to 1945 and incorporate earlier material back to 1610 through Wolf's compilations. Within the FARSuN project, the complete series was digitized and structured at the Royal Observatory of Belgium, integrated into the historical sunspot database, and is undergoing continued quality checks, metadata harmonization, and cross-comparisons with newly recovered records. This digitized dataset now provides four centuries of observer statistics and constitutes a key component of the historical foundation for the ongoing reconstruction of S_N V3.0.

- Refining data consistency and observer calibration (Wolf–Wolfer overlap).
- Finalizing metadata alignment for FAIR publication.
- Resolving remaining ID and provenance issues across merged databases.



Zürich Tables (1945–1979)

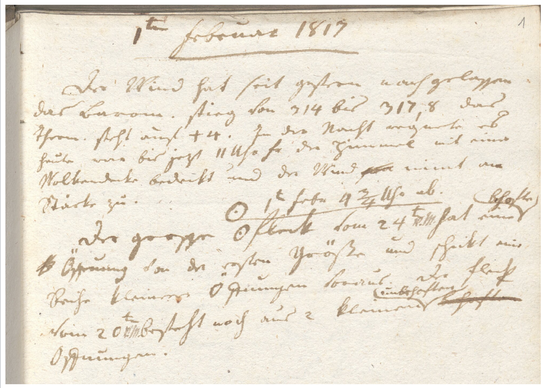
The Zürich Tables form the foundational record of daily sunspot number production at the Zürich Observatory between 1945 and 1979, detailing group and spot counts, the adopted Wolf number, and observing context. *Status*: 95% of the 2 000+ yearly tables are already digitized and the remaining late-1970s observer metadata are being encoded.

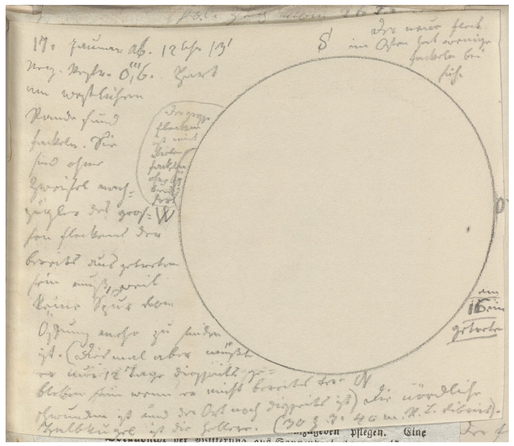
- Fully digitized for 1945–1979 with detailed metadata completed for 1945–1959.
- Quality control and observer metadata harmonisation underway to maintain traceability.

Gruithuisen Data (1817–1849)

The Gruithuisen manuscripts (1817–1849) capture daily sunspot notes, drawings, and weather context in German Kurrent script, supplying early-19th-century coverage vital for S_N V3.0 refinement. The entire image set is digitized and extraction via bespoke Transkribus-style models plus dual QC for NG/NS tallies is underway.

- Image sets and text snippets fully collected with preliminary CSV/XLSX extraction completed.
- Two-pass QC (survey + counting) with verified sunspot counts in progress.





C. H. Adams Data (1819–1823)

FARSuN has recovered 338 full-disk drawings and 1,056 daily entries (308 spot days, 748 spotless) from C. H. Adams (1819–1823), providing dense coverage with explicit spotless days. All spotless-day tallies are cross-checked against Wolf's compilations, anchoring early-1800s calibration experiments.

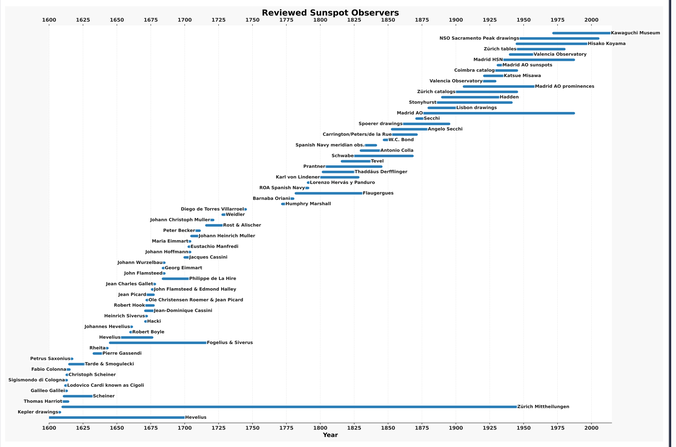
- Quality checks to align direct and reflection methods for consistent calibration.
- Orientation, and standardisation of drawings to enable future positional and area extractions.

Key Datasets

Augustin Stark Data (1813–1835)

The Stark dataset (1813–1835), comprising printed German descriptions of sunspot groups and chains, has been partially retrieved. Preliminary OCR and extraction tests show good potential, and ongoing work focuses on refining consistency, QC flags, and metadata integration into the FARSuN database. Trial extraction campaigns continue so that descriptive passages are tagged consistently before bulk ingestion.

Data Recovery Efforts Since 2010



A non-exhaustive set of early telescopic sunspot datasets recovered and published since 2010, showing their temporal coverage from the early 1600s to the modern era. These community efforts—including major contributions by Hayakawa, Arlt, Vaquero, Carrasco, and others—have greatly expanded the number of available raw observations, enabling independent cross-checks of overlapping observers.

Progress Highlights

- Citizen-science interface blueprint prepared for the upcoming transcription portal.
- Zürich pipeline combines Transkribus extraction with manual review to reach the 2026 target.
- Gruithuisen workflow couples Kurrent experts with validation scripts to catch missing symbols.
- Adams drawings cross-checked with *Mittheilungen* records to confirm spotless-day agreement.

Challenges

- Deciphering *Kurrent* manuscripts and annotating non-solar notes without losing context.
- Harmonising heterogeneous formats (drawings, tabular ledgers, weather marginalia).
- Quantifying observer bias and spotless-day reporting habits over centuries.
- Maintaining end-to-end provenance from scanned folio to released data product.

From Raw to S_N V3.0

1. Curate daily counts with provenance-rich metadata (observer, instrument, weather context).
2. Solve observer response functions to align group/spot scales across centuries.
3. Aggregate to monthly, yearly, and hemispheric indices with propagated uncertainties.
4. Benchmark against S_N V2.0 and satellite-era proxies to validate drift corrections.

Space Climate Impact

Citizen science
Interactive portal mobilises volunteers to transcribe *Kurrent* manuscripts and sustain STEM outreach.

Forecasting
Bias-free S_N V3.0 improves solar cycle prediction skill and operational space weather baselines.

Climate links
Validated historical forcing refines long-term climate models and Earth radiation budget studies.

Open science
FAIR-compliant releases with APIs and uncertainty bundles accelerate community reuse.

Poster QR Code

For more information on FARSuN project, datasets, and a digital copy of the poster, scan the QR code.

Contact: sidc-support@oma.be

Data Portal: https://www.sidc.be/silso/