Refining the Sunspot Number Series to SN V3

Challenges and Benefits for the Space Climate Community

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 International Sunspot Number (SN) remains affected by historical scale jumps such as the *Waldmeier break*, even after two major recalibrations. FARSuN (Findability and Accessibility of historical Raw Sunspot Numbers) tackles this by rebuilding the record from the ground up: collecting handwritten logs and tables from 1607–1980, digitizing them, and applying consistent metadata and validation.

Why Reconstruction?

The SN inherits observer-dependent scale changes (for example the 1947 "Waldmeier jump") that cannot be removed by tweaking coefficients alone. FARSuN pioneers a full reconstruction that replays the entire observing chain.

- **Recalibration** only rescales an existing index; legacy gaps, missing metadata, and hidden assumptions stay baked in.
- **Reconstruction** retrieves the primary artefacts (drawings, tabular logs, weather notes) and recomputes the index with transparent uncertainty tracking.
- Direct access to the raw sources lets us audit outliers, model observer response curves, and issue reproducible releases such as SN V3.

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 SN V3 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

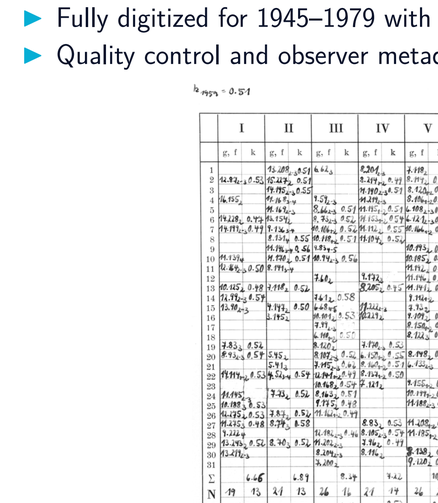
- Two-person verification for Adams and Gruithuisen counts before publishing derived indices.
- Cross-compare FARSuN transcriptions with Wolf's *Mittheilungen* and modern SILSO indices for drift detection.
- Metadata harmonisation ensures unique observer IDs, instrument notes, and weather context survive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

Key Datasets

Mittheilungen Dataset

The Mittheilungen journals, compiled by Rudolf Wolf and successors at the Zürich Observatory, form the backbone of historical sunspot observations, recording daily group and spot counts from 1848 to 1945 and extending back to 1610 through Wolf's compilations.

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



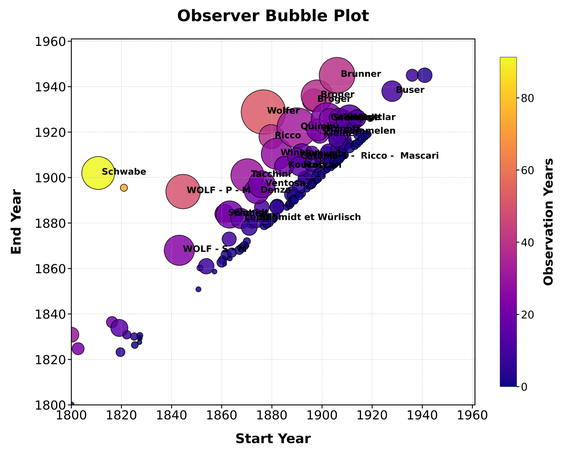
625) Alfred Wolfer, Beobachtungen der Sonnenflecken auf der Sternwarte in Zürich im Jahre 1890. (Fortsetzung zu 604.)

1890	1890	1890	1890	1890
I 1 1.1	II 14 1.1	III 1	5 1.2	V 10 1.11
– 2 1.1	– 16 0.0*	– 18 0.0	– 16 1.3	– 11 2.11
– 4 1.1	– 20 0.0	– 19 0.0	– 17 0.0	– 12 2.13
– 5 1.1	– 21 0.0	– 21 0.0	– 18 0.0	– 14 0.0
– 6 2.5	– 22 0.0	– 22 0.0	– 19 0.0	– 15 0.0
– 18 0.0	– 25 0.0	– 23 1.3	– 20 0.0*	– 16 0.0
– 19 1.3*	– 26 0.0	– 24 0.0	– 21 0.0	– 17 3.11
– 20 1.3*	– 27 0.0	– 26 0.0	– 22 0.0	– 18 2.11
– 24 0.0	– 28 1.1	– 27 0.0	– 23 0.0	– 19 2.6
– 25 0.0	III 1 1.1	– 28 0.0	– 24 0.0	– 20 3.6
– 26 0.0	– 2 0.0	– 29 0.0	– 25 1.1	– 22 2.5
– 27 0.0	– 3 1.1	– 30 0.0	– 26 0.0	– 23 1.1
– 28 0.0	– 4 1.6	– 31 0.0	– 27 0.0	– 24 1.1
– 29 0.0	– 5 1.6	IV 1 0.0	– 28 1.3	– 25 0.0
– 30 1.2	– 7 1.5	– 2 0.0	– 29 1.7	– 26 1.10
– 31 1.6	– 8 1.10	– 4 0.0*	– 30 1.11	– 27 0.0
II 1 1.3	– 9 1.10	– 5 0.0*	V 1 1.1	– 29 0.0
– 2 0.0*	– 10 1.16	– 6 0.0	– 2 0.0	– 30 0.0
– 3 0.0	– 11 1.11	– 7 0.0	– 3 0.0	– 31 0.0
– 4 0.0	– 12 1.6	– 9 0.0*	– 4 0.0	VI 1 0.0
– 5 0.0	– 13 1.3	– 10 0.0	– 5 0.0	– 2 0.0
– 10 0.0	– 14 1.3	– 12 2.10	– 6 0.0	– 3 0.0
– 11 0.0	– 15 1.1	– 13 2.8	– 7 0.0	– 4 0.0
– 12 0.0	– 16 0.0	– 14 1.1	– 9 1.2	– 5 1.5

NB. Die mit * bezeichneten Beobachtungen sind mit einem kleinern Fernrohr gemacht, welchem etwa der Factor 1,5 zukommt.

April 1891. **

(a) Typical yearly table(in the Mittheilungen)



(b) Observer coverage and overlaps

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.

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

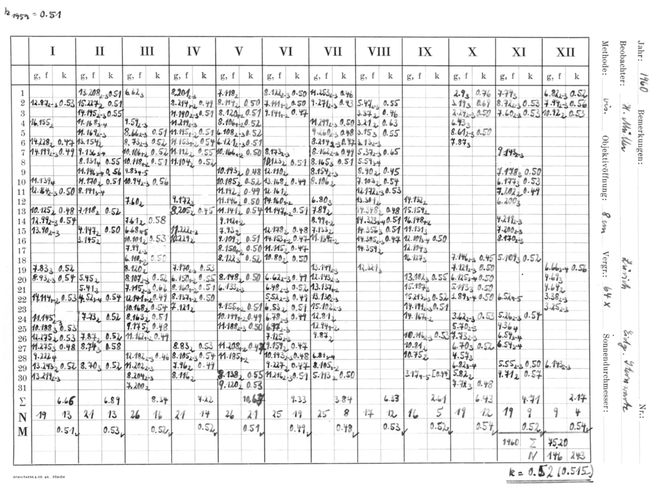
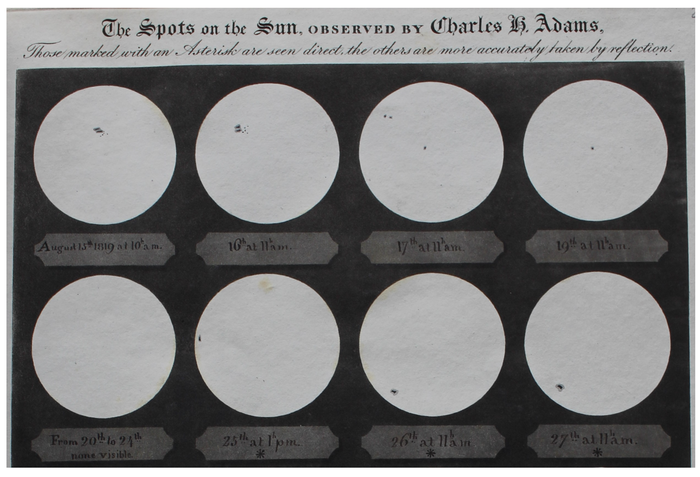


Figure: Facsimile of a typical handwritten yearly table

C. H. Adams Data

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.

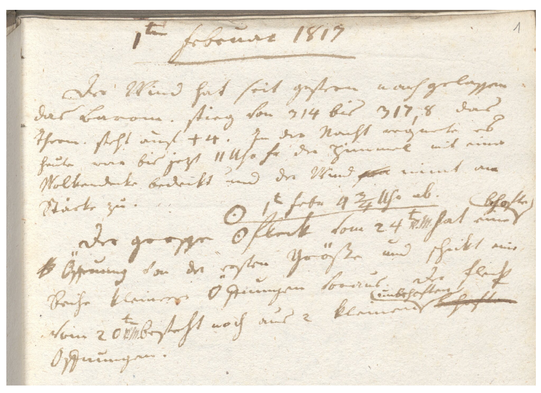
- Quality checks to align direct and reflection methods for consistent calibration.
- Orientation, and standardisation of drawings to enable future positional and area extractions.

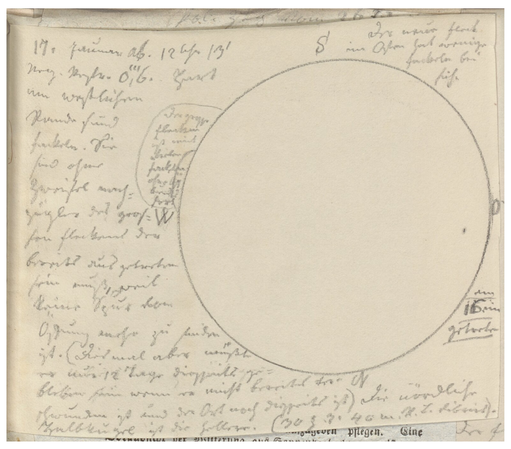


Gruithuisen Data

The Gruithuisen manuscripts (1817–1849) capture daily sunspot notes, drawings, and weather context in German Kurrent script, supplying early-19th-century coverage vital for SN V3 refinement.

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





Key Datasets

Augustin Stark Data

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.

Summary

Source	Period	Current Focus
Mittheilungen	1610–1945	Fully digitized and integrated; observer statistics produced from 400+ years of entries.
Zürich Tables	1945–1979	95% of 2 000+ tables digitized; metadata encoding for late-1970s observers underway.
Gruithuisen	1817–1849	100% digitized; Extraction using transkribus like models + dual QC for group/spot counts underway.
C. H. Adams	1819–1823	338 sketches extracted (1 056 entries); spotless days cross-checked against Wolf series.
Augustin Stark	1813–1835	Trial extraction are underway.

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.

From Raw to SN V3

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 SN V2 and satellite-era proxies to validate drift corrections.

Challenges

- Decode *Kurrent* manuscripts while preserving marginal notes.
- Harmonise drawings, ledgers, and weather annotations.
- Quantify observer bias and spotless-day reporting habits.
- Maintain provenance from scanned folio to released product.

Space Climate Impact

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

Forecasting
Bias-free SN V3 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, Different datasets, softcopy of the poster, scan this QR code.

Contact : sidc-support@oma.be

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

Version: Late 2025 snapshot