The Brussels period (1981 – 2014)
Towards a full recalculation

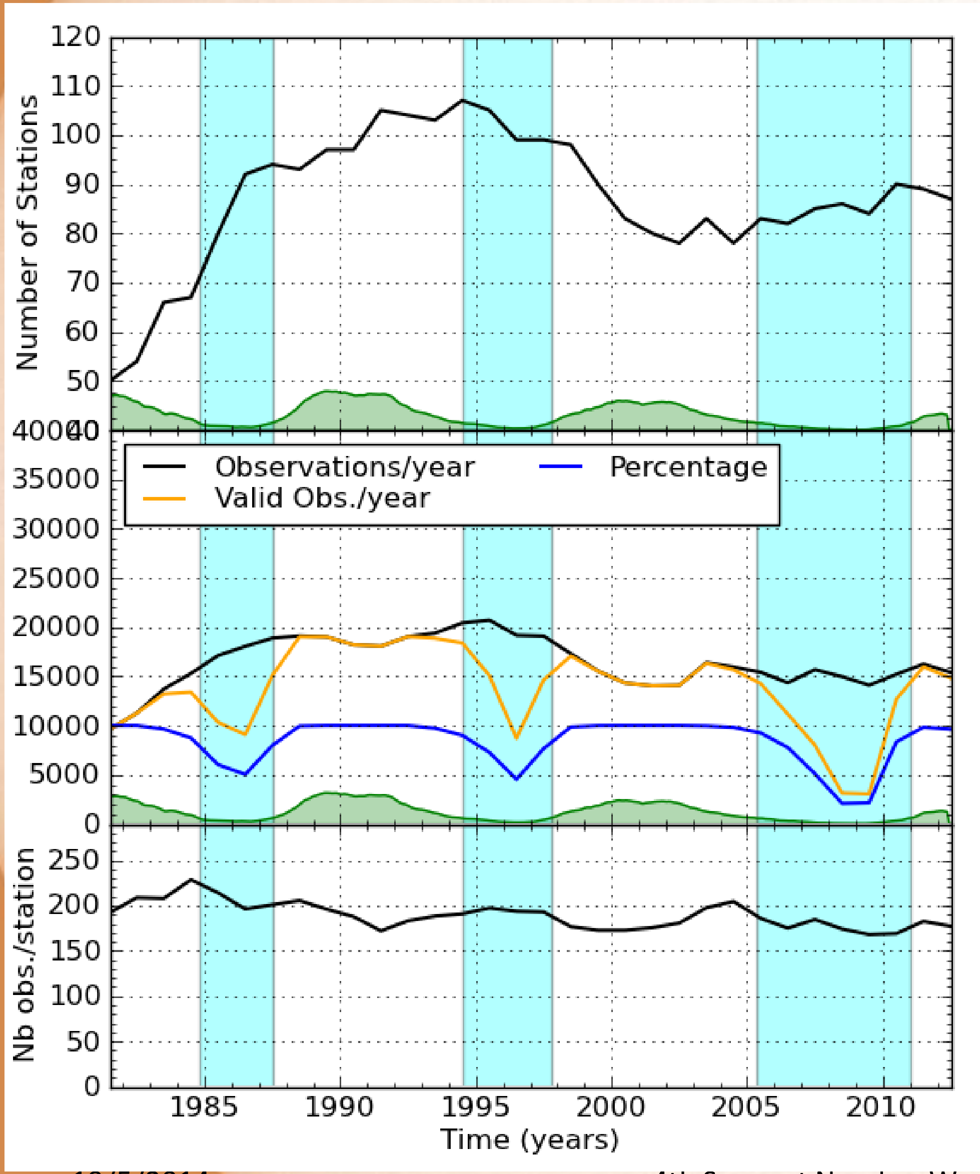
Frédéric Clette

World Data Center – SILSO
Sunspot Index and Long-term Solar Observations
Royal Observatory of Belgium

Outline

- The SILSO data archive
- Computation method
- Network-wide statistics: Locarno drift diagnostic
- Study of personal k coefficient of Specola observers
- Impact on relations with other indices
- Full recalculation of the Brussels SN: validation
- Results from alternate reference stations
- Conclusions : options for a new reference

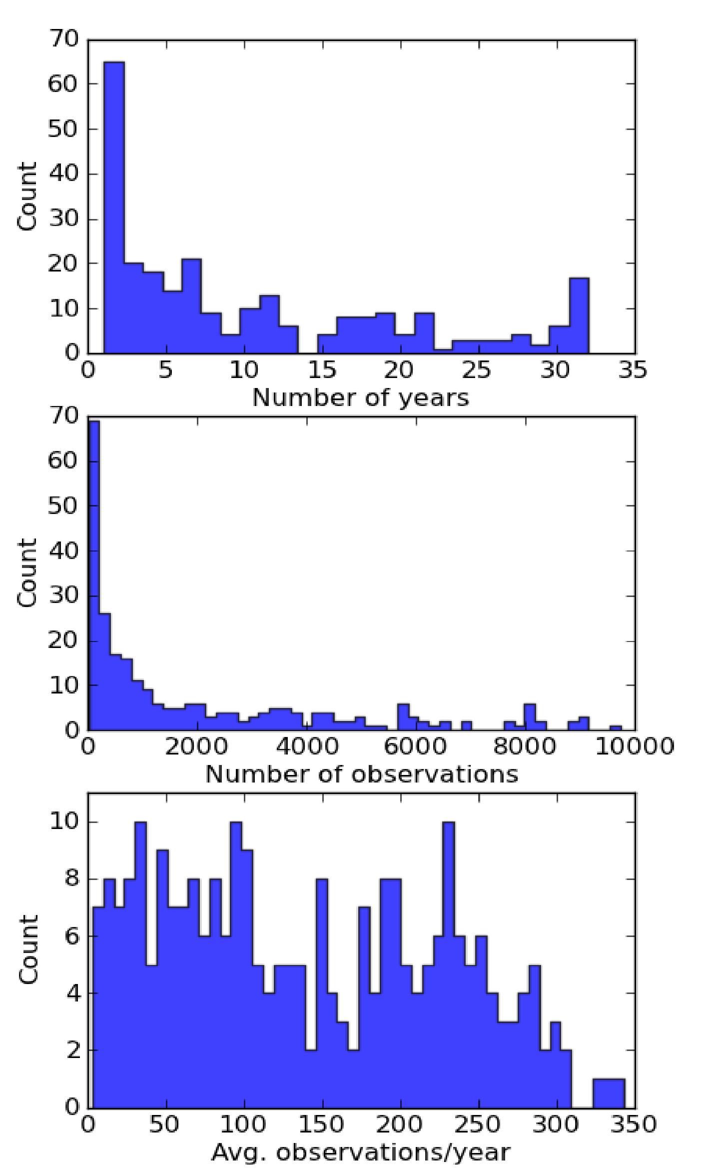
The SILSO data archive



Statistics over the last 34 years

- 269 contributing stations
- 30 countries
- At any time,
  ~ 80 - 100 stations
  - Initial recruiting
  - Stabilisation at 85-90
- ~180 observations/year/obs.
- 541,000 observations

The SILSO data archive



- 80 stations observed for more than 15 years
- 15 stations for more than 30 years
- Average number of observation per year:
  - Average 180 (~50%)
  - Peak at 230 obs/year (~66%):
    - Typical maximum weather-limited rate
  - A few stations close to 365 days/year (100%)
    - E.g. Egypt, Saudi Arabia

Processing method

SSN calculation method

- Scaling: average monthly k relative to pilot station:

$$k_s = \sum_d W_{LO}(d) / W_s(d)$$

- Daily SSN: statistics over whole network: $R_i = \sum_s k_s \cdot W_s$
- Elimination of anomalous daily Locarno values

- Daily scale:
  - Statistical cleaning of the data
  - ⇨ Reduced daily rms error
- Monthly scale and longer:
  - No use of past k coefficients (no network « memory »)
  - ⇨ R_i scaled to the monthly average Wolf Number from Locarno

Data elimination: outliers at 2 σ_k
Monthly mean station k_s:
first approximation (k_s, σ_k)

↓

Daily average SSN R:
first approximation (R, σ_R)
W_{LO} data elimination at 1 σ_R

↓

Data elimination: outliers at 1 σ_k
Monthly mean k: final (k_s, σ_k)

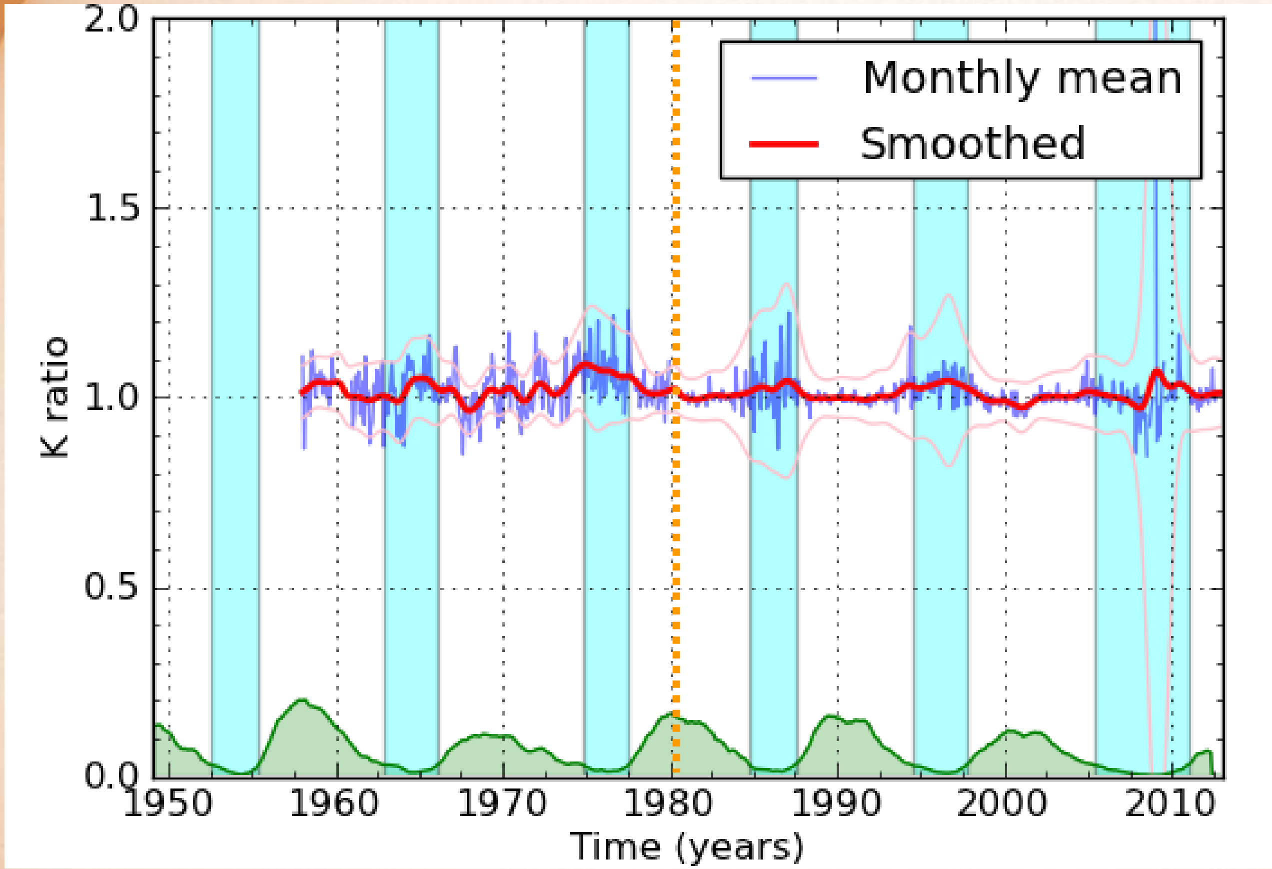
↓

Iteration: elimination at 1 σ_R
Daily average SSN R_i: final value (R_i, σ_{Ri})

Processing method: detailed mechanics

- Criteria for final iteration:
  - **Daily average standard deviation < 10%** (typ. primary trigger)
  - No new elimination at previous iteration
  - Number of remaining observations > 4 (safety limit)
- Substitution of a past k coefficent, when a monthly k cannot be computed:
  - Only if more than 80% of observations cannot be used for one station
  - ***Almost only during cycle minima (many null values)***
- Alternate «only positive » calculation:
  - If similar number of stations with null and non-null Wolf numbers ($22\% < N_{null} < 78\%$), **R_i = average only of non-null values**
  - ***Mainly during cycle minima (1 or a few days per month)***
- **2-step calculation:**
  - **Provisional number**: Day 1 of each month, subset of stations
    - Fraction of stations: from 40% (past) to 80% (currently)
  - **Definitive number:** 3-month delay, full recalculation with all stations
  - **Provisional values only replaced if new values differ by more than 5%**
    - Typ. 95% of provisional values are kept as definitive

Processing method: equivalence R_i - W_{LO}



- W_{LO}/R_i monthly rms variation: ~3 %
- Lower dispersion since 1981
- No significant jump at the Zürich-Locarno transition

⇨ **The choice of Locarno as pilot station allowed a seamless transition in 1981**

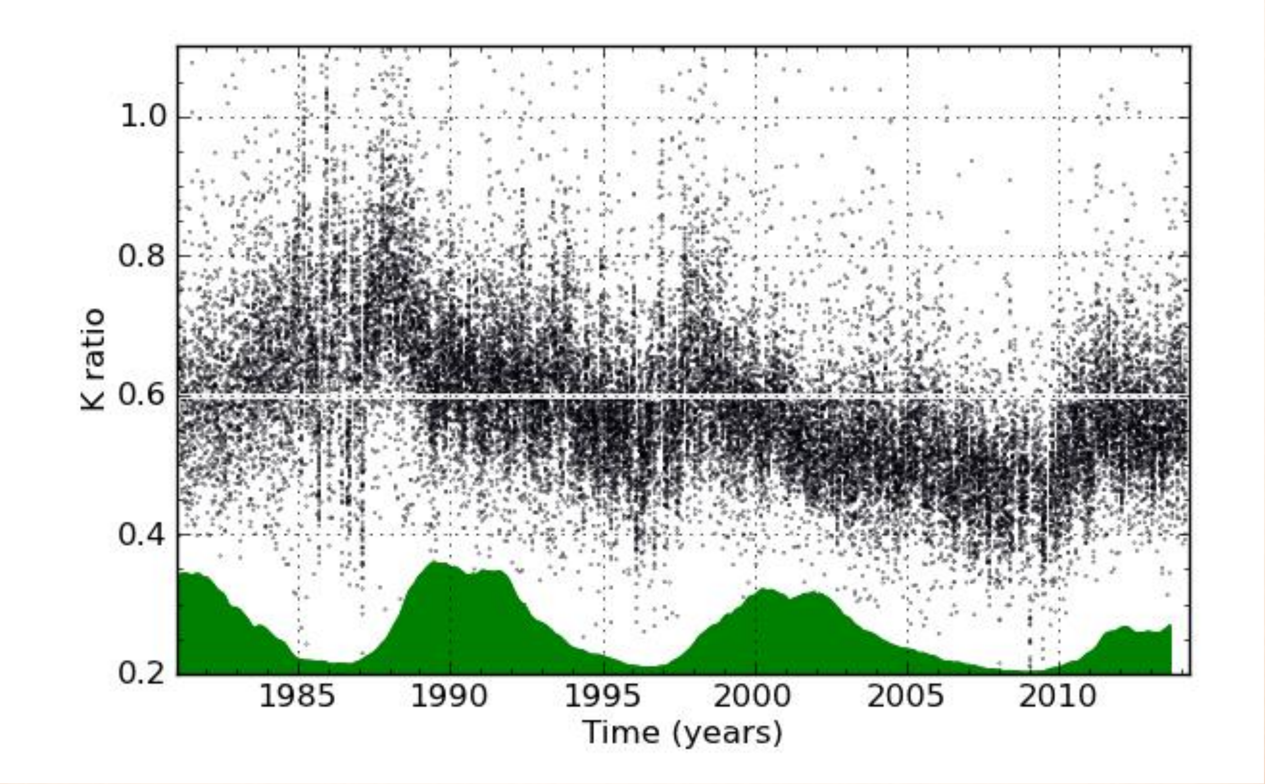
⇨ **On timescales > 1 month, R_i is fully tracking the pilot station**

Global k statistics

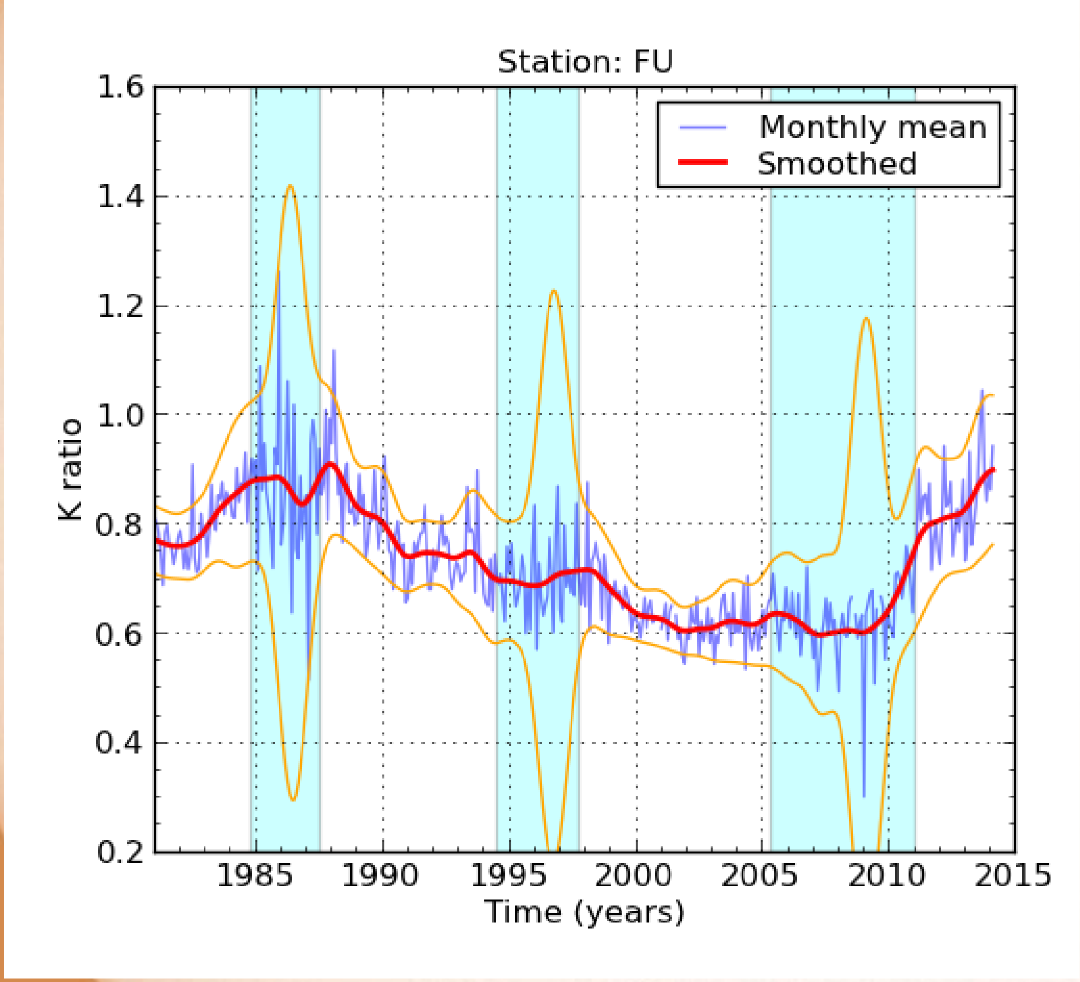
The power of a large network: global statistics

- Base data: 80 long-duration stations (> 200,000 observations)
- Monthly average k ratio

⇨ Clear common time variation of the k distribution



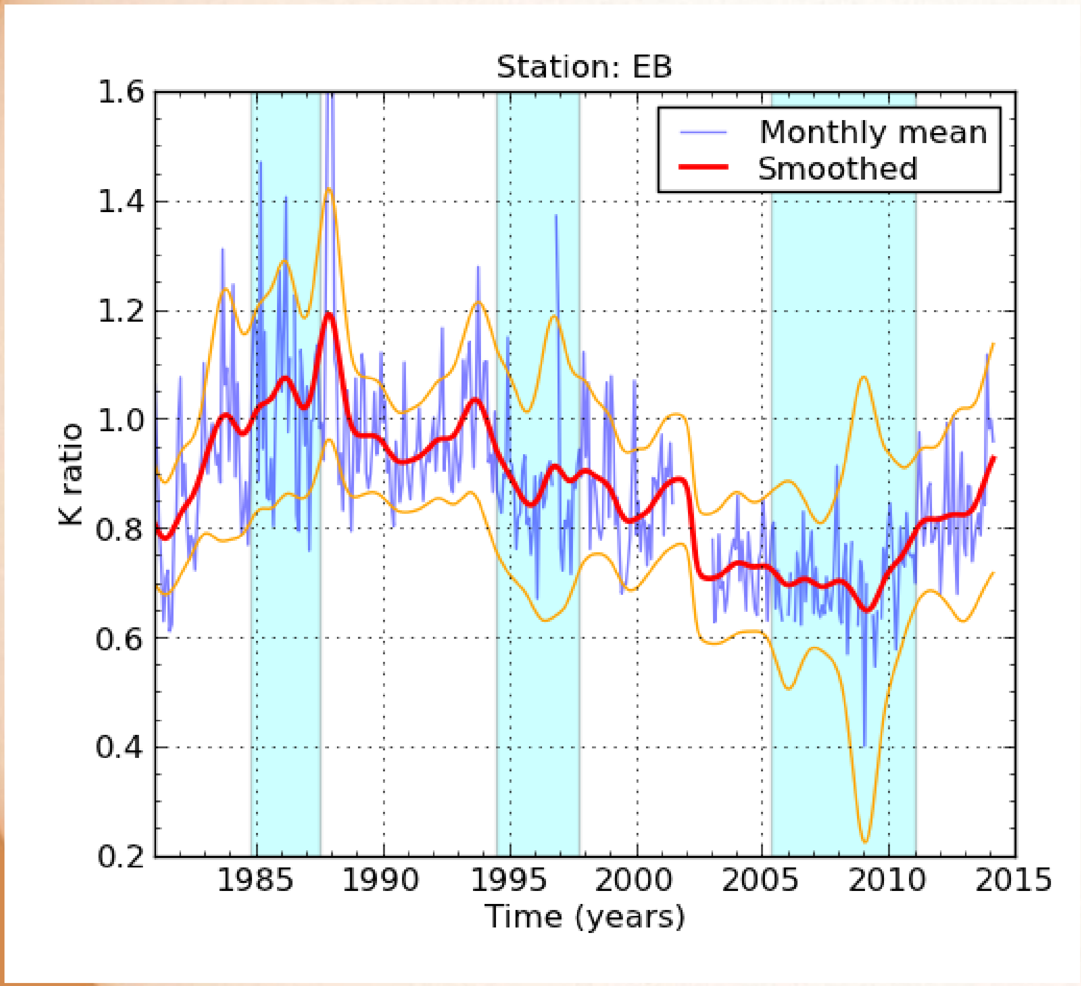
k variations: individual stations



Observer: Fujimori, Japan

- The same overall k variation is found for individual stations
  - Stations are **uncorrelated** (world wide dispersion, no direct interaction between stations)
  - All k ratios are referred to a **common reference**: Locarno
- Source of the drift = Specola, Locarno

k variations: individual stations

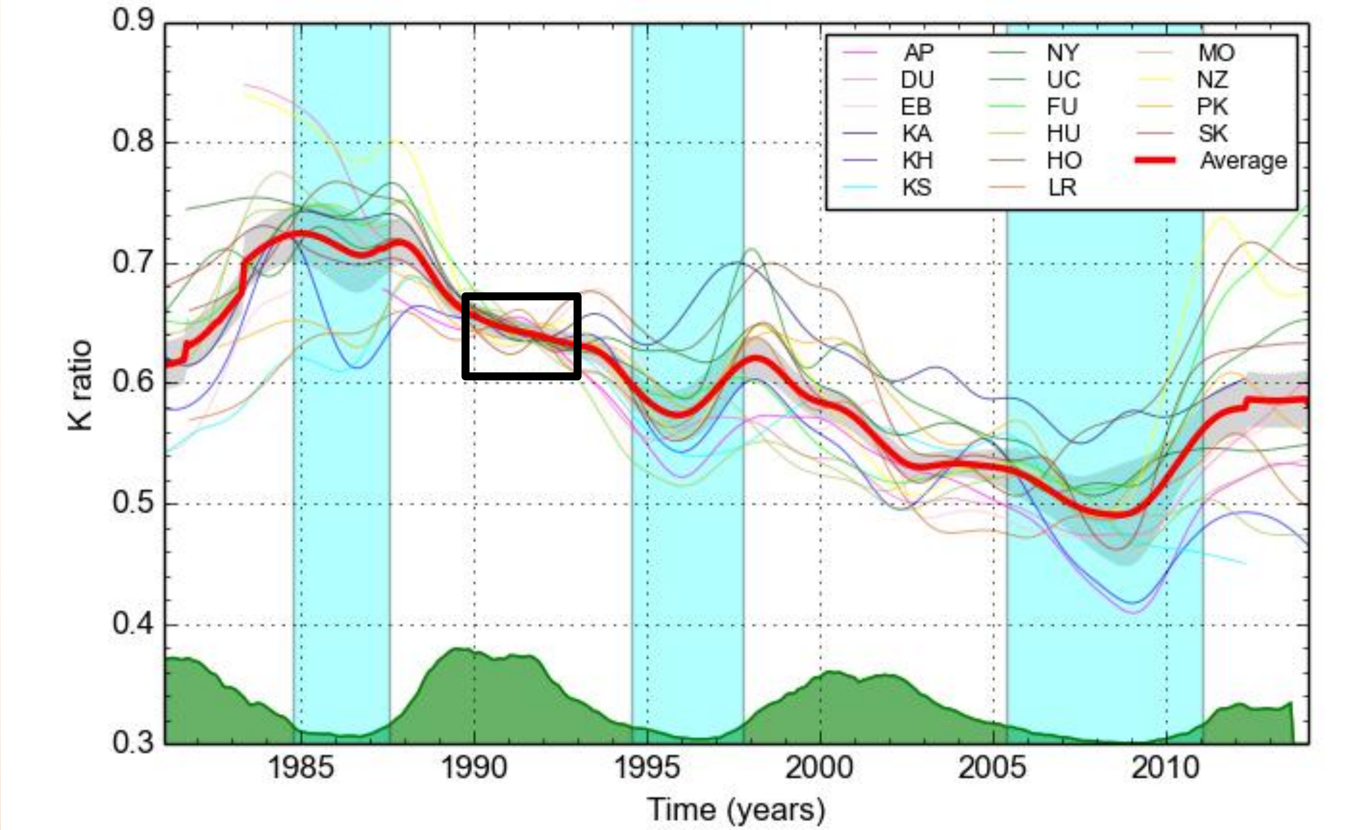


Observer: Ebro, Spain

- Rationale: a global network average will:
  - Reduce the uncorrelated random deviations of individual stations
  - Provide a precise determination of the common variation = Specola drift

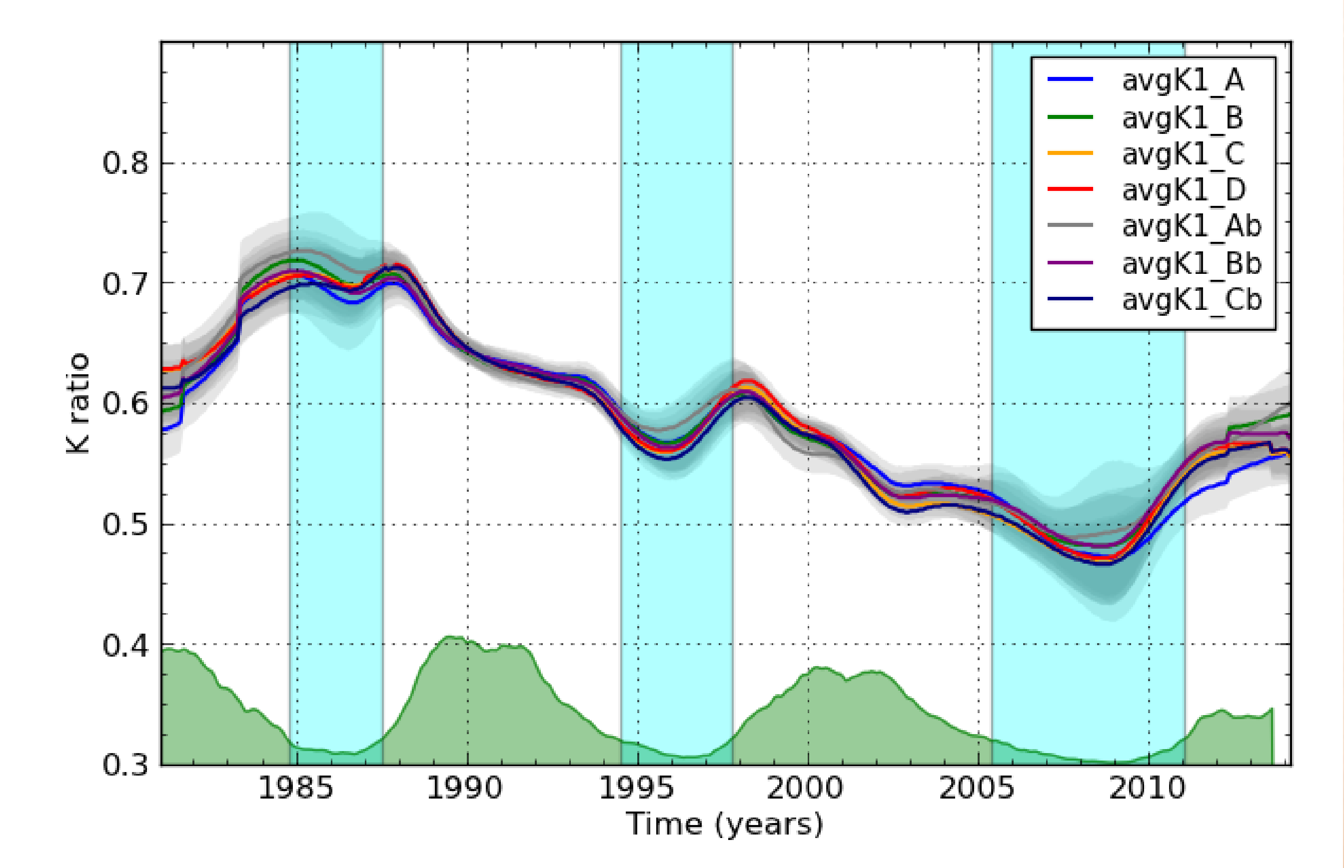
First step: normalization and first average

- Stations normalized to common scale over the interval 1990-1993:
  - Interval common to the largest number of stations
  - At maximum of cycle 22
  - Smaller common variation of k
  - ⇨ Gives an equal weight to all stations in the average
- Gaussian smoothing (FWHM = 20 months, σ = 12 months)



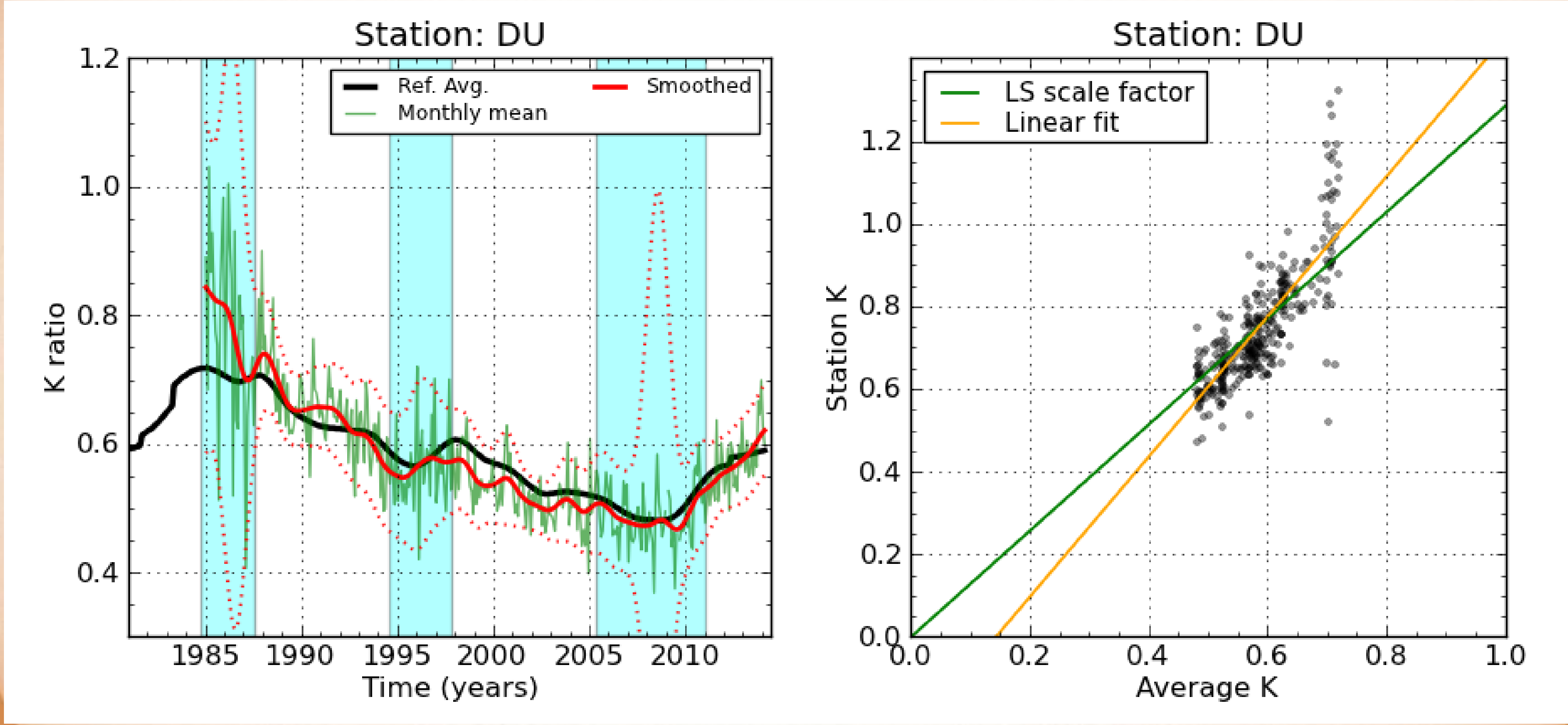
First step: normalization and first average

- Tests with different subsets of stations (duration, absence of large drifts, etc.): A, B, C : best, good, fair; D : poor (excluding C)
- **All average profiles agree +/- 10% max. regardless of the selection**
- All stations can now be least-square fitted to this average profile (entire interval 1981-2014)



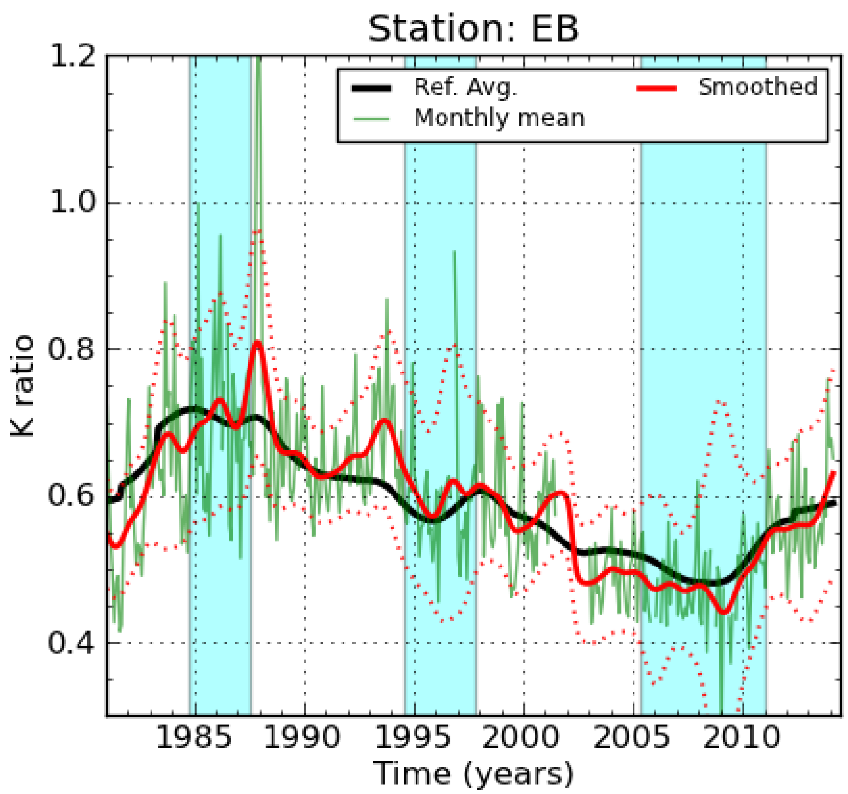
Linear fit to the average k

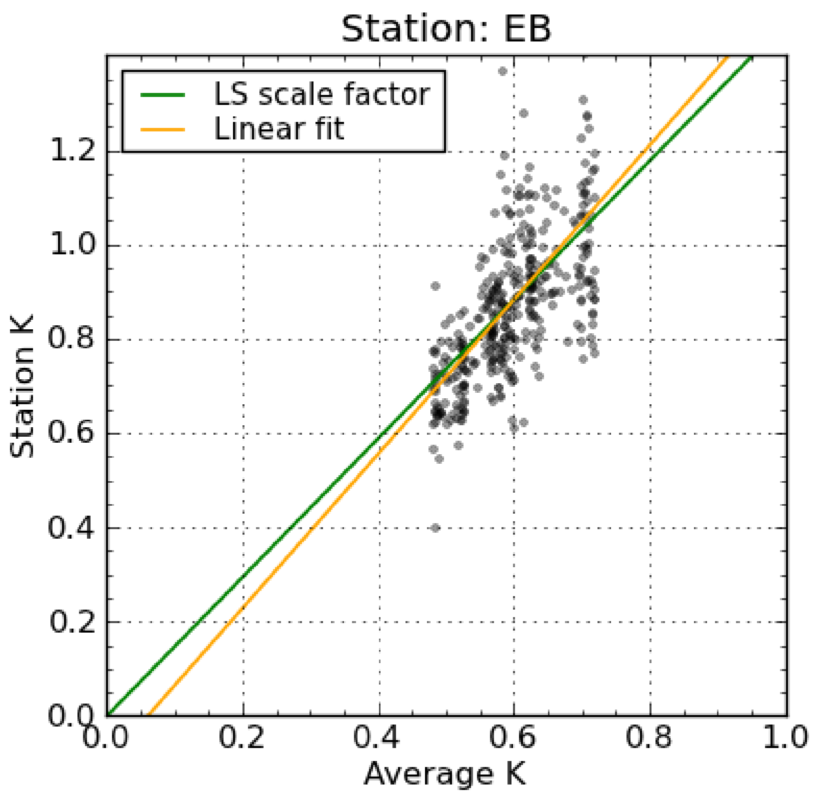
- F. Dubois, Belgium



Linear fit to the average k

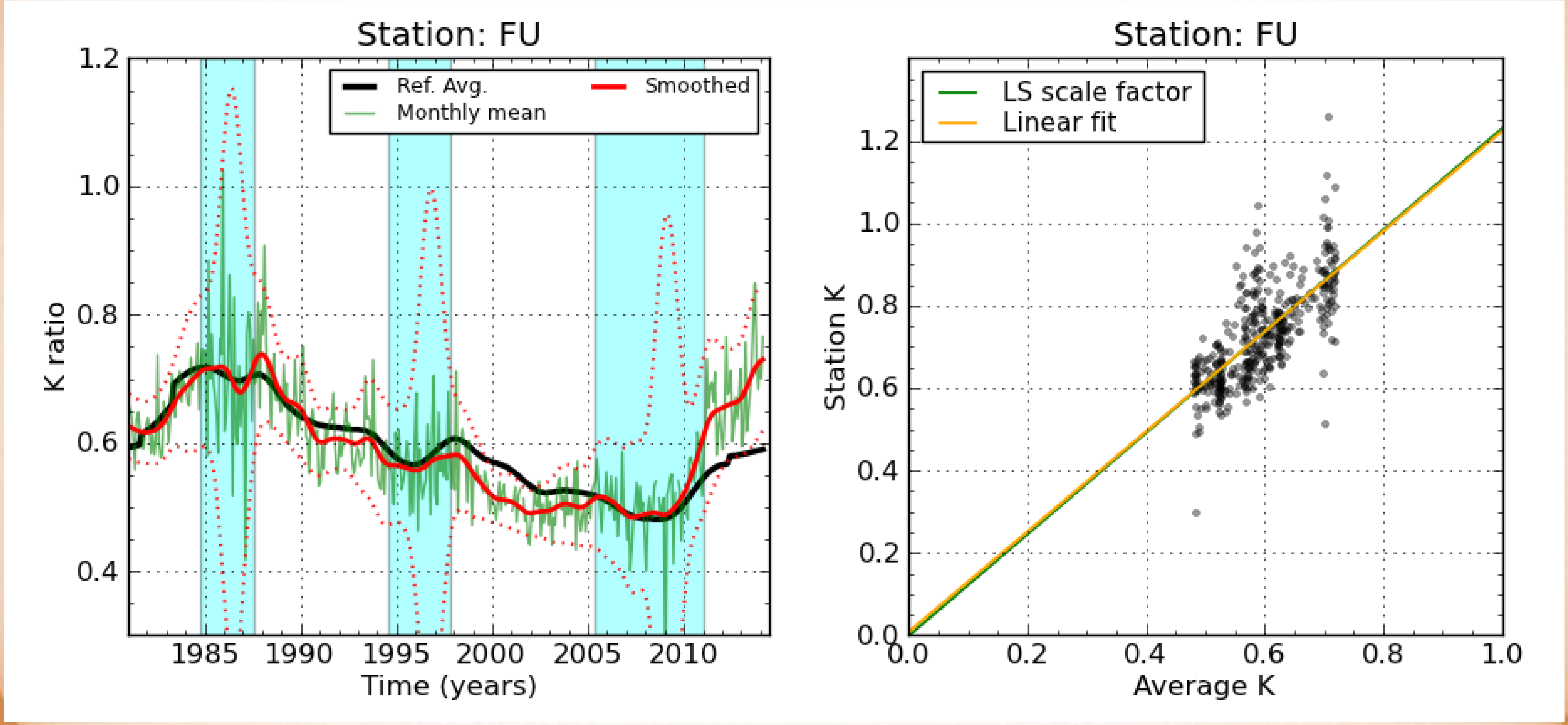
- Ebro, Spain





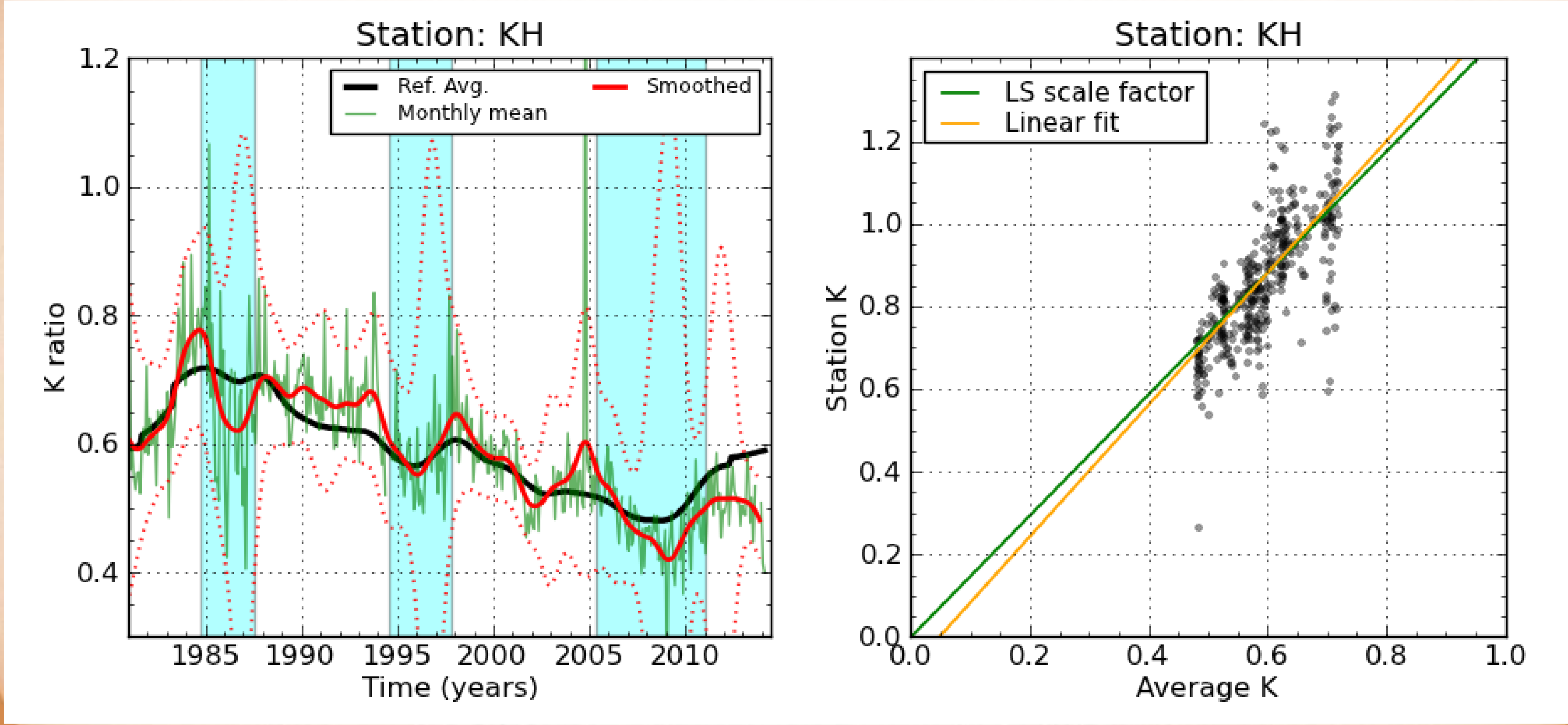
Linear fit to the average k

- K. Fujimori, Japan



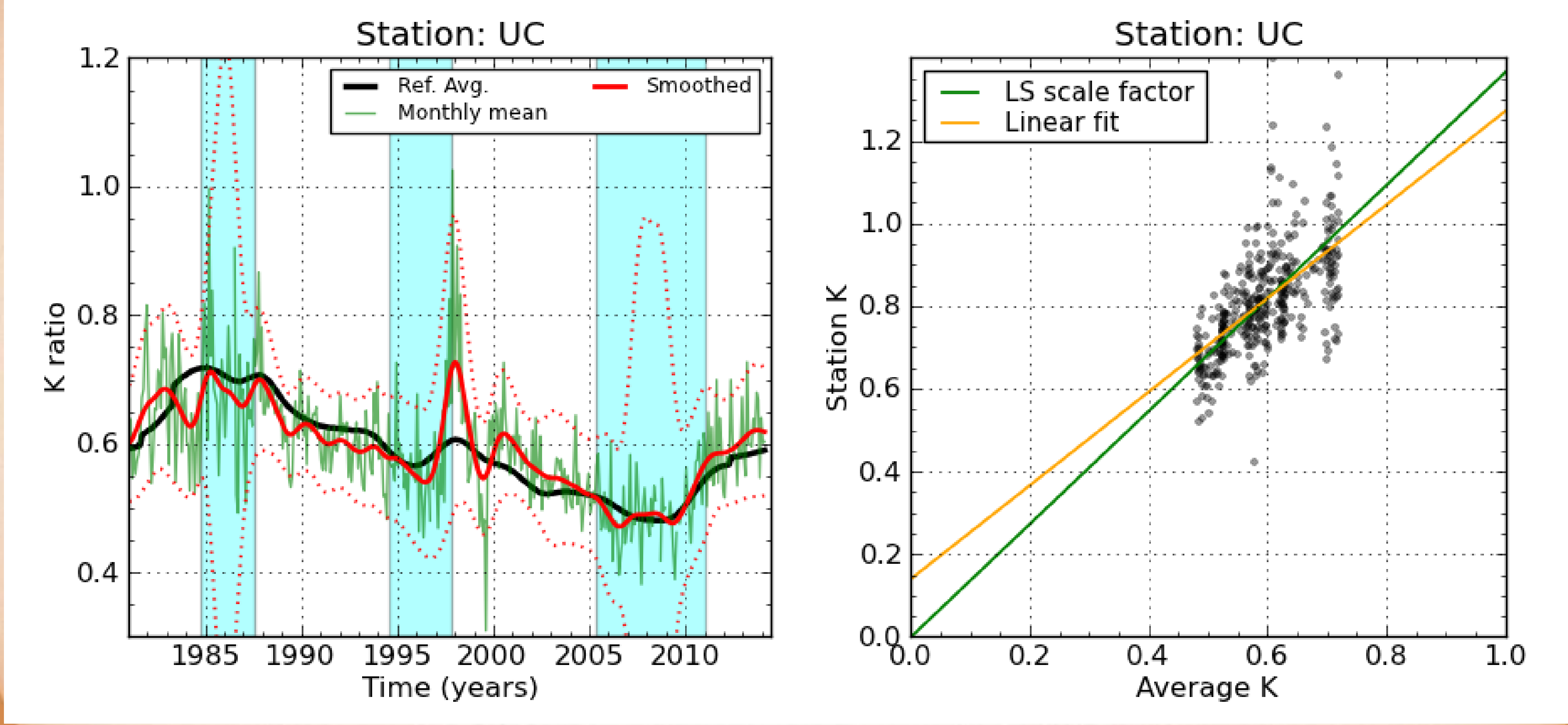
Linear fit to the average k

- Kandilli, Turkey



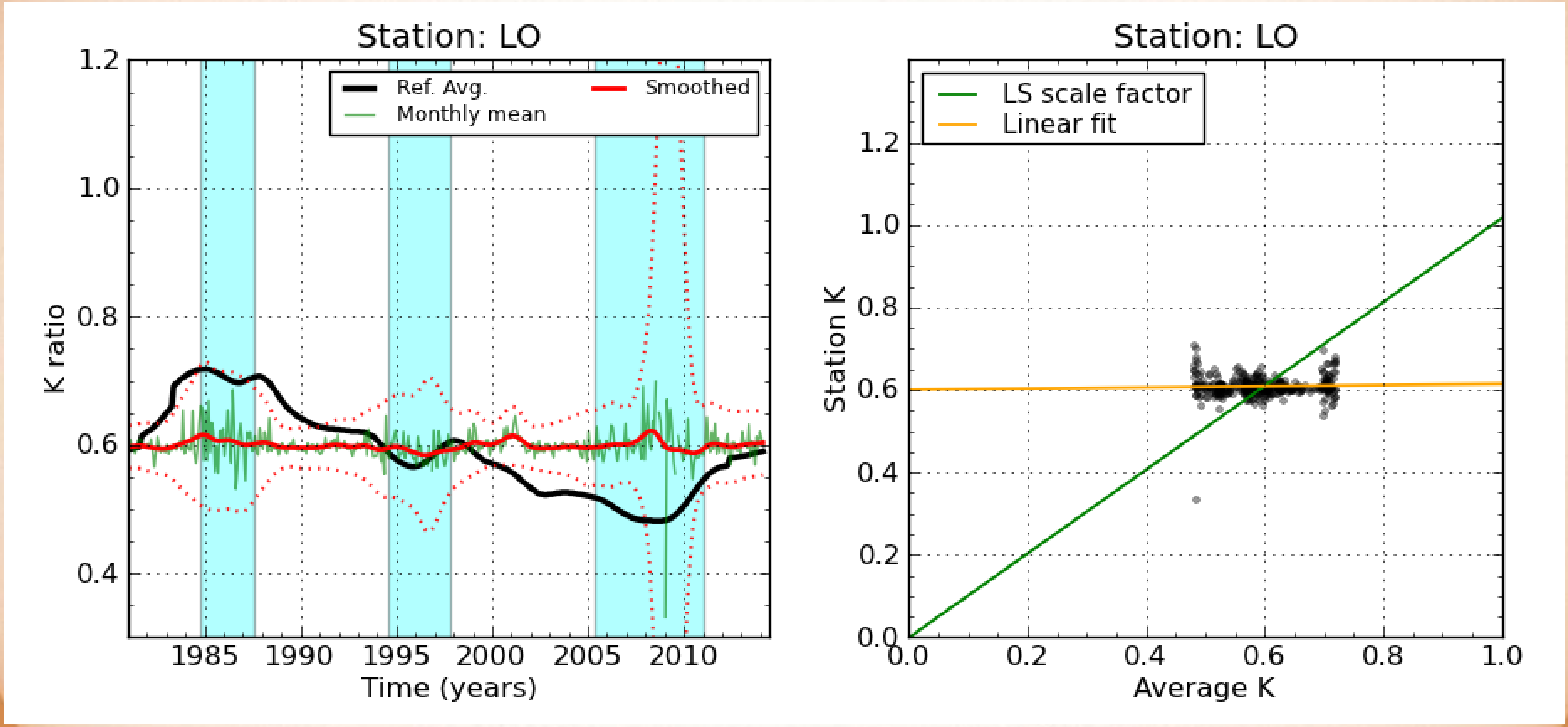
Linear fit to the average k

- Uccle, Belgium



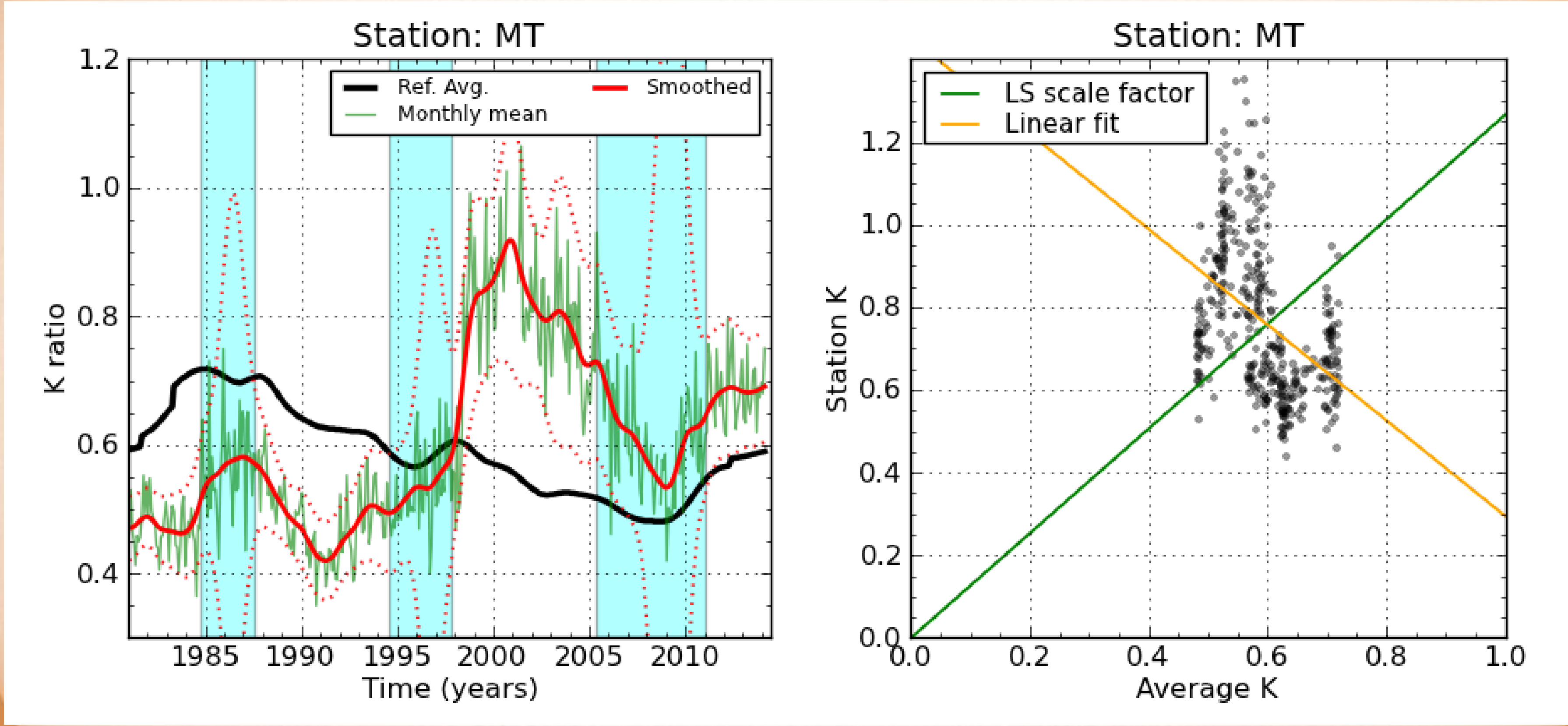
Linear fit to the average k

- Specola, Locarno, Switzerland



Linear fit to the average k

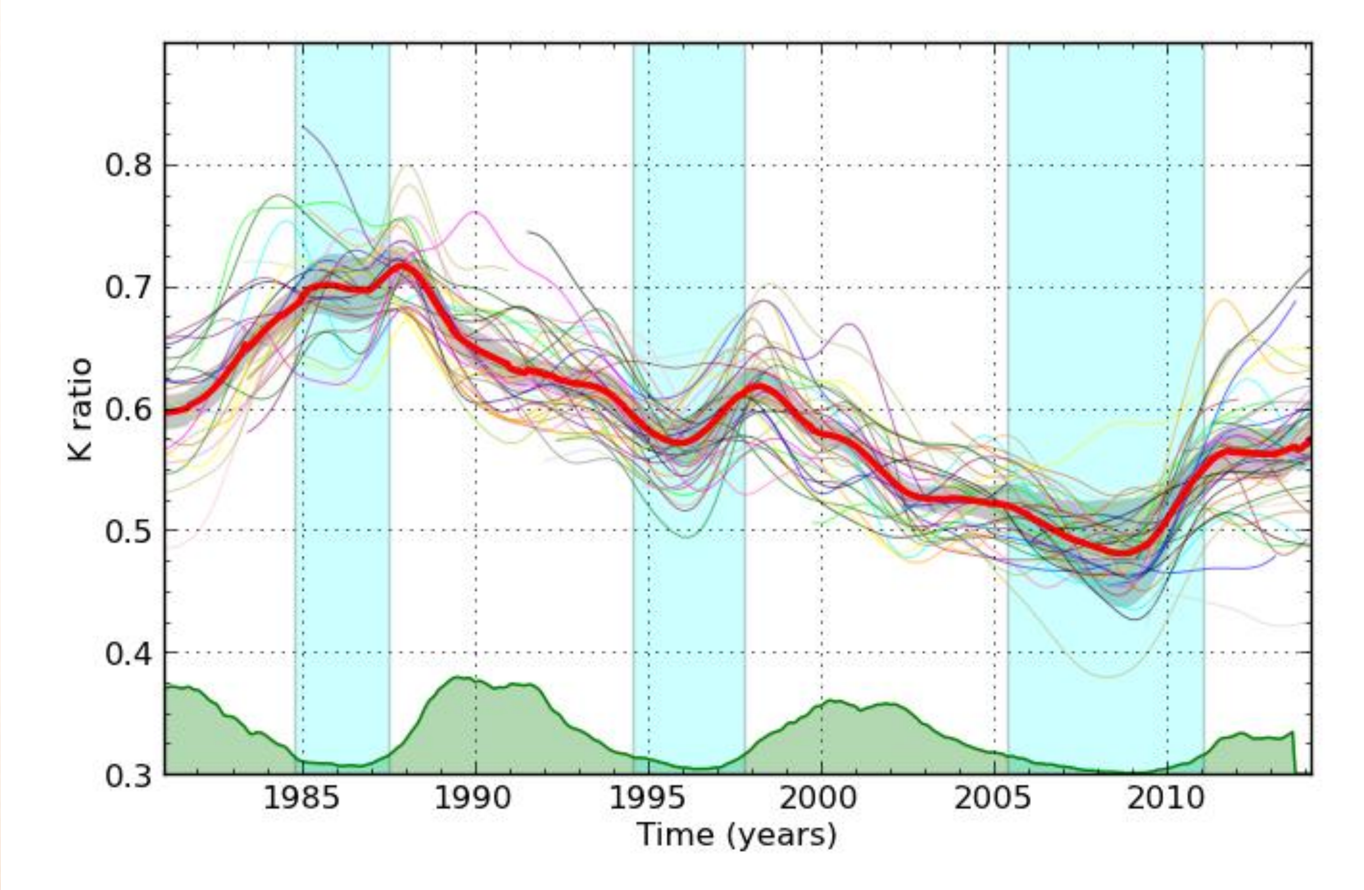
- Mitaka, Japan



Second step: normalization and average

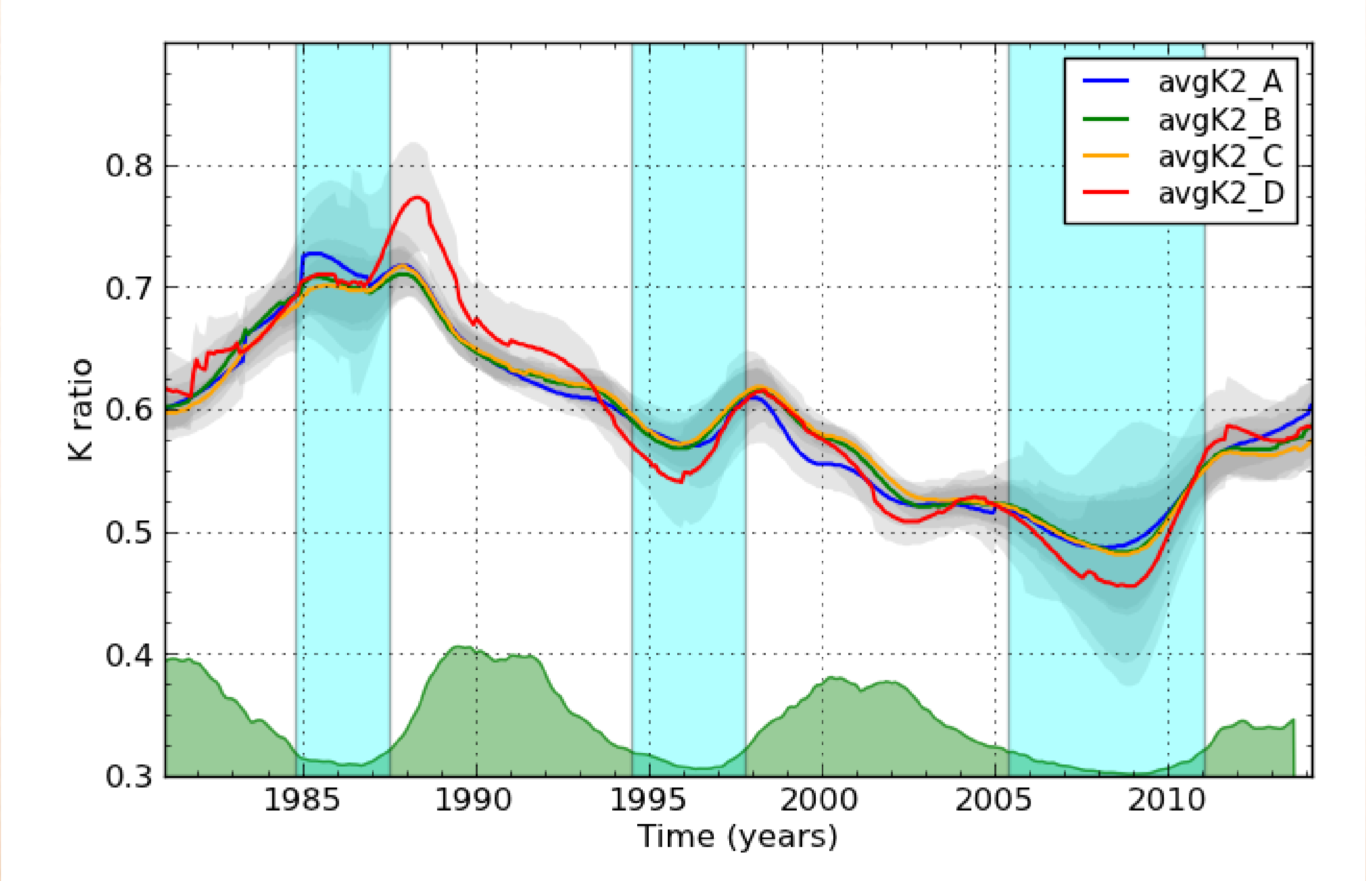
- Normalization over full duration for each station:
  - Least-square fit (intercept at origin) = multiplicative factor
- Global scale factor: average k in 1981 set to 0.6

(Zürich reference up to 1980)



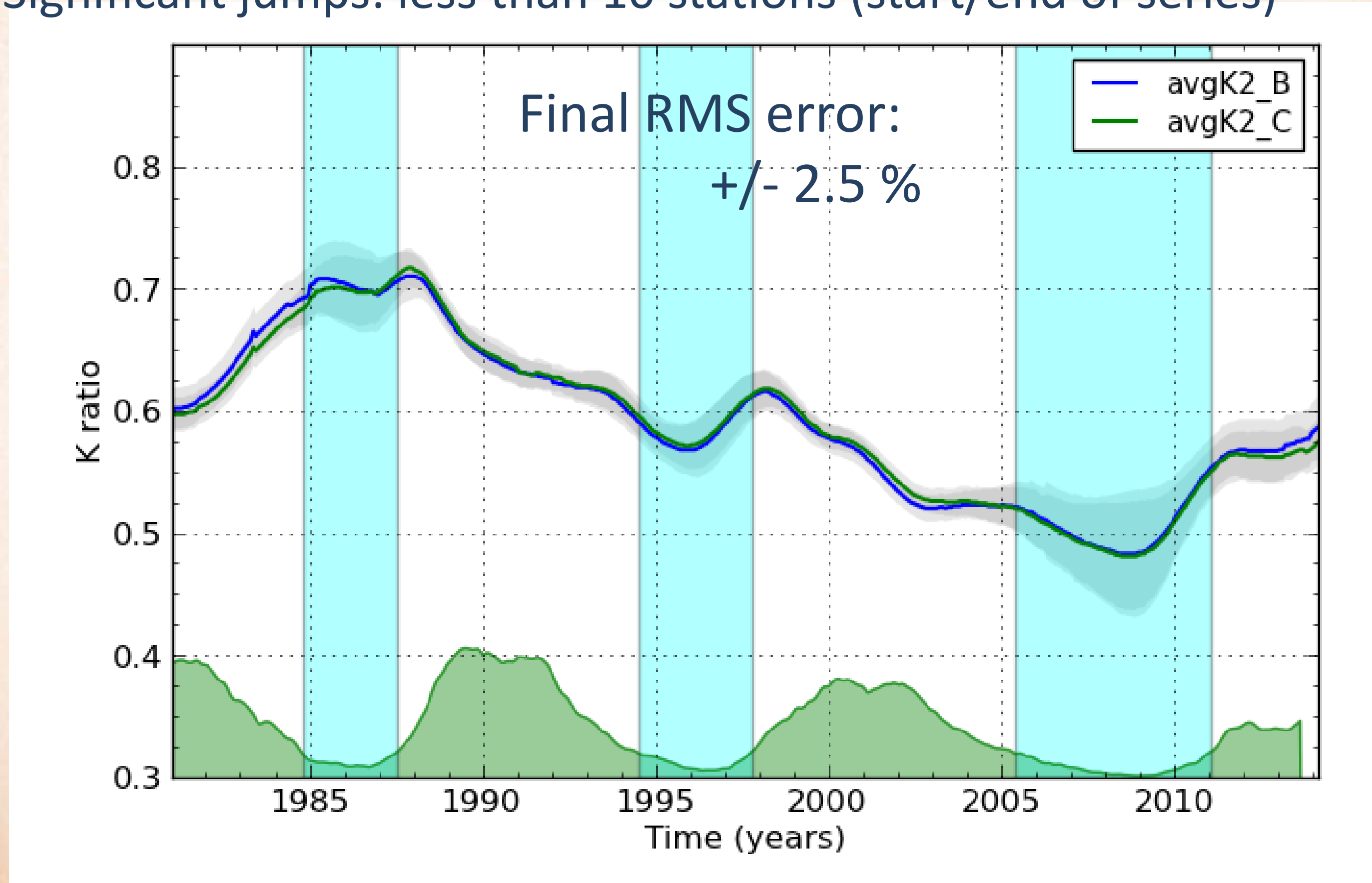
Second step: final average k correction

- Test with different subsets of stations
- Set D (quality: poor):
  - Disagrees by up to 20% with the other sets
  - Largest amplitude of the variations



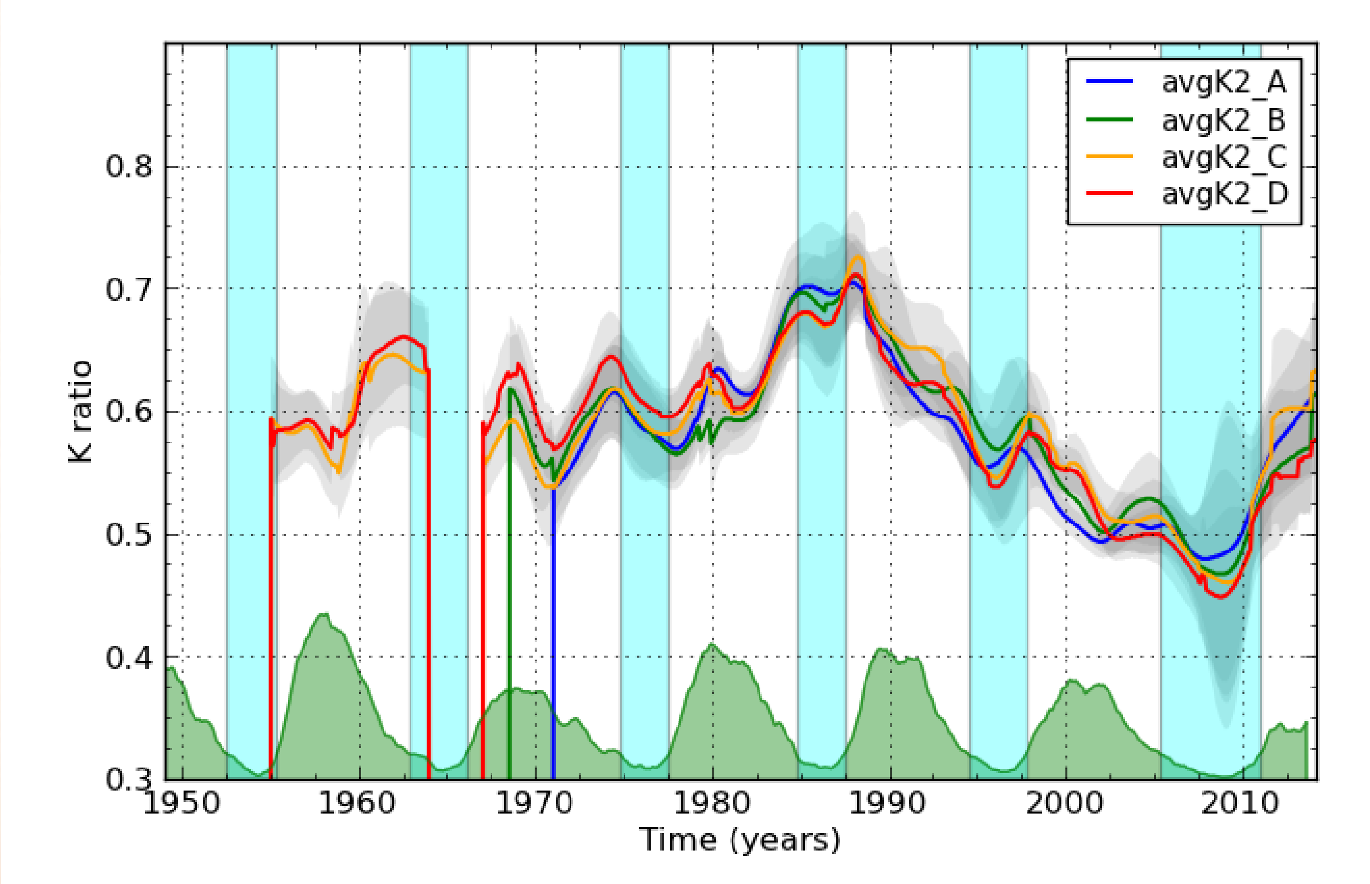
Second step: final average k correction

- Test with different subsets of stations
- Set A (quality: best):
  - Disagrees by up to 10% with the other sets
  - Significant jumps: less than 10 stations (start/end of series)



Extending to the Zürich period (1955 – 2014)

- Additional data:
  - Raw paper records from Zürich: 20 auxilliary stations
  - Most stations active before and after 1981 (Uccle, Kanzelhöhe, Catania, Kandilli, Fujimori): important links for the 1981 continuity

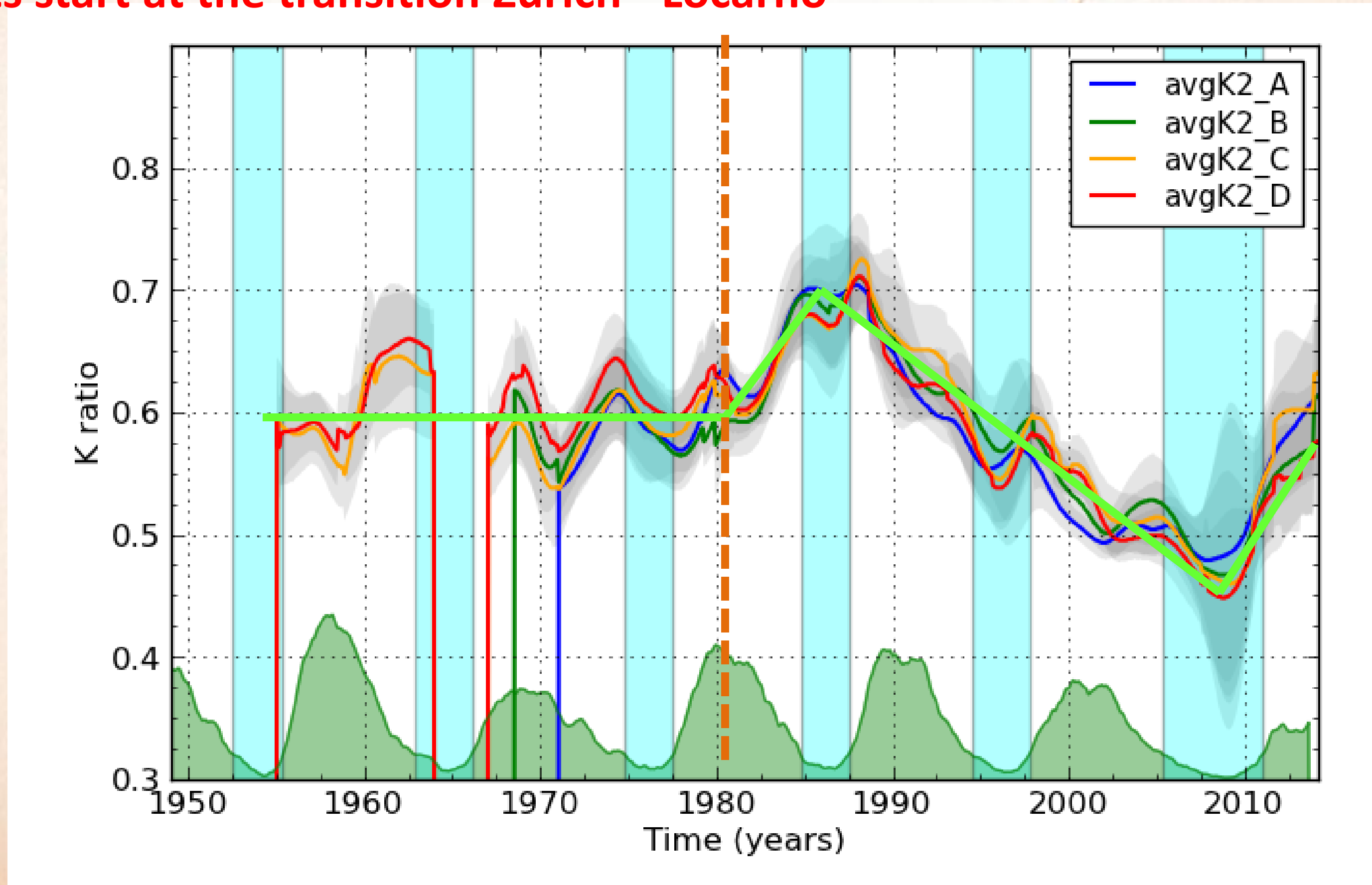


Extending to the Zürich period (1955 – 2014)

- Same statistical determination of the average k variations:
  - Results agree over the common 1981-2014 interval
  - RMS errors are larger: +/- 6% (few stations)

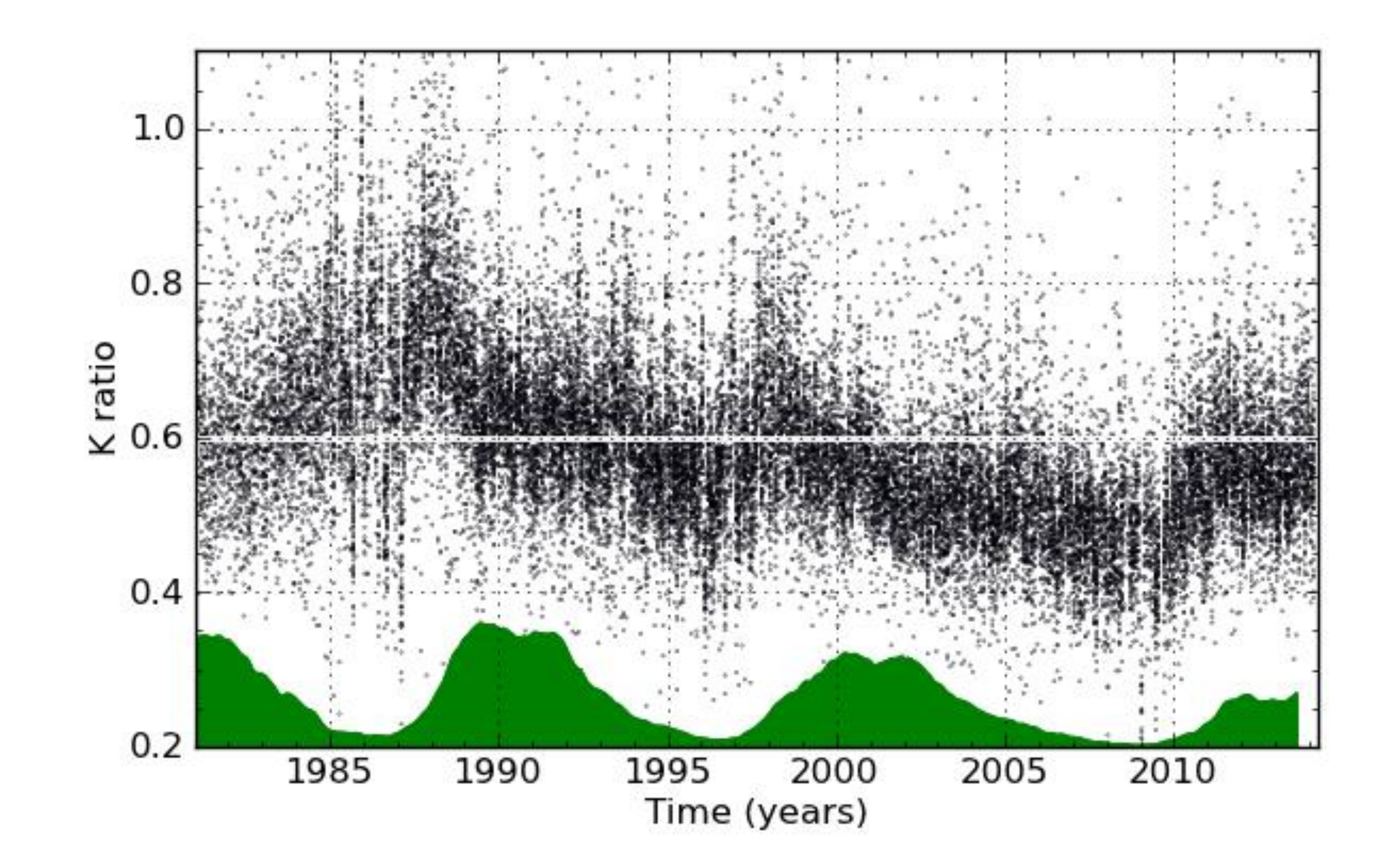
⇨ No global trend until 1980

⇨ Drifts start at the transition Zürich - Locarno

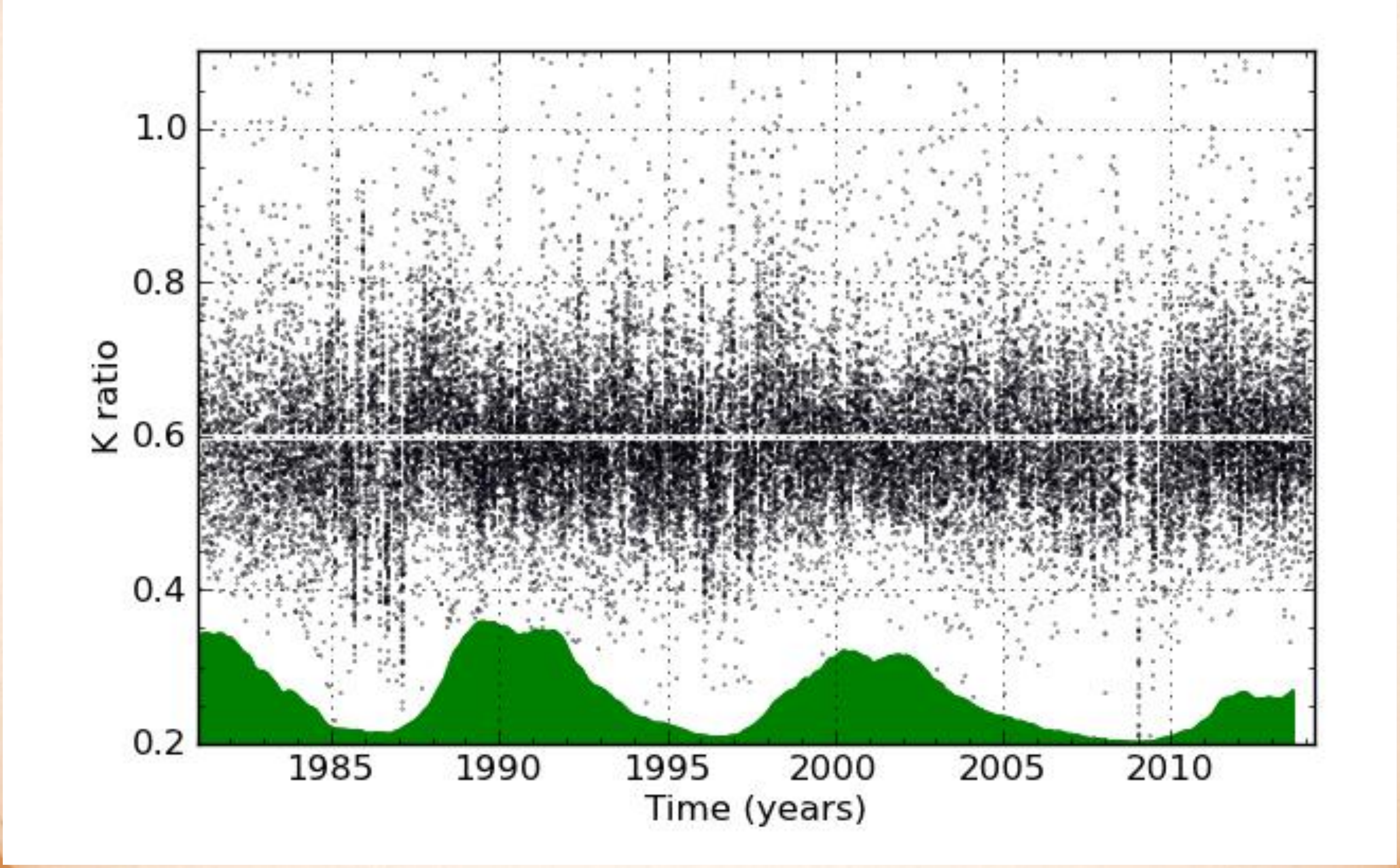


Corrected k series

Applying the correction

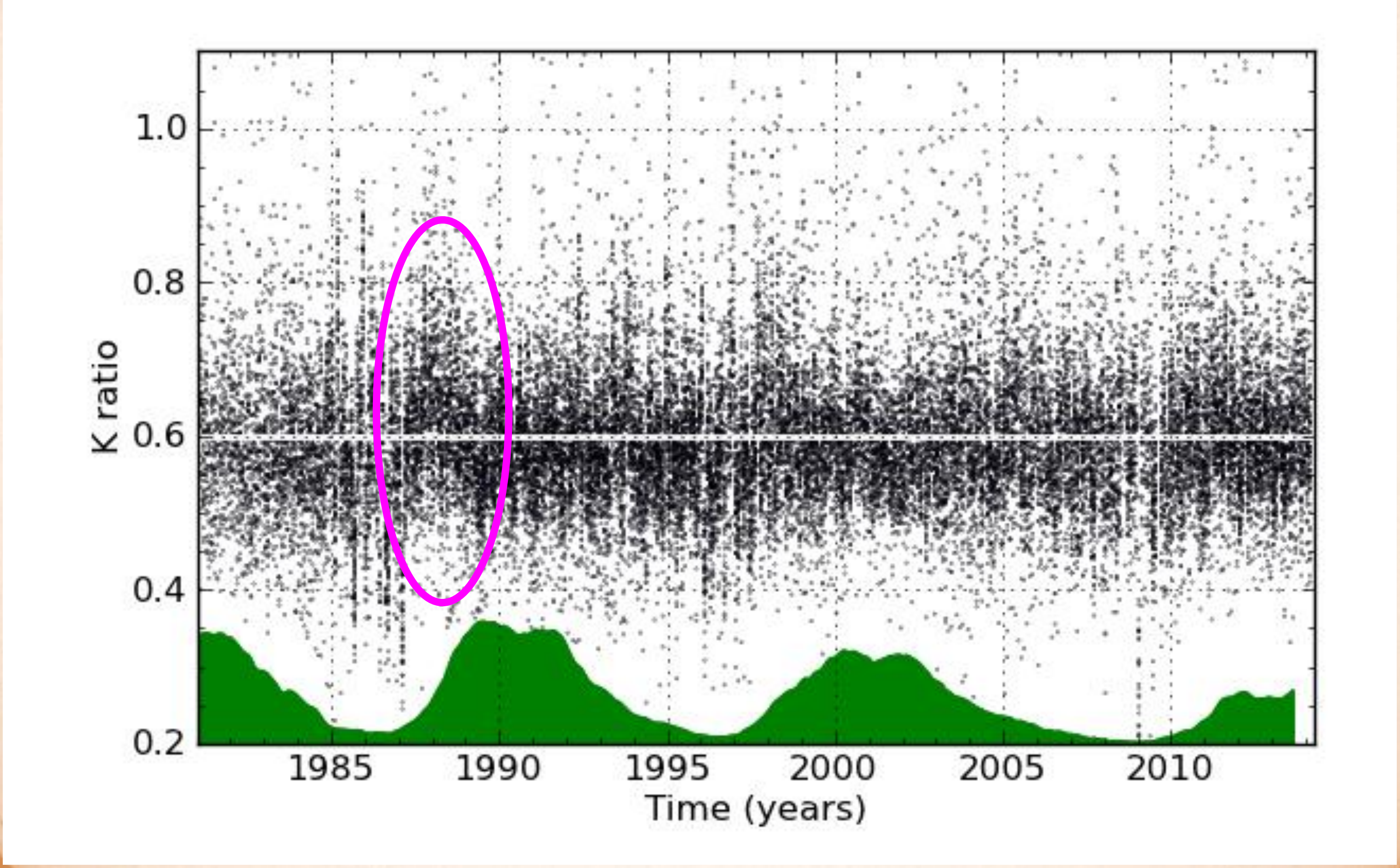


Applying the correction

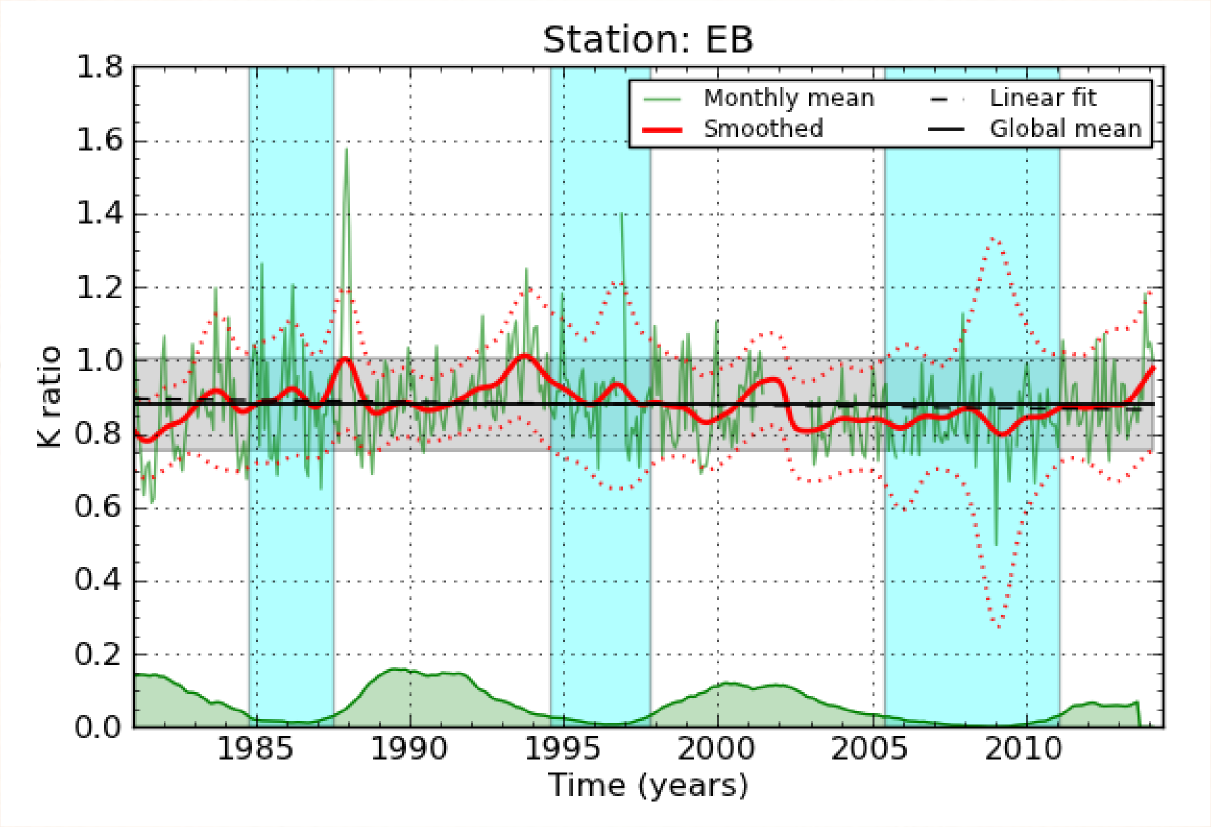


Applying the correction

- Sharp feature in 1988, not fully corrected
  - residual of temporal smoothing

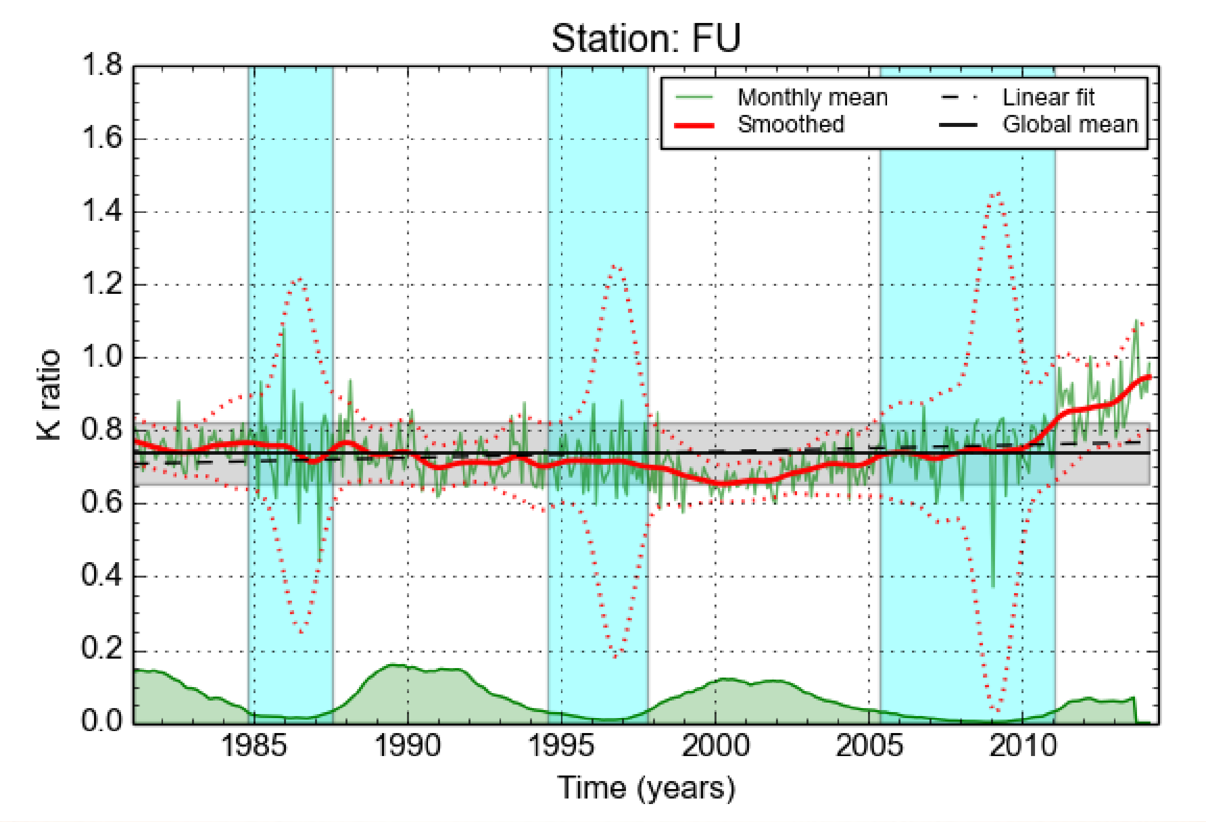


K coefficient after correction: sample stations



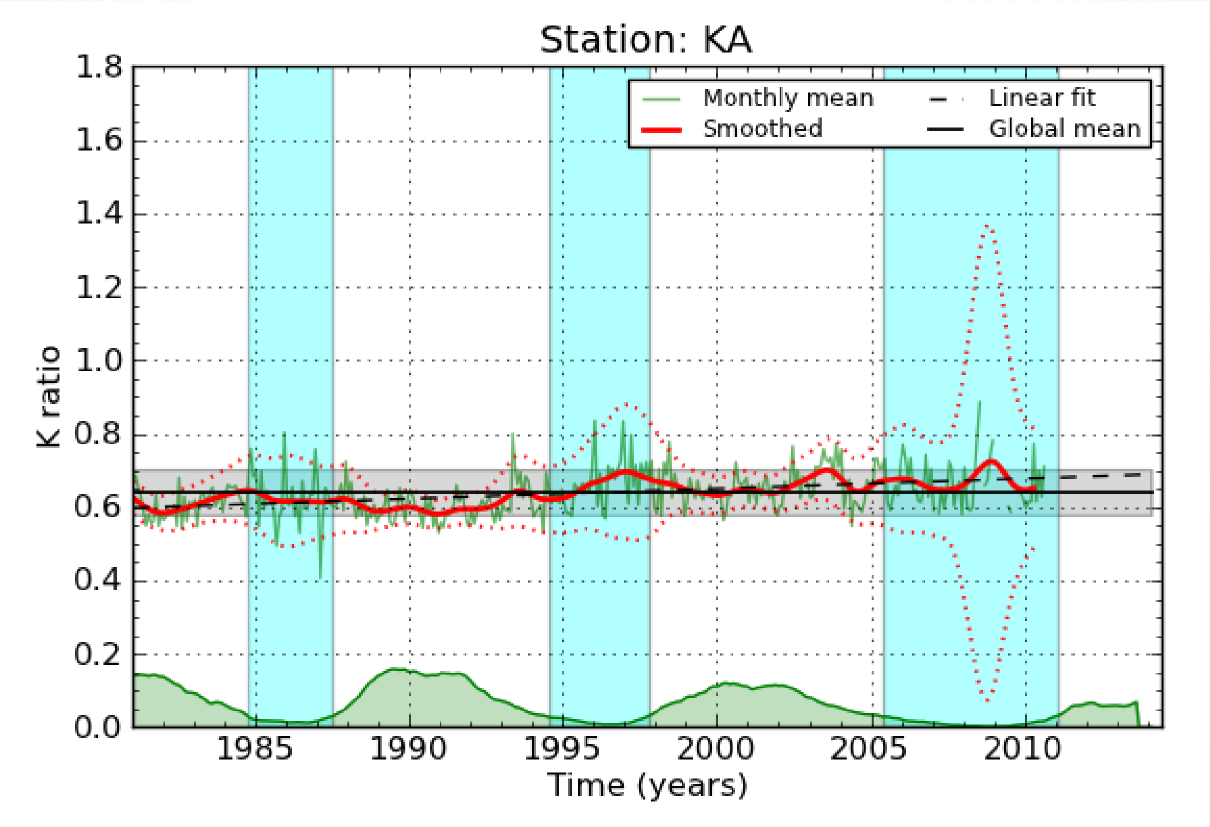
Ebro, Observatory, Spain

K coefficient after correction: sample stations



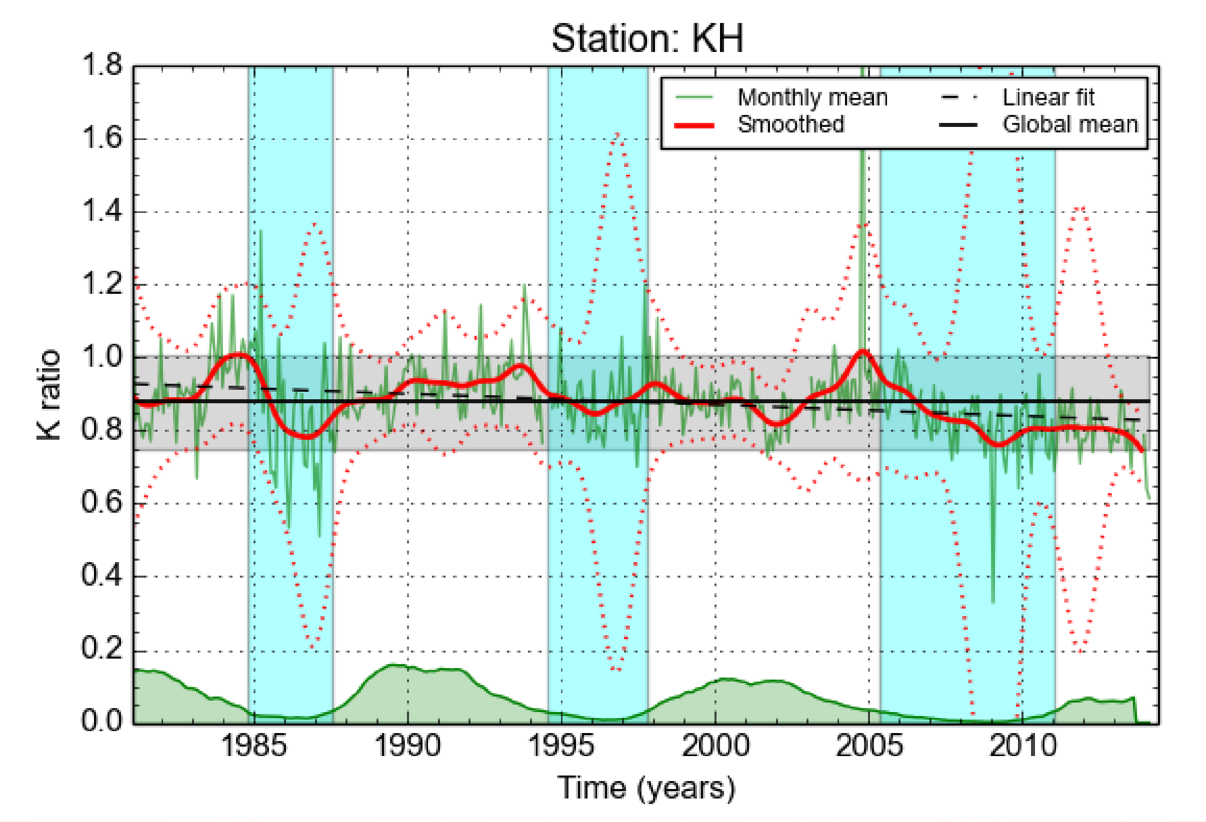
K. Fujimori, Japan

K coefficient after correction: sample stations



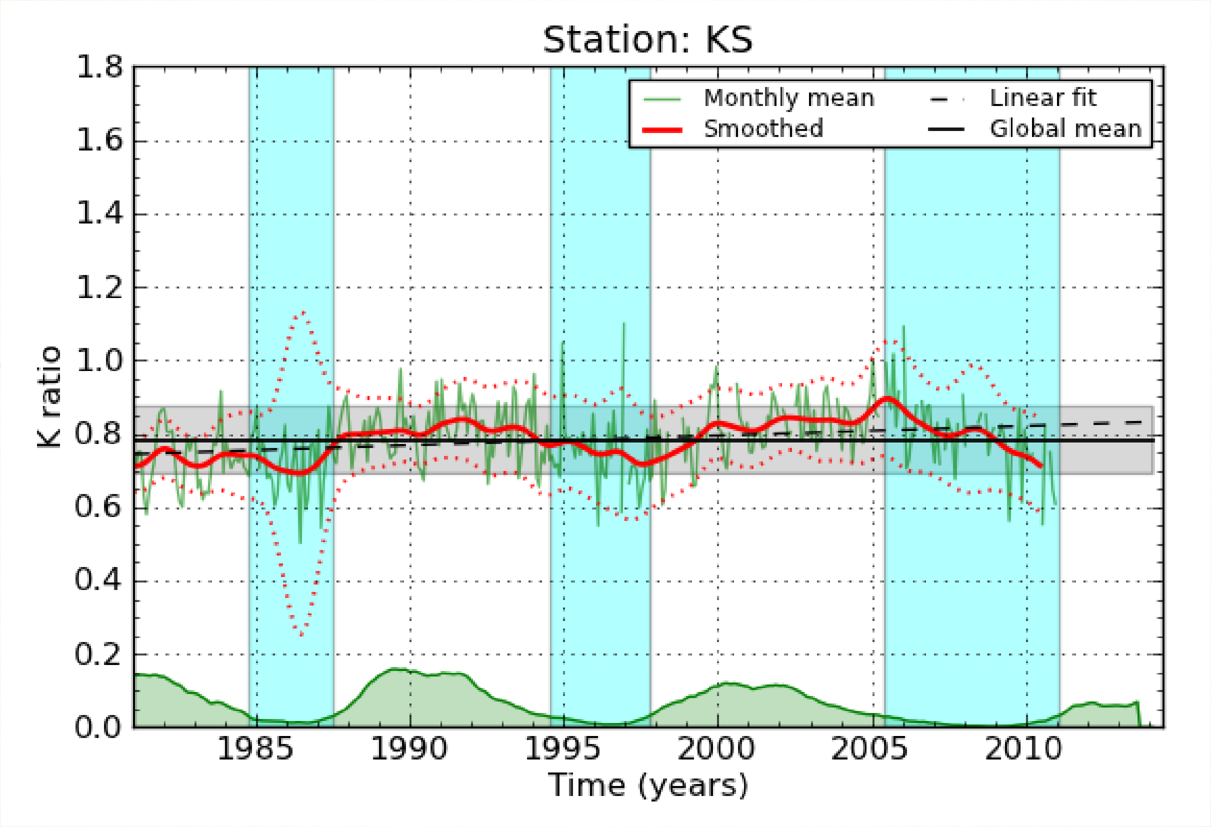
I. Kawagushi , Japan

K coefficient after correction: sample stations



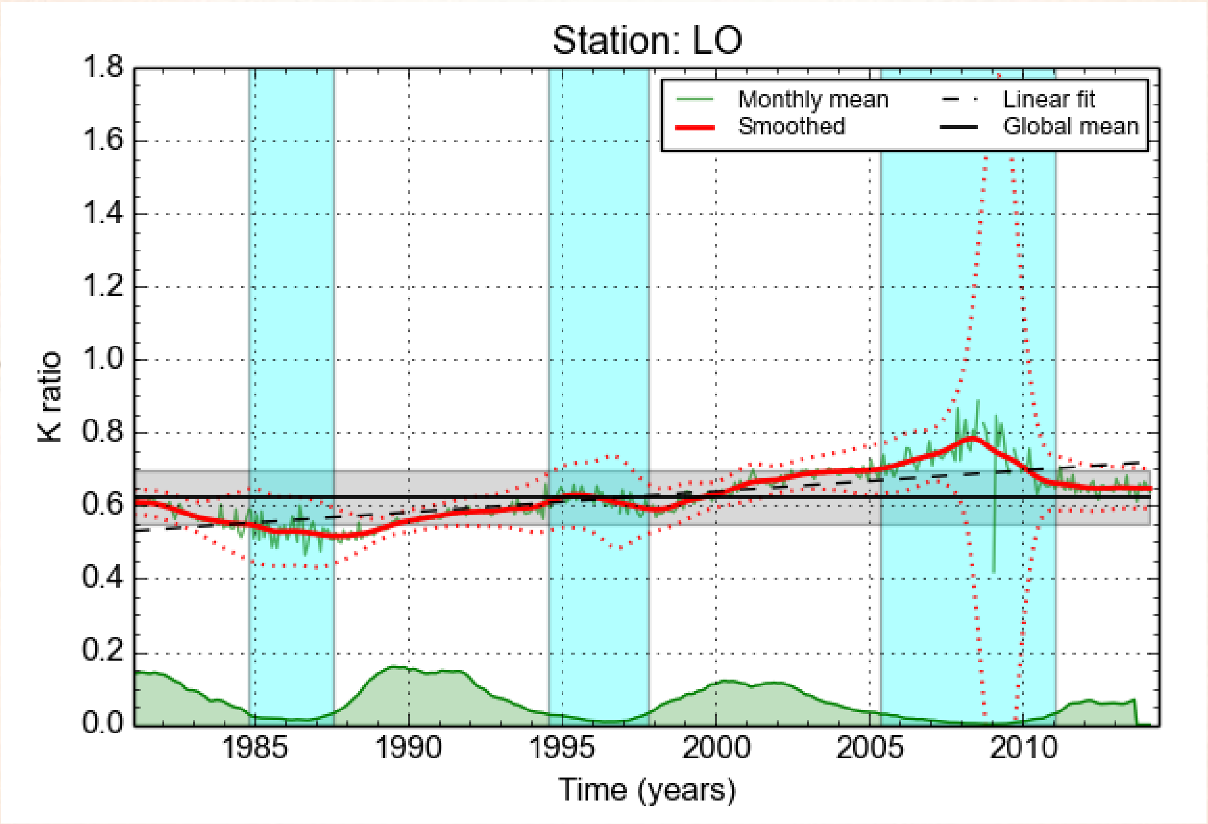
Kandilli Observatory, Turkey

K coefficient after correction: sample stations



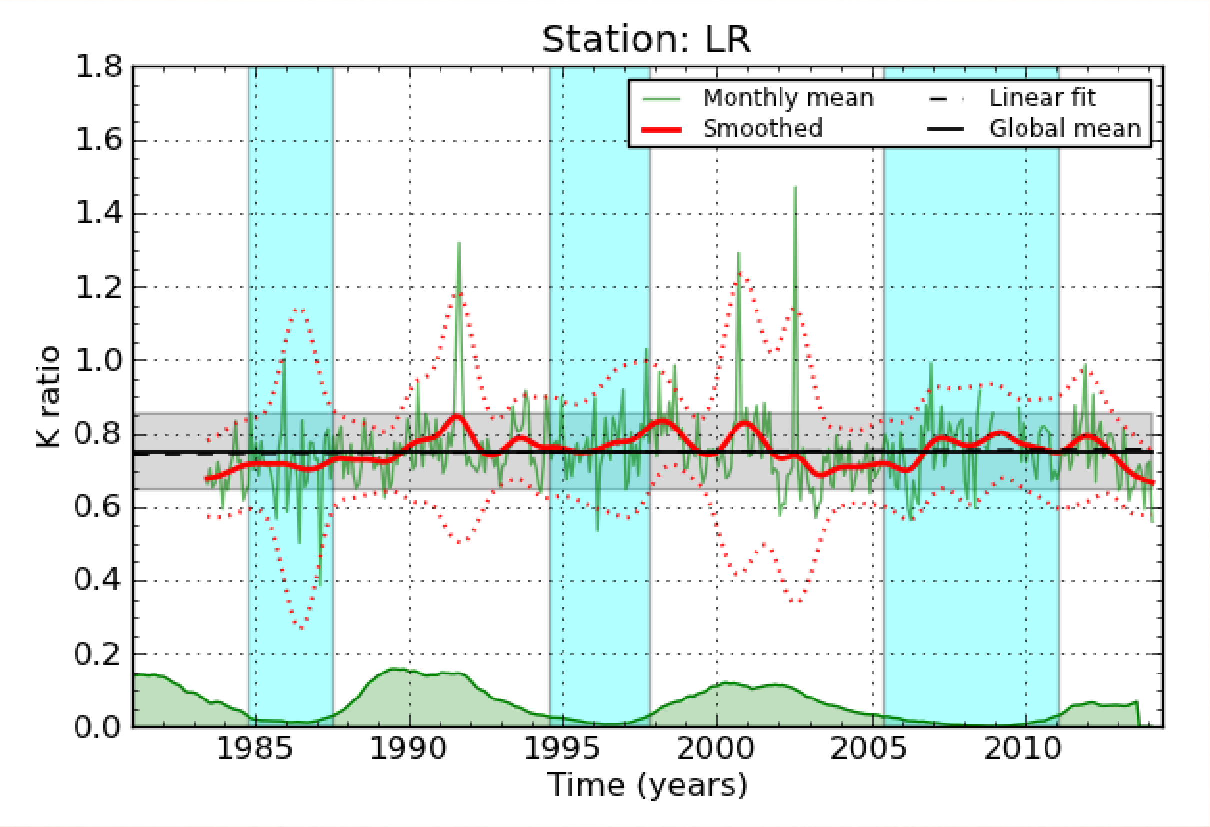
Kislovodsk Observatory, Russia

K coefficient after correction: sample stations



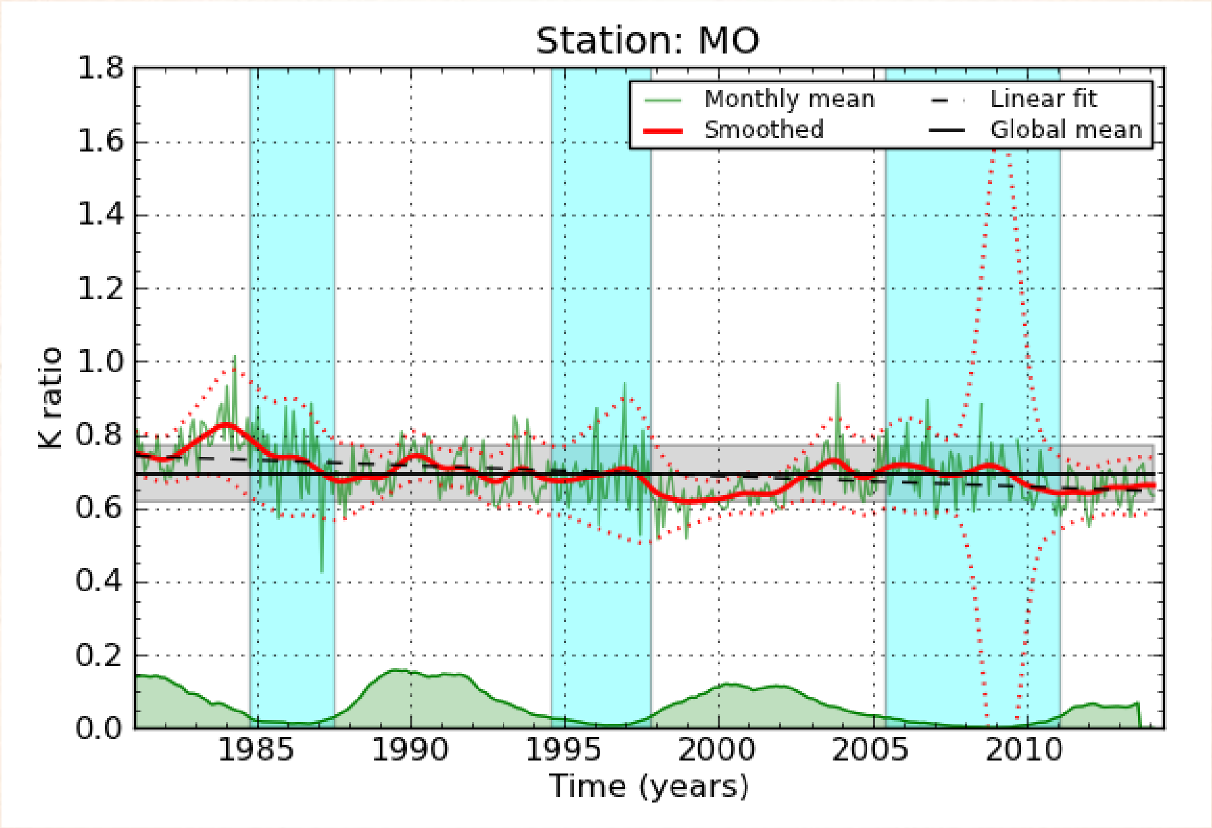
Specola Solare, Locarno, Switzerland

K coefficient after correction: sample stations



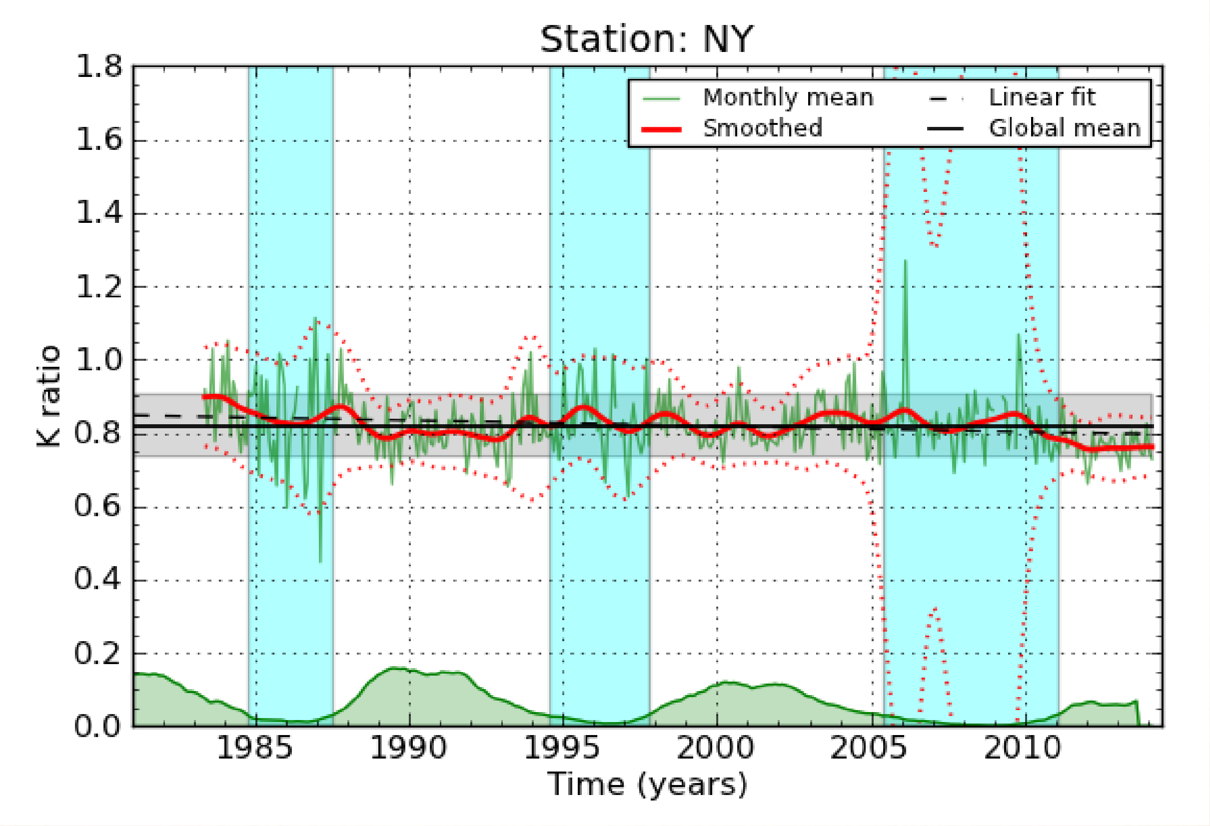
Learmonth Observatory, Australia

K coefficient after correction: sample stations



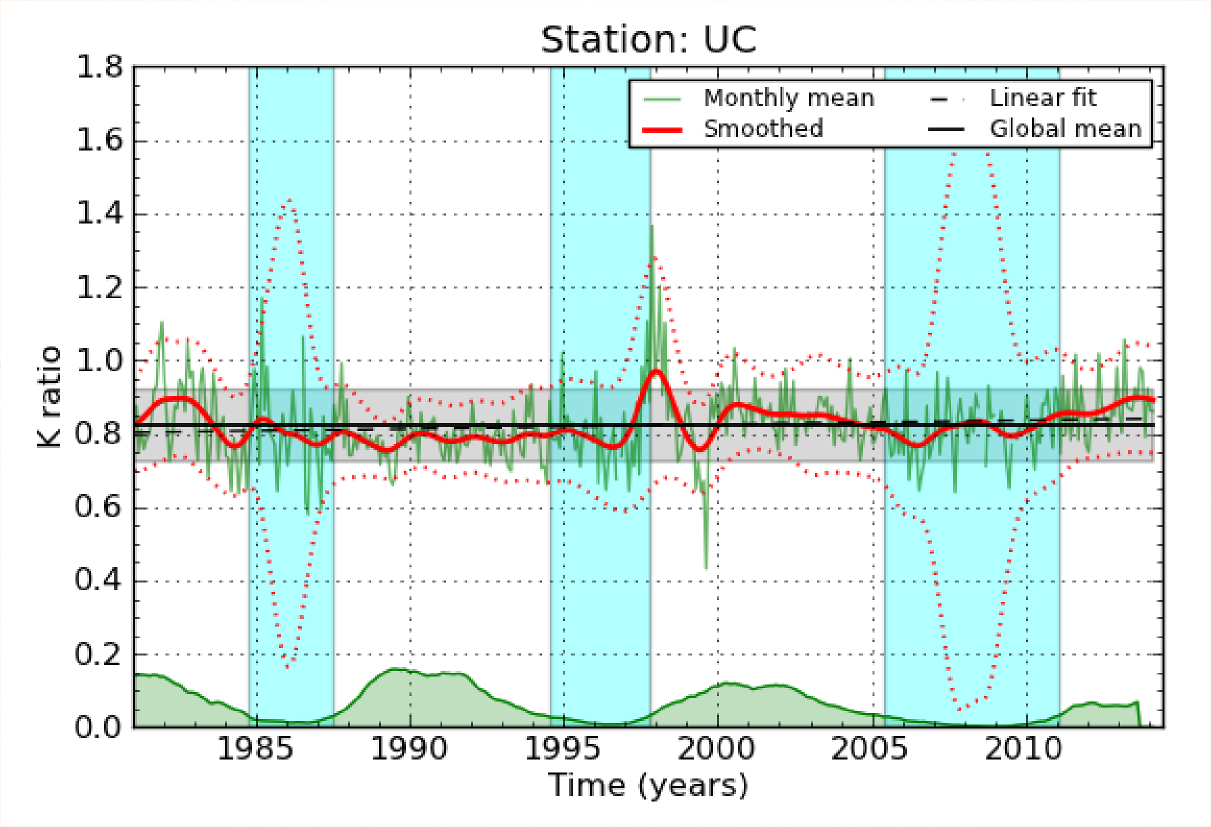
U. Mochizuki, Japan

K coefficient after correction: sample stations



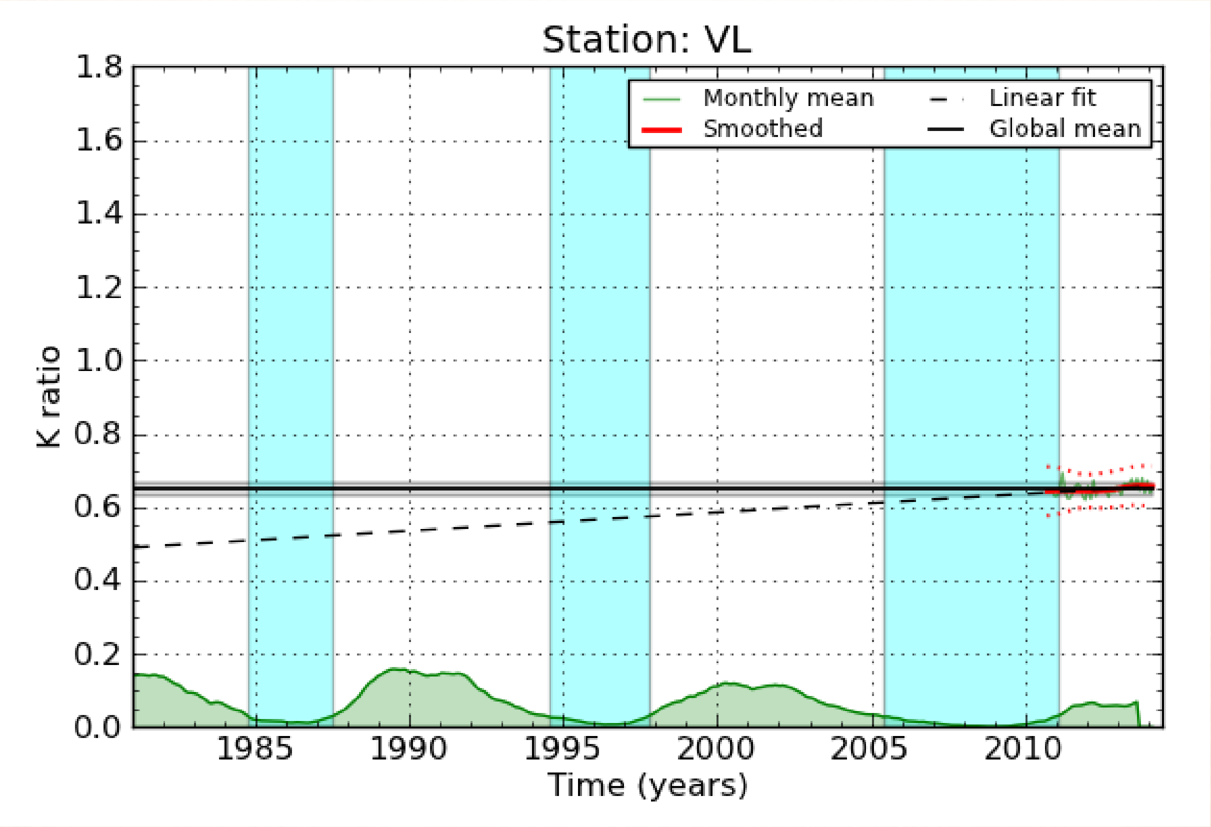
Nijmegen Public Observatory, Netherlands

K coefficient after correction: sample stations



Uccle, Belgium

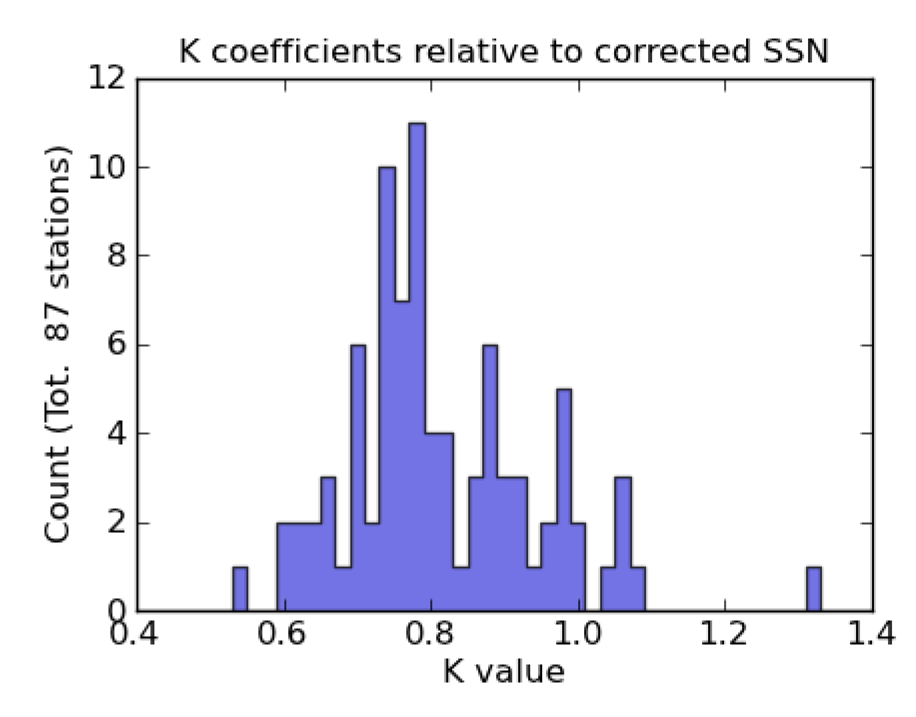
K coefficient after correction: sample stations



Valceresio de Bisuschio, Italy

K coefficients after corrections: distribution

- After correction, variations of k coefficients are reduced:
  - Constant k over full duration for many stations
  - The distribution of average k coefficients becomes meaningful
- Histogram of k coefficients for all stations with duration > 11 years (87 stations):

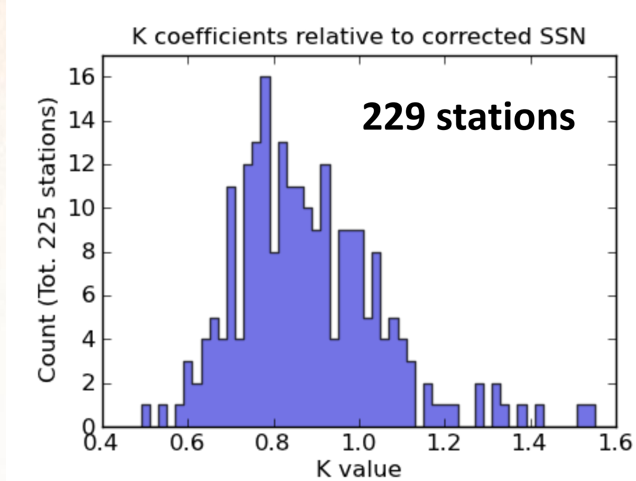


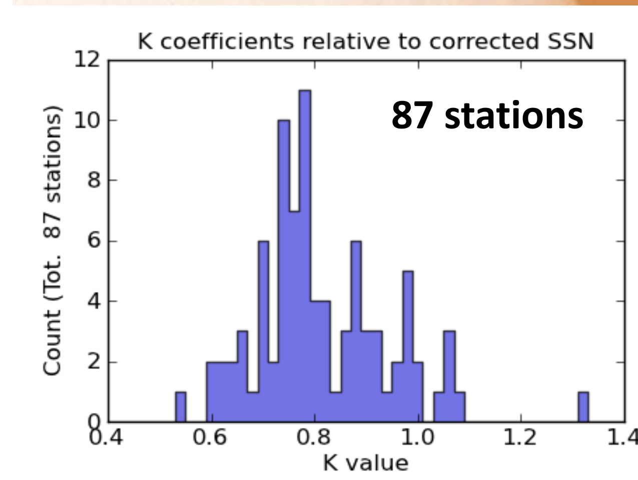
- Broad overall distribution (0.4 to 1.6) but …
- **Multiple discrete peaks**:

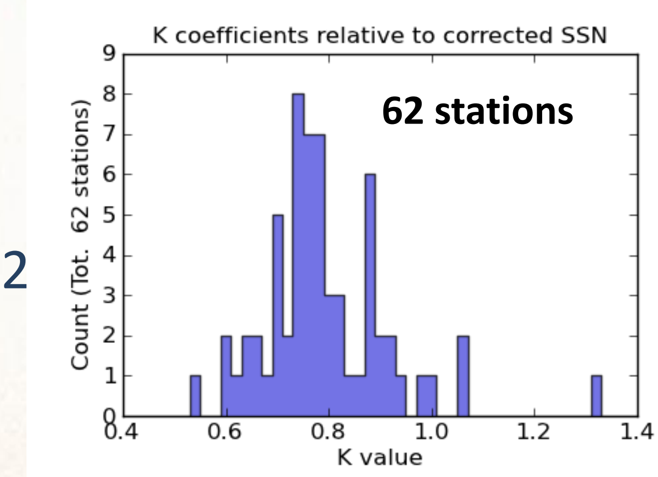
K vs R_i	K vs W_{LO}	Station group
0.6 – 0.65	1.0 – 1.1	Locarno
0.7	1.17	Mochizuki
0.75 – 0.82	1.25 – 1.37	Uccle, Kandilli, Catania, Fujimori, Learmonth
0.88	1.47	Ebro,Kandilli
0.98	1.63	Helwan, Pasternak
1.06	1.75	Culgoora

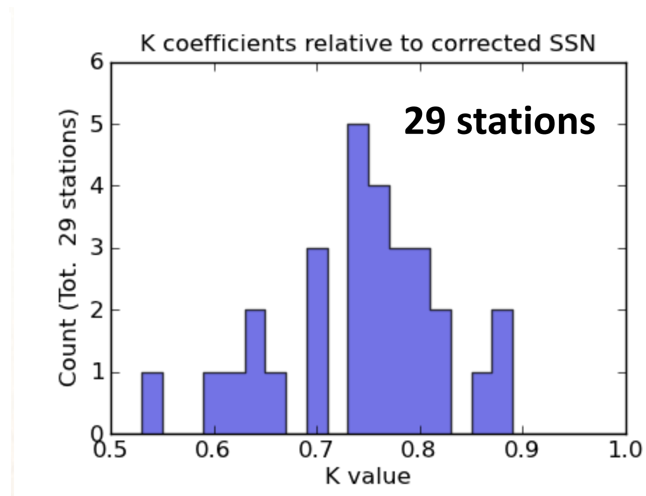
K distribution: dependence on quality

- Peaks merge only when all stations are included (low quality)
- Main peaks (0.6, 0.7, 0.75-0.82, 0.88) present for all other subsets
- Main range for high-quality stations: 0.75-0.82
- No good stations with k > 0.9







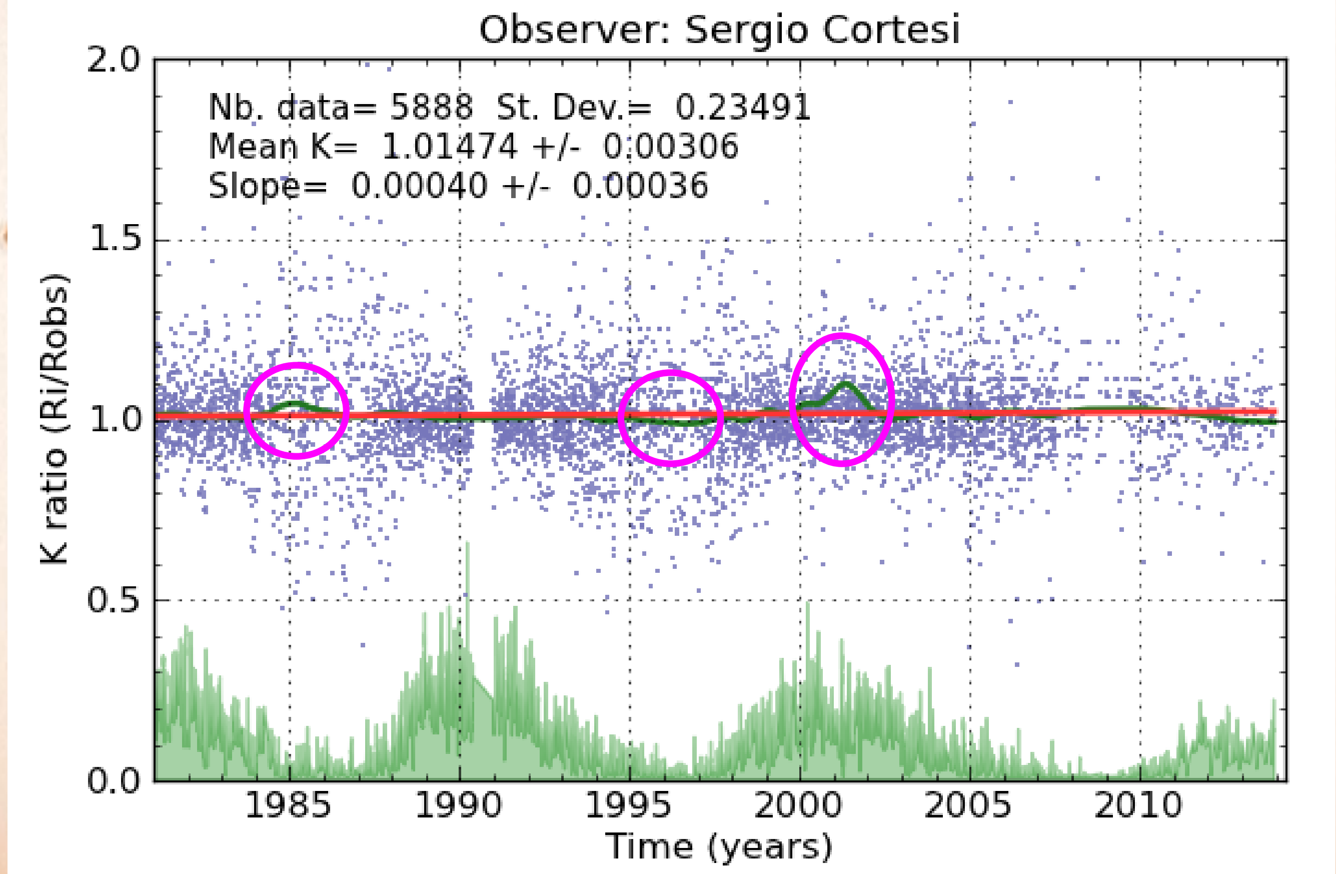


- k=0.72: equivalent to Locarno without weighted counts
- k=1.0: equivalent to Wolf’s original counts (small refractor)

Specola: personal k coefficients

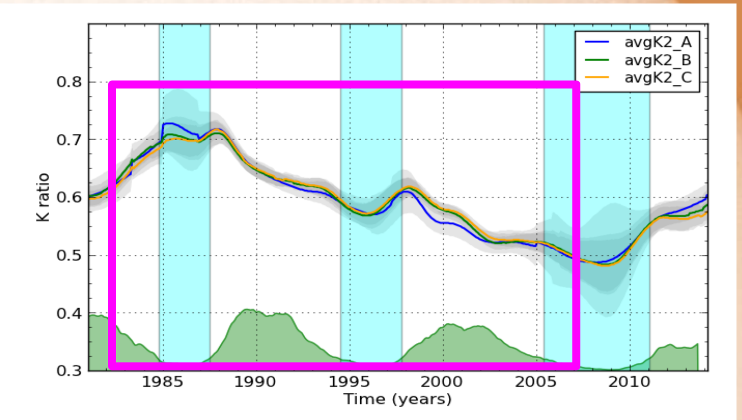
Locarno internal k coefficients: S. Cortesi

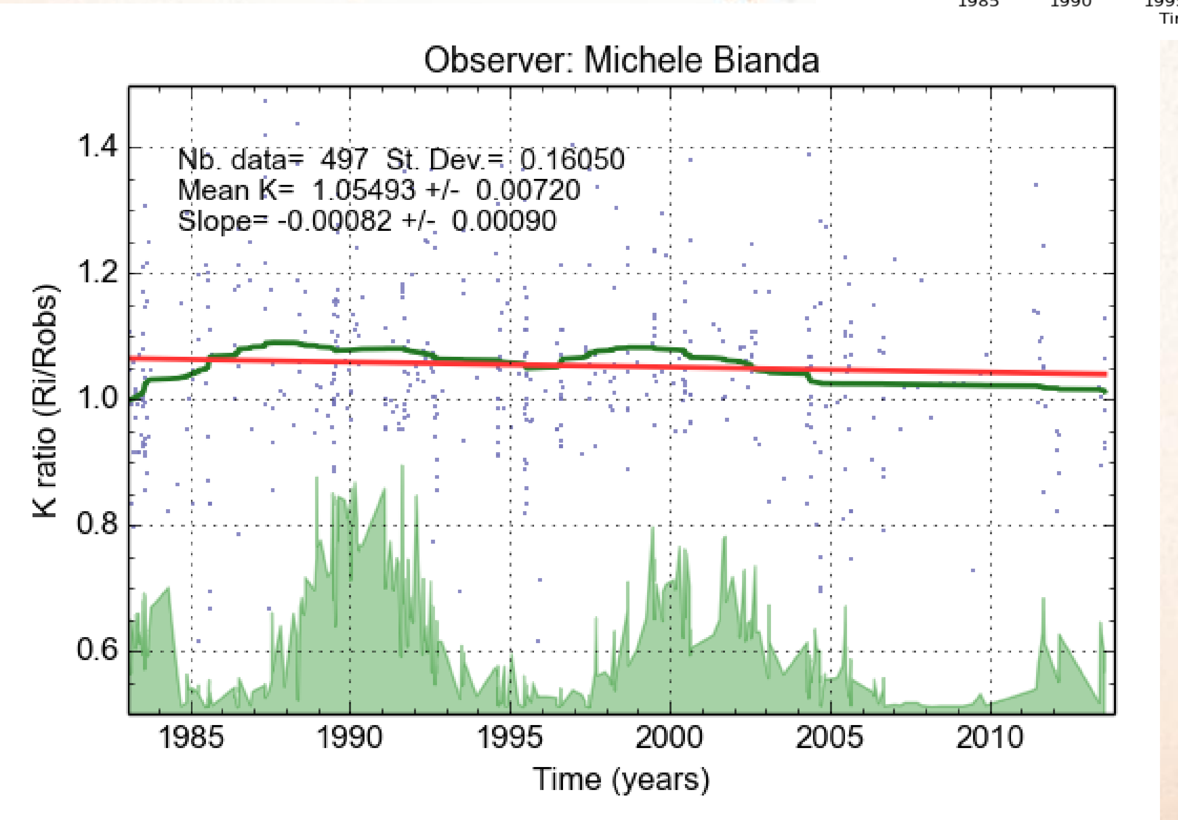
- Linear fit: no significant trend (cf. equivalence $W_{LO} = R_i$)
- 2-year running mean: only 3 temporary deviations
  - 1985, 1996: solar minima (non significant)
  - **2001: can only be explained by a flaw in the R_i calculation (SIDC, Brussels)**



Locarno internal k coefficients: M. Bianda

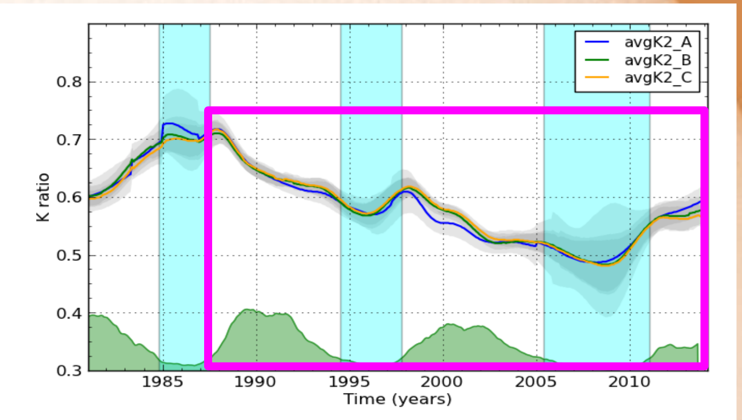
- Overall linear fit: negligeable trend
- Upward trend from 1983 to 1987
- Downward trend from 1987 to 2007
- Trend reversal from 1996 to 2000 (bump)
- Few values after 2007

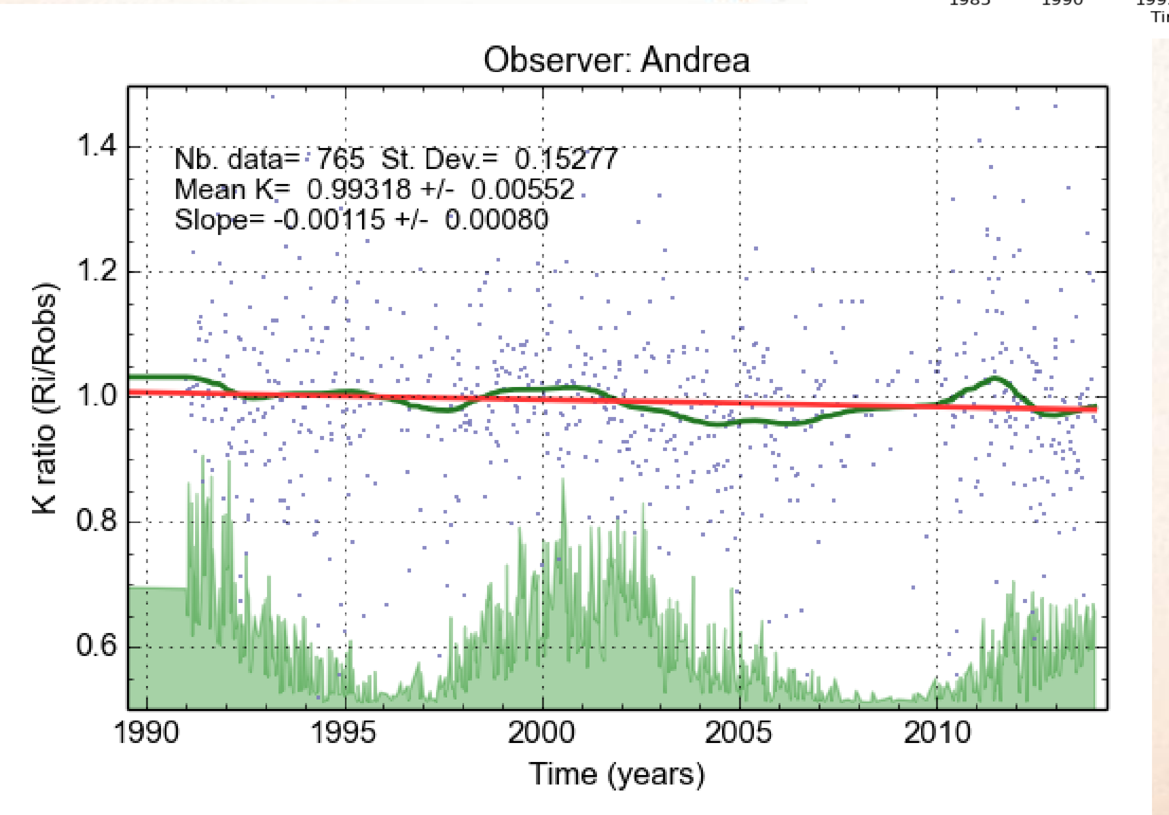




Locarno internal k coefficients: A. Manna

- Overal linear fit: no significant trend
- Downward trend from 1990 to 2005
- Upward trend from 2006 to 2012
- Positive bump in 2000





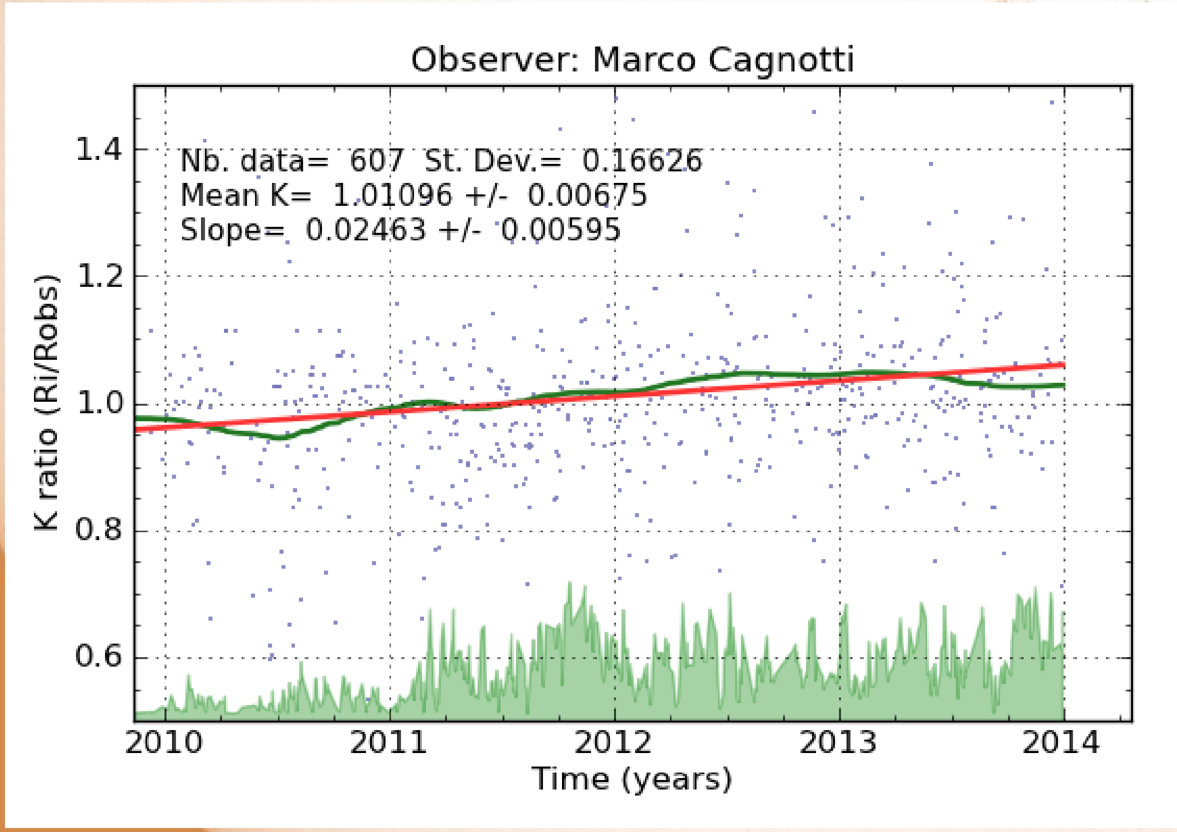
Locarno internal k coefficients: all observers

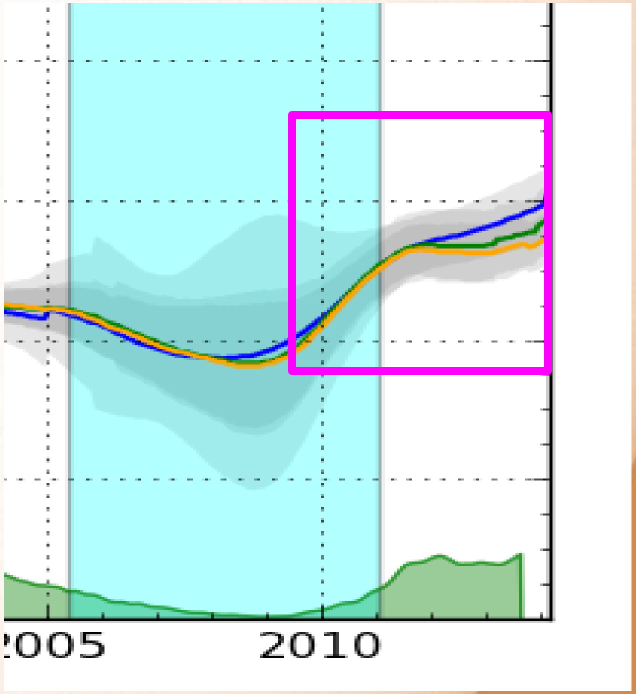
- Separate linear trends in short-duration observers:
  - Marginally significant
  - Match the SILSO average K (sign and slope value)
- S. Cortesi has the largest standard deviation: 0.234
  (0.10 to 0.16 for other observers)

Observer	Time interval	Slope	Error	Nb. Obs.	Std. Dev.	SILSO Avg. K slope
R. Cortesi	1981-1983	+0.008	0.014	148	0.105	+0.042
De Lorenzi	1981.5-1983	+0.009	0.036	8	0.051	+0.042
M. Cortesi	1983-1990	-0.004	0.006	387	0.212	-0.005
E. Alge	1995-2001	-0.010	0.017	24	0.138	-0.017
Pittini	1998 + 2001	-0.002	0.012	34	0.096	-0.017
E. Altoni	2002.5-2005.5	-0.018	0.018	43	0.103	-0.017
R. Ramelli	2004-2011	+0.014	0.013	34	0.146	+0.030
A.Cairoli	2010-2011	+0.044	0.058	8	0.055	+0.051
M.Cagnotti	2009.5-2014	+0.025	0.006	607	0.165	+0.030
	2010.5-2012.5	+0.050	0.006		0.165	+0.051
S. Cortesi	1981-2014	0.0004	0.0004	5888	0.234	
M. Bianda	1983-2014	-0.0008	0.0009	497	0.160	
A. Manna	1990-2014	-0.0012	0.0008	765	0.153	

The new primary observer: M. Cagnotti

- Significant upward trend: 0.025 +/- 0.006 yr^{-1}
- Trend is variable:
  - Sharp rise from 2010.5 to 2012.5
  - ~constant after 2012.5
- Matches the recent trend in the average k of the SILSO network

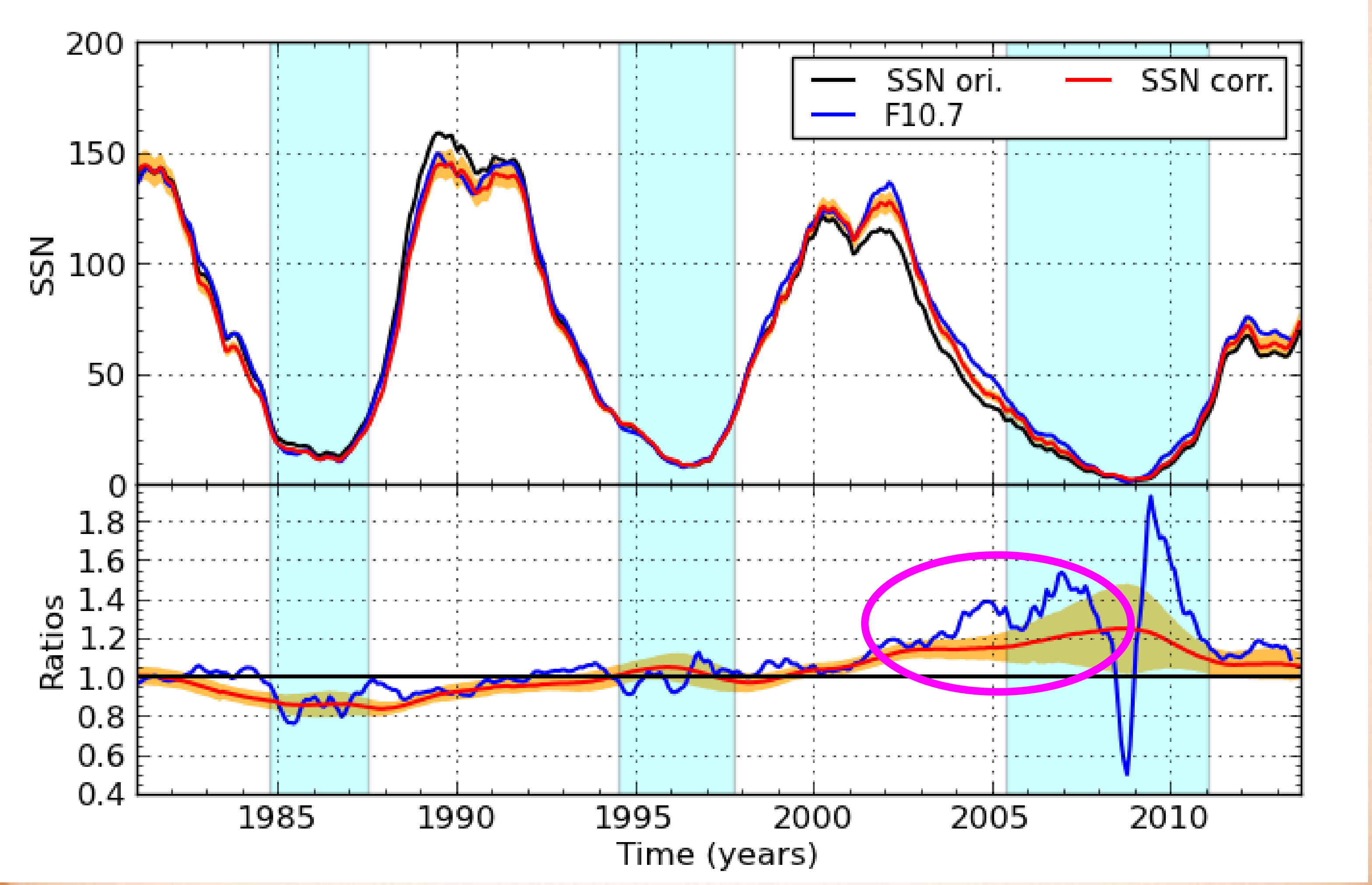




New SN versus other indices

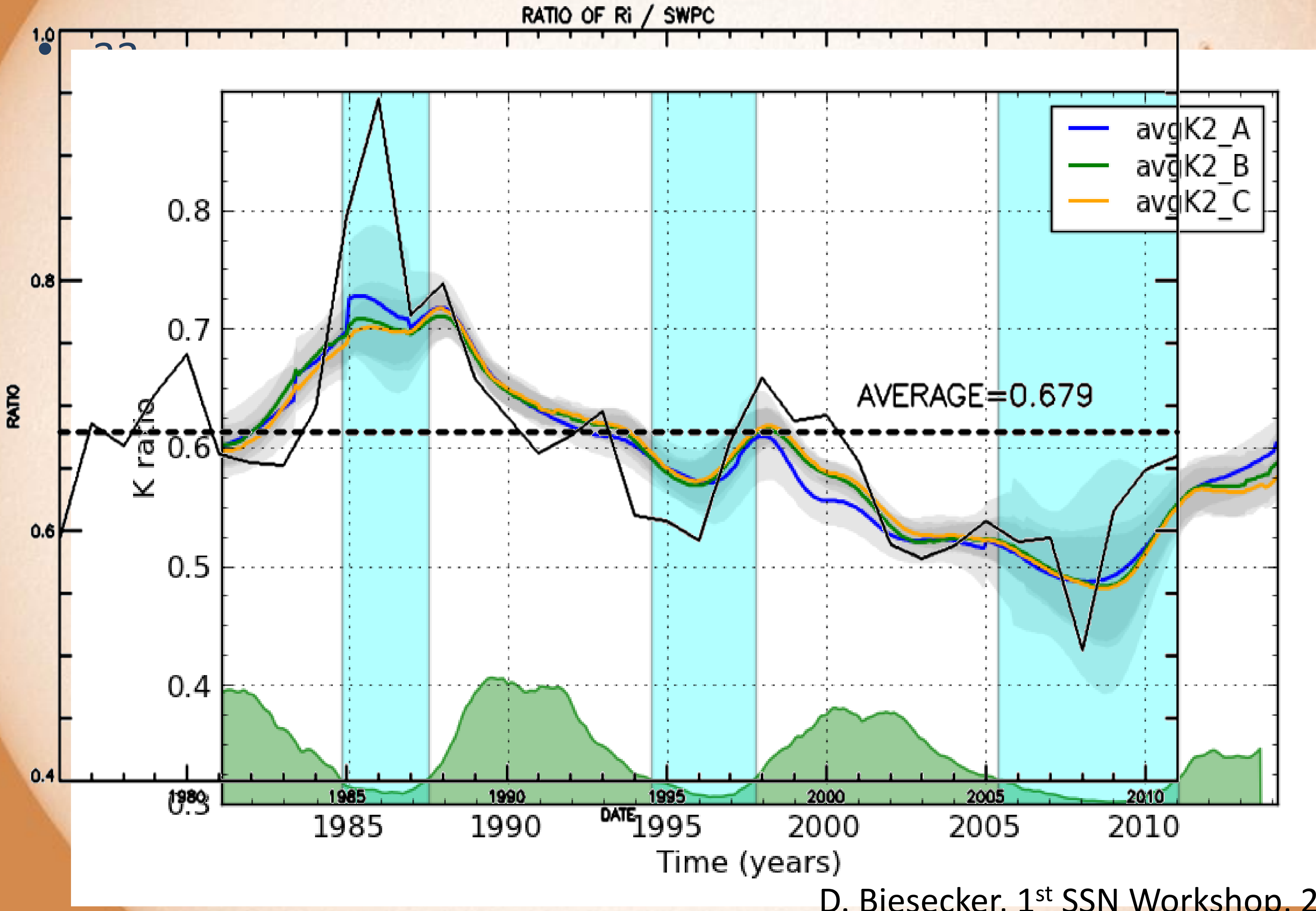
Cycle 23 anomalies: relation with $F_{10.7}$

- Discrepancies with other indices (R_A, $F_{10.7}$) are partly eliminated

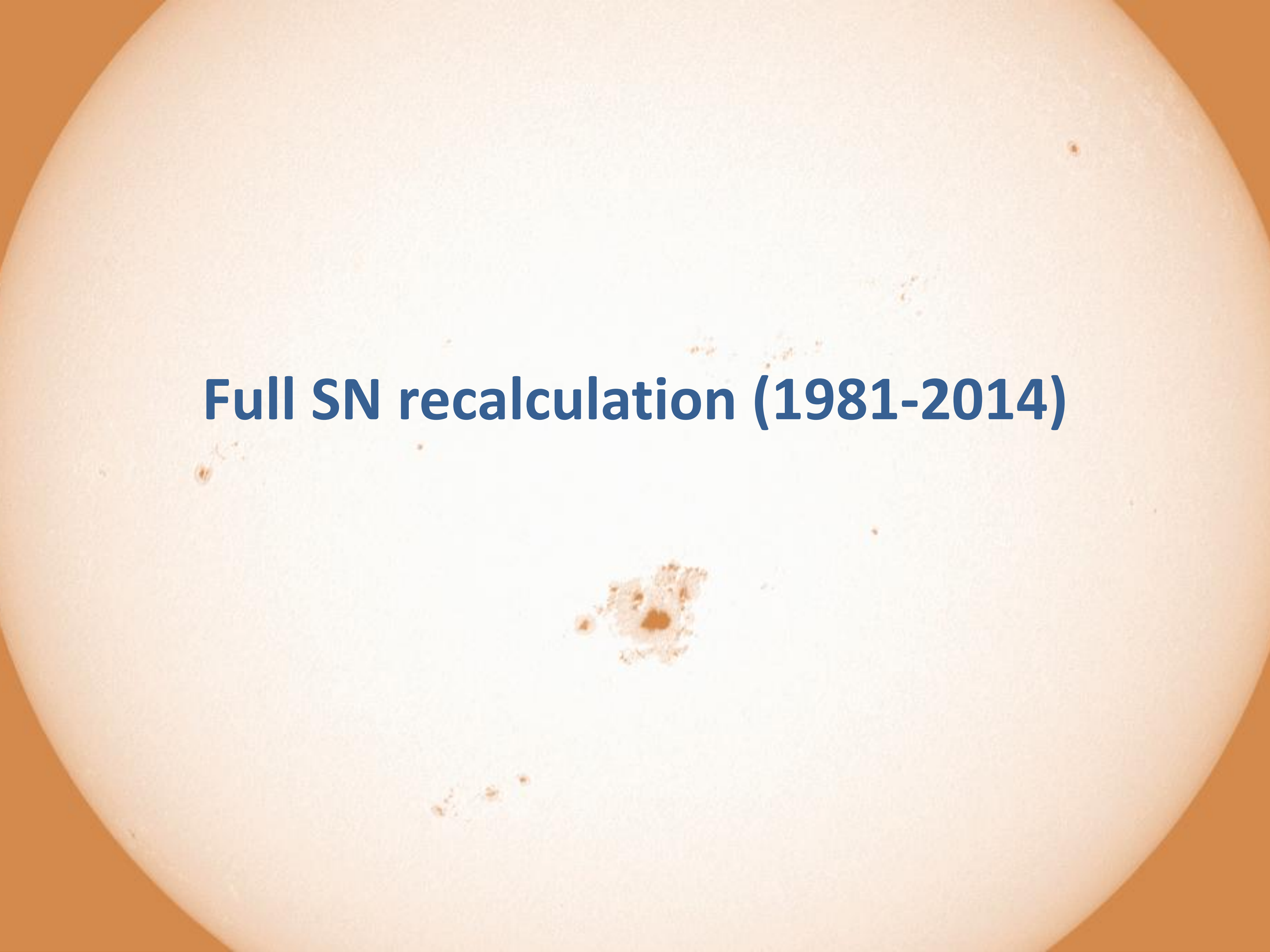


- **$F_{10.7}$ still deviates significantly after 2002 !**

Boulder sunspot Number: same trends



D. Biesecker, 1st SSN Workshop, 2011



Full SN recalculation (1981-2014)

Full recalculation of the SN

- Previous diagnostics obtained by applying the average k correction to the official R_i series
  - Only global trends are corrected (t > 2 years)
  - Daily values are only improved approximations
- A full recalculation is needed: the requirements
  - Replicate in detail the **original method** (heritage FORTRAN programs)
  - Allow **bulk calculation** of 34 years (400 months) + the classical monthly calculation
  - Direct data retrieval from the SILSO observation **database**
  - Allow choosing **any reference station**
  - Provide **additional information** about the processing (number of rejections, iterations, etc.)
  - Allow **selective (de) activation** of specific steps in the calculation (daily value rejection, k coefficient substitution, alternate « positive only » average, etc.)
  - **On-the-fly plotting** of the output
  - Multi-platform

⇨ New program : **Python language (Numpy, Scipy, Matplotlib libraries)**, >1000 lines of code

Full recalculation of the SN

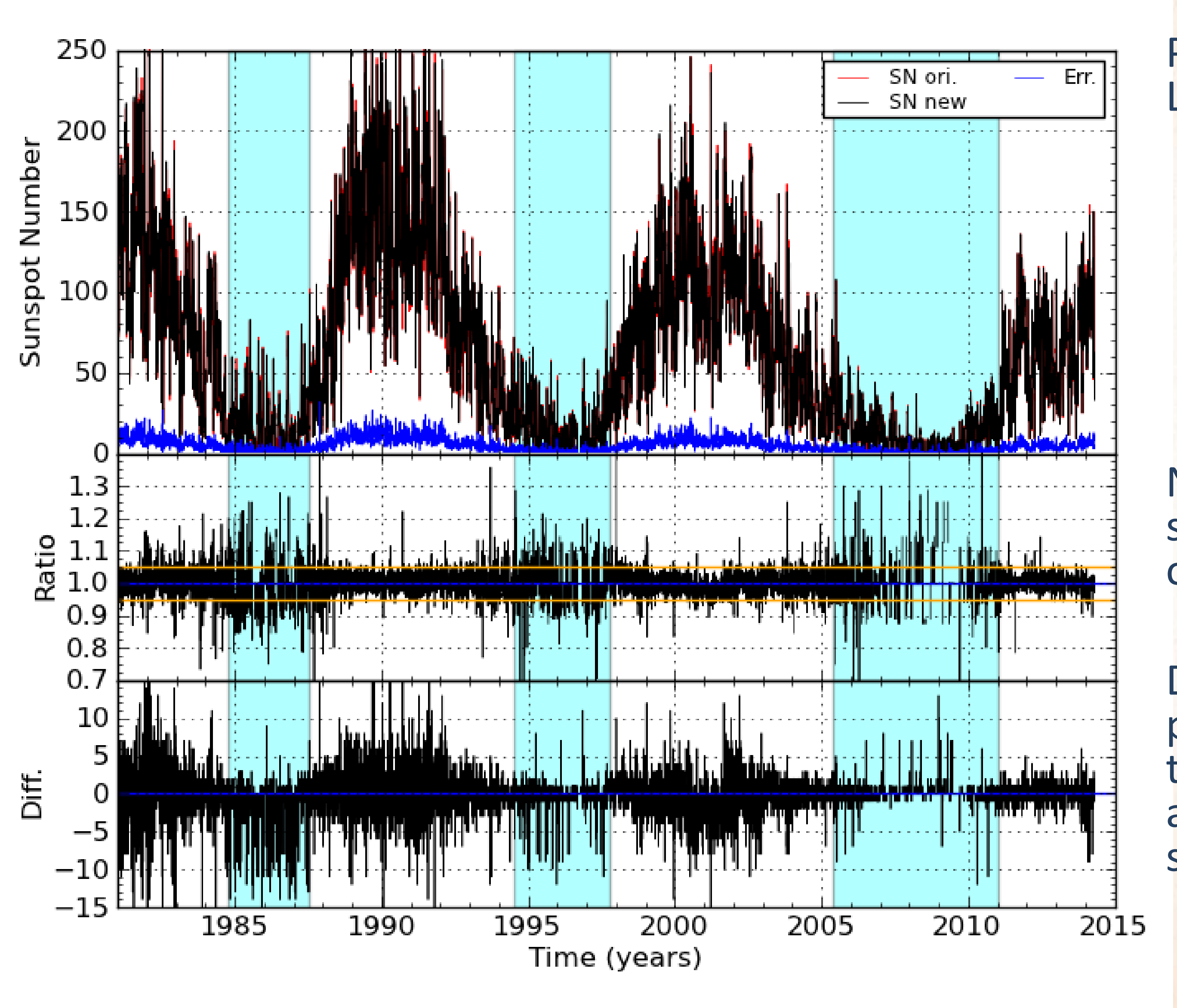
- New powerful tool for quality control and method validation
- Details of the calculation debugged and documented

- Caveats:
  - details of the calculations have probably changed over time (fine tuning)
    - K coefficients used for replacement
    - Thresholds for rejections and iterations
  - Calculation done on full database:
  - **No distinction between provisional and definitive datasets (5% tolerance).**

- Residual differences cannot be avoided on daily values

```
#=== Calculation of monthly mean k coefficients (First value) ==========
#  Elimination of stations for which a monthly k cannot be calculated
if selPrintMonth==True:
    print
    print 'K coefficients (initial values)'

nbKrepl=0
Krepl=[' ']*nbStaMon

for ista in range(0,nbStaMon) :
    if statValid[ista]>0 :
        N=statKnval[ista]
        if N>1 :
            sumV=0.;sumV2=0.
            for iday in range (0,lenMonth):
                if coefK[ista][iday]>0.:
                    sumV+= coefK[ista][iday]
                    sumV2+=coefK[ista][iday]*coefK[ista][iday]

            statKmon[ista]=sumV/N
            sumV2=(sumV2-(sumV*sumV/N))/(N-1)
            if sumV2<0. and sumV2>-1.e-10: sumV2=0.
            statKmonSD[ista]= mt.sqrt(sumV2)

            if selPrintMonth==True:
                print ('{:4d} {:02d} {:2s} | {:2d} {:3d} | {:5.3f} '+\
                       '{:5.3f} {:2d} {:3d} |').\
                       format(curYear,curMonth,statID[ista],\
                              statNobsVal[ista],statObsElim[ista],\
                              statKmon[ista],statKmonSD[ista],\
                              statKnval[ista],statKelim[ista])
        else:
#           If monthly mean k cannot be determined
#           (less than two daily values),
            if selKrepl==True:
#               If replacement mode, replace K with a past value
                for i in range(0,nbStats):
                    if statisID[i]==statID[ista]:
                        if pastK[i]>0:
#                           if valid k exists, makes the replacement
```

New SN vs official R_i: daily values

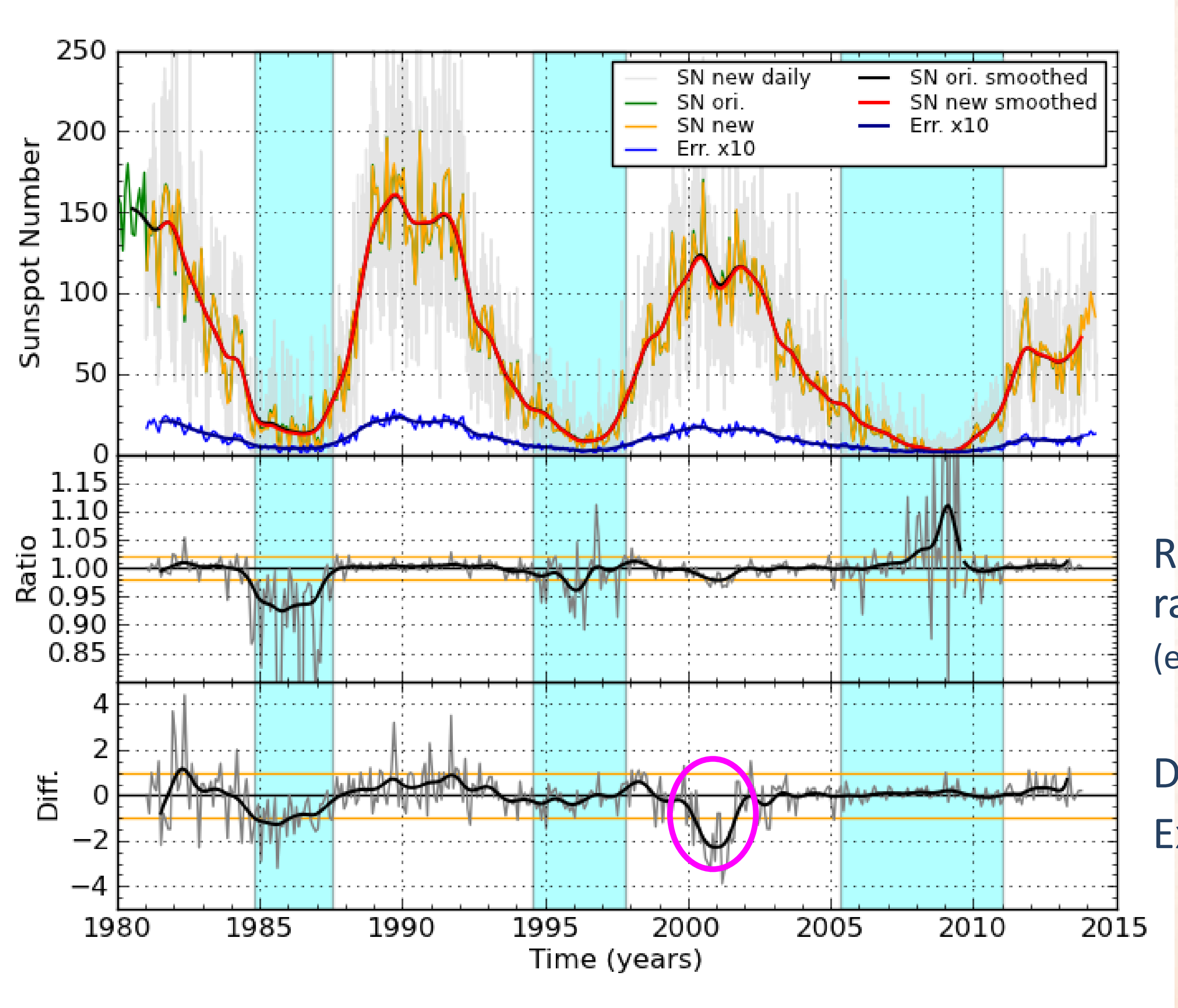


Pilot station: Locarno

New daily SN stays within 5% of original R_i

Differences proportional to the R_i value : always below the standard error

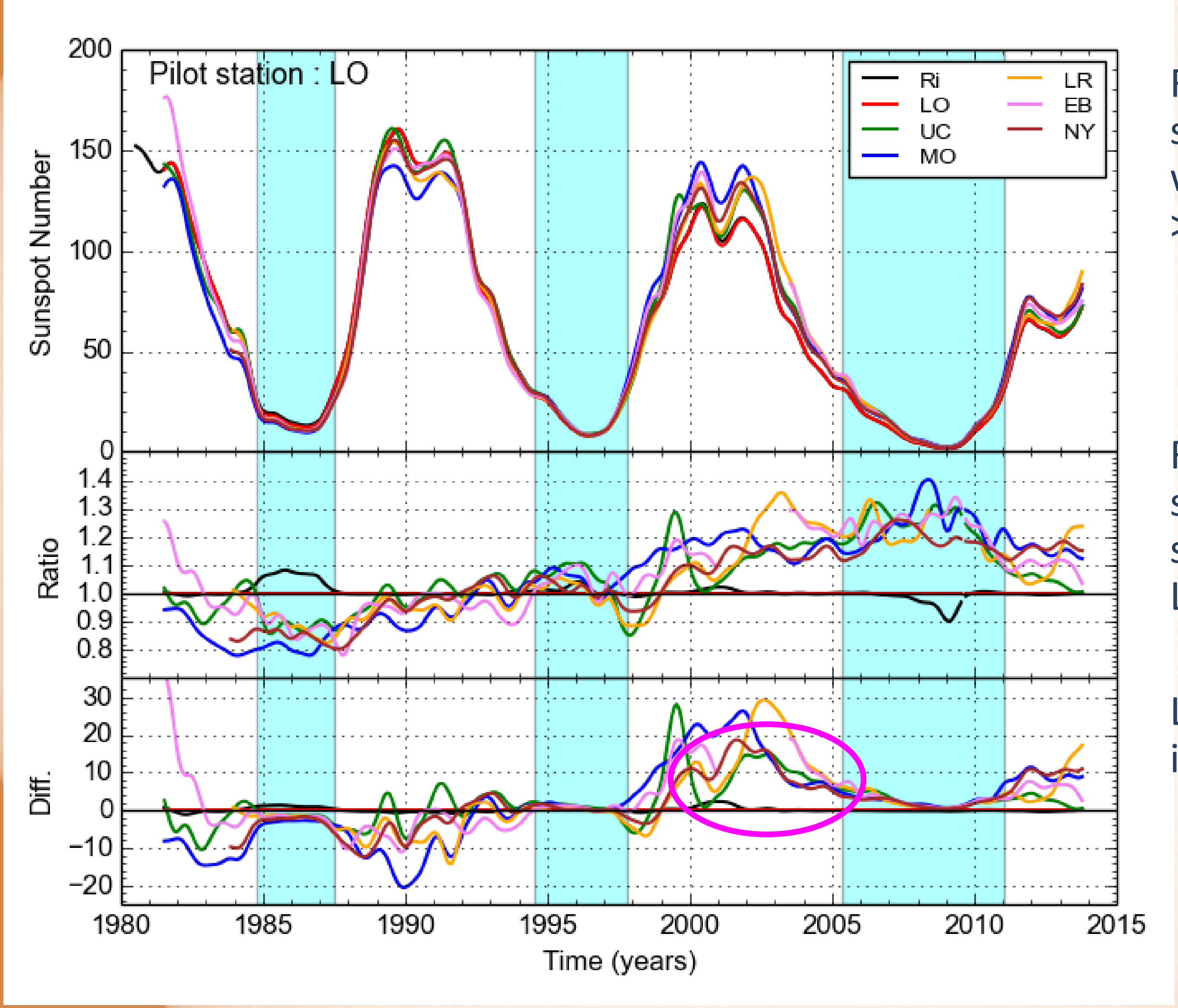
New SN vs official R_i: monthly mean values



RMS deviation in ratio < 1.5%
(except during minima)

Differences < 1 unit
Except 2000-2001
(cf. S.Cortesi k coefficient)

Multiple new SN series: other pilot stations

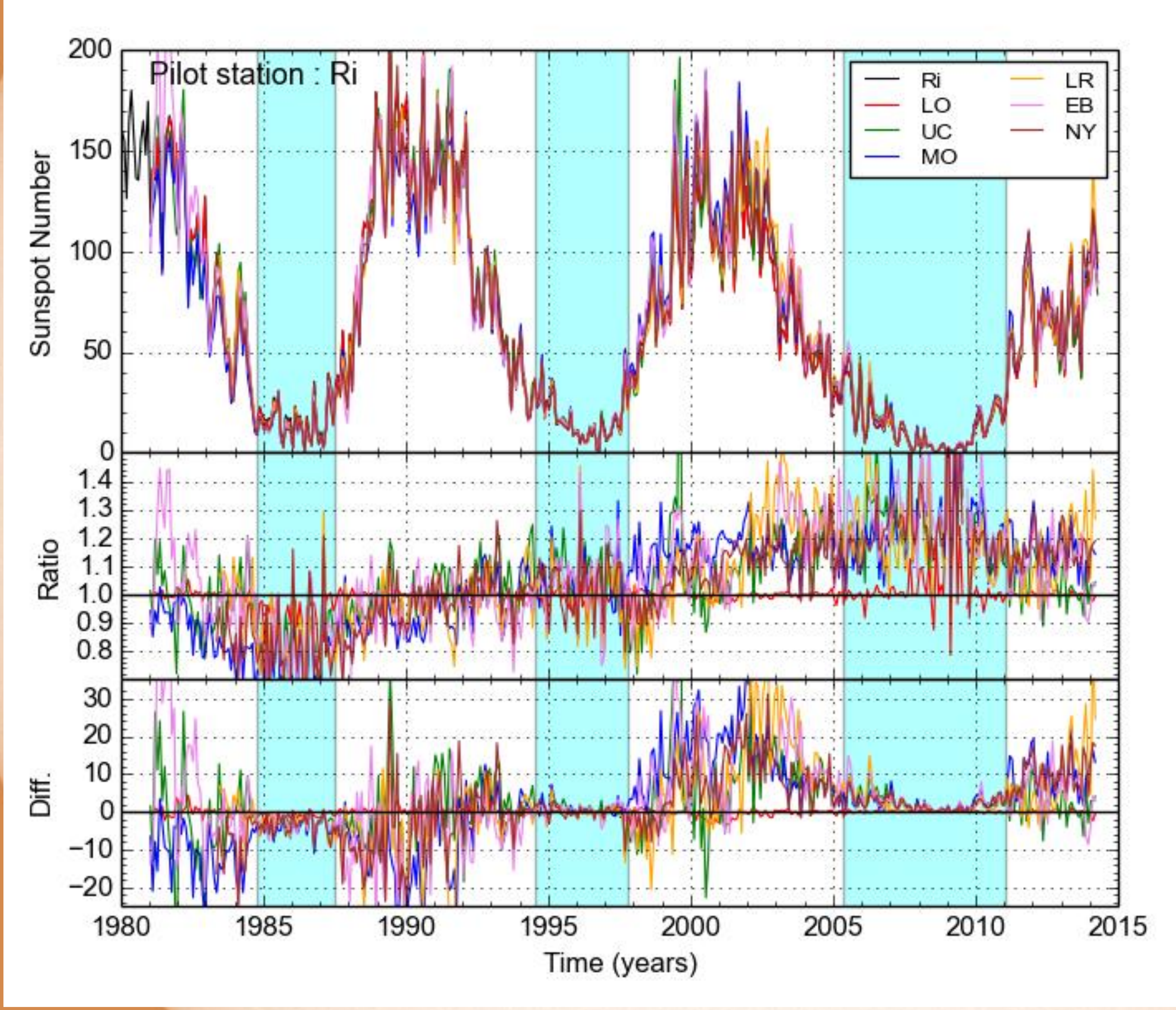


First run: only single stations with long duration > 25 years

Ratio: all new series show the same « S » drift vs LO

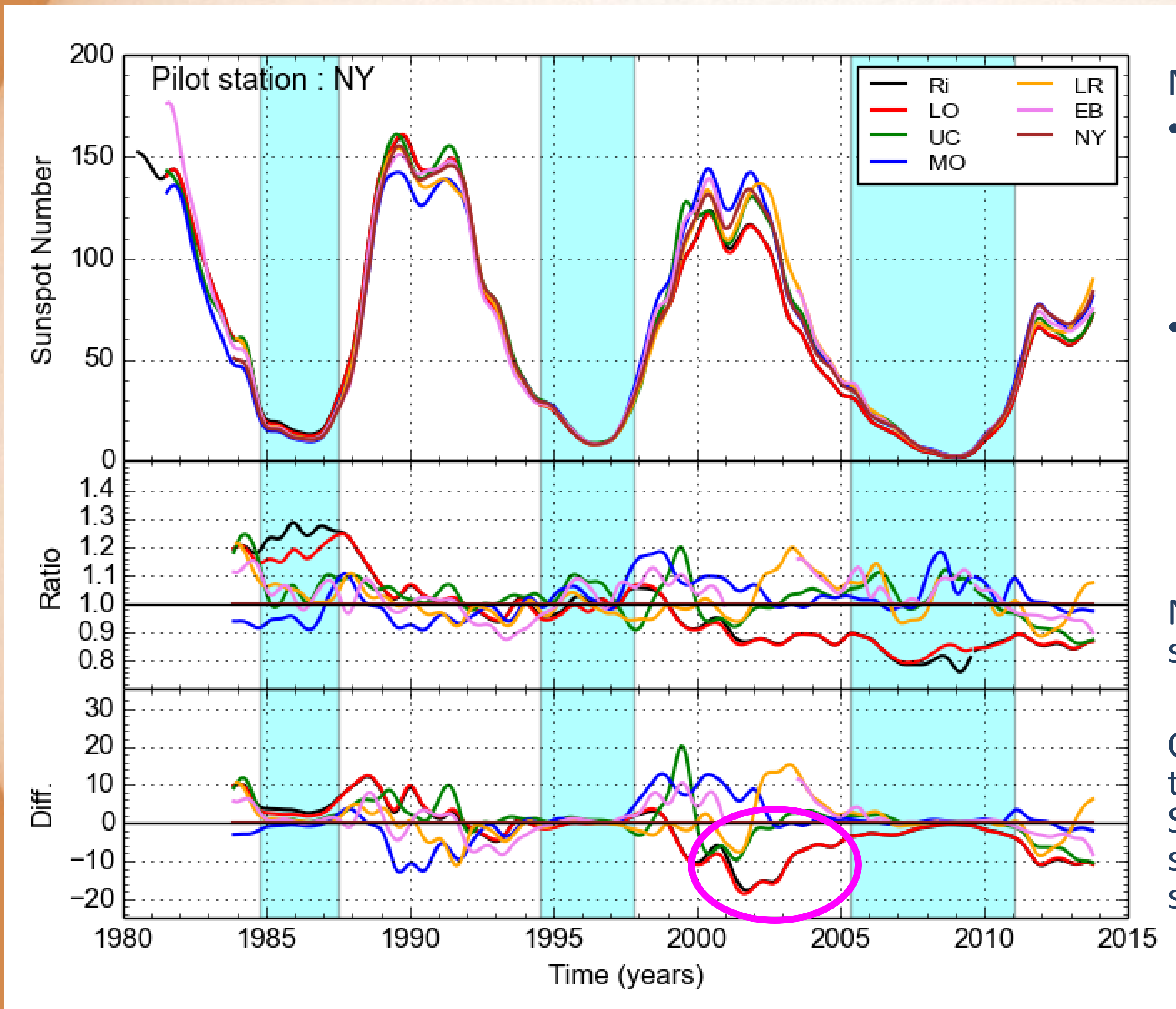
Largest differences in 2000 - 2004

Multiple new SN series: other pilot stations



Monthly mean SNs

Multiple new SN series vs NY (Nijmegen)



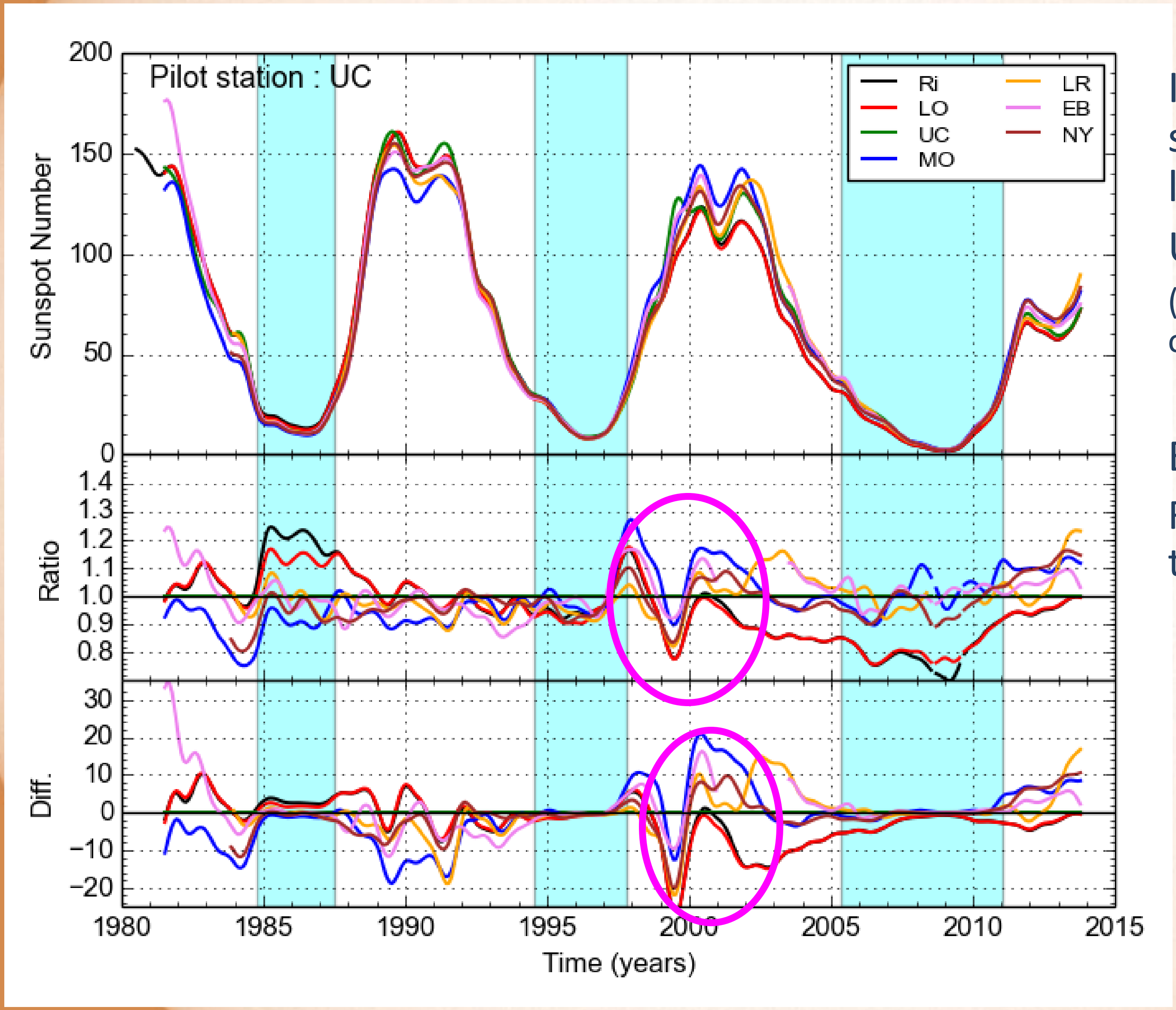
Main differences:

- **Cycle 23 raised: Cycles 22 and 23 have almost the same maximum activity**
- **Double maximum: Second peak raised versus first peak (peaks equal or second peak higher in cycle 23)**

No more trend for all series.

Only the orginal R_i and the new Locarno-based SN deviate systematically in the same way

Multiple new SN series vs UC (Uccle)



Individual stations can have local anomalies:

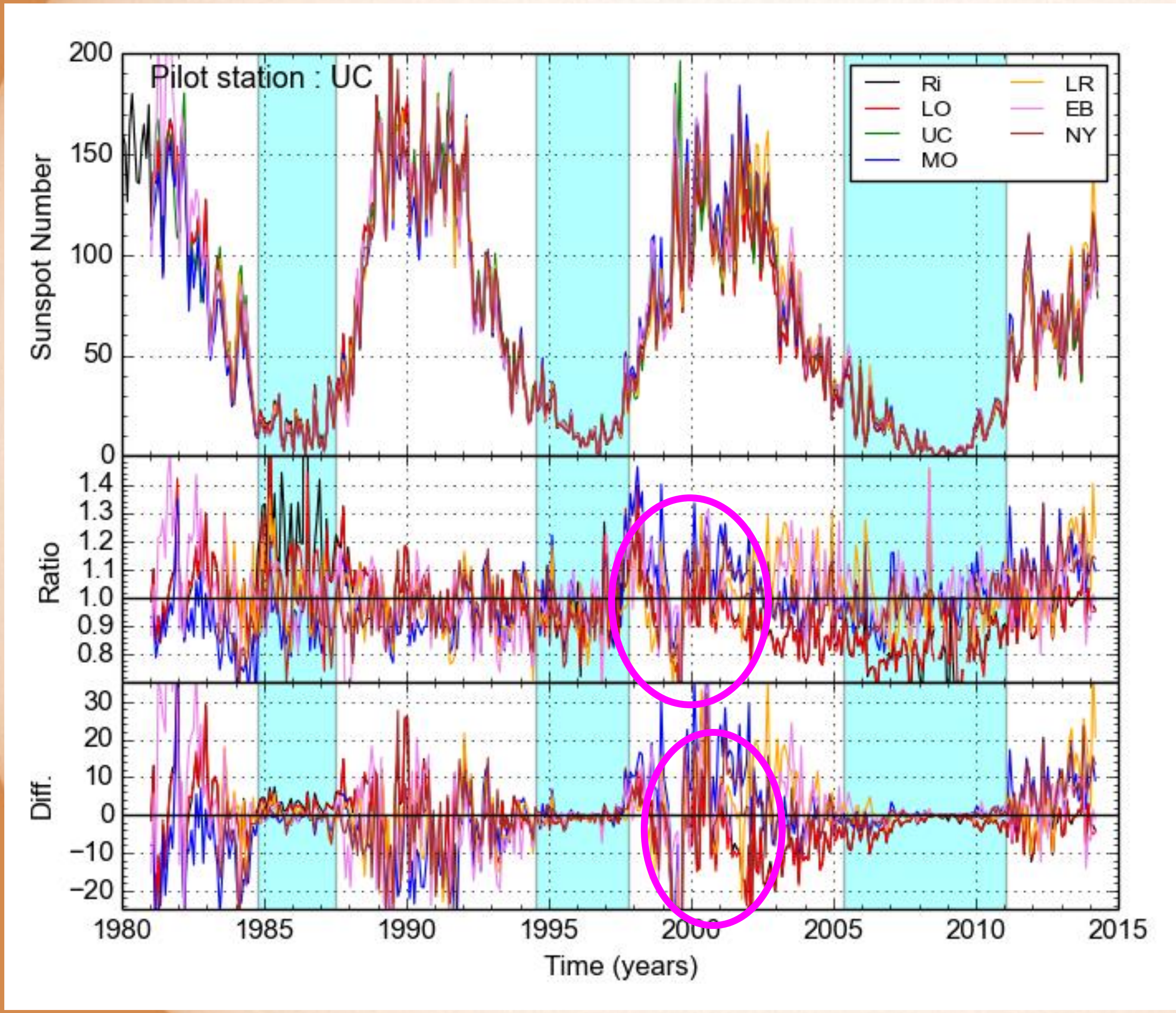
Uccle, 1998-2001

(presence of a bad observer)

Easy to identify:

Feature unique to that station

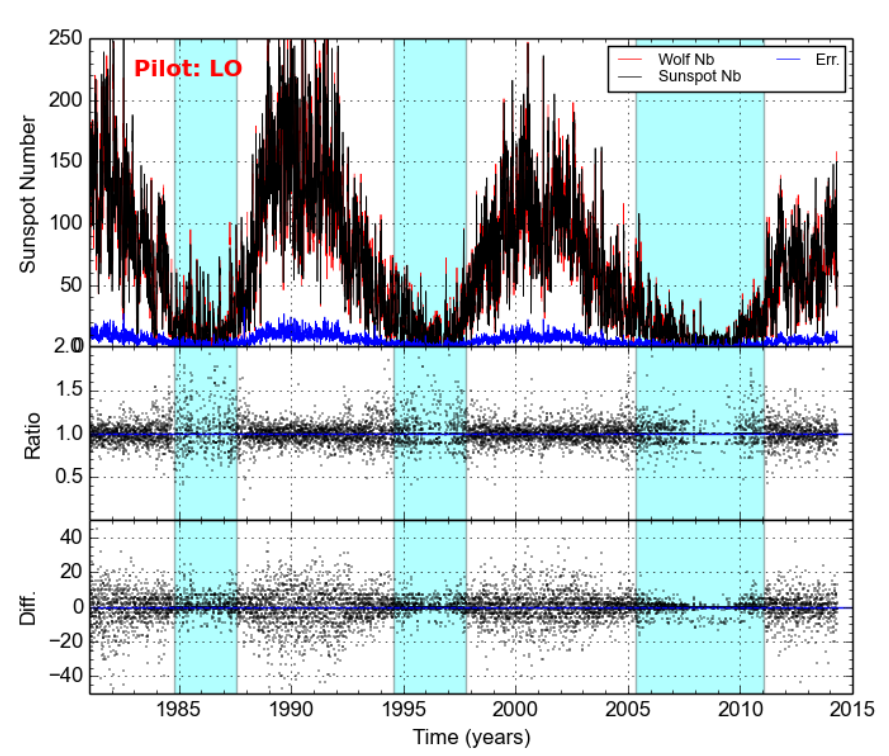
Multiple new SN series vs UC (Uccle)

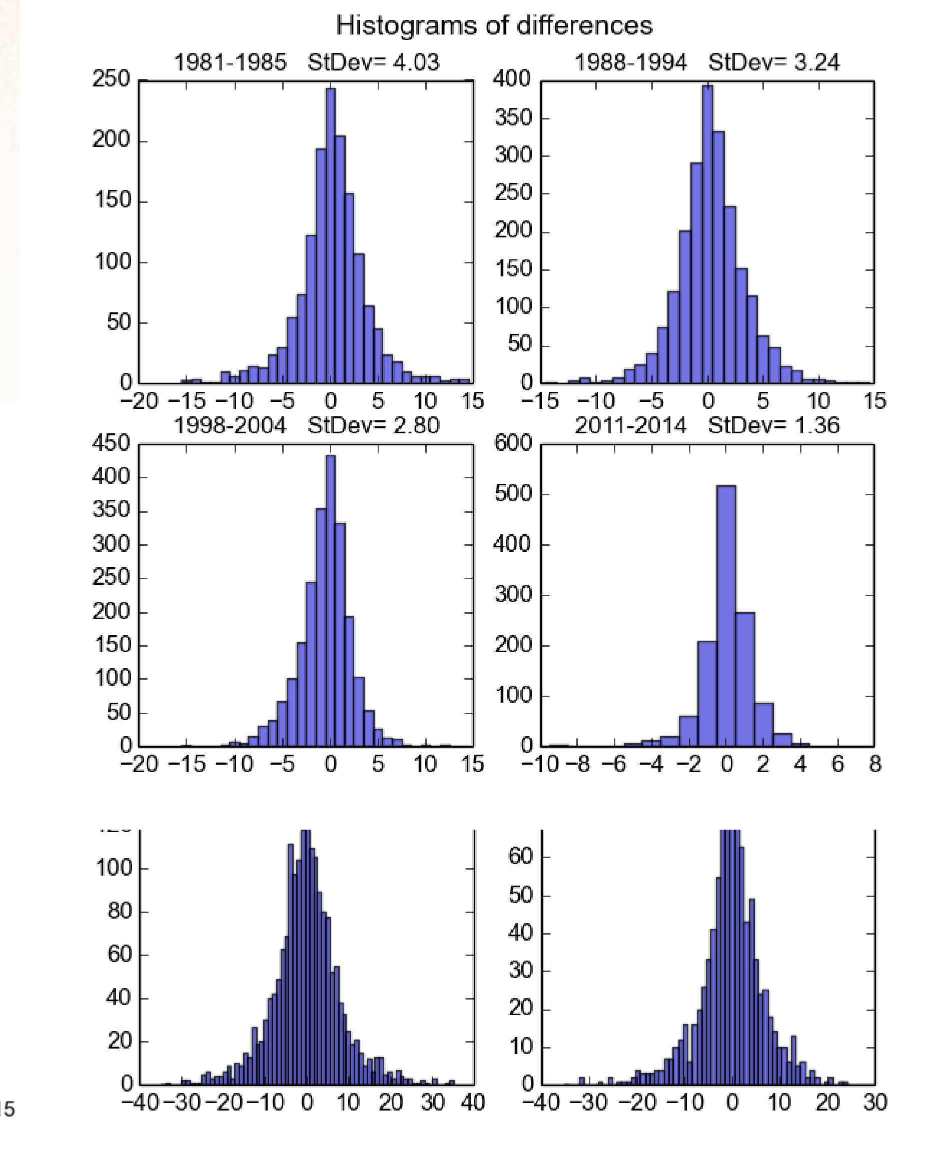


Monthly mean SNs

Raw pilot Wolf numbers vs resulting SN

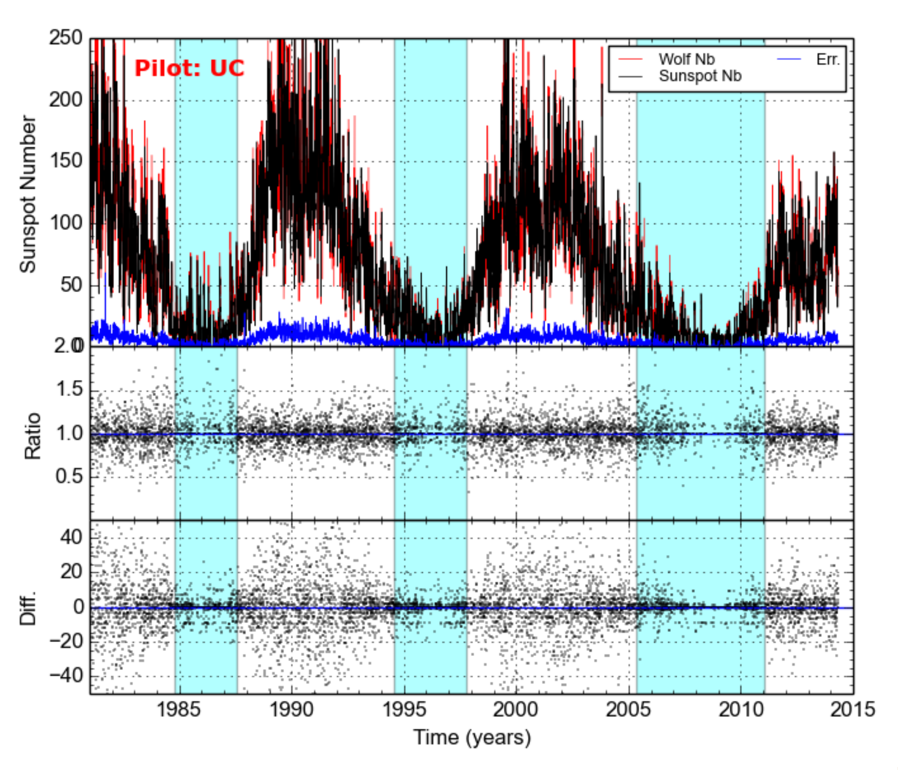
- New SN tracks the Wolf number of the pilot station
- **Locarno:** daily rms difference ~8 SN units (Poisson distribution)
- Daily rms difference old and new LO-based SN: 3 SN units

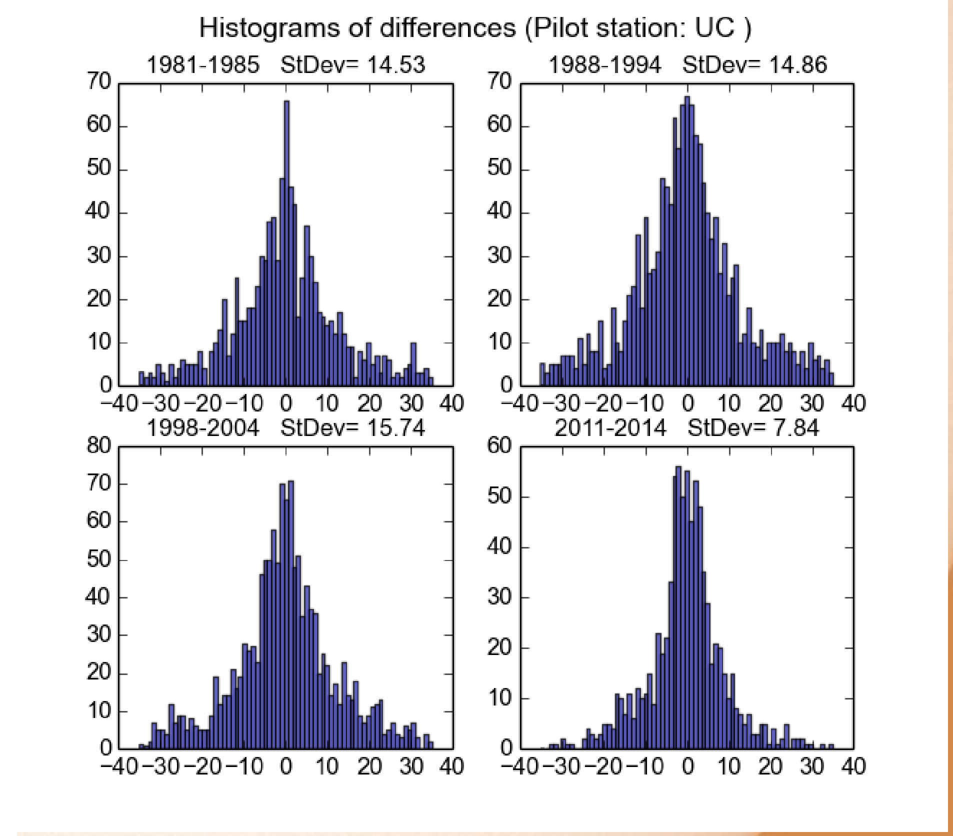




Raw pilot Wolf numbers vs resulting SN

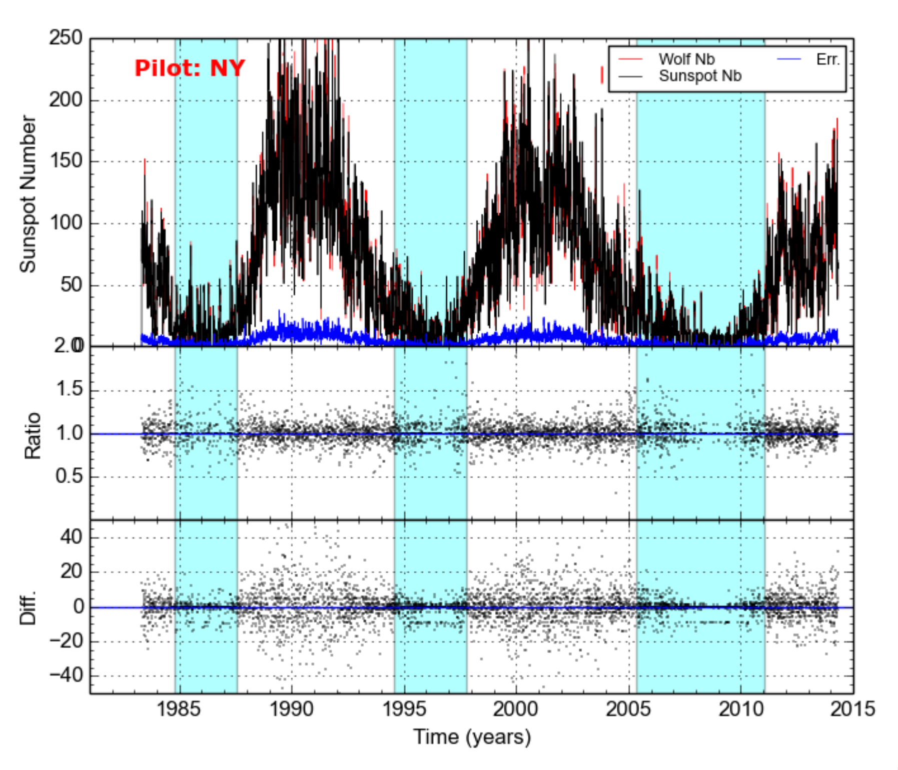
- **Uccle**: daily rms difference :
  ~ 15 units (10%)
- Professional observatories have multiple observers

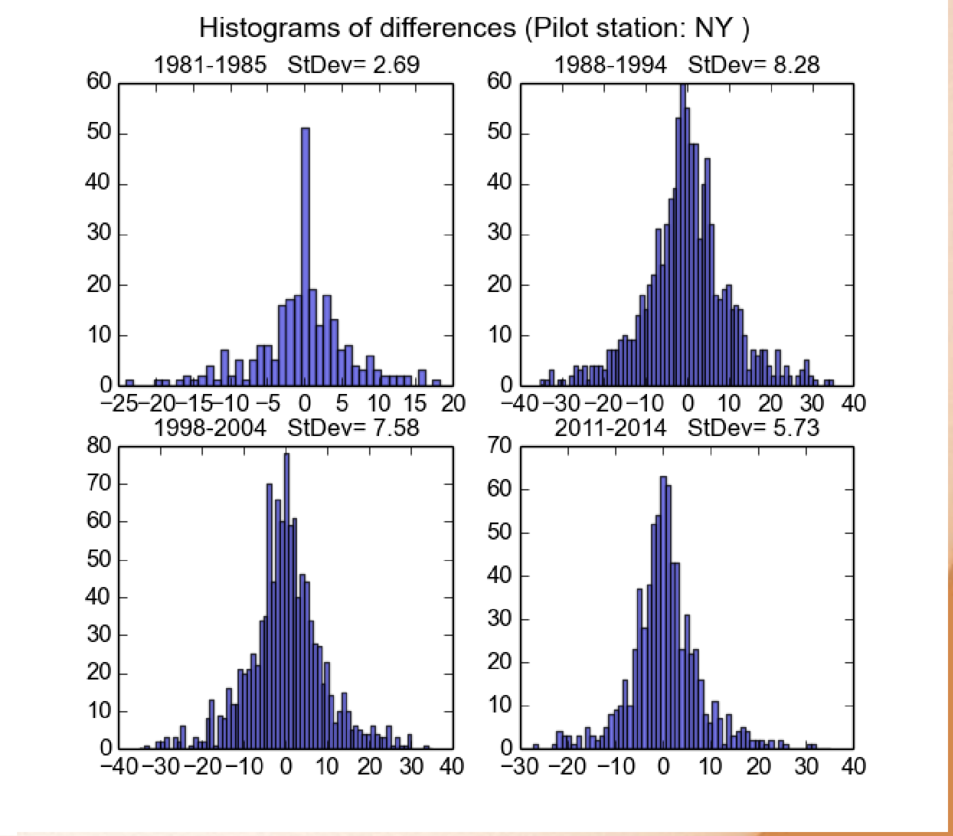




Raw pilot Wolf numbers vs resulting SN

- **Nijmegen**: single observer
- Daily rms difference :
  8 SN units (5%)



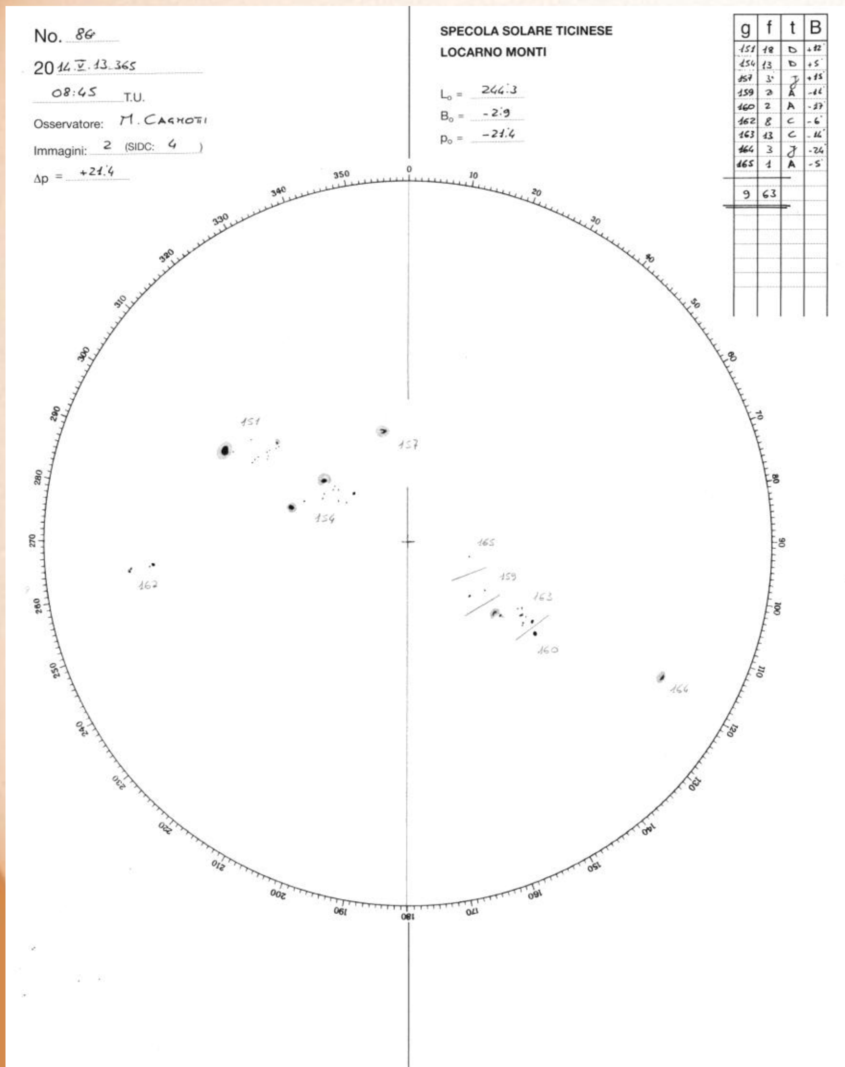


Conclusions and plans

Conclusions

- **The Brussels SN number faithfully tracked variable drifts in the Specola Wolf number**
- Various diagnostics:
  - Worldwide SILSO network
  - Parallel indices
  - Personal k coefficients of Specola Observers:
    - Bad news: deviations affected the numbers from the main observer
    - ⇨ **Good news: the new primary observer (M. Cagnotti) has a stable k**
- Recomputation of the Brussels SN:
  - All original data preserved (fully verified !)
  - New program: designed as a **versatile tool** for:
    - Continuous quality control
    - Statistical understanding of the method and the input data
  - ⇨ Program validated: no significant differences between equivalent calculations … except:
    - Temporary shift in 2000-2001 (2 SN units)

Conclusion: the next steps



- Revisiting the past Specola observations: reconstructing a new Locarno Wolf number series:
  - Combining series of individual secondary observers
  - Exploiting the Locarno sunspot drawing collection:
    - New unweighted counts based on Specola drawings
    - High quality drawings:
      - Extends the Zürich drawing series (1880-1980)
      - close reproduction of the projected image
      - Limited interpretation (weighting, lifetimes, etc.)

Choosing a new reference: the options

1. Revised Locarno Wolf numbers:
   - Requires an understanding of the internal k coefficient variations
   - Time needed for recounting from the drawings
   - Advantages:
     - Same reference as before (connection to Zürich)
     - Stable permanent observatory
2. Swtiching to another pilot station (temporarily? Permanently?)
   - A proper station must be identified: criteria provided by current global analysis:
     - Time coverage: Duration, continuity, duty cycle
     - Precision: daily rms dispersion (small, constant with time)
     - Accuracy: absence of long-term trend and short-term deviations
     - Commitment: guarantee of future continuation of observations
   - Advantage: more stable reference (immediately available)
   - Drawback: dependency on a single reference data source.

Choosing a new reference: the options

3. Multi-station reference:
   - **Principle: average of Wolf numbers from a core subset of good stations.**
   - Proper stations must be identified: same criteria as for single station

KZ	Kanzelhohe	A		395	0.7245	0.0795	11.0%	0.0007 +/- 0.0004	7893	Professionals
LO	Locarno	SW		394	0.6246	0.0736	11.8%	0.0057 +/- 0.0003	8683	Professionals
UC	Uccle	B		392	0.8223	0.0979	11.9%	0.0011 +/- 0.0005	7071	Professionals
KH	Kandilli	TK		383	0.8782	0.1277	14.5%	-0.0030 +/- 0.0007	7153	Professionals
EB	Ebro	ES		370	0.8803	0.1259	14.3%	-0.0009 +/- 0.0007	7762	Professionals
LR	Learmonth	AU		359	0.7508	0.1030	13.7%	0.0004 +/- 0.0006	7918	Professionals
KS	Kislovodsk	RU		333	0.7823	0.0916	11.7%	0.0026 +/- 0.0006	7716	Professionals
MO	U. Mochizuki	JP	A	393	0.6948	0.0767	11.0%	-0.0029 +/- 0.0004	7084	Amateurs
NY	Nijmegen	NL	A	364	0.8208	0.0845	10.3%	-0.0016 +/- 0.0005	4860	Amateurs
DU	F. Dubois	B	A	348	0.7614	0.0851	11.2%	-0.0021 +/- 0.0005	6171	Amateurs
SO	Sobota	SK	A	260	0.7062	0.0813	11.5%	0.0039 +/- 0.0007	4925	Amateurs
JB	J. Bourgeois	B	A	242	0.7507	0.0897	11.9%	0.0021 +/- 0.0009	3204	Amateurs
HP	Presov	SK	A	236	0.7530	0.0856	11.4%	0.0044 +/- 0.0009	3515	Amateurs
GA	G. Araujo	ES	A	160	0.6903	0.0665	9.6%	-0.0015 +/- 0.0013	3408	Amateurs
LM	L. Meeus	B	A	97	0.7434	0.0744	10.0%	-0.0066 +/- 0.0030	783	Amateurs
JY	J. Carels	B	A	95	0.7464	0.0832	11.1%	-0.0040 +/- 0.0036	1051	Amateurs
TH	T. Teague	UK	A	87	0.8967	0.1334	14.9%	-0.0031 +/- 0.0050	239	Amateurs
DM	Deman	B	A	85	0.7471	0.0838	11.2%	-0.0047 +/- 0.0008	726	Amateurs
GB	A. Gabriel	B	A	84	0.7451	0.0514	6.9%	-0.0097 +/- 0.0024	1478	Amateurs
BT	T. Broxton	UK	A	67	0.8030	0.0928	11.6%	-0.0077 +/- 0.0065	679	Amateurs
BW	A. Bowyer	UK	A	61	0.9475	0.1000	10.6%	0.0136 +/- 0.0082	510	Amateurs
BV	B.Vanslooten	NL	A	52	0.7777	0.0780	10.0%	0.0122 +/- 0.0081	988	Amateurs
WW	W. Wilson	USA	A	45	0.7959	0.0681	8.6%	-0.0143 +/- 0.0089	963	Amateurs
VL	Valceresio	IT	A	42	0.6484	0.0176	2.7%	0.0050 +/- 0.0025	1018	Amateurs
MQ	J. Quesada	ES	A	38	0.6054	0.0767	12.7%	-0.0038 +/- 0.0132	1013	Amateurs
JV	J. Alonso	ES	A	35	0.7603	0.0594	7.8%	-0.0218 +/- 0.0109	566	Amateurs
IU	Istanbul	TK	A	33	0.5997	0.0617	10.3%	0.0089 +/- 0.0130	804	Amateurs
ZF	T. Friedli	SW	A	22	0.6177	0.0802	13.0%	-0.0277 +/- 0.0302	139	Amateurs

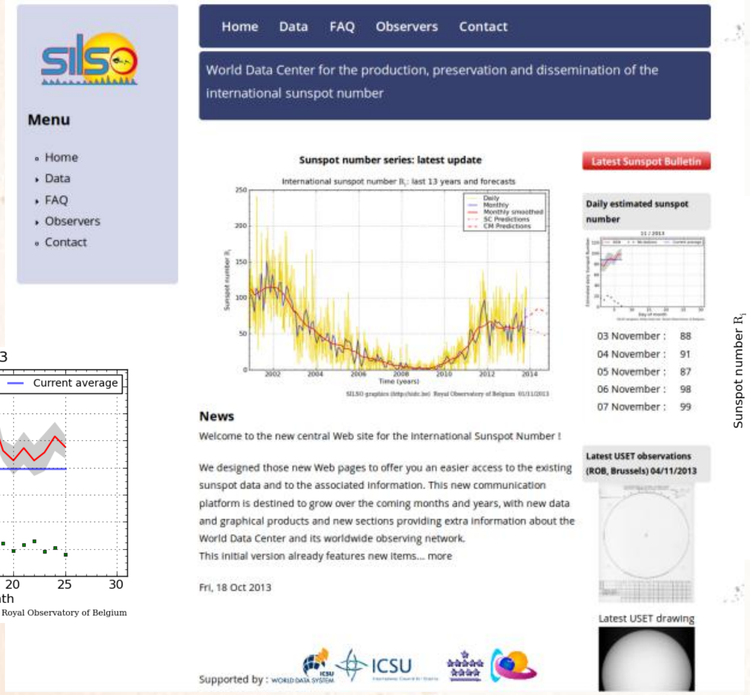
Choosing a new reference: the options

3. Multi-station reference: (cont.)
   - Current lesson:
     - All stations are stable only during limited intervals
   - ⇨ Use of **time-masks**: only valid parts of each series is included
   - Advantages:
     - **Cross-validation** between pilots stations (detection of anomalies at each station)
     - **Continuity** of the reference (no risk of disruption if a station stops operating)
   - Drawbacks:
     - **More difficult to interpret**: the reference contains a mix of contributions
     - **Possible sharp jumps** when a station enters or leaves the core group

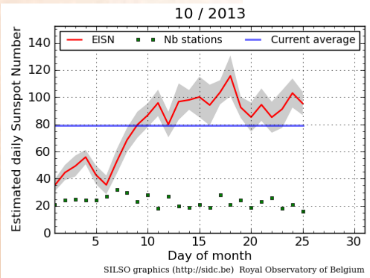
- Open questions:
  - What is the optimal number of stations?
    - Many: more difficult to keep track of events at each stations
    - Few: limited statistical validation
  - How to combine stations?
    - Average of raw numbers without k scaling (implies stations with very similar k)
    - Maintenance of a restricted k coefficient system within the core group
  - Only professional observatories or also dedicated amateurs?
  - Should we avoid stations contributing also to the AAVSO network ?

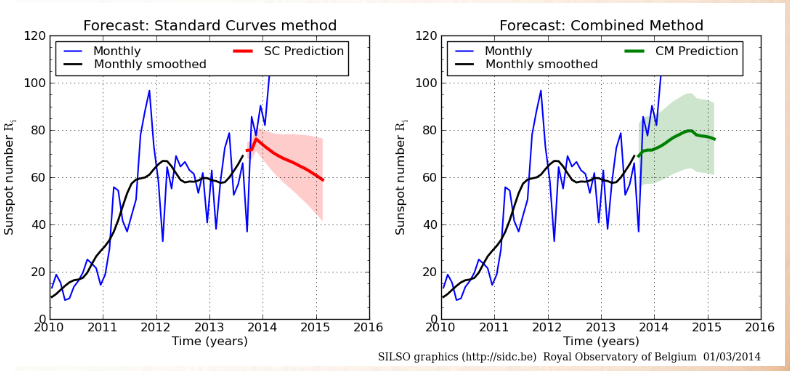
For the latest information, please visit …

WDC – SILSO Sunspot Index and Long-term Solar Observations



http://sidc.be/silso





http://ssnworkshop.wikia.com/wiki/Home