Composite analysis and Monte Carlo methods - an example with Forbush decreases and cloud cover

STRUKTURNI I INVESTICIJSKI FONDOVI

PoKRet

Europska Unija
Ulaganje u budućnost
Projekt je sufinancirala Europska Unija iz Europskog socijalnog fonda

ESF HUMAN RESOURCES DEVELOPMENT

SOLSTEL

hrzz
Hrvatska zaklada za znanost

Opservatorij Hvar
Hvar Observatory

Jaša Čalogović

Hvar Observatory
Faculty of Geodesy
Zagreb, Croatia

Short-term studies using Forbush decreases show conflicting results

- **positive correlations:**

Tinsley & Deen, 1991; Pudovkin & Vertenenko, 1995; Todd & Kniveton, 2001; 2004; Kniveton, 2004; Harrison & Stephenson, 2006; Svensmark *et al.*, 2009; Solovyev & Kozlov, 2009; Harrison & Ambaum, 2010; Harrison et al. 2011; Okike & Collier, 2011; Dragić et al. 2011; 2013; Svensmark et al., 2012; Zhou et al. 2013; Aslam & Badruddin, 2015

- **negative correlations:**

Wang *et al.*, 2006; Troshichev *et al.*, 2008

- **no correlations or inconclusive results:**

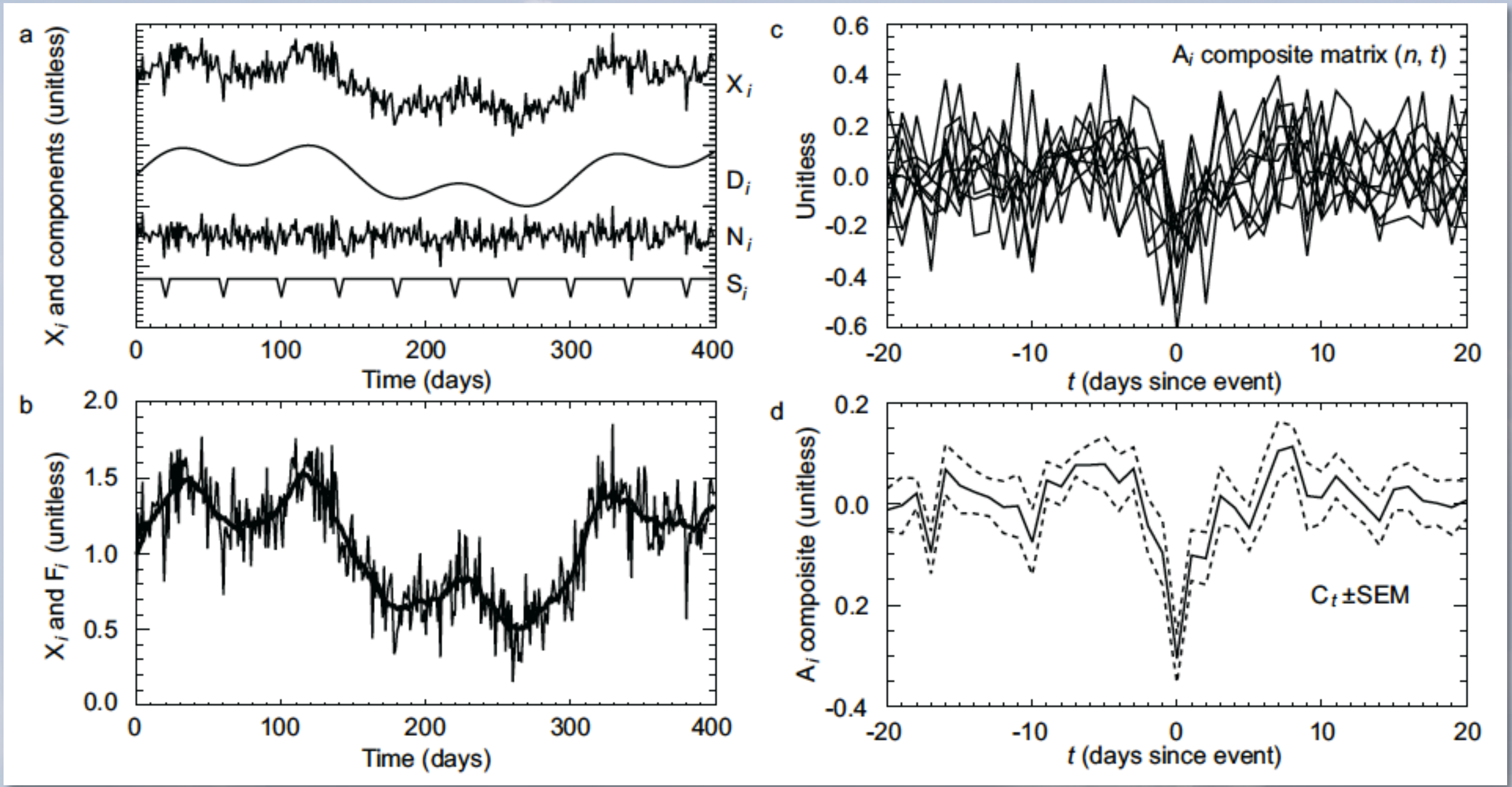
Pallé & Butler, 2001; Lam & Rodger, 2002 ; Kristjánsson *et al.*, 2008 ; Sloan & Wolfendale, 2008; Laken *et al.*, 2009; Čalogović *et al.*, 2010; Laken & Kniveton 2011; Laken et al., 2012; Erlykin and Wolfendale, 2013

Possible reasons

- there is no relationship between cosmic rays and clouds
- a relationship is too weak to detect (signal to noise ratio)
- other solar parameters may interfere with the results (e.g. TSI, UV) – problem with **signal attribution**
- relationship exists but it is constrained by the atmospheric conditions at the time

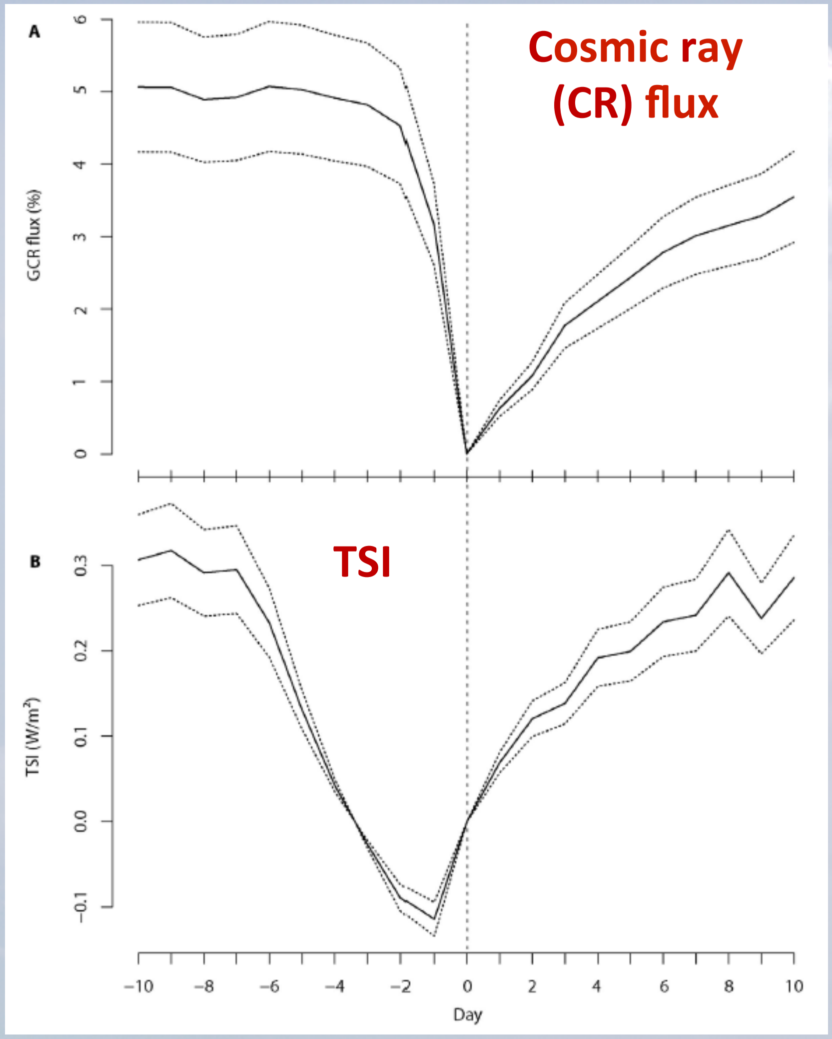
How to isolate the signal using composite?

Superposed epoch analysis or conditional sampling



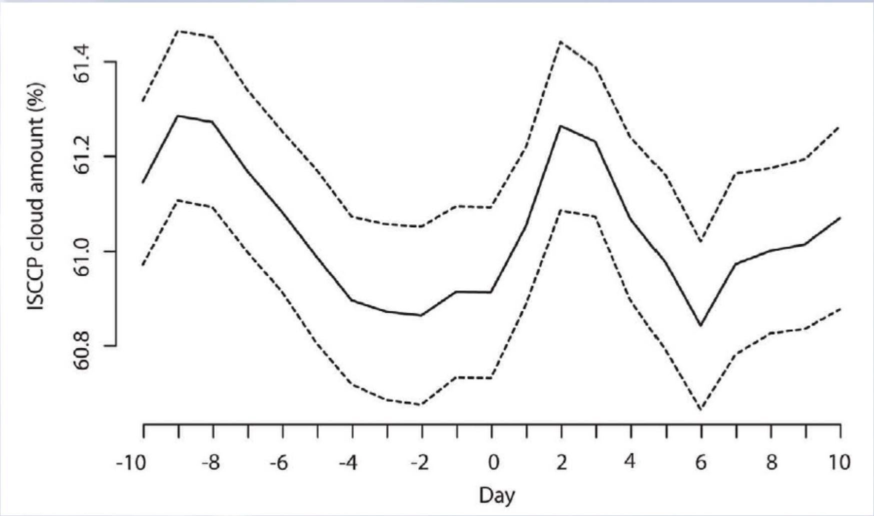
- Successive averaging of events (in time or space)
- Used to increase signal-to-noise ratio (SNR)
- Enable detection of small amplitude signal against large variability

TSI influences the cloud cover?



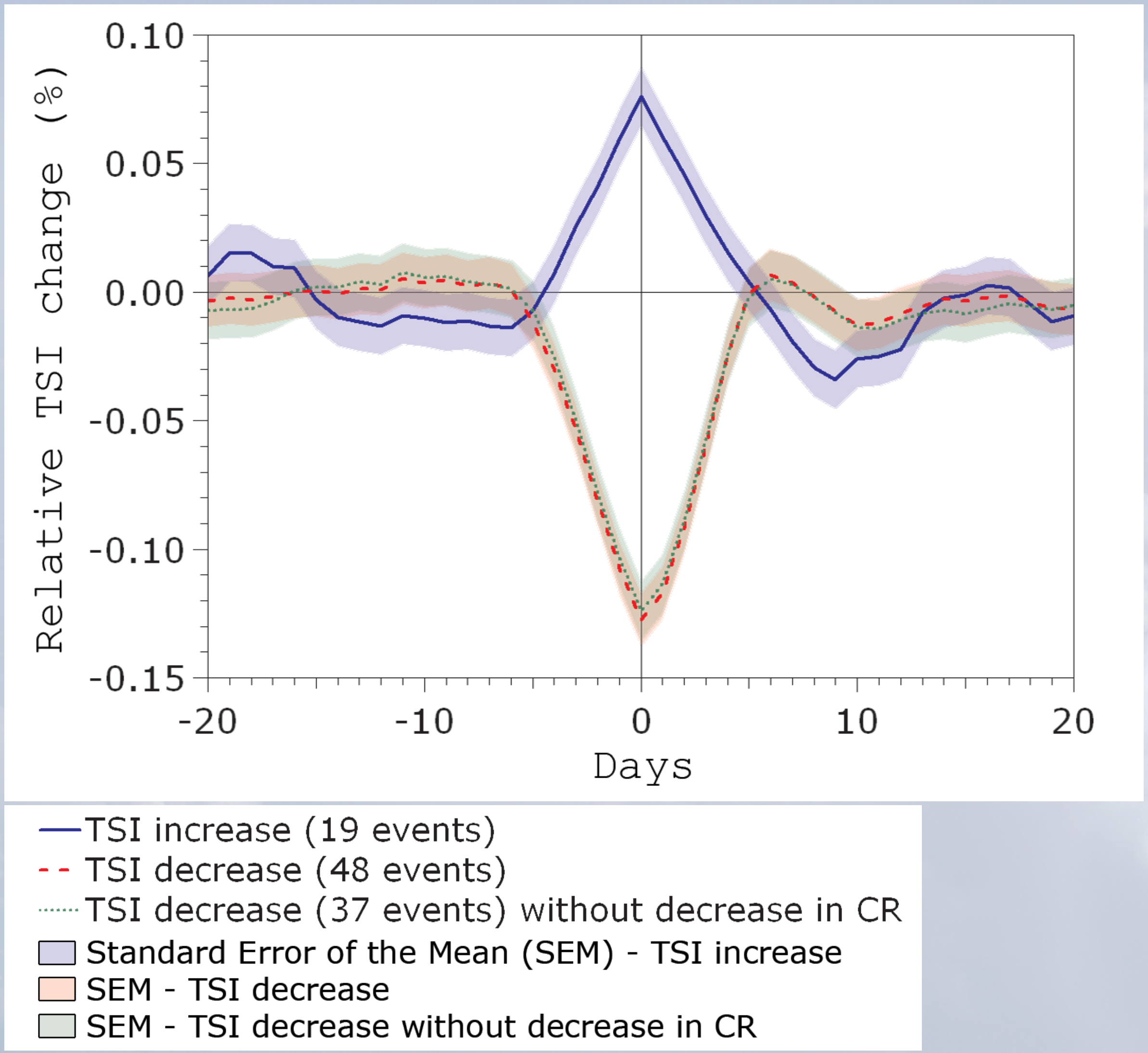
Laken et. al., 2011 (JGR)

ISCCP cloud cover



- Composite (superposed epoch) analysis of 123 Forbush decrease events
- cloud cover decreases about 2 days **before** the onset of Forbush decrease (CR flux)

TSI data and composite samples



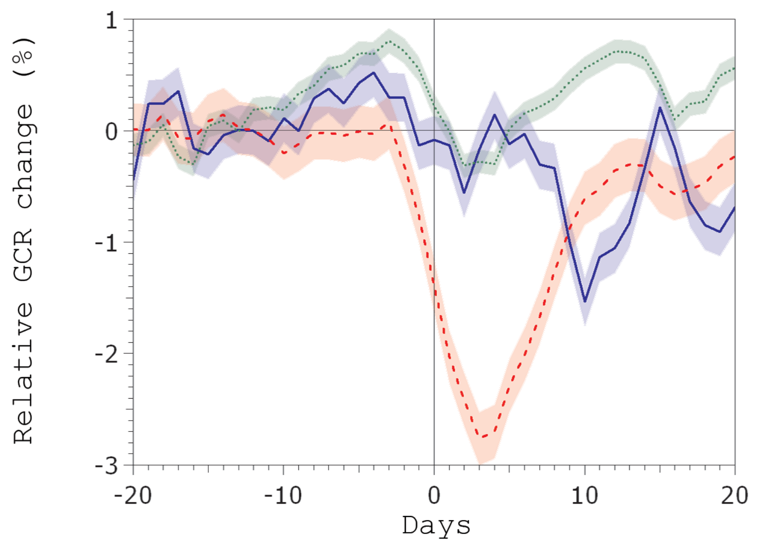
Laken & Čalogović, 2011 (GRL)

- Active Cavity Radiometer Irradiance Monitor (ACRIM) reconstruction, 1978-present, daily values

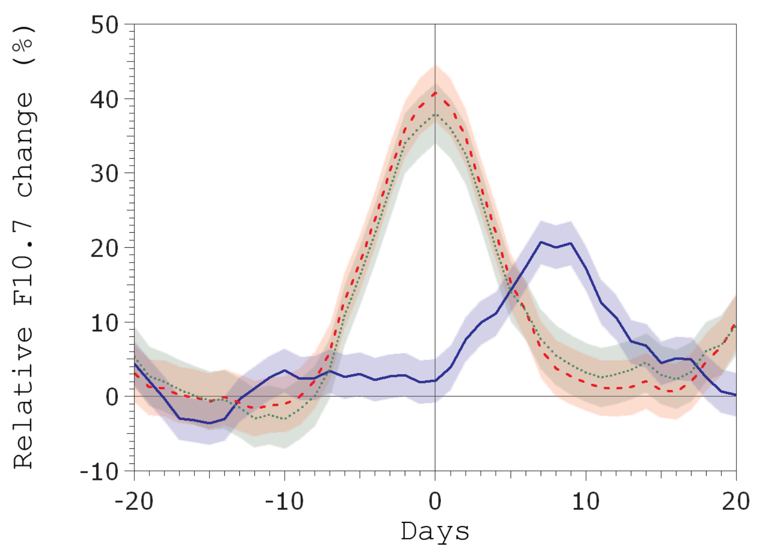
- 3 composite samples:
  - largest increases in TSI (19 events)
  - largest decreases in TSI (48 events)
  - largest decreases in TSI without significant CR variations (37 events)

GCR and F10.7 (EUV) composites

GCR

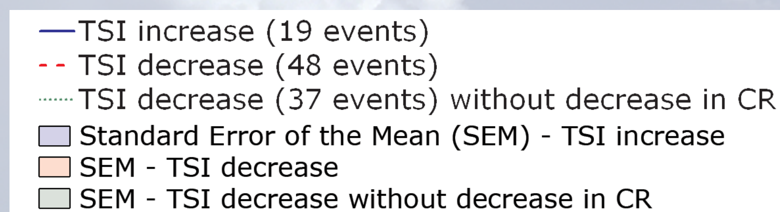


UV



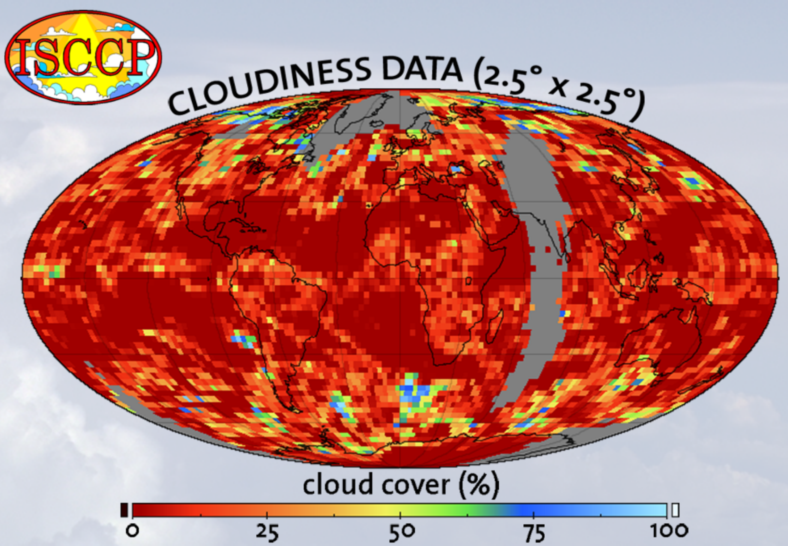
- CR flux data – Climax neutron monitor (R_c=2.99GeV)
- F10.7 (2800Mhz) data – proxy of extreme ultraviolet solar activity (EUV)

- all composites (TSI, CR, F10.7) correlated with corresponding cloud data using a lag of 20 days



Cloud data

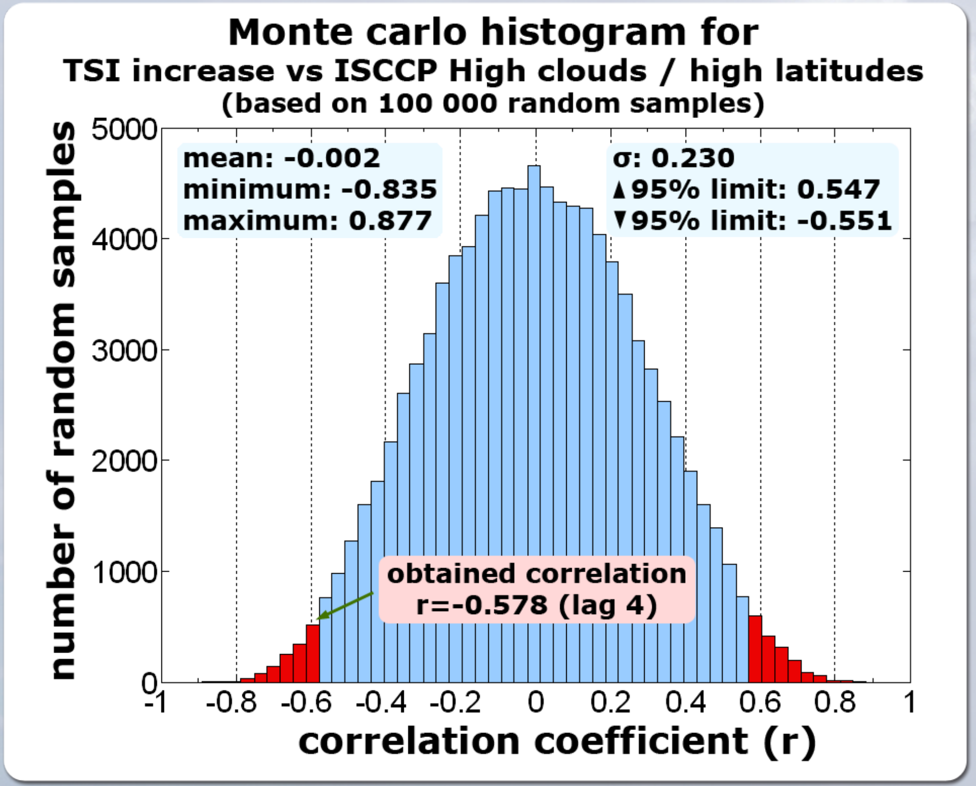
- International Satellite Cloud Climatology Project (ISCCP) D1 dataset, IR data, 1983-2008, temporal resolution 3h, equal-area grid (280x280km^2)
- 3 different altitude levels: high (>6.5km), middle (3.2 – 6.5km) and low (0 – 3.2km) clouds
- daily averaged



- area-averaging was applied <u>for different regions</u>:
  - global
  - low latitudes (<45°)
  - high latitudes (>45°)
  - regions over land
  - regions over ocean

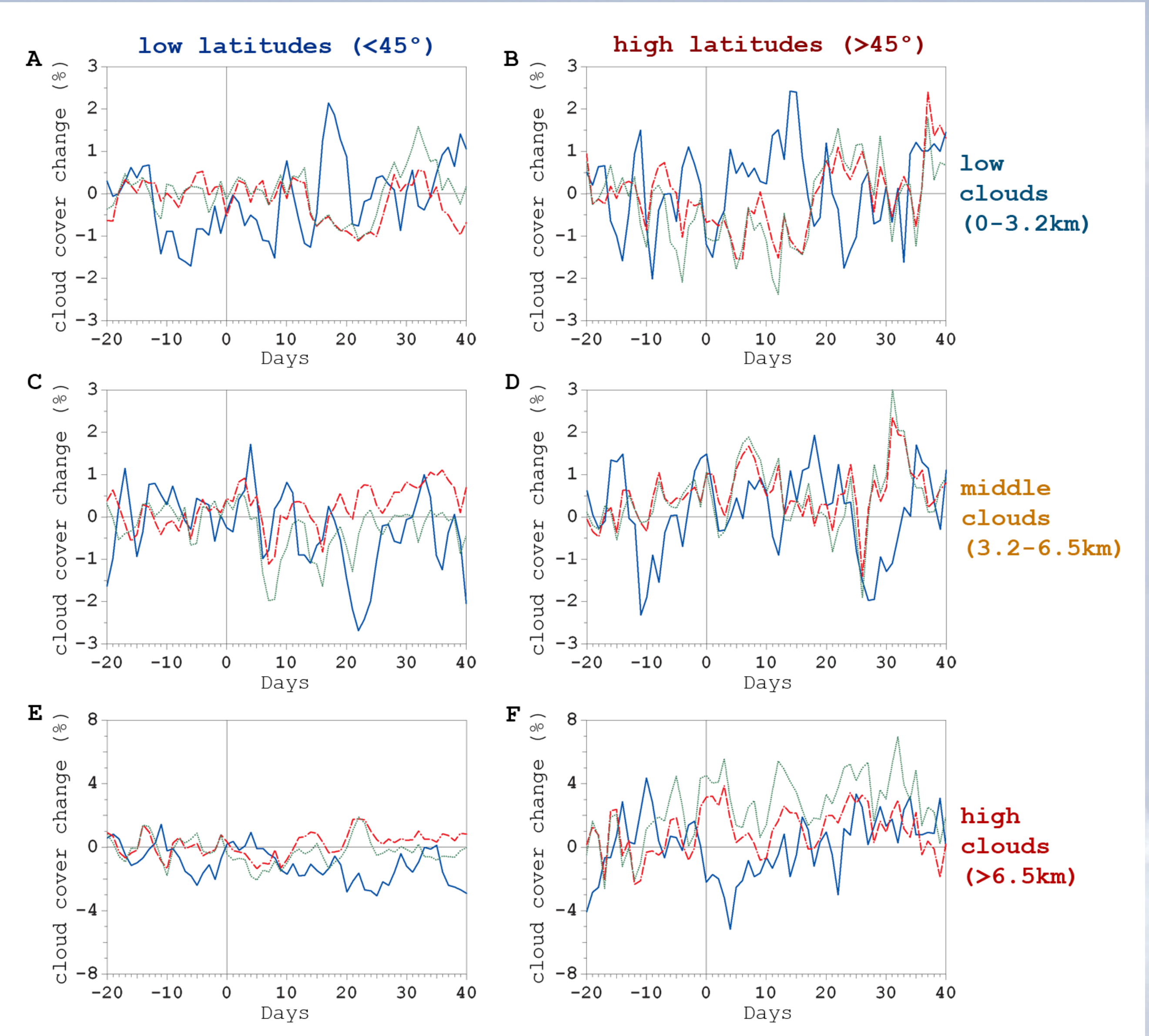
Monte Carlo tests

- employed to establish the threshold significance values for the correlation coefficients (r)
- for each parameter 100 000 randomly generated r



- Shapiro-Wilk test of normalcy: all r are normally distributed (W = 0.996, p = 4.8x10^{-10})
- statistical significance set by two-tailed 0.95 percentile MC generated r values

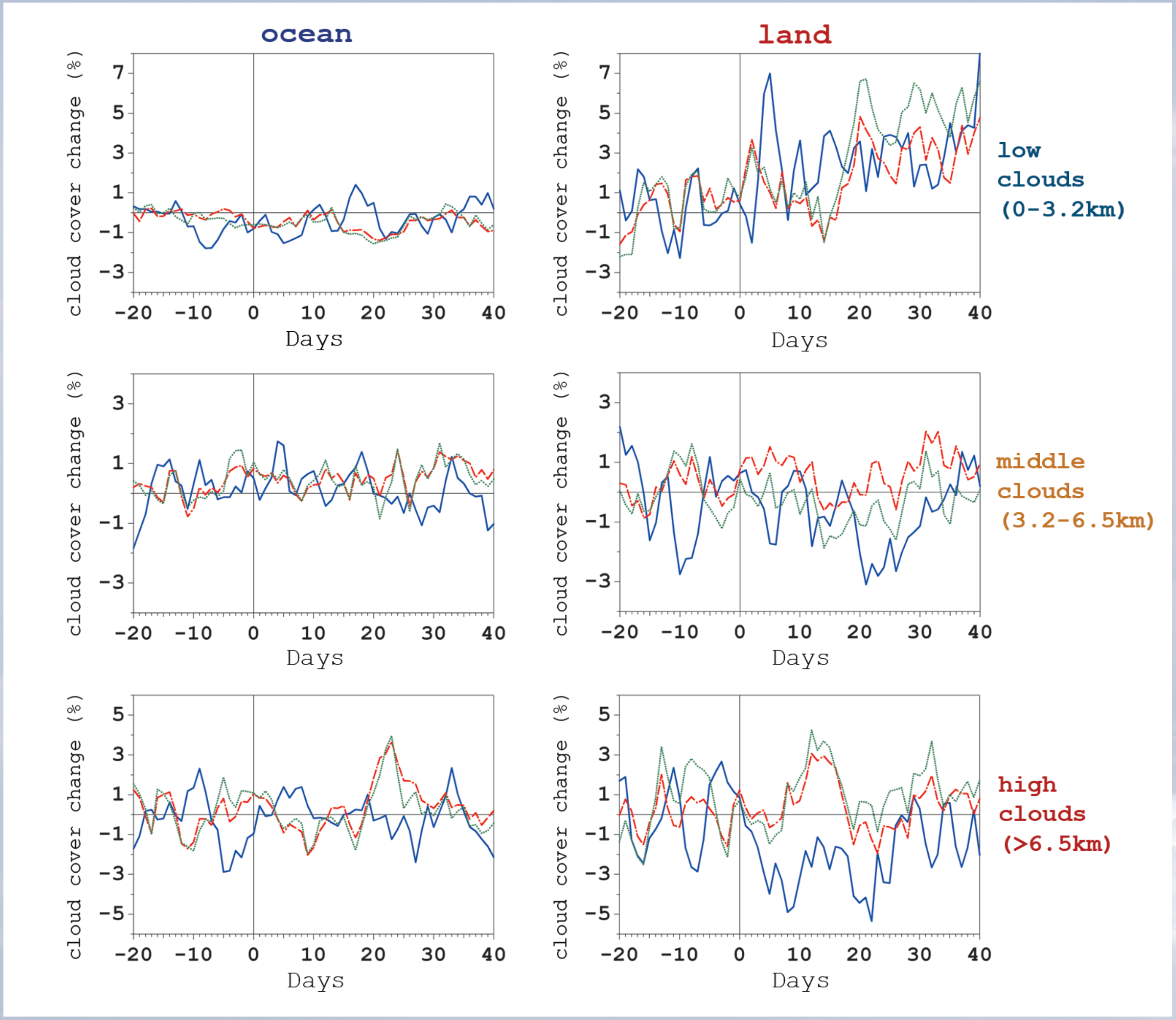
Cloud composites – low and high latitudes



Laken & Čalogović, 2011 (GRL)

- no significant correlations with TSI, CR and UV composites

Cloud composites – ocean and land

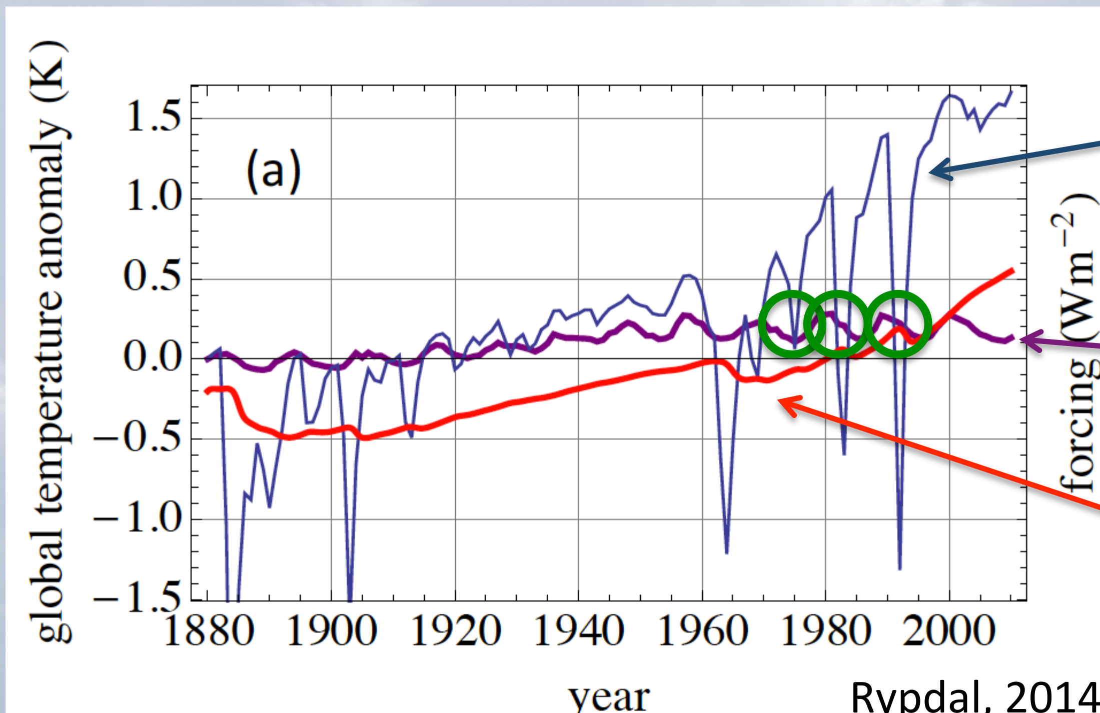


Laken & Čalogović, 2011 (GRL)

- no significant correlations with TSI, CR and UV composites

If some climate signal is found - it should be properly attributed to solar forcing

- Other external and internal factors influencing the climate parameters should be identified ➔ eg. attribution by multiple regression or models (if possible)
- Last few solar cycles coincidentally match with strong volcanic eruptions (volcanic forcing)

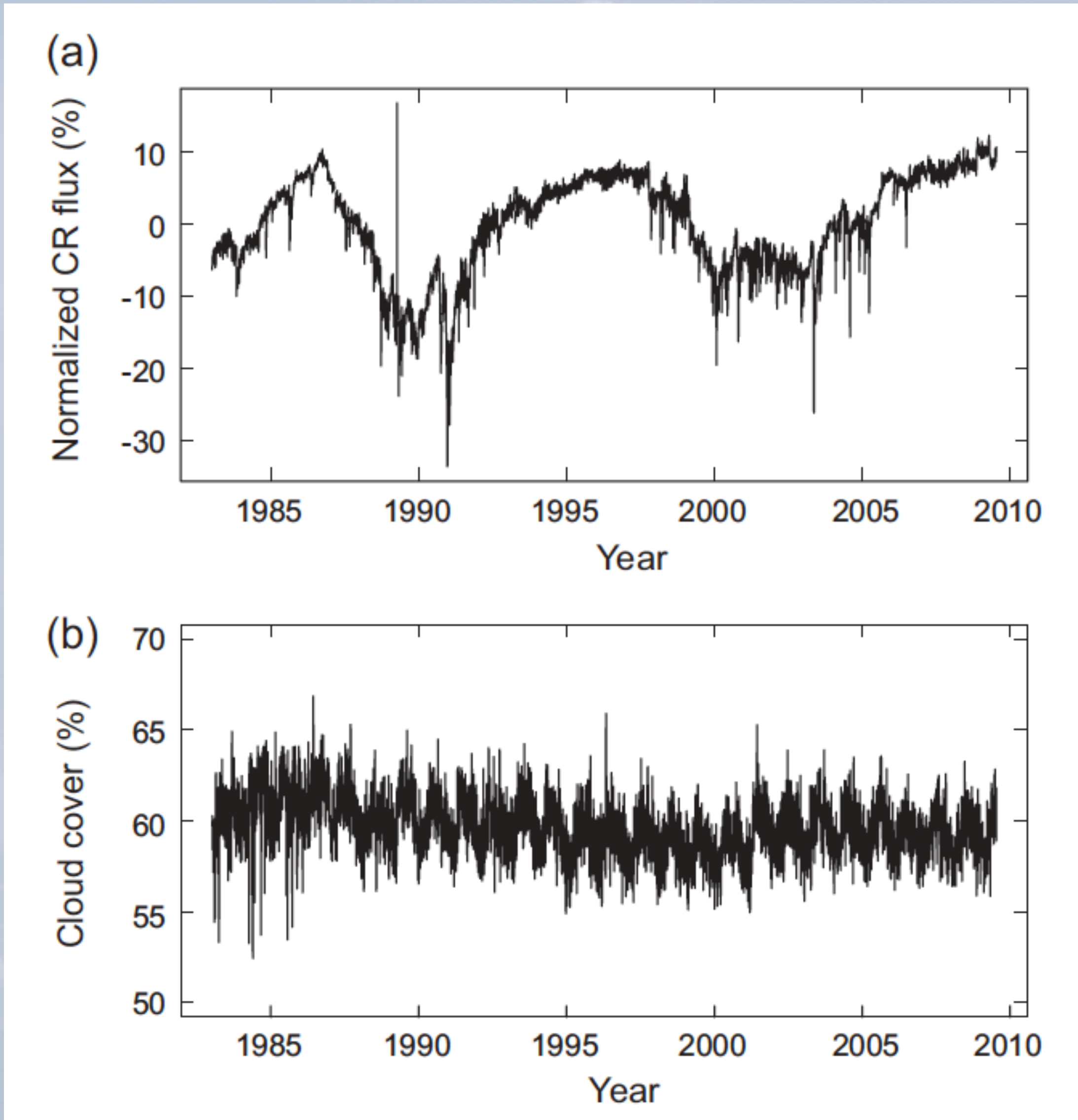


Possible methodological reasons for conflicting results

- unappropriate or no data filtering
- wrong statistical assumptions and/or improper use of statistical tools
- “quality” and properties of cloud datasets (**autocorrelated data**)

So how to test the CR-cloud link reliably ?

Cosmic ray flux and cloud data



Daily averaged normalized cosmic ray flux (%) calculated from Climax Colorado and Moscow neutron monitors

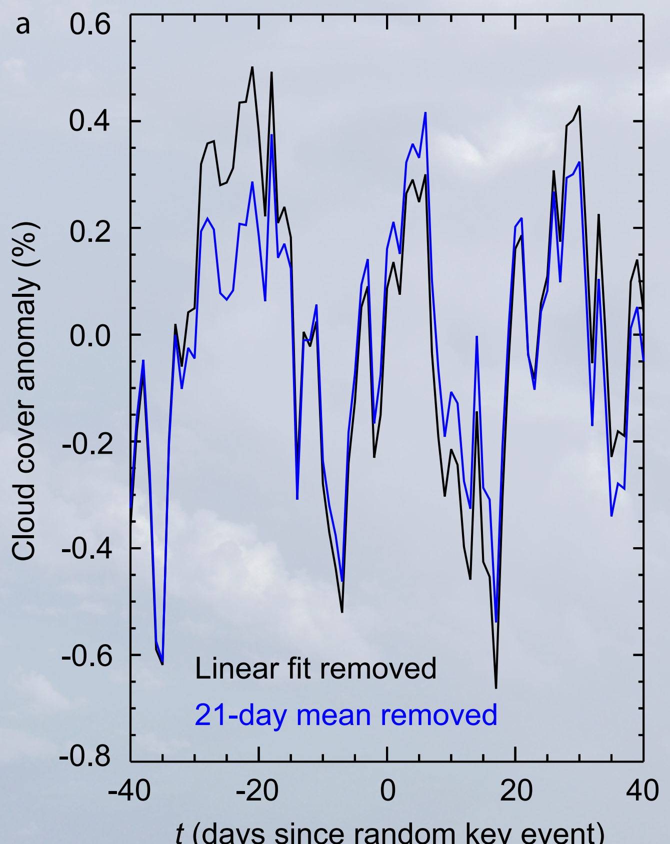
Global and daily averaged ISCCP D1 (IR-detected) cloud cover (%)

Composites should be made with anomalies rather than raw data…

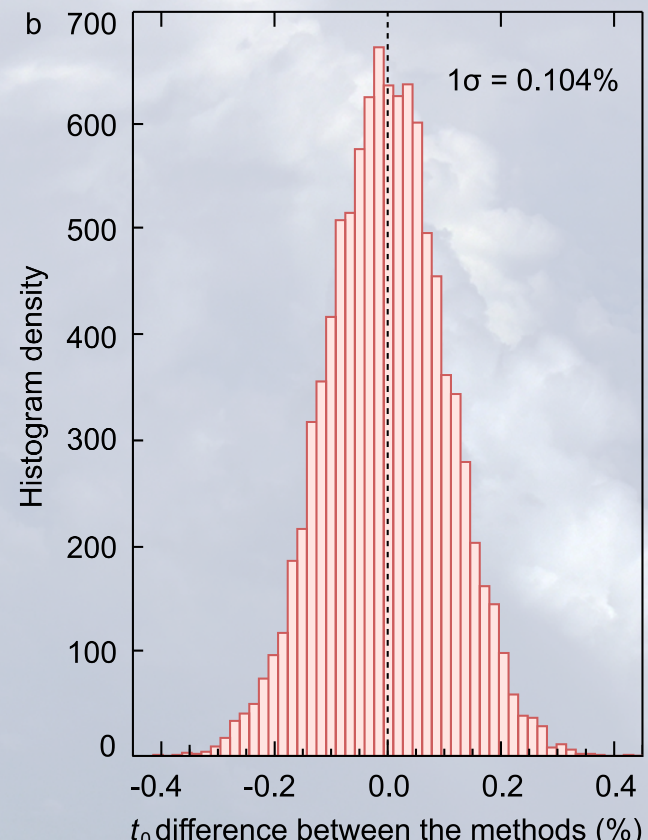
… to minimize variations in data unconnected with hypothesis testing (<u>high-pass filtering</u>)

Comparision of two methods to remove long-term variations (n=20): linear trends removed (black), and a only 21-day running mean removed

a) one random composite over $t_{\pm 40}$



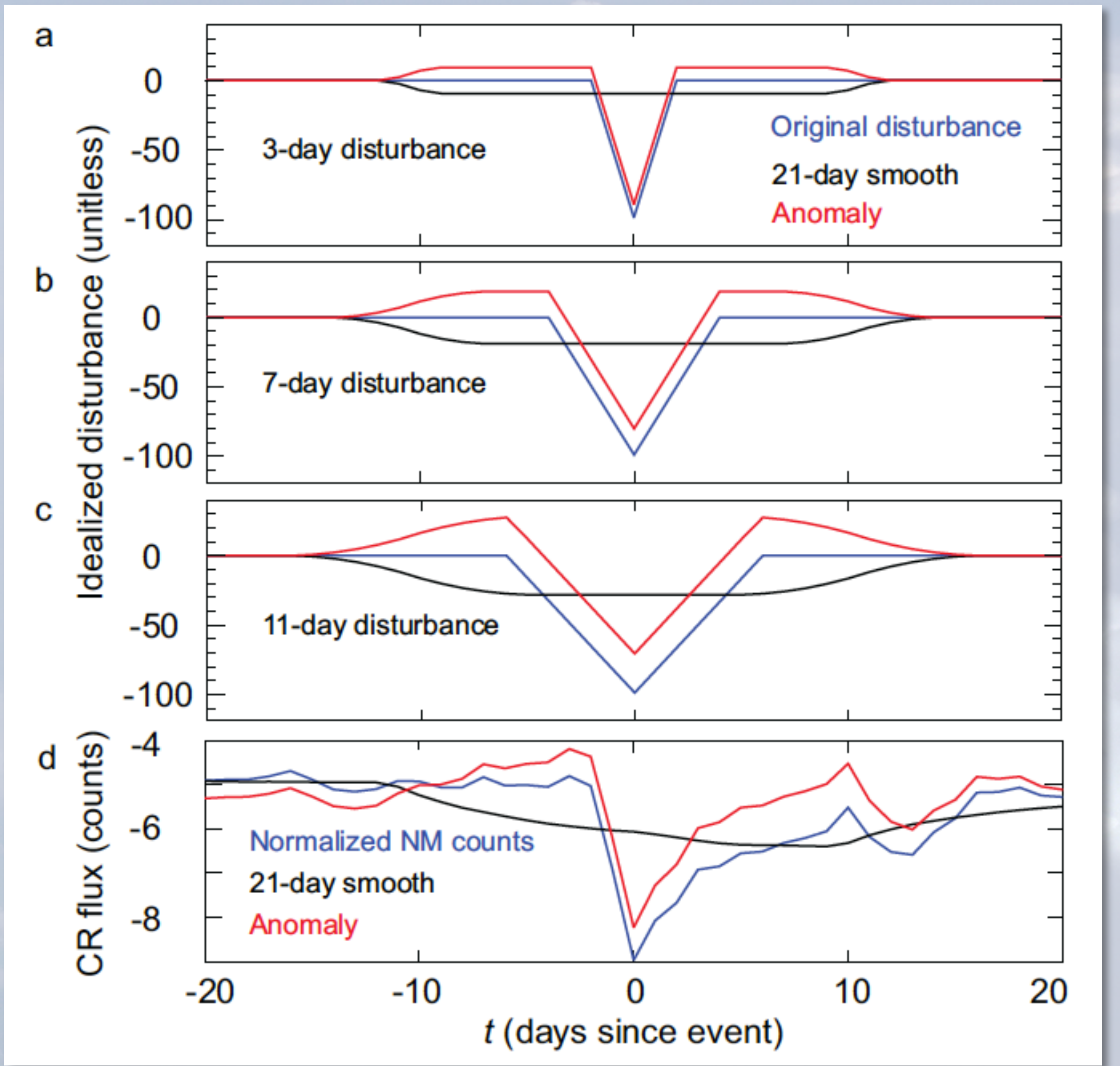
b) 100,000 composites at t_0



Differences (*b*), indicates remaining non-linear variations in composite from synoptic scale variability. <u>If not removed, **this will bias results of composites**</u>.

Proper selection of smooth filter width is needed to prevent signal attenuation (duration of searched signal is 1/3rd width of smooth filter)

Overshoot / undershoot effects by filtering the data with different filters (running mean)

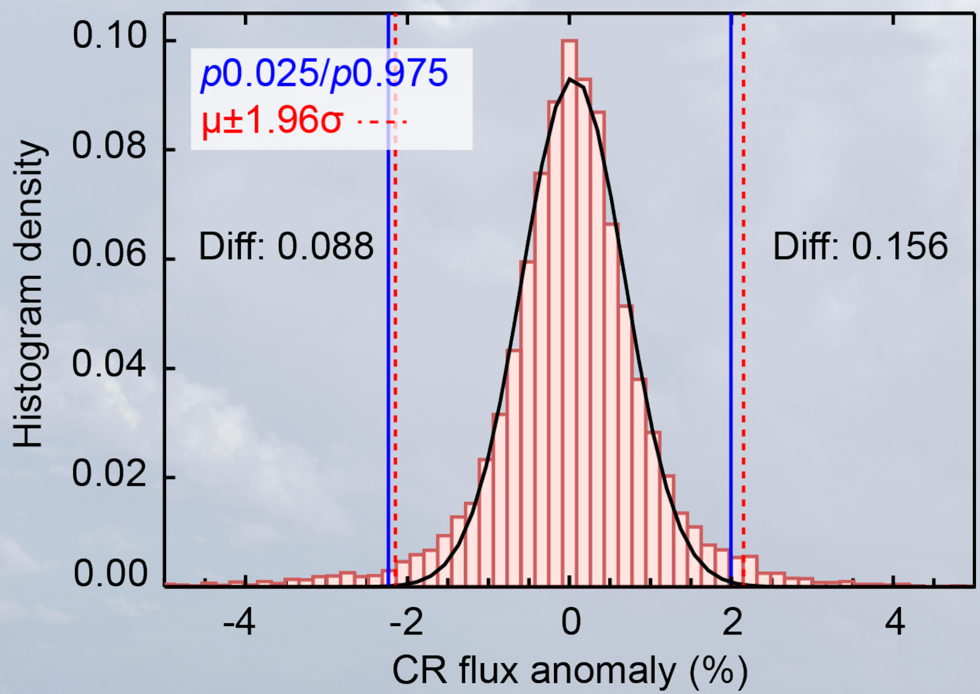


For deviations at timescales of aprox. 1/3rd the width of the smooth filter, disturbance attenuation is very small or neglible.

Calculate thresholds for statistical significance with Monte Carlo approach

By generating large populations of random events identical in design to a composite with real events, the probability (p) of obtaining a given value by chance in a composite with real events can be accurately known.

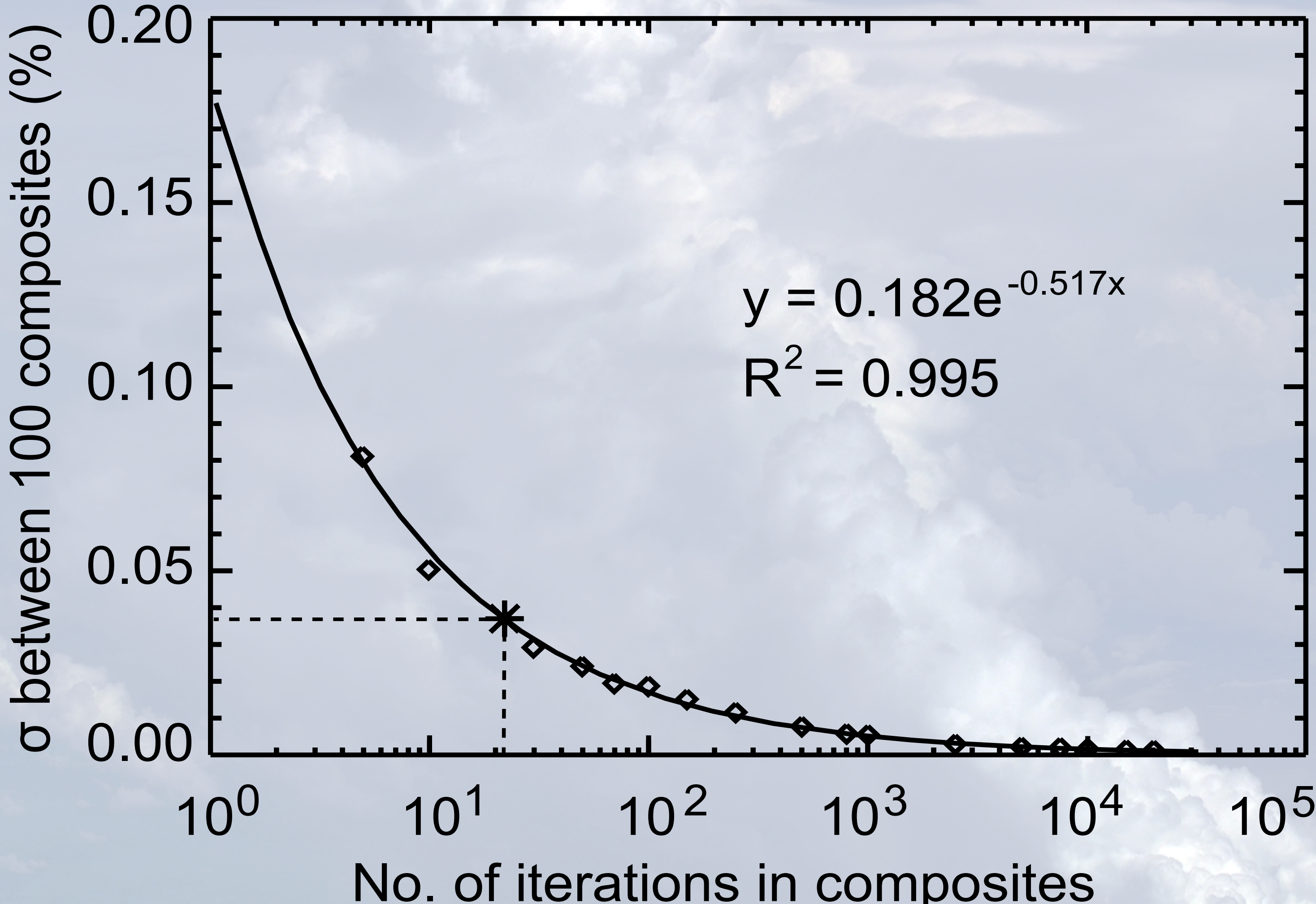
Distribution of daily anomalies



This has advantages over traditional tests (e.g. T/U tests), as it requires **no minimum sample size** or specific distribution, and it doesn't need **adjustment for autocorrelation**.

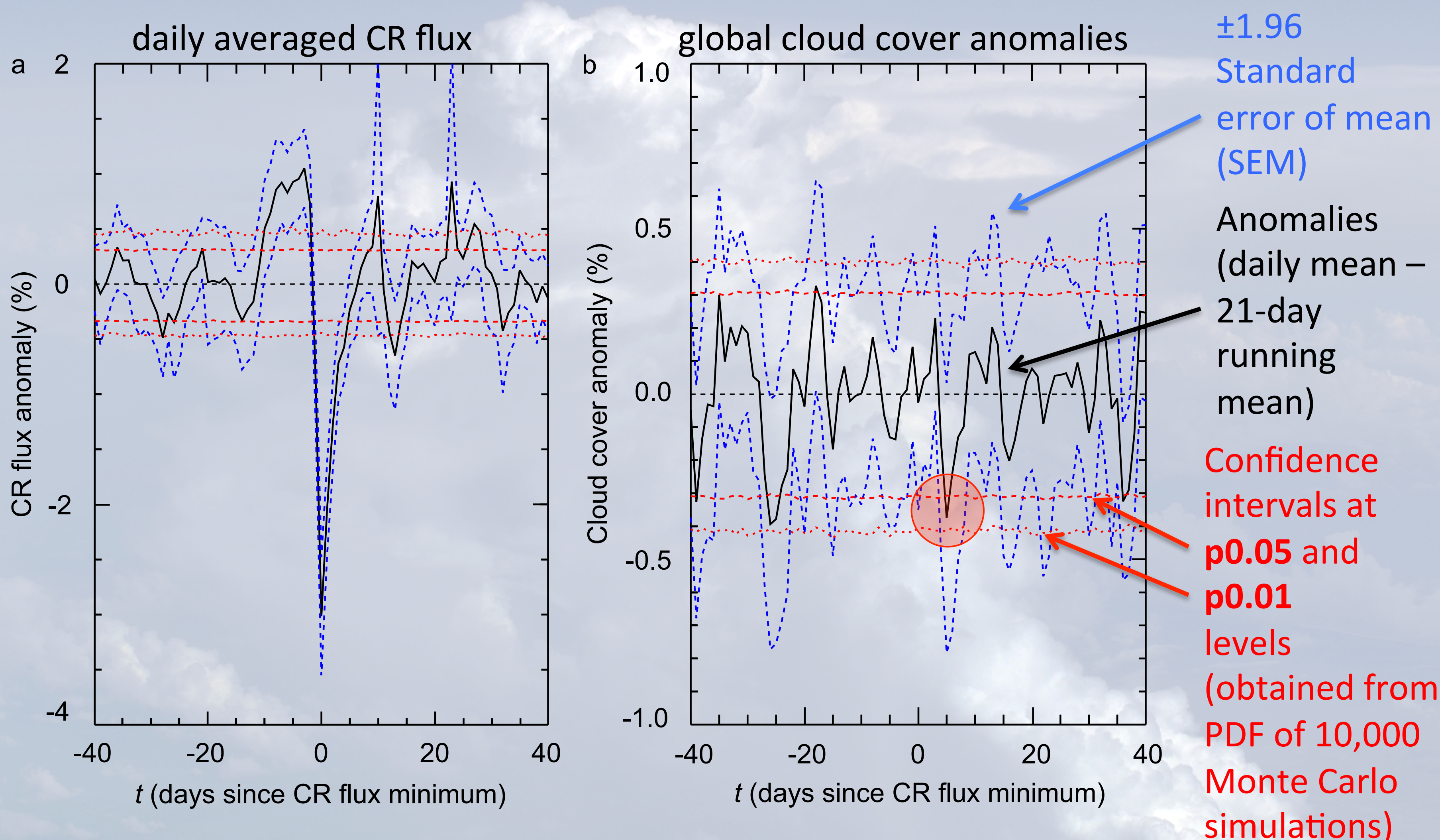
Laken & Čalogović, SWSC, 2013

How many Monte Carlo iterations are enough to get reliable significance intervals?



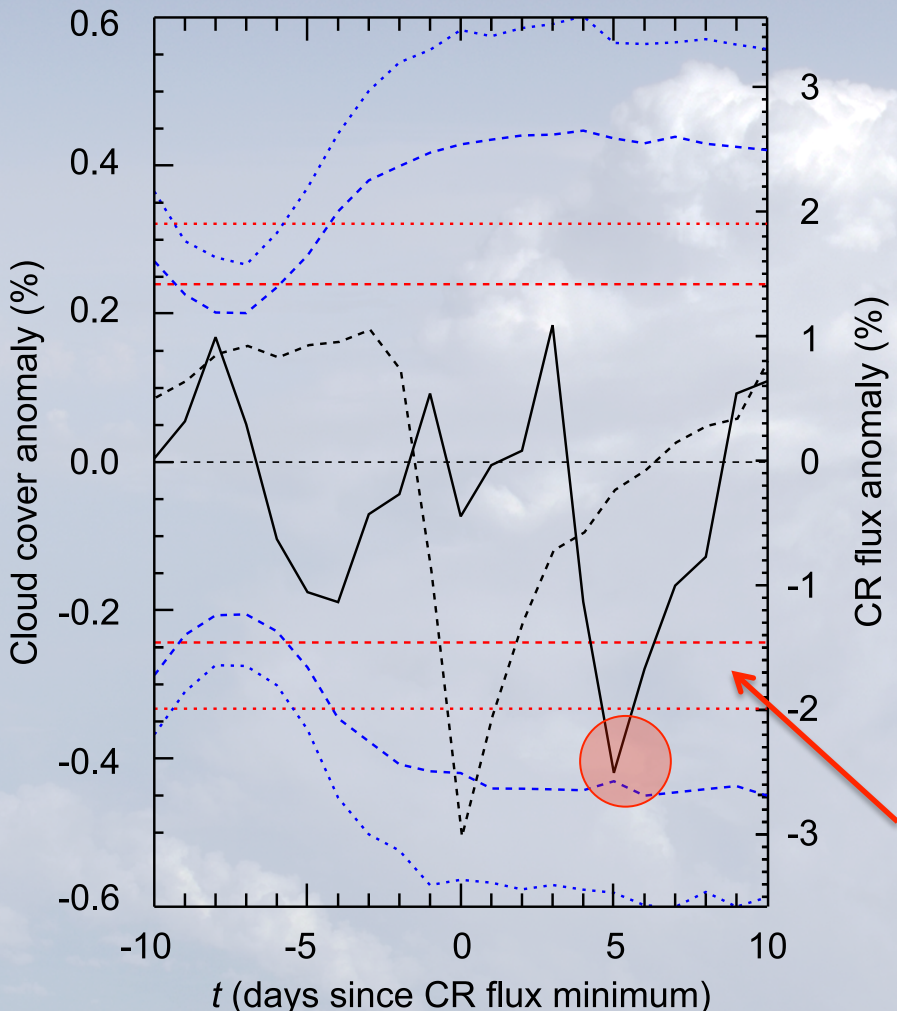
Laken & Čalogović, SWSC, 2013

Composites and significance intervals



Composites consists of 44 events

How to obtain a false positive

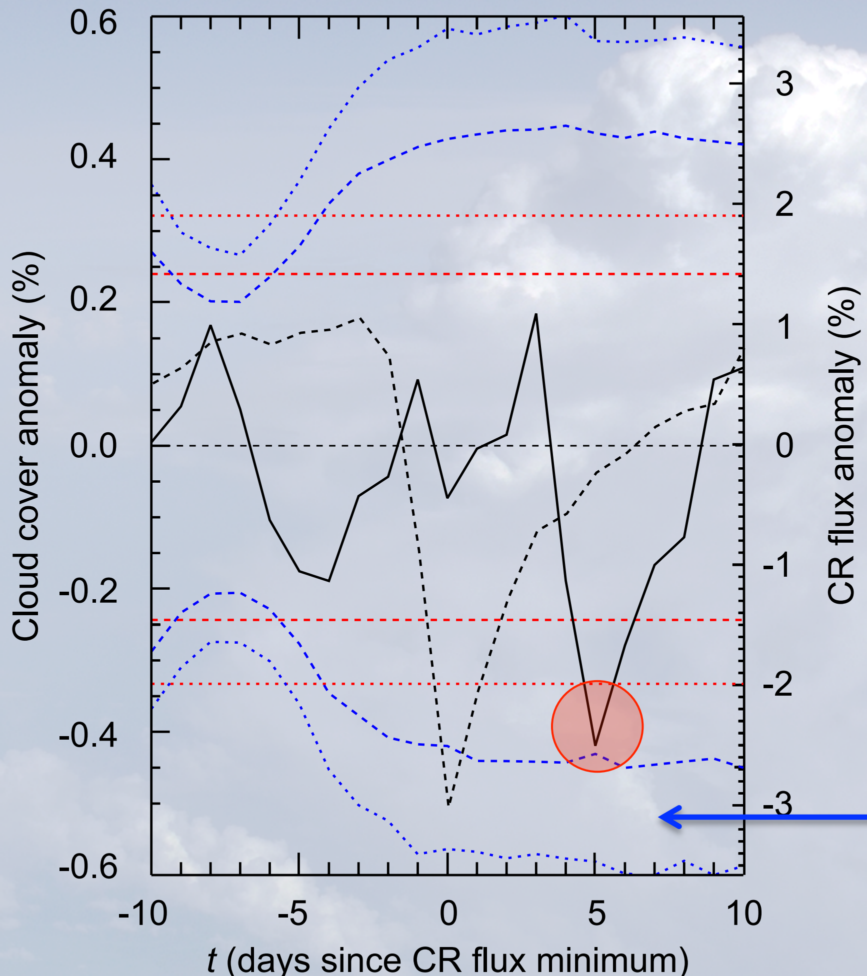


Cookbook…

- Identify a base or 'undisturbed' period before the key events, that represent 'normal conditions' (e.g. shown example uses t_{-10} - t_{-5})
- Calculate deviations against this 'undisturbed' period (i.e. subtract every *t* point from mean of 'normal conditions')
- Statistically compare the data to the 'undisturbed' period (e.g. T-test, or even MC from the base period [red lines p0.05 p0.01])

Normalization to base period reduces population variability towards base period, narrowing confidence intervals.

How to avoid a false positive



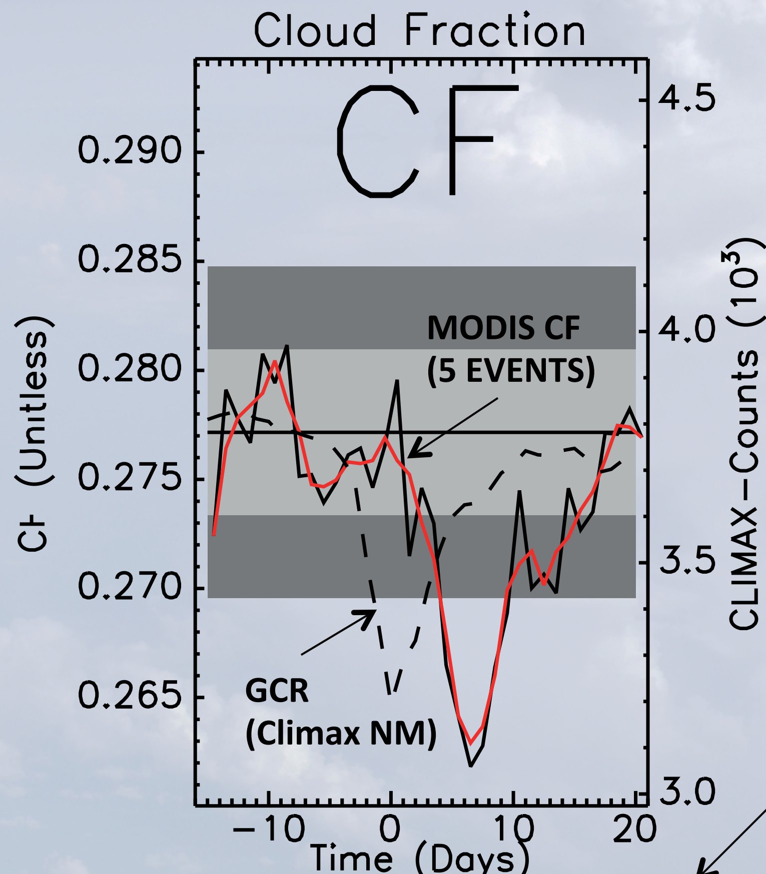
Overcoming bias with Monte Carlo (MC):

- Use confidence intervals from PDFs obtained with MCs, calculated **independently** for each *t* point
- Autocorrelation effects are automatically taken in to account (random samples in the MC all treated with an identical approach to the analyzed composite [blue lines p0.05 p0.01])

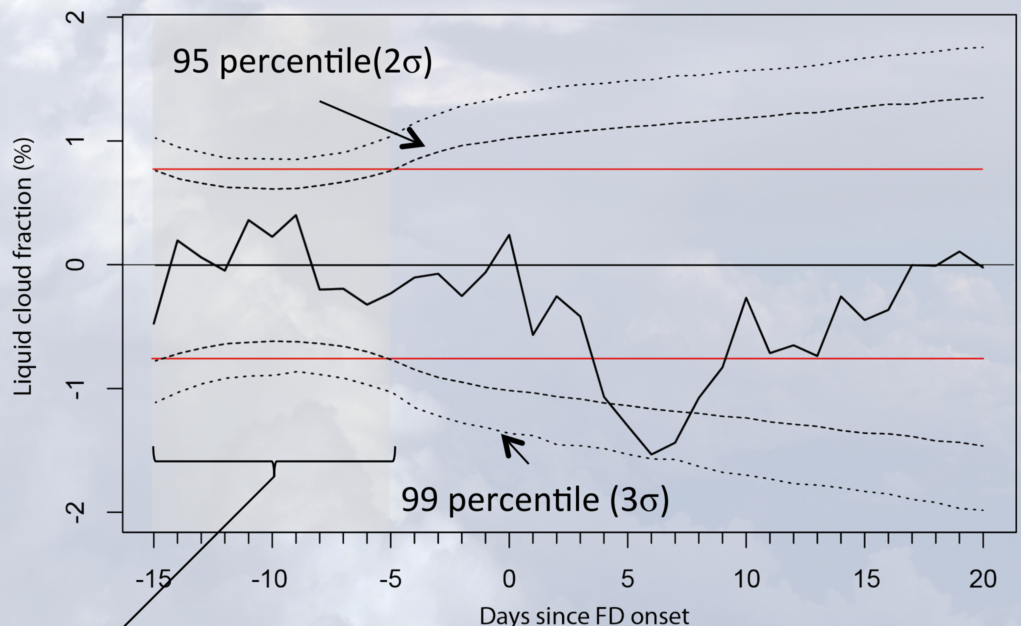
Two different results for t_{+5} (the above with a mean p<0,05 and the earlier, with a mean 0.01>p<0.05 so which is correct?

Big variability in the clouds can be often mixed with the expected signal!

Svensmark et al. 2012, ACPD



Laken, Čalogović, Beer and Pallé (2012), *ACPD*



Dashed/dotted lines show **correctly** adjusted 2 and 3σ level – calculated from 10,000 MC simulations

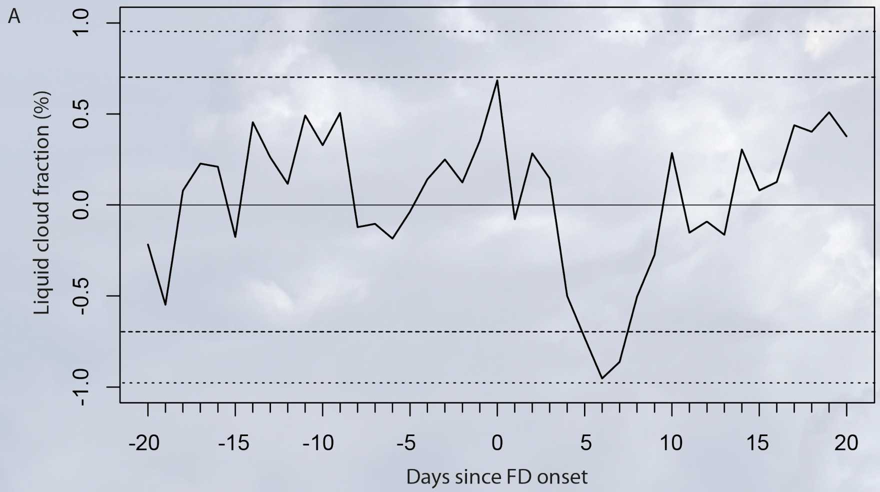
Data NORMALIZED between period of day -15 and day -5

Proper statistical tests (MC simulations) are needed to asses the correct statistical significance!

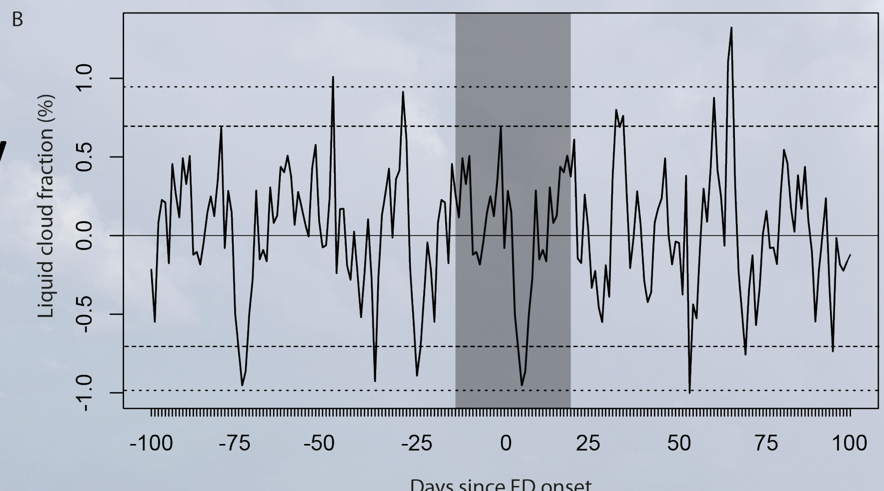
Extension to longer analysis periods reveals no unusual variability in clouds during Fd events

MODIS Liquid cloud fraction changes using 5 biggest Fd events from Svensmark et al. (2012)

±20 day analysis period



±100 day analysis period

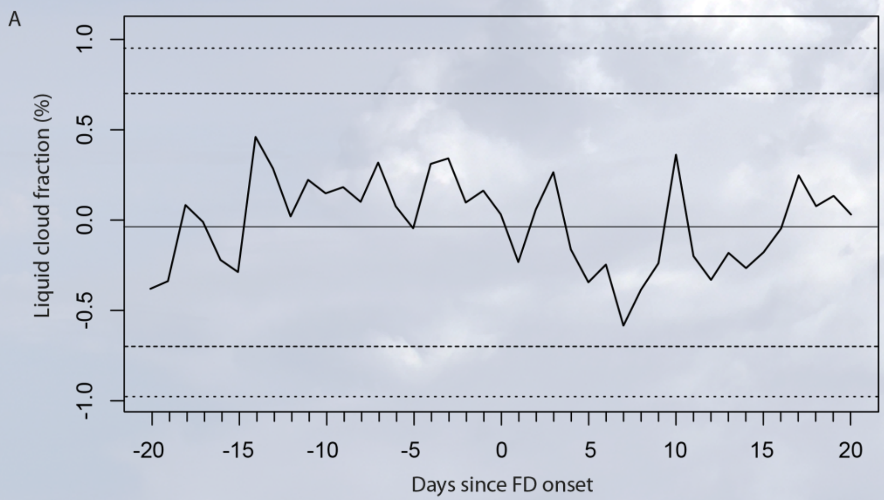


Values are anomalies from 21-day moving averages (i.e. mean of each day subtracted from 21-day moving average).

Dashed and dotted lines indicate the 95th and 99th (two-tailed) percentile confidence intervals respectively calculated from 100,000 Monte Carlo simulations.

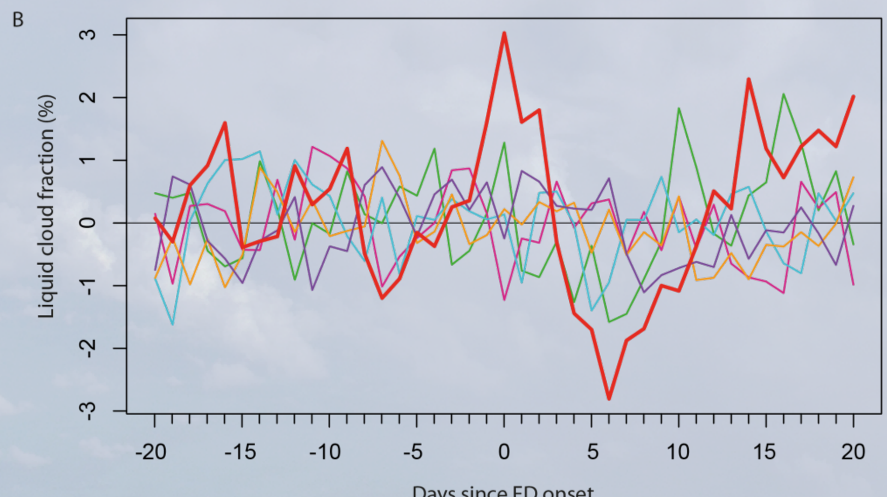
Laken, Čalogović, Beer and Pallé (2012), *ACPD*

Just one event (and eventually outlier) can influence the whole composite



MODIS cloud fraction composite for Fd events 1, 3, 4, 5, 6 ranked by Svensmark et al. 2012

By replacing the event 2 with event 6 there are no significant changes in the composite!



Individual 5 Fd events plotted against event 2 (19.1.2005) where is clear that all significance in Svensmark composite comes from event 2.

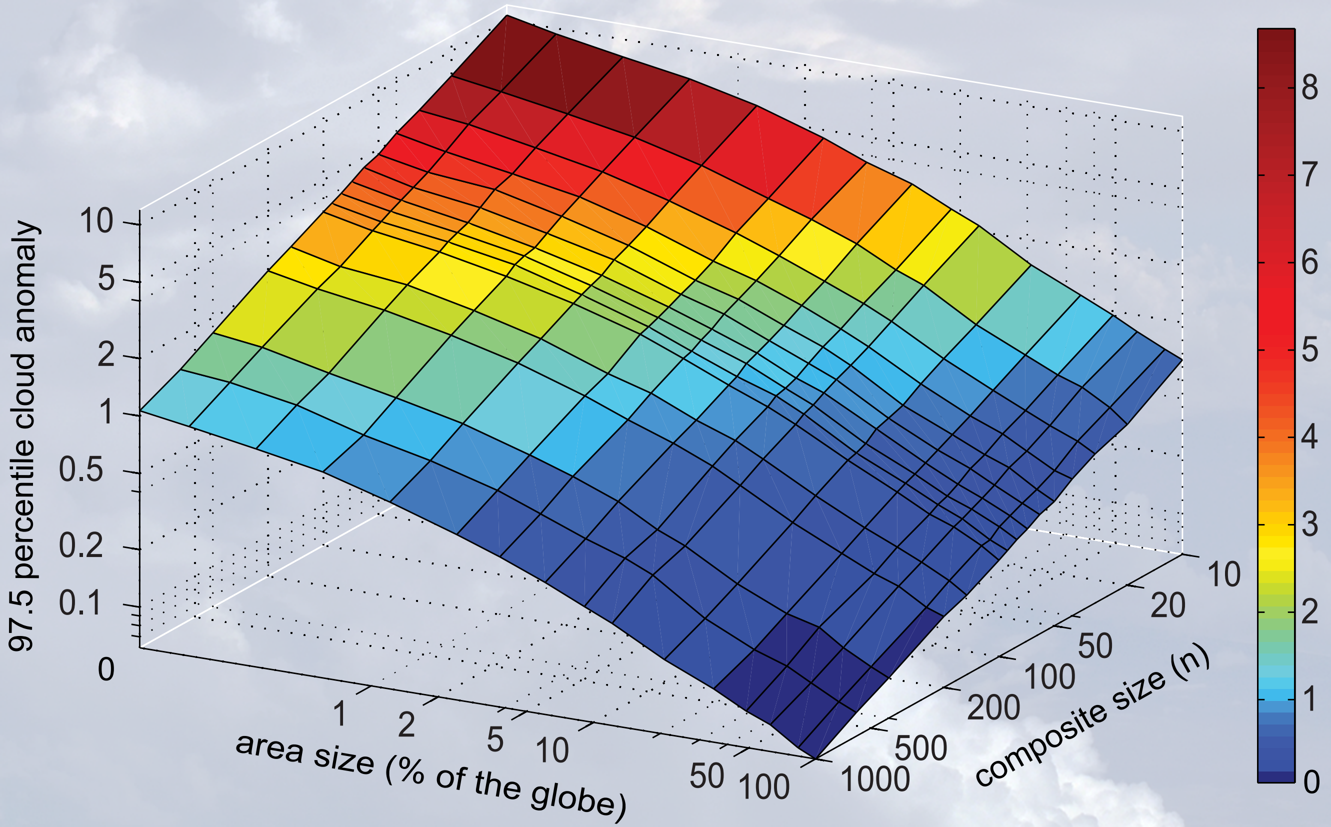
Laken, Čalogović, Beer and Pallé (2012), *ACPD*

Size of sample area and number of events impact the noise

Noise levels of data govern detectability of a signal. The noise varies with both the spatial area (*a*) that is averaged, and the number of composite events (*n*).

'*Noise*' indicated by <u>97.5th percentile values</u> from 10,000 random composites of varying *a* and *n* size.

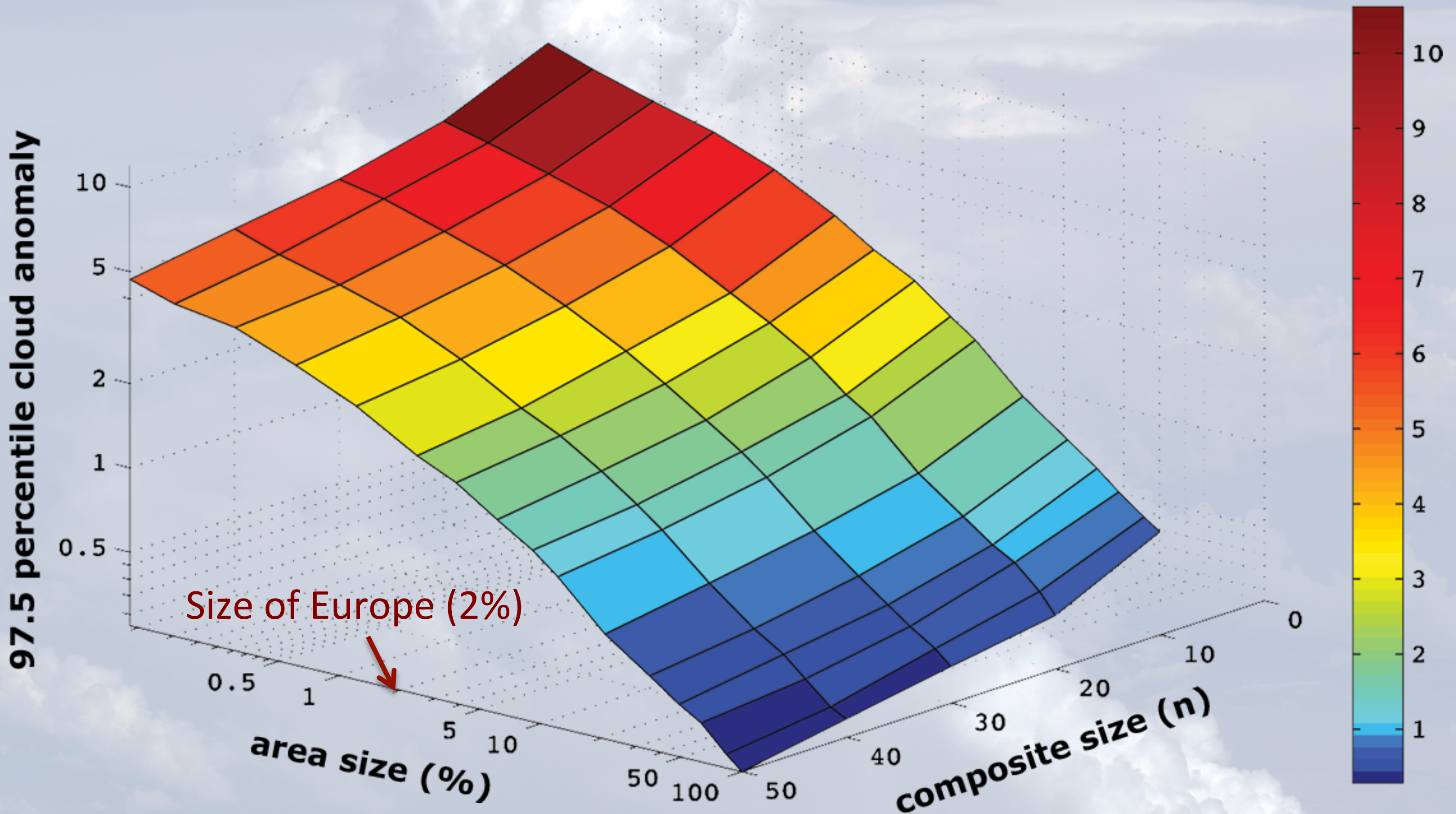
Each point of grid represents another independent set of 10,000 MC simulations



Laken & Čalogović, SWSC, 2013

possible to see how large *a* and *n* would need to be at minimum to see a hypothesized effect.

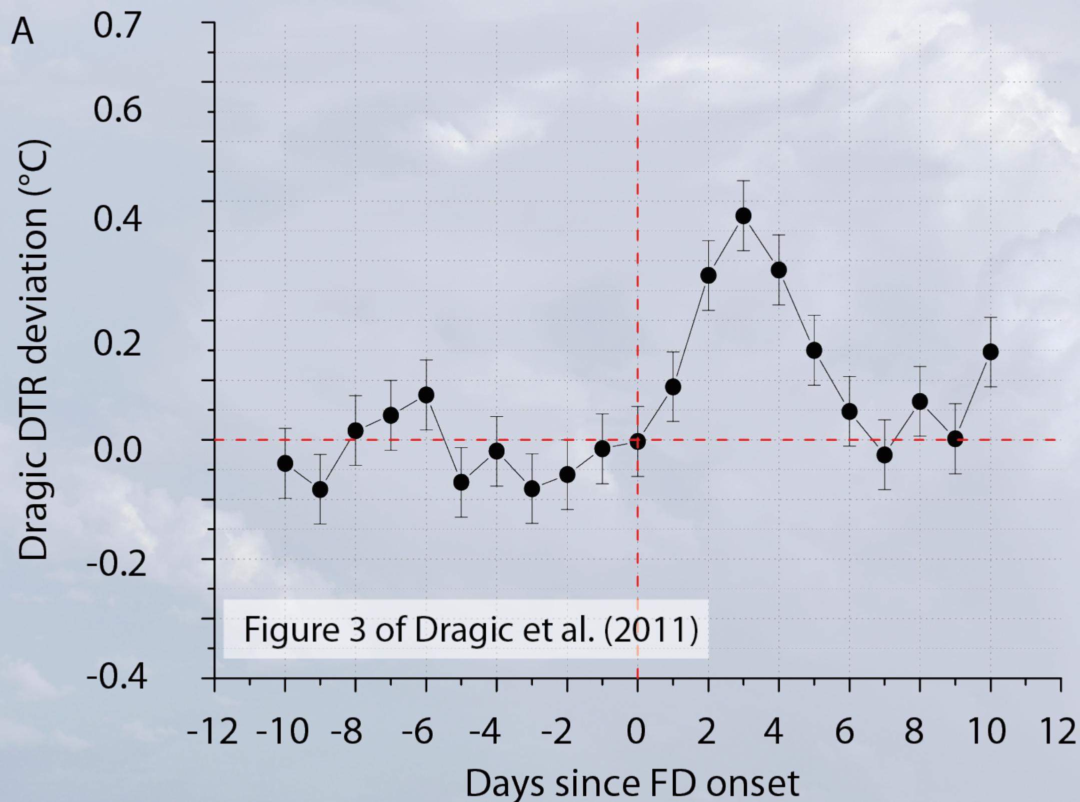
Majority of Fd studies use less than 50 events (n<50)



Studies using only strong Fd events have usually **less than 10 events**

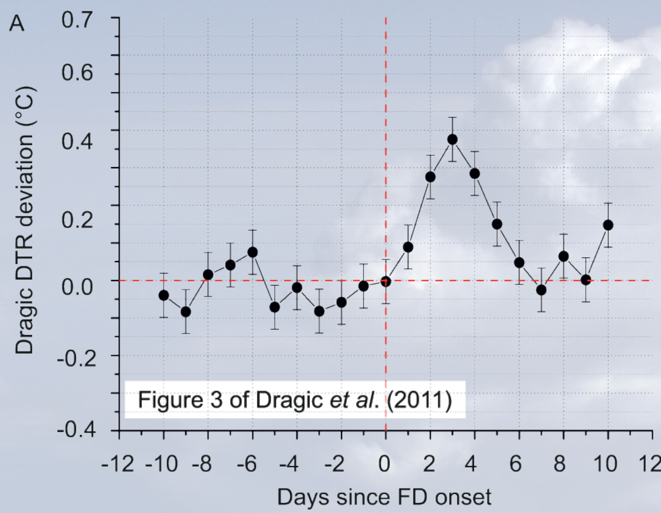
DTR shows response to Fd events?

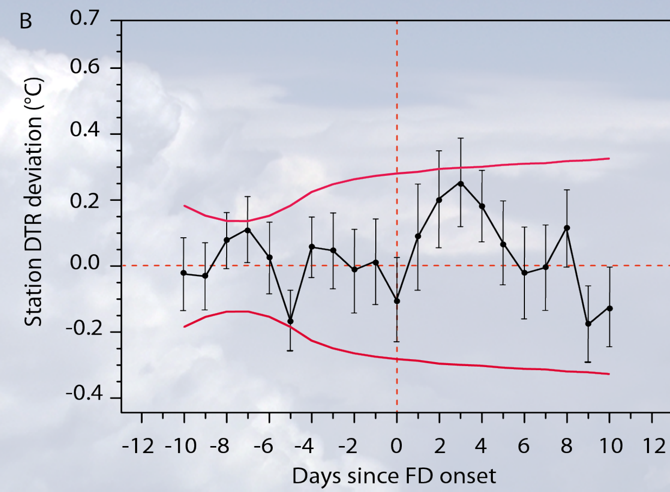
- Surface level Diurnal Temperature Range (DTR) → effective proxy for cloud cover (indirect cloud data)
- DTR has longer time span than satellite cloud observations → allows to have the larger number of Forbush events



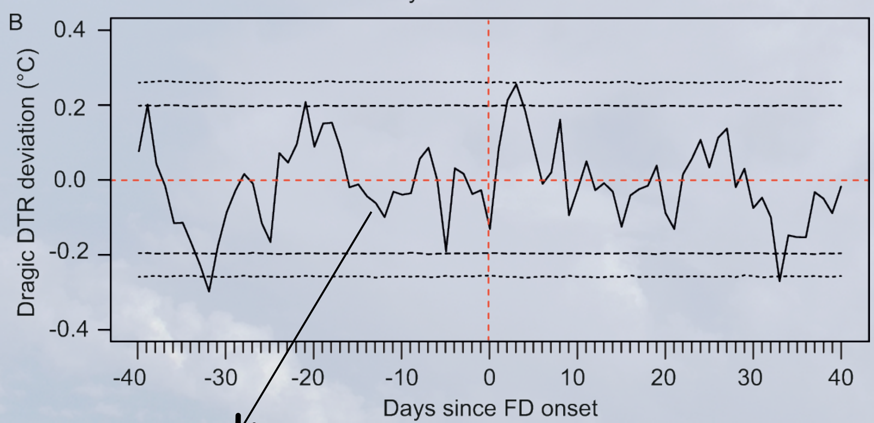
- Dragić et al. (2011) used composite of 37 Fd events (>7%) that show significant increase in DTR → support for GCR-cloud hypothesis

Analysis of Dragić et al. (2011) results





Dragić et al. Normalization of data in period from t_{-10} to t_{-5} and 99% significance intervals

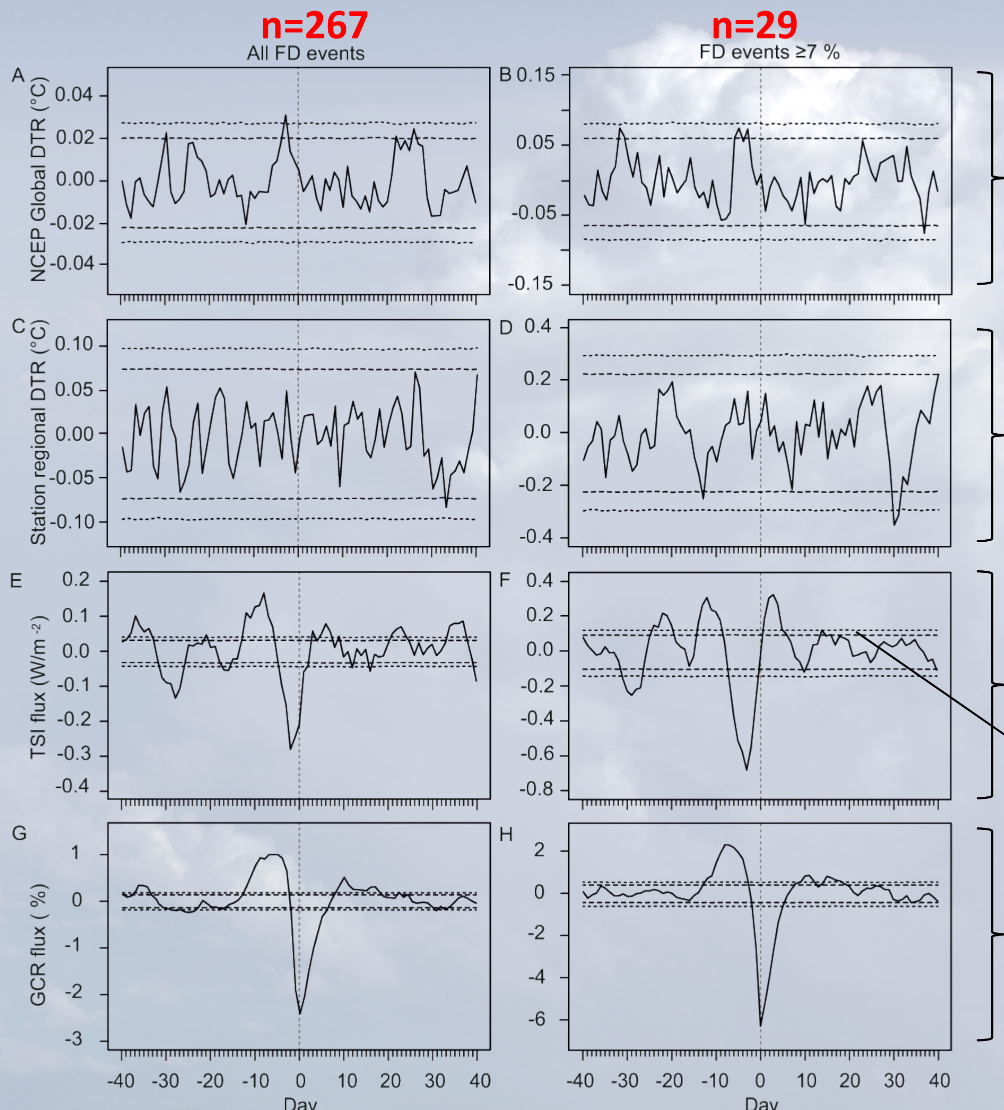


Significance intervals calculated from 100 000 Monte Carlo simulations (using 21-day running average)

Analysis of the same data as in Dragic et al. (DTR data and 37 Forbush events) shows that authors <u>didn't estimate correctly</u> statistical significance using t-test and certain statistical assumptions.

Laken, Čalogović, Shahbaz and Pallé (2012), JGR

Detaljna analiza pokazuje da nema odziva DTR-a tijekom Forbushevih smanjenja



NCEP/NCAR reanalysis data (60°N – 60°S, land-area pixels only)

DTR from 210 meteorological stations (77.7°N – 34.7°N, 179.4°W – 170.4°E)

TSI flux from the PMOD reconstruction

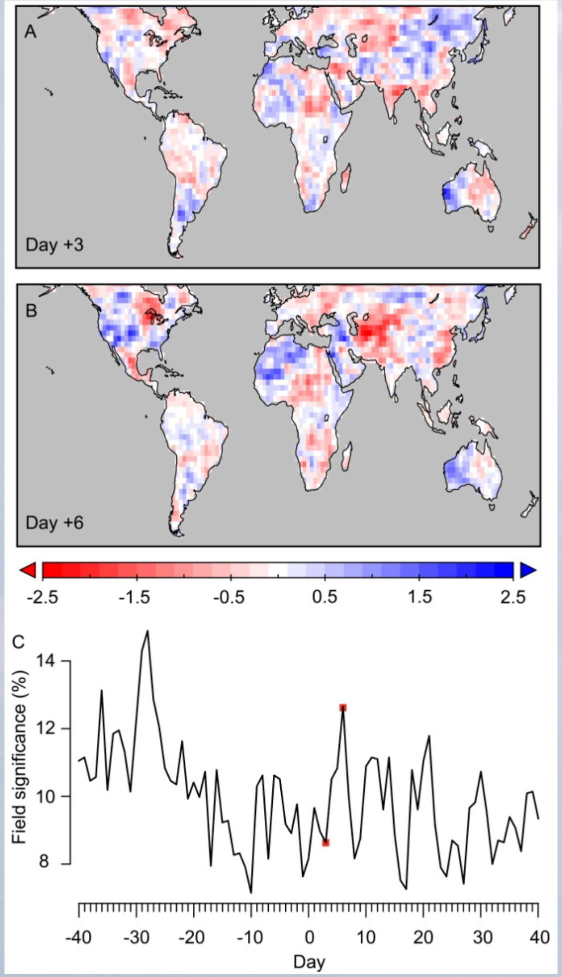
99th and 95th percentile confidence intervals (dotted and dashed lines) are calculated from 100,000 MC simulations

Climax/Moscow NM

Laken, Čalogović, Shahbaz and Pallé (2012), JGR

DTR shows no response to GCR or solar activity

Spatial distribution of DTR anomalies between day +3 and +6



Long term analysis (60 years of data) shows also that there is no significant periodicities in DTR data connected to the solar periodicities (e.g. 11-year, 1.68-year).

In conclusion, there is no evidence to support claims of a link between DTR and solar activity.

Laken, Čalogović, Shahbaz and Pallé (2012), JGR

Various issues that contributed to conflicting results of studies

- **Data filtering** - interference from variability in data at time scales greater than those concerning hypothesis testing, which may not necessarily be removed by accounting for linear trends over the composite periods
- **normalization procedures** which affect both the magnitude of anomalies in composites, and estimations of their significance
- the application of statistical tests unable to account for **autocorrelated data**
- **issues of signal-to-noise ratios** connected to spatio-temporal restrictions (e.g. by decreasing analyzed region size the searched signal may be buried in noise)

Identification of solar—terrestrial links has many difficulties

- Weather and climate are **highly variable** over all time-scales - only a small fraction of this variance (signal) could reasonably be ascribed to solar activity (rest is considered as noise).

- Statistical properties of climatic datasets are unstable (**non-stationary**) – significant correlations over short timescales may disappear.

- climatic data are spatially **autocorrelated** -> number of observations globally doesn't reduce uncertainty -> no good substitute for long duration datasets. **Problem:** modern satellite-era datasets only cover around <u>three solar cycles</u>.

Identification of solar—terrestrial links has many difficulties

- ***a posteriori* selection of data** ("cherry picking") – one sample may have a statistically significant correlation, but drawn from a larger quantity of data which doesn't show the same relationship.

- Exact (amplifying) mechanisms linking solar activity to climate are still poorly understood -> not always possible to evaluate them with models (<u>not testable = unscientific</u>)

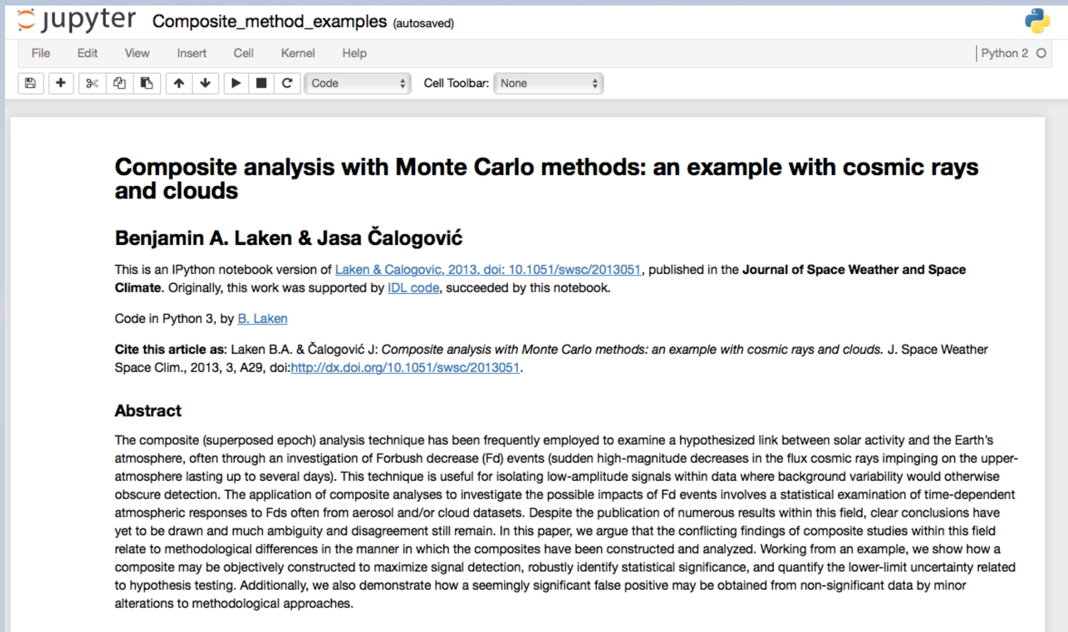
- Most **studies are purely statistical** -> tests of significance may be accompanied by ambiguities in data selection and treatment, applied methods, or assumptions— including human bias, autocorrelations, smoothing, and post-hoc hypotheses.

- Many of these issues already described by **Pittock 1979, 1978**

Open-access coding solutions

- Importance of reliable methods and statistical tests to overcome some of mentioned difficulties: **communal analysis approach**
- Implementation of robust significance testing (e.g. MC method)
- Python (completely free, all computer platforms)
- **iPython**: code in small editable units, descriptions and figures between code. Rapidly shared and replicated, runs in any internet browser
- Simple to run code on remote computers (cloud)
- Public **Git** repositories for instant download of analysis or upload tracked changes
- Allows even low skill programmers to follow the analysis. Viewed online, any system (only internet browser needed)
- Using **FigShare** (DOI number) code can be added as supplement to publications

iPhyton environment



Composite analysis with Monte Carlo methods: an example with cosmic rays and clouds

Benjamin A. Laken & Jasa Čalogović

This is an IPython notebook version of Laken & Calogovic, 2013, doi: 10.1051/swsc/2013051, published in the **Journal of Space Weather and Space Climate**. Originally, this work was supported by IDL code, succeeded by this notebook.

Code in Python 3, by B. Laken

Cite this article as: Laken B.A. & Čalogović J: *Composite analysis with Monte Carlo methods: an example with cosmic rays and clouds.* J. Space Weather Space Clim., 2013, 3, A29, doi:http://dx.doi.org/10.1051/swsc/2013051.

Abstract

The composite (superposed epoch) analysis technique has been frequently employed to examine a hypothesized link between solar activity and the Earth's atmosphere, often through an investigation of Forbush decrease (Fd) events (sudden high-magnitude decreases in the flux cosmic rays impinging on the upper-atmosphere lasting up to several days). This technique is useful for isolating low-amplitude signals within data where background variability would otherwise obscure detection. The application of composite analyses to investigate the possible impacts of Fd events involves a statistical examination of time-dependent atmospheric responses to Fds often from aerosol and/or cloud datasets. Despite the publication of numerous results within this field, clear conclusions have yet to be drawn and much ambiguity and disagreement still remain. In this paper, we argue that the conflicting findings of composite studies within this field relate to methodological differences in the manner in which the composites have been constructed and analyzed. Working from an example, we show how a composite may be objectively constructed to maximize signal detection, robustly identify statistical significance, and quantify the lower-limit uncertainty related to hypothesis testing. Additionally, we also demonstrate how a seemingly significant false positive may be obtained from non-significant data by minor alterations to methodological approaches.

```
p4b = figure(width=400, plot_height=400, title=None, tools=TOOLS)
p4b.quad(top=norm_hist, bottom=0, left=edges[:-1], right=edges[1:],
     fill_color="#ff9900", line_color="#ff9900", legend="MC population")
xvals_pdf = np.linspace(-0.1,0.1,num=100, endpoint=True)
p4b.line(xvals_pdf, gen_pdf(x_val=xvals_pdf, mu=np.average(fit_diffs), sigma=np.std(fit_diffs)),
         legend='PDF',color='black', line_width=1.5)
p4b.legend.location = "top_left"
p4b.xaxis.axis_label = "Diff. b/w two methods to calc. anomalies (at t₀)"
p4b.xaxis.axis_label_text_font_size = '10'
p4b.yaxis.axis_label = "Probability density"
p4b.yaxis.axis_label_text_font_size = '10'

fig4 = gridplot([[p4a, p4b]])
show(fig4)
del fit_diffs # free the memory
```

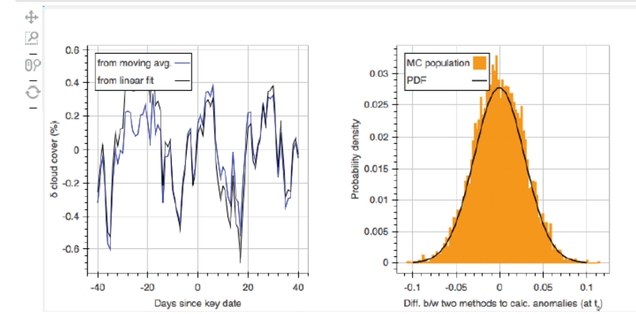


Figure 4 (a) Left panel: one random instance of $n = 20$ events, for both linearly detrended data (black line) and an anomaly from a 21-day moving average (blue line). Differences between these two curves show the possible influences of intermediate timescales on the data. (b) Right panel: the differences between these curves at t_0 for 10,000 random instances (of $n = 20$). The histogram shows a 1σ value of 0.1%, a non-trivial difference which may influence the outcome of a composite analysis.

- *nb. The text above reflects the original paper, but from the re-code it seems* $1\sigma = 0.01\%$. *Not sure yet where the larger deviation came from in the original, but may potentially be a bug.*

Notebook viewer on-line:
http://tinyurl.com/composite-methods

GitHub repository (download and run it locally):
https://github.com/benlaken/Composite_methods_LC13

Conclusions

- Satellite cloud estimates are fraught with limitations and calibration errors, meaning **long-term analysis is problematic at best**, and, as in the case of commonly used ISCCP data, is fundamentally flawed.

- Co-variance of solar-related parameters (UV, TSI, CR flux, solar wind) make **signal attribution difficult**.

- **Climate variability and volcanic activity**, operating over time-scales similar to the solar cycle, make disambiguating causes of cloud cover change difficult.

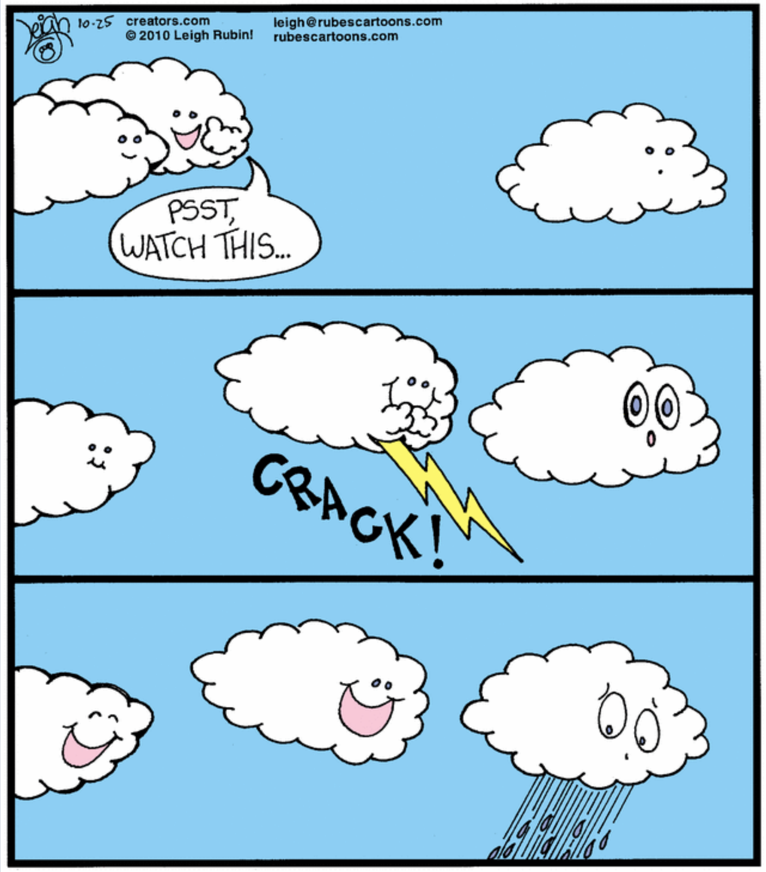
- Composite analysis of FD and GLE events is often compromised by the difficulties of statistical analysis of **autocorrelated data**. This is compounded by the application of inappropriate and black-box statistical tests.

- Changing **signal-to-noise ratios** connected to **spatio-temporal restrictions** in composites have generally not been sufficiently taken into account in composite studies, leading to widespread false-positive statistical errors.

Conclusions

- Methodological differences and inappropriate statistics in composite analysis can produce conflicting results. These are the likely source of discrepancies between cosmic ray – cloud composite studies.
- Present cloud datasets are limited to detect a small changes in cloud cover as well to detect the regional cloud changes (<several thousand km) due to the big natural cloud variability (noise). Thus, localized and/or small effect on cloud cover can't be completely excluded.
- **No compelling evidence** to support a global cosmic ray-link using the <u>satellite</u> cloud data (ISCCP, MODIS) with long- or short-term (Fd) studies.
- If cosmic ray-cloud relationship is second order (small and dynamic changes to cloud cover over certain regions) then it may be very difficult to detect it with currently available techniques and datasets.

Thank you for your attention!



We acknowledge the support of Croatian Science Foundation under the project 6212 „**Solar and Stellar Variability**“ and of European social fond under the project “**PoKRet**”.