AIP Leibniz-Institut für Astrophysik Potsdam

Data Analysis and Management for High Resolution Solar Physics

Carsten Denker, Christoph Kuckein, Meetu Verma, Sergio J. González Manrique, Andrea Diercke, Harry Enke, Jochen Klar, Horst Balthasar, Rohan E. Louis, and Ekaterina Dineva

→ Denker *et al.* 2017, ApJS, submitted

Motivation

- Follow dynamic processes in the photosphere and chromosphere
- Overcome the deleterious effects of the Earth's turbulent atmosphere
- High-cadence imaging and frame selection »lucky imaging«
  → Scharmer 1989, NATO Adv. Sci. Inst. (ASI) Ser. C 263, 161 | Kitai *et al.* 1997, PASJ 49, 513 | Law *et al.* 2009, ApJ 692, 924
- Real-time correction by adaptive optics (AO) systems
  → Rimmele 2000, Proc. SPIE 4839, 635 | Berkefeld *et al.* 2012, AN 333, 863
- Image restoration
  → Löfdahl *et al.* 2007, in Modern Solar Facilities – Advanced Solar Science, Universität Göttingen | Denker *et al.* 2015, in Encyclopedia of Optical and Photonic Engineering, CRC Press
- »Median Filter Gradient Similarity« (MFGS) image quality metric
  → Deng *et al.* 2015, SoPh 290, 1479
- Advances in sCMOS detector technology
- Large volume of high-cadence, large-format images
- Significant computational resources for image restoration and data archiving
  → Denker 2010, AN 331, 648
- Collaborative research environment
  → Denker *et al.* 2017, ApJS, submitted
- Identify best data sets for research
  → Denker et al. 2017, SoPh, submitted

E-Science Projects @ AIP

- German Astrophysical Virtual Observatory
  - Rave DB Query interface
  - MultiDark database
  - German SDSS mirror
- Digitization of astronomical photographic Plates
  - 85000 plates in the archives of Hamburger Sternwarte, Dr. Karl Remeis-Sternwarte Bamberg, and AIP.
  - 7000 solar plates of AOP and Einstein Tower
- DMPwerkzeug
  - Planning of research data management
- **Collaborative Research Environments (CREs)**
  - Collaborations enable efficient science with huge data sets.
  - Data, tools, and facilities to analyze and visualize these data are co-located.
  - CREs are dedicated, customized environments, with huge data storage and customized software and hardware encapsulated by user management through the collaboration.
  - Current CREs: CLUES collaboration, MUSE project, LOFAR solar key science project, GREGOR archive for GFPI, BIC, and HiFI

Daiquiri → escience.aip.de/daiquiri

- A framework for the publication of scientific databases
- Publication of research data
- Data publication becomes a key requirement demanded by funding agencies.
- Development of a new web framework, which is particularly designed to allow for different highly customizable web applications based on a common easily maintainable code base.
- Strive to make Daiquiri as Virtual Observatory (VO) compliant as possible.

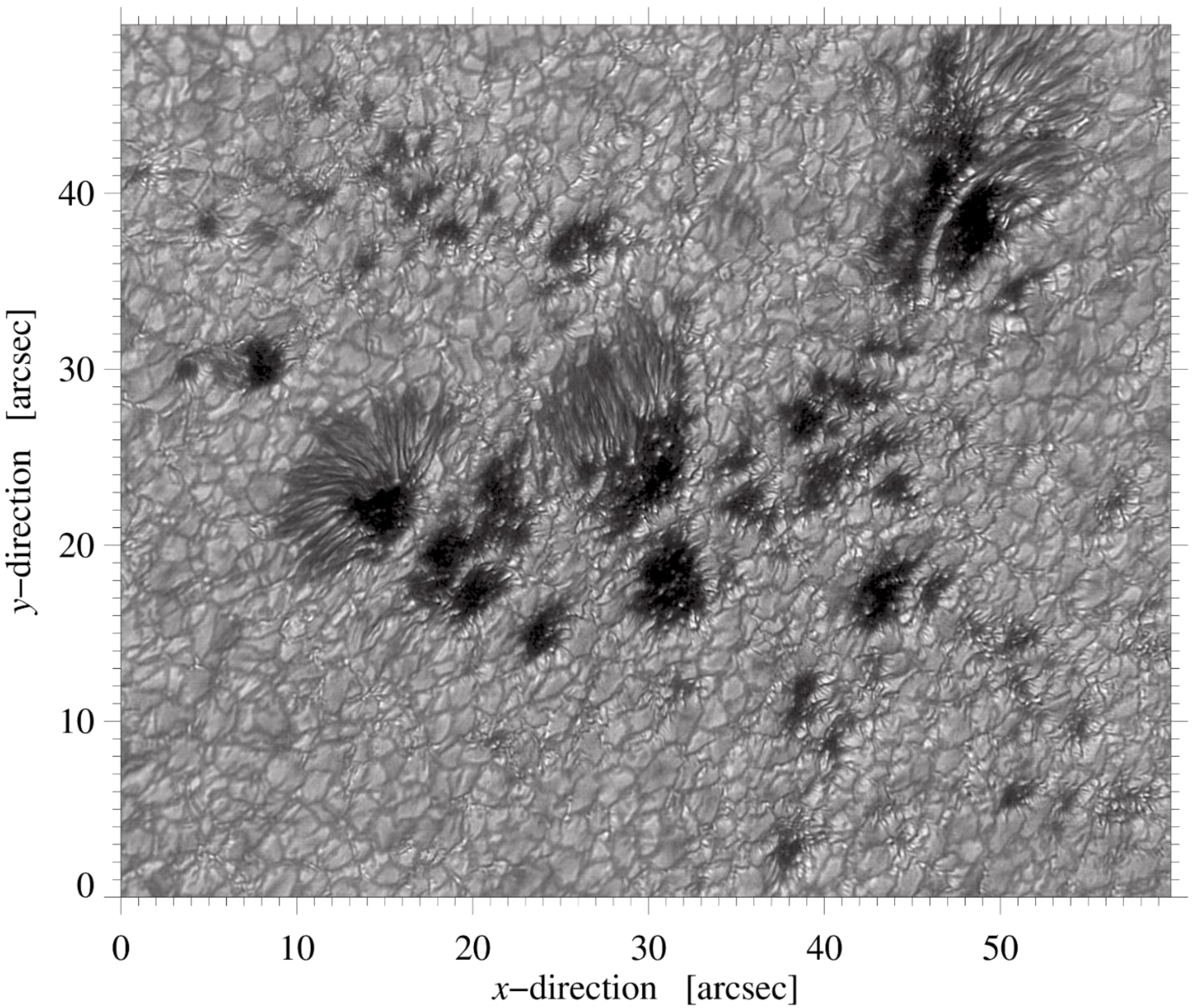
- Features
  - User management and user registration workflow → collaboration meeting
  - Contact messages to the support staff
  - SQL query interface with database browser and example queries
  - Customizable data query forms
  - SQL permission handling
  - Query management MySQL Job Queue
  - Parallel query engine PaQu
  - Database table viewer
  - Viewer for linked images
  - Synchronous and asynchronous downloads (MySQL, CSV, VOTable)
  - Database table and column meta data management
  - File links & downloads via database tables
  - WordPress integration

High-Cadence Imaging and Imaging Spectroscopy

- High-Resolution Fast Imager (HiFI)
  - Scientific Complementary Metal-Oxide-Semiconductor (sCMOS) technology
  - Two synchronized sCMOS detectors
  - Spectral regions: Ca II H λ396.8 nm, Fraunhofer G-band λ430.7 nm, and blue continuum λ450.55 nm
  - 16-bit images, 2560 × 2160 pixels
  - 50 (100) Hz data acquisition rate (HDTV 1920 × 1080 pixels)
  - FOV 65″ × 55″
  - Image scale 0.025″ pixel^{-1} or 18 km on the solar surface
  - Typical cadence Δt = 10 – 20 s with frame selection
  - Sustained write speed 660 MB s^{-1}

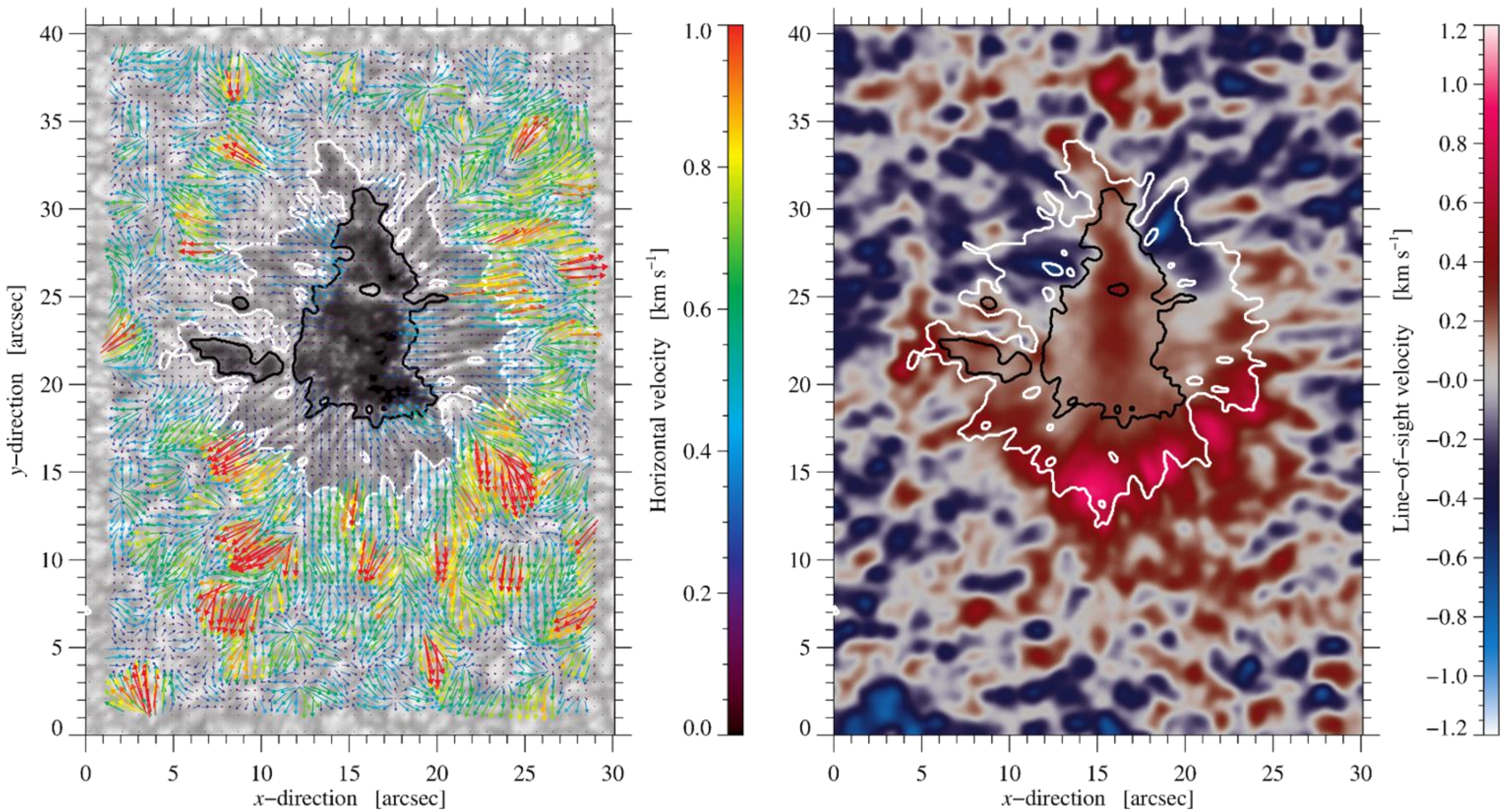
- GREGOR Fabry-Pérot Interferometer (GFPI)
  - Tunable, dual-etalon imaging spectrometer in collimated mounting
  - High finesse ($\mathcal{F}$ = 45 – 50)
  - High reflectivity ($R \approx 95\%$).
  - Coatings optimized for 530 – 860 nm
  - Spectral resolution $\mathcal{R} \approx 250\,000$
  - Imager QE CCD cameras
  - 12-bit images,1376 × 1040 pixels
  - 10 (20) Hz data acquisition rate (2×2-pixel binning)
  - FOV 55″ × 42″
  - Full-well capacity 18 000 e^-
  - Typical cadence Δt = 20 – 60 s for a spectral scan

HiFI Context Images



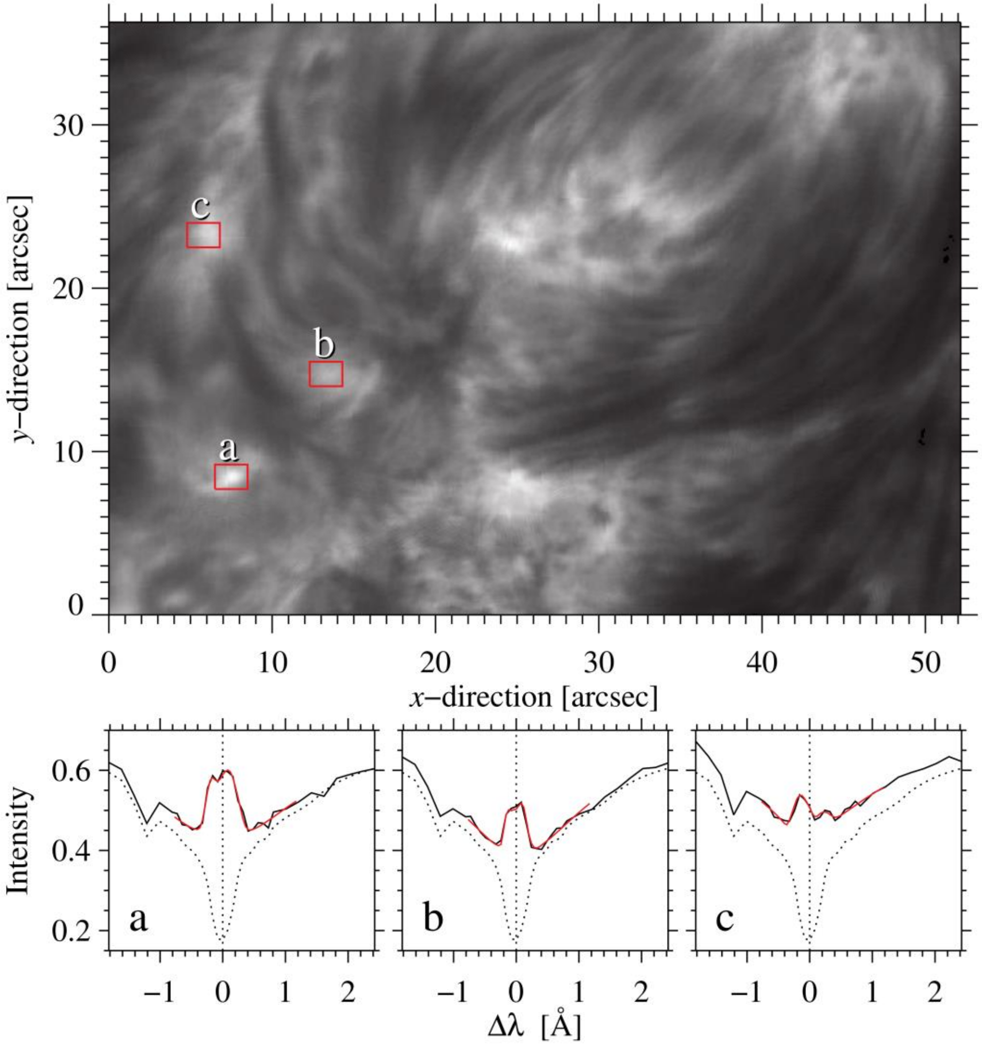
Blue continuum image λ450.55 nm of active region NOAA 12529 obtained with HiFI at 08:37 UT on 2016 April 11. The image was restored from a time-series of 100 short-exposure images with the speckle masking method implemented in KISIP.

3D Velocity Field around Sunspots



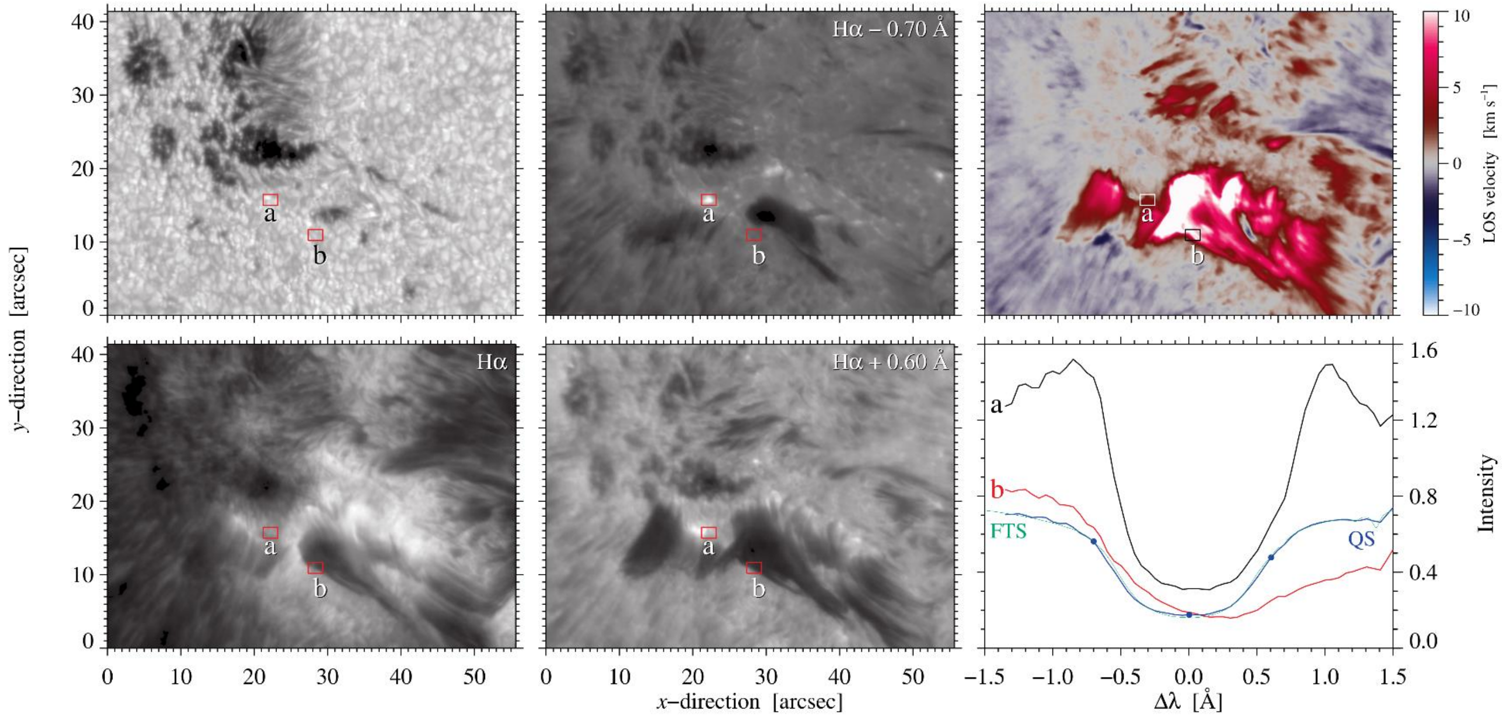
Average horizontal flow field (*left*) and LOS velocity (*right*) of active region NOAA 12597 observed on 2016 September 24. The horizontal flows were derived from a 40-minute time-series of restored GFPI broad-band images at 612.27 nm, whereas the Doppler velocities were determined from a restored narrow-band scan of the photospheric Fe I 617.33 nm line. → Verma *et al.* 2017, A&A, submitted

Chromospheric Micro-Flares



Filtergram at Δλ = −0.16 Å off the Ca II 8542.1 Å line center obtained with the GFPI on 2014 August 26 (*top*). The image was restored with MOMFBD. Selected Ca II intensity profiles (*bottom*) from areas 'a – c'. Observed profiles (*black*) are compared with results from the spectral inversions with NICOLE (*red*). The average quiet-Sun profile (*dotted*) serves as a reference.
→ Kuckein *et al.* 2017, A&A, in press

Chromospheric Downdrafts and Ellerman Bombs

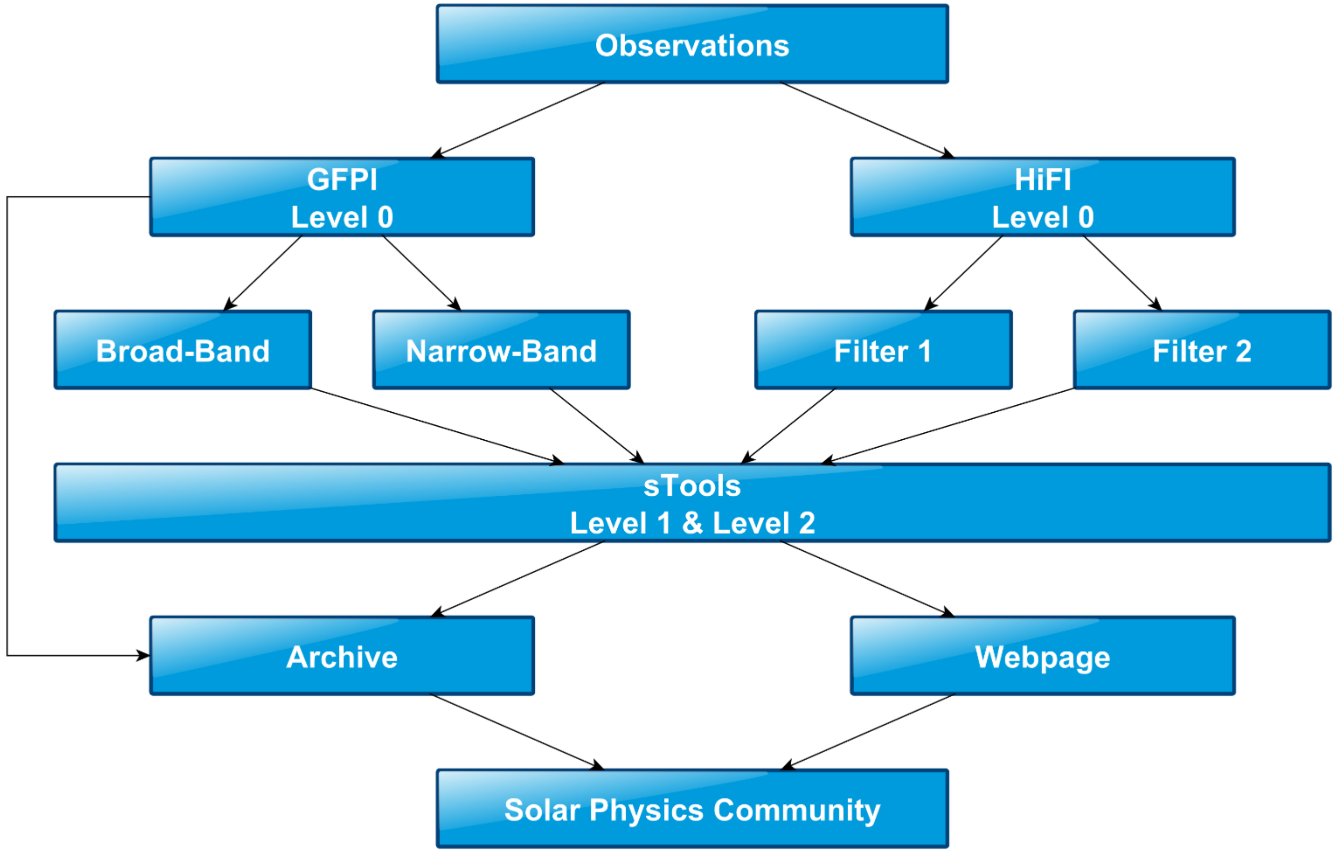


Imaging spectroscopy of active region NOAA 12139 observed on 2014 August 14 with the GFPI in the strong chromospheric absorption line Hα 6562.8 Å. The spectral scan was restored with MOMFBD: broad-band image (*top-left*) near the Hα spectral region, line-core intensity image (bottom-left), blue line-wing image at Δλ = −0.70 Å (*top-middle*), red line-wing image at Δλ = +0.60 (*bottom-middle*), and chromospheric Doppler velocity derived with the Fourier phase method (*top-right*), where blue and red colors represent up- and downflows, respectively.

Data Specific Challenges for GFPI and HiFI

- The campaign- and PI-oriented nature of the data, which results in different combinations of post-focus instruments with changing set-up parameters, e.g., selection of diverse spectral lines, dissimilar spectral and spatial resolution, and different cadences.
- Observations through Earth's turbulent atmosphere, which brings about data with fast-changing quality and necessitates image restoration, where different restoration algorithms introduce multiplicity in the major data levels.
- The complexity of data sets, which requires considerable efforts to condition the data products for a broader user base – even at the level of quick-look data.
- The readiness of personnel and financial resources for maintaining long-term access to data and for quality assurance beyond the typical funding cycle of third-party financial support.
- **AIP's mission includes the development of research technology and e-infrastructure as a strategic goal.**

Data Management



Overview of the data flow from observations, over data processing with sTools, to distribution of science-ready data sets to the solar physics community.

Results: Data Management

- Data Policy
  - GREGOR consortium agreed that in principle all data shall be publicly available.
  - Observing proposals contain proprietary information and original ideas of the PI and her/his team.
  - Data of PI-led observing campaigns will be embargoed for one year.
  - Embargo period can be extended another year for PhD thesis.
  - Quick-look data are available after storage at the GREGOR GFPI and HiFI archive and subsequent processing.
  - Cite GREGOR reference articles: Astron. Nachrichten 2012, **333**(9)

- GREGOR GFPI and HiFI Data Archive
  - Daiquiri framework
  - Elaborate permissions system → access restrictions and data security
  - User management and user registration work flows
  - Members receive CRE privileges, allowing them processing data on institute-owned computers and editing of webpages and blog entries. → organizational confirmation
  - Along with accounts for web portal, members obtain Secure Shell (SSH) login credentials for data access node.

gregor.aip.de News Instruments Observations Contact Member Login

GREGOR GFPI/BIC/HiFI Archive

This webpage provides access to the »quick-look« data products of GFPI (Levels 0.0 and 1.0), HiFi (Level 1.0), and BIC (Level 0.0). Access to the archived raw data and derived data products requires registration.

2017

June
GFPI Level 0.0: [02] [03] [04] [05] [06] [08] [09] [10] [16] [17] [18] [19] [20]
GFPI Level 1.0: [16] [17] [18] [19] [20]
HiFI Level 1.0: [16] [17] [18] [19] [20]

May
GFPI Level 0.0: [05] [07]
GFPI Level 1.0: [05] [07]
HiFI Level 1.0: [05] [07]

April
GFPI Level 0.0: [19] [20] [21]
HiFI Level 1.0: [04] [19] [20] [21]

March
HiFI Level 1.0: [28] [29]

2016

October
GFPI Level 0.0: [19] [27] [28]
GFPI Level 1.0: [27] [28]
HiFI Level 1.0: [27] [28]

September
GFPI Level 0.0: [21] [22] [23] [24] [26] [28] [29]
GFPI Level 1.0: [21] [22] [23] [24] [26] [28] [29]
HiFI Level 1.0: [21] [22] [23] [24] [26] [28] [29]

Search

Imprint & Data Protection Statement

Currently, reprocessing of all "quick-look" data (ready by end of October).

	GFPI	BIC	HiFI
2014	48	37	
2015	69	42	
2016	41	4	46
	8.3 TB	13.6 TB	12 TB

Number of data sets and stored data volume for GREGOR's instruments.

60 registered users (10 from outside the GREGOR consortium.)

Median Filter Gradient Similarity (MFGS)

(1) raw image A and (2) median-filtered image $B = F(A)$

(3) directional derivative $G = |G_x|$ (see Deng *et al.* 2015)

(3′) gradient magnitude $G_A = G(A)$ and $G_B = G(B)$

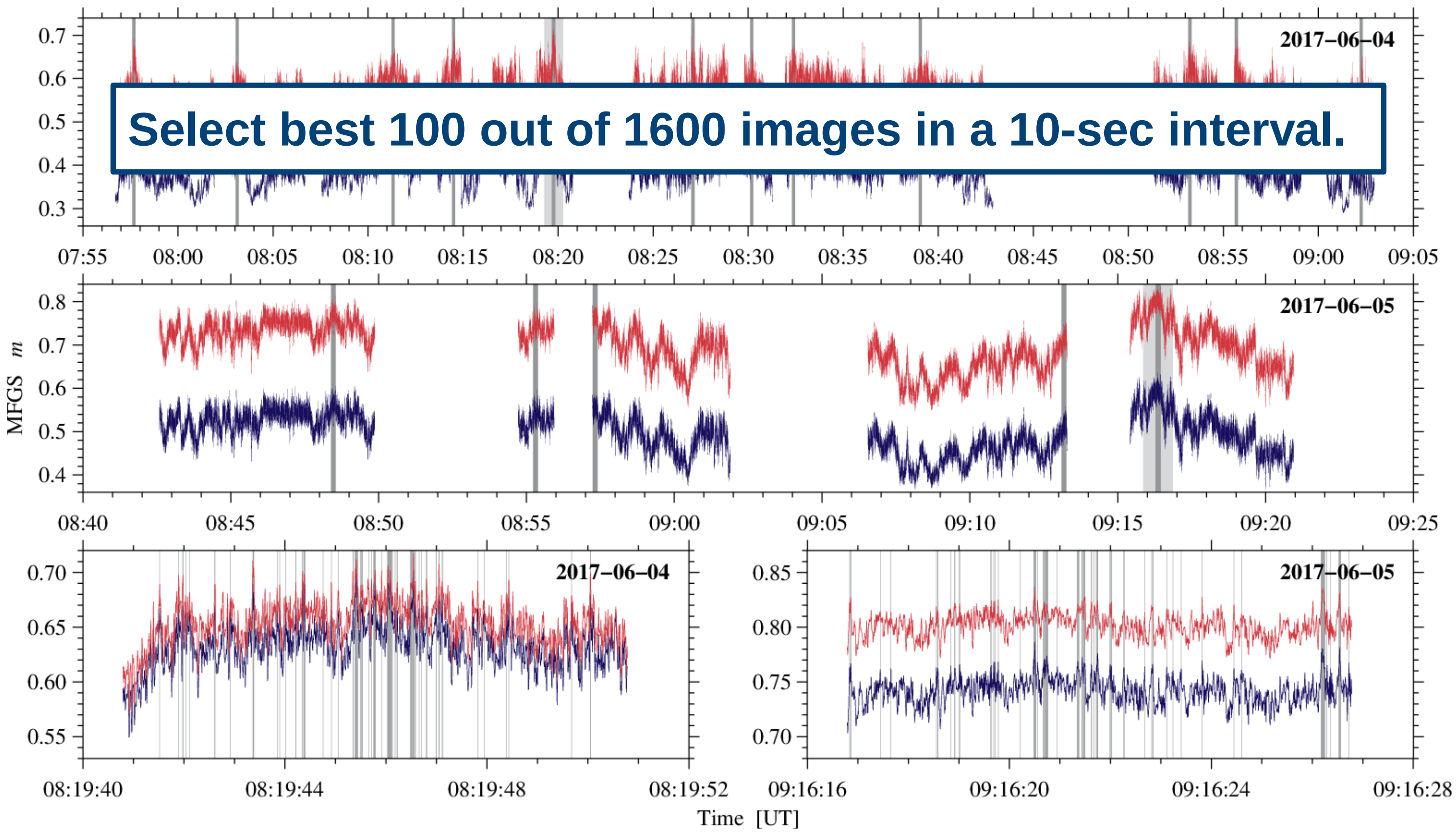
$G = \sqrt{G_x^2 + G_y^2}$ (Scharr operator)

$$G_x = \frac{1}{16}\begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix} \text{ and } G_y = \frac{1}{16}\begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix}$$

$$m,\ m' = \frac{2\left(\sum G_A\right)\left(\sum G_B\right)}{\left(\sum G_A\right)^2 + \left(\sum G_B\right)^2}$$ (see Deng *et al.* 2015)

$$m^* = \sum\left(\frac{2G_A G_B}{G_A^2 + G_B^2}\right)$$ (local, pixel-by-pixel implementation)

Temporal Evolution of Image Quality

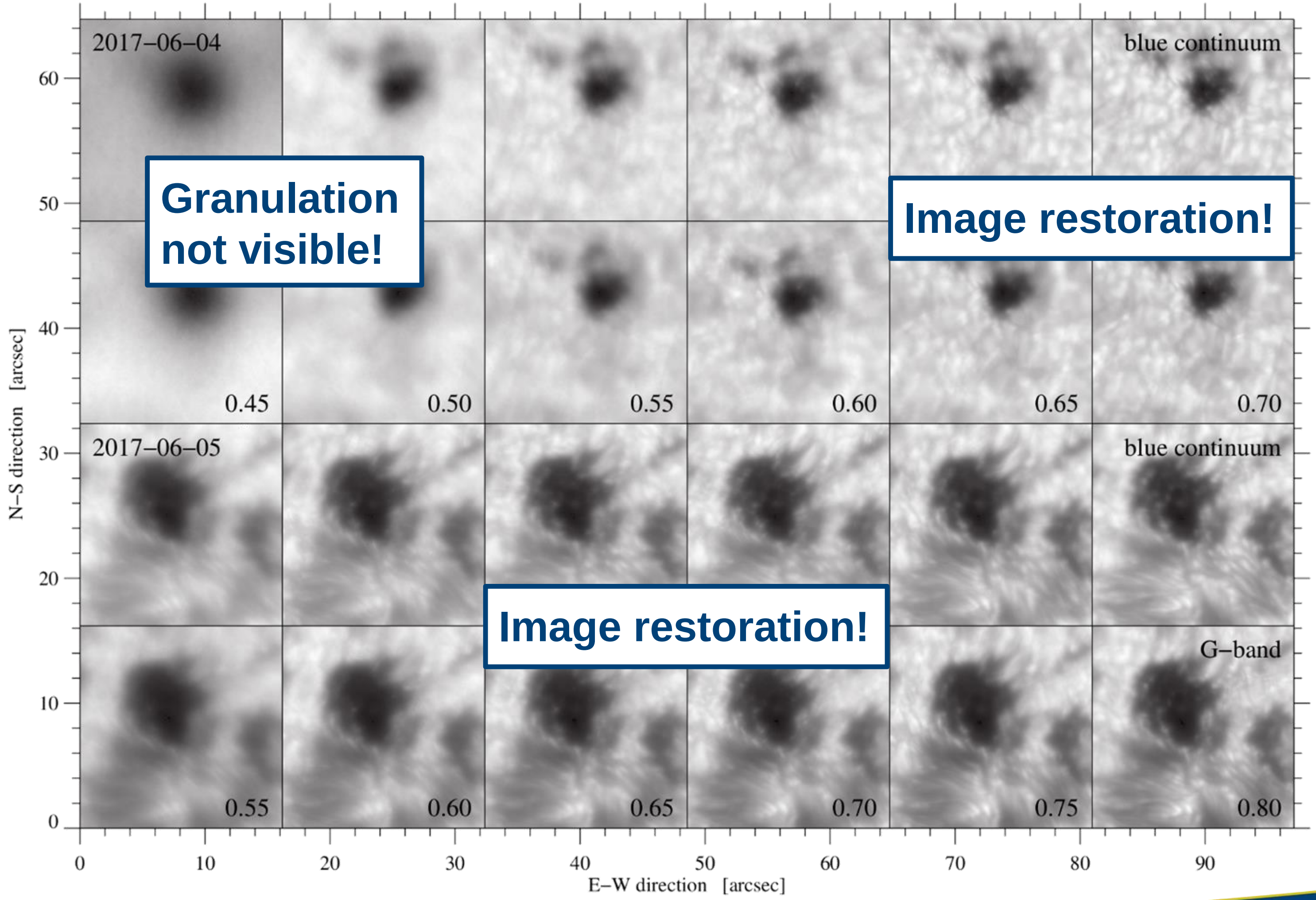


Blue continuum MFGS values are displaced by 0.15 downwards to facilitate better comparison.

No offset!

Fried-parameter $r_{0,\max}$ = 10 and 15 cm on June 4 and 5, respectively (GAOS total mode wavefront error). Image quality (MFGS) variations trace the seeing conditions, which vary on millisecond scales.

Visualization of Image Quality (MFGS)

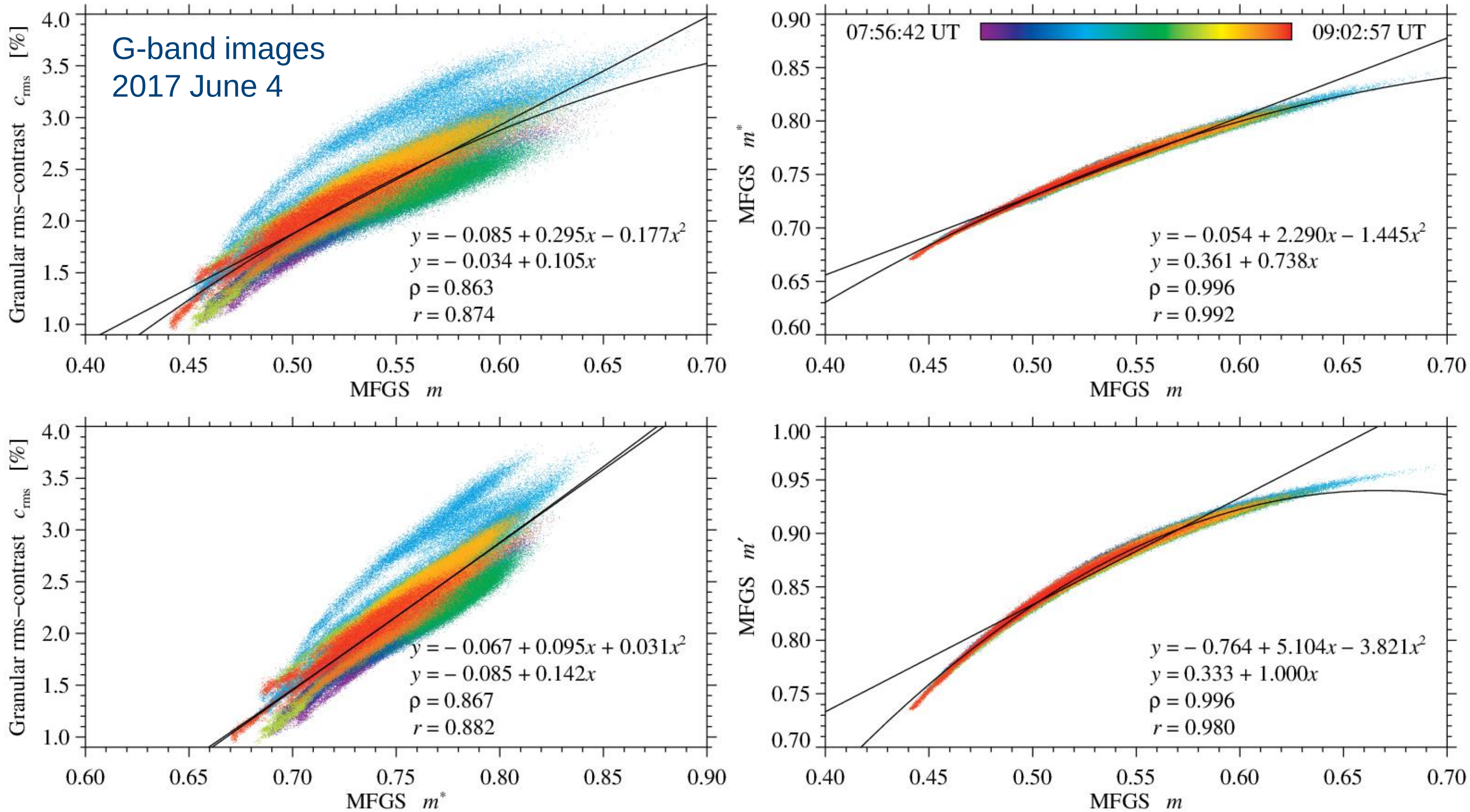


m > 0.65 fulfilled in 9 out of 277 possible 10-sec intervals.

m > 0.65 fulfilled in 143 out of 150 possible 10-sec intervals.

m > 0.75 in one third of the cases.

Correlations between Image Quality Metrics

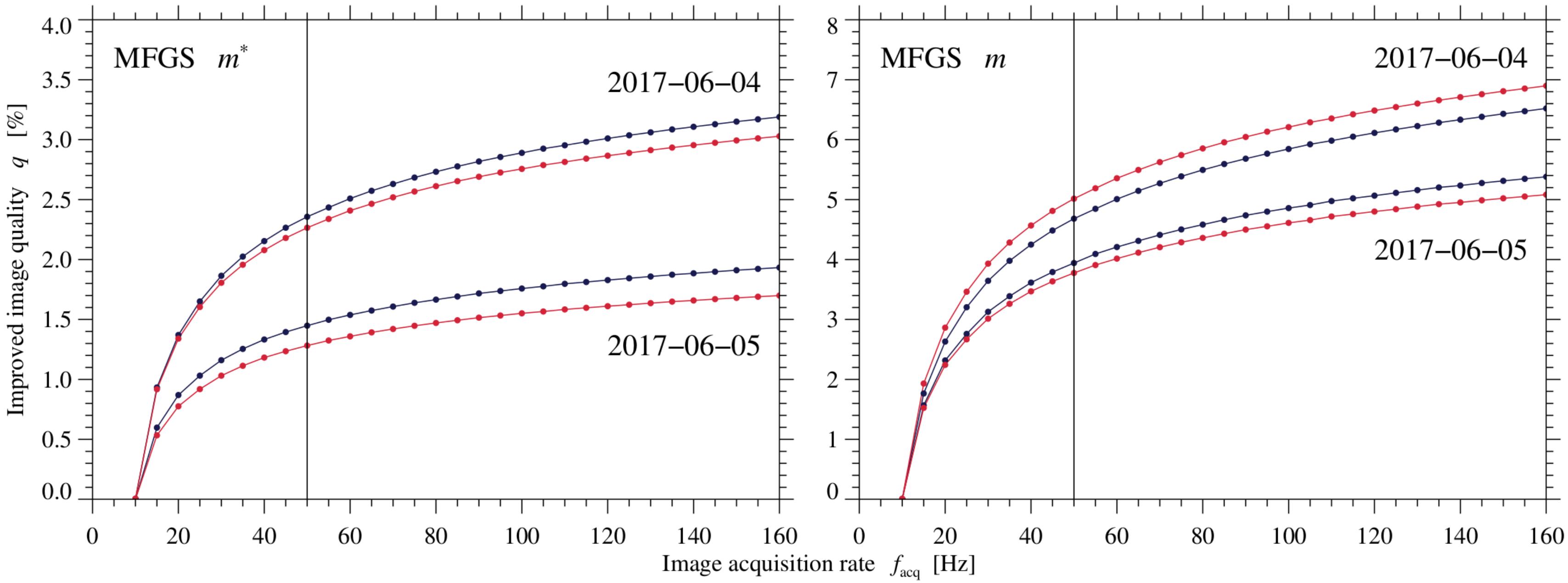


No granulation on 2017 June 5!

Contrast of pores and sunspots is one order of magnitude higher.

Monotonic relationship among image quality metrics, which are well represented by linear or parabolic models. The granular rms-contrast shows a much larger scatter than the MFGS metrics.

Impact of Image Acquisition Rate



$$q\left(f_{\mathrm{acq}}\right) = \frac{100\%}{N_{\mathrm{set}}} \sum_{i=1}^{N_{\mathrm{set}}} \frac{\bar{m}_i\left(f_{\mathrm{acq}}\right) - \bar{m}_i}{\bar{m}_i} \quad \text{with}$$

$$\bar{m}_i = \frac{1}{n_{\mathrm{set}}} \sum_{j=1}^{n_{\mathrm{set}}} m_{i,j} \quad \text{and} \quad \bar{m}_i\left(f_{\mathrm{acq}}\right) = \frac{1}{n_{\mathrm{sel}}} \sum_{j=1}^{n_{\mathrm{set}}} \mathrm{best}\left(n_{\mathrm{acq}}, m_{i,j}\right)$$

At higher frequencies, the image quality improves monotonically, with most of the gain at frequencies below f_{acq} = 50 Hz.

Image Restoration KISIP → Wöger & von der Lühe 2008, Proc. SPIE 7019, 70191E

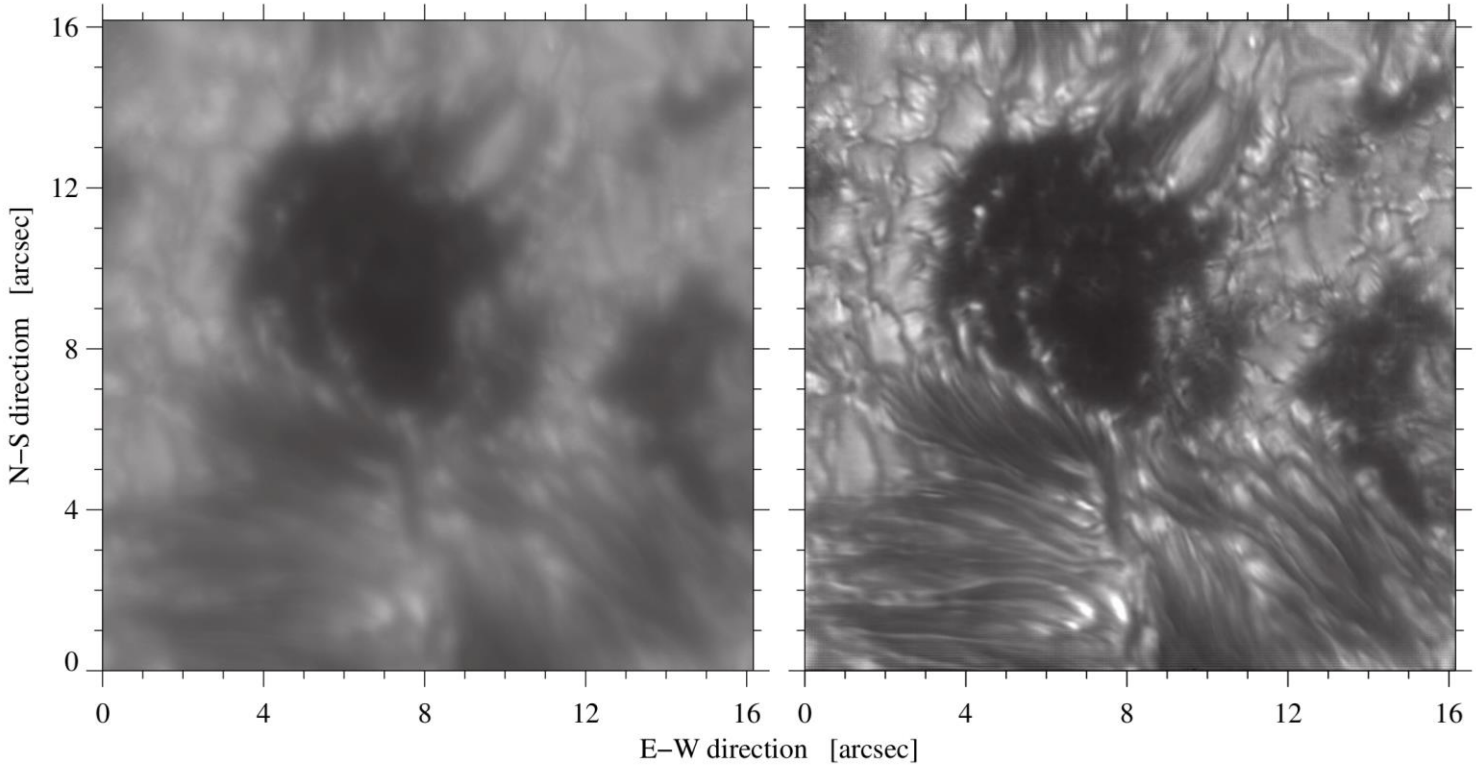


Image restoration affects both amplitudes and phases in the Fourier domain.

The restored 3×3 sub-images have a size of 256 × 256 pixels, which corresponds to 6.5″ × 6.5″.

G-band image m = 0.84 and restored image m = 0.99.

Obtaining sets of high-quality images is the goal of frame selection and input for post-facto image restoration (speckle masking technique).

→ Lohmann *et al.* 1983, Appl. Opt. 22, 4028 | Weigelt & Wirnitzer 1983, Opt. Lett. 8, 389 | de Boer 1993, PhD Thesis, Universität Göttingen | von der Lühe 1993, A&A 268, 374

Results: High-Cadence Imaging

- The granular rms-contrast is commonly used to characterize seeing conditions and image quality relying on the uniform and isotropic properties of granulation.
- Modifications of the MFGS algorithm uncover the field- and structure-dependency of this image quality metric.
- MFGS metrics outperform the rms-granular contrast.
- If the seeing conditions are moderately good, frame selection improves the image quality more than on days, when the seeing is already very good.
- Moments of good seeing occur in clusters but still show significant variations.
- Fast cadence and millisecond exposure times are still insufficient to reach the coherence time of daytime seeing.
- Frame selection significantly aids image restoration, if the data acquisition rate exceeds 50 Hz.
- Investigate the dependency of MFGS on wavelength, morphology, scale, and center-to-limb variation.
- Increase the duty cycle (photon-efficiency) of detectors.