The Solar Science Data Center and LOFAR

F. Breitling, G. Mann, C. Vocks

2009-01-27

1. Introduction

Solar astronomy is a multi-disciplinary field where knowledge and data of various experiments has to be combined and analyzed. In the framework of the solar key science project (SKSP) “Solar Physics and Space Weather with LOFAR” (Mann et al. 2008) the Astrophysical Institute Potsdam is developing a “Solar Science Data Center” (SSDC) where solar data of LOFAR and various other experiments is archived and provided to the astronomy community. LOFAR, the "Low Frequency Array", is the most advanced radio interferometer coming to operation. It will provide radio images of unprecedented high resolution in the frequency regime from 30-240 MHz. The images are expected to make an important contribution to solar science. Solar observations with LOFAR are coordinated through the SKSP. Here we discuss the amount of data that can be expected.

2. Yearly observation time

The data volume per year can now be calculated, if the yearly observation time is known. The possible observation time is determined by the visibility of the sun. It is assume that useful data can be taken for an altitude of more than 10° above horizon. At Exloo, where the core stations for the solar observations are located, a yearly average of 8 hours per day is possible (Fig. 1 and 2). So the total observation time per year t_y= 365*8*3600s=10x10^{6}s.

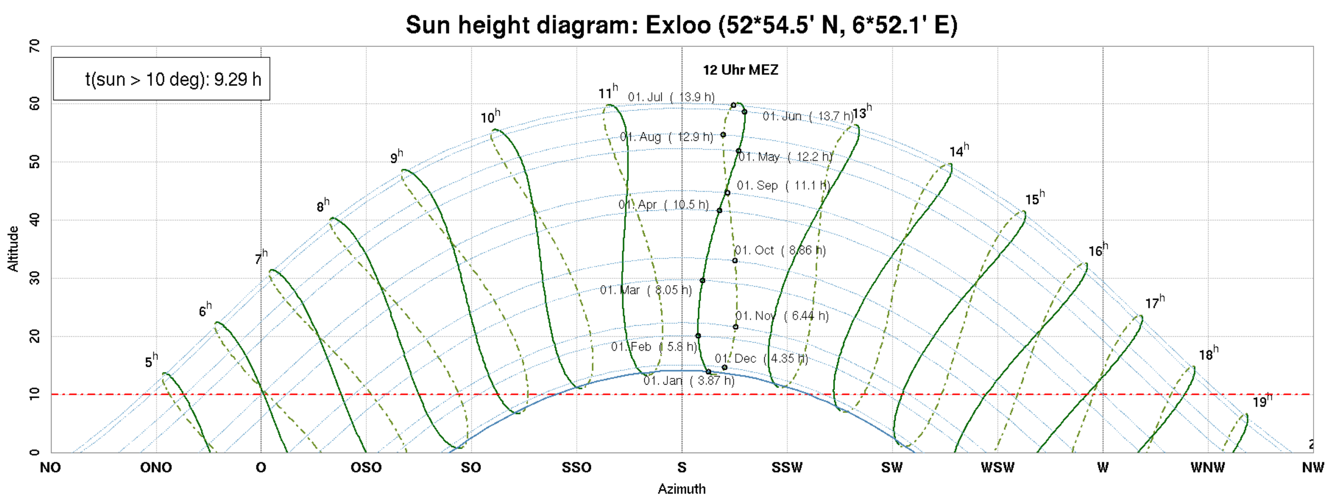

Illustration 1: Sun altitude diagram for one year at Exloo.

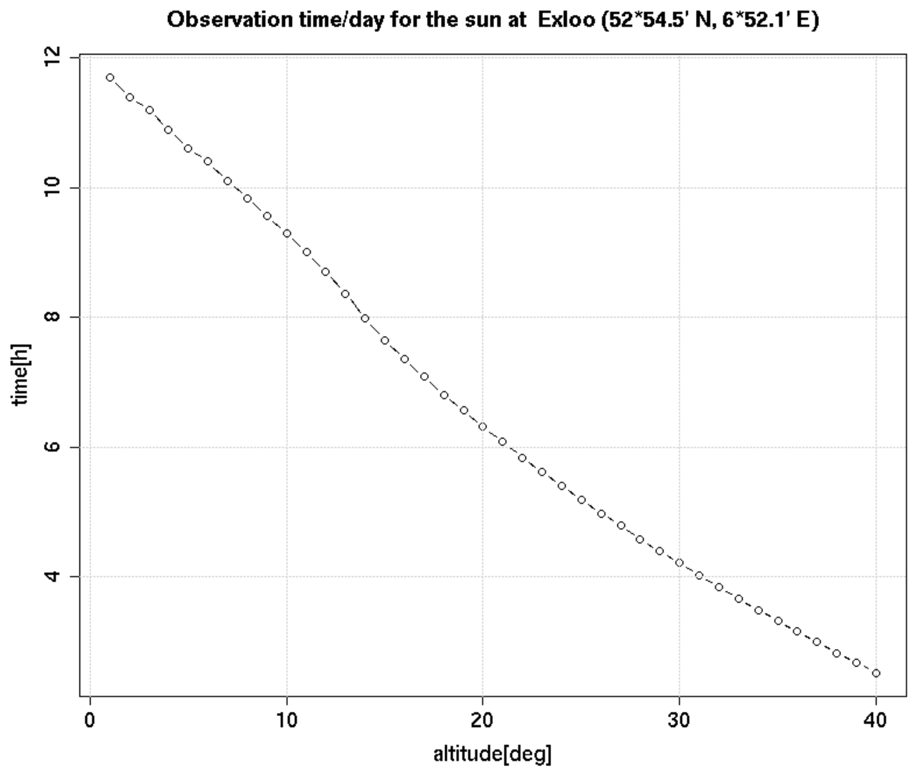

Illustration 2: Yearly average of observation time per day for the sun above a minimum altitude.

3. LOFAR Observation Modes

2.1 Spectrometer Mode

LOFAR can provide spectroscopic data at a millisecond rate. Spectroscopic data can be provided by any station independently of the core stations. Spectroscopic data with a time resolution of 0.01s is desirable for the investigation of fast acceleration processes in burst mode. In the frequency range from 30-240 MHz (170 MHz) LOFAR can monitor the intensity at a spectral resolution of 100 kHz, i.e. 170MHz/100kHz=1700 channels. An intensity value is represented by 2 bytes. With these parameters the corresponding data rate is

b_s=100/s *2B *1700=340kB/s.

In a year this leads to a data volume of B_s=340kB/s*10*10^6s= 3.4TB.

2.2 Burst Mode

For the study of eruptive processes such as bursts and coronal mass ejections observations at the 22 different frequencies are desirable to resolve the solar activity at different coronal depths. The frequencies are given in the table below.

Frequency Nr.	1	2	3	4	5	6	7	8	9
Frequency [MHz]	40	45	50	55	60	65	70	75	80

Frequency Nr.	10	11	12	13	14	15	16	17	18	19	20	21	22
Frequency [MHz]	120	130	140	150	160	170	180	190	200	210	220	230	240

The number of simultaneously observable frequencies at imaging mode (x) is determined by the number (n) of LOFAR stations to be correlated, the sample size of the station signal (b), the sample rate (r) and the bandwidth of the correlator (L) as

$$x\frac{(n-1)n}{2}rb \leq L \quad \Leftrightarrow \quad x \leq \frac{L}{0.5(n-1)nrb}$$

According to recent communication L=50Gb/s. For the imaging observation mode these parameters are listed in section 6.3.1 of the “LOFAR Architectural Design Document (2003-05-16)” as follows:

Parameter	n	b	r	L
Value	40	195*16b	100/s	50Gb/s

=> x ≤ 50Gb/s / [780 *100/s * 195*16b] ≤ 200.

Consequently 22 frequencies are far below LOFAR's capabilities.

In addition a high timing resolution is important. An image rate of 100/s has established as a standard of the field, but 10/s is still sufficient. The resolution of an image is limited by turbulences in the corona to a few 10''. Due to coronal emission, the diameter of the sun at radio wavelength (d_r) is about twice the diameter at optical wavelength, i.e. d_r≈60'≈3600''. With the resolution limit 360*360 different regions can be resolved. Therefore an image resolution of 1024*1024 pixels with 2 bytes per pixel are planed. This corresponds to 2 MB of data or 0.45MB, if lossless image compression is applied. So the data rate in burst mode

b_b= 0.45MB*22*10/s=100MB/s ≤ 1Gb/s.

This is less then the limiting bandwidth of the network connection from the LOFAR data center in Groningen of 1 Gb/s.
A ring buffer of 2 hours would require a size of $B_r=b_b*t_r$=100MB/s*2*3600s=720MB.
The amount of data collected in one year of observations with an average of 10% of observations done in burst mode results in $B_y=10\%*b_b*t_y$=0.1*100MB/s*10x10^7s= 100 TB.

2.2 Monitoring Mode

The Sun is an active star influencing the Earth's environment and our technical civilisation. That is usually called Space Weather. Therefore, the monitoring of the sun's activity is very important to deliver an input to the space weather forecast. For doing that a lower cadence of 1/min is sufficient. During these periods only observations at four frequencies are necessary, which are given at the table below.

Frequency Nr.	1	2	3	4
Frequency [MHz]	40	80	120	240

Therefore, during a major fraction of the observation time, the data rate reduces to

b_m= 0.4MB*4*1/60s=26kB/s.

However, the observation cadence should not be decreased too much in monitoring mode for two reasons: no reliable methods for an early detection of bursts exist, so that important information about the beginning of a burst would be lost; averaging over several images is more robust against fluctuations. Therefore a ring-buffer of two hours is set up, where data older than two hours is reduced to monitoring data, if no burst occurred.
Since bursts only last for a few minutes to hours 90% of time the data will be reduced to monitoring data. The resulting amount of data collected in one year is

B_m=0.9*b_m*t_y=0.9*26kB/s*10*10^6s= 0.23TB.

4. Estimated Data Volume per Year

During a burst, the spectrum contains important information and should be recorded. The additional amount of data is negligible. In monitoring mode the spectral data constitutes the major fraction but could be abandoned. A decision about whether spectral data should be stored in monitoring mode or not, will be made after first data has been analyzed. Here it is assumed that the spectrum is taken during all solar observations. The estimates are summarized in the table below.

Table 1: Comparison Data Volume per Year

Data type	Data rate [MB/s]	Data per year [TB/yr]	Number of files per hour
Burst mode	100	100	$720*10^3$
Monitoring mode	0.03	0.23	240
Spectrometer mode	0.34	3.4	1

5. Data Volume of Visibilities at the Cluster in Groningen

To obtain the final data products archived at the SSDC the visibility data from the correlator has to be processed. Here we calculate the amount of visibility data that will pile up during one hour at the correlator, if compute resources for the postprocessing are temporarily not available. The equation for the amount of visibility data (B_V) is
$B_V = N_s(N_s+1)/2 * N_{sb} * N_{ch/sb} * 4 * t/\tau_a * 64bits$,

where

N_s is the number of stations in the array,
N_{sb} is the number of subbands being used a typical value is ~200, but it can be up to 216 for the 16 bit mode, 432 for the 8 bit mode, and 864 for the 4 bit mode,
$N_{ch/sb}$ is the number of channels per subband, which defaults to 256 for most observations for RFI removal, and you need to have good calibration before you can average that down,
4 is the number of polarization cross-correlations, which is always 4 because LOFAR needs full polarization calibration,
t is the length of the observation,
τ_a is the correlator averaging time (duration of a single visibility) and
64bits is the size of a single visibility number (a complex float).

From the current LOFAR setup and the parameters for the burst mode we find the following values

Parameter	N_s	N_{sb}	$N_{ch/sb}$	t	τ_a
Value	30	22	256	1h	0.1s

So the required storage for one hour is

$B_V = 30(31)/2 * 22 * 256 * 4 * 3600s/0.1s * 64bits = 3.0TB$.