



---

BACHELORARBEIT AM MAX-PLANCK-INSTITUT  
FÜR SONNENSYSTEMFORSCHUNG GÖTTINGEN  
ABT. SOLARE VARIABLILITÄT UND KLIMA

---

**Effects of Spatial Resolution on Measurements of Total  
Solar Magnetic Surface Flux**

---

**Bedeutung von räumlicher Auflösung in der Messung des  
gesamten magnetischen Flusses an der Sonnenoberfläche**

---

Belal Kevin Raza aus Hamburg-Bergedorf  
belal-kevin.raza[at]stud.uni-goettingen.de

Erstgutachterin: Dr. Natalie Krivova  
MPI für Sonnensystemforschung Göttingen  
Solare Variabilität und Klima

Zweitgutachter: Prof. Dr. Stefan Dreizler  
Georg-August-Universität Göttingen  
Institut für Astrophysik

Abgabedatum: 24.04.2017



# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data, Noise and Processing</b>	<b>4</b>
2.1	HMI Line-of-Sight Magnetograms . . . . .	4
2.2	Synoptic Chart Construction . . . . .	8
2.2.1	Total Unsigned Flux at higher Latitudes . . . . .	9
2.2.2	Foreshortening Effect . . . . .	9
2.3	Reducing the Spatial Resolution of Synoptic Charts . . . . .	10
2.4	Active Region/Quiet-Sun Separation Techniques . . . . .	12
2.4.1	Intersection Algorithm . . . . .	14
2.4.2	Approximations in the Separation Process . . . . .	15
<b>3</b>	<b>Theoretical Explanation of Magnetic Flux Cancellation</b>	<b>17</b>
<b>4</b>	<b>Analysis</b>	<b>21</b>
4.1	Integration Time and Noise . . . . .	22
4.2	Comparing Measurements of Total Unsigned Flux from HMI,MDI-A and KP-A Observations . . . . .	24
4.2.1	HMI Synoptic Chart Magnetograms . . . . .	24
4.2.2	Creating artificial MDI and KP Synoptic Charts . . . . .	25
4.2.3	MDI-A and KP-A Data . . . . .	26
4.3	Effects of Spatial Resolution . . . . .	29
4.3.1	Empirical Model . . . . .	31
4.3.2	Spatial Resolution and Statistical Moments . . . . .	34
<b>5</b>	<b>Semi-Empirical Model for the 1st and 2nd Statistical Moment of Degraded Synoptic Charts</b>	<b>36</b>
5.1	Motivation . . . . .	36
5.2	Testing for Log-Normality . . . . .	36
5.3	Dynamics of Statistical Moments . . . . .	38
5.4	1D Approximation . . . . .	39
5.5	From 1D to 2D . . . . .	44
5.6	Estimation of Model Parameters . . . . .	47
<b>6</b>	<b>Conclusion</b>	<b>54</b>

<b>7 Appendix</b>	<b>57</b>
7.1 HMI Data . . . . .	57
7.2 MDI-A Data . . . . .	58
7.3 KP-A Data . . . . .	59
7.4 HMI/KP - Total Unsigned Flux Loss . . . . .	60
7.5 Covariance . . . . .	61
7.5.1 Vertical Covariance . . . . .	61
7.5.2 Model Parameters . . . . .	62
7.6 Intersection Algorithm IDL Program Code . . . . .	63

## List of Figures

1	Noise Array - Solar Disc . . . . .	5
2	Noise Radial Dependance . . . . .	6
3	Cycle 2097 Synoptic Chart . . . . .	8
4	Cycle 2147 Synoptic Chart . . . . .	8
5	Effect of Image Degradation . . . . .	11
6	Test Region . . . . .	13
7	Test Region - Polarity . . . . .	13
8	Logical Mask Comparison Intersection Algorithm - RMSD Method . .	16
9	Flux Cancellation Visualized . . . . .	19
10	Noise Distributions . . . . .	23
11	Lost signal in longitudinal band . . . . .	28
12	Ratio HMI/Degraded Active Region Magnetic Flux . . . . .	29
13	Ratio HMI/Degraded quiet-Sun Magnetic Flux . . . . .	30
14	Ratio HMI Degraded . . . . .	31
15	Fit on Ratio HMI-Degraded Total Magnetic Flux . . . . .	32
16	Variance vs Total Resolution . . . . .	35
17	Horizontal Covariances . . . . .	35
18	Histogram . . . . .	37
19	Log-Normality Test . . . . .	37
20	Mean Model Test . . . . .	45
21	Horizontal Covariances . . . . .	48
22	Eta . . . . .	50
23	Pixel-by-Pixel Mean . . . . .	51
24	Model Test . . . . .	53
25	Vertical Covariance Cycle 2147 . . . . .	61

## List of Tables

1	Average Noise Level . . . . .	5
2	Noise Pixels . . . . .	6
3	Flux Distribution at Meridian . . . . .	9
4	Total Flux - Higher Latitudes . . . . .	9
5	Residual-Sum-of-Squares Noise . . . . .	22
6	HMI-MDI Loss . . . . .	27
7	Flux Loss Model Parameters . . . . .	33
8	Distance Counting . . . . .	44
9	Size Magnetic Elements . . . . .	49
10	Estimation of alpha . . . . .	50
11	HMI Data . . . . .	57
12	MDI-A Data . . . . .	58
13	KP-A Data . . . . .	59
14	HMI/KP-A . . . . .	60
15	Covariance Model Parameter . . . . .	62
16	Covariance Model Parameter . . . . .	62

## Abstract

Boden- und satellitengestützte Messungen des magnetischen Flusses der Sonnenoberfläche sind im großem Maße von der zur Verfügung stehenden Auflösung der Messinstrumente abhängig. Denn, Messungen des magnetischen Feldes anhand des Zeeman Effektes sind anfällig für Auslöschung von gegenpoligem magnetischen Fluss auf der Subauflösungsskala. Dieser Effekt führt zu einer signifikanten Unterschätzung des tatsächlichen realen magnetischen Fluss auf der Sonnenoberfläche. HMI line-of-sight Magnetogramme werden in dieser Arbeit genutzt, um diesen Effekt anhand von künstlicher Herabsetzung der Auflösung zu untersuchen und qualitativ und quantitativ zu bestimmen. Dabei wird ein neuer selbst entwickelter Algorithmus genutzt, um quiet-Sun und active regions voneinander zu trennen und unabhängig voneinander zu untersuchen. Es wird gezeigt, dass räumliche Auflösung sich marginal auf Messungen vom magnetischen Fluss von active regions auswirkt. Messungen vom magnetischen Fluss von quiet-Sun Regionen sind signifikant abhängig von der Auflösung. Ein empirisches Modell wurde gefunden welches den Effekt interpoliert. Die experimentelle Betrachtung wird abgeschlossen mit dem Ergebnis, dass etwa 10% vom magnetischen Fluss in quiet-Sun Regionen von Magnetogrammen mit MDI Auflösung nicht detektiert wird. Dem experimentellen Teil folgt eine theoretische Betrachtung zum Verhalten von statistischen Momenten unter variabler Auflösung.

## **Abstract**

Ground and satellite based measurements of the solar magnetic surface flux are highly dependent on the available spatial resolution of the measuring instrument. This has its cause in magnetic flux cancellation on sub-resolution scales of Zeeman based measurements. This effect leads to a significant underestimation of the measured magnetic flux on the solar surface. HMI line-of-sight magnetograms are used to determine this effect qualitatively and quantitatively, by artificially decreasing the resolution of the magnetograms. It will be revealed, that spatial resolution is of negligible importance in the measurement process of active region flux, whereas it has significant impact in the measurement of quiet-Sun flux. An empirical model has been found to interpolate the dependence of missing magnetic flux on spatial resolution. The analysis is then concluded by a theoretical approach to understand the effect of variable resolutions on statistical moments.

# 1 Introduction

Various scientists have made efforts in studying the influence of solar magnetic flux on solar irradiance variability. It is therefore of vital importance to question the results from instrumental observations. In her 1992 paper [1], Karen Harvey stated that the total magnetic flux in the quiet-Sun and the active regions shows temporal seasonality. She was able to show that at solar activity minimum the majority of magnetic flux resides in the quiet-Sun, whereas at the solar activity maximum most of the flux would be present in the active regions. These results have been concluded by analysing Kitt Peak (KP) synoptic charts (SC) from cycle 21 and 22, from the National Solar Observatory. In 1994 Harvey [2] continued her work using an improved method to separate active regions and quiet-Sun.

Solanki and Krivova [3] used the line-of-sight magnetograms covering Carrington Rotation (henceforth denoted as rotation) 1915 and 1975)from the Michelson-Doppler-Imager (MDI, spatial resolution of  $4''$ ), to show that Harvey's results were too conservative in the case of the quiet-Sun. They concluded that spatial resolution plays a key role in the measurements of total solar magnetic flux and that KP observations were missing a non-negligible amount of magnetic flux emerging in the quiet-Sun. The amount of weak flux missing in KP observations was quantified to be at least 1/3 compared to MDI observations. This effect can be attributed to magnetic flux cancellation (or to be more precise: cancellation of Stokes V signals) on sub-resolution scales and is now known to be the fundamental flaw of all Zeeman based observations of the solar magnetic field [4].

Both studies showed that the magnetic field in the quiet-Sun cannot be underestimated compared to other more visible and larger magnetic structures. Despite their importance, precise measurements of its magnitude are difficult, because the typical size of a single magnetic element is smaller than the resolution limit of modern day instruments. Graham, Danilovic and Schüssler found observational evidence for a lower bound of one order of magnitude smaller than 200 km for magnetic structuring in the photosphere [5]. They derived a numerical estimation for the mean unsigned flux density in the quiet-Sun of about 50G.

I will compare this estimation to results which are based on HMI line-of-sight synoptic chart observations for rotation 2097 and 2147 (year 2010 and 2014) used in this analysis. HMI is part of the SDO satellite, which was launched into space in 2010. It has a spatial resolution of 1 arcsecond and resolves structures at the scale of approximately 350km. We can therefore expect a priori that HMI will miss flux due to flux cancellation. In this work, I will process the raw data in form of full solar disc magnetograms, remove its noise, perform an active region/quiet sun

separation with a newly developed method and analyse the effect of spatial and temporal resolution on physical and statistical variables by degrading the resolution in a controlled manner. To compare the results directly to Solanki and Krivova's work, I will change the calibration of the data based on the empirical MDI-HMI transformation equation by Liu *et al.* [9].

This work is then concluded by a semi-empirical model which describes the effect of binning and temporal resolution on the first two statistical moments.

All computations in this project were done in IDL and its matrix notation is carried over into this thesis.<sup>1</sup>

---

<sup>1</sup>Important IDL procedures and functions used in this thesis will be released on *GITHUB* at the 30th of April 2017, <https://github.com/zebilly/B.Raza-B.Sc.-Thesis-IDL-Procedures>.

## 2 Data, Noise and Processing

### 2.1 HMI Line-of-Sight Magnetograms

Two sets of HMI full disc line-of-sight magnetograms with a cadence of 45s covering Carrington Rotation 2147 (11.02.2014-11.03.2014/29 days) and 2097 (20.5.2010-17.06.2010/28 days) are being used in the analysis<sup>1</sup>. These two distinct rotations represent the sun both at its activity minimum (2010) and activity maximum (2014). HMI magnetograms are of size 4096x4096 pixels with a pixel size of 0.5" and measure the magnetic field density within the projected pixel area in units of Mx/cm<sup>2</sup> (or equivalently G). It is therefore not required to account for the fact that the actual solar surface area represented by each pixel grows with increasing latitude, when converting magnetic flux density to magnetic flux. This conversion solely depends on the pixel area at disc center, which can be calculated by inserting an angle of 0.5" and a radius of 1AU into the formula for the arc length of a circle. This yields a result for the pixel area of  $1.3 \cdot 10^{15} \text{ cm}^2$ .

The noise distribution on the magnetograms is assumed to follow a  $\mathcal{N}(0, \sigma^2)$  law. It is therefore possible to reduce the overall noise level by averaging over several measurements in each day (integration times of 45s, 315s, 1215s, 3600s). This results in four magnetograms for each day in both sets and a total of 228 magnetograms. In general, noise can complicate and ultimately bias the analysis of total magnetic flux considerably. Noise inside magnetograms has the property to change the signs from weak magnetic signals and will therefore lead to a decrease in net signal. It is therefore likely, that the effect of spatial resolution on the total flux will be overestimated. Krivova and Solanki pointed this out in their paper and hence removed magnetic signals below the noise level from the MDI magnetograms used in their analysis. For this reason, all magnetic signal in the HMI magnetograms falling below the noise level should be removed before moving on with the analysis.

Every HMI pixel has its own corresponding noise level. Consequently  $\sigma$  will not be a scalar but a 2D array with the same size as the magnetograms. A graphical visualization of the noise array is given in figure 1. Dark tones reflect low noise levels, while brighter ones represent pixels with higher noise levels. The shade of the edge of the solar disc indicates the transition to the background. Every HMI pixel detecting signal from the solar disc has its own unique noise level corresponding to an entry in the noise array. It is important to note that the solar disc on the magnetogram arrays and on the noise array are not aligned. The position of the solar disc moves inside the magnetograms. Both discs need to be aligned first to

---

<sup>1</sup>All magnetograms have been provided as .fits files by the institute.

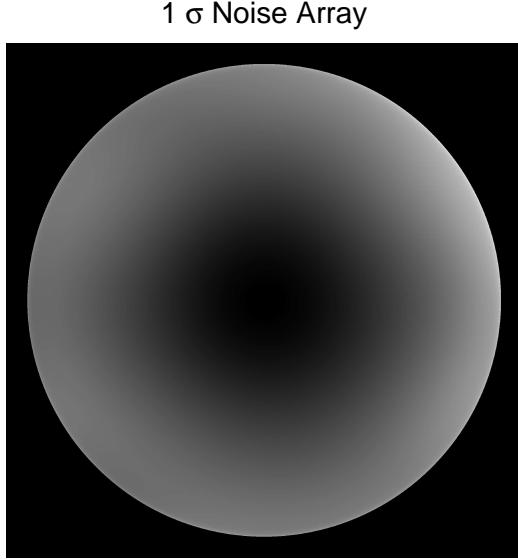


Figure 1: Greyscale illustration of the noise matrix. The solar disc is embedded in black background. Dark tones inside the disc represent pixels with low noise level. Pixels towards the limb are dominated by higher noise level. It is apparent that HMI's noise level increases with growing centre-limb distance.

assign each magnetogram pixel its corresponding noise value. I will use K.L. Yeo's RSAT.PRO IDL function to align both solar discs. The procedure extracts the exact position of the solar disc in the field-of-view from the data files (position of the solar disc in field of view is included in the files). It will then adjust the position accordingly.

According to the common known relation

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad (1)$$

we have to multiply the  $1\sigma$  noise matrix by a factor of  $\sqrt{7}$ ,  $1$ ,  $\sqrt{7/27}$ ,  $\sqrt{7/80}$  for the different integration times. The  $\sqrt{7}$  factor arises because the original noise array was originally defined for an integration time of 315s. The adjusted noise arrays are then used to remove magnetic signals falling below their corresponding noise threshold.

The average noise levels calculated by equation 1 are tabulated in table 1. It

Integration Time	45s	315s	1215s	3600s
Average Noise Level [G]	$17.3 \pm 2.5$	$6.5 \pm 0.9$	$3.3 \pm 0.5$	$1.9 \pm 0.3$

Table 1: Average noise level for different integration times.

	Solar Cycle 2097	Solar Cycle 2147
45s	83.9%	79.6%
315s	67.1%	62.4%
1215s	48.6%	44.8%
3600s	34.7%	31.8%

Table 2: Percentage of pixels affected by noise for different integration times and activity levels.

is apparent that magnetograms with a lower integration time will be more strongly affected by noise. Higher noise levels will lead to a decrease of net signal because more pixels will fall below the threshold. We can also deduce that mostly pixels from quiet regions on the Sun will be affected because of the overall low noise levels. This effect is tabulated in table 2. Around 80% of all 45s magnetogram pixel can be regarded as noise and are therefore set to zero. The numbers drop to around 33% for 3600s magnetograms. Furthermore, more pixels on rotation 2097 are thresholded than on rotation 2147. This effect can be related to the different activity levels on both rotations, which leads to an increase of the average signal detected at maximum activity. Figure 2 determines how the noise is distributed in the instrument.

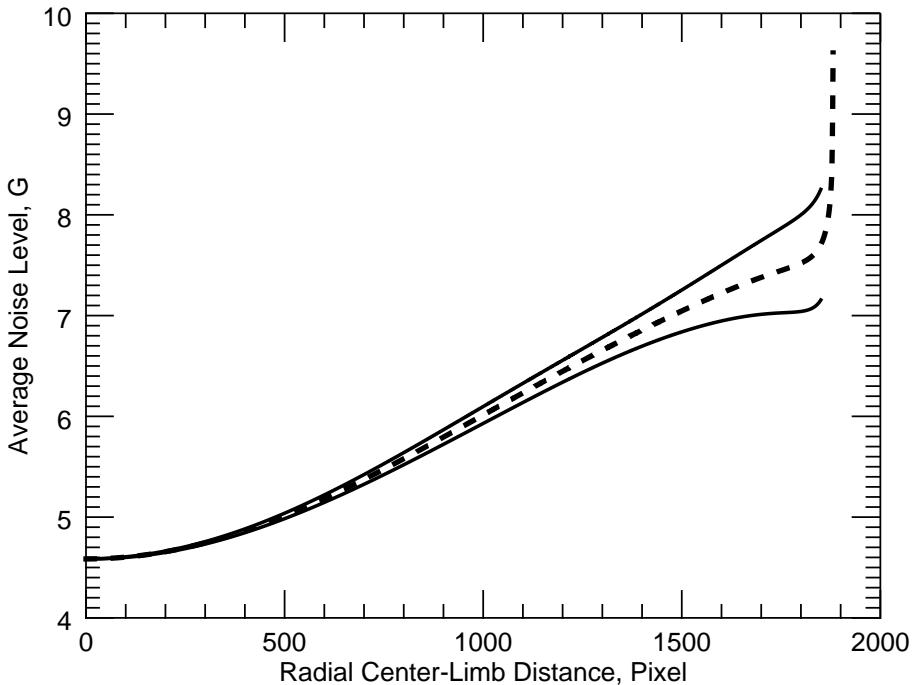


Figure 2: Dashed line: Average noise level as function of radial distance to the solar disc center. The solid lines represent the  $1\sigma$  confidence interval.

HMI can give a nearly exact estimation for the noise in the proximity of the solar disc center. The confidence interval widens for distances greater than 500 pixels. The average noise level itself increases with constant slope in the interval ( $500 < D < 1500$ ). It then increases asymptotically for the outermost regions of the solar disc. This information will be used at a later processing stage to cut off all pixels beyond a distance of  $\sim 1600$  in north and south direction (corresponds to a NS latitude of  $60^\circ$ ) <sup>2</sup>

---

<sup>2</sup>The noise array has been provided by the institute.

## 2.2 Synoptic Chart Construction

The next steps involves synoptic chart creation. These type of magnetograms are maps of the entire solar surface created by merging central meridian data from multiple observations. The number of magnetograms used in this process can vary from 27 to 29, depending on the length of the specific Carrington Rotation. Because of the effect described in chapter 2.2.1, only data between the North-South 60°latitudes have been included in this analysis. The pixel/latitude threshold can be calculated by simple trigonometric relations. The calculation requires the length of the Sun's radius expressed in pixel. The necessary parameter is included in the header of the original data files under the variable name R\_SUN. This parameter differs for both rotations. Whereas HMI measures a radius of  $1921 \pm 4$  pixels in rotation 2147, it measures a slightly smaller value of  $1873 \pm 3$  in rotation 2097. This discrepancy has to be considered in the polar region exclusion process. Two of these simplified charts are shown in figure 3 and 4. Black and white pixels in those figures reflect regions with an unsigned flux greater than 50G. Dark tones are defined for negative values and bright/white pixels reflect pixels with positive polarity. The orange or reddish regions portray observations of magnetic activity near 0.

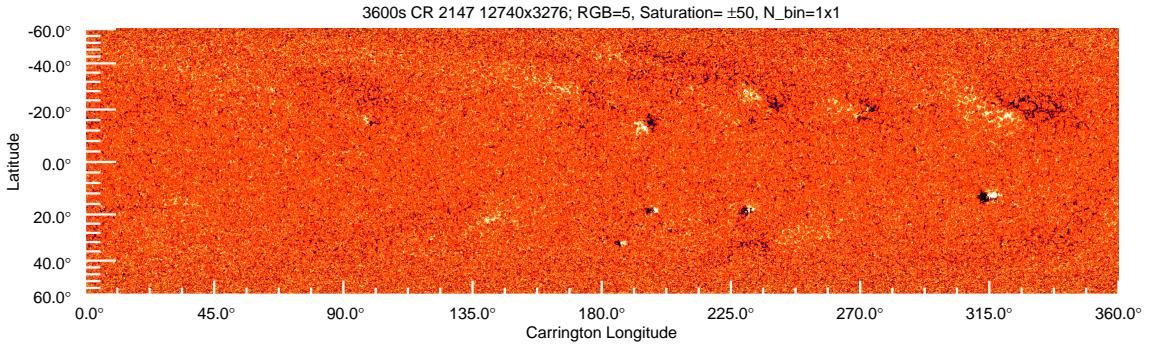


Figure 3: Cycle 2097 (2010) simplified synoptic chart with 3600s integration time.

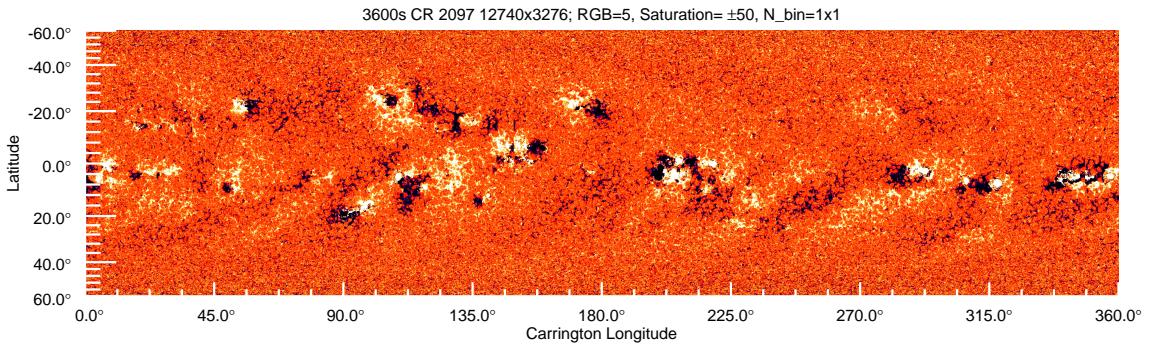


Figure 4: Cycle 2147 (2014) simplified synoptic chart with 3600s integration time.

### 2.2.1 Total Unsigned Flux at higher Latitudes

$\frac{\Phi_{S60} + \Phi_{N60}}{\Phi_{Central}}$	Cycle 2097	Cycle 2147
45s	0.71	0.51
315s	0.11	0.07
1215s	0.10	0.06
3600s	0.10	0.06

Table 3: Ratio of the total unsigned flux at higher latitudes to central regions for various integration times and rotations.

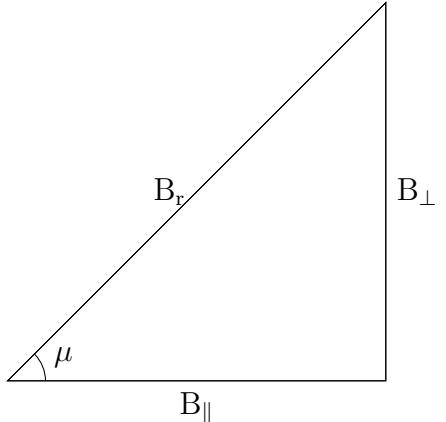
$\Phi [10^{22}\text{Mx}]$	45s	315s
North 60°-90°	10.0	1.3
Central	26.0	28.6
South 60°-90°	8.5	1.7

Table 4: Total unsigned flux in each region for 45s and 315s integration time in rotation 2097.

Table 3 demonstrates that there is a physical explanation of why a cut off at higher latitudes is reasonable. The data values in the table represent the ratio of the total unsigned flux in the cut off regions to total unsigned flux in the central region. The ratio is nearly constant for higher integration times and varies slightly for different activity levels of the sun. It is immediately apparent that the 45s ratios seem to be too high. The explanation for this ratio is given in table 4. HMIs 45s line-of-sight magnetograms detect up to an order of magnitude more flux in the polar regions than magnetograms with higher integration time. Both polar regions combined harbour nearly as much magnetic flux as the central region. This magnetic flux seems to disappear completely in the 315s synoptic charts. The magnetic flux in the central region can be interpreted to be invariant with increasing integration time. It is for this reason that a cut off at North-South 60° was chosen.

### 2.2.2 Foreshortening Effect

HMI line-of-sight magnetograms only measure the longitudinal/parallel component of the magnetic field. Dividing the absolute value of each pixel by its corresponding heliocentric angle,  $B_r = B_{\parallel}/\cos(\mu)$ , results in the radial component. This equation holds under the assumption that the magnetic field vector is radial everywhere. Cases where this assumption is violated, e.g. at sunspots [7], are neglected in this analysis.



## 2.3 Reducing the Spatial Resolution of Synoptic Charts

An effective way of image degradation is needed to study the effect of spatial resolution on properties of synoptic charts. IDL provides several build in functions, which can be used to control the size of arrays, e.g. CONGRID or REBIN. I chose the REBIN function because it works in exactly the way it is needed for this analysis. Consider the case where we want to resize the  $\mathbb{E}_{4 \times 4}$  unity matrix

$$\mathbb{E}_{4 \times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

into a 2x2 matrix. The REBIN function acts as follows on the input array

$$\mathbb{E}_{4 \times 4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\text{REBIN}} \begin{pmatrix} \sum \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \sum \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \sum \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \sum \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

It is important for the REBIN function to choose a target resolution equal to integral factors of the original dimensions. Otherwise it is not possible to choose well-defined summation boxes (binning box). This fact restricts the degradation parameters considerably and makes them dependent on the size of the original synoptic chart. However, it is possible to control this problem by cutting the array into a favourable size. In this analysis, a size of 12740x3276 pixels was chosen as array size. The degradation factors can then be identified as integral factors of 12740 and 3276.

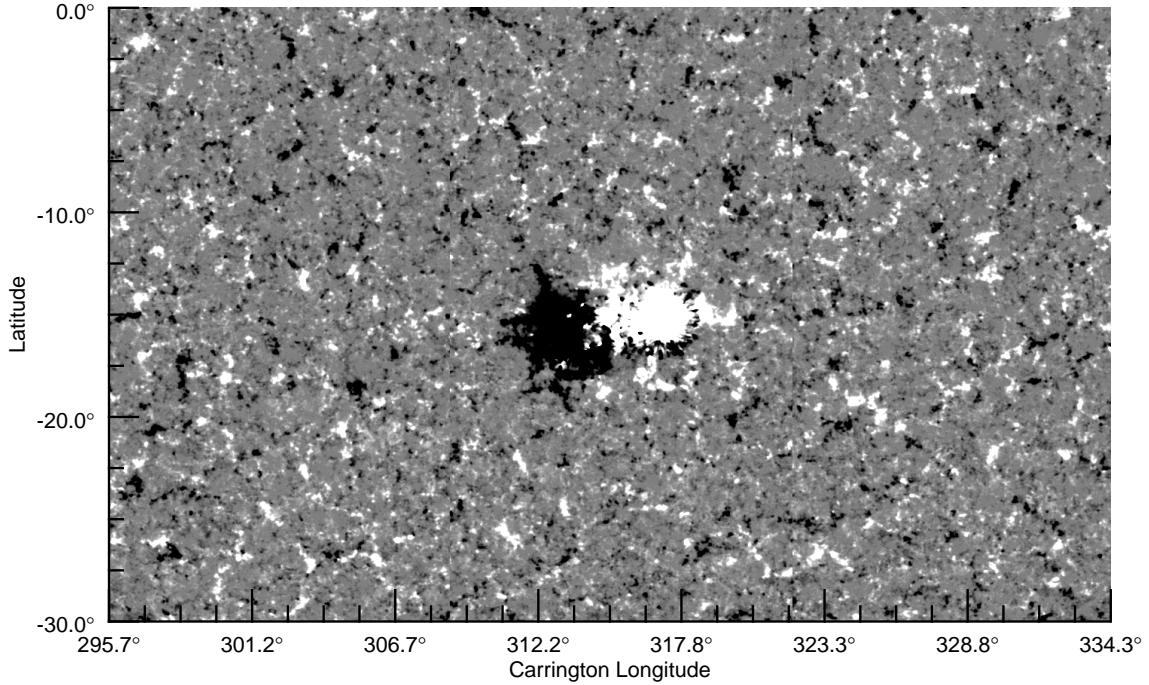
The following set of factors for the x-dimension

$$X = \{2, 4, 5, 7, 10, 13, 14, 20, 26, 28, 35, 49, 52, 65, 70, 91, 98, 130\} \text{ and}$$

$$Y = \{2, 3, 4, 7, 9, 12, 13, 14, 18, 21, 26, 39, 42, 52, 63, 78, 84, 91\}$$

for the y-dimension are used in this analysis. This results in  $|X| \cdot |Y| = 361$  different synoptic charts. The effect of this process is visible when comparing the KP sized image to the original HMI observation in figure 5. Nearly all visible information about small scale magnetic structures has been destroyed in the KP sized image by the degradation process.

CR 2097, 3600s, HMI Resolution



CR 2097, 3600s, KP Resolution

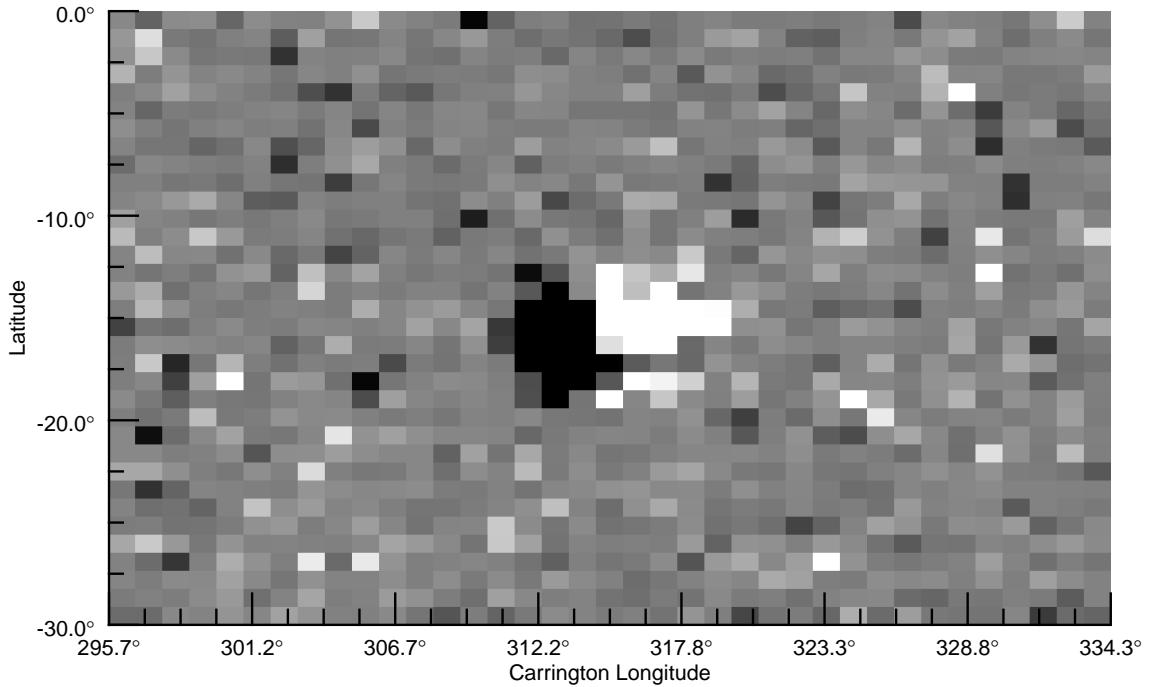


Figure 5: Comparison of high-resolution and degraded synoptic chart. An optimal image detail has been selected to illustrate the effect of spatial degradation on small-scale magnetic elements and active regions. Top: Original HMI resolution; bottom: Kitt Peak synoptic chart resolution. Binning Box Size:

## 2.4 Active Region/Quiet-Sun Separation Techniques

The solar magnetic field is divided into various structures, each with different properties. Active regions are defined by strong unipolar magnetic fields. They can be immediately identified as large scale structures in bright or dark tones on magnetograms depending on the polarity. Quiet-Sun structures are much smaller in size and intensity. It is for this reason that quiet-Sun components will be affected more heavily by binning processes like image degradation (see figure 5, bottom). Active region pixels are more robust against these kind of mathematical operations because they consist of large unipolar patches. It is therefore of interest to analyse the effect of spatial degradation on those both structures separately.

Various separation methods, of which the author is aware of, exist for this purpose, e.g. thresholding, in Karen Harvey's 1992 paper [1], her improved technique, the Root-Mean-Square-Difference method used in her 1994 paper [2], or the Erode-Dilate method of Haagenar, Schrijver and Title [8]. Krivova and Solanki used Harvey's RMS-Difference method in their analysis. The simplest way would be to use a threshold  $\Phi_{\text{thr}}$  and set all pixels with  $\Phi < \Phi_{\text{thr}}$  to 0 and everything non-zero to 1. Unfortunately the hereby created logical mask does not represent the true active regions we are interested in. Active regions are not solely defined by their pixel intensity but are part of much larger structure. These structures are much larger than indicated by isolines of high magnetic flux.

Consider the case where you want to find a mountain using a topographic map. Thresholding at a specific isohypse is clearly the wrong method because it is important to take the foot of the mountain into consideration. Mathematically, we are interested in the isoline where the active region structure switches polarity. Comparing figure 6 with a map which displays the polarity states in the test region (see fig7) confirms this statement. The two active regions at the center of the polarity map appear to be much larger in size. Thresholding can therefore only deliver an approximation of the boundary of an active region. This chapter will introduce a new self-developed algorithm for active region/quiet-Sun separation and compare it qualitatively to Harvey's RMS-Difference method.

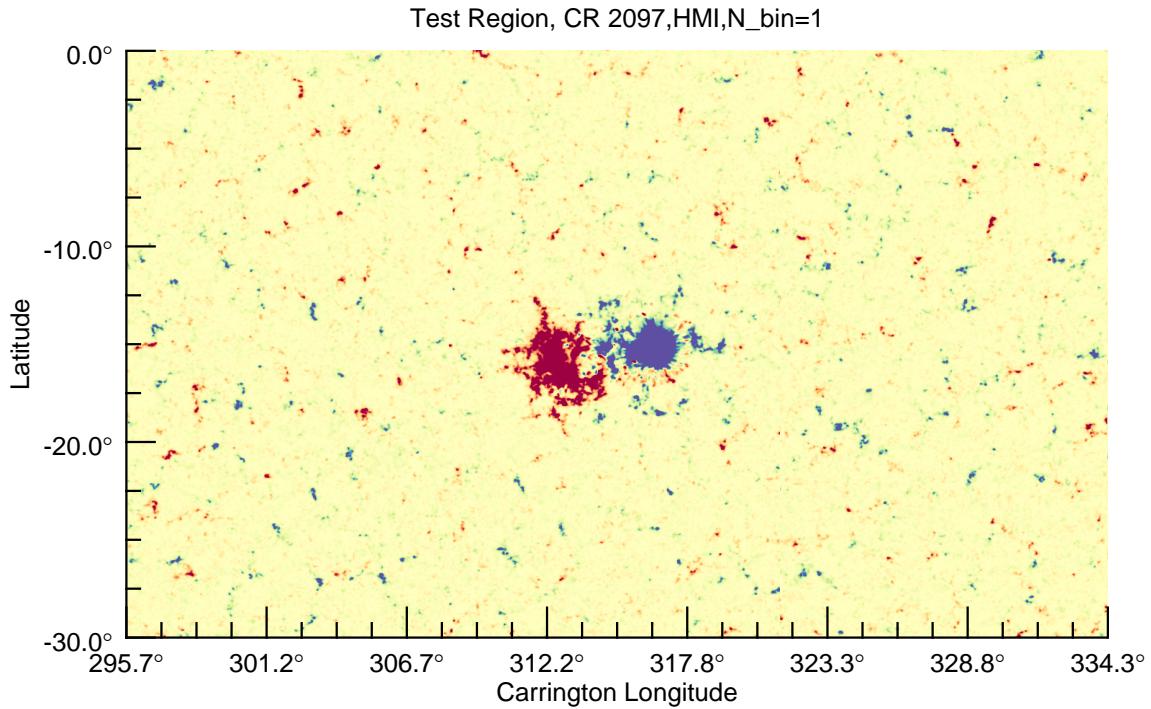


Figure 6: Test Region in the 3600s Cycle 2097 Synoptic Chart; RGB=74; red=negative, blue=positive, yellow=0; saturation at  $\pm 50$ G.

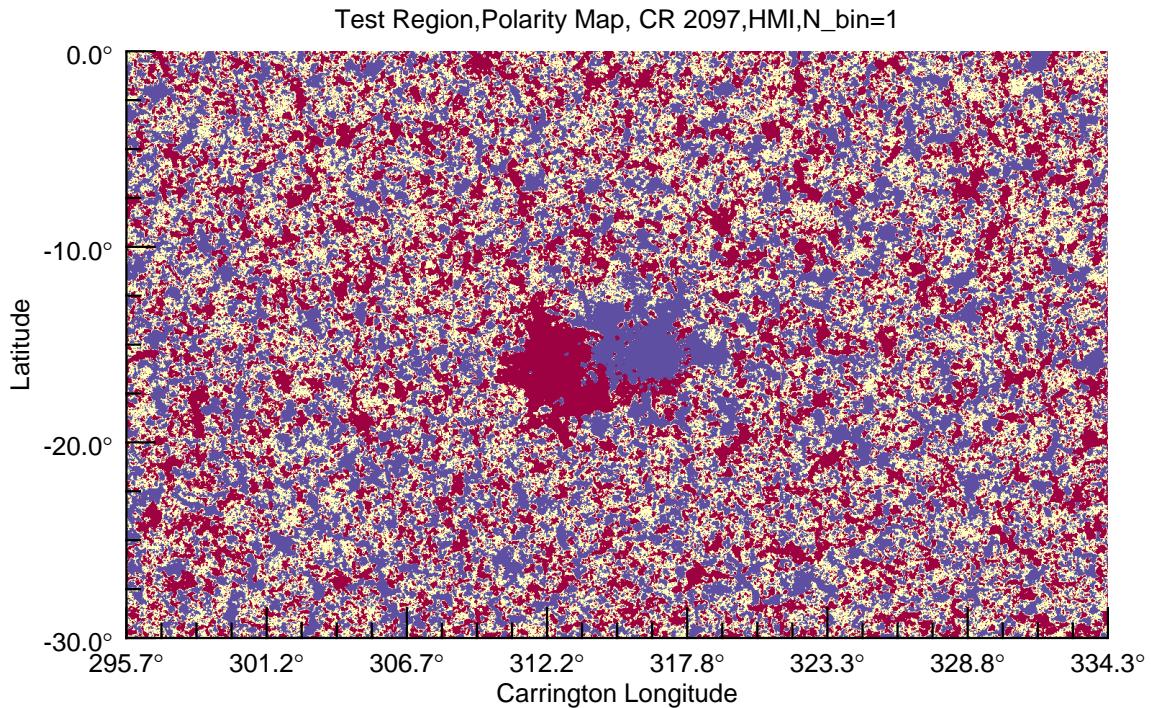


Figure 7: Polarity Map of the Test region; RGB=74, red = positive polarity, blue = negative polarity, yellow = no information / removed noise

### 2.4.1 Intersection Algorithm

The first attempt to create a logical mask for active region separation utilized Karen Harvey's Root-Mean-Square-Difference method. The amplitude variation inside a sub-array of size  $K \times K$  is calculated by taking the square root of the sum of the absolute quadratic difference of the center pixel with each pixel inside the sub-array

$$M_{i,j}(k) = \sqrt{\sum_{p=i-\lfloor \frac{k}{2} \rfloor}^{i+\lfloor \frac{k}{2} \rfloor} \sum_{q=j-\lfloor \frac{k}{2} \rfloor}^{j+\lfloor \frac{k}{2} \rfloor} (y_{i,j} - y_{p,q})^2}. \quad (2)$$

The RMSD map is then smoothed over another sub-array of size  $S \times S$  and thresholded to find the locations of active region pixels. This method was tested on the same region as in figure 5. The calculated logical mask is visualized in figure 8[Bottom].

Comparing the logical mask derived by Harvey's method with figure 7 shows that the RMSD method overestimates the size of the active region considerably. For this reason, a new method was developed that does not depend on kernels but merely on the choice of a specific threshold. For that I went back to the definition of active regions. As already defined, we are interested in the whole unipolar patch which surrounds pixels of high magnetic intensity. The new method will therefore identify all pixels belonging to these patches.

The algorithm needs to be applied separately on negative and positive polarities. It is therefore necessary to create two new arrays, each containing only flux of positive or negative polarity (target array plus, TAP, and target array minus, TAM). The following steps can then be applied on a synoptic chart (target array) to create a logical mask to identify active regions:

1. Create empty array (MASK) of same size as the target array.
2. Start with positive polarity: Find all regions  $s_i$  with  $\Phi \geq \Phi_{\text{thr}} \Rightarrow S = \{s_i\}$  in TAP.
3. Find all regions  $p_i$  with  $\Phi > 0 \Rightarrow P = \{p_i\}$  in TAP.
4. If  $s_i \cap p_i \neq \emptyset$ , set  $p_i$  to true, otherwise delete  $p_i$
5. Set  $\text{MASK}(p_i)$  to 1.
6. Rerun step 2-5 for TAM.

The resulting logical mask after applying the intersection algorithm on the test region is shown in figure 8[Top]. The numerically most stable results have been

obtained with a threshold of  $0.5 \cdot \max(|\Phi|)$ . The corresponding IDL code is presented in chapter 7.6 in the appendix.

### 2.4.2 Approximations in the Separation Process

Both methods introduced in the previous chapter are not running stable for large binning arrays. Both methods overestimate the size of active regions extremely. They can therefore not be used for synoptic charts of low resolution. Harvey's method fails to produce acceptable results because the kernel size needs to be adapted to the resolution of the synoptic chart. At a certain resolution limit the kernel size would theoretically shrink below the resolution limit. We can bypass this problem by reducing the size of  $\lfloor k/2 \rfloor$  and cap its minimal value to 1. The RMSD kernel can then be identified as a  $3 \times 3$  array in the case of extremely low resolution. However, it cannot shrink below this size. This results in an overestimation of identified active region pixels.

The intersection method relies on accurate data on the small-scale magnetic elements because the algorithm identifies contiguous areas of same polarity. As stated in subsection 2.3, the binning process in image degradation destroys information about the distribution of small scale magnetic elements. This information loss leads to the creation of large networks of unipolar flux stretching across the whole synoptic chart, making any attempt to calculate a logical mask impractical. However, it is possible to bypass this problem in a simple way. The position of active regions does not change in the binning process. We can therefore assign a rectangular box to each active region such that the box covers all of it. We can then transform the coordinates of the box to the new resolution. The new coordinates then define a shrunken box on the degraded synoptic chart. The area inside this box can then be identified as an active region. This box method will lead to an overestimation of active region area. The thus created systematic biases on calculated parameters in the analysis were found to be acceptable.

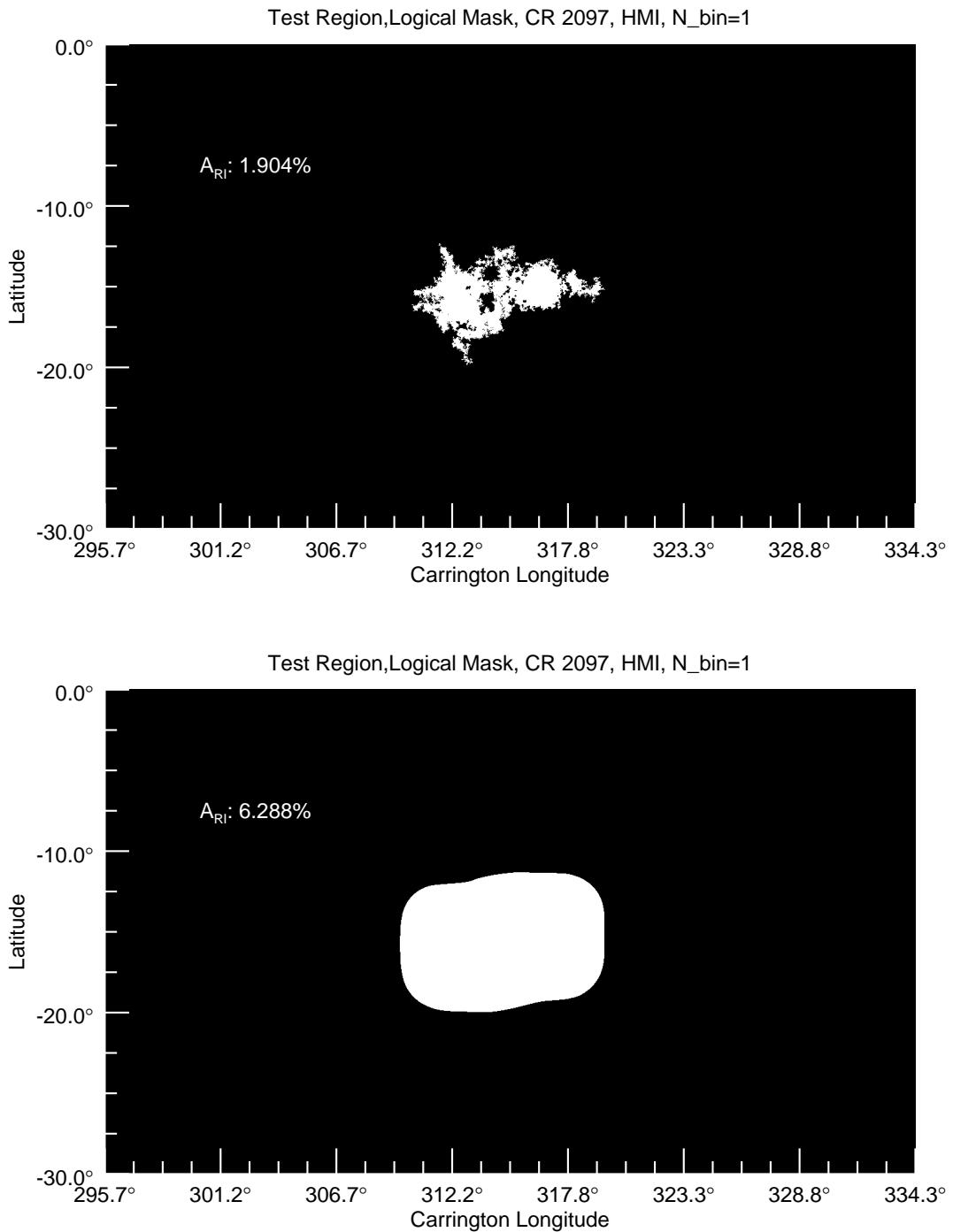


Figure 8: Comparison of the Intersection Algorithm (top) and Harvey's RMSD method (bottom). Top: The Intersection Algorithm was used to calculate the logical mask for the test region. White and black pixel mark ones and zeroes respectively. A total of 1.9% of active region pixels have been identified with this method. bot: Karen Harvey's Root-Mean-Square-Difference method applied on the test region results in a total of 6.3 % active region pixels.

### 3 Theoretical Explanation of Magnetic Flux Cancellation

I will first establish the theoretical framework of magnetic flux cancellation from the viewpoint of the observer/instrument. Flux cancellation as a result of pixel binning is of interest for this project. I will therefore neglect all physical processes that can potentially lead to signal loss inside the instrument, e.g. interferences inside the chip while varying the spatial resolution.

HMI resolves structures on the solar surface with a resolution of 1" or equivalently a pixel size of 0.5". In 1AU distance this corresponds to a resolution limit of approximately 350km. Experimental evidence of Hinode Observations with a resolution of 0.2" proved that magnetic structuring occurs on scales smaller than 200 km as presented in Graham *et al.*'s paper [5]. We can therefore expect that HMI resolution is not sufficient to observe small scale magnetic activity and will therefore suffer from magnetic flux cancellation on sub-resolution limit. This will then lead to a loss in net signal and bias calculations of the total solar magnetic flux and properties of the empirical pixel PDF.

The underlying mathematical structure of this problem can be found in the well known triangle inequality for n-dimensions

$$\left| \sum_{i=1}^n x_i \right| \leq \sum_{i=1}^n |x_i| \quad \forall x_i \in \mathbb{C} \quad (3)$$

How this inequality is connected to our missing flux? The HMI data is present in the form of NxM image arrays with real entries. Reducing the spatial resolution of magnetograms results in the reduction of the matrix size. The amount of dimension reduction depends on the pixel size of the degraded magnetogram. A target pixel size of 1 arcsecond x 1 arcsecond is realized by binning 2x2 sub arrays inside the original array together. These sub arrays are the new entries of the degraded image. The general mathematical operation equivalent to reducing its spatial resolution can therefore be identified as the summation of sub arrays. I will define the function  $\Upsilon$

$$\Upsilon(x, y) : \mathbb{R}^{N \times M} \rightarrow \mathbb{R}^{N/x \times M/y} \quad (4)$$

as the mathematical operation which degrades the dimensions of the original array by summation of sub arrays. The parameters x and y describe the pixel size of the degraded image in both dimensions. They are defined as

x = target pixel size in x dimension/0.5" and analogous for y. A target pixel size of

1" x 1" with x and y equal to 2, would result in a  $N/2 \times M/2$  array. This process can be illustrated as follows. The following matrix shall depict HMI magnetogram arrays of size  $N \times M$ :

$$\text{HMI} = \begin{bmatrix} y_{0,0} & y_{1,0} & \dots & y_{N-1,0} \\ y_{0,1} & y_{1,1} & \dots & \\ \vdots & & \ddots & \vdots \\ y_{0,M-1} & \dots & & y_{N-1,M-1} \end{bmatrix}$$

If N and M denote the x and y dimension of the array and x,y the corresponding degradation factors in x- and y-dimension then the general formula for the degraded matrix element  $y'_{i,j}$  is given by

$$\mathbb{R}^{x \times y} \rightarrow \mathbb{R}, \quad y'_{i,j} = \sum_{k=\frac{N}{x}i+1}^{\frac{N}{x}(i+1)} \sum_{l=\frac{M}{y}j+1}^{\frac{M}{y}(j+1)} y_{k,l} \quad (5)$$

Consider the case where  $x=2$  and  $y=2$  which corresponds to a spatial resolution of 2 arcseconds (equivalently a pixel size of 1" x 1"). The degradation process can be schematically described by the function

$$\begin{bmatrix} y_{0,0} & y_{1,0} & \dots & y_{N-1,0} \\ y_{0,1} & y_{1,1} & \dots & \\ \vdots & \ddots & & \vdots \\ y_{0,M-1} & \dots & & y_{N-1,M-1} \end{bmatrix} \xrightarrow{\Upsilon(2,2)} \begin{bmatrix} y'_{0,0} = \sum \begin{bmatrix} y_{0,0} & y_{1,0} \\ y_{0,1} & y_{1,1} \end{bmatrix} & y'_{1,0} & \dots & y'_{N/2-1,0} \\ y'_{0,1} & y'_{1,1} & \dots & \\ \vdots & \ddots & & \vdots \\ y'_{0,M/2-1} & \dots & & y'_{N/2-1,M/2-1} \end{bmatrix}$$

The first element of the new magnetogram was calculated by summing over all four pixels inside the first sub array. All following array elements have been calculated accordingly. Henceforth we will define  $N_{box}$  to be the number of elements inside these sub arrays (binning box). Note that  $\Upsilon$  is only well-defined if the dimensions of the degraded chart are integral factors of the original dimensions. This will be discussed more thoroughly in chapter 3. We can immediately tell by looking at the matrix equation above that the missing flux originates in the summation process inside the sub arrays. The total sum over all pixels can be identified as the net signal or as the total signed flux. Its value will be the same for the original magnetogram and the degraded one. However, the summation over the absolute values of all pixels behaves differently. In most cases the total unsigned flux of the original magnetogram will be bigger than for the degraded ones. This follows directly from the triangle inequality.

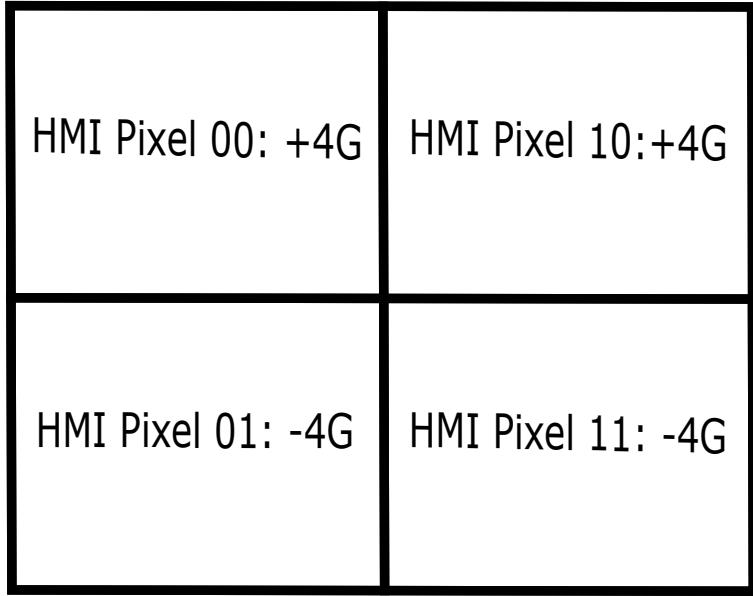


Figure 9: Four HMI Pixel can be binned together into a single pixel with double pixel size. Consider the case in which all four have the same absolute intensity but cancel each other to 0 if binned together because of opposite polarities. The new pixel would detect zero flux.

Magnetic flux cancellation will therefore always take place as long as the resolution scale is not on par with the scale of magnetic structuring. Note that the effect will be significantly more distinct if pixels with opposite polarities tend to be grouped together and weaker if the magnetic field tends to be uniform in polarity. This is the case for active regions on the solar surface. These regions can be identified with strong unipolar magnetic fields. The effect of flux cancellation in active regions can therefore be predicted a priori to be not as distinct as in the quiet-Sun.



## 4 Analysis

A total of 228 magnetograms have been used in the data processing to create 8 different synoptic charts, each one covering the solar surface at two distinct rotations with 4 different integration times. These two temporal dimensions lead to a high level of abstraction for the resulting data of interest. Therefore only selected results will be presented to qualitatively and quantitatively answer how spatial resolution affects statistical and physical properties of magnetic field observations. The first subsection will analyse the effect of integration time on synoptic charts and its connection to the distributions of noise signals. Studying the histograms of the noise distributions will reveal if the noise removal process was successful.

The second subsection of this analysis will use the enhanced resolution of HMI to reveal how much more flux is observed compared to artificial MDI and KP data (henceforth denoted as MDI-A (Artificial) and KP-A), created in the degradation process. Numerically determined values for the mean unsigned flux density in the quiet-Sun from HMI, MDI-A and KP-A synoptic charts are then compared to the lower-bound estimation of  $\langle \Phi_{QS} \rangle \approx 50$  G from Graham *et al.*'s numerical experiment.

The remainder of this section is about finding an empirical model which has the potential to explain how spatial degradation affects statistical and physical properties of synoptic charts.

## 4.1 Integration Time and Noise

It is important to study the effect of integration time on observations before analysing the results. Krivova and Solanki suspected that a fraction of real signal is hidden inside the removed noise. This effect would then lead to a deformation of the distribution function of the noise signal. This is immediately visible by comparing the distribution of the removed signals in figure 10. The 315s noise distribution function has a noticeable convex bulge on each side of its tails. This bulge increases in size for longer integration time. Under the assumption that noise is distributed normally, we can deduce that real signal was mistakenly identified as noise in the processing. The most probable cause for this effect can be determined by analysing the residual-sum-of-squares (RSS) for each Gaussian fit performed on the noise distributions. RSS values for fits on the regular distribution functions are extremely large. I have therefore normalized the histograms such that their area equals 1 (probability density). The calculated RSS values for the densities are listed in table 5. It is no coincidence that the RSS value for the 315s noise probability density is extremely small compared to the other values. It is important to recall how the 45s, 1215s and 3600s noise array were modified (see chapter 2.1). HMI measures line-of-sight magnetograms with a cadence of 45s. Magnetograms with higher integration time were created by consecutively averaging over several magnetograms, e.g. 7 for 315s, 27 for 1215s and 80 for 3600s. The noise array was then modified with an additional factor of  $1/\sqrt{n}$  to match magnetograms with different integration time. This factor was derived by the following law

$$\text{Var} \left( \frac{1}{n} \sum_{i=0}^n \mathbf{X}_i \right) = \frac{1}{n^2} \cdot \sum_{i=0}^n \text{Var} (\mathbf{X}_i) \quad (6)$$

with  $\mathbf{X}_i$  being random variables, e.g. pixel elements of magnetogram arrays. This law however is only valid in the case of uncorrelated random variables. The most probable cause for the non-normal distribution form of the noise signals lies in this assumption that the magnetogram elements are uncorrelated. To account for this

RSS	45s	315s	1215s	3600s
Cycle 2097	0.14	0.02	0.17	0.64
Cycle 2147	0.11	0.03	0.19	0.63

Table 5: Residual Sum of Squares (RSS) calculated for each gaussian fit.

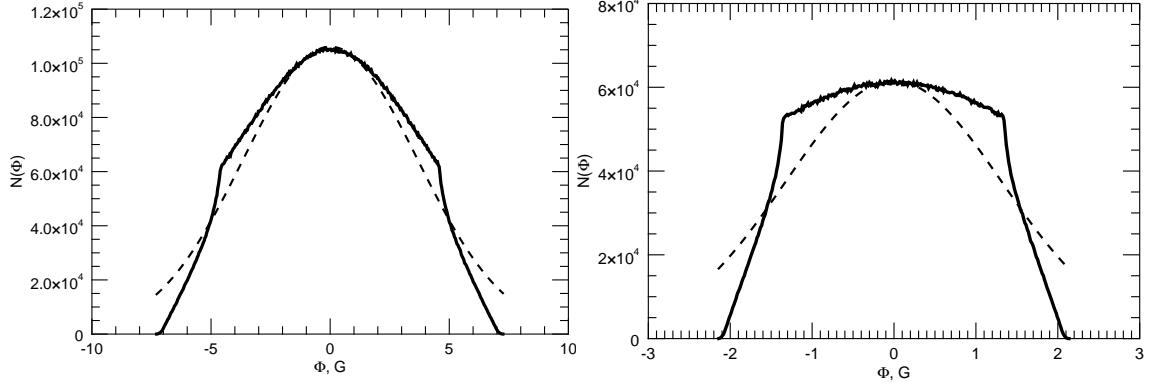


Figure 10: Noise distribution functions for rotation 2147 synoptic charts. Left: 315s integration time. Right: 3600s integration time

effect, it is necessary to use the more general covariance theorem

$$\begin{aligned} \text{Var} \left( \frac{1}{n} \sum_{i=0}^n \mathbf{X}_i \right) &= \frac{1}{n^2} \left( \sum_{i,j} \text{Cov} [\mathbf{X}_i, \mathbf{X}_j] \right) \\ &= \frac{1}{n^2} \cdot \sum_{i=0}^n \text{Var} (\mathbf{X}_i) + \frac{1}{n^2} \cdot \sum_{i \neq j} \text{Cov} [\mathbf{X}_i, \mathbf{X}_j] \end{aligned}$$

to calculate the correct noise array. The temporal covariance is determined by physical processes which governs the time evolution of magnetic flux in each pixel. The most convincing evidence for this hypotheses can be seen by comparing the listed 3600s RSS values. These values are by a factor of 4 to 5 larger than the rest of the noise fits, with the exception of rotation 2147 with an integration time of 1215s. The most probable cause for this effect, is the fact that the integration time of 1h is large compared to the typical lifetime of a solar surface granule.

Krivova and Solanki stated that magnetic flux redistributes itself in the intergranular lanes when new granules are formed, which leads to a measurable smearing effect in the magnetograms, which manifests itself in an underestimation of the effect of magnetic flux cancellation. This physical observation can explain the high RSS values for 3600s integration time magnetograms.

The temporal covariances between the consecutive magnetogram elements  $\mathbf{X}_{i,j}^k$  at time  $k$  and  $\mathbf{X}_{i,j}^{k+1}$  needs to be measured to judge if the hypothesis, that consecutive magnetogram elements are correlated, is true. This problem however lies outside the scope of this project because the temporal averaged magnetograms have not been processed by the author of this analysis. It is therefore not possible to precisely pin point the cause for the stated problem. Nevertheless it is immediately obvious that the average noise level is too low. A higher noise level would lead to less small scale

flux being removed which would shrink the size of the bulge on the distribution functions. Note that even the 315s integration time histogram in figure 10 (left plot) has a small but significant bulge. A possible cause for this deviation from a Gaussian could be that: i) the noise removal process is sensitive to small changes in the solar disc matching procedure. This would imply that the procedure used needs to be reworked and its parameters checked. ii) the noise matrix was not set up correctly or precise enough. Further investigation is needed to determine the underlying problem. Note that the excessive signal removal due to the presented effect leads to an underestimation of magnetic flux cancellation of unknown scale. The 315s measurements are therefore the most credible sources of data.

## 4.2 Comparing Measurements of Total Unsigned Flux from HMI,MDI-A and KP-A Observations

### 4.2.1 HMI Synoptic Chart Magnetograms

Important result obtained for both synoptic chart sets in the processing are listed in table 11(appendix; top: Cycle 2097 data; bottom: Cycle 2147 data) and are given in units of Gauss [G] (magnetic flux density). The total magnetic flux density during each rotation increases slightly for longer integration times. This is especially true for the total magnetic flux density in active regions, but behaves the other way around for quiet-Sun structures.  $\Phi_{QS}$  increases up to an integration time of 1215s but drops significantly at 3600s. This is explainable by recalling that the amount of pixels which can be regarded as noise signal shrinks for longer integration times (see table 2). Ergo, more total flux density will be detected with increasing integration time.

The drop in quiet-Sun flux density at 3600s integration time can be explained by the sudden increase of active region area. The separation technique used in this analysis is identifying active regions by affiliating pixels of high intensity to their corresponding patch of polarity. This unipolar patch consists of weak elements at its boundary because the magnetic field flips its sign at these areas. Consequently, the size of active regions will increase significantly for longer integration times and will include a non-negligible amount of pixels with weak intensity. This additional active region flux density is then missing in the quiet-Sun regions.

The proportion of active region flux density to quiet-Sun flux density however is negligible in the case of rotation 2097 where the Sun was at its activity minimum. Up to 96% of the total magnetic flux density at rotation 2097 emerges in the quiet-Sun. This situation reverses at activity maxima (see table 11; bottom) where the sun

contains  $2.1 \pm 0.1$  times as much flux density as at activity minimum. Nearly 60% of this measured flux at rotation 2147 has its origin in the quiet-Sun. A large fraction of the total solar magnetic flux density is measured in the active region covering up to 8-13% of the solar surface. The increase in total magnetic flux density cannot be attributed alone to the increase in active region area. Comparing the total flux density normalized to 1% at activity maximum and minimum reveals that active regions at activity maximum are nearly twice as strong as their counterpart. The quiet-Sun flux density normalized to 1% increases by a factor of 30% in the same interval. The amount of active region flux in total increases in average by a factor of 12.5.

The last row on both tables in 11 show the mean unsigned magnetic flux density in the quiet-Sun. The numerical experiments of Graham *et al.* predicted a mean unsigned magnetic flux density of approximately 50G. The arithmetically averaged value of  $\langle \Phi_{QS} \rangle$  is equal to  $20.8 \pm 12.1$  G. Ergo, HMI magnetograms miss on average at least  $58.4 \pm 24.4$  percent of quiet-Sun signal, though this result is highly dependant on the integration time and rotation. This result is consistent within the uncertainty with the estimation of Graham *et al.* that 80% of unsigned flux density remains undetected in the quiet-Sun at a resolution of 200 km.

However, it is important to note that the large measurements of  $\langle \Phi_{QS} \rangle$  result most probably from the choice to separate the sun into two structures, active regions and quiet-Sun. Whereas the definition of active region is definite, the quiet-Sun has been defined as the relative complement of the active regions pixels to the whole synoptic charts. This ambiguous definition of quiet-Sun regions is therefore the most probable cause for the large values of  $\langle \Phi_{QS} \rangle$  as observed in table 11. This effect will transfer itself to all subsequent measurements of  $\langle \Phi_{QS} \rangle$  at lower resolutions.

#### 4.2.2 Creating artificial MDI and KP Synoptic Charts

HMI and MDI synoptic charts do not differ only in their resolution but also in their different noise level. First of all, MDI magnetograms are not linked with a specific 2D noise array as HMI is. The noise level of MDI as it has been used in the analysis of Krivova and Solanki is a scalar and its average value for 1 minute magnetograms is typically about 20 G [3](and references therein). This average noise level is larger compared to the average noise level of HMI as indicated by figure 2. This excess noise has to be eliminated to effectively produce artificial MDI data. However the average  $1\sigma$  noise level of MDI (20G) needs to be converted because both instruments

have different calibrations. Liu *et al.* [9] and independently Chertok *et al.* [10]<sup>1</sup> have determined the coefficients of a linear model which describes the transformation of HMI to MDI data,

$$\text{MDI} = 1.40 \cdot \text{HMI}. \quad (7)$$

Inserting  $\text{MDI}=20$  G for MDI and solving for HMI results in a noise threshold of 14.3 G which is then applied on the HMI synoptic charts. The resulting charts are then degraded by a factor of 4x4 to match the resolution of MDI.

The average noise level of KP magnetogram is typically about 3G [2]. Artificial KP data can therefore be created by directly degrading HMI synoptic charts. However, to make this analysis consistent with the results from Krivova and Solanki's work, it was decided to use MDI-A data in the degradation process. MDI-A synoptic charts degraded by factor of 9x6 match the resolution of KP-A synoptic charts used in Krivova and Solanki's work.

#### 4.2.3 MDI-A and KP-A Data

Important results about the distribution of total unsigned flux density in different regions, as well as information about the active region and the mean unsigned flux density in the quiet-Sun, for both MDI-A and KP-A synoptic charts are listed in table 12 and 13 in the appendix. As a consequence of the separation technique used, the active region area covering the solar surface is increasing slightly with decreasing resolution but can be regarded as approximately constant. This has its origin in the transformation of the coordinates of the corners of the boxes covering the active regions at original resolution to MDI/KP resolution.

A direct comparison of the physical variables (total unsigned flux density and mean unsigned flux density) between HMI and MDI-A,KP-A data reveals the significant impact of instrument resolution on the measurements. MDI-A and KP-A synoptic charts are suffering from a significant amount of signal loss. The effect of signal loss is listed in greater detail in table 6. Table 6 gives an overview on how much total unsigned flux MDI-A measurements are missing compared to the enhanced resolution of HMI observations. Two temporal dimensions, integration time and activity rotation, will reveal how the signal loss is dependant on any of these two variables.

The second row in both rotations denote the missing active region flux in MDI-A magnetograms. It is immediately apparent that the reduced resolution of MDI does not lead to a significant loss of detected magnetic flux in these regions and is invari-

---

<sup>1</sup>Still in pre-print.

HMI/MDI-A	45s	315s	1215s	3600s
<b>CR 2097</b>				
Total	1.29	1.14	1.10	1.08
AR	1.01	1.01	1.01	1.01
QS	1.31	1.15	1.11	1.08
<b>CR 2147</b>				
Total	1.13	1.06	1.05	1.03
AR	1.02	1.01	1.01	1.01
QS	1.19	1.10	1.08	1.06

Table 6: Ratio of HMI magnetic flux  $\Phi_{\text{HMI}}$  to MDI-A magnetic flux  $\Phi_{\text{MDI-A}}$  for each region separately (Total,AR,QS).

ant under increasing integration time or level of activity. This result is consistent with the predictions from chapter 3. Degraded resolution is far more important for quiet-Sun fields as the third rows for both rotation indicate. This situation is furthermore illustrated in figure 11.

The 315s integration time HMI and MDI-A synoptic charts covering rotation 2047 have been separated into 28 vertical stripes (the components from which the synoptic charts have been constructed). The total unsigned magnetic flux from active regions (circles), quiet-Sun (diamonds) and both combined (asterisks) have then been calculated for each stripe. The results of the HMI stripes have then been divided by their corresponding results from the MDI-A stripes and were then plotted against the total unsigned flux from their corresponding region in the MDI-A stripe scaled by the amount of area covered <sup>1</sup>. The method behind this figure was first used in the paper of Krivova and Solanki. The ratios for the active regions are lying, as expected, near one, while ratios for the quiet-Sun and the total stripes are tightly packed together at ratios between 1.25 and 1.10. These regions are also containing an order of magnitude less magnetic flux than their active region counterparts. It is clearly visible that signal loss is strongly dependent on the level of activity. Furthermore, both quiet-Sun ratios from rotation 2097 and 2147 are affected greatly by integration time. In Chapter 2.2.1, a significant deviation between 45s and higher

---

<sup>1</sup>Consider the case in which an arbitrary vertical stripe is being made up of 90% quiet-Sun and 10% active region pixels. The amount of total unsigned flux from active regions is therefore scaled by a factor of 1/0.1, while the total unsigned flux from the quiet-Sun is being scaled by a factor of 1/0.9.

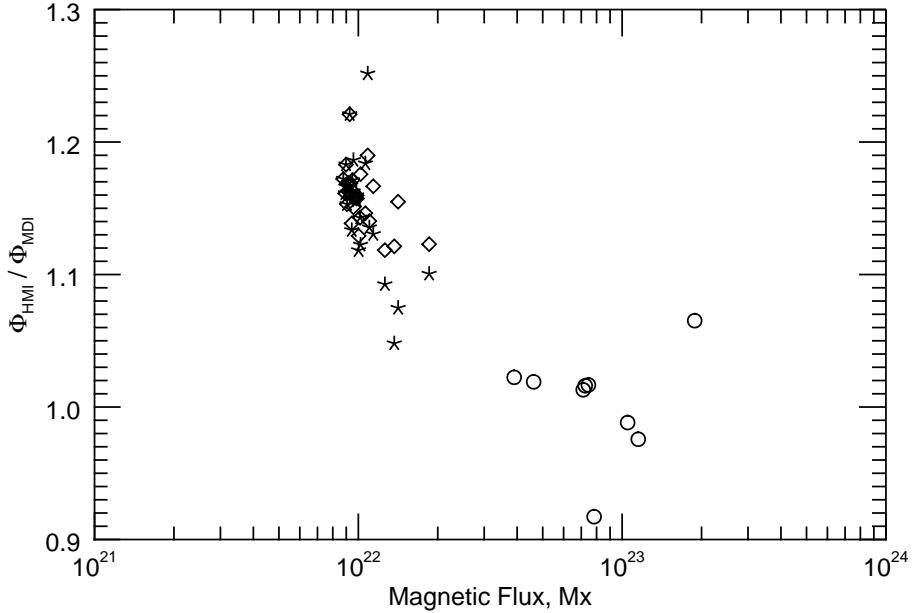


Figure 11: Ratio of HMI to MDI-A flux in their corresponding longitudinal bands for the total region, active regions and quiet-Sun.

integration times has been observed. Apparently 45s synoptic charts measure much stronger signals than their counterparts with larger integration time. This is the only possible explanation for the unintuitive and relatively large fraction of missing flux in the 45s column of table [6]. Furthermore, the results for the 1215s and 3600s column are probably too conservative because a significant amount of flux was most probably wrongly removed as discussed in chapter 4.1. Under this circumstance, that we can neglect the 45s measurements, we can assume that integration time does only play a minor role on how much signal is lost in the degradation process.

The signal loss in quiet-Sun flux leads to a significant over estimation of active region flux as a comparison of  $\Phi_{\text{AR}}/\Phi_{\text{QS}}$  values in table [11] and [10] reveals. The ratio of quiet-Sun to active region flux decreases on average by 10% compared to HMI resolution. This decreases by a factor of 2 at KP resolution (see chapter 7.3 appendix). Synoptic charts with KP resolution are missing approximately half of the existing magnetic flux in the quiet-Sun as it is observed by HMI. This effect however is not as distinctly strong as at activity maximum. The effect of signal loss for HMI-KP-A data is tabulated in table [14] in the appendix. KP-A synoptic charts measure about  $3.0 \pm 0.2$  times as much magnetic flux at activity maximum than at activity minimum, which is less than Harvey estimated for cycle 21 (factor of 4) and cycle 22 (factor of 5). This ratio decreases to  $2.3 \pm 0.2$  at MDI resolution and is, within the uncertainties, consistent with the results from Krivova and Solanki's MDI magnetogram analysis. Averaging over all integration times on both rotations

results in a signal loss in the quiet-Sun of KP-A synoptic charts of approximately  $(38 \pm 3)\%$  in rotation 2147 and up to  $(51 \pm 4)\%$  in rotation 2097.

MDI-A synoptic chart on the other hand have a significantly lower signal loss of  $(14 \pm 8)\%$  at activity minimum and  $(10 \pm 5)\%$  at activity maximum. This result is consistent with the estimation of Krivova and Solanki that MDI magnetograms are missing approximately 10% of the total magnetic flux. The measured average unsigned flux in the quiet-Sun at MDI and KP resolution suggests that  $(78 \pm 5)\%$  and  $(87 \pm 8)\%$  of the real signal is not being measured.

### 4.3 Effects of Spatial Resolution

The previous subsection revealed how degraded resolution affects measurements of total unsigned flux in different regions by comparing results from charts with original resolution to ones with MDI and KP resolution. However, it is not possible to find a general underlying model on how resolution degradation affects results with 2 data points. Therefore, HMI synoptic charts have been degraded by all combinations of  $X = \{1, 2, 4, 5, 7, 10, 13, 14, 20, 26, 28, 35, 49, 52, 65, 70, 91, 98, 130\}$  and  $Y = \{1, 2, 3, 4, 7, 9, 12, 13, 14, 18, 21, 26, 39, 42, 52, 63, 78, 84, 91\}$

in x and y direction, leading to a total of 361 different synoptic charts. Some

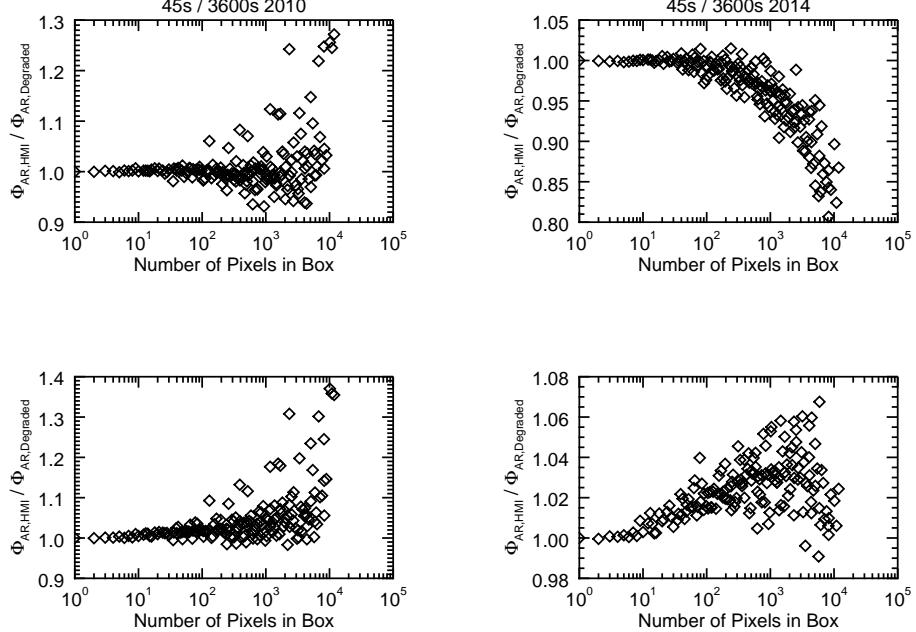


Figure 12: Figure shows the ratio of total active region flux at HMI resolution to active region flux at degraded resolution. From left to right: Cycle 2097 (year 2010) - Cycle 2147 (year 2014). From top to bottom: 45s integration time - 3600s integration time.

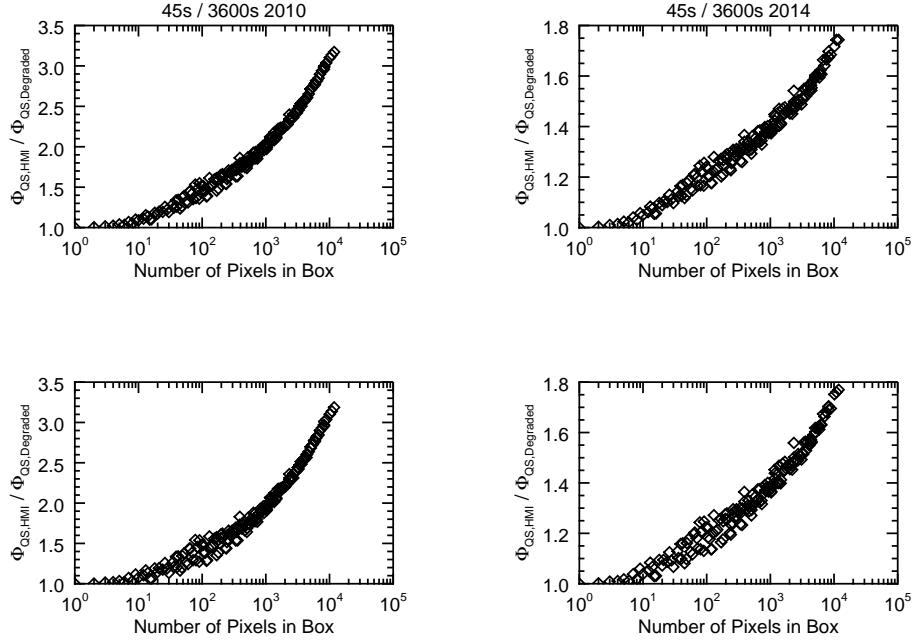


Figure 13: Ratio of total quiet-Sun flux at HMI resolution to active region flux at degraded resolution. From left to right: Cycle 2097 (year 2010) - Cycle 2147 (year 2014). From top to bottom: 45s integration time - 3600s integration time.

combinations of X and Y result in the same resolution, e.g. 2x1 and 1x2 but will undoubtedly deliver different results. This has its cause in the fact that small scale magnetic patches of same polarity are of the same scale as these binning boxes. However, this effect will decrease drastically at large values of X and Y, i.e. choosing a binning box of size 91x130 or 130x91 will make no difference, however a size of 130x1 or 1x130 will because the width of unipolar magnetic patches is of the same scale. In the previous subsection evidence was found that spatial resolution does not significantly affect the total unsigned flux in active regions. Does this fact hold true for all available resolutions?

Figure 12 displays in a scatter plot the ratio of active region flux from HMI resolution to the active region flux in the degraded resolution for all available resolutions. The control variable is in this case not the total resolution but the number of pixels inside the binning box ( $X \cdot Y$ ). The left plots depict the active region ratio in the year 2010 (rotation 2097) while the right side shows the situation in 2014 (rotation 2147). All four plots show a conic spreading of data points starting from the original resolution. Except of a dozen outliers in the 3600s plot from 2010, the active regions at highest available integration time show only weak deviations from one. The same arguments holds for the 45s integration time plot from 2010.

The top right plot depicts a totally different situation. A shape of an underlying process in the cloud of data points is clearly visible. Against all expectations, the

signal from active regions at lower resolutions is getting larger. A possible explanation for this exception is that the magnetic flux around active regions in the 45s rotation 2147 synoptic charts is distributed in such a way that only pixels of same polarity are getting binned together. Ergo, no magnetic flux cancellation takes place which implies that more and more magnetic flux amasses inside the active region of degraded charts. The outliers in the 2010 plots on the other hand are only appearing for large binning boxes. Furthermore it is known from table 11 that active regions only cover a marginal part of the solar surface. The active region flux is therefore more susceptible to large binning boxes.

Figure 13 displays the same situation but with quiet-Sun flux instead of active region flux. All four plots show an immediately visible dependence of quiet-Sun flux on the resolution. An increasing amount quiet-Sun flux is lost at low resolutions. The amount of missing flux is greater at activity minimum than at activity maximum which is consistent with the result from the previous subsection. However, the shape of the underlying curve which fits through the data points does not change. It is therefore possible to deduce the fact that both processes are governed by the same underlying function whose parameters are dependent on the rotation.

#### 4.3.1 Empirical Model

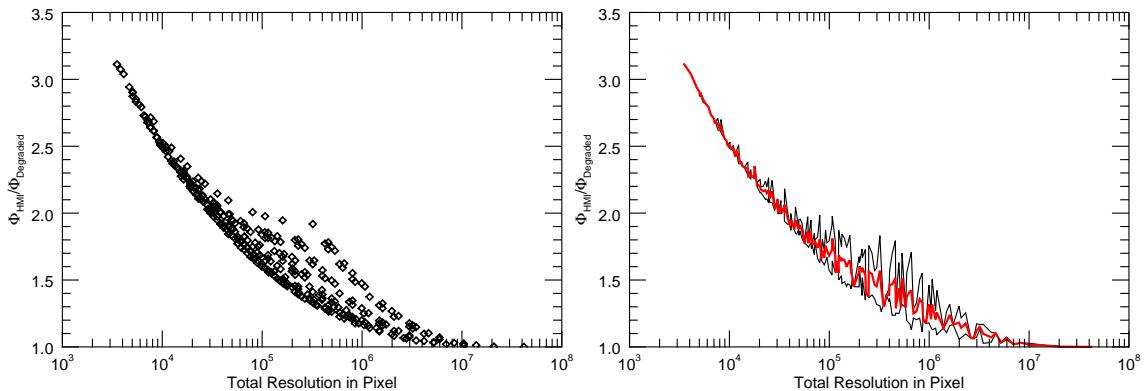


Figure 14: 315s integration time at rotation 297. Ratio of total unsigned flux at original resolution to degraded resolution. Left plot: Scatter plot at all available resolutions. Right plot: Red line - Averaging over all data points with same resolution. Black solid lines -  $1\sigma$  uncertainty range.

Figure 14 shows the ratio of the total unsigned flux at HMI resolution to the total unsigned flux at the degraded resolution. The left plot displays the ratio for all 361 binning boxes. The x-axis in this plot however denotes the total resolution in pixels, i.e. the ratio of the amount of pixels at HMI resolution to the amount of pixels in the binning box. As previously stated, some combinations of X and Y result in the

same resolution. The right plot illustrates the situation when all data points with same resolutions are averaged. The emerging uncertainties are illustrated as black lines. It is apparent that synoptic charts at a resolution range of  $10^5$  to  $10^7$  pixel are mostly affected by this uncertainty.

The missing signal from quiet-Sun regions and the total synoptic chart at degraded resolutions appear to have a common underlying law. The shape of the curves combined with the use of a semi-logarithmic scale imply that the underlying process can be identified to be of geometric nature, e.g.

$$f(x) = a \cdot x^b + c. \quad (8)$$

It is therefore possible to hypothesize that a geometric law describes the nature of image degradation. Proving this hypothesis could then lead to a complete explanation of the effect of image degradation.

A non-linear least square fit was performed on curve from figure 14 (right). The uncertainties have been used as weights in the fit. The resulting fit is displayed in figure 15. The direct comparison of the regression and the data points leads to the conclusion that the geometric model can indeed explain the effect of image degradation on the total magnetic flux. The geometric model is then fitted to all remaining integration times at rotation 2097 and 2147, to determine the model parameters for

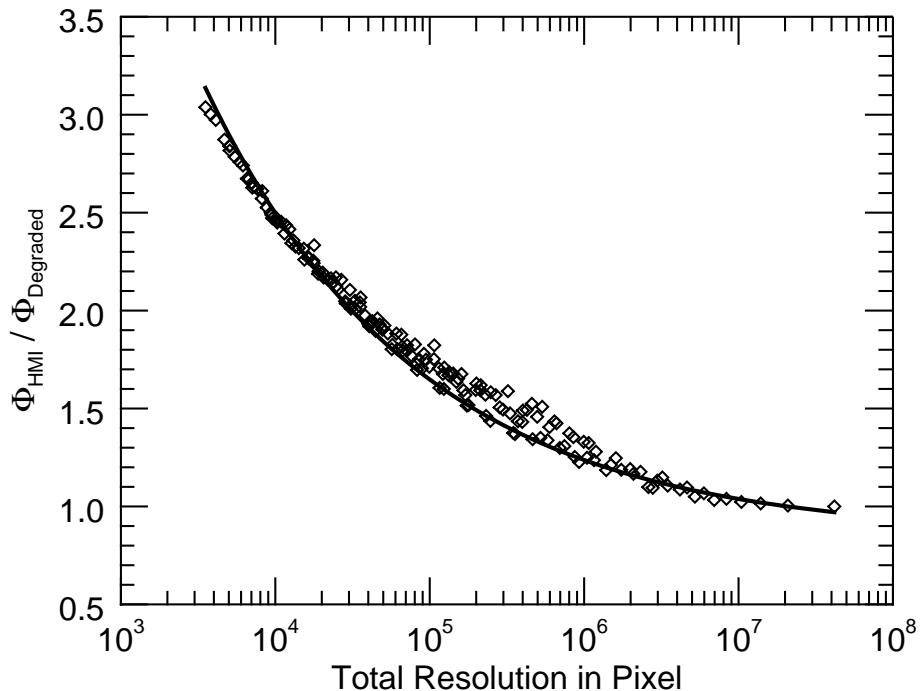


Figure 15: Geometric fit on the data. Diamonds: data points; Solid line: geometric model. Note that no error bars are included.

each rotation and each integration time. These parameters are listed in table 7. The parameter a has the most noticeable deviations. Its value at activity minimum is nearly 6-8 times as big as at activity maximum. It can therefore be interpreted as a key parameter in the process of signal loss. The parameter b describes the slope of the geometric model. A smaller exponent results in a steeper slope. Synoptic charts at activity minimum show a lower exponent as their counterparts at activity maximum. Therefore they are more susceptible to signal loss at lower resolutions. It can therefore be concluded that the parameter b contains information about the structuring of small scale magnetic elements. It will be larger if unipolar patches of magnetic elements are larger in size and decrease if those patches are smaller.

The constant c ensures that the curve converges to 1 at original resolution. The deviation of c from 1 in the estimated model parameters have their origin in the fact that the geometric was not restricted in its degrees of freedom, i.e. c was not restricted to one in the fitting process. Doing so let to convergence problems in the fitting process. Further improvement of the fitting methods, e.g. different algorithm or different software, could lead to more precise estimations.

Rotation 2097			Rotation 2147			
	a	b	c	a	b	
45s	30	-0.32	0.85	5.0	-0.23	0.89
	0.2	0.01	0.01	0.1	0.01	0.01
315s	42.8	-0.36	0.89	4.8	-0.22	0.89
	0.1	0.01	0.01	0.1	0.01	0.01
1215s	46.5	-0.36	0.89	4.5	-0.21	0.89
	0.1	0.01	0.01	0.1	0.02	0.01
3600s	43.6	-0.36	0.89	4.0	-0.21	0.87
	0.7	0.01	0.01	0.2	0.01	0.01

Table 7: Fit parameters of the geometric model  $f(x) = a \cdot x^b + c$  with their corresponding  $1\sigma$  confidence interval.

### 4.3.2 Spatial Resolution and Statistical Moments

Physical variables such as signal loss at a given resolution are a consequence of the spatial dynamics of the distribution function of a synoptic chart. Spatial dynamics denotes the fact that statistical parameters are dependent on the spatial resolution.

It is known from the previous subsections that quiet-Sun elements are covering most of the solar surface at any rotation and integration time. This implies that the empirical probability density function of a single pixel will have its maxima, which is denoted as  $A$ , centered at its median. The median can be considered to be approximately identical to the mean because in first order approximation it is expected that the amount of outgoing magnetic flux is identical to the amount of flux which flows back into the sun.

This distribution is furthermore characterized with a unique variance  $V$ . How are these parameters affected by spatial degradation? The huge amount of weak quiet-Sun flux implies that the probability for weak flux to cancel is much higher than for magnetic flux from larger quantiles of the distribution. It is therefore possible to predict without detailed knowledge of the distribution function that the maxima will shrink and corresponding with this the variance will increase (as seen in figure 10).

This effect was measured and is displayed in figure 16. A double log plot has been used to illustrate the spatial dynamics of the variance at rotation 2097 (red) and 2147 (blue). The exceptionally linear behaviour of both curves indicate that a geometric law fits onto the data points. However several attempts to fit a geometric model failed to converge. The constant shift in the double log plot can be explained by the fact that synoptic charts at activity maximum are measuring less weak flux compared to activity minimum but more flux with higher intensity. Figure 17 displays the effect of spatial resolution on measurements of the mean of the synoptic chart. The linear shape in the double log-plot are a strong evidence that the data follows a geometric law.

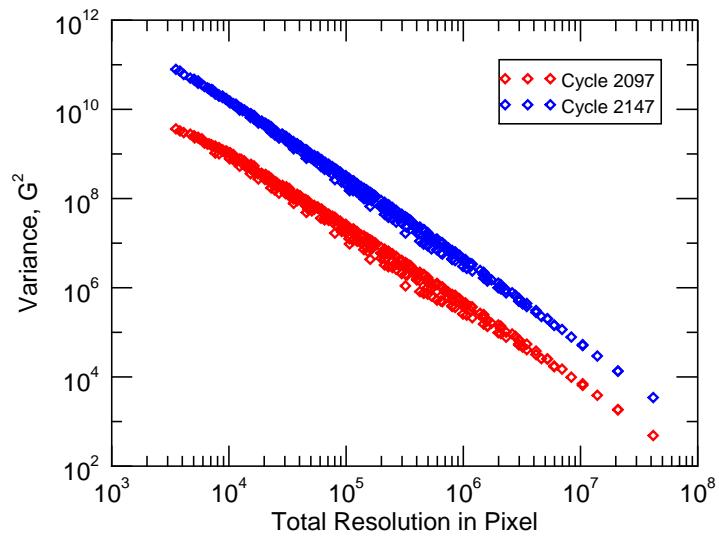


Figure 16: 315s synoptic chart. Variance at a given resolution. Red diamonds: Cycle 2097; Blue diamonds: Cycle 2147.

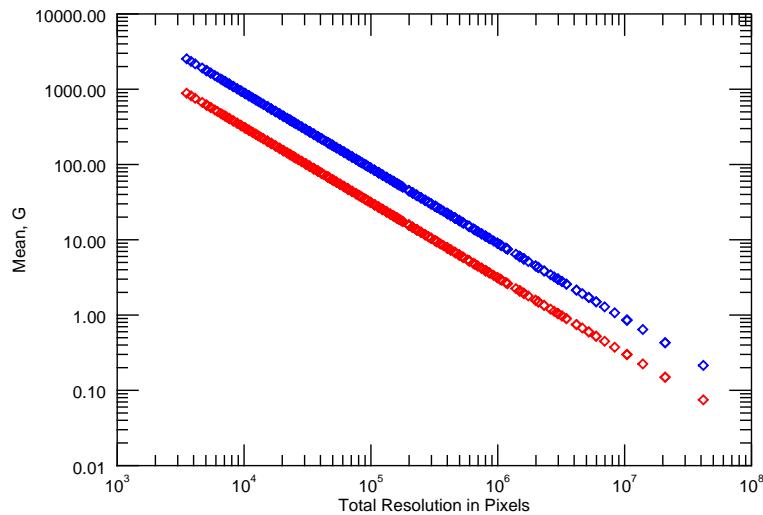


Figure 17: Mean of 315s synoptic charts at: Red diamonds: Cycle 2097; Blue diamonds: Cycle 2147. The rotation 2097 data has been multiplied by a factor of -1 to display the data on a double logarithmic plot.

## 5 Semi-Empirical Model for the 1st and 2nd Statistical Moment of Degraded Synoptic Charts

### 5.1 Motivation

Statistical moments describe the shape of a set of random variables. A theoretical model describing the dynamics of these moments can therefore determine the shape of the distribution functions at a given resolution. The total magnetic flux can then be calculated by integrating the distribution over the real line.

This procedure however relies heavily on the mathematical form of the distribution function and its underlying probability density function. Abramenko and Longcope [6] used MDI-HR mode magnetograms to measure the distribution functions of two active regions (NOAA 9077, NOAA 0061). They found out that only a log-normal model can explain the distribution of active region signals. A log-normal distribution is completely defined by the first two statistical moments, its mean and variance. Extending the findings of Abramenko and Longcope to quiet-Sun signals, i.e. verifying a log-normal behaviour for the whole magnetic signal, combined with a theoretical model of the spatial dynamics of the first and second statistical moment could then completely describe the effect of spatial resolution on magnetic flux measurements.

The spatial dynamics of the statistical moments can be derived theoretically by applying equation 5 on the general formula for the  $k$ th central moment  $\mu_k$ ,

$$\mu_k = \mathbb{E} \left( (yX - \mu)^k \right) = \frac{1}{n} \sum_{i=0}^n (X_i - \mu)^k. \quad (9)$$

I will predominantly focus on the derivation of the mean and variance as a function of resolution. This will be presented in great detail in chapter 5.3. Higher order moments have not been modelled because their theoretical derivations require complicated calculations of higher order joint cumulants.

### 5.2 Testing for Log-Normality

The graph in figure 18 displays the empirical probability density of a single pixel, constructed by normalising the histogram of real magnetic signal in the rotation 2147 315s synoptic chart. The bimodal structure of the pdf results from the noise removal process which removes most of the signal around 0G. Both halves of the density function are nearly identical to each other. Log-normality of the, e.g. right

half would then imply log-normality of the left and vice versa. The property of log-normality is given, if the logarithm of the data is normally distributed. Figure 19 compares the logarithmic distribution with a gaussian fit. The logarithmic distribution bears no definite resemblance with a gaussian distribution. We can therefore rule out that a log-normal model describes the distribution of synoptic chart data. No further possible distributions have been tested. This implies that it is currently not possible to calculate values for the total magnetic flux by integration. Ergo, it is not possible to analyse physical implications of the model.

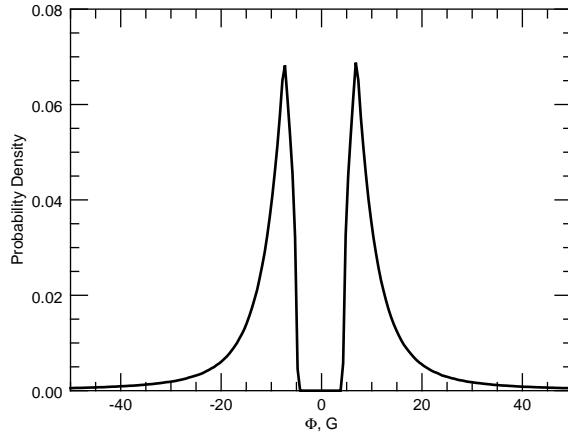


Figure 18: Black solid line: empirical probability distribution function of a single pixel in the rotation 2147 315s synoptic chart.

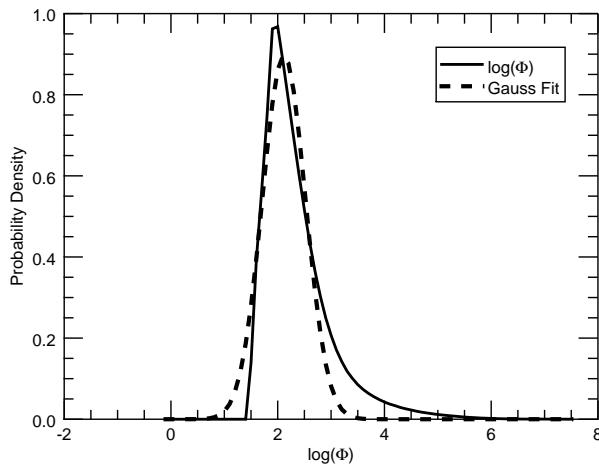


Figure 19: Test for log-normality. A set of data  $X$  can be interpreted to be distributed log-normal if  $\log(x)$  is distributed normally. Black solid line:  $\log(\Phi)$  for the right half of the whole pdf. Black dotted line: Gaussian fit on the distribution function of  $\log(\Phi > 0)$ .

### 5.3 Dynamics of Statistical Moments

The general equations of interest for the mean and empirical variance for a set of random variables  $y_i, i \in \{1, \dots, n\}$  are defined as<sup>1</sup>

$$\mu = \frac{1}{n} \sum_{i=1}^n y_i \quad (10)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2. \quad (11)$$

These definitions can be extended to the 2D case, e.g. magnetogram array of size NxM, by adding one more summation to the 1D case,

$$\mu = \frac{1}{N \cdot M} \sum_{i=1}^N \sum_{j=1}^M y_{i,j} \quad (12)$$

$$\sigma^2 = \frac{1}{N \cdot M} \sum_{i=1}^N \sum_{j=1}^M (y_{i,j} - \mu)^2. \quad (13)$$

This formulas can then be used to denote the empirical variance and mean of a arbitrary synoptic chart of dimension NxM. To compute these parameters at a given resolution of PxQ, we would have to use the following relations,

$$\mu(P,Q) = \frac{1}{P \cdot Q} \sum_{i=1}^P \sum_{j=1}^Q y'_{i,j} \quad (14)$$

$$\sigma^2(P,Q) = \frac{1}{P \cdot Q} \sum_{i=1}^P \sum_{j=1}^Q (y'_{i,j} - \mu(P,Q))^2. \quad (15)$$

Inserting equation 5 into both definitions completely defines the spatial dynamics of the first and second moment  $\mu_k(P,Q)$ ,

$$\mu(P,Q) = \frac{1}{P \cdot Q} \sum_{i=1}^P \sum_{j=1}^Q \sum_{k=\frac{N}{P}i+1}^{\frac{N}{P}(i+1)} \sum_{l=\frac{M}{Q}j+1}^{\frac{M}{Q}(j+1)} y_{k,l} \quad (16)$$

$$\sigma^2(P,Q) = \frac{1}{P \cdot Q} \sum_{i=1}^P \sum_{j=1}^Q \left( \sum_{k=\frac{N}{P}i+1}^{\frac{N}{P}(i+1)} \sum_{l=\frac{M}{Q}j+1}^{\frac{M}{Q}(j+1)} y_{k,l} - \mu(P,Q) \right)^2. \quad (17)$$

---

<sup>1</sup>The index notation in this chapter will continue to coincide with IDL's array notation, however the first element will be denoted as 1 instead of 0.

It is the purpose of the following subsections to give a theoretical explanation of the specific form of the graphs in the experimental findings, which have been presented in chapter 4.3.2, by recovering functions of the form  $\mu_k(P, Q) = \mathbb{F}(\mu_k(M, N), P, Q)$  from equations [14] and [15].

## 5.4 1D Approximation

The expansion of the square in equation 17 involves summation of up to 6 indices. Without loss of generality it is more convenient to start with the 1D case. This can be accomplished by reshaping the synoptic chart matrix  $Y$  of size  $N \times M$  into a vector of length  $P$  with  $P = N \cdot M$ , e.g.

$$Y = \begin{pmatrix} y_{1,1} & y_{2,01} & \dots & y_{N,1} \\ y_{1,2} & y_{2,2} & \dots & \\ \vdots & & \ddots & \vdots \\ y_{1,M} & \dots & & y_{N,M} \end{pmatrix} \rightarrow (y_1, y_2, \dots, y_P). \quad (18)$$

The degradation process of the vector  $Y$  to a lower resolution vector  $Y'$  of length  $x$  can then be modelled by summing vector elements in intervals of  $\{1, \frac{P}{x}\}$ , e.g.

$$Y' = \left( y'_1 = \sum \left( y_1, \dots, y_{\frac{P}{x}} \right), y'_2, \dots, y'_x \right). \quad (19)$$

The mean and empirical variance of the vector  $Y$  are defined as,

$$\mu = \frac{1}{P} \sum_{i=1}^P y_i \quad (20)$$

$$\sigma^2 = \frac{1}{P} \sum_{i=1}^P (y_i - \mu)^2, \quad (21)$$

whereas at a given resolution of  $x$  they are defined as

$$\mu'(x) = \frac{1}{x} \sum_{i=1}^x y'_i \quad (22)$$

$$\sigma^2(x) = \frac{1}{x} \sum_{i=1}^x (y'_i - \mu'(x))^2 \quad (23)$$

Inserting the 1D version of equation [5],

$$y'_i = \sum_{j=\frac{P}{x}i+1}^{\frac{P}{x}(i+1)} y_j \quad (24)$$

into equation [23] leads to

$$\sigma^2(x) = \frac{1}{x} \sum_{i=1}^x \left( \sum_{j=\frac{P}{x}i+1}^{\frac{P}{x}(i+1)} y_j - \frac{1}{x} \sum_{l=1}^x \sum_{j=\frac{P}{x}l+1}^{\frac{P}{x}(l+1)} y_j \right)^2 \quad (25)$$

Let us take a closer look at the summation process for the mean at resolution  $x$ . Consider the case where the vector  $Y$  has size 10 and we degrade its resolution by half, i.e.  $x=5$ . The mean of the vector  $Y'$  is then defined by the summation,

$$\mu(x) = \frac{y'_1 + y'_2 + y'_3 + y'_4 + y'_5}{5}. \quad (26)$$

Expressing the  $y'_i$  in terms of  $y_i$ , yields

$$\mu(x) = \frac{y_1 + y_2}{5} + \frac{y_3 + y_4}{5} + \dots + \frac{y_9 + y_{10}}{5} = \frac{10 \cdot \mu}{5} = \frac{P \cdot \mu}{x}. \quad (27)$$

We have therefore found the general function which describes the spatial dynamics of the mean of a 1D vector. Inserting this result back into equation [25] and expanding the square of the bracket, yields

$$\begin{aligned} \sigma^2(x) &= \frac{1}{x} \sum_{i=1}^x \left( \sum_{j=\frac{P}{x}i+1}^{\frac{P}{x}(i+1)} y_j - \frac{P}{x} \mu \right)^2 \\ &= \frac{1}{x} \sum_{i=1}^x \left( \left( \sum_{j(i)} y_j \right)^2 + \left( \frac{P}{x} \mu \right)^2 - 2 \frac{P}{x} \mu \sum_{j(i)} y_j \right). \end{aligned}$$

Drawing the summation over  $i$  into the bracket and making use of the general relation

$$\left( \sum_{j=0}^{n-1} Z_j \right)^2 = \sum_{j=0}^{n-1} Z_j^2 + \sum_{j \neq i}^{n-1} Z_i Z_j = \sum_{j=0}^{n-1} Z_j^2 + 2 \sum_{j < i}^{n-1} Z_i Z_j \quad (28)$$

leads to the simplified expression,

$$\sigma^2(x) = \frac{1}{x} \left( \sum_{i=1}^P y_i^2 + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} y_j y_k - \frac{P^2 \mu^2}{x} \right). \quad (29)$$

We can now express the summation over all  $y_i^2$  in terms of the variance at original resolution

$$\begin{aligned} \sigma^2 &= \frac{1}{P} \sum_{i=1}^P (y_i - \mu)^2 \\ \sigma^2 P &= \sum_{i=1}^P (y_i^2 + \mu^2 - 2y_i\mu) \\ \sum_{i=1}^P y_i^2 &= \sigma^2 P + P\mu^2 \end{aligned}$$

and insert this result into equation 29

$$\sigma^2(x) = \frac{1}{x} \left( \sigma^2 P + P\mu^2 \left( 1 - \frac{P}{x} \right) + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} y_j y_k \right). \quad (30)$$

Up to this point we have only simplified equation [25] into a more readable expression. Furthermore we have found a way in expressing the spatial dynamics of the variance in terms of its variance at original resolution. Starting from here, I will use approximations to find an algebraic expression, by beginning to apply the expectation operator  $\mathbb{E}$  on both sides of the equation. The operator  $\mathbb{E}$  ignores scalars by definition and only operates on the double sum of  $y_j y_k$ ,

$$\bar{\sigma}^2(x) = \frac{1}{x} \left( \sigma^2 P + P\mu^2 \left( 1 - \frac{P}{x} \right) + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} \mathbb{E}[y_j y_k] \right). \quad (31)$$

The property of linearity has been used here to apply  $\mathbb{E}$  directly on the product of  $y_j y_k$ . Applying the law of total expectation

$$\mathbb{E}[y_j y_k] = \mathbb{E}[y_j] \cdot \mathbb{E}[y_k] + \text{Cov}(y_j, y_k) \quad (32)$$

yields the following result,

$$\bar{\sigma}^2(x) = \frac{1}{x} \left( \sigma^2 P + P \mu^2 \left( 1 - \frac{P}{x} \right) + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} \mathbb{E}[y_j] \cdot \mathbb{E}[y_k] + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} \text{Cov}(y_j, y_k) \right). \quad (33)$$

Before continuing with further approximations, it is important to understand the double summation over  $j < k$ . What does it actually mean? The double sum has its cause in the squaring of

$$y_i'^2 = \left( \sum_{\frac{P}{x}i+1}^{\frac{P}{x}(i+1)} y_j \right)^2. \quad (34)$$

Consider the case in which  $P, x, i$  equal 8, 2, 1 respectively. Equation [34] would then read

$$y_1'^2 = (y_1 + y_2 + y_3 + y_4)^2, \quad (35)$$

which yields

$$y_1'^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + 2 \cdot (y_1(y_2 + y_3 + y_4) + y_2(y_3 + y_4) + y_3y_4). \quad (36)$$

We can therefore conclude that the double sum only acts on vector elements inside the binning interval and that the ensuing summation over all intervals does not increase the range of the multiplication. Note that the double sum contains  $P/x$  choose 2 ( $P/x ; 2$ ) elements. Let us continue at equation [33]. I will now introduce a subtle and important approximation. I will assume heuristically that

$$\mathbb{E}[y_j] \cdot \mathbb{E}[y_k] = \alpha \quad \forall j, k \quad (37)$$

It is immediately apparent that this approximation is not justifiable a priori. However, we already have information about key properties of the magnetic flux distribution, e.g. table 11 shows that active region pixels only cover a marginal part of the synoptic chart at rotation 2097 and up to 1/10th at rotation 2014. These experimental observations can indeed justify the approximations in equation 37, which we will assume heuristically at this point. The constant  $\alpha$  can then be calculated numerically.

Applying the approximation into the model results in the following expression

$$\bar{\sigma}^2(x) = \frac{1}{x} \left( \sigma^2 P + P\mu^2 \left( 1 - \frac{P}{x} \right) + 2x \cdot \left( \frac{\frac{P}{x}}{2} \right)^\alpha + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} \text{Cov}(y_j, y_k) \right). \quad (38)$$

I will now use the definition of the gamma function  $\Gamma(n) = (n-1)!$  combined with the definition of the binomial coefficient in terms of faculties to interpolate the third term inside the bracket,

$$\bar{\sigma}^2(x) = \frac{1}{x} \left( \sigma^2 P + P\mu^2 \left( 1 - \frac{P}{x} \right) + 2x \cdot \frac{\Gamma(\frac{P}{x} + 1)}{2 \cdot \Gamma(\frac{P}{x} - 1)}^\alpha + 2 \sum_{i=1}^x \sum_{j(i) < k(i)} \text{Cov}(y_j, y_k) \right). \quad (39)$$

Note that the formula is only well-defined for  $P/x > 2$ . The case  $P/x=1$  can be regarded as a special case where the double summation is equal to 0 which prevents the appearance of the binomial coefficient. Equation 39 leaves us with an unknown triple summation over covariances between two pixels. At this step, I will introduce a new approximation. Instead of considering specific pixel-by-pixel covariances of the form

$$\text{Cov}(y_1, y_2) + \text{Cov}(y_1, y_3) + \dots + \text{Cov}(y_2, y_3) \dots, \quad (40)$$

I will assume that the pixel-by-pixel covariance only depends on the euclidean distance between the observed pixels, i.e. pixel-by-pixel covariances can be interpolated by a function  $T$  which is dependant on the euclidean distance  $d$ :

$$\text{Cov}(y_1, y_2) = \text{Cov}(y_2, y_3) = \dots = T(d=1). \quad (41)$$

This results in a decoupling of the summation over  $i$  and the double summation over  $j$  and  $k$  in equation [39]. We can furthermore replace the double summation with a single one, which runs over the possible distances inside the binning interval. An additional factor of  $n_d$  counts how often a specific distance inside the binning interval occurs.

$$\sum_{i=1}^x \sum_{j(i) < k(i)} \text{Cov}(y_j, y_k) = x \sum_d n_d T(d) \quad (42)$$

The listed table entries in table 8 indicate that  $n_d$  can be modelled by a linear

Table 8: Counting the occurrence of specific distances for different binning intervals.

Binning Elements:	$\{1, 2\}$	$\{1, 2, 3\}$	$\{1, 2, 3, 4\}$	$\{1, 2, 3, 4, 5\}$
Distance: 1	1	2	3	4
2	-	1	2	3
3	-	-	1	2
4	-	-	-	1
5	-	-	-	-

function

$$n_d = \frac{P}{x} - d. \quad (43)$$

Thus, the spatial dynamic of the variance of reshaped synoptic charts is determined by

$$\begin{aligned} \bar{\sigma}^2(x) = & \frac{1}{x} \left( \sigma^2 P + P \mu^2 \left( 1 - \frac{P}{x} \right) + \right. \\ & \left. x \cdot \left( \frac{\Gamma(\frac{P}{x} + 1)}{\Gamma(\frac{P}{x} - 1)} \alpha + 2 \sum_{d=1}^{\frac{P}{x}-1} \left( \frac{P}{x} - d \right) T(d) \right) \right). \end{aligned} \quad (44)$$

## 5.5 From 1D to 2D

Extending the results from the 1D approximation process to the 2D case of matrices is a nearly trivial matter. Replacing  $x$  and  $P$  in  $\mu(x)$  (see equation [27]) with  $P \cdot Q$  and  $N \cdot M$  respectively results in

$$\mu_{\text{degraded}}(P, Q) = \frac{N \cdot M}{P \cdot Q} \cdot \mu = \frac{N_{\text{HMI}}}{N_{\text{degraded}}} \mu, \quad (45)$$

which determines the expectation value of a synoptic chart at a given resolution of  $P \cdot Q$ . A non-linear least square fit is performed on the data from figure 17 (rotation 2097 plot; red diamonds). The resulting fit is displayed in figure 20. A reduced chi-square value of 1.01 tells us that it is not possible to refute the hypothesis that equation 45 models the data.

This procedure of replacing  $x$  and  $P$  with their 2D counterparts has its limitations when being applied on the 1D equation for the variance (see equation [44]). It can be easily verified that all terms except the last one are independent of the shape of

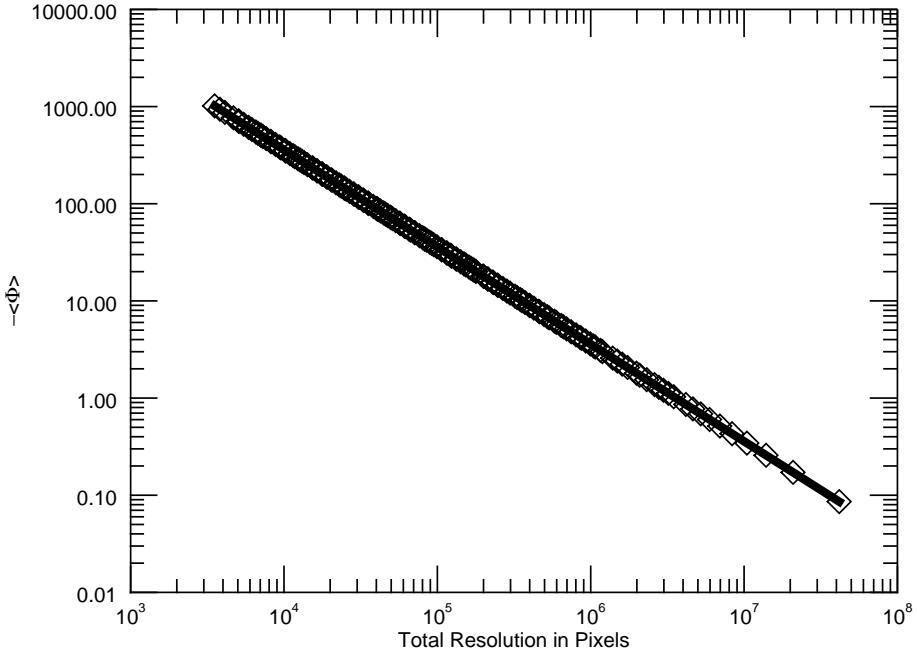


Figure 20: Comparing theoretically predicted model for the mean of 315s rotation 2147 synoptic chart with experimental data. A factor of -1 was multiplied on the fit and data points to allow the use of double logarithmic scales. Diamonds: Measured data; Solid black line: Theoretical Model

the binning box or interval. This means that the replacement of  $x$  and  $P$  with their counterparts is allowed for the first three terms. However, certain modifications to the last term containing a sum over covariances is needed. This has its cause in the two-dimensionality of the binning box compared to the 1D binning interval. This two-dimensionality leads to irrational inputs for  $T$ , e.g. the euclidean distance between  $y_{1,1}$  and  $y_{2,2}$  is  $\sqrt{2}$ . It is therefore necessary to find a function which counts the number of occurrences of a specific distance inside a binning matrix of arbitrary size. Without loss of generality we can consider the case of a binning matrix of size  $2 \times 3$  ( $A \times B$ ), e.g.

$$X = \begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ x_{1,3} & x_{2,3} \end{pmatrix} \quad (46)$$

Let  $D(x)$  be the symmetric distance matrix of  $X$  with size  $AB \times AB$ , whose  $i,j$ th element  $D_{i,j}$  is defined as the euclidean norm between the  $((j - 1 \bmod A) + 1), \lceil \frac{j}{A} \rceil$  and  $((i - 1 \bmod A) + 1), \lceil \frac{i}{A} \rceil$  element of  $X$ , e.g.

$$D_{i,j} = \sqrt{(((j - 1 \bmod A) + 1) - ((i - 1 \bmod A) + 1))^2 + \left(\left\lceil \frac{j}{A} \right\rceil - \left\lceil \frac{i}{A} \right\rceil\right)^2}. \quad (47)$$

By this definition, the distance matrix D of X becomes<sup>2</sup>,

$$D(X) = \begin{pmatrix} 0 & 1 & 1 & \sqrt{2} & 2 & \sqrt{5} \\ 1 & 0 & \sqrt{2} & 1 & \sqrt{5} & 2 \\ 1 & \sqrt{2} & 0 & 1 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 1 & 0 & \sqrt{2} & 1 \\ 2 & \sqrt{5} & 1 & \sqrt{2} & 0 & 1 \\ \sqrt{5} & 2 & \sqrt{2} & 1 & 1 & 0 \end{pmatrix}. \quad (48)$$

The summation over  $j < k$  in the variance equation implies that we only have to consider the top diagonal matrix without the principle diagonal of D. For an arbitrary binning box of size AxB, we would then only need to count how often a specific distance d occurs. This can be accomplished by counting unique elements in the distance matrix. However, a closed algebraic formula has been found which determines the number of occurrences of a specific distance  $d_{i,j}$  based on the size of the binning array, e.g.

$$\eta(A, B) = \sum_{j=0}^{B-1} \sum_{i=0}^{A-1} \begin{cases} j = 0 & \begin{cases} i = 0 & 0 \\ i \geq 1 & B \cdot (A - i) \cdot T(i) \end{cases} \\ j \geq 1 & \begin{cases} i = 0 & (B-j) \cdot A \cdot T(j) \\ i \geq 1 & (B-j) \cdot 2(A - i) \cdot T(\sqrt{i^2 + j^2}) \end{cases} \end{cases} \quad (49)$$

This formula was not derived theoretically but by trial-and-error<sup>3</sup>. However, it was tested on various test matrices and it was able to predict all appearing distances and their number of occurrences. Thus, we have found the final equation which describes the spatial dynamics of the variance of synoptic charts:

$$\bar{\sigma}^2(P, Q) = \frac{1}{P \cdot Q} \left( \sigma^2 N \cdot M + N \cdot M \mu^2 \left( 1 - \frac{N \cdot M}{P \cdot Q} \right) + P \cdot Q \cdot \left( \frac{\Gamma\left(\frac{N \cdot M}{P \cdot Q} + 1\right)}{\Gamma\left(\frac{N \cdot M}{P \cdot Q} - 1\right)} \alpha + 2\eta\left(\frac{N}{P}, \frac{M}{Q}\right) \right) \right). \quad (50)$$

---

<sup>2</sup>The first row of D defines the euclidean distances between  $x_{1,1}$  and all other elements of X, e.g. the distance of  $x_{1,1}$  with itself is equal to 0 etc. The second row defines the distances for the element  $x_{2,1}$ , the third row for the element  $x_{1,2}$  and so on.

<sup>3</sup>No analytical expression for  $\eta$  is needed when working with matrix based programming languages, i.e. it is possible to calculate  $\eta$  directly by inserting D into T.

## 5.6 Estimation of Model Parameters

A core problem of modelling the variance lies in the estimations of the free parameters  $\alpha$ , which requires information about sums over products of  $\mathbb{E}(y_j) \cdot \mathbb{E}(y_k)$ , and  $\eta$ , which involves a sum over covariances between  $y_j$  and  $y_k$ . A detailed examination of the properties of  $y_j$  and  $y_k$  requires a well-defined local neighbourhood or group of pixels from which the expectation value and covariance is then estimated. This group of pixels can be chosen to be rows and columns of the matrix, which introduces a two-dimensionality (vertical and horizontal direction) which needs to be considered in the modelling process. It is therefore necessary to measure both, the horizontal and vertical pixel-by-pixel covariance and check for possible spatial anisotropies in the data of interest.

This method of comparing matrix columns to columns and rows to rows is then used for the estimations of the parameters,  $\alpha$  and  $\eta$ , in the variance model. In the previous section the choice of replacing the sum over covariances between  $y_j$  and  $y_k$  with a function  $T$  depending on the euclidean distance between  $j$  and  $k$  has been introduced. The first step towards the estimation of  $\eta$  lies in finding a mathematical model which explains the average covariance between two pixels with euclidean norm  $d$ . The following steps were taken in the calculation of  $T$ . Let the synoptic chart be an arbitrary matrix of size  $A \times B$  ( $A$  columns,  $B$  rows). As it was previously motivated, it is necessary to determine  $T$  both vertically and horizontally.

To measure the horizontal part of  $T$ , it is necessary to create a matrix  $H$  of size  $A \times A$  where the  $H_{i,j}$  element denotes covariance between the  $j$ th and  $i$ th column. The same procedure is then used to create a matrix  $V$  of size  $B \times B$  for the vertical part of  $T$ . IDL's function CORRELATE with the synoptic chart matrix as input and the covariance parameter set to one, produces the required matrix  $H$ . Letting the transpose operator act on the input matrix produces the covariance matrix  $V$ . We then assign to each element of  $H$  and  $V$  its corresponding euclidean norm between  $i$  and  $j$ .

The final step involves averaging all values of  $H$  and  $V$  with the same euclidean distance. Figure 21 displays the horizontal part of  $T$  with the 315s synoptic charts from rotation 2097 and 2147 as input. A direct comparison between rotation 2097 and 2147 reveals that the pixel-by-pixel covariance is by a factor of 10 larger at rotation 2147. Furthermore, both curves converge asymptotically to a constant  $c$ . The dotted lines above and below the data points visualize the  $1\sigma$  confidence interval of the averaging process. The confidence tubes in both plots are decreasing asymptotically in size with increasing distance  $d$ .

However, the average pixel-by-pixel covariance at  $d=0$  is shifted more towards

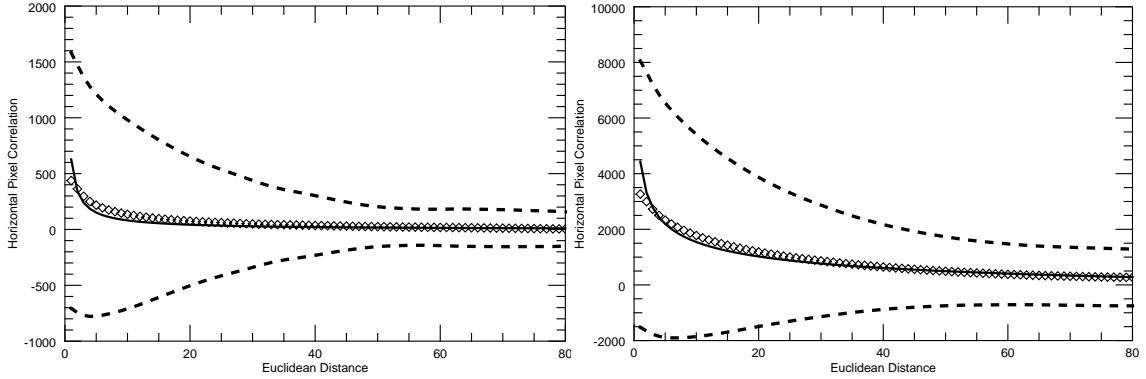


Figure 21: Pixel-by-pixel covariance as a function of euclidean distance between pixels. 315s Synoptic Chart; Left plot: Cycle 2097, Right plot: Cycle 2147. Diamonds: Horizontal pixel-by-pixel covariance; Solid Line: Geometric Fit; Dotted line:  $1\sigma$  confidence interval.

positive values than negative ones. A possible explanation for this behaviour is the fact that small scale magnetic elements in the HMI magnetograms are generally larger than one pixel. Combined with the fact that covariance measures how a change in one random variable would affect the other one, gives an explanation for this shift. Figure 25 in the appendix displays the vertical pixel-by-pixel covariances for the 315s integration time synoptic charts. The vertical covariances behave in similar way to their horizontal counterparts but with a much narrower confidence tube at greater distances.

A non-linear fit with a geometric model of the form

$$T(d) = ax^b + c. \quad (51)$$

was performed on the data. The model parameter for the horizontal and vertical data are listed in table 15 and 16 (see appendix). The relatively large uncertainty values for the estimated model parameters have their most probable cause in the large confidence intervals from figure [23]. Ignoring the uncertainties, i.e. fitting equation [51] on the data points with no weights, decreases the uncertainty range to 1%, while the model parameters stay approximately the same. However, the model parameters for the horizontal and vertical case lie in their mutual uncertainty range, which implicates that it is not possible to refute the hypotheses that an anisotropy in data exists.

The different exponents for both rotations are a hint on a fundamentally different structuring of the magnetic elements on the solar surface. The exponent at activity minimum is smaller than at activity maximum. This difference implies that  $T(d)$  decreases faster with increasing distance at activity minimum than at activity

	Integration Time		
Average Area [10 <sup>6</sup> km <sup>2</sup> ]	1215	3600	
CR 2047	3.14	6.72	
CR 2147	3.58	7.40	

Table 9: Average area covered by patches of magnetic elements of same polarity for both rotations. Only synoptic charts with the highest integration times have been used because of methodical constraints.

maximum. This, in turn, implies that magnetic elements of same polarity cluster themselves together into bigger patches at activity maximum. This effect is also directly observable by calculating the average size of unipolar patches in the synoptic charts at both rotations. The results of these calculations for two integration times are listed in table 9. As expected and consistent with the results of the covariance analysis, the average area covered by unipolar patches of magnetic elements is significantly smaller at activity minimum compared to activity maximum.

The fit parameters for the function T can now be inserted into the definition of  $\eta$  in equation [49]. The uncertainty of  $\eta$  which I will denote as  $\sigma_\eta$  can be calculated with the error law of Gauss,

$$\sigma_{f(p_i)} = \sqrt{\sum_{j=1}^n \left( \frac{df(p_i)}{dp_j} \sigma_{p_j} \right)^2}, \quad (52)$$

where  $f(p_i)$  denotes a function of parameters  $p_i$  and  $\sigma_{p_j}$  the corresponding uncertainties for these parameters.

Applying this law on the function T results in the following expression,

$$\sigma_T = \sqrt{(x^b \sigma_a)^2 + (abx^{b-1} \sigma_b)^2 + (\sigma_c)^2} \quad (53)$$

which is used to calculate the uncertainties of  $\eta$ . However, the large uncertainties of the fit parameters of T imply that the estimations of  $\eta$  will also suffer from extremely large uncertainties. These uncertainties follow a geometric law and are proportional to

$$\frac{\sigma_\eta}{\eta} \propto \left( \frac{N_{\text{Original}}}{N_{\text{Degraded}}} \right)^{1/b}. \quad (54)$$

The relative uncertainties are displayed in figure 22 (right plot).

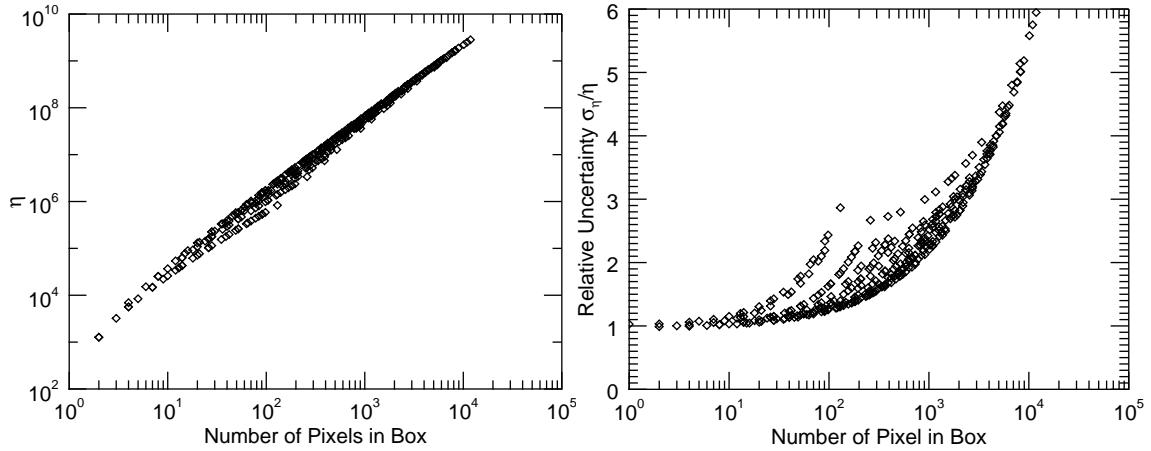


Figure 22: Estimation for model parameter  $\eta$ , rotation 2097 315s integration time data; Left plot:  $\eta$  as a function of binned pixel. Right plot:  $\sigma_\eta$  as a function of binned pixel

The last step involves determining the parameter  $\alpha$ , which we assumed heuristically to be a constant. This assumption is motivated by the graphs displayed in figure 23. These plots were created by calculating the expectation value for each row and column of the 315s integration time synoptic charts. The top side plots depict the vertical direction, whereas the bot side plots show the expectation values of the columns (horizontal).

The choice of  $\alpha = \text{const.}$  is justifiable, if the expectation value does not show strong dependence of image position (horizontal and vertical) and significant deviation from the mean of the whole chart  $\mu$ . This condition is fulfilled for the rotation 2097 synoptic charts, because the sun was at its activity minimum at that specific rotation. The situation is of a different kind at activity maximum, where graphs become more turbulent and erratic. This flaw transfers itself to the estimation of the parameter  $\alpha$  which is given in table 10. The  $1\sigma$  uncertainty of  $\alpha$  at activity maximum is exceptionally large.

	CR 2097		CR 2147	
	$\alpha$ [G]	$\sigma_\alpha$ [G]	$\alpha$ [G]	$\sigma_\alpha$ [G]
45s	0.47	4.02	5.78	77.0
315s	0.52	4.17	5.93	77.5
1215s	0.54	4.23	5.96	77.6
3600s	0.57	4.34	6.10	79.4

Table 10: Estimation of model parameter  $\alpha$

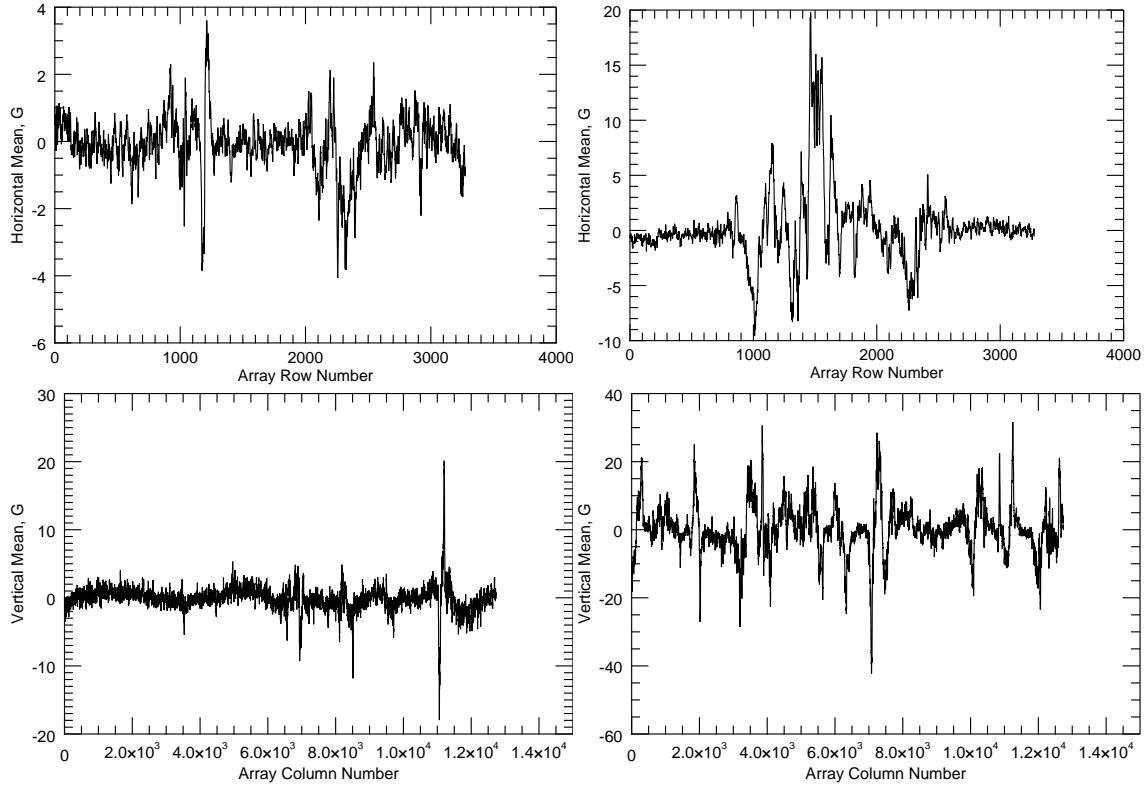


Figure 23: 315s Synoptic Chart; Left: Cycle 2097, Right: Cycle 2147

Here, the parameter  $\alpha$  has been calculated in a similar way to the function  $T$ , i.e.  $\mathbb{E}(y_j) \cdot \mathbb{E}(y_k)$  as a function of distance. The listed values in the table have been subsequently derived by averaging over all distances, horizontally and vertically. This rough procedure strips the free model parameter from all its spatial information but substantially simplifies the analysis. Furthermore, it can be interpreted from table 10 that the parameter  $\alpha$  with its corresponding factor of  $N_{Box}$  choose 2 is much smaller than  $\eta$ . Possible spatial dependences can therefore be neglected.

The estimations of  $\alpha$  and  $\eta$  can now be used as input for equation [50]. Figure 24 displays modelled and experimental data for both rotations at an integration time of 315s. The red solid line depicts the experimental data, gained from measuring the variance at a given resolution, while the blue line depicts the modelled data from equation [50] with the corresponding values of  $\alpha$  and  $\eta$  as input. The modelled data has exceptional resemblance with the experimental data at all resolutions. Nearly all local extrema in the experimental data are reproduced by the model. However, a small deviation from the original data can still be seen. This discrepancy has its most probable cause in the large uncertainties of  $\eta$  and  $\alpha$ , which is also the reason that no error bars have been included in the plots. Further investigation into more robust estimation methods of the model parameters are needed to improve model predictions with acceptable confidence intervals. One possible way of enhancing

the performance of the model would be to introduce spatial dependences of  $\eta$  and  $\alpha$ . Further empirical findings about the spatial distribution of covariances could justify to reverse the decoupling process in equation [41], which would complicate the estimation of the model parameters but could possibly lead to much more precise estimations of the variance.

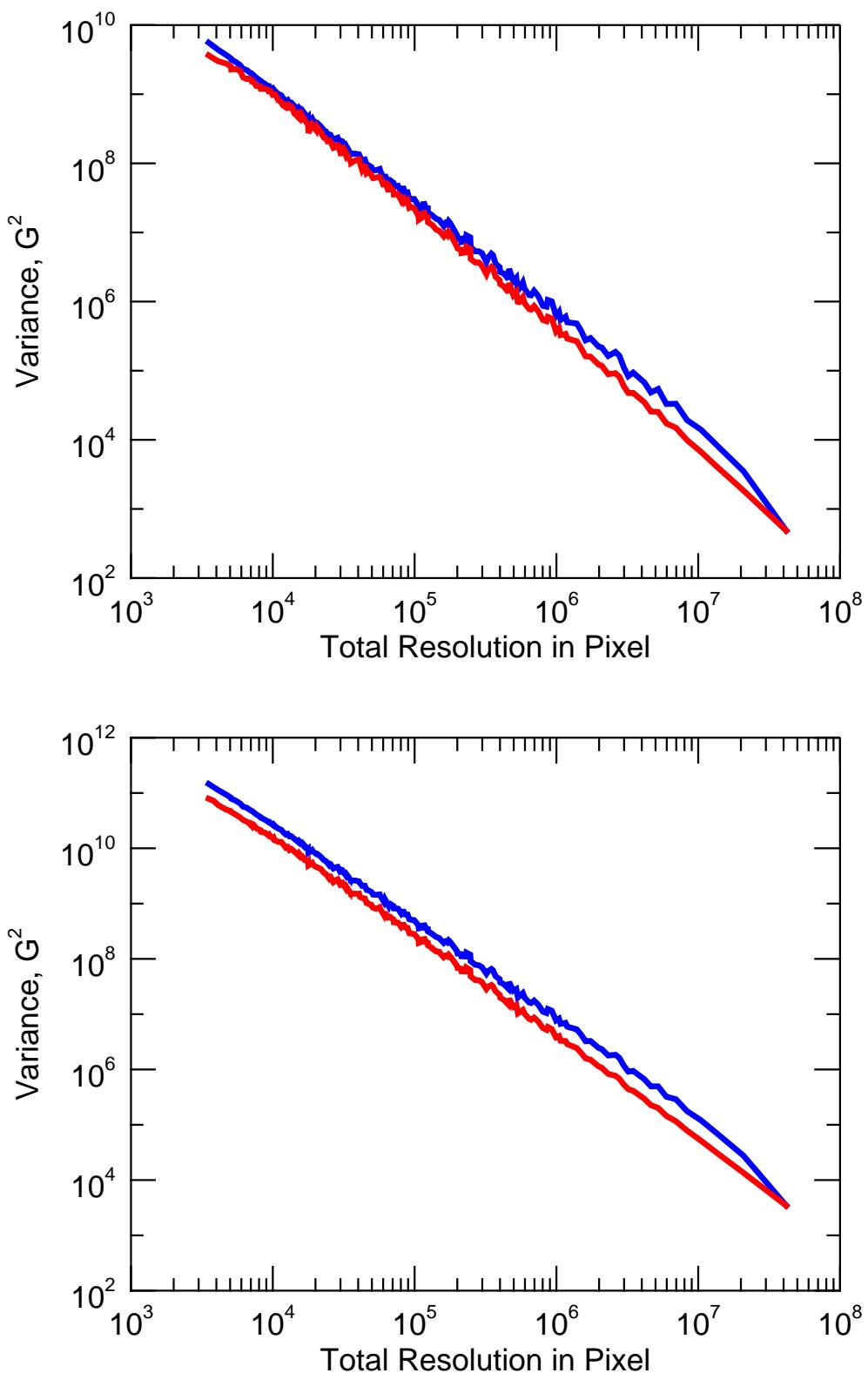


Figure 24: Comparing the modelled variance with experimental data from 315s integration time synoptic chart. Top figure: Cycle 2097; Bottom figure: Cycle 2147. Blue solid line: Model; Red solid line: Experimental Data

## 6 Conclusion

On the basis of HMI line-of-sight magnetograms, it was demonstrated that spatial resolution behaves as an important variable which needs to be properly accounted for in quantitative studies of the solar magnetic surface flux. Especially relations which are connected to the quiet-Sun flux are affected by non-sufficient resolution. HMI line-of-sight magnetograms with different integration times from two distinct rotations (2097 - activity minimum; 2147 activity maximum) have been used to create synoptic charts of the solar surface.

A new technique has been developed for this analysis which separates active regions more effectively than Karen Harvey's Root-Mean-Square-Difference method from 1994. The separated regions have then been analysed individually. It was shown that results obtained from HMI synoptic chart analysis are measuring twice as much quiet-Sun flux as ground based observations from NSO/KP, while measurements of active region flux are not deviating significantly from HMI measurements, even at extreme low resolutions.

It is estimated that MDI measurements of the quiet-Sun flux are missing approximately 10% of the signal which HMI measures. A comparison of the mean unsigned quiet-Sun flux as it is observed by HMI and Graham *et al.*'s numerical prediction of 50G reveals that even HMI is suffering from massive flux cancellation at sub-resolution scales.

HMI synoptic charts have been degraded to 361 different resolutions to find evidence for an underlying law which governs the process of signal loss. It was empirically shown that magnetic flux loss follows a geometric law of the form

$$\text{Signal Loss (Resolution)} = a \cdot \text{Resolution}^{-1/b} + c. \quad (55)$$

The model parameters  $a$  and  $b$  are heavily dependent on the level of activity of the solar magnetic field and to a minor degree on the specific integration time used. Especially the 45s synoptic charts gave rise to results, which were deviating significantly from synoptic charts measured at higher integration times.

A detailed analysis of synoptic chart noise signals, indicates that new corrections need to be introduced in the noise removal process. A significant amount of real signal fell below the noise threshold, which lead to an underestimation of the 45s, 1215s and 3600s integration time results in this analysis. Results obtained at these integration times can therefore only be seen as a lower bound. The 45s measurements have shown significant deviations from their counterparts with higher integration time.

It was shown that the empirical distribution function of the pixels does not follow a log-normal law. No further suitable distribution has been found. However, a semi-empirical model for the first and second statistical moment has been developed. This model can predict the mean and the variance at a given resolution and is dependent, in the case of the variance, on two free parameters.

These model parameters have been determined by a detailed pixel-by-pixel covariance analysis of the solar surface. Interestingly, the pixel-by-pixel covariance of synoptic charts is also governed by a geometric law. The first moment, the expectation value of the whole chart, was modelled successfully. The theoretical predictions for the second moment, the variance, are approximately the same as the measured data. The general trend and nearly all local extrema appearing in the predicted data are agreeing well with the measured data. Further investigation is needed to improve model parameters and close the gap between predictions and measured variance.

An optimized model could then be used as input for potential distributions to determine the shape of the distribution at a given resolution. Physical parameters such as total magnetic flux could then be calculated directly from the transformed distribution.

## References

- [1] K. L. Harvey, *Measurements of solar magnetic fields as an indicator of solar activity evolution*, Proceedings of the Workshop on the Solar Electromagnetic Radiation Study for Solar Cycle 22, held in Boulder, CO, 3-7 Jun. 1991. Edited by Richard F. Donnelly, 1992., p.113 (1992).
- [2] K. L. Harvey, *Irradiance models based on solar magnetic fields*, International Astronomical Union Colloquium **143**, 217 (1994).
- [3] Krivova, N. A. and Solanki, S. K., *Effect of spatial resolution on estimating the sun's magnetic flux*, A&A **417**, 1125 (2004).
- [4] J. O. Stenflo, *Solar magnetic fields*, Journal of Astrophysics and Astronomy **29**, 19 (2008).
- [5] M. J.P.Graham, S.Danlionic, *Turbulent magnetic fields in the quiet sun: Implications of hinode observations and small-scale dynamo simulations*, The Astrophysical Journal **693** (2009).
- [6] V. Abramenko and D. Longcope, *Distribution of the magnetic flux in elements of the magnetic field in active regions*, Astrophys. J (2005).
- [7] K. D. Leka, G. Barnes, and E. L. Wagner, *Evaluating (and improving) estimates of the solar radial magnetic field component from line-of-sight magnetograms*, Solar Physics **292** (2017).
- [8] H. J. Hagenaar, C. J. Schrijver, and A. M. Title, *The properties of small magnetic regions on the solar surface and the implications for the solar dynamo(s)*, The Astrophysical Journal **584**, 1107 (2003).
- [9] Y. Liu, J. T. Hoeksema, P. H. Scherrer, J. Schou, S. Couvidat, R. I. Bush, T. L. Duvall, K. Hayashi, X. Sun, and X. Zhao, *Comparison of line-of-sight magnetograms taken by the solar dynamics observatory/helioseismic and magnetic imager and solar and heliospheric observatory/michelson doppler imager*, Solar Physics **279**, 295 (2012).
- [10] I. Chertok, V. Grechnew, and A. Abunin, *An early diagnostics of the geoeffectiveness of solar eruptions from photospheric magnetic flux observations: The transition from soho to sdo*, ArXiv e-prints (2017).

## 7 Appendix

### 7.1 HMI Data

Table 11: Important results for, Top: Cycle 2097; Bottom: Cycle 2147

CR 2097 Data	45s	315s	1215s	3600s
$\Phi$	$2.3 * 10^8$ G	$2.6 * 10^8$ G	$2.7 * 10^8$ G	$2.7 * 10^8$ G
$\Phi_{\text{AR}}$	$1.3 * 10^7$ G	$1.7 * 10^7$ G	$1.9 * 10^7$ G	$2.5 * 10^7$ G
$\Phi_{\text{QS}}$	$2.2 * 10^8$ G	$2.4 * 10^8$ G	$2.5 * 10^8$ G	$2.4 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	16.9	14.4	13.3	9.7
AR area	0.30%	0.78%	1.28%	2.40%
$<  \Phi_{\text{QS}}  >$	33.0 G	17.8 G	12.0 G	9.1 G

CR 2147 Data	45s	315s	1215s	3600s
$\Phi$	$5.2 * 10^8$ G	$5.4 * 10^8$ G	$5.5 * 10^8$ G	$5.5 * 10^8$ G
$\Phi_{\text{AR}}$	$1.8 * 10^8$ G	$2.3 * 10^8$ G	$2.4 * 10^8$ G	$2.6 * 10^8$ G
$\Phi_{\text{QS}}$	$3.4 * 10^8$ G	$3.2 * 10^8$ G	$3.2 * 10^8$ G	$2.9 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	1.9	1.4	1.3	1.1
AR area	3.79%	7.96%	9.91%	13.40%
$<  \Phi_{\text{QS}}  >$	43.9 G	23.2 G	15.7 G	12.0 G

## 7.2 MDI-A Data

Table 12: Important results for, Top: Cycle 2097; Bottom: Cycle 2147

CR 2097 MDI-A Data	45s	315s	1215s	3600s
$\Phi$	$1.8 * 10^8$ G	$2.3 * 10^8$ G	$2.5 * 10^8$ G	$2.5 * 10^8$ G
$\Phi_{\text{AR}}$	$1.3 * 10^7$ G	$1.6 * 10^7$ G	$1.9 * 10^7$ G	$2.5 * 10^7$ G
$\Phi_{\text{QS}}$	$1.7 * 10^8$ G	$2.1 * 10^8$ G	$2.3 * 10^8$ G	$2.2 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	13.1	12.7	12.1	9.0
AR area	0.32%	0.81%	1.32%	2.44%
$<  \Phi_{\text{QS}}  >$	7.3 G	6.7 G	6.5 G	5.8 G

CR 2147 MDI-A Data	45s	315s	1215s	3600s
$\Phi$	$4.6 * 10^8$ G	$5.1 * 10^8$ G	$5.3 * 10^8$ G	$5.3 * 10^8$ G
$\Phi_{\text{AR}}$	$1.8 * 10^8$ G	$2.2 * 10^8$ G	$2.3 * 10^8$ G	$2.6 * 10^8$ G
$\Phi_{\text{QS}}$	$2.9 * 10^8$ G	$2.9 * 10^8$ G	$2.9 * 10^8$ G	$2.7 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	1.6	1.3	1.2	1.1
AR area	3.88%	8.10%	10.03%	13.53%
$<  \Phi_{\text{QS}}  >$	12.0 G	9.2 G	8.4 G	7.8 G

### 7.3 KP-A Data

Table 13: Important results for, Top: Cycle 2097; Bottom: Cycle 2147

CR 2097 KP-A Data	45s	315s	1215s	3600s
$\Phi$	$1.2 * 10^8$ G	$1.4 * 10^8$ G	$1.5 * 10^8$ G	$1.5 * 10^8$ G
$\Phi_{\text{AR}}$	$1.4 * 10^7$ G	$1.7 * 10^7$ G	$1.9 * 10^7$ G	$2.3 * 10^7$ G
$\Phi_{\text{QS}}$	$1.1 * 10^8$ G	$1.2 * 10^8$ G	$1.3 * 10^8$ G	$1.3 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	7.8	7.3	6.8	5.4
AR area	0.60%	1.20%	1.74%	288%
$<  \Phi_{\text{QS}}  >$	2.7 G	3.1 G	3.2 G	3.2 G

CR 2147 KP-A Data	45s	315s	1215s	3600s
$\Phi$	$3.9 * 10^8$ G	$4.1 * 10^8$ G	$4.2 * 10^8$ G	$4.2 * 10^8$ G
$\Phi_{\text{AR}}$	$1.7 * 10^8$ G	$2.1 * 10^8$ G	$2.2 * 10^8$ G	$2.4 * 10^8$ G
$\Phi_{\text{QS}}$	$2.2 * 10^8$ G	$2.0 * 10^8$ G	$2.0 * 10^8$ G	$1.8 * 10^8$ G
$\Phi_{\text{QS}}/\Phi_{\text{AR}}$	1.3	1.0	0.9	0.8
AR area	4.59%	8.98%	11.04%	14.55%
$<  \Phi_{\text{QS}}  >$	5.8 G	5.3 G	5.3 G	5.2 G

## 7.4 HMI/KP - Total Unsigned Flux Loss

	HMI/KP-A	45s	315s	1215s	3600s
<b>CR 2097</b>					
Total	2.10	1.92	1.90	1.80	
AR	0.94	1.04	1.03	1.07	
QS	2.25	2.05	2.03	1.94	
<b>CR 2147</b>					
Total	1.39	1.34	1.33	1.29	
AR	1.06	1.07	1.07	1.07	
QS	1.64	1.61	1.63	1.59	

Table 14: Ratio of HMI magnetic flux  $\Phi_{\text{HMI}}$  to KP-A magnetic flux  $\Phi_{\text{KP-A}}$  for each region separately (Total,AR,QS).

## 7.5 Covariance

### 7.5.1 Vertical Covariance

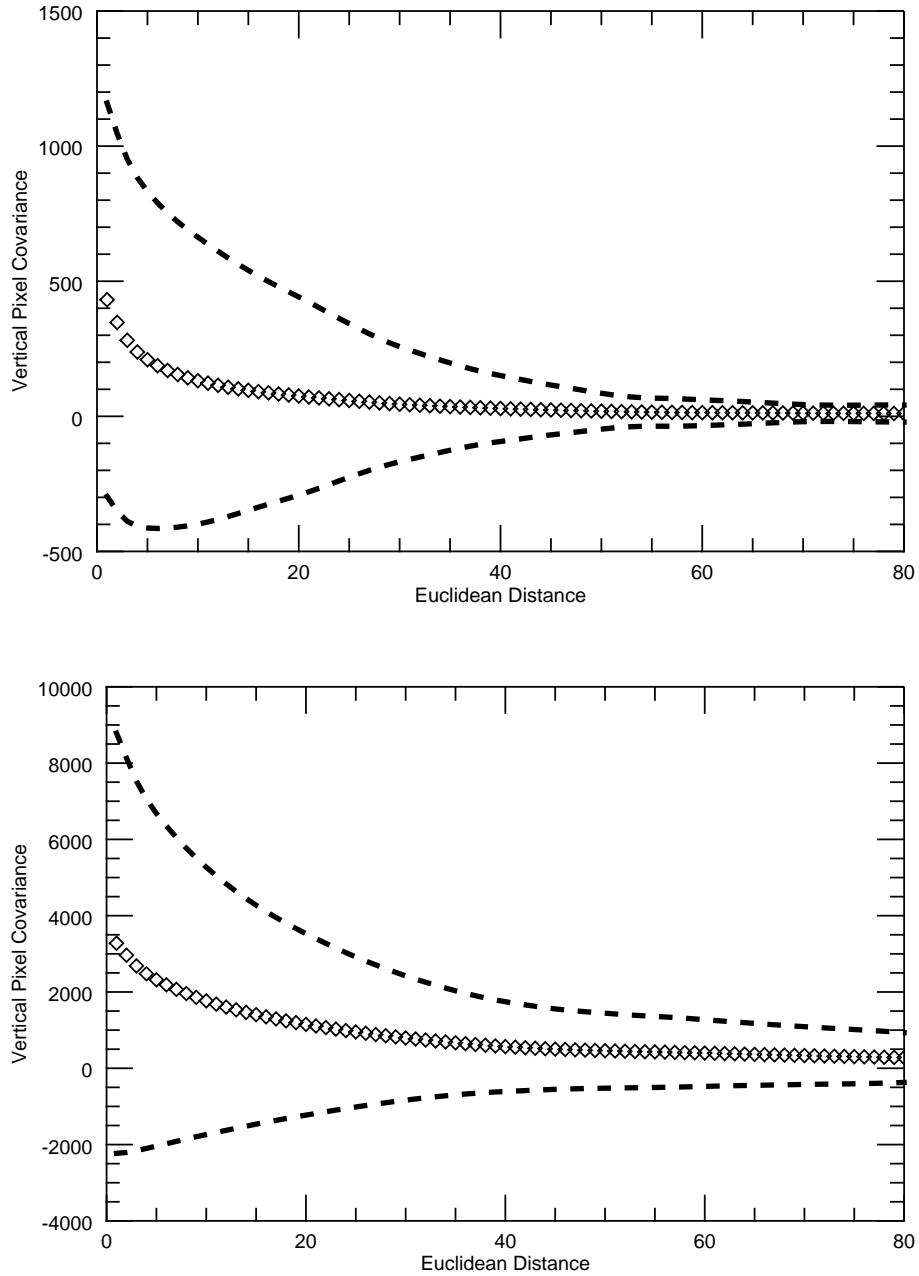


Figure 25: Pixel-by-pixel covariance as a function of euclidean distance between pixels. 315s Synoptic Chart; Top plot: Cycle 2097, Right plot: Cycle 2147. Diamonds: Vertical pixel-by-pixel covariance; Solid Line: Geometric Fit; Dotted line:  $1\sigma$  confidence interval.

### 7.5.2 Model Parameters

Table 15: Horizontal Covariance - Model Parameter Estimations

	CR 2097			CR 2147		
	a	b	c	a	b	c
45s	628	-0.91	-3.40	5577	-0.49	-368
	609	0.30	4.80	3311	0.23	295
315s	639	-0.90	-3.60	5543	-0.48	-381
	598	0.29	5.05	3256	0.23	308
1215s	639	-0.89	-3.71	5539	-0.48	-384
	595	0.29	5.13	3243	0.23	311
3600s	637	-0.88	-3.86	5634	-0.47	-398
	604	0.29	5.29	3287	0.23	322

Table 16: Vertical Covariance - Model Parameter Estimations

	CR 2097			CR 2147		
	a	b	c	a	b	c
45s	608	-0.87	-4.02	5409	-0.33	-983
	976	0.71	16.11	2233	0.30	1301
315s	631	-0.87	-3.90	5430	-0.33	-997
	969	0.68	16.20	2206	0.31	1321
1215s	628	-0.87	-3.85	5426	-0.32	-1006
	966	0.67	16.31	2188	0.31	1338
3600s	623	-0.86	-3.92	5533	-0.32	-1044
	604	0.68	16.87	3287	0.31	1402

## 7.6 Intersection Algorithm IDL Program Code

```
;+
; NAME:
;      INTERSECTIONEXTRACTION
;
; PURPOSE:
;      Active Regions defined by thresholding at a
; certain value are smaller than the real ones.
;      To find the real active regions one has to match
; patches of same polarity with the are defined by
; thresholding .
;      This mathematical operation is equivalent to the
; intersection of two sets .
;      The set of pixels  $S=\{s_i \mid \text{where } |y_{ij}| > thr\}$  is a
; superset of the set of patches with same polarity  $M$ :
;       $S \leq M$ . If  $S_i = M_i$  then  $M_i$  can be defines as an
; active region .
;      This function extracts the array indices of the pixel
; which are corresponding to active regions .
;      These indices are then used to create a logical mask
; for active region pixels
;
; CATEGORY:
;      Image Processing
;
; CALLING SEQUENCE:
;      result = intersectionextraction(chart , thr)
;
; INPUTS:
;      CHART: A synoptic chart magnetogram
;      THR: This function uses the extrema of the array as
; reference points for thresolding
;      THR=0.5 corresponds to threshold=0.5*max/min(chart) .
;      Other methods need changes in the code .
;
;
; OUTPUTS:
```

```

;      RESULT: A logical array with zeroes and ones, which
; can be used to find active regions.
;
; USES:
;      AR/QS Separation
;
; REQUIRED EXTERN IDL FUNCTIONS/PROCEDURES:
; UNDEFINE
;
; EXAMPLE:
;      data = randomn(seed, 10000, 10000 )
;      thr = 0.5
;      result = intersection(data, thr)
;
; MODIFICATION HISTORY:
;      Written by: Belal-Kevin Raza (belal-kevin.raza@stud.uni-
;-goettingen.de); 21.04.2017
;-

```

## FUNCTION INTERSECTIONEXTRACTION, chart , thr

```

sz=size(chart)

; begin with defining the thresholds.

thrp = thr*max(chart)
thrm=thr*min(chart)

; In the next step we will label all independent regions
; with the LABELREGION function.
chartthrplus=chart
chartthrminus=chart

```

```

chartthrplus [ where( chartthrplus lt thrp) ] =0
chartthrminus [ where( chartthrminus gt thrm) ] =0
regions_m=LABELREGION( chartthrminus ,/ALL_NEIGHBORS,/ULONG)
regions_p=LABELREGION( chartthrplus ,/ALL_NEIGHBORS,/ULONG)

; Count how many different regions have been found.
maxregp=max( regions_p )
maxregm=max( regions_m )

;Find position indices

mtable=lindgen( maxregm)+1
mlist=list()
for j=1l,maxregm do begin
  mlist.add, where( regions_m eq j )
endfor
m=orderedhash( mtable , mlist )

ptable=lindgen( maxregp)+1
plist=list()
for j=1l,maxregp do begin
  plist.add, where( regions_p eq j )
endfor
p=orderedhash( ptable , plist )

undefine , plist , mlist
undefine , chartthrminus , chartthrplus

;now do the same thing for the polarity patches
;polarity +-
;
plus=intarr(sz[1],sz[2])
plus[where(chart gt 0)] = 1

```

```

minus=intarr( sz [ 1 ] , sz [ 2 ] )
minus[ where( chart <= 0 ) ] = 1

regions_pp=LABEL_REGION( plus , /ALL_NEIGHBORS,/ULONG)
regions_mm=LABEL_REGION( minus , /ALL_NEIGHBORS,/ULONG)

maxpp=max( regions_pp )
maxmm=max( regions_mm )

mmlist=list()
mmtable=lindgen(maxmm)+1
for j=1,maxmm do begin

    mmlist.add, where( regions_mm == j )
endfor
mm=orderedhash( mmtable , mmlist )

pplist=list()
pptable=lindgen(maxpp)+1
for j=1,maxpp do begin
    pplist.add, where( regions_pp == j )
endfor
pp=orderedhash( pptable , pplist )

undefine , pplist , mmlist

;now do the set intersection.

```

```

maskadd=list()

for j=11,maxregp do begin
for i=11,maxpp do begin
if (SetIntersection(p[j],pp[i]) ne [-1]) then begin
maskadd.add,pp[i],/ extract
endif
endfor
endfor
undefine, pp,p

for j=11,maxregm do begin
for i=11,maxmm do begin
if (SetIntersection(m[j],mm[i]) ne [-1]) then begin
maskadd.add,mm[i],/ extract
endif
endfor
endfor
undefine, mm,m

```

*; build the mask*

```

mask=dblarr(sz[1],sz[2])
maskindices=maskadd.toarray(type='ulong')
mask[maskindices]=1
undefine, maskindices,maskadd

;DONE
return, mask
end

```

**Erklärung** nach §18(8) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbstständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 24. April 2017

(Belal-Kevin Raza)