

Observing the Sun: from start to finish.

PhD dissertation by

Pablo Santamarina Guerrero

Instituto de Astrofísica de Andalucía (IAA-CSIC)

Programa de Doctorado en Física y Matemáticas (FisyMat)
Universidad de Granada

A thesis submitted in fulfillment
of the requirements of the degree of
Doctor of Philosophy

July 4, 2024

PhD thesis supervised by

Dr. David Orozco Suárez

Dr. Julián Blanco Rodríguez



**UNIVERSIDAD
DE GRANADA**

ACKNOWLEDGEMENTS

Agradecimientos

RESUMEN

Resumen de la tesis

SUMMARY

Summary of the thesis

CONTENTS

1	TuMag’s design and calibration.	1
1.1	A brief introduction to spectropolarimeters.	1
1.1.1	Spectroscopy	2
1.1.2	Polarimetry	3
1.1.3	Imaging	5
1.1.4	What do spectropolarimeters tell us about the Sun?	5
1.2	The Tunable magnetograph: TuMag	6
1.3	Calibration of TuMag	7
2	Operation and data reduction.	9
2.1	TuMag’s Pipeline	9
2.1.1	Darks and flat fields	9
2.1.2	Blueshift	9
2.1.3	Demodulation and dual beam	9
2.1.4	Cross Talk	9
3	Challenges in data reduction. Etalon Cavity Map.	11
3.1	One device, two configurations.	11
3.1.1	Collimated configuration	12
3.1.2	Telecentric configuration	13
3.1.2.1	Telecentric imperfect configuration	15
3.2	Sunspot observation simulation	16
3.2.1	Simulated data.	16
3.2.2	Observations simulation.	16
3.3	Fitting algorithm	17
3.3.1	Initial approximations	17
3.3.2	Simulated observations	18
3.4	Fitting algorithm	19
3.4.1	Test scenarios and results	23
3.4.1.1	Impact of the noise level	24
3.4.1.2	Impact of the object approximation	25
3.4.2	The crossover case	27

4	Scientific exploitation.	31
4.1	Persistent Homology in Solar Magnetograms.	31
4.1.1	Persistent Homology	31
4.1.2	Persistent Images	34
4.1.3	Data	36
4.1.3.1	Quiet Sun observations	36
4.1.3.2	Active regions observations	36
4.1.4	Analysis and results	37
4.1.4.1	Quiet Sun results	40
4.1.4.2	Active region results	42
4.1.4.3	‘Interacting’ Diagram	45
4.1.5	Conclusions	47
5	Summary and conclusions	49
A	Profile derivatvies	51

CHAPTER 1

TUMAG'S DESIGN AND CALIBRATION.

In this chapter we take the first steps of the journey of developing an instrument to observe the Sun. We will define...

The SUNRISE III mission aims to study and establish the relations and couplings between the phenomena occurring at different layers of the Sun's surface. With this purpose in mind, three different post-focal instruments were included in the design, each of them responsible of observing at different regions of the spectrum. The SUNRISE UV Spectropolarimeter and Imager (SUSI, **REFERENCIA**), which will observe the spectra between 309 nm and 417 nm; The Sunrise Chromospheric Infrared spectroPolarimeter (SCIP, **REFERENCIA**), which will observe the near-infrared; and lastly, the Tunable Magnetograph (TuMag), which will observe three spectral lines in the visible, at 525.02 nm, 525.06 nm and 517 nm.

The design from scratch of an instrument such as this is very complex. There are many things that have to be meticulously designed and tested which span many fields of expertise, like optics, electronics, software, hardware, or thermal design. To avoid undue extension of this thesis, we will focus on the aspects of the design directly related to the **TO QUE**, that is, regarding the spectral, imaging and polarimetric capabilities of the instrument.

1.1 A brief introduction to spectropolarimeters.

Spectropolarimeters, as suggested by the name, are devices that measure the spectral and polarimetric properties of light, or in other words, that measure the polarization state of light as a function of wavelength. Their use is widely extended in astrophysics due to the huge amount of information about the light source we can infer from these properties.

In solar physics, it is common to encounter two distinct types of spectropolarimeters, distinguished by their approach to spectroscopy: slit-based spectrographs, such as SUSI and SCIP, and narrow-band tunable filtergraphs, like TuMag. The latter preserve spatial resolution by capturing two-dimensional images of the solar scene at the expense of sacrificing spectral resolution. Conversely, slit-based spectrographs provide excellent spectral resolution but have a limited spatial resolution.

Regardless of how spectroscopy is carried out, spectropolarimeters must be able to measure the polarization state of light. That is, they must be capable of determining the Stokes

parameters of the incident light. These four parameters, usually grouped in a pseudo-vector: $[I, Q, U, V]$, were defined by Stokes in Stokes (1851) as a mathematical formalism to completely define the polarization state of light. The first parameter, I , represents the total intensity; Q and U provide information about the intensity of linearly-polarized light, at 0° and 90° , respectively; and lastly, V , accounts for the intensity of circularly polarized light.

Excellent polarimetric sensitivity and spectral resolution are wasted if the optical capabilities of the instrument are not up to par. The design of these instruments must achieve diffraction-limited imaging, with a signal-to-noise ratio ensuring a polarimetric sensitivity of 1000 (typically), and the best spatial resolution the telescope allows, all without sacrificing spectral resolution and accomplishing this in the shortest possible time.

When designing the instrument, one must balance these three properties: spectral, optical, and polarimetric capabilities, trying to improve the performance in all of them without sacrificing too much. In the following sections, we will delve into each of these aspects in more detail.

1.1.1 Spectroscopy

Narrow-band tunable spectrographs play a significant role in this thesis. They will be extensively discussed in this chapter, particularly in relation to the design and calibration of TuMag, and again in Chapters 2 and 3 when addressing TuMag's pipeline and the correction of data produced by these instruments. Therefore, for the sake of simplicity, we will focus exclusively on this type of spectrographs from this point onward.

CAMBIAR ESTO.

Fabry-Pérot Interferometers (FPIs), also known as etalons (used interchangeably), represent one of the most prevalent forms of narrow-band tunable spectrographs. Composed by a resonant optical cavity formed by two distinct optical media, these devices allow only the passage of light with wavelengths corresponding to constructive interference within the cavity.

The transmission profile of an etalon, being produced by an interference phenomenon, is characterized by a series of narrow and periodic transmission peaks. The wavelengths at which this resonance peaks are located, their width, and their separation are determined solely by the physical properties of the etalon. In fact, it is not difficult to demonstrate (Bailén et al., 2019) that a resonant cavity produces a periodic transmission profile, with maxima occurring at a wavelength λ such that:

REVISAR -> VÁLIDO PARA TELECENTRIC??

$$\lambda = \frac{2nd \cos \theta}{m}, \quad (1.1)$$

where n is the refractive index of the medium inside the cavity, d is the distance between the mirrors, θ is the angle of incidence of the incoming light ray and m is the interferential order ($m \in \mathbb{Z}$).

With Eq. (1.1) in mind, it is clear that an etalon allows for tuning the wavelengths of the transmission peaks by either changing the distance between the mirrors or by altering

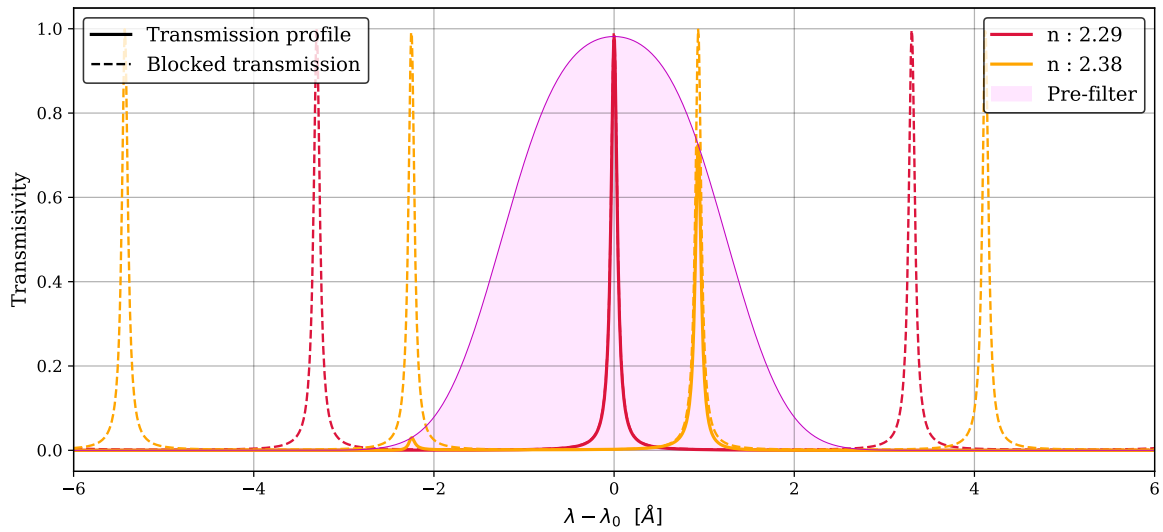


Figure 1.1 Transmission profiles of the same etalon with varying refractive indices (n). The dashed lines represent the original transmission profile, while the solid lines indicate the portion of the transmission profile that passes through the order-sorting pre-filter (shaded purple area).

the refractive index. Although changing the angle of incidence also results in a wavelength shift, it introduces other issues, such as ghost images or profile broadening in telecentric configurations, among other effects. Consequently, the angle is not used for wavelength tuning.

To tune to a single wavelength (or a very narrow band around it), it is necessary to isolate one transmission peak (main order). This is typically achieved by using a pre-filter that only allows light with wavelengths near the desired measurement region to pass through. This ensures that no light reaches the etalon that could pass through it due to interference orders other than the main one (secondary orders).

Figure 1.1 shows a simulation of the spectral behavior of this optical setup. The order-sorting pre-filter is shown with a shaded purple area and the unaltered transmission profile of the etalon is shown in dashed lines for different values of the refractive index. In solid lines, the resulting transmission profile is shown, that is, the transmission allowed through both the pre-filter and etalon at the same time.

1.1.2 Polarimetry

As previously noted, determining the polarization state of light requires the determination of the components of the Stokes vector. However, these parameters cannot be measured directly since we only know how to measure intensities. Since they are Thus, measuring the polarization of light always involves multiple measurements at once. Specifically, a number equal to the number of elements to be determined: four for the complete Stokes vector, or

two, if only the circular polarization and total intensity are to be measured. This is the root of the difficulties in measuring polarization, as the need for multiple measurements makes them much more susceptible to spurious effects compared to individual measurements. Cambiar que es un jaleo.

Mathematically, the effect on polarization of a linear and finite system can be treated as a combination of linear transformations on the Stokes vector and, therefore, can be represented by a matrix in \mathbb{R}^4 , known as the *Mueller Matrix*. Let \mathbf{M} be the matrix that describes these transformations, then the polarization state that reaches the detector follows:

$$\mathbf{I}_{out} = \mathbf{M}\mathbf{I}_{in}, \quad (1.2)$$

where \mathbf{I}_{in} and \mathbf{I}_{out} are the Stokes vectors of the light that reaches the instrument, and the detector, respectively. However, since we only know how to measure intensities, the actual quantity measured by our CCD is:

$$I_{obs} = m_{00}I_{in} + m_{01}Q_{in} + m_{02}U_{in} + m_{03}V_{in}, \quad (1.3)$$

where m_{0i} is the i -th element of the first row of the Mueller Matrix. This means that the intensity we measure is a linear combination of the different polarization states of the incoming light. To determine the values of the individual parameters I_{in} , Q_{in} , U_{in} , and V_{in} , further independent measurements are necessary, which can be achieved by modifying the Mueller matrix. In particular, it is easy to see that four independent measurements are required in order to construct a system of equations that allows us to determine the full Stokes vector. This process is known as modulation, and the four independent measurements are referred to as modulations.

If we denote each of the modulations by I_j with $j \in \{1, 2, 3, 4\}$, we can construct the following system of equations:

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \underbrace{\begin{pmatrix} m_{01}^1 & m_{02}^1 & m_{03}^1 & m_{04}^1 \\ m_{01}^2 & m_{02}^2 & m_{03}^2 & m_{04}^2 \\ m_{01}^3 & m_{02}^3 & m_{03}^3 & m_{04}^3 \\ m_{01}^4 & m_{02}^4 & m_{03}^4 & m_{04}^4 \end{pmatrix}}_{\mathbf{O}} \begin{pmatrix} I_{in} \\ U_{in} \\ Q_{in} \\ V_{in} \end{pmatrix} \quad (1.4)$$

where the superindex in m_{0i}^j denotes the values of the Mueller Matrix for each modulation. Through straightforward algebra, it is easy to see that the stokes vector of the incoming light can be determined by $\mathbf{I}_{in} = \mathbf{D}\mathbf{I}_{obs}$, where \mathbf{D} is the demodulation matrix, the inverse of the modulation matrix, \mathbf{O} , and \mathbf{I}_{obs} is the vector containing the 4 measured modulations. Accurately determining \mathbf{O} during the instrument calibration process is crucial, as the determination of the Stokes components depends entirely upon it.

1.1.3 Imaging

The high-resolution imaging that etalon-based instruments are capable of is one of the pivotal reasons for their extended use. The ability to capture a two-dimensional scene of the solar surface makes them ideal for studying solar plasma structures, which require resolutions close to 100 km on the solar surface. However, it is essential to achieve these resolutions while maintaining a sufficiently high signal-to-noise ratio to ensure the required polarimetric sensitivity.

Spectropolarimeters ultimately combine measurements in polarization, spectral, and spatial (image) domains. Consequently, the final observed intensity depends on all three properties simultaneously. By integrating the spectral behavior of the etalon and pre-filter with the polarimetric measurements, and taking into account the spatial dependence of these measurements, the observed intensity for a modulation j at any point of the focal plane η, ξ when the etalon is tuned at a wavelength λ_s is determined by:

$$I_j(\xi, \eta; \lambda_s) = g(\xi, \eta) \int_0^\infty T(\lambda) \iint O_j(\xi_0, \eta_0; \lambda) S(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s) d\xi_0 d\eta_0 d\lambda, \quad (1.5)$$

where $T(\lambda)$ accounts for the presence of the order-sorting pre-filter, $S(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s)$ accounts for the imaging response of the instrument when tuned at the wavelength λ_s , $g(\xi, \eta)$ represents a spatial gain factor that accounts for any wavelength independent pixel-to-pixel intensity fluctuations occurring in the focal plane, and $O_j(\xi_0, \eta_0; \lambda)$ is the intensity distribution of the incoming light for a modulation j and is given by:

$$O_j(\xi_0, \eta_0; \lambda) = m_{00}^j I_{in}(\xi_0, \eta_0; \lambda) + m_{01}^j Q_{in}(\xi_0, \eta_0; \lambda) + m_{02}^j U_{in}(\xi_0, \eta_0; \lambda) + m_{03}^j V_{in}(\xi_0, \eta_0; \lambda) \quad (1.6)$$

Determining the imaging response of the instrument can be quite complex, as it is influenced not only by their physical characteristics but also by their optical configuration, whether collimated or telecentric. In Chapter 2, we provide a detailed overview of the properties of each configuration, their differences, and the challenges involved in using these devices for data correction.

ADD noise Discussion?

1.1.4 What do spectropolarimeters tell us about the Sun?

Spectropolarimeters are often referred to as magnetographs (*e.g.*, TuMag), suggesting they measure magnetic fields directly. However, this is not entirely accurate. In astrophysics, the physical properties of the light source are inferred by correlating them with the observed properties of the light, rather than measuring them directly. By evaluating the polarization of sunlight at different wavelengths, spectropolarimeters enable us to infer the magnetic field and estimate plasma velocities on the solar surface.

The simplest calculation we can carry out that provides us with physical quantities of the Sun is that of the line-of-sight (LOS) velocities. Given the spectral shift of a specific

absorption or emission spectral line, $\Delta\lambda$, with respect to its rest position, λ_0 , the LOS velocities can be computed with the Doppler formula:

$$v_{LOS} = \frac{\Delta\lambda}{\lambda_0} c, \quad (1.7)$$

where c stands for the speed of light in vacuum.

The polarization properties of light come into play when determining the magnetic fields. Due to Zeeman and Hanle effects, the polarity and spectroscopy of spectral lines can be altered when formed in the presence of magnetic fields. Due to the Zeeman effect, the spectral lines widen or split into different polarized components when a strong magnetic field is present (del Toro Iniesta, 2003), such as in the surroundings of sunspots and active regions. In the other hand, the Hanle effect is sensitive to weaker fields, and can be used to study the magnetic structure of solar prominences or turbulent fields in the solar photosphere (Bianda et al., 1998).

One simple strategy to employ polarization and spectral data to derive the magnetic fields is through the center-of-gravity method. According to Uitenbroek (2003), the LOS strength of the magnetic field can be obtained through:

$$B_{LOS} = \frac{\lambda_+ - \lambda_-}{2} \frac{4\pi mc}{e g_L \lambda_0^2}, \quad (1.8)$$

where m and e are the electron mass and charge respectively, g_L stands for the Landé factor and λ_+ and λ_- are the centroids of the right and left circularly polarized line components, respectively, and are computed by:

$$\lambda_{\pm} = \frac{\int \lambda [I_{cont} - (I \pm V)] d\lambda}{\int [I_{cont} - (I \pm V)] d\lambda}, \quad (1.9)$$

where the subindex "cont" stands for the wavelength at the continuum. The vector magnetic field (*i.e.*, strength, azimuth and inclination), and not only the LOS strength can also be derived. However, the derivation of these quantities has to be achieved through inversions of the radiative transfer equation (RTE). The applicability of the different methods to carry out this inversion is an extensive topic as there may be some assumptions that can be applied in some cases but not in others, such as the weak-field or Milne-Eddington approximations, among others. For an extended discussion of this topic, we refer the interested reader to del Toro Iniesta & Ruiz Cobo (2016).

1.2 The Tunable magnetograph: TuMag

Among many examples of spectropolarimeters, one that has special relevance for this thesis is TuMag. The Tunable magnetograph (TuMag) is a wavelength-tunable spectropolarimeter that observes in the photosphere and the low chromosphere.

TuMag is a filtergraph that employs a solid LiNbO_3 -based FPI to scan the spectral lines. This kind of etalons are able to change the refractive index of the material inside the cavity by applying a voltage difference between the two mirrors. This design advantageously eliminates moving parts from the tuning process, though it requires high voltages for operation.

In order to be able to carry out observations of both the photosphere and low chromosphere, TuMag must be able to switch between different spectral lines. For this purpose, TuMag is equipped with three different pre-filters centered at the wavelengths of three different spectral lines: The FeI line at 5250.2 \AA , the FeI line at 5250.6 \AA , and the Mg line at 5172.7 \AA .

1.3 Calibration of TuMag

CHAPTER 2

OPERATION AND DATA REDUCTION.

2.1 TuMag's Pipeline

2.1.1 Darks and flat fields

2.1.2 Blueshift

2.1.3 Demodulation and dual beam

2.1.4 Cross Talk

CHAPTER 3

CHALLENGES IN DATA REDUCTION. ETALON CAVITY MAP.

INTRO. <https://link.springer.com/article/10.1007/s10509-023-04212-3>

3.1 One device, two configurations.

We established in section 1.1.3 that the observed intensity distribution at the coordinates ξ, η of the focal plane of any etalon-based instrument tuned to a wavelength λ_s obeys the following expression (Bailén et al., 2019):

$$I(\xi, \eta; \lambda_s) = g(\xi, \eta) \int_0^\infty T(\lambda) \iint O(\xi_0, \eta_0; \lambda) S(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s) d\xi_0 d\eta_0 d\lambda, \quad (3.1)$$

where $T(\lambda)$ accounts for the presence of an order-sorting pre-filter, $O(\xi_0, \eta_0; \lambda)$ represents the brightness distribution of the observed object at the point (ξ_0, η_0) , $S(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s)$ accounts for the imaging response of the instrument when tuned at the wavelength λ_s , and $g(\xi, \eta)$ represents a spatial gain factor that accounts for wavelength independent pixel-to-pixel intensity fluctuations occurring in the focal plane due to differences in the detectors' sensitivity.

The imaging response of the instrument coincides with the PSF of the instrument when the optical response is invariant against translations. In such a case, we can substitute the last two integrals by the convolution operator, but this is not strictly true in etalon-based instruments, where the response varies pixel to pixel either because of etalon irregularities or because of variations in the illumination across its clear aperture.

Deriving S typically requires determining the electric field of the polychromatic wave in the image plane. According to Bailén et al. (2019), this field can be calculated by summing all the electric fields ($E^{(t)}$) across the pupil, such that the electric field at any point (ξ, η) of the image plane follows the expression:

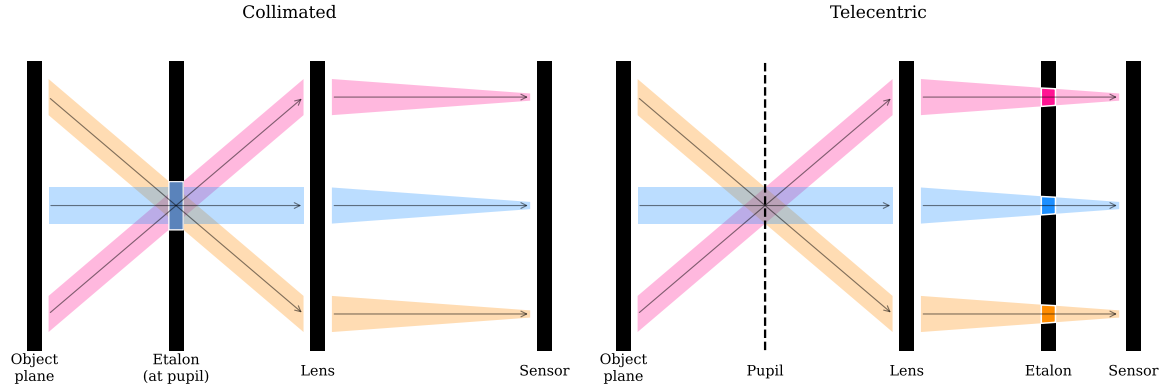


Figure 3.1 Schematic representation of the two optical setups of an FPI, collimated (left) and telecentric (right). The different colors represent distinct light rays originating from various points on the object plane. The white boxes in the etalon highlight the sections that are traversed by the light rays.

$$E_{im}^{(t)}(\xi, \eta) = \frac{1}{\pi R_{pup}^2} \int \int_{pupil} E^{(t)}(x, y) e^{-ik(\xi x/f + \eta y/f)} dx dy, \quad (3.2)$$

where (x, y) are the coordinates in the pupil, f stands for the focal length, and R_{pup} is the radius of the pupil. It can also be shown that the vector electric fields transmitted by the etalon can be expressed as:

$$E^{(t)}(\xi, \eta) = \frac{\sqrt{\tau}}{1 - R} \frac{e^{i\delta/2} - R e^{-i\delta/2}}{1 + F \sin^2(\delta/2)} E^{(i)}, \quad (3.3)$$

where τ is the transmission factor for normal incidence, R stands for the reflectivity of the etalon mirrors, F is a factor defined as $F \equiv 4R(1 - R)^{-2}$, $E^{(i)}$ are the incident electric fields, and δ stands for the phase differences between the transmitted and incident rays.

The transmission profile of the etalon (S) is defined as the average ratio between the transmitted and incident intensities, calculated from the complex conjugate of the corresponding electric fields. The shape of δ varies depending on the illumination setup of the etalon, whether collimated or telecentric. Consequently, each configuration has unique solutions and characteristics and is differently affected by inhomogeneities (defects) in the etalon's properties.

3.1.1 Collimated configuration

Collimated mounts are characterized by having the etalon located at the pupil plane and therefore receive a collimated beam from each point of the observed object. As illustrated in the schematic on the left side of fig. 3.1, light from any point on the object will fall on the same area of the etalon. Consequently, any local defects on the etalon crystals or on the plates' parallelism is averaged all over the clear aperture, thus making the optical quality

constant along the FoV. However, the angle of incidence of the light beam varies along the FoV, thus shifting the spectral transmission profile.

Analytical solutions of equation (3.2) are imposible to obtain if δ has a dependence on the pupil coordinates, as is the case of collimated etalons with a presence of local defects. In that case, the transmsion profile of the etalon has to be evaluated numerically.

However, in order to study the spectral behavior of the transmission profile, we can, as a first-order approximation, disregard the spatial PSF and focus solely on the spectral transmission profile (ψ). Under this assumption, the phase difference between the incident and transmitted rays of an ideal collimated etalon can be expressed as:

$$\delta(\xi, \eta, \lambda) = \frac{4\pi}{\lambda} n d \cos(\theta(\xi, \eta)), \quad (3.4)$$

where n stands for the refractive indx of the etalon cavity, d is the distance between the mirrors and θ is the angle of incidence. In such a case, it can be shown that the spectral transmission profile follows the expression:

$$\psi(\xi, \eta, \lambda) = \frac{\tau}{1 + F \sin^2(\delta(\xi, \eta, \lambda)/2)}. \quad (3.5)$$

The spectral behaviour of the transmission profile, such as the spectral position of the resonance peaks and the distance between them (the free spectral range), is encoded in the parameter δ , which is a function of the refractive index of the etalon cavity, the distance between mirrors, and the angle of the incident beam. The reflectivity R of the mirrors determines the width of the resonance peaks through the parameter F , $F \equiv 4R(1 - R)^{-2}$.

Local defects in the collimated configuration are averaged out, which means that d and n respectively represent the mean values of the thickness and refractive index across the clear aperture of the FPI, and thus remain constant for every pixel. Yet, they produce a broadening of the transmission profile and worsen the optical quality of the instrument. However, the spatial dependence of ψ naturally arises from θ , which vcaries from pixel to pixel.

Assessing the spatial PSF of the FPI is more challenging, as it can only be determined analytically for monochromatic light and in the absence of defects. We will not delve into the equations for this specific scenario as it lies beyond the scope of this thesis. However, interested readers are referred to the work of Bailén et al. (2019), where this topic is extensively discussed.

3.1.2 Telecentric configuration

In the telecentric configuration, the etalon is placed very close to an intermediate focal plane, while the pupil is focused at infinity. As shown in the sketch on the right side of fig. 3.1, in this setup, the etalon is illuminated by cones of rays that are parallel to each other, thus passing through different sections of the interferometer. Local inhomogeneities (defects or cavities) on the etalon produce differences in the transmission profile across the

FoV, which are directly mapped into the image plane. This means that the optical response and the transmission profile shift locally on the image sensor.

In telecentric configurations, δ always depends on the coordinates of the pupil, even in the absence of defects, since each point in the etalon sees a cone of rays coming from different parts of the pupil. Thus, as stated before, solutions to equation (3.2) are to be found numerically. Nonetheless, if we neglect the spatial PSF once again, we can derive an analytical expression for an ideal telecentric etalon, where all light cones impact perpendicularly.

After some messy algebra and clever variable changes, one can recast eq. (3.2) in terms of the radial coordinates of the pupil and analytically solve the equations. The resulting spectral transmission profile of an ideal telecentric etalon is given by (Bailén et al., 2021):

$$\Psi(\xi, \eta, \lambda) = \Re [E(a(\xi, \eta, \lambda), b(\xi, \eta, \lambda))]^2 + \Im [E(a(\xi, \eta, \lambda), b(\xi, \eta, \lambda))]^2, \quad (3.6)$$

with $E(a(\xi, \eta, \lambda), b(\xi, \eta))$ being:

$$E(a(\xi, \eta, \lambda), b(\xi, \eta)) = 2\sqrt{\tau} \left\{ \frac{1}{\alpha_1} [\arctan(\gamma_1) - \arctan(\gamma_2)] + i \frac{1+R}{1-R} \frac{1}{\alpha_2} \left[\ln \left(\frac{(1+\gamma_3)^2 + \gamma_4^2}{(1-\gamma_3)^2 + \gamma_4^2} \right) - \ln \left(\frac{(1+\gamma_3)^2 + \gamma_5^2}{(1-\gamma_3)^2 + \gamma_5^2} \right) \right] \right\}, \quad (3.7)$$

where, the auxiliary functions are defined as:

$$\begin{aligned} a(\xi, \eta, \lambda) &\equiv \frac{2\pi}{\lambda} n(\xi, \eta) d(\xi, \eta), \\ b(\xi, \eta) &\equiv \frac{1}{8n(\xi, \eta)^2 (f\#)^2}, \\ \alpha_1 &\equiv 2ab\sqrt{F}, \\ \alpha_2 &\equiv 2\alpha_1\sqrt{F+1}, \\ \gamma_1 &\equiv \sqrt{F} \sin a, \\ \gamma_2 &\equiv \sqrt{F} \sin(a[1-b]), \\ \gamma_3 &\equiv \sqrt{\frac{F}{F+1}}, \\ \gamma_4 &\equiv \frac{\tan(a/2[1-b])}{\sqrt{F+1}}, \\ \gamma_5 &\equiv \frac{\tan(a/2)}{\sqrt{F+1}}. \end{aligned} \quad (3.8)$$

The parameter a has the same role as δ for the collimated case. However, the dependence on the image plane coordinates in this case is caused by potential variations in n and/or d , as each light beam traverses different sections of the etalon. These variations

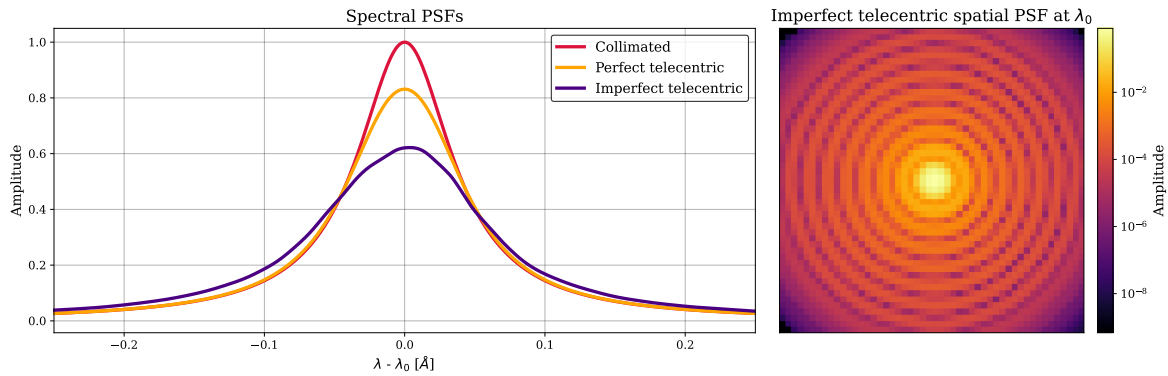


Figure 3.2 *Left*: central peak of the etalon's spectral transmission profile for the three different configurations. *Right*: Spatial PSF of the imperfect telecentric etalon at λ_0 . The parameters of the etalon are $R = 0.92$, $n = 2.29$, $d = 251 \mu\text{m}$, $f\# = 56$, $\theta = 0^\circ$ (collimated and perfect telecentric), and $\Theta = 0.3^\circ$ (imperfect telecentric).

constitute the "cavity map" of a telecentric etalon and must always be taken into account during data reduction processes.

The parameter b accounts for the contribution of the focal ratio, $f\#$, and has an impact on the spectral resolution and the apodization of the pupil as seen from the etalon (Beckers, 1998). Thus, the resolution is now affected by both F and $f\#$, through the parameters a and b .

3.1.2.1 Telecentric imperfect configuration

The equations shown in Sect. 3.1.2 are valid whenever the incident cone of rays is perpendicular to the etalon mirrors. We refer to this situation hereinafter as "perfect telecentrism". However, real instruments are likely to present deviations from such an ideal case. These deviations can be caused by an intentional tilt of the etalon to suppress ghost images on the detector (Scharmer, 2006), by an accidental tilted angle of incidence caused by deviations from the ideal paraxial propagation of rays within the instrument, or simply because of misalignment of the optical components. In the three cases, the incident cone of rays is no longer perpendicular to the etalon, and hence, we consider these scenarios to have imperfections in the telecentrism degree. One important consequence of the loss of telecentrism is an asymmetrization of the transmission profile that must be accounted for when modeling the instrument response.

The transmission profile in this case is influenced by the angle of incidence of the chief ray at each point of the clear aperture of the etalon, in addition to the parameters mentioned in the previous sections. Unfortunately, the equations for the transmission profile in these configurations are much more complicated than in the ideal case, and can no longer be analytically solved. We must revisit equation (3.2) and solve the integrals numerically.

Figure 3.2 shows on the left the transmission profile corresponding to the three different scenarios: collimated illumination of the etalon, perfect telecentrism, and imperfect

telecentricism. The etalon parameters have been selected to coincide with those of SO/PHI's etalon. In both the collimated and perfect telecentric configurations, a normal incidence ($\theta = 0$) scenario is shown, whereas in the imperfect telecentric case, we assumed an angle of incidence of the chief ray, Θ , of 0.3° . The parameter a has been adjusted slightly in order to tune the transmission profile at λ_0 .

Regarding the properties of each profile, note that the telecentric configurations achieve lower peak transmissions than the collimated case. In addition, the telecentric profiles are wider due to the different incidence angles across the illuminating cone of rays. Such a broadening increases with decreasing f-ratios. Lastly, non-normal incidence of the chief ray in the telecentric configuration further widens and shifts ($\sim 4 \text{ m\AA}$ for $\Theta = 0.3^\circ$) the profile, making it asymmetrical.

3.2 Sunspot observation simulation

INTRO sunspot. Challenges in data reduction, unknown effects etc.

3.2.1 Simulated data.

blabla

3.2.2 Observations simulation.

All the instruments built around the use of an etalon as a wavelength filtering element operate in a very similar way. They scan a spectral line by tuning the etalon (by changing the distance between mirrors and/or modifying the refractive index) to a desired number of wavelengths along the spectral line. At each spectral position, the solar scene is recorded. The measured intensity is approximately given by Eq. (3.1), with the etalon's transmission profile centered at the desired wavelength.

We have carried out two sets of simulations, one with a perfect telecentric configuration, where the incidence is normal, and one with an angle of incidence of the chief ray is 0.5° . For each of these sets, a simulated observation of the sunspot has been carried out employing 45 wavelengths for the scan of the line. A flat-field observation has also been computed with the same configuration than the one employed for the sunspot observation. For the flat-field observations we have generated an *ideal* flat-field where the average profile computed in quiet-sun was replicated at every pixel of the generated image.

Although, in practice, it is not often possible to fully characterize the pre-filter, we assumed it has a rectangular shape centered at the wavelength of the observed spectral line (λ_0) and a width of $2\Delta\lambda$ such that only one order of the etalon passes through. With this consideration, equation (3.11) can be written as follows:

$$I(\xi, \eta; \lambda_s) = g(\xi, \eta) \int_{\lambda_0 - \Delta\lambda}^{\lambda_0 + \Delta\lambda} \iint O(\xi_0, \eta_0; \lambda) \mathcal{S}(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s) d\xi_0 d\eta_0 d\lambda. \quad (3.9)$$

For every tuned wavelength, the intensity in a given pixel is given by equation (3.9), where the observed object is the sunspot simulation or the generated *ideal* flat-field.

The imaging response of the FPI is computed by solving the integrals in equation (3.2). Given that we will account for the PSF in these simulations, numerical methods are necessary to evaluate the integrals. The computational burden is significant, as the integrals must be evaluated $N_{pixel}^2 N_\lambda N_{mods} N_{sets} > 10^8$ times¹ for both sunspot and flat field observations. This extensive computation results in excessively long simulation times. To address this issue, a neural network was developed to reduce the computational time from **X to X**. The neural network was trained with known solutions to the integrals for a fixed set of etalon parameters, ensuring that its outputs are accurate to within 0.01% for parameters within the training set.

In real-world scenarios, variations in pixel sensitivities (gain) and etalon imperfections (cavity map) are unavoidable. Including the effect of the gain into the simulation is straightforward as it only acts as a multiplicative factor that modifies the final intensity. In contrast, the effects of the cavity map are more cumbersome. Pixel-to-pixel variations in etalon defects shift the transmission profile of the Fabry-Perot Interferometer (FPI) and must be accounted for when computing the imaging response. Both the gain and cavity map introduced in the simulations have been selected to resemble the real-case scenario of the SO/PHI instrument. The gain map (Figure XXX a) utilized in the simulations is derived from a flat field of the High Resolution Telescope (HRT) instrument, while the cavity map employed (Figure XXX a) is sourced from the cavity map of PHI's etalon.

3.3 Fitting algorithm

3.3.1 Initial approximations

As an initial approximation, we will disregard the spatial PSF and assume that the spatial dependence can be represented by a Dirac delta function to simplify the equations. Therefore, if we assume that the image response of the FPI follows the expression:

$$S(\xi_0, \eta_0; \xi, \eta; \lambda - \lambda_s) = \delta(\xi_0 - \xi, \eta_0 - \eta) \Psi(\xi, \eta, \lambda - \lambda_s), \quad (3.10)$$

equation (3.1) simplifies into:

$$I(\xi, \eta; \lambda_s) = g(\xi, \eta) \int_0^\infty T(\lambda) O(\xi, \eta; \lambda) \Psi(\xi, \eta; \lambda - \lambda_s) d\lambda. \quad (3.11)$$

The explicit shape of Ψ is different depending on the optical configuration of the instrument, that is, collimated or telecentric.

¹ $N_{pixels} = 561, N_\lambda = 45, N_{mods} = 2, N_{sets} = 2$

3.3.2 Simulated observations

We carried out a series of simulations of a spectral line observation in different conditions. We used the Kitt Peak FTS-Spectral-Atlas as the reference (Brault & Neckel, 1987) and, specifically, the Fe I spectral line at 6173.3 Å. Each observation was composed of N_λ wavelengths, where the measured intensity was recorded. At every wavelength λ_s , the corresponding transmission profile of the etalon Ψ^{λ_s} was computed, and the "observed" intensity $I_{\text{obs},i}^{\lambda_s}$ corresponding to a specific spatial location (ξ, η) , represented hereinafter by the pixel i , was calculated using Eq. (??). Additionally, we took into account the presence of additive Gaussian noise. This noise does not necessarily respond to any parameter fluctuation within our analytical expressions or photon noise but comes from any unexpected variations that may not have been modeled in the theoretical scheme.

Additionally, we included the presence of defects arising from irregularities or inhomogeneities on either the cavity thickness d , the refractive index n , or from deviations of the angle of incidence θ . In order to simulate this, we introduced a relative perturbation Δa into the etalon equation that accounts for any local deviation of the value of a with respect to its nominal value. This parameter changes from pixel to pixel differently for the collimated and telecentric configurations. In the former, the profile shifts across the FoV only because of the different incidence angles of the light beam on the etalon. In the latter, local variations of n and/or d are mapped directly onto the detector. We also note that variations in the incidence angle must be considered as well when the degree of telecentricism varies along the detector. Analytically, the parameter a at each i -th pixel is given by $a'_i = a\Delta a_i$, where $a = (2\pi/\lambda)nd \cos \theta$ is constant along the FoV. Note that rewriting the equations for the transmission profiles (sections 3.1.1 and 3.1.2) using this definition of a is straightforward. In collimated configurations, the parameter δ (eq. (3.4)) is simply $2a$. In perfect telecentric configurations, θ is always 0, so the given transmission profile is already expressed in terms of a .

We let $n_i^{\lambda_s}$ be the noise contribution at the i -th pixel and wavelength λ_s . Thus, the observed intensity at that pixel when the etalon is tuned at λ_s , $I_{\text{obs},i}^{\lambda_s}$ is given by

$$I_{\text{obs},i}^{\lambda_s} = g_i \frac{\int_{\lambda_0-\Delta\lambda}^{\lambda_0+\Delta\lambda} O(\lambda) \Psi^{\lambda_s}(\lambda, \Delta a_i) d\lambda}{\int_{\lambda_0-\Delta\lambda}^{\lambda_0+\Delta\lambda} O(\lambda) \Psi^{\lambda_c}(\lambda, \Delta a_i) d\lambda} + n_i^{\lambda_s}, \quad (3.12)$$

with λ_c being the continuum wavelength. From a practical point of view, the integration limits are set in such a way that only a single resonance (or order) of the etalon is included with the limits, thus, acting akin to the sorting pre-filter commented on previously. We note that the denominator strictly corresponds to the intensity at the continuum of the line in the absence of the transmission profile or if the continuum wavelength is far enough from the spectral line. In any other case, the transmission should be taken into account as well to normalize the observations to the local continuum, which is necessary since we work with relative measurements. An example of a spectral line measurement is displayed in Fig. 3.3.

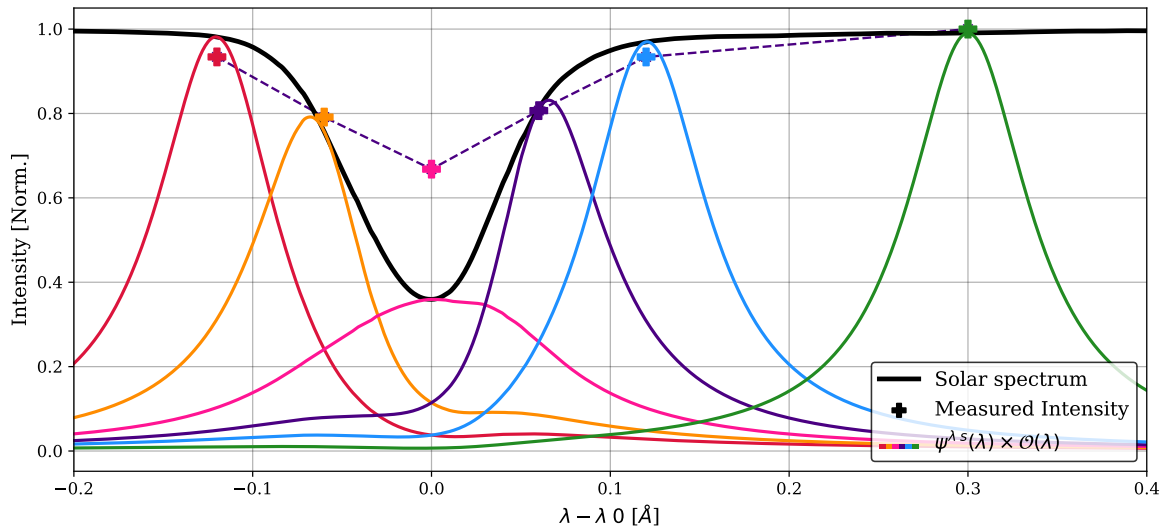


Figure 3.3 Simulated observation of the Fe I spectral line ($\lambda_0 = 6173.3\text{\AA}$) using a collimated mount and a total of $N_\lambda = 6$ wavelengths that have been equally distributed along the spectral line, with the exception of the continuum measurement (light blue), which is selected at 300 mÅ from the blue of the line core. The measured intensity is the result of computing the value given by Eq. 3.12 at each wavelength and with $g = 1$.

For both the collimated and telecentric configurations, we modeled etalon and gain imperfections over a $100 \times 100 \text{ px}^2$ image. Pixel-to-pixel variations in the sensor efficiency were modeled following a random spatial distribution, as shown in Fig. 3.4 (top panel). Additionally, we included a set of pixels with very low gain values, which represent a group of dead pixels or dust grains.

We modeled the etalon defects as changes in Δa in such a way that the maximum displacement reaches 3 pm. The spatial distribution of the values of Δa follows an increasing radial distribution, as shown in Fig. 3.4 (bottom panel). Such a spatial distribution coincides with the expected one in collimated etalons due to the change in the incidence angle across the FoV. Telecentric mounts do not exhibit a spatial distribution of their defects such as this, but using the same spatial distribution in the two cases allowed us to compare the performance of the method for both setups in a systematic way. Since Δa accounts for relative perturbations, it is by definition an adimensional parameter. However, to grant it a physical meaning, we express the values of Δa in Å, representing the associated shift of the transmission profile with respect to the original position determined by a .

3.4 Fitting algorithm

We have developed an algorithm able to extract the distribution of the etalon defects and the gain map from data taken by etalon-based instruments, which enables the correction of the

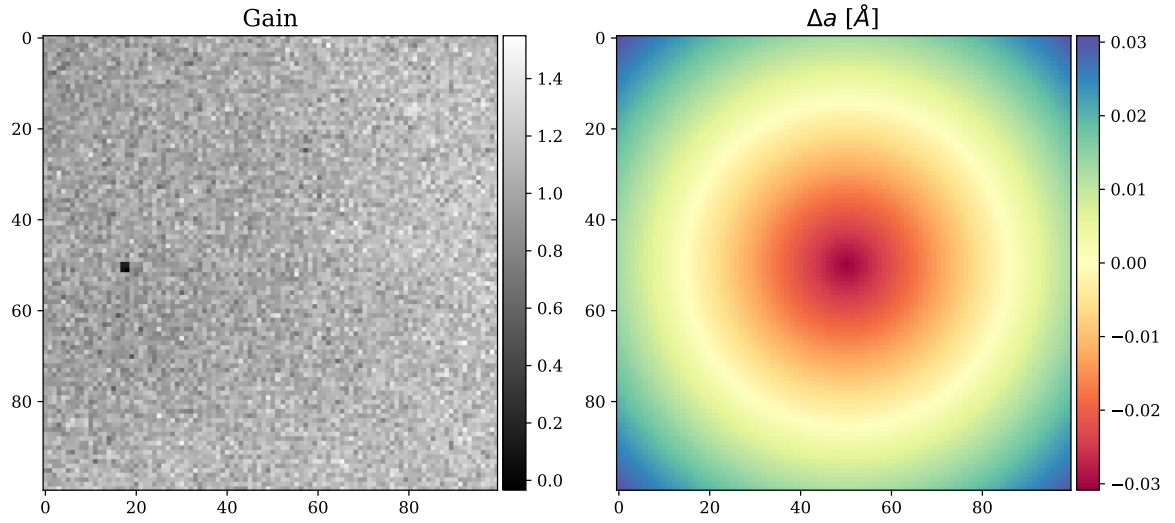


Figure 3.4 Input maps introduced when simulating the observations. The left panel represents the gain generated as white Gaussian noise, with values ranging from 0.8 to 1.2. A dust speck was introduced by creating a group of four pixels with low values of $g = 0.2$ for the gain. The right panel shows the spatial distribution of the defects in the etalon. The distribution follows a radial pattern starting from the center of the FoV. The defects vary from 0 % deviation to up to 5×10^{-4} %, which corresponds to a shift of 3 pm. Both possible directions for the deviations have been considered. The sign of the deviation is negative at the very center, which introduces a redshift, while it is positive at the corners, causing a shift of the profile into the blue.

two contributions separately. The algorithm works by minimizing a given merit function that depends on the gain and the etalon defects.

In particular, we have defined an error metric, ε^λ , at each tuned wavelength, computed by comparing the measured intensity with the theoretical prediction. If we let $I_{i,\text{obs}}^{\lambda_s}$ be the measured intensity at a given i pixel for an etalon tuned to the wavelength λ_s , the error metric at each wavelength is given by

$$\varepsilon^{\lambda_s}(\Delta a_i, g_i) = I_{i,\text{obs}}^{\lambda_s} - g_i \frac{\int_{\lambda_p}^{\lambda_q} O(\lambda) \Psi^{\lambda_s}(\lambda, \Delta a_i) d\lambda}{\int_{\lambda_p}^{\lambda_q} O(\lambda) \Psi^{\lambda_c}(\lambda, \Delta a_i) d\lambda}. \quad (3.13)$$

The merit function we employed is then the quadratic summation of the error metric over all tuned wavelengths:

$$f(\Delta a_i, g_i) = \sum_{s=0}^{N_\lambda} \left(\varepsilon^{\lambda_s}(\Delta a_i, g_i) \right)^2. \quad (3.14)$$

Both the camera gain and the defects of the etalon change from one pixel to another, which is why we address each pixel independently, but they remain constant at every wavelength. Hence, the transmission profile of the etalon varies at different points of the FoV; but at a given pixel, it is constant at all tuned wavelengths. Therefore, the algorithm is able to better obtain the etalon properties as we increase the number of wavelengths.

Figure 3.5 shows the derivatives of the error metric, Eq. (3.13), as a function of wavelength, that is, before computing the summation over s of the merit function, with respect to the gain, the reflectivity, and Δa . The curve corresponding to the Δa derivative is different from the others, whereas the derivatives of both the gain and the reflectivity exhibit similar shapes. Hence, variations in either the reflectivity or the gain introduce similar changes in the merit function, which can produce a trade-off between these two parameters, especially when the spectral line is sampled in only a few points. Given that discrepancies arising from errors in reflectivity are assimilated by gain maps, we did not take into account reflectivity errors when computing our simulations, as they have no impact on cavity map calculations.

A few key aspects arise when analyzing the merit function and its applicability on real data. The first point to bear in mind is that the shape of the object, $O(\lambda)$, is not known a priori. Therefore, we needed to provide a guess for it. The method works by assuming that differences between the prediction and the observation are caused exclusively by the etalon defects or the gain. If the object used during the fitting process differs considerably from the real one, the prediction and observation will have differences that will erroneously be identified as etalon defects or gain variations. This is the main source of errors for the method when applied to real data. Two approaches can be followed in order to address this issue. The first one consists of assuming the solar atlas profile as the object. This is a good approximation, provided the data to which the algorithm is applied to lack information about solar structure, either because they are observations of long integration times of the quiet sun or produced by averaging several quiet sun observations (flat fields). If this

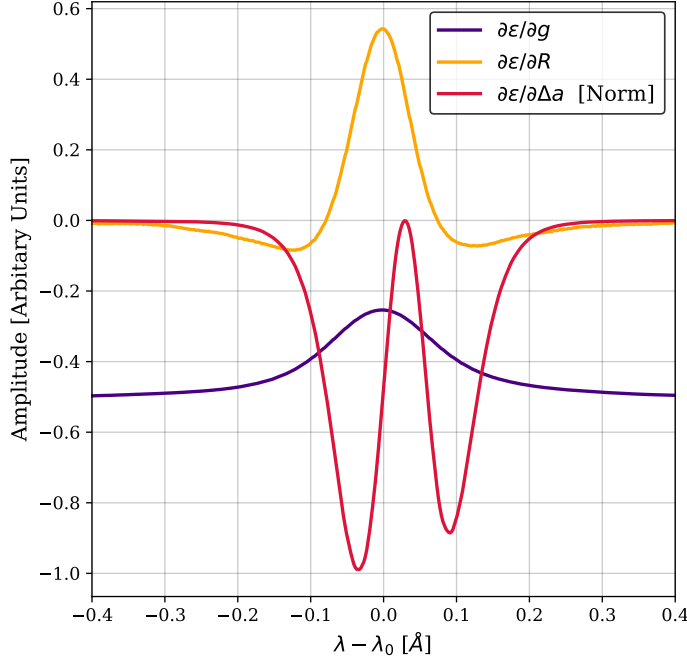


Figure 3.5 Derivatives of the error metric as a function of wavelength. The derivative with respect to Δa has been normalized in order to fit the three curves in the same plot.

condition cannot be met, this approach is not valid. The second approach involves deriving an approximated object from the data themselves by deconvolving the mean profile of the observation with the etalon's transmission profile. This approach can account for any difference the real object may have with the solar atlas and thus has a greater resemblance to the real object. Nevertheless, the process of deconvolving is prone to errors when the sampling is insufficient and can introduce additional noise into the problem. We have tested both approaches to compare their performances on different scenarios in order to assess when to use one or the other.

We employed Newton's method to minimize the merit function, Eq. (3.14), as it has been proven to quickly converge (in five iterations or fewer, usually). The method begins by assuming an initial guess for the gain g_j and Δa_j parameters. Then, provided the initial guess is sufficiently close to the solution and that the merit function is continuous and differentiable, the gain g_j and Δa_j encoded in the vector, \mathbf{x}_j , can be updated iteratively at each iteration, j , as

$$\mathbf{x}_{j+1} = \mathbf{x}_j - \mathcal{H}^{-1} \mathcal{J}^T f(\mathbf{x}_j), \quad (3.15)$$

where \mathcal{H} and \mathcal{J} are the Hessian and Jacobian matrices of the merit function f , respectively, calculated for $\mathbf{x}_j = [g_j, \Delta a_j]^T$, and T stands for the transpose. Hence, the transmission profile of the etalon and its derivatives have to be computed for every wavelength and every pixel at each iteration. This can be computationally costly, especially when using imperfect telecentric configurations, where numerical integrals are involved. All derivatives needed for the algorithm are calculated analytically, except when simulating imperfect telecentricism. A detailed formulation of these derivatives is provided in the appendix.

Regarding the object $O(\lambda)$, if we assume it is given by the solar atlas, no additional

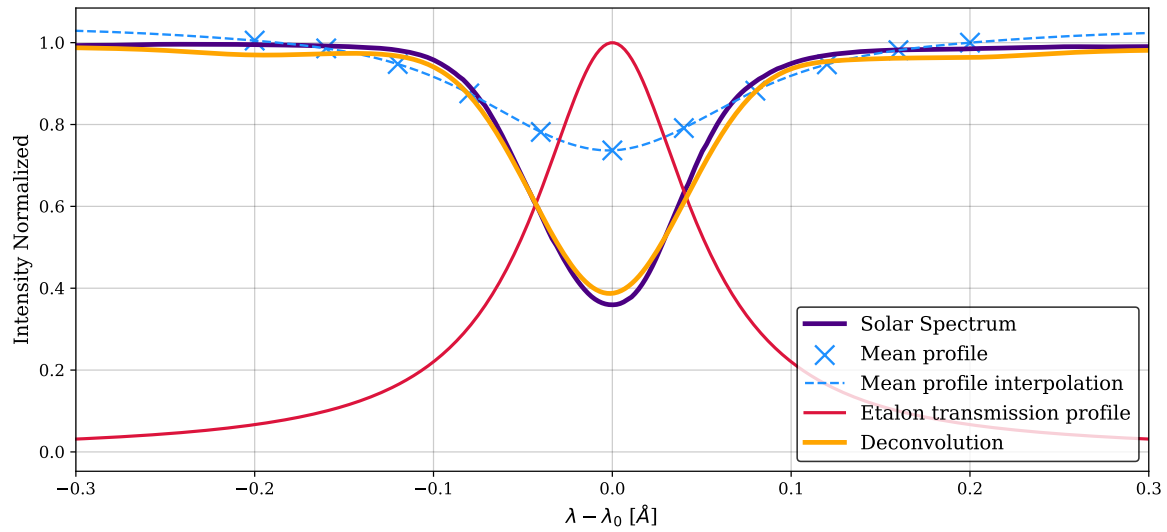


Figure 3.6 Deconvolution of the object with a measurement of the Fe I spectral line using $N_\lambda = 9$. All points of the FoV have been used to compute the average profiles (blue crosses). The deconvolution (orange) is the result of deconvolving the mean profile interpolation (dashed line) with the displayed etalon transmission profile (red).

computations are needed. However, when using the deconvolution approach, the object has to be calculated in each iteration. In this case, the algorithm works as follows: First, we compute the average profile across the whole FoV, and we force the continuum intensity to be the same on both sides of the spectral range to reduce the boundary effects of the deconvolution. This step is only necessary in case the spectral line is sampled in only a few positions, as is the case of the SO/PHI, IMAx, or TuMag instruments, where only a continuum point, either at the red or the blue side of the spectrum, is recorded. Both the object and transmission profile require a good spectral sampling to accurately compute the integrals of Eq. (3.13). Second, a cubic spline interpolation is applied to the generated average profile to artificially improve the spectral sampling, if necessary. Finally, the interpolated profile is then deconvolved by means of a Wiener filter with the etalon's transmission profile. The result of this deconvolution is the object, $O(\lambda)$, used in the minimization algorithm. The deconvolution of the object is done every time the etalon defects are updated in order to improve the resemblance of the deconvolved object to the real one. Figure 3.6 shows an example of this process in a simulated observation using nine scanned points and a collimated configuration. The deconvolution manages to reproduce the original signal, with only some minor differences in the line core and the beginning of the wings.

3.4.1 Test scenarios and results

The aim of the simulations carried out in this section was to characterize the role of the noise $\delta_i^{\lambda_s}$, the spectral sampling, the selection of the object $O(\lambda)$, and the accuracy of the

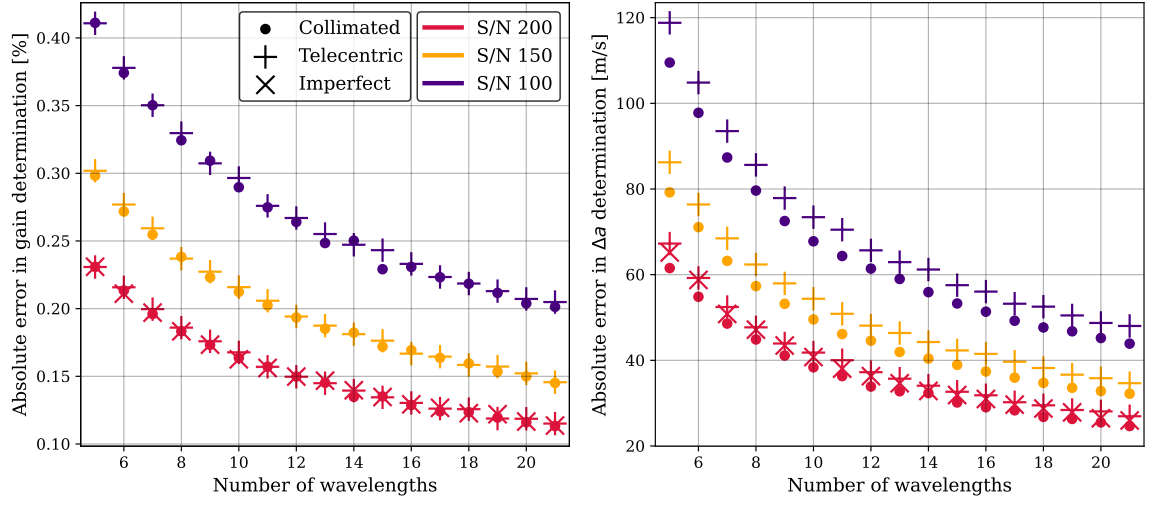


Figure 3.7 Absolute errors of the gain (left) and etalon defect (right) derivations averaged over all the FoV. The number of wavelengths corresponds to the parameter N_λ of wavelengths used to scan the profile.

method for both the collimated and telecentric configurations. All simulations were run for different choices of the number of scanned wavelengths, ranging from $N_\lambda = 5$ to $N_\lambda = 21$.

3.4.1.1 Impact of the noise level

We first assumed that the spectrum of the observed object is given by the solar atlas. This way, all errors in the derivation of the gain and etalon defects only come from the noise introduced into the measurement. We refer to this as the "ideal case." Since we were combining different measurements taken at different wavelengths, we considered a worst-case scenario and simulated three different signal-to-noise ratios: 100, 150, and 200.

We restricted imperfections in the telecentrism to arise only for one scenario, $S/N = 200$, since simulating imperfections requires a high computational effort due to the lack of a theoretical expression for both the transmission profile and its derivatives. We also assumed that the degree of telecentrism (0.3°) is known in this case.

Figure 3.7 shows the average absolute error in g (left panel) and in Δa (right panel) over the whole FoV as a function of the wavelength sampling, N_λ . The error in g is expressed as a percentage of its real value. Errors in Δa are given in meters per second since they are mostly responsible for shifting the profile. Errors in Δa can be translated into velocity errors by computing the doppler velocity (Eq. (1.7)) associated to the spectral shift of the transmission peak produce by the error in Δa .

All the scenarios exhibit a similar behavior as far as their dependence on the spectral sampling is concerned, namely, the absolute errors decrease monotonically when the wavelength sampling increases. The reason for this is simply that a larger number of wavelength samples increases the amount of available information that the algorithm can use, thus making the fitting for g and Δa more precise. These results highlight the importance

of properly sampling the targeted spectral line. A modest sampling of only $N_\lambda = 5$ can produce errors as large as 120 ms^{-1} in the worst-case scenario ($S/N = 100$).

The noise level also plays an important role in the accuracy of the results. Scenarios with a lower S/N always have larger errors, for a given N_λ , in both the gain and Δa computations. The difference in the performance of the algorithm due to the noise also changes with the spectral sampling; scenarios with a poor spectral sampling suffer from larger differences in the accuracy between the different S/N (50 ms^{-1} for $N_\lambda = 5$ between $S/N = 200$ and $S/N = 100$) than those with higher samplings (35 ms^{-1} for $N_\lambda = 21$).

The optical configuration of the etalon has a very small impact on the accuracy of the algorithm. Results for the three setups are very similar, particularly in the gain calculation, for which the results are almost identical for all configurations. Retrieval of Δa is slightly better for the collimated mount, though.

Figure 3.8 shows the spatial distribution of the errors in the retrieval of Δa for different choices of N_λ . There are no signs of a radial distribution in the maps shown in the figure, contrary to the actual distribution of the Δa parameter, as shown in Fig. ??, bottom panel. This means that the precision of the method is similar no matter the amplitude of the defects, that is, we achieve the same accuracy in the retrieval of defects associated with shifts of 3 pm ($\sim 1450 \text{ ms}^{-1}$, near the corners of our FoV), which correspond to cavity errors of around 1.5 nm or incidence angles of approximately 0.4 degrees, and in the retrieval of regions where no defect is present (radius of 20 pixels from the center of the FoV approximately). Instead of a radial distribution, the errors follow a Gaussian-like distribution (shown at the bottom panels in 3.8) similar to the one followed by the noise contribution.

The standard deviation of the errors for both the gain and Δa computations are also reduced with an increase in spectral sampling. The last row of Fig. 3.8 displays the error distributions in the calculation of Δa for the three optical configurations and different spectral samplings. These results illustrate how the three configurations yield practically identical results and how the distribution narrows as N_λ increases, thereby improving the results. Specifically, the standard deviation decreases from 50 ms^{-1} for $N_\lambda = 5$ to 20 ms^{-1} for $N_\lambda = 21$. In the case of the gain determination, the standard deviation ranges between 0.2% and 0.1% for the scenarios with the poorest and highest spectral sampling, respectively.

3.4.1.2 Impact of the object approximation

To infer the error of the algorithm when the object is unknown, we compared the performance of the ideal case, that is, when the object used to generate the observations is known, with the one achieved when deconvolving the object from the data. Only the collimated setup was simulated in order to focus exclusively on the errors introduced by the deconvolution. The data has been degraded by Gaussian noise with an $S/N = 200$ in both scenarios.

Figure 3.9 shows the results for the two approaches. Interestingly, the error in the gain for the deconvolution approach does not decrease with a larger number of wavelengths, unlike the ideal case. Nevertheless, the average error of the calculation is below 0.4% , with a dispersion (1σ) of $\pm 0.3 \%$. The deconvolution approach is prone to higher errors when

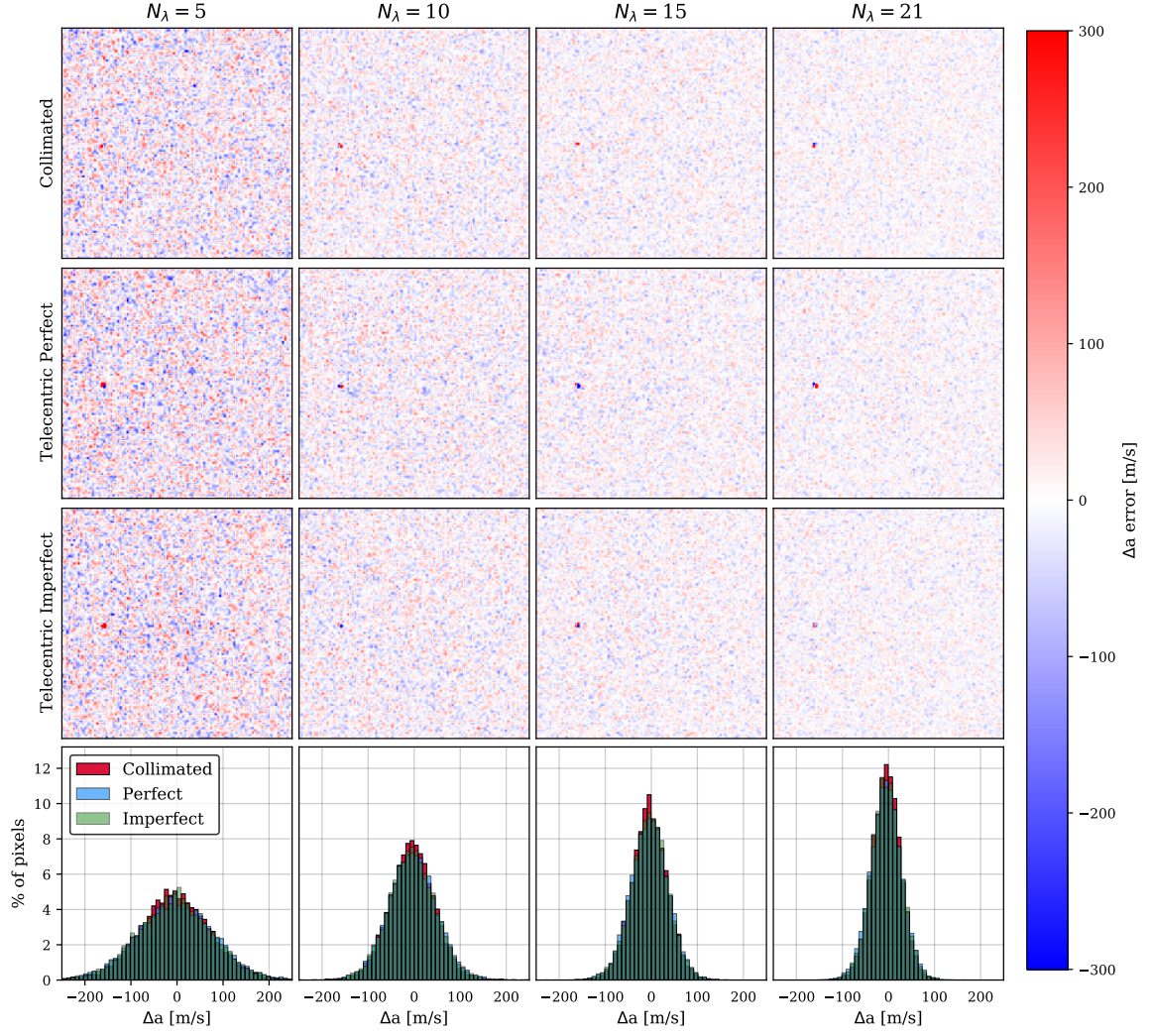


Figure 3.8 Distribution of the errors in the Δa computation for the three configurations (first three rows) and different spectral samplings (columns). In the bottom panels of each column, the error distribution for the corresponding spectral sampling is shown for the three configurations.

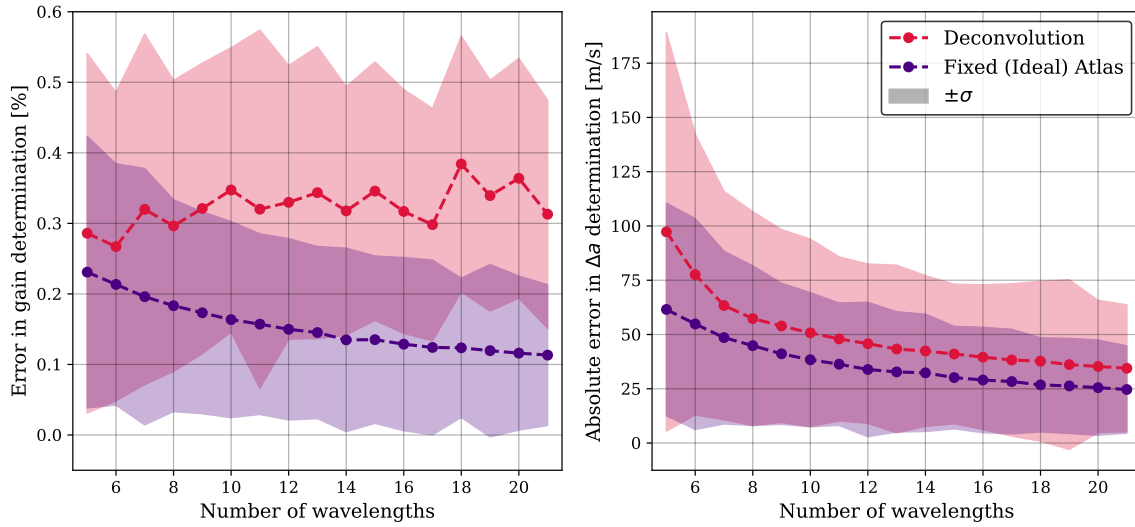


Figure 3.9 Errors in gain determination and etalon properties averaged over all the FoV with a signal-to-noise ratio of 200 and a collimated configuration.

deriving the gain due to the normalization of the profiles. The reason for this is two-fold. first, if the continuum is far enough from the spectral line, the normalization is strictly the integral over the transmission profile because the object is flat along the integration interval. However, this is not strictly true since the wings of the transmission profile can reach the spectral line (see Fig. ??), hence modifying the normalization of the profile when the object changes at each iteration. Second, should the continuum intensity of the derived object vary with respect to its real value due to the deconvolution process (e.g., due to boundary effects), there will be a shift in the intensity of the whole profile induced by the normalization process. These two effects seem to dominate the accuracy on the gain determination, regardless of the chosen sampling.

For Δa , the performance of the method is slightly worse than for the ideal case when using the deconvolution approach. Unlike the gain determination, errors in the Δa derivation show a strong dependence on the spectral sampling. Differences between both approaches range from 10 ms^{-1} to 40 ms^{-1} and increase with decreasing N_λ . The sensitivity with N_λ is especially high up to $N_\lambda = 8$. A modest increase of N_λ from five to six improves the determination of $\Delta a \sim 20 \text{ ms}^{-1}$, whereas at better spectral samplings, the difference between each simulation decreases more slowly, without any relevant improvement as the sampling increases. In any case, differences are all well within $\pm 1\sigma$.

3.4.2 The crossover case

The fact that the sensitivity of the model to the gain and to the Δa parameters are different guarantees (to some extent) that the parameters can be separated from each other. The treatment of the problem is very different between etalon configurations, and therefore full

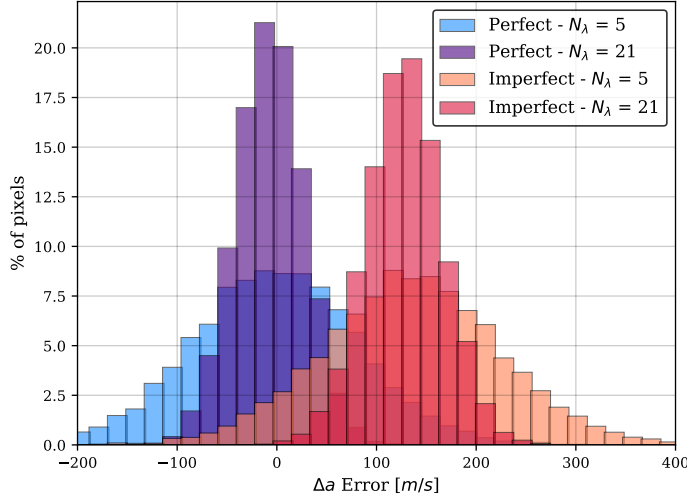


Figure 3.10 Distribution of the errors in the determination of Δa for the crossover scenarios (different configuration in the observation generation and minimization algorithm) for both perfect and imperfect configurations. Only results for the two extreme spectral samplings ($N_\lambda = 5$ and $N_\lambda = 21$) are shown.

knowledge of the setup is critical. However, this is not always feasible due to the unavoidable presence of errors, misalignment, and imperfections on the instrument. Approximations to describe the optical setup are also common in the pipeline of an FPI instrument because they reduce computational efforts. For instance, telecentric mounts are usually simplified as collimated setups, as the f-numbers employed in solar instruments are usually very large. Imperfections of telecentrism are commonly neglected, too. In this section, we analyze the impact of assuming an incorrect etalon mounting. To do so, we repeated the previous exercise, starting from a perfect and imperfect telecentric configuration but assuming that the transmission profile shape corresponds to a collimated one.

In this exercise, we assumed that we have an instrument with an FPI in a telecentric mount, as in the previous sections, in both perfect and imperfect configurations and an $S/N = 200$. We also considered that the object is given by the spectral solar atlas. The shift of the perfect telecentric transmission profile with respect to the collimated one was corrected using Eq.52 from Bailén et al. (2019) to avoid the emergence of spurious velocity signals. Imperfections in the telecentrism shift the profile more. This additional displacement was left uncorrected intentionally so we could study its effects.

Figure 3.10 shows the error distributions for Δa when the model assumes a collimated configuration for $N_\lambda = 5$ and $N_\lambda = 21$ and for both perfect and imperfect configurations. The amplitude and dispersion of the error distributions are very similar for the two mounts and are comparable to the results obtained in the ideal case (Fig. 3.8, bottom panel). The main difference between the perfect and imperfect scenarios is a shift of 130 ms^{-1} for the reason mentioned above. We note that this shift can easily be accounted for since it is a known and measurable effect.

The similarity in the error distributions for the calculation of Δa in the two scenarios demonstrates that the error incurred when assuming a collimated etalon does not significantly impact the determination of the cavity maps of the etalon. This is because changes in Δa mostly induce a shift of the transmission peak by an equal amount in both cases.

We note, however, that the amplitude, width, and shape of the transmission profile differ

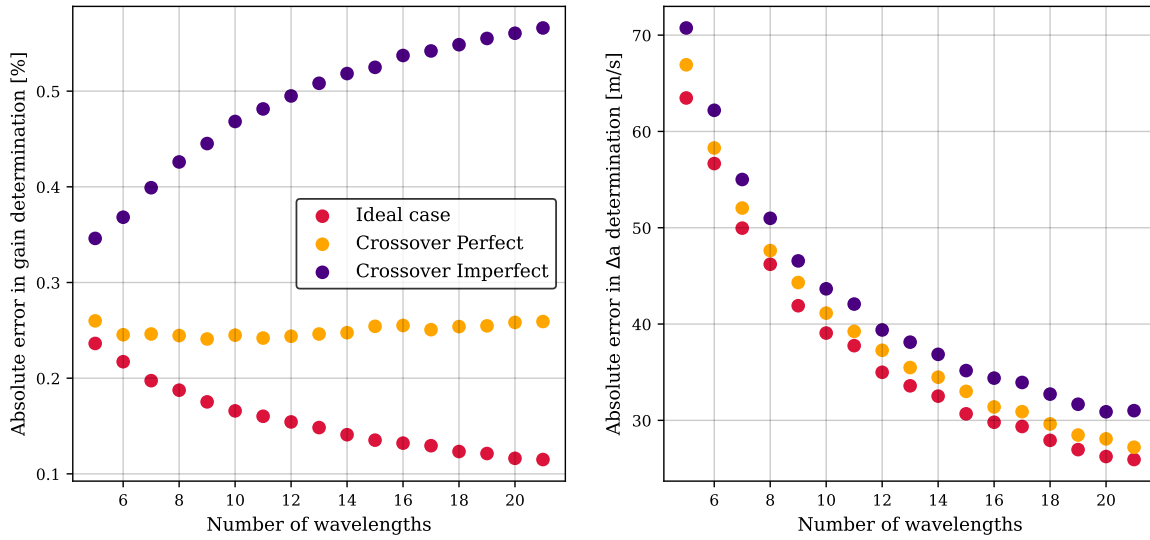


Figure 3.11 Average errors of the gain (left) and etalon defect (right) calculations over all the FoV for the two crossover cases and the standard case (also shown in Fig. 3.7 as the collimated case with $S/N = 200$) for reference. All Δa errors have been computed by correcting differential offsets of the transmission profile between the different mounts.

significantly between the telecentric and collimated configurations, leading to an expected higher error in gain calculation. Figure 3.11 shows the absolute errors in gain and Δa calculations for both crossover scenarios and the ideal case after correcting the wavelength shift between the different mounts. For Δa , the performance of the method is very similar in the three setups, as also observed earlier. This behavior is nevertheless anticipated since the properties selected for simulating the imperfect etalon were chosen to mirror those of the SO-PHI etalon, which were adjusted to closely resemble the behavior of a collimated etalon to the greatest extent possible.

The differences are larger for the gain determination. Not only are the errors higher in the crossover cases, but the trend is entirely different. Instead of decreasing when increasing the number of wavelengths, gain errors remain the same for the perfect case and increase with the number of samples for the imperfect case. Similar to the deconvolution case (Fig. 3.9), the difference between transmission profiles introduces an error in the normalization process that systematically affects the rest of the measurements. This effect becomes more pronounced as the number of wavelengths increases, given that this error is introduced more frequently, and it is even more prominent in the imperfect case, as not only are the profiles different in this scenario, but they are also asymmetric. This asymmetry results in an imbalance in the measurement of the profile, as one wing of the spectral line has a higher transmissivity and is observed with greater intensity than the other.

Our results suggest that assuming an etalon in a collimated configuration for instruments with telecentric mounts can be a good first-order approximation for cavity map calculations, provided that the level of asymmetry of the transmission profile is known. However, achieving an accurate knowledge of the degree of telecentrism is often challeng-

ing in real instruments, as it usually varies across the FoV. Meanwhile, the results highlight that this approximation leads to a considerable increase in the error in the gain determination, which increases when increasing the spectral sampling. This contrasts with the standard philosophy of solar instrumentation, which requires a high number of points to better scan the spectral line.

CHAPTER 4

SCIENTIFIC EXPLOITATION.

4.1 Persistent Homology in Solar Magnetograms.

4.1.1 Persistent Homology

Persistent homology stands out as a prominent technique within the topological data analysis toolkit, primarily for its capacity to capture the shape and distribution information of a dataset. The algorithm is rooted in the mathematical framework of homology groups. In topology, these groups measure the number of n -dimensional holes in a data set, or in other words, the number of connected components for a 0^{th} dimensional analysis, holes or rings for a 1^{st} dimensional analysis, spherical voids for the 2^{nd} dimensional analysis, and so on.

The primary objective of persistent homology is not only to compute the homology groups of a given dataset but also to study how they vary at different scales. To achieve this, the input data undergoes a process of division into a series of sequential subspaces, with each subspace encompassing the previous one. This sequential process, known as filtration, begins with a starting subspace comprising a single point from the original dataset. Subsequent subspaces are then constructed by incrementally adding points to the previous subspace until the final subspace includes all points of the original dataset.

After the filtration process is performed, persistent homology algorithms shift their focus to analyzing the evolution of topological features across the different subspaces. Specifically, they record the filtration value at which a new feature appears, meaning that it is absent in the previous subspace, and when it disappears, meaning that it is no longer present in the following subspaces. These two events are known as the birth and death of a topological feature, respectively.

In a nutshell, the n -dimensional persistent homology of a dataset with a given filtration can be described as the aggregation of all n -dimensional features (homology groups) that were created (birth) and subsequently eliminated (death) during the filtration process (Hensel et al., 2021).

When applying persistent homology on a greyscale image, our focus lies in filtering the data according to the pixel values. Multiple filtering approaches exist, with the most extended ones being sublevel and superlevel filtrations, both based on the concept of thresholding. In these filtrations, the image is cropped to a specific value, forming a subspace that

includes all pixels with values higher than this value in a superlevel filtration, or lower in a sublevel filtration. This cropping value (i.e. the filtration value) is systematically varied from the lowest to the highest values of the image, or vice versa, thus generating a different subspace for each value. As a result, the persistence homology analysis captures and examines the evolution of topological features across different thresholds, enabling insights into the image's structural properties at various scales (Barnes et al., 2021).

A more formal way of defining these filtrations can be done by considering an image as a discrete representation of a function f , defined over a two-dimensional space \mathbb{X} , such that:

$$f : \mathbb{X} \longrightarrow \mathbb{R} . \quad (4.1)$$

Let \mathbb{S}_ϕ be the subspace of \mathbb{X} for a filtration value of ϕ . In such a case, a filtration can be expressed as:

$$\mathbb{X} : \mathbb{S}_{\phi_0} \subset \mathbb{S}_{\phi_1} \subset \mathbb{S}_{\phi_2} \subset \dots \subset \mathbb{X} . \quad (4.2)$$

With this formulation, a topological feature with birth-death coordinates:

$$(B, D) = (\phi_I, \phi_{II}) , \quad (4.3)$$

corresponds to a feature that appears for the first time during the filtration process at the subspace \mathbb{S}_{ϕ_I} , and *persists* until the subspace $\mathbb{S}_{\phi_{II}}$, where it ceases to exist.

In a sublevel filtration, each subspace can be expressed as:

$$\mathbb{S}_\phi = f^{-1}((-\infty, \phi]) , \quad (4.4)$$

where ϕ_0 is selected as the lowest value for any given pixel and its value is increased until the subspace includes all pixels. On the contrary, in a superlevel filtration, the subspaces can be expressed as:

$$\mathbb{S}_\phi = f^{-1}([\phi, \infty)) , \quad (4.5)$$

where ϕ_0 is selected as the highest value for any given pixel and its value is decreased along the filtration.

Various methods exist for representing the information derived from a persistent homology analysis, including Betti numbers, persistence bars, and persistent diagrams (PDs) (Cohen-Steiner et al. 2005, Aktas et al. 2019), among many others. For this study, we will utilize the PDs as our chosen approach due to their straightforward interpretation and extended use. A n -dimensional PD is a multiset of Birth-Death pairs, (B_i, D_j) , with multiplicity k , where each pair measures the number (k) of n -dimensional components that have been born at the filtration subspace \mathbb{X}_i and died in \mathbb{X}_j , that is usually represented in a 2D scatter plot.

The process of generating a PD of a greyscale image is as follows. We start by selecting the filtration direction (sublevel or superlevel) and the dimension of the analysis (either 0 or 1). We initialize a threshold as the highest or lowest value from the image, depending on the choice of filtration. We then perform the filtration by systematically adjusting the threshold and creating a binary image for each threshold. This process divides the image into two sets: pixels with values above the threshold and pixels with values below it. The

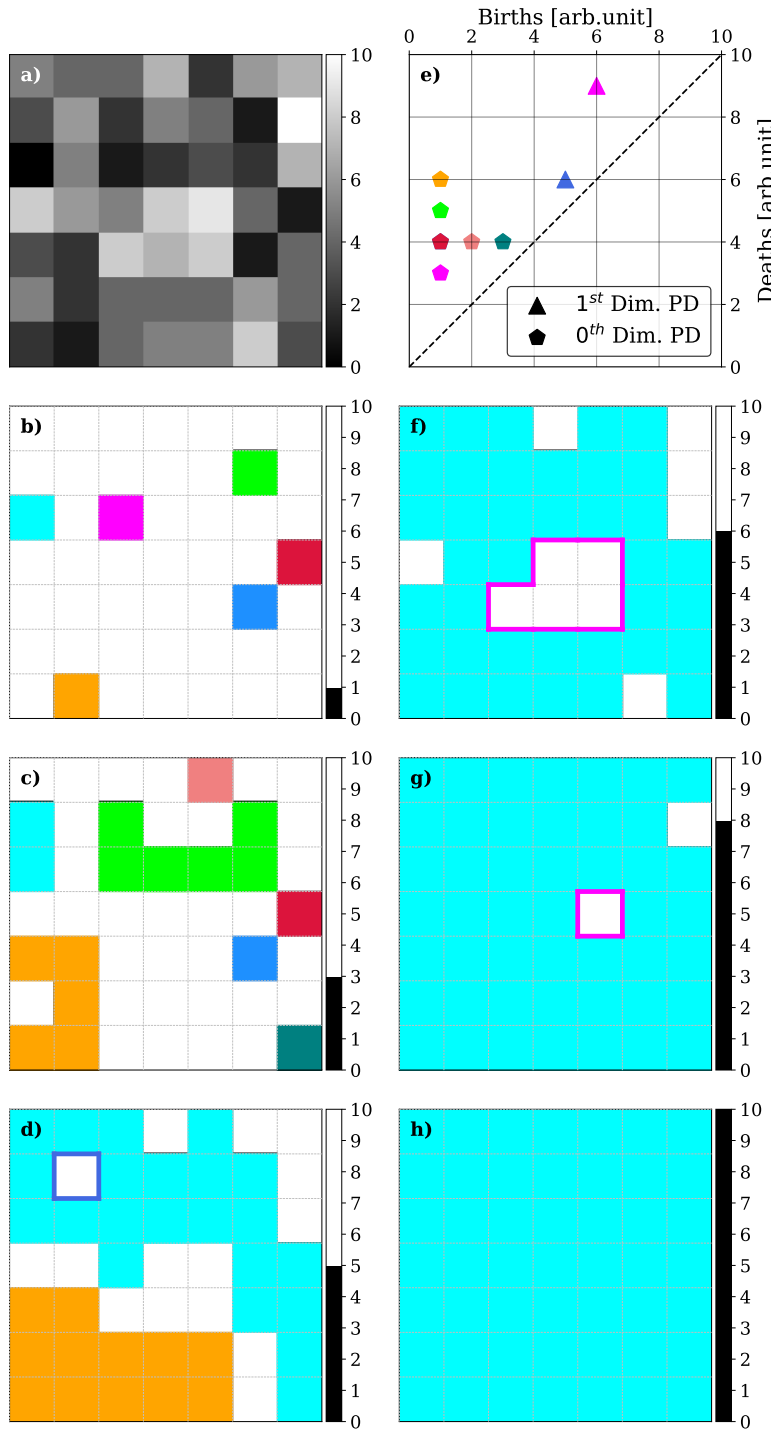


Figure 4.1 Sublevel filtration of a greyscale image and PDs of the 0^{th} and 1^{st} dimensions. Panel a) shows the input data. In panels b), c), d), f), g), and h) different snapshots of the filtration process are shown. The value of the filtration parameter, ϕ is shown in the color bar at the right of each image. Only pixels with a value lower than the filtration value (colored pixels) belong to the subspace shown in each snapshot. Different homology groups are represented with different colors at each snapshot. For connected components (0^{th} dimensional homology groups) the whole pixel is shown with the corresponding color. For rings, (1^{st} dimensional homology groups), only the border of each hole is colored. In panel e) the PDs of both dimensions are shown. The color of each point in the diagram is the same as the one used to plot the corresponding topological feature in other panels. An animation depicting the whole filtration process is also provided as part of the article.

choice of filtering determines which of the two sets makes up the subspace. We then look for the existing topological features within each of these subdivisions. The specific process by which these features are identified is detailed in the next paragraph, where the structures corresponding to both dimensions are illustrated using the example shown in Fig. 4.1. We repeat this process until the threshold reaches the opposite limit to that from which it started. Along this process, we follow the appearance, merging, and disappearance of connected components. When two components merge, the longer-lived one (*i.e.* the first to appear along the filtration process) absorbs the younger one, thus resulting in the death of the second (Edelsbrunner & Harer, 2022). We determine the birth and death for each component based on the thresholds at which these events occur. Finally, we construct a scatter plot where the horizontal axis represents the birth values and the vertical axis represents the death values. Each point on this plot corresponds to a persistence point, whose coordinates reveal the scales at which the corresponding topological feature is present.

An example of this process with a sublevel filtration is shown in Fig. 4.1. Panel a) displays the input data, panels b), c), d), f), g), and h) show some of the key steps of the filtration process, and, lastly, panel e) displays the PD for a 0^{th} and 1^{st} dimensional analyses. These plots illustrate how connected components and rings are born and then die as we increase the filtration level. As these components (shown in different colors) increase in size and come into contact with other components, one absorbs the other, thus resulting in the death of the second. This phenomenon is shown in panels b) to d), where we observe the progression of the components until only the blue and orange connected components remain. Additionally, the diagrams also reveal the appearance of two rings in the data (panels f) and g)). These rings are found when pixels that do not belong to the subspace are surrounded by a connected component, and die when those pixels are included in the component as the threshold increases (blue ring in panel f)). Finally, the PD (panel e)) displays the birth and death values (*i.e.* the filtration value) of all the features, of dimensions 0 and 1, that have been identified (birth) and subsequently eliminated (death) throughout the filtration process.

4.1.2 Persistent Images

The PD displayed in Fig. 4.1 contains only a limited number of points due to the simplicity of the input image. However, when analyzing real data, these diagrams can consist of hundreds or even thousands of birth-death pairs with high multiplicities, simply due to the size of the images. Additionally, features not only representing the genuine behavior of the data but also reflecting the distribution of noise appear on the diagrams. To address this complexity, several strategies have been developed to simplify the information from PDs, such as persistence curves (Chung & Lawson, 2019), persistence landscapes (Bubenik et al., 2015), or persistence images (PI) (Adams et al., 2017). In this study, we will focus on the latter, due to its noise filtering capabilities and because the representation of the results remains in a Birth-Death diagram, allowing for easy interpretation of the results, similar to a persistence diagram.

PIs are a condensed form of a persistence diagram, offering a concise and easy-to-

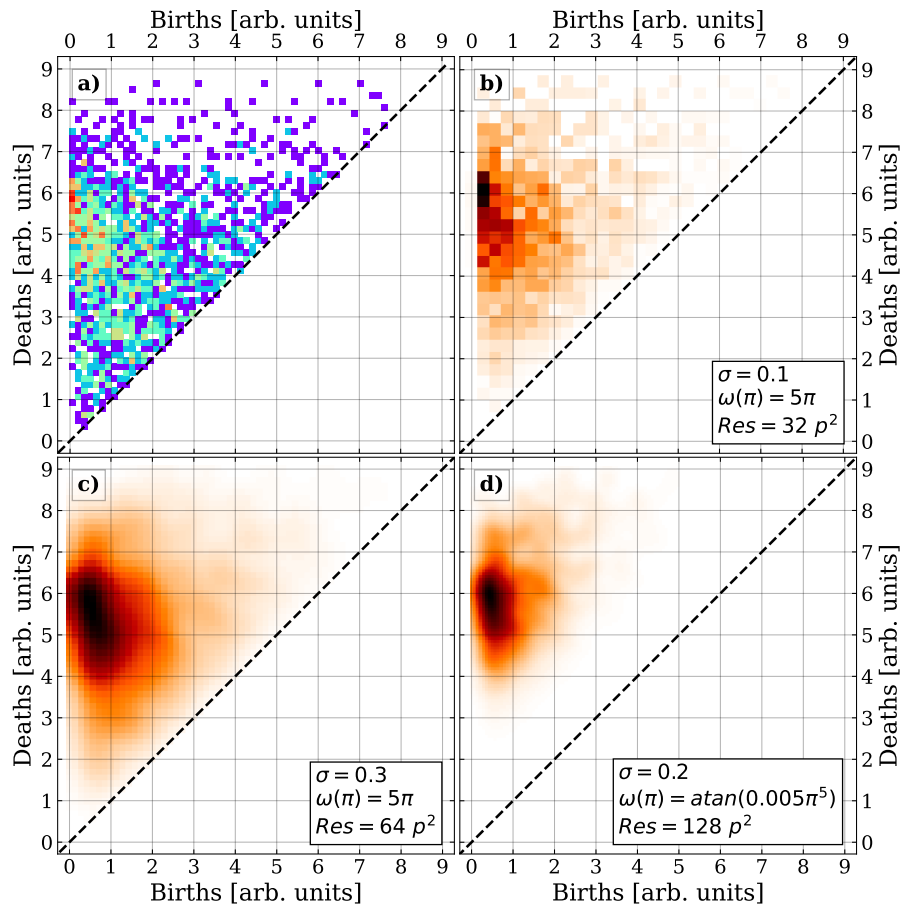


Figure 4.2 Panel a) shows an example of a persistence diagram as a 2D histogram. The color of each bin represents its multiplicity, with the red spots corresponding to higher values. Panels b), c), and d) show three different examples of PIs. The three parameters given in the legends of the PIs are: the standard deviation (σ) of the Gaussian kernel ($K(z)$), the weighting function, and the resolution of the image PI in pixels (p).

understand representation of its topological features. They capture the spatial distribution and persistence information of these features, allowing for the enhancement of the most relevant ones and filtering of the others. A PI is constructed using the concept of persistence. Each topological feature, represented by a point in a PD, has a persistence, π , of:

$$\pi = D - B \quad (4.6)$$

where (B, D) are the corresponding birth-death coordinates in the diagram. A feature with a large persistence is present at different scales in the data and therefore is more likely to represent the real behavior of the data. On the contrary, short-lived features are typically associated with the noise distribution and usually do not provide much information about the data.

When constructing a PI, a weighting function, $\omega(\pi)$, is employed to assign weights to

each point in the diagram, ensuring that longer-lived features have greater weights than shorter-lived ones. There are multiple choices for the shape of the weighting function, which are entirely dependent on the aims of the study and data type. The simplest example is often a linear or power-law relation ($\omega(\pi) = a\pi^b$), where a and b can be tuned to assign progressively higher weights to higher persistencies, thus focusing the study on the longer-lived components. On the other hand, if the objective is to filter out noise while assigning similar weights to all non-noise points so that all points have a similar relevance in the analysis, the chosen function is usually an arc-tangent.

The PI is then generated by dividing the persistence diagram plane into a grid with a desired resolution. Within each grid region (or pixel), the weighted features of the diagram within the region are added up using a kernel density estimation. The kernel function, $K(z)$, can be tuned to suit the nature and objectives of the analysis, with Gaussian functions being the most common approach.

The resultant PI is a 2D matrix, wherein each pixel corresponds to a specific area in the persistence diagram, and its value represents the cumulative weight of the topological features found within that area. In Figure 4.2, three examples of PI (panels b), c), and d)) are presented for the same persistence diagram (panel a)), where distinct choices of resolution, kernel function, and weighting function have been applied to each image.

All the PDs, PIs, and the rest of the analysis tools presented in this work, have been computed using the Homcloud python package (Obayashi et al., 2022).

4.1.3 Data

In this work, we study the results of applying persistent homology to different regimes of solar activity by applying the analysis to both quiet Sun and active region magnetograms.

4.1.3.1 Quiet Sun observations

The study of quiet Sun regions requires high magnetic spatial and temporal resolutions and sensitivities to be able to capture the small-scale evolution of the magnetic structures due to their weak signals and short time scales (Bellot Rubio & Orozco Suárez, 2019). For this reason, we employ observations taken by the Solar Optical Telescope (SOT; Tsuneta et al. 2008) aboard the *Hinode* satellite (Kosugi et al., 2007), a space-borne solar observatory. In particular, we employ observations from *Hinode*'s Operation Plan (HOP) 151. These observations consist of long (≥ 20 h) and mostly uninterrupted sequences of measurements of the Narrowband Filter Imager of the Na I D1 line at 5896 Å taken with a cadence of 50–70 s. The data correction of the selected observation sets has been carried out in (Gošić et al., 2014).

4.1.3.2 Active regions observations

We employ observations of active regions (ARs) taken by the Helioseismic and Magnetic Imager (HMI; Scherrer et al. 2012, Schou et al. 2012) on board the Solar Dynamics Obser-

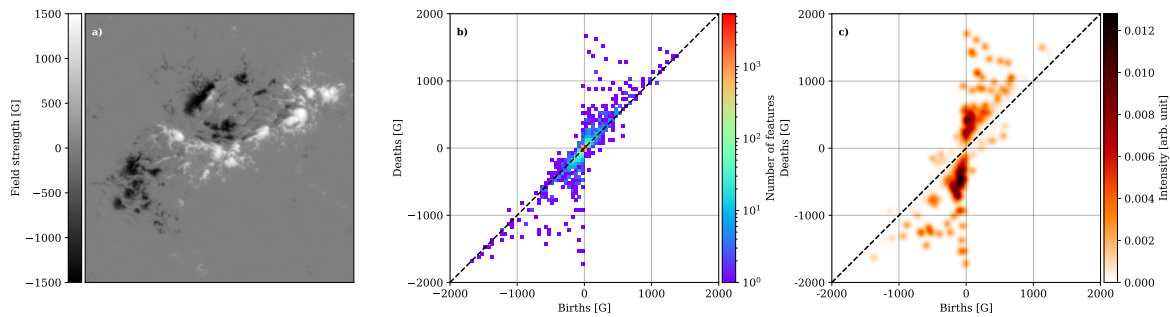


Figure 4.3 (a) SDO/HMI magnetogram taken on 2011-02-13 depicting an active region (NOAA AR 11158). (b) The corresponding PD combining superlevel and sublevel filtrations. (c) PI generated from the PD in panel b) with the following configuration: Resolution = 1000 pixels² (4 G per pixel), weighting function: $\omega(\pi) = \arctan(5 \times 10^{-8}\pi^3)$ and a gaussian kernel with $\sigma = 40$ G.

vatory (Pesnell et al., 2012). HMI provides a continuous observation of the Sun where a full-disk magnetogram, as well as Dopplergrams, are provided at all times. The full-disk, uninterrupted observations of HMI make it a very suitable instrument to study the evolution of active regions as the formation and development of active regions can be fully captured.

We focus the analysis on a series of newly-emerging ARs identified in (Toriumi et al., 2014a). In particular, we employed the 12-minute cadence observations taken during the period from May 2010 to June 2011, which corresponded to a period of low solar activity.

4.1.4 Analysis and results

The application of persistent homology to a specific dataset can vary depending on the aims of the study. Different dimensions of the analysis and various types of filtrations focus on distinct features within the data. It is crucial to have prior knowledge of the expected structures and relevant features to be captured in the analysis in order to determine the appropriate approach. In this section, we aim to outline the most appropriate approach for studying the particular case of solar magnetograms.

The solar magnetograms employed here represent the longitudinal component of the magnetic field on the photosphere and are typically presented as greyscale images, as shown in Figure 4.3, panel a). The polarity of the line-of-sight magnetic field is indicated by the sign of each pixel, where positive and negative signals correspond to field lines pointing towards and away from the observer. Applying a single filtration to a greyscale image only displays features corresponding to one polarity (positive or negative) in a PD. However, to conduct a comprehensive study of the magnetic field, both polarities are essential, thus necessitating the use of two separate filtrations with different filtration directions.

We determined that a combination of superlevel and sublevel filtrations with a 1st-dimensional persistent homology analysis was the most suitable approach for studying

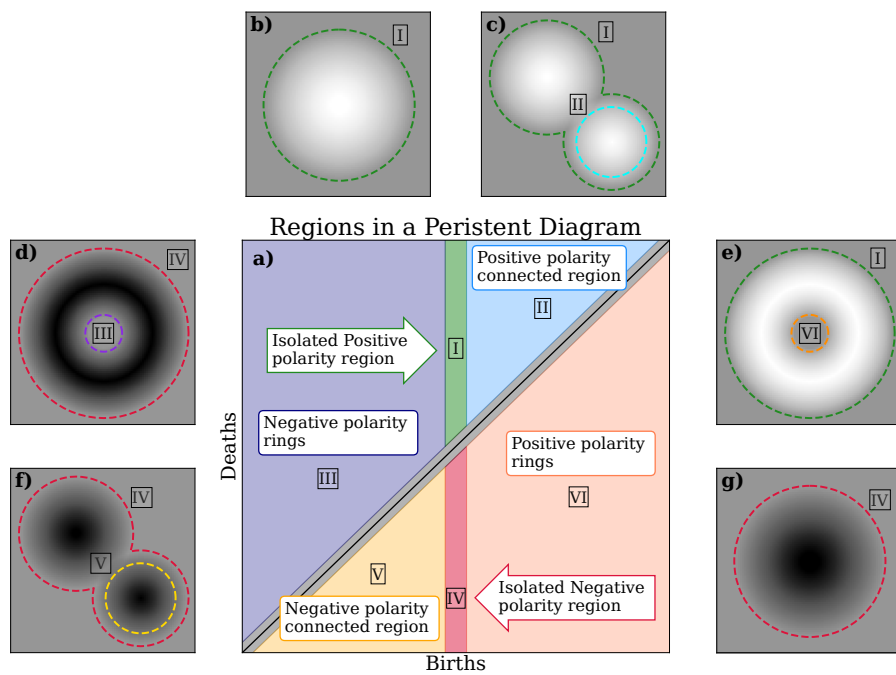


Figure 4.4 Schematic representation of the PD and the different regions (panel a)). The magnetic structures corresponding to the topological features found in the different regions are shown in panels b) to g), with each feature identified by a ring with the corresponding color.

solar magnetograms. This choice is based on two main reasons. Firstly, the 1st dimensional analysis allows us to identify the most prominent features in a PD, which is not the case in the 0th dimensional analysis where the strongest feature does not appear in the diagram as it never dies (see Fig. 4.1). Secondly, by combining superlevel and sublevel filtrations, we can display the results of both filtrations in a single diagram. Features corresponding to different filtrations will have persistencies with opposite signs (features found in a superlevel filtration will be born at higher filtration values than their death, resulting in a negative lifespan). This enables us to construct a PD in which all features displayed above the identity line (with positive persistencies) correspond to the sublevel filtration, and those below the line correspond to the superlevel filtration (see panel b) in Fig. 4.3).

The PDs, and consequently the PIs, offer valuable insights into the magnetic structures present in the magnetograms. The location of a topological feature on the diagram is completely determined by the properties of the corresponding magnetic structure. Specifically, this position is influenced by factors such as the maximum intensity of the magnetic field, its proximity to other magnetic structures, and its geometric shape, including the presence of holes or pores within the structure. These characteristics allow us to partition the diagram into distinct regions, where topological features within each region correspond to different types of magnetic structures.

We distinguished between six distinct regions in the diagram. Figure 4.4 illustrates these regions in panel a) and provides schematic representations of the corresponding magnetic structures in panels b) to g). The regions are as follows: first, topological features located above the identity line (positive persistencies) with birth values close to 0 (region I in the diagram). Features within this region represent isolated magnetic structures of positive polarity, that is, patches of positive magnetic flux fully enclosed by an absence of any magnetic field. The threshold defining what is considered “close to 0” is determined by the data’s properties. To identify isolated structures, we set the threshold as a function of the statistical properties of the background signal (i.e. areas of the magnetogram with little magnetic flux). Specifically, the limits for this region are set as $(-5\sigma_{bg}, 5\sigma_{bg})$, where σ_{bg} denotes the standard deviation of the background signal found in a 15×15 pixels box devoid of strong magnetic structures.

The second region (II in the diagram) comprises features above the identity line with positive birth values, representing connected structures with positive polarities, that is, positive magnetic field structures in contact with another positive structure but not fully merged. The third region (region III) contains topological features above the identity line with negative birth values, which corresponds to magnetic structures of negative polarity exhibiting a ring-like attribute, namely, structures with pores or holes. These three regions of the diagram have counterparts with negative persistencies. Thus, features associated with isolated structures of negative polarities are found in the region with a birth value close to 0 but below the identity line (region IV), features for connected negative structures are also located below the line but with negative birth values (region V). Lastly, features arising from positive magnetic structures with ring-like attributes are found below the identity line but with positive birth values (region VI).

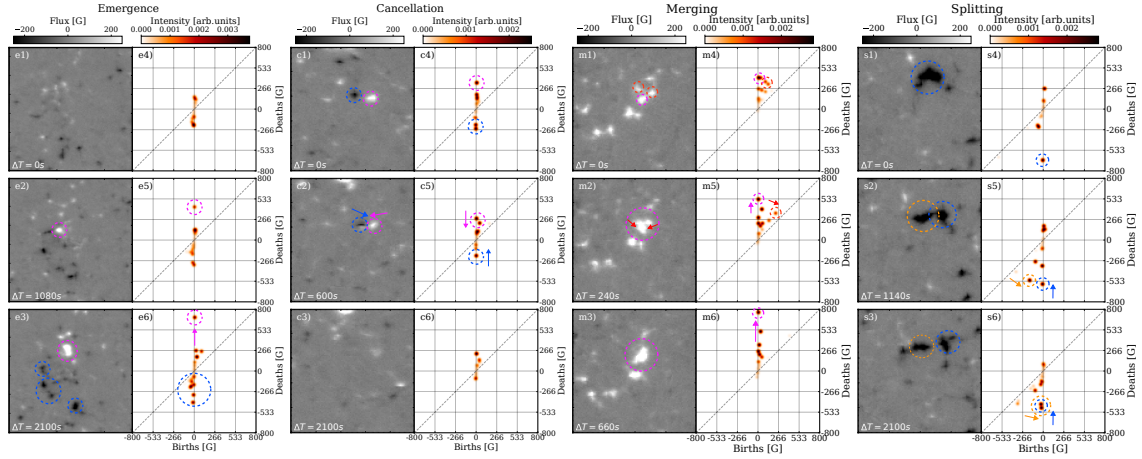


Figure 4.5 Examples of flux emergence (left column), flux cancellation (central left column), merging (central right column), and splitting (right column) events in the quiet sun. The time interval shown in the magnetograms is always with respect to the first frame (panel a) for emergence and g) for cancellation). The PIs correspond to the magnetogram to their left. The parameters for their computation are: resolution = 1000 pixels² (1.6 G per pixel), weighting function: $\omega(\pi) = \arctan(5 \times 10^{-7} \pi^3)$ and a Gaussian kernel with $\sigma = 16$ G.

4.1.4.1 Quiet Sun results

Quiet Sun regions are characterized by the presence of weak, small-scale magnetic field signals that exhibit rapid evolution. This rapid evolution leads to a multitude of signals in a single snapshot that evolve quickly from one frame to another. The small scale and rapid changes of the processes of the quiet Sun make data analysis techniques desirable for their study due to the complexity of such endeavors.

When examining the evolution of signals across the entire field of view, we observe a dynamic process characterized by numerous regions interacting destructively while new signals emerge throughout the whole region. Despite these continuous changes, the overall structure of the magnetogram appears stable, with consecutive snapshots exhibiting strikingly similar properties. This apparent equilibrium state is also evident when studying the PIs, as consecutive frames show minimal differences in their representations.

Due to the apparent equilibrium state of the overall structure of the magnetograms, it is necessary to narrow down the field to which we apply the analysis. We found that when the number of signals in the studied region is lower, we can observe flux cancellation and emergence events, as well as merging and splitting events, through the changes induced in the persistence diagram. In cancellation events, two regions of magnetic flux with opposite polarities interact destructively, nullifying each other. On the contrary, in flux emergence events, we observe signals of both polarities suddenly emerge from a region with little magnetic flux. In splitting and merging events, only one polarity is involved. Two or more distinct structures of equal polarity merge together in merging events, and a single structure is divided into two or more for splitting events.

In Figure 4.5, an example of an emergence event is shown in three snapshots through

the magnetograms (panels e1 to e3) and their corresponding PIs (panels e4 to e6). When we focus on the second snapshot (panel e2 and e5), we see that a new positive and isolated feature (birth ~ 0) that was not present in the previous snapshot has appeared in the PI and stands out from the rest of the signals (highlighted with a pink circle in both magnetogram and PI), while simultaneously, the density of negative polarity features also begins to increase. In the last frame, we see that in the case of positive polarity, the majority of the signal has concentrated in a single structure, as shown by the increase of the death value of the corresponding feature in the PI. A few connected features are also seen, but these are less significant. Meanwhile, the negative polarity signal has been distributed into multiple structures instead of concentrating in a single one, as evidenced by the absence of a prominent feature in the PI and the increased density of connected features (highlighted with blue circles both in the PI and the magnetogram).

A very similar analysis can be carried out to analyze cancellation events. The evolution of the persistence image is very similar to that of the emergence events but in the opposite direction. Figure 4.5 also shows an example of a flux cancellation event through magnetograms (panels c1 to c3) and PIs (panels c4 to c6). In the beginning, the PI shows the presence of features of opposite polarities. When the corresponding structures approach each other and begin to interact, the magnetic signal starts to decrease, which is observed in the PI as a simultaneous movement of the features towards the center of the diagram. This reduction in signal continues until both features reach the center of the diagram, which corresponds to the moment when they will have completely canceled each other out (panels c3 and c6).

For the events that only involve one polarity, namely merging and splitting events, the same behavior is seen in the PI for positive and negative features, but on their respective sides of the PI.

An example of a merging event of positive polarity structures is shown in Fig. 4.5, in panels m1) to m6). These events start with multiple isolated, or interacting structures (as shown in panels m1 and m4), that are moving towards each other. As the structures cluster, two movements are seen in the PI: firstly, the features corresponding to the structures with the weakest field (red features in panels m2 and m5) move towards the identity line; and secondly, the feature corresponding to the main structure (pink feature in panels m2 and m5) experiences an increase in its absolute death value due to the increase in magnetic flux coming from the rest of the structures. When the structures are fully merged, only a single feature appears in both the magnetogram and PI (pink feature in panels m3 and m6), in the isolated region (region I or IV, depending on the polarity).

This process is reversed for splitting events, as shown in panels s1) to s6) of Fig. 4.5. We see how an initially isolated feature in the PI (blue feature in panels s1 and s4) evolves into two (or multiple) features. When the process has started, but the two parts have not yet completely separated, a second feature appears in the diagram in the region corresponding to the connected structures (regions III or V in the diagram), as shown in panels s2) and s5). As the two structures continue to separate, this second feature gradually approaches the region for isolated structures (regions I and IV) as the magnetic field surrounding it in the magnetogram diminishes (panel s5). Eventually, when both structures are completely

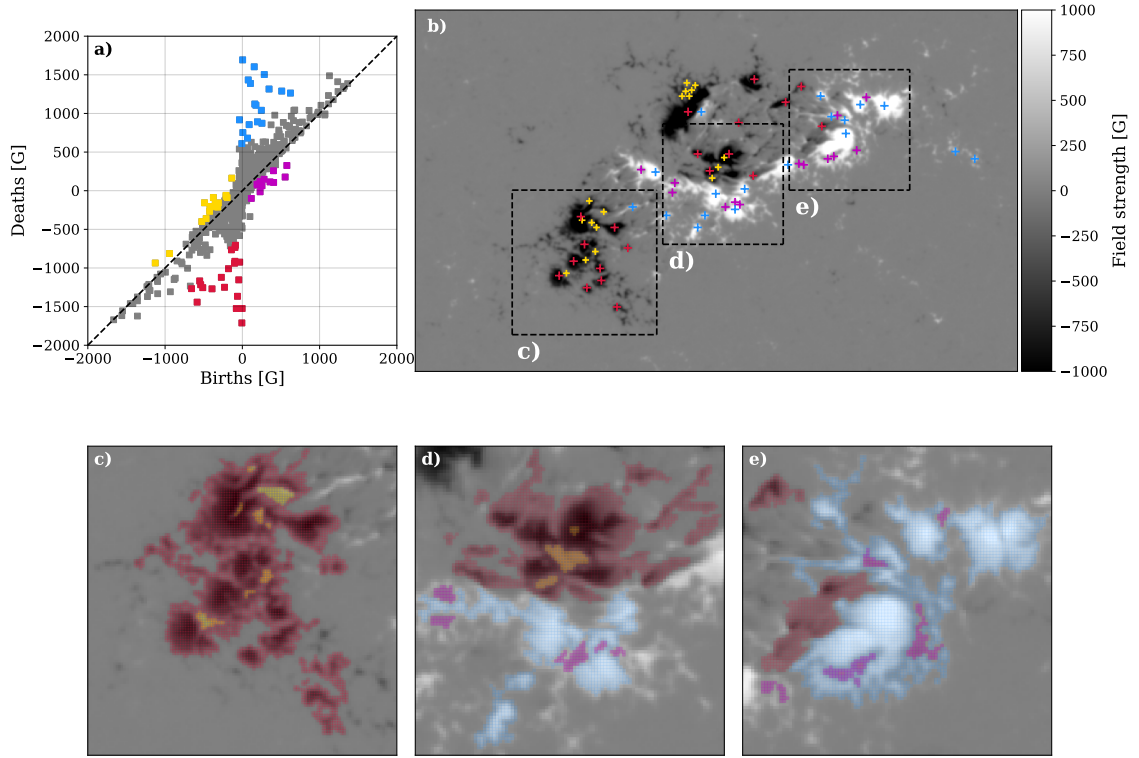


Figure 4.6 SDO/HMI magnetogram taken on 2011-02-13 at 06:34 UT depicting an active region (NOAA AR 11158) and the corresponding persistent diagram (panel a)). Panels c) to e) display a zoomed region of the active region, corresponding to the labeled region with the same letter in panel b). Features of different types are shown in different colors in the diagram. The corresponding structures are shown in the same colors in the magnetograms; as colored crosses in panel b) and as a colored transparent overlay displayed over the whole pixel in panels c) to e) to show the extent of the structure.

separated (i.e. with no signal around them), they will both appear on the diagram as two isolated features with a lower death value compared to the initial one (panels s3 and s5). This reduction occurs because the magnetic flux is distributed between the two structures.

4.1.4.2 Active region results

It is important to understand which types of magnetic structures can be identified through persistent homology and establish the correspondence between these structures and the position of the corresponding topological feature in a persistence diagram. To achieve this understanding, Fig. 4.6 displays both a magnetogram with complex morphology (panel b) and its corresponding persistence diagram (panel a), along with three zoomed-in regions of the magnetogram (panels c to e). In all panels, some topological features or their corresponding magnetic structures have been color-coded based on their types, or equivalently, based on their positions in the persistence diagram. In the complete magnetogram (panel

b), structures have been marked with a cross, indicating the pixel where the structure died during the filtration process. Meanwhile, in panels c to e, all pixels composing each structure have been colored. It is noteworthy that nearly all pixels appear colored because we have selected the most significant structures—those with longer lifetimes (Eqn. (4.6)). Consequently, these structures encompass all less significant structures that are absorbed and incorporated into the former during the filtration process.

The analysis of the persistence diagram allows us to deduce several properties of the magnetogram. Firstly, the persistence diagram provides a rapid assessment of the intensity of the magnetic flux since the death value of the topological feature coincides with the maximum flux (in absolute value) within the corresponding structure. In the case of Figure 6, we observe that several structures exhibit maximum (absolute) values surpassing 1500 G, with multiple structures falling within the range of 1000 G to 1500 G.

Secondly, we can infer how the magnetic signals are distributed by examining the number of isolated and connected structures in the diagram (structures highlighted in blue and red depending on their polarity in Fig. 4.6). The complex morphology of the structure displayed in the magnetogram is evident in the high number of connected structures (regions II and V in the diagram) and the absence of prominent isolated structures (regions I and IV).

Lastly, the presence or absence of ring-like structures provides insights into how the magnetic structures are connected. These features can only be found in regions where connected structures create highly complex morphologies with gaps between them, as illustrated in the magnified regions of the magnetogram in Figure 6 (panels c to e).

An example of how the three features allow us to classify ARs depending on their morphologies is shown in Figure 4.7, where three different ARs and their corresponding PIs are displayed. Although at first sight, the PIs appear to be very similar, especially the ones shown in panels d) and f), upon closer inspection, it is possible to find the differences when focusing on the three features mentioned previously. The first AR (panel a)), also shown in Fig. 4.6, shows a very complex morphology, where the magnetic field of both positive and negative polarities is distributed in multiple connected structures. This behavior is displayed in the PI through the high density in the isolated and connected features in equal proportions (i.e. with no prominent features) and with the presence of ring-like features in both polarities. In contrast, the second AR (panel b) shows a simpler magnetic structure with weaker signals. The PI for this case shows an absence of ring-like features in both polarities and a very low density in the regions for connected and isolated features. Lastly, panel c) shows an AR where the positive magnetic field is concentrated in one main area whereas the negative polarity magnetic field shows a fragmented structure. Comparing the corresponding PI with that of the initial case (panels f and d, respectively), a similarity is evident in the region corresponding to negative polarities, observed in both the density of connected components and the presence of ring-type structures. Nevertheless, when examining the positive polarity, it becomes evident that, unlike its negative counterpart, there are only a few prominent isolated structures and a notably low density of connected structures. Furthermore, this asymmetry between the positive and negative distributions is underscored by the absence of ring-like structures in the former.

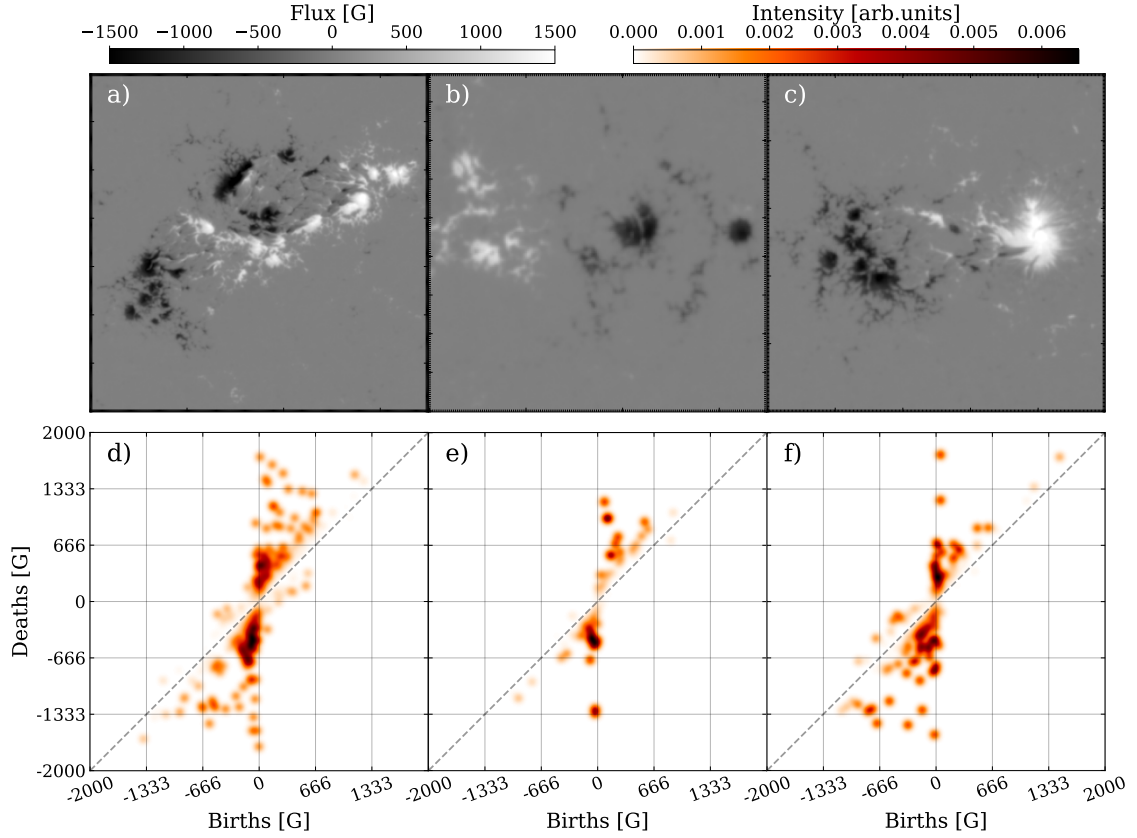


Figure 4.7 SDO/HMI magnetograms of three different active regions and their corresponding persistence images. a) NOAA AR 11158, date: 2011-02-13 at 06:34. b) NOAA AR 11098, date: 2010-08-12 at 20:58. c) NOAA AR 11072, date: 2010-05-22 at 20:58. All persistence images have been generated with the following parameters: Resolution = 1000 pixels² (4 G), weighting function: $\omega(\pi) = \arctan(5 \times 10^{-8} \pi^3)$ and a gaussian kernel with $\sigma = 40$ G.

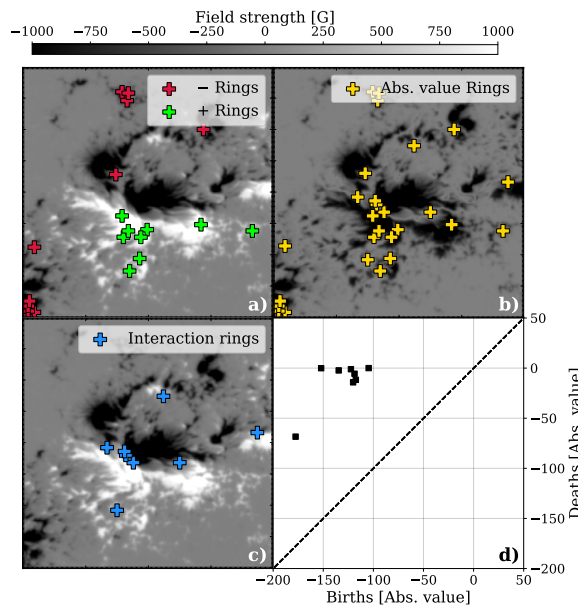


Figure 4.8 Depiction of the different steps carried out for the generation of the interaction diagram (panel d)) for a zoomed-in region of an SDO/HMI magnetogram of NOAA AR 11158 observed on the 2011-02-13 at 18:34. The crosses point to the position at which each of the ring-like structures die, which is always located inside the perimeter of the hole. Panel a) shows the original magnetogram and the position where rings of positive and negative polarities are found. Panel b) shows the magnitude of the field strength in negative values along with the rings found in this representation. Panel c) shows the original magnetogram again, but now featuring the interaction rings, identified when comparing the position of the rings found in the previous steps. Lastly, panel d) shows the ‘interaction diagram’ generated using only the interacting ring structures.

4.1.4.3 ‘Interacting’ Diagram

So far, we have shown how persistent homology is capable of identifying the various morphologies of active regions and the types of structures that can be identified through persistent images or diagrams. Nevertheless, these structures are either isolated or regions of the same polarity that interact with each other. Due to the nature of the filtration process, persistent homology is unable to detect structures where magnetic fields of opposite polarities form a joint structure. However, ARs where there is a significant interaction between magnetic fields of opposite polarities are of greater interest due to their association with flare production.

This issue is illustrated in the analysis carried out in the previous section of NOAA Active Region 11158 (see Fig. 4.6). Although we can capture the complexity of both positive and negative magnetic structures through the persistent diagram, the δ -spots present in the central region remain unnoticed. However, the large number of flare events associated with this region including an X2.2-class event are thought to be related to the abundance of these structures (e.g. Sun et al. 2012, Toriumi et al. 2014b). In this section, we describe a way to efficiently detect and quantify these structures using persistence homology, with only a few additional steps in the analysis.

We start by tracking the position in the magnetogram where rings are detected. Following this, we modify the magnetogram by inverting the sign of one polarity, ensuring that all pixels are either negative or positive. This way, we construct a second ‘magnetogram’ in which we only have information about the intensity, in absolute value, of the magnetic field.

This allows us to identify features formed by any combination of signals, whether they are of the same polarity or opposite. Using this new magnetogram, we repeat the analysis and record the positions in the magnetograms at which we find ring-like features. It is worth noting that in this analysis, all the rings identified in the initial step, which were formed by structures of equal polarities, are still detected, although their position in the persistent diagram may have changed due to the change in polarity, hence the relevance of tracking the pixel in the magnetograms. However, only in this second analysis can we identify rings formed between opposite polarities. By selectively considering the ring-like features exclusively identified in this second analysis, we can effectively identify and characterize the structures that are formed by the interaction between different polarities.

Figure 4.8 depicts an example of the interaction analysis for NOAA AR 11158, which exhibits a strong interaction between opposite-polarity fields. Panel a) displays the magnetogram, indicating also the positions where ring-like features have been identified for both positive and negative polarities. Only features with absolute persistencies greater than $5\sigma_{bg}$ and whose birth occurs in the range: $(5\sigma_{bg}, \infty)$, for positive polarities, and: $(-\infty, -5\sigma_{bg})$ for negative polarities, have been recorded. We now repeat the same analysis but using only the magnitude of the signal. To do this, we invert the positive polarity and, once again, register the positions of the ring-like features in this new magnetogram, as shown in panel b). As observed in panel c), the majority of the interaction rings (*i.e.* those exclusively identified in the second analysis) are located in areas characterized by strong interaction, where the δ -spots were found. In these areas, both polarities interact, resulting in the formation of ring-like structures comprising positive and negative magnetic fields due to their close proximity. These areas, such as δ -spots, are of particular interest, as they typically harbor magnetic cancellation, reconnection, and flux emergence. However, some points appear to be situated in uni-polar fields. These points, despite what may appear at first sight, are found by this analysis due to a structure that requires the other polarity to close completely and thus form a ring. It is noteworthy that the occurrences of such cases are quite limited when compared to the rings observed in highly interacting zones. While their presence does not necessarily indicate intense interaction, it does imply a certain level of interaction between the two polarities. These features can be represented in a persistence diagram in an analogous way to the standard results of a persistent homology analysis. This is what we have referred to as ‘interacting diagram’ and it is worth noting that only the magnitude of the birth and death coordinates are relevant parameters since the sign will be the one matching the polarity selected at the second step. We have chosen to invert the positive polarity so that these features have positive persistencies, as it is more common in persistent homology studies. However, it is important to emphasize that this decision is completely arbitrary and has no impact on the results of the analysis.

It is useful to determine the information conveyed by the interaction diagram regarding the structures themselves. By taking into account the position and quantity of the interaction rings, along with the temporal evolution of the diagram, we can discern the moment and location where these highly interacting structures develop. Therefore, interaction diagrams could be a new tool to identify, through their topological properties, the strong-gradient polarity inversion lines that characterize δ -spots. To achieve this, it is necessary

to incorporate into the analysis the temporal evolution of these structures and study the properties that can be extracted from the δ -spots through these diagrams, which goes beyond the scope of this work but represents the next (necessary) step to assess the predicting capabilities of persistent homology in the field of solar physics.

4.1.5 Conclusions

In this study, we investigate the most adequate approach for the application of persistent homology algorithms to the analysis of solar magnetograms. By combining different filtrations in a single one-dimensional persistent homology analysis, we can effectively capture structures corresponding to both polarities of the magnetic field. We have applied this analysis to observations of the quiet Sun and active regions, taken with both Hinode/SOT and SDO/HMI, respectively. Lastly, we have analyzed the results and identified the features of the data that can be found through persistent diagrams and images, and also show some examples of applications of the algorithms.

Our proposed approach to persistent homology algorithms involves the integration of sublevel and superlevel filtrations within a single analysis, enabling the creation of a comprehensive persistence diagram that encompasses features from both positive and negative magnetic structures. Through the examination of the positions of these identified features within the resulting persistence diagram, we can discern the diverse magnetic features present in the magnetograms. This approach has demonstrated its efficacy in capturing the intricate complexity of magnetic structures, with a particular emphasis on active regions. Through this method, we have achieved successful differentiation between the various morphologies present in active regions by analyzing the presence or absence of specific features in the corresponding persistence images.

On the other hand, the persistent images obtained from quiet Sun observations exhibit significant similarity to each other. This indicates a lack of overall evolution in the magnetic structures within these regions. In quiet Sun areas, small regions of magnetic flux interact with each other in small-scale events, while the overall structure remains relatively static. These small-scale events become more apparent in persistent images when the field of view is reduced. These small-scale events, such as flux emergence or cancellation, can be observed through persistent images as a joint movement of negative and positive features. In cancellation events, the features move toward the center of the image, while in emergence events, they move away from the center.

Additionally, we have successfully identified interactions between opposite-polarity magnetic fields by detecting ring-like features formed by these two polarities. To achieve this, we introduced a method for calculating an ‘interaction diagram’ that selectively displays features resulting from the interaction between polarities. This interaction diagram is generated by comparing the ring-like features identified in an analysis using only the absolute value of the signal with those found in the standard analysis. This approach enables us to detect the presence of δ -spots and quantify the level of interaction between polarities, which is one of the critical factors for the understanding and prediction of flare eruptions.

In conclusion, our application of persistent homology to solar magnetograms has pro-

vided a comprehensive and insightful framework for studying magnetic structures on the solar surface. The topological features derived from magnetograms serve as a foundation for classifying active regions based on their morphology and level of interaction, as certain topological features may have inherent connections to solar atmospheric phenomena. For instance, the presence of interaction rings in active regions might be correlated with flare production, while the interaction of signals from opposite polarities observed in a persistent diagram in the quiet Sun could be linked to small-scale reconnection events or the separation of signals associated with flux emergence. The exploration of these relationships and the assessment of the presented tools in achieving precise active region classification and their potential as predictive tools are topics of our upcoming research. Moreover, we have introduced new tools, such as the interaction diagram, which facilitates the detection and quantification of structures interacting with opposite polarities, like δ -spots, addressing a crucial aspect of flare prediction. The findings presented in this article lay a solid foundation for future studies, emphasizing the potential of persistence images as valuable inputs for machine learning algorithms and contributing to advancements in space weather forecasting.

Lastly, it is important to emphasize that in this study we have focused primarily on static images in order to provide a solid basis for future investigations. The next logical step in this study is to complete the analysis of active regions, which includes examining their temporal evolution. This approach allows for the simultaneous consideration of two key factors in understanding flare eruption processes: morphological complexity, whose analysis is intrinsic to persistent homology, and the study of their temporal evolution through the analysis of the evolution of persistence and interaction diagrams.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The conclusions are ...

APPENDIX A

PROFILE DERIVATIVES

BIBLIOGRAPHY

- Adams, H., Emerson, T., Kirby, M., et al. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research* **18**, 1 (2017).
- Aktas, M. E., Akbas, E., & Fatmaoui, A. E. Persistence homology of networks: methods and applications. *Applied Network Science* **4**, 1 (2019).
- Bailén, F. J., Suárez, D. O., & del Toro Iniesta, J. On fabry–pérot etalon-based instruments. i. the isotropic case. *The Astrophysical Journal Supplement Series* **241**, 9 (2019).
- Bailén, F. J., Suárez, D. O., & del Toro Iniesta, J. On fabry–pérot etalon-based instruments. iv. analytical formulation of telecentric etalons. *The Astrophysical Journal Supplement Series* **254**, 18 (2021).
- Barnes, D., Polanco, L., & Perea, J. A. A comparative study of machine learning methods for persistence diagrams. *Frontiers in Artificial Intelligence* **4**, 681174 (2021).
- Beckers, J. On the effect of narrow-band filters on the diffraction limited resolution of astronomical telescopes. *Astronomy and Astrophysics Supplement Series* **129**, 191 (1998).
- Bellot Rubio, L., & Orozco Suárez, D. Quiet sun magnetic fields: an observational view. *Living Reviews in Solar Physics* **16**, 1 (2019).
- Bianda, M., Solanki, S., & Stenflo, J. Hanle depolarisation in the solar chromosphere. *Astronomy and Astrophysics*, v. 331, p. 760-770 (1998) **331**, 760 (1998).
- Brault, J., & Neckel, H. Spectral atlas of solar absolute disk-averaged and disk-center intensity from 3290 to 12510 Å. *unpublished (tape-copy from KIS IDL library)* (1987).
- Bubenik, P., et al. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16**, 77 (2015).
- Chung, Y., & Lawson, A. Persistence curves: a canonical framework for summarizing persistence diagrams. arxiv. *arXiv preprint arXiv:1904.07768* (2019).
- Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. Stability of persistence diagrams. *Proceedings of the twenty-first annual symposium on Computational geometry*, 263–271 (2005).

- del Toro Iniesta, J. C. (2003), Introduction to spectropolarimetry (Cambridge university press).
- del Toro Iniesta, J. C., & Ruiz Cobo, B. Inversion of the radiative transfer equation for polarized light. *Living Reviews in Solar Physics* **13**, 4 (2016).
- Edelsbrunner, H., & Harer, J. L. (2022), Computational topology: an introduction (American Mathematical Society).
- Gošić, M., Rubio, L. B., Suárez, D. O., Katsukawa, Y., & del Toro Iniesta, J. The solar inter-network. i. contribution to the network magnetic flux. *The Astrophysical Journal* **797**, 49 (2014).
- Hensel, F., Moor, M., & Rieck, B. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence* **4**, 681108 (2021).
- Kosugi, T., Matsuzaki, K., Sakao, T., et al. The hinode (solar-b) mission: An overview. *Solar Physics* **1**, 3 (2007).
- Obayashi, I., Nakamura, T., & Hiraoka, Y. Persistent homology analysis for materials research and persistent homology software: Homcloud. *journal of the physical society of japan* **91**, 091013 (2022).
- Pesnell, W., Thompson, B., & Chamberlin, P. The solar dynamics observatory (sdo). *Solar Physics* **275** (2012).
- Scharmer, G. Comments on the optimization of high resolution fabry-pérot filtergraphs. *Astronomy & Astrophysics* **447**, 1111 (2006).
- Scherrer, P. H., Schou, J., Bush, R., et al. The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo). *Solar Physics* **275**, 207 (2012).
- Schou, J., Scherrer, P. H., Bush, R. I., et al. Design and ground calibration of the helioseismic and magnetic imager (hmi) instrument on the solar dynamics observatory (sdo). *Solar Physics* **275**, 229 (2012).
- Stokes, G. G. On the composition and resolution of streams of polarized light from different sources. *Transactions of the Cambridge Philosophical Society* **9**, 399 (1851).
- Sun, X., Hoeksema, J. T., Liu, Y., et al. Evolution of magnetic field and energy in a major eruptive active region based on sdo/hmi observation. *The Astrophysical Journal* **748**, 77 (2012).
- Toriumi, S., Hayashi, K., & Yokoyama, T. Statistical analysis of the horizontal divergent flow in emerging solar active regions. *The Astrophysical Journal* **794**, 19 (2014).
- Toriumi, S., Iida, Y., Kusano, K., Bamba, Y., & Imada, S. Formation of a flare-productive active region: observation and numerical simulation of noaa ar 11158. *Solar Physics* **289**, 3351 (2014).

- Tsuneta, S., Ichimoto, K., Katsukawa, Y., et al. The solar optical telescope for the hinode mission: an overview. *Solar Physics* **249**, 167 (2008).
- Uitenbroek, H. The accuracy of the center-of-gravity method for measuring velocity and magnetic field strength in the solar photosphere. *The Astrophysical Journal* **592**, 1225 (2003).



- | | | | |
|----|-----------------------|----|----------------------------------|
| 1 | one inch + \hoffset | 2 | one inch + \voffset |
| 3 | \oddsidemargin = -1pt | 4 | \topmargin = -7pt |
| 5 | \headheight = 14pt | 6 | \headsep = 19pt |
| 7 | \textheight = 631pt | 8 | \textwidth = 441pt |
| 9 | \marginparsep = 7pt | 10 | \marginparwidth = 88pt |
| 11 | \footskip = 30pt | | \marginparpush = 7pt (not shown) |
| | \hoffset = 0pt | | \voffset = 0pt |
| | \paperwidth = 597pt | | \paperheight = 845pt |



- | | | | |
|----|-------------------------------------|----|---|
| 1 | one inch + <code>\hoffset</code> | 2 | one inch + <code>\voffset</code> |
| 3 | <code>\evensidemargin = 13pt</code> | 4 | <code>\topmargin = -7pt</code> |
| 5 | <code>\headheight = 14pt</code> | 6 | <code>\headsep = 19pt</code> |
| 7 | <code>\textheight = 631pt</code> | 8 | <code>\textwidth = 441pt</code> |
| 9 | <code>\marginparsep = 7pt</code> | 10 | <code>\marginparwidth = 88pt</code> |
| 11 | <code>\footskip = 30pt</code> | | <code>\marginparpush = 7pt</code> (not shown) |
| | <code>\hoffset = 0pt</code> | | <code>\voffset = 0pt</code> |
| | <code>\paperwidth = 597pt</code> | | <code>\paperheight = 845pt</code> |