Coronal Holes, Solar High Speed Streams and Geomagnetic Storms: Statistical Relationships and Forecasts

Stefan Hofmeister

Master Thesis

Supervisor: Astrid Veronig

submitted on May 26, 2015 at the Institut für Physik, Karl-Franzens-Universität Graz
Acknowledgement

First of all I would like to thank my official supervisor Astrid Veronig and my unofficial supervisor Manuela Temmer for this great topic, for many good ideas, for the links to many interesting papers and for the great possibility of taking part at the European Space Weather Week 2014. I would like thank my dear colleagues Martin Reiss and Thomas Rotter for many interesting insights in second order image statistics, machine learning algorithms and automated forecast algorithms, and that they always remembered me that I have to finally complete my master’s thesis. I would like to thank Veronique Dellouille for my great stay at the Royal Observatory at Belgium and for a very interesting discussion about the comparableness of different datasets, Maximilan Burkard for an interesting discussion about reliability and objectivity in matters of the complex chain of data acquisition, data reduction and data analysis and the Team of the Alpbach Summer School for the interesting background knowledge on satellite missions. Last but not least I would like to thank my friends for remembering me that a life exists beyond this master’s thesis and my family for their support of any kind during my years of study. Now, that the years as student are over, it is time to apply the acquired knowledge to shape the world. Thanks to all of you.
Contents

1. Introduction

2. Sun, solar wind and geomagnetic storms – an overview
   2.1. Magnetohydrodynamics
   2.2. The Sun
       2.2.1. General parameters
       2.2.2. Structure
       2.2.3. Surface phenomena : Overview
       2.2.4. Coronal holes
       2.2.5. The solar cycle and αΩ dynamo
   2.3. Solar wind
       2.3.1. Structure of the solar wind
       2.3.2. Quiet state
       2.3.3. Perturbations
       2.3.4. Solar wind acceleration
   2.4. Interaction of solar wind and magnetosphere
       2.4.1. The Magnetosphere
       2.4.2. Current systems
       2.4.3. Geomagnetic storms
       2.4.4. Impacts on the Earth

3. Solar wind and geomagnetic storm models
   3.1. Solar wind: MHD models
   3.2. Solar wind: Flux tube expansion model
   3.3. Solar wind: Coronal holes in slice based algorithm
   3.4. Geomagnetic storms: Coronal holes in slice based algorithm
3.5. Geomagnetic storms: Empirical model by Burton, McPherron and Russel ............................................. 41

4. Evaluation parameters .................................................. 45
   4.0.1. Root mean square error ........................................ 45
   4.0.2. Forecast correlation coefficient ................................ 45
   4.0.3. Confusion matrix and true skill statistics .................... 47

5. Satellites, observatories, data and data reduction .................. 49
   5.1. Satellites and Observatories ..................................... 49
      5.1.1. SDO ....................................................... 49
      5.1.2. ACE ..................................................... 51
      5.1.3. KSO ..................................................... 53
      5.1.4. Kyoto World Data Center for Geomagnetism ................ 54
   5.2. Datasets .................................................................. 55
      5.2.1. SDO/AIA-193 ............................................... 55
      5.2.2. SDO/HMI-los ............................................... 55
      5.2.3. ACE/SWEPAM .............................................. 56
      5.2.4. ACE/MAG .................................................. 57
      5.2.5. KSO/\(H\alpha\) ............................................... 57
      5.2.6. WDC-C2/Dst ................................................. 57
   5.3. Data reduction ....................................................... 58

6. Extraction of coronal holes .............................................. 59
   6.1. Extraction of coronal holes and filaments in EUV images .... 61
   6.2. Statistics of coronal holes and filaments ......................... 64
      6.2.1. Dataset and analysis ........................................ 64
      6.2.2. Manual classification of coronal holes and filaments ... 65
      6.2.3. Statistics of coronal hole parameters ....................... 68
      6.2.4. Statistics of filaments parameters .......................... 71
      6.2.5. Classification features of coronal holes and filaments ... 72
   6.3. Automatic classification of coronal holes and filaments ....... 87
      6.3.1. Parameters .................................................. 87
      6.3.2. A first classification rule .................................... 90
6.3.3. Supervised classification

7. Correlation of high speed streams and Dst index with coronal holes
   7.1. Dataset
   7.2. Arrival times
   7.3. Solar wind velocity
   7.4. Solar wind density
   7.5. Solar wind magnetic field
   7.6. Dst-index

8. Simulated forecast
   8.1. Forecast tool
      8.1.1. Structure
      8.1.2. Forecast algorithms
   8.2. Forecast results
      8.2.1. Overview
      8.2.2. Evaluation procedure
      8.2.3. Forecast rate of solar high speed streams
      8.2.4. Prediction accuracy of high-speed stream peak velocities
      8.2.5. Accuracy of solar wind magnetic field density
      8.2.6. Accuracy of Dst index

9. Discussion

A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12
1. Introduction

The Sun continuously emanates supersonic particles moving radially away, the so-called solar wind. The time-steady solar wind consists of the slow solar wind with speeds of 250 to 400 km/s and high speed streams with speeds up to 800 km/s. Whereas the origin of the slow solar wind is still uncertain, high speed streams arise from coronal holes, the lowest density regions of the Sun. Coronal mass ejections are short-time events of the solar wind. They arise from massive bursts on the Sun and consist of a cloud of particles up to a total mass of $10 \cdot 10^{12}$ kg which propagates at speeds up to 3000 km/s into the interplanetary space. Earth is protected against solar wind particles by its magnetosphere. The strength of Earth’s magnetic field arises on the one hand from its magnetic dipole moment, on the other hand by the induced magnetic field of reflected solar wind particles. Variations of this induced magnetic field can be measured at the surface and are known as „geomagnetic storms“. The most violent geomagnetic storms arise from coronal mass ejections, but also some medium storms arise from high speed streams.

Studying the solar wind has economic and scientific reasons. It affects the altitude of satellites, which has to be regularly adjusted and can cause breakdowns at times of heavy solar wind bombardment. In addition it can cause power breakdowns at very heavy geomagnetic storms. Although this kind of geomagnetic storms appears statistically once every 100 years, a recent study showed that the rebuild of industry would need at least three month. The scientific case lies in understanding our solar system and the acceleration and interaction processes of supersonic plasma streams with the Earth’s magnetic field.

Although models for acceleration of solar streams and its interaction with Earth’s magnetosphere are already developed, reliable forecasts of solar wind
1. Introduction

parameters and geomagnetic storms are still not existent. This is not a surprise if we think of the complexity of the systems sun, solar wind and Earth’s magnetosphere on each own, even without interactions.

In this thesis we focus on predicting solar wind parameters and the strength of geomagnetic storms by empirical relations, neglecting the complexity of the subsystems. We investigate the correlations of high speed stream proton velocities, proton densities and magnetic field densities, their arrival times at Lagrangian point L1 and the strength of geomagnetic storms determined by the Dst index to the areas, latitude, magnetic field density and open magnetic flux of coronal holes on the Sun.

The thesis is structured as follows. Chapter 2 gives an overview on the Sun, the solar wind and Earth’s magnetosphere. Chapter 3 presents current solar wind models and prediction models of geomagnetic storms. Chapter 4 presents evaluation parameters for prediction models. Chapter 5 presents the datasets used and the data reduction performed. Chapter 6 deals with the extraction of coronal holes in extreme ultraviolet images (EUV) and especially with the discrimination of coronal holes and filaments, which both appear dark in EUV images. Chapter 7 investigates the correlation of solar high speed stream parameters and the strength of geomagnetic storms with characteristic parameters of coronal holes. In Chapter 8 the correlations obtained are applied to a dataset of four years of period from 2011 to 2014 in order to evaluate their use as prediction models. Chapter 9 concludes with a discussion of the results.
2. Sun, solar wind and geomagnetic storms – an overview

The existence of a link between the state of Earth’s magnetosphere and phenomena on the Sun is undoubted and a basic understanding of various processes involved is well established. Nevertheless, modelling these processes often appears to be quite difficult due to the lack of in-situ measurements. In this chapter we present the chain of observed phenomena and involved processes which lead to geomagnetic storms and influence their strength. After a short introduction to magnetohydrodynamics which is commonly used to model processes on the Sun and in interplanetary space, we start with the Sun as the source of the solar wind. Next we discuss acceleration processes of the solar wind plasma and its shape and characteristics in interplanetary space. We conclude this chapter with the solar wind interactions with Earth’s magnetosphere.

2.1. Magnetohydrodynamics

This section refers to the textbooks by Parks (1991) and Foukal (2004).

Magnetohydrodynamic (MHD) is the theory of plasmas in magnetic fields. A plasma is defined as ionized, quasi-neutral gas. Quasi-neutrality holds if the specific length of the system is much larger than the Debye-length, i.e. the shielding length of electric fields of ions, if the number of electrons within the Debye sphere is much larger than 1 and if the specific time scale of the
system is larger than the Langmuir frequency, i.e. the oscillation frequency of electrons around ions.

In principle a plasma can be modelled by the sum of all particles moving in an electric field generated by all other particles and being driven by external forces (kinetic theory). If one sums up all ions of one kind to a fluid, the number of equation reduces to the magnetohydrodynamics equations for each kind (multi-fluid models). If one does not distinguish between the different kind of particles, especially between electrons and ions, and if infinite conductivity is assumed you end up at ideal MHD.

MHD is described by the Navier-Stokes equations of fluids (which contain a continuity equation, a momentum equation, an energy equation and an equation of state) combined with the Maxwell equations. Thus the dynamics is described by the kinematics of the fluid and by the electromagnetic forces of the Maxwell equations. The plasma-β gives an estimate whether kinetic or magnetic effects dominate and is defined as the ratio of the plasma pressure \( p \) to the magnetic pressure \( B^2/2\mu_0 \),

\[
\beta = \frac{2\mu_0 p}{B^2}.
\] (2.1)

In MHD two important effects appear: MHD waves and confinement of magnetic fields with plasma particles. MHD waves divide into magneto-acoustic waves and Alfvén waves and are an important mechanism for energy transport and plasma heating. Confinement restricts particle movement to magnetic field lines. Because we will later refer on confinement, we deduce it briefly: We start with Ohm’s law,

\[
\mathbf{E} + \mathbf{v} \times \mathbf{B} = \sigma \mathbf{j} = \frac{\sigma}{\mu_0} \nabla \times \mathbf{B},
\] (2.2)

where \( \mathbf{E} \) is the electric field, \( \mathbf{B} \) the magnetic field, \( \mathbf{v} \) the velocity field, \( \mathbf{j} \) the current density field, \( \sigma \) the conductivity and \( \mu_0 \) the magnetic permeability. This we apply to Faraday’s law

\[
\nabla \times \mathbf{E} = -\frac{d\mathbf{B}}{dt}
\] (2.3)

and get

\[
\frac{d\mathbf{B}}{dt} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{\sigma}{\mu_0} \nabla^2 \mathbf{B}.
\] (2.4)
The change of magnetic field contains a convection term (left) and a dissipative term (right). If no dissipation occurs, i.e. by infinite conductivity, the magnetic field moves with the particles. Thus a particle at one time on a magnetic field line stays on it for its whole lifetime, i.e. a particle cannot switch from one to another field line, and different field lines cannot cross. Plasma and field are said to be „frozen in“. The plasma-$\beta$ indicates if the particle movement is constrained by the field line or if the field line is dragged by the particles.

The concept of confinement has important consequences to solar physics. Because plasma particles and magnetic field lines are frozen in, we are able to observe magnetic field lines by taking images of the corresponding plasma distributions. If a magnetic field line lies in a quiet environment, it stabilizes the plasma configuration. But if two points of a magnetic field line get close enough, e.g. by twisting, it can reconnect in an energetic more favourable way and cause massive bursts.

The success of MHD depends on its preconditions, i.e. if the system of particles can be considered as a one-fluid plasma. The preconditions of MHD are met quite well in the solar photosphere, however they are not kept strictly at the corona and within the solar wind. Because of the low density, thermodynamic equilibrium probably does not hold for electrons to ions as well as within ions. Thus treating electrons and ions as one species is not allowed and the equation of state - which usually relies on thermodynamic equilibrium - has to be changed. Therefore quantitative results of MHD should be seen as approximation in solar system physics. Qualitative statements on confinement of particles and magnetic fields remain fully valid. They only rely on ionized particles and a high conductivity, but not on thermodynamic equilibrium.

2.2. The Sun

In this section, we present the basic structure, processes and phenomena of the Sun. All phenomena are the result of a chain of underlying processes, thus many phenomena are linked to each other. We ask the reader to pay attention to the links between processes and phenomena, as these are also responsible
2. Sun, solar wind and geomagnetic storms – an overview

for solar wind acceleration. Sections 2.2.1 to 2.2.5 refer to the textbook by Foukal (2004).

2.2.1. General parameters

A spectroscopic comparison of the Sun to other stars classifies the Sun as an ordinary G2V main sequence star. The mass and the distance from the Sun can be calculated by Kepler’s third law

\[
\left(\frac{a_1}{a_2}\right)^3 = \frac{m_{\text{sun}} + m_2}{m_{\text{sun}} + m_1} \left(\frac{P_1}{P_2}\right)^2, \tag{2.5}
\]

saying that the quadratic periods of revolutions \(P_1\) and \(P_2\) of two bodies orbiting the Sun are related to the third power of their semi-major axes \(a_1\) and \(a_2\). The masses \(m_1\) and \(m_2\) can be derived from orbit disturbances, the periods of revolution can be measured directly. By calculating the distance of the two bodies, which was first done by triangulation and later by measurement of the runtime of radar echos sent to other planetary bodies, their distance to the Sun and the mass of the Sun can be calculated. The distance Earth-Sun is 149.6 \(\cdot\) 10^6 km, the mass 1.989 \(\cdot\) 10^{30} kg. By measuring the angular diameter the solar diameter was calculated to 6.959 \(\cdot\) 10^5 km. By measuring the solar constant - the solar flux outside of Earth’s atmosphere, 1367 W/m^2 - the luminosity was calculated to 3.84 \(\cdot\) 10^{26} W. Observing features on the solar surface resulted in a sideric rotation period of about 25 days at the equator, a differential rotation and a solar cycle of about 11 years. Assuming that the whole solar system evolved at the same time, the age of the Sun could be estimated to 4.6 \(\cdot\) 10^9 years by analysing radioactive components of meteorites. The declared values are from Foukal (2004).

2.2.2. Structure

Solar interior

In order to determine the solar interior, models assuming hydrostatic and energetic equilibrium were built to derive the chemical composition and the inner structure of the Sun. We assume that the Sun consists of about 92%
2.2. The Sun

hydrogen, 8% helium and a small amount of heavier elements (Foukal, 2004). By helioseismology and by measuring the neutrino flux resulting from nuclear fusion in the core the models were confirmed and enhanced. Helioseismology observes regular surface patterns in dopplergrams, which are interpreted as steady wave nodes of waves running through the interior of the Sun. The distance of the nodes to each other depends on the wavelength, which depends on depth and density (analogous to seismology).

The interior of the Sun can be divided into the core, the radiation zone, the tachocline and the convection zone.

- Core: The core is the energetic source of the Sun. It contains about 50% of the solar mass (Foukal, 2004) and reaches to about 0.2 to 0.25 solar radii (R⊙; García et al., 2007). At temperatures of 1.56 \times 10^7 K, a pressure of 2.3 \times 10^{16} Pa and a density of 1.5 \times 10^5 kg/m^3 nuclear fusion occurs (Foukal, 2004). Most of the energy is produced in the pp I-chain: 4 hydrogen ions fuse to 1 helium ion, 2 neutrinos, γ-radiation and release 26.731 MeV on energy (Foukal, 2004).

- Radiation zone: The radiation zone extends to about 0.75 R⊙, the temperature is less than 1 \cdot 10^7 K, all atoms are ionized by the high temperature and the γ-radiation (Foukal, 2004). At the same time Compton-scattering at electrons takes place, which reduces the energy of the γ-photons and increases the infrared wavelength. Because of the high density the mean free path of a γ-ray photon is only a few centimetres and the energy transports slow. A single photon needs statistically about 100,000 years to leave the radiation zone (Mitalas and Sills, 1992).

- Tachocline: The tachocline is a very thin region between the radiation and convection zone. While the radiation zone and core rotate as a rigid body, the convection zone rotates differentially. It serves as an overshooting region of convection cells and is thought to be the origin of the solar dynamo.

- Convection zone: The convection zone reaches from the radiation zone to the solar surface. At the boundary to the radiation zone the temperature
2. Sun, solar wind and geomagnetic storms – an overview

is about $2 \cdot 10^6$ K and hydrogen is only partially ionized (Foukal, 2004). The absorption coefficient is thus highly increased because part of the radiation energy is used to re-ionize hydrogen and energy transport by radiation becomes ineffective. Instead plasma is heated, expands and ascends, i.e. convection takes place.

Solar atmosphere

The solar atmosphere is optically thin, thus direct observation is possible. By comparing predicted line profiles of atmospheric models - which include transition rates of highly ionized atoms - with line profiles of measured spectrograms, the structure of the atmosphere was determined. The atmosphere can be divided into the photosphere, the chromosphere, the transition region and the corona.

- Photosphere: The photosphere has a thickness of about 500 km, a density of $10^{-2}$ g/m$^3$ and an effective temperature of 5778 K (Foukal, 2004). Because of the low density the photosphere becomes optically thin, i.e. photons are no longer absorbed efficiently and can escape the Sun. The photosphere is the layer of the Sun that we observe in the visible wavelength range.

- Chromosphere: The chromosphere has a thickness of about 2000 km and a reversed temperature gradient: after decrease in temperature to about 4000 K, it increases again to about 10000 K. The density drops below $10^{-6}$ g/m$^3$ (Foukal, 2004).

- Transition Region: The transition region has a thickness of only a few hundred kilometres and two jumps in temperature increasing the temperature from about $10^4$ K to $10^6$ K. Ionisation starts again. The density drops to $10^{-8}$ g/m$^3$ (Foukal, 2004).

- Corona: The corona has temperature of a few $10^6$ K, the density decreases with distance to the photosphere (Foukal, 2004). Highly ionized atoms exist again. It merges to the interplanetary space.
2.2. Surface phenomena: Overview

Surface textures in the solar atmosphere can be observed at a variety of wavelengths, which correspond to different temperatures and heights. We group the solar atmospheric features into two classes: phenomena related to convection in the convection zone and phenomena related to active regions, which have especially strong magnetic fields.

Phenomena related to convection

- Granulation: The photosphere shows an irregular pattern. Bright areas (granulum) are surrounded by dark lines (intergranulum). Their diameter is about 1000 km, their lifetime 5 to 10 min and the difference in temperature of granulum and intergranulum is 200 K (Foukal, 2004). Granulation is an effect of overshooting convection cells: hot plasma ascents, flows to the edges, cools and descents.

- Supergranulation: In photospheric dopplergrams averaged over a few hours a second photospheric pattern, known as supergranulation, is visible. The cells have a diameter of about 30 to 35 Mm, lifetimes of 20 to 40 hours and a horizontal flow field (Foukal, 2004). The edges coincide with the magnetic network seen in magnetograms.

- Chromospheric network: The chromosphere shows a pattern with cells of a diameter of 20 to 40 Mm (Foukal, 2004). These cells coincide spatially and temporally with the cells of the supergranulation.

- Spiculae, Mottles: At the limb of the Sun bright spiculae with an diameter of about 750 to 1500 km and an extension of about 7500 km appear (Foukal, 2004). They show a mass flow to the corona of 100 times of the solar wind, thus most of the mass has to descend again. Mottles are dark structures on the disk located at the edges of the network. Although a definite assignment of spiculae and mottles is not possible because of their short lifetime, it is thought that they are the same phenomena.
2. Sun, solar wind and geomagnetic storms – an overview

Phenomena related to active regions

- Sunspots: Sunspots are dark regions in the photosphere with a diameter of about 2500 to 100 000 km. They consist of an umbra with a vertical magnetic field of about 3 to 4 kG. The strong field hinders the mass flow from the convection zone and the plasma cools to a temperature of about 4000 K. The umbra is surrounded by a penumbra with a more or less horizontal field of about 1 kG and a temperature 500 K less than the quiet photosphere (Foukal, 2004). Sunspots always consist of at least two spots with different polarity. Their connection line is in east-west direction slightly inclined to the equator. Leading spots of different hemispheres have different polarities, changing their sign with each solar cycle. Sunspots only appear at mid and low latitudes.

- Faculae: Faculae appear as bright points with a typical diameter of 200 km and a high magnetic field of 1 kG (Stix, 2002). They appear around sunspots, but already exist when the sunspot is not visible and still exist when the sunspot vanished.

- Plages: Plages are extended bright areas in the chromosphere. They are associated with faculae in the photosphere and also have an increased temperature. It is assumed that the magnetic flux tubes of faculae expand laterally in the chromosphere because of the lower density and form the extended plages.

- Fibrils: Fibrils are dark, long structures with typical diameter of about 2000 km and a length of 15 000 km (Foukal, 2004). They appear next to plages and sunspots and connect areas of different polarities. They are thought to be magnetic flux loops at low altitudes.

Further phenomena in the corona

- Filaments: Filaments are elongated, dark features in the corona which appear often, but not exclusively next to active regions. The plasma has a temperature of less than 7000 K, but an increased density of a
factor of 100 compared to the surrounding corona (Foukal, 2004). They appear next to magnetic inversion lines and are thought to be stabilized by Lorentz forces.

- Coronal loops: Coronal loops are bright loops connecting opposite polarities, i.e. they are „closed“ magnetic flux tubes. They are the main structuring element of the corona.

- Coronal streamers: Coronal streamers are the farest ranging phenomena over the limb. They are the border of „open“ magnetic field lines, i.e. of field lines which close at large distances to the Sun. Helmet-streamers, bright arcs with a long radial needle, are thought to be a source of the slow solar wind.

- Coronal holes: Since coronal holes are a main part of the thesis, they are described in detail in Section 2.2.4.

2.2.4. Coronal holes

Coronal holes appear as long-living, dark features in EUV and X-Ray images (Figure 2.1). They are the lowest density regions in the corona, with sometimes sharp, sometimes diffuse boundaries. They have rapidly expanding „open“ flux tubes and are thought to be the origin of high speed solar wind streams. At images taken in the He-I line (10 830 Å, originates in the chromosphere) they appear as bright features due to their reduced absorption coefficient. In the photosphere and low chromosphere they are not visible.

While coronal holes are predominantly found at the poles at the minimum of the solar cycle, at the maximum they appear at all latitudes. Many are located at the edges of active regions. At magnetograms of the photosphere below coronal holes single flux tubes with a magnetic field density up to a few $10^3 \text{ G}$ can appear. The average magnetic field has an almost unipolar magnetic field distribution with a mean magnetic field density of about 8 G and a ratio of open to absolute magnetic flux of about 77 $\pm$ 14 % (Wiegelmann and Solanki, 2004). They rotate more like a rigid body in contrast to the underlying photosphere, which leads to permanent magnetic reconnection processes.
2. Sun, solar wind and geomagnetic storms – an overview

Figure 2.1.: Image of the Sun, taken in EUV (193 Å) on 2011/02/28. A large coronal hole (dark) is visible next to the central meridian. From Rotter et al. (2015).

at the boundaries. The abundance of ions matches the abundance in the photosphere. Because the plasma is almost collisionless due to the low density, the Maxwell distribution for each species of ions does not hold. The temperature of electrons was measured to 0.8 MK, of protons to 3 MK. The temperature of O$^{5+}$ is anisotropic with 200 MK in radial direction and 20 MK perpendicular to the radial direction (Cranmer, 2009). The extraordinary high temperature of ions is interpreted as heating by collisionless Alfvén wave damping (cyclotron resonance).

The radiation originating within the boundaries of coronal holes is blue shifted, which corresponds to a continuous mass outflow which carries away both mass and energy. Due to their temporal correlations to high speed streams they are identified as the source of high speed streams. Coronal holes contain plumes, i.e. streamers which have a slightly increased density and decreased
temperature and which are correlated to photospheric faculae, and the darkest of them are surrounded by coronal streamers (Cranmer, 2009; Wiegelmann, Thalmann, and Solanki, 2014).

2.2.5. The solar cycle and $\alpha\Omega$ dynamo

The solar cycle has a period of 11 years in average. At the beginning of a cycle, i.e. solar minimum, sunspots only appear at a latitude of about $\pm 30^\circ$, the number of sunspots show a minimum, coronal holes are mainly found at the Sun’s polar regions and the solar constant is minimal. Within the inclining phase to solar maximum the number of sunspots and the solar constant increase and get maximal, sunspots appear at lower latitudes and coronal holes appear at all latitudes. At the declining phase the number of sunspots and the solar constant decrease and the latitude of sunspots move to the solar equator. The start of the next solar cycle is characterized by the appearance of sunspots at latitudes of $\pm 30^\circ$ with opposite magnetic polarities to the preceding cycle.

The solar cycle is explained by the Sun’s dynamo process. The Sun’s magnetic field cannot be a remnant of its formation, thus it has to be regenerated by transforming mechanical energy into magnetic energy. We assume that the dynamo works in the lower part of the convection zone by twisting and winding magnetic field lines.

We start with a poloidal magnetic field. The differential rotation will wind the magnetic field up, increasing the magnetic flux density and forming a toroidal magnetic field. This is known as $\Omega$-effect. Because a magnetic flux applies a magnetic pressure, the density in flux robes decreases. As soon as the density reaches a certain value, it starts rising. While ascending the flux tube extends due to the lower pressure of surrounding. Due to the extension and the rotation Coriolis force acts, skewing the flux rope to a poloidal direction in a way that the magnetic flux component is contrary to the original poloidal magnetic field and that the direction changes on both hemispheres in an opposite way. Within this model sunspots are explained as ascending flux robes which penetrate the photosphere. This dynamo can explain the change of polarities of sunspots between solar cycles, the change of latitudes of sunspots within one
cycle and why the leading sunspot is nearer to the solar equator. Although the idea of this model is quite accepted, there are nevertheless many problems to be solved.

2.3. Solar wind

Parker (1958) stated that the Sun’s corona would have a non-vanishing pressure at infinite distance if it is in hydrostatic equilibrium. Thus he concluded that the corona expands. The concept of solar wind can be seen as this expansion process and was confirmed by in-situ measurements from several satellite missions (Figure 2.2). It consists of two time-steady kinds of supersonic particle flows which drag the frozen-in magnetic field of the Sun along and shape the interplanetary space: the slow solar wind and high speed streams (Figure 2.3). In addition coronal mass ejections have to be considered as massive disturbances in the steady solar wind flow. In Figure 2.4 a three month time line of the proton velocity, proton density and magnetic field density of the solar wind at Lagrangian point L1 is plotted along with a butterfly diagram of coronal holes and the Dst index, which is an indicator for geomagnetic disturbances.
2.3. Solar wind

In Section 2.3.1 we present the general structure of the solar wind, in Section 2.3.2 we describe the slow solar wind and the shape of the interplanetary magnetic field as quiet state, in Section 2.3.3 we discuss high speed streams and CME’s as perturbations and conclude with a discussion of solar wind acceleration mechanisms in Section 2.3.4.

2.3.1. Structure of the solar wind

The orbiter Ulysses investigated solar wind streams in a plane perpendicular to the solar equatorial plane. The measured solar wind velocities and the polarity of the interplanetary magnetic field is plotted in Figure 2.2. During solar minimum at medium and high latitudes high speed streams dominate the interplanetary space. Slow solar wind streams only appear near the equatorial plane. The polarity of the interplanetary magnetic field is strongly correlated to Sun’s hemispheres. At the subsequent solar maximum slow solar wind
2. Sun, solar wind and geomagnetic storms – an overview

Figure 2.4.: From top to bottom: Solar wind proton velocity, butterfly diagram of coronal holes, proton density, absolute magnetic field density, longitudinal and latitudinal incident angle of magnetic field measured at L1, and Dst-index from 2011/01/01 to 2011/04/01. A typical coronal mass ejection (CME), high speed stream (HSS) and period of slow solar wind (SSW) are marked.
2.3. Solar wind

Streams as well as high speed streams appear at all latitudes and both polarities of the interplanetary magnetic field appear at both hemispheres. At the subsequent solar minimum slow solar wind streams again only appeared next to the equatorial plane and the polarity of the interplanetary magnetic field is reverse to the preceding solar minimum. Therefore the latitudinal distribution of solar wind streams and the polarity of the interplanetary magnetic field is dependent on the solar cycle.

The space probes WIND and ACE investigate solar wind parameters at Lagrangian point L1, therefore one can derive the distribution of solar wind streams in the equatorial plane. Figure 2.3 sketches these results. In the equatorial plane the solar wind consists of slow solar wind streams which are disrupted by high speed streams and coronal mass ejections. At the boundary between slow solar wind streams and high speed streams the plasma is compressed and a shock front forms. The distribution of slow solar wind streams, high speed streams and shock fronts rotate with Sun’s angular velocity around the Sun (Meyer-Vernet, 2007).

2.3.2. Quiet state

Slow solar wind

The slow solar wind is a steady, highly variable radial flow of charged particles at a typical proton velocity of 300 km/s. It contains about 94% hydrogen ions, 4% helium ions, heavier ions and the corresponding free electrons. The proton density is about 7 to 10 cm$^{-3}$, the proton temperature $4 \cdot 10^4$ K (Rucker, 2001). Because of the low density, the different ion species are not expected to be in thermal equilibrium. A proton scatters statistically only about once on its way to the Earth (Meyer-Vernet, 2007). It accelerates to a proton velocity of about 300 km/s within 30 solar radii. The origin of the slow solar wind is still not clear, but it is supposed that it either arises from the boundaries between coronal holes and coronal streamers or from narrow plasma sheets at the tops of streamer cusps (e.g. Helmet streamers). In addition there are evidences that the slow solar wind also arises from small coronal holes at active solar cycles (Cranmer, 2009). The slow solar wind defines the quiet state of the
2. Sun, solar wind and geomagnetic storms – an overview

Figure 2.5.: Heliospheric current sheet (schematic). The tilted and waved shape remembers on a ballerina skirt. From Wikipedia\(^1\).

magnetosphere.

**Magnetic structure**

A particles moving radially away from the Sun always carries the magnetic field along (confinement in MHD). Because of the rotation of the Sun and therefore of the footpoint of the magnetic field line it is winded in a Archimedian spiral. The angle between the field line and the radial direction is given by

\[
\phi = \arctan \left( \frac{\omega R}{v_{\text{plasma}}} \right),
\]

where \( \omega \) is Sun’s rotation period, \( R \) the distance to the Sun and \( v_{\text{plasma}} \) the particle’s velocity (Parker, 1958). For the slow solar wind the angle is about \( 45^\circ \) to \( 55^\circ \) at Earth, at large distances the magnetic field is almost toroidal (Rucker, 2001).

Now we assume a rotating magnetic dipole moment of the Sun. The border between the magnetic hemispheres is field-free, a current sheet forms known as heliospheric current sheet. The magnetic field axis is inclined about \( 7^\circ \) (dependent on the solar cycle) against the rotation axis of the Sun and influences periodically the quiet magnetic field distribution. The Archimedian spiral is undulated resulting in a ballerina-like heliospheric current sheet (Figure 2.5).
In addition we have to take multipole moments arising from the solar surface into account, which complicate the undulation.

Magnetic flux tubes arising from the solar surface expand with the distance to the Sun, filling the available space and transforming into a indistinct magnetic field at large distances. At L1 we see fast changes of the direction of the interplanetary magnetic field. These are interpreted as steep flanks of magnetic flux tubes travelling through the measuring satellite. The granulation cells in Sun’s photosphere are considered as the footpoints of the fluxtubes (Borovsky, 2008).

### 2.3.3. Perturbations

**High speed streams (HSS)**

High speed streams originate in superradially expanding flux tubes within coronal holes and have proton velocities up to 800 km/s. They contain about 95% hydrogen ions, 5% helium ions, heavier ions and free electrons, which corresponds more to the abundances of the photosphere than the abundances of the slow solar wind (Rucker, 2001). At L1 the proton density is decreased to about 3 cm$^{-3}$, the proton temperature increased to about 3 · 10^5 K (Cranmer, 2002). Thermodynamic equilibrium between the species does not exist. High speed streams are accelerated to at least half of their terminal velocity within two solar radii. The single coronal hole flux tubes have magnetic field densities up to 2 kG and arise from the dark lanes between photospheric granulation cells which lie in the bright lanes of the supergranular network (Cranmer, 2009). Because coronal holes are long-lasting phenomena with lifetimes up to a few solar rotations, high speed stream have an impact on the shape of the heliospheric current sheet which cannot be neglected. They also cause some of the medium geomagnetic storms.

2. Sun, solar wind and geomagnetic storms – an overview

Corotating interaction regions (CIRs)

High speed streams of long-lasting coronal holes compress the slow solar wind ahead and form corotating interaction regions. There particles as well as magnetic field lines cumulate and create a forward and backward pressure wave which steepens up to a forward and reverse shock. Corotating interaction regions are tilted in the same way like the heliospheric current sheet and are already well formed at L1. Because of the accumulated density and magnetic flux they strengthen geomagnetic storms caused by high speed streams (Gosling and Pizzo, 1999).

Coronal mass ejections (CMEs)

Coronal mass ejections expel magnetized plasma of about $10^{12}$ kg at speeds up to several $10^3$ km/s into the interplanetary space (Meyer-Vernet, 2007). Due to the small diameter of Earth only a small part of all CMEs hit the Earth. They often appear near sunspots which have a complex magnetic structure and can be the result of erupting filaments. Twisting and shearing of magnetic loops, which is done by plasma motion in the photosphere, where kinetic motion is the dominant factor compared to magnetic energy, build up energy. As soon as two magnetic field lines of different polarity get close enough, they brake up and reconnect in an energetic more favourable way, dissipating the free energy by acceleration of the plasma. The expansion into interplanetary space is self-similar. CMEs causes the strongest geomagnetic storms. At solar minimum about 1 CMEs per week can be observed, in contrast to solar maximum with about 3 CMEs per day (Meyer-Vernet, 2007).

2.3.4. Solar wind acceleration

The issue of solar wind acceleration is still not solved. At the moment we distinguish wave and turbulent driven models from reconnection and loop-opening models, although a superposition of both models is most likely.
2.4. Interaction of solar wind and magnetosphere

Wave/turbulence driven models

In these models, Alfvén waves are generated at the outer boundary of the convection zone and travel to higher heights inside of flux tubes. At a certain height they are reflected and form strong magnetohydrodynamical turbulent cascades. The energy is dissipated on its whole way up, the maximum height is defined by the scale height of dissipation. If the scale height is low, Alfvén waves mainly heat the corona but accelerate more particles within the corona, forming the slow solar wind. If the scale height is high, the energy is mainly transformed into kinetic energy outside of the corona, forming the faster, less dense high speed streams. The scale height depends on the radial dependency of the magnetic flux density and thus on the expansion factor of the flux tube (Cranmer, 2009).

Reconnection/loop-opening models

In these models, microflares arising from magnetic reconnection in the lower atmosphere accelerate the solar wind. They appear as stochastic events, the maximum solar wind velocity depends on the rate magnetic flux emerges and cancels. Thus the boundary of large coronal holes are predestined for microflares due to their rigid rotation. Coronal holes float above the differential rotating plasma of the photosphere, forcing photospheric flux tubes to break up in front of the coronal hole and close again at the back, which corresponds to a permanent magnetic reconnection process (Cranmer, 2009).

2.4. Interaction of solar wind and magnetosphere

Earth’s magnetosphere is the region in which Earth’s magnetic field dominates. Earth’s magnetic field is a superposition of Earth’s dipole field and a magnetic field which is induced by current systems within Earth’s magnetosphere and by solar wind particles. The extension and shape of the magnetosphere is mainly dependent on the solar wind. Fast changes of the shape are known
2. Sun, solar wind and geomagnetic storms – an overview

Figure 2.6.: Structure of Earth’s magnetosphere (schematic). The solar wind arrives from left, the interplanetary magnetic field from top left. The magnetic field lines of Earth’s magnetic field are drawn in red, the magnetopause in turquoise. From NASA.

as geomagnetic storms. We describe the structure of the magnetosphere in section 2.4.1, the main current systems in section 2.4.2, geomagnetic storms as interaction of the magnetosphere with the solar wind in section 2.4.3 and its impact on Earth in section 2.4.4. Section 2.4.1 to 2.4.3 refer to the text book by Rucker (2001).

2.4.1. The Magnetosphere

The magnetosphere is the region in which Earth’s magnetic field is dominant, and is highly variable. Its inner boundary is the ionosphere, its outer boundary the magnetopause, which is defined as the area at which the magnetic pressure of Earth’s magnetic field and all by the solar wind induced magnetic fields equals the kinetic pressure of the incoming solar streams (Figure 2.6). At the magnetopause the normal component of the magnetic field has to be zero. The

2.4. Interaction of solar wind and magnetosphere

required magnetic field at the subsolar point of about 87 nT to withstand the solar wind is mostly induced by the solar wind itself, only 31 nT come from Earth’s dipole field (Rucker, 2001). Because of the frozen-in magnetic field of solar wind streams solar wind particles cannot penetrate Earth’s magnetic field. They are deflected and abrupt decelerated to subsonic speed to form a shock front about 2 \( R_{\text{earth}} \) in front of the magnetopause. Thereby the sun-facing magnetosphere is compressed to about 4.5 to 20 \( R_{\text{earth}} \). The magnetotail - the sun-far side of the magnetosphere, consisting of mainly „open“ field lines - is dragged by the solar wind to distances up to 3000 \( R_{\text{earth}} \) (Rucker, 2001).

At the magnetosheath - the region between shock front and magnetopause - the solar wind particles are thermalized. The particle density and temperature reaches a maximum, the magnetic field is highly variable. Between the sun-faced and sun-far side of the magnetosphere lies the cusp - a region of zero magnetic field, allowing solar wind particles to penetrate.

2.4.2. Current systems

The structure of the magnetosphere is determined by Earth’s dipole field, the parameters of the solar wind and a complex current system. This current system brings a negative feedback to the pressure of the solar wind by magnetic induction and form a reservoir of absorbed energy.

- The Chapman-Ferraro-currents are formed by deflected solar wind particles near the sun-faced magnetopause. Positive and negative charged solar wind particles are deflected in opposite directions, therefore an electric field and a current is induced. The thereby induced magnetic field is opposite directed to the magnetic field of the solar wind and strengthens the magnetic field of Earth.

- The ring-current consists of trapped particles within the magnetosphere. Particles trapped perform simultaneously three kinds of motion: gyration along a field line of Earth’s dipole field, reflection on a field line and circulation in the east-west-plane around the Earth. The area of motion is known as van-Allen belt. By averaging the motion in time gyration
2. Sun, solar wind and geomagnetic storms – an overview

Figure 2.7.: Dst-index from 1950 to 2010, averaged over each week of all years.

and reflection can be disregarded and a ring current results.

- The tail-current is induced by the magnetic field of the magnetotail. The magnetotail consists of „open“ magnetic field lines of both polarities and a neutral current sheet between them. Looking at the cross section of the magnetotail (which looks like the letter Θ), the magnetic fields in the north and south half have different signs and produce reverse currents, adding in the neutral current sheet. These currents can be excited easily and serve as energy cache at high solar wind activity.

- The Birkeland-currents are driven by an electric field, which results from the passing solar wind. The solar wind particles are deflected at the bow shock, the interplanetary magnetic field splits up ions and electrons into different directions. This leads to a positive effective charge passing at the dawn side and a negative electric charge passing at the dusk side and so to an effective electric field. The Birkeland-currents are generally closed in the ionosphere at the poles at a high of about 150 km by the Pedersen-currents. The Pederson-currents once again produce a hall current known as polar electrojet.

2.4.3. Geomagnetic storms

The magnetic field of Earth is a superposition of Earth’s dipole field and the magnetic field induced by the solar wind. While Earth’s dipole field is quite
stable, the induced field is very variable, depending mainly on solar wind pressure and the southward component of the interplanetary magnetic field which drives the energy input into the magnetosphere by magnetic reconnection processes. A southward interplanetary magnetic field component statistically increases the chance of reconnection of interplanetary field lines with field lines of Earth’s dipole field. Variations of Earth’s magnetic field due to strong changes of solar wind parameters are called geomagnetic storms. They are classified in pulsations with periods of $0.2 \, \text{s} \leq T \leq 600 \, \text{s}$ corresponding to the oscillation period of the longest, closed day-side magnetic field line and variations at oscillation periods $T > 600 \, \text{s}$. Variations due to reconfiguration processes within the magnetosphere are called substorms. These especially happen if the positive and negative lobe of the magnetotail reconnects in some distance. A part of by these reconnection processes accelerated plasma particles is accelerated towards the Earth and absorbed by the ring current, causing a geomagnetic substorm. Geomagnetic storms can be measured on Earth’s surface. They usually induce magnetic fluctuations less than 200 nT at a background field of 31 000 nT at the equator (Rucker, 2001).

A geomagnetic storm consists of three phases:

1. Initial stage: A shock front within the solar wind hits the magnetopause and compresses it. The Chapman-Ferraro-currents are strongly increased. The information propagates with the speed of the fast MHD-compression wave and needs about one minute to ground.

2. Main stage: The ring-current increases within 1 to 2 days, the magnetosphere expands, the magnetic disturbances increase. The source of the ring-current particles is not completely clear. A southward orientated interplanetary magnetic field increases the chance for reconnection with Earth’s dipole field and therefore increase the amount of trapped particles in the van-Allen belt. An increased particle flow of the solar wind leads to stronger Birkeland-currents. Reconnection in the magnetotail driven by the increased solar wind leads to plasma acceleration towards the Earth, causing substorms.
2. **Sun, solar wind and geomagnetic storms – an overview**

3. Recovery phase: The ring-current decays slowly, depending on the strength of the geomagnetic storm.

Statistically geomagnetic storms appear most frequently in April and October (Figure 2.7 top) which is explained by the Russel-McPherron-effect (Russell and McPherron, 1973). The incident angle of the interplanetary magnetic field is about $-45^\circ$ in the ecliptic plane, but the orientation of Earth’s magnetic dipole field axis changes its orientation within each day due to rotation and within each year due to revolution. The southward magnetic field component results of projection of the interplanetary magnetic field to the direction of Earth’s dipole field axis, and thus depends on the time of year. Therefore the occurrence rate and strength of geomagnetic storms also depends on the time of day and year.

To classify geomagnetic storms a variety of geomagnetic indices were defined. The common indices Dst, Kp, AE are based on variations of the induced magnetic field measured at the surface of Earth at different latitudes. The Dst index is defined as the average change of magnetic field measured at four low-latitude observatories and is mainly influenced by the ring current. The Kp index is derived by the average change at several mid-latitude observatories. The AE index is derived by high-latitude observatories and is mainly influenced by the auroral electrojet (Gonzalez et al., 1994). In this thesis we use the Dst index.

### 2.4.4. Impacts on the Earth

Geomagnetic storms have manifold effects on Earth. The most common is an increased appearance of polar lights and a degradation of terrestrial radiocommunications. At long overhead power lines strong compensating currents can appear, which can cause a power breakdown. Satellites are affected directly by the current induction due to the magnetic disturbance and indirectly by an increased air drag due to the expansion of the atmosphere by heating it up. Most of the time geomagnetic storms have a low impact on humanity, but the consequences of an extreme strong geomagnetic storm would be devastating nowadays. In 1989 in Quebec an overload of the power
2.4. Interaction of solar wind and magnetosphere

supply system due to a geomagnetic storm resulted in an 9-hour power breakdown, a Dst-index of 589 nT was measured. In 1921 a geomagnetic storm caused compensating currents in the power lanes ten times as strong as in 1989 with an minimum Dst-index of $-825$ to $-900$ nT (Kappenman, 2006). In 1859 a geomagnetic storm was reported with polar lights visible even in Rom, Havanna and Hawai. Today this storm is estimated at an Dst-index of $-1760$ nT (Li et al., 2006; National Research Council of the National Academies, 2008).

We have a quite complex chain of events, i.e. from strong magnetic flux tubes to sunspots to coronal holes and streamers to solar wind streams to coupling with Earth’s magnetosphere, but we do not understand the details of chain links. We think that we know the basics of how the solar dynamo is working, but the details are still unknown. We do not know by what the corona is heated exactly, in which way the solar wind is accelerated and how the solar wind is influenced on its way to Earth. We understand the coupling solar wind - Earth quite well, although also here not all problems are solved in detail. In principle we have a lot of observations, but verified models are still missing.
3. Solar wind and geomagnetic storm models

In this chapter we give an overview over currently existing models which focus on forecasting the solar wind proton velocity and on geomagnetic storms. There are three basic types of physical forecast models: Simulations of the system, empirical models which try to explain the underlying processes and statistical relationships. Section 3.1 deals with MHD models which simulate Sun’s corona and the propagation of the solar wind. Section 3.2 presents an empirical model which explains the solar wind velocity and density as acceleration processes by Alfvén waves in expanding flux tubes originating in Sun’s photosphere. Section 3.3 presents a statistical model which correlates the solar wind velocity with the area coronal holes cover within a meridional slice, and Section 3.4 presents a statistical model which correlates the strength of geomagnetic storms with the area coronal holes cover within a meridional slice. Section 3.5 deals with an empirical model which forecasts the strength of geomagnetic storms when solar wind parameters as input are assumed to be known.

3.1. Solar wind: MHD models

The most powerful way of forecasting is simulating the whole system by fundamental physical equations. The accuracy of simulations depends on how well the boundary conditions are determined, on the used temporal and spatial resolution and on the used set of physical equations, i.e. on how much the physical system was simplified. The used algorithm depends particularly
3. Solar wind and geomagnetic storm models

on the problem which has to be solved. In the following we discuss a model for simulating the heliosphere in 3-D within about 5 AU, which was developed by Linker and Mikić (1997), Mikić et al. (1999) and Riley, Linker, and Mikić (2001).

The concept is based on the resistive MHD equations, which are solved simultaneously in spherical coordinates:

\[ \nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J}, \]  
\[ \frac{1}{c} \frac{\delta \mathbf{B}}{\delta t} = -\nabla \times \mathbf{E}, \]  
\[ \mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} = \eta \mathbf{J}, \]  
\[ \frac{\delta \rho}{\delta t} + \nabla \cdot (\rho \mathbf{v}) = 0, \]  
\[ \rho \left( \frac{\delta \mathbf{v}}{\delta t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = \frac{1}{c} \mathbf{J} \times \mathbf{B} - \nabla p + \rho \mathbf{g} + \nabla \cdot (\nu \rho \nabla \mathbf{v}), \]  
\[ \frac{\delta p}{\delta t} + \nabla \cdot (p \mathbf{v}) = (\gamma - 1)(-p \nabla \cdot \mathbf{v} + S), \]

where \( \mathbf{B} \) is the magnetic field density, \( \mathbf{J} \) the electric current density, \( \mathbf{E} \) the electric field, \( \mathbf{v} \) the plasma velocity, \( \rho \) the plasma mass density, \( p \) the gas pressure, \( \mathbf{g} \) the acceleration due to gravitation, \( \gamma \) the ratio of specific heats, \( \eta \) the plasma resistivity, \( \nu \) the kinematic viscosity and \( S \) the energy source terms.

The heliosphere is split into two regimes: an inner regime up to a distance of 30 \( R_\odot \) and an outer regime up to 5 AU. At the inner region the solar corona is simulated, at the outer the heliosphere. For the corona a polytropic index of \( \gamma = 1.05 \) is assumed, reflecting the slowly varying temperature. As lower boundary conditions synoptic line-of-sight magnetic field maps are used and a uniform plasma density and temperature are presumed. These initial conditions are propagated in time by solving the system of equations until a steady state is reached. Then the derived parameters at the upper boundary of the corona are used as inner boundary conditions of the heliospheric model. The density at the inner boundary is derived under the assumption of momentum flux balance, the temperature by thermal pressure balance. The magnetic flux is gained by magnetic flux conservation and the velocity is estimated by the
3.2. Solar wind: Flux tube expansion model

photospheric flow field transformed to the inner boundary along the coronal magnetic field lines. The polytropic index is increased to $\gamma = 1.5$. Differential rotation is neglected as well as pick-up ions, which decrease the reliability of the simulation to about 5 AU.

Gressl et al., 2014 evaluated this model for a 1-year period during low solar activity and found that the solar wind speed is in general met well with Pearson correlation coefficients of simulated to observed values of $r_P = 0.57$. The magnetic sector structure is also well reproduced, but for the magnetic field density and the temperature too small values are calculated. All distributions of modelled and measured solar wind parameters differ significantly. The predicted arrival times of high speed streams yield typically uncertainties of the order of one day. Thus the algorithms reproduce partially the solar wind conditions well, but it has only a limited applicability for reliable forecasts.

3.2. Solar wind: Flux tube expansion model

Levine, Altschuler, and Harvey (1977) studied five coronal hole appearing at three carrington rotations in 1973 and found that the flux tube expansion factor of coronal holes correlate with the maximum velocity of high speed streams (Figure 3.1). Wang and Sheeley (1990) investigated the expansion factors of flux tubes. They found by potential field source surface magnetic field extrapolations (PFSS) which included the heliospheric current sheet, that flux tubes which expand at least up to 2.5 solar radii have the highest expansion factor at Earth, and vice versa (Figure 3.2). Because solar wind particle densities depend on the cross section of flux tubes due to conservation of mass and because high speed streams have a lower solar wind density than the slow solar wind, they claimed that high speed streams arise from flux tubes that have a high flux tube expansion factor at Earth. Two years later they published a theoretical explanation for the acceleration of high speed streams in flux tubes with a high expansion factor at Earth (Wang and Sheeley, 1991), which is based on solar wind acceleration by Alfvén waves. The theoretical model by Suzuki (2006) described in the following is based on these results.
3. Solar wind and geomagnetic storm models

Consider an open magnetic flux tube anchored at the solar surface. Energy conservation within this flux tube results in

\[ \nabla \left[ \rho v \left( \frac{v^2}{2} + \frac{\gamma}{\gamma - 1}RT - \frac{GM_\odot}{r} \right) - \frac{1}{4\pi} (v \times B) \times B + F_c \right] + q_R = 0, \quad (3.7) \]

where \( T \) is the temperature, \( M_\odot \) the mass of the Sun, \( F_c \) the conductive flux, \( q_R \) the radiative cooling, \( r \) the distance from the Sun, \( R \) the gas constant, and \( G \) the gravitational constant. The first term corresponds to kinetic energy, the second to thermal energy, the third to gravitational energy, the fourth to the Poynting flux, the fifth to conductive flux and the sixth to radiative cooling.

In the next step we evaluate Eq. 3.7 with the boundary conditions \( r_0 = 1R_\odot \) and \( r_1 = 1 \text{ AU} \). We assume the cross section of the flux tube to expand with
3.2. Solar wind: Flux tube expansion model

Figure 3.2.: Flux tube expansion factors versus the radial distance in solar radii, extrapolated for a photospheric distribution $B_r(R_\odot, \Theta) = \cos^7 \Theta$, a current-free source-surface at 2.5 solar radii (dotted curves) and a source-surface at 2.5 solar radii which respects the heliospheric current sheet (solid curves) for footpoint colatitudes $\Theta$ of 0° (defining the polar direction), 20°, 27° and 31°. From Wang and Sheeley (1990).
3. Solar wind and geomagnetic storm models

Figure 3.3.: Maximum solar wind speed versus flux tube expansion factors at 1 AU. The solid curve represents the flux tube expansion model for a coronal temperature of 1 MK and \( \langle \delta B_{\perp} \delta v_{\perp} \rangle = 8.3 \times 10^5 \text{ G cm s}^{-1} \), the dotted-dashed line for \( T = 1.5 \text{ MK} \) and \( \langle \delta B_{\perp} \delta v_{\perp} \rangle = 8.3 \times 10^5 \text{ G cm s}^{-1} \) and the dashed line for \( T = 1 \text{ MK} \) and \( \langle \delta B_{\perp} \delta v_{\perp} \rangle = 5.3 \times 10^5 \text{ G cm s}^{-1} \). Observed data are from Kojima et al. (2004). From Suzuki (2006).

A super-radial function \( r^2 f(r) \). Then the divergence of an arbitrary vector \( \mathbf{A} \) can be written as

\[
\nabla \cdot \mathbf{A} = f^{-1} r^{-2} \frac{d(f r^2 A_r)}{dr}.
\]

(3.8)

We split the Poynting flux term into two terms,

\[
- \frac{1}{4\pi} (\mathbf{v} \times \mathbf{B}) \times \mathbf{B} = \frac{1}{4\pi} (-B_r \delta B_{\perp} \delta v_{\perp} + v_r \delta B_{\perp}^2),
\]

(3.9)

where the subscript \( r \) denotes the radial and \( \perp \) the tangential components.

The first term corresponds to the shear of the magnetic field (Alfvén waves), the second term to advection of magnetic energy. If we further combine the compressible parts of the energy input by convection into a heating term,

\[
F_H = \frac{v_r \delta B_{\perp}^2}{4\pi} + \rho v^2 \frac{v^2}{2} + \rho v \frac{\gamma}{\gamma - 1} RT,
\]

(3.10)
and assume that the kinetic energy is the dominant part at 1 AU we get

\[
\rho v_r^2 f_{\text{total}} \frac{v^2}{2} \bigg|_{r=1 \text{ AU}} = \left[ \frac{r^2}{4 \pi} \left( -\frac{B_t \langle \delta B_\perp \delta v_\perp \rangle}{r} + F_H - \rho v \frac{GM}{r} \right) \right] \bigg|_{r=R_\odot} - \int_{R_\odot}^{1 \text{ AU}} dr \ r^2 f_{qR}, \tag{3.11}
\]

where angle brackets denote time averages. Dissipation of Alfvén waves is slow, thus they propagate a long way and contribute to heating and acceleration of the solar wind. In contrast, compressible waves and turbulences summarized in \( F_H \) contribute to the heating of the chromosphere and the low corona. Thus we can combine the heating and radiative loss term to an effective temperature \( T_C \) of the corona and finally get

\[
\rho v_r^2 f_{\text{total}} \frac{v^2}{2} \bigg|_{r=1 \text{ AU}} = \left\{ \frac{r^2}{4 \pi} \left( -\frac{B_t \langle \delta B_\perp \delta v_\perp \rangle}{r} + \rho v \left( \frac{\gamma}{\gamma - 1} RT_C - \frac{GM}{r} \right) \right) \right\} \bigg|_{r=R_\odot}. \tag{3.12}
\]

By rearranging the terms we can calculate the solar wind speed at 1 AU as (Suzuki, 2006)

\[
v_{1 \text{ AU}} = \left\{ 2 \left[ -\frac{R_\odot^2}{4 \pi \rho v_r^2} \frac{B_t \langle \delta B_\perp \delta v_\perp \rangle}{f_{\text{total}} \langle \delta B_\perp \delta v_\perp \rangle} + \frac{\gamma}{\gamma - 1} \frac{RT_C}{R_\odot} - \frac{GM}{R_\odot} \right] \right\}^{0.5}. \tag{3.13}
\]

Eq. 3.13 contains the required anticorrelation of solar wind speed at Earth and flux tube expansion factor at Earth which leads to low densities at high velocities and explicitly describes solar wind acceleration by Alfvén waves.

The prediction capability of Eq. 3.13 for predicting the solar wind speed at 1 AU is plotted in Figure 3.3. There a constant mass flux at 1 AU of \( \langle \rho v \rangle_{1 \text{ AU}} = 5.4 \cdot 10^{-16} \text{ g cm}^{-2} \), a velocity amplitude at the surface of \( \delta v_\perp = 1 \text{ km/s} \), a tangential magnetic flux density of \( B_\perp = 110 \text{ G} \), a corrected value for \( \langle \delta B_\perp \delta v_\perp \rangle = 8.3 \cdot 10^5 \text{ G cm s}^{-1} \) due to the reflection of Alfvén waves in the chromosphere, a coronal temperature of \( 1 \cdot 10^6 \text{ K} \) and an increased adiabatic value of \( \gamma = 1.1 \) due to thermal conduction is assumed. The modelled values match the statistical trend of the measured values well, but have a high scatter.
3. Solar wind and geomagnetic storm models

Figure 3.4.: Transition of a coronal hole in March, 2005 measured by GOES-SXI (a). The meridional slice of [-10°, 10°] and the contour of the coronal hole within the slice are outlined in white. The corresponding fractional area of the coronal hole within the slice versus the day of year is plotted in (b), the daily averaged solar wind velocity measured by ACE at Lagragian point L1 versus the day of year in (c). From Vršnak, Temmer, and Veronig (2007a).

3.3. Solar wind: Coronal holes in slice based algorithm

Wang and Sheeley (1990) claimed that we can expect a relationship between the flux tube expansion factor and the area of coronal holes. Flux tubes expand to larger distances to the Sun in order to fill the available space. A flux tube in the middle of a coronal hole will expand less up to 2.5 solar radii than a flux tube at the border of a coronal hole because it is surrounded by expanding other flux tubes. The flux tube expansion factor depends on the amount of flux tubes surrounding a flux tube, and therefore on the area of the coronal hole. Because the solar wind speed depends on the flux tube expansion factor, it should also depend on the area of coronal holes.

Vršnak, Temmer, and Veronig (2007a) compared the solar wind speed with
3.3. Solar wind: Coronal holes in slice based algorithm

Figure 3.5.: Measured solar wind speed at ACE versus the fractional area coronal holes cover within a meridional slice of [-10°, 10°] for day 25 to 125 in 2005 (top). Observed and predicted solar wind speeds at L1 for a considered east (E, [-40°, -20°]), mid (M, [-10°, 10°]) and west (W, [20°, 40°]) meridional slice (bottom). From Vršnak, Temmer, and Veronig (2007a).
3. Solar wind and geomagnetic storm models

Figure 3.6.: Fractional area coronal holes cover within a slice of [-7.5°, 7.5°] around the central meridian versus solar wind proton velocity, at a fixed time delay of 4.29 days (top). The bounding box and the prediction function of the slice based algorithm are plotted in red. Observed versus forecasted solar wind proton velocities for 2011/01/01 to 2013/12/31. From Rotter et al. [2015].
3.3. Solar wind: Coronal holes in slice based algorithm

the fractional areas coronal holes cover within meridional slices of $[-40^\circ, -20^\circ]$, $[-10^\circ, -10^\circ]$ and $[20^\circ, 40^\circ]$ for day 25 to 125 in year 2005, a period where three large low to mid latitude coronal holes separated by a longitudinal angle of about $120^\circ$ were apparent and where almost no CMEs occurred. These three coronal holes caused a regular pattern of eleven high speed streams within the 100 days. They found that the fractional area $A_{\text{slice}}$ coronal holes cover within the slices represented well the pattern of the solar wind speed $v$ measured at L1 at delays of 6, 4 respectively 2 days (Figure 3.4). In order to predict the solar wind speeds, they used the linear function

$$v(t) \text{[km/s]} = 350 + 900 \cdot A_{\text{slice}}(t^*)$$

(3.14)

with $t - t^* = 2, 4, 6$ days (Figure 3.5). The mean root square deviation of predicted to observed velocities is 76 km/s. Note that the slope used for prediction is more than double than the slope of the linear least-square fit of the velocity - fractional area distribution (Figure 3.5), and that these results were evaluated only for day 25 to 125 in 2005, a time where almost no CMEs occurred.

On basis of these results Rotter et al. (2012) developed a self-adapting forecast algorithm. To predict a solar wind velocity at time $t$ a dataset containing the time series of solar wind velocities and fractional areas of $[t - 3 \text{ month}, t]$ is created. All velocities greater than 800 km/s are excluded as they most probable belong to CMEs. The time delay $\Delta t$ between the transit time of coronal holes through the central meridian and the arrival time of the high speed stream is estimated by searching for the highest correlation coefficient between the two time series. The pairs of proton velocities $v(t_n)$ and fractional areas $A_{\text{slice}}(t_n - \Delta t)$ are plotted (Figure 3.6 top). At that plot a bounding box is searched for. The diagonal of the bounding box is used as empirical relationship between the proton speed at L1 and the fractional area coronal holes cover within the meridional slice.

Although it does not seem to be reasonable to get an empirical relationship out of this wide spread data, shown in the top panel of Figure 3.6, the results for predicted velocity peaks are reasonable. The accuracy of predicted high speed stream velocity peaks is plotted in the bottom panel of Figure 3.6, the Pearson correlation coefficient of predicted and observed peak velocities is 0.77
3. Solar wind and geomagnetic storm models

for 2011, 0.72 for 2012 and 0.40 for 2013. This indicates that either most of the peak velocities are next to the diagonal on the top plot or that the self-adapting algorithm covers another dependency.

3.4. Geomagnetic storms: Coronal holes in slice based algorithm

Because of the good results at the prediction of solar wind speeds by utilizing the fractional area coronal holes cover within a meridional slice Vršnak, Temmer, and Veronig (2007b) also correlated the Dst value with the fractional area coronal holes cover within a slice of [-10, 10] degrees around the central meridian of the Sun. They used a similar dataset to the solar wind speed, i.e. the fractional area coronal holes cover within the slice and Dst values between day 25 and 125 of year 2005.

Their comparison of fractional areas coronal holes cover within the slice to the daily averaged Dst values (Figure 3.7) led to the empirical equation

\[ Dst(t) = -65 \cdot [A_{\text{slice}}(t^*)]^{0.5}, \quad (3.15) \]

with the fractional area \( A_{\text{slice}} \) and \( t - t^* = 4 \) days. However coronal holes of negative polarities were underestimated and positive coronal holes overestimated. Due to the Russel-McPherron effect coronal holes of negative polarities have a southward \( B_z \) component in GSM coordinates at spring equinox which increases the Dst drop. Taking the Russel-McPherron effect into account, their empirical relationship changed to

\[ Dst = (-65 \pm 25 \cos \lambda) \cdot [A_{\text{slice}}(t^*)]^{0.5}. \quad (3.16) \]

\( \lambda \) is the ecliptic longitude, the plus sign applies for coronal holes of positive polarity, the minus sign for negative polarity.

The deviations of calculated to measured Dst drops were always less than 30 % of the measured Dst values.
3.5. Geomagnetic storms: Empirical model by Burton, McPherron and Russel

Burton, McPherron, and Russell (1975) found an empirical relationship relating the Dst index to solar wind parameters at 1 AU by simply utilising energy

www.antarctica.ac.uk/SatelliteRisks/
3. Solar wind and geomagnetic storm models

Figure 3.8.: 12-hour forecast of Dst-Index by a variation of the Burton-Russell-McPherron equation (blue), and observed Dst-index (black). From Antarctica\(^1\).

conservation and the definition of the Dst index. Their relationship found was improved by O’Brien and McPherron (2000) and is still used as well-working „nowcast“ algorithm.

The Dst index is an indicator of the current state of the magnetosphere. It is defined as average value of the magnetic flux density $H$ measured at several mid-latitude stations,

$$
Dst = \frac{1}{N} \sum_{n=1}^{N} H_n. \tag{3.17}
$$

Changes of $Dst$ originate in changes of currents in the magnetopause (index $m$) and ring currents (index $r$) which induce magnetic fields $H_m$ and $H_r$. Thus the $Dst$ can be written as

$$
Dst = -(H_m^q + H_r^q) + (H_m^d + H_r^d), \tag{3.18}
$$

where $q$ refers to the quiet state and $d$ the disturbed state. The quiet state can simply be summed up to a constant $c$. The magnetopause currents are proportional to the square root of the dynamic pressure $P_d$ of the solar wind, resulting in

$$
Dst = H_r^d + b\sqrt{P_d} - c. \tag{3.19}
$$

At quiet times the decay of an already excited ring current is proportional to its strength. The energy input into the ring current can be approximated by
3.5. Geomagnetic storms: Empirical model by Burton, McPherron and Russell

a function of the interplanetary electric field $F(E)$. Thus we can write

$$\frac{d}{dt} H^d = F(E) - a H^d,$$  \hspace{1cm} (3.20)

where $F(E)$ is defined as

$$F(E) = \begin{cases} 
  d(E - E_c) & E > E_c \\
  0 & E < E_c 
\end{cases},$$  \hspace{1cm} (3.21)

where $E_c$ is a critical value which has to be determined. Altogether we get for the change of $Dst$

$$\frac{d}{dt} Dst = F(E) - a(Dst - b\sqrt{P_d} + c) + \frac{d}{dt}(b\sqrt{P_d}).$$  \hspace{1cm} (3.22)

Burton, McPherron, and Russell (1975) determined the constants to $a = 1.296 \ h^{-1}$, $b = 16 \ nT \ (nPa)^{-0.5}$, $c = 20 \ nT \ (nPa)^{-0.5}$, $d = -5.4 \ nT \ m \ (h \ mV)^{-1}$ and $E_c = 0.5 \ mV \ m^{-1}$ (Burton, McPherron, and Russell, 1975).

O’Brien and McPherron (2000) suggested that the decay time $\tau = 1/a$ is a function on the interplanetary electric field due to charge exchange losses and that the interplanetary electric field is purely induced by the southward component of the interplanetary magnetic field $B_s$ in GSM coordinates and the solar wind velocity $v$. Then the equations change to

$$\frac{d}{dt} H^d = F(E) - \frac{H^d}{\tau}$$  \hspace{1cm} (3.23)

$$\tau(h) = 2.40 \cdot e^{9.74/(4.69 + vB_s)}$$  \hspace{1cm} (3.24)

$$E = vB_s = \begin{cases} 
  |vB_s| & B_s < 0 \\
  0 & B_s \geq 0 
\end{cases}.$$  \hspace{1cm} (3.25)

They recalculated the parameters $b = 7.26 \ nT \ (nPa)^{-0.5}$, $c = 11 \ nT \ (nPa)^{-0.5}$, $d = -4.4 \ nT \ m \ (h \ mV)^{-1}$ and $E_c = 0.49 \ mV \ m^{-1}$.

Although this empirical model is quite old in the sense of that we nowadays understand the underlying magnetospheric processes much better, it is still used in several variations because of its good predictive power. Nowadays many other modern models exist – especially neural network models – with a high potential of forecasting. But all the models have one restriction in
3. Solar wind and geomagnetic storm models

common: they require to know the solar wind parameters near Earth, which results in a lead time of only a few hours.

Note that this empirical model is build on energy conservation and does not claim from where the sources exactly come from. Vasyliunas (2006) showed that the source and decay terms can also be interpreted as reconnection processes. Therefore this model is still in agreement with ongoing results of research.
4. Evaluation parameters

In this chapter we present the evaluation parameters we use in this thesis. In order to evaluate a forecast, we have to evaluate the prediction accuracy and the prediction rate. For that purpose we present the root mean square error, the forecast correlation coefficient, the confusion matrix and the true skill statistics.

4.0.1. Root mean square error

We define a dataset $X$ with observed values $x_i$, a dataset $Y$ with predicted values $y_i$, the deviations $e_i = y_i - x_i$ and denote the number of elements in each dataset as $n$.

The root mean square error (RMSE) is defined by

$$RMSE = \sqrt{\frac{1}{n} \sum_{i} e_i^2}. \quad (4.1)$$

At an Gaussian distribution of deviations 68% of the data lie within 1 $RMSE$ and 95% lie within 2 $RMSE$. Since most of the time errors are more Gaussian than uniform, the RMSE should be preferred to the also often used mean absolute error (Chai and Draxler, 2014).

4.0.2. Forecast correlation coefficient

The Pearson correlation coefficient is defined by

$$r_P = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}. \quad (4.2)$$

The Pearson correlation coefficient correlates the two datasets by looking at the deviations of each dataset from its mean value. If the deviations are
4. Evaluation parameters

Figure 4.1.: Correlation coefficient of slope $r_{\Delta \beta}$ versus slope of the regression line $\beta$.

proportional to each other, the two datasets are linearly dependent and we get a correlation coefficient of $\pm 1$ (Khattar, 2008).

At evaluating a forecast we are not interested if the predicted and observed values are linear dependent, but if they are equal. Because we were not able to find a correlation coefficient which satisfies this constraint, we define one on our own.

The slope $\hat{\beta}$ and the angle of the slope $\beta$ of the linear regression is given by

$$\hat{\beta} = r_{xy} \frac{S_x}{S_y} \quad \hat{\beta} \in \mathbb{R}_\infty,$$

$$\beta = \arctan \hat{\beta} \quad \beta \in \left[-\frac{\pi}{4}, \frac{3\pi}{4}\right], \quad (4.4)$$

where $S_x$ and $S_y$ are the standard deviations of $x$ and $y$. Note that this domain of $\beta$ covers all possible slopes of the regression line, i.e. we do not lose generality. By choosing $\hat{\beta}$ to be in the affinely extended real numbers $\mathbb{R}_\infty$ we avoid problems at $\beta = \pi/2$.

The deviation of the slope to a slope of 45°, which correspond to a one-to-one correspondence, is

$$\Delta \beta = \left| \frac{\pi}{4} - \beta \right| \quad \Delta \beta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \quad (4.5)$$

A measure on the deviations is then

$$r_{\Delta \beta} = 1 - \frac{\Delta \beta}{\frac{\pi}{4}} = 1 - \left| 1 - \frac{4}{\pi} \beta \right|. \quad (4.6)$$
\[
\begin{array}{c|cc}
\text{Real} & \text{Predicted} & \text{Positive} & \text{Negative} \\
\text{true} & \text{TP} & \text{TN} \\
\text{false} & \text{FP} & \text{FN} \\
\end{array}
\]

Figure 4.2.: Confusion matrix (left) and true positive rate vs. false positive rate (right, from Reiss et al., 2014b).

\( r_{\Delta\beta} \) is 1 for \( \beta = 45^\circ \), -1 for \(-45^\circ \) and 0 for \( 0^\circ \) and \( 90^\circ \) (Figure 4.1).

We now define the "forecast correlation coefficient" by

\[
 r_{fc} = |r_P| \cdot \left( 1 - \left| 1 - \frac{4}{\pi} \beta \right| \right). 
\]

This correlation coefficient is 1 for a perfect correlation with a slope of \( 45^\circ \), -1 for a perfect anti-correlation with a slope of \(-45^\circ \), 0 for a slope of \( 0^\circ \) and \( 90^\circ \) and respects both the spreading (by the Pearson correlation coefficient) and the slope.

Note that a fixed offset between predicted and observed values does not change the coefficient. This could be taken into account by changing \( \bar{y} \) to \( \bar{x} \) at Pearson’s correlation coefficient, but is here neglected. A fixed offset is easy to compensate in forecast algorithms, considering it would complicate the interpretation of the forecast correlation coefficient.

\[4.0.3. \text{Confusion matrix and true skill statistics}\]

The evaluation of the prediction rate is usually treated as a dichotomous binary classification problem, i.e. if a species of event is predicted or not and if it incidents or not. The raw counts can be summarized in a four-cell contingency table, i.e. the confusion matrix (Figure 4.2 left). We define true and false positive (TP / FP) as events which are predicted and which occur / do not
4. Evaluation parameters

occur, and true and false negative (TN / FN) as events which are not predicted and which do not occur / occur. We define the prediction rates $tpr$, $tnr$, $fpr$, $fnr$ as

\[
tpr = \frac{TP}{TP + TN} \quad (4.8)
\]
\[
tnr = \frac{TN}{TP + TN} \quad (4.9)
\]
\[
fpr = \frac{FP}{FP + FN} \quad (4.10)
\]
\[
fnr = \frac{FN}{FP + FN}. \quad (4.11)
\]

Because $tpr + tnr = 1$ and $fpr + fnr = 1$, only two of the four prediction rate parameters are needed to determine distinctly the prediction rate.

Therefore the prediction rates can be graphically presented in a $tpr$ - $fpr$ plot. (Figure 4.2 right). The higher the $tpr$ and the lower the $fpr$, the better the prediction rate.

The True Skill Statistics (TSS), also known as Informedness, combines the two prediction rate parameters to a single parameter:

\[
TSS = tpr - fpr. \quad (4.12)
\]

This parameter ranges from -1 to 1, where -1 declares that all events were incorrectly predicted, 0 that we are not at all able to predict events and 1 that all events are correctly predicted (Powers, 2007).
5. Satellites, observatories, data and data reduction

In this chapter we present the satellites and observatories which provide the data used in this study, basic information on the data processing, the datasets we used and data reduction. The datasets used are EUV images of the corona (SDO/AIA-193), line-of-sight magnetograms of the photosphere (SDO/HMI-lsos), Hα images of the chromosphere (KSO/Hα), solar wind proton velocity, density, temperature and magnetic field density measured at Lagrangian point L1 (ACE) and the Dst index derived from four low-latitude stations along Earth’s equator (WDC-C2/Dst).

5.1. Satellites and Observatories

5.1.1. SDO

In the frame of NASA’s ‘Living with a star’ program the Solar Dynamics Observatory (SDO; Pesnell, Thompson, and Chamberlin, 2012) was launched on February 11, 2010 to investigate the following scientific questions:

- How is the 11-year solar cycle working?
- How is magnetic flux in active regions created, concentrated and distributed on the surface?
- How does small-scale magnetic reconnection influence the large-scale topography of magnetic fields, how important is it for coronal heating and solar wind acceleration?
5. Satellites, observatories, data and data reduction

Figure 5.1.: SDO (left) and ACE (right) schematic. From Caltech\(^1\) and NASA\(^2\).

- Where do variations of the EUV irradiance come from and are they related to the solar cycle?
- What magnetic field configurations lead to CMEs and flares?
- Is it possible to determine the structure of the solar wind at Earth by the magnetic field configuration and the atmospheric structure near the solar surface?
- Is it possible to make reliable space weather and space climate forecasts?

To give answers to these questions, SDO was equipped with three scientific instruments:

- The Atmospheric Imaging Assembly (AIA, Lemen et al., 2012) are four 20 cm-telescopes with 4k x 4k pixel CCDs observing the Sun at 7 wavelengths in EUV (304 Å, 171 Å, 193 Å, 211 Å, 335 Å, 94 Å, 131 Å) and 3 in UV/visible (white light, 1700 Å, 1600 Å). These wavelengths correspond to different ion transitions at different temperatures and different height above the solar surface. Images are made every 12 seconds at an resolution of 1.5 arcsec.

- The Helioseismic and Magnetic Imager (HMI, Hoeksema et al., 2014) is a magnetometer observing at the Fe I line (6173 Å) with an resolution

http://www.srl.caltech.edu/ACE/Gallery/gallery.html
sdo.gsfc.nasa.gov/assets/img/site/spacraft.zip
of 1 arcsec per pixel. It consists of two camera systems. The Doppler camera calculates dopplergrams and line-of-sight magnetograms at an cadence of 45 seconds by measuring the Doppler shift and Zeeman effect at two polarization states, whereas the Vector camera calculates vector magnetograms and line-of-sight magnetograms at an cadence of 720 seconds by measuring four polarization states of the light and calculating the Stoke’s parameters.

- The Extreme Ultraviolet Variability Experiment (EVE) is a multiple grating spectrograph to measure the EUV irradiance of 0.1 to 100 nm at an resolution of 0.1 nm and an cadence of 10 seconds.

In order to be able to transmit the enormous amount of data to Earth, SDO was launched into an inclined geosynchronous orbit. An K-Band-antenna combined with the geosynchronous orbit allows a continuous downlink of 130 megabits per second, which results in the requested 1.5 terabyte of scientific data per day. Its dimensions are about 4.5 m × 2.2 m × 2.2 m and it weights about 3100 kg, from which 290 kg are payload and 1450 kg are fuel (Figure 5.1). This amount of fuel will enable SDO to remain in its orbit over the nominal mission life time of 5 years up to 10 years. The solar panels of 6.6 m² produce about 1450 W of power. The mission costs about 856 million dollar, the first 5 years of operation inclusive (Pesnell, Thompson, and Chamberlin, 2012).

5.1.2. ACE

The Advanced Composition Explorer (ACE; Stone et al., 1998) was launched on August 25, 1997 to investigate the composition and evolution of the solar system, the interstellar medium and the galaxy as well as the acceleration processes of plasma particles. Particles from the Sun are accelerated by the solar wind, flares and CME’s, whereas matter from the local interstellar medium is thought to be accelerated at the solar wind termination shock and galactic cosmic rays by supernovae shock waves. Particles of the local interstellar medium sample the present date, galactic cosmic rays were accelerated $10^7$ years ago and solar matter stored in the Sun represents the composition of
5. Satellites, observatories, data and data reduction

matter of $4.6 \cdot 10^9$ years ago. Each of the three species are located in different energy regimes, and can thus be measured independently. These data will give important input to theoretical models of the evolution of the universe.

The scientific aims of the ACE mission focus on:

- the elemental and isotopic composition of matter.
- the origin of the elements.
- the evolutionary processes.
- the formation of the solar corona.
- the acceleration of the solar wind.
- the particle acceleration and transport in nature.

ACE is equipped with nine instruments to achieve its mission goals:

- The Cosmic Ray Isotope Spectrometer (CIRS) measures the composition of galactic cosmic rays in an energy range of 100 to 600 MeV/nucleon.
- The Solar Isotope Spectrometer (SIS) measures the composition of particles accelerated at large solar events to low energy galactic cosmic rays at an energy range of 10 to 100 MeV/nucleon.
- The Ultra Low Energy Isotope Spectrometer (ULEIS) measures the composition of solar wind particles to local interstellar particles in an energy range of 0.02 to 10 MeV/nucleon.
- The Solar Energetic Particle Ionic Charge Analyser (SEPICA) measures the charge state and kinetic energy from 0.2 to 3 MeV/nucleon.
- The Electron, Proton and Alpha Monitor (EPAM) characterizes the dynamic behaviour of matter accelerated by flares and interplanetary shocks in an energy range of 0.03 to 5 MeV/nucleon.
- The Solar Wind Ion Mass Spectrometer (SWIMS) is intended to measure the isotopic composition of all solar wind states at a resolution of $M/dM > 100$. 
5.1. Satellites and Observatories

- The Solar Wind Ion Composition Spectrometer (SWICS) measures the elemental and ionic charge state of all solar wind states.

- The Solar Wind Electron, Proton and Alpha Monitor (SWEPAM) measures the three-dimensional behaviour of electrons from 1 to 900 eV and ions from 0.26 to 35 keV.

- The Magnetometer (MAG) is a twin triaxial flux gate magnetometer to measure the behaviour of vector magnetic fields from 0.001 to 65 536 nT.

ACE is situated at Lagrangian point L1 which is located about $1.6 \cdot 10^6$ km ahead of Earth in the Sun-Earth line in order to not be affected by Earth’s magnetosphere. Its shape is a cylinder of an diameter of 2 m and a length of 1.9 m, with extended magnetometer booms it has a wing span of 8.3 m (Figure 5.1). The weight at launch was 785 kg, whereof 152 kg were payload and 189 kg were fuel. To optimize fuel consumption ACE is spin-stabilized at 5 rpm. This allows ACE to operate at L1 until 2024. In 2011 the costs were announced at 107 million dollar plus 4 million dollar per year (NASA, 2012; Stone et al., 1998).

5.1.3. KSO

Kanzelhöhe Observatory for Solar and Environmental research of the University of Graz (KSO; Pötzi et al., 2015) was founded in 1943. It lies at the mountain „Gerlitzen“ in Austria at a height of 1526 m above sea level. It’s main purposes are the observation of the Sun and climate monitoring, especially the

- physics of flares,

- activity cycle of the Sun,

- transfer of solar radiation through Earth’s atmosphere,

- natural and anthropogenic effects on temperature, humidity and ozone balance.
5. Satellites, observatories, data and data reduction

The observatory is equipped with a

- refractor for sunspot drawings,
- photospheric camera (PhoKa) for continuum observations,
- $\text{H}\alpha$ telescope for observations of the chromosphere,
- $\text{Ca II K}$ telescope for observations of the chromosphere,
- visual camera (MetCam) and semi-automatic meteorological station for measurements of humidity, temperature, precipitation, wind speed, amount of snow and sunshine duration,
- set of radiometers and pyrometers for measurements of UV-A, UV-B and photosynthetic active radiation.

5.1.4. Kyoto World Data Center for Geomagnetism

In 1957 the University of Kyoto established the World Data Center for Geomagnetism (WDCC2, 2014), which was reformed in 1975 to a new research facility. Besides being world data center, they established a graduate school with following main research topics:

- electric currents generated by the ionospheric dynamo,
- structure and dynamics of Earth’s magnetosphere,
- short period variations of the geomagnetic field,
- magnetic storms and substorms,
- electromagnetism in the Earth and at the sea floor,
- geomagnetic secular variations.

On data they provide

- magnetograms,
5.2. Datasets

- magnetic indices: AE, Dst, K, Kp, Kn, Ks, Km, Ap, aa, ASY, SYM,
- tellurigrams,
- magnetic field data of satellites: IMP, MAGSAT, GOES, and others.

5.2. Datasets

5.2.1. SDO/AIA-193

We obtained the AIA-193 images from the Virtual Solar Observatory (VSO) and the Joint Science Operations Center (JSOC). The AIA-193 dataset are 16 Mpx filtergrams with a resolution of 1.5 arcsec per pixel and a full width half maximum (FWHM) of 6.3 Å centred at a wavelength of 193 Å, taken every 12 seconds. The emission line arises from Fe XII transitions at a temperature of about 1.26 MK. The used dataset is level 1.5, i.e. overscan, dark currents, flat field, bad pixels and spikes are removed, north is at top of the image, pixel scale was interpolated to 0.6 arcsec per pixel and the image was aligned that way that the center of the Sun is exact at the center of the image (Lemen et al., 2012).

5.2.2. SDO/HMI-los

We obtained the HMI-los images from JSOC. HMI provides two line-of-sight magnetogram datasets at cadences of 45 seconds and 720 seconds. For a better signal-to-noise ratio we use the 720 seconds dataset. The line-of-sight and vector magnetic field is calculated by measuring the polarization states of light and calculating the Stokes parameters. HMI takes one 16 Mpx image at one of six wavelength bands (each 76 mÅ in the Fe I line at 6173.34 Å) at one of six polarization states (I ± V, I ± Q and I ± U, where I, Q, U and V are the Stokes parameters) every 3.75 seconds. The wavelength band is selected by tuning the final stage of a Lyot filter and two Michelson interferometers. Thus

1VSO: http://virtualsolar.org
2JSOC: http://jsoc.stanford.edu
5. *Satellites, observatories, data and data reduction*

One dataset consisting of images at six wavelength at six polarization states is achieved every 135 seconds. From this dataset dark images, flat fields and bad pixels are removed. Ten datasets in series are interpolated by a temporal Wiener interpolation on a regular temporal grid of 45 seconds, resulting in 25 frames per wavelength and polarization state. These are averaged by a boxcar function with $\cos^2$ apodized edges and a FWHM of 720 seconds. 23 frames have a weight greater than 0, the central nine frames have a weight of 1.0. Of these 36 averaged resulting frames the Stokes parameters and finally the line-of-sight (and vector) magnetic field is calculated. In difference to the 720 seconds dataset the 45 seconds line-of-sight dataset measures the six wavelength bands at only two polarization states and calculates the line-of-sight magnetic field without averaging (Hoeksema et al., [2014]).

5.2.3. *ACE/SWEPAM*

We obtained the SWEPAM dataset from Caltech\(^3\). SWEPAM consists of two instruments: SWEPAM-Ion and SWEPAM-Electron. We have restricted ourself to the SWEPAM-Ion instrument. The SWEPAM-Ion instrument consists of three parts: a spherical section electrostatic analyser, a channel electron multiplier and a counting unit. The fan-shaped field of view of the aperture (0 degree to 65 degree azimuthal angle) rotates about the spin axis of the spacecraft. The electrostatic analyser behind the aperture has a high-voltage biased plate so that only ions at a narrow energy per charge ratio at a given azimuthal angle can enter (energy per charge resolution is ± 5 %, range is 0.26 to 36 keV within 40 logarithmic steps in 64 seconds in track mode, 200 steps every 32 minutes in search mode). The channel electron multiplier consists of 16 channels at a azimuthal-angle-separation of 5 degrees, whereby only the 12 inner channels are used. The solar wind is characterized by this with 732 pixels: 12 azimuthal angles (aperture) times 61 polar angles (spin state), each having 40 energy levels (electrostatic analyser). From these the 92 pixels best representing the actual solar wind state are transmitted. The Real Time Solar Wind dataset only consists of the best 33 pixels. Our dataset contains the

\(^3\)Caltech: [http://www.srl.caltech.edu/ACE/ASC/](http://www.srl.caltech.edu/ACE/ASC/)
5.2. Datasets

The absolute value of the proton speed and proton density of an hourly averaged level 3 dataset (McComas et al., 1998).

5.2.4. ACE/MAG

We obtained the MAG dataset from Caltech. The MAG instrument consists of two 3-axial orthogonal ring-core fluxgate sensors. For noise-reduction from ACE electronics these are mounted on booms 4.19m from the center of the spacecraft. The instrument is capable to measure magnetic fluxes and magnetic flux gradients of $\pm 0.001$ to $\pm 65536$ nT at a resolution of 12 bit and a cadence of 24 Hz (Smith et al., 1998).

5.2.5. KSO/H$\alpha$

We obtained the H$\alpha$ images from KSO. The KSO H$\alpha$ telescope is a 12 cm refractor with a Lyot filter observing the full sun at 656.3 nm with a FWHM of 0.07 nm. Images are taken with a 4 MPix 12 bit CCD at a cadence of 6 seconds using frame selection (Pötz et al., 2015).

5.2.6. WDC-C2/Dst

We obtained the Dst dataset from WDC-C2. The Dst index is derived by the average magnetic field variations measured at Honolulu, San Juan, Hermanus and Kakioka. These four observatories are roughly evenly distributed along Earth’s equator. To determine a baseline for each observatory the magnetic field of the five quietest days of each month of the four preceding and the current year is developed into a temporal power spectrum to an order of 2 in order to determine long term changes. Then the power spectrum of the five quietest days of the current month is calculated to define the quiet short-term daily variations, i.e. the daily variations which do not arise from solar wind disturbances like CMEs or high speed streams. The Dst index is finally derived by subtracting the daily measured values from a baseline which is

---

4KSO: http://www.kso.ac.at
5WDC-C2: http://wdc.kugi.kyoto-u.ac.jp/dstdir/
5. Satellites, observatories, data and data reduction

estimated by the long and short-term quiet changes, and averaging over the
four observatories (Sugiura and Kamei, 1991).

5.3. Data reduction

AIA-193 data are already level 1.5, so the images are already corrected for dark
current, flat field, overscan, bad pixel mask, center alignment and rotation. We
further normalized the images to an exposure time of one second.

HMI-los images are also already level 1.5, but can be rotated by 180°. We
derotated the images that north is top (analogous to AIA-193 images) and
rescaled the images to fit the AIA-193 images. We further rotated the Sun in
the image to the reference time of the corresponding AIA-193 image. Because
we investigate coronal holes in this theses, we assumed rigid rotation. The
maximum temporal rotation was 12 minutes.

Ha images are level 1. We derotated the images that north is top, rescaled
the images to fit the AIA-193 images and derotated the Sun to the recording
time of the corresponding AIA-193 image by assuming rigid rotation. The
maximum temporal derotation was up to 12 hours. Further image reduction
on Ha images was not performed because we use these images only as visual
reference images.

ACE data and the Dst index are already fully processed, so we do not have
to do any data reductions on it.
6. Extraction of coronal holes

In this chapter we deal with the extraction of coronal holes out of EUV images of the sun. Because besides coronal holes also filaments appear as dark regions in EUV images (Cranmer, 2009, Labrosse et al., 2010), we classify the coronal holes and filaments afterwards.

In order to identify the dark regions, we use an intensity-based thresholding technique on EUV images developed by my dear colleague Thomas Rotter (Rotter et al., 2012, Rotter et al., 2015). To automatically classify the coronal holes and filaments afterwards, an extraction algorithm and an algorithm for automated classification of coronal holes and filaments was developed by my dear colleague Martin Reiss, Ruben de Visscher (of the Royal Observatory of Belgium) and myself (Reiss et al., 2014a, Reiss et al., 2014b).

Section 6.1 deals with the identification and extraction of dark regions in solar EUV images. In Section 6.2 we deal with the creation of a test dataset in which coronal holes and filaments are labelled manually and examine their statistical characteristics. These was done by myself. Section 6.3 deals with the creation of an automated classification tool. For this task, my colleague Martin Reiss built a systematic dataset of parameters out of the test dataset, containing shape parameters and first and second order statistics derived from both AIA-193 and HMI-los images. This dataset was used as input into machine learning algorithms by Ruben de Visscher to find the best parameters for automatic classification.
6. Extraction of coronal holes

Figure 6.1.: Sample AIA-193 image from 2011/02/01 (top) and corresponding intensity distribution of the solar disk pixels (bottom). The coronal holes outlined in the AIA-193 image result in a second peak (red) in the histogram.
6.1. Extraction of coronal holes and filaments in EUV images

Coronal holes and filaments both appear as dark regions in EUV images. When there are large dark regions on the solar disk, we see an extra low-intensity peak in the histogram of the EUV image (Figure 6.1). This peak allows us to estimate the EUV intensity of the dark regions, and therefore allows to define an intensity threshold for the segmentation of dark regions. A not published study of T. Rotter on appropriate thresholds in AIA-193 images resulted in thresholds of 0.30 to 0.38 times the median of the intensity of the solar disk for coronal holes which are located up to an angular distance of 60° from the disk center. Because of the increased optical depth near the limb due to projection effects this threshold is increased by a factor of 1.6 for coronal holes with an angular distance of more than 60° from the disk center (Rotter et al., 2012).

To create a binary map for dark structures, we decided to use a threshold of 0.38 times the median of the intensity of the solar disk. This threshold slightly increases the area of the structures detected, but reduces the effect of fragmentation. To further remove small coronal hole regions and artefacts in the binary map, we used a median filter with a kernel of 16×16 pixels.

The single objects segmented in the binary map were copied into single binary maps by the following procedure: First the lines of the binary map are searched for an object pixel. As soon as an object pixel is found, the position of this pixel is saved and the pixel erased. A 3×3 pixel window around the saved pixel position is searched for further object pixels. Each found object pixel within this window is again saved, erased and serves as new starting point for the 3×3 pixel window. These pixel positions saved by the recursive procedure are used to create a new binary map with the single object. Then the lines of the original binary map are searched for object pixels of further objects. This procedure segments the single regions „hard“ in the sense of that it does not recognize some associated, but spatial separated regions as one object.

To remove segmentation artefacts close to the limb and remaining artefacts of the full disk binary map, all binary maps with less than 100 object pixels...
are deleted. The remaining binary maps were used as pattern to cut the corresponding AIA-193 and HMI-los images.

Finally for all objects determined shape parameters and first and second order image statistics are calculated (see section 6.3.1), maps containing the latitude and longitude of each pixel are created and the images and data are rearranged into a more handy data product (Figure 6.2).
6.1. Extraction of coronal holes and filaments in EUV images

Figure 6.2.: Data products of dark objects identified in AIA-193 EUV images:
For each date a file was created containing a list of structures, which contain the raw images, the full disk binary map and the extracted dark objects.
6. Extraction of coronal holes

6.2. Statistics of coronal holes and filaments

6.2.1. Dataset and analysis

Coronal holes and filaments were extracted for a three-year period from 2011/01 to 2013/12 at a cadence of one image per day. We restricted the dataset to objects with an area of more than $1 \times 10^{10}$ km$^2$, with a center of the binary object within $\pm 30^\circ$ in heliospheric longitude and latitude, and for which AIA-193 images from VSO, HMI-los images from JSOC and H-alpha-images from KSO were available. This dataset covers 82% of the time range under study.

For each object we calculated the

- de-projected area: $A = \sum_i A_i / (\cos \varphi_i \cos \lambda_i)$,
- latitude: $\varphi = 1/A \sum_i \varphi_i A_i / (\cos \varphi_i \cos \lambda_i)$,
- mean AIA-193 intensity: $I_{193} = 1/n \sum_i I_{193,i}$,
- total magnetic flux: $\Phi = \sum B_i A_i$,
- mean magnetic flux density: $B = 1/n \sum_i B_i$,
- relative open magnetic flux: $\Phi_{\text{open/abs}} = \sum_i B_i A_i / \sum_i |B_i A_i|$,
- positive to absolute open magnetic flux: $\Phi_{\text{pos/abs}} = (\Phi_{\text{open/abs}} + 1)/2$,
- standard deviation, skewness and kurtosis of $B_i$,
- number of fluxtubes at 10 G, 30 G and 50 G: we used the thresholding technique (similar to Section 6.1) with $|B_i| < 10$ G, 30 G, respectively 50 G to segment the magnetogram of the object into single magnetic objects and counted their number,
- symmetry: we rotated and deflected the binary object and calculated the average percentage overlap, which is an indicator for symmetry (see Sect. 6.3.1).
6.2. Statistics of coronal holes and filaments

$A_i$ denotes the area of a pixel, $\varphi_i$ the heliographic latitude of a pixel, $\lambda_i$ the heliographic longitude of a pixel, $I_{193,i}$ the AIA-193 intensity of a pixel, $B_i$ the magnetic field density of a pixel and $n$ the total amount of object pixels. Note that the AIA-193 intensity was normalized to an exposure time of 1 s during data reduction.

To further analyse the magnetic flux distribution, we recalculated the parameters six times, whereby only pixels with an absolute magnetic flux density greater then 10 G, 30 G and 50 G and less then 10 G, 30 G and 50 G were taken into account. These parameters are denoted with the indices 10, 30, 50, -10, -30 and -50.

6.2.2. Manual classification of coronal holes and filaments

In general filaments are visible in Hα images and have a elongated shape and a negligible open magnetic flux (Mackay et al., 2010), whereas coronal holes often have a rounded shape, a large open magnetic flux and are not visible in Hα images (Cranmer, 2009). Therefore we used the appearance in Hα images, the shape and the magnetic flux as classifiers.

For each object a file was generated containing the full-disk AIA-193 image, the full-disk and cut-out HMI-los image, the full-disk and cut-out Hα-image, the date, the position on disk, the area, the mean magnetic flux density and the ratio of positive to absolute magnetic flux (Figures 6.3, 6.4).

Then manual classification into coronal holes and filaments took place by the following rules in this order:

1. If an elongated structure in Hα was apparent, the object was classified as filament.

2. If no structure in Hα was apparent and if the shape of the object was rounded and the ratio of positive to absolute magnetic flux greater then 0.6 or less than 0.4, the object was classified as coronal hole.

3. If the structure was not classified yet, we consulted the Hα images within
6. Extraction of coronal holes

Figure 6.3.: Page of dataset used for labelling coronal holes and filaments: sample coronal hole. From top left to bottom right: AIA-193 image, distribution of the magnetic field strength in the segmented coronal hole object, HMI-los magnetogram, zoomed magnetogram of object, Hα image and zoomed Hα image of object. The zoomed magnetogram shows predominantly negative magnetic field densities (dark pixels). In the zoomed Hα image a filament is not visible.
6.2. Statistics of coronal holes and filaments

Figure 6.4.: Page of dataset used for labelling coronal holes and filaments: sample filament. From top left to bottom right: AIA-193 image, distribution of the magnetic field strength in the segmented coronal hole object, HMI-los magnetogram, zoomed magnetogram of object, Hα image and zoomed Hα image of object. The zoomed magnetogram shows that the object lies at an inversion line of the magnetic field. In the zoomed Hα image a filament is clearly visible.

Datum : 2011–11–14  
Position N : 26.16 Grad  
Position W : −20.03 Grad  

Area : $1.10\times10^8$ km$^2$  
$B$ : 0.31  
$\Phi_{\text{pos}}/\Phi_{\text{obs}}$ : 0.53
6. Extraction of coronal holes

±3 days. If a structure was apparent there, the object was classified as filament.

4. If no structure was apparent in Hα within ±3 days and if the object had a rounded shape or a ratio of positive to absolute magnetic flux less than 0.4 or greater than 0.6, we classified the object as coronal hole.

5. If no structure was apparent in Hα within ±3 days, but the object had an elongated shape and a ratio of positive to absolute magnetic flux between 0.4 and 0.6, the object was not classified.

Coronal holes and filaments which were identified as one merged object by the extraction algorithm were also not classified.

The manual classification resulted in 349 coronal holes, 61 filaments and 14 not assignable objects.

6.2.3. Statistics of coronal hole parameters

The mean values and standard deviations of the parameters defined in Section 6.2.1 are given in Table 6.1. The histograms and the cumulative histograms of the parameters and the most significant dependencies are plotted in Figure 6.5 to 6.14. The histograms show clearly the distribution of the parameters, whereas the cumulative histograms allow to estimate the number of coronal holes which lie in a given parameter range. The following remarks represent only the quintessence of the figures. Therefore we encourage the reader to examine the figures before reading on.

Figure 6.5 shows the distribution of the AIA-193 intensity (top) and of the absolute magnetic field density (bottom) of all coronal hole pixels. The AIA-193 intensity distribution of all pixels peaks at 27 DNs, the mean intensity is 38 DNs and only 1% of all pixels have an intensity greater than 88 DNs. The mean value of the absolute magnetic field density is 5.8 G, 1% of all pixels have an absolute magnetic field density of more than 64 G, and still 0.1% of all pixels have more than 183 G.

Figure 6.6 and 6.7 top show the distribution of the areas, symmetry parameters and mean AIA-193 intensities of coronal holes. 82% of all coronal
6.2. Statistics of coronal holes and filaments

holes have areas less than $5 \cdot 10^{10}\text{ km}^2$. The largest coronal hole has an area of $1.6 \cdot 10^{11}\text{ km}^2$. The symmetry parameter is centred at about 0.5. 75\% of all coronal holes have a symmetry parameter larger than 0.4 which means that most are shaped roundly. The mean AIA-193 intensity is 40 DNs, only 10\% of all coronal holes have mean intensities of more than 50 DNs.

Figures 6.8 to 6.10 show the distributions of the total magnetic flux, the mean magnetic flux densities and the relative open magnetic flux of coronal holes. The total magnetic flux ranges from $4.5 \cdot 10^{18}$ to $4.6 \cdot 10^{21}\text{ Mx}$, the mean total magnetic flux is $8.86 \cdot 10^{20}\text{ Mx}$. The mean value of the absolute values of the mean magnetic flux densities is 2.69 G, less than 10\% of all coronal holes have less than 1 G. The histogram of the relative open magnetic flux has two broad peaks at about −0.4 and 0.5. Only 4\% have a relative open magnetic flux between −0.1 and 0.1, 40\% have a relative open magnetic flux of less than −0.5 or more than 0.5. Therefore almost all coronal holes have a significant open magnetic flux.

Figure 6.11 shows the distribution of the first order statistics of the magnetic field densities of coronal holes. The standard deviations show a broad distribution ranging from 7 to 26 G at a mean standard deviation of 14.28 G. Because the mean magnetic flux density of coronal holes is only 2.69 G, this indicates that the distribution of magnetic flux densities in coronal holes is not uniform but widely distributed. The skewness has two clear peaks at ±7, which reflects that coronal holes have a dominant polarity. The mean kurtosis is 99.67.

Figures 6.7 and 6.10 mid and bottom show the distributions of the relative areas and the relative total magnetic flux if only pixels with $|B_i| > 10\text{ G}$ ($30\text{ G}$) are taken into account. At 10 G the effective area shrinks to about 10\%, but they still contain about 80\% of the total magnetic flux. Rising this threshold to 30 G reduces the area to about 2.5\%, but they still contain about 60\% of the open magnetic flux. Therefore most of the open magnetic flux arises from only a small fraction of the area.

Figure 6.12 top shows $\Phi_{\text{threshold}}/\Phi$ versus the threshold. $\Phi_{\text{threshold}}/\Phi$ is the fraction of the total open magnetic flux of pixels with $|B_i| > \text{threshold}$ on the
6. Extraction of coronal holes

total open magnetic flux. 76% of the open magnetic flux arises from pixels with more than 20 G, 44% from pixels with more than 50 G and still 33% of pixels with more than 80 G.

Figure 6.12 mid and bottom show the distributions of $\Phi_{-10}/\Phi_{10}$ and $\Phi_{-30}/\Phi_{30}$ of coronal holes. At $\Phi_{10}$ ($\Phi_{30}$) only pixels with $|B_i| > 10$ G ($30$ G) were taken into account, at $\Phi_{-10}$ ($\Phi_{-30}$) only pixels with $|B_i| < 10$ G ($30$ G). Both ratios are almost always positive. Therefore the polarities of the average low magnetic field densities and of the average high magnetic field densities are always the same. The dominant polarity holds at both the high-value and the low-value magnetic field density pixels.

Figures 6.13 and 6.14 shows the number of fluxtubes in coronal holes versus the total area, the area covered by fluxtubes and the total magnetic flux within the fluxtubes. The number of fluxtubes depends linearly on the area of the coronal hole. The spreading depends on the thresholds at which the fluxtubes were extracted. A threshold of 10 G belong to an almost perfect linear dependency, against a threshold of 50 G has a medium spreading. The number of fluxtubes also depends linearly on the area covered by fluxtubes and the total magnetic flux within all fluxtubes, whereby an extraction threshold of 10 G has medium spreading and a threshold of 50 G results in an almost perfect dependency.

Figure 6.15 shows the mean magnetic field density, the total magnetic flux and the relative open magnetic flux versus the area, Figure 6.16 the latitude versus the mean magnetic flux density, the total magnetic flux and the area. In the time range under study, large coronal holes appeared predominantly at the northern hemisphere of the sun and and had predominantly negative polarities. They always have a medium or large relative open magnetic flux. In contrast small coronal holes appear at both hemispheres with both polarities and can have small, medium and large relative open magnetic fluxes. Small relative open magnetic fluxes predominantly appear next to the solar equator.
6.2.4. Statistics of filaments parameters

Since the extraction algorithm is tailored to extract coronal holes, we often only extracted fragments of filaments. Nevertheless we are able to make some basic statements.

The detected areas are all smaller than \(7 \cdot 10^{11} \text{km}^2\) (Figure 6.7). The histogram of the symmetric parameter peaks at about 0.3 indicating that they have a elongated shapes (Figure 6.6).

The mean AIA-193 intensity of all filament pixels is 51 DNs, the most common 47 DNs, only 1\% of all pixels have an intensity greater than 97 DNs (Figure 6.5). The mean AIA-193 intensity of all extracted filaments is 53 DNs (Figure 6.6). Since we have used a thresholding extraction algorithm, we can be sure that the mean AIA-193 intensity of filaments is at least 53 DNs.

The mean absolute magnetic flux density of all filament pixels is 4.5 G, 1\% of all pixels have an magnetic flux density of more than 38 G, 0.1\% more than 104 G (Figure 6.5). The absolute mean magnetic flux density is 0.36 G, only 3\% have more than 1 G, the relative open magnetic flux shows a single broad peak at about 0.0 (Figures 6.9, 6.10). 31\% of all filaments have an absolute relative open magnetic flux of more than 0.1. The open magnetic flux is always less then \(6.7 \cdot 10^{20} \text{Mx}\) (Figure 6.8). This indicates that filaments have an almost negligible open magnetic flux at the photosphere, but we cannot definitely confirm it because of the fragmentary extraction.

Taking only pixels above 10 G into account, the area shrinks to about 8\% and the open magnetic flux decreases to 75\%. At an threshold of 30 G the area is about 1\%, the open magnetic flux is 40\% (Figures 6.7, 6.10). The polarities of the mean magnetic flux density of pixels having less than 10 G and pixels having more than 10 G do not depend on each other, also the polarities at an threshold of 30 G do not depend on each other, indicating that an predominant polarity of all pixels does not exist (Figure 6.12). Without a predominant polarity the fraction of the open magnetic flux at a threshold to the open magnetic flux should be seen as upper boundary.

The number of fluxtubes depends linearly on the area of filaments. The area covered by fluxtubes and the absolute magnetic flux within all fluxtubes of a
filament depend linearly on the number of fluxtubes. The total magnetic flux within all fluxtubes is independent of the number of fluxtubes and always quite small. This implies that filaments have an almost equal amount of fluxtubes with different polarities. In general the spreading is less than that of coronal holes (Figures 6.13, 6.14).

Filaments appear at all latitudes independent of their area. On the northern hemisphere they have predominantly positive polarities, on the southern hemisphere predominantly negative polarities (Figure 6.16).

### 6.2.5. Classification features of coronal holes and filaments

The characteristics of coronal holes and filaments (Sect. 6.2.3 and 6.2.4) show some significant differences which can be used as classifiers.

Filaments have a higher mean AIA-193 intensity of $(60 \pm 7)$ DNs (mean ± standard deviation) to coronal holes of $(40 \pm 10)$ DNs, therefore they appear brighter. Their symmetrie parameter is with $0.31 \pm 0.11$ less than that of coronal holes with $0.48 \pm 0.24$, so they have a more elongated shape (Figure 6.6). The mean magnetic flux density is $(0.36 \pm 0.31)$ G, i.e. well below than that of coronal holes with $(2.69 \pm 1.61)$ G (Figure 6.9). The relative open magnetic flux of filaments has a single peak at 0.0, the one of coronal holes has two peaks at $-0.4$ and 0.5. When we take only pixels with more than 10 G into account, the filament peak remains at 0.0 and gets broader while the two peaks of coronal holes move to $-0.6$ and 0.6, which separates coronal holes and filaments even better (Figure 6.10). Coronal holes have a dominant polarity over the magnetic field density scala, while filaments do not (Figure 6.12). In the period of time under study small coronal holes appeared at all latitudes at all polarities, large coronal holes predominantly appear at the northern hemisphere with negative polarity, filaments appear at the northern hemisphere with predominantly positive polarity and at the southern hemisphere with predominantly negative polarity (Figures 6.15, 6.16).

Most of these parameters do not overlap within the standard deviations, nevertheless they overlap in the statistics (Table 6.1). Thus they are suited for...
6.2. Statistics of coronal holes and filaments

classification, but we have to find the best way to combine them. This issue is tackled in the next section.
6. Extraction of coronal holes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coronal Holes</th>
<th>Filaments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{193}$</td>
<td>40.11 6.95</td>
<td>52.69 7.57</td>
</tr>
<tr>
<td>Symmetry</td>
<td>0.48 0.24</td>
<td>0.31 0.11</td>
</tr>
<tr>
<td>Area</td>
<td>3.34e10 2.86e10</td>
<td>2.24e10 1.35e10</td>
</tr>
<tr>
<td>Area$_{10}$/Area</td>
<td>0.113 0.026</td>
<td>0.087 0.010</td>
</tr>
<tr>
<td>Area$_{30}$/Area</td>
<td>0.029 0.013</td>
<td>0.014 0.004</td>
</tr>
<tr>
<td>Area$_{50}$/Area</td>
<td>0.015 0.008</td>
<td>0.006 0.003</td>
</tr>
<tr>
<td>$\text{abs}(\Phi)$</td>
<td>8.86e20 8.66e20</td>
<td>0.83e20 1.10e20</td>
</tr>
<tr>
<td>$\Phi_{10}/\Phi$</td>
<td>0.82 0.32</td>
<td>0.74 14.55</td>
</tr>
<tr>
<td>$\Phi_{30}/\Phi$</td>
<td>0.57 0.38</td>
<td>0.41 4.28</td>
</tr>
<tr>
<td>$\Phi_{50}/\Phi$</td>
<td>0.41 0.41</td>
<td>0.15 2.52</td>
</tr>
<tr>
<td>$\text{abs}(B)$</td>
<td>2.69 1.61</td>
<td>0.36 0.31</td>
</tr>
<tr>
<td>$\text{abs}(B_{10})$</td>
<td>18.65 8.25</td>
<td>3.70 2.57</td>
</tr>
<tr>
<td>$\text{abs}(B_{30})$</td>
<td>51.84 16.25</td>
<td>15.84 10.95</td>
</tr>
<tr>
<td>$\text{abs}(B_{50})$</td>
<td>77.73 20.77</td>
<td>29.87 18.94</td>
</tr>
<tr>
<td>$\Phi_{\text{open}}/\Phi_{\text{abs}}$</td>
<td>0.42 0.09</td>
<td>0.03 0.03</td>
</tr>
<tr>
<td>$\Phi_{\text{open},10}/\Phi_{\text{abs},10}$</td>
<td>0.64 0.20</td>
<td>0.17 0.11</td>
</tr>
<tr>
<td>$\Phi_{\text{open},30}/\Phi_{\text{abs},30}$</td>
<td>0.79 0.19</td>
<td>0.29 0.19</td>
</tr>
<tr>
<td>$\Phi_{\text{open},50}/\Phi_{\text{abs},50}$</td>
<td>0.84 0.18</td>
<td>0.38 0.25</td>
</tr>
<tr>
<td>$\text{Stdev(abs(B))}$</td>
<td>14.28 4.25</td>
<td>9.27 1.90</td>
</tr>
<tr>
<td>$\text{Skew(abs(B))}$</td>
<td>6.59 2.08</td>
<td>2.77 2.62</td>
</tr>
<tr>
<td>$\text{Kurt(abs(B))}$</td>
<td>99.67 51.13</td>
<td>87.41 87.13</td>
</tr>
<tr>
<td>$\Phi_{\text{abs}}$</td>
<td>1.96e21 1.70e21</td>
<td>1.02e21 0.65e21</td>
</tr>
<tr>
<td>$B_{\text{abs}}$</td>
<td>5.88 1.28</td>
<td>4.52 0.40</td>
</tr>
</tbody>
</table>

Table 6.1.: Mean value and standard deviation of the following parameter distributions derived for coronal hole and filaments: mean AIA-193 intensity, symmetry parameter, area, total magnetic flux, mean magnetic field density, relative open magnetic flux, first order statistics of the magnetic field densities, total of the absolute magnetic flux and mean value of the absolute magnetic field density. For the definition of the parameters see Section 6.2.1.
6.2. Statistics of coronal holes and filaments

Figure 6.5.: Distribution of AIA-193 intensities (top) and absolute value of the magnetic field densities (bottom) derived from all pixels identified as CH pixels (black) and filament pixels (red). The left panel shows the histogram, the right panel the reverse cumulative histogram.
6. Extraction of coronal holes

Figure 6.6.: Distribution of mean AIA-193 intensities (top) and shape parameters (bottom) derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6.2. Statistics of coronal holes and filaments

Figure 6.7.: Distribution of areas (top) and fractional areas where only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G (bottom) were taken into account, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6. Extraction of coronal holes

Figure 6.8.: Distribution of total magnetic flux (top) and fractional total magnetic fluxes where only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G (bottom) were taken into account, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6.2. Statistics of coronal holes and filaments

Figure 6.9.: Distribution of the mean magnetic field density (top) and the mean magnetic field densities when only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G (bottom) were taken into account, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6. Extraction of coronal holes

Figure 6.10.: Distribution of relative open magnetic flux (top) and relative open magnetic fluxes when only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G (bottom) were taken into account, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6.2. Statistics of coronal holes and filaments

Figure 6.11.: Distribution of standard deviations (top), skewness (mid) and kurtosis (bottom) of the magnetic field densities, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6. Extraction of coronal holes

Figure 6.12.: Fraction of the total magnetic flux, when only pixels with an absolute magnetic field density above a threshold were taken into account, versus the threshold (top). Distribution of the ratios $\Phi_{-10}/\Phi_{10}$ and $\Phi_{-30}/\Phi_{30}$, derived from all CHs (black) and filaments (red). The left panel shows the histogram, the right panel the cumulative histogram.
6.2. Statistics of coronal holes and filaments

Figure 6.13.: Scatter plots of the number of magnetic flux tubes derived at a magnetic threshold of 10 G (top), 30 G (mid) and 50 G (bottom) versus the area (right) and the area of pixels which have an absolute magnetic field density of more than 10 G, 30 G and 50 G (left), derived from all coronal holes (black) and filaments (red).
6. Extraction of coronal holes

Figure 6.14.: Scatter plots of the number of magnetic flux tubes derived at a magnetic threshold of 10 G (top), 30 G (mid) and 50 G (bottom) versus the absolute total (left) and total (right) magnetic flux of pixels which have an absolute magnetic field density of more than 10 G, 30 G and 50 G (left), derived from all coronal holes (black) and filaments (red).
6.2. Statistics of coronal holes and filaments

Figure 6.15.: Scatter plots of the mean magnetic field densities (top), total magnetic flux (mid) and relative open magnetic flux (bottom) versus the areas, derived from all coronal holes (black) and filaments (red).
6. Extraction of coronal holes

Figure 6.16.: Scatter plots of the latitude versus the relative open magnetic flux (top), areas (mid) and mean magnetic field densities (bottom), derived from all coronal holes (black) and filaments (red).
6.3. Automatic classification of coronal holes and filaments

6.3.1. Parameters

In Section 6.2.5 we showed that some parameters are suited to distinguish well, but not distinct between coronal holes and filaments. Based on these results my colleague Martin Reiss created an extended set of parameters which will be used in the next section to find the best classification method. This extended set of parameters includes the magnetic flux imbalance, shape measures and first and second order image statistics of both AIA-193 and HMI-los images (Reiss et al., 2014b).

Magnetic flux imbalance

We define the parameter Magnetic Flux Imbalance as

\[ MFI = 0.5 \cdot \text{abs} \left( \frac{\Phi}{\Phi_{\text{abs}}} \right) , \]  

(6.1)

where \( \Phi \) is the total magnetic flux and \( \Phi_{\text{abs}} \) is the sum over the absolute values of the magnetic flux densities of each pixel.

Shape measures

We use two different shape descriptors: symmetry analysis and direction-dependent shape analysis.

- Symmetry analysis: Performing geometrical operations like rotation or reflection at a complete symmetric object should result in the original object. Measuring the overlap of the rotated to the original object gives us a measure on the symmetry of the shape. Be \( I(x, y) \) the original binary image with values \( \in \{0, 1\} \), \( O_i \) an operator performing the operation \{rotate image by \( i \cdot 90^\circ \mid i \in \{0, 1, 2, 3\}\}, R_j \) an operator performing the operation \{reflect image, identity\} and \( N \) the total amount of pixels in the image.
6. Extraction of coronal holes

Then we calculate the symmetry parameter by

\[ S_y = \frac{1}{8N} \sum_{i,j} \sum_{x,y} O_i(R_j(I(x,y))) \times I(x,y) \]  (6.2)

The image multiplication “\( \times \)” is done pixel by pixel.

- Direction-dependent analysis: Knowing the number of pixels in all possible directions, starting at the center of the image, you can reconstruct the shape. So the function \( f(\phi) \), which gives the number of pixels in direction \( \phi \), is an unique attribution to the shape. The direction-dependent measure is calculated by

\[ D = \text{stddev} \left( \frac{f}{\max(f)} \right). \]  (6.3)

First order image statistics

First order image statistics are based on the probability distribution of pixel values. The probability to find a pixel with value \( i \) is given by

\[ P(i) = \frac{n(i)}{N}, \]  (6.4)

where \( n(i) \) is the number of pixels with value \( i \) in the image and \( N \) is the total number of image pixels. The following first order image statistic parameters were calculated:

- Mean: \( \mu = \sum_i i P(i) \)
- Variance: \( \sigma^2 = \sum_i (i - \mu)^2 P(i) \)
- Standard Deviation: \( \sigma \)
- Energy: \( E_1 = \sum_i P(i)^2 \)
- Entropy: \( S_1 = -\sum_i P(i) \log P(i) \)
Second order image statistics

Second order image statistics are based on the probability distribution of *pixel arrangements*. The probability to find a pixel with value \(i\) and a second pixel with value \(j\) in a fixed direction and distance of \(i\) is given by

\[
P_{\Phi,d}(i,j) = \frac{n_{\Phi,d}(i,j)}{N},
\]

(6.5)

where \(n_{\Phi,d}(i,j)\) is the number of pixels with value \(i\), which have an associated pixel with value \(j\) in a fixed direction \(\Phi\) and a fixed distance \(d\) of \(i\), in the image and \(N\) is the total number of pixels in the image. The corresponding matrix \(P_{\Phi,d}^{i,j}\) of all arrangement probabilities is called co-occurance matrix. We calculated the co-occurance matrix for each of the eight directions \(n \cdot 45^\circ, n \in [0, 7]\) and a distance of one pixel. The final co-occurance matrix was calculated by averaging the eight co-occurance matrices.

With the definitions

\[
\begin{align*}
p_x(i) &= \sum_j P(i, j) \\
p_y(j) &= \sum_i P(i, j) \\
p_{x+y}(i) &= \sum_{x+y=i} P(x, y) \\
p_{x-y}(i) &= \sum_{x-y=i} P(x, y) \\
\mu_x &= \text{mean}(p_x) \\
\mu_y &= \text{mean}(p_y) \\
\sigma_x &= \text{stddev}(p_x) \\
\sigma_y &= \text{stddev}(p_y)
\end{align*}
\]

the following second order image statistic parameters were calculated:

**Energy:**

\[
H_1 = \sum_i \sum_j P(i, j)
\]

**Contrast:**

\[
H_2 = \sum_j (i - j)^2 P(i, j)
\]
6. Extraction of coronal holes

Correlation:
\[ H_3 = \sum_i \sum_j (i - \mu_x)(j - \mu_x)/(\sigma_x \sigma_y) \]

Variance:
\[ H_4 = \sum_i \sum_j (i - \mu)^2 P(i, j) \]

Homogeneity:
\[ H_5 = \sum_i \sum_j P(i, j)/(1 + (i - j)^2) \]

Sum Average:
\[ H_6 = \sum_{i=2}^{2N} ip_{x+y}(i) \]

Sum Variance:
\[ H_7 = \sum_{i=2}^{2N} (i - H_8)^2 p_{x+y}(i) \]

Sum Entropy:
\[ H_8 = -\sum_{i=2}^{2N} ip_{x+y}(i) \log p_{x+y}(i) \]

Entropy:
\[ H_9 = -\sum_i \sum_j P(i, j) \log P(i, j) \]

Difference Variance:
\[ H_{10} = \text{var}(p_{x-y}) \]

Difference Entropy:
\[ H_{11} = -\sum_{i=0}^{N-1} p_{x-y}(i) \log p_{x-y}(i) \]

Information measures of correlation (I):
\[ H_{12} = (H_9 - HXY)/(\max(HX, HY)) \]

Information measures of correlation (II):
\[ H_{13} = \sqrt{1 - \exp(-2(HXY^2 - H_9))} \]

6.3.2. A first classification rule

My colleague Martin Reiss used the program Weka\footnote{Waikato environment for knowledge analysis (Weka): http://www.cs.waikato.ac.nz/ml/weka} a machine learning suite programmed in Java, to find a first classification rule out of the extended parameter dataset defined in Section 6.3.1. All 349 coronal holes and 61 filaments were used as training dataset (without using a validation dataset) and Decision Tree as performing algorithm. The resulting decision tree for automated classification is shown in Figure 6.17 and 6.18. This decision tree classified 98.6% of the coronal holes as coronal holes and 82.0% of the filaments as filaments, resulting in a true skill statistics of 0.81. Note that no validation dataset was used, so this decision tree could overfit the data. In fact the real true skill statistics will probably be less.

Because no better classification algorithm was ready until the end of this
6.3. Automatic classification of coronal holes and filaments

This decision tree was used for automated classification of filaments and coronal holes in the datasets of Chapters 7 and 8.

6.3.3. Supervised classification

Ruben de Visscher applied and compared four supervised classification algorithms to find automatically the best decision rule between coronal holes and filaments, namely Linear Support Vector Machine, Support Vector Machine, Decision Tree and Random Forest, and tuned it to contain as much coronal holes as possible in exchange for a less prediction rate of filaments.

- The Linear Support Vector machine tries to find a hyperplane that separates the datasets as much as possible by minimizing a loss function.

- The key idea of the Support Vector Machine is the same, but it maps the parameters in a nonlinear way. A gaussian, sigmoid, polynomial and linear kernel was used.

- The Decision Tree tries to minimize the entropy of the dataset by if-then-else decisions.

- The Random Forest further tries to minimize the entropy by using a set of Decision Trees which prediction rates are averaged.

The dataset containing the 349 coronal holes and 61 filaments was split up randomly into a training dataset and a validation dataset, containing 75\% and 25\% of the data. The training dataset was further split up into five folds. Each combination of four of the five folds was once again used as training dataset and the fifth as validation dataset in order to find the best classifier combination for each supervised classification algorithm and thereby in particular to not prefer one of the algorithms. Having found the best combination of classifiers the complete training dataset was used to train the classifiers. The trained classifiers were validated in the validation dataset. We get one true positive ratio (TPR), the ratio of the coronal holes that were predicted as coronal holes, and a false positive ratio (FPR), the ratio of filaments that
6. Extraction of coronal holes

Figure 6.17.: Decision tree for automated classification of coronal holes and filaments, created by Weka.
6.3. Automatic classification of coronal holes and filaments

Figure 6.18.: Scatter plot of the symmetry parameter versus the absolute value of the relative open magnetic flux, derived for all coronal holes (black) and filaments (red). The classification border between coronal holes and filament as derived by Weka is drawn in green.
6. Extraction of coronal holes

Figure 6.19.: Results of the supervised classification algorithms: density plots of the true positive ratio (TPR) versus the false positive ratio (FPR) for Decision Tree (a), Linear SVM (b), Random Forest (c) and SVM (d). From Reiss et al., 2014b.

were predicted as coronal holes, for each supervised classification algorithm. The overall performance was evaluated by performing these steps 100 times.

Figure 6.19 shows the results for each of the four supervised classification algorithms. For each of the four algorithms the density of the true positive ratio to the false positive ratio are plotted, indicating that the Linear Support Vector machine performs best. The true skill statistics is $0.90 \pm 0.07$ for the Support Vector machine, $0.94 \pm 0.05$ for the Linear Support Vector Machine, $0.87 \pm 0.10$ for Random Forest and $0.80 \pm 0.15$ for Decision Tree. These values indicate that an almost distinct classification is possible. However an associated decision rule
has not been extracted until the end of this work.
7. Correlation of high speed streams and Dst index with coronal holes

The existence of a link between coronal holes and high speed streams as well as between high speed streams and geomagnetic storms is well established and many models were proposed. Nevertheless at the moment we are not able to make reliable predictions for high speed stream parameters and moreover for geomagnetic storms. This is on one hand related to the fact that we are not able to observe the physical properties of coronal holes and solar high speed streams at all heights, on the other hand because it is difficult to interpret all observations due to the complexity of the systems.

Vršnak, Temmer, and Veronig (2007a) and Rotter et al. (2012) have shown that a simple relationship between the fractional area coronal holes cover within a meridional slice and the solar wind proton velocity and Dst index exists (see chapter 3). Borovsky (2008) has shown that we do not measure an indistinct magnetic field but single flux tubes at L1 which are likely to originate in Sun’s photosphere. Thus we expect that not the area of coronal holes in the slice, but each coronal hole on its own creates a high speed stream, even though the speed of the high speed stream can depend on the total number of coronal holes near the slice.

This brings us to study the statistical relationship between single coronal holes, high speed streams and geomagnetic storms. We created a dataset of 49 coronal holes at times of low system complexity, i.e. where only one large coronal hole was near the central meridian, and correlated these parameters
Section 7.1 presents the dataset, Section 7.2 analyses the arrival times of the peaks in the solar wind velocity, magnetic field, density and in the geomagnetic storm index Dst. Section 7.3 deals with the solar high speed stream peak velocities, section 7.4 with the peak densities of the shock fronts, section 7.5 with the peak magnetic field densities of the shock fronts and section 7.6 with the drops of Dst.

7.1. Dataset

The dataset contains 49 selected coronal holes at latitudes up to ±60° which occurred in the period 2011/01 to 2013/12 (Figures 7.1 and 7.2). The selection criterion was a low overall system complexity and a distinct correspondence of coronal holes to high speed streams. Therefore we took coronal holes at times where only one large coronal hole appeared near the solar central meridian and where definite peaks in the solar wind data measured at Lagrangian point L1 appeared. The solar wind peaks were compared to the Richardson-Cane-Lis\(^1\) (Cane and Richardson, 2003; Richardson and Cane, 2010) to exclude ICME’s. We especially tried to also include small coronal holes. The distribution of our dataset on latitude and area of coronal holes is given in Figure 7.3. Anyway we have to assume that we did not select a perfect statistical representative sample. Therefore we are aware of that discrepancies on the statistical values which we derive to the reality are possible. These statistical values have to be checked in a validation dataset. This is partly done in Chapter 8.

Coronal holes: We calculated for all coronal holes the area (\(A_{\text{CH}}\)), latitude (\(\varphi_{\text{CH}}\)), magnetic flux density (\(B_{\text{CH}}\)), total magnetic flux (\(\Phi_{\text{CH}}\)), relative open magnetic flux (\(\Phi_{\text{CH,open/abs}}\)) and the times when the coronal hole first touched the central meridian (\(t_{\text{CH,front}}\)), when the center crossed the central meridian (\(t_{\text{CH,center}}\)) and when the coronal hole has left the central meridian (\(t_{\text{CH,back}}\)).

\(^1\)http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm
Figure 7.1.: Coronal holes in dataset. The detected coronal holes of the AIA-193 images are outlined. The data set contains the largest coronal hole at the central meridian.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.2.: Coronal holes in dataset. The detected coronal holes in the AIA-193 images are outlined. The dataset contains the largest coronal hole at the central meridian.
7.1. Dataset

The temporal resolution of the coronal hole images was six hours.

The areas of coronal holes were de-projected from 2 dimensional flat images. The magnetic flux density and all corresponding magnetic parameters are not corrected for projection effects. We also did the analysis for corrected magnetic parameters under the assumption of a radial magnetic field, but achieved better results with the uncorrected magnetic line of sight parameters. We derived the latitude and longitude of each coronal hole as the center of gravity of the de-projected area. The exact definition of the parameters is given in Section 6.2.1.

High speed streams and Dst: Of the corresponding high speed streams measured at L1 we picked the peak velocity ($v_{\text{max}}$), the peak absolute magnetic flux density at the shock front ($B_{\text{max}}$), the corresponding vector magnetic field components ($B_x$, $B_y$, $B_z$ in GSM), the peak density at the shock front ($\rho_{\text{max}}$), the time when the velocity starts rising ($t_{\text{rise}}$), the time of peak velocity ($t_{\text{vmax}}$), the time when the velocity reaches again the quiet state ($t_{\text{fall}}$), the time of peak absolute flux density ($t_B$) and the time of peak density ($t_{\rho}$). Of the corresponding geomagnetic storms we picked the minimum in Dst index ($D_{\text{st}}$) and the time of minimum Dst index ($t_{\text{Dst}}$). The temporal resolution for all picked data was six hours. The definitions of the points in time is illustrated in Figure 7.4.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.4.: Definition of parameters $t_{\text{CHfront}}$, $t_{\text{CHcenter}}$, $t_{\text{CHback}}$, $t_{\text{vrise}}$, $t_{\text{vmax}}$, $t_{\text{vfall}}$, $t_B$, $t_\rho$ and $t_{\text{Dst}}$, indicated by vertical lines. From top to bottom: butterfly diagram of coronal holes, solar wind proton velocity, solar wind proton density, solar wind absolute magnetic field density, solar wind proton temperature and Dst index. Black parts of the curves correspond to ICMEs.
7.2. Arrival times

We analysed the solar wind magnetic field in the Heliocentric Earth Equa-
torial Coordinate System (HEEQ). The x-y plane is the equatorial plane of
the Sun. The z-axis is the Sun’s rotation axis, the x-direction is defined as
the intersection of Sun’s equatorial plane with the central meridian as seen
from Earth and the y-axis is perpendicular to the x- and z-axis. The hourly
ACE magnetic field components are first transformed from GSM to HEEQ
coordinates and then averaged within \( \pm 3, \pm 6 \) respectively \( \pm 12 \) hours in order
to consider the change of orientation of GSM to HEEQ due to the rotation of
Earth’s magnetic field axis around its rotation axis. Out of these components
the magnetic magnitude \( B_t \), magnetic latitude \( B_\varphi \) (\(-90 \) to \( 90^\circ \)) and magnetic
longitude \( B_\lambda \) (\(-180 \) to \( 180^\circ \)) are calculated, where \( B_\lambda = 0^\circ \) refers to the x-
direction towards the Sun and \( B_\varphi = 90^\circ \) refers to the Sun’s rotational north
pole. The incident angle of the slow solar wind is about \( B_\lambda = -45^\circ \).

7.2. Arrival times

The Arrival time is composed of the travel time and the starting time. We
define a longitudinal position \( X \) within the coronal hole as the position which
is \( X \ (\in [0,1]) \) times the longitudinal width of the coronal hole behind the
longitudinal front of the coronal hole. The starting time \( t_{CH,X} \) is defined as
the time when the position \( X \) crosses the central meridian. The travel time is
defined as the delay of arrival time to starting time:

\[
d_{\text{travel}} = t_{\text{arrival}} - t_{CH,X}. \tag{7.1}\]

Figure 7.5 shows \( d_{v_{\text{max}}} \), i.e. the travel time of \( v_{\text{max}} \), for assumed starting
times \( t_{CH, \text{front}}, t_{CH,0.36} \) and \( t_{CH,\text{mid}} \) versus \( v_{\text{max}} \), the data are coloured by the
longitudinal width of coronal holes. The travel times depend linearly on \( v_{\text{max}} \)
and have a wide, but slightly different spreading. The linear fit for the distribu-
tion at \( X = 0 \) (corresponding to \( t_{CH,\text{front}} \)) results in significantly longer travel
times than the pure radial travel time with constant speed \( d = 1 \text{AE}/v_{\text{max}} \), the
Mean Root Square Error (RMSE) is 1.29 days. The linear fit systematically
underestimates the measured travel time of coronal holes which have a large
7. Correlation of high speed streams and Dst index with coronal holes

longitudinal width and systematically overestimates the travel time of coronal holes which have a small longitudinal width. The linear fit for $X = 0.36$ results in slightly longer travel times than the pure travel time at a RMSE of 1.20 days, the fit for $X = 0.5$ coincide well with pure travel time at a RMSE of 1.26 days. For $X = 0.36$ and $X = 0.5$ the spreading is evenly distributed for both large and small longitudinal widths. We further investigated the spreading by varying $X$ from 0 to 1 and calculating the RMSE of the data to the linear fits (Figure 7.6 top), which resulted in a minimum of RMSE at $X = 0.36$.

The different spreading has a simple explanation: Think of a systematic false chosen point of origin $X_{\text{false}}$. Then we have to add to the travel time of $v_{\text{max}}$ the time delay between the false point of origin crossing the central meridian $X_{\text{false}}$ to the real point of origin $X_{\text{real}}$ crossing the central meridian. This additional delay is dependent on the longitudinal width because of our definition of $X$. If we assume that $X = 0$ is a false point of origin, coronal holes of large longitudinal width will get a larger additional time delay than coronal holes with a small longitudinal width, it natively divides the travel times. In contrast a well chosen point of origin will have similar distributions of the travel time for both large and small longitudinal width of coronal holes. Therefore the spreading is a measure on the distance between the assumed point of origin and the real point of origin, and thus we suppose that the statistical real point of origin is $X = 0.36$. The statistical arrival time is then given by

$$t_{v_{\text{max}}} = t_{\text{CH, 0.36}} + (7.66 - 0.0066 \cdot v_{\text{max}}[\text{km/s}])[\text{days}]. \quad (7.2)$$

The RMSE of the observed arrival times to calculated arrival times is 1.2 days, typical travel times are 2 to 6 days.

Figure 7.7 top shows $d_{\text{rise}}$, i.e. the travel time of $v_{\text{rise}}$, versus $v_{\text{max}}$. The starting time was determined to be $X = 0.25$ by varying $X$ and looking for the minimum root mean square deviations (Figure 7.6 mid). The travel time of $v_{\text{rise}}$ shows a similar distribution like the travel time of $v_{\text{max}}$, but is about one day shorter. The delay between $t_{\text{rise}}$ and $t_{v_{\text{max}}}$ is given in Figure 7.7 bottom and confirms this. The statistical arrival time $t_{\text{rise}}$ was determined to

$$t_{\text{rise}} = t_{v_{\text{max}}} - (0.691 - 0.00096 \cdot v_{\text{max}}[\text{km/s}])[\text{days}]. \quad (7.3)$$
7.2. Arrival times

The RMSE of observed to calculated arrival times is 0.6 days, the velocity typically starts rising about 1.25 days before $v_{\text{max}}$ arrives.

Figure 7.8 top shows $d_{v_{\text{fall}}}$, i.e. the travel time of $v_{\text{fall}}$, versus $v_{\text{fall}}$, the statistically best starting time was determined to $X = 0.36$ (Figure 7.6 bottom). Higher $v_{\text{fall}}$ result in shorter travel times $d_{v_{\text{fall}}}$. The statistical arrival time of $v_{\text{fall}}$ was determined to

$$t_{v_{\text{fall}}} = t_{\text{CH,0.36}} + (13.5 - 0.0159 \cdot v_{\text{fall}}[\text{km/s}])[\text{days}].$$  \hspace{1cm} (7.4)

The RMSE of observed to calculated arrival times is 1.3 days, typical travel times are 5 to 10 days.

Figure 7.8 mid shows $d_{v_{\text{fall}}}$ for $X = 0.36$ versus the longitudinal width of coronal holes. Here a clear dependency is not visible, which is quite surprising. If a high speed stream originates over the whole longitudinal width of a coronal hole, we would expect that the slowest parts of high speed streams arise from the borders. Therefore we would expect that the slowest part of a high speed stream arriving at L1 would arise from the trailing edge of coronal holes. If we then choose a starting time of $X = 0.36$, we have to add to the travel time $d_{v_{\text{fall}}}$ the delay $t_{\text{CH,1.0}} - t_{\text{CH,0.36}}$, i.e. the time the trailing edge of the coronal hole needs to arrive at the solar central meridian. This additional delay is dependent on the longitudinal width of the coronal hole, therefore we would expect to see a dependence of $d_{v_{\text{fall}}}$ on the longitudinal width for a starting time $t_{\text{CH,0.36}}$. This dependency does not exist. To review this result we plotted the travel time $d_{v_{\text{fall}}}$ versus the longitudinal width at an starting time of $t_{\text{CH,1.0}}$, i.e. the trailing edge of the coronal hole (Figure 7.8 bottom). If $v_{\text{fall}}$ arises from $X = 1.0$, we should not see any dependency on the longitudinal width. But here a clear dependency exist. As a consequence either high speed streams do not originate over the whole area of coronal holes, or an effect exists which exactly reverses the dependency of the travel time on longitudinal width.

Because the travel times of $v_{\text{max}}$ and $v_{\text{rise}}$ have similar a dependency on $v_{\text{max}}$, we investigate their relationship. Figure 7.9 top shows the deviations of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ to its linear fit (corresponding to Figure 7.5 mid) versus the deviations of $d_{v_{\text{rise}}}$ vs. $v_{\text{max}}$ to its linear fit (corresponding to Figure 7.7 top). The deviations are strongly correlated with a Pearson correlation coefficient ($r_P$)
7. Correlation of high speed streams and Dst index with coronal holes

of 0.85. Thus if the velocity starts rising earlier than the linear fits predicts, the maximum velocity also arrives earlier. This result was expected because of the causal relationship between $v_{\text{max}}$, shock front and $v_{\text{rise}}$. Figure 7.9 bottom shows the deviations of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ to its linear fit (corresponding to Figure 7.5 mid) versus the deviations of $d_{v_{\text{fall}}}$ vs. $v_{v_{\text{fall}}}$ to its linear fit (corresponding to Figure 7.8 top). A dependency of deviations also exists here with $r_P = 0.52$.

Although there exists many possible explanations (e.g. energetic coupling between $v_{\text{max}}$ and $v_{\text{fall}}$), we would like to point to the fact that the correlation of deviations can be easily explained if high speed streams do not originate over the whole area of coronal holes but over a small singularity in the coronal hole. Then the deviations in travel time would correspond to the longitudinal distance of the singularities to the statistical points of origin. Moreover $v_{\text{fall}}$ would not arise from the trailing edge of coronal holes, but also from this singularity which would explain the unexpected dependency of $d_{v_{\text{fall}}}$ on the longitudinal width.

Note that we did not assume a constant travel velocity until now, although we have used linear fits on the scatter plots of the travel time vs. velocity. The linear fits consist of a linear dependency of the travel time on the velocity and an offset travel time. This offset travel time can include acceleration processes.

We now investigate if in general an assumption of a radial high speed stream with constant travel velocity and neglected acceleration time is justifiable. Therefore we assume a constant travel velocity $v_{\text{max}}$ and $v_{\text{fall}}$ and reckon back to their times of origin on the Sun:

$$t_{\text{sun}} = t_{v_{\text{max}}(v_{\text{fall}})} - 1 \text{AU}/v_{\text{max}}(v_{\text{fall}}).$$

(7.5)

With the estimated time of origin on the Sun and the knowledge of the position of the source coronal hole on the Sun we derive the point of origin X in the coronal hole. Figure 7.10 shows the longitudinal points of origins X for $v_{\text{max}}$ versus the longitudinal widths and $v_{\text{max}}$. Most of the points of origin lie within the coronal holes, smaller coronal holes have a wider spreading of points of origin. Thus the assumption of constant travel is justifiable for $v_{\text{max}}$. Figure 7.11 show the points of origin of $v_{\text{fall}}$ versus the longitudinal widths, $v_{\text{fall}}$ and $v_{\text{max}}$. Most points lie at values between 1 and 3, i.e. behind the trailing edge.
7.2. Arrival times

of coronal holes. Whereby a dependency of points of origins on \(v_{\text{fall}}\) does not exist, they depend on \(v_{\text{max}}\) and on the longitudinal width: the higher \(v_{\text{max}}\) and the smaller the longitudinal width, the farer off the points of origin. This implies that the assumption of constant travel velocity is not valid for \(v_{\text{fall}}\), thus that acceleration processes have to be taken into account into the travel time. The dependency of the points of origin \(X\) on \(v_{\text{max}}\) implies that the acceleration processes of \(v_{\text{fall}}\) are involved with \(v_{\text{max}}\).

Finally we have a look at the peak times \(t_{\rho}, t_B\) and \(t_{\text{Dst}}\). Figure 7.12 top and mid show the delay of \(t_{\rho}\) and \(t_B\) to \(t_{v_{\text{max}}}\). Both arrive about 1.25 days earlier than \(v_{\text{max}}\) and have a slight dependency on \(v_{\text{max}}\). Figure 7.12 bottom show the delay of \(t_{\rho}\) to \(t_B\). A clear dependency is not visible, the delay is usually less than 0.5 days. The statistical arrival times of \(t_{\rho}\) and \(t_B\) were fitted to

\[
    t_B = t_{v_{\text{max}}} + (0.314 - 0.002 \cdot v_{\text{max}}[\text{km/s}])[\text{days}], \quad (7.6)
\]

\[
    t_{\rho} = t_{v_{\text{max}}} + (1.25 - 0.0037 \cdot v_{\text{max}}[\text{km/s}])[\text{days}]. \quad (7.7)
\]

The RMSE of measured to calculated arrival times are 0.8 days for both \(t_B\) and \(t_{\rho}\).

Figure 7.13 shows the delay of \(t_{\text{Dst}}\) to \(t_{v_{\text{max}}}\) versus \(v_{\text{max}}\) (top), \(\rho_{\text{max}}\) (mid) and \(B_{\text{max}}\) (bottom). The delay is typically less than 1 day. It is statistically not dependent on \(v_{\text{max}}\) and \(b_{\text{max}}\), but has a slight dependency on \(\rho_{\text{max}}\). The statistically peak time \(t_{\rho}\) was determined to

\[
    t_{\text{Dst}} = t_{v_{\text{max}}} + (0.273 - 0.048 \cdot \rho_{\text{max}}[\text{cm}^{-3}])[\text{days}] \quad (7.8)
\]

with a RMSE of measured to calculated arrival times of 0.8 days.

Table 7.1 summarizes the dependencies, typical arrival times and root mean square deviations of \(t_{v_{\text{max}}}, t_{v_{\text{rise}}}, t_{v_{\text{fall}}}, t_B, t_{\rho}\) and \(t_{\text{Dst}}\).
7. Correlation of high speed streams and Dst index with coronal holes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( t_{\text{start}} )</th>
<th>Delay</th>
<th>RMSE</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_{v\text{max}} )</td>
<td>( t_{\text{CH0.36}} )</td>
<td>(+ (7.66 - 0.0066 \cdot v_{\text{max}}))</td>
<td>1.2</td>
<td>2 to 6</td>
</tr>
<tr>
<td>( t_{v\text{rise}} )</td>
<td>( t_{v\text{max}} )</td>
<td>(+ (-0.691 - 0.00096 \cdot v_{\text{max}}))</td>
<td>0.6</td>
<td>-2 to -0.5</td>
</tr>
<tr>
<td>( t_{v\text{fall}} )</td>
<td>( t_{\text{CH0.36}} )</td>
<td>(+ (13.5 - 0.0159 \cdot v_{\text{fall}}))</td>
<td>1.3</td>
<td>5 to 10</td>
</tr>
<tr>
<td>( t_{\text{B}} )</td>
<td>( t_{v\text{max}} )</td>
<td>(+ (0.314 - 0.0020 \cdot v_{\text{max}}))</td>
<td>0.8</td>
<td>-2 to 0</td>
</tr>
<tr>
<td>( t_{\rho} )</td>
<td>( t_{v\text{max}} )</td>
<td>(+ (1.25 - 0.0037 \cdot v_{\text{max}}))</td>
<td>0.8</td>
<td>-2 to 0</td>
</tr>
<tr>
<td>( t_{\text{Dst}} )</td>
<td>( t_{v\text{max}} )</td>
<td>(+ (0.273 - 0.048 \cdot \rho_{\text{shock}}))</td>
<td>0.8</td>
<td>-1 to 1.5</td>
</tr>
</tbody>
</table>

Table 7.1.: Arrival times of \( t_{v\text{max}}, t_{v\text{rise}}, t_{v\text{fall}}, t_{B}, t_{\rho} \) and \( t_{\text{Dst}} \). The arrival times are divided into starting times and delays. The RMSE gives the root mean square deviations of the measured to the calculated delays. Range gives the typical range of delays in days.
7.2. Arrival times

Figure 7.5.: Dependence of travel time $d_{\text{max}}$ on $v_{\text{max}}$ at a starting time of $t_{\text{CHfront}}$ (top), $t_{\text{CH0.36}}$ (mid) and $t_{\text{CHmid}}$. The black line corresponds to a travel time $1 \text{ AE}/v_{\text{max}}$, the red line to the linear fit to the data. Colours denote the longitudinal width of coronal holes.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.6.: Root mean square deviations (RMSE) of the travel times $d_{v_{\text{max}}}$ (top), $d_{v_{\text{rise}}}$ (mid) and $d_{v_{\text{fall}}}$ (bottom) to their linear fits versus the normalized starting positions $X$ within coronal holes.
7.2. Arrival times

Figure 7.7.: Dependence of travel time $d_{\text{rise}}$ on $v_{\text{max}}$ at a starting time of $t_{\text{CH, 0.25}}$ (top). Delay of $t_{\text{rise}}$ to $t_{\text{vmax}}$ versus $v_{\text{max}}$ (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.8: Dependence of travel time $d_{\text{vfall}}$ on $v_{\text{max}}$ (top) and longitudinal width of the coronal hole (mid and bottom). Starting times are $t_{\text{CH, 0.36}}$ (top and mid) and $t_{\text{CH, 1}}$ (bottom).
7.2. Arrival times

Figure 7.9.: Deviations of the travel times $d_{v_{\text{max}}}$ to the linear fit of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ at a starting time $t_{CH, \ 0.36}$ versus the deviations of the travel times $d_{v_{\text{rise}}}$ to the linear fit of $d_{v_{\text{rise}}}$ vs. $v_{\text{max}}$ at a starting time $t_{CH, \ 0.25}$ (top). Deviations of the travel times $d_{v_{\text{max}}}$ to the linear fit of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ at a starting time $t_{CH, \ 0.36}$ versus the deviations of the travel times $d_{v_{\text{fall}}}$ to the linear fit of $d_{v_{\text{fall}}}$ vs. $v_{\text{max}}$ at a starting time $t_{CH, \ 0.25}$ (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.10.: Starting position of $v_{\text{max}}$ in the coronal holes, when a constant radial travel velocity is assumed. The starting positions are normalized on the longitudinal width of the coronal holes (leading edge: $X = 0$, trailing edge: $X = 1$). The y-axis are the longitudinal width of the coronal holes (top) and $v_{\text{max}}$ (bottom). The dashed vertical lines indicate the leading and trailing edge of the coronal holes.
7.2. Arrival times

Figure 7.11.: Starting position of $v_{\text{fall}}$ in the coronal holes, when a constant radial travel velocity is assumed. The starting positions are normalized on the longitudinal width of the coronal holes (leading edge: $X = 0$, trailing edge: $X = 1$). The y-axis are the longitudinal width of the coronal holes (top), $v_{\text{fall}}$ (middle) and $v_{\text{max}}$ (bottom). The dashed vertical lines indicate the leading and trailing edge of the coronal holes.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.12.: Delay of $t_\rho$ to $t_{v_{\text{max}}}$ (top), $t_B$ to $t_{v_{\text{max}}}$ (mid) and $t_B$ to $t_\rho$ (bottom) versus $v_{\text{max}}$. 
7.2. Arrival times

Figure 7.13.: Delay of $t_{\text{Dst}}$ to $t_{v_{\text{max}}}$ versus $v_{\text{max}}$ (top), $\rho_{\text{max}}$ (mid) and $B_{\text{max}}$ (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

7.3. Solar wind velocity

According to the model of Wang and Sheeley (1991) the solar wind peak velocity is related to the flux tube expansion factor of the „open“ flux tube corresponding to the coronal hole. The flux tube expansion factor is expected to be related with the area of coronal holes (Wang and Sheeley, 1990). Based on these results we compare the area of coronal holes with the associated solar wind peak velocities.

Figure 7.14 top shows the peak velocities $v_{\text{max}}$ plotted against the areas $A_{\text{CH}}$ of coronal holes. The data points are wide spread. The color of the data points - indicating the latitudes $\varphi_{\text{CH}}$ of coronal holes - divide the the data into several branches.

Figure 7.14 bottom contains the same data which are divided into four panels with segments of coronal hole latitude ($0^\circ$–$15^\circ$, $15^\circ$–$30^\circ$, $30^\circ$–$45^\circ$, $45^\circ$–$60^\circ$). We denote the four subsets of data with index $i \in \{1, 2, 3, 4\}$. Each of the four plots corresponding to one latitude segment show a linear dependence of solar wind peak velocity on coronal hole area. The four corresponding regression lines $v_{\text{regression},i} = v_{\text{offset},i} + m_i \cdot A_{\text{CH},i}$ were calculated with the constraint that the offset velocities $v_{\text{offset},i}$ are identical.

To further investigate the dependence of $v_{\text{max}}$ on coronal hole latitude, we plotted the slopes $(v_{\text{max}} - v_{\text{offset}})/A_{\text{CH}}$ on $|\varphi_{\text{CH}}|$ (Figure 7.15 top). The slopes show a linear dependency on coronal hole latitude.

We chose as overall fitting function of the peak velocities as function on the coronal hole area and latitude a function of type $v_{\text{max}} = a + (b + c \cdot |\varphi_{\text{CH}}|) \cdot A_{\text{CH}}$ and got:

$$v_{\text{fit}}[\text{km/s}] = 467 + (2.5 \cdot 10^{-9} - 4.25 \cdot 10^{-11} \cdot |\varphi_{\text{CH}}|) \cdot A_{\text{CH}}[\text{km}^2]. \quad (7.9)$$

Figure 7.15 bottom shows the estimated peak velocities $v_{\text{fit}}$ versus the measured peak velocities. The distribution of the data shows that the chosen fitting function was reasonable, although high measured velocities are slightly underestimated and slow velocities overestimated. The RMSE is 68.7 km/s, the Pearson correlation coefficient $r_P$ 0.69, the slope of the regression line 25.5° and the forecast correlation coefficient $r_{fc}$ 0.39.
7.3. Solar wind velocity

The cumulative histogram of the deviations $v_{\text{max}} - v_{\text{fit}}$ is shown in Figure 7.16 top. More than 50\% have deviations of less than 41 km/s, more than 75\% deviations of less than 59 km/s. Figure 7.16 mid and bottom show the deviations versus $\Phi_{\text{CH}, \text{open/abs}}$ and versus $v_{\text{max}}$. The mid panel indicates that coronal holes with a high relative open magnetic flux have statistically less deviations than coronal holes with a small open magnetic flux. The bottom plot shows that the fit mostly underestimates high maximum velocities and overestimates small maximum velocities. However this effect cannot be removed, because we cannot use the $v_{\text{max}}$ as input parameter and because the deviations do not depend on $v_{\text{fit}}$.

We related the highest speeds of solar high speed streams to coronal holes near the solar equator, whereas Ulysses measured higher solar wind velocities over the poles than over the equator at solar minimum (see Sect. 2.3.1). But Ulysses measured the slow solar wind and high speed streams during its orbits around the Sun, i.e. the solar wind speeds at different ecliptic latitudes, whereas we measured only high speed stream speeds arising from mid- and low-latitude coronal holes at L1, i.e. in the ecliptic. Thus these results are not mutually exclusive. Robbins, Henney, and Harvey (2006) divided the solar disk into segments of 14° longitude and 30° latitude, calculated for each segment the fractional area coronal holes cover, and correlated the fractional areas with the solar wind speeds at L1 for a 11 year period from 1992 to 2003. They also found that coronal holes near the solar equator produce higher high speed streams velocities at L1 than coronal holes at higher latitudes.

We cannot say if high speed streams arising from mid-latitude coronal holes in general have lower peak velocities than high speed streams from low-latitude coronal holes, or if we only measure lower peak velocities at L1. If we expect the highest velocity in the center of the high speed stream flux tube with lower velocities near the edges, our result could mean that the velocity ACE measures depends on ACE’s latitudinal position in the flux tube. ACE’s position in the flux tube could be correlated to the latitude of the corresponding coronal hole.

If the latitude dependency arises from the position ACE measures in flux tubes, we also expect a slight statistical dependency on the distance to Sun
7. Correlation of high speed streams and Dst index with coronal holes

and on the ecliptic latitude of ACE and so a slight dependency on the day of year (DOY). ACE’s position in the flux tube should be affected by the distance to the Sun for coronal holes on the northern and southern hemisphere in the same way and thus cause a statistical dependency of measured solar wind peak velocities on the day of year. In contrast ACE’s position in the flux tube should be affected by its ecliptic latitude for coronal holes on the northern and southern hemisphere in contrary ways, thus we also expect a dependency of the spreading of the measured solar wind peak velocities on the day of year.

We plotted the deviations of Figure 7.15 bottom to its fit against the day of year (Figure 7.17 top). The distribution is widely spread and shows a slight, but not significant dependency on the day of year. We fitted a sinus function \( a \cdot \sin(\text{DOY} + b) \) to the data and found the minimum of deviations at \( b = 80 \) days. Although the large spreading prohibits a distinct statement, the found fit with \( b = 80 \) days meets our expectations: the statistic shift of deviations is small near the spring and autumn equinox, at the points where Earth’s orbit is in the equatorial plane of the Sun and large at summer solstice and winter solstice, where the distances and latitude are turning points. A dependency of the spreading on the day of year is not visible.

We applied a new fit which is also dependent on the day of year to the peak velocities and got

\[
v_{\text{fit}}[\text{km/s}] = 470 + (2.28 \cdot 10^{-9} - 3.78 \cdot 10^{-11} \cdot |\varphi_{\text{CH}}[^\circ]| - 3.62 \cdot 10^{-10} \cdot \sin((\text{DOY} + 80)/365 \cdot 2\pi)) \cdot A_{\text{CH}}[\text{km}^2].
\]  

Figure 7.17 bottom shows the estimated peak velocities \( v_{\text{fit}} \) versus the measured peak velocities. The RMSE is 65.2 km/s, \( r_P \) is 0.71, the slope of the regression line is 37° and \( r_{fc} \) is 0.58. Therefore by considering the dependency on the day of year the RMSE and \( r_P \) increased slightly, the slope of the regression line and \( r_{fc} \) increased significantly.

The cumulative histogram of the deviations \( v_{\text{max}} - v_{\text{fit}} \) is shown in Figure 7.18 top. More than 50% have deviations of less than 40 km/s, more than 75% deviations of less than 62 km/s. Figure 7.18 mid and bottom show the deviations versus \( \Phi_{\text{CH, open/abs}} \) and \( v_{\text{max}} \) similar to Figure 7.16. The mid panel
7.3. Solar wind velocity

indicates that coronal holes with a high relative open magnetic flux have statistically less deviations than coronal holes with a small open magnetic flux. The bottom plot shows that the fit still mostly underestimates high maximum velocities and overestimates small maximum velocities.
Figure 7.14.: Solar wind peak velocities versus the area of coronal holes, the latitudes of coronal holes are coloured (top). In the four bottom panels the data are divided into four segments, each covering a range of 15° latitude, and plotted against the normalized coronal hole area. The four linear fits (red lines) were calculated with the constraint that the ordinate value of the four fits are identical.
7.3. Solar wind velocity

Figure 7.15.: Latitude of coronal holes versus \((v_{\text{max}} - 485)/A_{\text{CH}}\) (top). Calculated solar wind peak velocities using Eq. 7.9 versus measured solar wind peak velocities (bottom). The linear fits are drawn in red. The dashed line corresponds to a one-to-one correspondence.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.16.: Cumulative histogram of the deviations of measured velocities to calculated velocities using by Eq. 7.10 (top). Relative open magnetic flux of coronal holes versus the deviations (mid) and \( v_{\text{max}} \) versus deviations (bottom).

124
7.3. Solar wind velocity

Figure 7.17.: Deviations of measured velocities to the calculated velocities using Eq. 7.9 versus the day of year (top). Calculated solar wind peak velocities using Eq. 7.10 versus measured solar wind peak velocities (bottom). The linear fit is drawn in red. The dashed line corresponds to a one-to-one correspondence.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.18.: Cumulative histogram of the deviations of measured velocities to calculated velocities using Eq. 7.10 (top). Relative open magnetic flux of coronal holes versus the deviations (mid) and $v_{\text{max}}$ versus deviations (bottom).
7.4. Solar wind density

Figure 7.19 show the scatter plots of the peak densities $\rho_{\text{max}}$ of the shock front versus the inverse magnetic flux density $1/B_{\text{CH}}$ (top), inverse area $1/A_{\text{CH}}$ (mid) and relative open magnetic flux $\Phi_{\text{CH,open/abs}}$ (bottom) of the coronal holes. A fit to these data seems not reasonable, though constraints on the peak density can be made:

$$\rho_{\text{max}} \ [1/\text{cm}^3] < 40 - \frac{22}{B_{\text{CH}} \ [\text{G}]} \quad (7.11)$$

$$\rho_{\text{max}} \ [1/\text{cm}^3] < 40 - \frac{8 \cdot 10^{11}}{A_{\text{CH}} \ [\text{km}^2]} \quad (7.12)$$

Thus high peak densities require a medium to high magnetic flux density and a large area. They do not appear at low magnetic flux densities or small areas. But a large area and high magnetic flux density do no enforce a high peak density. For prediction of density at least one more missing parameter is necessary.

Quite surprising is that the four highest peak densities appear at medium values of relative open magnetic flux (Figure 7.19 bottom), since the relative open magnetic flux like the magnetic flux density and the area mostly indicate the „strength“ of coronal holes. It is not clear if these high density values at medium relative open magnetic flux are outliers or not.

A dependence on the magnetic field density seems reasonable because a larger magnetic flux density induces a smaller plasma $\beta$ and thus a larger influence of electrodynamic effects. Since particles and magnetic field lines are confined and different magnetic field lines cannot penetrate, a larger magnetic field results in less penetration of different solar wind streams and thus into a higher compression of the proceeding solar wind. A dependence on the area seems reasonable because a large area induce a large volume of particles which can be accelerated.

Consider that the constraints are not definite but only observational. It can be that high speed streams do have data points outside the constraints, but that the corresponding parameter combination appears rarely. It can also be that the dataset is not representative on all high speed streams - the dataset
7. *Correlation of high speed streams and Dst index with coronal holes*

was selected manually. The statement of constraints should be interpreted as probability function.
7.4. Solar wind density

Figure 7.19: Scatter plots of peak density versus inverse magnetic flux density (top), inverse area (mid) and relative open magnetic flux (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

7.5. Solar wind magnetic field

The total magnetic flux of coronal holes should be conserved when the corresponding magnetic flux tubes expand into interplanetary space, therefore we should be able to predict the total magnetic flux and the average magnetic field density at L1 by observations of the magnetic flux of coronal holes. However at L1 we measure in-situ only the magnetic field density of a very small section of the flux tube, which is likely not representative for the average magnetic field density in the flux tube. We therefore try to correlate the total magnetic flux of the coronal hole not only with the peak magnetic field density measured at L1, but also with the magnetic field density measured at L1 averaged over ±3, ±6 and ±12 hours.

First we define the magnetic polarity of the solar wind which corresponds to the sign of $B_t$. Until now the sign of $B_t$ is ambiguous: $B_t, B_\lambda$ and $B_\varphi$ (in HEEQ) are derived from $B_x, B_y$ and $B_z$ (in GSM), though $(B_t, B_\lambda, B_\varphi)$ is identical with $(-B_t, B_\lambda + 180^\circ, -B_\varphi)$. We therefore have to define a further constraint to set the sign of $B_t$.

Figure 7.20 shows the magnetic field density and incident angles measured at L1 one day after $t_{v_{\text{max}}}$, i.e. one day after the shock front (maximum velocity) passed, and averaged over ±6 hours. At this time we assume to be unaffected by compression and deflection effects of the shock. We defined the polarity of $B_t$ by postulating that $B_\lambda = (-45 \pm 90)^\circ$, where $B_\lambda = -45^\circ$ is about the incident angle of the slow solar wind. The top panel shows the magnetic field density $B_t$ versus the total magnetic flux $\Phi_{CH}$ of the corresponding coronal holes. The polarities of coronal holes and the measured magnetic flux densities coincident quite well and the absolute magnetic flux density is always greater than 2 G. The middle panel shows the magnetic latitude $B_\varphi$ against the latitude $\varphi_{CH}$ of coronal holes. A dependency of solar wind magnetic latitude to coronal hole latitude is not apparent. The bottom panel shows the magnetic longitude $B_\lambda$ against the solar wind peak velocity $v_{\text{max}}$. Because of the constraint the incident angle is always between $(-45 \pm 90)^\circ$. A wide spread dependence on the velocity is apparent, the corresponding fit agrees well with the theoretical value of $B_{\lambda,\text{theo}} = \arctan(\omega_{\text{sun}} \cdot 1\text{AE}/v_{\text{max}})$ (Eq. 2.6).
7.5. Solar wind magnetic field

where $\omega_{\text{sun}}$ is the rotation velocity of the Sun. Because of the well agreement of magnetic polarities and of the statistical longitudinal incident angles with the theoretical values we assign these polarities to the high speed streams.

Figure 7.21 shows the magnetic flux density and the incident angles versus the total magnetic flux and latitude of coronal holes and the solar wind velocity at $t_{\text{vmax}}$ averaged over $\pm 6$ hours. The top panel shows that the polarities of coronal holes and magnetic flux densities still agree well and that the absolute magnetic flux is also always greater than 2 G. A dependence of the magnetic flux density of the solar wind on the total magnetic flux of coronal holes is not apparent. The middle panel shows the solar wind magnetic latitude against the latitude of coronal holes. A dependence is not visible. The bottom panel shows the magnetic longitude against the solar wind peak velocity. Most of the data are between $(-45 \pm 90)^\circ$, but one data point is at about $-180^\circ$ indicating that the polarity has changed.

Figure 7.22 shows the maximum magnetic flux density and the incident angles versus the total magnetic flux and latitude of coronal holes and the solar wind velocity at $t_{\text{B}}$, i.e. at the time of maximum magnetic field density in the shock, averaged over $\pm 6$ hours. The top panel shows the polarities of coronal holes and magnetic flux, which still quite agree. The offset of absolute magnetic flux has increased to about 5 G, the spreading has decreased drastically allowing to fit the data. The middle panel show the solar wind magnetic latitude against the latitude of coronal holes. A dependence is not apparent. The bottom panel shows the magnetic longitude against the solar wind peak velocity. Four data points have magnetic longitudes out of $(-45 \pm 90)^\circ$ indicating that polarities have changed. A dependence of solar wind magnetic longitude on the peak velocity is still visible.

Altogether the spreading of $B_t$ vs. $\Phi_{\text{CH}}$ is large in the high speed stream $(t_{\text{vmax}} + 1 \text{ day})$ and at the shock front $(t_{\text{vmax}})$ and medium at the peak magnetic field density in the shock $(t_{\text{B}})$. At all three points in time we have a statistical dependency of $B_\lambda$ on $v_{\text{max}}$ which coincide well with the theoretical value. The solar wind magnetic latitude and the latitude of coronal holes are uncorrelated. Therefore the magnetic field density in the shock represents the total
magnetic flux of coronal holes best. This can be explained by accumulation of the magnetic flux of coronal holes in the shock. We further investigate the dependence of the magnetic flux density in the shock on the total magnetic flux of coronal holes.

Figure 7.23 shows the magnetic field density at L1 at $t_B$, averaged over 0 hours (top panel), ±6 hours (middle panel) and ±12 hours (bottom panel) versus the total magnetic flux of coronal holes. The panel with an average of ±6 hours has the lowest spreading, followed by the panel with an average over ±12 hours with a slightly higher spreading and the not-averaged panel with a significant higher spreading.

In order to fit the dependency of the maximum magnetic field density averaged over ±6 hours on the total magnetic flux of coronal holes, we first removed the offset of 5 G (Figure 7.24 top), fitted the dependency and added the offset again. We got

$$B_{\text{fit}} \ [G] = -0.155 + 5 \cdot \text{sign}(\Phi_{\text{CH}}) + 1.63 \cdot 10^{-21} \cdot \Phi_{\text{CH}} \ [\text{Mx}].$$ (7.13)

Figure 7.24 mid shows the calculated magnetic field densities using Eq. 7.13 versus the measured magnetic field densities averaged over ±6 hours. The calculated magnetic field densities mostly lie near the measured magnetic field densities. The Pearson correlation coefficient $r_P$ is 0.86, the forecast correlation coefficient $r_{fc}$ is 0.69 and the RMSE is 4.8 G. The mean value of the absolute magnetic flux density of the measured data is 7.9 G. Figure 7.24 bottom shows the distribution of the deviations of calculated and measured values. More than 55% of the deviations are less than 2 G, more than 75% less than 5 G.

Note that the high correlation coefficients are mostly a result of that we divided $B_t$ into two polarities. In this process we divided $B_t$ into two spacial separated subsets, which enforces high correlation coefficients. The Pearson correlation between $|B_t|$ and $|\Phi_{\text{CH}}|$, i.e. without separation of the subsets, is only 0.18.

For the subsequent study of geomagnetic storms in Section 7.6 we further investigate the southward magnetic field component at L1 in GSM coordinates at the time of maximum magnetic field density in the shock $t_B$. Figure 7.25 shows $B_z$ averaged over 0 hours (top panel), ±3 hours (middle panel) and
$\pm 12$ hours (bottom panel) versus $\Phi_{CH,z}$, which is the $z$-component of the total magnetic flux of coronal holes and which is derived under the assumption that the total magnetic flux passes at L1 with the magnetic incident angles $\Phi_\lambda = -45^\circ$ and $\Phi_\varphi = 0^\circ$. The not-averaged panel shows a wide-spread dependence, which gets better at an average of $\pm 3$ hours and much better at an average of $\pm 12$ hours. Because the strength of geomagnetic storms are strongly affected by the peak southward magnetic field density, averaging over $\pm 12$ hours seems to be too much at a usual shock duration of 1 to 1.5 days. We therefore further investigate the average over $\pm 3$ hours.

Figure 7.26 top shows $B_t$ at $t_B$ averaged over $\pm 3$ hours versus $\Phi_{CH}$. The polarities mostly coincide and a offset of 4 G is apparent. We calculate $B_{t,\text{fit}}$ for the averaged data over $\pm 3$ hours analogue to Eq. 7.13 by

$$B_{t,\text{fit}} [G] = -0.155 + 4 \cdot \text{sign}(\Phi_{CH}) + 1.63 \cdot 10^{-21} \cdot \Phi_{CH} [\text{Mx}]$$ (7.14)

and derive the $z$-component $B_{z,\text{fit}}$ by projection under the assumption of $B_\lambda = -45^\circ$ and $B_\varphi = 0^\circ$.

Figure 7.26 mid shows $B_z$ versus $B_{z,\text{fit}}$. The dependence of $B_z$ on $B_{z,\text{fit}}$ is not steady. If we reduce the offset of 4 G to 0 G (Figure 7.26 bottom) the dependence becomes steady again. Thus, although an offset of 4 G is apparent for $B_t$, no offset is the better choice for $B_z$. This can be due to the used incident angle of $-45^\circ$, which is only an estimate. Note that the slope of the regression line of the $B_{z,\text{fit}} - B_z$ plot is with 9$^\circ$ much too low to use Eq. 7.14 as forecast. A perfect forecast, i.e. a one-to-one correspondence of $B_z$ and $B_{z,\text{fit}}$, would have a slope of regression line of 45$^\circ$.

We corrected Eq. 7.14 in that way that the slope of the regression line of Figure 7.26 bottom matches the one-to-one correspondence and finally got

$$B_{t,\text{fit}} [G] = -0.155 \cdot \text{sign}(\Phi_{CH}) + 10.65 \cdot 10^{-21} \cdot \Phi_{CH} [\text{Mx}].$$ (7.15)

$B_{z,\text{fit}}$ is achieved by projecting $B_{t,\text{fit}}$ to $B_{z,\text{fit}}$ under the assumption of $B_\lambda = -45^\circ$ and $B_\varphi = 0^\circ$. $B_{z,\text{fit}}$ versus $B_z$ is plotted in Figure 7.27 top. The Pearson correlation coefficient $r_P$ is 0.40, the forecast correlation coefficient $r_{fc}$ is 0.40 and the RMSE is 10.6 G. Figure 7.27 bottom shows the histogram of the
Correlation of high speed streams and Dst index with coronal holes

deviations $|B_z - B_{z,\text{fit}}|$. 36% of the deviations are less than 3 G, 51% less than 5 G and 83% less than 10 G. Because the mean value of the measured absolute magnetic flux density peaks is only $|B_z|$ is 2.7 G, the RMSE is in comparison with the mean value of the measured absolute magnetic flux density too high for using Eq. 7.15 as forecast for $B_z$, although the correlation coefficients are quite satisfying.
7.5. Solar wind magnetic field

Figure 7.20.: Solar wind magnetic field density at L1 measured 24 hours after $t_{v_{\text{max}}}$ versus total magnetic flux of coronal holes (top), solar wind magnetic latitude versus latitude of coronal holes (mid) and solar wind magnetic longitude versus $v_{\text{max}}$ (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.21.: Solar wind magnetic field density at L1 measured at $t_{v_{\text{max}}}$ versus total magnetic flux of coronal holes (top), solar wind magnetic latitude versus latitude of coronal holes (mid) and solar wind magnetic longitude versus $v_{\text{max}}$ (bottom).
7.5. Solar wind magnetic field

Figure 7.22.: Solar wind magnetic field density at L1 measured at $t_B$ versus total magnetic flux of coronal holes (top), solar wind magnetic latitude versus latitude of coronal holes (mid) and solar wind magnetic longitude versus $v_{\text{max}}$ (bottom).
Figure 7.23.: Solar wind magnetic field densities at L1 measured at $t_B$ versus total magnetic flux of coronal holes. The solar wind magnetic field density was averaged over ±0 hours (top), ±6 hours (mid) and ±12 hours (bottom).
7.5. Solar wind magnetic field

Figure 7.24: Dependence of the solar wind magnetic flux density at L1 measured at $t_B$ and averaged over ±6 hours on total magnetic flux of coronal holes, after an offset of 5 G has been removed (top). Calculated magnetic flux density using Equation (7.13) versus measured magnetic flux densities (mid). Cumulative histogram of the deviations of measured to calculated magnetic flux densities (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.25.: Scatter plots of $B_z$ in GSM coordinates measured at L1 at $t_B$ and averaged over 0 hours (top), ±3 hours (mid) and ±12 hours (bottom) versus the $B_z$-component of the total magnetic flux of coronal holes. To derive the $B_z$-component of the total magnetic flux of coronal holes, a magnetic longitude of $-45^\circ$ and a magnetic latitude of $0^\circ$ was assumed.
Figure 7.26: $B_t$ averaged over ±3 hours versus total magnetic flux of coronal holes (top). Calculated magnetic flux densities $B_{z,fit}$ using Eq. 7.14 under the assumption of an offset of 4 G (mid) and 0 G (bottom) versus measured $B_z$-components in GSM at L1 at $t_B$ averaged over ±3 hours.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.27.: Calculated magnetic field densities $B_{z, \text{fit}}$ using Eq. 7.15 versus measured $B_z$-components in GSM at L1 at $t_B$ averaged over ±3 hours (top). Cumulative histogram of the deviations $|B_z - B_{z, \text{fit}}|$ (bottom).
7.6. Dst-index

The Dst index corresponds mainly to the state of the ring-current system of Earth’s magnetosphere, which is influenced by the solar wind. If we analyse high speed streams as disturbances of the slow solar wind, we should accordingly not analyse the Dst index itself but the change of Dst index which arise from the high speed stream. To estimate the current “quiet” state of Dst index, i.e. the current state without high speed stream disturbances but with disturbances of CMEs, we fitted a parabola over the maxima of Dst index at a time period of ±2 weeks around the Dst minima, whereby we weighted the points of time with $1/\left(\max(Dst_{\pm 2 \text{ weeks}}) - Dst\right)^{0.75}$. Failed fits were corrected manually, which was necessary for 4 of the 49 selected Dst drops. Then we subtracted the Dst drop from the quiet state to estimate the Dst change due to the high speed stream, which we denote as $\Delta Dst$ (Figure 7.28).

According to Section 3.5 we expect a dependency of the change of Dst on solar wind velocity, density and the southward component of the interplanetary magnetic field. The dependence of $\Delta Dst$ versus the $v_{\text{max}}$, $\rho_{\text{max}}$ and $B_z$ measured at L1 are plotted in Figure 7.29. The $\Delta Dst$ values show a wide spread dependency on both the velocity and the southward magnetic flux density. Large Dst drops appear predominantly at high velocities and high negative southward magnetic field densities. In addition the largest Dst changes appear at small and high – but not medium – solar wind densities. Note that the solar wind density and velocity are correlated: low solar wind densities appear usually at high solar wind velocities and high solar wind densities at slow velocities (Schwenn, 1983). Figure 7.30 top shows the changes of Dst versus its delay to the solar wind peak velocity. High Dst drops appear most of the time before the solar wind peak velocity arrives, thus they are induced by the shock. Therefore it is justifiable to correlate the Dst drops with only the magnetic field and density peaks in the shock front and the maximum velocity without regarding the temporal distributions.

Next we expand the correlations from parameters of high speed streams to the corresponding coronal hole parameters.

Figure 7.30 bottom shows $\Delta Dst$ versus $\Phi_{\text{CH},z}$ at L1. $\Phi_{\text{CH},z}$ is the z-
7. Correlation of high speed streams and Dst index with coronal holes

component of the total magnetic flux of coronal holes in GSM coordinates, which is derived under the assumption of magnetic incident angles of $\Phi_{\varphi} = 0^\circ$ and $\Phi_{\lambda} = -45^\circ$. The changes of Dst show a wide spread, linear dependency on $\Phi_{\text{CH}, z}$. Changes of Dst are statistically higher at polarities of coronal holes which belong to a negative $\Phi_{\text{CH}, z}$ and therefore to a negative southward magnetic field component.

Figure 7.31 top shows the dependency of $\Delta Dst$ on the area $A_{\text{CH}}$, the polarities of $\Phi_{\text{CH}, z}$ are coloured. A clear dependence on the area is not visible. In Figure 7.31 bottom the same data are divided into four panels with segments of coronal hole latitude ($0^\circ$–$15^\circ$, $15^\circ$–$30^\circ$, $30^\circ$–$45^\circ$, $45^\circ$–$60^\circ$). In each segment the change of Dst has a clear, but wide spread linear dependency on the area. The slopes of the corresponding fits become smaller at higher latitudes. To further investigate the dependence of $\Delta Dst$ on the latitude of coronal holes $\varphi_{\text{CH}}$, we plotted $\Delta Dst / A_{\text{CH}}$ versus $\varphi_{\text{CH}}$ (Figure 7.32 top). A dependency on the latitude is clearly visible, large Dst per area drops only appear near the solar equator.

Altogether we can observe a dependency of changes of Dst on the area, latitude and the southward component of the transformed total magnetic flux of coronal holes. Therefore we chose a fitting function of type

$$\Delta Dst_{\text{fit}} = (a + b \cdot A_{\text{CH}}) \cdot (c + d \cdot |\varphi_{\text{CH}}|) \cdot (e + f \cdot \Phi_{\text{CH}, z}).$$

The associated least-square fit resulted in very low correlation coefficients of less than 0.25 between $\Delta Dst$ and $\Delta Dst_{\text{fit}}$. We therefore adjusted the parameters manually, whereby we tried to maximise the forecast correlation coefficient $r_{fc}$, and got

$$\Delta Dst_{\text{fit}}[G] = (A_{\text{CH}} [\text{km}^2]) \cdot (-6.88 \cdot 10^{-10} + 9.93 \cdot 10^{-12} \cdot |\varphi_{\text{CH}}[^\circ]|) \cdot (1 - 2 \cdot 10^{-22} \cdot \Phi_{\text{CH}, z} [\text{Mx}]). \quad (7.16)$$

Figure 7.32 mid shows the calculated Dst drops using Eq. 7.16 versus the measured Dst drops. The Pearson correlation coefficient $r_P$ of the measured and calculated Dst drops is 0.53 and the slope of regression line is $37^\circ$ resulting in an forecast correlation coefficient $r_{fc}$ of 0.44. The RMSE is 16.5 G, the
7.6. Dst-index

mean value of measured Dst drops is $-34\, \text{G}$. The cumulative histogram of the absolute deviations of the estimated to the measured Dst drops are plotted in Figure 7.32 bottom. 46\% of all deviations are less than 10 G, 75\% are less than 20 G. At the stated values two outlier with $(\Delta Dst_{\text{fit}} = 42\, \text{G}, \Delta Dst = -37\, \text{G})$ and $(\Delta Dst_{\text{fit}} = -46\, \text{G}, \Delta Dst = -107\, \text{G})$ were excluded.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.28.: 10 samples for the determination of Dst drop: the plots show the
time line of the Dst indices two weeks around the selected Dst
minimum. A parabola (red lines) was fitted weighted to the Dst
indices two weeks around the selected Dst minimum (red dashed
line). The Dst drop is the difference between parabola and Dst
minimum. The black horizontal line corresponds to a Dst index
of 0.
Figure 7.29.: Distribution of Dst drop against solar wind maximum velocity (top), maximum density in the shock (mid) and then $B_z$-component of the maximum magnetic field density in the shock (bottom).
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.30.: Dependence of Dst drop on its delay to $t_{vmax}$ (top) and on $\Phi_{CH,z}$ (bottom). $\Phi_{CH,z}$ is the z-component of the total magnetic flux of coronal holes in GSM coordinates, which is derived under the assumption of magnetic incident angles of $\Phi_\varphi = 0^\circ$ and $\Phi_\lambda = -45^\circ$ in HEEQ coordinates and transformation into GSM coordinates.
Figure 7.31.: Dst drops versus the area of coronal holes. Data points with a negative southward magnetic field $\Phi_{\text{CH}, z}$ are coloured black, with a positive southward magnetic field red. In the four bottom panels the data are divided into four segments of $15^\circ$ coronal hole latitude each and plotted against the normalized coronal hole areas.
7. Correlation of high speed streams and Dst index with coronal holes

Figure 7.32.: Dst drops per area versus latitude of coronal holes (top). Calculated Dst drops using Eq. 7.16 versus measured Dst drops (mid), the dashed line corresponds to a one-to-one correspondence. Cumulative histogram of the deviations of measured to calculated Dst drops (bottom).
8. Simulated forecast

In this chapter we implement the results of Chapter 6 and 7 to a new developed forecast tool in order to evaluate their usefulness for real-time prediction. We ran the forecast tool for a period of four years from 2011/01 to 2014/12 at a cadence of six hours to simulate real-time prediction and compare our results to the statistical forecast algorithms of Rotter et al. (2015) and Vršnak, Temmer, and Veronig (2007b).

Section 8.1 describes the structure of the new developed forecast tool and the implemented forecast algorithms. In specific we implemented a peak velocity, a peak magnetic field density and a peak Dst index forecast based on the results of Chapter 7 and the statistical forecast algorithms for solar wind velocity by Rotter et al. (2015) and Dst index by Vršnak, Temmer, and Veronig (2007b). Section 8.2 presents the forecast results, i.e. the four-year time line of predicted and measured solar wind parameters and Dst index, an overview over the performance of the forecast tool in general, the evaluation procedure and the evaluation results of prediction rate of high speed streams and prediction accuracy of solar wind peak velocity, peak magnetic field density and peak Dst index.

8.1. Forecast tool

8.1.1. Structure

The forecast tool analyse the distribution of coronal holes on the solar disk to predict solar wind parameters. It comprises the data acquisition, the data preparation, the forecast itself and the output refinement. The main development objective was a modular programmed tool in the Interactive Data Lan-
8. Simulated forecast

Program: Forecast

Diagram:

- Header
  - Create Job Parameters
    - Look for existing data
    - Create new folders and files
  - Download Data
    - Download AIA-193
    - Download HMI-los
    - Download ACE
    - Download Dst
    - Check images
  - Extract Coronal Holes
    - Create bimarymap
    - Cut AIA-193 and HMI-los images
    - Calculate first and second order image statistics
    - Calculate shape parameters
    - Remove filaments
  - Calculate Forecast
    - Calculate solar wind velocity
    - Calculate solar wind magnetic field density
    - Calculate Dst index
  - Create Output
    - Create image of sun with coronal holes marked
    - Create forecast plots
    - Upload forecast

Figure 8.1.: Structure of forecast tool.
8.1. Forecast tool

guage[1](IDL), which can be run by different users with different configurations and different forecast algorithms at the same time.

Figure 8.1 represents the structure of the new developed forecast tool. It consists of 45 new developed routines which are distributed into five main modules: the initialisation, the download of AIA-193 images, HMI-los images, ACE data and Dst index, the extraction of coronal holes, the forecast algorithms and the creation of refined output data. The modules as well as the single routines were programmed as independent from each other as possible, which allows to swap the forecast algorithms and to easily implement future modifications. Each instance of the program is controlled by a user defined header file.

In the following we shortly describe the configuration options in the header file and the content of the five main modules.

**Header file**

The header parameter define (amongst others)

- the time range and cadence of the forecast,
- if the program runs as real-time forecast or as simulated forecast,
- if a new dataset shall be started or an old one continued,

- the image resolution which shall be used to identify coronal holes,
- if filaments classification shall be used and filaments shall be excluded,
- the program names of forecast algorithms to be used,

- the folders for locally available data,
- if not locally available data shall be downloaded,

8. Simulated forecast

- if the images shall be tested for quality, and if broken images shall be deleted,
- if already prepared data of previous runs shall be used or be recalculated,
- if refined output data shall be produced,
- the folders and file names for output data,
- if output data shall be uploaded to a server,
- the login information to upload output data to the server.

Initialisation

The header file is loaded and made available for all routines. All routines are compiled. If a new dataset shall be started, all existing temporary files of previous runs are deleted. Not existing folders are created.

Download of data

The defined folders are searched for locally available AIA-193 and HMI-los images, ACE data and Dst indices. Not locally available AIA-193 and HMI-los images are downloaded from JSOC, ACE data from Caltech and the Dst index from WDC-C2. The images are checked for the exposure time, the minimum and maximum values, and the mean values of the solar disk and of off-limb pixels in order to reject images which are underexposed, which are blurred and which have a partial solar occlusion. If an image was rejected, the next in time available image was downloaded and again checked.

Extraction of coronal holes

Coronal holes as well as filaments appear as dark structures in AIA-193 images. At first the AIA-193 and HMI-los images are prepared, i.e. they are normalized to the exposure time, rotate north-up, the scale of the HMI-los images is adapted to the scale of AIA-193 images, the sun in HMI-los images
is rotated to the recording time of the AIA-193 images, and they are re-binned to the requested resolution. Then the intensity-based thresholding technique described in Section 6.1 at an threshold of 0.38 times the median of the AIA-193 image pixels is applied to the AIA-193 images to create a binary map of dark objects. The dark objects are cut out in the binary map and in the corresponding AIA-193 and HMI-los images. All dark objects with less than 100 pixels are excluded. Next the classification algorithm for coronal holes and filaments described in Section 6.3.2 is applied to the dark objects in order to exclude filaments. The remaining dark objects are assumed to be coronal holes. Finally the coronal holes are used to create a full-disk binary map of coronal holes. The full-disk binary map as well as the AIA-193 and HMI-los images of the extracted coronal holes are saved locally.

Calculation of forecast

The full-disk binary map as well as the extracted coronal holes are used to calculate the solar wind forecast. The exact forecast algorithms are described in Section 8.1.2. The forecast results are saved locally.

Forecast output

AIA-193 images of the sun with coronal holes outlined and time-lines of the predicted and measured solar wind parameters and Dst indices are created and, if requested, uploaded to a publicly available server.

8.1.2. Forecast algorithms

We use the statistical relationships between coronal holes parameters, high speed streams parameters and the Dst index of chapter 7 to predict the solar wind peak velocity, the peak magnetic flux density in the shock and the peak Dst index from single coronal holes. In addition we use the statistical forecast algorithms developed by Rotter et al. (2015) and Vršnak, Temmer, and Veronig (2007b) to predict the time-lines of solar wind velocity and Dst index from the area coronal holes cover within a meridional slice. We denote the forecast
8. Simulated forecast

algorithms which use the single coronal holes as source as coronal hole based forecast algorithms (CHA), the forecast algorithms which use the area coronal holes cover within a meridional slice as coronal holes in slice based forecast algorithms (CHSA).

**CHA:**

**Solar wind peak velocity:**

We predict a high speed stream peak velocity from each coronal hole crossing the central meridian with an area of more than 1000 pixels. To predict the peak velocity, we use the relationship of Section 7.3, which correlates the solar wind peak velocity with the area which is corrected for projection effects and with the latitude of the coronal hole, but refitted the data with a fixed offset velocity of 400 km/s to replace the original offset velocity of 470 km/s to account also for small coronal holes. This replacement satisfies the fact that small coronal holes—which were not included in the dataset of Section 7.3—generally do not produce solar wind velocities as high as 470 km/s and reduces drastically the amount of erroneous predictions of high speed streams arising from small coronal holes, however the goodness of fit gets slightly worse. The re-fitted relation used to predict the peak velocities is given by

\[
v_{\text{forecast}}[\text{km/s}] = 400 + (3.34 \times 10^{-9} - 5.25 \times 10^{-11} \cdot |\varphi_{\text{CH}}[\degree]| - 2.17 \times 10^{-10} \cdot \sin((DOY + 80)/365 \cdot 2\pi)) \cdot A_{\text{CH}}[\text{km}^2].
\]  
(8.1)

\(A_{\text{CH}}\) is the area of the coronal hole crossing the central meridian and \(\varphi_{\text{CH}}\) the latitude of the coronal hole. The arrival time of the peak velocity is calculated by (see Sect. 7.2)

\[
t_{v, \text{forecast}} = t_{\text{CHmid}} + (7.66 - 0.0066 \cdot v_{\text{max}}[\text{km/s}])[\text{days}].
\]  
(8.2)

t_{\text{CHmid}} refers to the recording time of the image in which the distance of the center of the coronal hole to the central meridian was less than 3\degree. To avoid multiple peaks from the same coronal hole, we used a
simple tracking algorithm which checks if a center of a coronal hole in
the previous images was within a longitudinal distance of less than 3° to
the central meridian and within a latitudinal distance of less than 12° to
the center of the current coronal hole. If the center of a coronal hole is
more than one time within the meridional slice, we selected the highest
predicted peak velocity.

**Solar wind peak magnetic field density:**

We use Eq. 7.13 of Section 7.5 to predict the solar wind peak magnetic
flux densities averaged over ±6 hours \(B_t\) from the total magnetic flux
of each coronal holes with an area of more than 1000 pixels crossing the
central meridian:

\[
B_{\text{forecast}} [\text{G}] = -0.155 + 5 \cdot \text{sign}(\Phi_{\text{CH}}) + 1.63 \cdot 10^{-21} \cdot \Phi_{\text{CH}} [\text{Mx}].
\]  

(8.3)

The total magnetic flux \(\Phi_{\text{CH}}\) of coronal holes is calculated from the line-
of-sight photospheric magnetic flux of the single coronal holes. According
to Eq. 7.2 the arrival time of the peak magnetic flux density in the shock
is calculated to (see Sect. 7.2)

\[
t_{B, \text{forecast}} = t_{\text{CHmid}} + (7.974 - 0.0086 \cdot v_{\text{max}}[\text{km/s}]) [\text{days}].
\]  

(8.4)

\(v_{\text{forecast}}\) is estimated by the CHA algorithm for the solar wind peak veloc-
ity. \(t_{\text{CHmid}}\) is derived analogue to the CHA algorithm for the solar wind
velocity, but the highest magnetic field density peak was chosen.

**Peak Dst index:**

We use equation 7.16 of Section 7.6 to predict the maximum Dst drops
from each coronal hole with an area of more than 1000 pixels crossing
the central meridian:

\[
\text{Dst}_{\text{forecast}} [\text{G}] = \text{Dst}_{\text{quiet}} + (A [\text{km}^2]) \cdot (-6.88 \cdot 10^{-10} + 9.93 \cdot 10^{-12} \cdot |\varphi_{\text{CH}}[\circ]|) \cdot (1 - 2 \cdot 10^{-22} \cdot \Phi_{\text{CH}, z} [\text{G}]).
\]  

(8.5)

The \(z\)-component of the total magnetic flux in GSM coordinates \(\Phi_{\text{CH}, z}\)
is derived from the total magnetic flux \(\Phi_{\text{CH}}\) by projection under the
8. **Simulated forecast**

assumption of magnetic incident angles $\Phi_\lambda = -45^\circ$ and $\Phi_\phi = 0^\circ$ in HEEQ coordinates. The area is corrected for projection effects. To estimate $D_{st\,\text{quiet}}$ the Dst values of the preceding seven days are sorted in ascending order and the value at a position of $0.95 \cdot N$ is chosen, where $N$ is the number of Dst values. The delay is chosen to be the same as for the solar wind maximum velocity:

$$t_{D_{st,\,\text{forecast}}} = t_{CH\text{mid}} + (7.66 - 0.00661 \cdot v_{\text{max}}[\text{km/s}]) \text{ [days].} \quad (8.6)$$

$v_{\text{forecast}}$ is estimated by the CHA algorithm for the solar wind peak velocity. $t_{CH\text{mid}}$ is derived analogue to the CHA algorithm for the solar wind velocity, but the lowest Dst index peak was chosen.

**CHSA:**

**Solar wind velocity:**

We use the forecast algorithm developed by Rotter et al. (2015) (see Sect. 3.3) to predict the time-line of solar wind velocities. The fractional (not projection corrected) area of coronal holes within a slice of $\pm 7.5^\circ$ around the central meridian $A_{\text{Slice},\,7.5}$ is presumed to be proportional to the solar wind velocity. To forecast the velocity at a time $t$, a dataset of the fractional areas and measured solar wind velocities of the preceding three Carrington rotations is used. The time delay $\Delta t$ between fractional areas in the slice and solar wind velocities is calculated by searching for the maximum of the cross correlation coefficient between the fractional areas and the solar wind velocities within 5.5 days. If the time delay is less than 3.5 days or greater than 5.5 days, the time delay is set to 4.25 days. The fractional areas are shifted in time by the time delay to get the tuple $(A_{\text{Slice},\,7.5}, v)$. All fractional areas less than 0.1 as well as the six smallest and highest values of area and velocities are withdrawn, the remaining velocities are plotted versus the corresponding fractional areas. The diagonal of the bounding box of the velocity - fractional area plot defines the linear prediction function $v_{\text{forecast}}(t + \Delta t) = a \cdot A_{\text{Slice},\,7.5}(t) + b$. If the predicted velocity is greater than 800 km/s, the value is set to 800 km/s.
### Dst index:

We use the relationship of Vršnak, Temmer, and Veronig (2007b) which was derived for times of low solar activity to predict the time-line of Dst indices:

\[
Dst_{\text{forecast}}(t + 4 \text{ days}) [G] = (-65 \pm 25 \cos(\lambda_{\text{Earth}}(t))) \cdot (A_{\text{Slice, } 10}(t))^{0.5},
\]

where \( A_{\text{Slice, } 10} \) is the fractional area coronal holes cover within a longitudinal slice of \([-10^\circ, 10^\circ]\) around the central meridian, \( \lambda \) is the ecliptic longitude of Earth and the plus (minus) sign belongs to coronal holes of positive (negative) polarities. However the polarity in this equation is not defined distinctly if more than one coronal hole is within the slice.

We assign to the polarity the sign of the total photospheric magnetic flux coronal holes cover within the slice. The ecliptic longitude of Earth is derived by \( \lambda_{\text{Earth}}[^\circ] = DOY - 81 \), which is a rough but proper estimate for forecasting. The time delay between the measured fractional area coronal holes cover within the slice and the predicted Dst index is presumed to be four days.

### 8.2. Forecast results

#### 8.2.1. Overview

We let the forecast algorithms run on 2011/01 to 2014/12 at a cadence of six hours, an image resolution of 1024 \( \times \) 1024 pixels and with activated filament identification and exclusion. For the CHSA algorithms we started the forecast at 2010/10 in order to have the necessary preceding three Carrington rotations of data on 2011/01. To evaluate the impact of filament identification and exclusion to the forecasts, we also calculated the forecasts without filament exclusion as reference. The output is plotted in Figures 8.2 to 8.14. The first line of each page is a butterfly diagram of coronal holes (green and red) and filaments (grey), recorded within a longitudinal slice of \([-3^\circ, 3^\circ]\) around the central meridian. Coronal holes with a total magnetic flux of positive (neg-
8. Simulated forecast

ative) polarity are drawn in green (red). The second line shows the fractional area coronal holes cover within a longitudinal slice of $[-7.5^\circ, 7.5^\circ]$ around the central meridian, the third line the solar wind velocity, the fourth line the absolute value of the interplanetary magnetic field density at L1 and the fifth line the Dst index. The measured values are plotted in black, the CHSA forecast in green, the CHA forecast in red and times of ICMEs according to Richardson-Cane’s ICME-list\(^2\) (Cane and Richardson, 2003; Richardson and Cane, 2010) in purple. The CHSA forecast is natively a steady line, since a fractional area in the slice and therefore a predicted value can be assigned to each point in time. In contrast the CHA algorithms give discrete peaks, since it is based on single coronal holes crossing the central meridian. Note that the number of predicted peaks of the peak velocity, the peak magnetic field density and the peak Dst indices by CHA are identical, but that some predicted peaks with a very low magnitude cannot be resolved visually in the time-lines.

In total we extracted and classified 99,998 objects as coronal holes and 99,974 objects as filaments on the solar disk (Table 8.1). 98% of all extracted coronal holes and 99% of all filaments have areas less than $5 \cdot 10^{10}$ km\(^2\). These small objects appeared predominantly near the solar equator, whereas larger objects appeared at most at a latitude of about 60°. The ratio of filaments to coronal holes with areas less than $5 \cdot 10^{10}$ km\(^2\) is 1.02, the ratio with areas of $5 \cdot 10^{10}$ to $2 \cdot 10^{11}$ km\(^2\) is 0.11 and the ratio with areas greater than $2 \cdot 10^{11}$ km\(^2\) is 1.03.

The high ratio of filaments to coronal holes at large areas indicates that our classification algorithm does not work properly at large coronal holes. A cross-check of the large objects classified as filaments revealed that indeed all of the 31 large filaments are polar coronal holes. Having a more detailed look on the butterfly diagram of coronal holes and filaments in Figures 8.2 to 8.14, some objects which are both green or red and black, so which are classified both as filaments and coronal holes, attract our attention. This change of classification can happen if an object changes its properties like its shape and indicates that the classification algorithm does not work perfectly. This effect mostly appears at polar coronal holes, but also a few mid and low latitude objects were twin

\(^2\)http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm
8.2. Forecast results

Extracted Coronal Holes

<table>
<thead>
<tr>
<th>Area [km$^2$] / Latitude [$^\circ$]</th>
<th>&lt;15</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5e10</td>
<td>61737</td>
<td>8853</td>
<td>11724</td>
<td>5253</td>
<td>8505</td>
<td>1925</td>
</tr>
<tr>
<td>1e11</td>
<td>219</td>
<td>127</td>
<td>282</td>
<td>381</td>
<td>256</td>
<td>0</td>
</tr>
<tr>
<td>1.5e11</td>
<td>28</td>
<td>65</td>
<td>69</td>
<td>269</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>2e11</td>
<td>12</td>
<td>23</td>
<td>35</td>
<td>106</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>2.5e11</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Extracted Filaments

<table>
<thead>
<tr>
<th>Area [km$^2$] / Latitude [$^\circ$]</th>
<th>&lt;15</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>75</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5e10</td>
<td>89829</td>
<td>1552</td>
<td>2944</td>
<td>1531</td>
<td>3571</td>
<td>307</td>
</tr>
<tr>
<td>1e11</td>
<td>10</td>
<td>14</td>
<td>17</td>
<td>79</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>1.5e11</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>32</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2e11</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2.5e11</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 8.1.: Number of extracted coronal holes (top) and filaments (bottom), dependent on their area and latitude.

classified. If we recall that the dataset of which the classification algorithm was created only contains low latitude objects this outcome does not surprise.

The butterfly diagram also shows nicely the reversal of polarities of polar coronal holes near solar maximum in 2013: From 2011/01 to 2012/01 all coronal holes at the southern pole had positive polarities and at the northern pole negative polarities. Until 2012/06 most coronal holes at both poles had positive polarities, coronal holes near the solar equator negative polarities. Then the polarities started mixing. At 2014/06 most coronal holes at the southern pole had negative polarities, at the northern pole positive polarities.
Figure 8.2.: Solar wind and storm forecast for the period 2011/01 - 03. From top to bottom: butterfly diagram of coronal holes and filaments, fractional area coronal holes cover within a slice of $[-7.5^\circ, 7.5^\circ]$ around the central meridian, solar wind velocity, absolute magnetic field density and Dst index. In the butterfly diagram coronal holes with positive polarity are drawn in green, with negative polarity in red and filaments in black. Measured values are plotted in black, CHSA forecasts in green, CHA forecasts in red and times of ICMEs in purple.
Figure 8.3.: Same as Figure 8.2, but for 2011/04 - 06.
8. Simulated forecast

Figure 8.4.: Same as Figure 8.2, but for 2011/07 - 09.
8.2. Forecast results

Figure 8.5.: Same as Figure 8.2, but for 2011/10 - 12.
8. Simulated forecast

Figure 8.6.: Same as Figure 8.2, but for 2012/01 - 03.
8.2. Forecast results

Figure 8.7.: Same as Figure 8.2, but for 2012/04 - 06.
Figure 8.8.: Same as Figure 8.2, but for 2012/07 - 09.
8.2. Forecast results

Figure 8.9.: Same as Figure 8.2 but for 2012/10 - 12.
8. Simulated forecast

Figure 8.10.: Same as Figure 8.2, but for 2013/01 - 03.
8.2. Forecast results

Figure 8.11.: Same as Figure 8.2, but for 2013/04 - 06.
8. Simulated forecast

Figure 8.12.: Same as Figure 8.2, but for 2013/07 - 09.
Figure 8.13.: Same as Figure 8.2, but for 2013/10 - 12.
8. Simulated forecast

Figure 8.14.: Same as Figure 8.2, but for 2014/01 - 03.
8.2. Forecast results

Figure 8.15.: Same as Figure 8.2, but for 2014/04 - 06.
8. Simulated forecast

Figure 8.16.: Same as Figure 8.2, but for 2014/07 - 09.
8.2. Forecast results

Figure 8.17.: Same as Figure 8.2, but for 2014/10 - 12.
8. Simulated forecast

8.2.2. Evaluation procedure

Our objective is to evaluate the forecast output, i.e. how accurate we are able to predict high speed stream peaks and how many peaks we miss. We use the forecast correlation coefficient, the Pearson correlation coefficient and the root mean square error to evaluate the accuracy and count the right, the wrong and the erroneous predicted peaks. However the significance of the evaluation parameters depend on the way you extract and assign your peaks.

Before we define the evaluation procedure, we present some problems we face. Figure 8.18 shows four selected time ranges of coronal holes on the disk and the corresponding solar wind velocities. The first row shows two medium coronal holes plus a few small coronal holes. The corresponding solar wind velocity shows one large peak with many small peaks at decaying solar wind velocity. How many peaks do we see in fact? A definitive answer to this question is not possible. The second row shows a large, a medium and two small coronal holes near solar central meridian at 2012/01/24, two high measured solar wind velocity peaks on 01/25 12 UT and 01/27 12 UT, two low measured solar wind velocity peaks on 01/29 6 UT and 01/31 0 UT, two the high predicted solar wind velocity peaks at 01/28 18 UT and 01/31 0 UT and two low predicted velocity peaks at 01/25 18 UT and 01/31 12 UT. We can assume that the two large measured velocity peaks belong to the two large predicted peaks with an unusual high error in time of four days. It can also be that the large predicted peaks belong to the two small measured peaks and that we missed at least one large peaks. Anyway a distinct assignment seems not to be possible. Row three shows a single coronal hole on the disk, but two solar wind velocity peaks in the measured solar wind velocity. Here we are not able to say if it is in fact one measured peak which has a large dip on its top for any reason, if we missed a coronal hole completely or if the extracted coronal hole is in fact two coronal holes grown together. Row four shows many small coronal holes near the equator, at least three measured solar wind velocity peaks of about 400 km/s, a predicted peak of 450 km/s and three predicted peaks of about 400 km/s. If we state that high speed streams have velocities of at least 450 km/s we conclude that we erroneously predicted a not existing
Figure 8.18.: Problems of evaluation: Each row shows a binary map of coronal holes on the solar disk (left) the measured solar wind velocity (right, black line) and the predicted solar wind peak velocities by CHA (right, red lines).
peak. In fact three peak are visible which only did not reach 450 km/s. Thus we did not erroneously predict a peak but the measured solar wind velocity peak only did not reach our self-made limit of 450 km/s.

We made the experience that a completely correct extraction and assignment of peaks is impossible. We decided to determine and assign the peaks visual instead of automated. Because of the complexity of the task we think that the visual assignment results in less erroneous assignments than an automated assignment, however the objectivity suffers from visual assignment. The extraction and assignment of peaks was done in the following way:

**Extraction of peaks** All measured solar wind velocity peaks with a maximum velocity of more than 100 km/s higher than the surrounding solar wind velocity are extracted as peaks and denoted as „distinct peaks“ . In addition we matched the number of coronal holes with the number of solar wind velocity peaks, i.e. for a given time \( t \) we counted all solar wind velocity maxima within \([t - 5 \text{ days}, t + 5 \text{ days}]\) and the number of coronal holes within \([t - 10 \text{ days}, t]\). We excluded all coronal holes for which CHA did not predict at least 420 km/s to exclude small coronal holes. If the number of coronal holes was larger than the number of distinct velocity peaks, we added as many of the highest measured solar wind velocity peaks till we matched the number of coronal holes.

We picked all predicted solar wind velocity peaks from CHA and CHSA which peak velocities were at least 100 km/s higher than the surrounding predicted velocity. To each distinct measured solar wind velocity peak at time \( t \) we picked the highest measured and predicted solar wind absolute magnetic field density within \([t - 3 \text{ days}, t]\) and the lowest measured and predicted Dst peak within \([t - 2 \text{ days}, t + 3 \text{ days}]\). Finally we excluded all peaks at times of ICMEs (according to Richardson-Cane’s ICME list).

**Assignment of peaks** For a given time \( t \) we counted the measured solar wind peaks with a solar wind velocity of at least 450 km/s and the CHSA peaks without velocity restriction within \([t - 5 \text{ days}, t + 5 \text{ days}]\). We matched the number of peaks by selecting the highest peaks and assigned the peaks in time, i.e. the first CHSA peak was assigned to the first measured solar
wind peak, etcetera. We denote the matched peaks as true positive peaks (TP). We counted the predicted peaks with velocities of at least 450 km/s which could not be assigned to the measured peaks and denote them as false positive peaks (FP). We counted the measured peaks with a velocity of at least 450 km/s which could not be assigned to a predicted peak and denote them as false negative peaks (FN). The measured and predicted Dst and magnetic field density peaks were assigned to the corresponding measured distinct solar wind velocity peak. The same procedure was done for the CHA algorithm.

The selected peaks used for evaluation of CHSA are printed in Figures 8.19 to 8.20, the peaks used for evaluation of CHA in Figures 8.21 to 8.22. TP peaks used for evaluation are drawn as black, green and red crosses, FP peaks as black stars and FN peaks as red and green stars. In total we used 183 peaks for evaluation of the CHA algorithm and 191 peaks for evaluation of the CHSA algorithm.
Figure 8.19.: Peaks used for evaluation of CHSA from 2011/01 to 2012/12: the measured solar wind velocity averaged over ±6 hours is drawn in black, the forecast by CHSA averaged over ±6 hours in green. Crosses denote the selected TP peaks of the un-averaged data, black stars the FP peaks and green stars the FN peaks.
8.2. Forecast results

Figure 8.20.: Same as 8.19 but for 2013/01 to 2014/12.
8. Simulated forecast

Figure 8.21.: Peaks used for evaluation of CHA from 2011/01 to 2012/12: the measured solar wind velocity averaged over ±6 hours is drawn in black, the selected peaks used for evaluation of the forecast by CHA in red. Crosses denote the selected peaks of the un-averaged data.
8.2. Forecast results

Figure 8.22.: Same as 8.21 but for 2013/01 to 2014/12.
8. Simulated forecast

8.2.3. Forecast rate of solar high speed streams

Each high-speed stream has one distinct velocity peak. Therefore we evaluate the rate of predicted high speed streams of CHA and CHSA by counting the numbers of TP, FP and FN solar wind velocity peaks, i.e. the number of correct predicted, not predicted and erroneous predicted velocity peaks. Table 8.2 shows the raw numbers for both algorithms, split into velocity ranges of 450 to 550 km/s, 550 to 650 km/s and larger than 650 km/s. In total we counted 172 velocity peaks with a peak velocity of more than 450 km/s in the in-situ measurements (= TP + FP), whereof 103 peaks had maximum velocity of 450 to 550 km/s („small“), 53 had maximum velocities of 550 to 650 km/s („medium“) and 16 had peak velocities greater than 650 km/s („strong“). The CHA algorithm predicted correctly 84% of the small peaks, 91% of the medium peaks and 94% of the strong peaks, the CHSA algorithm 84% of the small peaks, 81% of the medium peaks and 88% of the strong peaks. In addition the CHA algorithm predicted 12 peaks erroneously (5 small, 2 medium, 4 strong), the CHSA algorithm 19 peaks (12 small, 5 medium, 2 strong).

Figure 8.23 shows the latitude versus the area of coronal holes belonging to the TP peaks of CHA, the corresponding measured peak velocities are coloured. The scatter plot reveals that most of the measured TP high speed stream peaks can be explained by coronal holes with a heliospheric latitude within ±60°.

To investigate the FN peaks, we examined the full disk binary maps of coronal holes created by the forecast tool and the corresponding Hα images of KSO of four days before of each erroneously predicted peak (Table 8.3). Of the 11 erroneously predicted peaks of CHA 3 peaks could be distinctly assigned to filaments which were erroneously classified as coronal holes by the classification algorithm and 3 peaks to polar coronal holes which extend to low latitudes. Of the 19 erroneously predicted peaks of CHSA 3 peaks could be assigned to filaments, 6 peaks to polar coronal holes and 3 peaks to many small coronal holes crossing the central meridian at almost the same time which induces a medium fractional area coronal holes cover within the meridional slice. The cause of the remaining erroneously predicted velocity peaks could not be determined.
8.2. Forecast results

<table>
<thead>
<tr>
<th>Peak v [km/s]</th>
<th>CHA</th>
<th>CHSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP 450 ≤ $v_{\text{max}}$ &lt; 550</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>FP 450 ≤ $v_{\text{max}}$ &lt; 550</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>FN 450 ≤ $v_{\text{forecast}}$ &lt; 550</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>TP 550 ≤ $v_{\text{max}}$ &lt; 650</td>
<td>48</td>
<td>43</td>
</tr>
<tr>
<td>FP 550 ≤ $v_{\text{max}}$ &lt; 650</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>FN 550 ≤ $v_{\text{forecast}}$ &lt; 650</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>TP $v_{\text{max}}$ &gt; 650</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>FP $v_{\text{max}}$ &gt; 650</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>FN $v_{\text{forecast}}$ &gt; 650</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 8.2.: Number of TP, FP and FN peaks for the CHA and CHSA algorithms. The numbers are split into measured ($v_{\text{max}}$), respectively predicted ($v_{\text{forecast}}$) peak velocities in the range of 450 to 550 km/s, 550 to 650 km/s and larger than 650 km/s.

Note that the CHSA algorithm creates more erroneously predicted peaks due to polar coronal holes than the CHA algorithm. This is because the CHA algorithm has included a dependency on the latitude of coronal holes which reduces predictions from polar coronal holes. In contrast the CHSA algorithm predicts peaks by measuring the fractional area coronal holes cover within a longitudinal slice around the central meridian, therefore it generates peaks from polar coronal holes crossing the central meridian. At the same time the latitude dependence of the CHA algorithm results in that large polar coronal holes which extend down from mid to low latitudes create the largest erroneous peaks in the CHA algorithm ($v_{\text{forecast}} > 1000$ km/s). Because of their large latitudinal extension, these coronal holes are not seen as polar but as mid latitude objects with a huge area, which results in strong erroneously predicted peaks by CHA.
8. Simulated forecast

Figure 8.23.: Latitude versus area of the coronal holes belonging to the TP peaks of CHA. The measured peak velocities of TP peaks are coloured.

<table>
<thead>
<tr>
<th>Cause</th>
<th>Algorithm</th>
<th>Numbers</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filaments</td>
<td>CHA</td>
<td>3</td>
<td>2012/04/04, 2013/03/09, 2013/04/08</td>
</tr>
<tr>
<td></td>
<td>CHSA</td>
<td>3</td>
<td>2012/04/04, 2013/03/09, 2013/04/08</td>
</tr>
<tr>
<td>Polar coronal holes</td>
<td>CHA</td>
<td>3</td>
<td>2012/01/01, 2013/12/27, 2014/10/07</td>
</tr>
<tr>
<td></td>
<td>CHSA</td>
<td>6</td>
<td>2012/01/01, 2013/06/03, 2013/07/07, 2013/12/27, 2014/03/09, 2014/10/07</td>
</tr>
<tr>
<td>Many small objects near the central meridian</td>
<td>CHSA</td>
<td>3</td>
<td>2012/12/09, 2013/02/16, 2014/12/18</td>
</tr>
</tbody>
</table>

Table 8.3.: Numbers and dates of FN peaks of CHA and CHSA for which the cause for erroneous predictions could be determined.
8.2. Forecast results

8.2.4. Prediction accuracy of high-speed stream peak velocities

Table 8.4 shows the calculated evaluation parameters for the TP peak velocities of the CHSA and CHA algorithms, i.e. the Pearson correlation coefficient, the forecast correlation coefficient, the mean value of deviations, the standard deviation and the root mean square error of predicted to measured high-speed stream peak velocities, for the whole time range as well as per year. The Pearson correlation coefficient of CHSA for the whole time range is 0.58, the forecast correlation coefficient is 0.51, the mean deviation is $(−48 \pm 97)$ km/s and the root mean square deviation 108 km/s. The Pearson correlation coefficients for the individual years are 0.44 for 2011, 0.66 for 2012, 0.51 for 2013 and 0.70 for 2014, the mean deviation and standard deviations are $(−7 \pm 97)$ km/s for 2011, $(−86 \pm 74)$ km/s for 2012, $(−19 \pm 100)$ for 2013 and $(−93 \pm 85)$ for 2014. The correlation coefficients are reasonable, but the predicted values are especially for 2012 and 2014 statistically about 90 km/s too low.

Figure 8.24 shows the scatter plots of the predicted versus the measured high speed stream peak velocities of the CHSA algorithm for the whole time range as well as per year. All plots show a clear correlation between predicted and measured peak velocities at a high spreading. The top panel of Figure 8.25 shows the histogram of the deviations of measured to predicted velocity peaks. The histogram is not centred at zero, but shifted to higher values, which shows again that the predicted peak velocities are statistically too low. In total, 32% of all deviations are less than 50 km/s and 54% are less than 100 km/s.

Table 8.5 shows the evaluation parameters of the arrival time of high speed stream peak velocities for the CHSA and CHA algorithm, i.e. the mean deviation, the standard deviation and the root mean square error of measured to predicted arrival times. The mean deviation and the standard deviations for CHSA are $−5 \pm 25$ hours for the whole time range, $−3 \pm 24$ hours for 2011, $−9 \pm 18$ hours for 2012, $−7 \pm 30$ hours for 2013 and $−1 \pm 28$ hours for 2014. Therefore the predicted peak velocities are slightly predicted too early.

The middle panels of Figure 8.25 shows the deviations in arrival time of measured to predicted peak velocities versus the peak velocities. The scatter
plot is wide spread and does not show a statistical dependency on the peak velocities. This was unexpected because the CHSA algorithm uses a slowly varying, almost constant travel time of about 4 days, whereas the travel time of the peak velocity depends on the peak velocity according to Section 7.2. Therefore we expected a dependence of deviations of arrival time on the peak velocity. We cannot tell why the CHSA algorithm does not show the expected dependence. The bottom panel of Figure 8.25 shows the histogram of the deviations in arrival time. The arrival times are predicted statistically 5 hours too late. 75% of the predicted arrival times have deviations less than 1 day, 93% have deviations less than 2 days.

For the CHA algorithm the Pearson correlation coefficient of predicted to measured peak velocities for the whole time range is 0.44, the forecast correlation coefficient is 0.27, the mean deviation is $(-46 \pm 90)$ km/s and the root mean square deviation 102 km/s (Table 8.4). The Pearson correlation coefficients for the individual years are 0.42 for 2011, 0.54 for 2012, 0.67 for 2013 and 0.16 for 2014, the mean deviation and standard deviations are $(-58 \pm 93)$ km/s for 2011, $(-71 \pm 74)$ km/s for 2012, $(-32 \pm 69)$ km/s for 2013 and $(-29 \pm 123)$ km/s for 2014. The correlation coefficients are reasonable for 2011 to 2013, the predicted values are statistically somewhat too low for each year.

Figure 8.26 shows the predicted versus the measured high speed stream peak velocities of the CHA algorithm for the whole time range as well as per year. The data are coloured in two folds. Red data points belong to peaks which were also used in the dataset of Chapter 7, black points represent new peaks. Almost all red data points lie near the diagonal which represents a perfect forecast. These data points present the goodness of fit of the velocity - area - latitude dependence of section 7.3 and are not suitable for evaluation. The black data points have a wider spreading and belong to statistically too low predicted peak velocities in 2011 to 2013. Some of them complement the distribution of the red data points, but some also do not fit to the distribution. In 2014—the only year we can really use for evaluation—about 23 data points lie near to the diagonal, 2 are predicted much too high and about 3 are predicted too
low. We can argue that the distribution of 2014 does not show any correlation as well as that 23 data lie near the one-to-one correspondence with 5 outliers, therefore we cannot give a final judgement on 2014. Anyway, the black data in 2011 to 2013 show that the statistical relationship of Section 7.3—from which the CHA algorithm was developed—does not hold in general. The top panel of Figure 8.27 shows the histogram of the deviations of measured to predicted velocity peaks. In total, 40% of all deviations are less than 50 km/s and 71% less than 100 km/s.

For the CHA algorithm the mean deviation and the standard deviations of the arrival times of predicted to measured peak velocities are $-6 \pm 36$ hours for the whole time range, $-6 \pm 30$ hours for 2011, $-12 \pm 20$ hours for 2012, $-3 \pm 35$ hours for 2013 and $-3 \pm 46$ hours for 2014 (Table 8.5). Therefore the predicted peak velocities are slightly predicted too early.

The middle panels of Figure 8.27 shows the deviations in arrival time of measured to predicted peak velocities versus the peak velocities. The scatter plot is wide spread and does not show a statistical dependency on the peak velocities which was expected because the CHA algorithm considers the dependency of the arrival time of peak velocity on the peak velocity. The bottom panel of Figure 8.27 shows the histogram of the deviations in arrival time. The arrival times are predicted statistically 6 hours too late. 60% of the predicted arrival times have deviations less than 1 day, 86% have deviations less than 2 days.

Altogether the CHSA algorithm shows a clear correlation of predicted to measured peak velocities in each year, but the data is wide spread. The CHA algorithm predicts peak velocities which were not used in the dataset of Chapter 7 too low, therefore the dependencies for the peak velocity of Chapter 7 do not hold in general. Both algorithm predict the arrival times slightly too early, the deviations of predicted and measured arrival time are independent on the peak velocity.
8. Simulated forecast

### Evaluation: CHSA peak velocities

<table>
<thead>
<tr>
<th>Year</th>
<th>$r_{Pearson}$</th>
<th>$r_{Forecast}$</th>
<th>$\Delta v$ [km/s]</th>
<th>$\sigma$ [km/s]</th>
<th>RMSE [km/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>0.44</td>
<td>0.30</td>
<td>-7</td>
<td>±97</td>
<td>96</td>
</tr>
<tr>
<td>2012</td>
<td>0.66</td>
<td>0.62</td>
<td>-86</td>
<td>±74</td>
<td>112</td>
</tr>
<tr>
<td>2013</td>
<td>0.51</td>
<td>0.39</td>
<td>-19</td>
<td>±100</td>
<td>100</td>
</tr>
<tr>
<td>2014</td>
<td>0.70</td>
<td>0.62</td>
<td>-93</td>
<td>±85</td>
<td>125</td>
</tr>
<tr>
<td>2011 - 2014</td>
<td>0.58</td>
<td>0.51</td>
<td>-48</td>
<td>97</td>
<td>108</td>
</tr>
</tbody>
</table>

### Evaluation: CHA peak velocities

<table>
<thead>
<tr>
<th>Year</th>
<th>$r_{Pearson}$</th>
<th>$r_{Forecast}$</th>
<th>$\Delta v$ [km/s]</th>
<th>$\sigma$ [km/s]</th>
<th>RMSE [km/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>0.42</td>
<td>0.33</td>
<td>-58</td>
<td>±93</td>
<td>107</td>
</tr>
<tr>
<td>2012</td>
<td>0.54</td>
<td>0.37</td>
<td>-71</td>
<td>±74</td>
<td>99</td>
</tr>
<tr>
<td>2013</td>
<td>0.67</td>
<td>0.54</td>
<td>-32</td>
<td>±69</td>
<td>78</td>
</tr>
<tr>
<td>2014</td>
<td>0.16</td>
<td>0.01</td>
<td>-29</td>
<td>±123</td>
<td>123</td>
</tr>
<tr>
<td>2011 - 2014</td>
<td>0.44</td>
<td>0.27</td>
<td>-46</td>
<td>±90</td>
<td>102</td>
</tr>
</tbody>
</table>

Table 8.4.: Evaluation of CHSA (top) and CHA (bottom) peak velocities: Pearson correlation coefficient, forecast correlation coefficient, mean deviation, standard deviation and root mean square deviations of measured to predicted high-speed streams velocity peaks.
8.2. Forecast results

Evaluation: CHSA arrival times of peak velocities

<table>
<thead>
<tr>
<th>Year</th>
<th>$\Delta v$ [hours]</th>
<th>$\sigma$ [hours]</th>
<th>RMSE [hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>-3</td>
<td>±24</td>
<td>24</td>
</tr>
<tr>
<td>2012</td>
<td>-9</td>
<td>±18</td>
<td>19</td>
</tr>
<tr>
<td>2013</td>
<td>-7</td>
<td>±30</td>
<td>30</td>
</tr>
<tr>
<td>2014</td>
<td>-1</td>
<td>±28</td>
<td>28</td>
</tr>
<tr>
<td>2011 - 2014</td>
<td>-5</td>
<td>±25</td>
<td>26</td>
</tr>
</tbody>
</table>

Evaluation: CHA arrival times of peak velocities

<table>
<thead>
<tr>
<th>Year</th>
<th>$\Delta v$ [hours]</th>
<th>$\sigma$ [hours]</th>
<th>RMSE [hours]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>-6</td>
<td>±30</td>
<td>31</td>
</tr>
<tr>
<td>2012</td>
<td>-12</td>
<td>±20</td>
<td>23</td>
</tr>
<tr>
<td>2013</td>
<td>-3</td>
<td>±35</td>
<td>35</td>
</tr>
<tr>
<td>2014</td>
<td>-3</td>
<td>±46</td>
<td>47</td>
</tr>
<tr>
<td>2011 - 2014</td>
<td>-6</td>
<td>±36</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 8.5.: Evaluation of CHSA (top) and CHA (bottom) arrival times of peak velocities: Pearson correlation coefficient, forecast correlation coefficient, mean deviation, standard deviation and root mean square deviations of predicted to measured arrival times of high-speed stream peak velocities.
8. Simulated forecast

Figure 8.24.: CHSA: Predicted versus measured high-speed stream velocity peaks. At the four bottom plots the data was split up to the individual years. The straight line represents the one-to-one correspondence.
8.2. Forecast results

Figure 8.25.: CHSA: Histograms of the deviations of measured to predicted high-speed stream velocity peaks (top). Deviations of the measured to predicted arrival times of the velocity peaks versus $v_{\text{max}}$ (mid), and histogram of the deviations of measured to predicted arrival times of the velocity peaks (bottom).
Figure 8.26.: CHA: Predicted versus measured high-speed stream velocity peaks. At the four bottom plots the data was split up to the individual years. Red data points belong to peaks which were also used in the dataset of Chapter 7. The straight line represents the one-to-one correspondence.
Figure 8.27.: CHA: Histograms of the deviations of measured to predicted high-speed stream velocity peaks (top). Deviations of the measured to predicted arrival times of the velocity peaks versus \( v_{\text{max}} \) (mid), and histogram of the deviations of measured to predicted arrival times of the velocity peaks (bottom).
8. Simulated forecast

8.2.5. Accuracy of solar wind magnetic field density

Table 8.6 show the evaluation parameters for the peak magnetic field density forecast of the shocks of high-speed streams by CHA, i.e. the Pearson correlation coefficient of predicted to measured magnetic field densities, the Pearson correlation coefficient, the mean deviation, the standard deviation and the root mean square error of absolute predicted to absolute measured magnetic field density and the number of peaks which we predicted with correct polarity. The polarity of the solar wind magnetic field density is derived under the assumption that the magnetic incident angle $B_\lambda$ is between $(-45 \pm 90)^\circ$ and $B_\varphi$ is between $\pm 90^\circ$. Figure 8.28 shows the predicted versus the measured magnetic field densities. All parameters and plots are presented per year as well as for the whole time range.

The Pearson correlation coefficient for the whole time range is 0.46 and we predicted the correct polarity of magnetic field density peaks for 70\% of the data. Since we predict the correct polarity for a bigger part of the data, the data in the plots are mostly distributed in the upper right (positive polarities) and lower left (negative polarities). Since the Pearson correlation coefficient is a measure of the distribution of the data to its regression line and the data is mostly distributed in two opposed sectors in the plot, the Pearson correlation coefficient cannot be small. Therefore the medium value of 0.46 arises from the correctly predicted polarities. It does not state that we are able to predict the magnitude of the magnetic field density.

To evaluate the predicted magnitude of the magnetic field density peaks, we calculated the Pearson correlation coefficient of the absolute predicted magnetic field density peaks to the absolute measured magnetic field density peaks and got $r_P = 0.15$. The mean value of the deviations of absolute predicted to absolute measured magnetic field density is $(-0.36 \pm 2.40)$ G. From this follows that we are not able to predict the magnitude of magnetic field density peaks in general.

If we look at the individual years, the year 2013 attracts our attention. While 2011, 2012 and 2014 have Pearson correlation coefficients of the absolute magnetic field density of -0.09, 0.09 and 0.07, the year 2013 has a Pearson
### 8.2. Forecast results

Evaluation: CHA peak magnetic field densities

| Year      | $r_{\text{Pearson}}(B)$ | $r_{\text{Pearson}}(|B|)$ | $\Delta |B|$ [G] | $\sigma$ [G] | RMSE($|B|$) [G] | Pol |
|-----------|-------------------------|----------------------------|-------------|-------------|----------------|-----|
| 2011      | 0.63                    | -0.09                      | -0.59       | $\pm 2.8$   | 2.9            | 76% |
| 2012      | 0.40                    | 0.09                       | -0.62       | $\pm 2.5$   | 2.5            | 67% |
| 2013      | 0.58                    | 0.46                       | -0.39       | $\pm 2.0$   | 2.0            | 75% |
| 2014      | 0.34                    | 0.07                       | -0.69       | $\pm 2.3$   | 2.4            | 63% |
| 2011 - 2014 | 0.46                  | 0.15                       | -0.36       | $\pm 2.4$   | 2.4            | 70% |

Table 8.6.: Evaluation of CHA peak magnetic field densities: Pearson correlation coefficient of the measured to the predicted magnetic field density peaks, Pearson correlation coefficient, mean deviation, standard deviation, root mean square deviations of the absolute predicted to the absolute measured magnetic field density peaks and the percentage of correctly predicted polarities.

A correlation coefficient of 0.46. Therefore the CHA forecast of the magnetic field density peaks achieved reasonable results in year 2013, but bad results in 2011, 2012 and 2014.
Figure 8.28.: CHA: Predicted versus measured magnetic field density peaks. At the four bottom plots the data was split up to the individual years. Red data points belong to peaks which were also used in the dataset of Chapter 7. The straight line represents the one-to-one correspondence.
8.2. Accuracy of Dst index

The top panel of Table 8.7 show the evaluation parameters of predicted to measured Dst peaks for CHSA, i.e. the Pearson correlation coefficient, the forecast correlation coefficient, the mean deviation, the standard deviation and the root mean square error of predicted to measured Dst peaks, Figure 8.29 show the predicted versus the measured Dst peaks. All parameters and plots are presented per year as well as for the whole time range. For the whole time range the CHSA algorithm achieved a Pearson correlation coefficient of 0.21, a forecast correlation coefficient of 0.08 and a mean deviation of \((-15 \pm 28)\) G. The corresponding plot shows a very wide-spread data distribution which is statistically located at well too low predicted peak values and reveals that the forecast is not reasonable. If we look to the individual years, the results get better. The Pearson correlation coefficients are 0.33 for 2011, 0.26 for 2012, 0.35 for 2013 and -0.01 for 2014, the forecast correlation coefficients are 0.23 for 2011, 0.12 for 2012, 0.17 for 2013 and 0.00 for 2014. The scatter plots of 2011 to 2013 show that the data distribution indeed show a wide spread correlation between predicted and measured Dst peaks in each year, but that the orientation of the data distribution differs for each year. We therefore suppose that an adaptive algorithm similar to the CHSA algorithm for the solar wind velocity could achieve much better results.

The bottom panel of Table 8.7 show the evaluation parameters of predicted to measured Dst peaks for CHA, Figure 8.30 shows the predicted versus the measured Dst peaks. The Pearson correlation coefficient for the whole time range is 0.16, the forecast correlation coefficient 0.03 and the mean deviation is \((12 \pm 23)\) G. The Pearson correlation coefficients of the individual years are 0.18 for 2011, 0.06 for 2012, 0.28 for 2013 and 0.07 for 2014. Since also in the corresponding scatter plots clear correlations are not visible, we conclude that the CHA forecast for peak Dst indices does not work in general.
Table 8.7.: Evaluation of CHSA (top) and CHA (bottom) peak Dst index: Pearson correlation coefficient, forecast correlation coefficient, mean deviation, standard deviation and root mean square deviations of measured to predicted Dst peaks.
8.2. Forecast results

Figure 8.29.: CHSA: Predicted versus measured Dst peaks. At the four bottom plots the data was split up to the individual years. The straight line represents the one-to-one correspondence. The red line represents an assumed orientation of the data distribution.
Figure 8.30.: CHA: Predicted versus measured Dst peaks. At the four bottom plots the data was split up to the individual years. Red data points belong to peaks which were also used in the dataset of Chapter 7. The straight line represents the one-to-one correspondence.
9. Discussion

In Chapter 6 we investigated the characteristics of coronal holes and filaments with a center of the binary object within ±30° in heliospheric longitude and latitude for a 3-year period from 2011 to 2013. We found that coronal holes have a mean AIA-193 intensity of (40.11 ± 6.95) DNs, a mean absolute magnetic field density of (2.69 ± 1.61) G, a relative open magnetic flux of (42 ± 9) %, that more than 60% of the open magnetic flux arises from less than 2.5% of the coronal hole area and that the pixel distribution of coronal holes has a dominant polarity over the magnetic field density scale, i.e. the polarity of coronal hole pixels with a magnetic field density of less than 10 G (30 G) is always the same polarity as the polarity of coronal hole pixels with a magnetic field density of more than 10 G (30 G). The number of flux tubes, determined by a thresholding technique on the line-of-sight magnetic field maps of coronal holes, is linear dependent on the area of coronal holes. In general coronal holes appeared at both hemispheres at both polarities. Coronal holes with a small relative open magnetic flux only appeared next to the solar equator, large coronal holes appeared predominantly on the northern hemisphere and had negative polarities. Note that the polarities of polar coronal holes have changed between the years 2012/06 and 2014/06 according to the butterfly diagram of coronal holes in Chapter 8. Therefore the given parameters belong to the time of the polarity reversal of coronal holes near solar maximum.

The characteristics of filaments revealed that they had a mean AIA-193 intensity of (52.69 ± 7.57) DNs, a mean magnetic flux density of (0.36 ± 0.31) G, a relative open magnetic flux of (3 ± 3) %, and that filaments on the northern hemisphere had predominantly positive polarities and on the southern hemisphere negative polarities. They therefore appear brighter in AIA-193 images, have a smaller magnetic flux density, a negligible relative open magnetic flux
and a different localisation dependent on their polarity than coronal holes. Note that the extraction algorithm for dark objects in AIA-193 images was tuned to detect coronal holes, therefore the results of filaments should be seen as estimate.

Based on these results, my colleagues Martin Reiss and Ruben de Visscher showed that an automated classification of coronal holes and filaments is possible (TSS ≈ 0.90; Reiss et al., 2014b), but a final classification rule has not been extracted until the end of this work.

In Chapter 7 we investigated the statistical correlations between the peak velocity of high speed streams at L1, the peak density and peak magnetic field density in the corresponding shock at L1, their travel times, the peak Dst drop and its time of appearance with coronal hole parameters on the Sun. The dataset contained 49 coronal holes and the corresponding measured solar wind data at L1, whereby at each time only one large coronal hole was allowed to appear near the central meridian, in a 3-year period from 2011 to 2013. To determine the travel times and the points of origin of high speed streams, we postulated Sun’s central meridian as starting position of high speed streams travelling through L1. We determined the statistical point of origin of the peak velocity and the point of origin of the smallest trailing velocity in the measured time line of a high speed stream at L1 at \( X = 0.36 \), where \( X \) is the position behind the leading edge of the coronal hole normalized on its longitudinal width (\( X = 0 \): leading edge, \( X = 1 \): trailing edge). The travel time of the peak velocity depends on the peak velocity and lies typically between 2 and 6 days, the RMSE of our fit is 1.2 days. The travel time of the smallest velocity depends on the smallest velocity and lies between 5 and 10 days, the RMSE of our fit is 1.3 days. If we choose the trailing edge as point of origin of the smallest velocity, we found that its travel time depends on the longitudinal width of coronal holes and concluded that it is unlikely that the point of origin is the trailing edge. The peak density and magnetic field density in the shock arrives about 1 day earlier than the peak velocity, the RMSE of our fit is 0.8 days. We estimated the points of origin of the peak velocity and smallest velocity from the measured arrival times and the velocities for each coronal hole.
by assuming a constant radial velocity and found that most of the points of origin lie within the coronal hole for the peak velocity, but behind the trailing edge of coronal holes for the smallest velocity. Therefore we assume that the acceleration time of solar wind particles plays an underpart on the total travel time for particles belonging to the peak velocity, but plays a substantial part for particles belonging to the smallest velocities. The time of appearance of the peak Dst drop does not depend statistically on the solar wind absolute magnetic field density nor on the peak velocity, but on the peak density.

The peak velocity of solar high speed streams depends on the area and latitude of coronal holes. Coronal holes at the equator have statistically the highest peak velocity per area ratio, at about 60° the peak velocity per area ratio becomes zero. We also found a not-significant dependency of the peak velocity on the day of year: the peak velocity per area ratios are somewhat lower near winter solstice where the distance Sun-Earth is largest and somewhat higher near summer solstice where the distance is lowest. Our fit resulted in a RMSE of 65.2 km/s, the Pearson correlation coefficient between the fitted and measured peak velocities is 0.71, the forecast correlation coefficient 0.58. The deviations of fitted and measured velocity peaks are dependent on the relative open magnetic flux of coronal holes, a relative open magnetic flux results in less deviations.

We did not find a dependency of the peak density in the shock of high speed streams on coronal hole parameters.

The solar wind magnetic field density averaged over ±6 hours shows a dependence on the total magnetic flux of coronal holes, the Pearson correlation coefficient of our fitted to the measured magnetic field densities is 0.86, the forecast correlation coefficient is 0.69. The polarity of the solar wind magnetic field density coincide with the polarity of coronal holes for 88% of the data. However the Pearson correlation coefficient of the absolute magnetic field density of the solar wind to the absolute total magnetic flux of coronal holes is only 0.18, therefore an estimate on the magnitude of magnetic field density remains difficult. We are not able to predict the magnetic incident angles of the solar wind at L1. We also tried to estimate the z-component of the solar
9. Discussion

wind magnetic field density in GSM coordinates by deriving the "z-component" of the total magnetic flux of coronal holes under the assumption of magnetic incident angles $\Phi_\lambda = -45^\circ$ and $\Phi_\varphi = 0^\circ$ and achieved a Pearson correlation coefficient of 0.40, however the spreading is much too high that this fit can be used as forecast (RMSE = 10.6 G).

The peak Dst drop depends on the area, latitude and magnetic field density of coronal holes. A manual fit by maximising the forecast correlation coefficient between fitted and measured Dst drops resulted in a Pearson correlation coefficient of fitted to measured Dst drops of 0.53 at an RMSE of 16.5 G.

In Chapter 8 we applied the results of Chap. 7 as forecast for a 4-year period from 2011 to 2014 and compared the results to the forecast algorithms by Rotter et al. (2015) and Vršnak, Temmer, and Veronig (2007b), which predicted the solar wind velocity and Dst index time line by measuring the fractional area coronal holes cover within a longitudinal slice around the central meridian. For the CHA forecast algorithms, which are based on the results of Chap. 7 we found that solar wind peak velocities peaks which were not used in Chapter 7 are often predicted too low. The Pearson correlation coefficient of predicted to measured velocity peaks is 0.44, the forecast correlation coefficient 0.27 and the mean deviation is $(-46 \pm 90)$ km/s. For the year 2013 we achieved especially good results with a Pearson correlation coefficient of 0.67, a forecast correlation coefficient of 0.54 and a mean deviation of $(-32 \pm 69)$ km/s. The accuracy in time was within $(-6 \pm 36)$ hours. However the scatter plots of the predicted to measured peak velocities showed that the dependency found in Chapter 7 do not hold in general. We predicted 70% of the polarities of the solar wind magnetic field density peaks correct, but failed in general in the magnitude. The Pearson correlation coefficient of the absolute predicted magnetic field density to the absolute measured magnetic field density averaged over ±6 hours is 0.15, the mean deviation is $(-0.36 \pm 2.40)$ G. In the year 2013 we achieved especially good results, the Pearson correlation coefficient of the absolute predicted and measured data is 0.46 at a mean deviation of $(-0.39 \pm 2.00)$ G. The forecast of Dst index also failed, the Pearson correlation coefficient between predicted and measured Dst indices is 0.16, the mean
deviation is $(13 \pm 23)$ G.

For the adaptive CHSA solar wind velocity forecast algorithm by Rotter et al. (2015) we found a Pearson correlation coefficient of predicted to measured peak velocities of 0.58, a forecast correlation coefficient of 0.51 and a mean deviation of $(-48 \pm 97)$ km/s. The accuracy in time was $(-5 \pm 25)$ hours. For the CHSA Dst index forecast algorithm by Vršnak, Temmer, and Veronig (2007b) the Pearson correlation coefficient was 0.21, the forecast correlation coefficient 0.08 and the mean deviation $(-15 \pm 28)$ G. For the individual years we found a Pearson correlation coefficient of 0.33 for 2011, 0.26 for 2012, 0.35 for 2013 and -0.01 for 2014. Although the Dst forecast in general failed, it attracts our attention that forecast achieved better results in the individual years 2011, 2012 and 2013 than in the whole time range. We therefore suppose that an adaptive Dst forecast algorithm similar to the adaptive velocity forecast algorithm by Rotter et al. (2015), i.e. an algorithm which adapts itself to the current state of the Sun, could achieve much better results.

In summary we found clear dependencies of the high speed stream peak velocity and peak Dst index on the coronal hole area, latitude and total magnetic flux at times where only one large coronal hole was near the central meridian. However these results do not hold at times where more coronal holes are near the central meridian. Since the coronal-holes-in-slice based algorithm by Rotter et al. (2015) achieved quite reasonable results and since we found that the high-speed stream peak velocity is dependent on the latitude of coronal holes, we expect that high-speed stream peak velocities at L1 do not arise from polar coronal holes, but that polar coronal holes amongst others increase the peak velocities arising from mid- and low-latitude coronal holes.

This could be explained by the flux tube expansion models (see Chap. 3): even if the flux tubes arising from polar coronal holes do not run through L1, they could increase the flux tube expansion factor of flux tubes arising from nearby mid- and low-latitude coronal holes and therefore increase whose peak velocities. We therefore assume that the high speed stream peak velocities and the corresponding Dst peaks in general depend on the area, latitude, total magnetic flux and the distribution of surrounding coronal holes. We assume
9. Discussion

that we can greatly improve the statistical high speed stream forecast which is based on the coronal hole parameters of single coronal holes by also taking the distribution of coronal holes near the central meridian into account.

In order to further increase the physical significance of the statistical forecast algorithm, a study of flux tube expansion factors on the area of coronal holes seems reasonable. This study could finally allow to consolidate the results of statistical and empirical high-speed stream forecasts.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.1.: Coronal holes, filaments, solar wind parameters and geomagnetic storm index for 2010/10 to 2010/12. From top to bottom: Butterfly diagram of coronal holes, solar wind proton velocity, ion temperature, proton density, magnetic field density, Dst index, latitudinal incident angle of magnetic field density and longitudinal incident angle of magnetic field density. The butterfly diagram was recorded within a meridional slice of $\pm 3^\circ$. Coronal holes of positive polarities are drawn in green, of negative in red and filaments in black. Times of ICMEs according to Richardson-Cane’s ICME list are drawn in black. High speed stream parameters were measured by ACE at L1. The grey lines of the magnetic incident angles are the averages over $\pm 12$ hours.
Figure A.2.: Same as A.1. but for 2011/01 to 2011/03.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.3.: Same as A.1, but for 2011/04 to 2011/06.
Figure A.4.: Same as A.1. but for 2011/07 to 2011/09.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.5.: Same as A.1, but for 2011/10 to 2011/12.
Figure A.6.: Same as A.1, but for 2012/01 to 2012/13.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.7.: Same as A.1, but for 2012/04 to 2012/06.
Figure A.8.: Same as A.1, but for 2012/07 to 2012/09.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.9.: Same as A.1, but for 2012/10 to 2012/12.
Figure A.10.: Same as A.1 but for 2013/01 to 2013/03.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.11.: Same as A.1, but for 2013/04 to 2013/06.
Figure A.12.: Same as A.1 but for 2013/07 to 2013/09.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.13.: Same as A.1, but for 2013/10 to 2013/12.
Figure A.14.: Same as A.1 but for 2014/01 to 2014/03.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.15.: Same as A.1, but for 2014/04 to 2014/06.
Figure A.16.: Same as A.1 but for 2014/07 to 2014/09.
A. Map of coronal holes and solar wind parameters and Dst index from 2010/10 to 2014/12

Figure A.17.: Same as A.1, but for 2014/10 to 2014/12.
List of Figures

2.3. Parker spiral, ambient (slow) solar wind, high speed stream and shock front (schematic). From: Jian et al. (2006) . . . . . . . . 15
2.4. Solar wind parameters and Dst index from 2011/01 to 2011/04. 16
2.6. Structure of Earth’s magnetosphere (schematic). From NASA (http://www.nasa.gov/mission_pages/sunearth/multimedia/magnetosphere2_prt.htm) . . . . . . . . . . . . . . . . . . . 22
2.7. Dst-index from 1950 to 2010, averaged over each week of all years. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

3.1. PFSS extrapolation of the Sun for 1973/06/27. From Levine, Altschuler, and Harvey (1977) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.2. Flux tube expansion factors versus the radial distance in solar radii. From Wang and Sheeley (1990) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
3.3. Maximum solar wind speed versus flux tube expansion factors at 1 AU. From Suzuki (2006) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.4. Transition of a coronal hole in 2005/03, fractional area coronal holes cover within the slice and measured solar wind velocity at L1. From Vršnak, Temmer, and Veronig (2007a) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

229
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Measured solar wind speed at ACE versus the fractional area coronal holes cover within a meridional slice for day 25 to 125 in 2005. Predicted and observed solar wind speeds. From Vršnak, Temmer, and Veronig (2007a)</td>
<td>37</td>
</tr>
<tr>
<td>3.6</td>
<td>Fractional area coronal holes cover within a meridional slice versus solar wind proton velocity (top). Observed versus forecasted solar wind proton velocities for 2011/01/01 to 2013/12/31. From Rotter et al. (2015)</td>
<td>38</td>
</tr>
<tr>
<td>3.7</td>
<td>Dst index versus fractional area coronal holes cover within a meridional slice (top). Observed and by slice based algorithm predicted Dst values (bottom). From Vršnak, Temmer, and Veronig (2007b)</td>
<td>41</td>
</tr>
<tr>
<td>3.8</td>
<td>12-hour forecast of Dst-Index by a variation of the Burton-Russel-McPherron equation. From Antarctica (<a href="http://www.antarctica.ac.uk/SatelliteRisks/">www.antarctica.ac.uk/SatelliteRisks/</a>)</td>
<td>42</td>
</tr>
<tr>
<td>4.1</td>
<td>Forecast correlation coefficient: Correlation coefficient of slope versus slope of the regression line.</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Confusion matrix (left) and true positive rate vs. false positive rate (right, from Reiss et al., 2014b).</td>
<td>47</td>
</tr>
<tr>
<td>5.1</td>
<td>SDO (left) and ACE (right) schematic. From Caltech (<a href="http://www.srl.caltech.edu/ACE/Gallery/gallery.html">http://www.srl.caltech.edu/ACE/Gallery/gallery.html</a>) and NASA (sdo.gsfc.nasa.gov/assets/img/site/spacecraft.zip).</td>
<td>50</td>
</tr>
<tr>
<td>6.1</td>
<td>AIA-193 image from 2011/02/01 and corresponding intensity distribution of solar disk pixels.</td>
<td>60</td>
</tr>
<tr>
<td>6.2</td>
<td>Data products of dark objects identified in AIA-193 EUV images.</td>
<td>63</td>
</tr>
<tr>
<td>6.3</td>
<td>Page of dataset used for labelling coronal holes and filaments: sample coronal hole.</td>
<td>66</td>
</tr>
<tr>
<td>6.4</td>
<td>Page of dataset used for labelling coronal holes and filaments: sample filament.</td>
<td>67</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>Distribution of AIA-193 intensities and absolute value of the magnetic field densities derived from all pixels identified as CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>Distribution of mean AIA-193 intensities and shape parameters derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>Distribution of areas and fractional areas where only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G were taken into account, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>Distribution of total magnetic flux and fractional total magnetic fluxes where only pixels with an absolute magnetic field density greater than 10 G and 30 G were taken into account, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>Distribution of the mean magnetic field density and the mean magnetic field densities when only pixels with an absolute magnetic field density greater than 10 G (mid) and 30 G were taken into account, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.10</td>
<td>Distribution of relative open magnetic flux and relative open magnetic fluxes when only pixels with an absolute magnetic field density greater than 10 G and 30 G were taken into account, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.11</td>
<td>Distribution of standard deviations, skewness and kurtosis of the magnetic field densities, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>Fraction of the total magnetic flux, when only pixels with an absolute magnetic field density above a threshold were taken into account, versus the threshold. Distribution of the ratios $\Phi_{-10}/\Phi_{10}$ and $\Phi_{-30}/\Phi_{30}$, derived from all CHs and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.13</td>
<td>Scatter plots of the number of magnetic flux tubes derived at a magnetic threshold of 10 G, 30 G and 50 G versus the area and the area of pixels which have an absolute magnetic field density of more than 10 G, 30 G and 50 G, derived from all coronal holes and filaments.</td>
<td></td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>6.14</td>
<td>Scatter plots of the number of magnetic flux tubes derived at a magnetic threshold of 10 G, 30 G and 50 G versus the absolute total and total magnetic flux of pixels which have an absolute magnetic field density of more than 10 G, 30 G and 50 G, derived from all coronal holes and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.15</td>
<td>Scatter plots of the mean magnetic field densities, total magnetic flux and relative open magnetic flux versus the areas, derived from all coronal holes and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.16</td>
<td>Scatter plots of the latitude versus the relative open magnetic flux, areas and mean magnetic field densities, derived from all coronal holes and filaments.</td>
<td></td>
</tr>
<tr>
<td>6.17</td>
<td>Decision tree for automated classification of coronal holes and filaments, created by Weka.</td>
<td></td>
</tr>
<tr>
<td>6.18</td>
<td>Scatter plot of the symmetry parameter versus the absolute value of the relative open magnetic flux, derived for all coronal holes and filaments. The classification border between coronal holes and filament as derived by Weka is drawn in green.</td>
<td></td>
</tr>
<tr>
<td>6.19</td>
<td>Results of the supervised classification algorithms true positive rate vs. false positive rate.</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>EUV images of the Sun, for each coronal holes under study.</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>EUV images of the Sun, for each coronal holes under study.</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>Statistical relationships of coronal holes and high speed streams: Latitude versus area of the coronal holes under study.</td>
<td></td>
</tr>
<tr>
<td>7.4</td>
<td>Definition of parameters $t_{CHfront}$, $t_{CHcenter}$, $t_{CHback}$, $t_{vrise}$, $t_{vmax}$, $t_{fall}$, $t_B$, $t_p$ and $t_{Dst}$.</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>Dependence of travel time $d_{vmax}$ on $v_{max}$ at a starting time of $t_{CHfront}$, $t_{CH0.36}$ and $t_{CHmid}$</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>Root mean square deviations (RMSE) of the travel times $d_{vmax}$, $d_{vrise}$ and $d_{fall}$ to their linear fits versus the normalized starting positions $X$ within coronal holes.</td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>Dependence of travel time $d_{vrise}$ on $v_{max}$ at a starting time of $t_{CH 0.25}$. Delay of $t_{vrise}$ to $t_{vmax}$ versus $v_{max}$.</td>
<td></td>
</tr>
</tbody>
</table>
7.8. Dependence of travel time $d_{v_{\text{fall}}}$ on $v_{\text{max}}$ and longitudinal width of the coronal hole. Starting times are $t_{\text{CH}, 0.36}$ and $t_{\text{CH}, 1}$.

7.9. Deviations of the travel times $d_{v_{\text{max}}}$ to the linear fit of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ at a starting time $t_{\text{CH}, 0.36}$ versus the deviations of the travel times $d_{v_{\text{rise}}}$ to the linear fit of $d_{v_{\text{rise}}}$ vs. $v_{\text{max}}$ at a starting time $t_{\text{CH}, 0.25}$. Deviations of the travel times $d_{v_{\text{max}}}$ to the linear fit of $d_{v_{\text{max}}}$ vs. $v_{\text{max}}$ at a starting time $t_{\text{CH}, 0.36}$ versus the deviations of the travel times $d_{v_{\text{fall}}}$ to the linear fit of $d_{v_{\text{fall}}}$ vs. $v_{\text{max}}$ at a starting time $t_{\text{CH}, 0.25}$.

7.10. Starting position of $v_{\text{max}}$ in the coronal holes, when a constant radial travel velocity is assumed. The starting positions are normalized on the longitudinal width of the coronal holes (leading edge: $X = 0$, trailing edge: $X = 1$).

7.11. Starting position of $v_{\text{fall}}$ in the coronal holes, when a constant radial travel velocity is assumed. The starting positions are normalized on the longitudinal width of the coronal holes (leading edge: $X = 0$, trailing edge: $X = 1$).

7.12. Delay of $t_{\rho}$ to $t_{v_{\text{max}}}$, $t_{B}$ to $t_{v_{\text{max}}}$ and $t_{B}$ to $t_{\rho}$ versus $v_{\text{max}}$.

7.13. Delay of $t_{\rho}$ to $t_{v_{\text{max}}}$ versus $v_{\text{max}}$, $\rho_{\text{max}}$ and $B_{\text{max}}$.


7.15. Latitude of coronal holes versus $(v_{\text{max}} - 485)/A_{\text{CH}}$. Calculated solar wind peak velocities using Eq. 7.9 versus measured solar wind peak velocities.

7.16. Cumulative histogram of the deviations of measured velocities to calculated velocities using by Eq. 7.10. Relative open magnetic flux of coronal holes versus the deviations and $v_{\text{max}}$ versus deviations.

7.17. Deviations of measured velocities to the calculated velocities using Eq. 7.9 versus the day of year. Calculated solar wind peak velocities using Eq. 7.10 versus measured solar wind peak velocities.
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.18</td>
<td>Cumulative histogram of the deviations of measured velocities to calculated velocities using Eq. 7.10. Relative open magnetic flux of coronal holes versus the deviations and ( v_{\text{max}} ) versus deviations.</td>
</tr>
<tr>
<td>7.19</td>
<td>Scatter plots of peak density versus inverse magnetic flux density, inverse area and relative open magnetic flux.</td>
</tr>
<tr>
<td>7.20</td>
<td>Solar wind magnetic field density at L1 measured 24 hours after ( t_{v_{\text{max}}} ) versus total magnetic flux of coronal holes, solar wind magnetic latitude versus latitude of coronal holes and solar wind magnetic longitude versus ( v_{\text{max}} ).</td>
</tr>
<tr>
<td>7.21</td>
<td>Solar wind magnetic field density at L1 measured at ( t_{v_{\text{max}}} ) versus total magnetic flux of coronal holes, solar wind magnetic latitude versus latitude of coronal holes and solar wind magnetic longitude versus ( v_{\text{max}} ).</td>
</tr>
<tr>
<td>7.22</td>
<td>Solar wind magnetic field density at L1 measured at ( t_B ) versus total magnetic flux of coronal holes, solar wind magnetic latitude versus latitude of coronal holes and solar wind magnetic longitude versus ( v_{\text{max}} ).</td>
</tr>
<tr>
<td>7.23</td>
<td>Solar wind magnetic field densities at L1 measured at ( t_B ) versus total magnetic flux of coronal holes. The solar wind magnetic field density was averaged over ( \pm 0 ) hours, ( \pm 6 ) hours and ( \pm 12 ) hours.</td>
</tr>
<tr>
<td>7.24</td>
<td>Dependence of the solar wind magnetic flux density at L1 measured at ( t_B ) and averaged over ( \pm 6 ) hours on total magnetic flux of coronal holes, after an offset of 5 G has been removed. Calculated magnetic flux density using Eq. 7.13 versus measured magnetic flux densities. Cumulative histogram of the of deviations of measured to calculated magnetic flux densities.</td>
</tr>
<tr>
<td>7.25</td>
<td>Scatter plots of ( B_z ) in GSM coordinates measured at L1 at ( t_B ) and averaged over 0 hours, ( \pm 3 ) hours and ( \pm 12 ) hours versus the ( B_z )-component of the total magnetic flux of coronal holes.</td>
</tr>
<tr>
<td>Figure Number</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>7.26.</td>
<td>$B_{t}$ averaged over ±3 hours versus total magnetic flux of coronal holes. Calculated magnetic flux densities $B_{z,fit}$ using Eq. 7.14 under the assumption of an offset of 4 G and 0 G versus measured $B_{z}$-components in GSM at L1 at $t_{B}$ averaged over ±3 hours.</td>
</tr>
<tr>
<td>7.27.</td>
<td>Calculated magnetic field densities $B_{z,fit}$ using Eq. 7.15 versus measured $B_{z}$-components in GSM at L1 at $t_{B}$ averaged over ±3 hours. Cumulative histogram of the deviations $</td>
</tr>
<tr>
<td>7.28.</td>
<td>10 samples for the determination of Dst drop.</td>
</tr>
<tr>
<td>7.29.</td>
<td>Distribution of Dst drop against solar wind maximum velocity, maximum density in the shock and then $B_{z}$-component of the maximum magnetic field density in the shock.</td>
</tr>
<tr>
<td>7.30.</td>
<td>Dependence of Dst drop on its delay to $t_{v_{max}}$ and on $\Phi_{CH,z}$.</td>
</tr>
<tr>
<td>7.31.</td>
<td>Dst drops versus the area of coronal holes.</td>
</tr>
<tr>
<td>7.32.</td>
<td>Dst drops per area versus latitude of coronal holes. Calculated Dst drops using Eq. 7.16 versus measured Dst drops. Cumulative histogram of the deviations of measured to calculated Dst drops.</td>
</tr>
<tr>
<td>8.1.</td>
<td>Structure of forecast tool.</td>
</tr>
<tr>
<td>8.2.</td>
<td>Solar wind and storm forecast for the period 2011/01 - 03.</td>
</tr>
<tr>
<td>8.3.</td>
<td>Same as Figure 8.2 but for 2011/04 - 06.</td>
</tr>
<tr>
<td>8.4.</td>
<td>Same as Figure 8.2 but for 2011/07 - 09.</td>
</tr>
<tr>
<td>8.5.</td>
<td>Same as Figure 8.2 but for 2011/10 - 12.</td>
</tr>
<tr>
<td>8.6.</td>
<td>Same as Figure 8.2 but for 2012/01 - 03.</td>
</tr>
<tr>
<td>8.7.</td>
<td>Same as Figure 8.2 but for 2012/04 - 06.</td>
</tr>
<tr>
<td>8.8.</td>
<td>Same as Figure 8.2 but for 2012/07 - 09.</td>
</tr>
<tr>
<td>8.9.</td>
<td>Same as Figure 8.2 but for 2012/10 - 12.</td>
</tr>
<tr>
<td>8.10.</td>
<td>Same as Figure 8.2 but for 2013/01 - 03.</td>
</tr>
<tr>
<td>8.11.</td>
<td>Same as Figure 8.2 but for 2013/04 - 06.</td>
</tr>
<tr>
<td>8.12.</td>
<td>Same as Figure 8.2 but for 2013/07 - 09.</td>
</tr>
<tr>
<td>8.13.</td>
<td>Same as Figure 8.2 but for 2013/10 - 12.</td>
</tr>
<tr>
<td>8.14.</td>
<td>Same as Figure 8.2 but for 2014/01 - 03.</td>
</tr>
</tbody>
</table>
List of Figures

8.15. Same as Figure 8.2 but for 2014/04 - 06 .......................... 175
8.16. Same as Figure 8.2 but for 2014/07 - 09 .......................... 176
8.17. Same as Figure 8.2 but for 2014/10 - 12 .......................... 177
8.18. Problems of evaluation: four samples ......................... 179
8.19. Peaks used for evaluation of CHSA from 2011/01 to 2012/12 182
8.20. Same as 8.19 but for 2013/01 to 2014/12 .......................... 183
8.21. Peaks used for evaluation of CHA from 2011/01 to 2012/12 184
8.22. Same as 8.21 but for 2013/01 to 2014/12 .......................... 185
8.23. Latitude versus area of the coronal holes belonging to the TP peaks of CHA. The measured peak velocities of TP peaks are coloured ......................... 188
8.24. CHSA: Predicted versus measured high-speed stream velocity peaks .................................................. 194
8.25. CHSA: Histograms of the deviations of measured to predicted high-speed stream velocity peaks. Deviations of the measured to predicted arrival times of the velocity peaks versus $v_{\text{max}}$, and histogram of the deviations of measured to predicted arrival times of the velocity peaks ........................................ 195
8.26. CHA: Predicted versus measured high-speed stream velocity peaks .................................................. 196
8.27. CHA: Histograms of the deviations of measured to predicted high-speed stream velocity peaks. Deviations of the measured to predicted arrival times of the velocity peaks versus $v_{\text{max}}$, and histogram of the deviations of measured to predicted arrival times of the velocity peaks ........................................ 197
8.28. CHA: Predicted versus measured magnetic field density peaks .................................................. 200
8.29. CHSA: Predicted versus measured Dst peaks ....................... 203
8.30. CHA: Predicted versus measured Dst peaks ....................... 204
A.1. Coronal holes, filaments, solar wind parameters and geomagnetic storm index for 2010/10 to 2010/12 ............... 212
A.2. Same as A.1 but for 2011/01 to 2011/03 ............... 213
A.3. Same as A.1 but for 2011/04 to 2011/06 ............... 214
List of Figures

A.4. Same as A.1, but for 2011/07 to 2011/09. ......................... 215
A.5. Same as A.1, but for 2011/10 to 2011/12 ......................... 216
A.6. Same as A.1, but for 2012/01 to 2012/13 ......................... 217
A.7. Same as A.1, but for 2012/04 to 2012/06 ......................... 218
A.8. Same as A.1, but for 2012/07 to 2012/09 ......................... 219
A.9. Same as A.1, but for 2012/10 to 2012/12 ......................... 220
A.10. Same as A.1, but for 2013/01 to 2013/03 ....................... 221
A.11. Same as A.1, but for 2013/04 to 2013/06 ....................... 222
A.12. Same as A.1, but for 2013/07 to 2013/09 ....................... 223
A.13. Same as A.1, but for 2013/10 to 2013/12 ....................... 224
A.14. Same as A.1, but for 2014/01 to 2014/03 ....................... 225
A.15. Same as A.1, but for 2014/04 to 2014/06 ....................... 226
A.16. Same as A.1, but for 2014/07 to 2014/09 ....................... 227
A.17. Same as A.1, but for 2014/10 to 2014/12 ....................... 228
List of Tables
6.1. Mean value and standard deviation of the distributions of characteristic parameters of coronal hole and filaments. . . . . . . . 74
7.1. Statistical high speed stream arrival times: tvmax , tvrise , tvfall , tB ,
tρ and tDst . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
8.1. Number of extracted coronal holes and filaments by the forecast
algorithm, dependent on their area and latitude. . . . . . . . . . 161
8.2. Evaluation of forecast rate of high speed streams: number of
TP, FP and FN peaks for the CHA and CHSA algorithms. . . . 187
8.3. Numbers and dates of FN peaks of CHA and CHSA for which
the cause for erroneous predictions could be determined. . . . . 188
8.4. Evaluation parameters of CHSA and CHA peak velocities. . . . 192
8.5. Evaluation parameters of CHSA and CHA arrival times of peak
velocities. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 193
8.6. Evaluation parameters of CHA peak magnetic field densities. . . 199
8.7. Evaluation parameters of CHSA and CHA peak Dst index. . . . 202

239


Bibliography


[3] T. Chai and R. R. Draxler. “Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature”. In: Geoscientific Model Development 7 (June 2014), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014


Bibliography


Bibliography


Bibliography


Bibliography


246


