1st Advances in Solar Physics Euroconference
Advances in the Physics of Sunspots
ASP Conference Series, Vol. 118, 1997
B. Schmieder, J.C. del Toro Iniesta, M. Vázquez, eds.

Empirical Modelling and Thermal Structure of Sunspots

Sami K. Solanki

Institute of Astronomy, ETH-Zentrum, CH-8092 Zürich, Switzerland

Abstract. An overview of empirical models of sunspot thermal structure is presented. First a few remarks are made on empirical modelling in general. After that the literature on the various types of models, 1- and 2-component models of the umbra and penumbra, is reviewed. The evidence for or against the dependence of umbral temperature on various parameters like sunspot size, magnetic field strength and phase of the solar cycle is critically reviewed.

1. Introduction

The aim of this paper is to provide an overview of the empirical modelling efforts aimed at determining the thermal structure of sunspots. Previous reviews bearing on this subject have been given by Schröter (1971), Muller (1987, 1992), Solanki (1990), Maltby (1992a, b) and Sobotka (1996).

A knowledge of sunspot thermal structure is important for a variety of reasons. For example, it is needed for accurate line profile calculations, which in turn underly studies of sunspot dynamics and magnetism. Also, the empirically derived thermal stratification can constrain theoretical models of energy transport mechanisms. Questions that may be addressed by comparing empirical with theoretically predicted thermal stratifications concern the factor by which convective energy transport is inhibited in umbrae and penumbrae as a function of height, the layers at which mechanical (e.g., wave) energy transport and deposition become important, mechanical heating rates in sunspots compared to other solar magnetic features, etc. Although investigations aimed at answering these questions are still in their infancy their importance will certainly increase in the future.

The umbra and penumbra are always modelled separately, with umbrae grabbing the lion's share of the modelling effort. This review is structured accordingly, with most of the space devoted to umbrae.

2. Empirical Modelling

The main objective of empirical modelling is to determine the physical state of a solar feature (i.e. the magnetic and velocity vectors, temperature, pressure, etc. as a function of 3 spatial coordinates and time) from a given set of observations, i.e. from one or more of the Stokes parameters I, Q, U, V as a function of wavelength, location on solar surface and time. Of course, the final objective is a physical understanding of solar structures and processes. This, however, involves an additional step, namely the comparison with theoretical predictions.

Empirical modelling generally involves radiative transfer calculations and model-atmosphere construction under some set of constraints (e.g., hydrostatic equilibrium, pressure balance across magnetic field lines, LTE, choice of the magnetic geometry, number of atmospheric components). In the past the modelling process has generally been carried out on a trial-and-error basis. For example, an initial guess of the model atmosphere is constructed and used to calculate line profiles. These are then compared with observations. If the correspondence between synthetic and observed profiles is insufficient the model is changed. This process is repeated until a satisfactory fit is achieved.

More recently, so-called inversion techniques have become increasingly popular. Note that in general only linear dependences can be analytically inverted (Craig & Brown 1986), in the sense that from a set of known integrated quantities (such as the Stokes profiles) the integrand (contribution function) can be determined as a function of the free parameter (optical depth). Unfortunately, the quantities of interest — temperature, magnetic field, velocity — are rarely related linearly to the observed quantities. Hence the inversion of Stokes parameters is basically an automated version of the trial-and-error approach described above. In general, some minimisation technique is employed to find a minimum of the χ^2 hypersurface (χ^2 is proportional to the squared difference between observed and synthetic Stokes parameters). Nevertheless, the various Stokes inversion codes have been reasonably successful. [1] It should be borne in mind that even the best inversion code can only derive the information present within the data and is restricted to deducing the free parameters within the framework of the prescribed model (recall that there are *always* prescribed model constraints, even if they are only implicit). The reality and physical significance of the inverted model atmosphere thus depends on the data and the employed model.

An important parameter of the solar feature to be modelled is its size relative to the spatial resolution element. If it is spatially resolved then 1-component, i.e. purely plane-parallel modelling is sufficient. In the case of solar magnetic features, however, usually 2 or more components are required.

Consider first the case of magnetic elements, i.e. the small-scale magnetic features forming plages and the network. Since the individual magnetic elements are not resolved the simplest description is in terms of a 2-component model. Then the observed Stokes profiles (which are averaged over the spatial resolution element) can be written as

$$I_{\text{obs}} = \alpha I_{\text{m}} + (1-\alpha) I_{\text{s}} \,, \tag{1}$$

$$P_{\text{obs}} = \alpha P_{\text{m}} \,, \tag{2}$$

[1] One well known inversion code has been writted by the HAO group and is named the "Stokes Profile Inversion Routine" or SPIR. The IAC group calls their inversion technique SIR for "Stokes Inversion with Response-functions". Yet another possible name is "Stokes Advanced Magnetic Inversions". Working out the resulting abbreviation is left as an excercise to the reader.

where $I_{\rm m}$ is the intensity profile arising from the magnetic feature, $I_{\rm s}$ is the intensity from the surroundings, α is the magnetic filling factor (i.e. the fraction of the surface area within the resolution element covered by magnetic field) and $P = Q, U$, or V represents the 2 linearly and the circularly polarized Stokes parameter. It is clear from Eq. (1) that in the context of this simple model $Q_{\rm obs}$, $U_{\rm obs}$ and $V_{\rm obs}$ are formed in the magnetic element only.

Consider now a sunspot. Although a sunspot as a whole is resolved, it contains considerable unresolved fine structure. The closest one comes to a single-component situation is in the dark umbral core, which is often spatially resolved. However, its low intensity means that stray light from the much brighter surroundings can influence the observations. If a part of the stray light comes from the penumbra then it may also affect the polarized Stokes parameters:

$$I_{\rm obs} = \alpha I_{\rm u} + (1-\alpha) I_{\rm s} \,, \quad (3)$$

$$P_{\rm obs} = \alpha_{\rm p} P_{\rm u} + (1-\alpha_{\rm p}) P_{\rm s} \,. \quad (4)$$

Here $\alpha_{\rm p} \geq \alpha$, where $\alpha_{\rm p}$ is the "filling factor" derived from the polarized radiation. It can be larger than α because a part of the stray light may be unpolarized.

Finally, in the penumbra and the brighter parts of the umbra at least 3 components are present, a bright component denoted in the following by a subscript b (bright filaments in the penumbra, dots in the umbra), a dark component denoted by a subscript d (dark filaments and the umbral background, respectively) and the stray light. In the simplest case we can therefore write

$$I_{\rm obs} = \alpha_{\rm b} I_{\rm b} + \alpha_{\rm d} I_{\rm d} + (1-\alpha_{\rm b}-\alpha_{\rm d}) I_{\rm s} \,, \quad (5)$$

$$P_{\rm obs} = \alpha_{\rm b,p} P_{\rm b} + \alpha_{\rm d,p} P_{\rm d} + (1-\alpha_{\rm b,p}-\alpha_{\rm d,p}) P_{\rm s} \,. \quad (6)$$

Again $\alpha_{\rm b,p} \geq \alpha_{\rm b}$ and $\alpha_{\rm d,p} \geq \alpha_{\rm d}$. Clearly, a good stray-light correction is central to the correct modelling of sunspot thermal structure. Although the polarized Stokes parameters are less affected by stray light than the intensity, the situation is nevertheless more complicated than in plages.

3. Single Component Umbral Models

Umbrae have been modelled as either one- or two-component structures (the stray-light fraction of the total radiation is in general not counted as a separate component). The total number of empirical umbral models, in particular of 1-component models, is large, although not all models are independent of each other; many have grown out of older models by way of minor or major changes. In the present section I discuss single-component models. The 2-component models are the subject of Sect. 5.

It is well known that sunspot umbrae are not homogeneous (e.g., Sect. 4.2). Hence single-component models can only describe a part of the umbra (or an average over it). Usually such models are meant to describe the dark umbral cores, which are thought to be relatively homogeneous.

In addition to the assumption of homogeneity another important assumption underlying almost all single-component models is universality, i.e. the assumption that a single model is valid for all umbrae, or at least all umbral cores.

This assumption has been advocated by Albregtsen & Maltby (1981a), who contend that the brightness of the dark cores of large umbrae (with diameters $d_u \gtrsim 10''$) is independent of size or other umbral parameters and depends only on the phase of the solar cycle. This assumption is scrutinized in Sect. 4.

We need to distinguish between models restricted to the photospheric layers of umbrae and those including the chromosphere or transition zone. I'll therefore treat purely photospheric models separately.

3.1. Models of the umbral photosphere

Models of the umbral photosphere are based on LTE modelling of continuum contrast, weak spectral lines, or in some cases the wings of strong spectral lines. Both the centre-to-limb variation (CLV) and the wavelength dependence of the continuum intensity (I_c) or continuum contrast ($\phi_u = I_{c,u}/I_{c,\text{phot}}$) umay be used, although the latter provides information only on a very limited height range unless a very large wavelength range is considered. Note that I_c is often obtained from a relatively broad-band measurement and may be significantly affected by line blanketing.

Early models were generally based on the CLV of continuum contrast and belonged to either the class of rarified (Michard 1953, Van t'Weer 1963, Fricke & Elsässer 1965, Zwaan 1965) or the non-rarified models (Mattig 1958, Jakimiec 1965, Zwaan 1965). Near the $\tau = 1$ level the gas pressure in rarified models is generally reduced by some factor (2–10) below that given by hydrostatic equilibrium. The $\tau_c = 1$ level lies 2000–3000 km deeper inside sunspots than in the quiet sun according to these models and it is not obvious how they can be reconciled with Wilson-effect measurements, which reveal Wilson depressions of 400–800 km (e.g., Gokhale & Zwaan 1972). Rarified models are also expected to be unrealistically bright near the solar limb (Jensen & Maltby 1965). Such models were seemingly required to reproduce the observations, mainly due to the neglect of the contribution of line opacity to the total opacity. The non-rarified models usually build on the assumption of hydrostatic equilibrium. At least in the photospheric layers such models are expected to be far more realistic. Early umbral models are discussed in detail by Bray & Loughhead (1964).

Models since the end of the 1960s have been constructed by Hénoux (1969), who derived his temperature stratification from continuum measurements and the Na I D line wings and checked it using equivalent widths of medium-strong lines (cf. Hénoux 1968); Wittmann & Schröter (1969), from the CLV of I_c at different wavelengths between 4680 Å and 7900 Å; Mattig (1969), who chose a simple linear $T(\tau)$ to reproduce the CLV of his I_c observations; Stellmacher & Wiehr (1970), who modified Hénoux's model to reproduce $I_c(\lambda)$ and line observations made by the authors; Dicke (1970), from the temperature vs. field strength relation derived from observations by von Klüber (1947) under the assumption that the field is truly cylindrical below the solar surface (this model covers the low photosphere only); Yun (1971), based on the CLV of I_c for different $\lambda \lesssim 1.6$ μm and on the wings of Na I D; Webber (1971), from molecular lines of MgH and TiO; Kneer (1972), from the profiles of three Fe I lines and continuum (this model turns out to be very similar to that of Stellmacher & Wiehr 1970); Kjeldseth-Moe & Maltby (1974b), based on I_c observations (this model explains differences between measured I_c values of different sunspots by temperature fluctuations in the upper photosphere); Zwaan (1974, 1975), constructed to reproduce the wavelength dependence of I_c taken from different sources, Stellmacher & Wiehr (1975), based on Na I D lines wings, Fe I 5434 Å, 2 infrared C I lines and $I_c(\lambda)$; Kollatschny et al. (1980), from the wings of Ca II 8542 Å and other strong lines, as well as $I_c(\lambda)$; and Boyer (1980), a model based on the equivalent widths of 147 TiO lines, which is similar in $T(\tau)$ to the Stellmacher & Wiehr (1975) model.

3.2. Chromospheric models

These models are based on either strong spectral lines in the visible or on lines in the UV. The radiative transfer is always carried out in NLTE. Since the calculation of upper atmospheric lines usually also requires a knowledge of the photospheric thermal structure almost all chromospheric models include a photospheric part. In many cases this is simply taken from one of the photospheric models listed in Sect. 4.1. In this section I also include models of the transition zone and the corona.

Early observations (Hale 1892) showed that the Ca II H and K lines have emission cores, suggesting that sunspots also possess chromospheres (see Linsky & Avrett 1970 for a review). First exploratory models of sunspot umbral chromospheres were constructed by Baranovsky (1974a, b; cf. Staude 1981) and later by Kneer & Mattig (1978), based on general properties of the Ca II H, K and IR lines, and by Teplitskaya et al. (1978), based on the inversion of Ca II H and K line profiles.

This early work set off a spate of modelling, which resulted in a model based on Ca II H, K and 2 IR triplet lines, Hα and Na I D (Yun et al. 1981, Beebe et al. 1982), an umbral chromospheric and transition zone model based on the cores of Ca II H and K, Mg II h and k, H I Lyα and β and C IV observations made with the OSO 8 satellite (Lites & Skumanich 1981, 1982), a model due to Staude (1981), which agrees closely with the Stellmacher & Wiehr (1975) model in the photosphere, the Teplitskaya et al. (1978) model in the chromosphere and reproduces HRTS and OSO 8 data (see Staude et al. 1983, 1984), a model of the transition zone above an umbra that is based on HRTS spectra of FUV lines of various ions formed mainly in the temperature range 2×10^4–2×10^5 (Nicolas et al. 1981), and an investigation of the coronal temperature structure above umbrae by Foukal (1981).

In their chromospheric layers the various models can differ significantly. For example, Staude (1981) finds an almost steady increase of the temperature throughout the chromosphere, while Lites & Skumanich (1981, 1982) deduce a flat chromospheric temperature plateau. All models show, however, an increased height of the umbral temperature minimum relative to the quiet sun, needed to explain, e.g., the narrow Ca II H and K emission peaks in umbrae.

The efforts of a number of different modellers have been synthesized into a single model by Avrett (1981) and Staude et al. (1983). Avrett's "Sunspot sunspot model" of the umbral photosphere, chromosphere and transition zone combines the low photospheric model of Albregtsen & Maltby (1981b) with the upper photospheric and chromospheric parts of the Lites & Skumanich (1981, 1982) model and the transition region model of Nicolas et al. (1981). Staude et al. (1983), cf. Staude (1981), produced a comprehensive umbral model covering the full height range from the photosphere to the corona. It is based on a large set of observations (radio, optical, EUV, x-ray), but takes its photospheric structure largely from Stellmacher & Wiehr (1975), while its lower chromosphere is similar to that of Teplitskaya et al. (1978).

Since the mid 1980s the largest effort in single component umbral modelling has been invested into testing and improving the model of Avrett (1981). First Maltby et al. (1986) improved Avrett's model in the deep layers using photometric data obtained at Oslo. Their set of 3 models (each for a different phase of the solar cycle, see Sect. 4.3) has become something of an industrial standard.

Lites et al. (1987) compared the observed Stokes V profile shape of the Mg I intercombination line at 4572 Å with the synthetic profile resulting from the Maltby et al. (1986) atmosphere (for the atomic model of Altrock and Canfield 1974). Since the synthetic profiles show a strong inversion in the Stokes V core whereas the observations do not, they concluded that the chromospheric temperature rise must begin at a considerably greater height than in the Maltby et al. model. The conclusion of Lites et al. (1987) has been criticized by Mauas et al. (1988), who argued that the true transition probability of the Mg I intercombination line is significantly smaller than the value employed by Lites et al. (1987). Briand & Solanki (1995) obtained good fits to quiet-sun profiles of this line with $\log(gf)$ values close to those of Mauas et al. (1988), so that the criticism may be well-founded.

Caccin et al. (1993) constructed a modified version of one of the Maltby et al. (1986) model with a steeper temperature gradient in the photosphere, resulting in a minimum temperature of 2500 K. Their modification is based on fits to Na I D and K I 7699 Å profiles. The Caccin et al. (1993) model was in turn slightly modified by Severino et al. (1994), who raised the minimum temperature to 2900 K, without otherwise significantly changing the model. The temperature in the middle and upper photosphere of the Maltby et al. (1986) model is constrained mainly by the CO first overtone lines at 2.35 μm. These were used to fix the minimum temperature at 3400 K, giving a very flat temperature stratification through most of the photosphere. Support for the steeper photospheric temperature gradient in umbrae comes from Ayres (1996), who constructed a modified LTE version of the Maltby et al. (1986) model (i.e. a version without a chromosphere) that reproduces fundamental band CO lines at 4.67 μm. Since these lines are stronger than the first overtone lines at 2.3 μm they are more sensitive to upper photospheric temperatures. In the photospheric layers his model turns out to be much closer to that of Caccin et al. (1993) and Severino et al. (1994) than the original Maltby et al. (1986) model. The only recent model that agrees well with that of Maltby et al. (1986) in the upper photosphere is that of Collados et al. (1994), which, however, is not well constrained in the upper photosphere by the data it is based on (Ruiz Cobo, private communication).

Severino et al. (1994) point out the good agreement between the photospheric temperature stratification of their model and a non-grey radiatve equilibrium model. The Maltby et al. (1986) umbral photosphere, on the other hand, corresponds more closely to a grey radiative equilibrium. Rüedi et al. (1996) confirm that the (non-grey) radiative equilibrium atmospheres of Kurucz

(1991) reproduce observations of Ti I lines at 2.2 μm much better than the Maltby et al. (1986) model.

It thus appears that at least in photospheric layers the energetics of sunspot umbrae are determined by radiation. The fraction of the total energy flux transported by radiation is fairly large in all observable layers of umbrae (greater than 70% for $\log \tau < 1$). In contrast, this fraction drops rapidly with depth in the quiet photosphere (e.g., Maltby et al. 1986, Maltby 1992a, Collados et al. 1994). This implies that convection, although definitely present (as pointed out already by Zwaan 1974, 1975), plays a much smaller role in sunspot umbrae than in the quiet sun and stresses the need to properly incorporate radiation into simulations of magnetoconvection aimed at understanding the umbra.

4. Dependence of the Umbral Temperature on Other Parameters

4.1. Dependence on sunspot size

Let us now address the question whether the umbral brightness and temperature depends on umbral size or not, since it touches on a basic assumption underlying almost all single component models, namely whether it is possible to describe all umbrae (or at least those above a certain size) by a single thermal model. An excellent review of the relevant work up to the end of the 1980s has been given by Maltby (1992a).

Early observations suggested that large sunspots are darker than small sunspots (continuum observations prior to 1963 are listed and discussed by Bray and Loughhead 1964). Such observations were often insufficiently corrected for stray light produced in the earth's atmosphere (seeing) and in the instrument (scattered light). Zwaan (1965) first pointed out that the amount of stray light affecting the umbra increases rapidly with decreasing umbral size. Observations made since then have usually been corrected for stray light with much greater care. For sunspots with umbral diameters greater than 8–10″ subsequent observations did not reveal any significant dependence of umbral core brightness on sunspot size (e.g., Zwaan 1965, Rossbach & Schröter 1970, Albregtsen & Maltby 1981a).

More recently, however, evidence against this result has been mounting. Kopp & Rabin (1992) present observations that show a clear relationship between the umbral brightness at 1.56 μm and sunspot size after correction for stray light (note that their correction is much larger for small sunspots than for large, in agreement with the expectations). All their observations were obtained on the same day, but the umbral contrast $\phi_{\rm u} = I_{\rm u}/I_{\rm phot}$ shows a dispersion as large as over the whole solar cycle according to Albregtsen & Maltby (1978). In Fig. 1 the Kopp & Rabin results (triangles) are shown together with measurements of the minimum brightness at 1.56 μm of 2 sunspots (filled squares and stars) made by Solanki et al. (1992) and Rüedi et al. (1995), by the latter on 3 different days. The linear dependence on sunspot diameter is evident.

Martínez Pillet & Vázquez (1993) confirm this result for 7 sunspots observed within 2 weeks of each other in 1989 at similar $\mu = \cos\theta$ values (θ is the heliocentric angle). They correct carefully for stray light. Collados et al. (1994) have inverted spectra from the darkest parts of 3 sunspots of this sample, two large spots and a smaller one. The 2 larger spots (umbral diameters

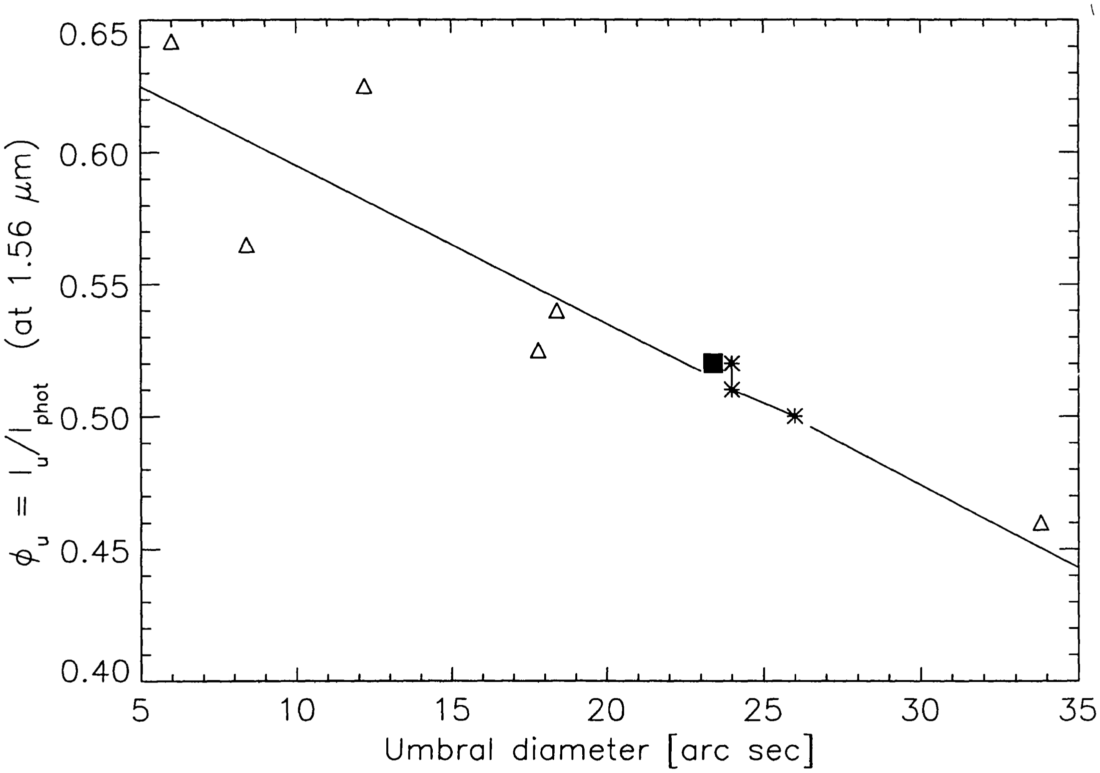

Figure 1. Umbral normalized continuum intensity $\phi_u = I_u/I_{phot}$ at 1.56 μm plotted vs. umbral diameter. Plotted are the lowest stray light corrected ϕ_u values of 8 umbrae. The 6 umbrae represented by triangles were observed by Kopp & Rabin (1992), the umbra represented by the filled square by Solanki et al. (1992). The 3 stars represent an umbra observed on 3 different days by Rüedi et al. (1995). The solid line indicates the trend.

of 22″) give similar results, $T(\tau = 1) = 3940$ K, while the small spot returns $T(\tau = 1) = 5030$ K. The stray-light problem in the small spot was avoided by fitting only Stokes V (note, however, that Stokes V may still be contaminated by polarized stray light from the penumbra).

Stellmacher & Wiehr (1988) found that 2 small spots with $d_u = 7''$ have $T_{\rm eff}$ values 600 K higher than model M4 of a large sunspot umbral core (Kollatschny et al. 1980). They take stray light into account in their modelling of lines which have mutually opposite temperature sensitivity.

A final line of evidence for a dependence of T on d_u comes from the nearly universal relationship between B and T, with T being lower when B is larger (see Sect. 4.2). Since the maximum field strength $B_{\rm max}$ of sunspots scales with their size (Brants & Zwaan 1982, Kopp & Rabin 1992) this relationship implies that larger spots should also be darker.

4.2. Dependence on magnetic field strength

Alfvén (1943) first predicted a relationship between the vertical component of the magnetic field, B_z, or alternatively the field strength, B, and temperature, T, at a given geometrical height. It was later extensively investigated, both theoretically and observationally. Much of the observational work prior to 1992 has been reviewed in the introduction to their paper by Martínez Pillet & Vázquez (1993).

Investigators have either considered the relationship between the maximum field strength $B_{\rm max}$ of a sunspot and the associated minimum temperature (or continuum brightness), or the B vs. T relationship at different locations within a single sunspot. I first review investigations of the former type.

Cowling (1957) considered the theoretical aspects of this relationship. Deinzer (1965) compared the observations of Stumpff (1961) via a relation given by Houtgast and van Sluiter (1948), with the predictions of his self-similar sunspot model (cf. Schlüter & Temesvary 1958), which includes a mixing length formalism to describe energy transport. The model qualitatively reproduced the data. Dicke (1970) provided an alternative formulation of the theoretical prediction and used the data of von Klüber (1947) to test it. Maltby (1977) published a clear derivation. He also compared the well calibrated and stray-light corrected intensity measurements of Ekmann & Maltby (1974) and Ekmann (1974) with routine field strength measurements made at Rome observatory (which are uncorrected for stray light). Unsurprisingly he found only a small variation of T, but a large variation of $B_{\rm max}$ from spot to spot. Chou (1987) presented a $B_{\rm max}$ vs. $\phi_u(\lambda = 6100\text{Å})$ relationship for stable spots, as well as showing that growing sunspots depart from this relationship. His stable spot relationship falls between those of Deinzer (1965) and Dicke (1970), but must be treated with caution since some of his measurements are obviously systematically incorrect (e.g., he claims $B_{\rm max} \approx 600$ G and $\phi_u \approx 0.6$ for sunspots having $d_u \approx 6$–$10''$. This $B_{\rm max}$ is much too low and the ϕ_u probably too large).

The most careful and thorough studies of this type has been made by Kopp & Rabin (1992) and Martínez Pillet & Vázquez (1990, 1993), the former at 1.56 μm, the latter around 6300 Å. Their results show a clear relationship, as illustrated for the infrared data in Fig. 2, in which data from the same spots is plotted as in Fig. 1.

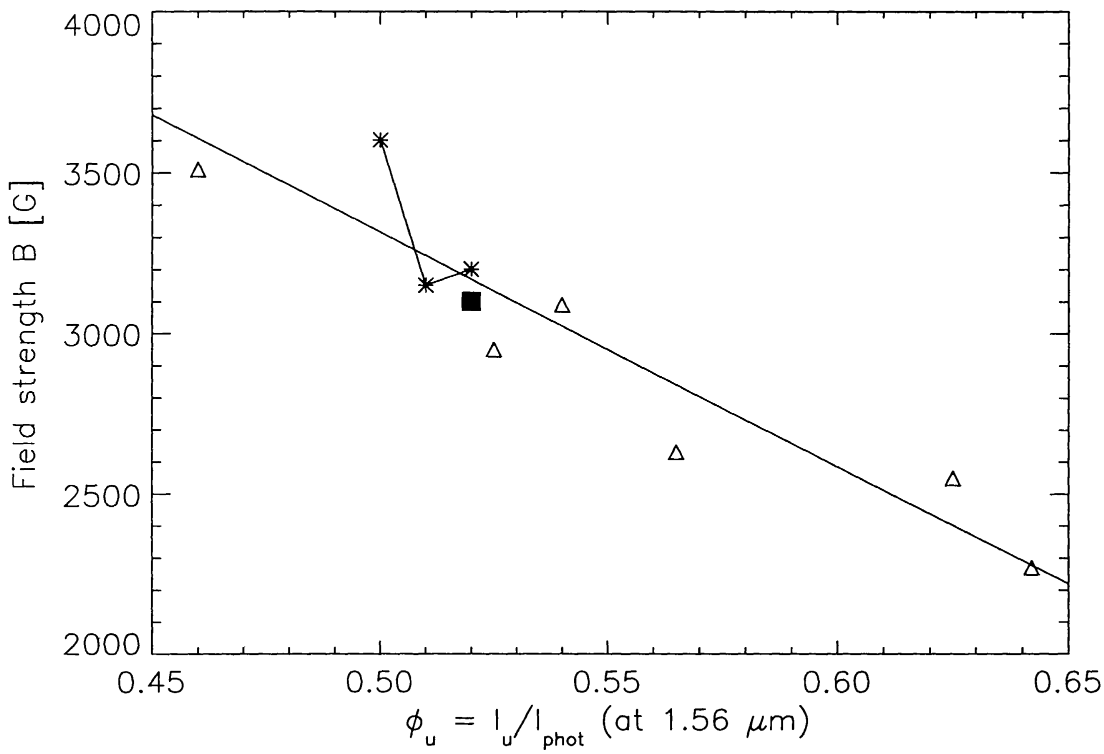

Figure 2. Umbral field strength B vs. umbral normalized continuum intensity ϕ_u. Plotted is the the maximum value of B minimum value of ϕ_u of each of the sunspots represented in Fig. 1 (same symbols).

The first to investigate the T vs. B relationship for different locations within a single sunspot was Von Klüber (1947), who published figures of B and I_c along slices through sunspots. Abdussamatov (1971, cf. 1973) plotted I_c vs. B for a sunspot pair. Although both his intensity and magnetic-field measurements are affected by stray light, his results for the preceding spot look similar to the best infrared observations and show the separate signatures of the umbra and the penumbra, including the relatively sharp transition in brightness between the two.

Gurman & House (1981) carry out a Milne-Eddington inversion of Fe I 6302.5 Å to determine the magnetic parameters at a large number of locations in 4 sunspots. They find a linear relationship between B and I_c for all their data sets. Although they correct for stray light their relationship remains linear all the way from $I_c \approx 0.3$ to 1.0 and from $B = 1800$ G to 0 G. In the absence of stray light and a correctly measured B one would expect to see a break in the relationship at the umbral boundary. Lites et al. (1991a) conclude from their high spatial resolution observations (of Stokes I only) that on scales larger than those of umbral dots B is inversely correlated with continuum intensity. Martínez Pillet & Vázquez (1990, 1993) and Del Toro Iniesta et al. (1991) have also presented such relationships. In particular Martínez Pillet & Vázquez observed at many locations in 8 sunspots and carried out a thorough stray-light analysis. They found a nearly linear relationship between B^2 and T throughout the umbra and into the penumbra. Again no break is seen at the umbral boundary. Note that the scatter in B^2 vs. T between different sunspots is not significantly larger than the scatter between individual positions in one spot. A linear trend of B^2 vs. T is also presented by Lites et al. (1993), but for the umbra only of a small, symmetric sunspot.

Using infrared data Kopp & Rabin (1992) and Solanki et al. (1993) also find, like the rest of the investigators, that the brightness of the umbra is clearly a function of location, with the darkest part usually coinciding with the strongest magnetic field (which is particularly well determined using infrared data). This position need not coincide with the centre of the sunspot, but may lie quite close to the edge of the penumbra (Solanki et al. 1993), which suggests that the brightness variation is not due to stray light. In addition, the strengths of various spectral lines were also monitored. Since different spectral lines exhibit very different correlations with I_c, stray light can be ruled out as the source of these strength variations.

Hence the umbra does *not* have a unique temperature. Rather, it is a function of position within the umbra. Furthermore, the minimum umbral temperature varies from one sunspot to another, and closely follows the maximum field strength. Both quantities are linear functions of sunspot size.

The first unambiguous sign of a change in the B–T relationship across the umbral boundary was provided by Kopp & Rabin (1992). They found one linear relationship between B and T in the umbra, but another (steeper) linear relationship in the penumbra. The break at the umbral boundary is even more clearly visible in the data of Solanki et al. (1993). They find, however, that the vertical component of the magnetic field, B_z, varies almost linearly with T across the whole sunspot. This suggests that older studies showing a linear relationship between B and T may have been measuring B_z rather than B. The break in the

B vs. T relationship at the umbral boundary is also clearly visible in the work of Balthasar & Schmidt (1993). According to their high-resolution observations the B in the penumbra changes by nearly a factor of 2, while T remains almost constant. Hence, unlike the umbra, the penumbra appears to be almost free of radial large-scale variations of T. There are, of course, considerable small-scale variations (visible as penumbral filaments) and according to Grossmann-Doerth & Schmidt (1981) there probably are large-scale azimuthal variations as well.

Let me now outline how the T vs. B relationship may be used to estimate the Wilson depression and the curvature forces within a sunspot. By integrating the radial component of the MHD force-balance equation (over the radial coordinate r) one obtains

$$P_0(z) - P(r,z) = \frac{1}{8\pi}\left(B_z^2(r,z) + F_c(r,z)\right), \quad (7)$$

where P_0 is the gas pressure outside the sunspot, P the gas pressure in the sunspot, B_z the vertical component of the magnetic field and F_c an integral describing the magnetic curvature forces. Eq. (7) is valid for a given geometrical height z. B_z is directly observed at some optical depth τ corresponding to an unknown z, $P(r,\tau)$ is a more or less unique function of the temperature stratification in the sunspot (in hydrostatic equilibrium), while $P_0(z)$ is known from standard quiet sun atmosphere and convection zone models. With this information it is possible, in a first step, to determine the Wilson depression $Z_W(r)$ (Martínez Pillet & Vázquez 1990, 1993) and its variation across the sunspot (Solanki et al. 1993) if one assumes that $F_c = 0$. Basically, one searches for the depth in the quiet sun at which P_0 equals the combined measured magnetic and gas pressure inside the sunspot. This depth is approximately the Wilson depression. In a following step one can compare the $Z_W(r)$ derived in this manner with other measurements of this quantity and therefrom set limits on the curvature forces and magnetic gradients in the sunspot. Two results of such an analysis are: Curvature forces are of a similar magnitude as gas pressure gradient driven forces, and at $\tau_c = 1$ the Gas pressure in a sunspot is always larger than the magnetic pressure (i.e. plasma $\beta > 1$).

The sensitivity of the T vs. B relationship to magnetic curvature forces is illustrated by Rüedi et al. (1995), who analysed it on 3 different days for the same (relatively young) sunspot and found evidence for a relaxation process. This is visible in Fig. 2 in which this sunspot is marked by the three stars. The symbol lying furthest from the solid line represents the observations of the first day, when the sunspot also showed extreme magnetic gradients ($dB/dz \approx 1400$ G/arc s) corresponding to currents almost as large as the largest seen in a flaring δ-spot (Zirin & Wang 1993). On the other days no such large gradients are seen and the T vs. B relationship of this spot also looks more similar to that of other mature sunspots.

4.3. Dependence on the solar cycle

One of the most surprising discoveries related to sunspots has been the dependence of the umbral core brightness of large sunspots ($d_u > 8''$) on the phase of the solar cycle found by Albregtsen & Maltby (1978), cf. Albregtsen & Maltby (1981a). Sunspots are darkest early in the cycle. According to Albregtsen et al. (1984) the solar cycle variation is approximately of the same magnitude as the difference in brightness between sunspots at high and low latitudes (5–35°). Since the average latitude of sunspots changes over the solar cycle it is conceivable that the solar cycle dependence is mainly a latitude dependence. Albregtsen et al. (1984) have therefore carefully corrected for this effect and found that the cycle dependence clearly holds also when ϕ_u is extrapolated to $\mu = 1$. Also, Maltby (1992a) has pointed out that the scatter of ϕ_u vs. latitude is larger than ϕ_u vs. solar cycle phase. Unfortunately, no other group has made a similar study. Confirmation of these intriguing results appears particularly important in the light of the recently found dependence of sunspot brightness on size.

Two explanations have been put forward. Schüssler (1980) proposed that the umbral brightness may be influenced by the age of the sub-photospheric flux tube (formed and wound up at the bottom of the convection zone over the solar cycle). Yoshimura (1983), on the other hand, suggests that it is rather the depth in the convection zone at which the flux tube is formed which is responsible for the umbral temperature. He proposes that this depth varies over the solar cycle. Current thinking, however, anchors the flux tubes in the overshoot layer below the convection zone, making the second explanation less plausible.

5. Umbral 2-component Models

The temperature is a function of position within the umbra. This follows both from low-resolution observations (see, e.g., Sect. 4.2) and high-resolution observations, which often show umbral dots (see the reviews by Muller 1992 and Sobotka 1996). Although some umbrae appear to be free of dots and reveal only filaments (Livingston 1991) the majority are thought to be composed of bright umbral dots and a dark umbral background.

Hence the obvious next step in umbral modelling is the construction of 2-component models. The advantage of this simple scheme is that by varying the filling factors of the two components it is possible to produce a whole range of brightnesses. This strength is at the same time also a weakness of 2-component models due to the fact that umbral dots are generally not spatially resolved and there is an ongoing debate regarding their true temperature and field strength (or the height variation of these quantities). Since the temperature of the hot component cannot be determined with certainty, the filling factor is equally uncertain (the measured intensities cannot distinguish between the two parameters). Due to the unknown $B(z)$ of umbral dots, it is not clear how the filling factor varies with height. Also unknown is whether hot features in the upper atmosphere are at all related to hot features in the lower photosphere.

Early 2-component models of the umbra have been constructed by Makita (1963) and Obridko (1968). More recent 2-component models are due to Adjabshirzadeh & Koutchmy (1983), Staude et al. (1983, 1984), Obridko (1985), Obridko & Staude (1988), Pahlke & Wiehr (1988), Sobotka (1988) and Sobotka et al. (1993). Note that the models of Staude et al. (1983, 1984) possess 2-components in the higher layers only. In the photosphere and chromosphere they are single component models.

The most comprehensive 2-component model is the working model of Obridko & Staude (1988). They propose that different mixtures of cool and hot gas describe the observed umbral core and the umbral dots. In the umbral core the hot and bright component must occupy a volume fraction of β = 5–10% in order to reproduce the observations of Albregtsen & Maltby (1981a), while in the dots β = 50% in order to reproduce the high-resolution observations of Wiehr & Stellmacher (1985). The filling factor of the secondary component varies with height according to horizontal pressure balance assuming a weak secondary-component field at the $\tau_c = 1$ level and a height-independent field of 3000 G in the umbral core. The basic philosophy behind this choice is that the thermal structure follows field lines.

In conclusion, 2- or more component models are important and some good models of this type have been constructed. Nevertheless, a basic uncertainty remains and the prime need at the moment is for observations that allow the properties of umbral dots to be determined, not just in continuum-forming layers, but at all heights in the atmosphere.

6. Penumbral Models

The most striking feature of the penumbra are the fibrils. It is therefore not surprising that the major fraction of the effort devoted to the study of penumbral brightness has concentrated on the fine structure. I.e. the effort has focussed on the horizontal variation of the temperature rather than on its vertical stratification. Consequently, empirical penumbral models are rare.

The first penumbral model was that of Kjeldseth-Moe & Maltby (1969). It is a single-component scaled quiet-sun model (with $\delta\theta = 0.055$, where $\theta = 5040/T$) that can reproduce weak spectral lines and observations of penumbral continuum intensity (normalized to the quiet sun value) $\phi_{\rm pen}(\lambda) = I_{\rm pen}(\lambda)/I_{\rm phot}(\lambda)$. It is only meant to be a model of the lower photospheric layers of the penumbra.

Models that include not just the photosphere, but also the chromosphere have been published by Yun et al. (1984) and Ding & Fang (1989). These models are based on observations of strong spectral lines (Ca II H and K, two of the Ca II IR triplet lines and the Na I D doublet in the case of Yun et al. and Ca II H and K, the Ca II IR triplet, Hα and Hβ in the case of Ding & Fang). Most of the effort has gone into modelling the chromospheric layers. In particular the authors have not taken into account weak lines or continuum observerations to reliably fix photospheric temperature and hence both models are too cool in the photosphere. A promising way of producing a better penumbral model is to consistently combine the chromospheric part of such a model (that of Ding & Fang 1989 appears less arbitrary) with a photospheric penumbral model, such as that of Kjeldseth-Moe & Maltby (1969).

Judging from white-light images of penumbrae a 2-component penumbral model seems obvious. Such a model has indeed been constructed by Kjeldseth-Moe & Maltby (1974a), with one component describing the bright filaments and the other the dark. It is a straightforward extension of the Kjeldseth-Moe & Maltby (1969) model to 2 components, being partly based on the same data. The weighted sum of the $\phi_{\rm pen}$ of both components is compared with the observed, low-spatial-resolution data of Maltby (1972). The temperature stratification of both components is scaled down from the photosphere. The difference between the 2 components is given by the continuum observations of

Muller (1973), which indicate $I_c \approx 0.3$–$0.7 I_{\rm phot}$ and a filling factor of 0.6 for the dark component, and $I_c \approx 0.7$–$1.0 I_{\rm phot}$ with a filling factor of 0.4 for the bright. These observations were criticized by Grossmann-Doerth & Schmidt (1981). Their observations showed that the actual distribution of brightness in a penumbra is single peaked. Collados et al. (1987) argue that both points of view are justified. A single peaked distribution can also be produced by the sum of two distinct brightness components, if each component exhibits a sufficiently wide distribution of brightness.

Finally, Del Toro Iniesta et al. (1994) have inverted many line profiles (801 to be precise) of Fe I 5576.1 Å. The profiles were constructed by scanning a narrow-band filter through the line. For each spatial location a temperature structure is obtained. The scatter of the temperature of all the models around the mean temperature stratification varies between 200 and 500 K at different heights, with the scatter being smallest in the low and mid photosphere, i.e. in the layers best constrained by the observations.

Del Toro Iniesta et al. (1994) have also constructed average temperature stratifications at different distances from the umbra. The temperature generally increases somewhat as the distance to the umbra increases (which may well be the result of stray light). They have also created a model of the average penumbral temperature stratification by combining the $T(\tau)$ of 411 of their models. This average is very similar to the Kjeldseth-Moe and Maltby models. Interestingly the standard deviation of the individual Del Toro Iniesta et al. models is smaller than the difference between the 2 Kjeldseth-Moe & Maltby (1974a) components, suggesting that the spatial resolution of the observations underlying the Del Toro Iniesta et al. (1994) models was lower than of Muller's observations.

7. Conslusion

Observations of sunspot brightness and spectra at high or low resolution are numerous and quite a number of these have found their way into the various models of sunspot temperature. These in turn have evolved to the point that a single umbral model (or a pair of them in the case of 2-component models) can now reproduce a large variety of observations. In the penumbra the situation is less satisfactory. Currently no model exists which gives a good representation of both the photospheric and the chromospheric layers. Inversions of large data sets (such as pioneered by Del Toro et al. 1994, but ideally based on more spectral lines) and detailed numerical simulations whose output can be directly compared with observations may well be the direction in which this field will move in the future.

Acknowledgments. I thank A. Brkovic for helping with Figs. 1 and 2, as well as K. Harvey for giving new meaning to an old abbreviation.

References

Abdussamatov H.I., 1971, Solar Phys. 16, 384

Abdussamatov H.I., 1973, Bull. Astron. Inst. Czech 24, 118

Adjabshirzadeh H., Koutchmy S., 1983, Astron. Astrophys. 122, 1

Albregtsen F., Maltby P., 1978, Nature 274, 41

Albregtsen F., Maltby P., 1981a, Solar Phys. 71, 269

Albregtsen F., Maltby P., 1981b, in Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 127

Albregtsen F., Jorås P.B., Maltby P., 1984, Solar Phys. 90, 17

Alfvén H., 1943, Arkiv. Mat. Astron. Fysik 29, No. 11

Altrock R.C., Canfield R.C., 1974, Astrophys. J. 194, 733

Ayres T.R., 1996, in Stellar Surface Structure, K.G. Strassmeier, ed., Kluwer, Dordrecht, IAU Symp. 176, 317

Avrett E.H., 1981, in The Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 235

Balthasar H., Schmidt W., 1993, Astron. Astrophys. 279, 243

Baranovsky E.A., 1974a, Izv. Krimsk. Astron. Obs. 49, 25

Baranovsky E.A., 1974b, Izv. Krimsk. Astron. Obs. 51, 56

Beebe H.A., Baggett W.E., Yun H.S., 1982, Solar Phys. 79, 31

Boyer R., 1980, Astron. Astrophys. Suppl. Ser. 40, 277

Brants J.J., Zwaan C., 1982, Solar Phys. 80, 251

Bray R.J., Loughhead R.E., 1964, Sunspots, Chapman and Hall

Briand C., Solanki S.K., 1995, Astron. Astrophys. 299, 596

Caccin B., Gomez M.T., Severino G., 1993, Astron. Astrophys. 276, 219

Chou D.-Y., 1987, Astrophys. J. 312, 955

Collados M., del Toro Iniesta J.C., Vázquez M., 1987, in The Role of Fine-Scale Magnetic Fields on the Structure of the Solar Atmosphere, E.-H. Schröter, M. Vázquez, A.A. Wyller, eds., Cambridge University Press, Cambridge, p. 214

Collados M., Martínez Pillet V., Ruiz Cobo B., del Toro Iniesta J.C., Vázquez M., 1994, Astron. Astrophys. 291, 622

Cowling T.G., 1957, Magnetohydrodynamics, Interscience, New York

Craig I.J.D., Brown J.C., 1986, Inverse Problems in Astronomy. A Guide to Inversion Strategies for Remotely Sensed Data, Adam Hilger, Bristol

Deinzer W., 1965, Astrophys. J. 141, 548

Del Toro Iniesta J.C., Martínez Pillet V., Vázquez M., 1991, in Solar Polarimetry, L. November, ed., National Solar Obs., Sunspot, NM, p. 224

Del Toro Iniesta J.C., Tarbell T.D., Ruiz Cobo B., 1994, Astrophys. J. 436, 400

Dicke R.H., 1970, Astrophys. J. 159, 25

Ding M.D., Fang C., 1989, Astron. Astrophys. 225, 204

Ekmann G., 1974, Solar Phys. 38, 73

Ekmann G., Maltby P., 1974, Solar Phys. 35, 317

Foukal P., 1981, in The Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 391

Fricke K., Elsässer H., 1965, Z. Astrophys. 63, 35

Gokhale M.H., Zwaan C., 1972, Solar Phys. 26, 52

Grossmann-Doerth U., Schmidt W., 1981, Astron. Astrophys. 95, 366

Gurman J.B., House L.L., 1981, Solar Phys. 71, 5

Hale G.E., 1892, Astron. Astrophys. 11, 310

Hénoux J.C., 1968, Ann. Astrophys. 31, 511

Hénoux J.C., 1969, Astron. Astrophys. 2, 288

Houtgast J., van Sluiter A., 1948, BAN 10, 325

Jakimiec J., 1965, Acta Astron. 15, 145

Jensen E., Maltby P., 1965, Astrophys. Norveg. 10, 17

Kjeldseth Moe O., Maltby P., 1969, Solar Phys. 8, 275

Kjeldseth Moe O., Maltby P., 1974a, Solar Phys. 36, 101

Kjeldseth Moe O., Maltby P., 1974b, Solar Phys. 36, 109

Kneer F., 1972, Astron. Astrophys. 18, 39 (in german)

Kneer F., Mattig W., 1978, Astron. Astrophys. 65, 17

Kollatschny W., Stellmacher G., Wiehr E., Falipou M.A., 1980, Astron. Astrophys. 86, 245

Kopp G., Rabin D., 1992, Solar Phys. 141, 253

Kurucz R.L., 1991, in Precision Photometry: Astrophysics of the Galaxy, A.G. Davis Philip, A.R. Upgren, K.A. Janes, eds., L. Davis Press, Schenectady

Linsky J.L., Avrett E.H., 1970, Publ. Astron. Soc. Pacific 82, 169

Lites B.W., Skumanich A., 1981, in Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 152

Lites B.W., Skumanich A., 1982, Astrophys. J. Suppl. Ser. 49, 293

Lites B.W., Skumanich A., Rees D.A., Murphy G.A., Carlsson M., 1987, Astrophys. J. 318, 930

Lites B.W., Bida T.A., Johannesson A., Scharmer G.B., 1991a, Astrophys. J. 373, 683

Lites B.W., Elmore D.F., Seagraves P., Skumanich A., 1993, Astrophys. J. 418, 928

Livingston W.C., 1991, Nature 350, 45

Makita M., 1963, Publ. Astron. Soc. Japan 15 145

Maltby P., 1972, Solar Phys. 26, 76

Maltby P., 1977, Solar Phys. 55, 335

Maltby P., 1992a, in Sunspots: Theory and Observations, J.H. Thomas, N.O. Weiss, eds., Kluwer, Dordrecht, p. 103

Maltby P., 1992b, in Surface Inhomogeneities on Late-Type Stars, P.B. Byrne, D.J. Mullan, eds., Springer-Verlag, Berlin, p. 124

Maltby P., Avrett E.H., Carlsson M., Kjeldseth-Moe O., Kurucz R.L., Loeser R., 1986, Astrophys. J. 306, 284

Martínez Pillet V., Vázquez M., 1990, Astrophys. Space Sci. 170, 75

Martínez Pillet V., Vázquez M., 1993, Astron. Astrophys. 270, 494

Mattig 1958, Z. Astrophys. 44, 280

Mattig W., 1969, Solar Phys. 6, 413

Mattig W., Kneer F., 1978, Astron. Astrophys. 65, 11

Mauas P., Avrett E.H., Loeser R., 1988, Astrophys. J. 330, 1008

Michard R., 1953, Ann. Astrophys. 16, 217

Muller R., 1973, Solar Phys. 32, 409

Muller R., 1987, in The Sun, Proc. 10th European Regional Astron. Meeting of the IAU, L. Hejna, M. Sobotka, eds., Astron. Inst. Czechoslovak Acad. Sci., Prag, p. 15

Muller R., 1992, in Sunspots: Theory and Observations, J.H. Thomas, N.O. Weiss, eds., Kluwer, Dordrecht, p. 175

Nicolas K.R., Kjeldseth-Moe O., Bartoe J.-D.F., Brueckner G.E., 1981, in Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 167

Obridko V.N., 1968, BAC 19, 186

Obridko V.N., 1985, Sunspots and Complexes of Activity (in Russian), Nauka, Moskva

Obridko V.N., Staude J., 1988, Astron. Astrophys. 189, 232

Pahlke K.-D., Wiehr E., 1990, Astron. Astrophys. 228, 246

Rossbach M. Schröter E.H., 1970, Solar Phys. 12, 95

Rüedi I., Solanki S.K., Livingston W., 1995, Astron. Astrophys. 302, 543

Rüedi I., Solanki S.K., Livingston W., 1996, these proceedings

Schlüter A., Temesvary S., 1958, in Electromagnetic Phenomena in Cosmical Physics, B. Lehnert, ed., Reidel, Dordrecht, IAU Symp. 6, 263

Schröter E.H., 1971, in Solar Magnetic Fields, R.F. Howard, ed., Reidel, Dordrecht, IAU Symp. 43, 167

Schüssler M., 1980, Nature 288, 150

Severino G., Gomez M.T., Caccin B., 1994, in Solar Surface Magnetism, R.J. Rutten, C.J. Schrijver, eds., Kluwer, Dordrecht, p. 169

Sobotka M., 1988, Solar Phys. 124, 37

Sobotka M., 1996, these proceedings

Sobotka M., Bonet J.A., Vázquez M., 1993, Astrophys. J. 415, 832

Solanki S.K., 1990, in Solar Photosphere: Structure, Convection and Magnetic Fields, J.O Stenflo, ed., , IAU Symp. 138, 103

Solanki S.K., Rüedi I., Livingston W., 1992, Astron. Astrophys. 263, 339

Solanki S.K., Walther U., Livingston W., 1993, Astron. Astrophys. 277, 639

Staude J., 1981, Astron. Astrophys. 100, 284

Staude J., Fürstenberg F., Hildebrandt J., Krüger A., Jakimiec J., Obridko V.N., Siarkowski M., Sylwester B., Sylwester J., 1983, Acta Astron. 33, 441

Staude J., et al. 1984, Astron. Zh. 61, 956

Stellmacher G., Wiehr E., 1970, Astron. Astrophys. 7, 432

Stellmacher G., Wiehr E., 1975, Astron. Astrophys. 45, 69

Stellmacher G., Wiehr E., 1988, Astron. Astrophys. 191, 149

Stumpff P., 1961, Z. Astrophys. 52, 73 (in german)

Teplitskaya R.B., Grigoryeva S.A., Skochilov B.G., 1978, Solar Phys. 56, 293

Van t'Weer, 1963, Ann. Astrophys. 26, 185

Von Klüber H., 1947, Z. Astrophys. 24, 121

Webber J.C., 1971, Solar Phys. 16, 340

Wiehr E., Stellmacher G., 1985, in High Resolution in Solar Physics, R. Muller, ed., Lecture Notes in Physics Vol. 233, Springer Verlag, Berlin, p. 369

Wittmann A., Schröter E.H., 1969, Solar Phys. 10, 357

Yoshimura H., 1983, Solar Phys. 87, 251

Yun H.S., 1971, Solar Phys. 16, 379

Yun H.S., Beebe H.A., Baggett W.E., 1981, in Physics of Sunspots, L.E. Cram, J.H. Thomas, eds., National Solar Obs., Sunspot, NM, p. 142

Yun H.S., Beebe H.A., Baggett W.E., 1984, Solar Phys. 92, 145

Zirin H., Wang H., 1993, Nature 363, 426

Zwaan C., 1965, Rech. Astr. Obs. Utrecht 17, part 4

Zwaan C., 1974, Solar Phys. 37, 99

Zwaan C., 1975, Solar Phys. 45, 115