Chapter 3

Role of Nonlinearities and Stochasticities in the long term modulation of Solar Cycles

3.1 Introduction

Sun is a magnetically active star whose activity is a result of the magnetic dynamo process operating in the Sun's convection zone as already discussed in previous chapters (also see, e.g., Charbonneau, 2020; Karak et al., 2014). Solar surface magnetic activity varies cyclicly with the main period of about 11 years (called the Schwabe cycle) or, considering inversion of the sign of its magnetic polarity, the 22-year Hale cycle. The physics of the dynamo mechanism is currently believed to be reasonably well understood. However, solar cyclicity is far from being a regularly ticking clock and experiences essential long-term variability at timescales longer than the Schwabe cycle. The solar cycles are not perfectly regular and vary in length, shape, and strength/intensity, or even can enter periods of almost inactive state, called grand minima of solar activity (e.g., Usoskin, 2017).

The standard index quantifying solar activity is related to sunspot numbers which are available from 1610 AD onward with the quality degrading backward in time. On one hand, this 410-year-long series exhibits a great deal of variability covering the range from an almost spotless period of the Maunder minimum between 1645 – 1715 AD (Eddy, 1976) to an epoch of very active Sun between 1940 – 2009 called the Modern grand maximum (Solanki et al., 2004; Usoskin et al., 2007). This great variability raises important questions, answers to which can put crucial observational constraints on the solar/stellar dynamo theory:

- Do the changes between the Maunder minimum and the Modern grand maximum cover the full possible range of solar variability?
- Do the grand minima and maxima represent special states of the solar dynamo or simply represent the tails of the distribution?
- How typical are these changes?
- Do the grand minima episodes appear periodically or randomly?
- What physical processes drive such changes?

The four-century-long sunspot number series is not sufficiently long to answer these questions, and a much longer dataset is needed to form a basis for the answers. Fortunately, solar activity can be reliably reconstructed from indirect natural proxy data (cosmogenic radioisotopes) on the timescale of 10 – 12 millennia, during the period of the Holocene with a stable warm climate on Earth, as discussed in Chapter 1. This reconstruction extends the solar-activity dataset by a factor of about 25 making it possible to perform a thorough statistical analysis of solar variability. In the following section, we discuss the implications of the long-term solar variability for the solar dynamo theory and our present level of understanding of the related physics.

3.2 Causes for long-term variations in the solar activity

With the above discussion of the solar dynamo, we now identify the causes of the cycle modulation. As the solar dynamo is nonlinear, it is natural to expect that the modulation in the solar cycle is caused by the back reaction of the flow on the magnetic field. Therefore, we first identify the nonlinearities in the dynamo models and check if they can lead to cycle modulations.

3.2.1 Nonlinearities in the dynamo

As we can see from Equation (1.4), the magnetic field can alter the flow directly through the Lorentz force. The Lorentz force can come from the mean magnetic field and the mean current (which is popularly known as the Malkus-Proctor effect (Malkus and Proctor, 1975) in the mean-field context) and from the fluctuating magnetic field and the current. The mean magnetic field can also alter the anisotropic convection which is responsible for transporting angular momentum and maintaining differential rotation and meridional flow in the Sun (Kitchatinov et al., 1994b). This effect is also called micro-feedback. When these Lorentz feedbacks of the magnetic fields are included in the flow, we expect a long-term modulation in the flow and the magnetic cycle.

In mean-field models, the magnetic feedback is captured by considering a direct Lorentz force of the mean magnetic field in the zonal flow (e.g., Bushby, 2006) and/or by a quenching term in the Λ effect (e.g., Küker et al., 1999). Cycle modulations in these systems can generally happen in two ways. In the first one, the magnetic energy of the primary mode (the equatorial symmetric or antisymmetric) can oscillate due to the energy exchange between the flow and the magnetic field via the nonlinear Lorentz feedback. In this case, a considerable amount of modulation in the differential rotation is observed. In the second case, a small magnetic perturbation on the differential rotation can slowly change one dominant dynamo mode into another. In this case, the magnetic field parity can change (between equatorially symmetric (quadrupole) and antisymmetric (dipole)) without producing a large change in the differential rotation. These two mechanisms are respectively coined as Type II and I modulations. Mean-field models have demonstrated that nonlinear back reaction of magnetic field on large-scale flow through these types of modulations can induce a variety of modulation patterns in the cycle amplitude, including grand minima and parity modulations which do not leave a strong imprint in differential rotation (e.g., Beer et al., 1998; Bushby, 2006; Knobloch et al., 1998; Weiss and Tobias, 2016). Both types of modulation can arise in a model, however, as the observed differential rotation shows a tiny variation over the solar cycle, we expect the Type II modulation is less likely to occur in the Sun. Even for Type I modulation, a detailed comparison of the magnetic field and the flows in these models with the observations is missing (also see Section 7 of Charbonneau, 2020, for a discussion on this topic).

Next is the meridional flow, which is the second important large-scale flow in the Sun. As it arises due to a slight imbalance between the non-conservative centrifugal and buoyancy forces, we expect its large variation. In fact, the global simulations find a large variation in the meridional flow despite a small variation in the differential rotation (Karak et al., 2015). In Babcock–Leighton type dynamo models, meridional circulation plays a crucial role in transporting the field on the surface from low to high latitudes and down to the deeper CZ where the shear produces a toroidal field. The toroidal field is transported to the low latitudes via the equatorward return flow and possibly causes the equatorward migration of the sunspot belt. Thus, in these models, meridional circulation largely regulates the cycle period (Dikpati and Charbonneau, 1999; Karak and Choudhuri, 2011). It also affects the strength of the field as a weak meridional circulation allows the field to advect slowly and gives more time for diffusion (Yeates et al., 2008). Karak (2010) showed that when a variable meridional flow is used in a high diffusivity dynamo model to match the observed solar cycle periods, the amplitudes of the cycles are also modelled up to some extent (also see Hazra et al., 2015; Karak and Choudhuri, 2011, for modelling various aspects of solar cycle using variable meridional flow). In an extreme case, a largely reduced meridional circulation can trigger a Maunder-like grand minimum. In reality, how large the variation in the meridional flow occurred in the past remains uncertain. However, it is obvious that any changes in the flow can lead to modulation in the solar cycle.

Turbulent transport as parameterized by, for example, the turbulent diffusivity, Λ effect, and heat diffusion are also nonlinear because the Lorentz force of the small-scale as well as the large-scale dynamo-generated fields act on the small-scale turbulent flows. However, due to limited knowledge in the turbulence theory for solar parameter regions, we do not have a satisfactory model for the magnetic field-dependent form of the turbulent transport parameters; however, see Ruediger and Kichatinov (1993) and Kitchatinov et al. (1994a) respectively, for the magnetic field-dependent forms of α and η based on the quasi-linear approximation.

Finally, the toroidal to poloidal part of the dynamo loop involves some nonlinearities. When the generation of poloidal field is due to the classical α effect, there is a well-known α quenching of the form $1/\left(1+(B/B_{\rm eq})^2\right)$ with $B_{\rm eq}$ being the equipartition field strength. However, this type of α quenching tries to make a stable cycle rather than producing irregularity in the cycle. In the Babcock–Leighton dynamo, the generation of the poloidal field from the toroidal one also involves several nonlinearities. Here we discuss the following three potential candidates for these.

- **Flux loss due to magnetic buoyancy**

I have discussed this nonlinearity elaborately in the Chapter 2, hence it is summarized very briefly here. The magnetic buoyancy as proposed by Parker (1955a) plays a critical role in the emergence of BMRs on the solar surface. As the shearing of the poloidal field due to differential rotation intensifies the strength of the toroidal field, there comes a point where the magnetic energy density of the toroidal flux tubes becomes greater than the kinetic energy of the local convective plasma inside the CZ, as a result, the flux tubes become buoyant and start rising through the CZ, eventually giving birth to the sunspots. Following this process, the strength of the magnetic field gets locally reduced as a part of it rises due to buoyancy and the flux tube becomes inefficient to produce further sunspots for some time (however see a counter-argument by Rempel and Schüssler, 2001). The sharp rise in the flux loss once the toroidal field strength exceeds a certain value clearly indicates a nonlinear mechanism in the solar dynamo. Incorporating this mechanism of toroidal flux loss due to buoyancy in a simple manner, Biswas et al. (2022) showed that this nonlinear process plays a critical role in limiting the growth of the solar dynamo which is a potential mechanism to explain why different solar cycles rise differently depending on their strength but all the solar cycles decay with similar statistical properties (see Figure 2.4). They found that introducing the flux loss in the dynamo simulations makes all the cycles (irrespective of their strength) have a similar amount of toroidal flux with similar latitudinal distribution which is critical to reproduce the long-term features of the latitudinal distribution of the sunspots (Cameron and Schüssler, 2016; Waldmeier, 1955). However, also see Cameron and Schüssler (2016) and Talafha et al. (2022a) for an alternative explanation of the universal decay of the solar cycle using cross-equatorial diffusion.

- **Latitude quenching**

It has been recently found that when BMRs appear in low latitudes, the leading polarities from both hemispheres get efficiently cancelled at the equator. This leads to the following polarities of the BMRs efficiently getting carried to the poles and contributing to the polar field, see Figure 3.1. On the other hand, BMRs appearing in the high latitudes do not exhibit efficient cross-hemisphere cancellation and thus do not contribute significantly to the polar field (Jiang et al., 2014a; Karak and Miesch, 2018). It is seen that strong cycles produce more BMRs at high latitudes. In other words, the average latitude of the BMRs is

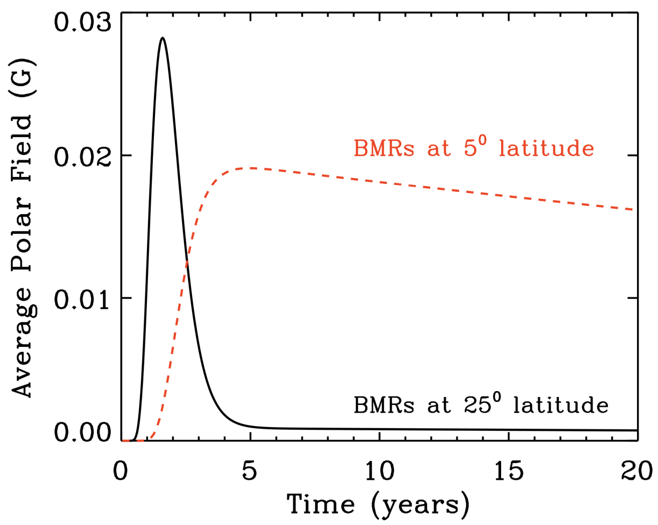

Fig. 3.1 Demonstration of latitude quenching: Temporal evolution of the net polar flux generated from two BMRs deposited symmetrically in two hemispheres at different latitudes (Karak, 2020). Image Credit: Biswas et al. (2023b)

high for the strong cycles (Mandal et al., 2017; Solanki et al., 2008). Hence for a strong cycle, most of its BMRs emerging at high latitudes would be less efficient in polar field production and vice versa for the weak cycles. This mechanism, so-called the *latitude quenching* (Petrovay, 2020) may help to stabilize the growth of the magnetic field in the Sun (Jiang, 2020).

Introducing a latitude-dependent threshold on the BMR emergence condition into a 3D Babcock–Leighton dynamo simulation, Karak (2020) showed that latitude quenching can regulate the growth of a magnetic field when the dynamo is not too supercritical.

- **Tilt quenching**

The tilt angle of BMR plays a crucial role in generating the poloidal field in the Sun. For a given latitude, the amount of generated poloidal field increases with the increase of tilt. The thin flux tube model for the sunspot formation suggests that the tilt of the BMR is produced due to a torque acting on the diverging flows produced from the apex of the rising flux tube which forms the BMR (D'Silva and Choudhuri, 1993; Fan et al., 1994). Thus, if

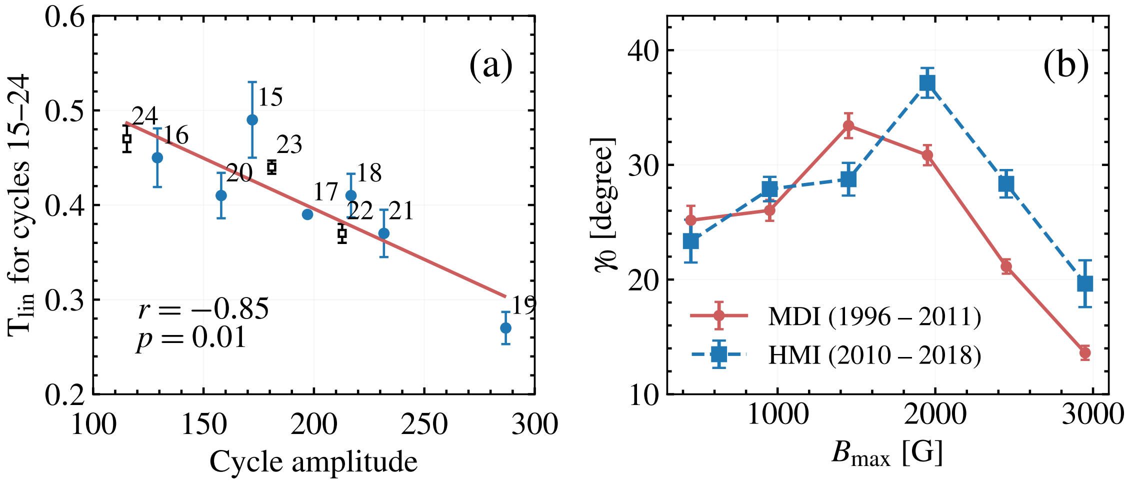

Fig. 3.2 Demonstration of tilt quenching: (a) Tilt coefficient (mean tilt normalized by the mean latitude) vs the cycle strength (Jiao et al., 2021); also see Dasi-Espuig et al. (2010). (b) The slope of Joy's law vs the maximum field strength in the BMR (Jha et al., 2020). Image Credit: Biswas et al. (2023b)

the magnetic field of the sunspot-forming flux tube is strong, then it will rise quickly and the Coriolis force will get less time to induce tilt. In a strong cycle, the toroidal magnetic field is strong and the number of BMRs with strong magnetic field tends to be high (Jha et al., 2020). Thus, we expect the mean tilt in that cycle to be smaller. A lesser tilt will produce less poloidal field and the next cycle will be weak. Hence, this can be another potential mechanism for stabilizing the growth of the magnetic cycle through the reduction of tilt which is known as the *tilt quenching*.

The observational evidence of tilt quenching is limited. Dasi-Espuig et al. (2010); Jiao et al. (2021) showed that there is a statistical anti-correlation between the cycle-average tilt of the sunspots with the cycle strength (Figure 3.2a). On the other hand, Jha et al. (2020) examined the variation of BMR tilt with the strength of its magnetic field within a cycle. They found a non-monotonous dependence of the tilt with the BMR field strength as seen in Figure 3.2(b). For weak field strengths, the tilt first increases, however at sufficiently strong field strengths, the BMR tilt starts to decrease rapidly.

3.2.2 Stochastic effects in the dynamo

The solar convection zone is turbulent and thus the turbulent quantities (such as α effect) are subject to fluctuate around their means. Hoyng (1993) showed that as there are finite numbers of convection eddies along the longitudes in the sun, the fluctuations of the turbulent transport coefficients can be larger than their means. There is a long history of including the stochastic noise in the α effect in the mean-field dynamo models. Most of these studies find long-term modulations in the cycle and grand minima in a certain parameter range of the dynamo number (Brandenburg and Spiegel, 2008; Choudhuri, 1992; Gómez and Mininni, 2006; Moss et al., 2008; Ossendrijver and Hoyng, 1996; Ossendrijver et al., 1996).

In Babcock–Leighton dynamo also stochastic fluctuations are unavoidable. The toroidal to poloidal part of this model primarily involves stochastic fluctuations due to the following effects.

- **Scatter around Joy's law**

Observations find that the tilts of the BMRs "statistically" increase with the increase of latitude, which is known as Joy's law. However, a large number of BMRs do not follow this relation strictly, as seen by a huge scatter around the mean trend in Figure 3.3. In fact, there are many BMRs which are having negative tilt which are called the anti-Joy type BMRs. There are often some BMRs that do not obey the Hale polarity rule which are known as the anti-Hale type BMRs (McClintock et al., 2014). These anti-Joy and anti-Hale BMRs (so-called 'anomalous' BMRs), having opposite tilts (negative in the northern hemisphere) and polarity orientations respectively are responsible for generating opposite polarity fields (with respect to the expected polarity) and lead to large fluctuations in the polar field (Hazra et al., 2017; Jiang et al., 2014a; Mordvinov et al., 2022a; Nagy et al., 2017). The randomness in their presence in the solar cycle makes the perturbation

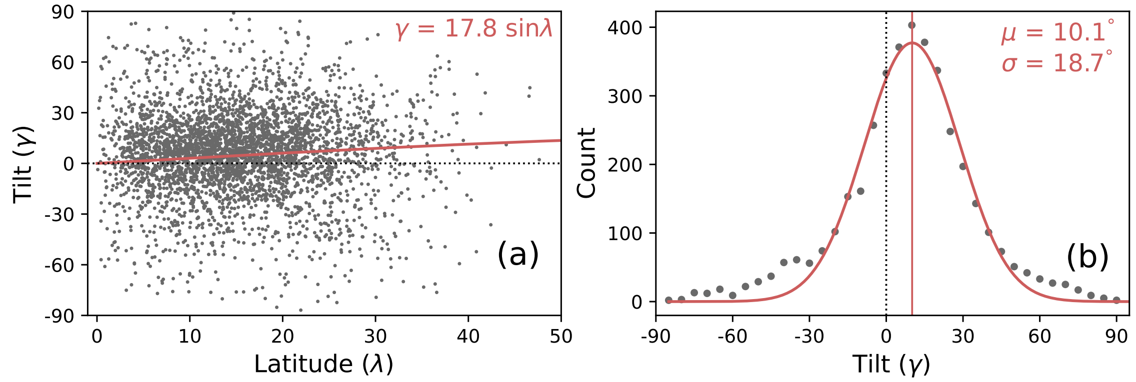

Fig. 3.3 (a) Scatter of BMR tilt around Joy's law (solid line). (b) The tilt distribution with fitted Gaussian (solid line). Here the tilt angles of BMRs are computed by tracking the MDI line-of-sight magnetograms for September 1996 – December 2008. For the details of the tracking algorithm see Sreedevi et al. (2023). Image Credit: Biswas et al. (2023b)

in the polar field vary stochastically. In later chapters, we will discuss more about their impact on the polar field and on the predictability of the solar cycle.

- **Variations in the BMR eruption rates**

There are spatial and temporal variations in the BMR eruptions. BMRs near the equator are much more efficient in generating poloidal field in the Sun because for them the leading polarity can easily connect with the opposite polarity flux from the other hemisphere (Cameron et al., 2013; Jiang et al., 2014a; Karak and Miesch, 2018; Karak, 2020; Mordvinov et al., 2022a). Thus variation in the latitudinal position can produce variation in the generated poloidal field (Baumann et al., 2004). Next, the rate of BMR eruption is not the same throughout the cycle—there is a distribution. Thus, the rate of generation of the poloidal field is also not the same (Karak and Miesch, 2017). Furthermore, the flux contents of the BMR also has a distribution and thus a wrongly tilted BMR with *high flux* can disturb the polar field in the sun considerably (Nagy et al., 2017; Pal et al., 2023).

In summary, the randomness involved in the BMR properties (originated due to the turbulent nature of the convection) produces variation in the poloidal field. Although the sun produces thousands of spots in a cycle, only a few spots are produced (on average) per day. This leads to variations in the polar field comparable to its mean value. In the next section, we shall demonstrate some illustrative results from stochastically driven Babcock–Leighton dynamo models.

3.3 Babcock–Leighton dynamo models for the long-term variation

As discussed above, the generation of the poloidal field in the Babcock–Leighton dynamo models involves some randomness. Thus, in axisymmetric dynamo models, these randomnesses were captured by adding a noise term in the poloidal source (e.g., Charbonneau and Dikpati, 2000). Long-term modulations, including Gnevyshev-Ohl/Odd-Even rule (Charbonneau, 2001; Charbonneau et al., 2007) and grand minima (Charbonneau et al., 2004; Choudhuri and Karak, 2009; Passos et al., 2014, 2012) are naturally produced in these models. Variations within the cycle, like the amplitude-period anti-correlation (Charbonneau and Dikpati, 2000; Karak, 2010) and Waldmeier effect (Biswas et al., 2022; Karak and Choudhuri, 2011) are also reproduced. Karak et al. (2018) showed that a large variation in the Babcock–Leighton process can change the polar field abruptly and this can lead to double peaks in the following cycle. While in most of the studies, the level of fluctuations was tuned to produce the observed variation of the solar cycle including a reasonable number of grand minima, Choudhuri and Karak (2012) and Olemskoy and Kitchatinov (2013) made some estimate of the fluctuations in the Babcock–Leighton process from observations. Choudhuri and Karak (2012) found the correct frequency of grand minima as observed in the cosmogenic data for the last 11,000 years. Olemskoy and Kitchatinov (2013) showed that the statistics of grand minima are consistent with the Poisson random process, indicating the initiation of grand minima to be independent of the history of the past minima.

In recent years, cycle modulations were, in particular, produced by including the variations in the BMR properties in two comprehensive models, namely, 2×2D (Lemerle and Charbonneau, 2017) and 3D dynamo models (Karak and Miesch, 2017). In Figure 3.4, we show cycles from the 3D dynamo model presented by Karak and Miesch (2017). As seen in Figure 3.4(a), the variation in the BMR emergence rate and the flux distribution produce little variation in the solar cycle. When the variation around Joy's law tilt is included, it produces a large variation, including suppressed magnetic activity like the one seen during Dalton minimum and Maunder minimum as shown in Figure 3.4b (the regions shaded in green). Here, the grand minima are identified in the same manner as done in the observed data (Usoskin et al., 2007), i.e., the modelled-sunspot data are first binned in 10 years window and smoothed and then a grand minimum is considered when the smoothed data fall below 50% of the average at least for two cycles.

In Figure 3.5, we present a detailed view of a grand minimum. We find that some of the observed features of the Maunder minimum (hemispheric asymmetry, gradual recovery, slightly longer cycle) are reproduced in this figure. We note that during this grand minimum, some BMRs are still produced, the number of which is a bit larger than that was observed during Maunder minimum (Carrasco et al., 2021; Usoskin et al., 2015; Vaquero et al., 2015; Zolotova and Ponyavin, 2016). However, we should keep in mind that the observations during Maunder minimun were limited (due to the poor resolving power of the 17th-century telescopes) to detect the small BMRs (e.g., Vaquero and Vázquez, 2009); and only big sunspots could be detected at that time. In our Babcock–Leighton dynamo model, few BMRs erupt which produces a poloidal field at a slow rate through the Babcock–Leighton process and the model emerges from the grand minimum episode. It is the downward magnetic pumping included in our model that helps to reduce the magnetic flux loss through the surface and recovers the model from grand minima (Cameron et al., 2013; Karak and Cameron, 2016).

There have been suggestions that during Maunder-like extended grand minima, the Babcock–Leighton process may not operate due to few observed sunspots, and α effect (Parker, 1955a) is the best candidate for this as it efficiently operates in sub-equipartition field strength (Karak and Choudhuri, 2013; Ölçek et al., 2019; Passos et al., 2014). We observe that our model also fails to recover when it enters a deep grand minimum and stops producing BMRs due to the fall of the toroidal field below the threshold for BMR formation. However, this happens very rarely. While it is a critical question to answer what mechanism dominates in recovering the Sun from an extended grand minimum, it is expected that Babcock–Leighton process becomes less efficient during this phase and the α effect certainly helps in recovering the Sun from grand minima.

Dynamo models with stochastic fluctuations also produce grand maxima. Our model presented in Figure 3.4b also produces a few grand maxima shown by the regions shaded in red. Similar to the grand minima, grand maxima are also computed based on the smoothed sunspot number, but here the threshold is taken as 150% of the long-term mean. Systematic studies of grand maxima using dynamo models are limited (however, see Inceoglu et al., 2017; Karak and Choudhuri, 2013; Olemskoy and Kitchatinov, 2013). Kitchatinov and Olemskoy (2016) showed that at the beginning of the cycle, if the generation of the poloidal field is reversed (say due to the emergence of some wrongly tilted BMRs), then it will amplify the existing polar field, instead of reversing it. This increase in the magnetic field can lead to a grand maximum. Another mechanism of grand maxima was given by Ölçek et al. (2019), who showed that when the deep-seated α effect is coupled with the surface Babcock–Leighton source, then these two sources more or less contribute equally to generate a strong poloidal field through a sort of constructive interference.

Finally, for the secular and supersecular modulations (modulations beyond 11-year periodicity, e.g., Gleissberg cycle, Suess/de Vries cycle, Eddy cycle, and 2400-year Hallstatt cycle; Beer et al., 2018), there are limited studies available in the literature. In a

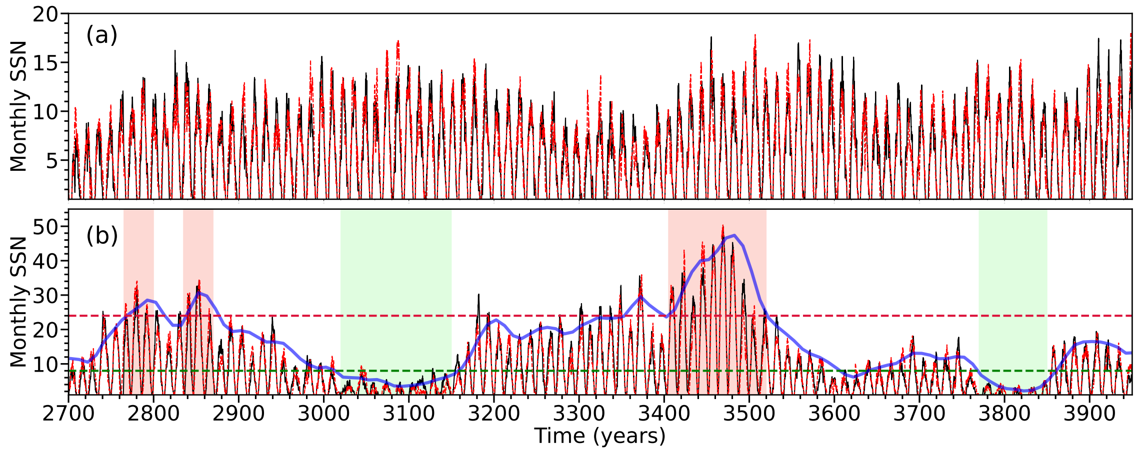

Fig. 3.4 Time series of the monthly BMR number from a 3D dynamo model of Karak and Miesch (2017) (a) without tilt scatter around Joy's law and (b) with scatter of $\sigma_\delta = 18^\circ$ (close to the observed value). The black/red curves indicate the north/south hemispheres. The blue curve in panel (b) is the smoothed curve of the cycle trajectories, and the green and red dashed horizontal lines indicate the thresholds for the grand minima and grand maxima, respectively. The green and red shaded regions indicate the grand minima and grand maxima episodes, respectively. Image Credit: Biswas et al. (2023b)

simplified $\alpha\Omega$ dynamo model coupled with the angular momentum equation, Pipin (1999) found the Gleissberg cycle as a result of the re-establishment of differential rotation after the magnetic feedback on the angular momentum transport. Cameron and Schüssler (2017) modelled the overall power spectrum of solar activity using a generic normal form model for a noisy and weakly nonlinear limit cycle, and Cameron and Schüssler (2019) showed that the long-term modulations beyond the 11-year cycle are consistent with the realization noise, thus casting doubt whether secular and supersecular modulations are connected to the intrinsic periodicities of the solar dynamo.

3.4 Conclusion

In conclusion, here a brief review of the theoretical perspectives to explain the observed features in the framework of the dynamo models is presented. It is discussed that the nonlinearities in the dynamo, including the effects of the flux loss due to magnetic buoyancy

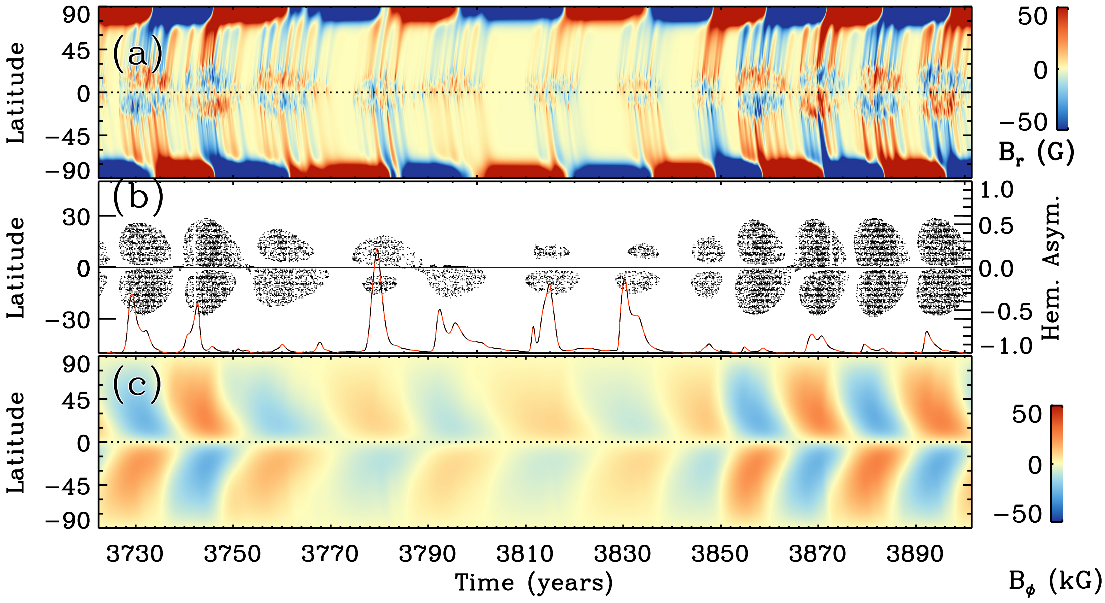

Fig. 3.5 Zoomed-in view of a grand minimum presented in Figure 3.4. Evolution of (a) the surface radial field (b) BMR eruptions and hemispheric asymmetry of the toroidal field (black/red curve), and (c) the toroidal field at the bottom of the convection zone. Image Credit: Biswas et al. (2023b)

as well as latitude and tilt quenching, help to stabilize the solar dynamo, rather than producing variability in the solar cycle. Primary causes of the solar cycle variability are the stochastic fluctuations in the dynamo which are inherent in different processes such as a large scatter of the BMR's tilts around Joy's law, and variability in the BMR eruption rates and locations. On one hand, while modern dynamo models are able to reproduce, with a reasonable ad-hoc tuning of the parameters, the observed features of solar variability, the exact role of those factors is not clear, and some discrepancies between the model results and the data still remain. On the other hand, the progress in the accuracy of models is significant, and we keep gaining knowledge of the processes driving solar variability with the new data acquainted and new models developed.